



Faculty of Sciences

Department of Computer Science & Information Systems

**ONTOLOGY-DRIVEN SEMANTIC ENRICHMENT FRAMEWORK
FOR MAXIMAL CREATION OF OPEN DATA VALUE
(THE CASE OF BOTSWANA NATIONAL PERFORMANCE
INDICATORS)**

By

OARABILE SEBUBI

Student ID number: 14100595

BIS (UB), MSc Information Systems (BIUST)

mmaseb@hotmail.com, (+267)-71754629

<https://orcid.org/0000-0001-8816-2722>

A Dissertation Submitted to the Faculty of Science in Partial Fulfillment of the
Requirements for the Award of the Degree of Doctor of Philosophy in Information
Systems of BIUST

Supervisor(s): Prof. Irina Zlotnikova; Dr. Hlomani

Department of Computer Science & Information Systems

Faculty of Science, BIUST

zlotnikovai@biust.ac.bw, (+267)-76618598; hlomanihb@biust.ac.bw, (+267)-75025393

November, 2022

DECLARATION REGARDING THE WORK AND COPYRIGHT

Candidate (please write in caps or type)

OARABILE SEBUBI

Student ID: 14100595

Thesis Titled:

Ontology-Driven Semantic Enrichment Framework for Maximal Creation of Open Data Value

(The Case of Botswana National Performance Indicators)

I, the **Candidate**, certify that the Thesis is all my own original work and that I have not obtained a degree in this University or elsewhere on the basis of any of this work.

(If the thesis is based on a group project, then the student must indicate the extent of her / his contribution, with reference to any other theses submitted or published by each collaborator in the project, and a declaration to this effect must be included in the thesis)

This dissertation/thesis is copyright material protected under the Berne Convention, the Copyright and Neighbouring Rights Act, Act. No. 8 of 2000 and other international and national enactments, in that behalf, on intellectual property. It must not be reproduced by any means, in full or in part, except for short extracts in fair dealing; for researcher private study, critical scholarly review or discourse with an acknowledgement, without the written permission of the office of the Postgraduate School, on behalf of both the author and the BIUST.

Signed:



Date: 30-Nov-22

Primary Supervisor (please write in caps or type)

Prof. Irina Zlotnikova

I, the Candidate's **Primary Supervisor**, hereby confirm that I have inspected the above titled thesis and, to the best of my knowledge, it is based on the original work of the candidate.

Signed:



Date: 30-Nov-22

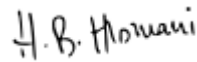
CERTIFICATION

The undersigned certifies that she has read and hereby recommends for acceptance by the Faculty of Science a dissertation titled: *Ontology-Driven Semantic Enrichment Framework for Maximal Creation of Open Data Value (The Case of Botswana National Performance Indicators)*, in fulfilment of the requirements for the degree of Doctor of Philosophy in Information Systems of the BIUST.



Prof. Irina Zlotnikova
(Supervisor)

Date: 30-Nov-22



Dr. H. Hlomani
(Co-Supervisor)

Date: 30-Nov-22



ACKNOWLEDGEMENTS

Firstly and foremost, I acknowledge the valuable role played by BIUST in sponsoring this research work. The supervisory roles of Prof. Irina Zlotnikova and Dr. H. Hlomani and their valuable contributions are highly appreciated. Also, the constructive comments from the Department of Computer Science and Information Systems team during the conceptual phases of this research (research concept and proposal stages) are also appreciated.

Secondly, I give gratitude to Statistics Botswana for their role in this research. Special thanks to SB's Statistician General, Dr. Burton Mguni for granting us permission to conduct research in SB and opening up on the future OD presentation and representation needs that served as the foundation for this research. He also provided us with useful documentations on Africa Agenda, the domesticated SDG Agenda, SDGs National Baseline Indicator Framework and numerous SDGs reporting documentations. He also introduced us to Statistics and SDGs online resources. These resources are at the core of this research. Moreover, he invited us to the stakeholder meeting where this research benefited from the collaborative dialogues of resourceful participants conceptualising on the SDGs National Baseline Indicator Framework. I also thank Ms. Grace Mphetolang for her assistance in our SB visits.

Thirdly, I am also grateful for the skills that the University of Cape Town through their Ontology training program (4th Interdisciplinary School on Applied Ontology – ISAO 2018) and the University of Pretoria with their Research Data Science training program (2020 CODATA-RDA School of Research Data Science) contributed to this research.

Lastly, and most importantly, I acknowledge that this work would not have been possible without the Lordship of my God.

THANK YOU, CHRIST JESUS, MY LORD!!!



TABLE OF CONTENTS

DECLARATION REGARDING THE WORK AND COPYRIGHT.....	i
CERTIFICATION	ii
ACKNOWLEDGEMENTS.....	iii
TABLE OF CONTENTS.....	iv
LIST OF FIGURES	ix
LIST OF TABLES	xii
LIST OF APPENDICES	xiii
LIST OF LINKS TO ONLINE RESOURCES.....	xiv
LIST OF ABBREVIATIONS	xv
LIST OF ODSE FRAMEWORK CONCEPTS DEFINITIONS	xvi
ABSTRACT	1
CHAPTER ONE: RESEARCH INTRODUCTION	2
1.1 Introduction to Research	2
1.2 Research Background.....	1
1.3 Research Problem Statement.....	6
1.4 Research Questions	7
1.5 Research Objectives	7
1.6 Research Contribution	1
1.7 Dissertation Structure.....	3
CHAPTER TWO: RESEARCH LITERATURE	5
2.1 Introduction	5
2.2 Open Data	5
2.2.1 Open Data as the Future Driver	5
2.2.2 Open Data Value (Knowledge)	6
2.2.3 Open Data Value (Knowledge) Reuse Potential.....	6
2.2.4 Open Data Value (Knowledge) Challenges.....	8
2.3 Ontology Potential for Open Data Value (Knowledge) Creation	1
2.3.1 Ontology	1
2.3.2 Web Ontologies	1
2.3.3 Performance Indicator Ontologies	3
2.4 Linked Open Data (LOD) Potential for Open Data Value (Knowledge) Creation.....	15
2.5 Semantic Enrichment Potential for Open Data Value (Knowledge) Creation.....	17



2.5.1	Semantic Enrichment Vocabularies	17
2.5.2	Semantic Enrichment Methods.....	18
2.5.3	Semantic Enrichment Models.....	19
2.5.4	Semantic Enrichment Approaches	19
2.5.5	Semantic Enrichment and the Formation of OD Value (Knowledge) Attributes ...	21
2.5.6	Semantic Enrichment Frameworks	26
2.6	Conclusion.....	36
CHAPTER THREE: RESEARCH METHODOLOGY.....		40
3.1	Introduction	40
3.2	Description of the Design Science Research Methodology for Information Systems Research	40
3.2.1	Activity 1: Problem Identification and Motivation	41
3.2.2	Activity 2: Definition of Objectives for a Solution	41
3.2.3	Activity 3: Design and Development.....	41
3.2.4	Activity 4: Demonstration.....	42
3.2.5	Activity 5: Evaluation.....	42
3.2.6	Activity 6: Communication	43
3.3	The Adapted Design Science Research Methodology.....	43
3.3.1	Phase 1: Framework Needs Identification.....	45
3.3.2	Phase 2: Framework Design and Development.....	46
3.3.3	Phase 3: Framework Demonstration	47
3.3.4	Phase 4: Framework Evaluation	53
3.3.5	Phase 5: Framework Finalisation	59
3.4	Research Limitations and Challenges.....	64
3.5	Conclusion.....	65
CHAPTER FOUR: FRAMEWORK NEEDS IDENTIFICATION		68
4.1	Introduction	68
4.2	Identification of Semantic Consumption Needs	68
4.3	Identification of Semantic Publication Needs.....	72
4.3.1	The Open Data Program Needs.....	72
4.3.2	The Open Data Presentation Needs	73
4.3.3	The Open Data Representation (Ontology) Needs.....	82
4.4	Conclusion.....	85
CHAPTER FIVE: FRAMEWORK DESIGN AND DEVELOPMENT.....		87



5.1	Introduction	87
5.2	The Design and Development of the Framework	87
5.2.1	Step 1: Review of Research Literature.....	87
5.2.2	Step 2: Identification of Semantic Enrichment Framework Entities, Elements and Relations.....	88
5.2.3	Step 3: Generation of the Conceptual Semantic Enrichment Framework	90
5.3	The Preliminary Semantic Enrichment Framework.....	90
5.3.1	Stage 1: Open Data Pre-processing	91
5.3.2	Stage 2: Open Data Semantic Enrichment Processing.....	92
5.3.3	Stage 3: Open Data Post-processing	93
5.4	Conclusion.....	94
CHAPTER SIX: FRAMEWORK DEMONSTRATION 1 (SEMANTIC ENRICHMENT APPROACH)		95
6.1	Introduction	95
6.2	Demonstration 1: The Selection of Generic Ontologies.....	95
6.2.1	Phase 1: Analysis of Source Data and Metadata Resources for Ontology Search Literals	95
6.2.2	Phase 2: Identification of Ontology Enrichment Requirements.....	97
6.2.3	Phase 3: Search for Potential Ontology Targets.....	97
6.2.4	Phase 4: Selection of Suitable Ontology Targets.....	99
6.3	Demonstration 2: The Development of the Custom Domain Ontology (NPIOnto)	100
6.3.1	Specification Phase	100
6.3.2	Knowledge Acquisition Phase	101
6.3.3	Conceptualisation Phase.....	103
6.3.4	Formalisation Phase.....	109
6.3.5	Evaluation Phase.....	123
6.4	Conclusion.....	127
CHAPTER SEVEN: FRAMEWORK DEMONSTRATION 2 (SEMANTIC ENRICHMENT PROCESSING)		129
7.1	Introduction	129
7.2	Stage 1: Open Data Pre-processing.....	129
7.2.1	Pre-processing Input Component.....	129
7.2.2	Pre-processing Data Re-organisation, Cleaning and Transformation Component	129
7.2.3	Pre-processing Output Component.....	131
7.3	Stage 2: Open Data Semantic Enrichment Processing	132



7.3.1	Data Enrichment Component	132
7.3.2	Semantic Enrichment Output Component	136
7.4	Stage 3 Open Data Post-processing	140
7.4.1	Resource Integration Component.....	140
7.4.2	Resource Publication and Consumption Component.....	141
7.5	Conclusion.....	156
CHAPTER EIGHT: FRAMEWORK EVALUATION AND FINALISATION		158
8.1	Introduction	158
8.2	Framework Verification	158
8.3	Framework Validation	160
8.3.1	The ODSE Pre-processing (Stage 1) Validation [Flat Text Resources].....	163
8.3.2	The ODSE Processing (Stage 2) Validation [Disintegrated Knowledge Resources]	165
8.3.3	The ODSE Post-processing (Stage 3) Validation [Integrated Knowledge Resources]	166
8.4	Framework Finalisation.....	170
8.4.1	Framework Redesign Elements.....	170
8.4.2	Framework Adjustments.....	177
8.5	Conclusion.....	180
CHAPTER NINE: RESEARCH CONTRIBUTION.....		182
9.1	Introduction	182
9.2	Stage 1: ODSE Pre-processing.....	183
9.2.1	Pre-processing Input (ODSE Component 1.0)	184
9.2.2	Pre-processing Implementation (ODSE Component 2.0)	185
9.2.3	Pre-processing Output (ODSE Component 3.0).....	187
9.3	Stage 2: ODSE Processing.....	190
9.3.1	ODSE Approach (ODSE Component 4.0).....	190
9.3.2	Data Enrichment (ODSE Component 5.0)	193
9.3.3	Enrichment Output (ODSE Component 6.0).....	194
9.4	Stage 3: ODSE Post-processing.....	195
9.4.1	Knowledge Resources Structuring (ODSE Component 7.0)	195
9.4.2	Knowledge Resources Deployment (ODSE Component 8.0).....	196
9.5	The ODSE Framework Implementation Support Tool Feature.....	197
9.5.1	The ODSE Scoping Models	197



9.5.2	The ODSE Assessment Models	204
9.6	Conclusion.....	215
CHAPTER TEN: RESEARCH CONCLUSIONS, RECOMMENDATIONS, FUTURE WORKS AND PUBLICATIONS.....		217
10.1	Conclusion	217
10.2	Recommendations	219
10.3	Future Mapping.....	221
10.4	Publications and Collaborations	222
REFERENCES.....		224
APPENDICES.....		235
Appendix 1: Africa Agenda Datasets Screenshots.....		235
Appendix 2: SDG Agenda Datasets Screenshots.....		237
Appendix 3: Vision Agenda Datasets Screenshots		239
Appendix 4: SDG Agenda Performance Datasets Screenshot.....		240
Appendix 5: Source Data Inconsistency Prevalence Summary		240



LIST OF FIGURES

CHAPTER TWO

<i>Figure 2. 1: Ontology Expressiveness Spectrum in [62] adapted from McGuinness (2003)</i>	2
<i>Figure 2. 2: SDGIO Class Hierarchy</i>	5
<i>Figure 2. 3: Class SDGIO_00000000 and Instances</i>	6
<i>Figure 2. 4: RDF Representation of SDGIO_00000035</i>	7
<i>Figure 2. 5: RDF Representation of SDGIO_00000057</i>	7
<i>Figure 2. 6: Class SDGIO_00000002</i>	8
<i>Figure 2. 7: SDGIs superclass (SDGIO_00000001) and SDGIs superclass (SDGIO_00000003) ...</i>	9
<i>Figure 2. 8: RDF Representation of SDGIO_00020000</i>	9
<i>Figure 2. 9: SDGIO Empty Data Properties Window</i>	10
<i>Figure 2. 10: SDGIO Object Properties Window</i>	11
<i>Figure 2. 11: Tim Berners-Lee's 5-Star scheme for LOD Publications [75]</i>	15
<i>Figure 2.12: CSV Sample</i>	16
<i>Figure 2.13: RDF Sample</i>	16
<i>Figure 2. 14: Semantic Richness of Knowledge Organisation Systems [50]</i>	17

CHAPTER THREE

<i>Figure 3. 1: Summary of ODSE Framework Validation Method</i>	54
<i>Figure 3. 2: Semantic Usability Assessment Model</i>	57

CHAPTER FOUR

<i>Figure 4. 1: Open Data Evolution Model [120]</i>	69
<i>Figure 4. 2: SDGs Data Interface Screen [5]</i>	74
<i>Figure 4. 3: Data View Options (SDG 16) [5]</i>	74
<i>Figure 4. 4: Basic Dashboard View Functionalities – Open Data Access Option (SDGI 16.9.1 Extract) [5]</i>	75
<i>Figure 4. 5: OD Export in Flat Text Format – Open Data Access Option (SDGI 16.9.1 Extract) [5]..</i>	75
<i>Figure 4. 6: Advanced Dashboard View Functionalities – Column Chart (SDGI 16.9.1) [5]</i>	76
<i>Figure 4. 7: Advanced Dashboard View Functionalities – API Snippet (SDGI 16.9.1) [5]</i>	78
<i>Figure 4. 8: OD Access Option – PDF View (SDGI 16.9.1) [5]</i>	78
<i>Figure 4. 9: Vision Agenda – Pillar 4 Sample (SDGI 16.9.1) [10]</i>	79
<i>Figure 4. 10: Open Data Access Option – XML View (SDGI 16.9.1) [5]</i>	80
<i>Figure 4. 11: SDG Agenda – Vision Agenda Sample Mapping (SDGI 16.9.1) [10]</i>	84



<i>Figure 4. 12: Africa Agenda – SDG Agenda Sample Mapping (SDGI 16.9.1) [8]</i>	84
--	----

CHAPTER FIVE

<i>Figure 5. 1: The Preliminary Ontology-Driven Semantic Enrichment Framework [119]</i>	91
---	----

CHAPTER SIX

<i>Figure 6. 1: Africa Agenda Indicator Model</i>	104
<i>Figure 6. 2: Sustainable Development Goals Indicator Model</i>	105
<i>Figure 6. 3: Vision Indicator Model</i>	106
<i>Figure 6. 4: National Development Plan Indicator Model</i>	107
<i>Figure 6. 5: NPIOnto Conceptual Model [129]</i>	109
<i>Figure 6. 6: Formalised NPIOnto Knowledge Model Sample Extract</i>	110
<i>Figure 6. 7: NPI Annotation Properties Extract (SDG 1)</i>	111
<i>Figure 6. 8: NPI Disaggregation Extract (SDGI 16.9.1)</i>	113
<i>Figure 6. 9: Fragment of Reasoner Inference for SDGI</i>	118
<i>Figure 6. 10: NPIOnto Inverse and Transitive Object Property Extract</i>	120
<i>Figure 6. 11: Fragment of NPIOnto knowledge base Instance for SDGI 15.5.1</i>	122

CHAPTER SEVEN

<i>Figure 7. 1: Resource Reconciliation in OpenRefine - Generic ODSE Approach</i>	133
<i>Figure 7. 2: Resource Reconciliation in OpenRefine – Domain ODSE Approach</i>	134
<i>Figure 7. 3: Resource RDFization in OpenRefine - Generic ODSE Approach</i>	135
<i>Figure 7. 4: RDFization in OpenRefine - Domain ODSE Approach</i>	136
<i>Figure 7. 5: Generic Knowledge Resource (SDGI 16.9.1 Extract)</i>	137
<i>Figure 7. 6: Domain Knowledge Resource (SDGI 16.9.1 Extract)</i>	138
<i>Figure 7. 7: Knowledge Resources per Agenda</i>	141
<i>Figure 7. 8: SPARQL Endpoint Publication (SDGI 16.9.1 Extract)</i>	142
<i>Figure 7. 9: Generic Knowledge Resource Mining (SDGI 16.9.1)</i>	144
<i>Figure 7. 10: Sample Generic Knowledge Resource Mining Output 1 (SDGI 16.9.1)</i>	145
<i>Figure 7. 11: Sample Generic Knowledge Resource Mining Output 2 (SDGI 16.9.1)</i>	146
<i>Figure 7. 12: Domain Knowledge Resource Mining 1 (SDGI 16.9.1)</i>	149
<i>Figure 7. 13: Domain Knowledge Resource Mining 2 (SDGI 16.9.1)</i>	150
<i>Figure 7. 14: Sample Domain Knowledge Resource Mining Output 1 (SDGI 16.9.1)</i>	151
<i>Figure 7. 15: Sample Domain Knowledge Resource Mining Output 2 (SDGI 16.9.1)</i>	151
<i>Figure 7. 16: Sample Domain Knowledge Resource Mining Output 3 (SDGI 16.9.1)</i>	152

CHAPTER EIGHT

<i>Figure 8. 1: ODSE Framework Validation Scenario (SDGI 16.9.1)</i>	162
--	-----



CHAPTER NINE

<i>Figure 9. 1: Final Ontology-Driven Semantic Enrichment Framework</i>	<i>183</i>
<i>Figure 9. 2: NPIOnto Scope Model</i>	<i>200</i>
<i>Figure 9. 3: Semantic Capabilities Relational Model</i>	<i>202</i>
<i>Figure 9. 4: Semantic Enrichment Assessment Model</i>	<i>206</i>
<i>Figure 9. 5: Semantic Richness Assessment Model</i>	<i>208</i>
<i>Figure 9. 6: Semantic Mining Assessment Model</i>	<i>211</i>



LIST OF TABLES

CHAPTER THREE

<i>Table 3. 1: Semantic Capabilities Assessment Model.....</i>	<i>55</i>
<i>Table 3. 2: Adapted Design Science Research Methodology Application Summary.....</i>	<i>66</i>

CHAPTER FIVE

<i>Table 5. 1: Semantic Enrichment Elements.....</i>	<i>88</i>
--	-----------

CHAPTER SIX

<i>Table 6. 1: Source Datasets.....</i>	<i>96</i>
<i>Table 6. 2: SDGIO Targets.....</i>	<i>97</i>
<i>Table 6. 3: Potential Generic Ontology Targets.....</i>	<i>98</i>
<i>Table 6. 4: Selected Generic Ontology Targets.....</i>	<i>99</i>
<i>Table 6. 5: Summary of NPIOnto Knowledge base Matrix Extract.....</i>	<i>123</i>
<i>Table 6. 6: Summary of NPIOnto Knowledge Model and Knowledge base Evaluations</i>	<i>126</i>
<i>Table 6. 7: Summary of Matrix for NPIOnto Object Properties.....</i>	<i>127</i>

CHAPTER SEVEN

<i>Table 7. 1: Datasets Inconsistencies.....</i>	<i>130</i>
<i>Table 7. 2: Summary of Semantic Enrichment Approach Concepts Specifications</i>	<i>139</i>

CHAPTER EIGHT

<i>Table 8. 1: Preliminary ODSE Framework Process Evaluation</i>	<i>159</i>
<i>Table 8. 2: Summary of Design Features for Semantic Valuation Models</i>	<i>174</i>
<i>Table 8. 3: Summary of Potential Application of the ODSE Framework Implementation Support Tools</i>	<i>178</i>

CHAPTER NINE

<i>Table 9. 1: Semantic Publication and Consumption Needs - Drivers of the ODSE Framework Design</i>	<i>189</i>
--	------------



LIST OF APPENDICES

APPENDIX 1: AFRICA AGENDA DATASETS SCREENSHOTS

<i>Table AFR 1: Africa Agenda Indicator Dataset Screenshot.....</i>	<i>235</i>
<i>Table AFR 2: Africa Agenda Target Dataset Screenshot.....</i>	<i>235</i>
<i>Table AFR 3: Africa Agenda Priority Dataset Screenshot.....</i>	<i>236</i>
<i>Table AFR 4: Africa Agenda Goal Dataset Screenshot</i>	<i>236</i>
<i>Table AFR 5: Africa Agenda Aspiration Dataset Screenshot</i>	<i>237</i>
<i>Table AFR 6: Africa Agenda Implementation Plan Dataset Screenshot.....</i>	<i>237</i>
<i>Table AFR 7: Africa Agenda Dataset Screenshot</i>	<i>237</i>

APPENDIX 2: SDG AGENDA DATASETS SCREENSHOTS

<i>Table SDG 1: SDG Indicator Dataset Screenshot.....</i>	<i>237</i>
<i>Table SDG 2: SDG Target Dataset Screenshot.....</i>	<i>238</i>
<i>Table SDG 3: SDG Goal Dataset Screenshot</i>	<i>238</i>
<i>Table SDG 4: SDG Agenda Dataset Screenshot</i>	<i>238</i>

APPENDIX 3: VISION AGENDA DATASETS SCREENSHOTS

<i>Table VIS 1: Vision Indicator Dataset Screenshot</i>	<i>239</i>
<i>Table VIS 2: Vision Outcome Dataset Screenshot.....</i>	<i>239</i>
<i>Table VIS 3: Vision Pillar Dataset Screenshot.....</i>	<i>239</i>
<i>Table VIS 4: Vision Agenda Dataset Screenshot.....</i>	<i>240</i>

APPENDIX 4: SDG AGENDA PERFORMANCE DATASETS SCREENSHOTS

<i>Table SDGI 1: Consolidated SDG Agenda Performance Datasets Screenshot.....</i>	<i>240</i>
---	------------

APPENDIX 5: SUMMARY OF SOURCE DATA INCONSISTENCY PREVALENCE

<i>Figure Appx 5. 1: Source Data Inconsistencies by Type.....</i>	<i>240</i>
---	------------



LIST OF LINKS TO ONLINE RESOURCES

The table below lists the links to resources (mostly, samples) that form input to the research as well as the output resources at the end of each of the three processing stages for the Ontology-Driven Semantic Enrichment (ODSE) Framework. These resources, linked to specific sections in the dissertation, have been uploaded to Figshare and are accessible from specific links for each resource category provided below.

Framework Process	Resource Type	Resource Web Link	Resource Web Link URL
ODSE Pre-processing (Stage 1) Resources	Input Resources (Raw OD Resources)	ODSE Framework Package 1 - Sample Raw OD for SDGI 16.9.1 ODSE Framework Package 2 - Raw OD Resources for SDG Agenda ODSE Framework Package 3 - Raw OD Resources for Vision Agenda ODSE Framework Package 4 - Raw OD Resources for Africa Agenda	https://figshare.com/s/8382f76f471144242070 https://figshare.com/s/177199fb05051428a987 https://figshare.com/s/ef5328b22072375dc733 https://figshare.com/s/5570670048ebb56da4eb
	Re-organised and Transformed Resources	ODSE Framework Package 5 - Pre-processed OD Resources for all Agendas	https://figshare.com/s/4ec6350b3bc89b2d118d
ODSE Processing (Stage 2) Resources	Individual Knowledge Resources	National Performance Indicator Ontology (NPIOnto)	https://figshare.com/articles/dataset/National_Performance_Indicator_Ontology_NPIOnto/20318562
		ODSE Framework Package 6 - Sample Knowledge Resources for SDG16	https://figshare.com/s/0c16ae71055e1794dfe8
ODSE Post-processing (Stage 3) Resources	Integrated Knowledge Resources	ODSE Framework Package 7 - Integrated Knowledge Resources	https://figshare.com/s/9b2ced2171da21eb7262
	Integrated Knowledge Resources Evaluation	ODSE Framework Package 8 - Integrated Knowledge Resources Evaluation Cases	https://figshare.com/s/0ba7ee666d1ee181a2c3



LIST OF ABBREVIATIONS

AFRI(s)	Africa Agenda Indicator(s)
API	Application Programming Interface
LOD	Linked Open Data
NDP	National Development Plan
NDPI(s)	National Development Plan Indicator(s)
NPI(s)	National Performance Indicator(s)
NPIOnto	National Performance Indicator Ontology
OD	Open Data
ODSE	Ontology-Driven Semantic Enrichment
OWL	Web Ontology Language
RDF	Resource Description Framework
SDG(s)	Sustainable Development Goal(s)
SDGI(s)	Sustainable Development Goals Indicator(s)
SDGIO	Sustainable Development Goals Indicator Ontology
SPARQL	Simple Protocol and RDF Query Language
UNSDG(s)	United Nations Sustainable Development Goal(s)
VISI(s)	Vision Indicator(s)



LIST OF ODSE FRAMEWORK CONCEPTS DEFINITIONS

Domain ODSE approach	ODSE approach which makes use of domain ontologies
Domain ontology	A term that identifies a custom domain ontology
Enriched Linked Open Data	Interlinked ontology enriched data
Generic ODSE approach	ODSE approach which makes use of a combination of generic ontologies
Generic ontology	A term that embraces both the top-level/upper/foundational/reference ontologies, as well as any unrelated domain ontology
Intelligent Linked Open Data	Interconnected pieces of data that holds within itself all the information that describes the syntax and semantic features needed for its understanding, interpretation and intelligible use
NPI triple-store	The designated triple-store for the storage of interlinked NPI in the form of subject-predicate-object
ODSE approach	The ODSE component that pertains to the choice and implementation of applicable semantic value creation approach for semantic value creation needs (semantic publication and consumption needs) at hand
ODSE assessment model	OD valuation models in the form of qualitative tools for the assessment of knowledge resources for OD value (knowledge) creation and use capabilities
ODSE component	A unique grouping of ODSE main activities according to functionality
ODSE framework	A combination of ODSE stages, approaches, processes, components, main activities, sub activities, scoping models and assessment models, into one solution that maps the pathway toward addressing the quest for semantic enrichment frameworks with potential for semantic value (knowledge) creation maximization
ODSE framework implementation support tool	ODSE scoping and assessment models for Botswana development agenda domain scoping and for qualitative evaluation of semantic value for knowledge resources
ODSE framework pillar	The overarching functional focus for the ODSE processing
ODSE main activity	ODSE Activity which represent tasks that can be broken down into other smaller tasks
ODSE Process	A unique grouping of different components with different functionalities that holistically implements a unique task
ODSE scoping model	Reference models that represent the Botswana development agenda domain scope
ODSE stage	A unique overarching process that defines the ODSE process linkages through specification of process implementation order, i.e., ODSE stage 2 identifies the second implementation process
ODSE sub-activity	ODSE Activity which represent the lowest granularity of the ODSE processing
Ontology constructs	Knowledge representation entities pertaining to classes (representing domain concepts), properties (representing concepts attributes and relations) and axioms (representing statements of truth or facts)
Ontology-driven enrichment	The creation of the semantic layer through the mapping of NPIs to ontology constructs, i.e., classes, properties and axioms in NPIOnto
Open Data pool	Nation-wide raw OD collections
Public knowledge resource pool	Nation-wide knowledge resource collections



RDF-driven integration	The semantic transformation process that converts disparate flat text resources into a more flexible, data integration structure, the RDF triple structure form of subject-predicate-object
Semantic capabilities value-added	The extra semantic value created through semantic enrichment in addition to raw OD (flat text) value
Semantic enrichment	The creation of semantic value (semantic value creation)



ABSTRACT

This research adopts a pragmatic paradigm in addressing the gap of reviewed semantic enrichment frameworks that lack mechanisms to assess the degree of semantic value-added¹ to flat text resources in terms of knowledge (semantic context) and semantic capabilities (semantic structure). Currently, due to this knowledge gap, it is difficult to drive the semantic value creation process toward specific enrichment output and to evaluate the output as well. This is what this research refers to as, the *semantic valuation challenge*.

To address the semantic valuation challenge, the research proposes the semantic value creation framework with mechanism for semantic valuation, the *Ontology-Driven Semantic Enrichment Framework* (ODSE Framework). The framework development adapted the Design Science Research Methodology for Information Systems. The framework leverages Linked Data principles in the conversion of flat text resources to knowledge resources². It further augments this basic semantic value creation process with qualitative semantic valuation models to be applied to the semantic enrichment added layers (syntax and semantic layers) for semantic value assessment, specifically, assessment of the degree of knowledge creation potential and usability effectiveness of a given knowledge resource on the basis of semantic attributes. The semantic valuation models are also to aid semantic enrichment decision making for Open Data³ (OD) publishers and consumers with the need for richer formats on applicable semantic enrichment approach to adopt for specific knowledge resource output target. The demonstration of the framework was conducted to determine the semantic enrichment approach with potential to maximal semantic value-adding potential for more knowledge creation and usability effectiveness.

Though the framework demonstration was specific to the Botswana development agenda domain, the final ODSE framework has been redesigned for reusability in any semantic value creation implementation for all types of flat text data transformations to semantic formats. Also, the ODSE Framework implementation support tools are complementary to any semantic enrichment framework implementation in any application domain.

¹ Semantic value-added is the extra semantic value created through semantic enrichment in addition to raw OD value.

² Knowledge resources are semantically enriched web resources/datasets.

³ Open Data is data availed for free public use, re-use and redistribution.

CHAPTER ONE: RESEARCH INTRODUCTION

1.1 Introduction to Research

Globally, the Open Government Data movement is demanding for Web publication⁴ of government data for free public access, with less reuse constraints [1]. In response, most governments are pushing for the publication of OD but despite these efforts, OD is still minimally used due to publications that are often not in reusable formats [2]. This results from flat text publications that are difficult to access, understand, integrate and interoperate with [2], [3]. To start with, OD is published at diverse places and in different platforms such as national websites and regional databases. Also, it is availed in diverse data formats such as unstructured, semi-structured, and structured forms [2]. Above that, it is availed in diverse publication formats, such as web files, portals and as Application Programming Interface⁵ (API) links [4]. Furthermore, most OD publications are limited only to current data views, thus, making it impossible to track data changes over time [2]. Moreover, the technical expertise needed to exploit OD are not often available to consumers [2]. Consequently, the issues mentioned above compromise semantic capabilities and knowledge creation potential of OD resources, thus posing a need for a transition to rich semantic formats.

To contextualise the problem scope, an investigation on semantic publication and consumption needs specific to Botswana was conducted with Statistics Botswana. Statistics Botswana collects and distributes national data via Statistics Botswana data portal [5]. The specific data of interest in this research pertains to the development agenda domain that constitute the Botswana Integrated Indicator Framework, comprising [6], [7], [8], [9], [10]:

- 1) Vision 2036 Agenda or 2036 Agenda, the national agenda, referred to in this dissertation as Vision Agenda,
- 2) National Development Plan (NDP) 11 or 2023 Agenda, the national agenda, referred to in this dissertation as NDP Agenda,

⁴ Publication in the context of this research refers to data uploaded to a web portal, mostly, in flat text formats.

⁵ Application Programming Interface is a software that enables communication between two applications.



- 3) Africa Agenda 2063 or 2063 Agenda, the continental agenda, referred to in this dissertation as Africa Agenda, and
- 4) 2030 Sustainable Development Goals (SDGs) Agenda or 2030 Agenda, the global agenda, referred to in this dissertation as SDG Agenda.

Firstly, semantic publication needs specific to Botswana development agenda domain pertain to the OD program, the OD presentation and the OD representation (ontology) needs. In the context of this research, the OD program needs pertain to the national OD program implementation goals. The OD presentation needs pertain to future publication needs drawn from the limitations of the current Web presentation structure and format. The OD representation needs pertain to semantic modelling needs for OD drawn from source data (flat text OD publications) and corresponding metadata (data about data).

The future publication need for Statistics Botswana [5] is to avail performance data on the four development agendas specified above for free public access, use and reuse. Performance data for the development agendas is expressed in performance indicators – statistical tools that measure, structure and express performance [11], [12]. This research refers to performance indicators in the domain of development agendas as National Performance Indicators (NPIs).

Of these four agendas, only SDG Agenda has data published as OD and Botswana's aspiration is to implement the consolidated indicator framework for the sharing of SDG Agenda data with the other three development agendas [13]. However, OD for SDG Agenda is published in diverse file formats such as Hypertext Mark-up Language⁶ (HTML), image, tabular/graph, spreadsheet, Extensible Markup Language (XML) and as API links. Furthermore, performance indicators are expressed in aggregate form and the relating metadata is availed in Adobe Acrobat Reader (PDF) format. Moreover, the data is presented separately for each performance indicator [5]. The diverse presentation and content structures limit OD accessibility, understandability, interoperability, integrability and intelligibility capabilities. That is, the 'as-is' publication of OD is in a form that does not yield valuable use, which in turn renders the OD value (knowledge) creation and use potential less effective. The current semantic limitations also compromise the establishment of the consolidated indicator framework. Thus, semantic models for

⁶ Hypertext Mark-up Language stands for, a set of markup symbols or codes inserted into a file for Internet display.



enhancement of knowledge creation and usability capabilities are needed to facilitate effective OD value creation, usage and sharing across the development agenda domain. Secondly, the semantic publication needs described above are to address Botswana's semantic consumption need expressed by *sustainable economic development* pillar of the Vision Agenda (Pillar 1). This need pertains to OD support for the implementation of the future vision of Botswana's social and economic development driven by knowledge. Botswana's OD program goals articulates that OD can be leveraged to this effect [14]. Botswana's main need is to transition from the current dependence on natural resources to a knowledge-driven society founded on a knowledge-driven economy for the attainment of sustainable development and prosperity for all by 2036 [6]. OD is one of the resources that lay dormant, yet with unlimited value that appreciates with use, yielding increasing returns with increased use, unlike natural resources that depreciates in value and yields decreasing returns with extensive use [15]. So, by nature, raw flat text OD resources are an embodiment of unlimited economic potential which can be unleashed through semantic solutions that build semantic capabilities to revolutionise the creation, extraction, distribution and consumption of OD value.

Overall, semantic publication and consumption needs constitute the semantic value creation needs. On the one hand, semantic publication needs pertain to the OD program need, the need for rich semantic formats for OD publications; the OD presentation need, the need for unified publications; and the OD representation (ontology) need, the need for the implementation of a consolidated indicator framework for development agendas. On the other hand, the semantic consumption need pertain to OD support for Botswana's national agenda, the need for knowledge resources to drive the knowledge economy and society. Therefore, for the semantic publication needs, semantic solutions are needed to support the implementation of standardised and unified publication structure and format, as well as, the implementation of the consolidated indicator framework for the four development agendas. For the semantic consumption needs, semantic solutions must support the creation of knowledge resources for the implementation of the economic aspiration of a transformation into a knowledge society founded on a knowledge-driven economy and the creation of rich semantic formats for reusability of the current OD publications. All in all, these needs serve as the basis for the development



of semantic solutions to drive a shift from the current flat text publication and consumption to semantic publication and consumption of OD.

To enable semantic consumption, this research proposes a shift toward semantic publication centred on the enrichment of raw OD with semantic structure, capabilities and context for increased OD value (knowledge) creation and use potential. That kind of a shift necessitates the development of semantic value creation models to drive the transition from raw OD publication and consumption to semantic publication and consumption, leading to the overarching research question that points to the main research deliverable, the **main research question** that this research addresses, was formulated as:

What semantic enrichment framework could best support the implementation of semantic publication and consumption of OD with potential for OD value (knowledge) creation maximisation?

Actions on the main research question sparked an inquest for semantic value creation implementation objective to guide research activities. To answer the main research question, the general objective of this research was:

To develop the semantic enrichment framework that could best support the implementation of semantic publication and consumption of OD with potential OD value (knowledge) creation maximisation.

This objective, in turn, sparked an inquest for architectural and functional design elements for semantic value creation maximisation. In response, this research proposes a three-stage flat text transformation process for semantic value creation with in-built functional capability design for semantic value creation maximisation and assessment as well, the ODSE framework. Chapters One-Eight discuss the motivation, objectives, methodology, and outline a step-by-step account of the conceptualisation, design, development, demonstration, evaluation and finalisation of the ODSE framework. The final ODSE framework is overviewed in Section 1.6 and fully presented in Chapter Nine.



1.2 Research Background

The research background starts with a contextual clarification of core research concepts. Firstly, OD value is synonymous to *semantic value* and both denote the value of data, demonstrated or determined by the *knowledge* extracted from it. Knowledge refers to actionable insights or as defined by Stehr [16], the capacity to act. In this regard, *OD/semantic value* is synonymous to *knowledge*. Therefore, *OD/semantic value creation* is synonymous to *knowledge creation*, the enhancement of structure and meaning to improve semantic capabilities. The *OD/semantic value creation* process is synonymous to *semantic enrichment* process. The end product of *OD/semantic value creation* process or *semantic enrichment* process is identified as, *knowledge resource* – semantically enriched OD resource. The *OD/semantic value mining* (harvesting or extraction) is synonymous to *knowledge mining* (harvesting or extraction) for structure and meaning value-added to flat text resources. Lastly, *OD/semantic value use* is synonymous to *knowledge use*. So, these concepts are used interchangeably in this dissertation.

Prior to the development of the ODSE Framework, numerous semantic enrichment frameworks were reviewed for architectural and functional design elements, with potential for semantic value creation maximisation and were discovered to be similar in the following ways:

- 1) They comprise the semantic enrichment input, processing, and output components – the semantic value creation processing scope,
- 2) The semantic enrichment process has an ontology component and adopts the LOD standard, and
- 3) The semantic enrichment objective is either focused on semantic annotation to improve semantic meaning or on semantic representation to improve semantic structure or on both, with limited or without any consideration for the assessment of the resulting semantic value.

The third aspect represents the research gap that this study addresses. Specifically, reviewed frameworks were found to be limited in answering the main research question pertaining to the development of the semantic enrichment framework with maximal OD value creation potential. Determining the effectiveness of the framework in semantic value creation involves the assessment of semantic enrichment input, processing and output for the entire framework to assess the degree of achievement of the framework

intent, of which in this research pertain to the degree of semantic capabilities value-added to flat text OD resources (resulting semantic value). Therefore, though framework evaluation is of great essence in this research case, reviewed frameworks either partially includes or completely excludes the semantic value assessment. Below is a detailed discussion of reviewed semantic enrichment frameworks.

The first identified research gap concerns framework designs with either partial or complete exclusion of framework evaluation, thus, making it difficult to assess how the processing output fulfils the purpose of the framework. One of the reviewed frameworks in this category is the *Fuzzy-Ontology-Enrichment-based Framework for Semantic Search* proposed by Baazaoui-Zghal and Ghezala [17] which integrates the domain ontology enrichment and the fuzzy ontology building in the information retrieval process. The framework comprises three components, pertaining to OD input, processing and output being: 1) the fuzzy information retrieval component; 2) an incremental ontology enrichment component; and 3) an ontology repository component. The focus of this framework is on semantic representation, specifically, the semantic enrichment of data structure to support search operations. However, the authors do not specify the mechanism for assessment of the impact of components' combination on the final output, as well as, assessment of the degree of framework effectiveness.

The other framework without specification of framework evaluation is *the Baquara² Knowledge-based Framework for Semantic Enrichment and Analysis of Movement Data* by Fileto et al. [18]. This framework addresses the lack of formal semantics for trajectories (movement) data. The framework comprises ontology constructs for the description of movement segments and annotation of movement data with corresponding objects and concepts described in ontology collections. The framework produces semantically enriched movement data compliant with an ontology that enables movement analysis queries on the basis of application and domain knowledge. The framework components comprise: 1) the data structures and abstractions component; 2) the Baquara² ontology component; and 3) the semantic enrichment and analysis of movement data. The focus of this framework is on semantic annotation, enriches the meaning of data. However, framework evaluation is partially described. The paper mentions that framework



evaluations were query-based, but, specifications of analysed semantic attributes, functionalities and additional metadata are not made explicit.

There is the *Framework for Ontology Enriched Semantic Annotation of CCTV Video* by Vrusias et al. [19]. This framework links textual and visual semantics for video sequence annotations to enable the semantic transcoding⁷ of Closed Circuit Television video footages. The video annotation component entails the segmentation of video sequences into moving objects, described by trajectories and blobs. Thereafter, machine learning techniques are used to mine the visual semantics (i.e., actors, events and locations) which are thereafter annotated with Closed Circuit Television ontology constructs and finally, summaries of video shots are produced in text format [19]. The focus of the framework is also on semantic annotation, however, it does not factor in the framework evaluation, meaning, there is no assessment of the semantic enrichment input, processing and output functions to determine the degree of the framework effectiveness.

The other framework with a complete exclusion of framework evaluation is the *Semantic Enrichment and Fusion of Multi-Intelligence Data* by Hull et al. [20] to facilitate easy identification of meaningful and situation-specific information. This is an automatised processing for the generation of semantic tags to improve the net-centric system's discovery services by enriching sensor data and intelligence reports. This framework pertains to semantic enrichment of data extracts from criminal reports with the military ontology metadata tags for identification of vehicle theft events. This process is performed in three steps as follows. Step 1 involves the identification and parsing of the various fields of data streams. The process begins with the transformation of raw source data which can be textual or binary and structured, semi-structured or unstructured, into XML format. This step identifies the major structural features of the source data, including header, body and footer features [20]. Step 2 involves the extraction of concepts and relationships found in those fields and the generation of the appropriate XML tags. Each structural feature is scanned so that domain features can be identified and annotated. This step produces XML annotations for salient features identified in the source data. No annotations are produced if the data stream does not describe a related

⁷ Transcoding is the process of indexing video shots with the appropriate semantic labels to enable easy access to video shots that are relevant given a specific semantic search.



concept or relationship [20]. Step 3 involves the translation of XML tags into OWL representations. This step maps feature extracts to concepts and relationships defined in target ontologies. For instance, in an application that describes the identification of vehicle theft events in messages (incident reports), the XML representation for a vehicle theft event is a <VEHICLE-THEFT> element. In the target ontology, this concept could be mapped to: a subclass of ACTION-EVENT called *Looting* or a subclass of ACTION-EVENT called *Robbery*. If it is clear that the theft was directly from a person rather than an organisation because of the use of the verb *carjack* or *hijack* or if the car was stolen directly *from* a person, the *Robbery* subclass may be mapped to *group* or *location*. All other instances map to the *Looting* concept [20].

Overall, Step 1 deals with the transformation of input data, thus, fits into semantic enrichment pre-processing while both Steps 2 and 3 describe the semantic enrichment process. It is clear that the focus of the processing is on data disambiguation. Above that, the focus of the framework is clearly on semantic representation, but framework evaluation is excluded. Therefore, it is unknown what specific semantic attributes, functionalities and additional metadata are expected to be on the data as a result of the semantic enrichment and there is no mechanism to assess the extent to which those semantic attributes, functionalities and additional metadata contribute to the identification of vehicle theft events.

Also, in the category of frameworks with a complete exclusion of semantic enrichment impact assessment is *An Ontology-based Framework to Support Performance Monitoring in Public Transport Systems* proposed by Storti et al. [21]. This framework is founded on the ontology of performance indicators and their formulas, business objectives, dimension analysis and the Transmodel⁸ relations. The other component is a reasoning framework which enables logic capabilities for collaborative designer interactions. This framework is focused on semantic annotation and representation, but, it is difficult to assess its effectiveness without framework evaluation mechanism.

The second research gap regards the lack of semantic enrichment frameworks that address the enhancement of semantic capabilities for OD value (knowledge) creation potential maximisation. Though, there are few frameworks that make mention of the

⁸ Transmodel is public transport information systems for the European reference data model.



capabilities to be created, the semantic enrichment processes do not explicitly reflect on how those capabilities are to be realised. One such is the *Framework for Semantic Enrichment of Sensor Data* by Moraru and Mladenović [22]. The purpose of this framework is the automatisation of semantic enrichment of sensor descriptions and measurements. The framework provides meaningful data to improve the usability and accessibility of sensor data and measurements and comprises: 1) the *sensor descriptions and measurements component* which involves analysis of the sensor descriptions and measurements to identify associated semantic concepts. The sensor descriptions contain the metadata defining the sensor characteristics while the sensor measurements contain numerical values quantifying the changes of sensor properties; 2) the *ontology component* involves selection of the most appropriate ontologies out of the existing ontologies, that which describe sensor characteristics and provide context for sensor measurements. This also pertains to the extension of the selected ontologies with concepts specific to the application domain; 3) the *enrichment component* comprises of descriptions enrichment, and measurements pre-process and enrichment. The sensor descriptions are enriched with semantic concepts and the sensor measurements are processed to generate new features enriched with semantic concepts; 4) the *semantic repository of sensor data component* deals with the storage of the enriched sensor descriptions and measurements; and 5) the *data consumer component* deals with the data consumption modes such as query end-points, semantic browsers and inference engines. In addition to automising semantic enrichment process, the framework aims to provide meaningful data and to increase the effectiveness of sensor networks by enhancing the usability and accessibility of sensor data [22]. The framework is focused on achieving both semantic annotation and representation; however, assessment of enhanced features is not explicitly factored into the framework and this creates a void in understanding the degree of understandability, usability and accessibility effected through semantic enrichment.

Overall, semantic enrichment frameworks facilitate the creation of semantic value through structural and contextual semantic enhancements to raw OD resources. Determining the effectiveness of semantic value creation frameworks involve the evaluation of the entire input, processing and output framework functions for its



effectiveness. This is qualitatively measured by the resource's demonstration of the degree of semantic attributes, functionalities and additional metadata value-added to raw OD (resulting semantic value) through the semantic value creation processing. However, as semantic enrichment frameworks with consideration for semantic value assessment are lacking, the main research question remains partly addressed and answered. The answer fully addresses the semantic value creation process but is silent on the architectural and functional design elements for semantic value creation maximisation. This justified the need for frameworks that facilitate the assessment of created semantic value in addition to facilitating the creation of semantic value. This is the research gap that the ODSE Framework addressed.

1.3 Research Problem Statement

There is mounting global pressure for governments to open performance data for national, regional, continental and global development agendas on the Web of Data. For this data to yield value to consumers, its attributes, functionalities and additional metadata must be semantically enriched. The focus of semantic enrichment frameworks reviewed so far are on the creation of semantic value, with either partial consideration or complete lack of consideration for the assessment of resulting semantic value. This makes it difficult to assess both the semantic value-adding and value-added potential resulting from the semantic value creation framework, thus, resulting in a knowledge gap for OD publishers and consumers aspiring to maximise on the potential value of OD. Therefore, the problem that this research addresses is lack of semantic enrichment frameworks that enrich semantic capabilities with a focus on the maximisation of OD value (knowledge) creation potential.



1.4 Research Questions

To address the research problem at hand, the main research question stated in Section 1.1, was broken down into five sub-questions that point to the implementation of the main research deliverable, formulated as follows:

- **RQ1:** What are the semantic publication and consumption needs for the preliminary semantic enrichment framework design?
- **RQ2:** What preliminary semantic enrichment framework is derived from the combination of potential semantic enrichment processes, approaches, capabilities, techniques, technologies and their relations?
- **RQ3:** What are the findings of the preliminary semantic enrichment framework demonstrations?
- **RQ4:** What is the completeness, cohesiveness and effectiveness of the preliminary semantic enrichment framework?
- **RQ5:** What is the reusability potential of the semantic enrichment framework finalised on the basis of the preliminary framework demonstration and evaluation findings?

1.5 Research Objectives

The five research questions above were answered through five specific research objectives, formulated as follows:

- **RO1:** To identify the semantic publication and consumption needs for the preliminary semantic enrichment framework design,
- **RO2:** To design and develop the preliminary framework,
- **RO3:** To demonstrate the preliminary framework,
- **RO4:** To evaluate the preliminary semantic enrichment framework completeness, cohesiveness and effectiveness, and
- **RO5:** To finalise the preliminary semantic enrichment framework for reusability.



1.6 Research Contribution

The main contribution of this research is a conceptual design artifact, the Ontology-Driven Semantic Enrichment Framework (ODSE Framework) which embodies the semantic value creation solution with design features for semantic value creation maximisation and evaluation. The ODSE Framework models the data transformation process for a transition from flat text OD publication and consumption to semantic publication and consumption, with a focus on maximising the value published and extracted for use and reuse. Specifically, it models the conversion of flat text resources published in formats with limited semantic capabilities such as pdf, HTML, spreadsheet, and image formats into knowledge resources with semantically enhanced capabilities (rich semantic formats) to facilitate semantic publication and consumption. The focus of the ODSE Framework is on the maximisation of knowledge creation with potential for cascading impact on knowledge mining. The ODSE Framework has eight components distributed across three processing stages pertaining to semantic enrichment input pre-processing, semantic enrichment processing and the resulting output processing. Following is a detailed overview of the ODSE Framework.

The *ODSE pre-processing stage* (Stage 1) outlines the pre-processing input, implementation and output processing and comprises three components. The first component (Component 1.0) is the *pre-processing input component* which involves the gathering of semantic enrichment input (source data and metadata resources). This pertains to the collection of OD publication resources in its raw form (flat text OD resources) for the development agendas, as well as, the collection of OD publication and consumption specification resources pertaining to the development agendas and the OD programs. The second component (Component 2.0) is the *pre-processing implementation component* which deals with the processing of raw OD resources in regard to OD analysis, cleaning, transformation and integration. This component also involves the identification of semantic value creation needs in terms of semantic publication needs (OD program, presentation and representation needs) and the consumption needs (OD development agendas support needs). The last component of Stage 1 (Component 3.0) is the *pre-processing output component* comprising the processed raw OD resources and the identified semantic publication and consumption needs.

The *ODSE processing stage* (Stage 2) outlines the selection of the appropriate ODSE approach, the actual data enrichment processing and the description of semantic value creation output (knowledge resources). The *ODSE processing* comprises three components. The first component (Component 4.0) is the *ODSE Approach component* (Ontology-Driven Semantic Enrichment Approach component) which is at the centre of the semantic value creation and assessment processing. Specifically, the ODSE Component 4.0 pertains to the choice and implementation of the applicable semantic value creation approach for the semantic publication and consumption needs at hand. This research proposes two ODSE approaches, the *generic ODSE approach* which makes use of generic ontologies and the *domain ODSE approach* which makes use of custom domain ontologies in the semantic value creation processing. So, the ODSE approach is defined by the scope of ontologies (*generic vs domain*) chosen for the semantic value creation processing. On the one hand, the term *generic ontology*, in this research context, embraces both the top-level/upper/foundational/reference ontologies, as well as any unrelated domain ontology. The process for choosing generic ontologies involves: 1) Analysis of source data and metadata resources for ontology search literals; 2) Identification of ontology enrichment requirements; 3) Search for potential ontologies; and 4) Selection of suitable ontologies. On the other hand, the term *domain ontology*, in the context of this research, identifies a custom domain ontology. The adapted METHONTOLOGY approach for the development of custom domain ontologies entails: 1) Specification of ontology goal, scope and intended purpose; 2) Analysis of data sources for ontology concepts, attributes and relationships; 3) Ontology conceptualisation; 4) Ontology formalisation; and 5) Ontology evaluation. The second component (Component 5.0) is the *data enrichment component* which pertains to *Resource Reconciliation* and *Resource RDFization*. Resource Reconciliation entails ontology upload or the creation of Web links to local server, followed by the process that replaces the local metadata values with Web ontology values. The second data enrichment sub-process, Resource RDFization, involves the creation of flat text resource links to Uniform Resource Locators of ontology targets for semantic value creation and the generation of Resource Description Framework (RDF) structured data, triples, in the form of *Subject-Predicate-Object*. The last component of Stage 2 (Component 6.0) is the *enrichment output*



component which provides a description of the semantic value creation output for both the generic and domain knowledge resources.

The *ODSE post-processing stage* (Stage 3) outlines knowledge resources structuring for ease of knowledge distribution and use, as well as, knowledge resources deployment for publication and consumption. The *ODSE post-processing* comprises two components. The first component (Component 7.0) is the *knowledge resources structuring component* which pertains to the categorisation of knowledge resources by functionality support needs for packaging (publication and consumption needs), followed by the creation of knowledge resource packages through triples integration (publication and consumption packages). The last component of Stage 3 (Component 8.0) is the *knowledge resources deployment component* which pertains to the availing of knowledge resources for use. Knowledge resources are first uploaded to knowledge mining environments such as semantic publication libraries, query endpoints, semantic browsers, and inference engines for use and that is followed by the execution of knowledge mining operations on knowledge resources. A detailed discussion on the ODSE Framework is presented in Chapter Nine.

1.7 Dissertation Structure

Firstly, Chapter One introduces the research and provides an outline of the research background which is followed by the research problem statement, questions, and objectives. Chapter Two presents literature review on OD value (knowledge) creation and the potential of a combination of ontology, LOD and semantic enrichment in the creation of OD value and the potential impact on knowledge extraction and use. Chapter Three covers the semantic enrichment framework development methodology, and research limitations and challenges. Chapter Four presents the semantic value creation needs for the semantic enrichment framework which encompasses the semantic publication and consumption needs. Chapter Five covers the design and development of the preliminary semantic enrichment framework. Chapter Six presents the first preliminary semantic enrichment framework demonstration pertaining to the choice of the semantic value creation approach in relation to the selection of *generic* ontologies and the development of the custom domain ontology. Chapter Seven presents the second



preliminary semantic enrichment framework demonstration which pertains to the implementation of the ODSE pre-processing, processing and post-processing stages. Chapter Eight covers the preliminary semantic enrichment framework evaluation and finalisation. Framework evaluation pertains to the assessment of the preliminary semantic enrichment framework processes' completeness, cohesiveness, and effectiveness while the framework finalisation pertains to the generalisation of the final conceptual design artifact for enhanced reusability potential. Chapter Nine presents the final ODSE Framework, the reusable conceptual design artifact and the ODSE Framework implementation support tools. Lastly, Chapter Ten closes with discussions on research conclusions, recommendations, future works and publications.



CHAPTER TWO: RESEARCH LITERATURE

2.1 Introduction

This research pertains to the development of the Ontology-Driven Semantic Enrichment Framework (ODSE Framework) with potential to maximise OD value (knowledge) creation and consequently knowledge extraction potential. So, the literature review explores the OD concept from four angles: 1) OD as the future driver; 2) OD value (knowledge); 3) knowledge reuse potential; and 4) knowledge creation challenges. That is followed by the exploration of the potential of the combination of ontology, LOD and semantic enrichment in the creation and use of knowledge. The purpose of the literature review was to further explore the knowledge research gap that the study addresses and to also explore semantic enrichment process modelling features from existing semantic enrichment frameworks. It was also to identify applicable semantic enrichment technologies for the design and development of the ODSE Framework.

2.2 Open Data

Data refers to any qualitative or quantitative statements or numbers that are (or assumed to be) fact-based. Data may be classified as raw or primary data (i.e., direct from measurement), or derivative of primary data as calculation output but not yet analysis or interpretation output [23]. Data is open when it is publicly available for free use, reuse, and redistribution [24]. Therefore, OD is a shared, public good/resource [25]. OD excludes data subject to valid privacy, security or privilege limitations, as governed by national statutes [25].

2.2.1 Open Data as the Future Driver

The global future lies in the creation of knowledge-based societies and economies that are by nature founded on shared knowledge [26], [27]. Filipović et al. [28] argues that to facilitate the establishment of knowledge societies and economies, the environment for knowledge sharing has to be favorable and if the highest effectiveness is to be reached, knowledge, as a global public good should be shared without limitations – Open Knowledge. The Open Knowledge International [29] considers knowledge to be open if all have access to the information they need to understand and shape the world around



them. OD has high potential to be leveraged for this effect. OD creates open knowledge bases of knowledge resources/goods of countries, thus, increasing the chances of generating new knowledge by using the accumulated knowledge and then distributing that knowledge to wider economic and social networks [30]. Therefore, at the heart of open knowledge is OD [31].

2.2.2 Open Data Value (Knowledge)

Generally, data is a resource with little intrinsic value. Data value is created when data is put to use and it infinitely increases with use [32], [15]. The distinguishing characteristic of data is that while most resources depreciates in value and yield decreasing returns with extensive use, data appreciates in value and yields increasing returns with increased use [15]. This makes it one of the most valuable assets. However, it is almost impossible to value [15]. That is why Moody & Walsh [15] refers to data as a valuable but unvalued asset. The European Commission [3] has reported that most of the EU28+ countries have difficulty assessing the impact of OD, with some effects visible while in overall an estimation of the impact of OD remains a challenge. Due to the difficulty or impossibility of quantifying data value through formal methodologies, OD or semantic value is qualitatively derived from use impact.

2.2.3 Open Data Value (Knowledge) Reuse Potential

The high value of OD is attributed to its quality of being infinitely sharable, thus, culminating in high potential for reuse [25]. Reuse of OD as a public resource has the potential to increase economic and social value. The economic value of OD pertains to the value added to the public realm through the generation of economic activity while the social value of OD pertains to value added to the public realm by contributing to social cohesion and well-being [33]. Both the social and economic value of OD is the key driver in knowledge-based economies and societies.

Knowledge-Based Economies

A knowledge-based economy is an economy in which knowledge plays an important role as the main engine for economic growth [30]. Stehr [16] defines knowledge as an effectual vehicle for economic value creation [28]. The Knowledge Economy Index 2015 and the Global Innovation Index 2018 reveal that high income countries are knowledge-



based economies [34, 35]. According to the Organization for Economic Cooperation and Development, knowledge-based economies are directly based on the production, distribution and use of knowledge and information [36]. Moreover, the key driver of knowledge-driven economic transformation is knowledge that is made available to all, thus, culminating into a state of open knowledge. Therefore, as data is one of the sources of knowledge, OD plays a key role in driving the shift towards knowledge-based economies so as to drive innovation, growth and employment [37], [38]. In-fact, OD environments support the creation of knowledge networks that are the source of innovation [39]. A knowledge-based economy, an economy based on the sharing and diffusion of knowledge, provides an opportunity for governments to increase the well-being of their people [28]. It is also a panacea for sustainable development as it propels economic growth and the sustainability of that growth as well [40]. Therefore, the ability to transform OD into knowledge, and knowledge into activity and innovation, enables economic prosperity, inclusiveness and sustainability [26].

The economic value of OD generated through innovation involves the creation of new products, services and models of service delivery, or engagements with citizens; economic growth; and transparency [41]. The European Commission estimates the economic value of opening-up and reusing government data to €40 billion annually [42]. Also, OD provides opportunities to develop business models and technology platforms for innovation brokering services that connect ideas, solutions, technologies, structured knowledge repositories and networking along the value chain to address innovation needs [39].

Knowledge-based Societies

The OD impact goes beyond economic transformation. It plays a big role in the attainment of a knowledge society. In the process of transforming the economy, the entire society is transformed as well into a knowledge society. According to Stehr [16], the development of a knowledge society is a gradual process that is shaped by the way that new technologies, data, needs, and imaginations interact and are configured to produce new possibilities and innovations. Therefore, the way in which the OD movement is mobilising for data openness is a key feature of the emergence of a knowledge society [43]. Wessels et al. [43] further argues that key in the OD movement mobilising a knowledge society is



the strong conviction of the value of OD utilisation within the wider society, the attention to the institutional aspects of making data open in an inclusive way, and a focus on the technological infrastructure.

Knowledge societies thrive on shared knowledge and this necessitates an open approach to knowledge [26]. That is, a knowledge society is one where knowledge is open to all. Its central value is openness, which means that data, information and knowledge are seen as a shared societal asset. Therefore, a knowledge society is one organised through digital networks in which knowledge serves as the basis for production, innovation and consumption [43]. This enables free use of data by all to engage and participate in economic, social, political and cultural projects. Thus, a knowledge society is one in which everyone can create, access, utilise, share and disseminate information and knowledge, so that individuals, communities, and peoples are empowered to improve their quality of life and to achieve their full potential [44]. The major characteristic of a knowledge society is the free flow of knowledge, and with the continuous free flow of information and knowledge, society is able to dynamically adapt to constant changes in the economic environment [45].

2.2.4 Open Data Value (Knowledge) Challenges

The OD/semantic value challenge is of great significance in the Information Systems research discipline as, currently, there is mounting global pressure for governments to open performance data for development agendas on the Web of Data, as stated in the introductory section. The big challenge is that, OD is not by default open knowledge [46]. The Asia Foundation states that OD, in most cases, is not in formats that developers can use to create applications [47]. That is, when released in diverse flat text formats, OD is not availed to the public in any meaningful form and therefore unusable. Consequently, if data is not used, it cannot yield economic benefits and it becomes a liability as no value is extracted from it [15]. That is, publishing datasets to a national portal as a public resource does not necessarily mean that OD is being used. Thus, simply declaring datasets to be open does not necessarily culminate in practical and valuable use [3], [48]. Most OD publications are difficult to use and reuse due to standards, semantics, platforms, formats, definitions and schema heterogeneities that result in difficulties pertaining to OD access, interoperability, and metadata aggregation [3], [42], [49]. Specifically, the barriers to the



publication and use of OD pertain to: availability and access, find-ability, usability, understandability, quality, linking and combining data, comparability and compatibility [41]. Most of these barriers compromise OD use and reuse, thus, limiting the capability to derive value from the data.

The realisation of OD benefits depends on the: 1) organisation of data for ease of location (accessibility); 2) comprehensiveness of data for analysis (intelligibility); 3) ease of assessment of data reliability and data producer competency; and 4) data format to support diverse use and reuse purposes (usability) [23]. Furthermore, augmentation of OD with proper background information and explanatory metadata is required to produce open knowledge. To produce open knowledge, OD has to be semantically structured and linked to support data integration, interoperability and reuse [46]. Specifically, it has to undergo transformation in which the datasets are cleansed and then converted to machine-readable formats. That is when they can be exposed to the public in ways and formats that enable software developers to easily integrate into applications [50]. That is, open knowledge is what OD becomes when it is useful, usable and put to use [29]. This, therefore, necessitates the creation of OD value in terms of semantic attributes, functionalities and additional metadata to facilitate easy and effective use and reuse of the data. In short, for flat text OD resources to yield value to consumers, it must be semantically enriched.

The semantic enrichment of data introduces the second challenge, the semantic valuation challenge. Currently, there is lack of semantic assessment tools to inform semantic enrichment decision making. Also, though semantic enrichment creates semantic value (knowledge), there is lack of mechanism for semantic valuation. This makes it difficult to determine the semantic value-adding potential of the semantic enrichment process, as well as, the semantic value-added potential resulting from the semantic value creation implementation. This has resulted in a knowledge gap for OD publishers and consumers with a focus on knowledge creation and use.



2.3 Ontology Potential for Open Data Value (Knowledge) Creation

The focus of this research is on web ontologies that play a central role in the semantic web. This section starts with a description of the ontology concept, followed by discussions on the web ontology.

2.3.1 Ontology

An ontology is a conceptual representation/model of an observed phenomenon or reality through formalised, explicit description of concepts and their corresponding attributes and relations for a given domain [51], [52]. Ontologies include class hierarchies, relationships between classes and instances, class attribute definitions and axioms that specify constraints [53]. Ontologies have advantages regarding the expressiveness and relationship definition and offers flexible modelling and automated reasoning capabilities [11]. Ontologies are best suited for defining a common vocabulary to facilitate knowledge sharing in a given domain across heterogeneous systems or environments [12]. That is, ontologies define a way to exchange semantic information between people and systems [54].

Ontologies can also be classified according to their scope. *Top-level ontologies*, also known as upper or foundational ontologies, describe domain or problem independent concepts (e.g., space, time, object, etc.). The second category, *domain ontologies*, or *domain-specific* ontologies represent the particular meanings of terms as they apply to a specific domain. That is, they describe entities and other information related to a given domain (e.g., education, health, finance, etc.) [55], [56]. The third category, *application ontologies*, describes ontologies developed for specific use or application focus and pertains to multiple domain modelling [57].

2.3.2 Web Ontologies

Web ontologies are frameworks encompassing both the terminological and assertional knowledge. They are specified in standard web ontology languages (RDF and OWL) and published on the Web for use and extension in open environments [58]. Web ontologies are the core frameworks of the semantic web that enable intelligent agents to access and link data in a more intelligent way [59]. Web ontologies can be in the form of visual or coded representations of the domain of discourse, or both.

The visual or conceptual representation of a knowledge domain constitutes lightweight ontologies [60]. Lightweight ontologies are mostly in taxonomy form as directed graphs comprising nodes and edges. Concepts are represented by the nodes and relations among concepts are represented as edges. Taxonomy reflects inheritance or subsumption relationships (i.e., *is-a*, *is-a-superclass-of*, *is-a-subtype-of*, and *is-a-subclass-of*). The other graph structure provides additional domain details and may use relations such as *is-parent-of*, *part-of*, and *located-in* [61]. Taxonomies are minimally defined and less formal, thus limiting their understanding, sharing, use, and impact to only humans. The coded or formal representation of a knowledge domain constitutes heavyweight ontologies that maximise on language expressivity (range of constructs for explicit representation); semantics (clarity of meaning in a language specification); and mathematical rigour (integrity and soundness provided by a language) by combining both visual and coded representations [60]. Figure 2.1 illustrates the difference between lightweight and heavyweight ontologies in terms of their levels of expressiveness.

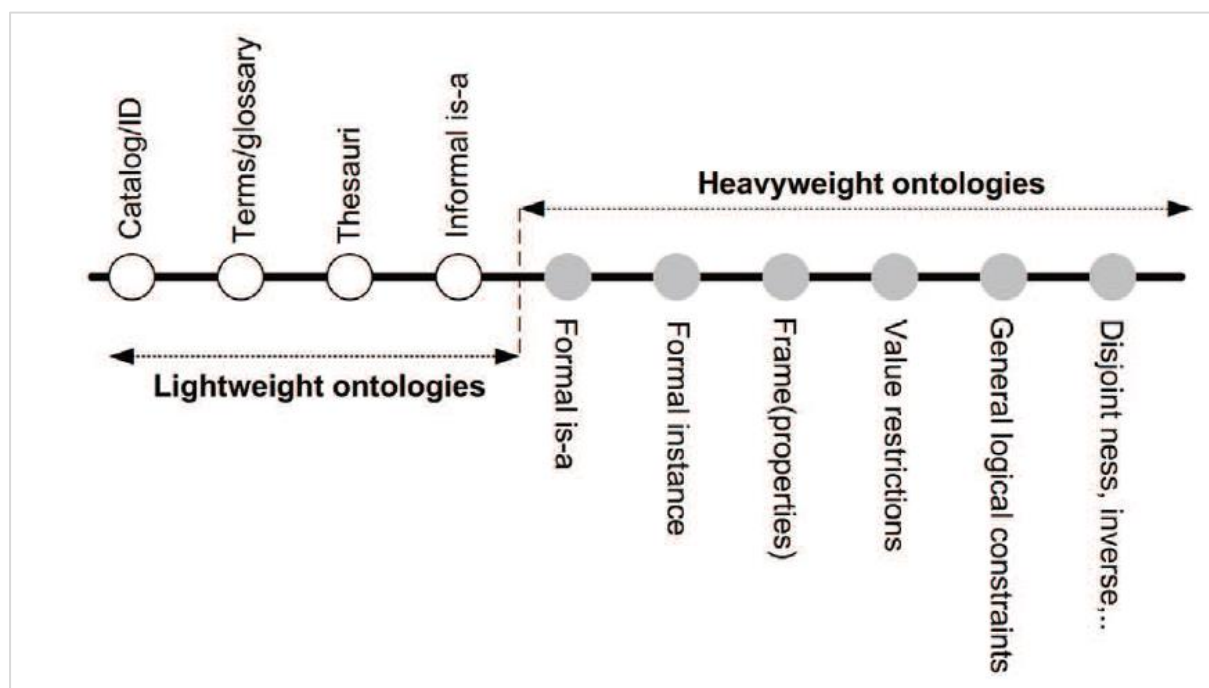


Figure 2. 1: *Ontology Expressiveness Spectrum in [62] adapted from McGuinness (2003)*

Unlike lightweight knowledge models, heavyweight models express relationships and enable users to interlink concepts in diverse ways [63].



In general, ontologies have high potentials in enhancing semantic capabilities of OD to improve the quality or value of resulting knowledge resources. Ontologies offer numerous benefits. They constitute a repository of interlinked concepts that represent concrete and abstract real-world objects for a given application domain, thus enabling the creation of knowledge bases [51], [64], [65]. With an ontology, organisational knowledge is encapsulated to build a set of knowledge bases that embeds details pertaining to rules, policies, processes, definitions and relationships [65]. In addition, they offer flexible modelling capabilities by simplifying the drawing of frameworks, patterns, and computer programs for process automation, thus, enhancing both the explanation and understanding of simple and complicated processes [53], [11]. Above that, they serve as references for communication, problem solving and knowledge-sharing for both human and machine-based tasks [52], [64]. They are used to create semantically rich data functionalities [53], [66]. Furthermore, storing data with respect to an ontology enables reuse and maintenance of knowledge within and between domains and disciplines [11], [52], [53]. Moreover, ontology, with its rich expressiveness, supports automated reasoning through the use of inference rules [11], [67], [59].

The next section explores the application of ontologies in the description of performance indicators with the aim of identifying the descriptions that best fit performance indicators for the development agenda domain.

2.3.3 Performance Indicator Ontologies

Several existing ontology models were reviewed to identify candidate ontologies for use or extension in this research. This review focused on performance indicator ontologies in the development agenda domain, specifically, the Sustainable Development Goals Interface Ontology (SDGIO) and on other generic performance indicator ontologies.

Sustainable Development Goals Interface Ontology (SDGIO)

In the development agenda domain, the United Nations Environment Programme has developed the Sustainable Development Goals Interface Ontology (SDGIO) to enable the logical representation, definition and linkage of SDGs entities to relating terminology in glossaries and other web resources [68]. Specifically, the SDGIO acts as a semantic bridge between SDG concepts by providing clarification on the nature of and interlinkages between 1) the SDGs, their targets, and indicators and 2) the large array of entities they



refer to [69]. The SDGIO was developed for global use so as to ensure that the SDG indicators are fully consistent across the SDGs [68]. The SDGIO clarifies the semantics behind some of the key terms used in the SDG indicator definitions to provide a comprehensive representation of the semantics of the SDG process [70].

Some of the key concepts integrated into the SDGIO are: access, adverse, adequate, appropriate, basic, benefit, efficient, effective, hazard, informal, infrastructure, integration, promote, resilience, resource, risk, sustainable, and vulnerability. The modelling approach for these terms is as follows. The semantics used in the SDGIO are defined with respect to a core set of universal terms. These core terms are generalised for application to a variety of levels of granularity and specialised domains. Below, is an illustration of the treatment of risk, hazard, vulnerability, and resilience in the SDGIO [68]:

- All entities have a set of innate vulnerabilities, tied to their physical makeup.
- A vulnerability is the potential for an entity to undergo change, usually negative, as a result of being exposed to a hazard.
- A hazard, then, is an entity which can cause the realisation or manifestation of one or more vulnerabilities in another entity.
- Risk is seen as the potential for an entity to realise or manifest a vulnerability as a result of being exposed to one or more hazards.
- Using logical relationships, the SDGIO seeks to clarify entities considered vulnerable to given set of hazards and deduce the risks they are exposed to.
- Quantifications of entities such as risks can then be associated with the specific forms of risk affecting entities vulnerable to particular hazards.
- Estimations of the chance of a particular risk manifesting, say, risk of flooding, can be precisely represented then by distinguishing between the entities relevant to the risk and information about them.
- From the semantics such as these, resilience can be approached as the capacity of an entity to recover its previous composition, qualities, and functionality following some process wherein one or more of that entity's vulnerabilities were realised.

In addition to the above literature review on the SDGIO, an analysis of the SDGIO to determine its suitability to the Botswana case study was carried out. The analysis was



focused on constructs applicable to this research case study, pertaining to SDGIO representation of SDGs and related targets and indicators. Below is the analysis.



Figure 2. 2: SDGIO Class Hierarchy

Note: The term *class* identifies a concept and the term *subclass* identifies a sub-concept.

Class SDGIO_00000000 (*Sustainable Development Goal*): As depicted in Figure 2.2, *Sustainable Development Goal* concept is a subclass of IAO_00000005 (*objective specification concept*) which is a subclass of IAO_00000033 (*directive information entity*)



concept). The term *directive information entity* refers to an information content entity whose concretisations indicate to their bearer how to realise them in a process [71]. This hierarchical structure contextualises goal description as an objective specification of some directive information entity. Also, as depicted in Figure 2.3, class `SDGIO_00000000` holds values for SDGs descriptions. That is, this class has individual instances which are the descriptions of SDG Agenda goals.

