



# Ontology-Driven Semantic Enrichment Framework for Open Data Value Creation

RESEARCH PAPER

OARABILE SEBUBI

IRINA ZLOTNIKOVA

HLOMANI HLOMANI

\*Author affiliations can be found in the back matter of this article

ubiquity press

## ABSTRACT

The reviewed semantic enrichment frameworks lack mechanisms to assess the degree of semantic value added to flat-text resources in terms of knowledge and semantic capabilities. This complicates the tasks of driving the semantic value creation process toward the specific enrichment output and evaluating the output. In addressing this gap, we propose the semantic value creation solution, which converts flat-text resources to knowledge resources. Namely, we propose the ontology-driven semantic enrichment (ODSE) framework, with a mechanism for semantic valuation. The framework's development involved adopting the design science research methodology for information systems. The developed framework leverages linked data principles for knowledge creation. This framework was demonstrated to determine the semantic capabilities enabled by the syntax additions, as well as the knowledge enabled by the semantic additions to flat-text resources, along with its potential impact on knowledge creation, mining, and resource-usability effectiveness. The ODSE framework is reusable in semantic value creation implementations that transform a flat text to semantic formats.

## CORRESPONDING AUTHOR:

**Irina Zlotnikova**

Department of Computer Science and Information Systems, Botswana International University of Science and Technology, Palapye, Botswana (BW), Southern Africa

[zlotnikovai@biust.ac.bw](mailto:zlotnikovai@biust.ac.bw)

## KEYWORDS:

ontology-driven framework; semantic enrichment framework; open data value creation; semantic value creation; knowledge creation; semantic enrichment effectiveness

## TO CITE THIS ARTICLE:

Sebubi, O, Zlotnikova, I and Hlomani, H. 2023. Ontology-Driven Semantic Enrichment Framework for Open Data Value Creation. *Data Science Journal*, 22: 40, pp. 1–18. DOI: <https://doi.org/10.5334/dsj-2023-040>

## 1 INTRODUCTION

Despite its numerous benefits, open data (OD) remains mostly an afterthought across nations and governments (Godhwani & Kuzev 2023). According to the 2022 Global Data Barometer survey, a study conducted across 109 countries, there has been a growth of only 10.6 percent in the volume of government datasets that are truly 'open' since the previous survey, conducted in 2016 (D4D 2022). The advent of the COVID-19 pandemic heightened the need for such a shift to publication of data when the whole world was closed. Throughout the pandemic, there was an apparent need for digital services that would have been developed as part of an open data ecosystem with the ultimate goal of enabling the creation of new added value (Seljan et al. 2022). The report on the state of open data by Digital Science (2022) advocates for data to be more discoverable, usable, and citable. These sentiments are traced back to the FAIR (findable, accessible, interoperable, and reusable) data principles (Wilkinson et al. 2016).

The reusability limitations of OD can be attributed to flat-text formats that are difficult to access, understand, interoperate, and integrate. Consequently, the intelligibility and reusability potential of current OD publications is compromised, and, ultimately, the potential of semantic value creation is rendered less effective. Semantic limitations pertaining to OD accessibility, understandability, interoperability, integrability, and intelligibility capabilities impact negatively on the potential for knowledge creation, mining, and use. These limitations prompt a shift to semantic publications for more clarity on the semantic structure and context.

The needs of the current research were identified as follows. First, we conducted the review to identify OD support needs for Botswana Vision 2036 (further referred to as 'the Vision Agenda') and National Development Plan 11 (further referred to as 'the NDP Agenda') by analysing relevant documents (Vision 2036 Presidential Task Team 2016; Government of Botswana 2017; Sebubi et al. 2019a). In addition to that, we reviewed the Botswana OD readiness assessment survey documenting the roadmap to future OD publications for Botswana's OD program goals (World Bank Group 2015). Second, we carried out the review of OD presentation needs pertaining to an assessment of the current OD publication system, in terms of the OD presentation structure and format. We based the review on OD from the Statistics Botswana data portal and other metadata for the agendas that constituted the Botswana integrated indicator framework. The documents we reviewed were the Vision Agenda, the NDP Agenda, Africa Agenda 2063 (further referred to as 'the Africa Agenda'), and the global and domesticated 2030 Sustainable Development Goals Agenda (further referred to as 'the SDG Agenda') (Statistics Botswana 2021; Vision 2036 Presidential Task Team 2016; African Union 2017; Statistics Botswana 2018a; Government of Botswana 2017). Third, we identified the OD representation needs from all the above documents and the SDGs national baseline indicator framework proposal (Statistics Botswana 2018b). We identified additional OD representation needs from the published literature on performance indicator semantic representations.

We addressed the semantic value creation needs through the ontology-driven semantic enrichment (ODSE) framework modelling needs, measures, and functionality design specifications we propose in this research. The framework modelling needs, measures, and functionality design specifications address the semantic value creation needs. Table 1 summarises the semantic value creation needs in terms of semantic publication and use. The needs are presented in relation to the ODSE framework modelling needs and measures, and functionality design specifications.

The ODSE framework is proposed to guide the transformation of OD with limited semantic capabilities and knowledge creation potential into that enriched with semantic capabilities and knowledge. The scope of this paper includes the detailed description of the proposed ODSE framework and the Botswana case study demonstration. The qualitative evaluation of the ODSE framework was performed using the semantic usability assessment model adopted from Berners-Lee (2010) and modified to include effectiveness of semantic enrichment. Semantic value assessment helps to identify knowledge resources with the highest degree of semantic value creation and, therefore, with the highest potential of OD reuse. Four examples of potential applications of the semantic usability assessment model are also considered within the scope of this paper.

SEMANTIC VALUE CREATION NEEDS	ODSE FRAMEWORK MODELLING NEEDS	ODSE FRAMEWORK MODELLING MEASURES	ODSE FRAMEWORK FUNCTIONALITY DESIGN
<b>Use Need</b>			
OD support need for the Botswana Vision Agenda	The modelling need to support the generation of knowledge resources for the implementation of the Vision Agenda aspirations (knowledge-driven economy and knowledge-based society)	Modelling measures to leverage OD for the generation of knowledge resources for the implementation of the Vision agenda aspirations	Functionality design incorporating semantic enrichment, pre-processing, processing, and post-processing for the creation, structuring, and sharing of knowledge resources
<b>Publication Needs</b>			
OD program need	The modelling need to support the generation of rich semantic formats for reusability of the current OD publications	Modelling measures to enhance semantic capabilities of flat-text resources	Functionality design incorporating ontology-driven semantic enrichment processing
OD presentation need	The modelling need to support the implementation of standardised and unified publication structure and format	Modelling measures to create standard and unified structure and format	Functionality design incorporating a combination of resource reconciliation and RDFisation for semantic enrichment processing
OD representation (ontology) need	The modelling need to support the implementation of the consolidated indicator framework for the four development agendas	Modelling measures to address the semantic representation and integration need for the four development agendas	Functional design based on custom domain knowledge representation model (ontology) for the semantic integration of the Botswana integrated indicator framework

**Table 1** The summary of OD semantic use and publication needs.

## 2 RELATED WORKS

Most of the reviewed semantic value creation frameworks completely omit the semantic enrichment impact assessment processing. One example is the framework by Vrusias et al. (2007) for ontology-enriched semantic annotation of close circuit television video (CCTV). This framework links textual and visual semantics for video sequence annotations to enable the semantic transcoding of CCTV video footage. The processing of video annotation 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 then annotated with CCTV ontology constructs. Finally, summaries of video shots are produced in text format (Vrusias et al. 2007). The focus of the framework is on semantic annotation; however, it does not factor in the impact assessment of the semantic enrichment to determine its effectiveness.

