A linked open data architecture for contemporary historical archives

Alexandre Rademaker[1], Suemi Higuchi[2], and Dário Augusto Borges Oliveira[2]

[1] IBM Research and FGV/EMAp
[2] FGV/CPDOC

Abstract. This paper presents an architecture for historical archives maintenance based on Open Linked Data technologies and open source distributed development model and tools. The proposed architecture is being implemented for the archives of the Center for Teaching and Research in the Social Sciences and Contemporary History of Brazil (CPDOC) from Getulio Vargas Foundation (FGV).

1 Introduction

The Center for Teaching and Research in the Social Sciences and Contemporary History of Brazil (CPDOC) was created in 1973 and became an important historical research institute, housing a major collection of personal archives, oral histories and audiovisual sources that document the country memory.

CPDOC is a vibrant and diverse intellectual community of scholars, technicians and students, and has placed increasing emphasis on applied research in recent years, working in collaborative projects with other schools and institutes, aiming at extending the availability and scope of the valuable records it holds. It is part of Getulio Vargas Foundation (FGV), a prestigious Brazilian research and higher education institution founded in 1944, considered by Foreign Policy Magazine to be a top-5 "policymaker think-tank" worldwide [24].

Thanks to the donation of the personal archives of prominent Brazilian figures from the 1930s onward, such President Getulio Vargas himself, CPDOC started to develop its own methodology for organizing and indexing documents. By the end of the 1970s, it was already recognized as a reference among research and historic documentation centers. In 1975, the institute launched its Oral History Program (PHO), which involved the execution and recording of interviews with people who participated in major events in Brazilian history. In 1984, CPDOC published the Brazilian Historical-Biographical Dictionary (DHBB) [1], a regularly updated reference resource that documents the contemporary history of the country. In the late 1990s, CPDOC was recognized as center of excellence by the Support Program for Centers of Excellence (Pronex) of the Brazilian Ministry of Science and Technology.

This year, celebrating 40 years of existence, CPDOC received the support of the Brazilian Ministry of Culture (MinC), which provided a fund of R$ 2.7 million to finance the project "Dissemination and Preservation of Historical Documents". This project has the following main goals: (1) digitizing a significant

amount of textual, iconographic and audiovisual documents; (2) updating the dictionary DHBB; and (3) prospecting innovative technologies that enable new uses for CPDOCs collections.

The advances in technology offer new modes of dealing with digital contents and CPDOC is working to make all data available in a more intelligent/semantic way in the near future, offering swift access to its archives. In collaboration with the FGV School of Applied Mathematics (EMAp), CPDOC is working on a project that aims to enhance access to documents and historical records by means of data-mining tools, semantic technologies and signal processing. At the moment, two applications are being explored: (1) face detection and identification in photographs, and (2) voice recognition in the sound and audiovisual archives of oral history interviews. Soon it will be easier to identify people in the historical images, and link them to the entries in CPDOC archives. Additionally, voice recognition will help locate specific words and phrases in audiovisual sources based on their alignment with transcription – a tool that is well-developed for English recordings but not for Portuguese. Both processes are based on machine learning and natural language processing, since the computer must be taught to recognize and identify faces and words.

CPDOC also wants its data to constitute a large knowledge base, accessible using the standards of semantic computing. Despite having become a reference in the field of organization of collections, CPDOC currently do not adopt any metadata standards nor any open data model for them. Trends for data sharing and interoperability of digital collections pose a challenge to the institution to remain innovative in its mission of efficiently providing historical data. It is time to adjust CPDOC's methodology to new paradigms.

In Brazilian scenario many public data is available for free, but very few are in open format following the semantic web accepted standards. Examples in this direction are the Governo Aberto SP [12], the LeXML [22] and the SNIIC project [3].

