ISSN 2278-3091
Volume 9, No.4, July – August 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse216942020.pdf
https://doi.org/10.30534/ijatcse/2020/216942020

Towards a knowledge graph for open healthcare data

Kawtar YOUNSI DAHBI[1], Hind LAMHARHAR [2], Dalila CHIADMI[3]
Mohammadia School of Engineers (EMI), Mohammed V University in Rabat, Morocco,
younsidahbi.kawtar@gmail.com
[2]Faculty of dental medicine of Rabat, Mohammed V University in Rabat, Morocco, hd.lamharhar@gmail.com
[3] Mohammadia School of Engineers (EMI), Mohammed V University in Rabat, Morocco, Chiadmi@emi.ac.ma

ABSTRACT

Open Government data present valuable knowledge that supports innovation and value creation. Despite the efforts deployed by different countries to open government data, their usage and exploitation remain limited due to several challenges such as data discoverability, processability, and integration. Used by major search engines such as Google, Knowledge graphs have emerged as a potential supportive technology that can handle the OGD usage challengeS. In this, work we propose to construct the KG4OGD, a knowledge graph that drives the transformation of open government raw data to valuable knowledge. To support the construction of the KG4OGD, we propose a knowledge representation model for both OGD Metadata and content as a schema model for the KG4OGD. The model enhances the descriptive and contextual dataset's background and aims to improve the KG coherence and consistency. We implemented our model for the public healthcare domain as a high interest has arisen toward this domain due to the international pandemic context.

Key words: Open Government Data, Knowledge graphs, Data integration, Ontologies, Public healthcare
.

1. INTRODUCTION

Public administrations produce and collect a wide range of data to deliver expected services and perform their functions [1]. Opening up this data has a significant impact on fostering innovation, improving transparency, public accountability, collaboration, and improving citizens 'quality of life [1], [1]. However, despite the proliferation of Open Government Data (OGD) initiatives, the usage of this data remains limited by numerous challenges that hinder the expected value creation. OGD Datasets are provided across diverse and multiple data portals with limited metadata [3], [4] which hinder data discoverability [5], [6]. Data is not provided in a machine-readable format, it is available in structured, semi-structured and non-structured format with lack of semantics which makes it complex and time-consuming to handle and not easy to use [2], [5], [6]. Furthermore, data provided by a single dataset are usually insufficient to respond to the user request, it is necessary to combine data that may be available across multiple disparate and heterogeneous datasets to provide the user with the desired information [7]. Integrating data from several sources is hampered by the heterogeneous data formats, schemas, and semantic models [7], [24]. To tackle these challenges, many previous works have explored the usage of knowledge graphs (KG) for diverse domains. [8] – [12].KG can be defined as a graph of data intended to accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent relations between these entities [12]. Knowledge graphs have become prevalent to be one of the most efficient and effective knowledge integration approaches [13], [14]. They mine information from structured, semi-structured, and unstructured data disparate and heterogeneous data sources [14], [15] and leverage them to knowledge represented as a graph. For their potential, knowledge graphs are used by major search engines such as Google and Bing.

In our work, we propose an approach to drive the transformation of open government raw data into a Knowledge graph that we call the KG4OGD. The KG4OGD will enable us to extract valuable knowledge related to the government data, infer new insights, and enable a large plethora of smart applications. These applications cover for example advanced semantic search, question answering systems, knowledge inference, and data recommendation. The proposed KG4OGD enables the representation of both the OGD dataset's content and metadata. This is important to promote the efficiency of data discovery and offer contextual background knowledge for data. To support the construction of the KG4OGD, our approach proposes a knowledge representation model based on ontologies as a schema for the KG4OGD. Constructing the KG4OGD based on a semantically enhanced schema model aims to improve its internal coherence, consistency, and support further knowledge extraction, reasoning, and inference. The proposed knowledge representation model is based on two main ontologies: First, *the GovDataset ontology* which is an upper-level ontology that represents the OGD metadata level. It enhances the semantic description of OGD datasets with descriptive and contextual features. Second, we propose the usage of *the GovDomain ontology*, a domain-specific ontology that incorporates main domains concepts, entities, and relations. It represents the OGD content level and aims to provide a knowledge representation formalism to enhance the semantic description of OGD data content

Our approach is generic and domain-independent. For simulation, we implement the proposed approach for the Public healthcare domain in the Moroccan context. In fact, the evolution of the COVID-19 pandemic has increased OGD users (citizens, social activists, experts, and academics) interest to public healthcare data. Therefore, we selected this domain as a use case for our approach.

The rest of this paper is structured as follow, in section 2, we present the background of our research, in section 3 we present related works and compare them according to we present our approach, in section 5, we present our use case related to the public healthcare in the Moroccan context, we finally conclude and present our future works.