The other framework with a complete exclusion of the impact assessment component is the semantic enrichment and fusion of multi-intelligence data framework (Hull et al. 2009). This framework pertains to semantic enrichment of data extracts from criminal reports with the military ontology metadata tags for the identification of vehicle theft events (Hull et al. 2009). This process is performed in three steps. First, the source data is transformed into extensible markup language (XML) to identify raw data structural features. Second, concepts and relationships found in those fields are extracted, and XML tags are generated. In addition, structural features are scanned for the identification and annotation of domain features. Third, XML tags are translated into web ontology language (OWL) representations, and the extracted features are aligned to the concepts and relationships specified in target ontologies. Although the focus of the framework is on semantic representation, the entire process omits the framework's assessment component. Therefore, it is unknown what specific capabilities are expected to be on the data as a result of the semantic enrichment, and how to determine the extent to which those capabilities contribute to the identification of vehicle theft events.

There are frameworks with partial inclusion of the semantic enrichment impact assessment processing. One of these frameworks is the fuzzy-ontology-enrichment-based framework for

semantic search, which integrates the domain ontology enrichment and the fuzzy ontology building in the information retrieval process (Baazaoui-Zghal & Ghezala 2014). This framework comprises three components: 1) the fuzzy information retrieval component, 2) an incremental ontology enrichment component, and 3) an ontology repository component (Baazaoui-Zghal & Ghezala 2014). The focus of this framework is semantic representation. Specifically, its focus is the semantic enrichment of data structure to support search operations. However, the authors do not specify the semantic capabilities that served as the basis for the determination of the framework's effectiveness and validity. Therefore, the assessment does not consider the impact of components' combination on the final output. The other framework with partial inclusion of the impact assessment component is the Baquara2 knowledge-based framework for semantic enrichment and analysis of movement data (Fileto et al. 2015). This framework addresses the lack of formal semantics for trajectories (movement) data. This framework comprises ontology constructs for the description of movement segments and annotation of movement data with corresponding objects and concepts described in ontology collections. This framework produces semantically enriched movement data compliant with an ontology that enables movement analysis queries based on application and domain specific knowledge. The components of this framework include the data structures and abstractions component, the Baquara2 ontology component, and the semantic enrichment and analysis of movement data (Fileto et al. 2015). The focus of this framework is on semantic annotation (enriches the meaning of data). Though the framework's assessments were query-based, there are no specifications of analysed capabilities.

In addition, there are frameworks that make mention of the semantic enrichment capabilities to be achieved without explicitly reflecting on how those capabilities are to be realised. For instance, the purpose of the framework for semantic enrichment of sensor data is the automation of semantic enrichment of sensor descriptions and measurements (Moraru & Mladenic 2012). This framework is based on the semantic enrichment process comprising a component of sensor descriptions and measurements, an ontology component, an enrichment component, a component of semantic repository of sensor data, and a component of data consumer (Moraru & Mladenic 2012). The framework aims to provide meaningful data to increase the effectiveness of sensor networks by enhancing the usability and accessibility of sensor data. Though this framework is focused on achieving both semantic annotation and representation, the lack of semantic capability specification in framework assessment creates a void in understanding the degree of understandability, usability, and accessibility effected through the semantic enrichment.

Overall, the reviewed frameworks are similar in that they factor in the semantic enrichment input, processing, and output components. The semantic enrichment process has an ontology component and adopts the linked open data (LOD) standard. The semantic enrichment either focuses on semantic annotation to improve semantic meaning, or on semantic representation to improve semantic structure, or on both. This can be done through the addition of a contextual (semantic annotation) or structural (semantic structure) layer or a combination of both layers to flat-text resources. There is limited to no consideration given to assessing the resulting semantic value. Determining the effectiveness of a framework in semantic enrichment involves the evaluation of the enriched output for semantic capabilities and semantic value-added to raw OD (the resulting semantic value). Therefore, this poses a need to augment semantic value creation frameworks with semantic value assessment mechanisms.

### 3 METHODOLOGY

For this study, we adopted the design science research (DSR) methodology for information systems, which guides the design of artefacts to solve specific problems (Peffer et al. 2007). The artefact produced in this research is a framework. According to Mattsson and Bosch (1999), a framework is a reusable system/application design in the form of abstract class components and their relationships, representing the foundational structure for the development of systems and applications.

On **Phase 1**, we focused on identifying framework needs, particularly those related to the preliminary semantic enrichment framework design. We identified the needs through flat-text resource reviews conducted according to our content analysis method, involving: 1) the search

for certain terms, concepts, keywords, and themes in flat-text resources; 2) the gathering of relevant domain knowledge related to the topic at hand; and 3) the inference and extraction of the framework needs pertaining to semantic publication and use needs.

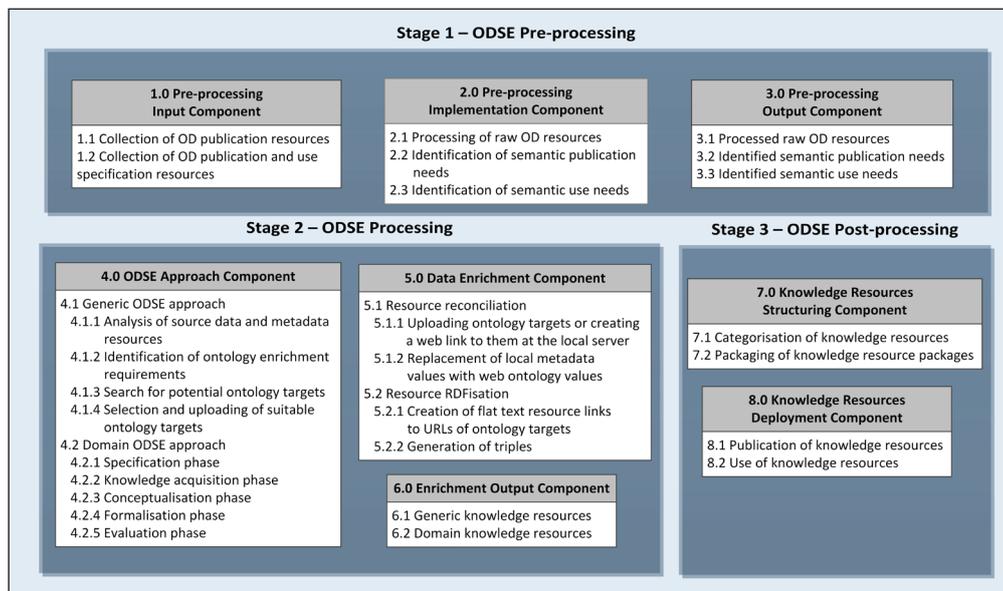
On **Phase 2**, we focused on framework design and development, based on the needs identified in Phase 1 and pertaining to the preliminary semantic enrichment framework artefact creation. We carried out this phase in three steps. The first step was the review of research literature. The second step was the identification of potential semantic enrichment processes, approaches, capabilities, techniques, technologies, and their relations. The third step was the generation of the conceptual semantic enrichment framework. We achieved this by grouping identified entities into components based on identified relationships and framework structuring, based on the combination of these components. The preliminary framework artefact comprised three stages contextualised to the semantic publication needs for the Botswana development agenda domain (Sebubi et al. 2019c).