In this sense, we present hereby a research project that reflects a change in the way CPDOC deals with archives maintenance and diffusion. The project is an ongoing initiative to build a model of data organization and storage that ensures easy access, interoperability and reuse by service providers. The project proposal is inspired by: (1) Open Linked Data Initiative principles [20]; (2) distributed open source development model and tools for easy and collaborative data maintenance; (3) a growing importance of data curating concepts and practices for online digital archives management and long-term preservation.

The project started with an initiative of creating a linked open data version of CPDOC's archives and a prototype with a simple and intuitive web interface for browsing and searching the archives was developed. The uses of Linked Open Data concept are conformed to the three laws first published by David Eaves [13] and now widely accepted: (1) If it can't be spidered or indexed, it doesn't exist;

[3] Sistema Nacional de Informaes e Indicadores Sociais, `http://culturadigital.br/sniic/`.

(2) If it isn't available in open and machine readable format, it can't engage; and
(3) If a legal framework doesn't allow it to be repurposed, it doesn't empower.

Among the project objectives we emphasize the construction of a RDF [23] data from data originally stored in a relational database and the construction of an OWL [17] ontology to properly represent the CPDOC domain. The project also aims to make the whole RDF data available for download similarly to what DBpedia does [6].

This paper reflects a real effort grounded in research experience to keep CPDOC as a reference institution in the field of historic preservation and documentation in Brazil.

2 CPDOC information systems

Figure 1 presents the current CPDOC database architecture. The archives are maintained in three different information systems that share a common relational database. Each of the systems has independent management and adopts idiosyncratic criteria concerning the organization and indexing of the information, which vary depending on the specifications of the content they host: personal archives documents, oral history interviews and the Brazilian Historical – Biographic Dictionary entries. CPDOC's web portal provides a query interface to archives data. In the following subsections we briefly describe each of the systems.

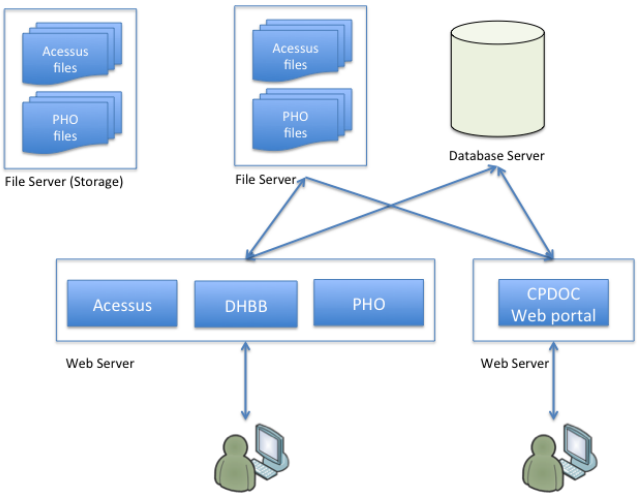

Fig. 1. CPDOC's current architecture

2.1 Personal Archives (Acessus)

This system is composed by personal files from people who influenced the political and social scenario of our country. These historical documents, in textual or audiovisual form, in form of handwritten and printed texts, diaries, letters, photographs, speeches or memos, represent much more than private memories: they are the registry of a whole collective memory.

Currently, more than 200 personal archives from presidents, ministers, military personal and other Brazil's important public figures compose the CPDOC's collections. Together, they comprise nearly 1.8 million documents or 5.2 millions pages. From this, nearly 700 thousands pages are in digital format and the expectance is to digitize all collections in the next few years. The collection entries metadata are stored in an information system called Acessus. It can be accessed through the institution's intranet for data maintenance or by internet for simple data query. Currently, allowed queries are essentially syntactic, i.e., restricted to keywords searches linked to specific database fields defined in an *ad hoc* manner. For those documents that are already digitized, two digital file versions were generated: one in high resolution aiming long-term preservation and another in low resolution for web delivery. High resolution files are stored in a storage system with disk redundancy and restricted access, while low resolution files are stored in a file server [4] (Figure 1).