2. BACKGROUND

2.1 Open Government Data: Definition, benefits, and challenges

Open Government Data can be defined as any data and information produced or collected by public administrations that can be freely used, re-used, and distributed for any purpose by anyone [2]. According to the OGD principles, Government data is considered open if it fulfills the following OGD main requirements [17]: Complete, primary, timely, accessible, machine-processable, non-discriminatory, non-proprietary, and license-free. As part of OGD national initiatives that drive the publication of OGD, OGD datasets are published on national portals with corresponding metadata. These datasets are related to various domains such as public sector budgeting and performance levels, business, education, health, statistics, and traffic [1]. Their publication aims to improve transparency accountability and value creation, foster innovation and economic development, and improve citizens' life's quality [1], [2]. However, the usage of OGD released datasets remains limited by several challenges. OGD portals provide only limited search mechanisms [3] and insufficient metadata to document available datasets [4]. This hinders data discoverability and makes it difficult and laborious for users to find relevant information tailored to their needs [5], [6]. Moreover, OGD datasets are not machine-processable, they lack semantic information [2] ,[5] ,[6] and presents heterogeneity in data formats, schemas, and models [7] which hamper data processability and integration.

2.2 Knowledge graphs

Since 2012 several definitions have been proposed for knowledge graphs. According to [18], a knowledge graph mainly describes real-world entities and their interrelations, organized in a graph. It defines possible classes and relations of entities in a schema, allows for interrelating arbitrary entities with each other and covers various topical domains. It is presented in the form of a semantic graph consisting of vertices (or nodes) and edges. The vertices represent concepts or entities. An entity is a physical object in the real world The edges represent the semantic relationships between concepts or entities [14]. Knowledge graphs can mine Knowledge from either structured, semi-structured [18] [19] or non-structured data sources [20] [21]. Extracted data is published in the KG according to a specific schema which improves their quality and consistency [20] [19] [22] or without a schema (schema-less Knowledge graphs). For their storage, several Knowledge graphs use the RDF standard [20] [23]. Knowledge graphs can be classified into two main categories: The first category concerns general-purpose knowledge graphs such as YAGO [20], DBpedia [19], Wikidata [22] and NELL [23] The Knowledge VAULT [21] and Microsoft's Satori. The second category is domain-specific Knowledge graphs that cover a particular domain of Knowledge. Knowledge graphs have been used for a variety of tasks, including question answering, relationship prediction, and searching for similar items [14] [18].

3. RELATED WORKS

In this section, we present related works that deal with the construction of a Knowledge graph based on Open Government data, in [25], authors propose a data platform that aims to enhance OGD discoverability. The platform automatically collects OGD metadata held by different government agencies. Metadata is then standardized according to Standard Terminology Dictionary and techniques for distance calculation are applied to construct the Knowledge graph. The approach was applied to the Korean OGD portal. In [26], the authors propose an approach to construct a KG based on OGD metadata to represent datasets relatedness. The KG Nodes represent either datasets, themes, or entities. The edges represent datasets relatedness and are constructed based on OGD Metadata using the SOM Algorithm. As a use case for the KG, authors propose the dataset's recommendations Authors in [27] proposed the construction of a knowledge graph related to the KADASTER Domain, datasets from three open base registers were selected to build the KG and semantic links were identified across datasets. For graph construction, the authors focus on the linking of datasets already published as Linked Data. They propose three use cases to show the KKG interest: an improved data browsing, multicriteria analysis for urban planning, and the development of local aware chatbots. In [28], the authors propose the construction of a city knowledge graph for Zaragoza's city to represent OGD published dataset. The construction approach covers specification modeling, semantic lifting, and publication of OGD using several general and specific vocabularies. Authors in [29] propose the construction of an administrative district-based KG that links administrative district data with relevant data like hospitals, schools, and traffic. They present a knowledge model for administrative districts to support the construction of the KG. In [30], authors propose the construction of a KG for the public procurement domain, the proposed approach covers semantic modeling of data based on two domain ontologies and their semantic lifting. Authors [31] propose the construction of a domain-specific KG related to the disaster management domain. The KG construction approaches combine both semantic lifting based on a domain ontology and KG embedding techniques

3.1 Discussion

All previous projects have demonstrated the interest of exploring knowledge graphs in the context of Open Government Data, we, further, compare then according to two main dimensions: KG scope and granularity and KG construction approach. Table 1 presents a summary of the comparative analysis

-*Knowledge graph scope and granularity*: The proposed KG can be classified into two main categories: general-purpose [25], [28], [26]and domain-specific KG [27], [29], [30], [31]. They either cover the representation of OGD Metadata [25], [26] or the representation of OGD Data (Dataset's content). Also, only descriptive and structural metadata is represented in [25], [26].