On **Phase 3**, we focused on framework demonstration, which involved the implementation of framework components based on potential framework entities we identified. The first demonstrations pertained to the ontology approach component, which involved the choice of generic ontologies and the development of national performance indicator (NPI) domain web ontology (Sebubi et al. 2019b). The second demonstrations pertained to the semantic enrichment pre-processing, processing, and post-processing per ontology approach and comparisons of the effectiveness of the resulting semantic enrichment output.

On **Phase 4**, we focused on framework evaluation, pertained to the assessment of the effectiveness of the semantic enrichment framework.

## 4 ONTOLOGY-DRIVEN SEMANTIC ENRICHMENT FRAMEWORK

The ODSE framework is organised into eight processing components that model the transition from static and disintegrated flat-text resources, to dynamic and interlinked knowledge resources for semantic publication and use. These components span across the three implementation stages, namely, ODSE’s pre-processing, processing, and post-processing, as depicted in Figure 1.



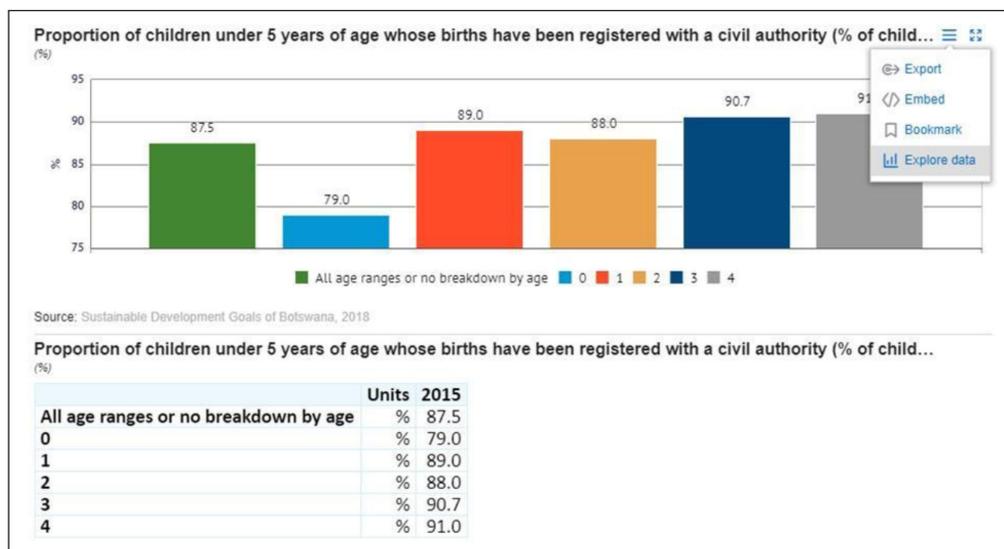
**Figure 1** The ontology-driven semantic enrichment (ODSE) framework.

### 4.1 STAGE 1: ODSE PRE-PROCESSING

This stage involves semantic enrichment input collection and preliminary processing. It has three sub-processes comprising the pre-processing input, implementation, and output components.

#### 4.1.1 ODSE Component 1.0 – the pre-processing input component

This component involves the gathering of semantic enrichment input comprising the source data and metadata resources. The first activity (Activity 1.1) pertains to the collection of OD publication resources, involving the identification and extraction of development agendas' raw OD resources to be enriched, specifically, the extraction of published data from secondary sources. For the Botswana case study demonstration, the published OD for the SDG agenda was extracted from the Statistics Botswana data portal. The set of raw OD comprised both the SDG indicator (SDGI) data and metadata in the form of performance values for the SDG agenda in raw (flat-text) format. For each of the 17 SDGs, the data was presented separately on the basis of the sustainable development goal indicator (SDGI). The metadata on SDGI attributes and mappings to related agendas were presented in raw form. These resources were presented at different levels of flat data structures. The SDGI data was presented in a diversity of HTML, spreadsheet, link (application programming interface snippet), semantic and PDF formats. The sample flat-text resource input shown in Figure 2 represents the basic dashboard view of the SDGI 16.9.1 data (proportion of children under five years of age whose births have been registered with a civil authority, by age).



**Figure 2** A sample open data access dashboard view (an extract from SDGI 16.9.1) (Statistics Botswana, 2021).

The second activity (Activity 1.2) of the pre-processing input component pertains to the collection of OD publication and use specification resources. The OD publication and use specification resources comprise relevant documentations on development agendas and OD programs. The process starts with the identification of potential sources of metadata for enrichment and identification of the motive for the enrichment so as to determine the semantic value to be added to the data. This involves the collection of documentations that are potential sources of OD and development agenda domain knowledge. These are mainly government strategic documents for OD and development agendas that constitute rich metadata sources for semantic enrichment (i.e., documentations on policies, strategic plans, OD programs, and development agenda initiatives). However, in most cases, these sources are available in diverse flat-text formats. They are also available on diverse platforms. Some ODs are published on the web and others in local data-management platforms. In the Botswana case study demonstration, the documentations were available in PDF and spreadsheet formats. These documentations comprised the NDP Agenda and Vision Agenda documentations, the Botswana OD Readiness Assessment Survey report, the SDGs National Baseline Indicator Framework Proposal, Botswana Domesticated SDGs, the Global SDGs Indicator List, and the Africa Agenda documentation. The raw OD resources for the sample SDGI 16.9.1 OD extract (Figure 2) can be accessed through the Input Resources link in Appendix A. We observed that, in most cases, raw input data was characterised by OD disparateness in terms of presentation formats, and OD heterogeneities in terms of access formats, platforms, and data definitions. This diversity makes it difficult to access, understand, interoperate, integrate, and make intelligible use of the data. This hampers semantic value creation capabilities and, consequently, knowledge

extraction capabilities. To address these limitations, the data has to undergo some processing described in detail in the next sections.

#### 4.1.2 ODSE Component 2.0 – the pre-processing implementation component

This component deals with the preliminary processing of raw OD resources and identification of semantic value creation needs, including the semantic publication and use needs. This component involves three activities, described as follows.

The first activity (Activity 2.1) is the pre-processing of raw OD resources, which involves data analysis, cleaning, implementation of some quality operations on the data, and the unification of the data. This allows to improve data quality and intelligibility capabilities to increase the value-adding potential of 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. The cleaning and transformation process involved the identification and correction of errors in the dataset and alignment of published data and metadata to the domesticated and global SDG agenda definitions. It also involved the integration of disparate data and the harmonisation of related data from different sources. This activity produces output formatted for semantic enrichment.

The second activity (Activity 2.2) is the identification of semantic publication needs. We extracted semantic publication needs from the OD program and development agenda documentations. The semantic publication needs were related 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 for the extraction of OD program goals, the deduction of OD presentation needs from the limitations of the current raw OD publications, and the identification of OD representation needs for ontology modelling.

The third activity (Activity 2.3) is the pre-processing implementation component, which pertains to the identification of semantic use needs. Semantic use needs are, in terms of OD use and reuse requirements, defined by the knowledge to be derived from the enriched output for further usage. We extracted these from the OD program and development agenda documentations. For the Botswana case study, identification of semantic use needs involved the assessment of the fourth pillar of Vision 2036, 'governance, peace and security', for OD-related support needs in the implementation of Vision 2036.

