

The Bibliography of the Italian Parliament goes linked open data¹

Lucia Panciera

Library of the Italian Chamber of deputies, <http://biblioteca.camera.it>
(panciera_l@camera.it)

Visit the new website in beta at <http://bprdemo.camera.it/>
(username: webtester password: THX:1138)

Abstract

This article presents a case study of a project, led by the Italian Chamber of deputies, to transform the *Bibliography of the Italian Parliament and Electoral Studies* into linked open data. The project involved migrating the previous database to a new platform which deals with native RDF format, creating a new faceted search interface, and publishing the dataset on the official open data portal. The linking to other datasets is underway, and some hypotheses are made about the future steps.

1. Introduction

The *Bibliography of the Italian Parliament and Electoral Studies* (BPR) is a database created and maintained by the Library of the Italian Chamber of deputies. It contains bibliographic references relating to the Italian Parliament, the post-1945 National Consultative Assembly and Constituent Assembly, and the general elections, since 1848.

During the project here described, which is being realized jointly by the Library and the Information technology Department, the original database was migrated to a new platform based on linked open data principles, and a new search interface was created. Data are now available for searching and download on the official open data portal of the Chamber of deputies, and the focus is currently on establishing links to other datasets, such as the parliamentary biographical database, the Virtual International Authority File (VIAF), and the Italian Nuovo soggettario.

The project is part of the more general adoption of linked open data at the Chamber of deputies and gave the Library an interesting opportunity to learn more about these principles and to get practical experience in their application to a relatively small but significant portion of its activity, making it more widely available, in the meantime.

2. The BPR

The BPR currently includes more than 20,000 bibliographic references to books and articles

¹ Text version of the poster presented at the international conference “[Faster, smarter and richer. Reshaping the library catalogue](#)”, Rome, 27-28 February 2014. Last updated on 03/06/2014.

published in periodicals, miscellaneous volumes, encyclopedias and selected websites.

It encompasses mainly juridical essays, but also research on political and organization science, on political sociology, as well as relevant historiographical literature.

The database is updated three times a year and can be consulted at <http://bpr.camera.it> and from within the historical portal of the Chamber of deputies (<http://storia.camera.it/bpr#nav>).

The bibliography aims to be exhaustive, especially for its extensive coverage of Italian publications; references to works written in other languages amount to about 3% of the total. Another main strength resides in the number of analytical entries from miscellaneous volumes (more than 25%).

The BPR is also a digital library of the Italian Parliament: it offers nearly 3,000 full texts, coming from books or articles in the public domain, or under copyright of the Chamber of deputies. In some other cases, copyright agreements have been reached with publishers to make texts available on the Internet. Many texts have been added with ad hoc digitization programs.

Each reference is assigned one or more classification codes, which are taken from a directory of over 100 classifications that are organized into seven major fields: General topics (A), Institutional relations (B), Sources of law and organizational aspects (C), The lawmaking function (D), Parliamentary oversight, policy-setting and information (E), Members of Parliament (legal and social status) (F), Elections (H).

3. Linked open data at the Chamber of deputies

The term ‘linked open data’ refers to a set of best practices, introduced by Tim Berners Lee [1], for publishing and interlinking structured data on the Web [2]. In recent years these practices have been widely adopted by libraries, museums and cultural institutions, as they can help increase the presence on the web of these communities and enhance discovery and reuse of collections, metadata element sets and vocabularies [3, 4, 5, 6].

Since 2010 the Chamber of deputies has begun to use standards and technologies related to linked open data and the semantic Web, for the purpose of innovation and transparency. On the occasion of the 150th anniversary of the Italian unification, the Chamber has undertaken a strategic and extensible project based on linked open data, integrating and organizing digital documents and information on the activity of the Chamber from multiple sources: websites and databases dedicated to legislatures of the Kingdom of Italy and the Italian Republic, archival materials, pictures and videos. Data about persons, activities and official documents of the

Chamber of deputies have been converted to RDF, and described in a specific ontology (OCD) integrated with more general ones. The project resulted in the launch of two Web portals in 2011 [7, 8]. <http://storia.camera.it> is the access point to the wealth of information about the Chamber of deputies since 1848; linked open data are published on <http://dati.camera.it>, so that they can be queried, downloaded, reused and repurposed. The portal includes a SPARQL endpoint that can return results in HTML, Spreadsheet, XML, JSON, Javascript, NTriples, RDF/XML e CSV.

The linked open data approach has also been applied to internal document management, in the implementation of an open documentation management system named PAD (Piattaforma aperta per la documentazione digitale = Open platform for digital documents). It has been entirely developed by open source technology and it is based on one information base that deals with digital documents in native RDF format and can publish data in real time on the SPARQL endpoint. PAD is progressively becoming the shared workspace for documentation activities, from information gathering and indexing to collaborative writing and publishing of dossiers and other documents. This approach ensures data persistence and allows for reuse, both internal and external, and different data representations. The BPR linked open data represents its most recent achievement.