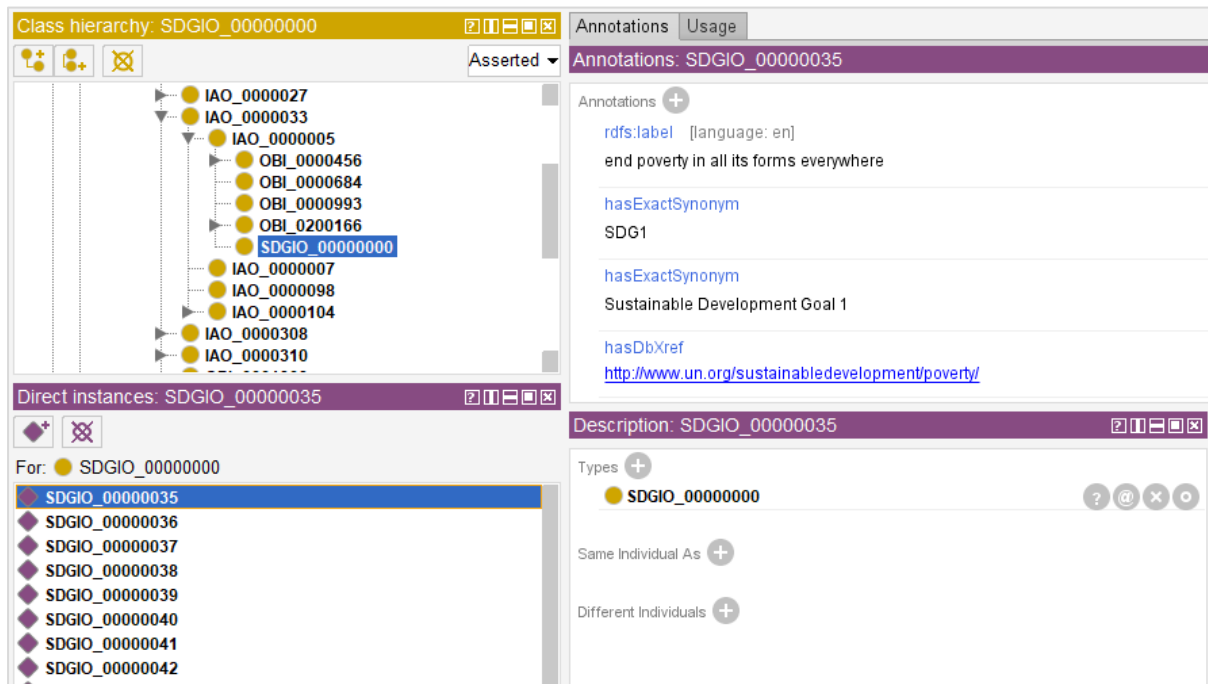


Figure 2. 3: Class `SDGIO_00000000` and Instances

As depicted in Figure 2.3, a sample `SDGIO_00000000` instance being `SDGIO_00000035` represents the description of SDG 1 which is, *end poverty in all its forms everywhere*. The `SDGIO` has no data and object property assertions (shown in Figure 2.9). Instead, each instance has the `rdfs:label` annotation and two more custom annotations being: `hasExactSynonym` with values for both the SDG description code and name; and `hasDbXref` with a reference to the UN SDGS data portal (i.e., <http://www.un.org/sustainabledevelopment/poverty/>). These metadata are reflected in RDF format as shown in Figure 2.4.



```
<http://purl.unep.org/sdg/SDGIO_00000035><http://www.w3.org/1999/02/22-rdf-syntax-ns#type><http://purl.unep.org/sdg/SDGIO_00000000"
<http://purl.unep.org/sdg/SDGIO_00000035><http://www.w3.org/1999/02/22-rdf-syntax-ns#type><http://www.w3.org/2002/07/owl#NamedIndividual"
<http://purl.unep.org/sdg/SDGIO_00000035><http://www.w3.org/2000/01/rdf-schema#label>"end poverty in all its forms everywhere"
<http://purl.unep.org/sdg/SDGIO_00000035><http://www.geneontology.org/formats/oboInOwl#hasDbXref><http://www.un.org/sustainabledevelopment/poverty/>
<http://purl.unep.org/sdg/SDGIO_00000035><http://www.geneontology.org/formats/oboInOwl#hasExactSynonym>"SDG1"
<http://purl.unep.org/sdg/SDGIO_00000035><http://www.geneontology.org/formats/oboInOwl#hasExactSynonym>"Sustainable Development Goal 1"
```

Figure 2. 4: RDF Representation of SDGIO_00000035

As reflected in Figures 2.3-2.4, the predicate `hasExactSynonym` relates the subject `SDGIO_00000035` (SDGI description) to two different objects being *SDG1* (SDGI code) and *Sustainable Development Goal 1* (SDGI name). If this property is reused, it will create duplicate values. To address this, a different property has to be defined to differentiate between these two relations.

Class SDGIO_00000001 (target description): The *target description* concept is defined as a class with individual targets as shown in Figure 2.2. This structure means that class `SDGIO_00000001` holds values for SDGs target descriptions. For instance, the first individual instance for class `SDGIO_00000001`, `SDGIO_00000057` provides description for target 1.1 as reflected in Figure 2.5. In addition, the individuals for this class do not have data properties which are the foundation of knowledge base structure but have two annotation properties: `rdfs:label` with the target description value; and `SDGIO_00000074` (target code – custom defined annotation property) with the target code value of 1.1.

```
<http://purl.unep.org/sdg/SDGIO_00000057><http://www.w3.org/1999/02/22-rdf-syntax-ns#type><http://purl.unep.org/sdg/SDGIO_00000001>
<http://purl.unep.org/sdg/SDGIO_00000057><http://www.w3.org/1999/02/22-rdf-syntax-ns#type><http://www.w3.org/2002/07/owl#NamedIndividual>
<http://purl.unep.org/sdg/SDGIO_00000057><http://www.w3.org/2000/01/rdf-schema#label>"By 2030 eradicate extreme poverty for all people everywhere currently measured
as people living on less than $1.25 a day"
<http://purl.unep.org/sdg/SDGIO_00000057><http://purl.unep.org/sdg/SDGIO_00000074>"1.1"
```

Figure 2. 5: RDF Representation of SDGIO_00000057

For this case study, target code (SDGIO class name) and description (SDGIO custom annotation) are needed to serve as variables for the actual data values. Therefore, this case study knowledge base structure is compatible to them (target code and description) defined as data properties for SDG target class.

Class SDGIO_00000002 (Sustainable Development Indicator): As reflected in Figure 2.6, *Sustainable Development Indicator* concept is modelled as a subclass of superclass



BFO_0000040 (material entity - an independent continuant that is spatially extended whose identity is independent of that of other entities and can be maintained through time) together with subclasses SDGIO_00010014 (medicine - a material entity that is used in the process of preventing, treating, diagnosing, or studying disease or disorder); SDGIO_00010024 (right bearer - an organism or organisation that is the bearer of a right.); SDGIO_00010026 (unemployed - a person or population that is not employed, but eligible for and currently seeking employment); SDGIO_00010028 (owner - a person or organisation that participates in an ownership process wherein they realise their role as owner by exercising their right to use, sell, rent, or gift an object); and SDGIO_00010030 (employed - a person or population that participates in employment processes during a specified time interval). SDGIO_00000002 is represented as an empty class.

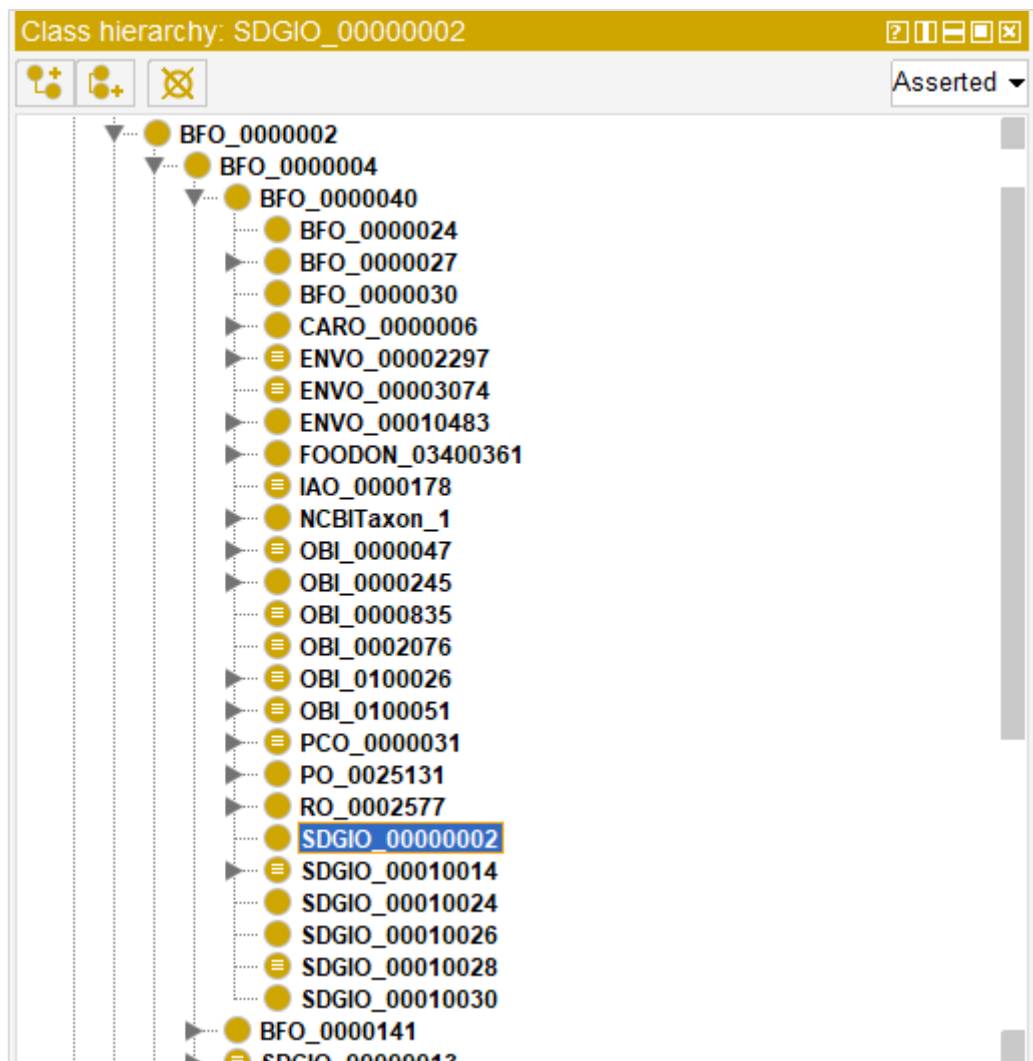


Figure 2. 6: Class SDGIO_00000002



For this case study knowledge base needs, Sustainable Development Indicator concept (SDGIO_00000002) has to serve as the superclass for all SDGs.

Class SDGIO_00000003 (Sustainable Development Goal Indicator Value): The Sustainable Development Goal Indicator Value concept is a superclass for all indicator classes. Each indicator is created as an empty class which is an independent subclass of SDGIO_00000003 (SDGI value).

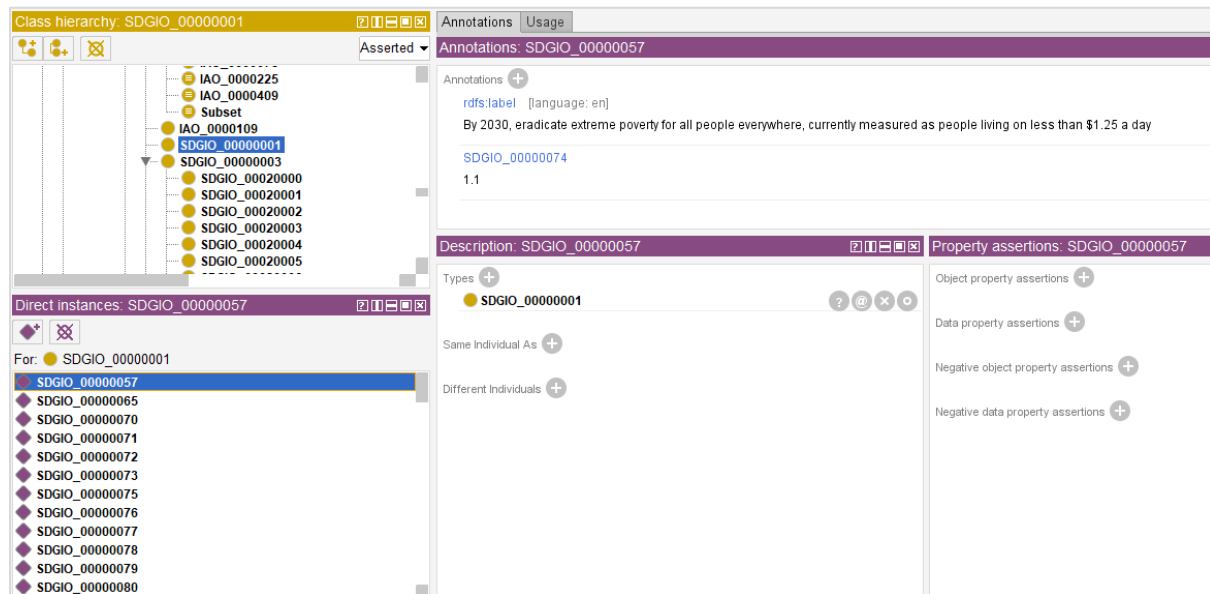


Figure 2. 7: SDGs superclass (SDGIO_00000001) and SDGs superclass (SDGIO_00000003)

Figure 2.7 depicts SDGs Indicator superclass (SDGIO_00000001) together with its instances and SDGs Indicator superclass (SDGIO_00000003) together with its subclasses. The SDGIO_00020000 (Indicator 1.1.1 description) is defined as a class, specifically, subclass of SDGIO_00000003. This design structure means that class SDGIO_00020000 holds values for instances of Indicator 1.1.1 descriptions. According to the RDF extract in Figure 2.8, the description for Indicator 1.1.1 of SDG Agenda is attached to the annotations as a value for the `rdfs:label` property and is not defined as an individual member of this class. So, class SDGIO_00020000 is without individuals.

```
<http://purl.unep.org/sdg/SDGIO_00020000><http://www.w3.org/1999/02/22-rdf-syntax-ns#type><http://www.w3.org/2002/07/owl#Class>
<http://purl.unep.org/sdg/SDGIO_00020000><http://www.w3.org/2000/01/rdf-schema#label>"Proportion of population below the international poverty line by sex, age,
employment status and geographical location (urban/rural)"
<http://purl.unep.org/sdg/SDGIO_00020000><http://purl.unep.org/sdg/SDGIO_00000003>"1.1.1"
<http://purl.unep.org/sdg/SDGIO_00020000><http://purl.unep.org/sdg/SDGIO_00000003>"1.1.1"
<http://purl.unep.org/sdg/SDGIO_00020000><http://www.w3.org/2000/01/rdf-schema#subClassOf><http://purl.unep.org/sdg/SDGIO_00000003>
```

Figure 2. 8: RDF Representation of SDGIO_00020000



Moreover, as reflected in Figure 2.8, other custom annotation properties are defined for this class being SDGIO_00000242 (UN SDG Indicator ID) with a value of 1.1.1; and SDGIO_00000227 (UNSD SDG Indicator Code) with a value of C010101.

Further analysis revealed that SDGIO has no data property definitions. Thus, its structure does not support the holding of data values for class instances as evidenced in Figure 2.9.

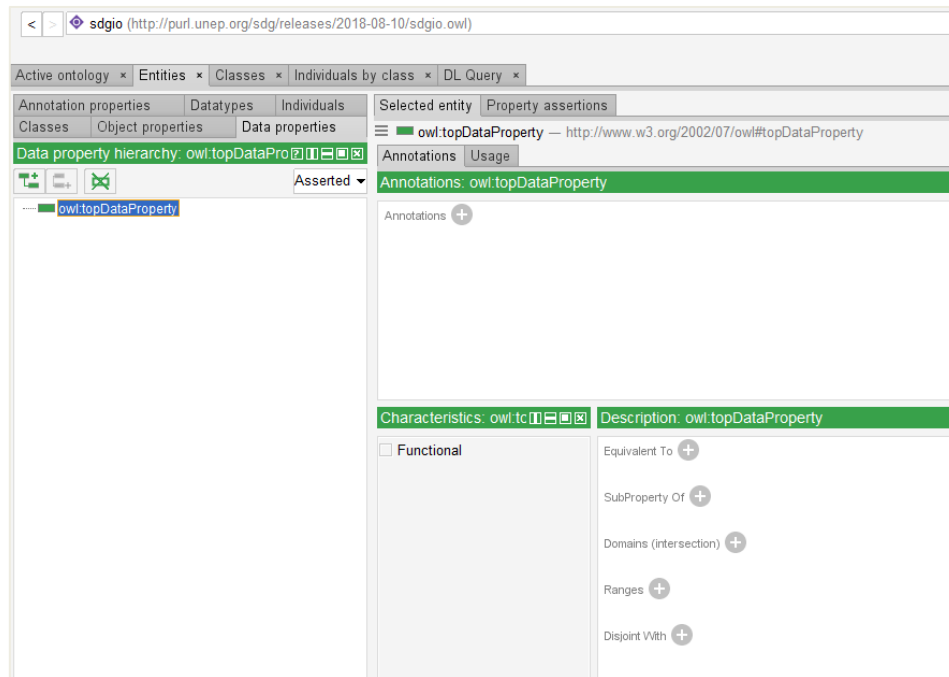


Figure 2. 9: SDGIO Empty Data Properties Window

Also, the defined object properties do not fit the Botswana case study application scenario as depicted in Figure 2.10.



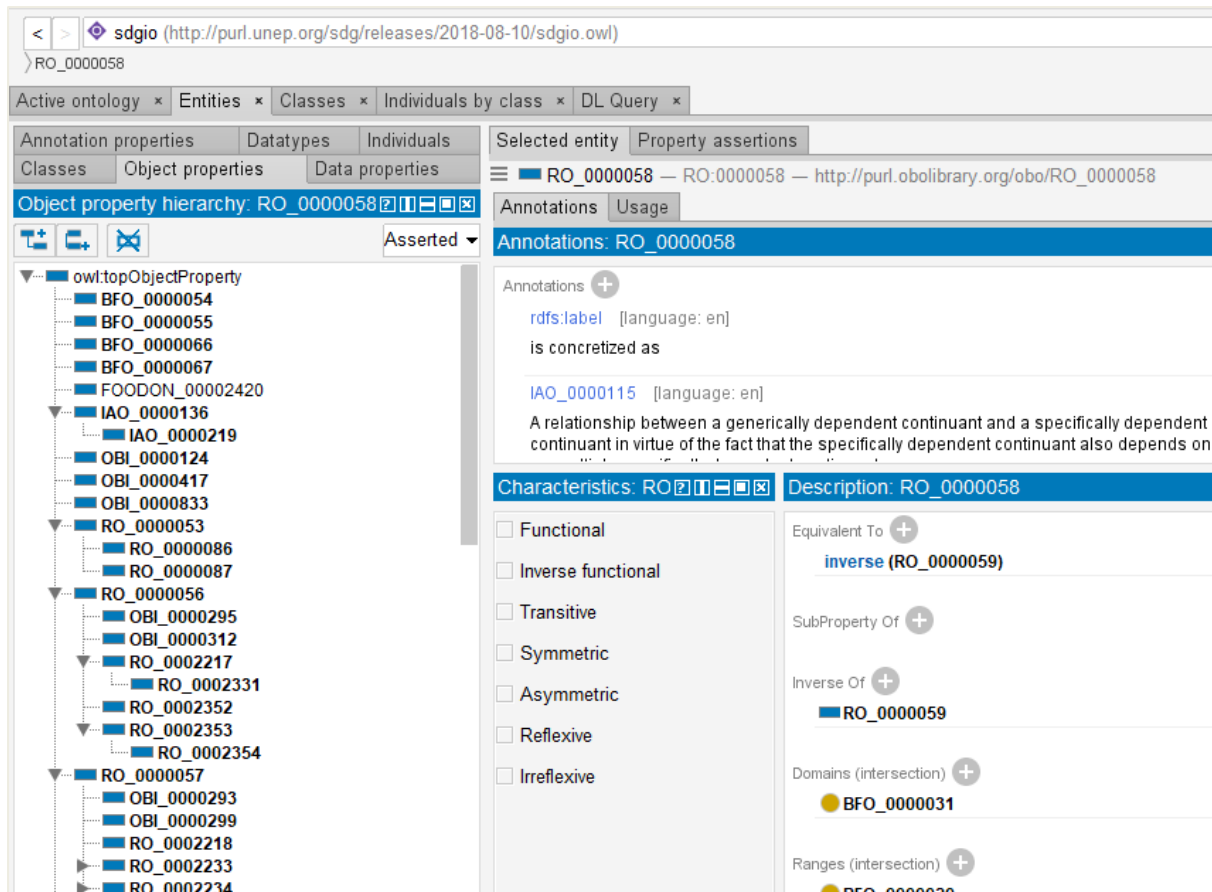


Figure 2. 10: SDGIO Object Properties Window

In overall, the review revealed that SDGIO modelling of concepts is based on annotation properties, backed up by the concepts' modelling approach explained in the SDGIO literature discussions preceding SDGIO analysis and Figures 2.9-2.10 of the SDGIO analysis. That is, the SDGIO focus is strictly on providing the definitions of concepts and semantic definitions based on a single-property type (annotation property type) and offers clarity relating only to that property type. The research problem at hand prompts for data and object property definitions, and axioms (definition of constraints) representing various reasoning capabilities for holistic definition and representation of the relationships and mappings across performance indicators within the Botswana Integrated Indicator Framework context. Furthermore, as an interface ontology with a focus on annotation modelling, the contextualisation of SDGIO concepts differ from the knowledge base contextualisation needs for this research in several ways:

- SDGIO contextualises SD target (SDGIO_00000001 - Figure 2.3) and SD goal descriptions (SDGIO_00000000 - Figures 2.2-2.3) as classes while the



contextualisation needs for this research prompts that they be defined as data properties for instances of SD target and SD goal classes,

- SDGIO contextualises descriptions of instances for SDGI entities as class entities (SDGIO_00000003 modelling - Figures 2.7-2.8) while the contextualisation needs for this research prompts that they be defined as data properties for instances of SDGI class, and
- SDGIO contextualises codes for SDGI, SDG target and SDG as annotation properties while the contextualisation needs for this research prompts that they be defined as data properties.

These contextual differences limit the reusability of SDGIO resources to the definitions of concepts, yet the knowledge base model has to be built on the attributes, relations, constraints, and actions on these concepts – data and object properties, and axioms.

Moreover, some annotation properties are associated with more than one object. For instance, `hasExactSynonym` is related to the same object with different values being *SDG1* (SDG code) and *Sustainable Development Goal 1* (SDG name). If this property is reused, it creates duplicate values when queried. To address this, a different property has to be defined to differentiate between these two objects.

The differences discussed above pose a need for domain specific ontology definitions that address the needs of the Botswana Integrated Indicator Framework. However, the ontology must be modelled to enable some linkages to SDGIO annotation properties for reuse of globally standardised meaning of concepts and to facilitate global interoperability in the Web of Data as well.

Generic Performance Indicator Ontologies

The review of other performance indicator ontologies, ontologies not specifically related to the development agenda domain, revealed that their focus is mostly on formal description of performance indicator definitions in terms of mathematical structures. One of this kind is KPIOnto, a logical representation of Key Performance Indicators proposed by Diamantini et al. [24]. KPIOnto integrates engineers' proposed definitions in a collaborative environment. At the core of KPIOnto is Indicator concept representing performance measures. Its attributes are indicator identifier, acronym, name, definition, unit of measurement, formula, aggregation function, business object, and dimension. The



Key Performance Indicator aggregative structure is represented by the `dimension` concept which is made up of sub-concepts, namely, `organisation` and `Time`. The third KPIOnto main concept is `formula` which represents Key Performance Indicator computations pertaining to the aggregation function, the way the formula is presented, the semantics of the formula, and references to its components [72]. One advantage of KPIOnto is that it provides an in-depth modelling of the Key Performance Indicator structure. It also enables the formal representation and manipulation of the formula structure in a collaborative environment. This further enhances the formal representation capabilities. The limitation of this model is that it is focused on the indicator concept. For a holistic representation of the `indicator` concept, it has to be presented in relation to other key performance constructs (i.e., goals, targets, etc.) in addition to the `dimension` and `formula` constructs.

This research reviewed a generic performance indicator ontology titled, *Defining Manufacturing Performance Indicators Using Semantic Ontology Representation* proposed by Chryssolouris, Pintzos and Matsas [11]. The core concept of the Manufacturing PPIs ontology is `PPIs` which is presented in relation to sub-concepts `PPI Generic Attributes`, `PPI Specific Attributes` and `PPI Formula`. The strength of this model is in its in-depth representation of the `indicator` concept in relation to its attributes. However, there is a need to represent the `indicator` concept in relation to other performance constructs for clarity in relations and operations, and interoperability within and across production performance entities.

The other generic performance indicator ontology reviewed is the *Ontology for Performance Measurement Indicators' Comparison* by Freitas et al. [73]. It addresses the challenges of indicator imprecision and vagueness. Their model is aimed at facilitating effective performance indicator comparisons. According to these researchers, complete knowledge about the origins of the data involved in the calculation of performance indicator value serves as the basis for more precise indicator comparison. This in turn improves the understanding of its meaning and helps in making decisions on the selection of suitable performance indicators for comparison and decision support tasks. They further argue that better understanding and comparison of indicators requires knowledge on performance indicator value calculations pertaining to entity relations and



performance indicator variables comprising description, value, relevance (weight), assessment criteria, formula and other performance indicator value attributes, time relations, and relations to other performance indicators [73]. These extended descriptions offer precise performance indicator understanding and facilitate effective performance indicator comparisons and decision support actions. However, the model partly addresses performance indicator disambiguation only.

The research also reviewed the *Process Performance Indicators* (PPIs) ontology named, PPINOT Ontology. Process Performance Indicators (PPIs) are a key asset for the measurement of the achievement of strategic and operational goals in process-oriented organisations. PPINOT addresses the challenge of PPIs definitions which pertains to either their informal definition in natural language or too formal definitions that render them difficult to understand. PPINOT provides the definition of PPIs in a way that is easy to learn; reuse; unambiguous and understandable. It further enables automated analysis and knowledge inferences regarding relationships between PPI definitions and other process elements [74].

Overall, the reviewed ontologies have research gaps pertaining to indicator modelling. This research addresses the semantic representation challenges pertaining to performance indicators' ambiguities and diverse definitions, performance indicators high levels of aggregation, performance indicators disparate publications, and performance indicators unidimensional presentation formats. However, most performance indicator ontologies are focused on formal description of performance indicator definitions based on mathematical structures. They model the indicator concept in relation to some data and object properties. The models do not reflect on the associated performance constructs such as goal, target, agenda, etc. In addition, the models are lacking in addressing the semantic needs of a consolidated indicator framework. So, they partly address performance indicators ambiguities and diverse definitions, and are limited in other semantic needs of this research. This, therefore, poses a need for ontology models that extend the above models with relating performance constructs and at the same time fulfilling the Botswana Integrated Indicator framework needs.



2.4 Linked Open Data (LOD) Potential for Open Data Value (Knowledge) Creation

The LOD standard is a semantic publication technique for the creation of semantic value. It was introduced by the inventor of the World Wide Web, Tim Berners-Lee [75] to make data accessible on the Web in machine readable formats and linked to other data. The LOD standard is implemented by converting flat files to a set of RDF triples to provide data about these resources and to link them to other web resources [76]. It addresses the challenges of dispersed and heterogeneous datasets on the Web of Data by creating interlinkages of OD in the World Wide Web [76]. The LOD standard follows a set of best practices for publishing and connecting structured data on the Web [50]. Figure 2.11 illustrates Tim Berners-Lee's [77] 5-star rating scheme for LOD publications.

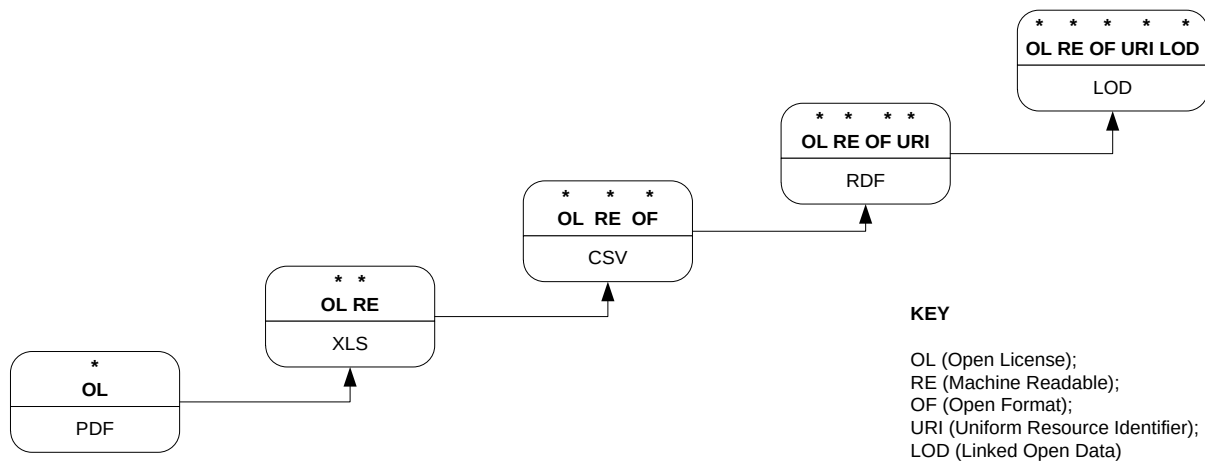


Figure 2. 11: Tim Berners-Lee's 5-Star scheme for LOD Publications [75]

The 1-Star is assigned to data made available on the Web in text, pdf, and tabular formats with an open licence [78], [79], [77]. The 2-Star is assigned to data made available on the Web in a proprietary machine-readable format which is well structured and predictable, and having well-established and known ways to query and use. The limitation of OD publications at this level is that access to the data requires specific software or systems [78], [77]. The 3-Star is assigned to data made available on the Web in non-proprietary machine-readable structure where standard rules apply to the data presentation structure. For example, a Comma Separated Value (CSV) file can be opened in any spreadsheet software and can also be imported into any software. A CSV file of BIUST University details could look like this:



```

Headings for university attributes
{universityName,location,country}
Data following the same pattern as the headings
{BIUST,Palapye,Botswana}

```

Figure 2.12: CSV Sample

As demonstrated in Figure 2.12, a CSV file format is simple and predictable when it comes to coding a program, to read it and to use it. It has the same number of columns (attributes or characteristics) separated by a comma and the first row is usually the column headings [78], [77]. Therefore, this data can be readily consumed by software programs from across the Web and developers can consume the data from within their programs or applications. Data at the 4-Star level possesses all the features at the lower levels and additionally complies with open standards such as RDF. RDF is a framework for describing resources on the web with the use of Uniform Resource Identifiers (URIs). A URI can either be an address to a static webpage location that contains the information, or an address plus a query to the database at that location. RDF breaks down data into a series of facts. For example, the CSV record in Figure 2.12, {BIUST,Palapye,Botswana} could be expressed in RDF triple form as:

```

(Subject) (Predicate) (Object)
:university :hasName :BIUST
:university :isLocatedIn :Palapye
:university :relatesToCountry :Botswana

```

Figure 2.13: RDF Sample

As demonstrated in Figure 2.13, a fact is expressed as a *triple* in the form of *Subject-Predicate-Object* [78], [77]. The facts are stored with corresponding classifications and relationships in a triplestore, and the triplestore represents collections of relationships, classifications, explicit and inferred facts and unstructured data for understanding and analysis at a much finer grain of detail [63]. Above that, RDF is not constrained to the structure imposed by the triplestore and does not become obsolete with the evolution of the data, as the data structure is expressed through the links within the data itself. In addition, when changes in the data structure occurs, they are reflected in the database through changes in the links within the data. Moreover, RDF enables the inference of new



facts from the existing data as well as the enrichment of the available knowledge by accessing LOD resources [80]. Data at the 5-Star level embraces all of the above features as well as linking the published data to other existing data to provide context. Linked RDF describes meaning by referring to the source of meaning on the Web, resulting in Linked Data. This is achieved by linking the web resource to the Web definition source [78], [77].

2.5 Semantic Enrichment Potential for Open Data Value (Knowledge) Creation

To start with, the value and impact of OD can be realised only if it is put to use; therefore, the core outcome of a successful OD initiative is to make data usable and semantic enrichment serves that purpose [50]. Semantic enrichment refers broadly to various technologies and practices used to add information at the semantic layer (semantic metadata) of resource data (content) to enable machines to connect and make sense of the data, and to also enable better search of content [50], [81], [50], [82], [83], [63]. These technologies, called Semantic Web Technologies, enable the consolidation of heterogeneous datasets from disparate sources into meaningful and valuable information [80]. They employ a data-centric architecture founded on a standardised model for data publishing and interchange, RDF, which enables the transition from OD to LOD by interlinking web resources. Furthermore, they offer the capability to import, transform, express, store and access heterogeneous data from multiple sources which are then integrated and interlinked as RDF statements into RDF triple-store [80].

2.5.1 Semantic Enrichment Vocabularies

Vocabularies constitute the Knowledge Organisation Systems and serve as *concept control* for semantic enrichment [50]. They control concept grouping and hierarchic relationships as discussed by Clarke and Harley [50] and shown in Figure 2.14.

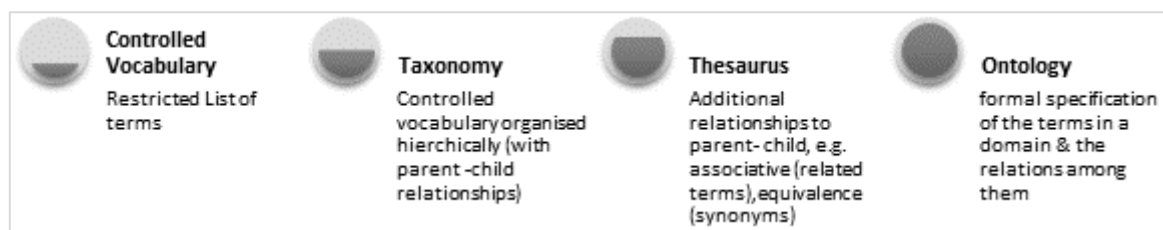


Figure 2. 14: Semantic Richness of Knowledge Organisation Systems [50]



According to Figure 2.14, the different Knowledge Organisation Systems used for content classification and organisation range in complexity from a simple controlled list of common terms to a highly complex ontology [50]. At the lower level of semantic richness is controlled vocabulary which is comprised of a restricted list of terms. That is superseded by a taxonomy which is a controlled vocabulary organised into parent-child relationships. Richer formats are thesaurus and ontology that add explicit structures, metadata, and definitions to content. A thesaurus includes additional relationships to parent-child such as associative (related terms) and equivalence (synonyms) relationships. The highest expressive Knowledge Organisation System is the ontology which provides the formal specification of terms in a domain and the relations among them [50]. An ontology is a conceptual representation of an observed phenomenon or reality through formalised, explicit description of concepts and their corresponding attributes and relations for a given domain [84]. Ontologies include class hierarchies, relationships between classes and instances, class attribute definitions and axioms that specify constraints [53], [66].

In short, the level of OD value (knowledge) is dependent on semantic capabilities that differ according to the level of semantic richness of the publication format. The highest level of semantic richness is the product of semantic enrichment which creates value by assigning meaning to text elements through the use of Knowledge Organisation Systems. The complexity of the Knowledge Organisation System determines the semantic value-adding potential. Knowledge Organisation Systems increase the value of OD by augmenting published data with corresponding semantic meaning, thus, enhancing their understanding and enabling advanced analytical capabilities. Of all Knowledge Organisation Systems, ontology offers the maximal enrichment value and this validates the adoption of the ODSE approach to semantic enrichment as having potential to maximise OD value. The next section explores various semantic enrichment methods.

2.5.2 Semantic Enrichment Methods

There are a diversity of semantic enrichment methods which can be categorised according to content type as either *structured*, *semi-structured*, and *unstructured* data. Semantic enrichment methods for structured data enhance the quality and discoverability of metadata with more contextualised meanings through the use of



vocabularies and other resources that have embraced LOD. Semantic enrichment methods for semi-structured data transform it into structured data through entity-based semantic analysis and annotation to extend access points. Semantic enrichment methods for unstructured data are focused on exploring it and generating knowledge bases to support knowledge discovery [83].

2.5.3 Semantic Enrichment Models

Semantic enrichment may be classified into four enrichment models: *tags*, *attributes*, *relations* and *ontologies* [85]. A *tag annotation* element is a keyword, or a term assigned to a resource that describes a particular resource property. Tags are at the bottom level and represent the easiest form of annotation. An *attribute annotation* element is a pair of two components, namely the name and value of the attribute. A *relation annotation* element is a pair of two components, the relation name and the related resource. The annotated resource is linked to relation with the relation name. The *ontology* model describes the metadata that align a resource or a part of it with some of its properties and characteristics description according to a formal conceptual model. Ontologies are at the top level and represent the complex form of annotation [85]. The ontology model of semantic enrichment is also known as *ontology-driven semantic enrichment*.

2.5.4 Semantic Enrichment Approaches

At the core of every nation and organisation is the aim of increasing the value of its assets, including data. OD value depends on the semantic richness of publication formats that can be enhanced through semantic enrichment with the use of ontologies and LOD technologies.

Ontology-Driven Semantic Enrichment

As stated in Subsection 2.5.1, ontologies, as part of the knowledge organisation systems have the highest enrichment value compared to controlled vocabularies, taxonomies and thesaurus. Therefore, ontologies are key drivers of semantic enrichment processing with potential for OD value creation maximisation. Ontology-driven semantic enrichment pertains to the process of adding meaning to web content by mapping it to terms defined in ontologies [53], [66]. Leveraging ontologies in semantic enrichment for OD value (knowledge) creation offers numerous benefits. With ontology-driven semantic



enrichment, unstructured text can be annotated with information that links related concepts together and clarifies the concepts ranking in a hierarchical family. This enables search for information on a certain concept to gather many different references of it as a class, achieving a recall impossible with unstructured text information [86]. Above that, semantic enrichment through shared ontologies make heterogeneous contents from diverse data providers semantically interoperable, thus, enabling data usage across data silos [83]. Also, it is difficult to understand whether the query results reflect the intended query, without a vocabulary definition that restrict potential interpretations of the used classes and roles or properties towards their intended meaning [79]. Furthermore, adding the ontology layer to openly published data subjects its description to logic and this enables reasoning capabilities on the data [11], [53]. Ontologies increase semantic interoperability among datasets, applications and the parties that exchange data on the semantic web [87], [88]. They are used to achieve both internal and external integrations and interoperability. They connect external resources to content and enable the exchange of information across applications [50].

Linked Open Data Driven Semantic Enrichment

The application of ontologies in semantic enrichment is according to the LOD standard, the data publishing technique that make data accessible on the web in machine readable formats and linked to other internal and external datasets [75]. Thus, LOD represents the interlinkages of OD in the World Wide Web to create semantic value. However, according to Janowicz et al. [79], this practice is limited in advancing the OD mandate as the mere practice of linking RDF triples does not necessarily make the data more (re)usable to humans or machines. Hitzler and Van Harmelen [88] have observed that some of the LOD work only just publishes the data, and ignores the value that can be derived from the data by annotating it with ontologies [88]. Above that, for maximal use and reuse, OD must be published in platform independent and machine readable formats, and made available to the public without restrictions that would impede its reuse [89, 48]. This can be achieved through the introduction of LOD platforms that enable data links to other related data of interest [76]. To realise the LOD benefits, LOD platforms act as hubs for publishing and sharing data. The LOD platform handles metadata and also allows for storing and keeping datasets to be used for applications and services. They foster free, unlimited data access



to the general public. They offer a point of datasets consolidation, minimising the efforts of developers and consumers in searching, accessing, downloading and sharing datasets of interest [90].

2.5.5 Semantic Enrichment and the Formation of OD Value (Knowledge)

Attributes

The OD/semantic value (knowledge) creation potential of semantically enriched OD resources is determined by the degree of formation of semantic enrichment, richness and mining attributes.

Semantic Enrichment Attributes

Semantic enrichment attributes, in the context of this research, refers to semantic value-added to flat text resources to enhance semantic capabilities at both the syntactic and semantic levels. The focus of this research is on ontology-driven semantic value (knowledge) creation. This pertains to the enhancement of semantic capabilities with the addition of the syntax (structure) and semantics (meaning) layers to flat text OD resources through the use of ontologies. Semantic capabilities of interest in this research are OD accessibility, understandability, interoperability, integrability and intelligibility. Ontologies offer high and flexible expressiveness, relationship definition, modeling and reasoning capabilities [11]. Therefore, they play an important role in enhancing semantic capabilities.

Firstly, *OD accessibility* pertains to the organisation of data for ease of discovery [41], [23]. Despite the OD abundance on the Web, its meaning is not constrained/linked to vocabulary repositories and this has resulted in the lack of expressivity in datasets thus, making it difficult to access it through search functionalities [50], [79]. For instance, the information needed to answer queries may be available on the Web, but may not be in RDF format or may be only partially available as RDF [91]. Semantic enrichment improves OD accessibility by adding a class of metadata to enable content discovery [50]. Semantic metadata provides the meaning of web content in a way that computers can process to find, filter, and connect [50]. That is, words transformed into an easy-to-search format enhances accessibility capabilities. The second challenge pertains to the limited availability of LOD on the Web. Though OD publications on the Web of Data are



expanding, there is too little Linked Data, thus limiting OD access [91]. Ontologies play an important role in enhancing OD accessibility capabilities and improving search and discovery of content for faster information retrieval and more accurate search result [50], [63].

Secondly, *OD understandability* pertains to the presentation of data for ease of comprehension, specifically, the expression of data in appropriate languages, symbols and units and in the specific context of use so it is easy to read and interpret [41], [23], [92]. This can be achieved through semantic enrichment. Semantic enrichment frees knowledge from the limitations of non-machine-readable texts it is locked into [63]. This process involves associating texts with semantically richer formats that explicitly express the structures, metadata, definitions, context and data relations to make it more meaningful, and enable machine readability and seamless processing [90, 63], [86]. With semantic enrichment, ambiguous, difficult-to-interpret, messy and heterogeneous content is converted into a neat set of machine-processable data pieces, thus bringing clarity to the knowledge content of an area which becomes the basis for communication, sharing and problem solving for both human and machine-based tasks [78]. Also, semantic enrichment enables different stakeholders in different sectors to make correct inferences on the data and to reach a consensus in the interpretation of the data as well. Ontologies play an important role in enhancing OD understandability capabilities. Data from diverse sources must be presented according to an agreed, formal representation that captures important details pertaining to the data context, data structure and data meaning. This requires a common data representation format that preserves the data context and meaning as content is shared among a diversity of stakeholders [93]. The common understanding of concepts can be provided through ontologies [11]. Ontology is used to make domain assumptions and terminological ambiguity explicit by providing clear definitions of domain concepts, their attributes, relations between concepts, relations between concepts and instances, and applicable constraints [52], [53]. This creates a shared understanding of concepts, thus, enhancing the efficiency of knowledge-based systems and the analysis of domain knowledge [11], [52].

Thirdly, *OD interoperability* pertains to data sharing capability demonstrated in the ability to intermix diverse datasets [41], [23], [94]. The ability to interoperate different



datasets so that they work together depends on the interpretive capabilities of relations between domain concepts [93], [94]. The interoperability of OD can also be described in terms of shared inferences. For instance, if a publisher shares OD for public consumption, the data publisher and the consumers can only be truly semantically interoperating if the data consumers not only know the literal content of the data, but also understands the implicit consequences of that data. These shared, implicit consequences are made explicit through vocabularies [88]. Ontologies play an important role in enhancing OD interoperability capabilities. Describing the data by means of ontologies increases interoperability among datasets, applications and the parties that exchange data on the semantic web [87], [88]. This facilitates knowledge sharing in a given domain across heterogeneous systems or environments. In addition, ontologies unify datasets based on the formal, explicit domain specifications of shared conceptualisations for interoperability [12]. By externalising a common understanding of the information structure, ontologies can be used to achieve both internal and external interoperability for information exchange among people, software agents and across applications [50], [52], [53].

Fourthly, *OD integrability* pertains to integration capability, the ability to interlink diverse datasets [41], [23], [94]. The value of OD is not in isolated files, but in consolidating all the relevant data into a single platform and providing easy linked access [2]. The integration of OD involves combining RDF structured data from diverse sources, on a syntactic level, into a unified view for actionable business intelligence [88], [95]. According to the *Pareto Principle* (80/20 rule), integrating 20 percent of OD is estimated to yield 80 percent of the value [15]. Combining OD with other pertinent data enables the extraction of specific and hidden insights from it [15]. For example, linking different data sources to ontologies enable the extraction of deeper insights, and the uncovering of hidden relationships [63]. For semantic enrichment, RDF links use vocabularies in the descriptions of the data. Ontologies and other classification systems (controlled vocabularies, taxonomies, and thesauri) can be used to achieve both internal and external integrations. They can also be used to connect external resources to content [50]. So, describing the data by means of ontologies increases integration among datasets and applications [87]. With the ontology-driven semantic integration, the abstraction level of



the ontology used for the enrichment determines the integration level of the ontology. With a high-level ontology, numerous concepts are combined into a single high-level concept, while with a custom domain ontology, the level of integration is closer to the lowest level of granularity [63].

Lastly, *OD intelligibility* pertains to the comprehensiveness of data for analysis and comparability [41], [23]. Ontologies play a central role in OD intelligibility by enhancing the analytical capabilities of knowledge resources [70]. Content semantically enriched with an ontology can be analysed to determine potential relationships between concepts and the resulting relationship graph can reveal some important hidden insights [96]. When compared to unstructured content which must be synthesised from scratch each time it is consumed, semantically enriched content has been annotated with its meaning, thus enabling users to move quickly to more intelligence-rich information activities [96].

Semantic Richness Attributes

In the context of this research, semantic richness refers to the semantic value-added to flat text resources in the form of structure and context. The semantically enriched data pieces are the building blocks of intelligent data [63]. Intelligent data is systematically processed, and presented in meaningful form for active sharing, as well as, further assessments and interpretations in the Web of Data [97]. These are knowledge resources that embodies, both structurally and contextually, all that an application needs to interpret the data [50]. The data is characterised by clarity – explicit definitions and measurements and is actionable – supports decision making and implementation [98].

Semantic Mining Attributes

Semantic mining, in the context of this research, refers to the extraction/derivation/harvesting of semantic value-added to flat text resources. That is, knowledge extraction/harvesting. Semantic enrichment builds semantic mining capabilities. It enhances discoverability with an additional layer of semantic metadata that reduces friction in the discovery, access, consumption, and synthesis of OD. Ontology-driven semantic enrichment enhances search capabilities. The main problem that arises when querying structured data lies in the differences in terms of quality and semantics that occur in the attribute values, e.g., a length may be expressed with different units of measurement. These variations may ultimately render query results very



imprecise or even meaningless. The semantic enrichment of structured data connects structural elements or data elements (attribute values) to semantic entities in an ontology [86]. This enables effective search, faster information retrieval and more accurate search results [63]. By mapping text with information that links related concepts together, ontologies clarify where these concepts sit in a hierarchical family. This enables a user looking for information about a certain concept to gather the many different references of it as a class, achieving a recall impossible with information in the unstructured text alone [96]. Moreover, ontologies disambiguate unstructured text to improve the accuracy of search results. For instance, an application that is not being able to differentiate AIDS (Acquired Immuno-deficiency Syndrome) from hearing aids will likely retrieve examples of both for an 'aids' keyword search while a system which can connect text to semantic entities of an ontology is cable of distinguishing between the concepts and achieve greater precision in search results [96].

Ontology-driven semantic enrichment enhances query capabilities. By associating attribute values with descriptors such as unit and scale, makes the data more meaningful. These descriptions can then be used to augment the queries with selection criteria or transformation functions to produce high quality results [86]. Furthermore, ontologies enable the expression of deductive queries that improve the accuracy of data analysis and also produce results not computable with standard queries [86], [93]. The abstraction level of the ontology used for the enrichment determines the level of query details. A high level ontology unifies many different concepts into a single high level concept, while more precise (domain specific) ontologies enable queries that are closer to the lowest level of granularity. Moreover, if the enrichment structure possesses a rich semantic structure, with subclass relations or more sophisticated axioms, it will support a more expressive query rewriting. For instance, a query about a given class could be rewritten as a query about its subclasses if the class concept is not directly represented in the data source [86]. For querying semi-structured data, the role of semantic enrichment is either to describe, unify or differentiate between the multiple naming schemes produced by users. If the names used in the data source associates with corresponding entities in an ontology, then the ontology's vocabulary can be used to express unified queries that can be



automatically rewritten into the dataset's vocabulary to produce a complex query on the data source [86].

2.5.6 Semantic Enrichment Frameworks

This section discusses semantic enrichment frameworks for ontology and LOD-driven processing of OD. Ontology-driven semantic enrichment describes the process of adding a layer of topical metadata in the form of ontology concepts to content to achieve the following: 1) to augment the source metadata with additional terms; 2) to enable machines to make sense of the data; and 3) to build connections to the data [50], [81]. In the LOD context, semantic enrichment pertains to the creation of links between the enriched resources and others, preferably from an existing reference dataset [99]. The reviews were focused on exploring semantic enrichment processing, the semantic capabilities enriched and the semantic enrichment output evaluation for the degree of semantic value-added to output resources.

The first reviewed semantic enrichment framework is the *Europeana Automatic Semantic Enrichment Framework* presented by Manguinhas et al. [99] in three stages. First is the *analysis stage* which involves the analysis of metadata fields in the original resource descriptions, the selection of potential resources to be linked to and the derivation of rules to match and link the original fields to the contextual resource. Second is the *linking stage* which involves the matching of metadata values to contextual resources' values, and the addition of contextual links based on values with equivalent relationships to the dataset. Third is the *augmentation stage* which involves the selection of values from contextual resources to be added to the original object description. The *Europeana Automatic Semantic Enrichment Framework* enhances access, integration and interoperability capabilities of Europeana objects with cultural heritage metadata [81]. It links Europeana objects and interconnects them to resources from external datasets by matching their metadata to external semantic data and controlled terms. For instance, the link <http://dbpedia.org/resource/Costume> means that the Europeana object was automatically enriched with the concept of *Costume* from the DBpedia dataset.

Therefore, the first stage of the *Europeana Automatic Semantic Enrichment Framework*, the *Analysis* stage, pertains to semantic enrichment pre-processing while the second and third stages the *Linking* and *Augmentation* stages, pertain to semantic enrichment



processing. So, it does not factor in the product of semantic enrichment processing, thus, leaving a void of knowledge on the process performance. Above that, the framework is explicit on the enriched semantic capabilities being semantic access, integration and interoperability but there is no specification of semantic capabilities evaluation. So, the focus of the framework is on semantic enrichment of structure and meaning and there is no consideration for the evaluation of the processing end-product.

The *Smart City Semantic Enrichment Process* was proposed by *Bellini et al.* [100]. The data engineering architecture processing component of the entire processing enhances OD interoperability capabilities through semantic enrichment that combines ontology and LOD principles and techniques as follows: 1) Data ingestion and quality improvement whereby input from diverse sources undergoes data quality improvement through ETL transformations; 2) Data mapping and indexing whereby the data is mapped to a smart city ontology. Thereafter, the data is indexed and undergoes reconciliation for semantic interoperability enhancement; 3) Exploiting and exploring smart city data whereby the output data is stored in an RDF-store for applications' access/exploitation through Simple Protocol and RDF Query Language (SPARQL). The framework provides as a separate process, the data verification, validation and reconciliation services implemented through a set of specific SPARQL queries and program tools for RDF link discovery [100].

Therefore, the first stage of the *Smart City Semantic Enrichment Process*, the *data ingestion and quality improvement*, pertains to semantic enrichment pre-processing; the *data mapping and indexing* stages pertain to semantic enrichment processing; and the *data exploiting and exploring* stage pertains to semantic enrichment post-processing. The semantic capability to be achieved through the enrichment is explicitly stated as data interoperability and the evaluation of this capability is provided as well. So, the focus of the framework is on enhancing semantic interoperability capability.

The *Linked Open Government Data Process* documented by Ding et al. [1], facilitates the enhancement of government data quality, accessibility, integrability and reusability capabilities through LOD practices. It is a Web-based open ecosystem that interlinks the original data owners (such as government agencies), data processing service providers (such as entity resolution services), and data consumers (enterprises and citizens). It is



pioneered by Data.gov and Data.gov.uk and emerged on the linked-data Web as a way of facilitating opening, linking, and reuse of Open Government Data. Linked data offers data representation using URIs and the Resource Description Framework and data access via HTTP. It enables incremental Open Government Data publishing according to Tim Berners-Lee's 5 Stars of Linked Data discussed above. The roadmap of Linked Open Government Data process has the following three data-processing stages: 1) In the open stage, government agencies publish Open Government Data online in reusable formats and maintain central Open Government Data catalogs for datasets access; 2) In the link stage, community participants such as academia and the industry enhance released Open Government Data quality; and 3) In the reuse stage, developers extract the published Open Government Data to build high-value applications [1]. These three stages synergistically enhance the raw open government data from data providers using the combination of machine and human power and deliver higher quality data to data consumers. All in all, The Linked Open Government Data project is focused on achieving the global *integration* of government data; enhancing data quality and access; and promoting data reuse.

Therefore, the first stage of the *Linked Open Government Data Process*, the *open* stage, pertains to semantic enrichment pre-processing while the *link* stage pertains to semantic enrichment processing. The third stage, the *reuse* stage pertains to semantic enrichment post-processing. In addition to enhancing data quality, the framework is aimed at achieving the accessibility, integrability and reusability capabilities. However, the framework scope does not cover the semantic enrichment evaluation for semantic value-added to the data.

The *OntoCore Semantic Enrichment Process* published by O'Shea and Ross [101] enhances unstructured and semi-structured data with relevant concepts and relations extracted from publicly available LOD repositories. The process involves an exploration of concepts held in existing data and their relations and integrating them to semantic web resources. The OntoCore process outline is as follows: 1) the textual data is fed into OntoCore for analysis, 2) the key semantic entities in the data are identified, 3) the relevant LOD is selected for augmenting the input data stream, and 4) the original text data is augmented with most relevant conceptual and relational data through document metadata [101].



The OntoCore project is concerned with enhancing the integration and reusability of both unstructured and semi-structured data on the Web and this is a crucial step in opening up new product potential from existing data.

From the discussions above, the analysis of textual data pertains to semantic enrichment pre-processing; the identification of key semantic entities and both the selection of relevant LOD for augmentation and the data augmentation pertains to semantic enrichment processing. Though the framework enriches data integration and reusability capabilities, there is no provision on how to assess the enrichment impact due to lack of the output evaluation component.

The *Folksonomy Ontology Enrichment Process* was advanced by Angeletou [102] to address the lack of explicit semantics for folksonomies. Folksonomies are Web 2.0 systems that enable content use in large-scale intelligent applications. The basic elements of a folksonomy are users, resources and tags. A resource is a folksonomy content object such as YouTube video while a tag can be any sequence of characters attached to a resource [102]. Tag semantics are often unknown, thus, hampering resource retrieval and integration in cross platform applications. Semantic enrichment transforms the flat folksonomy tag space into rich semantic representations. This involves the annotation of Folksonomy resources (raw text entities) with semantic entities defined in online ontologies, thesauri and other knowledge sources. The process creates a semantic layer on top of folksonomy tag spaces in a three-step process that entails the identification of the meaning of the tags (raw text entities) attached to a resource; obtaining the formal semantics that correspond to these tags; and attaching the formal semantics to the resource. The process consists of three phases as follows [102]:

- 1) Phase 1, *lexical process* determines the type of tags to be considered as the basis of semantic enrichment,
- 2) Phase 2, *semantic expansion* pertains to sense definition and disambiguation for the tags to address the challenge of different contextual meanings for the same tag. It also addresses the issue of semantic web sparseness in which ontologies might not contain concepts that are syntactically equivalent to a given tag but concepts labelled with one of its synonyms. To overcome this limitation, *semantic expansion*



is performed for each tag using a combination of thesauri and other knowledge sources, and

- 3) Phase 3, *semantic enrichment* identifies the semantic entities that are relevant for each tag by leveraging the results of Phases 1 and 2. The relevant *semantic web entities* are selected during the *entity discovery* step by querying online ontologies. Then, in the *entity selection* step the *semantic web entities* are filtered to identify the ones that correctly correspond to the tags. Finally the *relation discovery* step identifies relations between the *semantic web entities*. The output of the *Folksonomy Ontology Enrichment* process is semantically enriched tags.

From the above discussion, all the three stages, the *lexical process* (phase 1), *semantic expansion* (phase 2) and *semantic enrichment* (phase 3) pertain to semantic enrichment processing. The framework enhances semantic capabilities in terms of data understandability, integrability, and intelligibility. However, the evaluation of semantically enriched tags for semantic capabilities value-added is missing.

The *Linked Open Data Framework to Enhance the Discoverability and Impact of Culture Heritage* was proposed by Gustavo et al. [103], to enhance information retrieval. This framework uses open knowledge bases such as Wikidata and GeoNames for the enrichment and disambiguation of locations in text. The framework is aimed at facilitating the use of LOD by cultural heritage institutions to share or disseminate their collections and to enrich their metadata. This semantic enrichment framework addresses the challenges relating to the handling of increasing growth of datasets such as ingestion, management, querying and enrichment of the data faced by collaborating cultural heritage institutions. This semantic enrichment framework supports the implementation of refinements required for the recognition and extraction of implicit relationships expressed in natural language to improve the user experience in information retrieval. Though, the focus of the semantic enrichment framework is on semantic retrieval, there are no measures to evaluate the semantic value of the retrieved output.

The conceptual framework for *Enriching Mobility Data with Linked Open Data* was proposed by Ruback et al. [104]. The purpose of the framework is to guide the trajectory enrichment process that generates semantic trajectories. This process entails the partitioning of trajectories into homogeneous segments that are annotated with



contextual information. Enriching the representation of movement data with external contextual data results in semantic trajectories. This framework addresses the difficulty and error-prone processing of the semantic enrichment of movement data due to the lack of a comprehensive framework. The framework benefits from LOD both as a unifying formalism and as the source of contextual data and it takes advantage of the LOD principles in two aspects. Firstly, representing trajectories according to the LOD principles offers a strategy to incorporate trajectories in a way they can be easily shared and reused. Secondly, the Web of Data can be used as the main source of contextual information to enrich movement data. The proposed framework has been designed to create a publishable movement data repository by incrementally building a semantic trajectory ontology. This approach makes the process application-oriented, simple, reusable, flexible and general enough to cover many different domains. The proposed process covers the trajectory segmentation and the LOD enrichment up to the construction of a semantic trajectory repository to be used for the analysis. Semantic enrichment takes as input a raw trajectory and a number of LOD sources and builds a semantic trajectory repository. This process is driven by the use of ontologies and it is structured into two main steps [104].

Step 1 deals with segmentation which takes as input a raw trajectory and some particular segmentation criterion and splits the trajectory into segments, which are the entities to be enriched. This step is driven by a Segmented Trajectory Ontology, which identifies the different types of segmentation required by a specific application. The Segmented Trajectory Ontology represents the trajectory entities featuring the parts to be enriched [104].

Step 2 deals with the enrichment which matches (or enriches) trajectory segments (e.g., stops or moves) to the most appropriate semantic entities made available by the external contextual information through the Linked Data Mashup view. A typical example is the matching of stops with Points of Interest in the tourism Mashup View which can either be a religious building, museum or transport. The enrichment step involves two main tasks: 1) the creation or selection (if it already exists) of the Linked Data Mashup, which integrates data from the relevant selected linked data sources. The mashup aggregates entities from the DBpedia dataset and the OpenNER Linked Dataset which consists of a



repository for the tourism domain that covers the Tuscany region. The mashup reuses terms from DBPedia vocabulary, Friend of a Friend (FOAF), RDF schema (RDFs), vCard vocabulary, HOntology and Accommodation Ontology; and 2) Linking trajectory segments with Linked Data Mashup entities. In this step, the mashup view ontology specifies the concepts of the mashup view (i.e., the conceptual model), which is the common vocabulary for integrating data exported by the selected Linked Data sources. The outcome of the enrichment step is the semantic trajectory ontology and an RDF repository containing semantically enriched trajectories. The semantic trajectory ontology contains all concepts and properties of the segmented trajectory and mashup ontologies, and also the definition of the semantic links between them. In the proposed framework each step is driven by the incremental use of ontologies that can be easily adapted to meet the application needs [104].

From the above discussions, step 1 which deals with trajectory segmentation pertains to semantic enrichment pre-processing while step 2 which covers the LOD enrichment up to the construction of a semantic trajectory repository to be used for the analysis pertains to both semantic enrichment processing and post-processing. Like most reviewed frameworks, the enriched output evaluation is excluded making it difficult to assess the effectiveness of the framework.

The *Ontotext*⁹ *semantic enrichment process* [63] proposal comprises five steps that define how information is given well-defined meaning and becomes ready to join other data pieces, enabling new combinations, insights and cross-references: 1) text is extracted from articles, documents or any form of unstructured data; 2) after sentences are split, the important concepts are identified through dictionary word lists; 3) machine learning algorithms classify and disambiguate the identified concepts; 4) relationships between the entities are also identified; and 5) the facts and the original reference to the articles are indexed and stored with corresponding classifications and relationships in a triplestore. The triplestore represents collections of relationships, classifications, explicit and inferred facts and unstructured data that allow organisations to understand and interrogate their content and data at a much finer grain of detail [63].

⁹ A software organisation specialising in semantic technology



Therefore, step 1 which deals with text extraction pertains to semantic enrichment pre-processing. Steps 2, 3 and 4 that deals with the identification, classification and disambiguation of concepts, and the identification of concept relationships as well pertain to semantic enrichment processing. Lastly, step 5 deals with semantic enrichment post-processing functions. The framework is intended to build sharing and reuse capabilities but the framework does not factor in the evaluation of these capabilities.

The *Semantic Enrichment Process for Medical Ontologies* was proposed by Lee & Geller [105] for ontology enrichment to enhance semantic search. According to this framework, semantic enrichment process is as follows.

First, the process takes as input a local ontology and a global ontology and produces as output a Terminological Knowledge Base that has in its lower level the same concepts as the local ontology (a one-level ontology that consists of concepts and categories) and these concepts are assigned to semantic types from the global ontology (a Terminological Knowledge Base in which the upper level is organised as an *IS-A* hierarchy of semantic types and the lower level as an *IS-A* hierarchy of concepts which exhibits wide and deep coverage of the concepts for the defined ontology domain) [105].

Second is concept matching, a process which either finds for a concept (cl) of a local ontology a corresponding concept (cg) from a global ontology or for a category (al) of a local ontology a corresponding concept (cg) from a global ontology. The result is a pair (cl, cg) or (al, cg): furthermore, for the second kind of pair, a mapping function Match is defined such that Match (al) is equal to cg. Two concepts (or a category and a concept) are considered corresponding when they are identical according to a suitable string match, or similar enough as strings to warrant the assumption that they stand for the same real world (abstract or concrete) entity [105].

Third is semantic assignment, a process which creates a pair of a concept (cl) or category (al) from the local ontology and a semantic type Sl, which is a copy of a semantic type Sg from the global ontology, such that Sg is the semantic type of the cg that was found during Concept Matching: (cl, Sl) or (al, Sl) [105].

Fourth is assignment propagation, a process that “inherits” a semantic type Sl from a local category al to a local concept cl, provided that an *IS-A* relationship holds between cl and al, and provided that Sl was assigned to al during a step of semantic assignment.



Assignment propagation eliminate or reduce concept ambiguity. For instance, “cold” may be the disease “common cold” or a statement of temperature. COLD may even stand for Chronic Obstructive Lung Disease. If “cold” is assigned to the category “disease,” this ambiguity is eliminated. Thus, the semantic type of the category helps to better define the meaning of the concept [105].

So, this framework comprises semantic enrichment pre-processing and processing functions with a focus on building concept disambiguation capabilities.

The *Semantic Enrichment for Local Search Engine using Linked Open Data* framework was proposed by Alobaidi et al. [106], for enriching the web content of local search engines. The aim of this framework is to make local search indexing more effective and efficient. The framework process comprises two steps being entity segmentation and knowledge extraction.

Step 1, entity segmentation, has two modules, the Natural Language Processing module and the Name Entity Recognition module. Segmentation involves the process of tokenization and the process of identifying name entities. The processes begin with empty, raw text that is highlighted as sentences. The authors describe the sentences as mere raw material with potential value but of little semantic use. Tokenization and the recognition of names is implemented as follows. First, the Natural Language Processing module involves deep sentence analysis to produce a collection of lists of a structure tree. Then the lists are stemmed to reduce inflected words to their root. Thereafter, the lists are passed on to the name entity recognition module for disambiguation. Second, the Name Entity Recognition and disambiguation module performs the task of locating and labelling sequences of words in text and categorising these sequences into predefined classification, i.e., a name entity BIUST will be tagged and classified as university [106].

Step 2, knowledge extraction involves the use of LOD to extract knowledge from structured and unstructured data. The extracted knowledge presents the semantic implicit in that knowledge and also facilitates the function of local search engines. Knowledge extraction has two modules, the Entity URI Look-Up module and the Triples Extraction module. First, the Entity URI Look-Up module processes entity literals generated by the segmentation process. It identifies related entities from the LOD program, using properties such as `rdfs:label` (a lexical label for a resource),



skos:prefLabel (preferred lexical label for a resource) and skos:altLabel (an alternative lexical label for a resource). For instance, for the BIUST entity, this module uses the entity label to construct a SPARQL query to extract all the triples that contain the string "BIUST" as a part of its object. In turn, the module iterates through generated triples, extracts the URI subject, and builds the entity URI list that needs to be used in the next module under which extracted knowledge operate. Second, the Triple Extraction module accepts the entity URI list mentioned above from the Entity URI Look-Up Module, iterates that through the entity URI list, and construct the SPARQL query that uses each URI in the list as the subject, as well as the `rdfs:comment` (represents a description of the resource as a predicate). Then the module executes the SPARQL query against the LOD endpoint. Next, the module provides all the comments retrieved from the local search engine for consideration in the indexing process [106].

This framework processing is aimed at improving semantic search effectiveness and efficiency with its semantic enrichment pre-processing and processing functions, yet, it lacks the output evaluation to assess the impact of semantic enrichment processing on search performance.

An Ontology-Driven Framework towards Building Enterprise Semantic Information Layer was developed by Song et al. [107], to enhance semantic search. The semantic information layer refers to an information layer with semantics. This layer has interface to enable interaction with upper application layer and has access to lower data source layer. The semantic information layer represents knowledge among various systems to address the challenges of heterogeneous data stores and distributed systems on the basis of ontology extraction, enrichment and alignment [107].

Ontology extraction and enrichment are part of pre-process for the development of semantics. The first step is extraction of semantic and structure for the information stored in the knowledge sources in various formats, such as Relational Database and HTML files. The extracted ontology is regarded as *raw ontology*, which needs to be enriched and elaborated by reference ontology or manual adjustment. The raw ontology is extracted with defined rules, and then enriched with upper or reference ontology to obtain final ontology [107].



Ontology alignment is a process for ontology reuse that seeks to find semantic correspondence or reconciliation between a pair of ontology elements by identifying similarity. It involves the identification of correspondences and the setting up of connections. The *Match* process seeks the possible correspondences and *Align* process sets up the links with the constraints or criteria to reference ontologies [107].

Therefore, this framework comprises the semantic enrichment pre-processing (ontology extraction) and processing (ontology enrichment and alignment). It is limited in the output evaluation to determine the effectiveness of the enrichment processing.

Overall, the semantic enrichment frameworks described in this subsection are similar to those explored in the research background section (Section 1.2) in that they model the semantic enrichment pre-processing, processing and post-processing. Above that the semantic enrichment processing is on the basis of a vocabulary, mainly the ontology and the LOD technique. Furthermore, they support the implementation of semantic publications by enhancing OD accessibility, understandability, interoperability, integrability, and intelligibility capabilities. This has proved that these OD enrichment processes and capabilities are key to deriving value from published data and are created or enhanced through ontology and LOD-driven semantic enrichment. However, the semantic enrichment process either improves OD semantics or syntax or both, with the partial or complete exclusion of quality assessment of the resulting semantic value. This makes it difficult to establish, manage and control the degree and quality of OD value creation. Therefore these semantic enrichment frameworks are limited in addressing OD value (knowledge) creation maximisation and this is the research gap that this study addresses.

2.6 Conclusion

The OD review in Section 2.2 has revealed that knowledge resources/goods/assets are needed to drive the current global evolution to knowledge-based economies and societies. OD with its unlimited reusability, social and economic value is one of the raw resources with potential for the creation of high valued knowledge resources/goods/assets that appreciate in value with extended use. However, the big challenge is that OD is availed in raw form for public use and reuse. In its raw state, OD is



of less use and its value remains untapped in most cases. The value of data is determined by usability which depends on the formatting of data to enable its use and reuse, which in turn depends on the data structure. Therefore, the design focus for the proposed semantic enrichment framework was on developing a transformation process for the value to be unleashed and extracted for use. In OD processing, knowledge creation technologies, techniques, models and frameworks, i.e., ontologies, LOD, and semantic enrichment vocabularies, methods, models, approaches, attributes and frameworks are required to restructure OD for usable value.

A review of semantic enrichment (Section 2.5) revealed that the value of OD can be maximised to promote reuse by semantic enrichment which increases the quality and the potential use of datasets. For semantic enrichment effectiveness, it must be implemented through Semantic Web Technology combination comprising LOD and Ontologies. This, therefore, necessitates the development of frameworks that leverages on Semantic Web Technologies for semantic enrichment of OD to maximise its value and that can be achieved by building the framework around a combination of LOD and ontologies. As outlined in Section 2.5, in the LOD context, semantic enrichment pertains to the creation of links between the enriched resources and related resources in the Web of Data. Specifically, RDF links make use of vocabularies in the descriptions of the data.

The semantic enrichment necessitate that the LOD knowledge base be built on a vocabulary. The vocabulary of interest in this research is the ontology as it has the highest level of expressivity and enrichment value compared to other Knowledge Organisation Systems such as controlled vocabularies, taxonomies and thesaurus. Ontology-driven semantic enrichment pertains to the process of adding meaning to web content by mapping it to terms defined in ontologies. An extended review of ontologies (Section 2.3) revealed that they have the highest semantic enrichment value compared to all other vocabularies. A further review of existing ontologies revealed that they are limited in addressing the Botswana development agenda domain needs of the establishment of a consolidated indicator framework. To start with, there are very few studies that address semantic representation of performance indicators in general. None of these studies address performance indicators in the development agenda domain despite their critical importance in this era of OD in which countries are pushed to open up their data for



public consumption. This necessitates the development of a custom domain ontology. So, the semantic enrichment framework design included the development of a customised ontology with links to generic/reference ontologies.

A review of semantic attributes (Subsection 2.5.5) revealed that the realisation of OD value depends on the effectiveness of semantic capability features that enables OD accessibility, understandability, interoperability, integrability, and intelligibility. Therefore, semantic capabilities must be factored into semantic enrichment frameworks that leverage them for OD value creation. So, the design focus for the proposed semantic enrichment framework processing was on creating these OD capabilities. In addition, the five semantic capabilities constitute the *semantic enrichment attribute* variable for semantic enrichment assessment model. Furthermore, as semantic richness is grounded on structure and context, these two constitute *semantic richness attribute* variable for semantic richness assessment model. Moreover, from the semantic attributes discussions, it is clear that semantic mining concerns data access, linkages and knowledge extraction. So, these variables constitute the *semantic mining attribute* in the semantic mining assessment model. Moreover, as the review on semantic enrichment approaches has revealed that ontology-driven semantic enrichment plays an important role in enhancing semantic capabilities, the ontology component is the backbone of the framework.

Lastly, a review of existing semantic enrichment frameworks revealed crucial insights instrumental to the design and development of the proposed framework. As discussed in Subsection 2.2.1, OD is the future driver and as such the semantic enrichment frameworks for the conversion of raw OD resources to knowledge resources contribute significantly to the creation of valuable OD reuse potential. However, a review of existing semantic enrichment frameworks has revealed numerous research gaps on OD value enhancement through ontology and LOD-driven semantic enrichment. Firstly, the focus of ontology-driven semantic enrichment frameworks is on binding textual data to ontological constructs while that of LOD-driven semantic enrichment frameworks is on creating linkages between web resources. A review of these two types of semantic enrichment frameworks revealed a lack of both ontology and LOD-driven frameworks that could be used to address the identified publication needs for Botswana overviewed



in the research introduction. Secondly, according to literature review, no implementation or research related to either the publication of LOD or semantic representation of performance indicators applicable to Botswana has been carried out. This poses a need for semantic framework that addresses semantic enrichment and indicator consolidation needs for OD value maximisation. Thirdly, the reviewed diversities of semantic enrichment mention the impact on semantic capabilities and this points to the relevance of semantic capabilities in the creation of knowledge resources. However, they are either partially included or fully omitted in semantic enrichment frameworks that guide semantic enrichment implementations. Overall, the reviewed semantic enrichment frameworks are limited in answering the research question at hand, being:

What semantic enrichment framework could best support the implementation of semantic publication and consumption of OD with potential for OD value (knowledge) creation maximisation?

Therefore, this research proposes the semantic enrichment approach that leverages on the combination of the maximal expressiveness of ontologies and the integration capabilities of LOD to maximise on OD value (knowledge) creation potential. Specifically, the semantic enrichment transforms textual values to semantic values which are bound to ontology constructs and also linked to other values in the Web of Data. The insights from the research literature revelations were applied in the design and development, demonstration, and finalisation of the semantic enrichment framework (Chapters Five to Eight).



CHAPTER THREE: RESEARCH METHODOLOGY

3.1 Introduction

The design science approach, due to its focus on creations that address human needs, was preferred for this research which addresses human needs, semantic value creation needs. Design science specialises in the development and evaluation of Information Technology artifacts intended to solve identified organisational problems [108]. An artifact is any designed object which embeds a research contribution. This research adapted the *Design Science Research Methodology for Information Systems Research* proposed by Peffers, Tuunanen, and Rothenberger [108] for the suitability of its detailed step-by-step design and development orientation. This specific methodology involves the design and development of Information Systems artifacts for various purposes, including the attempt to: get a better understanding and/or improvement of the functioning of some Information Systems features, solve observed problems, make research contributions, evaluate the designs, and communicate the results to appropriate audiences [108]. Its detailed description and adaptation are discussed below.

3.2 Description of the Design Science Research Methodology for Information Systems Research

The Design Science Research methodology for Information Systems research comprises six activities implemented in the following order [108]:

- Activity 1 - Problem identification and motivation,
- Activity 2 - Definition of objectives for a solution,
- Activity 3 - Design and development,
- Activity 4 - Demonstration,
- Activity 5 - Evaluation, and
- Activity 6 - Communication.



3.2.1 Activity 1: Problem Identification and Motivation

Activity 1, *problem identification and motivation* involve the definition of a specific research problem and justification of the value of the proposed solution. The problem definition serves as the basis for the development of the artifact. For effective conceptualisation of the problem, it is broken into smaller pieces to facilitate an in-depth analysis from different dimensions for a balanced and holistic solution. Moreover, justifying the value of a solution motivates the researcher and the research audience to pursue the solution and to accept the results. It also helps to appreciate the researcher's understanding of the problem. Resources required for Activity 1 include knowledge of the problem and the importance of its solution [108].

