https://doi.org/10.2352/issn.2168-3204.2020.1.0.92

Linked Open Data Prototype of the Historical Archive of the European Commission

Mariana Damova; Mozaika; Sofia, Bulgaria

Abstract

In the age of the WWW consistent efforts have been made to make information from the archives available for the general public to facilitate access to the wealth of documentary history for research, consultation or education purposes. Linked Open Data (LOD)[1] provide a well suited framework to expose archival content to the general public while enriching it with content from other sources. This paper describes the creation of the first of its kind linked open data prototype to access data from the Historical Archive of the European Commission (HAS), carried out within ISA² programme of the European Commission [2]. We present the designed ontology based on ISAD(G)[3], ISAAR(CPF)[4] and RIC-CM [5] models and the business processes of HAS, and the created knowledge base from a sample of HAS data, re-using authority lists from the Publication Office [6] and EuroVoc [7] and allowing querying via SPARQL endpoint [8].

Introduction

The archives have been considered as part of the cultural heritage of the world and hence as relevant information providers for research on a given subject in concert with museums, libraries and galleries (GLAM sector). Relevant information about a single object of interest does not reside in a single preservation institution, but rather it is scattered in different institutions. They regard it from different perspectives and preserve specific aspects of it, hence the need to draw information from a number of institutional databases in order to obtain an overall view about it. Linked data technologies [9] are based on standards for data modeling and representation allowing for an unprecedented ease of integration of heterogeneous data sources, structuring of unstructured information at Web scale and enabling semantic interoperability, machine readable semantic data automatic reasoning and generation of new knowledge. These features have been explored in a series of cultural heritage and archival projects such as Europeana [10], Rijks museum collection [11], The National Archive of the United Kingdom [12] that have adopted linked data technologies to describe their collections. The European Commission's Publication Office has been maintaining a large amount of vocabularies describing different types of entities described in linked open data format. One of its aims is to develop, maintain and promote an integrated approach to interoperability in the EU and to contribute to the development of reusable IT solutions at European, national, regional and local levels of public administration. So, it has been of interest to produce solutions that re-use vocabularies and make use of the power of the semantic web and the linked open data. Publishing the HAS data as linked open data addresses this interest and allows to showcase a convincing use case of exposing archival data to the general public enriched with external content. That is why the ISA² programme of the European Commission has funded an action "Standard-based archival data management, exchange and publication" [13] that includes a work package about exploring the relevance of the application of linked open data in this context. This paper presents the creation of the linked data prototype of the Historical Archive of the European Commission. The paper is structured in the following way: we first outline the technological setting of the linked open data, then we present the adopted methodology and approach, followed by a description of the HAS ontology model, the HAS knowledge base and prototype implementation. We conclude with examples of queries and data exploration.

Technological setting

Semantic Web and Linked data technologies are based on standards for data modeling and representation suitable to meet the information management needs of the 21st century. Their advantages are in the cost of production, subsequent maintenance and efficiency of hardware resources utilization. These advantages can be summarized in the following points:

- RDF [14] – the basic data format of semantic technologies, is schema agnostic having both data and schema represented in the same RDF format. That is to say it is as easy to add data as to change their schema by adding RDF triples into a semantic database. This makes the cost of maintenance of semantic repositories over time quite low.
- Only the available information is being stored in the semantic databases, e.g. there is no waste of valuable hardware space because of the need to comply with the specification of the relational tables by leaving empty cells.
- Inference - new knowledge gets created based on the models that allow for the generation of new implicit facts out of the explicit ones. This makes the number of the explicitly introduced facts in the database to be one third or one fourth of the actually available facts for querying.
- Easy integration and interlinking between data from different and heterogeneous sources, as they are based on open standards, breaking the data siloes, c.f. Figure 1. This enables retrieval of information from different datasets with a single query.
- Easy combination with natural language processing tools to make possible to register the occurrence of conceptual entities in texts.

The information infrastructures based on these technologies allow for flexible querying, based not only on keywords, but also on semantic concepts, retrieving structured information about the searched objects and structured information from the text documents they are associated with.