2.2 Oral History Interviews (PHO)

The CPDOC collection of Oral History hosts currently more than 6.000 hours of recording, corresponding to nearly 2.000 interviews. More than 90% of it, video or audio, are in digital format. For the time being, two kinds of queries are available in the database: query by subject and query by interviewed. Each interview record holds a brief technical information and a summary with descriptions of the interview themes in the order they appear in the record. Only 10% of the interviews are transcribed, and to access the audio/video content the user is requested to come personally to CPDOC.

Currently, CPDOC is analyzing better ways of making this data available online, considering different aspects such as the best format, use policies, access control and copyrights.

As in the case of Acessus, the database actually stores only the metadata about the interviews, while digitized recorded audios and videos are stored as digital files in the file servers.

2.3 Brazilian Historical-Biographic Dictionary (DHBB)

The Brazilian Historical-Biographic Dictionary (DHBB) is certainly one of the main research sources for contemporary Brazilian politicians and themes. It contains more than 7.500 entries of biographic and thematic nature, i.e., people,

[4] `https://en.wikipedia.org/wiki/File_server`.

institutions, organizations and events records carefully selected using criteria that measure the relevance of those to the political history for the given period. The entries are written evenly, avoiding ideological or personal judgments. CPDOC researchers carefully revise all entries added to ensure the accuracy of the information and a common style criteria.

The DHBB's database stores few metadata concerning each entry, and the query is limited to keywords within the title or text.

3 Current Status

In this section we summarize the main problems identified in CPDOC's current infrastructure and daily working environment.

As described in Section 2, CPDOC's archives are maintained by three different information systems based on traditional relational data models. This infrastructure is hard to maintain, improve and refine, and the information is not found or accessed by standard search engines for two reasons mainly: (1) an entry page does not exist until it is created dynamically by an specific query; (2) users are required to login in order to make queries or access the digital files. Service providers do not access data directly and therefore cannot provide specialized services using it. Users themselves are not able to expand the queries over the collections, being limited to the available user interfaces in the website. Thereupon, data of CPDOC's collections is currently limited to what is called "Deep Web" [3].

The maintenance of current different information systems is very problematic. It is expensive, time demanding and ineffective. Improvements are hard to implement and therefore innovative initiatives are usually postponed. A relational database system is not easily modified, because relational data models must be defined *a priori*, i.e., before the data acquisition's stage. Moreover, changes in the database usually require changes in system interfaces and reports. The whole workflow is expensive, time consuming and demands different professionals with different skills from interface developers to database administrators.

Concerning terminology, CPDOC's collections do not follow any metadata standards, which hinders considerably the interoperability with other digital sources. Besides, the available queries usually face idiosyncratic indexing problems with low rates of recall and precision. These problems are basically linked to the *ad hoc* indexing strategy adopted earlier to define database tables and fields.

Finally, data storage is also an issue. Digitized Acessus's documents and Oral History's interviews are not stored in a single place, but scattered in different file servers. The CPDOC database only stores the metadata and file paths to the file servers, making it very difficult to ensure consistency between files, metadata information and access control policies.

4 The proposal

As discussed in Section 3, relational databases are often hard to maintain and share. Also, the idea of having in-house developed and closed source information systems is being increasingly replaced by the concept of open source systems. In such systems the responsibility of updating and creating new features is not sustained by a single institution but usually by a whole community that share knowledge and interests with associates. In this way the system is kept up-to-date, accessible and improving much faster due to the increased number of contributors. Such systems are usually compatible with standards so as to ensure they can be widely used.

Our objective is to propose the use of modern tools so CPDOC can improve the way they maintain, store and share their rich historical data. The proposal privileges open source systems and a lightweight, shared way of dealing with data. Concretely, we propose the substitution of the three CPDOC systems by the following technologies.