3.2.2 Activity 2: Definition of Objectives for a Solution

Activity 2, *definition of objectives for a solution* involves the formulation of solution objectives based on the problem definition and the knowledge of what is possible and feasible. Thus, solution objectives should be inferred rationally from the problem specification. Objectives can either be quantitative (for a quantitative research) or qualitative (for a qualitative research). Quantitative objectives provide specific terms that show how the proposed solution would be better than existing ones while qualitative objectives provide specific description of how the proposed artifact is expected to support solutions to problems not addressed. Resources required for Activity 2 include knowledge of the problem state and existing solutions, if any, and their efficacy [108].

3.2.3 Activity 3: Design and Development

Activity 3, *design and development* pertains to the determination of the functionality and architecture of the new artifact, followed by the creation of the actual artifact. A design artifact maybe in the form of constructs, models, methods, instantiations, and new properties of technical, social, and/or informational resources. Resources required for Activity 3 include knowledge of theory to serve as the basis for solution design and development [108].



3.2.4 Activity 4: Demonstration

Activity 4, *demonstration* involves the application of the artifact to solve one or more instances of the problem [108]. In the design science paradigm, the building and application of the designed artifact is paramount to acquiring knowledge and understanding of a problem domain and its solution [109]. The demonstration of the artifact could involve its use in experimentation, simulation, case study, proof, or other appropriate activity. Resources required for Activity 4 include effective knowledge of how to use the artifact to solve the problem [108].

3.2.5 Activity 5: Evaluation

Activity 5, *evaluation* involves the observation and measurement of how well the artifact supports a solution to the problem [108]. According to Venable, Pries-Heje and Baskerville [110], evaluation is what constitutes the *Science* in *Design Science*, without which results unsubstantiated design theory or hypothesis that some developed artifact will be useful for solving some problem or making some improvement. There are numerous aims for evaluation of either a design artifact or its instantiation or the formalised knowledge about it, including [110]:

- to establish its utility and efficacy for achieving its stated purpose;
- to compare it to other designed artifacts' ability to achieve a similar purpose;
- to explore the side effects or undesirable consequences of its use; and
- to identify its weaknesses and areas of improvement.

This activity involves comparing the objectives of a solution to actual observed results from use of the artifact in the demonstration. The evaluation pertains to the production of any appropriate empirical evidence or logical proof and could take many forms depending on the nature of the problem and the artifact. It could involve comparison of the artifact's functionality with the solution objectives from activity two above, objective quantitative performance measures, such as budgets or items produced, the results of satisfaction surveys, client feedback, or simulations. It could include quantifiable measures of system performance, such as response time or availability. At the end of this activity, the research nature determines the feasibility of iterating back to step three to improve the effectiveness of the artifact or to proceed to communication and leave



further improvement to subsequent projects. Activity 5 requires knowledge of relevant metrics and analysis techniques [108].

3.2.6 Activity 6: Communication

Activity 6, *communication* involves sharing the problem and its importance, the artifact, its utility and novelty, the rigor of its design, and its effectiveness to researchers and other relevant audiences. Activity 6 requires knowledge of the disciplinary culture [108].

3.3 The Adapted Design Science Research Methodology

The artifact produced in this research is a framework. The nature of this framework is best described by Mattsson and Bosch as a [111]:

Reusable design of a system comprising a set of abstract classes of components and their relationships, representing a foundational structure for the development of systems and applications.

To align with the qualitative nature of the research, the Design Science Research methodology was adapted for the design and development of the framework. For the definition of the knowledge scope, the genre of inquiry of Design science knowledge approach adopted from the Genre of Inquiry Framework is the *Idiographic Design (ID) Knowledge* which pertains to application of design knowledge and methods to produce an ideal artifact for a specific purpose [112]. This involves the generation of knowledge necessary for the research and development of an individual product and the artifact to be developed serves to embody or materialise this knowledge. The adapted methodology provides specific description of how and why the proposed artifact was developed and demonstrates the degree it is expected to support solutions to the semantic publication and consumption needs at hand. The research adopted a pragmatic paradigm involving the use of mixed methods in the practical exploration of semantic enrichment framework with potential for OD value (knowledge) creation maximisation. So, the adapted Design Science Research methodology was customised to the design and development needs of this research in numerous ways discussed below.

The first adjustment affected all framework activities. The activity categorisation of the research work was replaced with the phase categorisation. This adjustment was made to differentiate between the main research activities and research implementation activities



in research presentations. The research work spanned different incremental phases comprising the main research activities that represent the research objectives outlined in Section 1.5. Each phase involved a series of sub-activities implementing the main research activity (research implementation activities).

The other adjustments are specific to individual activities. The first activity pertaining to *problem identification and motivation* was customised to the framework scope to give the research work more focus. So, Phase 1, *framework needs identification* involved the identification of semantic publication and consumption needs for the preliminary semantic enrichment framework design. The needs that were identified in the development agenda domain embraced problem identification and motivation. The second activity pertaining to the *definition of objectives for the solution* was included in phase 1. The third activity pertaining to *design and development* of the artifact was adjusted to *framework design and development* and constitutes phase 2 of this research. The fourth activity pertaining to the *demonstration* of the artifact was presented as *framework demonstration* and constitutes phase 3 of this research. The fifth activity pertaining to the *evaluation* of the artifact was customised to *framework evaluation* and partly constitutes phase 4 of this research. To address the nature of this research which necessitates the preliminary framework to be refined for reusability in other related domains, phase 5 was added to include post-evaluation framework processing activities performed on the basis of evaluation results, *framework finalisation* activities. Framework finalisation constitutes phase 5 of this research. Lastly, the sixth activity pertaining to the *communication* of the artifact was replaced with what matters most for a research toward PhD qualification, the research contribution. Thus, the *communication* activity was replaced with the presentation of the main research contribution, the final research artifact, together with framework implementation support tools. Therefore, the adapted Design Science Research methodology for this research is as follows:

- Phase 1 - Framework needs identification,
- Phase 2 - Framework design and development,
- Phase 3 - Framework demonstration,
- Phase 4 - Framework evaluation, and
- Phase 5 - Framework finalisation.



Following is a detailed discussion on what is entailed in each methodology phase.

3.3.1 Phase 1: Framework Needs Identification

Phase 1, *framework needs identification* pertained to the identification of framework needs for the preliminary semantic enrichment framework design. This involved specification of framework scope and functionality modelling elements. The needs were identified through flat text resource reviews, i.e., pdf, Word, excel. All reviews were according to the *content analysis* method (RM1) which involved the following activities:

- The search for certain terms, concepts, keywords and themes in flat text resources, followed by,
- The gathering of relevant domain knowledge related to the topic at hand, and finally followed,
- The inference and extraction of framework needs, specifically, semantic publication and consumption needs.

On the one hand, the aim of content analysis for semantic consumption needs was to establish the OD support needs for the implementation of Botswana's national agendas. So, the review was conducted to identify OD support needs for the implementation of the Vision Agenda and the NDP Agenda from Vision 2036 (Pillar 1) and NDP 11 documentations [6], [7]. On the other hand, the aim of content analysis for semantic publication needs was to establish the future OD program, OD presentation and representation needs. First, the Botswana OD Readiness Assessment Survey Report which documents the road map to future OD publications was reviewed for Botswana's OD program goals [14]. Second, the review of OD presentation needs pertained to an assessment of the current OD publication system ('as-is' system) in terms of OD presentation structure and format. The review was based on Statistics Botswana's data portal and other metadata documentations for the agendas that constitute the Botswana Integrated Indicator Framework being Vision Agenda, NDP Agenda, Africa Agenda, and the global and domesticated SDG Agenda documentations [5], [6], [8], [9], [10], [113]. Third, the OD representation needs were identified from all the documentations above and the SDGs National Baseline Indicator Framework proposal documentation [13]. Additional OD representation needs were identified from a diversity of published



literature on performance indicator semantic representations. Phase 1 implementation is documented in Chapter Four.

3.3.2 Phase 2: Framework Design and Development

Phase 2, *framework design and development* pertained to the design and development of architectural structure and functionality for the conceptual model on the basis of the preliminary semantic enrichment framework design needs identified in phase 1. The generic conceptual framework development method by the Northcentral University [114] was adapted for the design and development of the preliminary semantic enrichment framework. The generic conceptual framework development method is as follows:

- Step 1 - Review of the body of literature to identify key concepts used by other studies, to determine what has been done by other researchers, and to determine what needs to be done;
- Step 2 - Identification of framework variables, concepts, theories, and existing frameworks explained in the relevant literature;
- Step 3 - Conceptualisation of how the identified variables, concepts, theories, and facets of existing frameworks integrate to shape the problem; and
- Step 4 - Creation of a graphic representation of the framework.

This method was adapted for the development of the preliminary semantic enrichment framework and applied as follows (RM2):

- Step 1 - Review of research literature on semantic enrichment entities comprising processes, approaches, capabilities, techniques, technologies and their relations;
- Step 2 - Identification of semantic enrichment framework entities, elements and their relations; and
- Step 3 - Generation of the conceptual semantic enrichment framework which involved the grouping of identified entities into components based on identified relationships and finally, the development of the framework from a combination of these components.



The preliminary semantic enrichment framework was contextualised to the semantic publication and consumption needs for the Botswana development agenda domain. Phase 2 implementation is documented in Chapter Five.

3.3.3 Phase 3: Framework Demonstration

Phase 3, *framework demonstration* involved the discovery of specifications on the semantic value creation approach, process and semantic valuation elements to be factored into the final framework design. The implementation of the framework components was on the basis of the identified potential framework entities. The main component of this inductive qualitative study consisted of the artifact construction to serve as proof of concept, in the form of a one cycle design [115]. This design comprised a composite artefact with two main components, the semantic enrichment approach component (Demonstration 1 – documented in Chapter Six) and the semantic enrichment implementation component (Demonstration 2 – documented in Chapter Seven) founded on the architectural design pillars of semantic value creation processing, semantic value creation maximisation approach and semantic valuation. Demonstration 1 made use of two methodologies, one involving the choice of generic ontologies and that which involved the development of a custom domain ontology.

Methodology for the Choice of Generic Ontologies (Demonstration 1)

The first demonstration pertained to the framework base component, the semantic enrichment approach component which involved the choice of *generic* ontologies and the development of a custom domain ontology. The selection of *generic ontologies* involved the identification of potential ontology targets to be used in the enrichment of semantic capabilities. In doing so, this research adapted the *Enrichment and Evaluation* methodology proposed by the Europeana Task Force [116].

The *Enrichment and Evaluation* methodology comprises the following steps [116]:

- *Analysis of source data* - gathering knowledge on issues regarding topic coverage, gaps, and data quality;
- *Identification of the enrichment requirements* - definition of expected results;
- *Finding datasets available on the Web*;



- *Selection of the enrichment targets* - choice of applicable targets on the basis of the following criteria:
 - *Availability and Access* - selection of targets on the basis of Web availability and compliancy with the Linked Data Standard of publication which requires that the targets be reusable under an open licence,
 - *Granularity and Coverage* - selection of targets that have the same coverage as that of the source data or that can complement the source data,
 - *Quality* - consideration of the quality of the target in terms of semantic and data modelling features,
 - *Connectivity* - selection of targets with incoming and outgoing links to other targets, and
 - *Size* - consideration of the *size* of potential targets.
- *Testing the selected target on a sample of source data* - verifying the target scope and quality to ensure that it covers the source data and does not introduce semantic ambiguities.

The *Enrichment and Evaluation* methodology is generic in nature and so it was customised to the chosen type of vocabulary for semantic enrichment, ontology. The adapted *Enrichment and Evaluation* methodology (RM3) was applied in four phases as follows:

- Phase 1: Analysis of source data and metadata resources for ontology search literals (constructs);
- Phase 2: Identification of ontology enrichment requirements;
- Phase 3: Search for potential ontology targets; and
- Phase 4: Selection of suitable ontology targets – the selection criteria was mainly based on the following four dimensions:
 - *Availability* - selection of suitable ontology targets on the basis of the following: 1) Web availability which involves the testing of potential ontology targets to ensure that they are technically available on the Web; 2) consideration of potential ontology targets' compliancy to the Linked Data best practices; and 3) consideration of the availability of proper ontology documentations,



- *Coverage* - selection of suitable ontology targets based on the precision of meaning and the degree of expressivity of the ontology properties,
- *Quality* - selection of suitable ontology targets based on the structure and representation of potential ontology targets. Selected ontologies must be self-descriptive and each property or concept must have a label, a definition and a comment, and
- *Value representation* - selection of suitable ontology targets on the basis of the more preferred semantic value creation variables – semantic structure (syntax) and semantic meaning (semantics).

Methodology for the Development of Custom Domain Ontology (Demonstration 1)

The second ontology implementation involved the development of a custom domain ontology (NPIOnto) according to the adopted METHONTOLOGY approach. The METHONTOLOGY approach has six phases pertaining to the specification, knowledge acquisition, conceptualisation, integration, implementation, and evaluation of the ontology as originally proposed by Fernández-López, Gómez-Pérez and Juristo [117]. The *specification phase* involves the production of either an informal, semi-formal or formal ontology specification document written in natural language, using a set of intermediate representations or using competency questions, respectively. METHONTOLOGY specification includes: a) the purpose of the ontology outlining its intended uses, scenarios of use, and end users; b) the level of formality of the implemented ontology; and c) the scope of the ontology defined by the set of terms to be represented, their characteristics and granularity [117]. The *knowledge acquisition phase* involves knowledge elucidation from different knowledge sources such as experts, books, handbooks, figures, tables and other ontologies. Knowledge is gathered through techniques such as brainstorming, interviews, formal and informal analysis of texts, and knowledge acquisition tools. The final product is a list of terms and their meaning [117]. The *conceptualisation phase* involves the structuring of domain knowledge into a conceptual model that describes the problem and its solution in terms of the domain vocabulary identified in the ontology specification activity. The first activity involves the building of a glossary of terms representing a collection of domain knowledge and its meanings i.e., concepts, instances, verbs and properties. That is followed by the grouping



of terms into closely related sets of concepts and verbs. For each set of related concepts and related verbs, a concepts classification tree and a verbs diagram is built [117]. The *integration phase* involves the identification of definitions already built into other ontologies for reuse. This is done by first, carrying out an inspection of existing ontologies to select those that better fit the conceptualisation needs at hand. This is to ensure that the sets of new and reused definitions are based upon the same set of basic terms. If existing ontologies are not appropriate for the ontology needs at hand, that necessitates the definition and implementation of a new meta-ontology in a formal language [117]. The *implementation phase* involves the choice of an environment that supports the custom ontology development and the integration of ontologies selected at the integration phase. The result of this phase is the ontology codified in a formal language [117]. Lastly, the *evaluation phase* involves the technical judgment of the ontologies. The evaluation term subsumes the terms verification and validation. Verification refers to the technical process that guarantees the correctness of an ontology while validation assesses the ontologies for correspondence to systems they represent [117].

The application of the METHONTOLOGY approach in this research was according to all the phases described above, except for one, the integration phase. This phase was removed as it is already covered in the demonstration 1 methodology outlined above which pertains to the choice of generic ontologies. In addition, the implementation phase was renamed to formalisation phase to be more specific on the type of implementation. Following, is the detailed description of the application of the METHONTOLOGY approach (RM4) in this research.

The *specification phase* involved the statement of the ontology purpose in terms of ontology goal, scope and intended use. The statement of the ontology goal pertained to the specification of the aims and objectives of the ontology. The statement of the ontology scope pertained to the definition of the ontology knowledge domain and the specification of the boundaries for domain coverage. The intended use of the ontology was described in terms of its potential applications. All these were derived mainly from the OD representation (ontology) needs identified in Phase 1. The implementation of this phase is documented in Subsection 6.3.1.



The *knowledge acquisition phase* pertained to the analysis of data sources gathered in Phase 1 for ontology concepts, attributes and relationships. This phase involved the collection and understanding of information from a wide range of sources central to the ontology domain. This phase served as a checkpoint to assess the goal and scope definition and to refine them. This phase was conducted mainly through content analysis method which involved the following four steps:

- Step 1: Identification of statements of interest from domain documentations,
- Step 2: Analysis of the statements to derive concepts and their relationships,
- Step 3: Identification of modelling specification for the concepts and their relationships from the metadata resources, and
- Step 4: Creation of flat text resources based on concepts per agenda performance variable.

The output of this phase comprised collections of domain knowledge on NPIs in flat text format (spreadsheets). The implementation of this phase is documented in Subsection 6.3.2.

The *conceptualisation phase* pertained to the initial structuring of the ontology on the basis of the information collected in the *specification* and *knowledge acquisition* phases. This phase focused on the conceptual definition of the ontology and involved the preliminary modelling of the ontology, specifically, the design and development of the lightweight ontology. That is, collections of concepts, their attributes and relation definitions were converted to visual knowledge model representation. Therefore, the output of the conceptual phase comprised the NPIOnto scope model and the lightweight version of the knowledge model. The implementation of this phase is documented in Subsection 6.3.3.

The *formalisation phase* involved the codification of both the consolidated knowledge model (schema) and the knowledge base (individuals or specific entities in the universe of discourse). That is, the light weight model was extended into heavy weight model. The formalised knowledge model comprises classes and relating object and data properties, individuals and axioms. This phase is documented in Subsection 6.3.4.

The *evaluation phase* involved the assessment of the effectiveness of ontology in two ways. The first evaluation, the design effectiveness evaluation was conducted to assess



the effectiveness of the ontology design with the use of metrics that evaluated both the knowledge model (schema) and the knowledge base (instances) designs (Subsection 6.3.5). The second evaluation, the functionality effectiveness evaluation was conducted to assess the performance effectiveness of the ontology when deployed in its intended settings (Subsection 7.4.2).

Methodology for Semantic Enrichment (Demonstration 2)

The second demonstration was according to the preliminary ODSE Framework (Chapter 5), pertaining to the ODSE pre-processing, processing and post-processing based on the ODSE approaches chosen for the demonstration (RM5). The ODSE pre-processing stage pertained to the gathering and transformation of input data. That was followed by the ODSE processing stage which involved the ontology and data enrichment processing. Last, was the ODSE post-processing which pertained to the output integration and deployment for mining. The knowledge mining (harvesting or extraction) approach involved comparisons of the effectiveness of the semantic value creation potential given a specific ODSE approach. The harvesting of semantic value from knowledge resources – knowledge mining was conducted through queries issued and evaluated over SPARQL endpoint. The mining method adopted the *black box testing* approach which involved the formulation of logical deductions as proof of the impact of semantic enrichment based on the mining output for a given mining case. The purpose of the evaluation of knowledge resources was to establish the semantic capabilities value-added to OD through semantic enrichment – semantic enrichment effectiveness. Specifically, the evaluation involved a comparison of the generic and the domain knowledge resources to determine the ODSE approach with maximal semantic capabilities value-adding and knowledge mining potential, demonstrated by the knowledge resource's degree of accessibility, understandability, interoperability, integrability and intelligibility capabilities.



3.3.4 Phase 4: Framework Evaluation

Phase 4, *framework evaluation* pertained to *framework verification* (implementation documented in Section 8.2) and *framework validation* (implementation documented in Section 8.3).

Framework verification, in this research context, pertains to the manual reviews of the framework through reconciliation of preliminary semantic enrichment framework elements and processes to the actual implemented elements and processes in-order to assess the level of framework completeness and cohesiveness. It was conducted according to the *ODSE Framework verification method* (RM6) according to which reviews of the completeness of the preliminary framework process elements involve the manual assessment of the degree of inclusion of all semantic enrichment entities comprising processes, approaches, capabilities, techniques, technologies and their relations for each framework stage. Reviews of the cohesiveness of framework elements involve manual assessment of processes for the degree of connectivity and flow of both the individual stage elements and the integrated elements per process for all stages.

Framework validation, in this research context, pertains to the assessment of the effectiveness of the semantic enrichment framework, *semantic value assessment*. As this research is of a qualitative nature, framework validation uses qualitative measurements, specifically, qualitative rating schemes based on knowledge resources' demonstration of certain traits/attributes and behaviour. That is, performance measurements are expressed in relation to use impact, specifically, they are derived from the logical analysis and deductions of the processes and data attributes and behaviour under given test case dynamics. This evaluation approach is based on the current qualitative approach where OD value is qualitatively derived from use and reuse impact - the *highly valuable, unvalued asset* challenge (discussed in Subsection 2.2.2) due to the difficulty or impossibility of quantifying data value through formal methodologies.

The *ODSE Framework validation method* (RM7) was developed to guide the validation of the ODSE framework. This method adopts a holistic approach to semantic value assessment that pertains to the integrated evaluation of the ODSE Framework as a single unit. The assumption for the integrated stages evaluation is that the effectiveness of the entire ODSE Framework takes into consideration all processing for each stage and per



semantic value creation approach. As illustrated in Figure 3.1, the ODSE Framework validation method embraces the entire ODSE Framework and involves a comparison of the semantic capabilities for the ODSE Framework input (Stage 1), processing (Stage 2) and output (Stage 3).

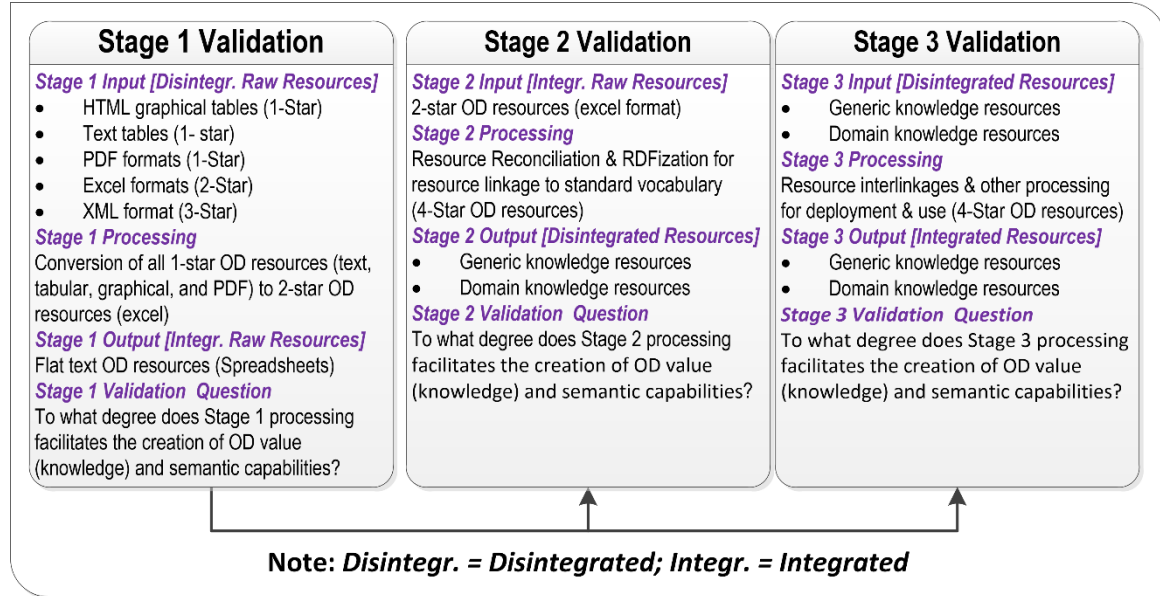


Figure 3. 1: Summary of ODSE Framework Validation Method

Figure 3.1, illustrates the holistic approach to semantic value assessment that pertains to the integrated evaluation of the ODSE Framework input, processing and output stages as a single unit. The assumption for the integrated stages evaluation is that the effectiveness of the entire ODSE Framework takes into consideration all processing for each stage per semantic enrichment approach. The semantic enrichment validation question per ODSE stage reflected in Figure 3.1, seeks to establish how the framework implementation output for each framework stage fulfils the ODSE Framework intent. The ODSE Framework intent is to create knowledge and semantic capabilities through the addition of syntax and semantics to augment flat text OD resources with metadata and to improve OD accessibility (structure, classification, organisation and presentation to enable search, query and discovery), understandability (disambiguation and clarity), interoperability (linkage compatibility), integrability (connectivity), and intelligibility (analysis and comparability - comprehensiveness of data for analysis). Therefore, the validations assesses the input and output of each ODSE stage for the formation of these five capabilities.



There are two alternative tools that this research devised for the assessment of effectiveness of each individual ODSE Framework stage processing. The first proposed candidate tool for ODSE Framework validation is the *semantic capabilities assessment model* (Table 3.1) provides a mechanism for assessment of the degree of OD/semantic publication and consumption effectiveness in terms of semantic capabilities (accessibility, understandability, interoperability, integrability and intelligibility) value-added to OD resources at each individual level of semantic value creation implementation. This semantic value assessment tool is a qualitative matrix in the form of a 5-point rating scheme of OD publication and consumption effectiveness, derived from a combination of semantic capabilities and Tim Berners-Lee's 5-Star deployment scheme for Linked Data illustrated by Figure 2.11 in Section 2.4. The model is shown in Table 3.1.

Table 3. 1: Semantic Capabilities Assessment Model

		Semantic Capabilities Accessibility, Understandability, Integrability, Interoperability & Intelligibility	
Open Data 5-Star Publication Rating	Open Data Publication Features	Open Data Publication Effectiveness Rating	Open Data Consumption Effectiveness Rating
1-Star (e.g., pdf)	Open License	Lowest	Lowest
2-Star (e.g., Excel)	Open License Machine Readable	Low	Low
3-Star (e.g., CSV, XML)	Open License Machine Readable Open Format	Moderate	Moderate
4-Star (URI link)	Open License Machine Readable Open Format Uniform Resource Identifier	High	High
5-Star (Linked)	Open License Machine Readable Open Format Uniform Resource Identifier Linked Data	Highest	Highest

The use of variables in Table 3.1 in the context of this research are described as follows:

- *Semantic capabilities* refer to attributes added to raw flat text OD through structural and contextual modifications to enhance some specific functional/operational qualities, demonstrated by the resource's degree of *accessibility, understandability, interoperability, integrability* and *intelligibility* capabilities,
- *Open Data 5-Star Publication Rating* is the 5-Star deployment scheme for Linked Data (detailed in Section 2.4),



- *Open Data Publication Features* are publication attributes that determines OD publication ratings,
- *Open Data Publication Effectiveness Rating* is defined in terms of the 5-Star deployment scheme for Linked Data,
 - Rates for different degrees of minimum level – lowest and low,
 - Rates for different degrees of maximum level – high and highest, and
- *Open Data Consumption Effectiveness Rating* refers to the knowledge creation, mining and use rating and is defined in terms of the *Open Data Publication Effectiveness Rating*. The rating scheme assumes that publication effectiveness has direct impact on consumption effectiveness – the *direct impact assumption*. On this premise, the publication effectiveness rates also represent consumption effectiveness rates.

As reflected in Table 3.1, the 5-Star publication scheme (column 1) described by publication features in column 2 is assigned qualitative rates for publication effectiveness (column 3) that also represent consumption effectiveness (column 4).

The second proposed candidate tool for ODSE Framework validation is the *semantic usability assessment* model, an extension of the *semantic capabilities assessment model* (Table 3.1). It adds the richness and mining value attributes to the enrichment value attributes for the *semantic capabilities assessment model*. Thus, the *semantic usability assessment* model provides the mechanism for assessment of semantic *value-adding* and *value-added* potential to OD resources at each level of semantic value creation implementation in terms of the enrichment, richness and mining value attributes. It is a qualitative matrix in the form of a 5-point rating scheme of OD publication and consumption effectiveness, derived from a combination of semantic usability attributes and Tim Berners-Lee's 5-Star deployment scheme for Linked Data.



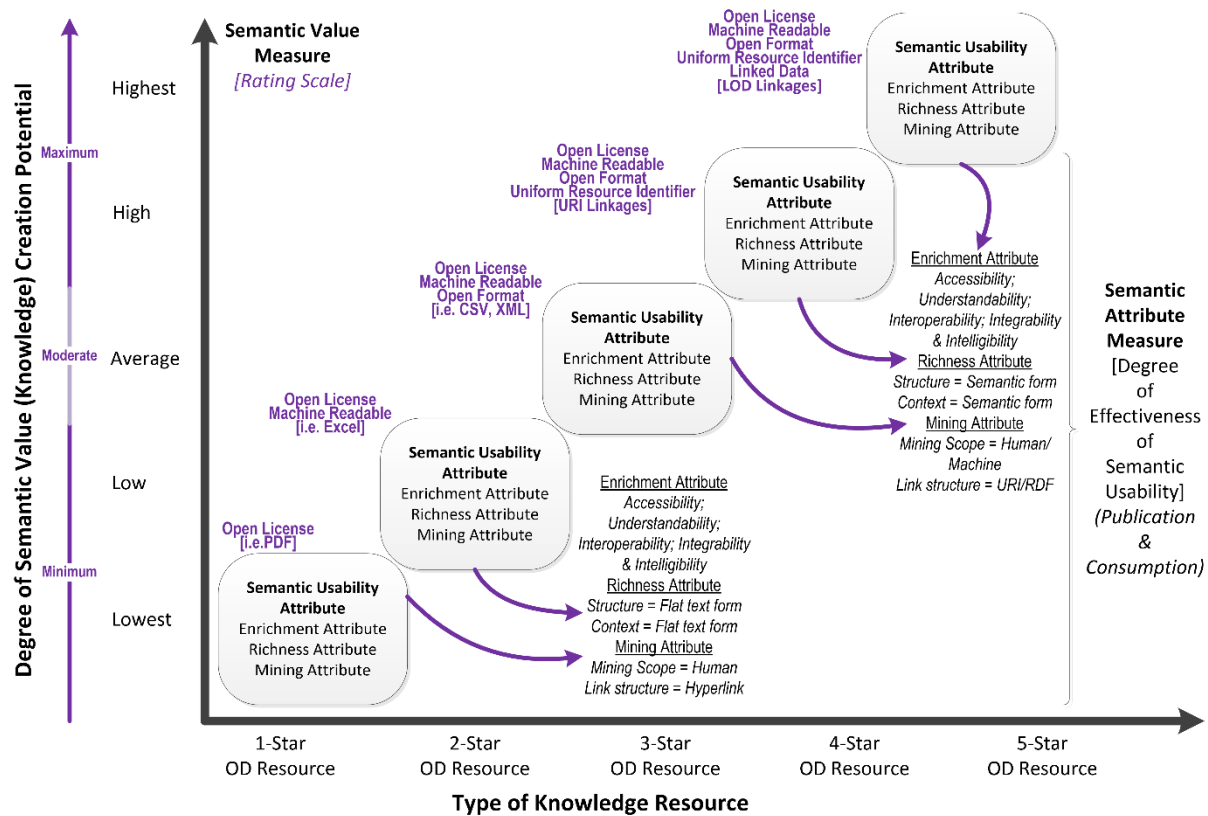


Figure 3.2: Semantic Usability Assessment Model

Overall, as depicted in Figure 3.2, this model is for the assessment of the degree of semantic usability impact on semantic value (knowledge) creation potential and usability effectiveness of a given OD resource in terms of the resources' enrichment, richness and mining attributes. The *semantic value* variable is a qualitative measurement value of OD publication rating defined in terms of the 5-Star deployment scheme for Linked Data comprising rates for different degrees of minimum level (lowest and low), moderate level (average) and maximum level (high and highest) rating scale. This rating scheme assumes that the publication rating has a direct impact on the consumption rating – the *direct impact assumption*. On this premise, the Linked Data publication rates also represent the Linked Data consumption rates. The semantic value variable is presented in relation to the independent measurement variable which pertains to the different types of OD publications, *type of knowledge resource*, expressed in terms of Tim Berners-Lee's 5-Star deployment scheme for Linked Data (1-, 2-, 3-, 4- & 5-star OD publications). The 5-Star ranking is on the basis of the resource fulfilling the semantic publication standard features of being published in the Web of Data: 1) having an open license, 2) being



machine readable, 3) being presented in an open format, 4) having URI linkages, and 5) having Linked Data linkages. The semantic value variable is also presented in relation to the dependent measurement variable that pertains to the *degree of semantic value (knowledge) creation potential*. Moreover, the semantic value variable is presented in relation to the dependent measurement variable that pertains to the *degree of effectiveness of semantic usability (semantic attribute measure)*. The semantic attribute measure comprises publication attributes that determine OD publication ratings and are assigned qualitative rates for publication effectiveness and constitutes the *semantic usability attribute* variable. The *semantic usability attribute* variable comprises the semantic enrichment, richness and mining attributes, the basis for assessment of the semantic measurement values for both the independent and dependent variables. The semantic usability attribute variable is the enabler of OD publication and consumption, thus, the determinant of semantic value variable. The enrichment attribute variable is expressed in the form of semantic capabilities, the functional/operational qualities added to raw flat text OD through structural and contextual modifications, demonstrated by the resource's degree of accessibility, understandability, interoperability, integrability and intelligibility capabilities. The richness attribute variable is expressed in terms of structure and context and the mining attribute variable in terms of mining scope and link structure form.

Drawn from this graphical model are five potential applications. First is the knowledge resource grading/rating application drawn from the relation between the semantic attribute variable and the knowledge resource variable. This application pertains to the assignment of rates to semantic usability attributes for a given type of knowledge resource. Second is the resource selection application drawn from the relation between the knowledge resource, semantic usability attribute and the knowledge creation dependent variables. The application pertains to the selection of knowledge resources with appropriate semantic usability attributes for a given target of semantic value (knowledge) creation rating. Third is the resource selection application drawn from the relations between the knowledge resource, semantic usability attribute and the semantic attribute measure variables. This application pertains to the selection of knowledge resources with appropriate semantic usability attributes for a given target of semantic



usability effectiveness rating. Fourth is the resource evaluation application drawn from the relation between the semantic attribute measure variable and the knowledge resource variable. The application pertains to the evaluation of the degree of effectiveness of semantic usability (both publication and consumption effectiveness) for a given type of knowledge resource. Last, is the resource evaluation application drawn from the relation between the knowledge creation dependent variable and the knowledge resource variable. This application involves the evaluation of the degree of semantic value (knowledge) creation potential for a given type of knowledge resource. For all potential semantic usability applications, semantic values are expressed in terms of the minimum, moderate and maximum rating scale.

3.3.5 Phase 5: Framework Finalisation

Phase 5, *framework finalisation* pertained to the framework redesign for application in similar scenarios. This involved the conformation of the preliminary ODSE Framework design for generic application in the development agenda domain (implementation documented in Section 8.4). So, from the three basic forms of reasoning underlying the adapted Design Science Research methodology, being, deduction, induction and abduction as proposed by [118], this research adopted the induction form. This is the bottom-up approach to reasoning that involves making general inferences or premises or conclusions from specific observations or applications or demonstrations. So, the final ODSE Framework was generalised on the basis of the Botswana case study demonstration and framework evaluation findings. This involved the preliminary ODSE Framework redesign into a generic model on the basis of the framework demonstration and evaluation (verification and validation) findings.

In addition, the ODSE Framework was augmented with seven models developed to serve as the ODSE Framework implementation support tools, comprising, three reference models and four assessment models. On the one hand, the domain scoping models were developed according to the generic conceptual framework development method described in Subsection 3.3.2. On the other hand, a pragmatic paradigm was adopted in the development of semantic valuation models to address the current lack of semantic value assessment tools for semantically enriched OD resources, knowledge resources. The proposed semantic valuation models were deduced from the semantic value creation



demonstration which was targeted at the three dimensions to semantic value creation represented by the three sub-questions above. The models were developed according to *Semantic Valuation Tools Development Method* (RM8), implemented in three steps as follows. The first step (**Step 1**) involved the semantic value creation demonstration. The demonstrations adopted the *Ontology-Driven Semantic Enrichment Framework* (ODSE Framework) [119] which embodies the semantic value creation solution with a provision for semantic usability assessment. On the one hand, the ODSE Framework was applied in the process of semantic value creation to build semantic structure and context in flat text resources so as to create knowledge (semantic product) and semantic capabilities (syntax product). Specifically, it was applied in the conversion of flat text resources published in formats with limited semantic capabilities (i.e., pdf, HTML, image and spreadsheet) into knowledge resources with semantically enhanced capabilities (rich semantic formats) to improve semantic publication and consumption effectiveness. On the other hand, the semantic usability assessment model (Subsection 3.3.4 – Figure 3.2) was applied in the evaluation of semantic publication and consumption effectiveness. The second step (**Step 2**) involved the identification of semantic value creation variables, attributes, relationships and capabilities from the knowledge mining findings (Step 1) to serve as the basis for the design of the models. That was followed by the third step (**Step 3**) which pertained to the actual design of the models, involving the graphical representations of semantic value creation variables, attributes, capabilities and relationships in the form of semantic value dependent and independent variables, as well as, semantic attribute measure variables. The detailed discussions of these steps are as follows.

Step 1: Semantic Value Creation Demonstration

Semantic value creation implementation was conducted to build semantic capabilities and augment knowledge in the resulting ontology enriched OD resources (knowledge resources), in the form of contextual and structural additions to flat text resources. Knowledge resources were produced according to the ODSE Framework described in the method section. That was followed by the exploration of semantic enrichment output (knowledge resources) through two test case scenarios with the use of the *semantic usability assessment* model. Finally, insights from the semantic value creation



demonstration findings were drawn out to serve as the basis for the development of the proposed semantic valuation models. Following is an overview of the entire processing.

Step 2: Identification of Semantic Value Creation Variables, Attributes, Relationships and Capabilities

From semantic value creation demonstration findings, semantic value creation variables, attributes, relationships and capabilities were deduced from three dimensions – semantic enrichment, richness and mining dimensions.

Semantic Enrichment Variables, Attributes, Relationships and Capabilities

Semantic value creation variables for the semantic enrichment assessment model were identified as follows. First, the *semantic value* variable is a qualitative measurement value for semantic enrichment comprising the minimum, moderate and maximum rating scales. Second, the semantic value independent variable, the variable relating to the method of enrichment is the *type of semantic enrichment approach* which comprises three options being, no semantic enrichment, generic ODSE approach and domain ODSE approach. Third, the dependent variables (publication and consumption measure) pertain to the *degree of semantic value (knowledge) creation potential* and to the *degree of semantic enrichment effectiveness* (semantic attribute measure). Fourth, the *semantic enrichment attribute* variable (semantic value determinant) pertains to semantic capabilities being accessibility, understandability, interoperability, integrability and intelligibility.

The semantic enrichment relations are as follows. The first relation is between the independent variable (type of semantic enrichment approach) and semantic attribute variable (semantic value determinant variable). The type of semantic enrichment approach determines the degree of formation of semantic enrichment attributes. The second relation is between the type of knowledge resource and semantic attribute measure variable. The knowledge resources differ according to the degree of effectiveness of semantic enrichment. The third relation is between semantic attribute variable and semantic attribute effectiveness measure variable. The level of semantic enrichment attributes determines the degree of effectiveness of semantic enrichment. That is, the minimum, moderate and maximum degrees of semantic capabilities is equal the minimum, moderate and maximum degrees of effectiveness of semantic enrichment.



The fourth relation is between the two dependent variables. The degree of effectiveness of semantic enrichment has a direct impact on the degree of semantic value (knowledge) creation potential. The fifth relation is between the independent variable and the two dependent variables. The type of semantic enrichment approach determines the degree of effectiveness of semantic enrichment and the degree of semantic value (knowledge) creation potential. All relations are expressed in terms of semantic richness attributes comprising semantic capabilities (accessibility, understandability, interoperability, integrability and intelligibility)

Semantic Richness Variables, Attributes, Relationships and Capabilities

Semantic value creation variables for the semantic richness assessment model were identified as follows. First, the *semantic value* variable is a qualitative measurement value for semantic richness comprising the minimum, moderate and maximum rating scales. Second, the independent variable, the variable relating to OD publication is *the type of knowledge resource* comprising raw Open Data resource, generic knowledge resource and domain knowledge resource. Third, the dependent variables (data publication and consumption measure) pertain to the *degree of semantic value (knowledge) creation potential* and the *degree of effectiveness of semantic richness* (semantic attribute measure). Fourth, the *semantic richness attribute* variable (semantic value determinant) is expressed in terms of structure (flat text vs semantic form) and context (flat text vs semantic form).

The semantic richness relations are as follows. The first relation is between independent variable (type of knowledge resource) and semantic attribute variable (semantic value determinant). The type of knowledge resource determines the degree of formation of semantic richness attributes. The second relation is between the type of knowledge resource and semantic attribute measure variable. The knowledge resources differ according to the degree of effectiveness of semantic richness. The third relation is between semantic attribute variable and semantic attribute effectiveness measure variable. The level of semantic richness attributes determine the degree of effectiveness of semantic richness. That is, the minimum, moderate and maximum degrees of structure and context is equal minimum, moderate and maximum degrees of effectiveness of semantic richness. The fourth relation is between the two dependent variables. The



degree of effectiveness of semantic richness has a direct impact on the degree of semantic value (knowledge) creation potential. The fifth relation is between the independent variable and the knowledge creation dependent variable. The type of knowledge resource determines the degree of semantic value (knowledge) creation potential. All relations are expressed in terms of semantic richness attributes comprising knowledge resource structure and context.

Semantic Mining Variables, Attributes, Relationships and Capabilities

Semantic value creation variables for the semantic mining assessment model were identified as follows. First, the *semantic value* variable is a qualitative measurement value for semantic mining comprising the minimum, moderate and maximum rating scales. Second, the independent variable (variable relating to OD publication) is the *type of knowledge resource* comprising flat text resources, generic knowledge resource and domain knowledge resource. Third, the dependent variables (data publication and consumption measure) pertain to the *degree of semantic value (knowledge) creation potential* and the *degree of effectiveness of semantic mining* (semantic attribute measure). Fourth is the *semantic mining attribute* variable (semantic value determinant) expressed in terms of mining scope (human vs machine); link structure form (hyperlink vs RDF); link coverage scope (wide vs narrow); and knowledge coverage scope (shallow vs deep). The semantic mining relations are as follows. The first relation is between the independent variable (type of knowledge resource) and semantic attribute variable (semantic value determinant). The type of knowledge resource determines the degree of formation of semantic mining attributes. The second relation is between the type of knowledge resource and semantic attribute measure variable. The knowledge resources differ according to the degree of effectiveness of semantic mining. The third relation is between semantic attribute variable and semantic attribute effectiveness measure variable. The level of semantic mining attributes determines the degree of effectiveness of semantic mining. That is, the minimum, moderate and maximum degrees of mining scope, link structure form, link coverage scope and knowledge coverage scope is equal minimum, moderate and maximum degrees of effectiveness of semantic mining. The fourth relation is between the two dependent variables. The degree of effectiveness of semantic mining has a direct impact on the degree of semantic value (knowledge)



creation potential. The fifth relation is between the independent variable and the knowledge creation dependent variable. The type of knowledge resource determines the degree of semantic value (knowledge) creation potential. All relations are expressed in terms of semantic mining attributes comprising mining scope (human vs machine); link structure form (hyperlink vs RDF); link coverage scope (wide vs narrow); and knowledge coverage scope (shallow vs deep).

Step 3: Graphical Representations of Semantic Value Creation Variables, Attributes, Relationships and Capabilities

This is the design stage which involved the integration of identified semantic value creation variables, attributes, relationships and capabilities into graphical representations. Three semantic valuation models were developed to represent semantic value variable (minimum, moderate and maximum) in relation to the independent variable (type of knowledge resource/semantic enrichment approach), dependent variable (the knowledge creation variable – *degree of semantic value creation potential* and the semantic attribute measure variable – *degree of effectiveness of semantic attribute*), semantic attribute variable and. The relationship between any of these variables and the semantic value is what this research terms *semantic valuation*.

Lastly, framework finalisation was followed by the presentation of the final research deliverable, the generalised ODSE Framework together with its implementation support tools in Chapter Nine.

3.4 Research Limitations and Challenges

The structure of the domain knowledge for national agendas is not readily compatible with knowledge representation. Also, the agenda documentations are not precise enough for transformation into description logics. Above that, the presentation of the data does not support system implementation, i.e., lack of unique identification codes for record identification. In addition, the research only deals with published OD. There may be some other unpublished data lying dormant in the hands of stakeholders. Furthermore, some of the published data on the SDG Agenda is from old statistics. In fact what is published is the baseline data ranges in years from 2006 to 2017. Moreover, data for most performance indicators is missing. For instance, SDGs 12, 14 and 17 have no data at all.



Lastly, the research scope is limited to the semantic structure and context of knowledge resources. That is, semantic attributes (enrichment attributes – accessibility, understandability, interoperability, integrability, intelligibility capabilities; richness attributes and mining attributes); functionalities; and implementation support tools are addressed at the semantics/meaning and syntax layers, not application layer.

3.5 Conclusion

The ODSE Framework development adapted the Design Science Research methodology that was implemented in five phases pertaining to the identification of semantic publication and consumption needs for the preliminary semantic enrichment framework design and its design and development, demonstration, assessment, and finalisation. The alignment of the adapted phases to the research questions, objectives, activities and implementation methods is outlined in Table 3.2.



Table 3. 2: Adapted Design Science Research Methodology Application Summary

Research Phase	Research Question (RQ)	Research Objective (RO)	Research Activity (RA)	Research Method (RM)
Phase 1: Framework needs identification <i>[Chapter Four]</i>	RQ1: What are the semantic publication and consumption needs for the preliminary semantic enrichment framework design?	RO1: To identify the semantic publication and consumption needs for the preliminary semantic enrichment framework design	RA1: Identification of semantic consumption needs RA2: Identification of semantic publication needs	RM1: Content Analysis Method
Phase 2: Framework design and development <i>[Chapter Five]</i>	RQ2: What preliminary semantic enrichment framework is derived from the combination of potential semantic enrichment processes, approaches, capabilities, techniques, technologies and their relations?	RO2: To design and develop the preliminary semantic enrichment framework	RA3: Design & development of the preliminary framework	RM2: Adapted Generic Conceptual Framework Development Method
Phase 3: Framework demonstration <i>[Chapter Six (Demo 1) & Chapter Seven (Demo 2)]</i>	RQ3: What are the findings of the preliminary semantic enrichment framework demonstrations?	RO3: To demonstrate the preliminary semantic enrichment framework	Demonstration 1 RA4: Selection of generic ontologies RA5: Development of custom domain ontology Demonstration 2 RA6: Semantic enrichment pre-processing, processing, and post-processing	RM3 (for RA4): Adapted Enrichment and Evaluation Methodology RM4 (for RA5): Adapted METHONTOLOGY Approach RM5 (for RA6): Preliminary ODSE Framework
Phase 4: Framework Evaluation <i>[Chapter Eight]</i>	RQ4: What is the completeness, cohesiveness and effectiveness of the preliminary semantic enrichment framework?	RO4: To evaluate the preliminary semantic Enrichment framework completeness, cohesiveness and effectiveness	Framework Verification RA7: Evaluation of the completeness of the preliminary framework RA8: Evaluation of the cohesiveness of the	RM6 (for RA7 & 8): ODSE Framework Verification Method RM7 (for RA9): ODSE Framework Validation Method



			preliminary framework Framework Validation RA9: Evaluation of the effectiveness of the preliminary framework	
Phase 5: Framework Finalisation <i>[Chapter Eight]</i>	RQ5: What is the reusability potential of the semantic enrichment framework finalised on the basis of the preliminary framework demonstration and evaluation findings?	RO5: To finalise the semantic enrichment framework for reusability	RA10: Redesign of the preliminary semantic enrichment framework for reusability based on the preliminary framework demonstration and evaluation findings	RM8: Semantic Valuation Tools Development Method



CHAPTER FOUR: FRAMEWORK NEEDS IDENTIFICATION

4.1 Introduction

This chapter presents Phase 1 in line with research activities specified in Table 3.2 that pertaining to the identification of semantic value creation needs in the form of semantic publication and consumption needs for the preliminary semantic enrichment framework design. The aim of this phase was to conduct framework scoping so as to provide justification and to assess the feasibility for the development of the ODSE Framework. It was also to establish the semantic value creation scope and capability modelling for the ODSE Framework design.

4.2 Identification of Semantic Consumption Needs

The identified semantic consumption needs are in relation to the OD support needs for the implementation of Botswana's national development agenda which outlines the national strategic direction for Botswana, the Vision Agenda. It expresses Botswana's aspiration of transforming into a knowledge society founded on a knowledge-driven economy [6], [120]. For this aspiration to be achieved, there is need for Botswana to evolve from the current natural resources-driven economy to a knowledge resources-driven economy. The core defining mark of knowledge-driven economies is that they depend on the creation, structuring, sharing, use, diffusion, and recreation of knowledge [120]. Therefore, the kind of economic and social transition aspired by Botswana is practically viable and can only be facilitated and sustained by knowledge resources. This in turn poses a need for Botswana to shift the social and economic investment focus from natural resources to knowledge resources.

Though OD constitutes raw resource for the creation of knowledge resources, there is lack of data-centric approaches toward knowledge resources' production that caters for Botswana's semantic consumption needs. This poses a need for knowledge creation models to leverage the current OD publications toward the implementation of the Vision Agenda aspiration of transitioning to a knowledge-driven economy and society. So, a detailed analysis of the Vision Agenda needs was carried out to figure out the degree of the potential viability, as well as, to device the best approach of leveraging OD to address



Botswana's consumption needs. The output was a conceptual model of OD evolution to open knowledge [120] shown in Figure 4.1.

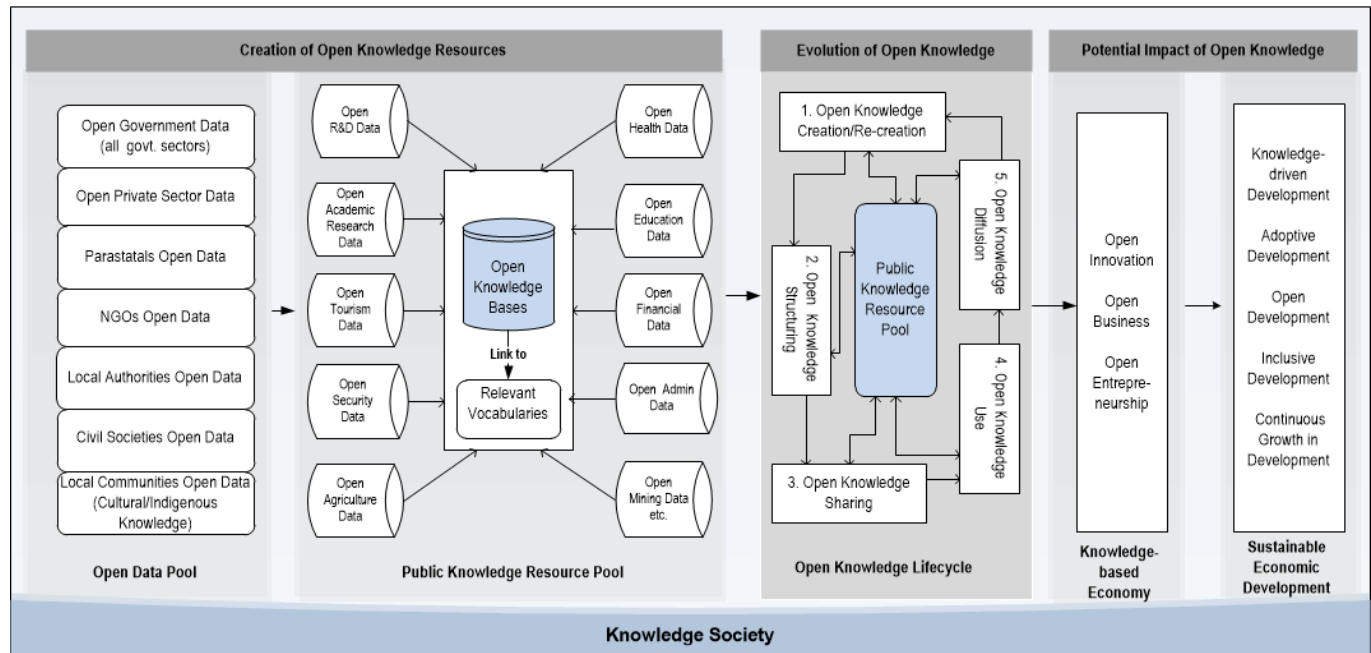


Figure 4. 1: *Open Data Evolution Model* [120]

Figure 4.1, the *Open Data Evolution Model* is the theoretical representation of the OD support needs for the implementation of Botswana's development agenda from a generic perspective to demonstrate the feasibility of OD potential in driving a knowledge economy and society, as well as, to map the OD pathway toward the achievement of a knowledge society founded on a knowledge-based economy. It was developed to serve as a reference model, a conceptual roadmap for Botswana's OD-driven economic and social transition and transformation implementations. To start with, it served as the source of theoretical justification for the implementation viability and the value-adding potential of OD in the transition to a knowledge-driven economy and society. Therefore, it provided motivation for the design and development of the preliminary ODSE Framework. Above that, it served as a point of reference in the conceptualisation, design, development and demonstration of the ODSE framework in numerous ways discussed below.

The *Open Data Evolution Model* depicts that the domain for NPI collections of *Open Government Data* relates to economic and social performance in relation to the development agenda performance data (*Open Data Pool* component). So, from these insights, the scope of ODSE input (raw OD resources) for collection was narrowed down

to the development agenda domain. In addition, having raw OD resources been associated with development agendas, screened down the search for ODSE metadata resources to government strategic documents for OD and development agendas. This composite model further influenced the modelling of the ODSE Framework through its three evolutionary processes, the main processes in Figure 4.1. These processes are modelled as the basic building blocks of the knowledge society establishment, comprising: 1) the raw OD resources evolutionary process, 2) the knowledge resources evolutionary process, and 3) the open knowledge diffusion evolutionary process.

First, the raw OD resources evolutionary process reflected in Figure 4.1 pertains to the *creation of open knowledge resources*. This is the preliminary stage of knowledge resources creation on a national scale, centred on raw OD resources. That is, the raw OD resources evolutionary process acts on raw OD resource collections. Specifically, this preliminary process creates the *Public Knowledge Resource Pool* (semantically enriched OD – knowledge resource collections) from the *Open Data Pool* (nation-wide raw OD gathering). The *Open Data Pool* represents potential semantic value to drive economic and social transformations but in unrefined and unusable form while the public knowledge resource pool represents refined OD as the economic and social powering machine/engine. The preliminary knowledge resources creation process necessitates OD framework designs to be centred on the gathering of raw OD input into the *Open Data Pool* component, followed by the processing that transforms the *Open Data Pool* (nation-wide raw OD gathering) into the *Public Knowledge Resource Pool* – (nation-wide knowledge resources gathering). This OD transformation process as depicted in Figure 4.1 is such that the different sectoral data on social and economic performance indicators is gathered from diverse, disparate sources, after which it undergoes cleaning, reconciliation, transformation and integration. So, this influenced the addition of the pre-processing resource cleaning, transformation and integration into the ODSE Framework architecture. Moreover, this process (the *creation of open knowledge resources*) narrowed down the scope of identifying semantic consumption needs to the creation of the public knowledge resource pool component which in turn raised numerous issues: 1) OD integration needs, specifically, semantic capabilities to integrate OD across economic and social structures; 2) OD enrichment needs, specifically, capabilities to enrich OD through



vocabularies; and 3) OD sharing needs, specifically, capabilities for open and integrated knowledge sharing platforms/repositories.

Second, the knowledge resources evolutionary process reflected in Figure 4.1 pertains to the *evolution of open knowledge*. The knowledge resources evolutionary process acts on the output of the raw OD resources evolutionary process, the *Public Knowledge Resource Pool*. That is, the *evolution of open knowledge* is the advanced modelling of knowledge resources. The knowledge resources evolutionary process circulates knowledge resources around the *Open Knowledge Lifecycle* for open knowledge creation/recreation, structuring, sharing, use and diffusion within the economic and social spheres of application (creation, circulation and use of value) – all of which are powered/driven by (circulates around) the *Public Knowledge Resource Pool*. For the preliminary ODSE Framework modelling, this process was used to determine the scope of investigation for semantic consumption needs comprising the economic and social needs, presented in the form of OD value (knowledge) creation cycle – the need for knowledge circulation within the economy and society. The proposal of the *public knowledge resource pool* is featured in the ODSE Framework and discussions is at the centre of this model. In addition, the ODSE Framework demonstrations and finalisation were implementing the semantic enrichment processes modelled according to the *Open Knowledge Lifecycle* component comprising the first three stages of open knowledge creation (ODSE Components 1.0-6.0), structuring (ODSE Component 7.0) and sharing (ODSE Component 8.0).

Third and final, the open knowledge diffusion evolutionary process reflected in Figure 4.1, explores the *potential impact of open knowledge*. The open knowledge diffusion evolutionary process acts on the output of the knowledge resources evolutionary process, the *open knowledge diffusion*. This process explores open knowledge diffusion in terms of OD potential support for the achievement of a knowledge economy and sustainable development. It first explores open knowledge diffusion impact on the realisation of a knowledge-based economy, specifically, the potential to drive open innovation, business, and entrepreneurship. It further explores the potential cascading effects of a knowledge-based economy on the realization of sustainable economic development – knowledge-driven, adoptive, open, inclusive, and continuous growth in development. For the preliminary ODSE Framework modelling, this process was used to



determine the scope of search for semantic consumption needs, comprising the economic and social needs, presented in the form of OD value (knowledge) saturation on the entire social and economic spheres of life – the need for knowledge saturation on the social and economic spheres of life.

4.3 Identification of Semantic Publication Needs

4.3.1 The Open Data Program Needs

In the context of this research, the OD program needs pertain to Botswana's OD program goals documented in the Botswana OD Readiness Assessment Survey by the World Bank Group [14] which provides a roadmap for future publications. This report unveiled that an OD program could make a significant contribution to Botswana's economic and innovation goals. The report cites evidence of a nascent and innovative 'apps economy' with ideas for ways in which OD could revolutionise key economic sectors, such as agriculture and tourism. In general, the vision of opening up government data is to facilitate government data reuse and this is crucial to the attainment of Botswana's economic and innovation goals that are outlined in the Vision and NDP Agendas [14]. Specifically, OD enables the production of new and valuable products and services and consequently, facilitates the generation of new business opportunities. It is also to be integrated into existing products and services [1], [25], [14]. Above that, OD is expected to yield improved service delivery and transparency benefits from readily availing the data to government decision makers and empowering them with skills needed for its effective use [14]. Moreover, the opening up of government data is aimed at improving the transparency of government and citizen engagement, and in facilitating the sharing of data processing costs with communities [1]. In the development agenda domain, public sector data is critical in tracking development progress and in making informed decisions on environmental, social, and economic policies, programmes and activities [121]. Furthermore, for national statistical systems, OD is expected to improve process efficiencies, facilitate access to new data sources, and increase outreach and build trust in statistics [121].

The realisation of all these potential OD applications prompts for the publication of OD in rich semantic formats. Therefore, the main focus of Botswana's OD program is OD



enrichment to enable its unlimited reusability potential accruing from its infinite sharing potential. The realisation of the potential OD applications overviewed above also necessitates Botswana to rise to the levels of OD readiness. The Botswana OD Readiness Assessment Survey [14] has concluded that Botswana is OD mature by: 1) having many of the key prerequisites for the successful implementation and benefit realisation from OD; 2) having digitised many of the important datasets in preparation for OD move; and 3) having the capability to easily extract data from other key systems through the e-government program for publication of OD [14]. In addition, the Open Data Open Science forum comprising key research and academic institutions and government was established to create consensus on policies, legal frameworks, governance institutions and strategies required to enable Botswana to take full advantage of the global trends in data-driven Open Science to advance key areas of national research importance and increase the competitiveness of key research institutions in Botswana [122]. So, Botswana has a strong network support for OD program implementations.

4.3.2 The Open Data Presentation Needs

In the context of this research, the OD presentation needs pertain to future publication needs drawn from the limitations of the current Web presentation structure and format. That is, OD presentation needs are rooted in the way the current data values are presented, structurally and contextually. An assessment of the current OD publication system (the ‘as-is’ system) was conducted on the basis of Statistics Botswana’s data portal and the metadata documentations for the four development agendas that constitute the Botswana Integrated Indicator Framework [6], [123], [8], [9] [113]. Of these four agendas, published data is on SDG Agenda as part of the *Open Data for Africa* initiative launched by African Development Bank to improve data management and dissemination in Africa. The initiative provides OD portals to African countries with interactive and easy-to-use tools for data visualisations and analysis [124]. This section provides a detailed description of the current OD publication and consumption formats for the ‘as-is’ system. The current publication and consumption formats were finally compared to the future semantic publication and consumption needs. This review pertain to both the OD presentation structure and format. Semantic publication needs are deduced from the



current OD publications in Statistics Botswana's data portal [5]. The SDGs are presented according to the globally defined SDGs structure depicted in Figure 4.2.



Figure 4. 2: SDGs Data Interface Screen [5]

As shown in Figures 4.2, the OD platform offers a point of datasets consolidation, minimizing the efforts of developers and consumers in searching, accessing, downloading and sharing the datasets [90]. With a consolidated view on all of the datasets that have been catalogued, consumers can easily navigate, identify, access and use data of interest [90]. OD is organised according to the 17 SDGs and for each SDGI, there are three views represented by the tabs in Figure 4.3 – browse data view, metadata view and XML view.

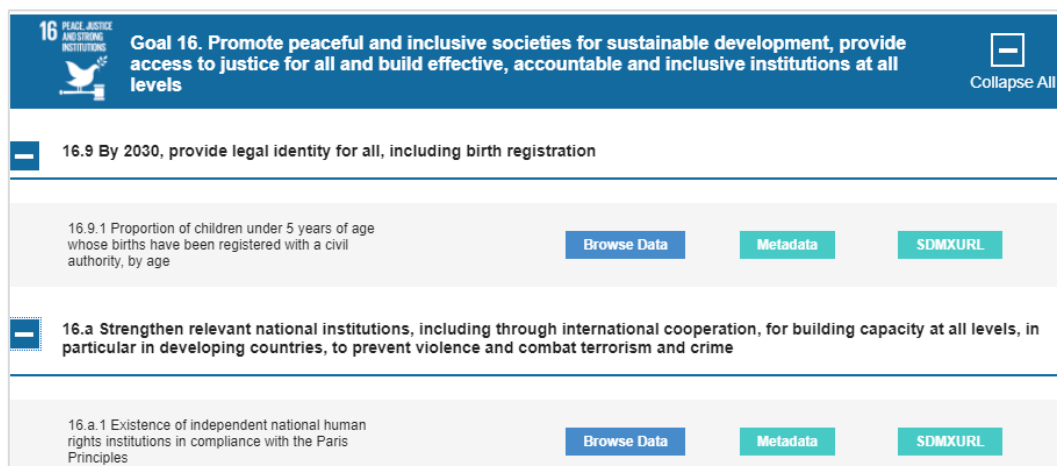


Figure 4. 3: Data View Options (SDG 16) [5]

Open Data View Option 1: Data Browsing View

The *Browse Data* tab opens the basic dashboard view of the data shown in Figure 4.4.

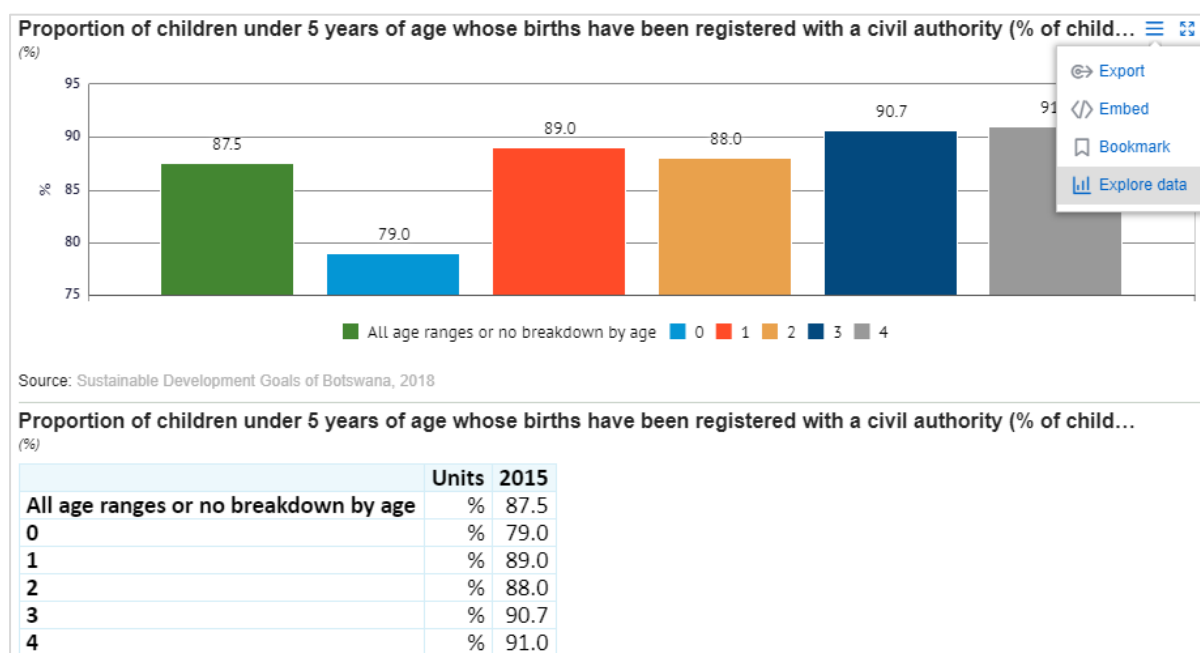


Figure 4. 4: Basic Dashboard View Functionalities – Open Data Access Option (SDGI 16.9.1 Extract) [5]

The basic dashboard view displays the data in the visual (graphical) format as depicted in Figure 4.4. There are five functionalities that can further be performed on the data. The data consumer may further explore, export, embed, bookmark, and or share the data. The data can be exported in textual format (Figure 4.5).

	A	B	C	D	E	F	G	H
1								
2	Proportion of children under 5 years of age whose births have been registered with a							
3	civil authority (% of children under 5 years of age)							
4	(%)							
5								
6								
7		Units	2015					
8	All age ranges or no breakdown by age	%	87.5					
9	0	%	79					
10	1	%	89					
11	2	%	88					
12	3	%	90.7					
13	4	%	91					
14								
15	Source: Sustainable Development Goals of Botswana, 2018							
16								
17								
18								

Figure 4. 5: OD Export in Flat Text Format – Open Data Access Option (SDGI 16.9.1 Extract) [5]



The *explore data* option provides the basic and general functionalities which include the users' ability to bookmark the page and to share the link to the data. In addition, it provides advanced functionalities in the form of flat text intelligibility capabilities pertaining to a set of rich functionalities demonstrated in Figure 4.6. These functionalities facilitate the viewing and extraction of the data.

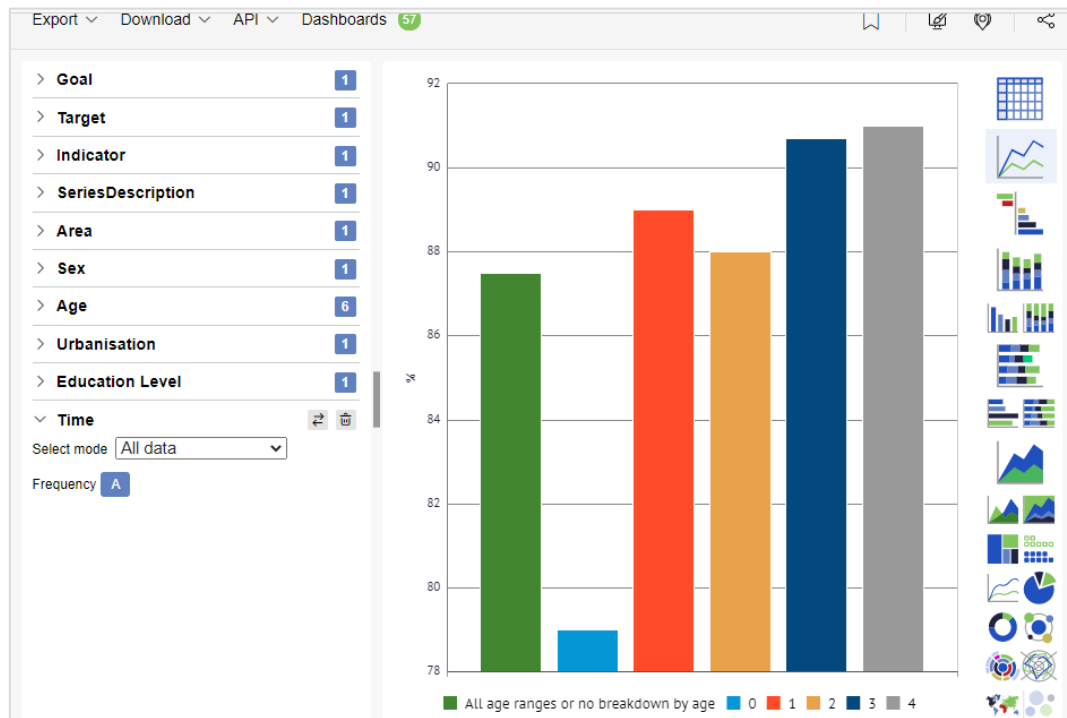


Figure 4. 6: Advanced Dashboard View Functionalities – Column Chart (SDGI 16.9.1) [5]

The data exploration dashboard provides views to a diversity of data visualisations for SDGIs as reflected in Figure 4.6. Firstly, *the Data View Functionalities* present SDGI performance in association with relating SDG target and SDG. The two are among the metadata fields that serve as filters for data views. As depicted on the left of Figure 4.6, the data can be filtered for viewing by SDG description, SDG target description, SDGI description, series description (represent the aggregation field and different categorical groups), area (Botswana), sex, age, urbanisation, education level and time. Despite the rich dashboard options on the right and the rich filter options on the left, they are all limited to a single SDGI and single filter view of the data for any given instance. That is, they offer a unidimensional view of the data. This limits the OD value (knowledge) creation capabilities.

Secondly, *the Data Extraction Functionalities* pertains to data export, download and API links. The export functionality offers the options to export the data in either of the following file formats: presentation slides; spreadsheets; pdf files; images; and as HTML, JavaScript Object Notation¹⁰ (Json), Python¹¹, R¹², C#¹³, and Statistical Data and Metadata Exchange (SDMX) snippets. The OD bound to presentation slides, images and pdf files lock-up data in unstructured format to the documents. Thus, it is difficult to get value from data values as the accessibility, understandability, interoperability, integrability, and intelligibility of the data is strictly limited to humans. Humans can view, copy and print the data. Above that, they can add/record the data into any data/information management system for further editing/conversion/transformation in any way and to any format. In addition, they can store the data in any device and share/distribute the data to anyone and in any way. That is, the data is accessible on the Web in a machine-unreadable and machine-unprocessable format. Furthermore, humans can easily read and understand these publications, but, intensive manual manipulations are required to process the data and make valuable use of. Thus, in cases where the human computing skills are limited, the OD reusability may be compromised.

The spreadsheet format limits OD accessibility, understandability, integrability, intelligibility and usability to humans and proprietary software/applications. Consumers have the flexibility to directly process the data with proprietary software and directly perform any desired transformations. Moreover, the data can be directly and easily converted or exported to other structured formats. The challenge is that the data is locked-up in documents and access, mining and reuse capabilities are dependent on the proprietary software. So, to unlock the OD potential, the data must be locked-out of proprietary software/applications.

Lastly, the data can be imported by external consumers through HTML, Json, Python, R, C#, and SDMX API links. For all these data access links, the portal provides code snippets (like the one in Figure 4.7) to be embedded in code files, specifying the data to be extracted on the basis of the selection. Thus, the snippets provide the capability for

¹⁰ Json is a standard text-based format for representing structured data based on JavaScript object syntax.

¹¹ Python is an interpreted, object-oriented programming language.

¹² R is an open source scripting language for predictive analytics and data visualisation.

¹³ C# is Microsoft's object-oriented programming language that is a hybrid of C++ and Visual Basic.



consumers of the document web to embed the code in their websites and applications to create external links to the source webpage.

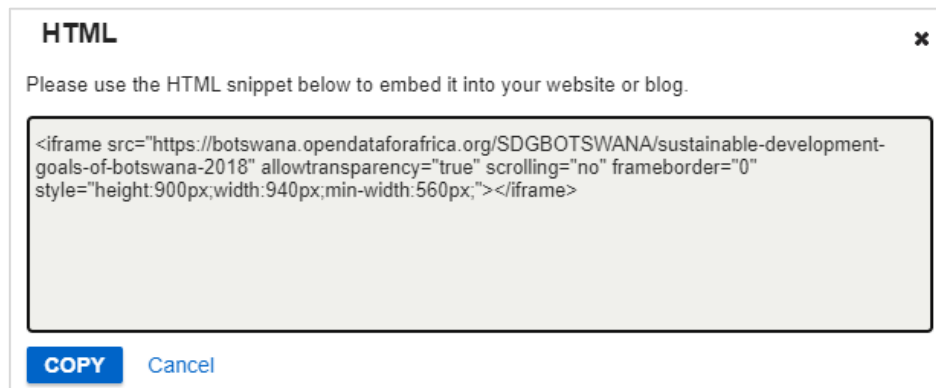


Figure 4. 7: Advanced Dashboard View Functionalities – API Snippet (SDGI 16.9.1) [5]

The download options are spreadsheet and CSV formats. The CSV format lacks the semantics thus compromising OD value creation, mining and reuse potentials.

Open Data View Option 2: SDGI Metadata View

The second data view option in Figure 4.3 represented by the *metadata* tab enables access to SDGI metadata in pdf file format (Figure 4.8). Moreover, the metadata is presented independent of the data it describes. Thus, accessibility, understandability, interoperability, integrability, intelligibility, and usability are limited to humans.

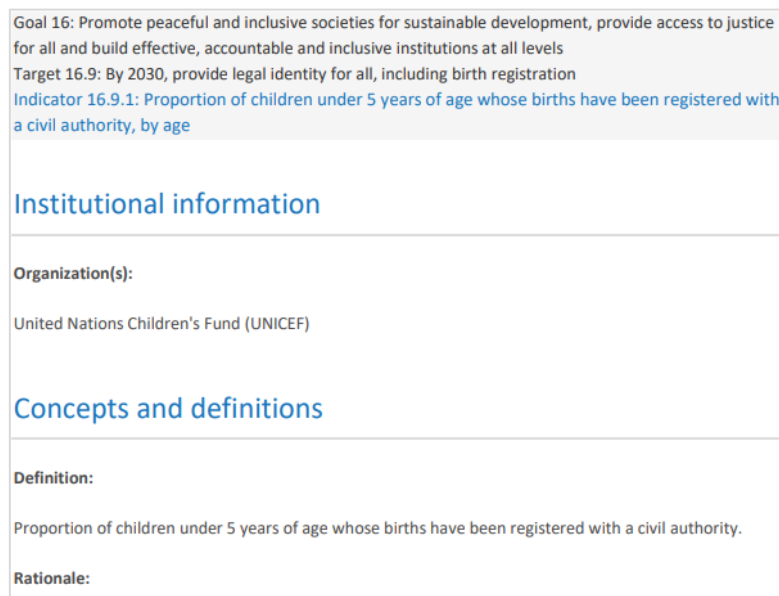


Figure 4. 8: OD Access Option – PDF View (SDGI 16.9.1) [5]



The corresponding metadata for other three development agendas are also in pdf files as shown in Figure 4.9, thus, posing some semantic limitations discussed above.