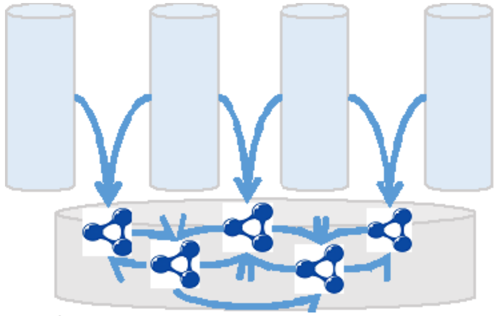


FIGURE 1 SEMANTIC DATA INTEGRATION BREAKING DATA SILOS

Prototype context setting

The LOD prototype of HAS is an implementation of such a semantic infrastructure that encompasses the HAS information that is of interest to the general public, e.g. the typical users of open access information – researchers, historians, political scientists politicians, students, scholars or citizens, jurists, journalists, public bodies, etc. but also archive professionals from other archival organizations. The main goal of the LOD prototype has been to show how linked data technologies can be applied in the context of the HAS and to showcase the possibilities and the effects of linking, enriching and re-using existing resources for the end users, but also for the creators of authority vocabularies and of archival records. For this reason, the LOD prototype's scope has covered a demonstration of semantic search and information acquisition, semantic integration of data from the HAS with datasets from the Publications Office, e.g. re-using already existing resources of the European Commission, and with general purpose information datasets such as DBPedia [15] and Wikipedia [16] for data enrichment.

To achieve this demonstration and to design the LOD prototype, we answered the following questions: 1. What part of the information available at the HAS is relevant for the end user? 2. How the information is to be exposed to the general public? 3. How the information to be exposed is retrieved from the HAS? To answer the first question, we analyzed the archives management process and the information available for the archived entities. For the second question, we analyzed the way archival information is being exposed by other archival institutions and what access and publications channels are already available within the European Commission. To answer the third question, we analyzed the way the relevant information is organized, stored and accessible. As a result, the LOD prototype has been set up as a site, a SPARQL end point that allows users, knowledgeable in the LOD query language SPARQL [17] to formulate queries, to inspect and navigate through the query results experiencing the effects of linked open data.

Methodology

To build a LOD prototype we need three components: 1. a conceptual model, 2. a knowledge base, and 3. a way to expose the linked data. The conceptual model in the LOD context has to be an ontology, reflecting the domain, the business processes and the usage scenarios of the intended LOD application. Following the best practices of building conceptual models for LOD applications we have scanned and re-used concepts from existing ontologies. They will be described below. Further, the ontology reflected the archiving standards ISAD(G) and ISAAR(CPF) on the one hand and addressed the specific business processes of the HAS, on the other. For the knowledge base, it is necessary to select and analyze the data to be included in it and to design extraction, transformation and loading (ETL) procedures to build it, further to determine data enrichment strategies to select and add revelant information about the objects in the knowledge base from related external sources. For this, we have developed converters of data into RDF for the HAS datasets, and chose RDF datasets available from the Publication Office, EuroVoc and DBpedia. Two sample datasets from the HAS had to be analyzed and converted in linked data format, and entities that could be semantically enriched identified. As already stated the chosen approach to expose the LOD data of HAS has been via SPARQL end point. We have used GraphDB [18], one of the well established semantic software platforms, to produce it.

HAS ontology

To design HAS ontology the method of conceptual analysis has been adopted, by first analyzing the available empirical evidence, then building conceptual model, and consequently implementing it as an OWL ontology using OntoClean [19] method. OntoClean is a method for defining the ontology elements following the rule to distinguish between objects and roles. The analysis of the exemplary data, provided by the Historical Archive of the European Commission, comprising a sample with records sets about the European Commission, and a sample with record sets about Roy Jenkins, supplemented with the analysis of several archive management models, e.g. RIC-CM, EDM [20], ISAD(G), ISAAR(CPF), preliminary version of RIC-O [21], [22] helped build a conceptual model that re-uses concepts from ISAD(G) and ISAAR models that have been reflected in RIC-CM as well. The different conceptual domains were represented by proper ontological models. Figure 2 shows how ontological models were built from different conceptual domains and source datasets and then merged into a single ontological model.

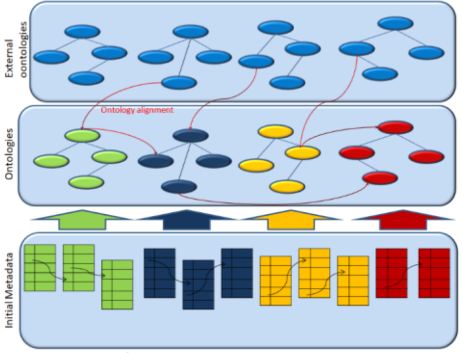


FIGURE 2 ONTOLOGY ENGINEERING APPROACH

The source concept in the LOD prototype ontology is the ISAD(G)'s "Unit of description". The "unit of description" instantiates different possible ISAD(G)'s "Levels of description", the concept that gathers all manifestations of archival content organization - Fonds, Series, File or Record Set, Item or Digital Object, Physical object, Record, or Transition. Figure 3 below shows the conceptualized relation between the "Unit of description" and the "Levels of description". It is worth noting that the concepts, classified as Levels of description in the HAS ontology belong to two different categories of concepts in the Archives management world, e.g to the archives metadata management like Transition, and to the content of the unit of description, such as Record, Record set, etc.