The Acessus data model comprises personal archives that contains one or more series (which can contain also other series in a stratified hierarchy) of digitalized documents or photos. The PHO system data model is basically a set of interviews grouped according to some defined criteria within the context given by funded projects. For instance, a political event could originate a project which involve interviewing many important people taking part on the event.

Therefore, Acessus and PHO systems can be basically understood as systems responsible for maintaining collections of documents organized in a hierarchical structure. In this way, one can assume that any digital repository management system (DRMS) have all the required functionalities. Besides, DRMS usually have desirable features that are not present in Acessus or PHO, such as: (1) data model based on standard vocabularies like Dublin Core [19] and SKOS [25]; (2) long-term data preservation functionalities (tracking and notifications of changes in files); (3) fine-grained access control policies; (4) flexible user interface for basic and advanced queries; (5) compliance with standard protocols for repositories synchronization and interoperability (e.g., OAI-PMH [21]); (6) import and export functionalities using standard file formats and protocols; and more.

In our proposal Acessus and PHO systems data and files are planned to be stored in an open source institutional repository software such as Dspace [5] or Fedora Commons Framework [6]. In this article we assume the adoption of Dspace with no prejudice of theoretical modeling.

The DHBB relational model can be summarized to a couple of tables that store metadata about the dictionary entries (stored in a single text field of a given table). The actual dictionary entries are created and edited in text editors outside the system and imported to it only after being created and revised.

The nature of its data suggests that DHBB entries could be easily maintained as text files using a lightweight human-readable markup syntax. The files would

[5] `http://www.dspace.org/`
[6] `http://www.fedora-commons.org`

be organized in an intuitive directory structure and kept under version control for coordinated and distributed maintenance. The use of text files [7] is justified by a couple of reasons. They are: easy to maintain using any text editor (tool independent) allowing the user to use the prefered text editor; conform to long-term standards by being software and platform independent; easy to be kept under version control by any modern version control system [8] since they are comparable (line by line); and efficient to store information [9].

The use of a version control system will improve the current workflow of DHBB reviewers and coordinators, since presently there is no aid system for this task, basically performed using Microsoft Word text files and emails. The adoption of such tool will allow file exchanges to be recorded and the process controlled without the need of sophisticated workflow systems, following the methodology developed by open sources communities for open source software maintenance. For instance, Git [10] is specially suited to ensure data consistency and keeps track of changes, authorship and provenance.

Many of the ideas here proposed were already implemented as a proof of concept to evaluate the viability of such environment in CPDOC. Figure 2 illustrates the necessary steps to fully implement our proposal. In the following text we briefly describe each step.

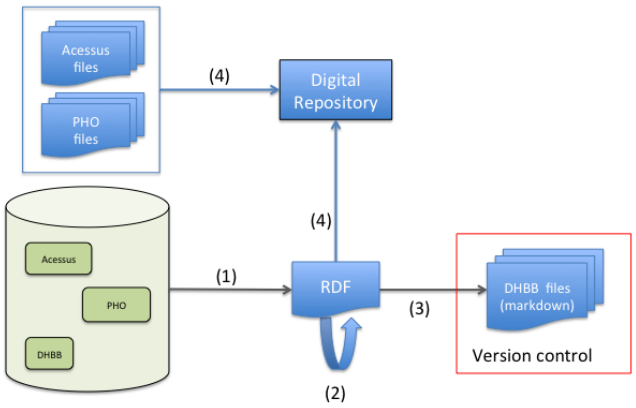

Fig. 2. Migrating from relational databases to the proposed model

Step (1) is implemented and the relational database was exported to RDF [23] using the open source D2RQ [5] tool. The D2RQ mapping language [10] allows the definition of a detailed mapping from the current relational model to a graph

[7] `http://en.wikipedia.org/wiki/Text_file`
[8] `https://en.wikipedia.org/wiki/Revision_control`.
[9] A text file of a DHBB entry has usually 20% the size of a file DOCX (Microsoft Word) for the same entry.
[10] `http://git-scm.com`.

model based on RDF. The mapping from the relational model to RDF model was already defined using the standard translation from relational to RDF model sketched out in [4]. The mapping created so far defers any model improvement to step (2) described below.