PILLAR 4: GOVERNANCE, PEACE AND SECURITY				
By 2036, Botswana will be a land of peace, freedom and progressive governance. Crime and corruption in our country will be reduced to minimal levels, with civil society organisations and the community being active participants in the national development process. Botswana will have an accessible and efficient judicial system and will be a peaceful and politically stable country that is an attractive destination for investment. Our country will remain a committed member of the international community whilst maintaining her international image. In order to achieve transformation and meet the expectations of Botswana the constitution and legal framework will be reviewed and reformed.				
Governance, Safety and Security				
Accountable, transparent & effective governance	Ranking of Botswana in the Worldwide Governance Indicators (percentile)	World Bank	85	71.7
	Ibrahim Index of African Governance (IIAG) ranking	Mo Ibrahim Foundation	TBD	3 (2014)
	Corruption Perceptions Index	Transparency International	TBD	28 (2015)
	Freedom in the World score	Freedom House	TBD	73 (2016)
Peaceful and secure nation	Incidence of offences against morality; against the person and against property, per 10,000 population	Botswana Police Service (BPS)	2% reduction a year	97 (2014) (Index = 100, average 2005-14)
	Global Peace Index (ranking)	IEP	TBD	31

Figure 4. 9: Vision Agenda – Pillar 4 Sample (SDGI 16.9.1) [10]

Open Data View Option 3: XML Data View

The third data view option in Figure 4.3 represented by the *SDMXURL* tab label enables access to the XML format of the data (Figure 4.10). XML is a tag-based meta-markup language for representing structured information. XML enables the specification of element and attribute names and how they should be interpreted. For instance, Figure 4.10 presents both the aggregate and disaggregated NPI values for SDGI 16.9.1 together with the publication metadata such as publication ID, date and version; as well as the XML standard specification and the published data source links. It also presents performance values in relation to corresponding SDG, SDG target, SDG indicator and series descriptions being area, sex, age, urbanisation, education, frequency, and year. The XML structure represents the syntax (data structure), thus, enabling machine readability.



However, it does not represent the semantics (meaning of concepts) and as such, is limited in machine processability. As a result, applications can hardly make semantic distinction among data published in distinct XML documents with the same attribute and property names but different set of rules [125]. For instance, the *target* concept has a diversity of contextualisation and Figure 4.10 is not specific on any. This poses a need for semantic representations that specify the interpretation of a concept for use by intelligent agents.



Figure 4. 10: Open Data Access Option – XML View (SDGI 16.9.1) [5]

Compared to the OD views discussed above, XML format presents the data in an open, machine-readable format with no platform proprietary dependency. However, the data is still disintegrated and each data piece does not have the metadata linked to it. This creates ambiguity in the interpretation of the data, thus, limiting OD understandability, intelligibility and usability.

Summary of Open Data Presentation Needs

The study of OD presentation needs for Botswana has demonstrated that the data is presented separately according to the 17 SDGs and for each goal, the data is presented separately on the basis of each SDGI. In addition, the data is presented in flat text format which is easily understood by humans who can easily draw relationships between the different data elements. However, this format is limited in machine accessibility and



understandability. Though these data is also published in XML format which support machine accessibility, XML data is limited in machine understandability. These diversities are reflected in sample datasets uploaded to figshare and available from the following links:

- [ODSE Framework Package 1 - Sample Raw OD for SDGI 16.9.1](#)
- [ODSE Framework Package 2 - Raw OD Resources for SDG Agenda](#)
- [ODSE Framework Package 3 - Raw OD Resources for Vision Agenda](#)

Overall, the data has to be presented in a way that defines the concepts and their relationships, and constraints as well as the computational elements pertaining to the data. In addition, the relating metadata are published in pdf format. Furthermore, the presentation structure has limitations as both the data and the metadata are availed in diverse platforms – the Web and other local (in-house) data management platforms. This diversity of OD presentation structures and formats make it difficult to access, understand, interoperate, integrate, and make intelligible use of the data. This hampers OD value (knowledge) creation and use. Moreover, the advanced view of the data is a rich set of dashboard options and filters that are limited to a unilateral access and view of the data for each particular instance. For the current OD to be of maximal value and effect, there is need to also publish it in reusable formats. In-fact, the Botswana OD Readiness Assessment Survey [14], has raised a need for Statistics Botswana to rapidly release reusable formats of what it already publishes. In addition, the current publications of flat text resources are highly aggregate in nature and thus need to be presented in a disaggregated semantic form for intelligible use.

This data served as the basis for the identification of the data enrichment needs as it specifies the type of information requiring enrichment. It also served to identify the gaps in the dataset such as missing information that would add context to the data and the identification of quality issues that would require normalisation of the data against an authority [116].



4.3.3 The Open Data Representation (Ontology) Needs

In the context of this research, the OD representation needs pertain to semantic modelling needs for OD drawn from source data (flat text OD publications) and corresponding metadata. That is, the OD representation needs are concerned with the restructuring (structure) and contextualisation (meaning) of the data.

National Performance Indicators (NPIs) have several characteristics. They have concrete values obtained through some method and specified in terms of a concrete measure and unit of measurement. They refer to a certain attribute of some entity. Above that, they satisfy some information need required in decision making [87]. Also, they quantify development agenda performance to measure the achievement of agenda goals. In addition, they enable easy tracking of fluctuations in performance of a given agenda between different periods. They also make it easier to compare different agenda performance for different entities. Furthermore, they enable the usage of agenda performance in decision making. Moreover, they enable easy assessment of the achievement of quantitative goals for a given agenda in transformation and improvement processes. However, despite NPIs potential value, OD value (knowledge) creation and use capabilities in most cases are limited by numerous semantic representation challenges pertaining to NPIs' ambiguities and diverse definitions, high levels of aggregation, disparate web publications, and unidimensional presentation formats.

NPIs Ambiguities and Diverse Definitions

As with most development agendas, terminologies for NPIs are devised by independent bodies with no formal representations. In most cases, a diversity of economic sectors produce data on different programmes during the implementations of development agendas. This results in a diversity of NPI value sources (contributors) who operate independently in different public sectors, parastatal organisations, and non-government organisations. This further culminates in heterogeneities in NPI definitions [11], [126]. For instance, the meaning behind terms used to define development agendas, their targets and indicators are often multifaceted due to the independent definitions from different stakeholders. The definition of several keywords used in the development agendas varies across usage scenarios, and some take on different meanings when applied to different entities, leading to ambiguities. This hinders efforts that attempt to



integrate NPIs and also hinder performance comparison and monitoring across the national, regional and global development agendas. This poses a need to represent these diverse shades of meaning in a coherent way to prevent confusion when handling data and developing policy actions as well as enhancing the discoverability and management of development agenda data across a diversity of knowledge domains [68].

NPIs Aggregate Nature

The NPIs, in most cases, represent aggregate and compound values derived either by computations on the basis of a given formula or by aggregation of raw values, or both [127]. The aggregate nature of NPIs compromises their understanding and as a result, they have minimal impact in communicating key issues, as well as being difficult to share [126], [128]. Most managerial misunderstandings during meetings result from different NPI understanding [127]. Moreover, the aggregate nature of NPIs results in the lack of knowledge on their interdependencies. This, in turn, results in them being treated as autonomous, thus, compromising the effectiveness of the design, development and management of NPI monitoring systems [126]. Also, this lack of knowledge on NPI interdependencies limits the reusability potential of the publications. This problem can be resolved by data disaggregation which is one of the systemic issues of SDG Agenda [8].

NPIs Disparateness

Despite the OD abundance on the Web, it is difficult to integrate the data due to lack of schema mappings between datasets [79]. It is also hard to link NPIs to relating strategic elements, such as performance objectives, goals and applicable data dimensions. This disparateness gives a scattered and incomplete view of performance insights. Furthermore, publication platforms are limited to the integration of NPIs within dashboard and scorecard applications. The review of future semantic publication needs in terms of OD interlinkages in the Web of Data (semantic integration needs) pertained to the SDGs National Baseline Indicator Framework proposal [13]. As mentioned above, of these four agendas, only SDG Agenda has data published as OD. This data is to be shared across agendas according to defined agenda mappings. On the one hand, the domesticated SDGs are defined with mappings to Vision Pillars [10]. The sample mapping for SDGI 16.9.1 is reflected in Figure 4.11.



1.4 SDGs Indicators mapped Vision 2016 Pillar 4:
"GOVERNANCE PEACE AND SECURITY"
Goal 16. Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels
Target 16.9 By 2030, provide legal identity for all, including birth registration
Indicator 16.9.1 Proportion of children under 5 years of age whose births have been registered with a civil authority, by age
Target 16.a Strengthen relevant national institutions, including through international cooperation, for building capacity at all levels, in particular in developing countries, to prevent violence and combat terrorism and crime
16.a.1 Existence of independent national human rights institutions in compliance with the Paris Principles
Goal 17. Strengthen the means of implementation and revitalize the Global Partnership for Sustainable Development

Figure 4. 11: SDG Agenda – Vision Agenda Sample Mapping (SDGI 16.9.1) [10]

On the other hand, the Africa Agenda performance indicators are designed with mappings to SDG Agenda performance indicators [8]. The sample mapping for SDGI 16.9.1 is reflected in Figure 4.12.

ASPIRATION 6. AN AFRICA WHOSE DEVELOPMENT IS PEOPLE DRIVEN, RELYING ON THE POTENTIAL OF THE AFRICAN PEOPLE			
Goal 17: Full Gender Equality in All Spheres of Life			
Priority area	Agenda 2063 Target	Indicator	Corresponding SDG Indicator
Priority Area 1. Women Empowerment	1. Equal economic rights for women, including the rights to own and inherit property, sign a contract, save, register and manage a business and own and operate a bank account by 2026	Proportion of total agricultural population with ownership or secure rights over agricultural land	5.a.1
		Share of women among owners or rights bearers of agricultural land by type of tenure.	5.a.1
	2. At least 30% of all elected officials at local, regional and national levels are Women as well as in judicial institutions	Proportion of seats held by women in national parliaments, regional and local bodies	5.5.1
Priority Area 2. Violence & Discrimination against Women and Girls	1. Reduce 2013 levels of violence against women and Girls by at least 20%	Proportion of women and girls subjected to sexual and physical violence	5.2.1
	2. Reduce by 50% all harmful social norms and customary practices against women and girls and those that promote violence and discrimination against women and girls	Proportion of girls and women aged 15 - 49 years who have undergone female genital mutilation/ cutting by age	5.3.2
	3. Eliminate all barriers to quality education, health and social services for Women and Girls by 2020	Proportion of children whose births are registered in the first year	16.9.1

Figure 4. 12: Africa Agenda – SDG Agenda Sample Mapping (SDGI 16.9.1) [8]



The SDGs National Baseline Indicator Framework expresses the need to map Vision Agenda performance indicators to Africa Agenda and SDG Agenda performance indicators to ensure that in pursuing national aspirations, Botswana, is at the same time meeting regional and global goals. It also proposes the mapping of Vision Agenda performance indicators to NDP Agenda performance indicators [13]. Currently, semantic integration is limited by stand-alone publications and a lack of semantic integration models for the Botswana Integrated Indicator Framework. Consequently, the reusability potential of the current OD publications is limited. Therefore, the semantic integration need pertains to the development of a consolidated indicator framework to support integrated NPI publications. So, models of semantic publishing need to factor in these interlinkages.

NPIs Unidimensional Presentation Formats

The Vision Agenda defines three dimensions for indicator tracking, namely national, regional and global [13]. In addition, the SDG Agenda necessitate a reflection on the economic, environment, and social dimensions of development. The NPI nature also demands for a reflection on the computational type of measure, which may either be a base measure, derived measure or an aggregate measure. Furthermore, the current platform is mostly based on dashboard and scorecard applications and offers a unidimensional view of the data. The multidimensional presentation of this data is needed for increased intelligibility capability.

4.4 Conclusion

Chapter Four documents how Research Objective 1, *identification of semantic publication and consumption needs for the preliminary semantic enrichment framework design*, addressed Research Question 1 expressed as:

What are the semantic publication and consumption needs for the preliminary semantic enrichment framework design?

This research question sought to establish the semantic value creation needs pertaining to the functionality modelling specifications for the ODSE Framework design, as well as, the design and development feasibility and scope establishment. The semantic value creation needs identification pertains to semantic publication and consumption needs.



There are three main ODSE Framework needs in relation to semantic publication. The first semantic publication need pertains to OD presentation formats, the ODSE Framework need to support the implementation of standardised and unified publication structure and format. Therefore, the ODSE framework processing was designed to create standard and unified structure and format. The second semantic publication need pertains to OD representation structure, the need to support the implementation of the consolidated indicator framework for the four development agendas. Thus, the ODSE framework processing was designed to address the semantic representation and integration of the four development agendas that constitute the Botswana Integrated Indicator Framework. The third semantic publication need pertains to OD program needs, the ODSE Framework need to support the generation of rich semantic formats for reusability of the current OD publications. Therefore, the ODSE Framework was designed to enhance semantic capabilities of flat text resources with ontology-driven semantic enrichment.

The ODSE Framework needs in relation to semantic consumption pertains to OD support for the implementation of the Vision Agenda, the ODSE Framework need to support the generation of knowledge resources for the implementation of the economic aspiration of a transformation into a knowledge society founded on a knowledge-driven economy. Therefore, the ODSE framework processing was designed to leverage OD for the generation of knowledge resources to support the implementation of the knowledge-driven economy and society. In addition, the semantic consumption need was modelled into the *Open Data Evolution Model* (Figure 4.1), a generic conceptualisation of the creation of knowledge resources from OD resources. Therefore, it is the conceptual roadmap to guide the ODSE Framework design, development, demonstration and implementations.

The next section, Chapter Five, presents the conceptual design and development of the preliminary semantic enrichment framework to set the foundation for implementations that address these needs.



CHAPTER FIVE: FRAMEWORK DESIGN AND DEVELOPMENT

5.1 Introduction

Chapter Five presents Phase 2 according to research activities specified in Table 3.2 concerning the design and development of the semantic enrichment framework.

5.2 The Design and Development of the Framework

The preliminary Ontology-Driven Semantic Enrichment Framework (ODSE Framework) shown in Figure 5.1 was developed according to the conceptual framework development method as follows:

- The conceptual framework development method (Subsection 3.3.2 – Phase 2: Framework Design and Development);
- Its adaptation (Subsection 3.3.2); and
- Its application (discussed in this section – Section 5.2).

The design and development of the framework was implemented in three steps pertaining to the review of research literature; the identification of semantic enrichment entities, elements, and their relations; and the generation of the conceptual semantic enrichment framework. Framework design and development conceptualisation and implementation was founded on identified publication and consumption needs presented in Chapter Four.

5.2.1 Step 1: Review of Research Literature

Step 1 involved a synthesis of existing knowledge on semantic enrichment pertaining to potential semantic enrichment processes, approaches, capabilities, techniques, technologies and their relations. Research literature was conducted to create an understanding of semantic enrichment entities comprising processes, approaches, capabilities, techniques, technologies and their relations. Thus, the literature review explored the OD concept from four angles: 1) OD as the future driver; 2) OD value (knowledge); 3) knowledge reuse potential; and 4) knowledge creation challenges. That was followed by an exploration of the potential of a combination of ontology, LOD and semantic enrichment in the creation and use of knowledge resources.



5.2.2 Step 2: Identification of Semantic Enrichment Framework Entities, Elements and Relations

Step 2 involved the design of the framework, specifically, the preliminary conceptualisation of the model by analysing the research literature reviewed in Step 1. To start with, a review of the OD value (knowledge) and reuse capabilities (Section 2.2) provided justification that motivated the development of the framework. In addition, the research literature has revealed numerous elements pertaining to semantic enrichment entities that became the basis for the framework design (shown in Table 5.1).

Table 5. 1: Semantic Enrichment Elements

Semantic Enrichment Entities	Entity Elements
Semantic Enrichment Processes	semantic integration, semantic web resource linking (Section 2.5), OD value (knowledge) creation (Sections 2.2-2.5)
Semantic Enrichment Approaches	ontology-driven semantic enrichment and LOD-driven semantic enrichment (Subsection 2.5.4)
Ontologies	Top-level/upper-level/foundational ontologies, domain ontologies, application ontologies (Subsection 2.3)
Semantic Attributes	enrichment, richness and mining attributes (Subsection 2.5.5), OD value (knowledge) reuse (Subsection 2.2.3)
Semantic Techniques	5-star Linked Data Standard, Web of Data, semantic web (Section 2.4)
Semantic Technologies	Ontology (Subsection 2.3), knowledge organisation systems/vocabulary (Subsection 2.5.1), semantic web technologies (OWL, RDF, triples, triple-store, LOD) – (Sections 2.4-2.5), OD value (knowledge) mining (Subsection 2.5.5)

Furthermore, the research literature has revealed numerous relationships between semantic enrichment entities and elements.

- The first relationship is between OD value (knowledge) creation (Section 2.2), ontologies (Section 2.3) and LOD (Section 2.4). The review discovered that the potential to maximise OD value creation is determined by the level of semantic richness which is created through the use of ontologies and the application of LOD semantic publication standards.
- The second relationship is between semantic enrichment and semantic capabilities (Section 2.5.2 and Subsection 2.5.5). Semantic enrichment may either be structural in nature for semantic enrichment of accessibility, interoperability and integrability capabilities; or contextual in nature for semantic enrichment of understandability capabilities; or may be both structural and contextual in nature for semantic enrichment of the intelligibility capabilities.



- The third relationship is between Tim Berners-Lee's 5-Star scheme for Linked Data (Section 2.4) and semantic attributes (Section 2.5.5). The 5-Star scheme is a publication standard that addresses semantic enrichment of semantic attributes. Thus, the two were combined to serve as the standard of semantic enrichment transformations and the evaluation standard for the validation of the ODSE framework (publication and consumption efficiency standard – Table 3.1 and Figure 3.2 discussed in detail in the methodology – Subsection 3.3.4).
- The fourth relationship is between the triples and the triple-store (Section 2.4). Triples are stored in triple-stores.
- The fifth relationship is between OD value (knowledge) creation, semantic enrichment and knowledge mining (Section 2.5.5) and knowledge reuse (Section 2.2.3). Semantic enrichment has a direct impact on knowledge creation potential which in turn has a direct impact on knowledge mining potential. The knowledge mining potential in turn, has a direct impact on knowledge reuse.

Moreover, the research literature has revealed numerous insights on semantic enrichment entities and elements. Firstly, the review revealed a maximal effect from an ontology-driven semantic enrichment model (Section 2.3, 2.5). This validated the application of ontology as the basis for the semantic enrichment process. So, the semantic enrichment model chosen for this research is the ontology model. The semantic enrichment framework was developed for the contextualisation of flat text resources with its metadata represented by ontological constructs. Also, the ontology scopes described in Subsection 2.3.1 served as the basis for the development of the ODSE approaches. Specifically, two ODSE approaches were devised to implement the ontology-driven semantic enrichment.

Secondly, as revealed above, there are semantic enrichment methods for structured, semi-structured, and unstructured data (Section 2.5.2). Two of these methods proved to be relevant to this research. The semi-structured data in text and pdf files were enriched through semantic enrichment methods that transformed it into structured form for enhanced data access. Semantic enrichment methods for structured data in spreadsheet files were used to enhance the quality and discoverability of metadata with more



contextualised meanings through the use of ontologies and they were also used to generate the knowledge base to support knowledge discovery.

Thirdly, the framework conceptualisation followed the LOD standard (Subsection 2.5.4 and Section 2.4) in mapping the transition from static and disintegrated OD on NPIs (flat resources) to intelligent LOD (knowledge resources) for enhanced OD value (knowledge) creation potentials. Also, based on the literature review, the LOD-driven semantic publication standards (i.e., RDF, OWL and Berners-Lee's 5-star deployment scheme for Linked Data) were adopted in the implementation and evaluation of the semantic enrichment framework.

Lastly the semantic enrichment framework processing was according to the three processes that were discovered to be the standard semantic enrichment process for existing frameworks (Subsection 2.5.6). So, the preliminary semantic enrichment framework was modelled around three stages comprising the semantic enrichment pre-processing, processing and post-processing.

5.2.3 Step 3: Generation of the Conceptual Semantic Enrichment

Framework

Step 3 pertained to the development of the framework, involving two activities. The first activity involved the classification of identified semantic enrichment entities and elements (Table 5.1) according to functionality that corresponds to either semantic enrichment pre-processing, processing or post-processing. That was followed by the consolidation of the components into a single, solid, holistic framework by structuring the flow of entities and elements' functionalities per processing stage on the basis of the insights and relationships between framework entities and elements specified in Step 2. The *Open Data Evolution Model* shown in Figure 4.1 and the OD program, presentation and representation needs were also considered in putting the pieces of the framework together, specifically, in the determination of the process flows. The final deliverable is the preliminary semantic enrichment framework described in the next section.

5.3 The Preliminary Semantic Enrichment Framework

The result of this phase is the preliminary conceptual design artifact of the Ontology-Driven Semantic Enrichment Framework shown in Figure 5.1.



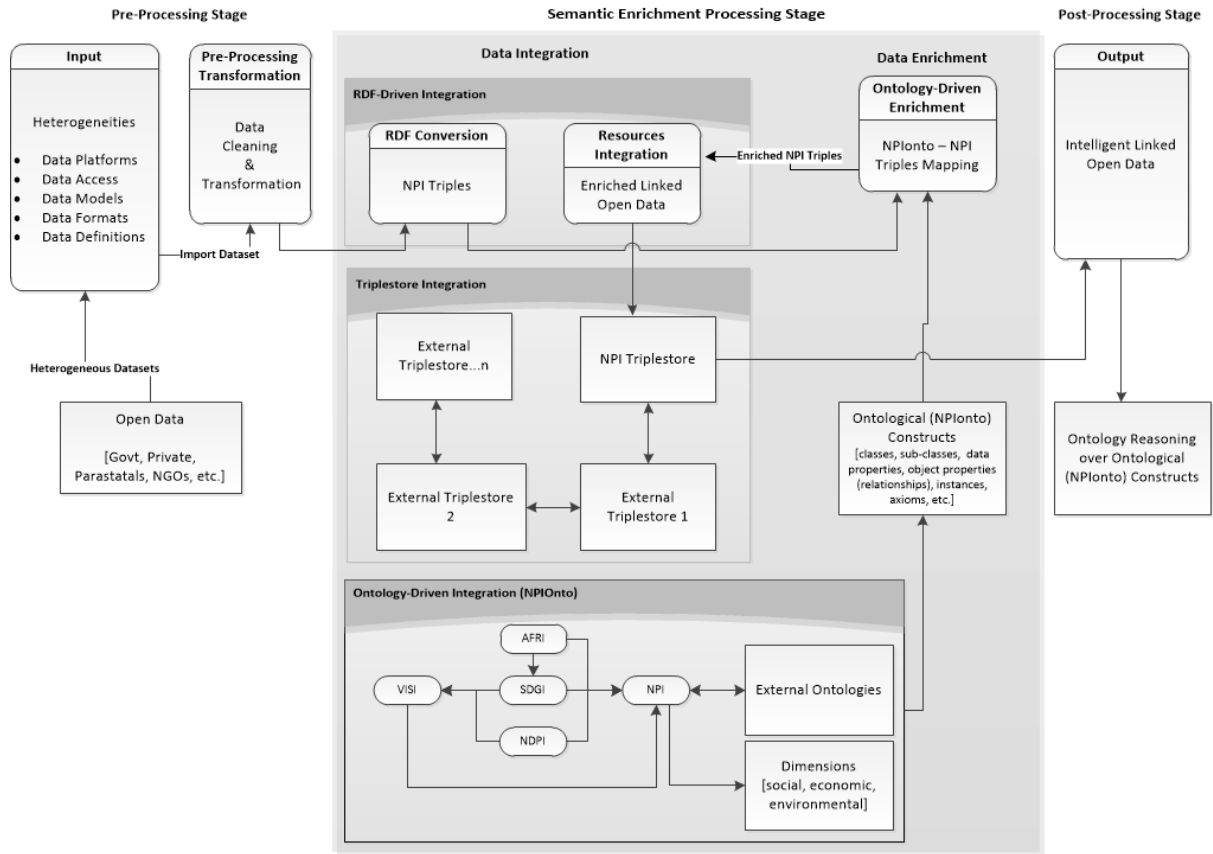


Figure 5. 1: The Preliminary Ontology-Driven Semantic Enrichment Framework [119]

As depicted in Figure 5.1, the preliminary framework is in three stages that are described in the following Subsections.

5.3.1 Stage 1: Open Data Pre-processing

The *Open Data pre-processing* stage prepares OD for integration and enrichment processing. It comprises the input and pre-processing transformation components. The *input component* represents OD published in its raw form. Generally, OD is characterised by a variety of heterogeneities. OD on the performance of development agendas is produced by different stakeholders in all economic sectors, i.e., government and private sectors, parastatals, non-government organisations, etc. These heterogeneities in data sources result in heterogeneous datasets in terms of data platforms (diversities in databases, applications, etc.), data formats (PDFs, spreadsheet and database tables, CSVs, etc.), data definitions (variations in performance indicator formulas and calculations), data access (standalone, Local Area Networks, cloud networks, etc.) and relational data models. In their heterogeneous forms, raw OD resources reporting on NPIs for the four



agendas that are the focus of this research were imported from multiple sources. The detailed discussions on the pre-processing input (the OD presentation needs discussions) is documented in Subsection 4.3.2. Under the *pre-processing transformation component*, raw data undergoes the transformation process entailing data cleaning and modification.

5.3.2 Stage 2: Open Data Semantic Enrichment Processing

The *Open Data semantic enrichment processing* stage comprises the data integration and enrichment components. The *data integration* component addresses the challenge of NPIs' disparateness. It is done at three levels: RDF-driven integration, triple-store integration, and Ontology-driven integration. *RDF-driven integration* pertains to RDF conversion, the semantic transformation process that converts disparate flat text resources into a more flexible, data integration structure, the RDF triple structure form of subject-predicate-object explained and demonstrated in Section 2.4. RDF stands for Resource Description Framework, a model for data publishing and interchange on the Web [80].

As the RDF conversion pertains to national performance indicator (NPI) data, the product is NPI triples, NPI data formatted in the subject-predicate-object structure. The NPI triples support multifunctional use as it allows data to be linked in comparison to traditional HML which represents a Web of Documents. The NPI triples are further interlinked into an RDF triple-store, a graph database that stores semantic facts as a network of objects with embodied or materialised links between them for more effective management of highly interconnected data [80]. Above that, triple-stores with their subject-predicate-object format, are more flexible compared to relational databases. The designated triple store for the storage of NPI triples is the NPI triple store. The data store triples interlinkage, the NPI triple-store represents the *triple-store integration* which pertains to the integration of the datasets for the four agendas that are the focus of this research. The NPI triple-store integration can further be extended through triple-store merging with other external triple-stores (Web triple-stores with triples relating to NPI triples) to allow for semantic queries that integrates a wide web knowledge base. The *ontology driven integration* pertains to the use of knowledge representation models to unify the datasets. NPIOnto is a knowledge model for the integration of NPIs for the four agendas with the following mappings: 1) SDG Indicator (SDGI), Africa Agenda Indicator



(AFRI), NDP Indicator (NDPI), and Vision Indicator (VISI) instances are mapped to NPI; 2) AFRI instances are defined with mappings to SDGI instances; 3) SDGI and NDPI instances are directly mapped to NVI instances; and 4) AFRI instances map to NVI instances through SDGI mappings.

Furthermore, the NPIOnto models relationships among NPI instances to give a better understanding of the dependencies that exist between them. It also models NPI's underlying attributes and mathematical structures to address the ambiguity challenges. Moreover, the NPIOnto represents knowledge regarding NPI's relation to performance targets, performance outcomes, priority areas, performance goals, thematic areas, aspirations, pillars, etc. The analytic dimensions (social, economic and environment) are factored in as well in the model to implement high-level data disaggregation.

The *Data Enrichment Component* involves enhancement of NPI triples with information about its meaning. This is accomplished by a process of mapping NPI triples to relevant NPIOnto constructs, knowledge representation entities pertaining to classes (representing domain concepts), properties (representing concepts attributes and relations) and axioms (representing statements of truth or facts) for explicit definitions of NPIs. This process turns the heterogeneous, ambiguous, and difficult-to-interpret NPI content into machine-processable data for reusability and for better search and presentation in a LOD environment. This component addresses the challenge of heterogeneities that relate to diverse data definitions by restricting meaning to specific ontology constructs. There are two processes performed here, pertaining to the identification of NPIOnto constructs in the form of classes, properties and axioms and finally, the actual implementation of the ontology-driven enrichment which pertains to the creation of the semantic layer through the mapping of NPIs to ontological constructs, i.e., classes, properties and axioms in NPIOnto. Resources integration, the interlinkages of ontology enriched data constitute the enriched Linked Open Data. The embedding of the semantics and syntax structure into the data enables intelligibility operations. Thus, enriched Linked Open data is referred to as intelligent Linked Open Data.

5.3.3 Stage 3: Open Data Post-processing

The last, third stage, is *Open Data post-processing*. The deliverable of this stage is *intelligent data* as the data holds within itself all the information that describes the



structural and contextual features needed for its understanding and interpretation. This data is machine-readable, easy to search, query, use, and reuse. It can also be linked to different data sources to get deeper insights and to uncover hidden relationships. With intelligent LOD, objects can be interconnected, thus facilitating the discovery of knowledge and the exchange of information.

5.4 Conclusion

Chapter Five documents how Research Objective 2, *design and development of the preliminary semantic enrichment framework*, addressed Research Question 2 being:

What preliminary semantic enrichment framework is derived from the combination of potential semantic enrichment processes, approaches, capabilities, techniques, technologies and their relations?

This research question served to establish the architectural structure and functionality for the conceptual design of the preliminary ODSE Framework (Figure 5.1). The preliminary ODSE Framework is a more specific conceptualisation that demonstrates the practical creation of knowledge resources from OD resources based on the conceptual reference model, the *Open Data Evolution Model* (Figure 4.1). Specifically, it models the creation of knowledge resources in the form of Intelligent LOD to increase the value-adding potential of OD publications and to increase the knowledge mining potentials. The framework is comprised of three interrelated stages: pre-processing, processing and post-processing. These stages are further broken down into five components¹⁴, comprising the pre-processing input, pre-processing transformation, data integration, data enrichment, and post-processing output. It is worth noting that the variables for the preliminary ODSE Framework are more specific to Botswana development agenda domain. This constrained variable scope was to enable ease of demonstrations. The final framework was generalised for different development agenda domain-specific applications on the basis of the demonstration findings. So, the next chapters discuss the preliminary ODSE Framework demonstrations, evaluation, finalisation and the presentation of the final framework as the research contribution.

¹⁴ The groupings of related processes and technologies that achieve a given unique functionality.



CHAPTER SIX: FRAMEWORK DEMONSTRATION 1 (SEMANTIC ENRICHMENT APPROACH)

6.1 Introduction

As the proposed semantic enrichment framework is ontology-driven, the first demonstration involved the establishment of one of the pillars of semantic enrichment processing, the semantic enrichment approach which is determined by ontology scope. Therefore, Chapter Six presents Phase 3 according to Demonstration 1 activities specified in Table 3.2. Phase 3 involved the selection of generic ontologies and the development of the custom domain ontology.

6.2 Demonstration 1: The Selection of Generic Ontologies

As stated in Subsection 3.3.3, the selection of generic ontologies adapted the *Enrichment and Evaluation Method* (RM3) proposed by the Europeana Task Force [116] comprising the following phases: 1) analysis of source data and metadata resources for ontology search literals; 2) identification of ontology enrichment requirements; 3) search for potential ontology targets; and 4) selection of suitable ontology targets.

6.2.1 Phase 1: Analysis of Source Data and Metadata Resources for Ontology Search Literals

The data sources comprised the published NPI values for the 17 SDGs and the metadata for the four agendas as listed in Table 6.1.



Table 6. 1: Source Datasets

Agenda Data/Metadata	Datasets	Fields Per Dataset
SDG Agenda data	Published OD	SDG Name, SDG Theme and SDG Description; SD Target Code and SD Target Description; SDGI Code and SDGI Description; SDGI Performance Measurement Description, SDGI Performance Measurement Unit, SDGI Performance Measurement Year; SDGI Aggregate Performance Value and SDGI Disaggregate Performance Value.
SDG Agenda metadata	SDGI metadata	SDGI Code, SDGI Name, SDGI Description, UNSD Code, SDGI Sub-Theme, SDGI Development Type, SDGI Measurement Type, SDGI Measurement Scope, SDGI Disaggregation Factors and SD Target Relation
	SD Target metadata	SD Target Code, SD Target Name, SD Target Description and SD Goal Relation
	SD Goal metadata	SD Goal Code, SD Goal Name, SD Goal Description, SD Agenda Relation and Vis Pillar Relation
	SD Agenda metadata	SD Agenda Code, SD Agenda Name, SD Agenda Description, SD Agenda Begin Year and SD Agenda End Year
Africa Agenda metadata	AFRI metadata	AFRI Code, AFRI Name, AFRI Description, AFRI Formula, AFRI Unit Measure, AFRI Disaggregation Factors, AFRI Data Collection Frequency, AFR Target Relation and SDGI Relation
	AFR Target metadata	AFR Target Code, AFR Target Name, AFR Target Description and AFR Priority Relation
	AFR Priority metadata	AFR Priority Code, AFR Priority Name, AFR Priority Description and AFR Goal Relation
	AFR Goal metadata	AFR Goal Code, AFR Goal Name, AFR Goal Description, and AFR Aspiration Relation
	AFR Aspiration metadata	AFR Aspiration Code, AFR Aspiration Name, AFR Aspiration Description, and AFR ImpPlan Relation
	AFR Implementation Plan metadata	AFR ImpPlan Code, AFR ImpPlan Name, AFR ImpPlan Description, AFR ImpPlan Begin Year, AFR ImpPlan End Year and AFR Agenda Relation
	AFR Agenda metadata	AFR Agenda Code, AFR Agenda Name, AFR Agenda Description, AFR Agenda Begin Year and AFR Agenda End Year
Vision Agenda metadata	VISI metadata	VISI Code, VISI Name, VISI Description, VISI Data Source, VISI Outcome Relation and NDPI Relation
	VIS Outcome metadata	VIS Outcome Code, VIS Outcome Name, VIS Outcome Description and VIS Pillar Relation
	VIS Pillar metadata	VIS Pillar Code, VIS Pillar Name, VIS Pillar Description and VIS Agenda Relation
	VIS Agenda metadata	VIS Agenda Code, VIS Agenda Name, VIS Agenda Description, VIS Agenda Begin Year and VIS Agenda End Year
NDP Agenda metadata	NDPI metadata	NDPI Code, NDPI Name, NDPI Description, NDPI Data Source and NDPI Outcome Relation
	NDP Outcome metadata	NDP Outcome Code, NDP Outcome Name, NDP Outcome Description and NDP Theme Relation
	NDP Theme metadata	NDP Theme Code, NDP Theme Name, NDP Theme NDP Theme Description and NDP Volume Relation
	NDP Volume metadata	NDP Volume Code, NDP Volume Name, NDP Volume Description and NDP Agenda Relation
	NDP Agenda metadata	NDP Agenda Code, NDP Agenda Name, NDP Agenda Description, NDP Agenda Begin Year and NDP Agenda End Year

Table 6.1 provides a detailed description of the contents of the data for each OD resource. All entity field names (column 3 of Table 6.1) served as keywords or categories representative of the source dataset for the different dimensions of the data that needed to be enriched or contextualised. So, the search for ontology targets was based on these keywords. The second aspect pertained to the analysis of the data structure which involved the identification of the properties used in the source data and metadata resources containing the literals that were to be enriched. The property literals identified are: *name, theme, description, unique identity, unit, year, relation, and value*.

6.2.2 Phase 2: Identification of Ontology Enrichment Requirements

In addition to OD representation (ontology) needs identified in the preliminary analysis of the data sources (Subsection 4.3.3), additional requirements supporting specific services or applications were identified. Further analysis of the source data and metadata resources provided insights on the requirements to be addressed through the selected ontology targets. With generic ontology targets of top-level/upper/reference ontologies, the objective of data enrichment was to make the NPI domain dataset more global for enhanced discoverability. With a custom domain ontology targets, the objective was to reach a balance between generality and unambiguity by making the dataset more specific. Also, the objective of custom domain ontology targets was to link resources to other similar resources and to controlled resources as well. So, these objectives served as the basis for the next stage of searching for potential ontology targets.

6.2.3 Phase 3: Search for Potential Ontology Targets

The search for potential ontology targets was based on enrichment requirements and identified literals (Subsections 6.2.1-6.2.2). The first search looked for SDG Agenda domain ontology (SDGIO) targets on the basis of the SDG constructs shown in Table 6.2.

Table 6. 2: SDGIO Targets

SDG Construct	SDGIO Property
Goal description	SDGIO:00000000 (SD Goal)
Target code	SDGIO:00000074 (UNSDG target ID)
Target description	SDGIO:00000001 (SD target)
SDGI code	SDGIO:00000242 (UN SDG Indicator ID)
SDGI description	SDGIO:00000002 (Sustainable Development Indicator)
Aggregate value	SDGIO:00000003 (SDG Indicator Value)



The second search looked for applicable generic ontologies. Numerous ontologies were identified with a diversity of contextualisation of these literals in Table 6.3.

Table 6. 3: Potential Generic Ontology Targets

Literal	Potential Ontology Targets
Name	• NCI Thesaurus OBO Edition - the words or language units by which a thing is known • SemanticScience Integrated Ontology - a name is a label used to identify an entity • eNanoMapper Ontology - the words or language units by which a thing is known
Theme	• BBC Ontology – this concept is used to represent intangible things or abstract concepts • dbpedia – theme property (dbp:theme)
Description	• Allotrope Merged Ontology Suite - has description relates an entity to an information content entity that is its description (has description) • eNanoMapper Ontology - a written or verbal account, representation, statement, or explanation of something • SemanticScience Integrated Ontology - a description is language entity in which elements of a language (formal or natural) are used to characterise an entity • Apollo Structured Vocabulary (Apollo-SV) - a textual entity that specifies characteristics of some entity
unique identity	• dbr:Unique-identifier
Unit	• Units of Measurement Ontology - a unit of measurement is a standardised quantity of a physical quality • Ontologized MIABIS (Minimum Information About Biobank data Sharing) - a relation between a value specification and its unit of measurement (has measurement unit label) • BioAssay Ontology - It is the definite magnitude of a physical quantity or of time. It has a quantity and a unit associated with it (unit of measurement)
Year	• Dublin Core Metadata Initiative (DCMI terms) - a point or period of time associated with an event in the lifecycle of the resource (date) • SemanticScience Integrated Ontology - a year is a period of time taken by a planet to make one revolution around the sun • Units of Measurement Ontology - a time unit which is equal to 12 months which in science is taken to be equal to 365.25 days
Relation	• Dublin Core Metadata Initiative (DCMI terms) - a related resource • OBO Ontology Library - an edge that links together terms
disaggregate value	• Dbpedia-de:Disaggregation • Allotrope Merged Ontology Suite - is a facet that relates an information content entity to a code that describes a category for classification (has Categorical Value) • SemanticScience Integrated Ontology - has measurement value is a relation between a quality/realisable and a measurement value (has measurement value)

Overall, Table 6.2 lists potential ontology targets from related domains, SDGIO targets, targets more specific to the SDG Agenda domain for more domain clarity. Potential ontology targets listed in Table 6.3 are from unrelated domains, the more generic set of targets to enable more semantic interoperability capabilities. This combination was meant to ensure that the datasets are understandable, integrable and interoperable.



6.2.4 Phase 4: Selection of Suitable Ontology Targets

The last stage pertained to the selection of suitable ontology targets from the list of potential targets in Subsection 6.2.3. The final selection encompassed all of the first ontology target group (SDGIO) as the only available SDG Agenda ontology constructs – potential ontology targets with no alternative options (Table 6.2). For the second ontology target group with numerous options (Table 6.3), the selection criteria was mainly based on availability, coverage and quality factors discussed in Subsection 3.3.3 above. In terms of availability, the ontology targets were selected on the basis of their web availability, compliancy to the Linked Data best practices. In terms of coverage, the selection was focused on ontology definitions that represent the development agenda knowledge domain with high precision of meaning and high expressivity of ontology properties. In terms of quality, ontology targets were selected on the basis of metadata clarity in relation to construct descriptions and the availability of clear labels and comments. The final selection is shown in Table 6.4.

Table 6. 4: Selected Generic Ontology Targets

Resource	Link	Data Field Value
dbpedia-de	http://de.dbpedia.org/resource	disaggregated value
Dbp	http://dbpedia.org/property/theme	theme
Dbr	http://dbpedia.org/resource/	unique identifier
Sio	http://semanticscience.org/resource/	name and unit
Dterms	http://purl.org/dc/terms/	Relation
Rdfs	http://www.w3.org/2000/01/rdf-schema#	Labels
Afo	http://purl.allotrope.org/ontologies/property#	description and categorical
SDGIO	http://purl.unep.org/sdg/	SD Goal, UNSDG target ID, SD target, UNSDG Indicator ID, Sustainable Development Indicator, SDG Indicator Value

Table 6.4 is the final selection of suitable generic ontology targets comprising the most applicable literals in Table 6.3 (one target option per literal) selected on the basis of availability, coverage and quality factors.



6.3 Demonstration 2: The Development of the Custom Domain Ontology (NPIOnto)

This section presents the National Performance Indicator Ontology (NPIOnto). The NPIOnto knowledge model represents domain knowledge in the field of Botswana development agendas to address the lack of semantics for NPIs and relating performance goals and other measurements. It also represents semantic integration of NPIs for the development agendas that form the Botswana Integrated Indicator Framework, comprising the Vision Agenda, NDP Agenda, Africa Agenda, and domesticated SDG Agenda. The ontology development was guided by the METHONTOLOGY approach [117]. As detailed in Subsection 3.3.3, the METHONTOLOGY approach outlines a set of guidelines on how the activities identified in the ontology development process should be conducted. These activities have been implemented in five phases of the adapted METHONTOLOGY approach, outlined in Subsection 3.3.3, as follows: a) specification, b) knowledge acquisition, c) conceptualisation, d) formalisation, and e) evaluation.

6.3.1 Specification Phase

The specification phase involved the statement of the ontology purpose in terms of goal, scope and intended use. The *goal of NPIOnto* is to provide a solution to Botswana's OD representation needs regarding NPI disambiguation, disaggregation, interoperability, integrability, multidimensionality, and reasoning capabilities. The *scope of NPIOnto* confines it to the following national, regional and global development agendas: Vision Agenda, NDP Agenda, Domesticated SDG Agenda and Africa Agenda. Finally, the *intended use of NPIOnto* is described in terms of its potential applications. NPIOnto is part of the ODSE Framework demonstration in which it serves as the basis for the conversion of flat text OD resources to knowledge resources so as to drive the transition from flat text OD publications to semantic publications for semantic consumption. To enable semantic consumption, this research, proposes a shift toward semantic publication centered on the enrichment of raw OD with semantic structure, capabilities and context for increased knowledge creation and mining potential. This shift necessitates the development of semantic models to drive the transition from raw OD publication and consumption (Web



of Documents) to semantic publication and consumption (Web of Data). In this regard, the competency questions (CQs) for NPIOnto are as follows:

- **CQ1:** What are the accessibility capabilities of OD resources enriched with NPIOnto?
- **CQ2:** What are the understandability capabilities of OD resources enriched with NPIOnto?
- **CQ3:** What are the interoperability capabilities of OD resources enriched with NPIOnto?
- **CQ4:** What are the integrability capabilities of OD resources enriched with NPIOnto?
- **CQ5:** What are the intelligibility capabilities of OD resources enriched with NPIOnto?
- **CQ6:** What is the overall semantic capabilities enrichment impact on semantic publication and consumption?

6.3.2 Knowledge Acquisition Phase

The knowledge acquisition phase involved the collection and understanding of information from a wide range of sources key to the ontology domain. To ensure that NPIOnto represents a shared and agreed upon conceptual framework, information gathering for the development of NPIOnto was mainly based on the analysis of standardised and detailed information from official documentations for the four development agendas. The information was extracted from the following sources: 1) Africa Agenda 2063 Core Indicators Profile Handbook [8], 2) Botswana Domesticated SDGs Indicator list [10], 3) Global SDG Indicator list for SDG Agenda [9], 4) Vision 2036 document [6], 5) NDP 11 Volume 1 document [7], and (6) Sustainable Development Goals (SDGs) National Baseline Indicator Framework document [13]. The process started by identifying NPIs and then establishing an understanding of their characteristics and relations to each other and to performance measurement goals. Above that, disaggregation factors were identified for each NPI. Specifically, this phase was conducted according to the following steps, illustrated on the basis of Figure 4.12 of the Africa Agenda documentation extract on OD representation needs in Subsection 4.3.3:



- 1) Step 1: Identification of statements of interest from domain documentations – the statements of interest were identified in the flat text resources of domain documentations, i.e., for Figure 4.12, the sample statement reads, *Aspiration 6. An Africa whose development is people driven, relying on the potential of the African people.*
- 2) Step 2: Analysis of the statements to derive concepts and their relationships. This pertained to the identification of subjects (domain entities), predicates (action words – object properties) and objects (range entities). For the sample statement in Step 1) above, the subject *aspiration* has two predicates (*has code* and *has description*) and two objects (*aspiration 6* and *An Africa whose development is people driven, relying on the potential of the African people*).
- 3) Step 3: Identification of modelling specification for the concepts and their relationships from the metadata resources – for instance, for the identified subject, predicates and objects in Step 2) above, two RDF triple statements are modelled as, ‘*aspiration has code aspiration code*’ and ‘*aspiration has description aspiration description*’. The subject represent class, predicates represent data and object properties and objects represent data values for data properties and range class (relating class) for the domain class (subject class).
- 4) Step 4: Creation of flat text resources based on concepts per agenda performance variable. For Step 3) above, the dataset for the subject aspiration was created with the aspiration code and description subjects reflected in row 7, columns B and D of *Table AFR 5: Africa Agenda Aspiration Dataset Screenshot* in Appendix 1.

The output of this phase comprised collections of relevant development agenda domain knowledge on NPIs in flat text format described in detail in Subsection 6.2.1. These indicator datasets constitute the NPI knowledge base for the development agenda domain contextualised to the Botswana Integrated Indicator Framework specifications. The knowledge model (NPIOnto) was designed on the basis of the metadata fields (specified in Subsection 6.2.1) for these datasets and represented by the dataset screenshots attached to the appendices section, ranging from appendix 1-4.



6.3.3 Conceptualisation Phase

The conceptualisation phase pertained to the initial structuring of the knowledge model (schema). This preliminary NPIOnto modelling phase involved the schema design and development. The information collected in the previous stage (Subsection 6.3.2) was converted to the visual representation of NPIOnto, reflecting objects in the development agenda domain, and their attributes and relations. The first four deliverables of the conceptual phase are conceptual models for the four development agendas. The fifth main deliverable is a consolidated conceptual model that integrates the four conceptual models on the basis of shared classes, object properties and data properties being: 1) the Africa Agenda Indicator (AFRI) Knowledge Model, 2) the Sustainable Development Goals Indicator (SDGI) Knowledge Model, 3) the Vision Indicator (VISI) Knowledge Model, 4) the National Development Plan Indicator (NDPI) Knowledge Model, and 5) the Consolidated Knowledge Model (NPIOnto Lightweight Version).

Africa Agenda Indicator (AFRI) Knowledge Model

The *Africa Agenda Indicator (AFRI) Knowledge Model* represents the Africa Agenda Indicator, `afri`, in terms of relating performance variables. Class `afri` is at the core of the AFRI knowledge model (Figure 6.1).



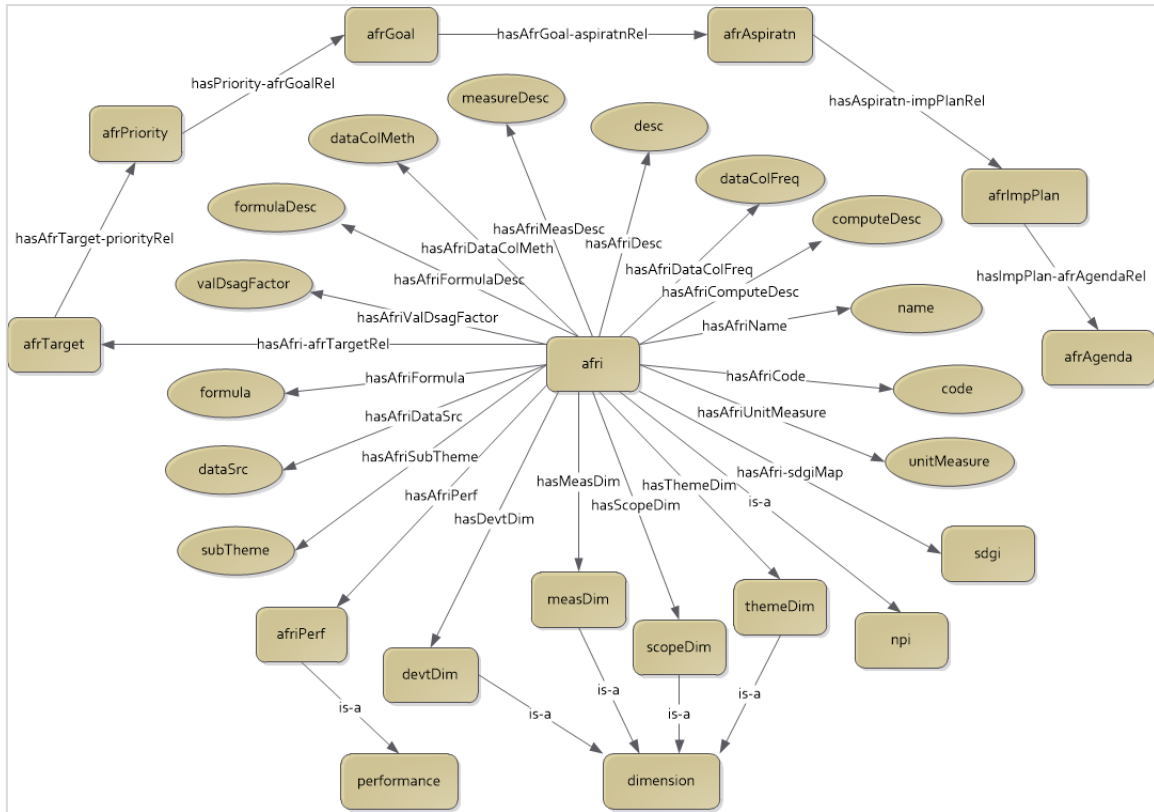


Figure 6. 1: Africa Agenda Indicator Model

Figure 6.1 presents class `afri` with an external relation to class `sdgi` through `hasAfri-sdgiMap` property. Internally, it has a direct relation with class `afriTarget` (target) through `hasAfri-afriTargetRel` property. It also reflects a series of relations building up from class `afri` relation to class `afriTarget` being:

- class `afriTarget` relation to class `afriPriority` (priority) through `hasAfriTarget-priorityRel` property,
- class `afriPriority` relation to class `afriGoal` (goal) through `hasPriority-afriGoalRel` property,
- class `afriGoal` relation to class `afriAspiratn` (aspiration) through `hasAfriGoal-aspiratnRel` property,
- class `afriAspiratn` relation to class `afriImpPlan` (implementation plan) through `hasAspiratn-impPlanRel` property, and
- class `afriImpPlan` relation to class `afriAgenda` (Africa Agenda) through `hasImpPlan-afriAgendaRel` property.

Sustainable Development Goals Indicator (SDGI) Knowledge Model

The Sustainable Development Goals Indicator (SDGI) Knowledge Model represents the Sustainable Development Goals Indicator, `sdgi`, in association with relating performance variables. At the core of the SDGI knowledge model is class `sdgi` (Figure 6.2).

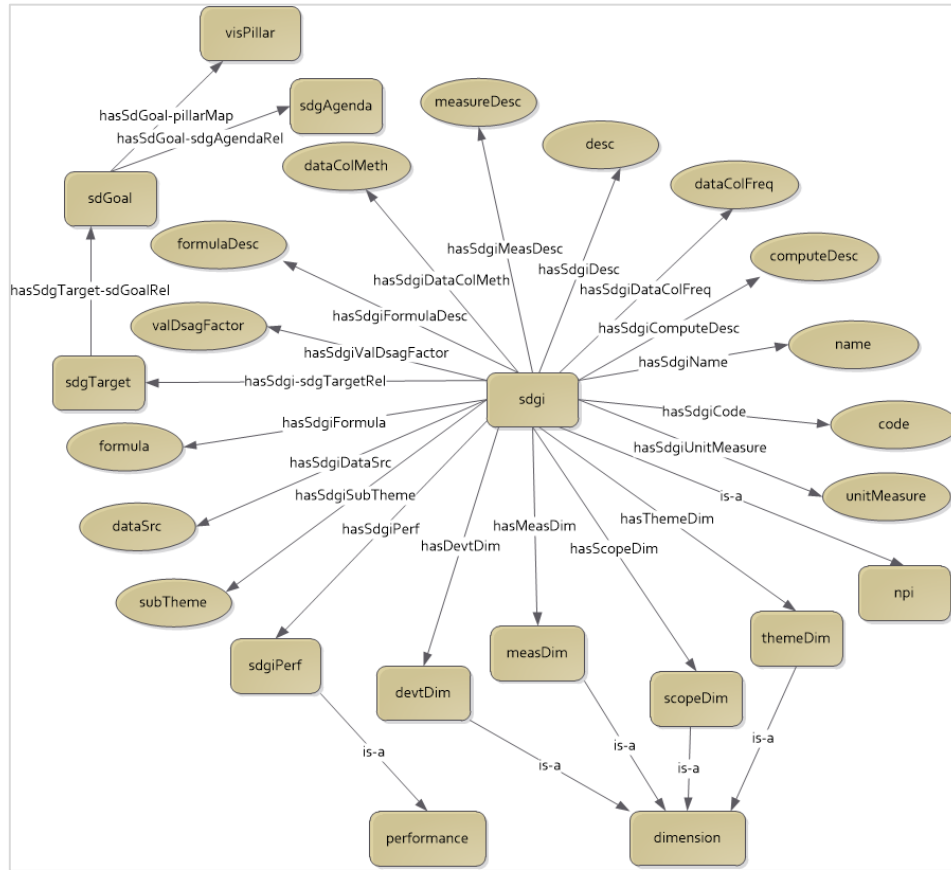


Figure 6. 2: Sustainable Development Goals Indicator Model

Figure 6.2 presents class `sdgi` with a direct relation to class `sdgTarget` through `hasSdgi-sdgTargetRel` property. It also reflects a series of relations building up from class `sdgi` relation to class `sdgTarget` being:

- class `sdgTarget` relation to class `sdGoal` (SDG) through `hasSdgiTarget-sdgGoalRel` property,
- class `sdGoal` relation to class `sdgAgenda` (SDG Agenda) through `hasSdGoal-sdgAgendaRel` property, and
- class `sdGoal` relation to class `visPillar` (Vision Pillar) through `hasSdGoal-pillarMap` property (external relation).



Vision Indicator (VISI) Knowledge Model

The *Vision Indicator (VISI) Knowledge Model* represents the Vision Indicator, `visi`, in connection with relating performance variables. Class `visi` is at the core of the VISI knowledge model (Figure 6.3).

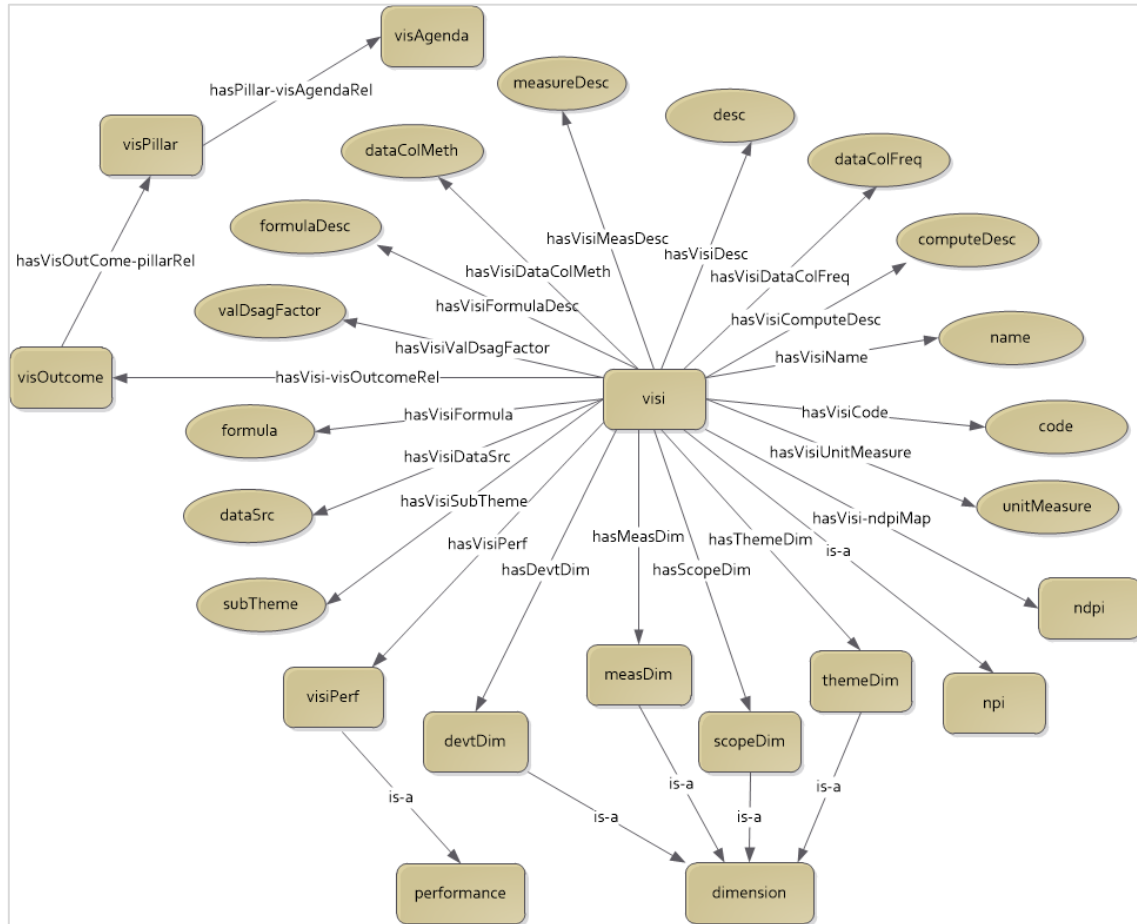


Figure 6.3: Vision Indicator Model

Figure 6.3 presents class `visi` with an external relation to class `ndpi` through `hasVisi-ndpiMap` property. Internally, it has a direct relation with class `visOutcome` (Vision Outcome) through `hasVisi-visOutcomeRel` property. It also reflects a series of relations building up from class `visi` relation to class `visOutcome` being:

- class `visOutcome` relation to class `visPillar` (Vision Pillar) through `hasVisiOutCome-pillarRel` property, and
- class `visPillar` relation to class `visAgenda` (Vision Agenda) through `hasPillar-visAgendaRel` property.



The *National Development Plan Indicator (NDPI) Knowledge Model* represents the National Development Plan Indicator, `ndpi`, in association with relating performance variables. At the core of the NDPI knowledge model is class `ndpi` (Figure 6.4).

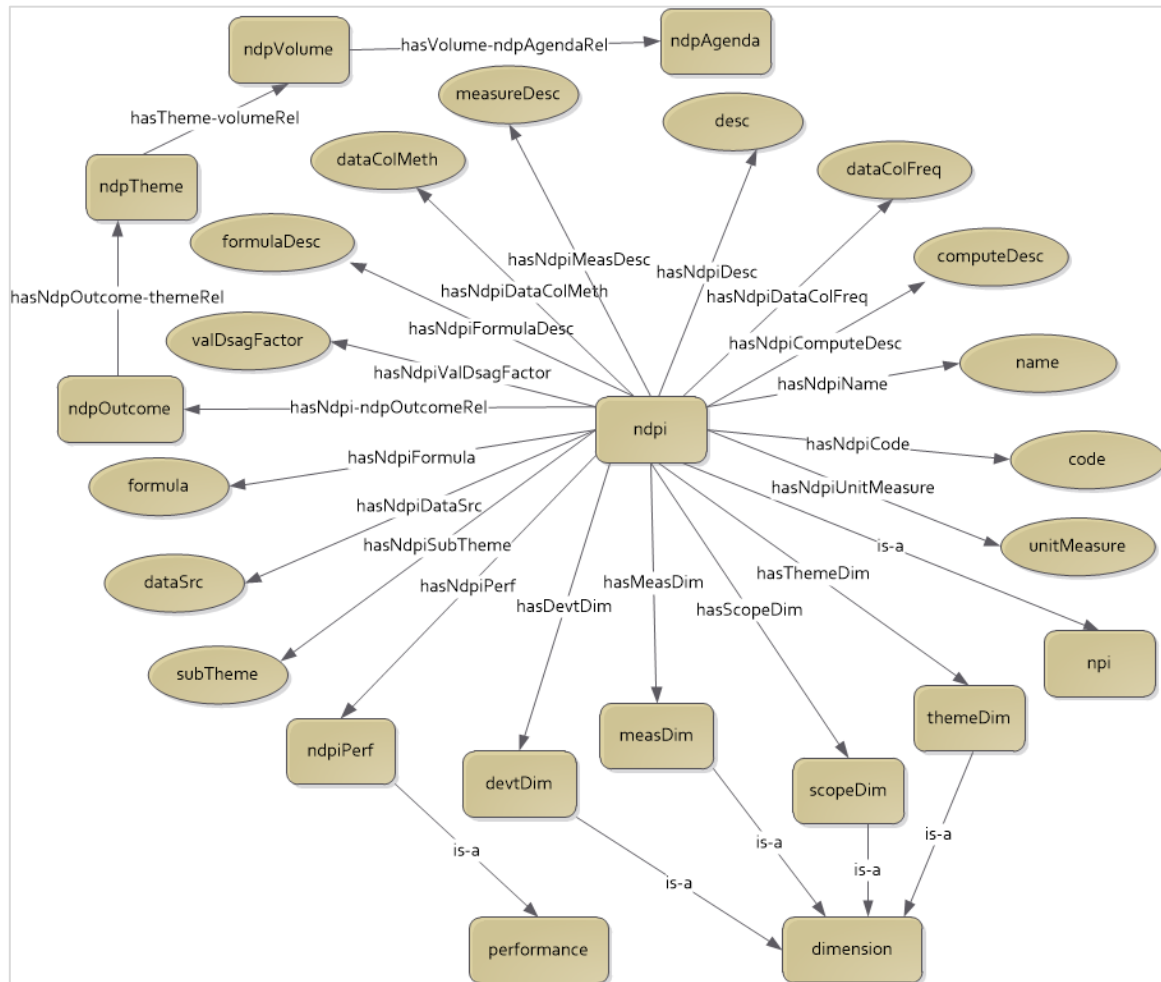


Figure 6.4 presents class `ndpi` with a direct relation to class `ndpOutcome` (NDP Outcome) through `hasNdpi-ndpOutcomeRel` property. It also reflects a series of relations building up from class `ndpi` relation to class `ndpOutcome` being:

- **class** `ndpOutcome` **relation to class** `ndpTheme` (**theme**) **through** `hasNdpOutcome-themeRel` **property**,
- **class** `ndpTheme` **relation to class** `ndpVolume` (**volume**) **through** `hasTheme-volumeRel` **property**, and



- class `ndpVolume` relation to class `ndpAgenda` (`ndp Agenda`) through `hasVolume-ndpAgendaRel` property.

In overall, the *Shared Classes and Properties* are as follows:

- all indicator classes (`afri`, `sdgi`, `visi`, and `ndpi`) share the following data properties: `code` (identity code), `name`, `desc` (description), `subtheme`, `unitMeasure` (unit of measurement), `measDesc` (measurement description), `computeDesc` (computation description), `formula`, `formulaDesc` (formula description), `dataSrc` (source of data), `dataColMeth` (data collection method), `dataColFreq` (data collection frequency), and `valDsagFactor` (value disaggregation factor);
- all indicator classes (`afri`, `sdgi`, `visi`, and `ndpi`) also share dimensional classes (`devtDim` - development type of dimensional attribute, `measDim` - measurement type of dimensional attribute, `scopeDim` - scope type of dimensional attribute, `themeDim` - thematic type of dimensional attribute) and their relating object properties (`hasDevtDim`, `hasMeasDim`, `hasScopeDim`, `hasThemeDim`);
- The four dimensional classes share *is-a* relation to superclass dimension. That is, they are all specialisations of the superclass dimension;
- All indicator classes (`afri`, `sdgi`, `visi`, and `ndpi`) share *is-a* relation to superclass `npi` (National Performance Indicator). That is, they are all specialisations of the superclass `npi`;
- All indicator performance classes (`afriPerf`, `sdgiPerf`, `visiPerf`, and `ndpiPerf`) share *is-a* relation to superclass performance. That is, they are all specialisations of the superclass performance.

It is on the basis of these shared classes, object properties and data properties that the consolidated model was drawn.

Consolidated Knowledge Model

Figure 6.5 is a detailed conceptual representation of the consolidated agenda models and their performance measurement constructs; as well as their relationship and dimensional attribute interlinkages within and across agendas as shown in Figure 6.5.



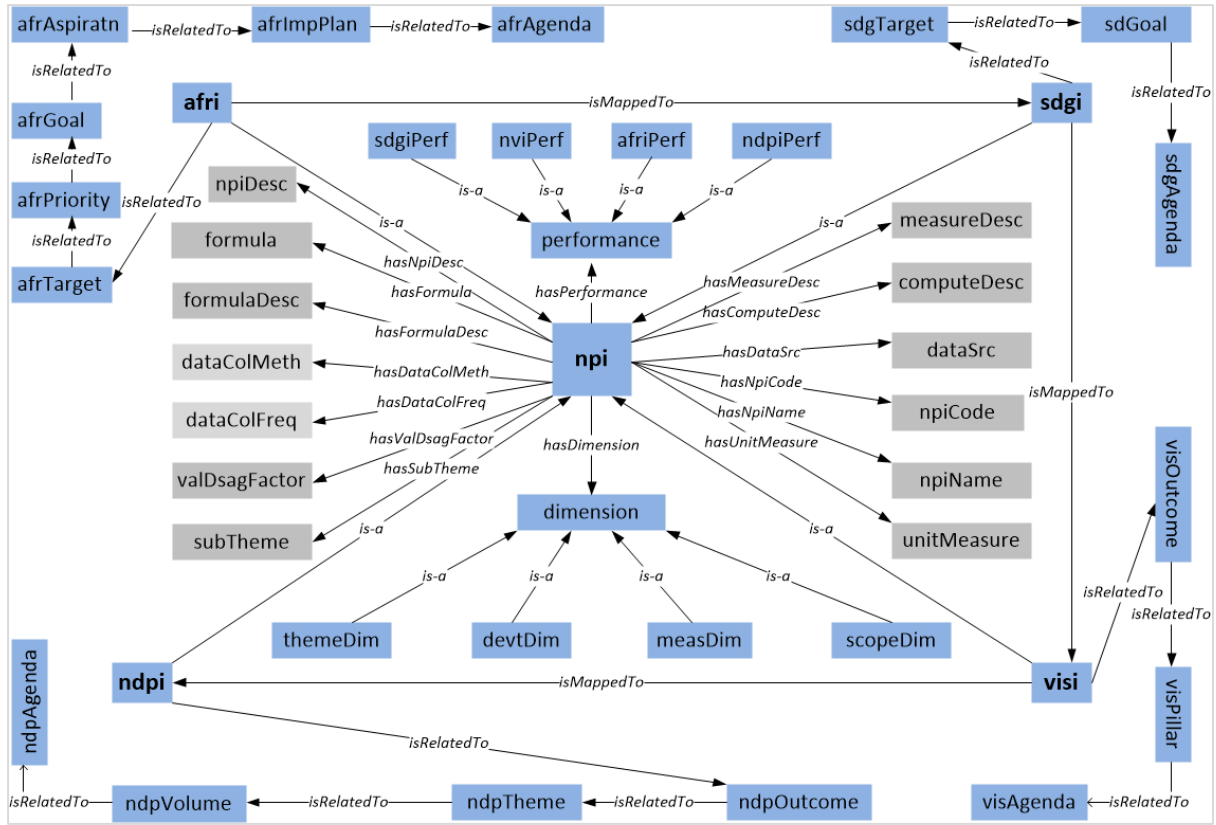


Figure 6. 5: NPIONto Conceptual Model [129]

As depicted in Figure 6.5, NPIONto conceptual model provides a snapshot of concepts, sub-concepts and relations in the development agenda domain within the Botswana Integrated Indicator Framework context. It is a conceptual model of the integration potential in Botswana’s development agenda domain. It provides a better understanding of NPIs underlying composition structures and other important performance measurement attributes. Furthermore, it represents relations between NPIs, as well as NPIs relations to performance goals and measures for enhanced NPIs comparison capabilities. Moreover, it served as the blueprint for the formalisation of NPIONto [129].

6.3.4 Formalisation Phase

The formalisation phase involved the implementation of the ontology in a formal language. This pertained to the codification, in Protégé OWL description logic expressions, of both the consolidated knowledge model/schema as shown in the first three columns of Figure 6.6 and the knowledge base as shown in the fourth column of Figure 6.6. The formalised NPIONto version is accessible through this figshare link,



[National Performance Indicator Ontology \(NPIOnto\)](#). OWL with its explicit formal semantics was used to capture knowledge on NPIs in a machine interpretable way. The following subsections describe NPIOnto formalisation.

Formalisation of the Knowledge Model (NPIOnto Schema)

The knowledge model is a semantic representation of the terminological knowledge which includes the terms used to describe data (classes/concepts); relations (facets/roles/slots/properties); restrictions; rules; and axioms [130]. Thus, the formalised NPIOnto knowledge model shown in Figure 6.6 comprises classes (superclasses and subclasses) and relating object and data properties (superproperties and subproperties).

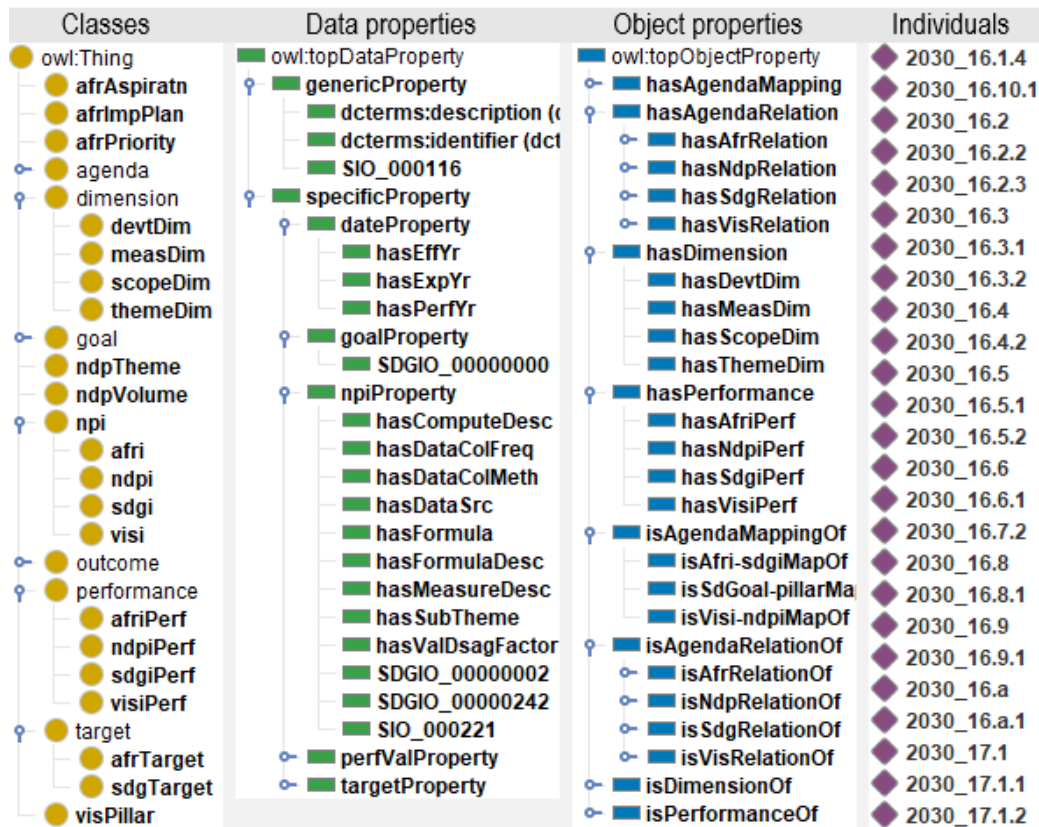


Figure 6. 6: Formalised NPIOnto Knowledge Model Sample Extract

The first column in Figure 6.6 is a hierarchical presentation of NPIOnto classes organised into a superclass-subclass hierarchy/taxonomy in which subclasses specialise or are subsumed by their superclasses. The second column of Figure 6.6, data properties, are also organised into a superproperty-subproperty hierarchy. Data properties describe



relations between a class instance and its data value. That is, they specify the data attributes and the corresponding data types. Object properties (third column of Figure 6.6) specify relations between classes and these relations are further inherited by class instances. That is, object properties describe relationships between instances of two classes. Object properties link instances in the domain class to instances in the range class. NPIOnto object properties are also organised in a taxonomy. The last set of NPI definitions are the annotation properties (sample shown in Figure 6.7) used to augment additional information or metadata to classes, individuals and object/data properties.