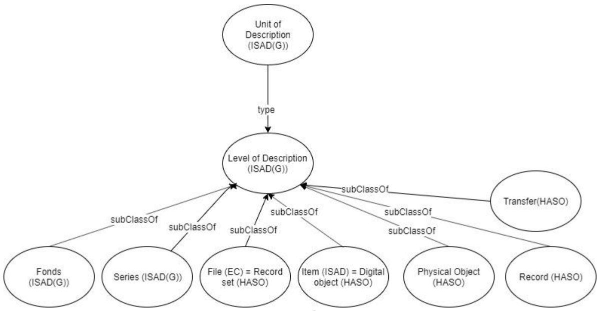


Figure 3 *Representation of ISAD's levels of description*

The analysis of the business processes as HAS has lead to a conceptualization of the relationship between a record, record set and digital object that links the record set and the record in an indirect way, via the concept of the digital object. Thus, the digital object being the carrier of the actual content is represented as central for this conceptual construct. Figure 4 below shows this conceptualization.

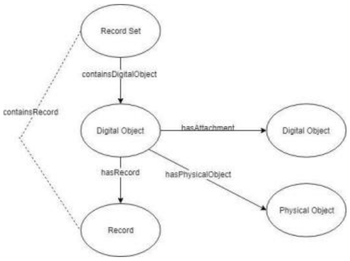


Figure 4 *Record – Record set – Digital object representation*

As far as the content related metadata of the "Level of description" are concerned the HAS ontology has covered the ISAD(G)'s standard by re-using the characterizations like identifier, title, extent, medium, place of holding, place of creation, clearance, accrual, but it has also provided with models allowing metadata enrichment and linking of objects occurring in the titles and other textual descriptions, as well as in the extents (cf. Figure 5).

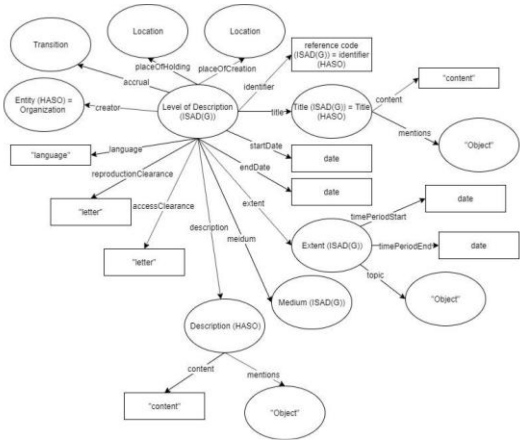


Figure 5 ***Level of description***

The creators of the "Levels of description" have been represented in HAS ontology in compliance with the ISAAR(CPF) model and with the business concepts description of the HAS. The concept "Entity" stands for an Agent that can be an Organisation, a Family or a Person. The model on Figure 6 shows how ISAAR(CPF) model is reflected covering a variety of aspects of "Entity", like identifier, name, function, location of operation, location of foundation, date of establishment, date of dissolution, address, email, related entities, record.

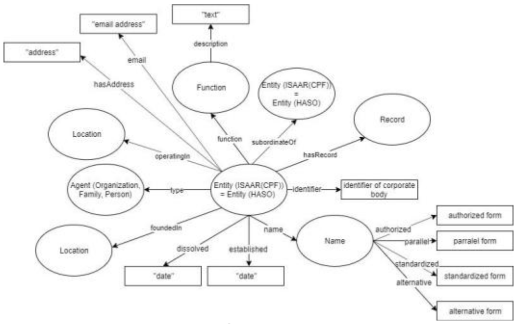


Figure 6 *Creator – Entity*

The model for Person re-uses the PROTON ontology [23] conceptualization, and the model for legal references partially re-uses ELI (European Legislation Identifier) ontology [24] and the Legal ontology of CNR [25]. With this respect HAS ontology adds further granularity to the representation of legal entities by introducing the concepts "Article" and "Paragraph" and by modeling the relationships between the Legal concepts, as shown on Figure 7.

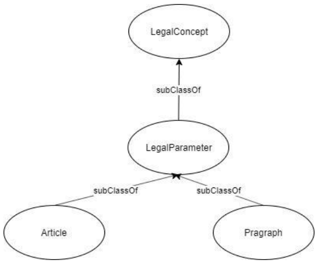


Figure 7 *Legal concept*

Overall HAS ontology, c.f. Figure 8, comprises 26 classes and 44 properties. Except for the original concepts, interpreting ISAD(G), ISAAR(CPF), RIC-CM, and the business processes and records at HAS, it re-uses concepts and relations from XSD Schema [26], SKOS [27], Dublin Core [28], [29], DBpedia ontology [30], PROTON ontology, Legal ontology, ELI.