4. The restructuring of the BPR database

After more than ten years since the setup of the original database, there was an urgent need to revise and update both the back end software and the end user search interface. In particular, the main objectives of the project were:

- to substitute the original Access database and to add more management features: in particular, to increase the reuse of data, in order to avoid repetitive tasks, and to strengthen controls over consistency and correctness;
- to redesign and update the end user interface, making it more agile and including faceted search;
- to increase the external projection of the BPR, permitting data harvesting and general data reuse.

The project was therefore included in the more general adoption of linked open data at the Chamber of deputies, resulting in the choice of the PAD system to manage citations and in the publication of data.

The whole process can be described as an iterative incremental life cycle model, like the one

proposed by B. Villazón-Terrazas and O. Corcho [9]; it is structured in five main activities and various sub activities, that are not necessarily performed in a fixed order but can be mixed and repeated as required.

1. Specification (analysis of the data sources, URI design, definition of the license): data in the Access database were analyzed in order to find peculiarities and select the more appropriate handling. As the original database had been created to facilitate the publication of the print edition, some fields contained mixed information, which needed to be disaggregated. All data on the portal are licensed under Creative Commons Attribution-ShareAlike 3.0. Version 4 is under consideration due to its coverage of sui generis database rights.

2. Modelling: the ontology was constructed paying special attention to the reuse of existing ones: Dublin Core Ontology, IFLA International Standard Bibliographic Description (ISBD) Ontology, Bibliographic Ontology (BIBO), the Ontology of the Chamber of deputies; the classification, that has a simple hierarchical structure, was represented using the SKOS vocabulary (*Fig. 1, 2*). The issue of FRBRizing the model will be addressed in a subsequent stage; for the moment, references to other versions of the same work are given in a note

3. Generation (transformation, data cleansing, linking): data from the original database were transformed to RDF; disaggregation of data was carried out using mapping rules based on ISBD punctuation. Several test migrations were run in order to refine rules and identify and correct inconsistencies and mistakes, resulting in a general quality improvement of data. This phase also involved the customization of the PAD platform functionalities of input, searching and displaying. As for linking, see § 6.

4. Publication: data are now available on the official open data portal of the Chamber of deputies (<http://dati.camera.it>) both for searching through the SPARQL endpoint and for download in the RDF/XML serialization. The new BPR website permits direct download of single records. In order to facilitate discovery, the dataset will be registered in the main specialized portals and search engine (<http://data.gov.it>, <http://publicdata.eu>, etc.).

5. Exploitation: this is still an open issue, as in many LOD projects; in order to foster consumption, the SPARQL endpoint will suggest various query examples; we would like to engage students and academics in the elaboration and visualization of data and we are also considering submitting our data to CulturaItalia and Europeana.

Italians do it better? How governability challenged representation in Italy's electoral debates (1946-2012) at OCD - Ontologia della Camera dei deputati

<http://dati.intra.camera.it/ocd/bpr/articolo.rdf/22710>



Property	Value
isbd:P1004	■ Italians do it better? How governability challenged representation in Italy's electoral debates (1946-2012)
dcterms:bibliographicCitation	■ Representation. - 48 (2012), n. 2, p. 151-168
dc:coverage	■ Periodo repubblicano
dcterms:creator	■ < http://dati.intra.camera.it/ocd/bpr/autore.rdf/9779 >
dcterms:isPartOf	■ < http://dati.intra.camera.it/ocd/bpr/periodico.rdf/e159 >
bibo:issue	■ 2
dcterms:issued	■ 2012
dc:language	■ en
bibo:pages	■ 151-168
dcterms:provenance	■ < http://dati.intra.camera.it/ocd/bpr/bpr >
ocd:rif_file	■ < http://dati.intra.camera.it/ocd/bpr/file.rdf/22710_3309_t >
dcterms:subject	■ < http://dati.intra.camera.it/ocd/bpr/ccd.rdf/C0000000064 > ■ < http://dati.intra.camera.it/ocd/bpr/ccd.rdf/C0000000075 >
dc:title	■ Italians do it better? How governability challenged representation in Italy's electoral debates (1946-2012)
dc:type	■ articolo
rdf:type	■ < http://purl.org/ontology/bibo/Article >