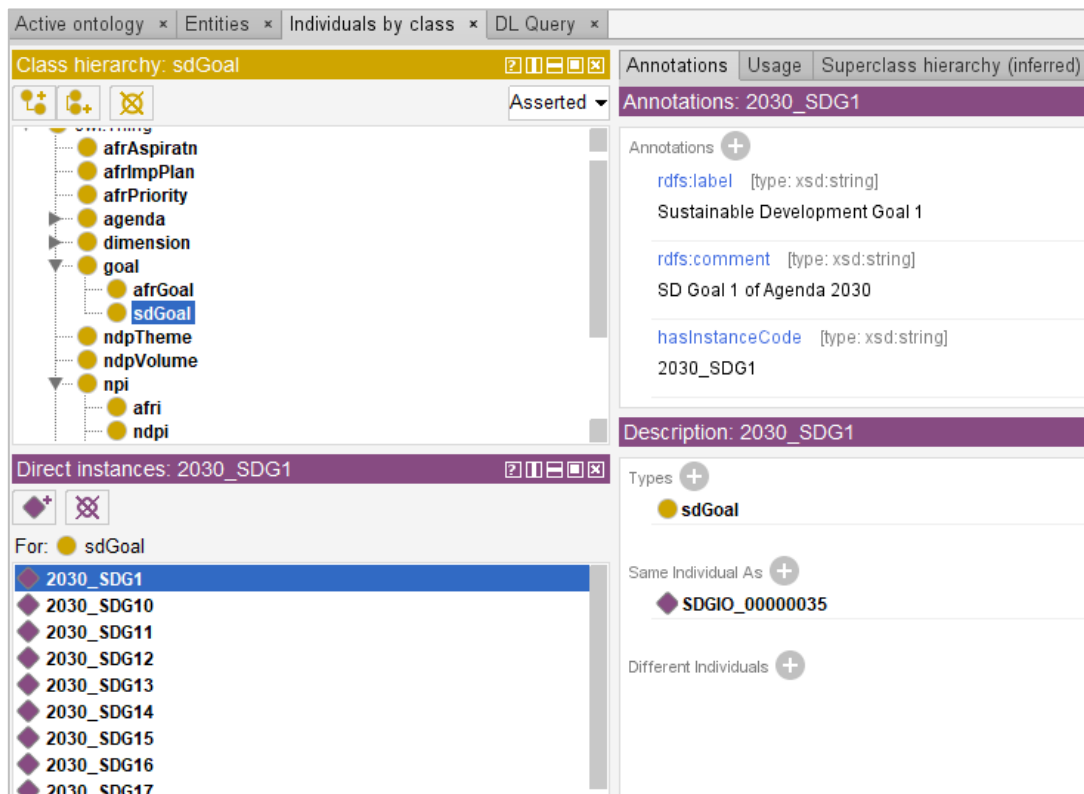


Figure 6. 7: NPI Annotation Properties Extract (SDG 1)

As reflected in Figure 6.7, all individuals, data and object properties use the `rdfs:label` and `rdfs:comment` annotation properties. The third annotation property (`hasInstanceCode`) is a custom property that identifies each unique instance. There is also the `dcterms:RightsStatement` property that has been used to define the openness of NPIOnto for free use, reuse and redistribution (the creative commons license).



Overall, the formalised NPIOnto model represents NPI definitions as a combination of NPI concepts/classes and their applicable data, object, and annotation properties. NPIOnto concepts and corresponding properties have been modelled to address the semantic representation needs in the Botswana development agenda domain that pertain to NPI disambiguation, disaggregation, integration, interoperability, multidimensionality, and reasoning capabilities. The modelling of each of these semantic capabilities is discussed in the following subsections.

NPI Disambiguation Capability Modelling

The superclass `npi` represents a quantitative measure for evaluating development agenda performance. Firstly, the `npi` definition encompasses the attribute descriptive semantics being `hasNpiCode` (unique identifier), `hasNpiName`, `hasNpiDesc` (description), and `hasNpiSubTheme`. Secondly, it encompasses the measurement descriptive semantics being `hasUnitMeas` (unit of measurement), `hasMeasureDesc` (measurement description), `hasComputeDesc` (computation description), `hasFormula`, and `hasFormulaDesc` (formula description). Thirdly, it encompasses the data descriptive semantics being `hasDataSrc` (data source), `hasDataColMeth` (data collection method), `hasDataColFreq` (data collection frequency), and `hasValDsagFactor` (value disaggregation factor). This provides `npi` clarity in terms of its make up (attributes), measurement and data composition. For contextual clarity all data properties are further clarified with the `xsd:string` data type which is further disambiguated for specific properties as follows:

- `npi code` property is further clarified with links to `dcterms:identifier` and `SDGIO:00000242` properties,
- all name properties are further clarified with links to `dcterms:title` and `SIO:000116` properties,
- All description properties are further clarified with links to `dcterms:description` and `AFX:0002802` properties. The specific description for `npi` property description is `SDGIO:00000002`,
- The sub-theme property is further clarified with links to `dcterms:theme`,
- The unit of measurement property is further clarified with links to `SIO_000221`,



- The value disaggregation property is further clarified with links to AFX:0002801,
- The data source property is further clarified with links to dterms:source, and
- The data collection frequency property is further clarified with links to dterms:Frequency.

NPI Disaggregation Capability Modelling

The model implements a lower-level data disaggregation as depicted in Figure 6.8.

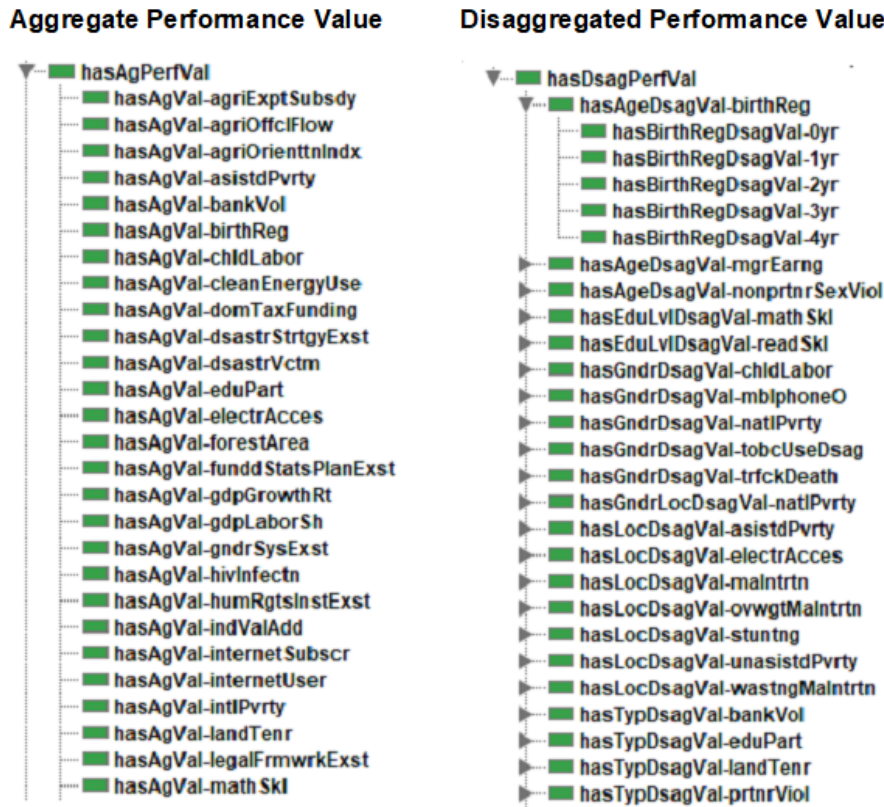


Figure 6. 8: NPI Disaggregation Extract (SDGI 16.9.1)

Figure 6.8 defines additional data properties that provide capabilities to express npi values in aggregate terms (`hasAgPerfVal`) and to also disaggregate the values to the lowest levels of granularity specified per agenda (`hasDsagPerfVal`). For instance, on the left column, `hasAgVal-birthReg` (birth registration aggregate performance value) is disaggregated by age (`hasAgeDsagVal-birthReg`). The specification of the disaggregated values are as follows (right column):

- `hasBirthRegDsagVal-0yr` (the less than one year category),
- `hasBirthRegDsagVal-1yr` (the less than two years category),



- `hasBirthRegDsagVal-2yr` (the less than three years category),
- `hasBirthRegDsagVal-3yr` (the less than four years category), and
- `hasBirthRegDsagVal-4yr` (the less than five years category).

NPI Interoperability Capability Modelling

The formalised NPI ontology models NPI interoperability in terms of shared data properties for shared classes introduced in Subsection 6.3.3. The shared classes and their shared data properties enhance interoperability by providing a common data representation format which preserves the data context and meaning as content is shared across agendas. The formalised NPI interoperability capability modelling is as follows:

- superclass `npi` (`hasNpiAtt`) data properties are shared between the `afri`, `sdgi`, `visi`, and `ndpi` subclasses,
- superclass `agenda` (`hasAgendaAtt`) data properties are shared between `afrAgenda`, `sdgAgenda`, `visAgenda`, and `ndpAgenda` subclasses,
- superclass `performance` (`hasPerfAtt`) data properties are shared between `afriPerf`, `sdgiPerf`, `visiPerf`, and `ndpiPerf` subclasses,
- superclass `dimension` (`hasDimensionAtt`) data properties are shared between `devtDim`, `measDim`, `scopeDim`, and `themeDim` subproperties,
- superclass `goal` (`hasGoalAtt`) data properties are shared between `afrGoal` and `sdGoal` subclasses,
- superclass `outcome` (`hasOutcomeAtt`) data properties are shared between `ndpOutcome` and `visOutcome` subclasses,
- superclass `target` (`hasTargetAtt`) data properties are shared between `afrTarget` and `sdgTarget`, and
- superclass `date` (`hasDateAtt`) data properties are shared between `agenda` and `afrImpPlan` classes.

These semantic interoperability promotes reusability of data properties across the Botswana Integrated Indicator Framework domain.

NPI Integration Capability Modelling

The formalisation of the integration capability involved the creation of linkages at both the individual and consolidated agenda levels based on shared object properties as



reflected in the third column of Figure 6.6 (Subsection 6.3.4). At the individual level for Africa Agenda, semantic integration is through object properties as follows:

- hasAfri-afrTargetRel object property links afri to afrTarget instances,
- hasAfrTarget-priorityRel object property links afrTarget to afrPriority instances,
- hasPriority-afrGoalRel object property links afrPriority to afrGoal instances,
- hasAfrGoal-aspiratnRel object property links afrGoal to afrAspiratn instances,
- hasAspiratn-impPlanRel object property links afrAspiratn to afrImpPlan instances, and
- hasImpPlan-afrAgendaRel object property links afrImpPlan to afrAgenda instances.

At the individual level for SDG Agenda, semantic integration is through object properties as follows:

- hasSdgi-sdgTargetRel object property links sdgi to sdgTarget instances;
- hasSdgTarget-sdGoalRel object property links sdgTarget to sdGoal instances; and
- hasSdGoal-sdgAgendaRel object property links sdGoal to sdgAgenda instances.

At the individual level for Vision Agenda, semantic integration is through object properties as follows:

- hasVisi-visOutcomeRel object property links visi to visOutcome instances,
- hasVisOutcome-pillarRel object property links visOutcome to visPillar instances, and
- hasPillar-visAgendaRel object property links visPillar to visAgenda instances.

At the individual level for NDP Agenda, semantic integration is through object properties as follows:

- hasNdpi-ndpOutcomeRel object property links ndpi to ndpOutcome instances,



- hasNdpOutcome-themeRel object property links ndpOutcome to ndpTheme instances,
- hasTheme-volumeRel object property links ndpTheme to ndpVolume instances, and
- hasVolume-ndpAgendaRel object property links ndpVolume to ndpAgenda instances.

At the consolidated agenda level, integration across agendas are as follows:

- hasAfri-sdgiMap object property links afri to sdgi instances,
- hasSdGoal-pillarMap object property links sdGoal to visPillar instances, and
- hasVisi-ndpiMap object property links visi to ndpi instances.

In addition, the superclass constructs (first column of Figure 6.6) integrate performance at the consolidated agenda level using subsumption relations as follows:

- agenda concept integrates the four agendas (Africa Agenda - afrAgenda, Vision Agenda - visAgenda, SDG Agenda - sdgAgenda, and NDP Agenda - ndpAgenda),
- goal concept integrates the development goals for the two agendas (Africa Agenda - afrGoal, and SDG Agenda - sdgGoal),
- outcome concept integrates the development outcomes for the two agendas (Vision Agenda - visOutcome, and NDP Outcome - ndpOutcome),
- target concept integrates the development targets for the two agendas (Africa Agenda - afrTarget, and SDG Agenda - sdgTarget),
- performance concept integrates performance values for four agendas (Africa Agenda performance - afriPerf instance, NDP performance - ndpiPerf, SDG performance - sdgiPerf, and Vision performance - visiPerf),
- dimension concept integrates the four dimensions (development dimension - devtDim, measurement type dimension - measDim, scope dimension - scopeDim, and theme dimension - themeDim), and
- npi concept (the main concept at the core of NPIOnto) further integrates all the consolidated indicator concepts (afri, sdgi, visi, ndpi, performance, and dimension) into a single unified representation. This integration is using subsumption (superclass-subclass or inheritance or is-a) relations for the afri,



`sdgi`, `visi`, `ndpi` concepts while the `hasPerformance` superproperty is used for the performance concept and the `hasDimension` superproperty for the dimension concept.

NPI Multidimensionality Capability Modelling

The multidimensional capability modelling is represented by the relation between the `npi` superclass. The dimension superclass represents the different perspectives of viewing, analysing, grouping, and querying data on NPI and relating performance measures. It has four subclasses being:

- `scopeDim` (an NPI can be applicable at the national, regional or global level),
- `devtDim` (an NPI can either address an economic, social, or an environmental issue),
- `themeDim` (an NPI relates to a thematic area which can either be an Africa Agenda, SDG Agenda, Vision Agenda, or NDP Agenda theme), and
- `measDim` (an NPI can either be a base, aggregate, or derived measure).

NPI Reasoning Capability Modelling

The reasoner makes inferences across the body of knowledge represented in an ontology [131]. The reasoning and logical inferencing is enabled by the following ontology specifications: class and property hierarchy (inheritance); property domain and range; relations between classes (subsumption, disjointness, equivalence, union, intersection); cardinality (minimum, maximum, exact number); equality (same as); richer typing of properties (object vs. datatype, specific datatypes); and characteristics of properties / special properties (transitive, symmetric, functional, inverse functional) [132]. Thus, this section discusses some of these reasoning features as applied in `NPIOnto`.

Firstly, *NPIOnto* subsumption relations pertain to superclass-subclass relations and the superproperty-subproperty relations as demonstrated in column 1 of Figure 6.6.



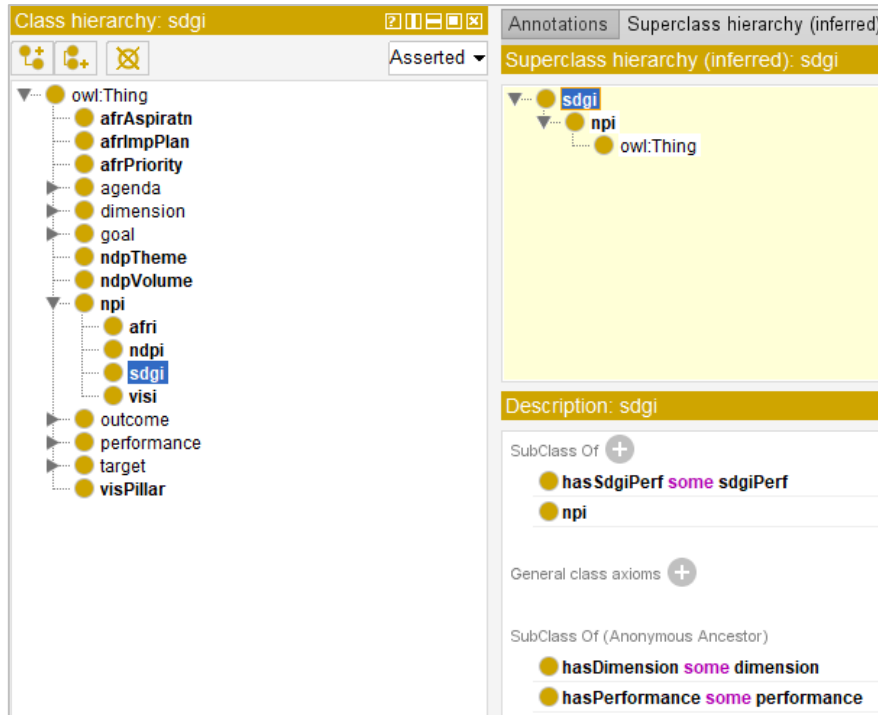


Figure 6. 9: *Fragment of Reasoner Inference for SDGI*

These relations enable subsumption reasoning over subclass and subproperty relations, and can be automatically computed by a reasoner as demonstrated in Figure 6.9. Executing the `sdgi` class query for superclasses automatically retrieves the direct superclass (`npi`) and the root superclass being `owl:Thing` (upper part of column 2). Based on the explicit definitions, the reasoner infers that class `sdgi` is a direct subclass of class `npi` as reflected in the left column and is also directly related to class `sdgiPerf` through a specific property link, `hasSdgiPerf`. It also infers that `sdgi` has indirect (ancestor) relationships to class `dimension` through `hasDimension` property and to class `performance` through `hasPerformance` property. That is, the two are directly related to `npi` and the reasoner deduces the subclass relations on the basis of the superclass relations.

Secondly, *NPIOnto classes' disjointness* involves a separation of a group of classes from one another to ensure that an individual which has been asserted to be a member of one of the classes in the group cannot be a member of any other class in that group. As there are no overlapping classes for NPIOnto, for effective reasoning capabilities, the first set of disjoint classes encompasses the six superclasses being `agenda`, `performance`,



dimension, goal, outcome and target. The second set of disjoint classes are all subclasses of the six superclasses to enforce that a superclass instance belongs to only one subclass. The third set of disjoint classes encompasses all other subclasses of the owl:Thing class being afrImpPlan, afrAspiratn, afrPriority, ndpTheme, ndpVolume, and visPillar to ensure that an individual (or object) cannot be an instance of more than one of these subclasses. The npi superclass is excluded from all sets of disjoint classes as its subclasses have equivalent mappings, i.e., *AFRI 1* (A1-G1-P1-T1-I1) is equivalent to *SDGI 8.1.1*.

Thirdly, *NPIOnto inverse object properties* enable effective reasoning capabilities, the definition of NPIOnto object properties includes also their inverses for more flexible and well-grounded inferences as reflected in the object properties' section (Figure 6.6). The inverse properties represent the swapped over property domain(s) and range(s) for object properties:

- isAgendaMappingOf is the inverse of hasAgendaMapping,
- isAgendaRelationOf is the inverse of hasAgendaRelation,
- isDimensionOf is the inverse of hasDimension (Figure 6.10), and
- isPerformanceOf is the inverse of hasPerformance.

Fourthly, *NPIOnto transitive object properties* include all superproperties to enable inheritance of the superproperties by the subproperties. If a property is transitive, and the property *P* relates individual *X* to individual *Y*, and also individual *Y* to individual *Z*, then it can be inferred that individual *X* is related to individual *Z* via property *P*.



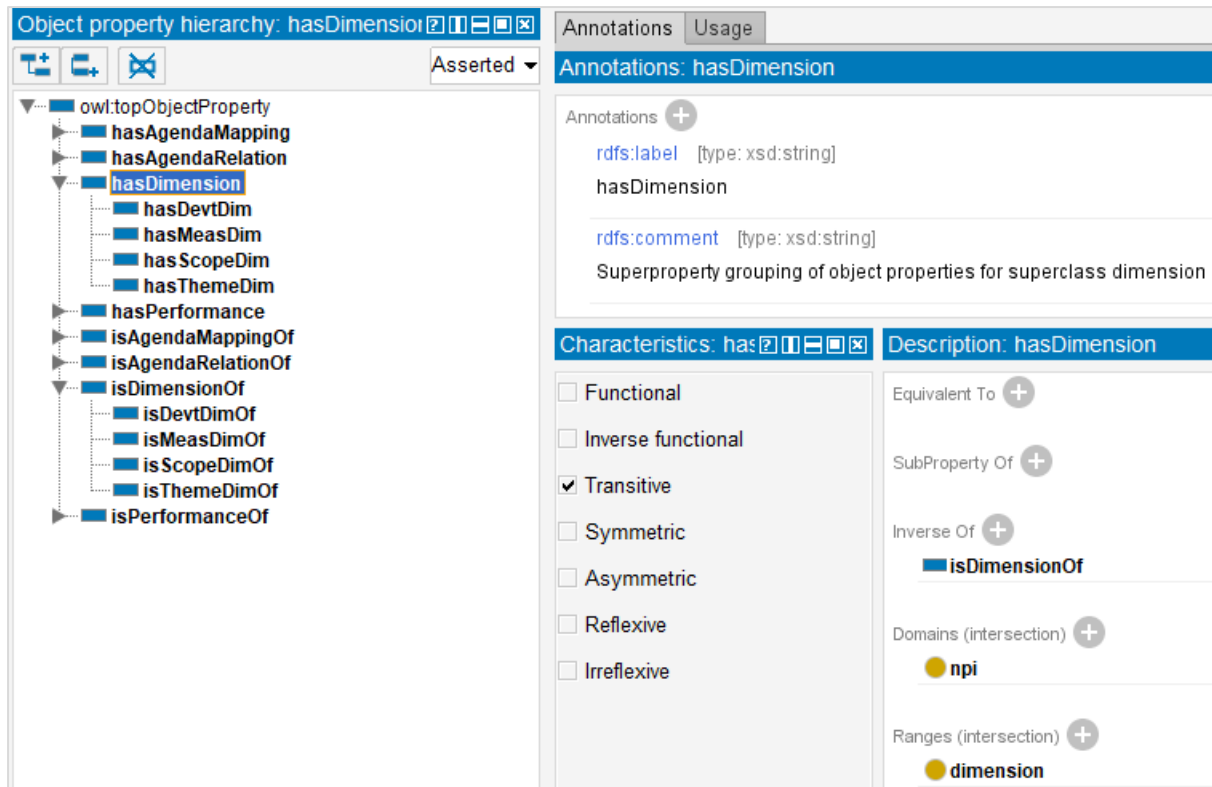


Figure 6. 10: NPIOnto Inverse and Transitive Object Property Extract

As reflected in Figure 6.10, if `hasDimension` property relates AFRI A to economic development dimension, and `hasDimension` property also relates the economic development dimension to SDGI S, then it can be inferred that AFRI A is related to SDGI S through the `hasDimension` property. That is, the two indicators from two different agendas are related in the sense that they share the same property. Therefore, they are interoperable and integrable through the shared property.

Fifthly, *NPIOnto domain and range restrictions* pertain to NPIOnto properties. The domain and range restrict the possible values (instances) of subject and object of a given property to members of a specific class or type. The domain restricts what kinds of resources or members of a class can be the subject of a given property in an RDF triple while the range restricts what kinds of resources/members of a class or data types (literals) can be the object of a given property in an RDF triple [132]. For instance, Figure 6.10 specifies the potential instance relations (inverse) for `isDimensionOf` property as pertaining to members of the two classes, `npi` and `dimension`.



Sixthly, *NPIOnto* quantifier restrictions definitions are mainly existential restrictions. An existential restriction is a logical restriction that applies to a class of individuals that have at least one given property relationship to individuals that are members of the class being restricted. The sample restrictions reflected in the second column of Figure 6.9 are summed up by the description logic expressions:

$$\exists \text{ hasDimension dimension} \quad (1)$$

$$\exists \text{ hasPerformance performance} \quad (2)$$

In these expressions *hasDimension* and *hasPerformance* are properties, $\exists$ is the quantifier for *at least one*, and class *dimension* and class *performance* are fillers. The first expression (Eq. (1)) describes a class of individuals that have at least one *hasDimension* relation to individuals that are members of the class *dimension*. The second expression (Eq. (2)) describes a class of individuals that have at least one *hasPerformance* relation to individuals that are members of the class *performance*. As performance differs per agenda, the second DL expression relations are further constrained to a specific agenda (Eqs. (3) and (4)).

$$\exists \text{ hasAfriPerf afriPerf} \quad (3)$$

In Eq. (3) the subclass *afri performance* is constrained to subclass *afriPerf* (*afri performance*) instance relations through subproperty *hasAfriPerf*.

$$\exists \text{ hasSdgiPerf sdgiPerf} \quad (4)$$

In Eq. (4) the subclass *sdgi performance* is constrained to subclass *sdgiPerf* (*sdgi performance*) instance relations through subproperty *hasSdgiPerf*.

$$\exists \text{ hasVisiPerf visiPerf} \quad (5)$$

In Eq. (5) the subclass *visi performance* is constrained to subclass *visiPerf* (*visi performance*) instance relations through subproperty *hasVisiPerf*.

$$\exists \text{ hasNdpiPerf ndpiPerf} \quad (6)$$

In Eq. (6) the subclass *ndpi performance* is constrained to subclass *ndpiPerf* (*ndpi performance*) instance relations through subproperty *hasNdpiPerf*.

These restrictions constrain the relations between concepts to specific domain entities for improved machine understandability of concepts relations.



Formalisation of the Knowledge Base (NPIOnto Knowledge Base)

The knowledge base comprises the knowledge model (classes, and data and object properties); individuals or instances of classes; and ground facts asserting the situation between these individuals (data property assertions). As depicted in Figure 6.11, individuals as specific entities in the universe of discourse, are tied to the knowledge model. That is, individuals conform to the schema structure.

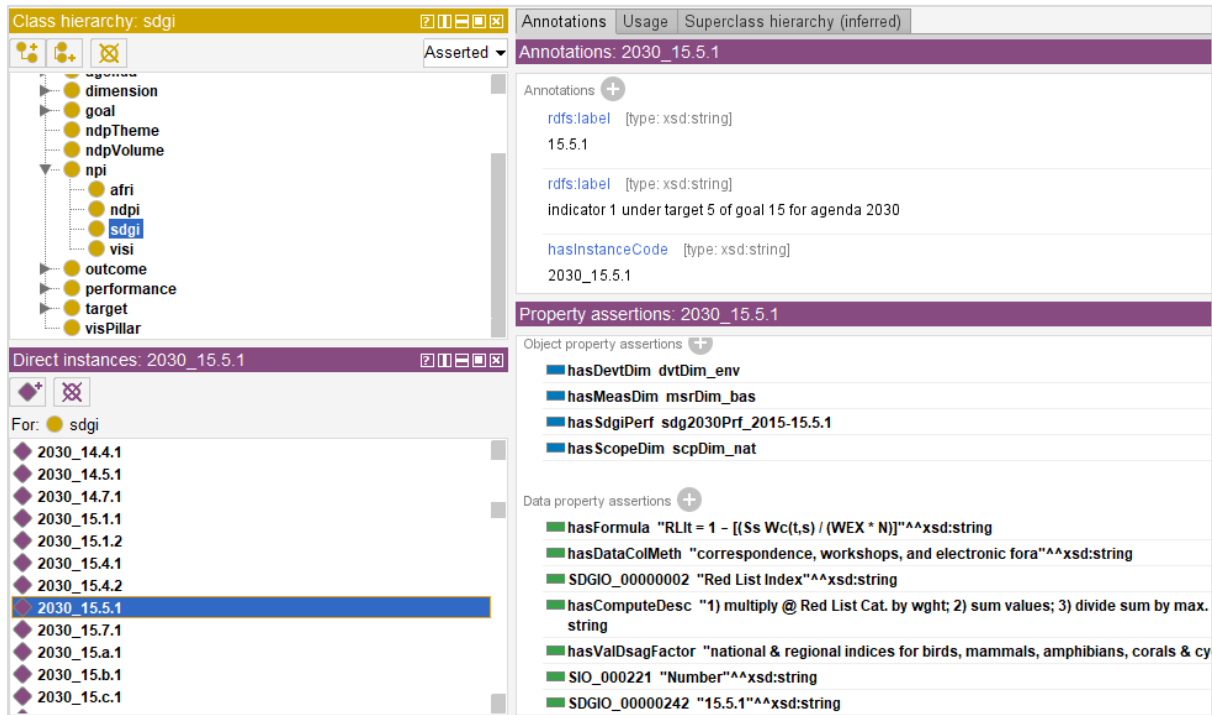


Figure 6. 11: Fragment of NPIOnto knowledge base Instance for SDGI 15.5.1

Figure 6.11 depicts the creation and description of individuals in terms of subclass *sdgi* and relevant annotation, data and object property assertions. Also, the NPIOnto knowledge base adopts the unique performance indicator reference codes to serve as individual names. As OWL does not use the Unique Name Assumption for individual names, this serves to avoid a situation where an individual could be associated with numerous diverse names.

Overall, Table 6.5 displays the NPIOnto summary extract in the form of metrics derived from the Protégé development environment.



Table 6. 5: Summary of NPIOnto Knowledge base Matrix Extract

Metrics		Data property axioms	
Axiom	10158	SubDataPropertyOf	199
Logical axiom count	6662	EquivalentDataProperties	0
Declaration axioms count	1061	DisjointDataProperties	0
Class count	35	FunctionalDataProperty	0
Object property count	70	DataPropertyDomain	230
Data property count	201	DataPropertyRange	207
Individual count	742	Individual axioms	
Annotation Property count	5	ClassAssertion	613
Class axioms		ObjectPropertyAssertion	1142
SubClassOf	28	DataPropertyAssertion	3780
EquivalentClasses	0	NegativeObjectPropertyAssertion	0
DisjointClasses	8	NegativeDataPropertyAssertion	0
GCI count	0	SameIndividual	172
Hidden GCI Count	0	DifferentIndividuals	0
Object property axioms		Annotation axioms	
SubObjectPropertyOf	62	AnnotationAssertion	2435
EquivalentObjectProperties	0	AnnotationPropertyDomain	0
InverseObjectProperties	35	AnnotationPropertyRangeOf	0
DisjointObjectProperties	0		
FunctionalObjectProperty	27		
InverseFunctionalObjectProperty	27		
TransitiveObjectProperty	16		
SymmetricObjectProperty	0		
AsymmetricObjectProperty	0		
ReflexiveObjectProperty	0		
IrreflexiveObjectProperty	0		
ObjectPropertyDomain	58		
ObjectPropertyRange	58		

These values are the basis for the computations of evaluation metrics for both NPIOnto knowledge model and knowledge base presented in the next section, Section 6.3.5.

6.3.5 Evaluation Phase

The evaluation phase pertains to the assessment of the effectiveness of ontologies [117]. The design effectiveness evaluation was conducted to establish the effectiveness of the ontology design with the use of metrics to evaluate both the knowledge model (schema) and the knowledge base (instances) designs. The Knowledge model metrics were used to evaluate ontology design and its potential for rich knowledge representation in terms of Attribute, Inheritance and Relationship Richness, as well as the Axiom/Class Ratio [133], [134]. The knowledge base metrics were used to evaluate the placement of instance data



within the ontology (Average Population) and the distribution of instances across classes (Class Richness) [134], [133].

Firstly, *the knowledge model (schema) metrics* used to evaluate NPIOnto knowledge model are Attribute Richness, Inheritance Richness, Relationship Richness, and Axiom/Class Ratio [133]. Attribute Richness measures the amount of information conveyed by the ontology. As a measure of ontology design quality, this measure is determined by the number of attributes (slots) per class which gives the amount of information pertaining to instance data [133]. The Attribute Richness for NPIOnto was computed using Eq. (7) adopted from [134]:

$$\text{Attribute Richness} = \frac{|ATT|}{|C|}, \quad (7)$$

Where ATT is the number of class attributes, and C is the number of classes.

Inheritance Richness measures the distribution of information across different levels of the ontology's inheritance tree or the fan-out of parent classes. An ontology with a low Inheritance Richness would be of a deep (or vertical) ontology which indicates that the ontology covers a specific domain in detail. An ontology with a high Inheritance Richness would be a shallow (or horizontal) ontology which indicates that the ontology represents a wide range of general knowledge with a low level of detail. The Inheritance Richness of the schema is defined as the average number of subclasses per class [133]. The Inheritance Richness for NPIOnto was calculated using Eq. (8) adopted from [133]:

$$\text{Inheritance Richness} = \sum C_i \in C \frac{|H^c(C1, C_i)|}{C}, \quad (8)$$

Where the number of subclasses of a class is $H^c(C1, C_i) = \text{no of subclasses of class } C_i$, and H is the number of inheritance relationships.

Relationship Richness measures the diversity of the types of relations in the ontology. An ontology that contains only inheritance relationships conveys less information than an ontology that contains a diverse set of relationships. Relationship Richness is represented as the percentage of the non-inheritance relationships between classes compared to all the possible connections that can include inheritance and non-inheritance relationships. The Relationship Richness for NPIOnto was calculated using Eq. (9) adopted from [133]:

$$\text{Relationship Richness} = \frac{P}{|ISA+P|}, \quad (9)$$



Where P is a number of non-Is-A relationships, and IsA is a number of Is-A relationships. Axiom/Class Ratio measures the ratio between axioms and classes [133]. This metric represents the relation between the classes containing attributes and all classes. As axioms are ontology statements that serve as the basis for reasoning, a positive high value indicates that the ontology is a rich source of reasoning. The Axiom/Class Ratio for NPIOnto was calculated using Eq. (10) adopted from [133]:

$$Axiom/Class\ Ratio = \frac{Axioms}{Classes} \quad (10)$$

The calculated values of the knowledge model metrics, as well as their interpretation, are found in Table 6.6.

Secondly, the knowledge base metrics used to evaluate NPIOnto knowledge base are Average Population and Class Richness [133]. Average Population measures the average distribution of instances across all classes. This measure is an indication of the number of instances compared to the number of classes. The Average Population of NPIOnto knowledge base was calculated using Eq. (11) adopted from [133]:

$$Average\ Population = \frac{|I|}{|C|}, \quad (11)$$

Where I is the number of instances in the knowledge base, and C is the number of classes defined in the ontology schema.

Class Richness measures how instances are distributed across classes. The number of classes that have instances in the knowledge base is compared with the total number of classes. If the knowledge base has a very low Class Richness, then it does not have data that exemplifies all the class knowledge that exists in the schema. On the other hand, a knowledge base that has a very high Class Richness would indicate that the data in the knowledge base represents most of the knowledge in the schema [133]. The Class Richness of NPIOnto knowledge base was calculated using Eq. (12) adopted from [133]:

$$Class\ Richness = \frac{|C'|}{|C|}, \quad (12)$$

Where C' is the number of non-empty classes, and C is the number of classes defined in the ontology schema.



The calculated values of the knowledge base metrics, as well as their interpretations are shown in Table 6.6 which provides the summary of NPIOnto evaluations using metrics displayed in Table 6.5 and described in the above section. The values in Table 6.6 were computed as follows. The Attribute Richness for NPIOnto was calculated using Eq. (7). The attribute values for NPIOnto are as follows (Table 6.5): 201 for data properties, 70 for object properties, 5 for annotation properties, and 35 for classes. This gives the Attribute Richness value of 7.89. On a scale of 8 for Attribute Richness, the value of 7.89 for slot count represents a very high amount of NPI domain knowledge conveyed by NPIOnto.

According to the calculation by Eq. (8) with values of the number of subclasses of a class and the number of inheritance relationships being 23 and 35 respectively (Table 6.5), the Inheritance Richness value for NPIOnto is equal to 0.8. The value of 0.8 on a scale of 8 indicates a very deep level of domain representation. Thus, NPIOnto covers the development agenda domain in detail.

Table 6. 6: Summary of NPIOnto Knowledge Model and Knowledge base Evaluations

S/N	Metrics	NPIOnto Values	Value Interpretation
1.	Attribute Richness	8.0	Very high value. Represents a very high amount of NPI domain knowledge conveyed by NPIOnto
2.	Inheritance Richness	1.06	Low value. Represents a deep (or vertical) ontology which indicates that NPIOnto covers the Development Agenda domain in detail
3.	Relationship Richness	0.66	Low value. Minimal diversity of relationships. Most of the relationships are inheritance relationships, thus NPIOnto conveys less information than an ontology that contains a diverse set of relationships
4.	Axiom/Class Ratio	287.4	Positive high value. The NPIOnto is a rich source of reasoning
5.	Average Population	21.2	Positive low value. Reflects an insignificant presence of instances in NPIOnto knowledge base insufficient to represent all the knowledge in the schema due to less publications
6.	Class Richness	0.6	Low value. Reflects the limited amount of information available online. Most classes are either empty or have less instances

The Relationship Richness value of 0.74 reflects a minimal diversity of relations. According to Table 6.5, of the 70 object properties the majority are inheritance relations with a `SubObjectPropertyOf` value of 62. Other relations are low or undefined due to lack of applicable data definitions as shown in Table 6.7.



Table 6. 7: Summary of Matrix for NPIOnto Object Properties

Object Property	Value
InverseObjectProperties	35
FunctionalObjectProperty	27
InverseFunctionalObjectProperty	27
TransitiveObjectProperty	16
EquivalentObjectProperties	0
DisjointObjectProperties	0
SymmetricObjectProperty	0
AsymmetricObjectProperty	0
ReflexiveObjectProperty	0
IrreflexiveObjectProperty	0

Given the values of 10150 for axioms and 35 for classes (Table 6.5), the value of Axioms/Classes Ratio is equal 290. The positive high value indicates that NPIOnto is a rich source of reasoning. Given the values 742 for individual count and 35 for class count (Table 6.5), the value of Average Population is equal to 21.2. The positive low number reflects an insignificant presence of instances in NPIOnto knowledge base insufficient to represent all the knowledge modelled by NPIOnto schema. This is a result of the publication of a limited amount of information. For instance, NPI performance data pertain only to SDG Agenda and is to be shared across the four agendas. So, the NPI performance classes for the other three agendas are currently empty. The Class Richness value for NPIOnto was as low as 0.6 due to the limited amount of information available online. As a result, most classes are either empty or have a few individual instances.

6.4 Conclusion

Chapter Six documents how Research Objective 3, *demonstration of the preliminary semantic enrichment framework*, addressed Research Question 3 expressed as:

What are the findings of the preliminary semantic enrichment framework demonstrations?

This research question sought to establish key issues pertaining to one of the main drivers of the semantic enrichment process modelling (ontology) to factor into the final ODSE Framework design. For the generic ODSE approach, the deliverable is a list of applicable generic ontologies selected for the semantic enrichment processing.

For a custom domain ontology, the development of NPIOnto was motivated by the lack of semantic models for NPIs and related performance measures that represent knowledge in the development agenda domain within the Botswana Integrated Indicator Framework



context. The model development involved acquiring and representing knowledge to enable both user and machine intelligibility for the maximisation of NPIs use and reuse potentials. The conceptual knowledge models were developed first, analysed and finalised for implementation. The knowledge model implementation involved its formalisation which was centred on achieving NPI disambiguation, disaggregation, interoperability, integration, multi-dimensionality and reasoning capabilities. The last phase pertained to the evaluation of the NPIOnto structure and content for the effectiveness of the ontology design through the knowledge model and knowledge base assessments. The evaluation of NPIOnto proved the effectiveness of its design. However, the knowledge base evaluations, specifically, the Average Population and Class Richness revealed the different dimensions of insufficient instances for the class population. This limitation prompts a need for improvement on the quantity of statistics. There is also a need for more specific mappings for Botswana national agendas. The Botswana domesticated SDG Agenda relations to Vision Agenda are generically defined by the mapping of the Botswana domesticated SDGs to Vision Agenda pillars. The challenge with implementing these generic mappings is that they have numerous performance measures associated with them. For instance, Botswana domesticated SDG Agenda generically maps SDG 16 with 9 targets and 14 indicators (Figure 4.11) to Vision Agenda Pillar 4 with 2 outcomes (Figure 4.9, column 1) and 6 indicators (Figure 4.9, column 2). So, these mappings must be expressed at the lowest level of granularity for clarity in data representation. The mapping for Botswana Vision to National Development Plan also needs to be clearly specified. Overall, this was the first set of evaluations and the second set of evaluations were focused on the effectiveness of the functionality of the NPIOnto when deployed in its intended setting as detailed in Chapter Seven, Subsection 7.4.2. NPIOnto was part of a series of demonstrations of the Ontology-Driven Semantic Enrichment framework (ODSE Framework) which involved the creation of OD value (knowledge) by enriching, through NPIOnto constructs, the flat text OD resources with semantic capabilities pertaining to NPI accessibility, understandability, interoperability, integrability and intelligibility. Overall, both the generic and custom domain ontologies are partial solutions to Research Question 3 and serve as the basis for the second part of the demonstrations in Chapter Seven.



CHAPTER SEVEN: FRAMEWORK DEMONSTRATION 2 (SEMANTIC ENRICHMENT PROCESSING)

7.1 Introduction

This chapter presents Phase 3 according to Demonstration 2 activities specified in Table 3.2, pertaining to semantic enrichment pre-processing, processing, and post-processing on the basis of the two ontology approaches discussed in Chapter 6 – the generic and domain ODSE approaches. The three stages are described in Chapter Five and graphically represented by Figure 5.1.

7.2 Stage 1: Open Data Pre-processing

The *Open Data pre-processing stage* description starts with an overview of the pre-processing input data discussed in detail in the OD presentation needs (Subsection 4.3.1). That is followed by a discussion of the transformations that were performed on the input data pertaining to data unification and cleansing, among others. The last discussion is on the final deliverable of Stage 1.

7.2.1 Pre-processing Input Component

As discussed in Subsection 4.3.2, input data comprises numerous spreadsheet files presented separately according to the 17 SDGs. For each goal, the data is presented separately on the basis of each SDGI. The metadata is also presented in a diversity of flat text data structures. Overall, the pre-processing input is characterised by OD heterogeneities in terms of access formats and platforms, and data definitions. To address these limitations, the data had to undergo some *pre-processing* detailed below.

7.2.2 Pre-processing Data Re-organisation, Cleaning and Transformation Component

To increase the value-adding potential of SDG Agenda resources, presented in diverse flat text formats, the data was consolidated into single files per goal. The metadata for relating agendas was also presented in separate spreadsheet formats. In re-organising this data, numerous inconsistencies prompted for specific cleaning and transformations. The



cleaning and transformation process involved the identification and correction of errors in the datasets and alignment of published data and metadata to the domesticated and global SDG Agenda definitions. All pre-processing cleaning and transformations were done manually. Table 7.1 summarises inconsistencies, together with transformations performed to address the inconsistencies.

Table 7. 1: Datasets Inconsistencies

NPI	Inconsistency	Inconsistency Type	Transformation	Cases
<i>SDG Indicators (SDGIs)</i>				
2.2.1	Stunting values include overweight & wasting values (for 2.2.2)	NPI & value inconsistency	Removed overweight & wasting values from 2.2.1 to 2.2.2	1
1.5.1; 1.5.3; 5.1.1; 5.c.1; 6.4.1; 11.5.1; 11.b.1	NPI values published in SB portal but NPIs missing in the Domesticated SDGI list	NPI & value inconsistency	NPIs added to the Domesticated SDGI list for consistent representation	7
2.4.1; 5.a.1; 14.5.1; 14.7.1; 16.1.2; 16.7.2; 16.10.1	SDGIs mapped to Africa Agenda Indicator but SDGIs missing in the Domesticated SDGI list	NPI mapping inconsistency	NPIs added to the Domesticated SDGI list for consistent representation	7
3.a.1	Tabular data attached is for 3.6.1	NPI & value inconsistency	Correct value alignment	1
4.c.1	NPI values for literacy & numeracy (for 4.6.1) instead of teacher training values	NPI & value inconsistency	No action	1
6.4.2	Missing Values	Missing values	No action	1
8.5.1	NPI values for manager earnings instead of male & female employees' values	NPI & value inconsistency	Changed SDGI definition to match value definition	1
	Conflicting age ranges of 15-35 & 15-17	Value disaggregation inconsistency	Removed 15-35 range & adjusted 15-17 to 15-19	1
9.c.1	NPI values for medium & high-tech industry value added (for 9.b.1) instead of mobile network coverage	NPI & value inconsistency	Did the correct value matching	1
11.5.1 & 13.1.1; 15.7.1 & 15.c.1; 1.5.3 & 11.b.1	Same NPIs with same values but different identity codes	NPI definition & value duplication	No action	3
5.2.1	Acronym GBV for value definition	Value ambiguity	used specific Gender Based Violence	1
	Inclusion of GBV for all genders while semantic definition specifies women & girls only	NPI & value inconsistency	Screened out all non-gender specific values	1
<i>Africa Indicators</i>				
A1 – G1 – P1 – I2; A1 – G1 – P1 – I52; A1 – G1 – P1 – I53	Missing target, both in the NPI specification list & NPI code	Incomplete Semantic definition	No action	3
<i>Vision Indicators</i>				



All PIs	No identity & relationship Codes' definitions	Incomplete semantic definition	Generated both identity & relationship codes	1
---------	--	-----------------------------------	---	---

These inconsistencies are summarised in terms of their prevalence rate (Appendix 5, Figure Appx 5. 1). The highly prevalent data inconsistency pertain to NPI and value definitions, followed by incomplete semantic definitions and duplicate NPI definitions and values. The less prevalent data inconsistencies pertain to missing values, NPI mapping inconsistencies, ambiguous values, and value disaggregation inconsistencies. The resulting integrated and transformed data is presented in the next section.

7.2.3 Pre-processing Output Component

The result of the pre-processing stage comprised cleaned and transformed data which is at this stage flat text resources (spreadsheet) consolidated per goal. Three other resources were produced from the unstructured data pertaining to the SDG Agenda metadata, the Africa Agenda metadata, and the Vision Agenda metadata. The pre-processing output are collections of relevant domain knowledge on NPIs in loose format as follows (sample screenshots of datasets could be found in the Appendices section).

- Africa Agenda Indicator Dataset (Appendix 1, Table AFR 1),
- Africa Agenda Target Dataset (Appendix 1, Table AFR 2),
- Africa Agenda Priority Dataset (Appendix 1, Table AFR 3),
- Africa Agenda Goal Dataset (Appendix 1, Table AFR 4),
- Africa Agenda Aspiration Dataset (Appendix 1, Table AFR 5),
- Africa Agenda Implementation Plan Dataset (Appendix 1, Table AFR 6),
- Africa Agenda Dataset (Appendix 1, Table AFR 7),
- SDG Indicator Dataset (Appendix 2, Table SDG 1),
- SD Target Dataset (Appendix 2, Table SDG 2),
- SD Goal Dataset (Appendix 2, Table SDG 3),
- SDG Agenda Dataset (Appendix 2, Table SDG 4),
- Vision Indicator Dataset (Appendix 3, Table VIS 1),
- Vision Outcome Dataset (Appendix 3, Table VIS 2),
- Vision Pillar Dataset (Appendix 3, Table VIS 3),
- Vision Agenda Dataset (Appendix 3, Table VIS 4), and



- Published SDGI data (SDG Agenda Performance data) for SDG 1 to SDG 17 resources (Appendix 4, Table SDGI 1).

The contents of these resources are described in detail in Subsection 6.2.1. Sample resources have been uploaded to figshare and are available from this link, [ODSE Framework Package 5 - Pre-processed OD Resources for all Agendas](#). These resources served as input for the next processing stage.

7.3 Stage 2: Open Data Semantic Enrichment Processing

The *Open Data semantic enrichment processing* is the OD value (knowledge) creation stage, comprised of both the integrated flat text resources from Stage 1. In order for this data to be more understandable, accessible, interlinked, interoperable and intelligible, it had to be semantically enriched. For increased value-adding potential, the LOD technique was adopted in the conversion of flat text resources to a rich, flexible and dynamic format on the basis of ontology constructs. The approach involved the transformation of data to semantic format and the assignment of URI reference for web interlinking.

7.3.1 Data Enrichment Component

This subsection presents the process of how the flat text resources were transformed to knowledge resources through LOD-driven semantic enrichment. The demonstration involved two major activities: 1) the demonstration of semantic enrichment process on the basis of generic ODSE approach and 2) the demonstration of semantic enrichment process on the basis of the domain ODSE approach. These ODSE approaches are introduced in Section 6.1. In both activities, semantic enrichment entailed the reconciliation and RDFization of flat text resources.

Resource Reconciliation

Resources reconciliation involves a comparison of local metadata values to those in controlled vocabularies to assess the degree of resource interoperability and integrability in the Web of Data. For the generic ODSE approach, the reconciliation involved the alignment of the domesticated SDG Agenda metadata to the global SDGs repository/registry (SDGIO) for standardisation. For example, the description metadata for `sdgi 1.3.1` is published as “*The Poor Receiving Government Assistance*” while the global



repository (SDGIO) gives the exact description as provided in the global SDGs list being, *“Proportion of population covered by social protection floors/systems, by sex, distinguishing children, unemployed persons, older persons, persons with disabilities, pregnant women, newborns, work-injury victims and the poor and the vulnerable”* [10], [5]. That was followed by pulling URIs from the SDGIO in matching cases. Therefore, the reconciled links provide a more accurate description for ease of global integration. The end product is standardised, interoperable and integrable product for enhanced consumption capabilities. However, the reconciliation process was successful for only three performance variables that are defined in the SDGIO, being, `goalDesc` (goal description); `targetDesc` (target description); and `sdgiDesc` (sdgi description) as shown in Figure 7.1. The reconciled values appear in blue colour.

goalName	goalTheme	goalDesc	targetCode	targetDesc	uniqSdgiCode	sdgiCode	sdgiDesc	measureDesc	measureUnit	measureYear	mat
Sustainable Development Goal 3	good health and well-being	ensure healthy lives and promote well-being for all at all ages Choose new match	3.1	By 2030, reduce the global maternal mortality ratio to less than 70 per 100,000 live births Choose new match	2030_3.1.1	3.1.1	Maternal mortality ratio Choose new match	Maternal mortality ratio	per 100,000 population	2015	
Sustainable Development Goal 3	good health and well-being	ensure healthy lives and promote well-being for all at all ages Choose new match	3.1	By 2030, reduce the global maternal mortality ratio to less than 70 per 100,000 live births Choose new match	2030_3.1.2	3.1.2	Proportion of births attended by skilled health personnel Choose new match	n/a	n/a		
Sustainable Development Goal 3	good health and well-being	ensure healthy lives and promote well-being for all at all ages Choose new match		http://purl.unep.org/sdg/SDGIO_00020029 http://purl.unep.org/sdg/SDGIO_00020029 Under-five mortality rate			Under-five mortality rate Choose new match	Under-five mortality rate, by sex (deaths per 1,000 live births)	per 1,000 live births	2011	

Figure 7. 1: Resource Reconciliation in OpenRefine - Generic ODSE Approach

As reflected in Figure 7.1, resource reconciliation connects local metadata to controlled vocabulary in the Web of Data. With the domain ODSE approach, NPIOnto extended the reconciliation scope as depicted in Figure 7.2 in which the reconciliation scope includes matching values for `goalName`; `goalTheme`; `targetCode` and `sdgiCode`. A high number of reconciliation matches, signifies the resource’s degree of interoperability and integrability capabilities.



goalName	goalTheme	goalDesc	targetCode	targetDesc	uniqSdgiCode	sdgiCode	sdgiDesc	measureDesc	meas
Sustainable Development Goal 3 Choose new match	Good Health & Well-being Choose new match	ensure healthy lives and promote well-being for all at all ages Choose new match	3.1 Choose new match	By 2030, reduce the global maternal mortality ratio to less than 70 per 100,000 live births Choose new match	2030_3.1.1	3.1.1 Choose new match	Maternal mortality ratio Choose new match	Maternal mortality ratio	per 100,000 population
Sustainable Development Goal 3 Choose new match	Good Health & Well-being Choose new match	ensure healthy lives and promote well-being for all at all ages Choose new match	3.1 Choose new match	By 2030, reduce the global maternal mortality ratio to less than 70 per 100,000 live births Choose new match	2030_3.1.2	3.1.2 Choose new match	Proportion of births attended by skilled health personnel Choose new match	n/a	n/a
Sustainable Development Goal 3 Choose new match	http://localhost/8080/npio/2030_G3 http://localhost/8080/npio/2030_G3 Sustainable Development Goal 3 Goal 3 of Agenda 2030				2030_3.2.1	3.2.1 Choose new match	Under-five mortality rate Choose new match	Under-five mortality rate, by sex (deaths per 1,000 live births)	per 1,000 births

Figure 7. 2: Resource Reconciliation in OpenRefine – Domain ODSE Approach

Overall, figures 7.1-7.2 demonstrate the scope of reconciliation by ontology type. The generic ontology (SDGIO) reconciliation yielded only three global matching values (Figure 7.1) while the domain ontology with more domain specific terms, extended the reconciliation scope to seven values (Figure 7.2).

Resource RDFization

Resource RDFization refers to the generation of the RDF representation of data that was originally in a different format or the conversion of data to RDF format [135]. Resource RDFization involves the addition of a semantic layer, the ontology layer, for its high potential in maximising semantic enrichment impact. So, in addition to the reconciliation of flat text resources with links to controlled vocabularies, i.e., the SDGIO, for semantic interoperability and integrability, the data values were enriched with mappings to those ontologies. The final product of the mappings is the triple structure that specifies the links between the data as depicted in Figure 7.3.



uniqSdgiCode URI	☐ X> sio:SIO_000116→	☐ goalName Cell
X dbr:Unique_identifier	X> dbp:Theme→	☐ goalTheme Cell
Add type	X> sdgio:SDGIO_00000000→	☐ goalDesc Cell
	X> sdgio:SDGIO_00000074→	☐ targetCode Cell
	X> sdgio:SDGIO_00000001→	☐ targetDesc Cell
	X> rdfs:label→	☐ uniqSdgiCode Cell
	X> sdgio:SDGIO_00000242→	☐ sdgiCode Cell
	X> sdgio:SDGIO_00000002→	☐ sdgiDesc Cell
	X> afo:AFX_0002802→	☐ measureDesc Cell
	X> sio:SIO_000221→	☐ measureUnit Cell
	X> dcterms:date→	☐ measureYear Cell
	X> sdgio:SDGIO_00000003→	☐ maternal_mortality_agg Cell
	X> sdgio:SDGIO_00000003→	☐ under5_mortality_agg Cell
	X> sdgio:SDGIO_00000003→	☐ HIV_infections_agg Cell
	X> sdgio:SDGIO_00000003→	☐ tuberculosis_agg Cell
	X> sdgio:SDGIO_00000003→	☐ trafficDeaths_agg Cell
	X> afo:AFX_0002801→	☐ trafficDeaths_male Cell

Figure 7. 3: Resource RDFization in OpenRefine - Generic ODSE Approach

As reflected in Figure 7.3, unique identifiers serve as the subjects, Uniform Resource Identifiers that serve as consistent addressable resource identity [136]. The other data values serve as objects, the values of the statements made in the triple which could be literal values such as date or number or another Uniform Resource Identifier in cases where the value of the resource is another resource [136]. The predicates are represented by mappings between subjects and objects. A predicate is a Uniform Resource Identifier that describes the relationship between the resource and the value of the triple [136].

The mappings specify data properties between the subjects and objects, and all the four resources are mapped to the chosen generic ontologies specified in Subsection 6.2.4. The resource mappings specify how each resource is to be addressed in the semantic web. Thus, the base Uniform Resource Identifier was specified for each resource to serve as a connection link to other resources both within an individual resource and to external web resources. For SDGI data, the mappings were drawn separately per goal. Enriching text values with meaning was meant to improve ease of understanding and interpretation of publications. For the domain ODSE approach, NPIOnto produced a more specific and rich semantic enrichment results from its specific/precise meanings and rich semantics depicted in Figure 7.4.



uniqSdgiCode URI	☐ X>npio:hasGoalName→	☐ goalName Cell
X npio:sdgi	X>npio:hasThemeDim→	☐ goalTheme Cell
Add type	X>npio:hasGoalDesc→	☐ goalDesc Cell
	X>npio:hasTargetCode→	☐ targetCode Cell
	X>npio:hasTargetDesc→	☐ targetDesc Cell
	X>npio:hasInstanceCode→	☐ uniqSdgiCode Cell
	X>npio:hasNpiCode→	☐ sdgiCode Cell
	X>npio:hasNpiDesc→	☐ sdgiDesc Cell
	X>npio:hasMeasureDesc→	☐ measureDesc Cell
	X>npio:hasUnitMeasure→	☐ measureUnit Cell
	X>npio:hasPerfYr→	☐ measureYear Cell
	X>npio:hasAgVal-matrnMrtalty→	☐ maternal_mortality_agg Cell
	X>npio:hasAgVal-u5Mrtalty→	☐ under5_mortality_agg Cell
	X>npio:hasAgVal-hivInfectn→	☐ HIV_infections_agg Cell
	X>npio:hasAgVal-tuberculosis→	☐ tuberculosis_agg Cell
	X>npio:hasAgVal-trfckDeath→	☐ trafficDeaths_agg Cell
	X>npio:hasTrfckDeathDsagVal-m→	☐ trafficDeaths_male Cell

Figure 7. 4: RDFization in OpenRefine - Domain ODSE Approach

Figure 7.4 reflects linkages to precise meanings and rich, domain specific semantics – NPIOnto constructs. For instance, there is clarity on aggregate values which are represented by meaningful data property labels.

Overall, the data enrichment process creates OD value (knowledge) through the addition of a semantic layer to flat text resources. This stage depicts a LOD-driven semantic model that integrates the data and its metadata (Section 4.3, Figures 4.2-4.12) for more contextual and conceptual clarity, and also for standardisation of meaning across the Web of Data. On the one hand, resource reconciliation creates the accessibility, interoperability and integrability capabilities in flat text resources by establishing linkages across the Web of Data. On the other hand, resource RDFization creates the accessibility, understandability, interoperability, integrability, and intelligibility capabilities to the data by defining a common meaning and representation format across the Web of Data. All in all, the data enrichment process serves to enable seamless access to the data and to establish a common understanding of the data for OD interoperability, integrability, and intelligibility. In addition, data enrichment process makes use of ontology constructs (classes, data properties and object properties) that create understandability and intelligibility (analytical and reasoning) capabilities.

7.3.2 Semantic Enrichment Output Component

The semantic enrichment output is the semantically enriched version of the disparate raw OD resource publications presented above, ranging from Figures 4.2-4.12 (Section 4.3) published in HTML image, text and graphic formats (Figures 4.2-4.4 and 4.6),



spreadsheet format (Figure 4.5), API link (Figure 4.7) and XML format (Figure 4.10) together with corresponding metadata in pdf format (Figures 4.8-4.9 and 4.11-4.12). These semantically enriched data pieces are the building blocks of intelligent data [63]. Intelligent data is what this research contextualises as, *knowledge resource*, data that is systematically processed, and presented in meaningful form for active sharing and for further assessment and interpretation in the Web of Data [97]. A knowledge resource holds within itself, both structurally and contextually, all that an application needs to interpret it [50]. It is characterised by clarity (explicit definitions and measurements) and is actionable (supports decision making and implementation) [98]. The output for the generic ODSE approach is referred to in this thesis as the generic knowledge resource. Figure 7.5 depicts a sample generic knowledge resource extract for Figure 4.4.

```

1 <http://www.statsbots.org/bw/data/sdgi/goal16/2030_16.9.1> a dbr:Unique_identifier;
2 dbp:theme "peace, justice and strong institutions";
3 afo:AFX_0002801 "79"^^<http://www.w3.org/2001/XMLSchema#int>;
4 "88"^^<http://www.w3.org/2001/XMLSchema#int>;
5 "89"^^<http://www.w3.org/2001/XMLSchema#int>;
6 "90.7"^^<http://www.w3.org/2001/XMLSchema#int>;
7 "91"^^<http://www.w3.org/2001/XMLSchema#int>;
8 afo:AFX_0002802 "Proportion of children under 5 years of age whose births have
9     been registered with a civil authority (% of children under 5 years of age)";
10 dcterms:date "2015"^^<http://www.w3.org/2001/XMLSchema#int>;
11 sdgio:SDGIO_00000000 "Promote peaceful and inclusive societies for sustainable development, provide
12     access to justice for all and build effective, accountable and inclusive institutions at all levels";
13 sdgio:SDGIO_00000001 "By 2030, provide legal identity for all, including birth registration";
14 sdgio:SDGIO_00000002 "Proportion of children under 5 years of age whose births have been registered
15     with a civil authority, by age";
16 sdgio:SDGIO_00000003 "87.5"^^<http://www.w3.org/2001/XMLSchema#int>;
17 sdgio:SDGIO_00000074 "16.9";
18 sdgio:SDGIO_00000242 "16.9.1";
19 sio:SIO_000116 "Sustainable Development Goal 16";
20 sio:SIO_000221 "%";
21 rdfs:label "2030_16.9.1" .

```

Figure 7. 5: Generic Knowledge Resource (SDGI 16.9.1 Extract)

Figure 7.5 reflects the use of generic ontologies for standardisation of meaning. The description of the data is as follows: sample instance of SDG Agenda published data (Line 1); SDGI instance code (Line 21); SDG description (Lines 11-12); SDG name (Line 19); SDG theme (Line 2); SDG target code (Line 17); SDG target description (Line 13); UN SDG Indicator ID (Line 18); SDGI description (Lines 14-15); measurement description (Lines 8-9); unit of measurement (Line 20); performance year (Line 10); SDGI aggregate value



(Line 16); and SDGI disaggregate values (Lines 3-7). In comparison, the output for the domain ODSE approach is the domain knowledge resource. Figure 7.6 presents a sample domain knowledge resource extract for Figure 4.4.

```

1 <http://www.statsbots.org/bw/data/sdgi/goal16/2030_16.9.1> a npio:sdgi;
2 npio:hasAgVal-birthReg "87.5"^^<http://www.w3.org/2001/XMLSchema#int>;
3 npio:hasBirthRegDsagVal-0yr "79"^^<http://www.w3.org/2001/XMLSchema#int>;
4 npio:hasBirthRegDsagVal-1yr "89"^^<http://www.w3.org/2001/XMLSchema#int>;
5 npio:hasBirthRegDsagVal-2yr "88"^^<http://www.w3.org/2001/XMLSchema#int>;
6 npio:hasBirthRegDsagVal-3yr "90.7"^^<http://www.w3.org/2001/XMLSchema#int>;
7 npio:hasBirthRegDsagVal-4yr "91"^^<http://www.w3.org/2001/XMLSchema#int>;
8 npio:hasGoalDesc "Promote peaceful and inclusive societies for sustainable development,
9     provide access to justice for all and build effective, accountable
10    and inclusive institutions at all levels";
11 npio:hasGoalName "Sustainable Development Goal 16";
12 npio:hasInstanceCode "2030_16.9.1";
13 npio:hasMeasureDesc "Proportion of children under 5 years of age whose births have been
14    registered with a civil authority (% of children under 5 years of age)";
15 npio:hasNpiCode "16.9.1";
16 npio:hasNpiDesc "Proportion of children under 5 years of age whose births have been
17    registered with a civil authority, by age";
18 npio:hasPerfYr "2015"^^<http://www.w3.org/2001/XMLSchema#int>;
19 npio:hasTargetCode "16.9";
20 npio:hasTargetDesc "By 2030, provide legal identity for all, including birth registration";
21 npio:hasThemeDim "peace, justice and strong institutions";
22 npio:hasUnitMeasure "%".

```

Figure 7. 6: Domain Knowledge Resource (SDGI 16.9.1 Extract)

The domain knowledge resource (Figure 7.6) reflects the use of custom domain ontology (NPIONto) for standardisation of meaning. The description of the data is as follows: sample instance of SDG Agenda published data (Line 1); SDGI instance code (Line 12); SDG description (Lines 8-10); SDG name (Line 11); SDG theme (Line 21); SDG target ID (Line 19); SDG target description (Line 20); UN SDG indicator ID (Line 15); SDGI description (Lines 16-17); measurement description (Lines 13-14); unit of Measurement (Line 22); performance year (Line 18); SDGI aggregate value (Line 2); and SDGI disaggregate values (Lines 3-7). Sample domain and generic knowledge resources for SDG16 have been uploaded to figshare and are available in this link, [ODSE Framework Package 6 – Sample Knowledge Resources for SDG16](#).

Overall, knowledge resources represent the unified presentation structure and format for raw OD publications. The unified data in Figures 7.5-7.6 was scattered as demonstrated by raw OD extracts in Figures 4.2-4.12 (Section 4.3). However, this research has observed



that the level of clarity differs on the basis of the degree of concepts specification (contextualisation) for a given semantic enrichment approach or concepts specification of the ontology/vocabulary used for semantic enrichment. As reflected in Figure 7.6, NPIOnto properties provide more precise meaning resulting from more clarity in domain descriptive properties. Above that, NPIOnto provides more clarity in NPI value disaggregation properties that differentiate between the NPI composite elements. In short, while the generic knowledge resource achieves standardised publication of data, the domain knowledge resource extends that by factoring in the precision of meaning as well. Table 7.2 presents a summary of semantic enrichment approach concepts specifications of the two semantic enrichment approaches for this case study.

Table 7. 2: Summary of Semantic Enrichment Approach Concepts Specifications

Entity	Attribute	Generic ODSE Specifications	Domain ODSE Specifications
SDGI	unique identifier	has similar contextual meaning to dbpedia property dbr:Unique_Identifier	has similar contextual meaning to property npio:hasInstanceCode & aligns to class npio:sdgi in npioOnto
	indicator code	has similar contextual meaning to object SDGIO_00000242 (UN SDG Indicator ID) in SDGIO ontology	has similar contextual meaning to property npio:hasNpiCode in npioOnto
	indicator name	has similar contextual meaning to Semantic Science property SIO:000116 (name)	has the same contextual meaning as that of property npio:hasNpiName in npioOnto
	indicator description	has the same contextual meaning as that of object SDGIO_00000002 (SDGI) in SDGIO ontology	has the same contextual meaning to npioOnto property npio:hasNpiDesc
	indicator theme	has similar contextual meaning to dbpedia property dbr:theme	has the same contextual meaning to npioOnto property npio:hasThemeDim
	target code	has the same contextual meaning as that of object SDGIO_00000074 (Target Code) in SDGIO ontology	has the same contextual meaning as that of npioOnto property npio:hasTargetCode
	target description	has the same contextual meaning as that of object SDGIO_00000001 (SD Target) in SDGIO ontology	has the same contextual meaning as that of npioOnto property npio:hasTargetDesc
	goal name	contextual meaning aligns to that in the Semantic Science resource registry for property SIO:000116 (name)	has the same contextual meaning as that of npioOnto property npio:hasGoalName
	goal description	has the same contextual meaning as that of object SDGIO_00000000 (SD Goal) in SDGIO ontology	has the same contextual meaning as that of npioOnto property npio:hasGoalDesc
	measure description	has the same contextual meaning as that of object AFX: 0002802 (has description) in AFO ontology	has the same contextual meaning as that of npioOnto property npio:hasMeasureDesc
SDGI Value	all values	are of integer data type as to w3 XMLSchema definition	are of integer data type according to w3 XMLSchema definition



aggregate value	has the same contextual meaning as that of object SDGIO_00000003 (SDGI Value) in SDGIO ontology	has the same contextual meaning as that of property npio:hasAgVal-birthReg in npioOnto
disaggregate value	represents the diverse disaggregation of the aggregate value and have the same contextual meaning as that of object AFX: 0002801 (has Categorical Value) in AFO ontology	represents value disaggregation that aligns to superproperty npio:hasAgeDsagVal-birthReg with similar meaning to one of the following subproperties: npio:hasBirthRegDsagVal-0yr; npio:hasBirthRegDsagVal-1yr; npio:hasBirthRegDsagVal-2yr; npio:hasBirthRegDsagVal-3yr; npio:hasBirthRegDsagVal-4yr
Unit	has similar meaning to that in the Semantic Science resource registry for property SIO:000221 (has unit)	has the same contextual meaning as that of property npio:hasUnitMeasure in npioOnto
Year	has the same contextual meaning as that of property dcterms:date in the Dublin Core Metadata registry	has the same contextual meaning as that of property npio:hasPerfYr in npioOnto

Table 7.2 reveals that the two ODSE approaches have different degrees of contextualisation. Concepts specification is general for the approach that is focused on data standardisation, the generic ODSE approach and is more specific for the approach that is focused on data clarity, the domain ODSE approach. This has cascading impact on OD value (knowledge) creation potential. For the generic ODSE approach, the knowledge is more generic while for the domain ODSE approach, the knowledge is more specific to the development agenda domain within the Botswana Integrated Indicator Framework.

7.4 Stage 3 Open Data Post-processing

The *Open Data post-processing stage* deals with further processing of the output of semantic enrichment (knowledge resources) presented in Figures 7.5-7.6. This stage pertains to knowledge resources integration and publication.

7.4.1 Resource Integration Component

Resource integration component involves the consolidation of individual resources. For the SDGI data, knowledge resources were generated separately for each SD Goal (column 1 of Figure 7.7). For the metadata, knowledge resources were also generated separately on the basis of performance definition constructs or elements per agenda (column 2-4 of Figure 7.7).



<i>Sustainable Devt Goals Indicator Data</i>	<i>Sustainable Devt Agenda 2030 Metadata</i>	<i>Africa Agenda 2063 Metadata</i>	<i>Vision 2036 Agenda Metadata</i>
 1. sdgiData-goal1	 1. indicator-sdgaMe	 1. indicator-afraMeta	 1. indicator-visaMeta
 2. sdgiData-goal2	 2. target-sdgaMeta	 2. target-afraMeta	 2. outcome-visaMeta
 3. sdgiData-goal3	 3. goal-sdgaMeta	 3. priority-afraMeta	 3. pillar-visaMeta
 4. sdgiData-goal4	 4. agenda-sdgaMeta	 4. goal-afraMeta	 4. agenda-visaMeta
 5. sdgiData-goal5		 5. aspiration-afraMeta	
 6. sdgiData-goal6		 6. impPlan-afraMeta	
 7. sdgiData-goal7		 7. agenda-afraMeta	
 8. sdgiData-goal8			
 9. sdgiData-goal9			
 10. sdgiData-goal10			
 11. sdgiData-goal11			
 12. sdgiData-goal12			
 13. sdgiData-goal13			
 14. sdgiData-goal14			
 15. sdgiData-goal15			
 16. sdgiData-goal16			
 17. sdgiData-goal17			

Figure 7. 7: Knowledge Resources per Agenda

These knowledge resources are a product of the generic ODSE approach to semantic enrichment. For ease of publication and evaluation, they were then integrated into one knowledge resource by combining all triple files depicted in Figure 7.7. For the domain ODSE approach to semantic enrichment, the post-processing input are the 17 SDGI resources (column 1) only as all the metadata files are components of NPIONto. These files together with NPIONto were also integrated into single knowledge resource. The sample integrated datasets have been uploaded to figshare and are available from this link, [ODSE Framework Package 7 – Integrated Knowledge Resources](#).

7.4.2 Resource Publication and Consumption Component

The knowledge resources produced from semantic enrichment, semantic capabilities and knowledge value-added OD, are the basis for knowledge extraction/mining and use. Therefore, they must be made readily available for consumption. Hence, resource integration is followed by resource publication which involves making the resources freely available for both human and machine consumption. This necessitates publication



in platforms that support the extraction of value (knowledge mining) from the data. This is achieved through queries that are issued and evaluated over SPARQL Endpoints [137].

Showing 1 to 30 of 30 entries (filtered from 38,690 total entries) Search: 2030_16.9.1 Show 50 entries

subject	predicate	object
<http://www.statsbots.org/bw/data/sdgi/SDG16/2030_16.9.1>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://dbpedia.org/resource/Unique_identifier>
<http://www.statsbots.org/bw/data/sdgi/SDG16/2030_16.9.1>	<http://dbpedia.org/property/theme>	"peace, justice and strong institutions"
<http://www.statsbots.org/bw/data/sdgi/SDG16/2030_16.9.1>	<http://purl.alkotrope.org/ontologies/property#AFX_0002802>	"Proportion of children under 5 years of age whose births have been registered with a civil authority (% of children under 5 years of age)"
<http://www.statsbots.org/bw/data/sdgi/SDG16/2030_16.9.1>	<http://purl.org/dc/terms/date>	"2015"^^xsd:integer
<http://www.statsbots.org/bw/data/sdgi/SDG16/2030_16.9.1>	<http://purl.unep.org/sdg/SDGIO_00000000>	"Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels"
<http://www.statsbots.org/bw/data/sdgi/SDG16/2030_16.9.1>	<http://purl.unep.org/sdg/SDGIO_00000001>	"By 2030, provide legal identity for all, including birth registration"
<http://www.statsbots.org/bw/data/sdgi/SDG16/2030_16.9.1>	<http://purl.unep.org/sdg/SDGIO_00000074>	"16.9"

Figure 7. 8: SPARQL Endpoint Publication (SDGI 16.9.1 Extract)

SPARQL is a semantic query language for databases. SPARQL Endpoint is a web service that conform to the SPARQL protocol (SPARQL protocol service) which enables both human and machine users to query a knowledge base via the SPARQL language [138], [139]. SPARQL queries are passed to a Uniform Resource Locator where a SPARQL service processes the query and returns results in a defined semantic format [138].

As reflected in the header of Figure 7.8, for this demonstration, the SPARQL server publication pertains to 38,690 triples comprising both data and metadata triples. This constitutes the NPI triple-store that serves as the *public knowledge resource pool* for inclusive knowledge mining as proposed in the preliminary ODSE Framework [120].

In this research case, through the execution of knowledge mining operations, the semantically enriched OD resources (knowledge resources) were deployed for exploration of the degree of semantic capabilities and the mining/extraction of semantic value (knowledge) added to flat text resources through ontologies. The mining was according to two knowledge resources mining cases, the generic and the domain knowledge resource mining cases, that demonstrate semantic value in terms of potential semantic capabilities (query formulation – syntax) and embedded knowledge (query output – semantics). The knowledge resources mining operations were conducted to



serve as the basis for the comparative analysis of resources. Specifically, the knowledge mining assessment was conducted to:

- assess the degree of knowledge value-added to raw OD resources, demonstrated by the degree of query/extraction/harvest output; and
- assess the degree of semantic capabilities value-added to raw OD resources, demonstrated by the degree of the knowledge resource' accessibility, understandability, integrability, interoperability, and intelligibility.

This knowledge mining assessment comprised the generic and domain knowledge resources' mining cases (Figures 7.9, 7.12-7.13) and corresponding knowledge harvest (Figures 7.10-7.11, 7.14-7.16) for the sample instance (SDGI 16.9.1). The scripts for the mining cases have been uploaded to figshare and are accessible from this link, [ODSE Framework Package 8 - Integrated Knowledge Resources Evaluation Cases](#). The knowledge harvesting was conducted through SPARQL queries against the SPARQL endpoint publications of knowledge resources. Following is the detailed discussion of the knowledge mining demonstration for the three assessments specified above.