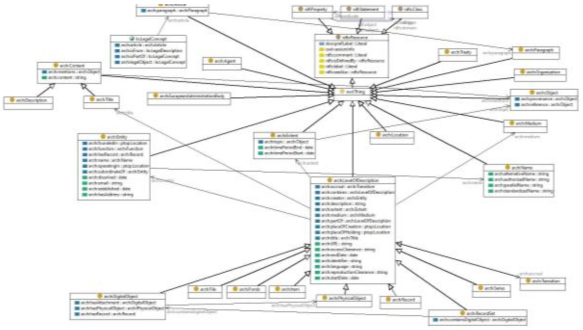


Figure 8 *HAS ontology*

A comparison with the recently published RIC-O shows that HAS ontology is compact and oriented towards description of the archive content for external users, whereas RIC-O blends the entire RIC-CM model. Further, a closer look at the conceptualizations in

HAS ontology and in RIC-O indicates that different approaches to ontology modeling have been adopted with respect to the definitions of concepts and properties. For example, RIC-O defines concepts from RIC-CM, describing relations as concepts, whereas HAS ontology specifies such concepts from RIC-CM as properties.

HAS knowledge base

The HAS knowledge base contains two samples of datasets from the HAS with records from the European commission (COM) and about Roy Jenkins. From them the fields with textual content from the HAS datasets – the titles, the comments and the keywords on the one hand and the fields describing the creators, referring to organizations referring to European institutions, places, people on the one hand and the keywords, encoded in the HAS datasets referring to people, places and legal references on the other have been selected to be included into the HAS knowledge base, because they correspond to the information of interest for the target end users, and because this selection allowed to showcase the benefits of adopting the LOD approach for exposing archival data. In addition, a number of datasets from the Publication office and from Eurovoc have been selected after analysis of the data suitable for enrichment. Thus, authority vocabularies from the Publication office and from EuroVoc, describing corporate bodies, people, locations, and legal datasets have been incorporated into the HAS knowledge base, e.g. re-used. They were also employed in the process of metadata enrichment. Where appropriate they were also associated with DBpedia or Wikipedia entries describing people, locations, organizations. The employment of the different datasets is seamless as the generated RDF triples referring to entities described in the re-used datasets, include their URIs directly. For example, the three RDF triples below, ex. 1-3, show how a record set title is linked to Roy Jenkins and the European Commission, using two different vocabularies, e.g the vocabulary Person of the Publication Office, and the resource of DBpedia. Further metadata enrichment has been achieved by linking the Publication Office entry about Roy Jenkins to the DBpedia entry about Roy Jenkins.

```
(1)  Archr:11597_title  arch:mentions popers:Roy_Jenkins .
(2)  popers:Roy_Jenkins owl:sameAs    dbpedia:Roy_Jenkins .
(3)  archr:11596_title  arch:mentions dbpedia:European_Commission .
```

Overall, the HAS knowledge base is composed of the following datasets: 1) HAS COM sample with records sets created by the European Commission 2) HAS Jenkins sample with record sets about Roy Jenkins, 3) Publication Office corporate bodies, 4) Publication office people, 5) Publication office place and country, 6) Publication office legal references, 7) EuroVoc people, places and corporate bodies, and reference to 8) DBpedia entries for people and places, 9) Wikipedia entries for institutions and people. The employed schemata are the semantic web stack, SKOS, DBpedia ontology, Protontop, DC, Legal ontology, HAS ontology.

The HAS datasets are available in several files comprising: the description of the schema of the dataset, the Datasets (csv), Main descriptive metadata; the creators and samples of keywords pertaining to the entries of the datasets. These separate files have been represented in a single semantic knowledge graph. The conversion of the HAS datasets into RDF has been done using semantic principles, resulting in a lean and straightforward set of semantic data that have been abstracted away from relational data keys.

The HAS knowledge base is designed as a reason-able view [31], e.g. all datasets and ontologies have been loaded into a single semantic repository and OWL [32] reasoning has been performed on them while loading. This generates implicit facts and increases amount of information available for querying. Further, the OWL reasoning rules have been augmented with additional inference rules that reflect the model, represented on Figure 4 above. They generate automatically facts based on the relation between Record set, Record and Digital Object, on the one hand increase the number of facts available for querying, and on the other hand reduce the effort for creating explicit facts for the HAS knowledge base, while properly reflecting the HAS business processes.