Step (2) is planned and represents a refinement of the graph data model produced in step (1). The idea is to produce a data model based on standard vocabularies like Dublin Core [19], SKOS [25], PROV [16] and FOAF [7]. The use of standard vocabularies will make the data interchangeable with other models and facilitate its adoption by service provides and users. It will also help us to better understand the database model and its semantics. In Section **??** we describe the refinement proposal in detail.

Step (3) is already implemented and deploys a text file for each DHBB entry. Each text file holds the entry text and metadata [11]. The files use YAML [2] and Markdown [18] markup languages to describe the metadata and entry content. YAML and Markdown were adopted mainly because both languages are human-readable markups for text files and are supported by almost all static site generators [12]. The use of a static site generator allows DHBB maintainers to have full control over the deployment of a DHBB browsable version.

Note that step (3) was actually implemented to use the initial version of the RDF produced in step (1). The code can be easily adapted to use the final RDF model produced by step (2).

In the planned step (4) the digital files and their metadata will be imported into a DRMS. This step is much more easily implemented using the RDF produced in step (2) than having to access the original database. It is only necessary to decide which repository management system will be adopted.

The proposed workflow architecture is presented in Figure 3. Recall that one of the main goals is to make CPDOC archive collections available as open linked data. This can be accomplished by providing data as RDF/OWL files for download and a SPARQL Endpoint [9] for queries. Since data evolve constantly, CPDOC teams would deliver periodical data releases. Besides the RDF/OWL files and the SPARQL Endpoint, we believe that it is also important to provide a lightweight and flexible web interface for final users to browse and query data. This can be easily done using a static website generator and Apache Solr [13] for advanced queries. As a modern index solution, Solr can provide much powerful and fast queries support when compared to traditional relational database systems. Note that the produced website, the RDF/OWL files and SPARQL Endpoints are complementary outputs and serve to different purpose and users.

Finally, it is vital to stress the main contrast between the new architecture and the current one. In the current CPDOC architecture the data is stored in relational databases and maintained by information systems. This means that any data modification or insertion is available in real time for CPDOC website

[11] It is out of this article scope to present the final format of these files.

[12] In this application we used Jekyll, `http://jekyllrb.com`, but any other static site generator could be used.

[13] `http://lucene.apache.org/solr/`

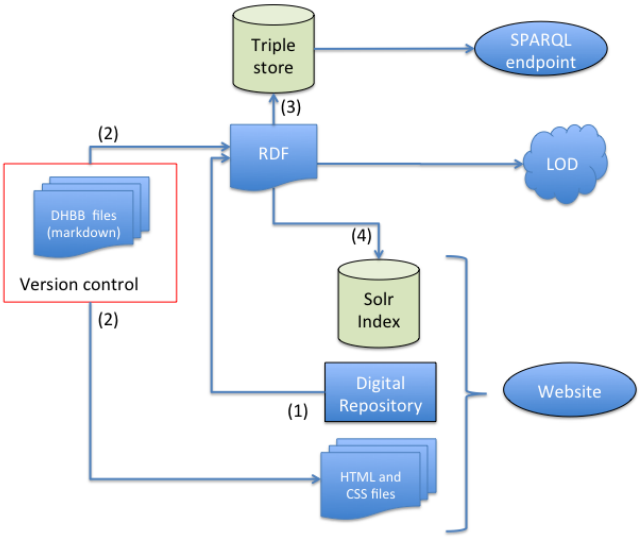

Fig. 3. The final architecture

users. However, this architecture has a lot of drawbacks as mentioned in Section 3, and also the nature of CPDOC data does not require continuous updates, which means that the cost of this synchronous *modus operandi* is not needed. Usually, CPDOC teams work on projects basis and therefore new collections, documents and metadata revisions are not very often released.