#### 4.1.3 ODSE Component 3.0 – the pre-processing output component

This component pertains to pre-processing implementation deliverables. The first output (Activity 3.1) is the preliminary processed raw OD resources. These resources represent the cleaned, transformed, and consolidated data. The pre-processed raw OD resources for the sample SDGI 16.9.1 OD extract ([Figure 2](#)) can be accessed through the Re-organised and Transformed Resources link in Appendix A. These resources provide input into the ODSE processing stage of minimally integrated, machine-readable data in the same publication formats and platforms for both published data and metadata sources. This enables semantic value creation capabilities and, consequently, knowledge extraction capabilities, making it easy to access, understand, interoperate, integrate, and make intelligible use of the data. For the Botswana case study demonstration, the preliminary processed raw OD resources comprised collections of domain knowledge on NPIs in flat-text formats. The second output is the OD semantic publication needs (Activity 3.2). For the Botswana case study demonstration, the main semantic publication needs were (1) the need for OD presentation formats, (2) the OD representation (ontology) need, and (3) the OD program need. These needs culminated in the three respective ODSE framework's needs to support (1) the implementation of the standardised and unified publication structure and format, (2) the implementation of the consolidated indicator framework for the four development agendas, and (3) the generation of rich semantic formats for reusability of current OD publications. The third output of the pre-processing output component (Activity 3.3) is the identified semantic use needs. For the Botswana case study demonstration, the main semantic use needs were the OD support needs for the Vision Agenda as shown in [Table 1](#).

## 4.2 STAGE 2: ODSE PROCESSING

This is the intermediate processing stage, which pertains to the actual implementation of the semantic value creation. We adopted the LOD technique 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: (1) the ODSE approach component, (2) the data enrichment component, and (3) the enrichment output component.

### 4.2.1 ODSE Component 4.0 – the ODSE approach component

This is the central processing component of the ODSE framework, the main driver of semantic value creation and assessment functions/operations. It involves the selection and implementation of semantic value creation and assessment approach that best fits specific semantic publication and use needs. The generic ODSE approach is based on the use of generic ontologies, whereas the domain ODSE approach is based on the use of custom domain ontologies.

The first activity of the ODSE approach component outlines the method for choosing applicable generic ontologies (Activity 4.1). The term ‘generic ontology’ in the context of this framework refers to top-level reference ontologies and any unrelated domain ontologies published as online resources. The method for selecting applicable generic ontologies involves four steps, described below.

**Step 1** consists of the analysis of source data and metadata resources for ontology search literals (Sub-activity 4.1.1), which involves identifying and assigning entity field names to data holding fields, as well as descriptions of data content. This is followed by refining entity field names to serve as keywords for online ontology search targets. Lastly, there is the analysis of entity field names to extract online search literals.

**Step 2** focuses on the identification of ontology enrichment requirements (Sub-activity 4.1.2), which involves the defining the objective of ontology enrichment needs to be addressed by ontology targets. For generic ontology targets in the Botswana case study, the objective was to create more global links to the NPI domain dataset for enhanced discoverability. NPI ontology (NPIOnto) is a custom domain ontology created for semantic representation of NPIs in the context of the Botswana development agenda domain (Sebubi et al. 2019b). With NPIOnto, the objective was to link more specific meaning to the dataset.

**Step 3** involves the search for potential ontology targets (Sub-activity 4.1.3), which is conducted on the basis of the literals identified in Step 1 and the ontology enrichment requirements identified in Step 2.

**Step 4** comprises the selection of suitable ontology targets (Sub-activity 4.1.4), which involves developing the selection criteria to serve as the basis for the selection of suitable ontology targets. In the Botswana case study demonstration, the selection criteria were mainly based on availability, coverage, quality, and value representation (semantic structure/syntax and meaning/semantics).

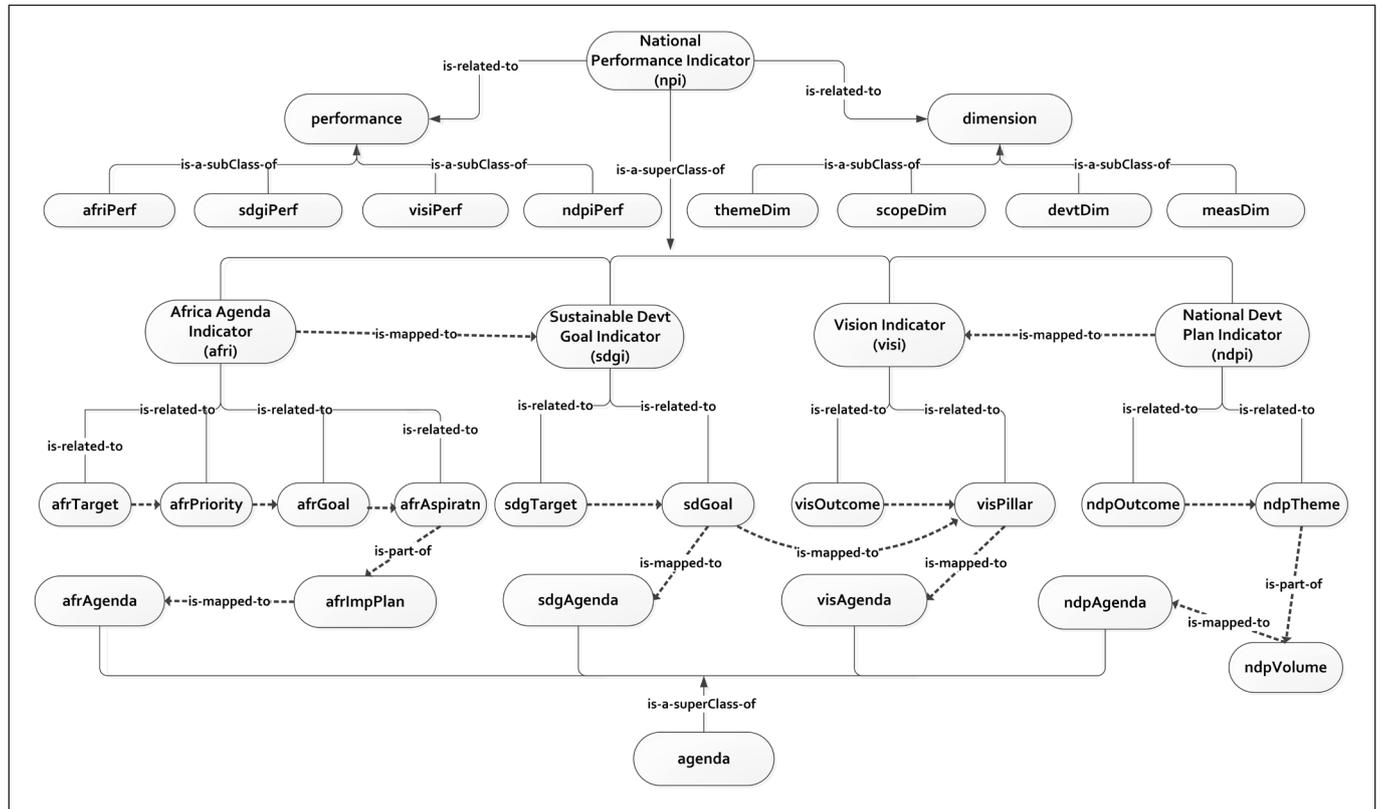
The second activity of the ODSE approach component outlines the method for the development of custom domain ontologies (Activity 4.2). This activity pertains to the development of a custom domain ontology, i.e., NPIOnto. The development of NPIOnto involved five steps, described below.