Generic Knowledge Resource Mining Case

The knowledge resource mining case, shown in Figure 7.9, mines the integrated generic knowledge resource (discussed in Subsection 7.4.1) for semantic value-added to flat text resources, in the form of embodied knowledge and semantic capabilities.



```

1 SELECT DISTINCT ?sdgiPerfYr ?sdgiCode ?unsdCode ?sdgTargetCode ?sdgTargetDesc ?sdGoalCode ?sdGoalDesc ?sdgUndbRef
2 ?measDesc ?unitMeas ?sdgiPerfVal ?visPillar_agendaCode ?visPillarPerfVal ?afriCode ?afriPerfVal ?sdgiDisaggVal
3 WHERE {
4   <http://www.statsbots.org/bw/data/sdgi/SDG16/2030_16.9.1>
5     <http://www.w3.org/2000/01/rdf-schema#label> ?sdgiDataInstance;
6     <http://purl.unep.org/sdg/SDGIO_00000242> ?sdgiCode;
7     <http://purl.unep.org/sdg/SDGIO_00000002> ?sdgiDesc;
8     <http://semanticscience.org/resource/SIO_000221> ?unitMeas;
9     <http://purl.allotrope.org/ontologies/property#AFX_0002802> ?measDesc;
10    <http://purl.org/dc/terms/date> ?sdgiPerfYr;
11    <http://purl.unep.org/sdg/SDGIO_00000003> ?sdgiPerfVal;
12    <http://purl.allotrope.org/ontologies/property#AFX_0002801> ?sdgiDisaggVal;
13    <http://purl.unep.org/sdg/SDGIO_00000001> ?sdgTargetDesc;
14    <http://dbpedia.org/property/theme> ?themeDim;
15    <http://dbpedia.org/property/theme> "peace, justice and strong institutions";
16    <http://purl.org/dc/terms/identifier> ?sdGoalCode;
17    <http://semanticscience.org/resource/SIO_000116> ?sdGoalName;
18    <http://purl.unep.org/sdg/SDGIO_00000000> ?sdGoalDesc.
19  ?sdgiMetaInstance <http://purl.unep.org/sdg/SDGIO_00000242> ?sdgiCode;
20    <http://semanticscience.org/resource/SIO_000116> ?sdgiName;
21    <http://dbpedia.org/property/theme> ?sdgiSubtheme;
22    <http://purl.org/dc/terms/relation> ?sdgTargetCode;
23    <http://purl.obolibrary.org/obo/NCIT_C115618> ?devtDim;
24    <http://purl.obolibrary.org/obo/NCIT_C115618> "Social".
25  ?sdgTargetMetaInstance <http://purl.unep.org/sdg/SDGIO_00000074> ?sdgTargetCode;
26    <http://semanticscience.org/resource/SIO_000116> ?sdgTargetName.
27  ?sdgiIndInstance <http://purl.unep.org/sdg/SDGIO_00000242> ?sdgiCode;
28    <http://purl.unep.org/sdg/SDGIO_00000227> ?unsdCode.
29  ?sdgiGoalInstance <http://www.geneontology.org/formats/oboInOwl#hasExactSynonym> ?sdGoalCode;
30    <http://www.geneontology.org/formats/oboInOwl#hasDbXref> ?sdgUndbRef.
31  ?sdGoalMetaInstance <http://purl.org/dc/terms/identifier> ?sdGoalCode;
32    <http://purl.org/dc/terms/relation> ?visPillar_agendaCode.
33  ?visiMetaInstance <http://purl.org/dc/terms/identifier> ?visPillar_agendaCode;
34    <http://semanticscience.org/resource/SIO_000116> ?visPillar_agendaDesc.
35  ?afriMetaInstance <http://purl.org/dc/terms/relation> ?sdgiCode;
36    <http://purl.org/dc/terms/identifier> ?afriCode;
37    <http://semanticscience.org/resource/SIO_000116> ?afriName;
38    <http://purl.allotrope.org/ontologies/property#AFX_0002802> ?afriDesc;
39    <http://purl.org/dc/terms/relation> ?afriTargetCode.
40  BIND(?sdgiPerfVal AS ?visPillarPerfVal) BIND(?sdgiPerfVal AS ?afriPerfVal) }

```

Figure 7. 9: Generic Knowledge Resource Mining (SDGI 16.9.1)

The output for the generic knowledge resource mining case represents selected variables specified in Lines 1-2 from the variable pool reflected in Figure 7.8. For image visibility, the sample output extract for this knowledge mining case is displayed in two separate interfaces (Figures 7.10-7.11). The first part of the output (Figure 7.10) displays the extracted metadata for SDGI 16.9.1 and the corresponding performance value extractions are displayed in the second part of the output (Figure 7.11).



	sdgiPerfYr	sdgiCode	sdgiDesc	unsdCode	sdgTarget	sdgTargetDesc	sdGoalCode	sdGoalDesc	sdgUndbRef	measDesc	unitMeas
1	"2015"^^xsd:integer	"16.9.1"	"Proportion of children under 5 years of age whose births have been registered with a civil authority, by age"	"C160901"	"16.9"	"By 2030, provide legal identity for all, including birth registration"	"SDG16"	"Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels"	"http://www.un.org/sustainabledevelopment/peace-justice/"	"Proportion of children under 5 years of age whose births have been registered with a civil authority (% of children under 5 years of age)"	"%"

Showing 1 to 1 of 1 entries

Figure 7. 10: Sample Generic Knowledge Resource Mining Output 1 (SDGI 16.9.1)

Figure 7.10 represents the semantic form of the metadata extract from statistics Botswana OD portal, specifically, `sdgiPerfYr`, `sdgiCode`, `sdgiDesc`, `sdgTarget`, `sdgTargetDesc`, `sdGoalCode`, `sdGoalDesc`, `unitMeas` and `measDesc` in graphical, spreadsheet and pdf form. The `unsdCode` and the `sdgUndbRef` values are extracted from the global indicator list for SDG Agenda in spreadsheet form as well. This data further links to two other agendas that form the Botswana Integrated Indicator Framework, Vision Agenda (`visPillar_agendaCode` and `visPillar_agendaDesc`) and Africa Agenda (`afriCode` and `afriDesc`) as reflected in the second part (Figure 7.11).



	sdgiPerfVal	visPillar_agendaCode	visPillar_agendaDesc	visPillarPerfVal	afriCode	afriDesc	afriPerfVal	sdgiDisaggVal
1	"87.5"^^xsd:int	<u>"agenda2030"</u>	"agenda 2030"	"87.5"^^xsd:int	"A6-G16-P2-T3-I51"	"The proportion of children whose births are registered within the first year"	"87.5"^^xsd:int	<u>"90.7"^^xsd:int</u>
2	"87.5"^^xsd:int	<u>"pillar4"</u>	"Governance, Peace & Security"	"87.5"^^xsd:int	"A6-G16-P2-T3-I51"	"The proportion of children whose births are registered within the first year"	"87.5"^^xsd:int	<u>"90.7"^^xsd:int</u>
3	"87.5"^^xsd:int	<u>"agenda2030"</u>	"agenda 2030"	"87.5"^^xsd:int	"A6-G16-P2-T3-I51"	"The proportion of children whose births are registered within the first year"	"87.5"^^xsd:int	<u>"79"^^xsd:integer</u>
4	"87.5"^^xsd:int	<u>"pillar4"</u>	"Governance, Peace & Security"	"87.5"^^xsd:int	"A6-G16-P2-T3-I51"	"The proportion of children whose births are registered within the first year"	"87.5"^^xsd:int	<u>"79"^^xsd:integer</u>
5	"87.5"^^xsd:int	<u>"agenda2030"</u>	"agenda 2030"	"87.5"^^xsd:int	"A6-G16-P2-T3-I51"	"The proportion of children whose births are registered within the first year"	"87.5"^^xsd:int	<u>"88"^^xsd:integer</u>
6	"87.5"^^xsd:int	<u>"pillar4"</u>	"Governance, Peace & Security"	"87.5"^^xsd:int	"A6-G16-P2-T3-I51"	"The proportion of children whose births are registered within the first year"	"87.5"^^xsd:int	<u>"88"^^xsd:integer</u>
7	"87.5"^^xsd:int	<u>"agenda2030"</u>	"agenda 2030"	"87.5"^^xsd:int	"A6-G16-P2-T3-I51"	"The proportion of children whose births are registered within the first year"	"87.5"^^xsd:int	<u>"89"^^xsd:integer</u>
8	"87.5"^^xsd:int	<u>"pillar4"</u>	"Governance, Peace & Security"	"87.5"^^xsd:int	"A6-G16-P2-T3-I51"	"The proportion of children whose births are registered within the first year"	"87.5"^^xsd:int	<u>"89"^^xsd:integer</u>
9	"87.5"^^xsd:int	<u>"agenda2030"</u>	"agenda 2030"	"87.5"^^xsd:int	"A6-G16-P2-T3-I51"	"The proportion of children whose births are registered within the first year"	"87.5"^^xsd:int	<u>"91"^^xsd:integer</u>
10	"87.5"^^xsd:int	<u>"pillar4"</u>	"Governance, Peace & Security"	"87.5"^^xsd:int	"A6-G16-P2-T3-I51"	"The proportion of children whose births are registered within the first year"	"87.5"^^xsd:int	<u>"91"^^xsd:integer</u>

Figure 7. 11: Sample Generic Knowledge Resource Mining Output 2 (SDGI 16.9.1)

Figure 7.11 displays the last part of Figure 7.10 output and therefore, comprises one record. Conversely, the data values marked red in column 2 represent duplicate extraction and the source of the duplication appears in Line 32 of Figure 7.9, the generic ontology concept representation limitation – lack of custom properties to differentiate SDG local relation to a specific SDG Agenda (*agenda2030*) and external agenda relation to Vision Agenda (*pillar4*). Thus, these 2 different relations share the same object property definition. The second duplication is reflected in column 8 and the source of this duplication appears in Line 12 of Figure 7.9. The second duplication results from the lack of unique disaggregation constructs to differentiate between the different value disaggregation. The five disaggregated values (Figure 4.4-4.5) share the same property definition resulting in the same property repeating with the same metadata property constructs.

All in all, the generic knowledge resource mining case (Figure 7.9) demonstrates semantic enrichment attributes in the form of semantic capabilities. The case starts with the selection of a sample instance (*sdgiDataInstance 16.9.1*) of a knowledge entity (*sdgi*) in Lines 4-5. That is followed by the extraction of published SDGI data connected



to this instance (Lines 6-18) and the extraction of other relating semantic metadata for SDGI data (Lines 19-40). So, the output displayed in Figures 7.10-7.11 is enabled by the semantic value-added to flat text resources in the form of semantic structure (syntax) to enhance semantic capabilities as follows.

This semantic value mining case demonstrates *accessibility* capabilities in terms of the degree of published SDGI data access to relating web resources. The semantic metadata resources relating to published SDGI data reflected in Figure 7.7 (Subsection 7.4.1) provide a diversity of semantic access points for the selected `sdgiDataInstance`. Specifically, `sdgiData` (published SDGI data) resource' capability to access relating semantic resources being SDGI (Line 19), SDG Target (Line 25), SD Goal (Line 31), Vision Pillar (Line 33), Africa Agenda Indicator (Line 35), and the SDGIO (Lines 27 and 29) metadata resources.

The *understandability* capabilities are demonstrated in terms of the degree of clarity effected with the use of generic ontologies. For instance, Lines 4-5 specifies the sample instance (SDGI 16.9.1) from the `sdgiData` resource, and the corresponding data is extracted as specified in Lines 6-18. This data is clarified with SDGI descriptive semantics (Lines 20 and 22), development dimension and dimension criteria specification semantics (Lines 23-24), relating SDG Target semantics (Line 26), SDG Agenda and Vision Agenda semantics (Line 34), Africa Agenda semantics (Lines 36-39), and SDGIO semantics (Lines 28 and 30).

The *interoperability* capabilities are demonstrated by the resource's degree of data sharing potential. The sharing of values across agendas is demonstrated in Line 40 where `sdgiPerfVal` (SDGI performance value) is shared among `afriPerfVal` (Africa Agenda Indicator performance value) and `visPillarPerfVal` (Vision Pillar performance value). Africa Agenda Indicator is mapped to SDGIs (Line 35), and SDGs are mapped to Vision Pillars (Line 32).

This mining case further demonstrates *integrability* capabilities. The semantic access points discussed above (accessibility capability) represent the linkage capabilities. Overall, the generic knowledge resource offers interlinkages to a wide web resource base comprising *w3.org*, *unep.org*, *allotrope.org*, *semanticscience.org*, *purl.org* (*dcterms*), *dbpedia.org*, *obolibrary.org*, and *geneontology.org*. In addition, the mining case



demonstrates the degree of resource linkages within the Botswana Integrated Indicator Framework and to external web resources. The `sdgiDataInstance` links to SDGI metadata through the `SDGIO:00000242` (Line 19); SDG target metadata through `SDGIO:00000074` (Line 25); SD Goal metadata; SDG agenda metadata and Vision Pillar metadata through `dcterms:identifier` (Lines 31, 33); SDGIO metadata through `hasExactSynonym` (Line 29) and `hasDbXref` (Line 30); and Africa Agenda metadata through `dcterms:relation` (Line 35) and the SDGIO (Lines 27 and 29) metadata resources.

Lastly, the *intelligibility* capabilities are demonstrated by the resource's degree to analyse performance based on specific criteria. For instance, the query extracts performance data that applies to a given development dimension and that is further screened by theme in Lines 21 and 23-24. For the sample instance, SDG 16 is of social type of sustainability and the theme is *Peace, Justice and Strong Institutions*. Thus, the multi-dimensional view of the data pertains to these two dimensions (Lines 14-15).

Domain Knowledge Resource Mining Case

The evaluation case reflected in Figures 7.12-7.13 (Lines 1-70) constitutes two parts of one query (the separation is for better image visibility). It mines the knowledge embodied in the integrated domain knowledge resource (Subsection 7.4.1) as the semantic value-added to flat text resources for SDGI 16.9.1 extract.



```

1 SELECT ?sdgiPerfDesc ?sdgiPerfYr ?sdgiPerfValDesc ?sdgiCode ?sdgiDesc ?formula ?dataSrc ?dataColMeth ?dataColFreq
2 ?unitMeas ?sdgiPerfVal ?valDisagg1Desc ?valDisagg1 ?valDisagg2Desc ?valDisagg2 ?valDisagg3Desc ?valDisagg3
3 ?valDisagg4Desc ?valDisagg4 ?valDisagg5Desc ?valDisagg5
4 WHERE {
5   <http://www.statsbots.org/bw/data/sdgi/goal16/2030_16.9.1> <http://localhost/8080/npio/hasInstanceCode> ?sdgiDataInstance;
6   <http://localhost/8080/npio/hasAgVal-birthReg> ?sdgiPerfVal;
7   <http://localhost/8080/npio/hasPerfYr> ?sdgiPerfYr;
8   <http://localhost/8080/npio/hasNpiCode> ?sdgiCode;
9   <http://localhost/8080/npio/hasNpiDesc> ?sdgiDesc;
10  <http://localhost/8080/npio/hasBirthRegDsagVal-0yr> ?valDisagg1;
11  <http://localhost/8080/npio/hasBirthRegDsagVal-1yr> ?valDisagg2;
12  <http://localhost/8080/npio/hasBirthRegDsagVal-2yr> ?valDisagg3;
13  <http://localhost/8080/npio/hasBirthRegDsagVal-3yr> ?valDisagg4;
14  <http://localhost/8080/npio/hasBirthRegDsagVal-4yr> ?valDisagg5;
15  <http://localhost/8080/npio/hasUnitMeasure> ?unitMeas.
16  ?sdgiInstance <http://localhost/8080/npio/hasInstanceCode> ?sdgiDataInstance;
17  <http://localhost/8080/npio/hasNpiName> ?sdgiName;
18  <http://localhost/8080/npio/hasSubTheme> ?subTheme;
19  <http://localhost/8080/npio/hasComputeDesc> ?compDesc;
20  <http://localhost/8080/npio/hasFormula> ?formula;
21  <http://localhost/8080/npio/hasFormulaDesc> ?formulaDesc;
22  <http://localhost/8080/npio/hasValDsagFactor> ?disaggFactor;
23  <http://localhost/8080/npio/hasDataSrc> ?dataSrc;
24  <http://localhost/8080/npio/hasDataColFreq> ?dataColFreq;
25  <http://localhost/8080/npio/hasDataColMeth> ?dataColMeth;
26  <http://localhost/8080/npio/hasDevtDim> ?devtDimInstance;
27  <http://localhost/8080/npio/hasMeasDim> ?measDimInstance;
28  <http://localhost/8080/npio/hasScopeDim> ?scopeDimInstance;
29  <http://localhost/8080/npio/hasSdgi-sdgTargetRel> ?sdgTargetInstance;
30  <http://localhost/8080/npio/hasSdgiPerf> ?sdgiPerfInstance.
31  ?sdgTargetInstance <http://localhost/8080/npio/hasTargetCode> ?sdgTargetCode;
32  <http://localhost/8080/npio/hasTargetDesc> ?sdgTargetDesc;
33  <http://localhost/8080/npio/hasSdgTarget-sdGoalRel> ?sdGoalInstance.
34  ?sdGoalInstance <http://localhost/8080/npio/hasGoalCode> ?sdGoalCode;
35  <http://localhost/8080/npio/hasGoalDesc> ?sdGoalDesc;
36  <http://localhost/8080/npio/hasSdGoal-sdgAgendaRel> ?sdgAgendaInstance;
37  <http://localhost/8080/npio/hasSdGoal-pillarMap> ?visPillarInstance;
38  <http://localhost/8080/npio/hasThemeDim> ?themeDimInstance.
39  ?sdgAgendaInstance <http://localhost/8080/npio/hasAgendaCode> ?sdgAgendaCode;
40  <http://localhost/8080/npio/hasAgendaDesc> ?sdgAgendaDesc.

```

Figure 7. 12: Domain Knowledge Resource Mining 1 (SDGI 16.9.1)

```

41 ?visPillarInstance <http://localhost/8080/npiOnto/hasPillarCode> ?visPillarCode;
42     <http://localhost/8080/npiOnto/hasPillarDesc> ?visPillarDesc.
43 ?afriInstance <http://www.w3.org/2002/07/owl#sameAs> ?sdgiInstance;
44     <http://localhost/8080/npiOnto/hasNpiCode> ?afriCode;
45     <http://localhost/8080/npiOnto/hasNpiDesc> ?afriDesc.
46 ?sdgiPerfInstance <http://localhost/8080/npiOnto/hasPerfCode> ?sdgiPerfCode;
47     <http://localhost/8080/npiOnto/hasPerfDesc> ?sdgiPerfDesc.
48 ?devtDimInstance <http://localhost/8080/npiOnto/hasDimCode> ?devtDimCode;
49     <http://localhost/8080/npiOnto/hasDimName> ?devtDim;
50     <http://localhost/8080/npiOnto/hasDimName> "social development".
51 ?measDimInstance <http://localhost/8080/npiOnto/hasDimCode> ?measDimCode;
52     <http://localhost/8080/npiOnto/hasDimName> ?measDim;
53     <http://localhost/8080/npiOnto/hasDimName> "base measure".
54 ?scopeDimInstance <http://localhost/8080/npiOnto/hasDimCode> ?scopeDimCode;
55     <http://localhost/8080/npiOnto/hasDimName> ?scopeDim;
56     <http://localhost/8080/npiOnto/hasDimName> "national scope".
57 ?themeDimInstance <http://localhost/8080/npiOnto/hasDimCode> ?themeDimCode;
58     <http://localhost/8080/npiOnto/hasDimName> ?themeDim;
59     <http://localhost/8080/npiOnto/hasDimName> "peace, justice and strong institutions".
60 ?sdgiIndInstance <http://purl.unep.org/sdg/SDGIO_00000242> ?sdgiCode;
61     <http://purl.unep.org/sdg/SDGIO_00000227> ?unsdCode.
62 ?sdgiGoalInstance <http://www.geneontology.org/formats/oboInOwl#hasExactSynonym> ?sdGoalCode;
63     <http://www.geneontology.org/formats/oboInOwl#hasDbXref> ?sdgUndbRef.
64     <http://localhost/8080/npiOnto/hasAgVal-birthReg> <http://www.w3.org/2000/01/rdf-schema#comment> ?sdgiPerfValDesc.
65     <http://localhost/8080/npiOnto/hasBirthRegDsagVal-0yr> <http://www.w3.org/2000/01/rdf-schema#comment> ?valDisagg1Desc.
66     <http://localhost/8080/npiOnto/hasBirthRegDsagVal-1yr> <http://www.w3.org/2000/01/rdf-schema#comment> ?valDisagg2Desc.
67     <http://localhost/8080/npiOnto/hasBirthRegDsagVal-2yr> <http://www.w3.org/2000/01/rdf-schema#comment> ?valDisagg3Desc.
68     <http://localhost/8080/npiOnto/hasBirthRegDsagVal-3yr> <http://www.w3.org/2000/01/rdf-schema#comment> ?valDisagg4Desc.
69     <http://localhost/8080/npiOnto/hasBirthRegDsagVal-4yr> <http://www.w3.org/2000/01/rdf-schema#comment> ?valDisagg5Desc.
70 BIND(?sdgiPerfVal AS ?visiPerfVal) BIND(?sdgiPerfVal AS ?afriPerfVal) }

```

Figure 7.13: Domain Knowledge Resource Mining 2 (SDGI 16.9.1)

The output for the domain knowledge resource mining represents selected variables specified in Lines 1-3 from the variable pool reflected in Figures 7.12-7.13. For visibility, the output is displayed in three separate interfaces (Figures 7.14-7.16). The first output (Figure 7.14) displays extracted metadata for SDGI 16.9.1 while the second output (Figure 7.15) displays corresponding performance value extractions. Finally, the corresponding disaggregation for performance value extractions are displayed in Figure 7.16.



sdgiPerfDesc	sdgiPerfYr	sdgiPerfValD	sdgiCode	sdgiDesc	unsdCode	sdgUnDbRe	formula	unitMeas	compDesc	dataSrc	dataColMeth	dataColFre
1 "Performance value for SDGI 16.9.1"	"2015"^^xsd:integer	"aggregate value for birth registration"	"16.9.1"	"Proportion of children under 5 years of age whose births have been registered with a civil authority, by age"	"C160901"	"http://www.un.org/sustainabledevelopment/peace-justice/"	"(under-5 registered births/total under-5 births)*100"	"%"	"No. of children under 5 yrs whose births are registered with relevant national civil authorities"	"Censuses, household surveys & national civil registration systems"	"Annual country consultation"	"annual"

Showing 1 to 1 of 1 entries

Figure 7. 14: Sample Domain Knowledge Resource Mining Output 1 (SDGI 16.9.1)

Figure 7.14 output illustrates the integration of raw OD publications with corresponding metadata from both the organisational and global reference resources through semantic enrichment processing. This figure represents data pulled from different flat text resources and in different formats integrated into a single knowledge resource through semantic enrichment processing. For instance, `sdgiPerfDesc` (Line 47), `sdgiPerfYr` (Line 7), `sdgiPerfValDesc` (Line 64), `sdgiCode` (Line 8), `sdgiDesc` (Line 9), `unitMeas` (Line 15) are NPIOnto custom (semantic) constructs for the definition of raw OD published in heterogeneous flat text formats. The `unsdCode` (Line 61) and the `sdgUnDbRef` (Line 63) are SDGIO semantic constructs. Finally, `formula` (Line 20), `compDesc` (Line 19), `dataSrc` (Line 23), `dataColMeth` (Line 25), `dataColFreq` (Line 24) are also NPIOnto custom constructs for semantic definition of published SDG metadata availed in pdf file format. Figure 7.15 output extends Figure 7.14 output.

sdgTargetCod	sdgTargetDes	sdGoalCode	sdGoalDesc	sdgiPerfVal	afriCode	afriDesc	afriPerfVal	visPillarCode	visPillarDesc	visiPerfVal
1 "16.9"	"By 2030, provide legal identity for all, including birth registration"	"SDG16"	"Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels"	"87.5"^^xsd:int	"A6-G16-P2-T3-I51"	"proportion of children whose births are registered within the first year"	"87.5"^^xsd:int	"pillar4"	"Governance, Peace & Security"	"87.5"^^xsd:int

Showing 1 to 1 of 1 entries

Figure 7. 15: Sample Domain Knowledge Resource Mining Output 2 (SDGI 16.9.1)



Figure 7.15 displays the sharing of data across agendas according to the mappings reflected in Figures 4.11-4.12. The SDG Agenda performance values in Figures 4.4-4.5 and metadata are semantically linked to Africa Agenda metadata and Vision Agenda metadata. The SDGI performance value (column 5) is shared across agendas according to the Botswana Integrated Indicator Framework mappings, specifically, Africa Agenda (column 8) and Vision Agenda (column 11). The last ten output variables, the five SDGI value disaggregation descriptions and corresponding disaggregated values, shown in Figure 7.16 demonstrate value disaggregation through semantic enrichment processing.

	sdgiPerfVal	valDisagg1Desc	valDisagg1	valDisagg2Desc	valDisagg2	valDisagg3Desc	valDisagg3	valDisagg4Desc	valDisagg4	valDisagg5Desc	valDisagg5
1	"87.5"^^xsd:float	"The 0 year disaggregated value for birth registration"	"79"^^xsd:integer	"The 1 year disaggregated value for birth registration"	"89"^^xsd:integer	"The 2 year disaggregated value for birth registration"	"88"^^xsd:integer	"The 3 year disaggregated value for birth registration"	"90.7"^^xsd:float	"The 4 year disaggregated value for birth registration"	"91"^^xsd:integer

Showing 1 to 1 of 1 entries

Figure 7. 16: Sample Domain Knowledge Resource Mining Output 3 (SDGI 16.9.1)

Figure 7.16 depicts value disaggregation in the form of semantic representation of the different categories/classes of data reflected in the raw OD publications (Figures 4.4-4.5). The disaggregation capabilities are custom defined in NPIOnto. All the data values in Figure 7.16 (columns 3, 5, 7, 9, 11) are shown in Figures 4.4-4.5 while value descriptions (columns 2, 4, 6, 8, 10) are NPIOnto custom constructs to accompany the values with more descriptive semantics.

All in all, in comparison to the flat text resources for the sample instance SDGI 16.9.1 (Chapter Four), the knowledge harvest shown in Figures 7.14-7.16 is a result of semantic capabilities added to the knowledge resource (Figure 7.6). The knowledge mining case demonstrates the knowledge resource' accessibility, understandability, interoperability, integrability, and intelligibility capabilities in Figures 7.12 (Lines 1-40) and 7.13 (Lines 41-70) as follows. The mining case starts with the selection of a sample instance `sdgiDataInstance` (16.9.1) of a knowledge entity (`sdgi`) in Line 5. That is followed by the extraction of published SDGI data connected to this instance (Lines 5-15) and the extraction of other relating semantic metadata for SDGI data (Lines 16-70). This mining case, thus, answers competency questions (CQs) stated in Subsection 6.3.1 as follows.



CQ1: What are the accessibility capabilities of OD resources enriched with NPIOnto?

The resource' *accessibility* capabilities are demonstrated by the degree of published SDGI data access to NPIOnto constructs. The NPIOnto constructs reflected in Figure 6.7 (classes and properties) provide a diversity of semantic access points for the `sdgiDataInstance`. For instance, the `sdgiData` resource accesses the NPIOnto constructs at four levels. The first level pertains to the `sdgiData` resource's access of SDG Agenda metadata resources (Lines 29, 30, 33 & 36). The second level of access extends to the metadata for relating agendas being Vision Agenda and Africa Agenda (Lines 37 & 43). The third level pertains to the `sdgiData` resource access of NPIOnto dimensions (Lines 26, 27, 28 & 38). The fourth level further extends access to relating resources in the Web of Data (Lines 61 & 63). Therefore, the accessibility evaluation has proved that OD, semantically enriched with an ontology, supports flexible and dynamic access to individual data values for both humans and machines than OD published in flat text format or in non-enriched semantic formats (e.g., XML).

CQ2: What are the understandability capabilities of OD resources enriched with NPIOnto?

The resource' *understandability* capabilities are demonstrated by the degree of NPIOnto disambiguation and disaggregation capabilities. A domain specific vocabulary such as NPIOnto allows for specific metadata definition in terms of attribute, measurement, data, and relationship descriptive semantics, as well as specific value disaggregation. Thus, domain specific knowledge resources, with their in-depth semantic enrichment value, offer maximal clarity when compared to flat text resources. The demonstration in terms of NPIOnto disambiguation and disaggregation capabilities is as follows. Regarding NPI disambiguation evaluation, the query extracts performance descriptions for `sdgiDataInstance` (SDGI 16.9.1) from the `sdgiData` resource (Lines 5-15). This performance data is first clarified with descriptive semantics for classes `sdgi` (Lines 17-30), `sdgTarget` (Lines 32-33), `sdGoal` (Lines 35-38), `sdgAgenda` (Line 40), `visPillar` (Line 42), `afri` (Lines 44-45), and `sdgiPerf` (Line 47). Secondly, it is clarified with disaggregate value descriptive semantics (Lines 64-69). Thirdly, it is clarified with diverse dimensional views (Lines 48-59), and, finally, with relating SDGIO metadata (Lines 61 and 63). Thus, this mining case creates a pool of knowledge that spans the



different agendas and relating web resources for holistic understanding of a given knowledge instance. Therefore, the understandability evaluation has proved that SDGI performance, enriched with SDGI metadata such as attribute, measurement, data, and relationship descriptive semantics, provides a holistic performance view. In addition to the data descriptive semantics, data disaggregation offers an extended level of SDGI performance disambiguation. A holistic view of SDGI performance and an extended level of its disambiguation both contribute to better OD clarity.

CQ3: What are the interoperability capabilities of OD resources enriched with NPIOnto?

The resource' *interoperability* capabilities are demonstrated by the resource's degree of data sharing potential. The interoperability evaluations have proved that semantic data, enriched with an ontology enables data sharing in support of the consolidated indicator framework implementation through linkages to NPIOnto and other external resources such as the SDGIO, w3.org, unep.org, and geneontology.org. It is also demonstrated by sharing `sdgiPerfVal` among `afriPerfVal` and `visPillarPerfVal` (Line 70).

CQ4: What are the integrability capabilities of OD resources enriched with NPIOnto?

The resource' *integrability* capabilities are demonstrated for the degree of NPIOnto constructs interlinkages within the Botswana Integrated Indicator Framework and to external web resources. The mining case demonstrates the `sdgiData` resource linkage to NPIOnto constructs at four levels.

The first level pertains to connections between `sdgiDataInstance` and SDG Agenda metadata resources and the links are as follows: `sdgiDataInstance` links to class `sdgi` instances through `hasInstanceCode` (Line 16); class `sdgi` instances link to class `sdgTarget` instances through `hasSdgi-sdgTargetRel` (Line 29); class `sdgTarget` instances link to class `sdGoal` instances through `hasSdgTarget-sdGoalRel` (Line 33); class `sdGoal` instances link to class `sdgAgenda` instances through `hasSdGoal-sdgAgendaRel` (Line 36); and class `sdgi` instances link to class `sdgiPerf` instances through `hasSdgiPerf` (Line 30).

The second level of interlinkages extends to the metadata for relating agendas being Vision Agenda and Africa Agenda. Instances of class `sdGoal` link to class `visPillar`



instances through `hasSdGoal-pillarMap` (Line 37) and instances of class `sdgi` link to class `afri` instances through the `sameAs` property (Line 43).

The third level of interlinkages connect `sdgiDataInstance` to NPIOnto dimensions, specifically, instances of class `devtDim` through `hasDevtDim` (Line 26); instances of class `measDim` through `hasMeasDim` (Line 27); and instances of class `scopeDim` through `hasScopeDim` (Line 28). Instances of class `sdGoal` link to class `themeDim` through `hasThemeDim` (Line 38).

The fourth level of interlinkages further extends connections to relating resources in the Web of Data. Instances of class `sdgi` link to SDGIO annotations on class `sdgi` instances through the `SDGIO:00000227` annotation (Line 61) for `unsdCode` while instances of class `sdGoal` link to SDGIO annotations on class `sdGoal` instances through `hasDbXref` annotation (Line 63) for `sdgUnDbRef`.

CQ5: What are the intelligibility capabilities of OD resources enriched with NPIOnto?

The resource *intelligibility* capabilities are demonstrated by the degree of NPIOnto enhanced analytical capability demonstrations. The domain knowledge resource mining case demonstrates the extension of the intelligibility capability of the generic knowledge resource mining case (Figure 7.9) through NPIOnto customised dimensions, specifically, the addition of the measurement and scope dimensions to the development and theme dimensions. So, this mining case further screens the output for the mining criteria of social development, related to SDG 16 theme (Lines 48-50, 57-59) to extract only performance instances that are *base* measures and of the *national* measurement scope (Lines 51-56).

CQ6: What is the overall semantic capabilities enrichment impact on semantic publication and consumption?

All in all, flat text OD publications are limited in accessibility, understandability, interoperability, integrability, and intelligibility capabilities. This hampers OD consumption, specifically, the knowledge creation, mining and use potential. NPIOnto can be leveraged through the LOD-driven semantic enrichment processing to address this flat text OD publications' ineffectiveness challenge by standardising and upgrading/enhancing the semantic capabilities of flat text OD resources. This research has proved that the deliverables of semantic enrichment, knowledge resources (i.e., OD



resources enriched with corresponding metadata and linkages across the Web of Data) have potential for increasing resources' accessibility, understandability, interoperability, integrability, and intelligibility capabilities to enable semantic publication and consumption.

7.5 Conclusion

Chapter Seven fulfils Research Objective 3 which involved the demonstration of the semantic enrichment framework. This chapter answers Research Question 3 which is:

What are the findings of the preliminary semantic enrichment framework demonstrations?

This research question sought to establish key issues pertaining to the entire semantic enrichment process modelling to factor into the final ODSE Framework design. The deliverable of this demonstration is the preliminary ODSE Framework which is the integration of three stages pertaining to ODSE pre-processing, processing and post-processing into a holistic semantic enrichment solution for the transition from static, flat text OD resources with limited semantic capabilities and value to flexible, knowledge resources with unlimited semantic capabilities and value.

The first stage, *ODSE pre-processing*, has three components that address the input data being the pre-processing input, implementation, and output components. This stage pertains to the extraction of OD from source which is a web platform and its preparation for semantic enrichment. These preparations involve unifying the data and conducting some quality operations on these data to produce the output that serve as the input to the semantic enrichment stage.

The second stage, ODSE processing, has three components being the ODSE approach, data enrichment, and enrichment output components. The ODSE approach deals with the choice of suitable generic ontologies (generic ODSE approach) and the development of custom domain ontologies/vocabularies (domain ODSE approach) for semantic enrichment. The semantic enrichment component starts with the choice of the ODSE approach which may either be a generic ODSE approach that makes use of generic ontologies or a domain ODSE approach that makes use of custom domain ontologies. Within the chosen ODSE approach is performed the data enrichment process entailing



the resource reconciliation and the resource RDFization. The end product are generic and domain knowledge resources which are both machine readable and processable.

The third stage, ODSE post-processing, has three components being the knowledge resources integration into single resource for ease of manipulation; knowledge resources publication to a SPARQL endpoint for query; and knowledge resource evaluation for semantic capabilities value-added to flat text resources. The mining of integrated knowledge resources for the degree of knowledge value-added to raw OD resources, revealed that the domain ODSE approach to semantic enrichment produces richer knowledge resources compared to the generic ODSE approach. Also, the mining of integrated knowledge resources for the degree of semantic capabilities value-added to raw OD resources, revealed greater degrees of the knowledge resource' accessibility, understandability, integrability, interoperability, and intelligibility for the domain ODSE approach.

Overall, as an extended analysis, the integrated knowledge resources mining output served as the basis for the evaluation of the overall effectiveness of the ODSE Framework, presented in Subsection 8.3.3.



CHAPTER EIGHT: FRAMEWORK EVALUATION AND FINALISATION

8.1 Introduction

Chapter Eight presents research activities specified in Table 3.2, pertaining to framework evaluation (Phase 4) and framework finalisation (Phase 5). To start with, the preliminary ODSE Framework (Chapter Five) was designed according to the adapted DSR methodology (Chapter Three) from a combination of reviewed literature on semantic enrichment (Chapter Two) and semantic publication and consumption needs (Chapter Four). It was then demonstrated (Chapters Six and Seven) so as to ascertain its completeness, cohesiveness and effectiveness (Chapter Eight). These three evaluation dimensions pertain to framework verification and validation. Framework verification embraces the evaluation of the completeness and cohesiveness of the framework while framework validation pertains to the evaluation of the effectiveness of the framework. Lastly, framework finalisation pertains to the generalisation of the framework for reusability on the basis of the preliminary framework demonstration.

8.2 Framework Verification

Framework verification was done to assess the degree of completeness of the elements that constitute the preliminary semantic enrichment framework processes and the cohesiveness of those process elements as well. The reviews on the completeness of the preliminary framework process elements pertained to the assessment of the degree of inclusion of semantic enrichment entities comprising processes, approaches, capabilities, techniques, technologies and their relations for each framework stage. The reviews on the cohesiveness of the framework elements pertained to assessment of process flows and involved the determination of the degree of connectivity and flow of both the individual stage elements and the integrated elements per process for all stages. Framework verification was executed according to the *framework verification method* explained in the methodology chapter (Subsection 3.3.4) pertaining to manual reviews of the framework by reconciling processes and corresponding elements to their actual demonstrations and identifying the differences. Table 8.1 provides a summary of the evaluations.



Table 8. 1: Preliminary ODSE Framework Process Evaluation

Framework Stage	Framework Process	Process Demonstration	Process Reconciliation
Stage 1 (OD Pre-processing)	Collection of Input	- Extraction of raw OD	1. Addition of the pre-processing output component.
	Pre-processing Transformations (Data Cleaning & transformation)	- Data editing - Data re-organisation - Data re-formatting - conversions from 1-star to 2-star publication format - spreadsheet files	
		Production and consolidation of Pre-processing Output	
Stage 2 (OD Semantic Enrichment Processing)	Data Integration - RDF-driven Integration (RDF resources) - Triplestore Integration (RDF Resource Integration) - Ontology-driven Integration (Integration of the 4 Agendas) Data Enrichment - Ontology-driven enrichment with the use of ontology constructs	Ontology Demonstrations - Choice of <i>Generic</i> Ontologies - Development of NPI Domain Ontology Data Enrichment Demonstrations - Resource Reconciliation - Resource RDFization	2. <i>Ontology-driven integration</i> – demonstrated with the development of NPI domain ontology 3. <i>RDF-driven Integration & ontology-driven semantic enrichment</i> – demonstrated with Resource RDFization 4. <i>Triplestore Integration</i> - demonstrated with the consolidation of semantic enrichment output 5. Renaming the ontology-driven integration component to ODSE approach component for process generalisation 6. Addition of raw OD reconciliation process
		Production and consolidation of semantic enrichment output	
Stage 3 (OD Post-processing)	Production of output (Intelligent Linked OD)	Resource structuring Resource deployment	7. Addition of knowledge resources structuring 8. Addition of knowledge resources deployment

The demonstrations brought some clarifications to the ODSE Framework stages that were key to the preliminary framework finalisation into a generalised semantic enrichment framework. Firstly, Stage 1 (OD pre-processing) verification revealed a need to factor in the pre-processing output component in the framework to serve as the link to the semantic enrichment processing stage. Secondly, Stage 2 (OD semantic enrichment processing) verification revealed a need to factor in the ODSE approach component that embraces both the development of a custom domain ontology and the selection of generic



ontologies. In addition, for more efficient semantic enrichment results, resource reconciliation was factored into the semantic enrichment process. Furthermore, for clarity, the development of NPI domain ontology demonstrated the ontology-driven integration. Moreover, the *Semantic Integration* and the *Data Enrichment* components in the preliminary framework were streamlined into one process as the two are implemented through one process – RDFization. Lastly, Stage 3 (OD post-processing) verification revealed a need for the ODSE Framework to reflect on how the final output is prepared (knowledge resources structuring) and availed (knowledge resources deployment) for public consumption.

8.3 Framework Validation

The ODSE Framework validation involved an assessment of the degree of framework effectiveness in semantic value (knowledge) creation and usability potential for the three stages summarised by Figure 3.1 in the methodology chapter. The *framework validation method* explained in the methodology chapter (Subsection 3.3.4) was adopted in the semantic value assessment aimed at establishing the semantic enrichment value added to flat text resources in terms of knowledge and semantic capabilities. The evaluations involved a comparison of embodied knowledge and semantic capabilities of the ODSE Framework input (Stage 1) to the ODSE Framework processing (Stage 2) and the ODSE Framework output (Stage 3). Therefore, the ODSE Framework validation comprised a combination of Stage 1 input and output - flat text resources discussed in Section 4.3, Stages 2 and 3 output – knowledge resources (semantically enriched OD resources discussed in Subsection 7.3.2 and 7.4), knowledge resource mining cases (presented in Figures 7.9 and 7.12 -7.13), and knowledge resource mining output (presented in Figures 7.10-7.11 and 7.14-7.16).

The ODSE Framework validation scenario narrowed down the evaluations to a sample performance indicator, SDGI 16.9.1. Therefore, the set of flat text OD resources for SDGI 16.9.1 comprises:

- 1 Figure 4.2 - sample published SDGI metadata in HTML image form,
- 2 Figure 4.3 - sample published SDGI metadata in HTML text form,



- 3 Figures 4.4 and 4.6 - sample published SDGI data and metadata in HTML graphic form (tabular/charts),
- 4 Figure 4.5 - sample published SDGI data and metadata in spreadsheet text form,
- 5 Figure 4.7 - sample published SDGI metadata in link form (API Snippet),
- 6 Figure 4.8 - sample published and unpublished SDGI metadata in pdf text form,
- 7 Figure 4.9 - sample published Vision Agenda metadata in pdf text form,
- 8 Figure 4.10 - sample published SDGI data and metadata in semantic form (XML),
- 9 Figure 4.11 - sample unpublished metadata on SDG Agenda-Vision Agenda mapping (pdf), and
- 10 Figure 4.12 - sample unpublished metadata on Africa Agenda-SDG Agenda mapping (pdf).

Performance values for SDGI 16.9.1, published in flat text OD format represented by Figures 4.4-4.6 and 4.10 contain SDGI 16.9.1 performance values while the metadata files, represented by all other figures listed above, contain SDGI attributes and relations in raw form. The SDGI 16.9.1 performance values pertain to children's birth registration details. The global agenda (SDG Agenda) values are to be shared across the other three relating development agendas: the local agendas being Vision Agenda and NDP Agenda and the regional agenda, Africa Agenda. The sharing of the values are to be implemented according to the mappings defined in Figures 4.11-4.12. Tracking or exchanging these data across these agendas for valuable use, necessitates communication among the different pieces of data contained in the resource files. The communication is enabled by semantic capabilities that support a common data format, classification, organisation, presentation, clarity, compatibility, connectivity, analysis, and comparability structure. Therefore, the data and relating metadata represented by all the figures above were semantically transformed, enriched and consolidated into integrated knowledge resources per ODSE approach. These resulted in additional resources in semantic format (knowledge resources) being:

- 11 Figure 7.5 - sample semantically enriched OD resource according to the generic ontology-driven semantic enrichment approach – generic ODSE approach (generic knowledge resource), and



12 Figure 7.6 - sample semantically enriched OD resource according to the custom domain ontology-driven semantic enrichment approach – domain ODSE approach (domain knowledge resource).

In summary, the data and metadata on SDGI 16.9.1 are contained in a diversity of resources comprising image, HTML, graphic, PDF, spreadsheet, API snippets, XML, domain knowledge and generic knowledge resources. Figure 8.1 presents this knowledge resources diversity in graphical form, representing the *ODSE Framework validation scenario*.

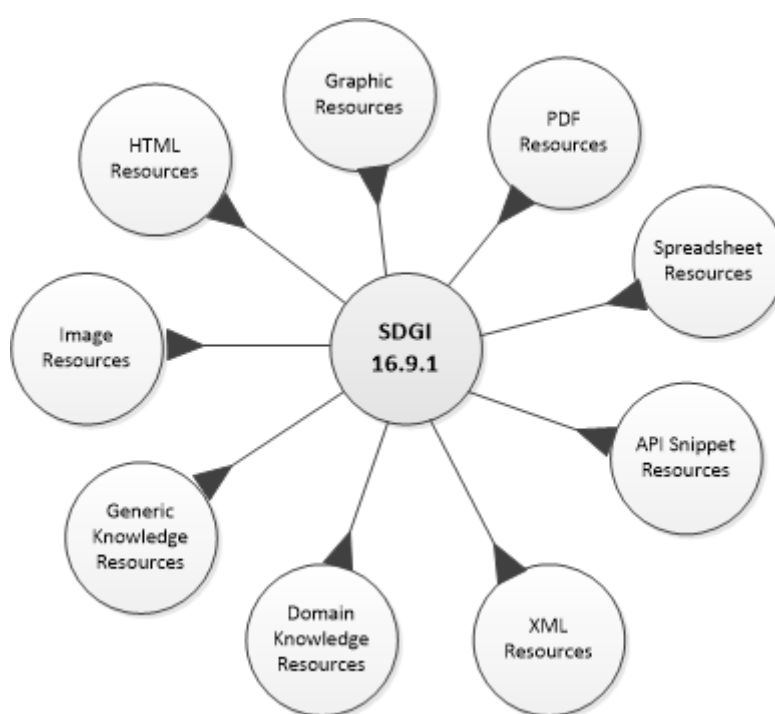


Figure 8. 1: *ODSE Framework Validation Scenario (SDGI 16.9.1)*

As shown in Figure 8.1, the *ODSE Framework validation scenario* is a comparative analysis of the potential OD value (knowledge) for all raw OD resources and corresponding created knowledge resources, in their diverse formats.

Overall, these resources diversity, together with the *semantic usability assessment* tool discussed in the methodology (Subsection 3.3.4, Figure 3.2) constitute the input to the evaluation. The ODSE Framework validation scenario is expressed by the following question:

Which OD resource has the potential to maximise OD value (knowledge) creation and usability capabilities?



As shown in Figure 3.1 (Subsection 3.3.4), this overarching question was broken down to each ODSE Framework stage assessment as follows:

- To what degree does the ODSE Framework Stage 1 processing facilitates the creation of OD value (knowledge) and semantic capabilities?
- To what degree does the ODSE Framework Stage 2 processing facilitates the creation of OD value (knowledge) and semantic capabilities?
- To what degree does the ODSE Framework Stage 3 processing facilitates the creation of OD value (knowledge) and semantic capabilities?

Formulating answers to these questions involved OD resources' (Figure 8.1) comparisons on the basis of semantic enrichment, richness and mining attributes per knowledge resource for the sample performance indicator instance (SDGI 16.9.1). Comparisons were implemented by assigning semantic values to the five types of OD publications based on semantic usability attributes (enrichment, richness and mining attributes) to determine the degree of effectiveness of semantic usability of a given OD resource at each framework stage.

8.3.1 The ODSE Pre-processing (Stage 1) Validation [Flat Text Resources]

The *Framework Validation for ODSE Stage 1*, the *ODSE Pre-processing* validation, involved an assessment of the degree of OD value (knowledge) and semantic capabilities creation that the ODSE Framework Stage 1 processing facilitates. The effectiveness assessment for Stage 1, ODSE pre-processing pertained to raw and pre-processed OD resources. This evaluation, applied to both the input data comprising disintegrated raw OD resources and the output data comprising integrated raw OD resources, revealed that the published SDGI data in tabular, HTML, and image formats (Figures 4.2-4.6) and metadata publications in pdf format (Figures 4.8-4.9 & 4.11-4.12) are at 1-Star level which translates to the lowest publication and consumption effectiveness. On the one hand, as the publication effectiveness is defined in terms of the 5-Star deployment scheme for Linked Data, 1-Star level publications translate to the lowest publication effectiveness. On the other hand, as the consumption effectiveness is defined in terms of OD value (knowledge) creation and usability potential per semantic capability, 1-Star publications translate to the lowest levels of OD value (knowledge) creation and usability potential in



terms of OD accessibility, understandability, interoperability, integrability and intelligibility capabilities. Moreover, as depicted in Figure 8.1, the data is downloadable in spreadsheet and XML formats, thus being upgraded to the 2-star (spreadsheet - Figure 4.5) and 3-star (XML – Figure 4.10) levels. In addition, all 1-Star publications were converted to 2-Star OD resources (spreadsheet format). All these resources' transformations offer minimal and moderate levels of publication and consumption effectiveness, respectively. In terms of consumption effectiveness, this translates to low and moderate levels of accessibility, understandability, interoperability, integrability, and intelligibility of the data. This is highly likely to hamper OD value (knowledge) creation and usability potential. Therefore, following is the answer to the sub-question:

To what degree does the ODSE Framework Stage 1 processing facilitates the creation of OD value (knowledge) and semantic capabilities?

Flat text resources belong to the category of 1-, 2-, 3-Star OD publications and thus embodies minimum to moderate degree of knowledge and offers minimum to moderate degrees of semantic capabilities creation potential. This is due to the differing degrees ranging from minimum to moderate formation of semantic value determination variables, semantic usability attributes. In terms of richness attributes, flat text resources range from lowest to average degree of mining capabilities due to resource's structure and context limitations of flat text formats. This in turn results in the lowest to average degree of OD accessibility, understandability, interoperability, integrability and intelligibility capabilities. This has a direct impact on mining capabilities (lowest to average) due to mining scope and structure limitations - human mining scope and hyperlinks structure. Overall, flat text resources offer minimal to moderate degree of semantic value (knowledge) creation potential and consequently minimal to moderate degree of semantic usability effectiveness. For maximal semantic enrichment, richness, and consequently, knowledge creation, mining/harvesting and reuse potential, the data must pass the level of machine readability (3-star level – XML) to also attain to levels of machine processability (4- and 5-star levels).



8.3.2 The ODSE Processing (Stage 2) Validation [Disintegrated Knowledge Resources]

The *Framework Validation for ODSE Stage 2, ODSE Processing* validation involved an assessment of the degree of OD value (knowledge) and semantic capabilities creation that the ODSE Framework Stage 2 processing facilitates. The effectiveness assessment for Stage 2, ODSE processing addressed the current challenge of OD publications ineffectiveness by standardising and upgrading the ODSE pre-processed output in flat text format through a combination of ontology and LOD-driven semantic enrichment. The SDGI data and corresponding metadata were reconciled and interlinked to corresponding resources within the Botswana Integrated Indicator Framework and the Web of data. The deliverables of semantic enrichment are knowledge resources (Figures 7.5-7.6) – OD resources enriched with corresponding metadata and linkages across the Web of Data. Knowledge resources are easily understood by humans who can easily manipulate the data. For machine, the data has to be presented in a way that defines the concepts and their relationships as well as the computational elements pertaining to the data. Therefore, knowledge resources present the data in a way that is accessible, understandable, interoperable, integrable, and intelligible to both humans and machines. In comparing the two knowledge extracts, it is clear that for both knowledge resources, the result of semantic enrichment is a highly flexible data, RDF Data, in the form of *subject-predicate-object* structure. This data has been validated through the process of resource reconciliation to be valid and existing in the Web of Data. Also, through resources reconciliation the metadata has been validated to be compliant with the standard controlled vocabularies (ontologies). In addition, the data is linked to web resources for contextual and conceptual clarity. That is, this semantically enriched resource extract contains the metadata needed to understand and interpret SDGI 16.9.1.

So, the validation at this level was applied to the output data comprising disintegrated knowledge resources. In terms of the richness attribute variable, knowledge resources have high degree of mining capabilities due to resource's semantic structure and context which in turn results in high degree of OD accessibility, understandability, interoperability, integrability and intelligibility capabilities. This has a direct impact on mining capabilities (high). Therefore, following is the answer to the sub-question:



To what degree does the ODSE Framework Stage 2 processing facilitates the creation of OD value (knowledge) and semantic capabilities?

The ODSE Framework Stage 2 processing embodies high degree of knowledge and offers high degree of semantic capabilities creation potential as well. Knowledge resources, in the form of Linked Data with resource linkages involving a diversity of resources available on the Web of Data, are an upgrade from 3-Star (moderate effectiveness – not so good, not so bad status) to 4-Star (high level of publication effectiveness and high level of accessibility, understandability, interoperability, integrability and intelligibility for consumption effectiveness) with the differing degrees of maximum formation of semantic value determination variables. Therefore, knowledge resources offer different maximal degree of semantic value (knowledge) creation potential and consequently different maximal degree of semantic usability effectiveness. However, at this stage, the semantic data is in disintegrated files which limits the value mining (semantic) capabilities to individual resources.

8.3.3 The ODSE Post-processing (Stage 3) Validation [Integrated Knowledge Resources]

The *Framework Validation for ODSE Stage 3, ODSE Post-processing* validation involved an assessment of the degree of OD value (knowledge) and semantic capabilities creation potential facilitated by the ODSE Framework Stage 3 processing so as to determine the ODSE approach with OD value maximisation potential in terms of the enrichment, richness and mining attributes.

Stage 3 processing resolved the challenges of disintegrated knowledge resources with the introduction of knowledge resources structuring. This involved the integration of these files according to publication and consumption needs, resulting in integrated knowledge resources. At this stage, the interlinked knowledge resources are hypothesised to be having the highest levels of publication and consumption effectiveness resulting from resources' integration. To prove or disprove this hypothesis, further analysis of the integrated knowledge resources mining cases was conducted on the basis of the evaluations made in Subsection 7.4.2. Therefore, the effectiveness assessment for Stage 3, *ODSE post-processing* pertained to the output data comprising integrated knowledge resources that were further mined through SPARQL queries for semantic capabilities value added in



terms of OD accessibility, understandability, interoperability, integrability and intelligibility (Subsection 7.4.2). Overall, in comparison to flat text and individual knowledge resources, integrated knowledge resources, in the form of Linked Data with resource linkages involving a diversity of resources available on the Web of Data, fit in the category of 5-Star OD publications with the differing degrees of maximal formation of semantic value determination variables. In terms of semantic richness, integrated knowledge resources offer the highest degree of mining capabilities due to integrated resource's semantic structure and context which in turn results in the highest degree of OD accessibility, understandability, interoperability, integrability and intelligibility capabilities. This has a direct impact on mining capabilities, differences in the maximal degree due to differences in the mining scope and structure. The demonstrations of the generic and domain knowledge resources mining cases above, revealed some insights on the differences in embodied knowledge and created semantic capabilities that account to the differing maximal degrees of knowledge creation and use capabilities.

Firstly, the *accessibility* comparison has proved that OD, semantically enriched with web resource links, supports flexible and dynamic access to individual data values by both humans and machines than OD published in flat text format or in non-enriched semantic formats (e.g., XML). To start with, the accessibility demonstration proved flat text resources to be limited in semantic access. The access differs according to presentation structure and data formats. Data published in flat text format is accessed most effectively by humans and this limits the knowledge creation and use potential to humans. Access to SDGI published data is also provided in machine readable format (XML) and the Json, Python, R, C#, and SDMX snippets are provided as access links. Despite this extended access, the data visualisation and analysis are still limited to only one publication at a time. In contrast, knowledge resources addressed these limitations by providing flexibility in data access through the RDF (triple) structure which is easily accessible to both humans and machines, thus increasing knowledge mining potential. However, generic knowledge resources maximise global access with its *wide web access points* while a domain knowledge resource maximise access to resources within a given domain (*deep web access points*). Therefore, a combination of both wide and deep web access



points has high positive impact on maximisation of knowledge creation and extraction potentials.

Secondly, the *understandability* comparison proved flat text resources to be limited in semantic clarity as only humans can easily draw relationships between the different data elements. In addition, the clarity of non-enriched semantic data presented in XML format is compromised as machines can read XML data but are limited in understanding without specification of meaning, thus hampering data processing and knowledge creation and use capabilities. Both ODSE approaches provide better understandability capabilities for they augment published SDGI data with descriptive semantics in terms of SDGI attributes, measurement attributes, dimensional attributes, and relating internal and external agenda constructs. However, of the two ODSE approaches, the generic ODSE approach has a shallow semantic representation potential compared to the domain ODSE approach with customised and additional data descriptive semantics. For instance, NPI disaggregation properties increased the understandability with a clear definition of the population distribution of performance elements/components/aggregates.

In addition, the shared generic categorical properties of the generic ODSE approach resulted in the extraction of duplicate records for matching values. For example, the values marked in red (Figure 8.4), points to the source of record duplication, where the generic relation property links SDG Agenda relations to Africa Agenda and Vision Agenda. In contrast, a custom domain ontology (NPIOnto) which models domain knowledge in detail enabled specifications of the disaggregation factors, as shown in Figure 6.8, for more machine understanding and relationship differentiation (Figure 8.9). Therefore, raw OD resources (flat text resources), unenriched semantic resources (XML) and generic knowledge resources (generic ontology enriched resources) are limited in knowledge mining potential as compared to domain knowledge resources (custom domain enriched resources).

Thirdly, the *interoperability* comparison proved that flat text resources offer limited interoperability potentials while knowledge resources enable easier sharing of values across the semantic web. Further comparison revealed that generic knowledge resources, with interlinkages to a wider web resource base are better suited in increasing knowledge mining scope.



Fourthly, the *integrability* comparison revealed that flat text resources offer minimal resource linkage capabilities due to their disparate and heterogeneous publication form. In contrast, knowledge resources have increased integrability capabilities, culminating into increased knowledge mining potential. However, the generic knowledge resource provides moderate integration capabilities while the domain knowledge resource provides high integration capabilities with NPIOnto's more precise integration points (classes, and data, object, and annotation properties). Above that, NPIOnto integration points enables interlinkages that fulfil the Botswana Integrated Indicator Framework need of performance sharing and consolidation across development agendas. Furthermore, for a domain knowledge resource output shown in Figures 8.7, 8.8-8.9, each relation was given a different name for mapping differentiation, thus eradicating the duplicate records extraction error depicted in Figures 8.3-8.4.

Lastly, the *intelligibility* comparison proved flat text resources to be limited in intelligible operations. Intelligibility capabilities of flat text resources are based on data visualisation tools that offer a rich dimensional view diversity (Figures 4.4. & 4.6) but this intelligibility is compromised by the unidimensional, disparate and heterogeneity of the data. Knowledge resources, with a clear definition of dimensional data attributes, transcends these limitations and offer the capability to produce intelligible insights. However, the generic knowledge resource intelligibility proved to be limited by the generality of knowledge representation while that of the domain knowledge resource featured additional custom dimensions that extended analytical views to include the development, theme, and measurement scope and type. In short, domain knowledge resources have high knowledge mining potential. Therefore, following is the answer to the sub-question:

To what degree does the ODSE Framework Stage 3 processing facilitates the creation of OD value (knowledge) and semantic capabilities?

Knowledge resources offer different highest degree of semantic value (knowledge) creation potential and consequently different highest degree of semantic usability effectiveness. The domain knowledge resources have the highest knowledge creation and use potential.

Overall, answers to these three sub-questions indirectly address the overarching question. As all indicators have the same relationship structure depicted in Figure 8.1,



the results of the sample performance indicator instance analysis were then generalised and presented as the overall framework validation findings and contributions. Therefore, following is the answer to the main ODSE Framework validation question:

Which OD resource has the potential to maximise OD value (knowledge) creation and usability capabilities?

Integrated, semantically enriched OD resources (integrated knowledge resources) have the potential to maximise OD value (knowledge) creation and usability capabilities at differing levels. Integrated domain knowledge resources have highest degrees of knowledge creation and usability capabilities.

8.4 Framework Finalisation

To conform the preliminary ODSE Framework for generic use, it was redesigned for reusability capabilities. This was achieved by factoring in the revision needs for the generalised framework identified from framework demonstration and evaluation (verification and validation).

8.4.1 Framework Redesign Elements

There are five main framework redesign elements identified from the demonstration and evaluation findings. The first framework redesign element pertains to the ODSE Framework architectural pillars. The ODSE Framework demonstration was centred on the creation of semantic value on the basis of semantic value creation approaches with potential for knowledge creation maximisation and with a consideration for the assessment of semantic value as well. Therefore, the final ODSE Framework design is founded on the pillars of semantic value creation processing, semantic value creation maximisation approach and semantic valuation. These framework foundational design elements were featured into the ODSE Framework functionality design.

The second framework redesign element pertains to the ODSE Framework approaches. The framework validation revealed important insights on the ODSE approaches pertaining to: 1) insights on semantic capabilities value-added through the two ODSE approaches, and 2) insights on the knowledge creation potential of ODSE approaches. So, framework validation justified the modelling of the ODSE approaches component as the



major determinant of the level of OD value (knowledge) creation potential and consequently, knowledge resource usability potentials. The two ODSE approaches were factored into the framework to create a diversity in the production of knowledge resources which differ in quality/value according to the levels of semantic richness of the ontology used. Framework validation through the OD publication and consumption effectiveness evaluations have further proved that the generic ODSE approach to ODSE processing and the resulting knowledge resources are more focused on attaining semantic interoperability in the Web of Data while the domain ODSE approach to ODSE processing and the resulting knowledge resources are more focused on attaining semantic richness. Therefore, the latter was discovered to be having the maximal potential for OD value (knowledge) creation maximisation, demonstrated by the resource's degree of accessibility, understandability, interoperability, integrability and intelligibility capabilities. However, to maintain a balance between semantic richness for maximal knowledge harvesting, and interoperability for web wide connectedness, the ODSE Framework was redesigned on the combination of these two ODSE approaches.

The third framework redesign element pertains to the formation of semantic capabilities. The ODSE Framework validations revealed that OD accessibility and interoperability precedes all other capabilities. This verified that the first process in data enrichment processing, *resource reconciliation* establishes the accessibility and interoperability potentials of related web resources. These are instrumental to the second stage of data enrichment processing, *resource RDFization*, which creates metadata linkages between relating web resources. The wider the metadata linkages across the semantic web, the stronger is the support for OD understandability and intelligibility operations and vice versa.

The last two framework redesign elements pertain to two categories of ODSE Framework implementation tools. Therefore, the fourth framework redesign element pertains to the first category which embraces three models with reusability potential that were produced in the course of the preliminary ODSE framework demonstration, the ODSE scoping models. These models, considered ODSE Framework implementation support tools, were factored into the redesigned version of the ODSE Framework as development



agenda domain reference models for future support of ODSE Framework implementations. The three comprises the following:

- *Open Data evolution model* (Figure 4.1 in Section 4.2), a theoretical representation of OD support needs for the implementation of Vision 2036, Sustainable Economic Development (Pillar 1). It models the transition from OD to Open Knowledge, centred around OD processing, distribution and use;
- *NPIOnto scope model*, the scoping model for NPI domain knowledge representation, a conceptual and semantic representation of NPIs and their properties for the development agendas that constitute the Botswana Integrated Indicator Framework; and
- *Semantic capabilities relational model* deduced from the findings of the knowledge resources evaluation findings.

As explained in the methodology chapter (Subsection 3.3.5) the first two models were developed according to the generic conceptual framework development method described in Subsection 3.3.2. The third model was developed according to the *Semantic Valuation Tools Development Method* (RM8) outlined in Subsection 3.3.5.

Lastly, the fifth framework redesign element pertains to the second category of framework implementation support redesign elements which embraces the ODSE assessment models. In addition to the ODSE scoping models, this research factored into the ODSE Framework design, four semantic valuation models to address the semantic valuation challenge pertaining to lack of semantic value assessment tools for knowledge resources. The models were developed to serve as the ODSE Framework implementation support tools and for semantic enrichment output (knowledge resources) assessment of resources' enrichment, richness, mining and usability worth or value. Three models, semantic enrichment, richness and mining assessment models were produced on the basis of the preliminary ODSE Framework evaluation findings according to the *Semantic Valuation Tools Development Method* (RM8) outlined in Subsection 3.3.5 and the fourth model, the *semantic usability assessment model* detailed in Subsection 3.3.4 (Table 3.1 & Figure 3.2) was produced as an extension of Tim Berners Lee's Linked Data publication standard discussed in Section 2.4.



- The *semantic enrichment assessment model* is a graphical representation of OD value (knowledge) creation potential of OD resources on the basis of enriched semantic capabilities for a given ODSE approach for the assessment of the impact of a given ODSE approach on the resulting knowledge resources and for the assessment of the semantic capabilities value-added to a given knowledge resource, as well as, for the assessment of the OD value (knowledge) creation potential of a given knowledge resource, demonstrated by the resource's degree of enriched semantic capabilities;
- The *semantic richness assessment model* is a graphical representation of relationship between semantic richness of OD resource and OD value (knowledge) creation potential to depict how a given type of OD publication resource impacts on OD extraction/harvesting mode for the assessment of how the semantic richness of a given resource is likely to influence its knowledge creation capabilities and for the assessment of the appropriate OD value (knowledge) harvesting tool for a given level of semantic richness of a knowledge resource;
- The *semantic mining assessment model* is a graphical representation of the linkage between the type of OD resource and the knowledge mining potential for the assessment of the ontology approach that best fits specific knowledge mining needs. For instance, with a given set of semantic use value requirements (knowledge harvest expectations), decisions can be made on the most applicable ODSE approach to knowledge production that will yield the desired knowledge harvest (semantic use value), i.e., whether domain or generic ODSE approach or a combination of the two ODSE approaches; and
- The *semantic usability assessment model* is a graphical representation of the qualitative matrix in the form of a 5-point rating scheme of OD publication and consumption effectiveness ratings derived from a combination of the semantic enrichment, richness and mining attributes and the 5-Star Scheme for Linked Data for the assessment of the publication and consumption effectiveness of a given type of OD resource.

The summary of semantic value assessment models is provided in Table 8.2.



Table 8. 2: Summary of Design Features for Semantic Valuation Models

MODEL VARIABLE	MODEL 1	MODEL 2	MODEL 3	MODEL 4
MODEL DESCRIPTION				
Model Name	Assessment Model for Open Data Value (Knowledge) Creation Capabilities per Semantic Enrichment Approach	Assessment Model for Open Data Value (Knowledge) Creation Capabilities per Open Data Resource	Assessment Model for Open Data Resource Mining Capabilities	Assessment Model for Open Data Publication and Consumption Effectiveness
Model Assessment Dimension	semantic enrichment assessment	semantic richness assessment	semantic mining assessment	semantic usability assessment
Model Assessment Scope	assessment of the degree of semantic enrichment impact on semantic value (knowledge) creation potential of a given OD resource	assessment of the degree of semantic richness impact on semantic value (knowledge) creation potential of a given OD resource	assessment of the degree of semantic mining impact on semantic value (knowledge) creation potential of a given OD resource	assessment of the degree of semantic usability impact on semantic value (knowledge) creation potential of a given OD resource
SEMANTIC VALUE CREATION VARIABLES				
Semantic Value Variable	expressed in terms of minimum, moderate and maximum rating scale	expressed in terms of minimum, moderate and maximum rating scale	expressed in terms of minimum, moderate and maximum rating scale	expressed in terms of minimum, lowest, low, moderate, high, highest and maximum rating scale
(Semantic Value Independent Variable) [data publication Variable]	type of semantic enrichment approach (no semantic enrichment, generic ODSE approach, domain ODSE approach)	type of OD resource (flat text resources, generic knowledge resource, domain knowledge resource)	type of OD resource (flat text resources, generic knowledge resource, domain knowledge resource)	type of OD resource (expressed in terms of Tim Berners-Lee's 5-Star deployment scheme for Linked Data (1-, 2-, 3-, 4- & 5-star OD publications) – publication variable)
(Semantic Value Dependent Variable 1) [data publication and consumption Measure Variable]	degree of semantic value (knowledge) creation potential (expressed in terms of semantic capabilities - accessibility, understandability, interoperability, integrability & intelligibility)	degree of semantic value (knowledge) creation potential (expressed in terms of structure - flat text vs semantic and context - flat text vs semantic)	degree of semantic value (knowledge) creation potential (expressed in terms of mining scope - human vs machine; link structure form - hyperlink vs RDF; link coverage scope - wide vs narrow; knowledge coverage scope shallow vs deep)	degree of semantic value (knowledge) creation potential (expressed in terms of a combination of semantic enrichment, richness and mining attributes)
(Semantic Value Dependent Variable 2) Semantic Attribute Measure Variable [Enabler of OD consumption Measure variable]	Degree of semantic enrichment Effectiveness (expressed in terms of semantic enrichment attributes that are in the form of semantic capabilities)	Degree of semantic richness Effectiveness (expressed in terms of structure and context)	Degree of semantic mining Effectiveness (expressed in terms of mining scope, link structure form, link coverage scope and knowledge coverage scope)	Degree of semantic usability Effectiveness (expressed in terms of a combination of semantic enrichment, richness and mining attributes)



(Semantic Value Determinant Variable) Semantic Attribute variable [Enabler of OD consumption variable]	semantic enrichment attributes (expressed in terms of semantic capabilities - accessibility, understandability, interoperability, integrability & intelligibility)	semantic richness attributes (expressed in terms of structure - flat text vs semantic and context - flat text vs semantic)	semantic mining attributes (expressed in terms of mining scope - human vs machine; link structure form - hyperlink vs RDF; link coverage scope - wide vs narrow; knowledge coverage scope shallow vs deep)	semantic usability attributes (expressed in terms of a combination of semantic enrichment, richness and mining attributes)
RELATIONS BETWEEN SEMANTIC VALUE CREATION VARIABLES				
Relation 1 Independent variable and semantic attribute variable relation	the type of semantic enrichment approach determines the degree of effectiveness of semantic enrichment attributes (expressed in terms of semantic enrichment attributes that are in the form of the five semantic capabilities)	the type of OD resource determines the degree of effectiveness of semantic richness attributes (expressed in terms of structure and context)	the type of OD resource determines the degree of effectiveness of semantic mining attributes (expressed in terms of mining scope, link structure form, link coverage scope and knowledge coverage scope)	the type of OD resource determines the degree of effectiveness of semantic usability attributes (expressed in terms of semantic usability attributes comprising semantic enrichment, richness and mining attributes)
Relation 2 OD resource variable and semantic attribute variable relation	OD resources differ according to the degree of effectiveness of semantic enrichment attributes (expressed in terms of semantic enrichment attributes that are in the form of the five semantic capabilities)	OD resources differ according to the degree of effectiveness of semantic richness attributes (expressed in terms of structure and context)	OD resources differ according to the degree of effectiveness of semantic mining attributes (expressed in terms of mining scope, link structure form, link coverage scope and knowledge coverage scope)	OD resources differ according to the degree of effectiveness of semantic usability attributes (expressed in terms of semantic usability attributes comprising semantic enrichment, richness and mining attributes)
Relation 3 Semantic attribute variable and semantic attribute measure variable relation	The level of semantic enrichment attributes determines the degree of effectiveness of semantic enrichment [that is, the minimum, moderate and maximum degrees of semantic capabilities = minimum, moderate & maximum degrees of effectiveness of semantic enrichment]	The level of semantic richness attributes determines the degree of effectiveness of semantic richness [that is, the minimum, moderate and maximum degrees of structure and context = minimum, moderate and maximum degrees of effectiveness of semantic richness]	The level of semantic mining attributes determines the degree of effectiveness of semantic mining [that is, the minimum, moderate and maximum degrees of mining scope, link structure form, link coverage scope and knowledge coverage scope = minimum, moderate and maximum degrees]	The level of semantic usability attributes determines the degree of effectiveness of semantic usability (publication and consumption) [that is, the minimum, moderate and maximum degrees of a combination of semantic enrichment, richness and mining attributes = minimum, moderate



			of effectiveness of semantic mining]	& maximum degrees of effectiveness of semantic usability]
Relation 4 Semantic attribute measure variable and dependent variable relation	the degree of effectiveness of semantic enrichment has a direct impact on the degree of semantic value (knowledge) creation potential (expressed in terms of semantic enrichment attributes that are in the form of the five semantic capabilities)	the degree of effectiveness of semantic richness has a direct impact on the degree of semantic value (knowledge) creation potential (expressed in terms of semantic richness attributes - structure and context)	the degree of effectiveness of semantic mining has a direct impact on the degree of semantic value (knowledge) creation potential (expressed in terms of semantic mining attributes - mining scope, link structure form, link coverage scope and knowledge coverage scope)	the degree of effectiveness of semantic usability has a direct impact on the degree of semantic value (knowledge) creation potential (expressed in terms of semantic usability attributes comprising semantic enrichment, richness and mining attributes)
Relation 5 independent variable and dependent variable relation	the type of semantic enrichment approach determines the degree of semantic value (knowledge) creation potential (expressed in terms of semantic enrichment attributes that are in the form of the five semantic capabilities)	the type of OD resource determines the degree of semantic value (knowledge) creation potential (expressed in terms of semantic richness attributes - structure and context)	the type of OD resource determines the degree of semantic value (knowledge) creation potential (expressed in terms of semantic mining attributes - mining scope, link structure form, link coverage scope and knowledge coverage scope)	the type of OD resource determines the degree of semantic value (knowledge) creation potential (expressed in terms of semantic usability attributes comprising semantic enrichment, richness and mining attributes)
Model Applications				
Model Application 1 [Resource Grading/Rating Application] Semantic attribute variable and OD resource variable relation	Assignment of rates to semantic enrichment attributes for a given type of OD resource (expressed in terms of minimum, moderate and maximum rating scale)	Assignment of rates to semantic richness attributes for a given type of OD resource (expressed in terms of minimum, moderate and maximum rating scale)	Assignment of rates to semantic mining attributes for a given type of OD resource (expressed in terms of minimum, moderate and maximum rating scale)	Assignment of rates to semantic usability attributes for a given type of OD resource (expressed in terms of minimum, lowest, low, moderate, high, highest and maximum rating scale)
Model Application 2 [Resource Selection Application] OD resources variable; semantic attribute variable and dependent variable relation	Selection of OD resources with appropriate semantic enrichment attributes for a given target of semantic value (knowledge) creation measure (expressed in terms of minimum, moderate and maximum rating scale)	Selection of OD resources with appropriate semantic richness attributes for a given target of semantic value (knowledge) creation measure (expressed in terms of minimum, moderate and maximum rating scale)	Selection of OD resources with appropriate semantic mining attributes for a given target of semantic value (knowledge) creation measure (expressed in terms of minimum, moderate and maximum rating scale)	Selection of OD resources with appropriate semantic usability attributes for a given target of semantic value (knowledge) creation measure (expressed in terms of minimum, lowest, low, moderate, high, highest and maximum rating scale)
Model Application 3 [Resource Selection Application]	Selection of OD resources with appropriate	Selection of OD resources with appropriate	Selection of OD resources with appropriate	Selection of OD resources with appropriate



OD resources variable; semantic attribute variable and semantic attribute measure variable relation	semantic enrichment attributes for a given target of semantic enrichment effectiveness rating (expressed in terms of minimum, moderate and maximum rating scale)	semantic richness attributes for a given target of semantic richness effectiveness rating (expressed in terms of minimum, moderate and maximum rating scale)	semantic mining attributes for a given target of semantic mining effectiveness rating (expressed in terms of minimum, moderate and maximum rating scale)	semantic usability attributes for a given target of semantic usability effectiveness rating (expressed in terms of minimum, lowest, low, moderate, high, highest and maximum rating scale)
Model Application 4 [Resource Evaluation Application] Semantic usability measure variable and OD resource variable relation	Evaluation of the degree of effectiveness of semantic enrichment for a given type of OD resource (expressed in terms of minimum, moderate and maximum rating scale)	Evaluation of the degree of effectiveness of semantic richness for a given type of OD resource (expressed in terms of minimum, moderate and maximum rating scale)	Evaluation of the degree of effectiveness of semantic mining for a given type of OD resource (expressed in terms of minimum, moderate and maximum rating scale)	Evaluation of the degree of effectiveness of semantic usability (both publication and consumption effectiveness) for a given type of OD resource (expressed in terms of minimum, lowest, low, moderate, high, highest and maximum rating scale)
Model Application 5 [Resource Evaluation Application] Dependent variable and OD resource variable relation	Evaluation of the degree of semantic value (knowledge) creation potential for a given type of semantic enrichment approach (expressed in terms of minimum, moderate and maximum rating scale)	Evaluation of the degree of semantic value (knowledge) creation potential for a given type of OD resource (expressed in terms of minimum, moderate and maximum rating scale)	Evaluation of the degree of semantic value (knowledge) creation potential for a given type of OD resource (expressed in terms of minimum, moderate and maximum rating scale)	Evaluation of the degree of semantic value (knowledge) creation potential for a given type of OD resource (expressed in terms of minimum, lowest, low, moderate, high, highest and maximum rating scale)

8.4.2 Framework Adjustments

To the first stage, *OD pre-processing*, the adjustment made pertained to the addition of the *pre-processing output component*. To the second stage, *OD semantic enrichment processing*, all the sub-processes were compressed into two components, *ODSE approach component* and *data enrichment component*. The ODSE approach component comprises the generic ODSE approach (semantic enrichment processing based on generic ontologies) and the domain ODSE approach (semantic enrichment processing based on custom domain ontologies). The data enrichment component comprises the *resource reconciliation* and the *resource RDFization* processes. To the third stage, *OD post-processing*, the framework was adjusted to reflect knowledge resources structuring and



deployment components. Lastly, Table 8.3 gives a summary of how the framework implementation support tools were factored into the ODSE Framework (Figure 9.1).