-*Knowledge graph construction approach:* Previous works propose two main KG construction approaches. First, the Similarity calculation based construction approach [25], [26] proposes NLP, features extraction, and distance calculation methods for entity and relation extraction. Features were mainly syntactical, [25] introduces a dictionary for terms standardizations. Second, Semantic lifting based construction approaches which focus on the semantic lifting of OGD. These works propose for the semantic interpretation and transformation of OGD datasets based on different knowledge representation formalisms like the usage of ontologies [30], [31] the usage of vocabularies [27], [29], the usage of several vocabularies [28].

Table 1: Overview of related works

	KG Scope	KG granularity	Knowledge graph construction	
			Entity and relation	Semantic representation model
[25]	General-purpose	OGD Metadata	NLP Features ' extraction Similarity Calculation (Jaccard distance)	
[29]	Domain specific	OGD Data	Semantic lifting	Vocabulary
[27]	Domain specific	OGD Data	Semantic lifting	Vocabulary
[28]	General-purpose	OGD Data	Semantic lifting	9 general and 16 specific vocabularies
[26]	General-purpose	OGD Metadata	-NLP Features ' extraction Similarity Calculation SOM Algorithm	
[30]	Domain specific Public procurement domain	OGD Data	Semantic lifting	Domain specific Ontology
[31]	Domain specific/Disaster domain	OGD Data	Semantic lifting + KG embedding	Domain specific Ontology

3.2 Remaining issues

Our comparative analysis of previous works depicts the following limitations and remaining issues:

- Focusing only on OGD metadata representation in the KG limited its potential, the OGD value creation relies on the OGD Content usage, exploration, and analysis
- Representing only descriptive and structural OGD metadata limits the background knowledge about datasets and thus hinder data discoverability, adding context-related and government-specific metadata can improve user's discovery of relevant content
- Representing metadata with only syntactical features limits its discovery, there is a need for a semantic enhanced model to capture OGD Metadata content
- For the KG construction, relying on vocabularies to semantically enhance OGD content can limit the KG potential as vocabularies lack of expressiveness and can’t allow further reasoning and inference
- The usage of several knowledge representation models can leads to a semantic heterogeneity that may hinder data integration and further responding to user’s query

4. THE PROPOSED APPROACH

To overcome the aforementioned limitations and bridge the gap of previous works, our approach proposed the construction of an OGD KG that we call the OGD-KG based on both the metadata and data level to better support data discoverability. To support the construction of the OGD-KG a knowledge representation model is proposed to structure, enrich, and semantically enhance OGD representation in the KG. Using a predefined knowledge representation model for KG construction improves its coherence, consistency, and further querying, reasoning, and inference capabilities. We chose ontologies as a knowledge representation formalism due to their powerful expressiveness.

We present in this section, our OGD knowledge representation model and its ontologies and the OGD-KG definition and formalism

4.1 The OGD knowledge representation model

Our proposed knowledge representation model presents OGD related knowledge according to two main dimensions:

·*Knowledge related to the OGD dataset's metadata:* This dimension focuses on the OGD metadata level. It aims to provide a rich background knowledge to describe OGD datasets and their context. It responds to the following question What is the dataset's about and what is its context?

·*Knowledge related to the OGD dataset's content:* This dimension focuses on the OGD Data level (dataset's content).it aims to provide rich background knowledge for concepts, entities, and relations represented in the dataset. Does it respond to the following question? What concepts, entities, and relations are represented in the OGD dataset?

We describe below both dimensions:

A-OGD Metadata knowledge representation model

OGD metadata knowledge representation model aims to provide rich background knowledge about OGD datasets to improve data discovery. It presents a model to represent both descriptive and contextual metadata about OGD datasets.

Enhancing descriptive metadata

Several international vocabularies and standards have already proposed a set of descriptive metadata for OGD datasets. As an example, the DCAT Vocabulary proposes properties such as title, description, and keywords. Our proposed knowledge representation model improves these metadata by providing government-specific properties like government data category and government data topic. We estimate that adding such metadata will enhance the semantic description of OGD datasets, especially that several international indexes have already identified keys OGD categories and topics.

Enhancing contextual metadata

Enriching OGD Datasets with contextual knowledge is important to support user's access to data relevant to their context and information needs and interests. Our knowledge representation model proposes a representation for the following OGD Dataset's related context's type: The temporal context, the geographic context, the usage context, the organizational context, and the legal context.