The results obtained so far encouraged us to propose a complete data model aligned with open linked data vocabularies, presented in detail in next section.

5 Improving Semantics

More than improving the current infrastructure for storing and accessing CPDOC data, we would like to exploit the semantic possibilities of such rich source of knowledge. One of the ways to do that is to embed knowledge from other sources by creating links within the available data. Since much of the data is related to people and resources with historical relevance, or historical events, some available ontologies and vocabularies can be used in this task.

The personal nature of the data allows us to use projects that are already well developed for describing relationships and bonds between people, such as FOAF [7] (Friend of a Friend) – a vocabulary which uses RDF to describe relationships between people and other people or things. FOAF permits intelligent agents to make sense of the thousands of connections people have with each other, their belongings and historical positions during life. This improves accessibility and generates more knowledge from the available data.

The analysis of structured data can automatically extract connections and, ultimately, knowledge. A good example is the use of PROV [16], which provides a

vocabulary to interchange provenance information. This is interesting to gather information of data that can be structurally hidden in tables or tuples.

The RDF graph model enables also the merging of data content naturally. The DBpedia project, for instance, allows users to query relationships and properties associated with Wikipedia resources, and users can link other datasets to the DBpedia dataset in order to create a big and linked knowledge knowledge base. CPDOC could link their data to DBpedia and then make their own data available for a bigger audience.

In the same direction, the use of lexical databases, such as the WordNet [14] and its Brazilian version OpenWordnet-PT [11], will allow us to make natural language processing of DHBB entries. Named entities recognition and other NLP tasks can automatically create connections that improve dramatically the usability of the content. Other resources like YAGO [28] and BabelNet [26] links Wikipedia to WordNet. The result is an "encyclopedic dictionary" that provides concepts and named entities lexicalized in many languages and connected with large amounts of semantic relations. Finally, the SUMO Ontology [27] could also be used to provide a complete formal definition of terms linked to WordNet. All of these lexical resources and ontologies will be further explored when we start the natural language processing of DHBB entries.

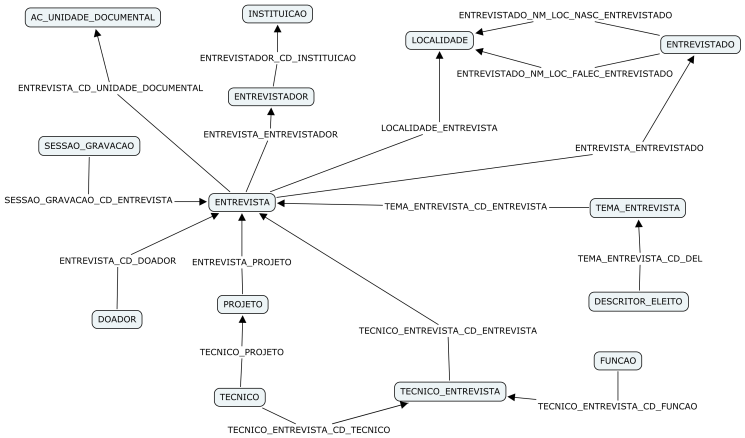

Fig. 4. PHO first RDF model

Figure 4 shows a fragment of the current RDF model produced by D2RQ (in step (1) of Figure 2) using the original CPDOC database relational model. This fragment shows only some PHO classes (derived from the tables) and some properties (derived from the foreign keys). Classes are written inside the boxes and properties are represented by the names in arrows that connect boxes.