Table 8. 3: Summary of Potential Application of the ODSE Framework Implementation Support Tools

ODSE MODEL	ODSE FRAMEWORK APPLICATION	
	ODSE ACTIVITY	DESCRIPTION
SCOPING TOOLS - ODSE SCOPING MODELS		
Open Data Evolution Model <i>(Figure 4.1)</i> Models the transition from Open Data to Open Knowledge, centred on OD processing, distribution and use.	1.1	To serve as a generic development agenda domain scoping tool to be used in the assessment of the scope of OD publication resources (raw OD resources) for collection.
	1.2	To serve as a generic development agenda domain scoping tool to be used in the assessment of the scope of OD publication and consumption specification resources for collection.
	2.1	To serve as a generic development agenda domain scoping tool to be used in the assessment of the pre-processing resource cleaning, transformation and integration needs.
	2.2	To serve as a generic development agenda domain scoping tool to be used in the assessment of the scope of search for semantic publication and consumption needs.
	2.3	To serve as a generic development agenda domain scoping tool for modelling the semantic value creation needs.
NPIOnto Scope Model <i>(Figure 9.2)</i> - A conceptual and semantic representation of NPIs and their properties for the development agendas that constitute the Botswana Integrated Indicator Framework. - It is a source of insights that address the domain scoping needs in the Botswana development agenda domain applications, with a specific focus on the Botswana Integrated Indicator Framework scope.	4.1	To serve as a potential ontology target for semantic enrichment and any other reuse applications.
	4.2	To serve as a reference model in the specification, initial structuring, conceptualisation (design), formalisation (development), evaluation and deployment of related NPI domain applications and systems.
	4.2.1	To serve as an NPI domain scoping tool that provides a common understanding of the development agenda domain.
	4.2.3	To serve as an NPI domain scoping tool for NPIOnto conceptualisation adaptations in similar domain applications.
	4.2.4	To serve as an NPI domain scoping tool in the extraction of NPI constructs (development agenda concepts and property definitions) for reuse or extension among OD publishers and consumers.
Semantic Capabilities Relational Model <i>(Figure 9.3)</i> Models relations between OD accessibility, understandability, interoperability, integrability and intelligibility	5.0	For the assessment of how to factor in semantic capabilities in the design and implementation of semantic enrichment process flows for applications that enhance semantic capabilities. It provides insights on semantic capabilities formation dependencies to support holistic semantic capabilities modelling for semantic enrichment processing applications.
ASSESSMENT TOOLS - ODSE ASSESSMENT MODELS		



Semantic Enrichment Assessment Model <i>(Figure 9.4)</i> Models OD value (knowledge) creation potential of OD resources on the basis of enriched semantic capabilities for a given ODSE approach	7.1	- For the assessment of the impact of a given ODSE approach on the resulting knowledge resources. - For the assessment of the semantic capabilities value-added to a given knowledge resource. - For the assessment of the OD value (knowledge) creation potential of a given knowledge resource, demonstrated by the resource's degree of enriched semantic capability demonstration.
Semantic Richness Assessment Model <i>(Figure 9.5)</i> - Models the relationship between semantic richness of OD resource and OD value (knowledge) creation potential to depict how a given type of OD publication resource impacts on OD extraction/harvesting mode - It gives insights on how the combination of a given OD resource structure (syntax) and context (semantics) is likely to impact on the richness of semantic meaning and mining/harvesting of that resource and the overall impact on OD value (created knowledge).	7.1 8.2	- For the assessment of how the semantic richness of a given resource is likely to influence its knowledge creation capabilities. - For the assessment of the appropriate OD value (knowledge) harvesting tool for a given level of semantic richness of a knowledge resource
Semantic Mining Assessment Model <i>(Figure 9.6)</i> - Models the linkage between the type of OD resource and the knowledge mining potential - It enables a balanced assessment of the key dimensional variables (data access, link structure, link coverage and knowledge coverage) of a given knowledge resource (that represent qualities to be embedded on the data through semantic enrichment processing so as to enhance the knowledge mining potential) to determine its knowledge mining capabilities	4.0	For the assessment of the ontology approach that best fits specific knowledge mining needs.
	4.1.2 4.2.1	For the identification and specification of ontology enrichment requirements to support the production of knowledge resources that meet the knowledge mining results. It enables the specifications of semantic enrichment output (semantic resources) in terms of semantic qualities (data access, link structure, link coverage and knowledge coverage) that determine the knowledge mining potential
	4.1.4	For the assessment of suitable ontology targets for a given knowledge mining operation. It gives insights on the potential impact of a given OD resource type (resource types differ in terms of the level of data access, link structure, link coverage and knowledge coverage) on OD value (knowledge) mining operations.
Semantic Usability Assessment Model <i>(Table 3.1 & Figure 3.2)</i> Models qualitative matrix in the form of a 5-point rating scheme of OD publication and consumption effectiveness ratings derived from a combination of the semantic capabilities and the 5-Star Scheme for Linked Data	2.1 3.1 6.0 7.0	For the assessment of the publication and consumption effectiveness of a given type of OD resource.



8.5 Conclusion

Chapter Eight documents how Research Objective 4, *evaluation of the preliminary semantic Enrichment framework completeness, cohesiveness and effectiveness*, addressed Research Question 4 expressed as:

What is the completeness, cohesiveness and effectiveness of the preliminary semantic enrichment framework?

Firstly, in terms of framework completeness and cohesiveness (framework verification), this research question sought to establish the degree of inclusion of all semantic enrichment entities and the degree of the connectivity and flow of both the individual stage elements and the integrated elements per process for all stages. Secondly, in terms of framework effectiveness (framework validation), this research question sought to establish the degree of framework effectiveness in maximising OD value (knowledge) creation potential and in determining the potential impact on knowledge resource usability potential. The framework validations pertained to functionality assessment through OD value (knowledge) mining for structure and meaning value-added to OD semantic capabilities per ODSE Framework stage.

To start with, framework verification has proved semantic enrichment processes, approaches, capabilities, techniques and technologies to be complete and satisfactorily working cohesively. However, some adjustments were found necessary to be effected for the generalised framework. Those adjustments were further validated to be feasible through framework validation.

Chapter Eight also documents how Research Objective 5, *finalisation of the semantic enrichment framework for reusability* addressed Research Questions 5 below:

What is the reusability potential of the semantic enrichment framework finalised on the basis of the preliminary framework demonstration and evaluation findings?

This research question sought to establish the degree of framework application in similar scenarios. This research has fully demonstrated with the ODSE Framework, the creation, structuring, sharing deployment and mining of semantic value. Due to the demonstration aspect of the ODSE Framework, it brought more light on the implementation specifics that played a key role in evolving the preliminary ODSE Framework design into a more generic ODSE Framework for its reusability. The framework was redesigned on the basis



of the demonstration and evaluation findings for generic use in the development agenda domain. In comparison, the preliminary ODSE Framework presented in Chapter Five (Figure 5.1) and demonstrated in Chapters Six and Seven is specific to Botswana's needs on elements and process flows while the final ODSE Framework represents the generalised redesign of these elements and process flows. The framework was redesigned as follows:

- Stage 1 - addition of the pre-processing output component,
- Stage 2 - compression of all sub-processes into two components such that the ODSE approach component comprises the generic ontologies (generic ODSE approach) and custom domain ontologies (domain ODSE approach) and the data enrichment component comprises resource reconciliation and RDFization processes, and
- Stage 3 - addition of resource processing and deployment components.

In addition, it factors into the three stages reusable implementation support tools, the scoping and assessment models. Therefore, the final ODSE Framework is a reusable design artifact presented in Chapter Nine as the main research contribution together with the ODSE Framework implementation support tools.



CHAPTER NINE: RESEARCH CONTRIBUTION

9.1 Introduction

This chapter presents the research output, the final version of the Ontology-Driven Semantic Enrichment Framework (ODSE Framework) that was developed as an endeavour to address the gap of reviewed semantic value creation frameworks. Reviewed frameworks were found to be focused on adding semantic value (use value) to raw OD resources in flat text formats to improve semantic capabilities through either:

- the addition of a layer of semantic meaning to raw OD resources (semantics enrichment focus); or
- the addition of a layer of semantic structure to raw OD resources (syntax enrichment focus); or
- the addition of a combination of the two layers.

However, they were found to be lacking mechanisms to facilitate and evaluate the degree of semantic value creation capabilities. Currently, this knowledge gap makes it difficult to drive the semantic value creation process toward yielding specific desired results, which in this case, pertains to the creation of knowledge resources with maximal knowledge creation potential. It also makes it difficult to evaluate the degree of semantic enrichment impact effected through the semantic value creation process. This knowledge gap has sparked a quest to find semantic solutions with mechanisms for facilitation and evaluation of the degree of semantic value creation capabilities of knowledge resources. To address the knowledge gap discussed above, the ODSE Framework, represented in graphic form as shown in Figure 9.1, is founded on the pillars of: 1) semantic value creation processing modelled by the ODSE Framework Components 1.0-8.0; 2) semantic value creation maximisation approach modelled by the ODSE Framework Component 4.0; and 3) semantic valuation modelled into all the ODSE Framework Components (all tools).



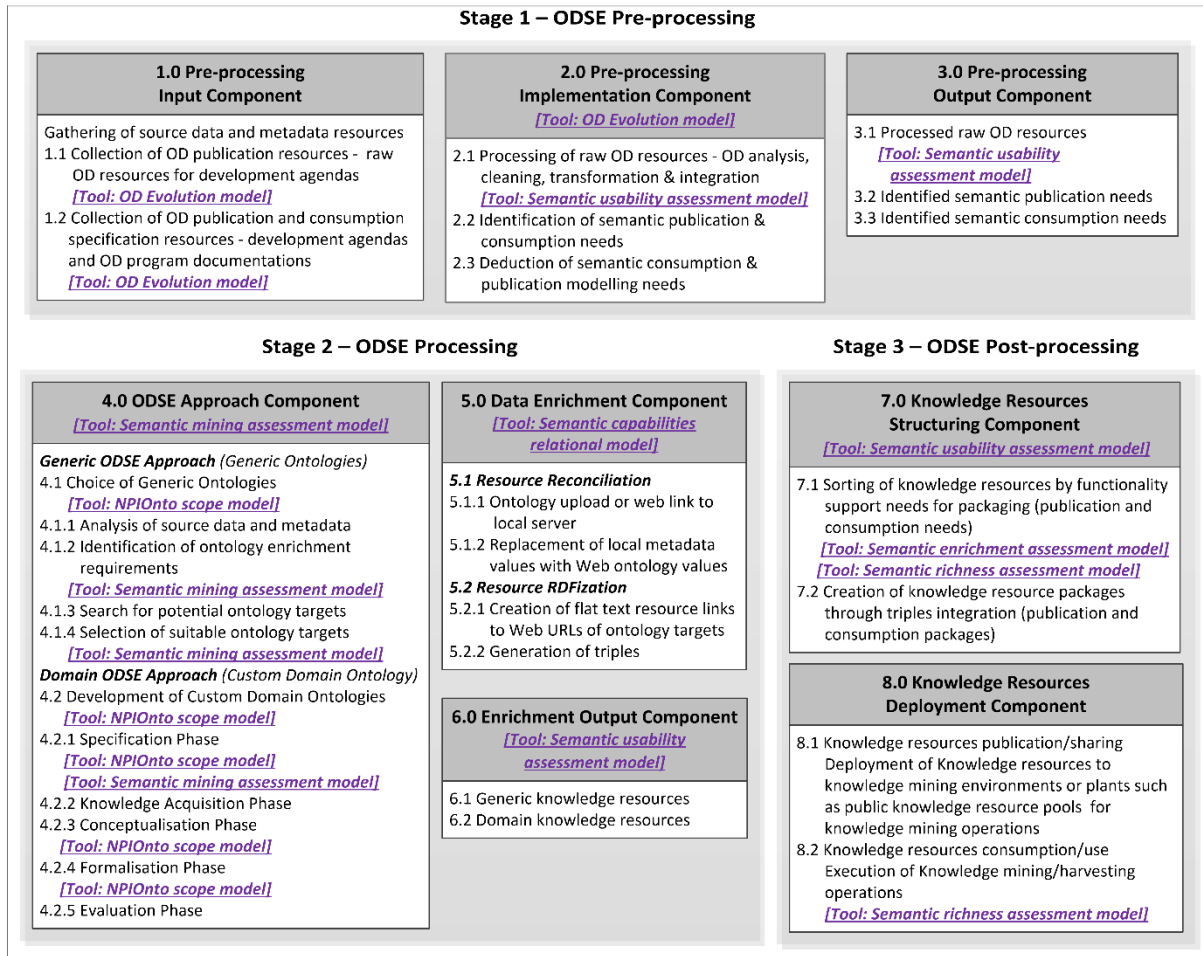


Figure 9. 1: Final Ontology-Driven Semantic Enrichment Framework

The ODSE framework is organised in three implementation stages pertaining to ODSE pre-processing, processing and post-processing. Following is a detailed framework discussion.

9.2 Stage 1: ODSE Pre-processing

The *ODSE pre-processing* is the preliminary stage which deals with the gathering of semantic enrichment input and its preliminary processing to produce preliminary output for further semantic enrichment processing. This stage has three sub-processes comprising the pre-processing input, pre-processing implementation, and pre-processing output components.



9.2.1 Pre-processing Input (ODSE Component 1.0)

The *pre-processing input component* involves the gathering of source data and metadata resources which pertains to the collection of OD publication resources together with corresponding OD publication and consumption specification resources.

Collection of OD publication resources (ODSE Activity 1.1)

The *pre-processing input component* pertains to the collection of OD publication resources (raw OD resources) for development agendas. So, this activity involves the identification and extraction of development agendas' raw OD resources to be enriched, specifically, the extraction of published data from secondary sources, i.e., web platforms. However, this research has observed that access to the data for most OD publications is in diverse formats and these formats are also at diverse levels of publication effectiveness, i.e., pdf - 1-Star publication providing the different degrees of minimum levels of publication effectiveness; and spreadsheet – 2-Star publication also providing the different degrees of minimum levels of publication effectiveness. In addition, some access formats are compatible (readable and understandable) only to humans (i.e., flat text resources) while others to both humans and machines (i.e., XML). Furthermore, some of the published data may have diverse definitions and availed in different platforms, for instance, the published data may be available in a web platform while the unpublished data pertaining to additional metadata may be residing in different platforms. In short, input data to be enriched is characterised by OD disparateness in terms of presentation formats and OD heterogeneities in terms of access formats, platforms, and data definitions. In the Botswana case study demonstration, published OD for the SDG Agenda was retrieved from Statistics Botswana data portal. The input data was found to be comprised of numerous spreadsheet files presented separately according to the 17 SDGs. For each goal, the data was presented separately on the basis of each SDGI. In addition, the metadata was found to be presented at different levels of flat data structures.

Collection of OD Publication and Consumption Specification Resources (ODSE Activity 1.2)

The *pre-processing input component* also involves the collection of OD publication and consumption specification resources of relevant documentations on development agendas and OD programs. This pertains to the identification of the motive for the



enrichment so as to determine the semantic value to be added to the data and the identification of potential sources of metadata for the enrichment. This involves the collection of documentations, mainly government strategic documents for OD and development agendas, i.e., policies, strategic plans and initiatives for OD programs and development agendas. The OD programs and development agenda documentations are sources of the OD and development agenda domain knowledge, thus, constituting the rich metadata sources for semantic enrichment. However, this research has discovered that the metadata sources are in most cases availed in diverse flat text formats. They are also availed in diverse platforms, i.e., some are published in the Web and others in local (in-house) data management platforms. In the Botswana case study demonstration, the documentations availed in pdf and spreadsheet formats, comprised of the current NDP Agenda and Vision Agenda documentations; the Botswana OD Readiness Assessment Survey report; SDGs National Baseline Indicator Framework Proposal; Botswana Domesticated SDGs and the global 2030 SDGs Indicator List.

Overall, the diversity of OD publication formats and platforms for both published data and metadata sources makes it difficult to access, understand, interoperate, integrate, and make intelligible use of the data. This hampers OD value (knowledge) creation capabilities and consequently, knowledge extraction capabilities. To address these limitations, the data has to undergo some processing detailed in the next sections.

9.2.2 Pre-processing Implementation (ODSE Component 2.0)

The *pre-processing implementation component* deals with the preliminary processing of source data and metadata resources, specifically, the processing of raw OD resources and the identification of semantic publication and consumption needs.

Processing of Raw OD Resources (ODSE Activity 2.1)

The *pre-processing implementation component* pertains to the processing of raw OD resources which involves data analysis; cleaning; quality operations on the data; and the unification of the data, to produce output formatted for semantic enrichment. The pre-processing OD cleaning transformation and integration are performed to improve data quality and intelligibility capabilities so as to increase the value-adding potential of the input data. In the Botswana case study demonstration, this component involved the resolution of inconsistencies pertaining to NPI and value definitions, incomplete



semantic definitions, missing values, NPI mapping inconsistencies, ambiguous values, and value disaggregation inconsistencies. It also involved the integration of disparate data and the harmonisation of related data from different sources.

At this stage, in addition to the processing of raw OD resources, the identification of semantic value creation and modelling needs is executed at **three levels** as follows.

Identification of Semantic Publication and Consumption Needs (ODSE Activity 2.2)

At the **first level**, the question, **what needs to be addressed**, is asked so as to focus the survey on discovering semantic consumption needs. The semantic consumption needs pertain to OD use and reuse requirements in terms of knowledge to be derived from the enriched output for further usage. For the Botswana case study, the identification of semantic consumption needs involved the assessment of Vision 2036 (Pillar 1) for OD-related support needs in the implementation of Vision 2036 (Pillar 1), in light of the question stated above. All in all, the semantic consumption needs map the pathway to the identification of semantic publication needs.

At the **second level**, the question, **what needs to be done, to attend to what needs to be addressed**, is asked so as to focus the survey on discovering semantic publication needs. The semantic publication needs relate to OD structure (syntax) and contextual (semantics) requirements. For the Botswana case study, the identification of semantic publication needs involved the assessment of agenda documentations and raw OD publications, in light of the question asked above to: 1) identify OD program goals; 2) deduct OD presentation needs from the limitations of the current raw OD publications; and 3) identify OD representation needs for ontology modelling. All in all, the semantic publication needs map the pathway to addressing the semantic consumption needs.

Deduction of Semantic Publication and Consumption Modelling Needs (ODSE Activity 2.3)

At the **third level**, the question, **how to do, what needs to be done to attend to what needs to be addressed**, is asked so as to focus the framework design on mapping how to traverse the pathway that addresses semantic consumption needs. This calls for further analysis of identified semantic publication and consumption needs to deduce the ODSE Framework modelling needs, modelling measures and functionality design elements.



9.2.3 Pre-processing Output (ODSE Component 3.0)

The *pre-processing output component* pertains to the deliverables of pre-processing implementations, comprising the preliminary processed OD resources, identified semantic publication and consumption needs and semantic publication and consumption modelling needs.

Processed Raw OD Resources (ODSE Activity 3.1)

The *pre-processing output component* comprises preliminary processed OD resources (flat text resources) representing the cleaned, transformed and consolidated data. These resources provide input for the ODSE processing stage. For the Botswana case study demonstration, the pre-processed output comprised collections of domain knowledge on NPIs in flat text format (sample screenshots of datasets shown in appendix 1-4). The contents of these resources are described in detail in Subsection 6.2.1. The *pre-processing output component* also comprises identified OD semantic consumption and publication needs discussed next.

Identified Semantic Consumption and Publication Needs (ODSE Activity 3.2)

The semantic consumption need was identified as the OD support need for vision agenda, the need for the ODSE Framework to support the generation of knowledge resources for the implementation of the economic aspiration of a transformation into a knowledge society founded on a knowledge-driven economy.

The semantic publication needs were identified as: 1) the OD program need for the ODSE Framework to support the generation of rich semantic formats for reusability of the current OD publications; 2) the OD presentation need for the ODSE Framework to support the implementation of standardised and unified publication structure and format; 3) the OD representation (Ontology) need for ODSE Framework to support the implementation of the consolidated indicator framework for the four development agendas.

Identified Semantic Consumption and Publication Modelling Needs (ODSE Activity 3.3)

To address the consumption need, the OD support need for vision agenda, the ODSE Framework functional design comprised the semantic enrichment pre-processing, processing and post-processing for the creation, structuring and sharing of knowledge



resources needed to drive a knowledge-driven economy and society. The ODSE Framework functional design to support semantic publication needs are as follows: 1); the OD program need for richer semantic formats was addressed by the introduction of ontology-driven semantic enrichment processing; 2) the OD presentation need for standardised and unified publications was addressed by the introduction of a combination of resource reconciliation and RDFization for semantic enrichment processing; and 3) the OD representation (ontology) need was addressed by the development of the custom domain knowledge representation model (ontology) for the semantic integration of the Botswana Integrated Indicator Framework, the NPIONto.

Overall, there is a dependency relationship between the semantic value creation needs and the development of the ODSE Framework. The identification of semantic consumption needs drove the formulation of semantic publication needs which in turn drove the development of the solution to the semantic consumption needs, the ODSE Framework.

Table 9.1 summarises identified semantic publication and consumption needs and generically outlines how they influenced the ODSE Framework modelling needs, design measures taken to address the modelling needs and consequently, its functionality design.



Table 9. 1: *Semantic Publication and Consumption Needs - Drivers of the ODSE Framework Design*

Identified Needs	ODSE Framework Modelling Needs	ODSE Framework Modelling Measures	ODSE Framework Functionality Design
Open Data Support Need for Vision Agenda (Consumption Need)	The ODSE Framework need to support the generation of knowledge resources for the implementation of the economic aspiration of a transformation into a knowledge society founded on a knowledge-driven economy	The ODSE framework leverages OD for the generation of knowledge resources to support the implementation of the knowledge-driven economy and society	The introduction of semantic enrichment pre-processing, processing and post-processing for the creation, structuring and sharing of knowledge resources
Open Data Program Need (Publication Need)	The ODSE Framework need to support the generation of rich semantic formats for reusability of the current OD publications	The ODSE framework enhances semantic capabilities of flat text resources	The introduction of ontology-driven semantic enrichment processing
Open Data Presentation Need (Publication Need)	The ODSE Framework need to support the implementation of standardised and unified publication structure and format	The ODSE framework creates standard and unified structure and format	The introduction of a combination of resource reconciliation and RDFization for semantic enrichment processing
Open Data Representation (Ontology) Need (Publication Need)	The ODSE Framework need to support the implementation of the consolidated indicator framework for the four development agendas	The ODSE framework addresses the semantic representation and integration need for the four development agendas	The development of the custom domain knowledge representation model (ontology) for the semantic integration of the Botswana Integrated Indicator Framework, the NPIONto



9.3 Stage 2: ODSE Processing

The *ODSE processing* is the intermediate ODSE Framework processing stage that pertains to the actual implementation of the OD value (knowledge) creation. The LOD technique was adopted in the conversion of flat text resources to a more rich, flexible and dynamic format on the basis of relevant ontologies for increased value-adding potential. The implementation of these transformations is organised in three ODSE processing components being, the *ODSE approach component*, the *data enrichment* component and the *enrichment output* component.

9.3.1 ODSE Approach (ODSE Component 4.0)

The *ODSE approach component* is the central processing component of the ODSE Framework, the main driver of semantic value creation and assessment. The ODSE approach component involves the selection and implementation of the semantic value creation and assessment approach that best fits the semantic enrichment challenge at hand as defined by the semantic publication and consumption needs (semantic value creation needs). Semantic enrichment based on the use of a combination of generic ontologies constitutes the generic ODSE approach while the domain ODSE approach is based on the use of custom domain ontologies. The *ODSE approach component* has two sub-processes described below.

Generic ODSE Approach

The first sub-process of the *ODSE approach component* outlines the method for choosing applicable generic ontologies, thus, pertains to the generic ODSE approach (ODSE activity 4.1). The term generic ontology in the context of this framework refers to the top-level/upper/foundational/reference ontologies, as well as any unrelated domain ontology. These are ontologies published as online resources. The method for selecting applicable generic ontologies introduced in Subsection 3.3.3 involves four steps outlined below. Step 1, the *analysis of source data and metadata resources for ontology search literals* (ODSE sub-activity 4.1.1), i.e., pre-processed OD resources and metadata documentations discussed in stage 1 involves the identification and assignment of entity field names to all data holding fields and the descriptions of data content as well. That is followed by the refinement of entity field names to serve as keywords for online ontology



search targets (i.e., *SDG name*). Lastly, is the analysis of the entity field names to extract the online search literals (i.e., *name*). Step 2, the *identification of ontology enrichment requirements* (ODSE sub-activity 4.1.2) involves the definition of the objective of ontology enrichment needs to be addressed by ontology targets, i.e., in the Botswana case study, for generic ontology targets, the objective was to create more global links to the NPI domain dataset for enhanced discoverability. With NPIOnto (custom domain ontology), the objective was to link more specific meaning to the dataset. Step 3, the *search for potential ontology targets* (ODSE sub-activity 4.1.3) is conducted on the basis of the literals identified in Step 1 and the ontology enrichment requirements identified in step 2, i.e., below are three potential ontology targets for the literal *name* and the semantic enrichment objective of enhanced discoverability:

- NCI Thesaurus OBO Edition - concept *name* pertains to the words or language units by which a thing is known,
- SemanticScience Integrated Ontology - concept *name* pertains to a label used to identify an entity,
- eNanoMapper Ontology - concept *name* pertains to the words or language units by which a thing is known.

Step 4, the *selection of suitable ontology targets* (ODSE sub-activity 4.1.4) pertains to the development of the selection criteria to serve as the basis for the selection of suitable ontology targets. In the Botswana case study demonstration, the selection criteria was mainly based on the following four dimensions:

- *Availability* - the selected ontology targets were tested to be technically available on the Web and published according to the Linked Data best practices and properly documented as well,
- *Coverage* - the selection was based on the precision of meaning and the degree of expressivity of the ontology target property,
- *Quality* - the selection was based on the structure and representation of the ontology target. Selected ontology targets were: 1) self-descriptive; 2) labels were provided for concepts and properties; and 3) definitions and comments on concepts and properties were provided,



- *Value representation* - the selection of ontology targets on the basis of the more preferred semantic value creation variables – semantic structure (syntax) and semantic meaning (semantics).

Therefore, for the literal *name*, option 2 (<http://semanticscience.org/resource/name>) was chosen for satisfying the criteria above.

Domain ODSE Approach

The second sub-process of the *ODSE approach component* outlines the method for the development of custom domain ontologies (ODSE activity 4.2), thus, pertains to the domain ODSE approach. The term domain ontology identifies a custom domain ontology like NPIs ontology for the development agenda domain in the case study of Botswana. This sub-process pertained to the development of NPI ontology (NPIOnto) for semantic representation of NPIs in the context of the Botswana development agenda domain. For this research case demonstration, the development of the ontology followed the adapted METHONTOLOGY approach and involved the following steps:

Step 1, the specification phase (ODSE sub-activity 4.2.1) involves the statement of the ontology purpose in terms of goal, scope and intended use. For the Botswana case study, the identified goal of NPIOnto was to address Botswana's OD representation needs regarding NPI disambiguation, disaggregation, interoperability, integrability and multidimensional analytical capabilities. The scope of NPIOnto covers the Botswana Integrated Indicator Framework comprising the Vision Agenda, NDP Agenda, domesticated SDG Agenda and Africa Agenda. NPIOnto is part of the ODSE Framework demonstration in which it is the basis for the conversion of flat text OD resources to knowledge resources. Step 2, the Knowledge acquisition phase (ODSE sub-activity 4.2.2) pertains to the identification of ontology concepts, their characteristics and relationships from source data and metadata resources gathered in Stage 1. The output at this stage comprised collections of development agenda domain knowledge on NPIs in loose format (spreadsheet) – appendix 1-4. Step 3, the conceptualisation phase (ODSE sub-activity 4.2.3) involves the design and development of ontology schema. For the Botswana demonstration case, the deliverables comprised the conceptual models for the four development agendas (Figures 6.1–6.4) and a consolidated conceptual model (Figure 6.5) that integrates the four individual conceptual models on the basis of shared classes,



object properties and data properties. Step 4, the formalisation phase (ODSE sub-activity 4.2.4) involves the codification of the consolidated knowledge model (schema) and the population of the knowledge base with instances. For the Botswana demonstration case, the output of the formalisation phase is the NPIOnto comprising the knowledge model (Figure 6.6) and the knowledge base (Figure 6.11). Step 5, the evaluation phase (ODSE sub-activity 4.2.5) involves the assessment of the effectiveness of ontologies. For the Botswana case study, this involved the schema effectiveness design evaluation with the use of knowledge model metrics and the knowledge base effectiveness design evaluation with the use of knowledge base metrics.

9.3.2 Data Enrichment (ODSE Component 5.0)

The *data enrichment component* pertains to the conversion of flat text resources (Step 1 output) to a more rich, flexible and dynamic format for increased semantic capabilities value-adding potential. The *data enrichment component* has two sub-processes, *resource reconciliation* and *RDFization*. It creates knowledge resources through the addition of semantic layer to flat text resources which constrains the resource metadata to standardised vocabularies through resource reconciliation and RDFization. Resource reconciliation establishes potentials of OD accessibility and interoperability to web resources while resource RDFization creates the accessibility, understandability, integrability, and intelligibility capabilities through linkages to ontological constructs, mainly, classes and properties. The semantic enrichment approach (ODSE approach) serves as the basis for the data enrichment implementations. The choice of either the generic ODSE approach or the domain ODSE approach depends on the semantic enrichment requirements – i.e., either global visibility maximisation or knowledge value maximisation. Regardless of whichever ODSE approach is chosen, semantic enrichment constrains the metadata of enriched resources to relevant vocabularies as described below.

Resource Reconciliation (ODSE Activity 5.1)

The first sub-process of the *data enrichment component*, the *resource reconciliation* pertains to the act of determining the potential for link creations by assessing the compatibility of flat text resources to relating resources in the Web of Data. An automatised resource reconciliation process involves the online comparison of the local metadata



values to those in controlled vocabularies to assess the extent of data accessibility and interoperability. For the manual process, resource reconciliation is accomplished by downloading the selected ontology targets (identified under the ODSE Component 4.0) and uploading them to the local server environment or creating Web link addresses to selected Web ontology targets in the local server environment (ODSE sub-activity 5.1.1). These processes apply to both the generic and domain ODSE approaches. That is followed by the replacement of the metadata for flat text resources with matching values in relating web repositories/registries of selected ontology targets as reflected in Figures 7.1-7.2 (ODSE sub-activity 5.1.2). For the domain ODSE approach, the metadata is replaced with specific domain constructs from the custom domain ontology. In both cases, the reconciled Uniform Resource Locators provide more accurate metadata descriptions that are standardised to controlled vocabularies in the Web of Data for ease of OD accessibility and interoperability.

Resource RDFization (ODSE Activity 5.2)

The second sub-process of the *data enrichment component*, the *Resource RDFization* involves the generation of the RDF representation of data (ODSE sub-activity 5.2.1). This process involves the addition of a semantic layer to the data, preferably the ontology layer. The ontology layer has potential for maximal semantic enrichment value resulting from its compositional combination of both the standardised structure and context. That is, it has complete semantic value creation features in its make-up structure. All in all, the resource RDFization process represents the RDF-driven semantic integration which implements the actual data mappings to controlled vocabularies. First, flat text resource links to the Uniform Resource Locators of selected ontology targets are created on the basis of the *subject-predicate-object* structure (Figures 7.3-7.4). That is followed by the generation of triples and the final product of the mappings is the triple structure that specifies the links between the data in the form of *subject-predicate-object* structure (ODSE sub-activity 5.2.2).

9.3.3 Enrichment Output (ODSE Component 6.0)

The *enrichment output component* comprises the final output of the *ODSE processing* stage, knowledge resources. For the generic ODSE processing, the output is the generic knowledge resource (ODSE activity 6.1 – Figure 7.5) and the domain knowledge resource



(ODSE activity 6.2 – Figure 7.6) for domain ODSE processing. Knowledge resources represent the upgraded version of flat text resources to interlinked RDF data through the LOD-driven semantic enrichment. Knowledge resources are systematically processed, and presented in structures that support knowledge harvesting. They are comprised of both the syntax (structural) and the semantic (contextual) capabilities needed to access, understand, interoperate, integrate, and support intelligible operations. The syntax is defined by the RDF structure which is in the form *subject-predicate-object* while the semantics are provided by the defined ontology linkages. Thus, knowledge resources provide contextual and conceptual clarity for more explicit expression of performance definitions and measurements. Above that, linking the data to ontology constructs such as classes and properties enable the application of axioms defined for those classes and properties on the data and this is what enables knowledge resources' support for intelligible operations. Thus, knowledge resources are both human and machine readable, processable and actionable as well.

9.4 Stage 3: ODSE Post-processing

The *ODSE post-processing* is the final stage that concludes the ODSE Framework processing with knowledge resources structuring and deployment for publication and consumption.

9.4.1 Knowledge Resources Structuring (ODSE Component 7.0)

The *knowledge resources structuring component* pertains to the categorisation and packaging of knowledge resources for distribution and use.

Categorisation of Knowledge Resources (ODSE Activity 7.1)

The *knowledge resources structuring component* starts with the screening of knowledge resources by functionality support needs (publication and consumption needs) for packaging, followed by the creation of knowledge resource packages through triples integration (publication and consumption packages). For the Botswana case demonstration (Subsection 7.4.1), knowledge resources created from the SDGI data were generated separately for each SD Goal. The knowledge resources created from the metadata were also generated separately on the basis of performance definition



constructs or elements per agenda. They were screened according to the adopted semantic enrichment processing approach (ODSE approach) to draw a separation between the products of generic ontology enrichment and NPIOnto enrichment for assessments of the ODSE approach with OD value maximisation potential. For ease of publication and evaluation, these files were integrated into unified knowledge resource per ODSE approach output.

Packaging of Knowledge Resources (ODSE Activity 7.2)

The *knowledge resources structuring component* ends with the creation of knowledge resource packages, both publication and consumption packages through triples integration. Here, individual knowledge resources are consolidated into composite resources by combining triples. The consolidation of knowledge resources is for ease of knowledge mining/harvesting/extraction operations. For the Botswana case study, the triples in individual knowledge resources (Figure 7.7) were unified per ODSE approach (triples integration) into integrated knowledge resources, the integrated generic knowledge resource and the integrated domain knowledge resource.

9.4.2 Knowledge Resources Deployment (ODSE Component 8.0)

The *knowledge resources deployment component* pertains to the publication or distribution of knowledge resources, as well as, the consumption or use of knowledge resources.

Publication of Knowledge Resources (ODSE Activity 8.1)

The *knowledge resources deployment component* marks the deployment stage of the ODSE Framework processing which pertains to availing knowledge resources for use. This involves the dumping or depositing of the publication and consumption knowledge resource packages to knowledge mining or harvesting environments for sharing. That is, deployment of knowledge resources for sharing. This involves making the knowledge resources freely available in the Web of Data. The process starts with the choice of the most suitable deployment environment for knowledge resources sharing and harvesting, environments that enable both human and machine accessibility, understandability, interoperability, integrability and intelligibility. Knowledge resources can be uploaded to semantic publication libraries, query endpoints, semantic browsers and inference engines for publication and consumption. For this research case study, the integrated



knowledge resource packages created in the knowledge resources structuring phase were uploaded to the SPARQL endpoint for value mining as reflected in Figure 7.8.

Consumption of Knowledge Resources (ODSE Activity 8.2)

The *knowledge resources deployment component* also marks the consumption stage of the ODSE Framework processing which involves the deployment of knowledge resources for use through knowledge mining or harvesting tools and techniques. For this research case study, OD value (knowledge) was mined from integrated knowledge resource packages through SPARQL queries.

9.5 The ODSE Framework Implementation Support Tool Feature

Figure 9.1 maps the ODSE Framework implementation support tools introduced in Subsection 8.4.2 to applicable ODSE processes for reuse in ODSE Framework implementations. These tools played a major role in driving both the development and demonstration of the ODSE Framework and are reusable in future framework implementations as well. The ODSE Framework implementation support tools are categorised into the ODSE scoping models and the ODSE assessment models.

9.5.1 The ODSE Scoping Models

The *ODSE scoping models* are domain reference models that represent the development agenda domain scope based on the agenda documentations and so they determine the value to be created. The detailed discussion of the three knowledge resources reference models follows.

Open Data Evolution Model

The first reference model is the *Open Data Evolution Model* (Figure 4.1), a conceptual design artifact that represents the generic view of semantic consumption needs for Botswana. This reference model comprises three high-level (abstract) sub-processes, being, the raw OD evolutionary process model, the knowledge resources evolutionary process model and the knowledge diffusion evolutionary process. Specifically, it models the transition from OD to open knowledge centered on open knowledge creation/recreation, structuring, sharing, use and diffusion. This model was developed to serve as a point of reference in Botswana's OD-driven economic transitions and



transformations. The usability of this model goes beyond the ODSE Framework development to future ODSE Framework implementations. That is, it is reusable in other application scenarios that leverages on OD for knowledge creation with a diversity of applications as a generic development agenda domain scoping tool. As reflected in Figure 9.1, the *OD evolution model* is applicable for reuse in the gathering of the source data and metadata resources. It is also reusable as the domain scoping reference model for Botswana development agendas that form the consolidated national indicator framework. It is also reusable in applications that leverage flat text OD for knowledge resources. To be specific, it is the candidate ODSE Framework implementation support tool for:

- ODSE activity 1.1 (collection of OD publication resources) as a generic development agenda domain scoping tool for the assessment of the scope of OD publication resources (raw OD resources) for collection;
- ODSE activity 1.2 (collection of OD publication and consumption specification resources) as a generic development agenda domain scoping tool for the assessment of the scope of OD publication and consumption specification resources to be collected;
- ODSE activity 2.1 (processing of OD raw resources – the cleaning, transformation and integration of OD) for the reconciliation of the ongoing pre-processing transformations to resource cleaning, transformation and integration requirements specifications; and
- ODSE activity 2.2 (identification of semantic publication and consumption needs) for:
 - assessment of the scope of search for semantic publication needs. The semantic publication scope it covers comprises OD integration needs pertaining to capabilities for OD integration across economic and social structures – semantic integration through vocabularies; OD enrichment needs pertaining to capabilities for OD enrichment through vocabularies (common representation of disparate, diversities of data); Enriched OD storage needs pertaining to capabilities for open and integrated data storages or knowledge repositories for the Vision Agenda; and Enriched



OD deployment pertaining to the need for OD publication and consumption platforms.

- assessment of the scope of search for semantic consumption needs. The model defines the consumption needs scope in economic and social terms, as pertaining to the need for knowledge circulation within the society (portrayed through the OD value (knowledge) creation cycle modelling) and the need for knowledge use and the potential resulting impact in the social and economic spheres (portrayed through the OD value (knowledge) use and resulting impact modelling).
- ODSE Activity 2.3 (deduction of semantic publication and consumption modelling needs) to serve as the basis for preliminary framework conceptualisation.

More details on the *Open Data Evolution Model* (Figure 4.1) are presented in Section 4.2 and published in [120].

NPIOnto Scope Model

The second reference model is the *NPIOnto Scope Model*, a high-level conceptual and semantic representation of the integrated NPIs and their properties for the development agenda domain. NPIOnto embodies insights that address the domain scoping needs in the Botswana development agenda domain applications, with a specific focus on the Botswana Integrated Indicator Framework scope. Figure 9.2 shows how NPIOnto integrates the four development agendas that are of interest in this study.



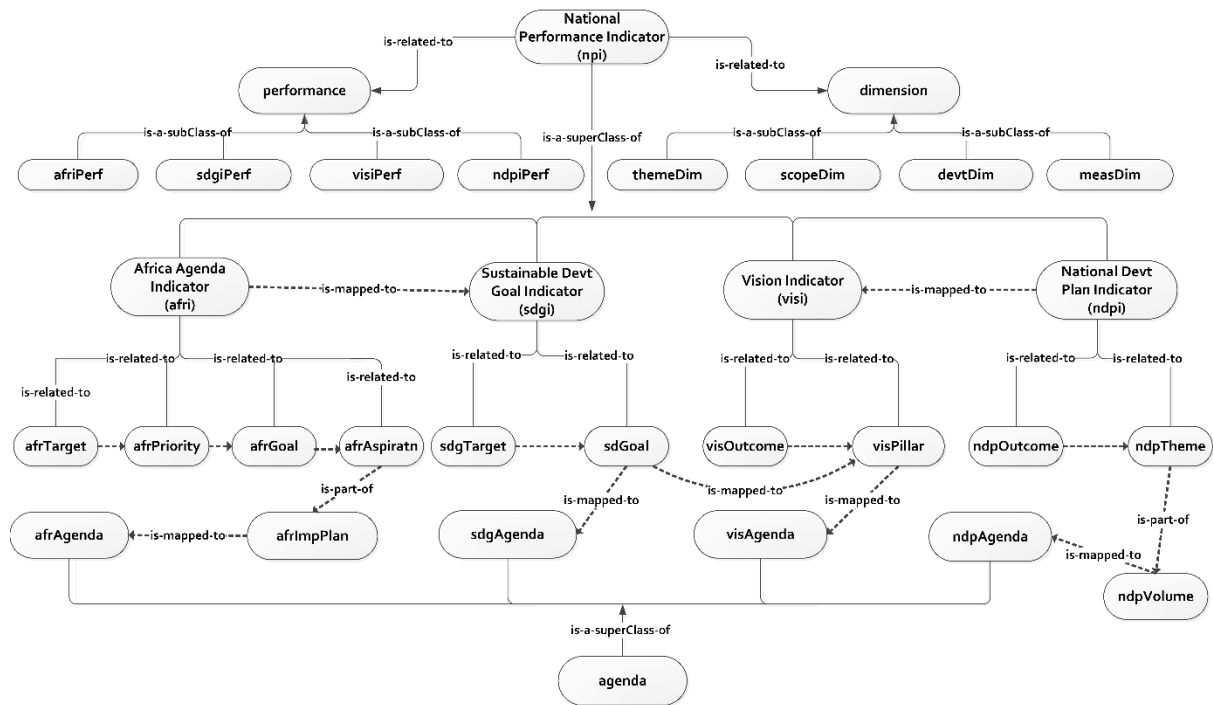


Figure 9.2: NPIOnto Scope Model

Figure 9.2 provides the backbone representation model for the consolidated indicator framework relations for the four agendas. Each agenda has both internal and external relations. At the top, Figure 9.2 depicts that the National Performance Indicator concept, *npi*, at the core of NPIOnto is-a-superClass-of the four agenda indicators, being Africa Agenda Indicator - *afri*, Sustainable Development Goal Indicator - *sdgi*, Vision Indicator - *visi* and National Development Plan Indicator - *ndpi*. Class *npi* is related to agenda performance and dimension. In turn, dimension is a superclass of theme - *themeDim*, scope - *scopeDim*, development - *devtDim* and measurement - *measDim* while performance is a superclass of Africa Agenda Indicator performance - *afriPerf*, Sustainable Development Goal Indicator performance - *sdgiPerf*, Vision Indicator performance - *visiPerf* and National Development Plan Indicator performance - *ndpiPerf*. At the bottom, Figure 9.2 depicts that the concept agenda is-a-superClass-of the four agendas: Africa Agenda - *afriAgenda*, Sustainable Development Goals Agenda - *sdgAgenda*, Vision Agenda - *visAgenda* and National Development Plan Agenda - *ndpAgenda*. Each agenda has relations between internal performance indicator variables.



For *afrAgenda*, the relations are as follows: 1) Africa Agenda Indicator - *afri* has a direct relation to Africa Agenda Target - *afrTarget*; 2) *afrTarget* has a direct relation to Africa Agenda Priority - *afrPriority*; 3) *afrPriority* has a direct relation to Africa Agenda Goal - *afrGoal*; 4) *afrGoal* has a direct relation to Africa Agenda Aspiration - *afrAspiratn*; 5) *afrAspiratn* has a direct relation to Africa Agenda Implementation Plan - *afrImpPlan*; and 6) *afrImpPlan* is directly related to the a specific Africa Agenda instance - *afrAgenda*.

For *sdgAgenda*, the relations are as follows: 1) Sustainable Development Goal Indicator - *sdgi* has a direct relation to SDG Agenda Target - *sdgTarget*; 2) *sdgTarget* has a direct relation to SDG Agenda Goal - *sdGoal*; 4) *sdGoal* is directly related to a specific SDG Agenda instance - *sdgAgenda*.

For *visAgenda*, the relations are as follows: 1) Vision Agenda Indicator - *visi* has a direct relation to Vision Agenda Outcome - *visOutcome*; 2) *visOutcome* has a direct relation to Vision Agenda Pillar - *visPillar*; 4) *visPillar* is directly related to a specific Vision Agenda instance - *visAgenda*.

For *ndpAgenda*, the relations are as follows: 1) NDP Agenda Indicator - *ndpi* has a direct relation to NDP Agenda Outcome - *ndpOutcome*; 2) *ndpOutcome* has a direct relation to NDP Agenda Theme - *ndpTheme*; 3) *ndpTheme* has a direct relation to NDP Agenda Volume - *ndpVolume*; and 4) *ndpVolume* is directly related to a specific NDP Agenda instance - *ndpAgenda*. Lastly, agenda relations between external performance indicator variables are as follows. Class *afri* maps to *sdgi* while *sdGoal* maps to *visPillar* and *visi* maps to *ndpi*.

This scoping model for NPI domain knowledge representation is one of the ODSE Framework implementation support tools to serve as: 1) a potential ontology target for reuse; 2) a reference model in the specification, initial structuring, conceptualisation, formalisation, evaluation and deployment of NPI domain related applications; and 3) an NPI domain scoping tool that provides a common understanding of the development agenda domain. Specifically, the *NPIOnto scope model* is the candidate ODSE Framework implementation support tool for:



- ODSE Component 4.0 with numerous applications as an assessment tool for generic application in the choice of generic ontologies (ODSE activity 4.1) or be used as a potential candidate for semantic enrichment ontology target and any other reuse applications.
- the development of custom domain ontologies (ODSE activity 4.2), specifically, to serve as a reference model in the specification, initial structuring, conceptualisation (design), formalisation (development) and deployment of related NPI domain applications and systems.
 - At the specification phase (ODSE activity 4.2.1), it is applicable as an NPI domain scoping tool that provides a common understanding of the development agenda domain.
 - At the conceptualisation phase (ODSE activity 4.2.3), it is applicable as an NPI domain scoping tool for NPIOnto conceptualisation adaptations in similar domain applications.
 - At the formalisation phase (ODSE activity 4.2.4), it is applicable as an NPI domain scoping tool in the extraction of NPIOnto constructs (NPI concepts and property definitions) for reuse or extension among OD publishers and consumers.

Semantic Capabilities Relational Model

The third reference model is the *semantic capabilities relational model*, drawn from the ODSE framework validation findings to represent the order of formation of semantic value attributes for the five semantic enrichment attributes (semantic capabilities) that of interest in this research. The framework validation findings have revealed that the five semantic capabilities are closely linked to each other and have a dependency relationship which directly impact on each other as depicted in Figure 9.3.

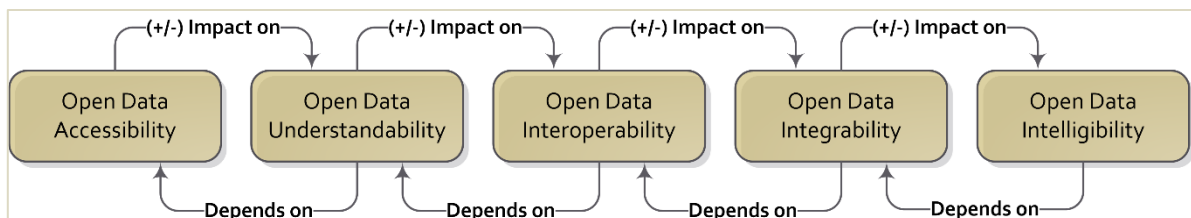


Figure 9. 3: Semantic Capabilities Relational Model



Therefore, the semantic capabilities relational model is a graphic representation of relations between OD accessibility, understandability, interoperability, integrability and intelligibility for insights on how to factor in semantic capabilities in the design and implementation of semantic enrichment process flows for applications that enhance semantic capabilities. This model provides insights on semantic capabilities formation dependencies to support holistic semantic capabilities modelling for semantic enrichment processing applications. The interpretation of Figure 9.3 is as follows:

- The resource accessibility capability precedes the understandability capability, thus, implying that OD inaccessibility hampers its understandability;
- The combination of resource accessibility and understandability capabilities, in turn, determine the interoperability capability, thus, implying that limitations in the accessibility and understandability capabilities lead to limitations of communication between raw OD resources;
- The combination of resource accessibility, understandability, and interoperability capabilities determine OD integrability capability, thus, implying that inaccessible, ambiguous, and interoperable data compromises its interlinking capability; and
- The combination of accessibility, understandability, interoperability, and integrability capabilities, in turn, drive intelligible operations. Above that, featuring in the knowledge resource modelling in Figure 9.3 into this analysis, reveals that, the wider is the linkages across the Web of Data, the stronger is the support for intelligent operations, and vice versa. In totality, the degree of semantic richness impacts on the degree of access to the data. That in turn determines the degree of OD understandability, interoperability, integrability, and intelligibility, and that has corresponding impact on the knowledge extraction potential.

As reflected in Figure 9.1, the *semantic capabilities relational model* is a candidate ODSE Framework implementation support tool for the *data enrichment component* (ODSE Component 5.0). As the ODSE Framework is centred on semantic capabilities enhancement, this semantic value assessment model is handy in the implementation of the *data enrichment component*, specifically, in the assessment of how to factor in



semantic capabilities in the design and implementation of semantic enrichment process flows for applications that enhance semantic capabilities.

9.5.2 The ODSE Assessment Models

In the context of this research, *semantic value* refers to knowledge. Ontology enriched open data resources are the embodiment of knowledge, hence, knowledge resources. The *ODSE assessment models* are *semantic valuation models* that qualitatively rate *semantic value* of ontology enriched open data resources (knowledge resources) for embodied knowledge, the product of semantic value creation or recreation implementation. The qualitative rating schemes, expressed in terms of relationships between different entities in the development agenda domain, were drawn on the basis of the preliminary ODSE Framework validation findings (Section 8.3). *Semantic valuation* in the context of this research is the relationship between the *semantic attribute* variable and the *semantic value measure*. The *semantic value measure* is a qualitative measure (i.e., minimum, moderate and maximum) for the *semantic attribute* variable that determines the degree of embodied knowledge and semantic capabilities (i.e., accessibility, understandability, interoperability, integrability and intelligibility) formation for knowledge resources. The *semantic attribute* variable is also the semantic impact assessment variable and the semantic impact adjusting variable. It can be enhanced through the addition of the syntax and semantic layer to flat text resources to achieve the semantic impact value per specific ontology enrichment requirement for a given semantic value creation need. In this regard, semantic value assessment tools enable the determination of semantic value creation. Therefore, the *semantic value assessment tools* are information-driven tools for informed adjustments to the semantic value variable from four different dimensions – semantic enrichment, richness, mining and usability dimensions. The design of the tools was according to the *Semantic Valuation Tools Development Method* (RM8) outlined in the methodology Subsection 3.3.5 and the design of the models was guided by the following overarching question:

What are candidate models for assessment of the degree of semantic value for a given knowledge resource?



As semantic value creation potential is determined by the degree of semantic enrichment, richness and mining capabilities, this overarching question was broken down into the following sub-questions (SQs):

- **SQ1:** *What is the candidate model for assessment of the degree of semantic enrichment impact on semantic value for a given knowledge resource?*
- **SQ2:** *What is the candidate model for assessment of the degree of semantic richness impact on semantic value for a given knowledge resource?*
- **SQ3:** *What is the candidate model for assessment of the degree of semantic mining impact on semantic value for a given knowledge resource?*

The detailed discussion of the three knowledge resources evaluation models follows

Semantic Enrichment Assessment Model

The first assessment model, the *semantic enrichment assessment* model (Figure 9.4), answers the first evaluation question (EQ1):

What is the candidate model for assessment of the degree of semantic enrichment impact on semantic value for a given knowledge resource?

The demonstrations have revealed that the degree of semantic enrichment impact on semantic value measure for a given knowledge resource depends on the ODSE approach used in the semantic enrichment processing. To start with, the ODSE processing is founded on ODSE approaches. The ODSE approach identifies the type of ontology for semantic enrichment and that differs according to the ontology scope. The generic ODSE approach makes use of generic ontologies with global linkage coverage for global visibility while the domain ODSE approach makes use of custom domain ontologies with richer expressiveness for richer semantic enrichment value. The two ODSE approaches have different degrees in semantic enrichment effectiveness and in the knowledge potential depending on the domain specification capability. The domain ODSE approach, due to its detailed and in-depth domain modelling, has been proved to have the highest knowledge creation potential, consequently, culminating in the maximisation of knowledge harvesting potential. This is best summarised by the *Semantic Enrichment Assessment Model* depicted in Figure 9.4.



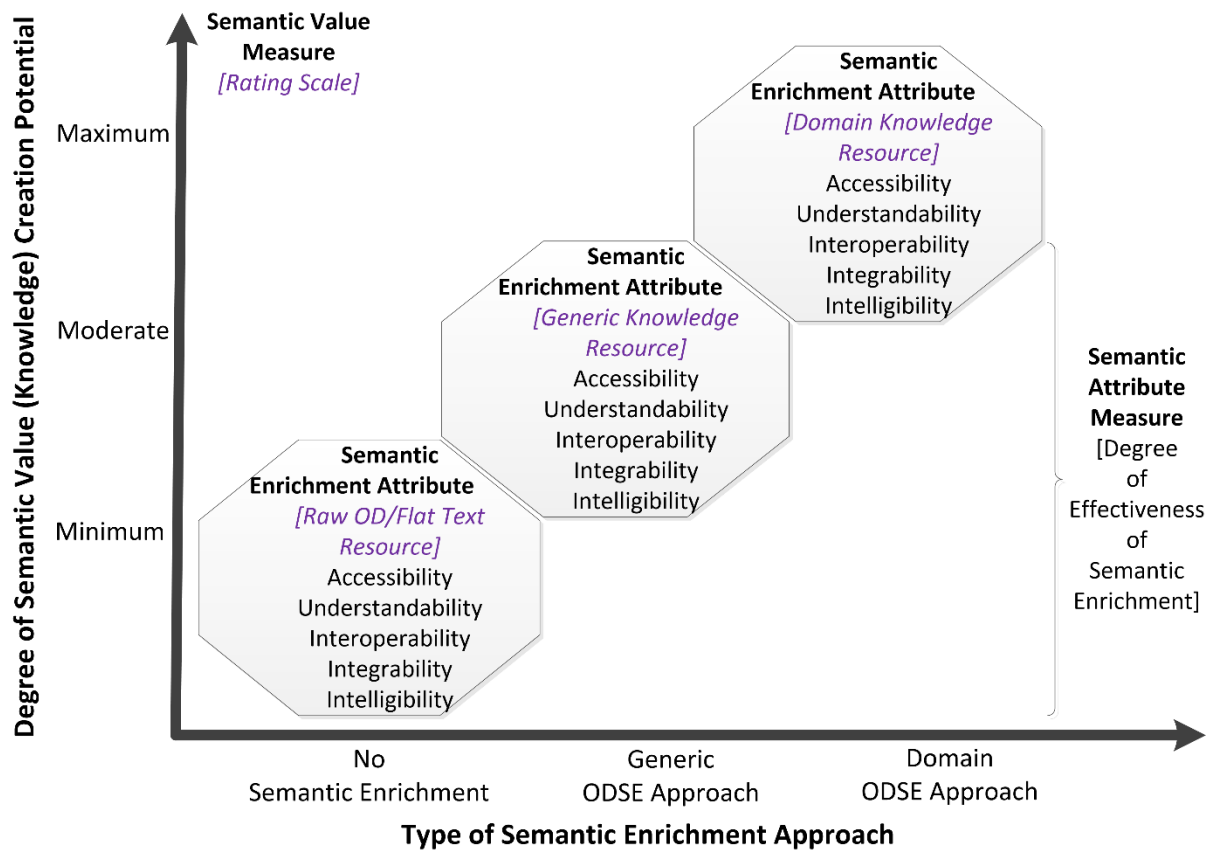


Figure 9. 4: Semantic Enrichment Assessment Model

Figure 9.4 represents knowledge creation capabilities per semantic enrichment approach, specifically, how the ODSE processing based on a given type of ODSE approach is likely to impact on the resulting knowledge resource. It further explores how the type of knowledge resource is likely to impact on the degree of semantic enrichment effectiveness, as well as, the degree of semantic value (knowledge) creation potential demonstrated by the resource’s degree of enriched semantic capabilities (accessibility, understandability, interoperability, integrability and intelligibility capabilities). The interpretation of Figure 9.4 is as follows:

- The non-enriched raw OD resources offer knowledge creation potential that ranges from minimum to moderate levels , determined by the resource’s degree of semantic capabilities;
- The generic ODSE approach to semantic enrichment processing produces generic knowledge resources offers knowledge creation potential that ranges from moderate to maximum levels, determined by the resource’s degree of semantic capabilities; and



- The domain ODSE approach to semantic enrichment processing produces domain knowledge resources with the differing degrees of maximum levels of knowledge creation potential, determined by the resource's degree of semantic capabilities.

The model depicts the kind of semantic enrichment suitable for global visibility maximisation and the one more suitable for knowledge harvesting maximisation. However, the choice of the ODSE approach for semantic enrichment processing is determined by the semantic enrichment intent. If the intent is to enhance semantic richness regardless of Web visibility, then the domain ODSE approach is most applicable but if it to enhance Web visibility regardless of semantic richness, then the generic ODSE approach is the best fit. Lastly, the two approaches can be combined to enhance semantic richness and Web visibility at the same time.

As reflected in Figure 9.1, the *semantic enrichment assessment model* is one of the candidate semantic valuation tools for ODSE Activity 7.1 (sorting of knowledge resources for packaging) applicable as a semantic value assessment tool for knowledge resources in four dimensions:

- in the assessment of the impact of a given ODSE approach on the resulting knowledge resources;
- in the assessment of the semantic capabilities value-added to a given knowledge resource;
- in the assessment of the OD value (knowledge) creation potential of a given knowledge resource, demonstrated by the resource's degree of enriched semantic capabilities and;
- to guide the knowledge resources classification for consumption and publication.

Semantic Richness Assessment Model

The second assessment model, the *semantic richness assessment* model (Figure 9.5), answers the second evaluation question (EQ2):

What is the candidate model for assessment of the degree of semantic richness impact on semantic value for a given knowledge resource?

The semantic enrichment demonstrations also proved that describing the data by means of ontologies increases knowledge creation capabilities as illustrated by Figure 9.5.



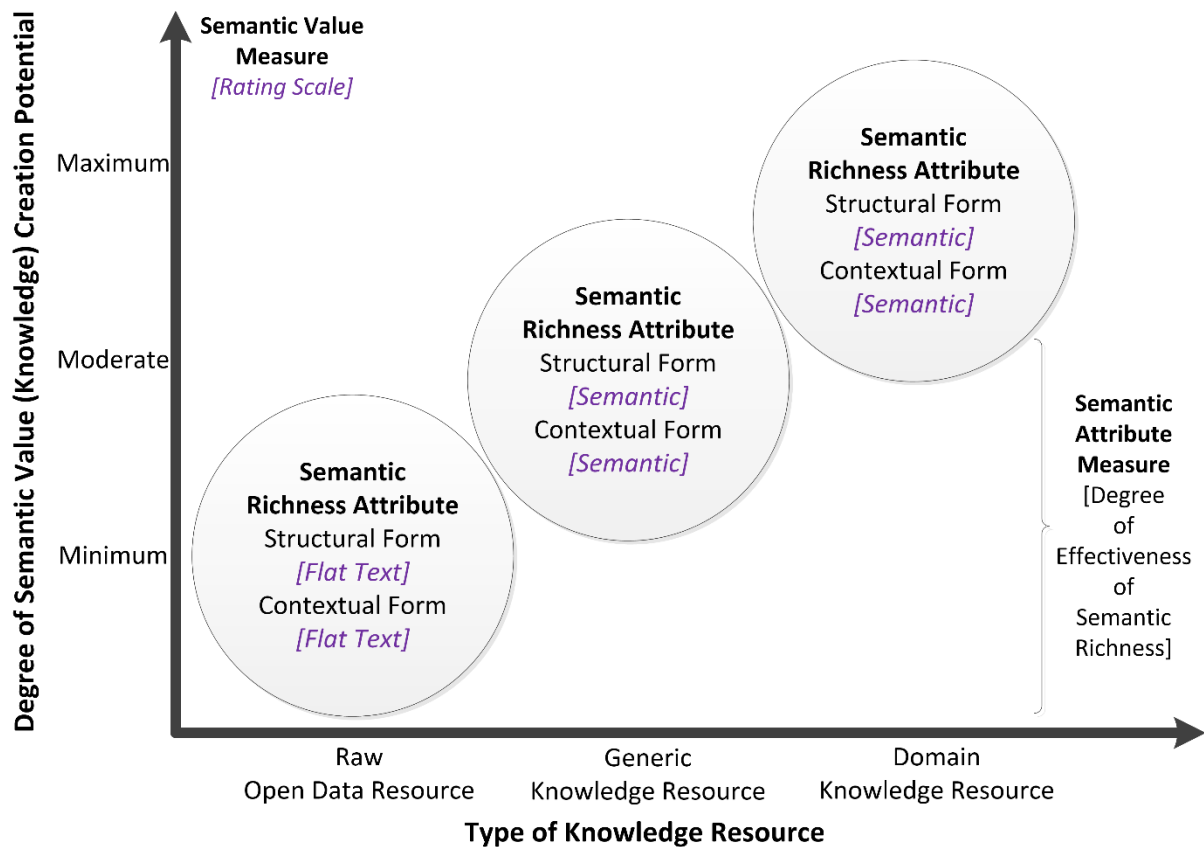


Figure 9.5: *Semantic Richness Assessment Model*

Figure 9.5 represents knowledge creation capabilities per knowledge resource. This model maps the potential impact of OD publication format on OD consumption capabilities. It gives insights on how the combination of OD resource publication structure (syntax/form) and context (semantics/meaning) impacts on the richness of semantic meaning and potential semantic value (created knowledge), and consequently, OD resource use (consumption) and, in overall, semantic value (knowledge) creation potential. The interpretation of Figure 9.5 is as follows:

- Raw OD resources, with minimum to moderate levels of semantic richness have low knowledge creation capabilities resulting from flat text structure and context that supports human consumption only;
- Generic knowledge resources, with moderate to maximum levels of semantic richness have high knowledge creation capabilities resulting from semantic structure and context that supports both human and machine consumption; and
- Domain knowledge resources, with the different degrees of maximum levels of semantic richness have the highest knowledge creation capabilities resulting from



semantic structure and context that supports both human and machine consumption.

Figure 9.5 shows that OD value (knowledge) appreciates with an increase in the richness of semantic meaning, and vice versa. The richness of semantic meaning results from the added semantic structure (syntax) and context (semantics) to raw OD that impact on semantic value (knowledge) mining/harvesting/extraction mode or type. Rich semantic formats maximise the semantic value (knowledge) mining by supporting both human and machine extraction capabilities. Moreover, the OD resource structure and context are determined by the type of OD resource. The OD resources with high value are presented in semantic format, the products of semantic enrichment processing.

Drawn from this graphical model are five potential applications. First is the knowledge resource grading/rating application drawn from the relation between the semantic attribute variable and the knowledge resource variable. This application involves the assignment of rates to semantic richness attributes for a given type of knowledge resource. Second is the resource selection application drawn from the relation between the knowledge resource, semantic richness attribute and the knowledge creation dependent variables. The application pertains to the selection of knowledge resources with appropriate semantic richness attributes for a given target of semantic value (knowledge) creation rating. Third is the resource selection application drawn from the relations between the knowledge resource, semantic richness attribute and the semantic attribute measure variables. This application involves the selection of knowledge resources with appropriate semantic richness attributes for a given target of semantic richness effectiveness rating. Fourth is the resource evaluation application drawn from the relation between the semantic attribute measure variable and the knowledge resource variable. The application pertains to the evaluation of the degree of effectiveness of semantic richness for a given type of knowledge resource. Last, is the resource evaluation application drawn from the relation between the knowledge creation dependent variable and the knowledge resource variable. This application involves the evaluation of the degree of semantic value (knowledge) creation potential for a given type of knowledge resource. For all potential semantic richness applications, semantic values are expressed in terms of the minimum, moderate and maximum rating scale.



As reflected in Figure 9.1, the *semantic richness assessment model* is the candidate ODSE Framework implementation support tool for:

- ODSE Activity 7.1 (sorting of knowledge resources for packaging) applicable in the assessment of how the semantic richness of a given resource is likely to determine its knowledge creation capabilities and to guide the knowledge resources classification for consumption and publication; and
- ODSE Activity 8.2 (knowledge resources consumption/use) applicable at the knowledge resources deployment stage to guide the execution of knowledge resources consumption/use/harvesting operations. Specifically, it is applicable for:
 - pre-consumption assessment of potential semantic value of a given knowledge resources
 - the assessment of how the semantic richness of a given OD knowledge resource is likely to influence/determine its knowledge creation capabilities,
 - the assessment of the appropriate knowledge harvesting tool for a given level of semantic richness of a knowledge resource.

Semantic Mining Assessment Model

The third assessment model, the *semantic mining assessment model* (Figure 9.6), answers the third evaluation question (EQ3):

What is the candidate model for assessment of the degree of semantic mining impact on semantic value for a given knowledge resource?

The other insight from the overall framework validation findings pertains to OD value (Knowledge) mining potential. The comparison of the knowledge mining potential for flat text resources to that of knowledge resources revealed that the mining potential depends on the semantic richness. In addition, the comparisons revealed that semantic richness of the domain knowledge resource highly impacts on the knowledge mining capabilities. The *Assessment Model for Open Data Resource Mining Capabilities* illustrating the link between the type of OD resource and the knowledge mining potential is shown in Figure 9.6.



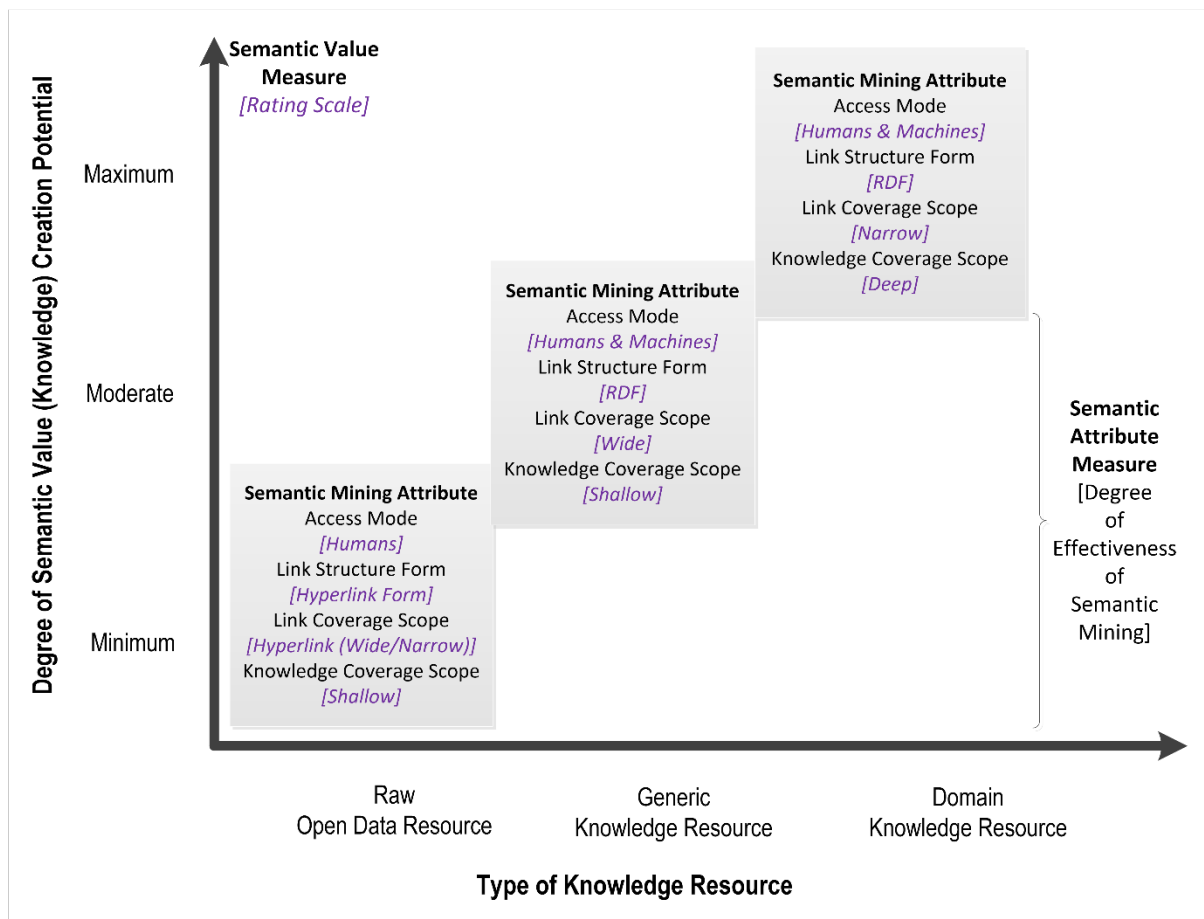


Figure 9. 6: Semantic Mining Assessment Model

Figure 9.6 represents OD resource mining capabilities per knowledge resource. The model gives insights on how the attributes of a given type of an OD resource is likely to impact on the degree of the knowledge mining/extraction/harvesting potential. This tool enables a balanced assessment of the key dimensional variables (data access, link structure, link coverage and knowledge coverage) of a given knowledge resource (that represent qualities to be embedded on the data through semantic enrichment processing so as to enhance the knowledge mining potential) to determine its knowledge mining capabilities. The interpretation of Figure 9.6 is as follows:

- For raw OD (flat text resources) access to the resources is most suitable to humans. The link structure and coverage are limited to hyperlinks. Therefore, the knowledge coverage is shallow. Flat text resources provide the different degrees of minimum levels of knowledge mining potential;
- For generic knowledge resources, access to the resources is suitable to both humans and machines. The RDF structure supports flexible knowledge mining



operations. However, the knowledge coverage is shallow due to the shallow coverage of the knowledge domain. Thus, generic knowledge resources provide the moderate level of knowledge mining potential; and

- For domain knowledge resources, access to resources is also suitable to both humans and machines and the flexible RDF structure supports flexible knowledge mining operations. The knowledge coverage is deep due to the in-depth coverage of the knowledge domain. Thus, domain knowledge resources provide the maximal level of knowledge mining potential.

Overall, the *Semantic Mining Assessment Model for Open Data Resource Mining Capabilities* is one of the ODSE Framework implementation support tools has numerous applications. It is applicable in the choice of ontology targets or in developments of new ontology targets, specifically, in the assessment of the ontology approach that best fits specific knowledge mining needs. It is reusable as an assessment tool in the: 1) selection of the ontology approach that best fits specific knowledge mining needs; 2) identification and specification of ontology enrichment requirements to support the production of knowledge resources that meet the knowledge mining results; and 3) selection of ontology targets for a given knowledge mining operation.

Drawn from this graphical model are five potential applications. First is the knowledge resource grading/rating application drawn from the relation between the semantic attribute variable and the knowledge resource variable. This application pertains to the assignment of rates to semantic mining attributes for a given type of knowledge resource. Second is the resource selection application drawn from the relation between the knowledge resource, semantic mining attribute and the knowledge creation dependent variables. This application involves the selection of knowledge resources with appropriate semantic mining attributes for a given target of semantic value (knowledge) creation measure. Third is the resource selection application drawn from the relation between the knowledge resource, semantic attribute and the semantic attribute measure variables. The application pertains to the selection of knowledge resources with appropriate semantic mining attributes for a given target of semantic mining effectiveness rating. Fourth is the resource evaluation application drawn from the relation between the semantic attribute measure variable and the knowledge resource



variable. This application involves the evaluation of the degree of effectiveness of semantic mining for a given type of knowledge resource. Fifth is the resource evaluation application drawn from the relation between the knowledge creation dependent variable and the knowledge resource variable. The application pertains to the evaluation of the degree of semantic value (knowledge) creation potential for a given type of knowledge resource. For all potential semantic mining applications, semantic value measures are expressed in terms of the minimum, moderate and maximum rating scale.

As reflected in Figure 9.1, the *semantic mining assessment model* is the candidate ODSE Framework implementation support tool for ODSE Component 4.0 with numerous applications:

- Generic application in the selection of the semantic enrichment approach (ODSE Component 4.0) that best fits the knowledge mining needs under consideration. This tool enables a balanced assessment of the key dimensional variables (data access, link structure, link coverage and knowledge coverage) of a given knowledge resource (that represent qualities to be embedded on the data through semantic enrichment processing so as to enhance the knowledge mining potential) to determine its knowledge mining capabilities.
- identification of ontology enrichment requirements (ODSE activity 4.1.2). Its assessment capabilities that supports the specification of semantic enrichment output (semantic resources) in terms of semantic qualities (data access, link structure, link coverage and knowledge coverage) that determine the knowledge mining potential, makes it handy in the identification and specification of ontology enrichment requirements to support the production of knowledge resources that meet the knowledge mining results.
- the selection of suitable ontology targets (ODSE activity 4.1.4), specifically, to serve as an assessment tool in the choice of ontology targets for a given knowledge mining operation. It gives insights on the potential impact of a given OD resource type (resource types differ in terms of the level of data access, link structure, link coverage and knowledge coverage) on OD value (knowledge) mining operations. Therefore, this model adds the fifth dimension of value extraction to the four described above (availability, coverage, quality and value representation). Value



extraction capabilities in a value-driven semantic enrichment solution becomes the basis for framework effectiveness evaluations.