Table 1 below shows the size of the HAS knowledge base. The implicit statements, e.g. the automatically generated facts, are more than twice as much as the explicitly introduced statements. This shows the power of inference and OWL reasoning to deliver new knowledge, as well as the potential of linked open data technologies to produce interesting instruments for content research and consultation.

Statements	Number of statements
Total statements	23 018 200
Explicit statements	9 714 718
Implicit statements	13 303 482
Expansion ratio	2.37

Table 1.

SPARQL End point

The SPARQL End point of the LOD prototype of the HAS is the interface between the end-users and the HAS knowledge base described above. It allows to query and to navigate through the linked data. The HAS ontology model, presented in figures 3, 4, 5, 6 and 7 makes explicit what kind of information is in the scope of the LOD prototype of HAS and has been made available for querying by the user. One can ask for information about Record sets as per their topics, about the Digital objects in English language that are part of a given Record Sets, about Record sets pertaining to a particular period of time or to particular legal reference – the legal document, the article and the paragraph. A Record Set about Roy Jenkins will lead not only to the Publication Office entry of Roy Jenkins, that provides data about the name of the politician in all European languages, but also to the DBpedia entry providing extensive information about this politicial. Similarly, if the topic of a given Record set is a city or a country, the entry for the city or the country is from EuroVoc or from the Publication Office, and give access to the EuroVoc or the Publication Office information available about them.

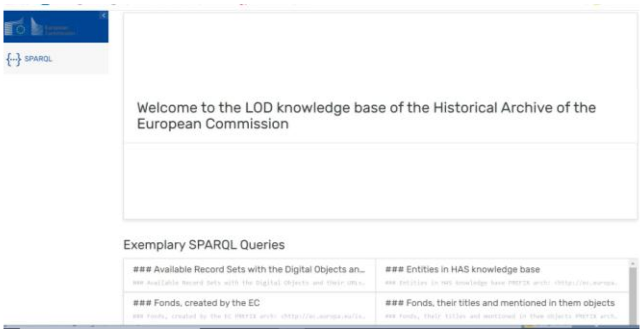
Figure 9 shows the opening page of the HAS SPARQL End point when it is being loaded.

HAS LOD knowledge base is loading...

Waiting for has.mozajka.co...

***FIGURE 9** OPENING PAGE OF HAS SPARQL END POINT*

The landing page is shown on Figure 10. It offers a number of exemplary queries that can be used as a starting point for the exploration and discovery of the HAS knowledge base.



Welcome to the LOD knowledge base of the Historical Archive of the European Commission

Exemplary SPARQL Queries

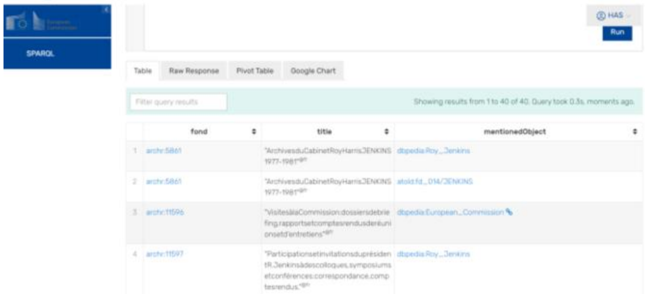
### Available Record Sets with the Digital Objects an... ### Available Record Sets with the Digital Objects and their URIs...	### Entities in HAS knowledge base ### Entities in HAS knowledge base PREFIX arch: <http://ec.europa...
### Fonds, created by the EC ### Fonds, created by the EC PREFIX arch: <http://ec.europa.eu/is...	### Fonds, their titles and mentioned in them objects ### Fonds, their titles and mentioned in them objects PREFIX arch...

FIGURE 10 *THE HAS SPARQL END POINT LANDING PAGE*

The following example shows a navigation flow possible because of the linked data architecture. The query "Topics of record sets, based on their titles":

```
PREFIX arch:
<http://ec.europa.eu/isa/ontology#>
select DISTINCT ?fond ?title ?object
where {
    ?fond arch:title ?titleNode .
     ?titleNode arch:mentions ?object .
    ?titleNode arch:content ?title .
} limit 100
```

returns the results of this query as shown on Figure 11. The entries under "fond" and the entries under "object" are links that will lead to further information about the entities they describe.