The temporal context aims to provide information including the temporal coverage and time validity for OGD datasets**.** ***The geographic context*** aims to provide information related to the geographic coverage of OGD datasets. Both temporal and geographic context has a significant impact on improving data access and discovery. As an example, a user who is looking for public expenditure for a specific city between two specific years will experience better discovering relevant data due to enriching datasets with temporal and geographic. ***The usage context*** aims to represent the main government-related concepts and entities (persons, locations, and organizations) describes in the OGD datasets. Besides, the usage context represents OGD value (social, economic, political …) and potential benefits (transparency, innovation, economic growth) both expected from the publication of the dataset. ***The organizational context*** aims to represent the public organization, responsible for publishing the dataset, Its structure, changes, and evolutions. It aims to support tracking a part of the dataset's provenance data which improves trust and believability. ***The legal context*** aims to represent datasets related legislation including laws and regulations.

B-OGD content knowledge representation model

OGD content knowledge representation model aims to provide rich background knowledge to semantically enhance the description of OGD datasets with government domain-specific concepts, relations, and entities. Unlike, the OGD metadata knowledge representation model which is generic and domain-independent, the OGD content knowledge representation model is domain-specific and straightly related to the dataset's domain.

4.2 The proposed ontologies for knowledge representation

Ontologies represent a powerful knowledge representation [38] formalism for semantic modeling and annotation They present a shared understanding of a domain of interest and capture its semantic content in a way that can be machine-processable. They present a high level of expressiveness and allow reasoning and knowledge inference. Our ontological model involves two main ontologies. The GovDataset Ontology and The GovDomain Ontology. Each ontology supports the representation of a specific dimension of our model respectively the OGD metadata knowledge representation dimension and the OGD content knowledge representation dimension.

A-The GovDataset Ontology

Reusing existing, established ontologies when building one's ontology is crucial for achieving interoperability. For this aim, the GovDataset Ontology was built around existing standards, ontologies and vocabularies, the GovDatset Ontology employs also the following vocabularies and ontologies:

- **Dublin Core [32]:** a vocabulary of terms that can be used to describe digital resources as well as physical resources
- **DCAT [33]:** an RDF vocabulary recommended by W3C designed to facilitate interoperability between data catalogs published on the Web. DCAT enables a publisher to describe datasets and data services in a catalog using a standard model and vocabulary that

facilitates the consumption and aggregation of metadata from multiple catalogs

- **The Organization ontology [34]:** a w3C recommendation designed to enable the publication of information on organizations and organizational structures.

- **The Legal ontology:** incorporates concepts and properties used to model the structure of legislative and law documents. The legal ontology was developed by our research team in previous works **[35]**.

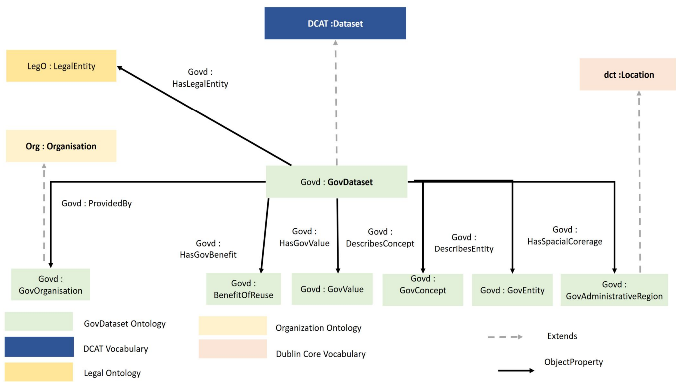


Figure 1: The main classes of the GovDataset Ontology

Figure 1 presents the main classes and properties of the GovDataset Ontology. The main class of the GovDataset Ontology is the GovDataset Class which extends the dcat: Dataset class and adds government-specific properties such as Govdatacategory and GovdataTopic. To represent the temporal context, we add two properties to the GovDataset class: GovdatasetTempCoverage and GovdatasetTempValidity to respectively represent the temporal coverage and temporal validity.

For the geographic context, the GovDataset ontology proposes the GovAdministrativeRegion class to represent the different geographic regions and their properties according to the national territory organization. It extends the Dublin core location class and It is instantiated according to the specific territory organization for each country. The GovDataset class is related to the GovAdministrativeRegion by the Hasgeographiccoverage object property. For the usage context, the GovDataset ontology proposes two classes the GovConcept and GovEntity classes that represent Government concepts and entities described by the dataset. These classes are related to the GovDataset class with the DescriblesGovConcept and DescribesGovEntity properties. Likewise, the GovValue and Govbenefit classes represent the government value and government benefits. Both classes are related to the GovDataset class by respectively hasGovValue and hasGovBenefit properties. For the organizational context, the Govdataset ontology proposes The main class of the GovOrganization ontology is the GovOrganisation class which extends the org organization class and is related to the Govdataset class by the property HasGovprovider .For the legal context, the Govdataset ontology reuses the legal entity class from the legal ontology and adds the relation haslegalcontext between the GovDataset Class and the legalentity class.