**Step 1** is the specification phase (Sub-activity 4.2.1), which involves the statement of NPIOnto purpose in terms of goal, scope, and intended use. 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 covered the Botswana integrated indicator framework comprising the Vision Agenda, the NDP Agenda, the domesticated SDG Agenda and the 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** is the knowledge acquisition phase (Sub-activity 4.2.2), which pertains to the identification of NPIOnto concepts, their characteristics, and relationships from source data and metadata resources gathered in Stage 1. The output of this stage comprised collections of development agenda domain knowledge on NPIs in spreadsheet format.

**Step 3** is the conceptualisation phase (Sub-activity 4.2.3), which involves the design and development of NPIOnto schema. The deliverables comprised the conceptual models for the four development agendas, and a consolidated conceptual model that integrates the four individual conceptual models on the basis of shared classes, object properties, and data properties. [Figure 3](#) presents the conceptual NPIOnto model.

**Figure 3** The conceptual NPIOnto model.



**Step 4** is the formalisation phase (Sub-activity 4.2.4), which involves codification of the consolidated knowledge model (schema) and the population of the knowledge base with instances.

**Step 5** is the evaluation phase (Sub-activity 4.2.5), which involves the assessment of the effectiveness of NPIOnto design and functionality. The schema effectiveness design evaluation was conducted with the use of knowledge model metrics and the knowledge base effectiveness design evaluation with the use of knowledge base metrics. We tested the functionality's effectiveness in the semantic enrichment demonstration.

#### 4.2.2 ODSE Component 5.0 – the data enrichment component

This component involves the conversion of flat-text resources (i.e., the output of Activity 3.1 – processed raw OD resources) into a more rich, flexible, and dynamic format for an increased semantic capabilities value-adding potential. This component has two activities, namely, resource reconciliation (Activity 5.1), and RDFisation (Activity 5.2). RDFisation is the generation of resource-description framework (RDF) structured data, in the form of subject-predicate-object triples. It creates knowledge resources through the addition of a semantic layer to flat-text resources, which constrains the resource metadata to standardised vocabularies through resource reconciliation and RDFisation. Resource reconciliation establishes potentials of OD accessibility and interoperability to web resources, while resource RDFisation creates the accessibility, understandability, integrability, and intelligibility capabilities through linkages to ontological constructs, mainly, classes and properties. Thus, 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. Regardless of the ODSE approach that is chosen, semantic enrichment constrains the metadata of enriched resources to relevant vocabularies.

The first activity of the data enrichment component is resource reconciliation (Activity 5.1). It determines the potential for link creations by assessing the compatibility of flat-text resources to related resources in the Web of Data. An atomised resource reconciliation process involves the online comparison of local metadata values to those in controlled vocabularies to determine the degree of data accessibility and interoperability. For the manual process, resource reconciliation is accomplished by downloading selected ontology targets (identified under Sub-activity 4.1.4) and uploading them to the local server environment or creating web link addresses to selected web ontology targets in the local server environment (Sub-activity 5.1.1). The resource reconciliation is followed by the replacement of the metadata for flat-text resources with matching values in related web repositories of selected ontology targets (Sub-activity 5.1.2). The replacement of metadata promotes a more universal understanding. In cases of no matching global values, the metadata is replaced with specific domain constructs from the custom domain ontology. In both cases, the reconciled uniform resource locators (URLs) provide more accurate metadata descriptions standardised to controlled vocabularies for ease of OD accessibility and interoperability.

The second activity of the data enrichment component, the resource RDFisation (Activity 5.2), generates the RDF representation of data. This activity adds 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. The ontology layer has complete semantic value creation features in its make-up structure. The resource RDFisation process represents the RDF-driven semantic integration, which implements the actual data mappings to controlled vocabularies. First, flat-text resource links to the URLs of selected ontology targets are created on the basis of the subject-predicate-object structure (Sub-activity 5.2.1). This is followed by the generation of triples (Sub-activity 5.2.2). The final product of the mappings is the triple structure that specifies links between the data in the form of the subject-predicate-object structure.

#### 4.2.3 ODSE Component 6.0 – the enrichment output component

This component pertains to the final output of the ODSE processing stage: knowledge resources (semantically enriched or actionable OD). For the generic ODSE processing, the output is the generic knowledge resources (Activity 6.1). A sample of the generic knowledge resource extract is depicted in [Figure 4](#).

**Figure 4** A sample of the generic knowledge resource (derived from the SDGI 16.9.1 OD extract).

```
<http://www.statsbots.org/bw/data/sdgi/goal16/2030_16.9.1> a dbr:Unique_identifier;
dbp:theme "peace, justice and strong institutions";
afo:AFX_0002801 "79"^^<http://www.w3.org/2001/XMLSchema#int>, "88"^^<http://www.w3.org/2001/XMLSchema#int>,
"89"^^<http://www.w3.org/2001/XMLSchema#int>, "90.7"^^<http://www.w3.org/2001/XMLSchema#int>,
"91"^^<http://www.w3.org/2001/XMLSchema#int>;
afo: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)";
dcterms:date "2015"^^<http://www.w3.org/2001/XMLSchema#int>;
sdgio: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";
sdgio:SDGIO_00000001 "By 2030, provide legal identity for all, including birth registration";
sdgio:SDGIO_00000002 "Proportion of children under 5 years of age whose births have been registered
with a civil authority, by age";
sdgio:SDGIO_00000003 "87.5"^^<http://www.w3.org/2001/XMLSchema#int>;
sdgio:SDGIO_00000074 "16.9";
sdgio:SDGIO_00000242 "16.9.1";
sio:SIO_000116 "Sustainable Development Goal 16";
sio:SIO_000221 "%";
rdfs:label "2030_16.9.1" .
```

[Figure 4](#) reflects the use of generic ontologies for standardisation of meaning. In comparison, the output for the domain ODSE processing approach is the domain knowledge resources (Activity 6.2). [Figure 5](#) demonstrates a sample of the domain knowledge resource.

```

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

```

The domain knowledge resource in [Figure 5](#) reflects the use of custom domain ontology (NPIOnto) for standardisation of meaning. These knowledge resources represent the unified presentation structure and the format for raw OD publications in machine-readable format. The generic and domain-specific knowledge resources for the sample SDGI 16.09.1 OD extract ([Figure 2](#)) can be accessed through the Individual Knowledge Resources link in Appendix A.

**Figure 5** A sample of the domain knowledge resource (derived from the SDGI 16.9.1 OD extract).

Overall, 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 comprise 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 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 enables 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.

### 4.3 STAGE 3: ODSE POST-PROCESSING

This final stage concludes the ODSE framework processing with knowledge resources structuring and deployment for publication and use.

#### 4.3.1 ODSE Component 7.0 – the knowledge resources structuring component

This component pertains to the final structuring of the knowledge resources, involving two activities. The first activity pertains to the categorisation of knowledge resources. This activity involves knowledge resources screening by functionality support needs (publication and use needs) for packaging. For the Botswana case demonstration, knowledge resources created from the SDGI data were generated separately for each SDG. 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 to differentiate between the products of generic ontology enrichment and NPIOnto enrichment for assessment of semantic value creation potential per ODSE approach. The second activity pertains to the packaging of knowledge resources. This activity involves the creation of both publication and use of knowledge resource packages through triples integration, which pertains to the consolidation of knowledge resources by combining

triples for ease of knowledge mining/harvesting/extraction operations. For the Botswana case study, the triples in individual knowledge resources were unified into the integrated generic and domain-knowledge resources. The generic and domain-specific knowledge resources for the sample SDGI 16.9.1 OD extract (Figure 2) can be accessed through the Integrated Knowledge Resources link in Appendix A.