- Potential application in the ontology specification phase of the custom ontology development process (ODSE activity 4.2.1) as an assessment tool in the identification and specification of ontology enrichment requirements to support the production of knowledge resources that meet the knowledge mining results.

Semantic Usability Assessment Model

The fourth assessment model, the *semantic usability assessment model* is a graphical representation of the qualitative matrix in the form of a 5-point rating scheme of OD publication and consumption effectiveness ratings derived from a combination of the semantic enrichment, richness and mining attributes and the Tim Berners Lee's 5-Star Scheme for Linked Data detailed in the methodology chapter (Subsection 3.3.4 – Table 3.1 and Figure 3.2). As reflected in Figure 9.1, these two are the ODSE Framework support tools for the overall assessment of the degree of semantic capabilities enrichment impact on semantic publication and consumption, specifically, for implementation of:

- ODSE activity 2.1 (processing of raw OD resources) for the assessment of the ODSE pre-processing input's publication and consumption effectiveness based on resources' semantic attributes;
- ODSE Activity 3.1 (processed raw OD resources) for the assessment of the publication and consumption effectiveness of pre-processed OD resources;
- ODSE Component 6.0 (enrichment output component) for assessment of the publication and consumption effectiveness of knowledge resources;
- ODSE Activity 7.1 (sorting of knowledge resources for packaging) applicable in the process of sorting knowledge resources for publication and consumption packaging.

In overall, the ODSE assessment models are qualitative tools for assessment of semantic value-adding potential as well as semantic value-added potential of knowledge resources re-created through semantic value creation implementations for semantic value capabilities and usability effectiveness. They draw potential relationships between different entities in the development agenda domain based on the demonstration and evaluation findings. The main purpose of the ODSE assessment models is to aid semantic



enrichment decision making for OD consumers with the need for semantically richer formats on the most applicable semantic value creation approach to adopt from four dimensions - semantic enrichment, semantic richness, semantic mining and semantic usability, so as to yield desired semantic value measure for a given type of knowledge resource.

9.6 Conclusion

This research proposes the semantic enrichment solution, the ODSE Framework, which is a composite of ODSE stages, ODSE processes, ODSE components, ODSE main activities, ODSE sub-activities, ODSE approaches, ODSE scoping models and ODSE assessment models. Following is a description of how all these parts fit together into one holistic conceptual artifact, the ODSE Framework.

First, are ODSE sub-activities that represent the lowest granularity of the ODSE processing, i.e., ODSE sub-activity 4.2.1. The ODSE sub-activities are implementing the ODSE main activities that represent tasks that can be broken down into other smaller tasks, i.e., ODSE activity 4.2. The ODSE main activities are grouped according to functionality into a single and unique component, the ODSE component, i.e., the input component (component 1.0) is a grouping of all input activities. Components can further be grouped together into different functionalities that implements unique ODSE processing, i.e., components 7.0 and 8.0, though having unique activities, all pertains to semantic enrichment post-processing, thus, belong to ODSE post-processing. The different ODSE processes are related through the implementation stage definition, ODSE stage 1 identified the first implementation process. The implementation of the ODSE processes are according to a chosen semantic enrichment approach. In addition, the implementation of ODSE processes are supported by the ODSE Framework implementation support tools, the ODSE scoping models and the ODSE assessment models. Moreover, the semantic value creation process is driven by semantic value creation needs that pertain to semantic consumption and publication needs. For instance, for the Botswana case study, the semantic consumption need pertains to OD support for Botswana's national agenda (Vision 2036 – Pillar 1), specifically, leveraging OD for the creation of knowledge resources to drive the transition from Botswana's dependence on



natural resources-driven economy to a knowledge resources-driven economy. This semantic consumption need is the building block for semantic publication needs. Semantic publication needs for the semantic enrichment framework pertain to: 1) OD program need for the re-usability of the current OD publications which prompts for enhancement of semantic capabilities; 2) OD presentation need for unified OD publications; and 3) OD representation (ontology) need for standardised publications and for the implementation of the Botswana Integrated Indicator Framework.

The semantic enrichment elements described above have a threefold functional focus. Some are focused on semantic value creation, others on semantic value creation maximisation and others on semantic valuation. These three functional divisions represent the three ODSE Framework pillars.

Lastly, the ODSE Framework combines all the semantic enrichment elements described above into one solution that maps the pathway toward addressing the quest for semantic enrichment frameworks with potential for semantic value (knowledge) creation maximisation. Overall, the ODSE framework is a template for more specific applications in future implementations that adopt or adapt it while its support tools are to aid its implementations.



CHAPTER TEN: RESEARCH CONCLUSIONS, RECOMMENDATIONS, FUTURE WORKS AND PUBLICATIONS

10.1 Conclusion

This dissertation concludes with an answer to the introductory, main research question stated in Section 1.1, formulated as:

What semantic enrichment framework could best support the implementation of semantic publication and consumption of OD with potential for OD value (knowledge) creation maximisation?

This research question ignited a quest to discover the functional design/architectural pillars of semantic enrichment frameworks with potential for semantic value (knowledge) creation maximisation. Semantic enrichment frameworks reviewed for an answer to this question (Section 1.2 and Subsection 2.5.6) were found to fully address the semantic value creation aspect of the main research question but limited in advancing the mechanism to enable facilitation and evaluation of the degree of semantic value creation capability of semantic value creation output (knowledge resources), i.e., whether minimum, moderate or maximum semantic value creation capabilities. So, the *Ontology-Driven Semantic Enrichment Framework for Maximal Creation of Open Data Value* (ODSE Framework) was created as an attempt to address this research gap by advancing the mechanism for the facilitation and evaluation of the degree of semantic value creation capability of knowledge resources. In this regard, the answer this research provides is:

The semantic enrichment framework that could best support the implementation of semantic publication and consumption of OD with semantic value maximisation potential is one where the pillar of semantic value creation processing is founded on the pillar of semantic value creation approach with potential for semantic value creation maximisation and supported by the pillar of semantic valuation.

The first ODSE Framework pillar, semantic value creation processing as reflected in Figure 9.1, starts with the identification of semantic value creation needs, modelled by ODSE Components 1.0-3.0. That is followed by the creation of semantic value through the addition of both the syntax and semantic layers and knowledge resources processing, modelled by ODSE Components 4.0-6.0. This pertains to the conversion of static, flat text



OD resources that are inaccessible, un-understandable, un-interoperable, un-integrable and un-intelligible to dynamic, open knowledge resources that are accessible, understandable, interoperable, integrable and intelligible. That is followed by the structuring of knowledge resources, implemented by Component 7.0. Lastly is the deployment of knowledge resources for sharing and knowledge mining/harvesting operations, implemented by component 8.0 and demonstrated in Section 7.4.2.

For the second ODSE Framework pillar, the semantic value creation maximisation approach, the semantic value creation process outlined above is driven by a combination of two semantic enrichment approaches with knowledge creation maximisation potential with cascading effects on resource mining and reusability potential. The domain ODSE approach maximises semantic attributes for improved resource capabilities demonstrated by NPIOnto's production of more specific and rich semantic enrichment results due to its specific/precise meanings and rich semantics (Figure 7.4). The generic ODSE approach maximises Web interoperability for improved resource visibility in the Web of Data (Figure 7.3). These capabilities were also demonstrated through resource reconciliation process in which the generic ontology (SDGIO) yielded only three matching global values (Figure 7.1) while the domain ontology (NPIOnto) with more domain specific terms, extended the reconciliation scope to seven values (Figure 7.2).

The third ODSE Framework pillar, the semantic valuation is implemented through semantic value assessment tools. Though this research addresses OD value, there is lack of semantic valuation models. In this regard, semantic valuation is an important ODSE Framework pillar that fits into this gap. So, semantic value assessment models were factored into the framework design as ODSE Framework implementation support tools. As depicted in Figure 9.1, the ODSE Framework has extended the semantic value creation focus of reviewed semantic enrichment frameworks by featuring in semantic value assessment into the syntax and semantics added layers to enable the assessment of the effectiveness of these two added enrichment layers. This enables informed decisions in effecting any changes that adjust both semantic publication and consumption capabilities. Moreover, as knowledge resources constitute digital economic investment resources/goods/assets with unlimited economic value that have lured the attention of global governments to shift their investments to, digital asset/resource/good value



determination becomes a critical investment factor. So, there must be a way for investors to determine the following: the semantic value measure to get for their knowledge resources investments; the expected output production capacity and value for their knowledge resources investments; and in overall, be able to qualitatively estimate the profitability of their knowledge resources investments in terms of knowledge use capabilities. All these pertain to semantic valuation of knowledge resources (knowledge resources quality measure) measured in terms of use impact.

10.2 Recommendations

The first recommendation pertains to the establishment of the public knowledge resource pool. As knowledge economies and societies thrive on shared knowledge, Botswana has to adopt an open approach to knowledge that would be core to the establishment, operational, maintenance and sustenance of the aspired economic and societal knowledge transformation and transition. Therefore, the creation of OD-based knowledge resources necessitates the establishment of a nationwide knowledge plant/mining environment in the form of a knowledge resources consolidation hub for government, economic, environment and social sectors. This research proposes that for sustainable, inclusive knowledge creation, publication, harvesting and sharing, knowledge resources/goods be scaled up into LOD platforms, interlinked knowledge bases that contain individuals and statements about them (triple-stores) to constitute the public knowledge resource pool, a nationwide knowledge plant in the form of web-based portal/platform for nationwide knowledge resources publication and consumption. In this case, then, knowledge resources would be deployed to this central processing hub.

The public knowledge resource pool has numerous potential tangible benefits. It will serve as a digital economic and social power hub within the global Web of Data so as to facilitate free, unlimited and unending data access to the general public. As a result, an economy based on the sharing and diffusion of knowledge provides an opportunity for governments to increase the well-being of their people. As knowledge resources' value appreciates with use, the wider the linkages, the deeper the semantics and the greater the semantic value measure. So, the greater the social participation in the economy, the greater the asset value. Consequently, the attainment of the key economic aspiration of



sustainable development - open, knowledge-driven, adaptive, and inclusive development, as well as, continuous growth in development. So, the public knowledge resource pool would serve as an unlimited engine for powering and sustaining the needs of the evolving future knowledge-driven economy and society in an open market environment where knowledge drives open innovation, business, and entrepreneurship creations/re-creations. Such an environment and platform would make it easy for developers and consumers to search, access, use, download and share knowledge resources of interest. It would also maximise the knowledge creation, mining and use potential as a result of the flexible data format and presentation structure (RDF). As the triples in the Web of Data can be connected to any other and there is no limit in access points for as long as there are similarities in either the subject, predicate or object. The public knowledge resource pool would also relieve consumers from the extensive and costly task of preliminary discovery and preparation of data. Moreover, it makes it easier to analyse and derive hidden insights in the integrated relationships and to manipulate data across all the consolidated resources. Moreover, for maximal effect, it must be built on the foundation of domain specific knowledge models like NPIOnto to enable the integration of both historical and current OD publications so as to expand the OD intelligibility capabilities to support trend analysis which is critical in strategic decision making and planning within the development agenda domain.

Despite this OD potential, both the Vision Agenda and NDP Agenda are silent on the potential of harnessing OD in the implementation, monitoring and evaluation of the development agendas. Therefore, this research recommends the next national development agendas to be explicit on the OD position and potential contribution to national agenda implementations.

The second recommendation pertains to OD representation needs. NPIOnto constrains the interlinked web resources to the development agenda domain vocabulary, specific to the consolidated indicator framework for agendas of interest in this case study. Specifically, NPIOnto is used in the reconciliation of text to web resources and the resource RDFization to enhance semantic capabilities with the aim of maximising the creation and use of knowledge resources. NPIOnto will serve as the foundation for the development of web applications and knowledge bases in the development agenda



domain. Furthermore, it can be used as a common reference model in the development of semantic web applications in order to improve communication, interoperability and automation of NPIs. Specifically, it can be used as a reference body of domain knowledge for the generation, integration, sharing and management of NPIs for Botswana's national, regional and global development agendas. It can also be implemented to serve as the basis for the design, development and management of NPI-based management, distribution, monitoring and evaluation systems for enhanced economic and social impact. However, despite these valuable NPIOnto potential, the evaluations of NPIOnto knowledge base, specifically, the Average Population and Class Richness revealed the different dimensions of insufficient instances for the class population. This limitation prompts a need for Botswana to improve on the quantity of statistics. The need for comprehensive data availability and quality has also been raised in the NDP Agenda which states that, "It is therefore necessary to improve on the quality of data collection, management, use, and dissemination at all levels [7]." Above that, the disintegration of NPIs at both the macro and micro level is also an issue of concern raised in the NDP Agenda. This poses a need for more specific mappings for Botswana national agendas. The domesticated SDG Agenda mappings to Vision Agenda are generically defined, and the domesticated SDGs are mapped to Vision Pillars. The challenge with implementing these generic mappings is that they have numerous performance measures associated with them. So, these mappings must be expressed at the lowest level of granularity for clarity in data representation. The mapping for Vision Agenda to NDP Agenda also needs to be clearly specified.

10.3 Future Mapping

The ODSE Framework is an application template which provides a generic view of the basic functions of semantic enrichment for knowledge creation maximisation in the development agenda domain. Thus, it can be customised to the needs of more specific value-based semantic enrichment applications that address specific semantic publication and consumption needs in the development agenda domain. Therefore, further researches in the development agenda domain are needed to either contribute to the improvement of this generic framework or to customise it to the needs of more specific



applications for other developing countries. In addition, researches are needed to extend the NPIOnto model beyond its current scope of the Botswana Integrated Indicator Framework. Above that, further research works are needed to evolve the ODSE Framework implementation support tools, especially the ODSE assessment models to address the current lack of OD valuation models. Moreover, the focus of the demonstrations were on semantic value creation, structuring and deployment for sharing and mining. This covers the open knowledge creation, structuring and sharing stages of the *Open Data Evolution Model* in Figure 4.1. The part remaining is putting the created semantic value into innovative and practical use. Thus, the use of resulting semantic value is yet to be determined in future studies. The last process in Figure 4.1, the knowledge diffusion evolutionary process, is a long term outcome that will be determined by the level of national commitment to the OD mandate.

10.4 Publications and Collaborations

This research involved numerous collaborations and publications as follows:

- 1) The *Open Data to Open Knowledge evolution model* that defined a roadmap for the development of the ODSE Framework was presented at the International Data Week 2018 (IDW 2018) held in Gaborone, Botswana from the 5th to the 8th of November 2018. The output of this participation is a paper entitled, *Open Data for Sustainable Development on a Knowledge-based Economy: The Case of Botswana* which is published in the Data Science Journal [120],
- 2) The preliminary ODSE Framework was presented at IST-Africa 2019 conference held in Nairobi, Kenya from the 8th to the 10th of May 2019. The output of this participation is a paper entitled, *Ontology-Driven Semantic Enrichment Framework for Value Maximisation: The Case of National Performance Indicators for Botswana* which is part of the IST-Africa 2019 Conference Proceedings published in IEEE Xplore [119], and
- 3) The conceptual model of the domain (NPI) ontology (NPIOnto) was presented at the Next Generation Computing Applications 2019 (NextComp 2019) conference held in Mauritius from the 19th to the 21st of September 2019. The output of this participation is a paper entitled, *A Lightweight Version of National Performance*



Indicator Ontology (NPIOnto) which is part of the NextComp 2019 Conference Proceedings published in IEEE Xplore [129].



REFERENCES

- [1] L. Ding, V. Peristeras and M. Hausenblas, "Linked Open Government Data," *IEEE Intelligent Systems*, no. 27, pp. 11-15, 2012.
- [2] Open Access Government, "The challenges posed by officially published open data," 28 August 2019. [Online]. Available: <https://www.openaccessgovernment.org/published-open-data/67170/>. [Accessed 02 May 2021].
- [3] European Commission, "Open Data Maturity in Europe 2015: Insights into the European state of play," data.europa.eu, 2015.
- [4] European Data Portal, "Open Data Goldbook for Data Managers and Data Holders: Practical," European Data Portal, 2018. [Online]. Available: <https://www.europeandataportal.eu/en/providing-data/goldbook/publishing-data>. [Accessed 4 July 2019].
- [5] Statistics Botswana, "Statistics Botswana Data Portal," Statistics Botswana, 2021. [Online]. Available: <http://www.statsbots.org.bw/data-portal>. [Accessed 21 September 2021].
- [6] Vision 2036, "Vision 2036: Achieving Prosperity for all," Government of Botswana, Gaborone, 2016.
- [7] Botswana Government, "National Development Plan (NDP) 11 Volume 1 (April 2017-March 2023)," Botswana Government, Gaborone, 2017.
- [8] African Union, "Agenda 2063: FTYIP 2013-2023 Core Indicators Profile Handbook," 2017.
- [9] UN, "Official Revised List of Global SDG Indicators_17.04.2017_Web," UN, New York City, 2017.
- [10] Statistics Botswana, "Botswana Domesticated Sustainable Development Goals Indicators," Statistics Botswana, Gaborone, 2018.
- [11] G. Pintzos, M. Matsas and G. Chryssolouris, "Defining Manufacturing Performance Indicators Using Semantic Ontology Representation," in *45th CIRP Conference on Manufacturing Systems*, Athens, Greece, 2012.
- [12] C. Diamantini, D. Potena and E. Storti, "SemPI: A semantic framework for the collaborative construction and maintenance of a shared dictionary of performance indicators," *Future Generation Computer Systems*, no. 54, pp. 352-365, 2015.



- [13] Statistics Botswana, "Sustainable Development Goals (SDGs) National Baseline Indicator Framework," Gaborone, 2018.
- [14] World Bank Group, "Botswana ODRA Report," 2015.
- [15] D. Moody and P. Walsh, "Measuring The Value Of Information: An Asset Valuation Approach," 1999.
- [16] N. Stehr, "Knowledge Societies," in *The Wiley-Blackwell Encyclopedia of Globalization*, Cambridge: Wiley-Blackwell, 2012.
- [17] H. Baazaoui-Zghal and G. H. Ben, "Fuzzy-Ontology-Enrichment-based Framework for Semantic Search," in *Proceedings of the 10th International Conference on Web Information Systems and Technologies (WEBIST-2014)*, 2014.
- [18] R. Fileto, C. May, C. Renso, N. Pelekis, D. Klein and Y. Theodoridis, "The Baquara knowledge-based framework for semantic enrichment and analysis of movement data," *Data & Knowledge Engineering*, vol. 98, pp. 104-122, July 2015.
- [19] B. Vrusias, D. Makris, J.-P. Renno, N. Newbold, K. Ahmad and G. Jones, "A Framework for Ontology Enriched Semantic Annotation of CCTV Video," *Eighth International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '07)*, June 2007.
- [20] R. D. Hull, D. Jenkins and A. McCutchen, "Semantic Enrichment and Fusion of Multi-Intelligence Data," Melbourne, FL 32901, 2006-2009.
- [21] E. Storti, F. Benvenuti, C. Diamantini and D. Potena, "An Ontology-based Framework to Support Performance Monitoring in Public Transport Systems," *Transportation Research Part C Emerging Technologies*, 2017.
- [22] A. Moraru and D. Mladenec, "A Framework for Semantic Enrichment of Sensor Data," *Journal of Computing and Information Technology - CIT 20*, no. 3, pp. 167-173, 2012.
- [23] The Royal Society, "Science as an open enterprise," The Royal Society, London, 2012.
- [24] Open Knowledge Foundation, [Online]. Available: <http://opendefinition.org/>.
- [25] T. Jetzek, M. Avital and N. Bjørn-Andersen, "Generating Value from Open Government Data," in *34th Int. Conf. on Information Systems*, Milan, 2013.
- [26] UNESCO, "Towards Knowledge Societies for Peace and Sustainable Development," Paris, 2013.



- [27] B. Melnikas and J. Samulevicius, "Sustainable Development and Knowledge Based Economy Creation Initiatives in Lithuania," in *5th International Vilnius Conference*, Vilnius, LITHUANIA, 2009.
- [28] J. Filipović, S. Devjak and G. Putnik, "Knowledge Based Economy: The Role of Expert Diaspora," *PANOECONOMICUS*, vol. 3, pp. 369-386, 2012.
- [29] Open Data Handbook, "Open Knowledge," Open Knowledge International, [Online]. Available: <http://opendatahandbook.org/>.
- [30] J. Ham, J.-N. Lee and B. Choi, "Effect of Government Data Openness on a Knowledge-Based Economy," *Procedia Computer Science*, no. 91, pp. 158-167, 2016.
- [31] Aceel Medium, "Open Data in the Knowledge Economy: Measuring the benenefits of open data in Lebanon," 2017.
- [32] M. Janssen, Y. Charalabidis and A. Zuiderwijk, "Benefits, Adoption Barriers and Myths of Open Data and Open Government," *Information Systems Management*, vol. 29, no. 4, pp. 258-268, 2012.
- [33] J. Benington, "From Private Choice to Public Value," in *Public Value: Theory and Practice*, Houndmills, Basingstoke: Palgrave MacMillan., Basingstoke, Palgrave MacMillan, 2011.
- [34] COMSA, "Comstat Data Hub," 2015. [Online]. Available: <http://comstat.comesa.int/onllghg/knowledge-economy-index>. [Accessed 21 November 2018].
- [35] Cornell University; INSEAD; WIPO , "Global Innovation Index 2018: Energising The World with Innovation," WIPO, 2018. [Online]. Available: http://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2018.pdf. [Accessed 21 November 2018].
- [36] OECD, "The Knowledge-based Economy," Paris, 1996.
- [37] OECD, "Assessing Government Initiatives on Public Sector Information: A Review of the OECD Council Recommendation," Paris, 2015.
- [38] M. A. Talaat, "Public Sector Information and Open Government Data," *JOURNAL OF SYSTEMS INTEGRATION*, no. 2, pp. 25-32, 2016.
- [39] I. Ramos, M. Cardoso, J. V. Carvalho and J. I. Graça, "An Action Research on Open Knowledge and Technology Transfer," *IFIP AICT 301*, p. 211-223, 2009.



- [40] O. Adedamola and O. Ejohwomu, "Knowledge Economy: A panacea for sustainable development in Nigeria," in *International Conference on Sustainable Design, Engineering and Construction*, 2016.
- [41] A. Zuiderwijk, N. Helbig, J. R. Gil-García and M. Janssen, "Special Issue on Innovation through Open Data - A Review of the State-of-the-Art and an Emerging Research Agenda: Guest Editors' Introduction," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 9, no. 2, pp. I-XIII, 2014.
- [42] F. A. Zeleti, A. Ojo and E. Curry, "Exploring the economic value of open government data," *Government Information Quarterly*, pp. 2-18, 2016.
- [43] B. Wessels, K. Wadhwa, R. L. Finn and T. Sveinsdottir, *Open Data and the Knowledge Society*, Amsterdam: Amsterdam University Press, 2017.
- [44] Civil Society, "Shaping Information Societies for Human Needs," 2003.
- [45] B. B. Edward, "The Concept of Sustainable Economic Development," *Environmental Conservation*, vol. 14, no. 2, pp. 101-110, June 1987.
- [46] Z. Wang, "Open Knowledge Management Model in Open Innovation Environment," in *IFLA WLIC 2017 – Wrocław, Poland – Libraries. Solidarity. Society. in Session S04 - Satellite Meeting: Knowledge Management Section.*, China, 2017.
- [47] Asia Foundation, November 2015. [Online]. Available: <https://asiafoundation.org/2015/11/04/ogp-global-summit-open-government-open-knowledge/>.
- [48] T. Roberts, "The Problem with Open Data," *Computerweekly*, February 2012. [Online]. Available: <https://www.computerweekly.com/opinion/The-problem-with-Open-Data/>. [Accessed 27 June 2022].
- [49] C. Böhm, M. Freitag, A. Heise, C. Lehmann, A. Mascher, F. Naumann, V. Ercegovac, M. Hernandez, P. Haase and M. Schmidt, "Govwild: Integrating open government data for transparency," in *Proceedings of the 21st International Conference Companion on World Wide Web*, New York, 2012.
- [50] M. Clarke and P. Harley, "How Smart Is Your Content? Using Semantic Enrichment to Improve Your User Experience and Your Bottom Line," *Science Editor*, vol. 37, no. 2, pp. 40-44, 2014.
- [51] A. D. Nicola and M. Missikoff, "A Lightweight Methodology for Rapid Ontology Engineering," *Communications of the ACM*, vol. 59, no. 3, pp. 79-86, 2016.



- [52] N. F. Noy and D. L. McGuinness, "Ontology development 101: a guide to creating your first ontology," 2001.
- [53] P. A. Martin, "Knowledge representation in CGLF, CGIF, KIF, Frame-CG and Formalized-English," in *10th International Conference on Conceptual Structures, Springer, Borovets,, Borovets, Bulgaria*, 2002.
- [54] J. V. Vuuren, L. Leenen and J. Zaaïman, "Using an ontology as a model for the implementation of the National Cybersecurity Policy Framework for South Africa," CSIR, Pretoria, 2014.
- [55] N. Guarino, "Formal Ontology in Information Systems," in *Proceedings of the 1st Intern. Conference*, 1998.
- [56] N. Santos, F. C. Campos and R. M. B. Villela, "Handbook of Research on Digital Libraries: Design, Development and Impact," IGI Global, Brazil, 2009.
- [57] J. Malone and H. Parkinson, "Reference and Application Ontologies," Ontogenesis, UK, 2010.
- [58] D. Vrandečić, "Ontology Evaluation," in *International Handbooks on Information Systems: Handbook on Ontologies*, Berlin, Springer-Verlag, 2009, pp. 293-313.
- [59] M. S. F. Farazi, "Faceted Lightweight Ontologies: A Formalization & some Experiments," Core.ac.uk, Italy, 2010.
- [60] T. Chungoora, "Practical Knowledge Modelling: Ontology Development 101," 2022. [Online]. Available: <https://www.udemy.com/practical-knowledge-modelling/>.
- [61] F. Giunchiglia and I. Zaihrayeu, "Lightweight Ontologies," Trento, 2007.
- [62] F. Kharbat and H. El-Ghalayini, "Building Ontology from Knowledge Base Systems," in *Data Mining in Medical and Biological Research*, Vienna, I-Tech, 2008, p. 320.
- [63] Ontotext, "Semantic Enrichment," Ontotext, May 2015. [Online]. Available: <https://ontotext.com/tag/semantic-enrichment/>.
- [64] M. J. Darlington and S. J. Culley, "Investigating ontology development for engineering design support," *Advanced Engineering Informatics*, no. 22, pp. 112-134, 2008.
- [65] S. Saeidlou, M. Saadat, E. A. Sharifi and G. D. Jules, "An ontology-based intelligent data query system in manufacturing networks," *Production & Manufacturing Research*, vol. 5, no. 1, pp. 250-267, 11 October 2017.



- [66] H. Dudycz and J. Korczak, "Process of Ontology Design for Business Intelligence System," *Information Technology for Management. Lecture Notes in Business Information Processing*, vol. 243, 2016.
- [67] L. Kalinichenko, M. Missikoff, F. Schiappelli and N. Skvortsov, "Ontological Modeling," in *Russian Conference on Digital Libraries*, Moscow, 2003.
- [68] United Nations Statistics Division, "Clarifying terms in the SDGs: representing the meaning behind the terminology," United Nations Environment Programme, New York City, 2015.
- [69] UNEP, "Ontologies: Sustainable Development Goals Interface Ontology (SDGIO)," EMBL-EBI, 10 October 2018. [Online]. Available: <https://www.ebi.ac.uk/ols/ontologies/sdgio.owl>. [Accessed 27 June 2022].
- [70] United Nations Statistics Division, "The Sustainable Development Goals Interface Ontology (SDGIO): Progress and peer review," United Nations Environment Programme, New York City, 2016.
- [71] TISAF, "Directive Information Identity," TISAF, [Online]. Available: https://www.tisaf.de/obi/directive_information_entity. [Accessed 30 November 2020].
- [72] C. Diamantini, L. Genga, D. Potena and E. Storti, "Collaborative Building of an Ontology of Key Performance Indicators," in *CoopIS'14*, 2014.
- [73] V. F. Junior, V. Uren, C. Brewster and A. L. Goncalves, "Ontology for Performance Measurement Indicators' Comparison," *International Journal of Digital Information and Wireless Communications (IJDIWC)*, vol. 6, no. 2, pp. 87-96, 2016.
- [74] A. Del-Río-Ortega, M. A. R. De-Reyna, D. A. Toro and A. Ruiz-Cortés, "Defining Process Performance Indicators by Using Templates and Patterns," in *10th Int. Conference, BPM 2012*, Tallinn, 2012.
- [75] B. Haslhofer and A. Isaac, "The Europeana Linked Open Data Pilot," in *Proc. Int'l Conf. on Dublin Core and Metadata Applications 2011*, 2011.
- [76] A. M. Fermoso, M. Mateos, E. M. Beato and R. Berjón, "Open linked data and mobile devices as e-tourism tools. A practical Approach to Collaborative e-Learning," *Computers in Human Behavior*, no. 51, pp. 618-626, 6 March 2015.
- [77] S. 2. NZ OPEN GOVERNMENT DATA AND INFORMATION PROGRAMME, "Applying the 5 Star Open Data model to your High value public data," 2012.
- [78] YDS, "A systematic review of open government data initiatives," 2015. [Online]. Available: <https://yourdatastories.eu/the-eu-opendata-handbook/>.



- [79] K. Janowicz, P. Hitzler, B. Adams, D. Kolas and C. Vardeman II, "Five Stars of Linked Data Vocabulary Use," *Semantic Web*, vol. 5, no. 3, pp. 173-176, 2014.
- [80] Ontotext, "Semantic Data Integration," Ontotext, 2018. [Online]. Available: <https://ontotext.com/knowledgehub/fundamentals/semantic-data-integration/>. [Accessed 17 April 2018].
- [81] Europeana, "Europeana semantic enrichment," Europeana, 5 November 2015. [Online]. Available: <https://pro.europeana.eu/page/europeana-semantic-enrichment>. [Accessed 6 July 2019].
- [82] University of Sheffield, NLP, "Module 6: Ontologies and Semantic Annotation," The University of Sheffield, 2016.
- [83] M. L. Zeng, "Semantic enrichment for enhancing LAM data and supporting digital humanities," *El profesional de la información*, vol. 28, no. 1, 2019.
- [84] A. Nicola and M. Missikoff, "A lightweight methodology for rapid ontology engineering," *Communications of the ACM*, vol. 59, no. 3, pp. 79-86, 1 Mar 2016.
- [85] T. Slimani, "Semantic Annotation: The Mainstay of Semantic Web," Taif University.
- [86] R. Lemmens, G. Falquet, S. De Sabbata, B. Jiang and B. Bucher, "Querying VGI by semantic enrichment," in *European Handbook of Crowdsourced Geographic Information*, London, Ubiquity Press, 2016, p. 185-194.
- [87] M. Poveda-Villalón, F. Radulovic and R. García-Castro, "Developing Ontologies for Representing Data about Key Performance Indicators," MADRID, SPAIN, 2014.
- [88] P. Hitzler and F. V. Harmelen, "A reasonable semantic web," *Semantic Web*, vol. 1, no. 1, pp. 39-44, 2010.
- [89] K. Reiche and E. Höfig, "Implementation of metadata quality metrics and application on public government data," in *COMPSAC workshops*, 2013.
- [90] E. Lapi, N. Tcholtchev, L. Bassbouss, F. Marienfeld and I. K. Schieferdecker, "Identification and Utilization of Components for a Linked Open Data Platform," in *IEEE 36th International Conference on Computer Software and Applications Workshops*, Berlin, 2012.
- [91] A. Polleres, A. Hogan, A. Harth and S. Decker, "Can we ever catch up with the web?," *Semantic Web*, vol. 1, no. 1, pp. 45-52, 2010.
- [92] ISO 25000 Software & Data Quality, "Understandability," [iso25000.com](https://iso25000.com/index.php/en/iso-25000-standards/iso-25012/122-understandability), 2020. [Online]. Available: <https://iso25000.com/index.php/en/iso-25000-standards/iso-25012/122-understandability>. [Accessed 25 February 2020].



- [93] L. Marco-Ruiz, K. Malm-Nicolaisen, R. Pedersen, A. Makhlysheva and P. Bakkevoll, "Ontology-based Terminologies for healthcare: Impact assessment and transitional consequences for implementation," Norwegian Centre for E-Health Research, 2017.
- [94] Open Data Handbook, "Open Data Interoperability and Integration," <https://opendatahandbook.org/guide/en/>, [Online]. Available: <https://opendatahandbook.org/guide/en/>. [Accessed 23 March 2021].
- [95] S. Pearlman, "What is Data Integration," Talend, 3 September 2019. [Online]. Available: <https://www.talend.com/resources/what-is-data-integration/>. [Accessed 25 February 2020].
- [96] Copyright Clearance Center, "Semantic Enrichment and the Information Manager," SciBite, 2017.
- [97] IGI Global, "What is Intelligent Data," IGI Global, [Online]. Available: <https://www.igi-global.com/dictionary/intelligent-data/36784>. [Accessed 9 December 2020].
- [98] Klaviyo Corporation, 5 September 2012. [Online]. Available: <https://www.klaviyo.com/blog/big-data-vs-intelligent-data-startups>. [Accessed 8 March 2021].
- [99] H. Manguinhas, "Europeana Semantic Enrichment Framework Documentation," Europeana, 2018.
- [100] P. Bellini, M. Benigni, R. Billero, P. Nesi and R. N., "Km4City ontology building vs data harvesting and cleaning for smart-city services," *Journal of Visual Languages and Computing*, no. 25, pp. 827-839, 2014.
- [101] OntoCore, "Centre for Applied Data Analytics Research (CeADAR)," Dublin Institute of Technology, [Online]. Available: www.ceadar.ie.
- [102] S. Angeletou, "Semantic Enrichment of Folksonomy Tagspaces," in *ISWC 2008*, Berlin Heidelberg, 2008.
- [103] C. Gustavo, E. Pilar, C. C. Rafael and M.-S. Manuel, "A linked open data framework to enhance the discoverability and impact of Cultural Heritage," *Journal of Information Science*, 2018.
- [104] L. Ruback, A. M. Casanova, A. Raffaetà, C. Renso and V. Vidal, "Enriching Mobility Data with Linked Open Data," in *IDEAS '16 Proceedings of the 20th International Database Engineering & Applications Symposium*, Montreal, 2016.
- [105] Y. Lee and J. Geller, "Semantic enrichment for medical ontologies," *Journal of Biomedical Informatics*, vol. 39, pp. 209-226, September 2006.



- [106] M. Alobaidi, K. Mahmood and S. Sabra, "Semantic Enrichment for Local Search Engine using Linked Open Data," in *WWW 2016 Companion*, Montréal, Québec, Canada, 2016.
- [107] F. Song, G. Zacharewicz and D. Chen, "An Ontology-Driven Framework towards Building Enterprise Semantic Information Layer," *Advanced Engeneering Informatics*, vol. 27, no. 1, pp. 38-50, January 2013.
- [108] K. Peffers, T. Tuunanen and A. M. Rothenberger, "A Design Science Research Methodology," *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45-78, 2007/8.
- [109] A. Hevner, S. March, J. Park and S. Ram, "Design Science in Information systms research," *MIS Quarterly*, no. 28, pp. 75-105, 2004.
- [110] J. Venable, J. Pries-Heje and R. Baskerville, "A comprehensive framework for evaluation in design science research:Design Science Research in Information Systems. Advances in Theory and Practice," *Advances in Theory and Practice*, p. 423-438, 2012.
- [111] M. Mattsson and J. Bosch, "Characterizing Stability in Evolving Frameworks," in *Proceedings of the 29th International Conference on Technology of Object-Oriented Languages and Systems*, 118-130.
- [112] R. Baskerville, M. Kaul and V. Storey, "Genres of inquiry in design-science research: Justification and evaluation of knowledge production," *Mis Quarterly*, no. 39, p. 541-564 , 2015.
- [113] Botswana Government, "National Development Plan (NDP) 11 Volume 1 (April 2017- March 2023," Botswana Government, Gaborone, 2017.
- [114] Northcentral University, "How to Make a Conceptual Framework," Applied Doctoral Experience, Northcentral University, 2 March 2022. [Online]. Available: <https://library.ncu.edu/Applied Doctoral Center/Section 1/Conceptual Framework>. [Accessed 15 June 2022].
- [115] A. Van der Merwe, A. Gerber and H. Smuts, "Mapping a Design Science Research Cycle to the Postgraduate Research Report," *ICT Education*, p. 293-308, 2017.
- [116] A. Isaac, H. Manguinhas, V. Charles and J. Stiller, "Selecting Target Datasets for Semantic Enrichment," Europeana, 2015.
- [117] M. Fernández-López, A. Gómez-Pérez and N. Juristo, "METHONTOLOGY: From Ontological Art Towards Ontological Engineering," in *Spring Symposium on Ontological Engineering of AAAI*, California, 1997.



- [118] C. Fischer and S. Gregor, "Forms of Reasoning in the Design Science Research," *Service-Oriented Perspectives in Design Science Research*, pp. 17-31, 2011.
- [119] O. Sebubi, I. Zlotnikova and H. Hlomani, "Ontology-driven Open Data Enrichment Framework for Value Optimisation: The Case of National Performance Indicators for Botswana," in *IST-Africa 2019 Conference Proceedings*, 2019.
- [120] O. Sebubi, I. Zlotnikova and H. Hlomani, "Open Data for Sustainable Development on a Knowledge-Based Economy: The Case of Botswana," *Data Science Journal*, pp. 1-6, 2019.
- [121] Open Data Watch, "Open Data to Support the Sustainable Development Goals".
- [122] Joint Minds Consult, "Participating in the Global Open Science/Open Data Space:Towards a Botswana National Data Forum and Strategy," 2016.
- [123] SDGs, "Statistics Botswana Portal," Statistics Botswana, [Online]. Available: <http://www.statsbots.org.bw/data-portal> .
- [124] E. Vladimir, "Open Data for Africa Platform".
- [125] H. Ericksson, "Digitizing Curricula: An Approach for Digital Usability," in *eChallenges e-2015 Conference Proceedings*, Vilnius, 2015.
- [126] A. Del-Río-Ortega, M. Resinas and A. Ruiz-Cortés, "Defining process performance indicators: An ontological approach," in *Proceedings of the 2010 International Conference on the Move to Meaningful Internet Systems - Volume Part I:OTM 10*, Berlin, 2010.
- [127] C. Diamantini, D. Potena and E. Storti, "Extended drill-down operator: Digging into the structure of performance indicators," *Concurrency & Computation: Practice & Experience*, vol. 28, no. 15, p. 3948–3968, November 2016.
- [128] G. Mitchell, "Problems and fundamentals of sustainable development indicators," *Sustainable Development*, vol. 4, no. 1, pp. 1-11, 1996.
- [129] O. Sebubi, I. Zlotnikova and H. Hlomani, "A Lightweight Version of National Performance Indicator Ontology (NPIonto)," in *NextComp 2019*, Mauritius, 2019.
- [130] H. Züllighoven and T. Slotos, "Chapter 2, The T&M object metamodel," in *Object-oriented construction handbook: Developing application-oriented software with the tools and materials approach*, Amsterdam (Netherlands), Elsevier Academic Press, 2005, pp. 17-56.
- [131] J. J. V. Vuuren, L. L., M. Grobler, K. Chan and Z. Khan, "Modelling the Cybersecurity Environment using Morphological Ontology Design Engineering," CSIR, Pretoria.



- [132] S. J. Miller, "Introduction to Ontology Concepts and Terminology," University of Wisconsin-Milwaukee, Lisbon, 2013.
- [133] Universitat Rostock, "OntoMetrics," Wirtschafs Informatik, September 2016. [Online]. Available: <https://ontometrics.informatik.uni-rostock.de/ontologymetrics/>. [Accessed 26 January 2021].
- [134] S. Tartir, I. B. Arpinar and P. A. Sheth, "Ontological Evaluation and Validation," *Computer Applications: Theory and Applications of Ontology*, pp. 115-130, 2010.
- [135] G. L. J. Castro, C. McLaughlin and A. Garcia, "Biotea: RDFizing PubMed Central in Support for the Paper as an Interface to the Web of Data," *Journal of Biomedical Semantics*, vol. 4(Suppl 1), no. S5, pp. 1-22, 2013.
- [136] Networked Planet, "networkedplanet.com," Turning flat data into semantic data with Open Refine and the RDF extension, 13 July 2015. [Online]. Available: <https://networkedplanet.com/blog/2015/07/13/turning-flat-data-into-semantic-data-with-open-refine-and-the-rdf-extension.html>. [Accessed 8 December 2020].
- [137] G. Montoya, I. Keles and K. Hose, "Querying Linked Data: An Experimental Evaluation of State-of-the-Art Interfaces," arXiv.org, Aalborg, 2019.
- [138] TopQuadrant, "How to publish linked data with topbraid live sparql endpoints," The Semantic Ecosystems Journal, 7 May 2010. [Online]. Available: <https://www.topquadrant.com/how-to-publish-your-linked-data-with-topbraid-live-sparql-endpoints/>. [Accessed 23 November 2020].
- [139] EU Open Data Portal, "EU Open Data Portal SPARQL endpoint," EU, 06 April 2017. [Online]. Available: <https://data.europa.eu/euodp/en/data/dataset/eu-open-data-portal-sparql-endpoint>. [Accessed 23 November 2020].



APPENDICES

Appendix 1: Africa Agenda Datasets Screenshots

Table AFR 1: Africa Agenda Indicator Dataset Screenshot

	A	B	C	D	E	F	G	H	I	J
1	Unique Indicator	Indicator Code	Indicator Name	Indicator Description	Formula	Unit of Measure	Disaggregation Factors	Data Collectn Freq	Target Relation	Sdgi Mapping
2	2023_I1	A1-G1-P1-T1-I1	Indicator 1 of 2023 implem	Gross National Income (GNI) per capita	Gross Domestic Product (GDP) + Net Income	US Dollars	rural/urban	annual	A1-G1-P1-T1	8.1.1
3	2023_I2	A1-G1-P1-I2	Indicator 2 of 2023 implem	Unemployment rate	unemployed/labour force	person(s)	Age, sex (m/f), vulnerability (disabled,	annual		8.5.2
4	2023_I3	A1-G1-P2-T6-I3	Indicator 3 of 2023 implem	Prevalence of underweight among children under 5	underweight/total No. of children who were weighted*100	percentage	Sex: Male, Female	annual, 3-5 yr	A1-G1-P2-T6	2.2.1
5	2023_I4	A1-G1-P2-T4-I4	Indicator 4 of 2023 _impPla	% of population with access to safe drinking water	No. of people who use an improved water source/total urban or rural population*100	percentage	urban/rural	annual, 3-5 yr	A1-G1-P2-T4	6.1.1
6	2023_I5	A1-G1-T4-I5	Indicator 5 of 2023 _impPla	% of household with access to electricity	Total No. of households that could be connected to the national electricity grid/total number of households*100	percentage	urban/rural	annual	A1-G1-T4	7.1.1
7	2023_I6	A1-G1-P4-T8-I6	Indicator 6 of 2023 _impPla	% of household using electricity	Households connected to national electricity grid/total No. of households*100	percentage	urban/rural	annual	A1-G1-P4-T8	
8	2023_I7	A1-G1-T4-I7	Indicator 7 of 2023 _impPla	% of population with access to the internet	Population which can access Internet/total eligible population*100	percentage	Age, sex (m/f), rural/urban, frequency of use	annual	A1-G1-T4	

Table AFR 2: Africa Agenda Target Dataset Screenshot

	A	B	C	D	E
1	unique Target	Target Code	Target Name	Target Description	Priority Relation
2	2023_A1-G1-P1-T1	A1-G1-P1-T1	target 1 under priority area 1 of goal 1 for aspiration 1 of 2023 implem	Increase 2013 per capita income by at least 30%	A1-G1-P1
3	2023_A1-G1-P1-T2	A1-G1-P1-T2	target 2 under priority area 1 of goal 1 for aspiration 1 of 2023 implem	Reduce 2013 unemployment rate by at least 25%	A1-G1-P1
4	2023_A1-G1-P1-T3	A1-G1-P1-T3	target 3 under priority area 1 of goal 1 for aspiration 1 of 2023 implem	Reduce youth and women unemployment rate by 2% per annum	A1-G1-P1
5	2023_A1-G1-P1-T4	A1-G1-P1-T4	target 4 under priority area 1 of goal 1 for aspiration 1 of 2023 implem	Reduce 2013 unemployment rate for vulnerable groups by at least 25%	A1-G1-P1
6	2023_A1-G1-P2-T5	A1-G1-P2-T5	target 5 under priority area 2 of goal 1 for aspiration 1 of 2023 implem	Reduce (improve) the 2013 Gini coefficient by at least 20%. Reduce income inequality	A1-G1-P2
7	2023_A1-G1-P2-T6	A1-G1-P2-T6	target 6 under priority area 2 of goal 1 for aspiration 1 of 2023 implem	Reduce stunting in children to 10% and underweight to 5%	A1-G1-P2
8	2023_A1-G1-P2-T7	A1-G1-P2-T7	target 7 under priority area 2 of goal 1 for aspiration 1 of 2023 implem	Reduce 2013 level of proportion of the population without access to safe drinking water	A1-G1-P2
9	2023_A1-G1-P4-T8	A1-G1-P4-T8	target 8 under priority area 4 of goal 1 for aspiration 1 of 2023 implem	Access and use of electricity and internet is increased by at least 50% of 2013 levels	A1-G1-P4
10	2023_A1-G2-P1-T1	A1-G2-P1-T1	target 1 under priority area 1 of goal 2 for aspiration 1 of 2023 implem	early childhood education is at least 300% of the 2013 rate	A1-G2-P1
11	2023_A1-G2-P1-T2	A1-G2-P1-T2	target 2 under priority area 1 of goal 2 for aspiration 1 of 2023 implem	enrolment rate for basic education is 100%	A1-G2-P1
12	2023_A1-G2-P1-T3	A1-G2-P1-T3	target 3 under priority area 1 of goal 2 for aspiration 1 of 2023 implem	Increase number of qualified teachers by at least 30% with focus on STEM	A1-G2-P1
13	2023_A1-G2-P1-T4	A1-G2-P1-T4	target 4 under priority area 1 of goal 2 for aspiration 1 of 2023 implem	universal secondary school (including technical high schools) with enrolment rate of 100%	A1-G2-P1
14	2023_A1-G3-P1-T1	A1-G3-P1-T1	target 1 under priority area 1 of goal 3 for aspiration 1 of 2023 implem	Increase 2013 levels of access to sexual and reproductive health services to women by at least 30%	A1-G3-P1
15	2023_A1-G3-P1-T2	A1-G3-P1-T2	target 2 under priority area 1 of goal 3 for aspiration 1 of 2023 implem	Reduce 2013 maternal, neo-natal and child mortality rates by at least 50%	A1-G3-P1
16	2023_A1-G3-P1-T3	A1-G3-P1-T3	target 3 under priority area 1 of goal 3 for aspiration 1 of 2023 implem	reduce the 2013 incidence of HIV/AIDS, malaria and TB by at least 80%	A1-G3-P1
17	2023_A1-G3-P1-T4	A1-G3-P1-T4	target 4 under priority area 1 of goal 3 for aspiration 1 of 2023 implem	Access to Anti-Retroviral (ARV) drugs is 100%	A1-G3-P1
18	2023_A1-G4-P1-T1	A1-G4-P1-T1	target 1 under priority area 1 of goal 4 for aspiration 1 of 2023 implem	Annual GDP growth rate of at least 7%	A1-G4-P1
19	2023_A1-G4-P2-T2	A1-G4-P2-T2	target 2 under priority area 2 of goal 4 for aspiration 1 of 2023 implem	Real value of manufacturing in GDP is 50% more than the 2013 level	A1-G4-P2

Table AFR 3: Africa Agenda Priority Dataset Screenshot

	A	B	C	D	E
1	unique Priority	Priority Code	Priority Name	Priority Description	Goal Relation
2	2023_A1-G1-P1	A1-G1-P1	priority area 1 under goal 1 of aspiration 1	Incomes, Jobs and Decent work	A1-G1
3	2023_A1-G1-P2	A1-G1-P2	priority area 2 under goal 1 of aspiration 1	Poverty, Inequality and Hunger	A1-G1
4	2023_A1-G1-P4	A1-G1-P4	priority area 4 under goal 1 of aspiration 1	Modern and Liveable Habitats and Basic Quality Services	A1-G1
5	2023_A1-G2-P1	A1-G2-P1	priority area 1 under goal 2 of aspiration 1	Education and STI driven Skills Revolution	A1-G2
6	2023_A1-G3-P1	A1-G3-P1	priority area 1 under goal 3 of aspiration 1	Health and Nutrition	A1-G3
7	2023_A1-G4-P1	A1-G4-P1	priority area 1 under goal 4 of aspiration 1	Sustainable inclusive economic growth	A1-G4
8	2023_A1-G4-P2	A1-G4-P2	priority area 2 under goal 4 of aspiration 1	STI driven Manufacturing /Industrialization and Value Addition	A1-G4
9	2023_A1-G4-P3	A1-G4-P3	priority area 3 under goal 4 of aspiration 1	economic diversification and resilience	A1-G4
10	2023_A1-G4-P4	A1-G4-P4	priority area 4 under goal 4 of aspiration 1	Hospitality/tourism	A1-G4
11	2023_A1-G5-P1	A1-G5-P1	priority area 1 under goal 5 of aspiration 1	Agricultural productivity and production	A1-G5
12	2023_A1-G6-P1	A1-G6-P1	priority area 1 under goal 6 of aspiration 1	Marine Resources and Energy	A1-G6
13	2023_A1-G7-P1	A1-G7-P1	priority area 1 under goal 7 of aspiration 1	Biodiversity, conservation, sustainable natural resource	A1-G7
14	2023_A2-G8-P2	A2-G8-P2	priority area 2 under goal 8 of aspiration 2	Political and economic integration	A2-G8
15	2023_A2-G10-P1	A2-G10-P1	priority area 1 under goal 10 of aspiration 2	Communications and Infrastructure Connectivity	A2-G10
16	2023_A3-G11-P1	A3-G11-P1	priority area 1 under goal 11 of aspiration 3	Democratic Values and Practices are the norm	A3-G11
17	2023_A3-G12-P1	A3-G12-P1	priority area 1 under goal 12 of aspiration 3	Institutions and Leadership	A3-G12
18	2023_A4-G13-P1	A4-G13-P1	priority area 1 under goal 13 of aspiration 4	Maintenance and Restoration of Peace and Security	A4-G13
19	2023_A4-G14-P1	A4-G14-P1	priority area 1 under goal 14 of aspiration 4	Institutional Structure for AU instruments on Peace and Security	A4-G14
20	2023_A4-G15-P1	A4-G15-P1	priority area 1 under goal 15 of aspiration 4	Operationalization of APSA Pillars	A4-G15
21	2023_A5-G16-P1	A5-G16-P1	priority area 1 under goal 16 of aspiration 5	Values and Ideals of Pan Africanism	A5-G16
22	2023_A6-G17-P1	A6-G17-P1	priority area 1 under goal 17 of aspiration 6	Women Empowerment	A6-G17
23	2023_A6-G17-P2	A6-G17-P2	priority area 2 under goal 17 of aspiration 6	Violence & Discrimination against Women and Girls	A6-G17
24	2023_A6-G18-P1	A6-G18-P1	priority area 1 under goal 18 of aspiration 6	Youth Empowerment and Children's Rights	A6-G18
agenda impPlan aspiration goal priority target indicator +					
READY					

Table AFR 4: Africa Agenda Goal Dataset Screenshot

	A	B	C	D	E
1	unique Goal	Goal Code	Goal Name	Goal Description	Aspiration Relation
2	2023_A1-G1	A1-G1	goal 1 under aspiration 1 of 2023 implement	A High Standard of Living, Quality of Life and Well Being for All	A1
3	2023_A1-G2	A1-G2	goal 2 under aspiration 1 of 2023 implement	Well Educated Citizens and Skills revolution underpinned by Science,	A1
4	2023_A1-G3	A1-G3	goal 3 under aspiration 1 of 2023 implement	Healthy and Well-Nourished Citizens	A1
5	2023_A1-G4	A1-G4	goal 4 under aspiration 1 of 2023 implementation	Plan for agenda 2063	A1
6	2023_A1-G5	A1-G5	goal 5 under aspiration 1 of 2023 implement	Modern Agriculture for increased productivity and production	A1
7	2023_A1-G6	A1-G6	goal 6 under aspiration 1 of 2023 implement	Blue/ ocean economy for accelerated economic growth	A1
8	2023_A1-G7	A1-G7	goal 7 under aspiration 1 of 2023 implement	Environmentally sustainable climate resilient economies and communities	A1
9	2023_A2-G8	A2-G8	goal 8 under aspiration 2 of 2023 implement	United Africa (Federal or Confederate)	A2
10	2023_A2-G9	A2-G9	goal 9 under aspiration 2 of 2023 implement	Key Continental Financial and Monetary Institutions established and	A2
11	2023_A2-G10	A2-G10	goal 10 under aspiration 2 of 2023 implement	World Class Infrastructure criss-crosses Africa	A2
12	2023_A3-G11	A3-G11	goal 11 under aspiration 3 of 2023 implement	Democratic values, practices, universal principles of human rights, justice	A3
13	2023_A3-G12	A3-G12	goal 12 under aspiration 3 of 2023 implement	Capable institutions and transformed leadership in place at all levels	A3
14	2023_A4-G13	A4-G13	goal 13 under aspiration 4 of 2023 implement	Peace, Security and Stability are preserved	A4
15	2023_A4-G14	A4-G14	goal 14 under aspiration 4 of 2023 implement	A Stable and Peaceful Africa	A4
16	2023_A4-G15	A4-G15	goal 15 under aspiration 4 of 2023 implement	A Fully Functional and Operational African Peace and Security Architecture	A4
17	2023_A5-G16	A5-G16	goal 16 under aspiration 5 of 2023 implement	African Cultural Renaissance is pre-eminent	A5
18	2023_A6-G17	A6-G17	goal 17 under aspiration 6 of 2023 implement	Full Gender Equality in All Spheres of Life	A6
19	2023_A6-G18	A6-G18	goal 18 under aspiration 6 of 2023 implement	Engaged and Empowered Youth and Children	A6
20	2023_A7-G19	A7-G19	goal 19 under aspiration 7 of 2023 implement	Africa as a major partner in global affairs and peaceful co-existence	A7
21	2023_A7-G20	A7-G20	goal 20 under aspiration 7 of 2023 implement	Africa takes full responsibility for financing her development	A7
22					
agenda impPlan aspiration goal priority target indicator +					
READY					



Table AFR 5: Africa Agenda Aspiration Dataset Screenshot

	A	B	C	D	E	F
1	unique Aspiration	Aspiration Code	Aspiration Name	Aspiration Description	Implementation Plan Relation	
2	2023_A1	A1	aspiration 1 of 2023 implementation plan	A prosperous Africa based on inclusive growth and sustainable development	FTYIP2023	
3	2023_A2	A2	aspiration 2 of 2023 implementation plan	An integrated continent, politically united	FTYIP2023	
4	2023_A3	A3	aspiration 3 of 2023 implementation plan	An Africa of good governance, democracy, respect for human rights, justice and the rule of law	FTYIP2023	
5	2023_A4	A4	aspiration 4 of 2023 implementation plan	A peaceful and secure Africa	FTYIP2023	
6	2023_A5	A5	aspiration 5 of 2023 implementation plan	Africa with a strong cultural identity, common heritage, values and ethics	FTYIP2023	
7	2023_A6	A6	aspiration 6 of 2023 implementation plan	An Africa whose development is people driven, relying on the potential of the African people	FTYIP2023	
8	2023_A7	A7	aspiration 7 of 2023 implementation plan	Africa as a strong and influential global partner	FTYIP2023	
9						
10						
11						
12						
13						
14						
15						
16						
17						

agenda impPlan aspiration goal priority target indicator +

READY

Table AFR 6: Africa Agenda Implementation Plan Dataset Screenshot

	A	B	C	D	E	F	G
1	unique Implementation Plan	Implementation Plan Code	Implementation Plan Name	Implementation Plan Description	Implementation Plan Beginning	Implementation Plan Ending	Agenda Relation
2	2023_FTYIP	FTYIP2023	First Ten Year Implementation Plan for Agenda 2063	Implementation plan for period 2013-2023	2013	2023	agenda2063

Table AFR 7: Africa Agenda Dataset Screenshot

	A	B	C	D	E	F
1	unique Agenda	Agenda Code	Agenda Name	Agenda Description	Agenda Beginning	Agenda Ending
2	2063_agenda	agenda2063	agenda 2063	africa agenda 2063	2013	2063

Appendix 2: SDG Agenda Datasets Screenshots

Table SDG 1: SDG Indicator Dataset Screenshot

	A	B	C	D	E	F	G	H	I	J	K	L
1	Unique Indicator	Indicator Code	Indicator Name	Indicator Description	unsdCode	Indicator Theme	Sub-theme	Indicator Pillar	Indicator Scope	Measure Type	Disaggregation Factors	Target Relation
2	2030_1.1.1	1.1.1	indicator 1 under target 1 of goal 1 for agenda 2030	Proportion of population below the international poverty line, by sex, age, employment status and geographical location (urban/rural)	C010101	No Poverty	below the internatic social		global		sex, age, employment status, geographic location (urban/rural)	1.1
3	2030_1.2.1	1.2.1	indicator 1 under target 2 of goal 1 for agenda 2030	Proportion of population living below the national poverty line, by sex and	C010201	No Poverty	below the national social		national		sex, age	1.2
4	2030_1.2.2	1.2.2	indicator 2 under target 2 of goal 1 for agenda 2030	Proportion of men, women and children of all ages living in poverty in all its dimensions according to national	C010202	No Poverty	poverty in all its dirn social		national		poverty dimension, population dimension (men/women/children)	1.2
				Proportion of population covered by social protection floors/systems, by sex, distinguishing children, unemployed persons,							sex, distinguishing children, unemployed persons, older persons, persons with disabilities, pregnant	

agenda goal target indicator +

READY

No new notifi



Table SDG 2: SDG Target Dataset Screenshot

	A	B	C	D	E
1	Unique Target	Target Code	Target Name	Target Description	Goal Relation
2	2030_1.1	1.1	target 1 under goal 1 of agenda 2030	By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than \$1.25 a day	goal1
3	2030_1.2	1.2	target 2 under goal 1 of agenda 2030	By 2030, reduce at least by half the proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions	goal1
4	2030_1.3	1.3	target 3 under goal 1 of agenda 2030	Implement nationally appropriate social protection systems and measures for all, including floors, and by 2030 achieve substantial coverage of the poor and the vulnerable	goal1
5	2030_1.4	1.4	target 4 under goal 1 of agenda 2030	By 2030, ensure that all men and women, in particular the poor and the vulnerable, have equal rights to economic resources, as well as access to basic services, ownership and control over land and other forms of property, inheritance, natural resources, appropriate new technology and financial services, including microfinance	goal1
6	2030_1.5	1.5	target 5 under goal 1 of agenda 2030	By 2030, build the resilience of the poor and those in vulnerable situations and reduce their exposure and vulnerability to climate-related extreme events and other economic, social and environmental shocks and disasters	goal1
				Ensure significant mobilization of resources from a variety of sources, including through enhanced development cooperation, in order to	

Table SDG 3: SDG Goal Dataset Screenshot

	A	B	C	D	E	F	G
1	Unique Goal	Goal Code	Goal Name	Goal Description	Agenda Relation	Pillar Mapping	
2	2030_01	goal 1	goal 1 of agenda 2030	End poverty in all its forms everywhere	agenda2030	pillar2	
3	2030_02	goal 2	goal 2 of agenda 2030	End hunger, achieve food security and improved nutrition and promote sustainable agriculture	agenda2030	pillar2	
4	2030_03	goal 3	goal 3 of agenda 2030	Ensure healthy lives and promote well-being for all at all ages	agenda2030	pillar2	
5	2030_04	goal 4	goal 4 of agenda 2030	Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all	agenda2030	pillar2	
6	2030_05	goal 5	goal 5 of agenda 2030	Achieve gender equality and empower all women and girls	agenda2030	pillar2	
7	2030_06	goal 6	goal 6 of agenda 2030	Ensure availability and sustainable management of water and sanitation for all	agenda2030	pillar3	
8	2030_07	goal 7	goal 7 of agenda 2030	Ensure access to affordable, reliable, sustainable and modern energy for all	agenda2030	pillar1	
9	2030_08	goal 8	goal 8 of agenda 2030	Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all	agenda2030	pillar1	
10	2030_09	goal 9	goal 9 of agenda 2030	Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation	agenda2030	pillar1	
11	2030_10	goal 10	goal 10 of agenda 2030	Reduce inequality within and among countries	agenda2030	pillar1	
12	2030_11	goal 11	goal 11 of agenda 2030	Make cities and human settlements inclusive, safe, resilient and sustainable	agenda2030	pillar1	
13	2030_12	goal 12	goal 12 of agenda 2030	Ensure sustainable consumption and production patterns	agenda2030	pillar3	
14	2030_13	goal 13	goal 13 of agenda 2030	Take urgent action to combat climate change and its impacts	agenda2030	pillar3	
15	2030_14	goal 14	goal 14 of agenda 2030	Conserve and sustainably use the oceans, seas and marine resources for sustainable development	agenda2030	pillar3	
16	2030_15	goal 15	goal 15 of agenda 2030	Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, conserve and enhance freshwater ecosystems and their services, and protect and restore the oceans and marine life	agenda2030	pillar3	
17	2030_16	goal 16	goal 16 of agenda 2030	Promote peaceful and inclusive societies for sustainable development, provide access to justice for all	agenda2030	pillar4	
18	2030_17	goal 17	goal 17 of agenda 2030	Strengthen the means of implementation and revitalize the Global Partnership for Sustainable Development	agenda2030	pillar4	

Table SDG 4: SDG Agenda Dataset Screenshot

	A	B	C	D	E	F
1	Unique Agenda	Agenda Code	Agenda Name	Agenda Description	Agenda Beginning	Agenda Ending
2	2030_agenda	agenda2030	agenda 2030	sustainable development goals agenda 2030	2015	2030
3						



Appendix 3: Vision Agenda Datasets Screenshots

Table VIS 1: Vision Indicator Dataset Screenshot

	A	B	C	D	E	F	G
1	Unique Indicator	Indicator Code	Indicator Name	Indicator Description	Data Source	Outcome Relation	NDPI Mapping
2	2036_P1-O1-I1	P1-O1-I1	indicator 1 under outcome 1 of pillar 1 for agenda 2036	Annual average rate of growth of real GDP	National Accounts	P1-O1	n/a
3	2036_P1-O1-I2	P1-O1-I2	indicator 2 under outcome 1 of pillar 1 for agenda 2036	Real Gross National Income (GNI) per capita	World Development Indicators	P1-O1	n/a
4	2036_P1-O2-I1	P1-O2-I1	indicator 1 under outcome 2 of pillar 1 for agenda 2036	Non-diamond export as a % of GDP	BIMTS, BoB	P1-O2	n/a
5	2036_P1-O2-I2	P1-O2-I2	indicator 2 under outcome 2 of pillar 1 for agenda 2036	Annual average rate of growth of non-diamond real GDP	National Accounts	P1-O2	n/a
6	2036_P1-O3-I1	P1-O3-I1	indicator 1 under outcome 3 of pillar 1 for agenda 2036	National wealth (assets)	Bank of Botswana	P1-O3	n/a
7	2036_P1-O4-I1	P1-O4-I1	indicator 1 under outcome 4 of pillar 1 for agenda 2036	Total Factor Productivity (TFP) growth	BNPC	P1-O4	n/a
8	2036_P1-O4-I2	P1-O4-I2	indicator 2 under outcome 4 of pillar 1 for agenda 2036	Global competitiveness ranking	WEF	P1-O4	n/a
9	2036_P1-O4-I3	P1-O4-I3	indicator 3 under outcome 4 of pillar 1 for agenda 2036	Doing business ranking	World Bank	P1-O4	n/a
10	2036_P1-O5-I1	P1-O5-I1	indicator 1 under outcome 5 of pillar 1 for agenda 2036	Unemployment rate	Various Surveys	P1-O5	n/a
11	2036_P1-O5-I2	P1-O5-I2	indicator 2 under outcome 5 of pillar 1 for agenda 2036	Number of new formal sector jobs per year	StB Employment	P1-O5	n/a
12	2036_P1-O6-I1	P1-O6-I1	indicator 1 under outcome 6 of pillar 1 for agenda 2036	Legatum Prosperity Index (ranking)	Legatum Inst.	P1-O6	n/a
13	2036_P2-O1-I1	P2-O1-I1	indicator 1 under outcome 1 of pillar 2 for agenda 2036	Share of population living under Poverty Datum Line (head-count poverty rate)	Household Survey	P2-O1	n/a
14	2036_P2-O1-I2	P2-O1-I2	indicator 2 under outcome 1 of pillar 2 for agenda 2036	Multidimensional Poverty Index (MPI)	including BCWIS,	P2-O1	n/a
agenda pillar outcome indicator + : ◀							
READY							

Table VIS 2: Vision Outcome Dataset Screenshot

	A	B	C	D	E
1	Unique Outcome	Outcome Code	Outcome Name	Outcome Description	Pillar Relation
2	2036_P1-O1	P1-O1	outcome 1 under pillar 1 of agenda 2036	High-income country	pillar1
3	2036_P1-O2	P1-O2	outcome 2 under pillar 1 of agenda 2036	Diversified economy	pillar1
4	2036_P1-O3	P1-O3	outcome 3 under pillar 1 of agenda 2036	Preserving national wealth	pillar1
5	2036_P1-O4	P1-O4	outcome 4 under pillar 1 of agenda 2036	Competitive & productive economy	pillar1
6	2036_P1-O5	P1-O5	outcome 5 under pillar 1 of agenda 2036	Adequate employment opportunities	pillar1
7	2036_P1-O6	P1-O6	outcome 6 under pillar 1 of agenda 2036	Broad-based prosperity	pillar1
8	2036_P2-O1	P2-O1	outcome 1 under pillar 2 of agenda 2036	Socially inclusive	pillar2
9	2036_P2-O2	P2-O2	outcome 2 under pillar 2 of agenda 2036	Gender equality	pillar2
10	2036_P2-O3	P2-O3	outcome 3 under pillar 2 of agenda 2036	Long & healthy lives	pillar2
11	2036_P2-O4	P2-O4	outcome 4 under pillar 2 of agenda 2036	Quality education	pillar2
12	2036_P3-O1	P3-O1	outcome 1 under pillar 3 of agenda 2036	Food security	pillar3
13	2036_P3-O2	P3-O2	outcome 2 under pillar 3 of agenda 2036	Water security	pillar3
14	2036_P3-O3	P3-O3	outcome 3 under pillar 3 of agenda 2036	Energy security	pillar3
15	2036_P3-O4	P3-O4	outcome 4 under pillar 3 of agenda 2036	Clean environment	pillar3
16	2036_P3-O5	P3-O5	outcome 5 under pillar 3 of agenda 2036	Biodiversity	pillar3
17	2036_P4-O1	P4-O1	outcome 1 under pillar 4 of agenda 2036	Accountable, transparent & effective governance	pillar4
18	2036_P4-O2	P4-O2	outcome 2 under pillar 4 of agenda 2036	Peaceful & secure nation	pillar4

Table VIS 3: Vision Pillar Dataset Screenshot

	A	B	C	D	E
1	Unique Pillar	Pillar Code	Pillar Name	Pillar Description	Agenda Relation
2	2036_Pillar1	pillar1	Sustainable Economic Development	pillar 1 of agenda 2036	vision2036
3	2036_Pillar2	pillar2	Human & Social Development	pillar 2 of agenda 2036	vision2036
4	2036_Pillar3	pillar3	Sustainable Environment	pillar 3 of agenda 2036	vision2036
5	2036_Pillar4	pillar4	Governance, Peace & Security	pillar 4 of agenda 2036	vision2036



Table VIS 4: Vision Agenda Dataset Screenshot

	A	B	C	D	E	F
1	Unique Agenda	Agenda Code	Agenda Name	Agenda Description	Agenda Beginning	Agenda Ending
2	2036_agenda	vision2036	agenda 2036	Botswana Vision 2036	2016	2036

Appendix 4: SDG Agenda Performance Datasets Screenshot

Table SDGI 1: Consolidated SDG Agenda Performance Datasets Screenshot

goalName	goalTheme	goalDesc	targetCode	targetDesc	sdIndicator	indicatorDesc	measureDe m
Sustainable Development	good health and we	Ensure healthy lives and promote	3.1	By 2030, reduce the global maternal mortality rat	3.1.1	Maternal mortality ratio	Maternal m p
Sustainable Development	good health and we	Ensure healthy lives and promote	3.1	By 2030, reduce the global maternal mortality rat	3.1.2	Proportion of births attended by skilled h	n/a
Sustainable Development	good health and we	Ensure healthy lives and promote	3.2	By 2030, end preventable deaths of newborns ar	3.2.1	Under-five mortality rate	Under-five p
Sustainable Development	good health and we	Ensure healthy lives and promote	3.2	By 2030, end preventable deaths of newborns ar	3.2.2	Neonatal mortality rate	n/a
Sustainable Development	good health and we	Ensure healthy lives and promote	3.3	By 2030, end the epidemics of AIDS, tuberculosis	3.3.1	Number of new HIV infections per 1,000 u	Number of p
Sustainable Development	good health and we	Ensure healthy lives and promote	3.3	By 2030, end the epidemics of AIDS, tuberculosis	3.3.2	Tuberculosis incidence per 100,000 popul	Tuberculosi p
Sustainable Development	good health and we	Ensure healthy lives and promote	3.3	By 2030, end the epidemics of AIDS, tuberculosis	3.3.3	Malaria incidence per 1,000 population	n/a
Sustainable Development	good health and we	Ensure healthy lives and promote	3.3	By 2030, end the epidemics of AIDS, tuberculosis	3.3.4	Hepatitis B incidence per 100,000 populat	n/a
Sustainable Development	good health and we	Ensure healthy lives and promote	3.3	By 2030, end the epidemics of AIDS, tuberculosis	3.3.5	Number of people requiring intervention	n/a
Sustainable Development	good health and we	Ensure healthy lives and promote	3.4	By 2030, reduce by one third premature mortalit	3.4.1	Mortality rate attributed to cardiovascular	n/a
Sustainable Development	good health and we	Ensure healthy lives and promote	3.4	By 2030, reduce by one third premature mortalit	3.4.2	Suicide mortality rate	n/a
Sustainable Development	good health and we	Ensure healthy lives and promote	3.6	By 2020, halve the number of global deaths and i	3.6.1	Death rate due to road traffic injuries	Death rate (n
Sustainable Development	good health and we	Ensure healthy lives and promote	3.7	By 2030, ensure universal access to sexual and re	3.7.1	Proportion of women of reproductive age	n/a
Sustainable Development	good health and we	Ensure healthy lives and promote	3.7	By 2030, ensure universal access to sexual and re	3.7.2	Adolescent birth rate (aged 10-14 years; a	n/a
Sustainable Development	good health and we	Ensure healthy lives and promote	3.9	By 2030, substantially reduce the number of dea	3.9.1	Mortality rate attributed to household an	n/a
Sustainable Development	good health and we	Ensure healthy lives and promote	3.9	By 2030, substantially reduce the number of dea	3.9.2	Mortality rate attributed to unsafe water, n/a	n/a
Sustainable Development	good health and we	Ensure healthy lives and promote	3.9	By 2030, substantially reduce the number of dea	3.9.3	Mortality rate attributed to unintentional	n/a
Sustainable Development	good health and we	Ensure healthy lives and promote	3.a	Strengthen the implementation of the World He	3.a.1	Age-standardized prevalence of current tr	Age-standa %
Sustainable Development	good health and we	Ensure healthy lives and promote	3.b	Support the research and development of vaccin	3.b.1	Proportion of the target population cover	n/a
Sustainable Development	good health and we	Ensure healthy lives and promote	3.b	Support the research and development of vaccin	3.b.2	Total net official development assistance	n/a
Sustainable Development	good health and we	Ensure healthy lives and promote	3.c	Substantially increase health financing and the	3.c.1	Health worker density and distribution	n/a
Sustainable Development	good health and we	Ensure healthy lives and promote	3.d	Strengthen the capacity of all countries, in partic	3.d.1	International Health Regulations (IHR) cap	n/a

Appendix 5: Source Data Inconsistency Prevalence Summary

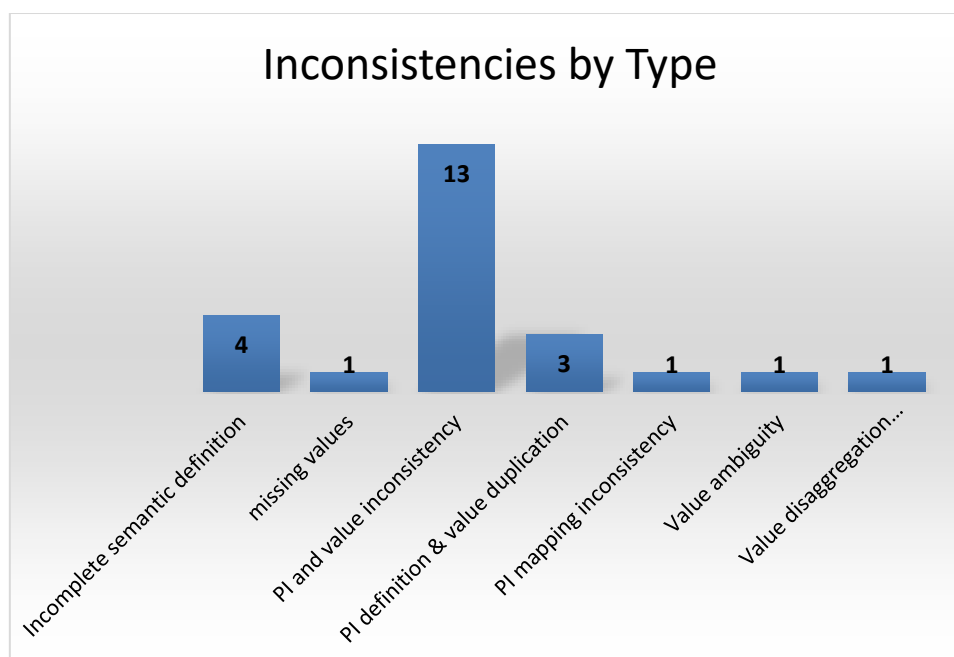


Figure Appx 5. 1: Source Data Inconsistencies by Type