B-The GovDomain Ontology

The GovDomain ontology contains the domain's types, concepts, and relations among them. Its objective is to provide an understanding and interpretation of government concepts and relations used in datasets in a unified manner. The GovDomain Ontology is a Rich ontology that defines the possible concepts in government domain and relations as well as the restrictions that hold on them.

The GovDomain ontology is domain related and it is instantiated according to the dataset's domain.

4.3 The KG4OGD formalism

Based on the Open Government Knowledge representation model, we define the OGD Knowledge graph(KG4OGD) as Follow:

KG4OGD =(Gmd,Gd,R)

- Gmd represents the OGD metadata graph
- Gd represents the OGD data graph
- R represents a set of relations between Gd and Gmd

A-The OGD Metadata graph

Supported by the GovDataset Ontology as a schema model, the OGD Metadata graph is an Upper-level knowledge graph constructed based on the semantic lifting of the OGD dataset's metadata.

The Metadata graph is structured around a set of MetaGovFacts.Each MetaGovFact is represented as an RDF triple that describes an instance of the Govdataset class.

We note : MetaGovFact=(S,P,O)

where:

- The subject S: Instance of a class in the GovDataset Ontology
- The predicate P: Objectpropety or DataProperty of the GovDataset Ontology
- The Object O: Literal or Instance of a class in the GovDataset Ontology

B-The OGD data graph

Supported by the GovDomain Ontology as a schema model, the OGD data graph is constructed based on the semantic lifting of the OGD dataset's content.

The data knowledge graph is structured around a set of Govfacts. Each GovFact is represented as an RDF triple that describes a government fact extracted from the datasets.

We note :GovFact =(S',P',O')

- The subject S': Instance of a class in the GovDomain Ontology
- The predicate P': Objectpropety or DataProperty in the GovDomain Ontology
- The Object O': Literal or Instance of a class in the GovDomain Ontology

Each Govfact is related to its original GovDataset

R = {(GovFact, Govdataset)|GovFact ∈ Gd, Instance of the class GovDataset represented in the Gmd

5. CASE STUDY

These days, the world is witnessing an unprecedented pandemic that could not be compared to any global sanitary crisis since the Spanish flu in the early 1920s. As for sure, how the international community is responding to COVID-19 pandemic is not the same as back then, and that is driven by too many factors such as globalization and ICT development. Every citizen has become more interested in data related to COVID-19 evolution. In parallel to this interest, more area of concern has arisen; public awareness is steered more toward healthcare domain. Citizens, social activists, experts, and academics are all interested in exploring and analyzing the health care system and sanitary data.

On a national scale, the Moroccan health ministry has implemented a portal to publish datasets related to health care infrastructure, resources, and services. This work is a continued effort of the OGD initiative undertaken by the Moroccan government in early 2013, and which has been supported by the FOIA (freedom of information act) regulation enacted in 2019.

The portal offers several datasets that present available healthcare infrastructure, resources, and services according to national and regional levels. However, dataset's usage is hindered by several limitations, we summarize some on them bellow:

- Datasets are published in diverse and heterogeneous data formats
- Datasets are not described with metadata
- Syntactic information such as attribute names and attribute types is insufficient to understand the meaning of the data
- Overlapping exists between two or more datasets or additional information, but the data is represented differently which makes the identification of data mapping difficult.

Table 2 presents the main datasets used in our case study .

Table 2: The main datasets used in our use case study related to the public healthcare domain

Publisher	Datasets	Frequency of update	Scope
Moroccan health ministry	Public healthcare infrastructures	Quarterly	National /Regional
Moroccan health ministry	Private heath infrastructures	Quarterly	National /Regional
Moroccan health ministry	Healthcare human resources	Quarterly	National /Regional
Moroccan health ministry	Heavy biomedical equipment and high-technology healthcare facilities	Quarterly	National /Regional

To overcome these limitations and explore the potential of these OGD datasets. We propose the implementation of our approach to construct the KG4OGD for the public healthcare domain . The approach involves the following steps:

- The implementation of the GovDataset Ontology
- The construction of the OGD Metadata graph
- The design, modelling and implementation of the GovDomain ontology for public healthcare
- The construction of the OGD Data graph

5.1 The GovDataset Ontology:

To implement the GovDataset Ontology, we use protégé [37] a tool for designing and modeling OWL ontologies. Figure presents the main classes of the ontology

The GovAdministrativeRegion class was instantiates according to the Moroccan territory administrative subdivision The GovOrganization class was instantiates according to the Moroccan public organizations' structure and subdivisions