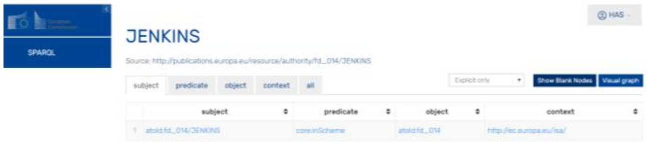


Showing results from 1 to 40 of 40. Query took 0.3s, moments ago.

	fond	title	mentionedObject
1	archr:5861	"ArchivesduCabinetRoyHarrisJENKINS 1977-1981"@fr	dbpedia:Roy_Jenkins
2	archr:5861	"ArchivesduCabinetRoyHarrisJENKINS 1977-1981"@fr	atold:fd_014/JENKINS
3	archr:11596	"VisitesàlaCommission:dossiersdebriefing,rapportsetcomptesrendusderéunionsetd'entretiens"@fr	dbpedia:European_Commission
4	archr:11597	"Participationsetinvitationsduprésident R.Jenkinsàdescolloques,symposiumsetconférences:correspondance,comptesrendus."@fr	dbpedia:Roy_Jenkins

FIGURE 11 *QUERY RESULTS*

Selecting the link of Joy Jenkins, the third raw of the third column under Object from the results table of Figure 11, we obtain the Publication Office's description of Jenkins that is shown in its full form on Figure 12, following the link, e.g. the URI of Jenkins from the Publication office.



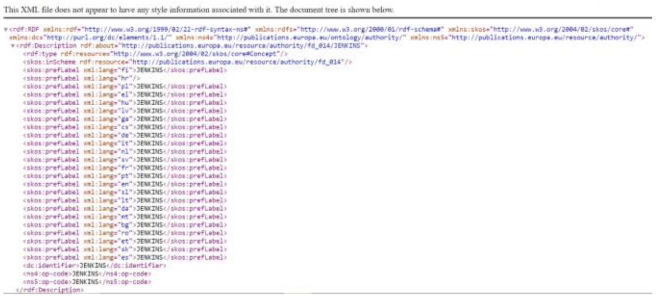
JENKINS

Source: http://publications.europa.eu/resource/authority/fd_014/JENKINS

	subject	predicate	object	context
1	atold:fd_014/JENKINS	core:inScheme	atold:fd_014	http://ec.europa.eu/isa/

FIGURE 12 *LINKED ENTRY OF ROY JENKINS FROM PUBLICATION OFFICE AUTHORITY LIST EXPLICIT STATEMENT*

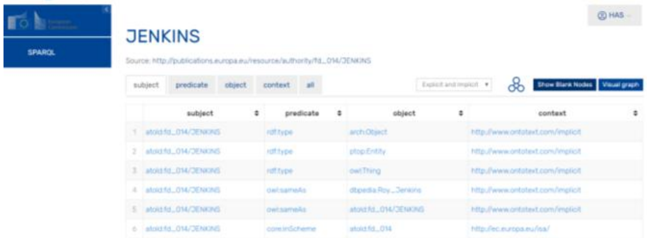
Clicking on the URI of Jenkins, the link in the first column, under Subject, shows the authority list entry of Jenkins in the Publication Office, c.f. Figure 13, where one can see information available about Roy Jenkins in the Publication Office. This is the name of Roy Jenkins, referred to in different languages.