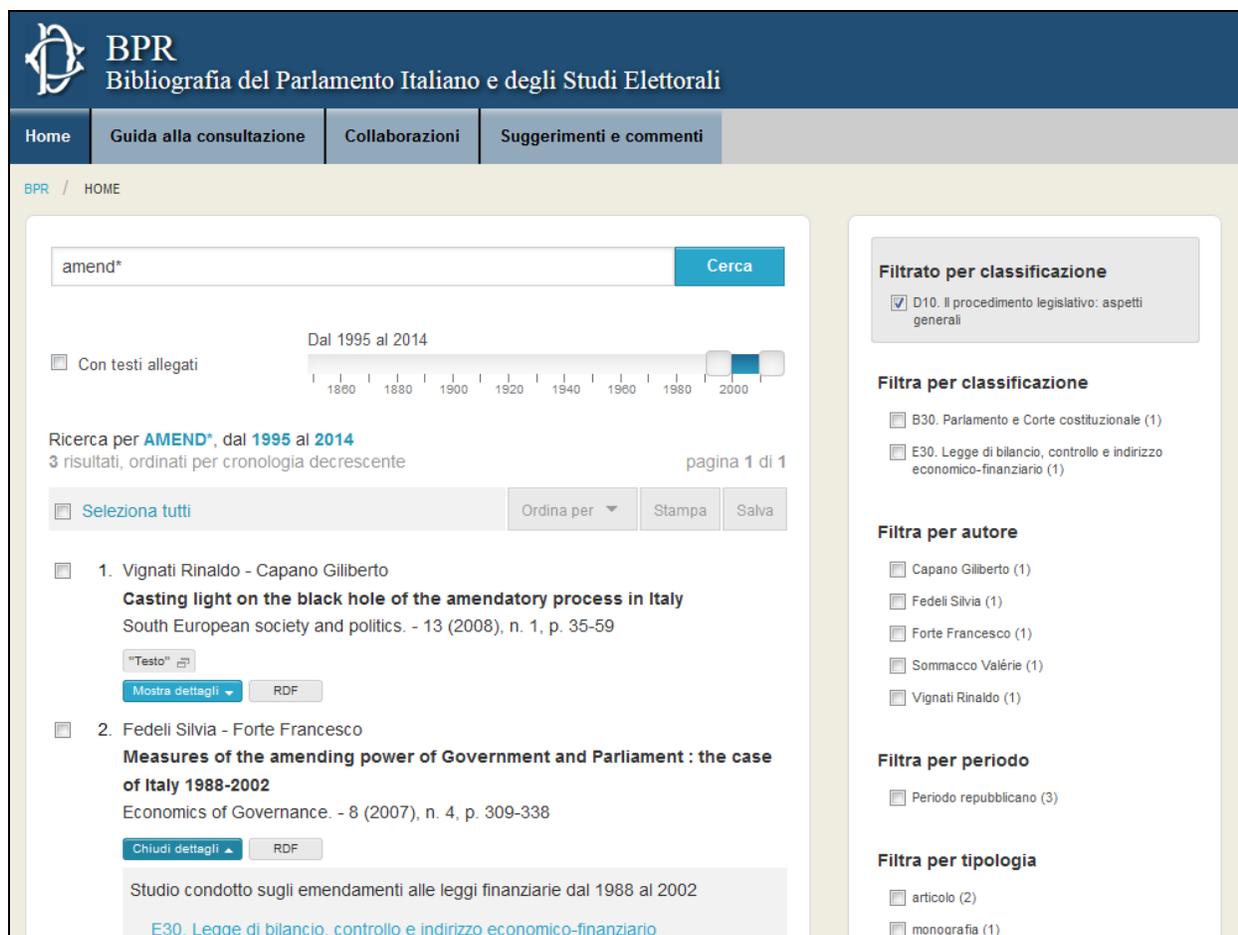
Fig. 1 – Properties of a bibliographic citation

Property	Value
skos:broader	■ < http://dati.intra.camera.it/ocd/bpr/ccd.rdf/C0000000056 >
skos:definition	■ La funzione legislativa in generale, la sede referente, la discussione degli emendamenti in generale. Comprende la continuità legislativa e la decadenza dei progetti di legge. Non comprende studi generali sulle fonti del diritto. Non comprende il potere presidenziale di rinvio delle leggi (in B10) e il tema dell'interferenza tra attività legislativa ed attività della Corte costituzionale (in B30). Per la discussione in generale e le questioni pregiudiziali vedi anche C14. Per l'ostruzionismo vedi C16. Per le audizioni interne al procedimento legislativo vedi anche E60
skos:in Scheme	■ < http://dati.intra.camera.it/ocd/bpr/ccd.rdf/C0000000000 >
rdfs:label	■ Il procedimento legislativo: aspetti generali
skos:notation	■ D10
dcterms:references	■ 70, 72

Fig. 2 – Properties of the classification

5. The new consultation website

The interface was completely redesigned, based mainly on faceted search. It now features a single search box, a date control bar and narrowing filters by classification codes, author, language, typology of documents, historical period. The classification is browsable and can be used to trigger subject searches. Search can be limited to include only citations with a full-text attached. Results can be ordered by different criteria and can be printed or downloaded (Fig. 3).