Figure 2 presents the implementation of the GovDataset Ontology

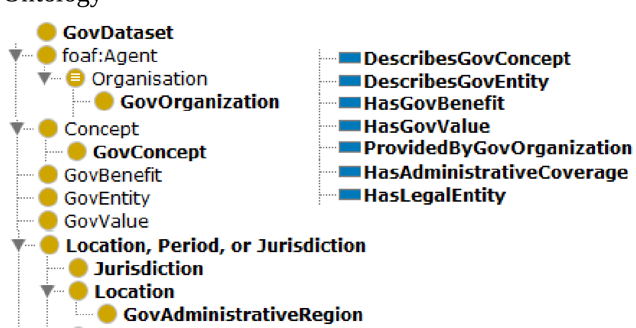


Figure 2: The main classes of the GovDataset ontology implemented in protégé tool

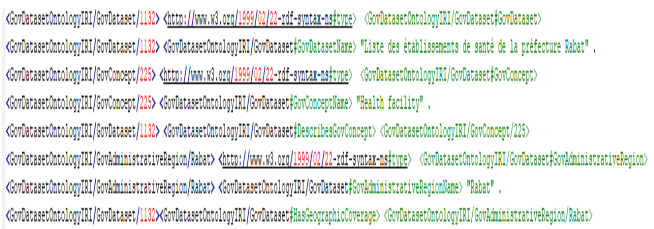
5.2 The OGD Metadata graph construction:

To construct the OGD Metadata graph, we perform the following steps for each selected dataset:

- Gathering and enriching available metadata.
- Adding contextual metadata (temporal, geographic, usage, organizational, legal)
- Semantic interpretation of metadata based on the Govdataset Ontology.
- MetaGovFact generation: Generation to RDF triples annotated with the GovDataset Ontology

This two last steps were supported by an open-source tool called Karma [38]. Karma a semi- automatic tool used to define mapping from structured sources to ontologies in order to build semantic descriptions (source models) and publish data as RDF.

Figure 3 present a set of MetaGovFacts that describes the dataset related to the list of health facilities of Rabat prefecture



```
<GovDatasetOntologyIRI/GovDataset/1132> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <GovDatasetOntologyIRI/GovDataset#GovDataset>
<GovDatasetOntologyIRI/GovDataset/1132> <GovDatasetOntologyIRI/GovDataset#GovDatasetName> "Liste des établissements de santé de la préfecture Rabat" .
<GovDatasetOntologyIRI/GovConcept/225> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <GovDatasetOntologyIRI/GovDataset#GovConcept>
<GovDatasetOntologyIRI/GovConcept/225> <GovDatasetOntologyIRI/GovDataset#GovConceptName> "Health facility" .
<GovDatasetOntologyIRI/GovDataset/1132> <GovDatasetOntologyIRI/GovDataset#DescribesGovConcept> <GovDatasetOntologyIRI/GovConcept/225>
<GovDatasetOntologyIRI/GovAdministrativeRegion/Rabat> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <GovDatasetOntologyIRI/GovDataset#GovAdministrativeRegion>
<GovDatasetOntologyIRI/GovAdministrativeRegion/Rabat> <GovDatasetOntologyIRI/GovDataset#GovAdministrativeRegionName> "Rabat" .
<GovDatasetOntologyIRI/GovDataset/1132><GovDatasetOntologyIRI/GovDataset#HasGeographicCoverage> <GovDatasetOntologyIRI/GovAdministrativeRegion/Rabat>
```

Figure 3: RDF triples annotated with the GovDataset Ontology representing the MetaGovFacts related to the list of health facilities of Rabat prefecture dataset

5.3 The GovDomain Ontology

We instantiated the GovDomain ontology by designing a domain-specific ontology related to public healthcare. the design and development involve experts in the public healthcare system and engineers. Its detailed description is behind the scope of the paper;

The ontology models concepts such as health facility, health region, health unit, healthcare staff and healthcare equipment Figure 4 presents the main concepts, objects properties and data properties of the ontology

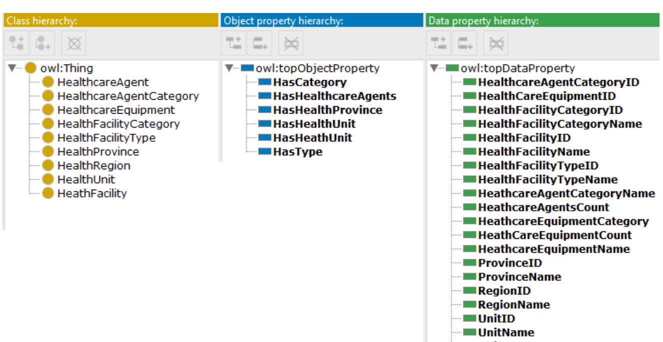


Figure 4: The main classes of the GovDomain ontology related to the Public health implemented in protégé tool

5.4 The OGD Data graph construction:

To construct the OGD data graph, we perform the following steps for each selected dataset:

- Dataset's semantic interpretation which involves attribute annotation, entities annotation and relation extraction
- GovFacts generation: by transforming the datasets into RDF triple annotated with the Govdomain Ontology

Both steps were supported by the usage of KARMA tool .