### 4.3.2 ODSE Component 8.0 – the knowledge resources deployment component

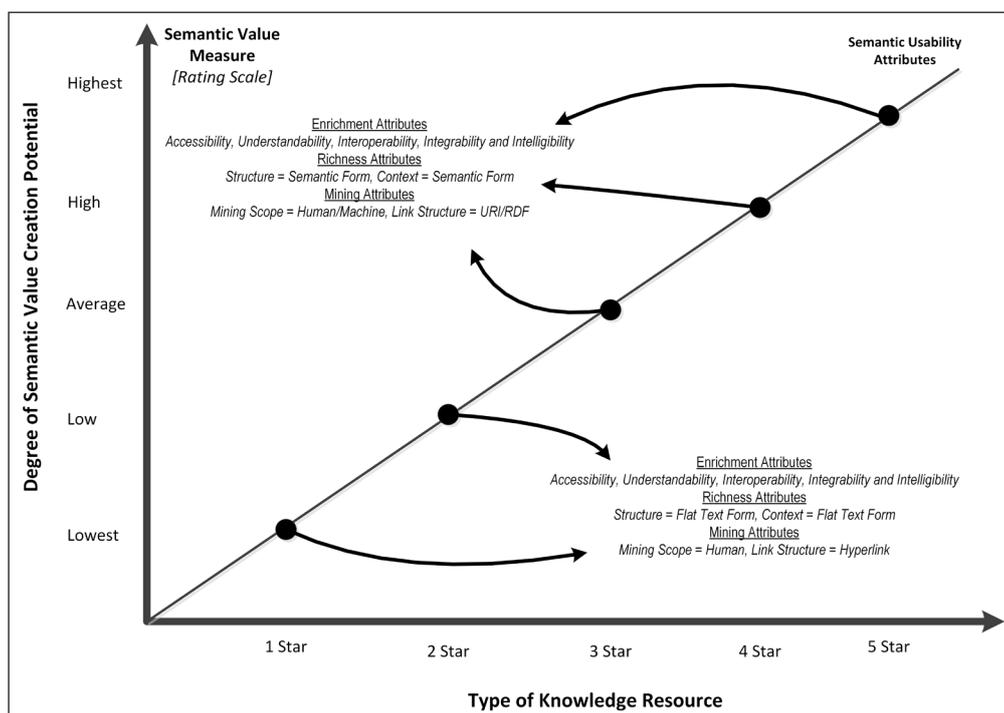
This component pertains to the publication and use of knowledge resources. While the two activities within this component use the simple protocol and RDF query language (SPARQL), each activity uses it for a different purpose. For Activity 8.1, SPARQL is used for publication, while for Activity 8.2 it is used for querying.

The first activity (Activity 8.1) involved in this component pertains to the publication of knowledge resources. This activity starts with the selection of the most suitable environments for sharing and harvesting knowledge resources. The selection criteria presume that environments must enable both human and machine accessibility, understandability, interoperability, integrability, and intelligibility. Knowledge resources can be uploaded into semantic publication libraries, query endpoints, semantic browsers, and inference engines. For the case study of Botswana, integrated knowledge resource packages were uploaded to the SPARQL endpoint for semantic value mining.

The second activity (Activity 8.2) pertains to the use of knowledge resources. This activity involves the deployment of knowledge resources for the use through knowledge mining tools and techniques. For the case study of Botswana, the semantic value was obtained from integrated knowledge resource packages through SPARQL queries.

## 4.4 THE SEMANTIC VALUATION COMPONENT

The ODSE framework addresses the gap of reviewed semantic value creation frameworks. These frameworks focus on adding semantic value to improve semantic capabilities with no provision to either determine the value-adding or value-added capability of the semantic-enrichment implementation. The ODSE framework is augmented with the semantic usability assessment model (Figure 6), a model for evaluation of OD publication and use effectiveness. This model provides a mechanism assessing the semantic value-adding and value-added potential to OD resources at each level of semantic value creation implementation in terms of enrichment, richness, and mining value. The semantic usability assessment model is derived from a combination of semantic usability attributes and Berners-Lee’s five-star deployment scheme for linked data (Berners-Lee 2010).



**Figure 6** The semantic usability assessment model based on the Berners-Lee’s five-star deployment scheme for linked data (2010) and modified by adding semantic enrichment effectiveness.

The model depicted in [Figure 6](#) is used for the assessment of the degree of semantic usability impact on semantic value (knowledge) creation potential of a given OD resource in terms of the knowledge resource enrichment, richness, and mining attributes. Semantic value is a qualitative measure of OD publication rating, defined in terms of the five-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. On this premise, the linked data publication rates also represent the linked data consumption rates. Semantic value is presented in relation to semantic value independent variable which pertains to the different types of OD publications, specifically, *type of knowledge resource*, expressed in terms of the Berners-Lee's five-star deployment scheme for linked data (one-, two-, three-, four-, and five-star OD publications). This five-star ranking is based on the semantic publication standard features pertaining to (1) open licensing, (2) machine readability, (3) open format, (4) URI linkages, and (5) linked data linkages ([Berners-Lee 2010](#)).

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, which are 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, measured by the resource's degree of accessibility, understandability, interoperability, integrability, and intelligibility demonstration. 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.

Below are four examples of potential applications drawn from the semantic usability graphical model in [Figure 6](#).

1. The knowledge resource rating application (Application 1) is 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.
2. The second potential application has two subtypes. Both subtypes are the resource selection applications drawn from the relations between the knowledge resource and semantic usability attributes. The difference between the two subtypes is the last component (the knowledge creation variables and the semantic attribute measure variables, respectively) The first subtype (Application 2.1) pertains to the selection of knowledge resources with appropriate semantic usability attributes for a given target of the semantic value creation rating. The second subtype (Application 2.2) pertains to the selection of knowledge resources with appropriate semantic usability attributes for a given target of semantic usability effectiveness rating.
3. The resource evaluation application (Application 3) is 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 use effectiveness) for a given type of knowledge resource.
4. The resource evaluation application is drawn from the relation between the knowledge creation dependent variable and the knowledge resource variable. This application involves the evaluation of the degree of the semantic value 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.

In terms of the semantic valuation applications in the ODSE framework (Figure 1), the semantic usability assessment model is one of the candidate tools for assessing ODSE pre-processing input's publication and use effectiveness based on resources' semantic capabilities (Activity 2.1). It is also applicable in assessing the publication and use effectiveness of pre-processed OD resources (Activity 3.1). It is also the most viable tool for assessing the publication and use effectiveness of knowledge resources (Component 6.0). Finally, it is applicable in implementing the knowledge resources structuring component, specifically in the process of sorting knowledge resources for publication and use packaging (Activity 7.1).

## 5 FRAMEWORK EVALUATION

The framework evaluation of ODSE involved assessing the degree of framework effectiveness in the semantic value creation and usability potential using the semantic usability assessment model, shown in Figure 6. The assessment of ODSE framework's effectiveness was performed by assigning the 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.