This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
▼<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:skos="http://www.w3.org/2004/02/skos/core#"
 xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:ns4="http://publications.europa.eu/ontology/authority/" xmlns:ns5="http://publications.europa.eu/resource/authority/">
 ▼<rdf:Description rdf:about="http://publications.europa.eu/resource/authority/fd_014/JENKINS">
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <skos:inScheme rdf:resource="http://publications.europa.eu/resource/authority/fd_014"/>
    <skos:prefLabel xml:lang="fi">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="hr"/>
    <skos:prefLabel xml:lang="pl">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="el">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="hu">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="lv">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="ga">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="cs">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="de">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="it">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="nl">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="sv">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="fr">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="pt">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="en">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="sl">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="lt">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="da">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="mt">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="bg">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="ro">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="et">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="sk">JENKINS</skos:prefLabel>
    <skos:prefLabel xml:lang="es">JENKINS</skos:prefLabel>
    <dc:identifier>JENKINS</dc:identifier>
    <ns4:op-code>JENKINS</ns4:op-code>
    <ns5:op-code>JENKINS</ns5:op-code>
  </rdf:Description>
```

FIGURE 13 *LINKED DATA ENTRY OF ROY JENKINS FROM THE PUBLICATION OFFICE AUTHORITY LIST*

So far, all results about Jenkins for the query mentioned above have been explicit facts, i.e. statements inserted directly into the HAS knowledge base from the converted data. But LOD technologies are able to apply formal logic to generate automatically new facts based on the ontologies, the semantic web technology stack, e.g. RDF, RDFs [33], OWL, and the explicitly entered data all introduced into the knowledge base. Figure 14 shows additional facts about Roy Jenkins that have been generated automatically. They are concepts classifying the object of Roy Jenkins into more general categories, such as Object, Entity Thing, and linked objects from other datasets also referring to Roy Jenkins from DBpedia that ensure the data enrichment. Thus, the inference makes 6 more facts about Roy Jenkins available for querying. Moreover, this example demonstrates the benefit of introducing the equivalence between the entity, describing Roy Jenkins, described in the Publication Office dataset and the entity, describing Roy Jenkins described in DBpedia, as this makes possible to obtain more information about Jenkins by clicking on the DBpedia entry about it, displayed in the query results.



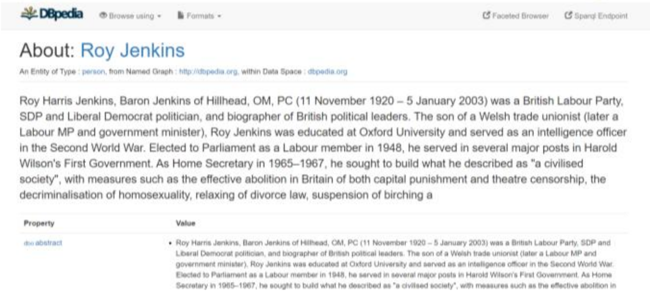
JENKINS

Source: http://publications.europa.eu/resource/authority/fd_014/JENKINS

	subject	predicate	object	context
1	atold:fd_014/JENKINS	rdf:type	arch:Object	http://www.ontotext.com/implicit
2	atold:fd_014/JENKINS	rdf:type	ptop:Entity	http://www.ontotext.com/implicit
3	atold:fd_014/JENKINS	rdf:type	owl:Thing	http://www.ontotext.com/implicit
4	atold:fd_014/JENKINS	owl:sameAs	dbpedia:Roy_Jenkins	http://www.ontotext.com/implicit
5	atold:fd_014/JENKINS	owl:sameAs	atold:fd_014/JENKINS	http://www.ontotext.com/implicit
6	atold:fd_014/JENKINS	core:inScheme	atold:fd_014	http://ec.europa.eu/isa/

FIGURE 14 *EXPLICIT AND IMPLICIT STATEMENTS ABOUT ROY JENKINS*

Figure 15 below shows the beginning of the DBpedia entry about Roy Jenkins. Thus, the HAS knowledge base entry for Roy Jenkins has been enriched with the information from the Publication Office available about it, and with the information from DBpedia available about it.



About: Roy Jenkins

An Entity of Type : person, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

Roy Harris Jenkins, Baron Jenkins of Hillhead, OM, PC (11 November 1920 – 5 January 2003) was a British Labour Party, SDP and Liberal Democrat politician, and biographer of British political leaders. The son of a Welsh trade unionist (later a Labour MP and government minister), Roy Jenkins was educated at Oxford University and served as an intelligence officer in the Second World War. Elected to Parliament as a Labour member in 1948, he served in several major posts in Harold Wilson's First Government. As Home Secretary in 1965–1967, he sought to build what he described as "a civilised society", with measures such as the effective abolition in Britain of both capital punishment and theatre censorship, the decriminalisation of homosexuality, relaxing of divorce law, suspension of birching a

Property	Value
dbo:abstract	• Roy Harris Jenkins, Baron Jenkins of Hillhead, OM, PC (11 November 1920 – 5 January 2003) was a British Labour Party, SDP and Liberal Democrat politician, and biographer of British political leaders. The son of a Welsh trade unionist (later a Labour MP and government minister), Roy Jenkins was educated at Oxford University and served as an intelligence officer in the Second World War. Elected to Parliament as a Labour member in 1948, he served in several major posts in Harold Wilson's First Government. As Home Secretary in 1965–1967, he sought to build what he described as "a civilised society", with measures such as the effective abolition in