6. CONCLUSION

To support OGD value creation, we proposed in this paper the construction of the OGD knowledge graph to integrate data from available across multiple and disparate OGD. To support the construction of the knowledge graph, we propose a knowledge representation model based on ontologies to support the description of OGD datasets and their usage context, the organizational and legal context and to semantically enhance data transformation to knowledge. We implemented our model for the public healthcare domain.

This is a part of our ongoing work to improve the value creation of OGD, in our next works we will focus on the automation of the KG4OGD' construction and Update and explore the potential of the KG to generate smart applications for citizens

REFERENCES

1. M. Janssen, Y. Charalabidis, A. Zuiderwijk,(2012). **Benefits, adoption barriers and myths of open data and open government.** *Information systems management, 29(4), 258-268.*
2. B. Ubaldi (2013). **Open government data: Towards empirical analysis of open government data initiatives.**
3. K. Y. Dahbi, H. Lamharhar, D. Chiadmi **Exploring dimensions influencing the usage of Open Government Data portals.** *In Proceedings of the 12th International Conference on Intelligent Systems: Theories and Applications (pp. 1-6).*
4. S. Neumaier, J. Umbrich, A. Polleres (2016). **Automated quality assessment of metadata across open data portals**. *Journal of Data and Information Quality (JDIQ), 8(1), 1-29.*
5. J. Attard , F. Orlandi, , S. Scerri, S. Auer (2015). **A systematic review of open government data initiatives**. *Government Information Quarterly, 32(4), 399-418.* https://doi.org/10.1016/j.giq.2015.07.006
6. K. Y. Dahbi, H. Lamharhar, D. Chiadmi .**Toward a user-centered approach to enhance Data discoverability on Open Government Data portals**. *In 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS) (pp. 1-5). IEEE.*