The model presented in Figure 4 depicts that D2RQ was not able to automatically improve much further the model. D2RQ was able to correctly translate relations N:M in the relational model, such as `entrevista_entrevistador` (origi-

nally a table in the relational model) to a property that connect directly instances of `entrevista` (interview) with instances of `entrevistador` (interviewer). Nevertheless, the N:M relation between `entrevista` and `tecnico` (technician) was kept as an intermediary class called `tecnico_entrevista` due to the existence of an aditional information in this N:M relation, the role (`funcao` class) of the interview technician. The relational model also seems to have some inconsistences. Although the connection of technician and interview is parameterized by different roles, the donator, interviewer and interviewed of an interview are modeled each one in a specific table. In this case interviewed, interviewer, donator and technician are all people that share a lot of common properties like name, address, etc, and could be modeled as people. These problems are all result of a "ad hoc" modeling process. The model defined this way only makes sense for CPDDOC team and it could hardly be useful outside CPDOC.

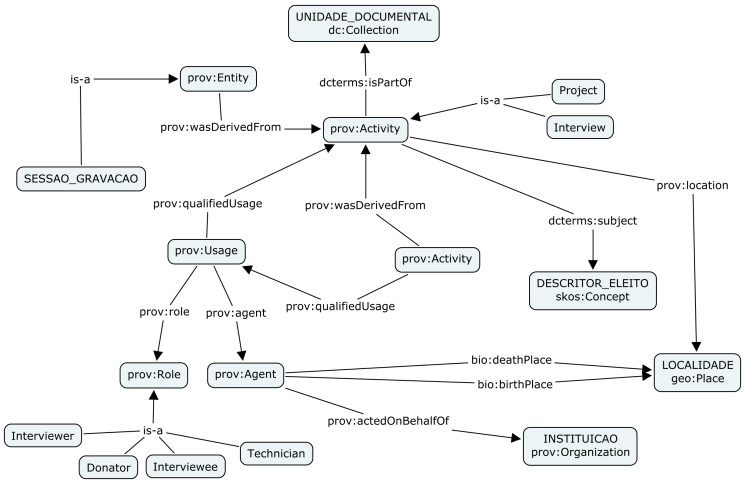

Fig. 5. PHO revised RDF model

Figure 5 shows how PHO model can be refined. The new model uses standard vocabularies and ontologies, making the whole model much more understandable and interoperable. In the Figure 5, `prov:Activity` was duplicated only for a better representation. The prefixes in the names indicate the vocabularies and ontologies used: `prov`, `skos`, `dcterms`, `dc`, `geo`, and `bio`. We also defined a CPDOC ontology that declares its own classes and specific ontology links, such as the one that states that a `foaf:Agent` is a `prov:Agent`. In this model, we see that some classes can be subclasses of standard classes (e.g. `Interview`), while some classes can be replaced by standard classes (e.g. `LOCALIDADE`).

6 Conclusion

In this paper we presented a new architecture for CPDOC archives creation and maintenance. It is based on open linked data concepts and open source methodologies and tools. We believe that even though CPDOC users would need to be trained to use the proposed tools such as text editors, version control software and command line scripts; this architecture would give more control and easiness for data maintenance. Moreover, the architecture allows knowledge to be easily merged to collections data without the dependency of database refactoring. This means that CPDOC team will be much less dependent from FGV's Information Management and Software Development Staff.

Many proposals of research concerning the use of lexical resources for reasoning in Portuguese using the data available in CPDOC are being carried out so as to improve the structure and quality of the DHBB entries. Moreover, the automatic extension of the mapping proposed in Section 5 can be defined following ideas of [8]. Due the lack of space, we do not present them in this paper.

Finally, we aim to engage a wider community and an open-source development process in order to make the project sustainable. As suggested by one of the reviewers, we must also learn from experiences of projects like Europeana [14] and German National Digital Library [15].