The effectiveness assessment for Stage 1 (ODSE pre-processing), pertaining to raw and processed OD resources (Activities 2.1 and 3.1), revealed that published SDGI data in tabular (HTML) and image formats, and metadata publications in PDF format are at the one-star level, which translates to the lowest publication and use effectiveness. The data is downloadable in spreadsheet and XML formats, thus being upgraded to the two-star (spreadsheet) and three-star (XML) levels, offering minimal and moderate levels of publication and use effectiveness, respectively. All flat-text resources belong to the category of one-, two-, and three-star OD publications, with differing degrees ranging from the minimum to moderate formation of semantic value determination variables and 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 link structure. Overall, flat-text resources offer the minimal to moderate degrees of the semantic value creation potential and, consequently, minimal to moderate degrees of semantic usability effectiveness. For the maximum semantic enrichment, richness, mining, and consequently, reuse potential, the data must pass the level of machine readability (three-star level-XML) and attain levels of machine processability (four- and five-star levels).

The effectiveness assessment for Stage 2, (ODSE processing), addressed the challenge of OD publications ineffectiveness by standardising and upgrading the ODSE pre-processed output through the LOD-driven semantic enrichment. The SDGI data and corresponding metadata were reconciled and interlinked to corresponding resources within the Web of Data and the Botswana integrated indicator framework.

The assessment model was applied to the deliverables of semantic enrichment, knowledge resources (OD resources enriched with corresponding metadata and linkages across the Web of Data) as reflected in the enrichment output component (Component 6.0). 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 the three-star (moderate) level to five-star OD publications with differing degrees of maximum formation of semantic value determination variables. In terms of richness attributes, knowledge resources offer the highest degree of mining capabilities due to the 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 (highest) due to differences in the mining scope and structure. Overall, knowledge resources offer the highest degrees of the semantic value creation potential and, consequently, the highest degrees of semantic usability effectiveness.

The effectiveness assessment for Stage 3 (ODSE post-processing), pertained to OD packages of integrated knowledge resources (Activity 7.1). These packages were further mined through SPARQL queries for semantic capabilities value added. The knowledge resources evaluation cases for the sample SDGI 16.9.1 OD extract (Figure 2) can be accessed through the Integrated Knowledge Resources Evaluation link in Appendix A.

Firstly, the accessibility evaluation 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).

Secondly, the understandability evaluation has proved that SDGI performance, enriched with SDGI metadata such as attribute, measurement, data, and relationship descriptive semantics, provides a holistic view. In addition to the data descriptive semantics, data disaggregation also 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. A designated domain-specific vocabulary allows for specific metadata definition in terms of attribute, measurement, data, and relationship descriptive semantics, as well as specific value disaggregation. Therefore, designated domain-specific knowledge resources, with their in-depth semantic enrichment value, offer maximal clarity when compared to the other three types of OD publications, namely, flat-text format, non-enriched semantic formats (XML), and generic knowledge resources.

Thirdly, the intelligibility evaluation has proved that both ontology approaches extend SDGI performance intelligibility with the definition of global resource linkages which provide additional dimensions of strategic significance, such as the measurement type and scope, theme, and development dimensions.

Lastly, the integrability and interoperability evaluations have proved that the semantic data enriched with links to web vocabularies enable data sharing to support the implementation of the Botswana integrated indicator framework.

## 6 STUDY LIMITATIONS AND FUTURE WORK

The present research encountered several limitations. First, 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 the record identification. In addition, this research only dealt with published OD.

Additionally, there may be some other unpublished data lying dormant in the hands of stakeholders. The research scope did not include quantitative evaluations; however, the proposed framework could be extended to include them. Lastly, the research scope was limited to the semantic structure and context of knowledge resources.

The duration of the NDP 11 plan, one of the main documentations of this research, was six years (April 2017-March 2023). However, the implementation of NPD 12 was deferred, and, in the interim, a two-year transitional national development plan (TNDP) was developed for the financial years 2023/24 to 2024/25. Once the new relevant data is available from the Government of Botswana and the Statistics Botswana data portal, this will be studied by applying the ODSE framework and methodology proposed in this paper.

## 7 CONCLUSION

The proposed ODSE framework provides the mechanism for facilitation and evaluation of the degree of knowledge resources' semantic value creation capability. The framework is founded on the semantic value creation approach with potential for semantic value creation maximisation and supported by the semantic valuation. The ODSE framework is designed for reusability in all types of flat-text data transformations to semantic formats in both the development agenda and generic domains. However, the proposed transition to Web of Data resource publications is

not a replacement, but rather an augmentation of Web of Documents publications to support both human and machine use. In addition, we recommend the establishment of a public knowledge resource pool for integration of government, economic, environment, and social sectors to drive open innovation, business, and entrepreneurship creations/recreations. For the public knowledge resource pool to serve as an unlimited engine for powering and sustaining the needs of the evolving knowledge-driven economy and the knowledge-based society, we further recommend that the publication knowledge base be built on the foundation of domain specific knowledge models (ontologies), like the NPIOnto. This would support the integration of both historical and current OD publications for support of strategic decision making and planning in the development agenda domain and any other domain of application.

## DATA ACCESSIBILITY STATEMENT

All data used and produced by this research is open, and, therefore, can be freely used, reused, and redistributed by anyone. Resources (mostly, samples) that form input to research as well as the output resources at the end of each of the three processing stages for the ODSE framework are uploaded to Figshare. Those are accessible through links for each resource category provided in Appendix A.

## APPENDIX A: LINKS TO ODSE RESOURCES

FRAMEWORK PROCESS	RESOURCE TYPE	RESOURCE WEB LINK	RESOURCE WEB LINK URL
ODSE Pre-processing (Stage 1) Resources	<i>Input Resources (Raw OD Resources)</i>	ODSE Framework Package 1 – Sample Raw OD for SDGI 16.9.1	<a href="https://figshare.com/s/8382f76f471144242070">https://figshare.com/s/8382f76f471144242070</a>
		ODSE Framework Package 2 – Raw OD Resources for SDG Agenda	<a href="https://figshare.com/s/177199fb05051428a987">https://figshare.com/s/177199fb05051428a987</a>
		ODSE Framework Package 3 – Raw OD Resources for Vision Agenda	<a href="https://figshare.com/s/ef5328b22072375dc733">https://figshare.com/s/ef5328b22072375dc733</a>
		ODSE Framework Package 4 – Raw OD Resources for Africa Agenda	<a href="https://figshare.com/s/5570670048ebb56da4eb">https://figshare.com/s/5570670048ebb56da4eb</a>
	<i>Re-organised and Transformed Resources</i>	ODSE Framework Package 5 – Pre-processed OD Resources for all Agendas	<a href="https://figshare.com/s/4ec6350b3bc89b2d118d">https://figshare.com/s/4ec6350b3bc89b2d118d</a>
ODSE Processing (Stage 2) Resources	<i>Individual Knowledge Resources</i>	National Performance Indicator Ontology (NPIOnto) ODSE Framework Package 6 – Sample Knowledge Resources for SDG16	<a href="https://figshare.com/articles/dataset/National_Performance_Indicator_Ontology_NPIOnto_/20318562">https://figshare.com/articles/dataset/National_Performance_Indicator_Ontology_NPIOnto_/20318562</a> <a href="https://figshare.com/s/0c16ae71055e1794dfe8">https://figshare.com/s/0c16ae71055e1794dfe8</a>
ODSE Post-processing (Stage 3) Resources	<i>Integrated Knowledge Resources</i>	ODSE Framework Package 7 – Integrated Knowledge Resources	<a href="https://figshare.com/s/9b2ced2171da21eb7262">https://figshare.com/s/9b2ced2171da21eb7262</a>
	<i>Integrated Knowledge Resources Evaluation</i>	ODSE Framework Package 8 – Integrated Knowledge Resources Evaluation Cases	<a href="https://figshare.com/s/0ba7ee666d1ee181a2c3">https://figshare.com/s/0ba7ee666d1ee181a2c3</a>

## FUNDING INFORMATION

PhD research by Dr Oarabile Sebubi was partially funded by Botswana International University of Science and Technology under the postgraduate research initiation grant programme (grant number S00120).