7. M. Mountantonakis, Y. Tzitzikas, (2019). **Large-scale semantic integration of linked data: A survey**. *ACM Computing Surveys (CSUR), 52(5), 1-40.*
8. H.Purohit , R. Kanagasabai, N. Deshpande (2019, January). **Towards Next Generation Knowledge Graphs for Disaster Management.** *In 2019 IEEE 13th International Conference on Semantic Computing (ICSC) (pp. 474-477). IEEE.*
9. Y. Jia, Y.Qi , H.Shang , R. Jiang, A. Li, (2018). **A practical approach to constructing a knowledge graph for cybersecurity**. *Engineering, 4(1), 53-60.*
10. M. Wang, Q. Zeng, W. Chen , J. Pan, H. Wu, C. Sudlow, D. Robertson, (2020). **Building the Knowledge Graph for UK Health Data Science.**
11. J. M. Gomez-Perez, J. Z .Pan, G. Vetere, , H. Wu, (2017). **Enterprise knowledge graph: An introduction**. *In Exploiting linked data and knowledge graphs in large organisations (pp. 1-14). Springer, Cham.*
12. T. Ruan , L. Xue, H . Wang, F. Hu, L. Zhao, J. Ding, (2016, October). **Building and exploring an enterprise knowledge graph for investment analysis.** *In International Semantic Web Conference (pp. 418-436). Springer, Cham.*
13. A. Hogan, E . Blomqvist, M .Cochez, C . d'Amato, G. de Melo, C. Gutierrez, , R. Navigli(2020). **Knowledge graphs**. *arXiv preprint arXiv:2003.02320.*
14. J. Yan, C. Wang, , W. Cheng, M. Gao, A. Zhou, (2018). **A retrospective of knowledge graphs**. *Frontiers of Computer Science, 12(1), 55-74.* https://doi.org/10.1007/s11704-016-5228-9
15. P. Haase, , D. M. Herzig, A. Kozlov, A. Nikolov, J. Trame, (2019). **metaphactory: A platform for knowledge graph management.** *Semantic Web, 10(6), 1109-1125.*
16. Z. Zhao, S. K. Han, I. M. So (2018). **Architecture of knowledge graph construction techniques**. *International Journal of Pure and Applied Mathematics, 118(19), 1869-1883.*
17. J. Tauberer , L. Lessig, (2007). **The 8 principles of open government data**. *http://www. opengovdata. org/home/8principles.*
18. H. Paulheim (2017). **Knowledge graph refinement: A survey of approaches and evaluation methods.** *Semantic web, 8(3), 489-508.*
19. S. Auer , C. Bizer , G. Kobilarov, J. Lehmann, R. Cyganiak, Z. Ives (2007). **Dbpedia: A nucleus for a web of open data.** *The semantic web (pp. 722-735). Springer, Berlin, Heidelberg*.
20. J. Hoffart , F. M. Suchanek, K. Berberich, G. Weikum (2013). **YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia**. *Artificial Intelligence, 194, 28-61.*
21. X. Dong , E. Gabrilovich, G. Heitz , W. Horn , N. Lao, K. Murphy, W.Zhang (2014, August). **Knowledge vault: A web-scale approach to probabilistic knowledge fusion.** *In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 601-610).*
22. D. Vrandečić, M. Krötzsch, (2014). **Wikidata: a free collaborative knowledgebase**. *Communications of the ACM, 57(10), 78-85.* https://doi.org/10.1145/2629489
23. P. H. Barchi, E. R. Hruschka (2014, December). **Never-ending ontology extension through machine** reading. In 2014 14th International Conference on *Hybrid Intelligent Systems (pp. 266-272). IEEE.*
24. I. Zaoui, F. Wadjinny, D. Chiadmi, L. Benhlima **Construction d'un profil utilisateur pour un médiateur de bibliothèques électroniques** *Proceeding of WOTIC, Agadir, Morocco, 2009*
25. J.W. Lee, J. Park (2019). **An Approach to** Constructing **a Knowledge Graph Based on Korean Open-Government Data**. *Applied Sciences, 9(19), 4095*.
26. A. Ojo, O. Sennaike, (2020, January). **Constructing Knowledge Graphs from Data Catalogues.** *In International Conference on Distributed Computing and Internet Technology (pp. 94-107). Springer, Cham.*
27. S. Ronzhin, E. Folmer, P. Maria, M. Brattinga, W. Beek, R. Lemmens, R., R. van't Veer(2019). **Kadaster knowledge graph: Beyond the fifth star of open data**. *Information, 10(10), 310*.
28. P. Espinoza-Arias, M. J. Fernández-Ruiz, V.Morlán-Plo, R.Notivol-Bezares, O. Corcho (2020). **The Zaragoza's Knowledge Graph: Open Data to Harness the City Knowledge**. *Information, 11(3), 129.* https://doi.org/10.3390/info11030129
29. H. Kim (2018). **Interlinking Open Government Data in Korea using Administrative District Knowledge Graph.** *Journal of Information Science Theory and Practice*
30. Soylu, A., Corcho, O., Simperl, E., Roman, D., Francisco, Y. M., Taggart, C., ... & Konstantinidis, G. (2018). **Towards integrating public procurement data into a semantic knowledge graph**.
31. H. Purohit, R. Kanagasabai, N. Deshpande (2019, January). **Towards Next Generation Knowledge Graphs for Disaster Management**. *In 2019 IEEE 13th International Conference on Semantic Computing (ICSC) (pp. 474-477). IEEE*
32. Dublin Core Metadata Initiative. (2012). **Dublin core metadata element set, version 1.1.**
33. World Wide Web Consortium. (2014). **Data Catalog Vocabulary (DCAT).**
34. World Wide Web Consortium. (2014).**The organization ontology**
35. L. Hind, D. Chiadmi, L. Benhlima, L. (2014). **How semantic technologies transform the e-government domain: a comparative study and framework**. *Transforming Government: People, Process, and Policy, 8(1), 49-75.*

36. A.Moujane, D .Chiadmi, L .Benhlima .**Semantic Mediation: An Ontology-Based Architecture Using Metadata** *Proceeding of the 4th International Multiconference on Computer Science and Information Technology (CSIT 2006) Amman, Jordan, April 5-7, 2006, pp. 317-326,*
37. N. F. Noy, M. Crubézy, R. W , H. Fergerson, Knublauch, , S. W. Tu, J. Vendetti, M. A. Musen, (2003). **Protégé-2000: an open-source ontology-development and knowledge-acquisition environment.** *In AMIA Annual Symposium proceedings. AMIA Symposium (pp. 953-953).*
38. S. Gupta, P. Szekely, C. A. Knoblock, A.Goel, M. Taheriyan, M. Muslea, (2012, May). **Karma: A system for mapping structured sources into the Semantic Web**. *In Extended Semantic Web Conference (pp. 430-434). Springer, Berlin, Heidelberg.* https://doi.org/10.1007/978-3-662-46641-4_40