References

1. Alzira Alves Abreu, Fernando Lattman-Weltman, and Christiane Jalles de Paula. *Dicionário Histórico-Biográfico Brasileiro pós-1930*. CPDOC/FGV, 3 edition, February 2010.
2. Oren Ben-Kiki, Clark Evans, and Ingy dot Net. Yaml: Yaml ain't markup language. `http://www.yaml.org/spec/1.2/spec.html`.
3. Michael K Bergman. White paper: the deep web: surfacing hidden value. *journal of electronic publishing*, 7(1), 2001.
4. Tim Berners-Lee. Relational databases on the semantic web. Technical report, W3C, 1998. `http://www.w3.org/DesignIssues/RDB-RDF.html`.
5. Christian Bizer and Richard Cyganiak. D2R server-publishing relational databases on the semantic web. In *5th international Semantic Web conference*, page 26, 2006. `http://d2rq.org`.
6. Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. Dbpedia - a crystallization point for the web of data. *Web Semantics*, 7(3):154–165, 2009.
7. Dan Brickley and Libby Miller. Foaf vocabulary specification. `http://xmlns.com/foaf/spec/`, 2010.
8. Isabel Cafezeiro, Edward Hermann Haeusler, and Alexandre Rademaker. Ontology and context. In *IEEE International Conference on Pervasive Computing and Communications*, Los Alamitos, CA, USA, 2008. IEEE Computer Society.
9. Kendall Grant Clark, Lee Feigenbaum, and Elias Torres. SPARQL protocol for RDF. Technical report, W3C, 2008.

[14] `http://www.europeana.eu`.

10. Richard Cyganiak, Chris Bizer, Jorg Garbers, Oliver Maresch, and Christian Becker. The D2RQ mapping language. http://d2rq.org/d2rq-language.
11. Valeria de Paiva, Alexandre Rademaker, and Gerard de Melo. Openwordnet-pt: An open brazilian wordnet for reasoning. In *Proceedings of the 24th International Conference on Computational Linguistics*, 2012.
12. Governo do Estado de São Paulo. Governo aberto sp. http://www.governoaberto.sp.gov.br, 2013.
13. David Eaves. The three law of open government data. http://eaves.ca/2009/09/30/three-law-of-open-government-data/.
14. Christiane Fellbaum, editor. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA, 1998.
15. Natalja Friesen, Hermann Josef Hill, Dennis Wegener, Martin Doerr, and Kai Stalmann. Semantic-based retrieval of cultural heritage multimedia objects. *International Journal of Semantic Computing*, 06(03):315–327, 2012.
16. Yolanda Gil and Simon Miles. Prov model primer. http://www.w3.org/TR/2013/NOTE-prov-primer-20130430/, 2013.
17. W3C OWL Working Group, editor. *OWL 2 Web Ontology Language Document Overview*. W3C Recommendation. World Wide Web Consortium, 2 edition, 2012.
18. John Gruber. Markdown language. http://daringfireball.net/projects/markdown/.
19. Dublin Core Initiative. Dublin core metadata element set. http://dublincore.org/documents/dces/, 2012.
20. Open Data Initiative. Open data initiative. http://www.opendatainitiative.org, 2013.
21. Carl Lagoze, Herbert Van de Sompel, Michael Nelson, and Simeon Warner. The open archives initiative protocol for metadata harvesting. http://www.openarchives.org/OAI/openarchivesprotocol.html, 2008.
22. LexML. Rede de informação informativa e jurídica. http://www.lexml.gov.br, 2013.
23. Frank Manola and Eric Miller, editors. *RDF Primer*. W3C Recommendation. World Wide Web Consortium, February 2004.
24. James McGann. *The Think Tank Index*. Foreign Policy, February 2009.
25. Alistair Miles and Sean Bechhofer. Skos simple knowledge organization system reference. http://www.w3.org/2004/02/skos/, 2009.
26. Roberto Navigli and Simone Paolo Ponzetto. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250, 2012.
27. Ian Niles and Adam Pease. Towards a standard upper ontology. In *Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001*, pages 2–9. ACM, 2001.
28. Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: A Core of Semantic Knowledge. In *16th international World Wide Web conference (WWW 2007)*, New York, NY, USA, 2007. ACM Press.