## COMPETING INTERESTS

The authors have no competing interests to declare.

Dr Oarabile Sebubi was a PhD student in Botswana International University of Science and Technology, Botswana. She wrote and edited this paper based on the findings of her completed PhD research.

Prof. Irina Zlotnikova was a principal PhD supervisor of Dr Oarabile Sebubi and contributed to writing and editing the paper.

Dr Hlomani Hlomani was a co-supervisor of Dr Oarabile Sebubi and contributed to writing and editing the paper.

## AUTHOR AFFILIATIONS

**Oarabile Sebubi**  [orcid.org/0000-0001-8816-2722](https://orcid.org/0000-0001-8816-2722)

Department of Computer Science and Information Systems, Botswana International University of Science and Technology, Palapye, Botswana (BW)

**Irina Zlotnikova**  [orcid.org/0000-0002-1865-8586](https://orcid.org/0000-0002-1865-8586)

Department of Computer Science and Information Systems, Botswana International University of Science and Technology, Palapye, Botswana (BW)

**Hlomani Hlomani**  [orcid.org/0000-0002-6961-5906](https://orcid.org/0000-0002-6961-5906)

Department of Computer Science and Information Systems, Botswana International University of Science and Technology, Palapye, Botswana (BW)

## REFERENCES

- African Union.** 2017. *Agenda 2063: FTYIP 2013–2023 Core Indicators Profile Handbook*. Addis Ababa, Ethiopia: African Union.
- Baazaoui-Zghal, H and Ghezala, HB.** 2014. Fuzzy-Ontology-Enrichment-based Framework for semantic search. In: *Web Information Systems and Technologies. The 10th International Conference*. Barcelona, Spain, 123–130. DOI: <https://doi.org/10.5220/0004923801230130>
- Berners-Lee, T.** 2010. *Linked data: Design issues*. Available at <http://www.w3.org/DesignIssues/LinkedData.html> [Last accessed 4 October 2023].
- D4D.net.** 2022. *Measuring the state of data for the public good. Global Data Barometer*. Available at <https://globaldatabarometer.org> [Last assessed 4 October 2023].
- Digital Science.** 2022. *The state of open data 2022, October 2022*. Available at [https://digitalscience.figshare.com/articles/report/The\\_State\\_of\\_Open\\_Data\\_2022/21276984](https://digitalscience.figshare.com/articles/report/The_State_of_Open_Data_2022/21276984) [Last accessed 4 October 2023].
- Fileto, R, May, C, Renso, C, Pelekis, N, Klein, D and Theodoridis, Y.** 2015. The Baquara knowledge-based framework for semantic enrichment and analysis of movement data. *Data and Knowledge Engineering*, 98: 104–122. DOI: <https://doi.org/10.1016/j.datak.2015.07.010>
- Godhwani, G and Kuzev, P.** 2023. *Open data as backbone of digital public infrastructure: prioritising high-value datasets for effective governance, July 2023*. Available at <https://t20ind.org/research/open-data-as-backbone-of-digital-public-infrastructure/> [Last accessed 4 October 2023].
- Government of Botswana.** 2017. *National Development Plan 11 Volume 1 (April 2017–March 2023)*. Gaborone, Botswana: Government of Botswana.
- Hull, RD, Jenkins, D and McCutchen, A.** 2009. *Semantic Enrichment and Fusion of Multi-Intelligence Data [White Paper]*. Melbourne, Australia: Modus Operandi, Inc.
- Mattsson, M and Bosch, J.** 1999. Characterising Stability in Evolving Frameworks. In: *Technology of Object-Oriented Languages and Systems, The 29th International Conference*, Nancy, France, pp. 118–130. DOI: <https://doi.org/10.1109/TOOLS.1999.779005>
- Moraru, A and Mladenic, D.** 2012. A Framework for semantic enrichment of sensor data. *Journal of Computing and Information Technology*, 10(3): 167–173. DOI: <https://doi.org/10.2498/cit.1002093>
- Peppers, K, Tuunanen, T, Rothenberger, AM and Chatterjee, S.** 2007. A design science research methodology. *Journal of Management Information Systems*, 24(3): 45–78. DOI: <https://doi.org/10.2753/MIS0742-1222240302>
- Sebubi, O, Zlotnikova, I and Hlomani, H.** 2019a. open data for sustainable development on a knowledge-based economy: The case of Botswana. *Data Science Journal*, 1–6. DOI: <https://doi.org/10.5334/dsj-2019-060>
- Sebubi, O, Zlotnikova, I and Hlomani, H.** 2019b. A Lightweight Version of National Performance Indicator Ontology (NPIonto). In: *NextComp 2019, Mauritius*, 21 September 2019, pp. 57–62. DOI: <https://doi.org/10.1109/NEXTCOMP.2019.8883619>

- Sebubi, O, Zlotnikova, I and Hlomani, H.** 2019c. Ontology-driven Open Data Enrichment Framework for Value Maximisation: The Case of National Performance Indicators for Botswana. In: *IST-Africa 2019*, Nairobi, Kenya, 8 May 2019, 1–8. DOI: <https://doi.org/10.23919/ISTAFRICA.2019.8764881>
- Seljan, S, Viličić, M, Nevistić, Z, Dedić, L, Grubišić, M, Cibilić, I, Kević, K, van Loenen, B, Welle Donker, F and Alexopoulos C.** 2022. Open Data as a Condition for Smart Application Development: Assessing Access to Hospitals in Croatian Cities. *Sustainability*, 14(19): 12014. DOI: <https://doi.org/10.3390/su141912014>
- Statistics Botswana.** 2018a. *Botswana Domesticated Sustainable Development Goals Indicators*. Gaborone, Botswana: Statistics Botswana.
- Statistics Botswana.** 2018b. *Sustainable Development Goals (SDGs) National Baseline Indicator Framework*. Gaborone, Botswana: Statistics Botswana.
- Statistics Botswana.** 2021. *Statistics Botswana Data Portal*. Retrieved from Statistics Botswana: <https://statsbots.org/bw/data-portal>.
- Vision 2036 Presidential Task Team.** 2016. *Vision 2036: Achieving prosperity for all*. Gaborone: Botswana Government.
- Vrusias, B, Makris, D, Renno, JP, Newbold, N, Ahmad, K and Jones, G.** 2007. A Framework for Ontology Enriched Semantic Annotation of CCTV Video. In: *Image Analysis for Multimedia Interactive Services, WIAMIS '07*. DOI: <https://doi.org/10.1109/WIAMIS.2007.4>
- Wilkinson, M, Dumontier, M, Aalbersberg, I, et al.** 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3: 160018. DOI: <https://doi.org/10.1038/sdata.2016.18>
- World Bank Group.** 2015. *Botswana ODRA Report*. Gaborone, Botswana: Botswana Government.

#### TO CITE THIS ARTICLE:

Sebubi, O, Zlotnikova, I and Hlomani, H. 2023. Ontology-Driven Semantic Enrichment Framework for Open Data Value Creation. *Data Science Journal*, 22: 40, pp. 1–18. DOI: <https://doi.org/10.5334/dsj-2023-040>

**Submitted:** 05 May 2023

**Accepted:** 06 October 2023

**Published:** 27 October 2023

#### COPYRIGHT:

© 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

*Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.