The screenshot displays the BPR (Bibliografia del Parlamento Italiano e degli Studi Elettorali) search interface. At the top, there is a navigation bar with links for Home, Guida alla consultazione, Collaborazioni, and Suggerimenti e commenti. Below this, a search box contains the query 'amend*' and a 'Cerca' button. A date range filter is set to 'Dal 1995 al 2014'. There are checkboxes for 'Con testi allegati' and 'Seleziona tutti'. The search results are displayed in a list format, with the first result being '1. Vignati Rinaldo - Capano Gilberto' and the second being '2. Fedeli Silvia - Forte Francesco'. The interface also features several faceted filters on the right side, including 'Filtrato per classificazione', 'Filtra per classificazione', 'Filtra per autore', 'Filtra per periodo', and 'Filtra per tipologia'. The search results are ordered by 'Ordina per' and include options for 'Stampa' and 'Salva'.

Fig. 3. - The new search interface

6. Linking to other datasets

The linking phase is still in implementation but several datasets were considered to be suitable as linking targets: firstly, the parliamentary biographical datasets already published by the Chamber of deputies, secondly, external datasets such as VIAF for name authorities, and the Italian Nuovo soggettario for classification, that are already important nodes in the LOD cloud and may serve as a bridge towards several other linked library data (catalog and authority records) but also more

References

- [1] T. BERNERS LEE, Linked Data - Design Issues, 2006. Retrieved from <http://www.w3.org/DesignIssues/LinkedData.html>
- [2] T. HEATH - C. BIZIER, Linked Data: Evolving the web into a global data space, Morgan & Claypool, 2011. Retrieved from <http://linkeddatabook.com/editions/1.0/>;
- [3] Library Linked Data Incubator Group Final Report, October 2011. Retrieved from <http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>
- [4] G. DUNSIRE, Linked Data, libraries and the Semantic Web, Library science talk, 2012. Retrieved from <http://www.gordondunsire.com/pubs/pres/linkeddatalibsemweb.ppt>
- [5] R. WALLIS, Library linked data progress, communication at the IFLA Conference, 2012. Retrieved from <http://www.slideshare.net/rjw/library-linked-data-progress>
- [6] J. EDELSTEIN, L. GALLA, C. LI-MADEO, J. MARDEN, A. RHONEMUS, N. WHYSEL, Linked Open Data for Cultural Heritage: Evolution of an Information Technology, Proceedings of the 31st ACM international conference on Design of communication, 2013, p. 107-112. Retrieved from <http://www.whysel.com/papers/LIS670-Linked-Open-Data-for-Cultural-Heritage.pdf>
- [7] E. CANDIA, Beyond Open Document towards Open Linked Data, communication at the international meeting “Achieving Greater Transparency in Legislatures through the Use of Open Document Standards”, Washington, February 2012. Retrieved from <http://dati.camera.it/it/video/World-e-Parliament-Conference-2012-session-A4.html>
- [8] G. BRUNO, La Camera dei deputati nel Web of Data : i Linked Open Data per la trasparenza ed efficienza nella gestione della documentazione della Camera (20-09-2012). Retrieved from <http://www.regesta.com/2012/09/20/la-camera-dei-deputati-nel-web-of-data/>
- [9] B. VILLAZÓN-TERRAZAS, O. CORCHO, Methodological Guidelines for Publishing Linked Data. Retrieved from http://delicias.dia.fi.upm.es/wiki/images/7/7a/07_MGLD.pdf
- [10] G. BERGAMIN, A. LUCARELLI, The Nuovo soggettario as a service for the linked data world, *JLIS*, Vol. 4, No. 1 (2013). Retrieved from <http://leo.cilea.it/index.php/jlis/issue/view/536>
- [11] M. KLEIN, A. KYRIOS, VIAFbot and the Integration of Library Data on Wikipedia, *Code4Lib Journal*, 22 (14 October 2013). Retrieved from <http://journal.code4lib.org/articles/8964>
- [12] T. AGNOLONI, G. PERUGINELLI, M.T. SAGRI, D. TISCORNIA, Annotation schema for legal doctrine: a case study on DoGi database, *JLIS*, Vol. 4, No. 1 (2013). Retrieved from <http://leo.cilea.it/index.php/jlis/article/view/5480>

Project team members:

From the Library of the Italian Chamber of deputies:

Fernando Venturini, *supervisor*, venturini_f@camera.it

Lucia Panciera, panciera_l@camera.it

Stefano Tabacchi, tabacchi_s@camera.it

From the Information technology Department of the Italian Chamber of deputies:

Elena Candia, *supervisor*, candia_e@camera.it

Danilo Narduzzi, narduzzi_d@camera.it

From Regesta.exe:

Diego Valerio Camarda, dcamarda@regesta.com

Silvia Mazzini, smazzini@regesta.com