FIGURE 15 *METADATA ENRICHMENT ABOUT ROY JENKINS*

Similar navigation and discovery chains, topics detection and representation can be experienced with other types of entities like legal references, places, departments at the European Commission, record sets, records, documents, etc. The SPARQL endpoint of the HAS prototype is available at: `http://has.mozajka.co.`

Conclusions

This paper presented the LOD prototype of the PHASE II : Work Package 4 of ISA² Action 2017.01. It outlined the elements of building it, starting from a suitable data model, selection of datasets to re-use, creation of the knowledge base, and the SPARQL endpoint and querying about different facets of the data. This LOD prototype is the basis for evaluating the adoption of LOD approach for exposing the HAS content to the general public. It is a good example of representing archival data in LOD format and demonstrating the capabilities for enrichment. This prototype showcases and emphasizes the advantages of the linked open data approach for exposing archival information to the general public. It is intended to be extended with the full scale of archival data and business processes from HAS and to be deployed as a real life user facing solution in the future.

The most prominent result of our research has been the HAS ontology that should be regarded as Archives domain domain ontology. It is an OWL ontology comprising originally defined 26 classes and 44 properties, and re-uses as per the best practices guidelines concepts and properties from other ontologies and schemata. The HAS knowledge base, implementing a reason-able view, also re-uses many available linked open data resources. Most importantly, the Publication Office datasets and EuroVoc have been put into action by adopting their vocabularies for achieving metadata generation and enrichment. The LOD prototype showed how these resources are intertwined within the produced semantic facts from the HAS datasets. It also showed the advantages of inference, providing complementary useful information to the users after applying OWL reasoning over the included ontologies and over model dependent inference rules. The representation of the archival content is semantic, and hence all keys from the relational tables, where it is currently stored, have been removed. The LOD prototype of the HAS successfully shows the potential of the linked open data technologies for exposing archival information to the public. It demonstrates the flexibility of the semantic representation to analyze text and structure the information in the texts so it becomes accessible to semantic querying. The LOD prototype of the HAS is a solid starting point for the full scale implementation of LOD in the European Commission, bringing all advantages of the semantic web architectures, search, linking, enrichment, and easy maintenance.

Acknowledgement

This work has been carried out in collaboration with representatives from the Historical Archive of the European Commission and from the European Commission, e.g. Emmanuel Dervaux, Julie Urbain, Lieven Baert, Sorin Bobeika, Theodor-Bogdan Dimofte within the PHASE II : Work Package 4 of ISA² Action 2017.01, and Hermand Pessek, software developer at Mozaika. Mozaika was subcontracted by Infeurope S.A. for this project.

References

[1] http://linkedata.org

[2] https://ec.europa.eu/isa2/home_en

[3] https://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition

[4] https://www.ica.org/en/isaar-cpf-international-standard-archival-authority-record-corporate-bodies-persons-and-families-2nd

[5] https://www.ica.org/en/egad-ric-conceptual-model-ric-cm-01pdf

[6] https://op.europa.eu/en/

[7] https://eur-lex.europa.eu/browse/eurovoc.html

[8] https://www.w3.org/TR/sparql11-protocol/

[9] John Domingue, Dieter Fensel, James A. Hendler "Handbook of Semantic Web Technologies", Springer Verlag, Heidelberg, Germany, 2011

[10] http://www.europeana.eu

[11] https://www.rijksmuseum.nl/en

[12] https://www.nationalarchives.gov.uk/

[13] https://ec.europa.eu/isa2/actions/facilitating-archive-management-across-europe_en

[14] https://www.w3.org/RDF/

[15] https://wiki.dbpedia.org/about

[16] https://www.wikipedia.org/

[17] https://www.w3.org/TR/sparql11-query/

[18] http://graphdb.ontotext.com/documentation/

[19] https://www.l2f.inesc-id.pt/~joana/prc/artigos/07a%20An%20overview%20of%20OntoClean%20-%20Guarino,%20Welty%20-%202004.pdf

[20] https://pro.europeana.eu/resources/standardization-tools/edm-documentation

[21] https://www.ica.org/en/ric-o-extended-call-for-reviewers

[22] https://www.ica.org/standards/RiC/ontology.html

[23] https://ontotext.com/documents/proton/Proton-Ver3.0B.pdf

[24] https://op.europa.eu/en/web/eu-vocabularies/eli

[25] http://www.loa-cnr.it/ontologies/IOLite.owl#

[26] http://www.w3.org/2001/XMLSchema#

[27] https://www.w3.org/2004/02/skos/

[28] https://www.dublincore.org/specifications/dublin-core/dces/

[29] http://purl.org/dc/terms/

[30] https://wiki.dbpedia.org/services-resources/ontology

[31] Mariana Damova, Dana Dannells."Reason-able View of Linked Data for Cultural Heritage" in Proceedings of S3T'2011, Burgas, Bulgaria, September 2011, pp. ,Advances in Intelligent and Soft Computing, ISSN: 1867-5662, Springer Verlag, Heidelberg, Germany, 2011.

[32] https://www.w3.org/TR/owl2-primer/

[33] https://www.w3.org/TR/rdf-schema/

Author Biography

Dr. Mariana Damova is the CEO of Mozaika Ltd. Her background is in natural language processing, Semantic Web Technologies and AI, with strong academic and industrial record in North America and Europe. She has successfully lead international interdisciplinary teams and content management projects carrying technological risk. Mariana holds a PhD from the University of Stuttgart and teaches Semantic Web at the New Bulgarian University and at Sofia State University. She regularly reviews for ACM ComputingReviews.com.