

University for Business and Technology in Kosovo

UBT Knowledge Center

UBT International Conference

2014 UBT International Conference

Nov 8th, 2:45 PM - 3:00 PM

Best Practices of Consuming Linked Open Data

Arbër Beshiri

South East European University, arber.beshiri@gmail.com

Joana Mita

South East European University

Anja Vangieli

South East European University

Follow this and additional works at: <https://knowledgecenter.ubt-uni.net/conference>



Part of the [Computer Sciences Commons](#)

Recommended Citation

Beshiri, Arbër; Mita, Joana; and Vangieli, Anja, "Best Practices of Consuming Linked Open Data" (2014). *UBT International Conference*. 53.

<https://knowledgecenter.ubt-uni.net/conference/2014/all-events/53>

This Event is brought to you for free and open access by the Publication and Journals at UBT Knowledge Center. It has been accepted for inclusion in UBT International Conference by an authorized administrator of UBT Knowledge Center. For more information, please contact knowledge.center@ubt-uni.net.

Best Practices of Consuming Linked Open Data

Arbër Beshiri¹, Joana Mita², Anja Vangjeli³

Department of Software Engineering and Telecommunication, Faculty of Contemporary Sciences
and Technologies, South East European University

arber.beshiri@gmail.com¹

Abstract. The term Linked Data is defined as a set of best practices for publishing and interlinking structured data on the web. These best practices were introduced by Tim Berners-Lee and are also known as principles. These best practices are used by the vast majority of data providers leading to the establishment of a global data space known as the web of data. In this paper will analyze and explore the technical principles of Linked Data, the best practices of using Linked Data, some deployed Linked Data applications and use cases to exploit the Web of Data.

Keywords: Linked Data, Resource Description Framework, SPARQL, Web of Data, Interlinking Data

1 Introduction

“Linked Data is defined as a set of best practices for publishing and joining structured data on the web [22]”. These best practices were introduced by Tim Berners-Lee and are used by the vast majority of data providers, leading to the establishment of a global data space known as the Web of Data. In recent times, the web is developed from a global information space of linked documents to a space where both documents and data are linked. The use of Linked Data best practices has taken us to a wider web with a space for global data deriving from different sources such as people, companies, films and music. In addition, Linked Data will enable the users to start browsing in one data source and then follow the links related to this data source. Linked Data can implement generic applications which operate over the complete data space. This is possible because of the use of Web standards and a common data model. Furthermore, the global space of data will allow for new types of applications to be enabled for example the Linked Data browsers and the Linked Data search engines. Linked Data browsers will permit users to browse in one data source and navigate along links into related data sources while the Linked Data search engines will search the global space of data by following links between data sources and make available query abilities over gathered data. The difference between Web 2.0 mashups and Linked Data is that the first works on a fixed set of data sources and finally operates on a global data space. Linked Data uses the Web to create typed links between data from different sources. The data published on the web is readable machine, with a clear meaning and it is related to other external data sets which might be linked to it. The web of data is created based on the contribution of a group of people and is useful in integrating and reusing data retrieved from different sources. In the nearby future, some of the more redefined proposals connected with the semantic web vision, may be implemented and become real [1], [22].

2 Technical Principles of Linked Data

The best practices introduced by Tim Berners-Lee in his web architecture note linked data are also known as principles. These principles are:

- Use URIs as names for things.

- Use HTTP URIs, so that people can look up those names.
- When someone looks up a URI, provide useful information, using the standards RDF or SPARQL.
- Include links to other URIs, so that we can discover more things.

The idea behind linked data is to take into account the general architecture of the World Wide Web (WWW) when sharing a structured data on the global space of data. In order to comprehend the linked data principles, it is fundamental first to understand the architecture of the World Wide Web as well as the classic document web. The document web is built on a small set of simple standards: Uniform Resource Identifiers (URIs) as globally unique identification mechanism, the Hypertext Transfer Protocol (HTTP) as universal access mechanism, and the Hypertext Markup Language (HTML) as a widely used content format. The web is constructed by setting hyperlinks between web documents that may be a feature of the different web servers. Hyperlinks allow the users to navigate many different servers as well as search the web. Hyperlinks provide sophisticated crawl capabilities on top of searched data. This shows that hyperlinks are essential in linking content from the variety of servers into one global data space. The web architecture has been very successful for the last 20 years because it has brought together simplicity with decentralization and openness. Linked data uses the web architecture to share data on the global scale [22]. The member countries of Open Government Partnership have discovered that linked data is a powerful tool for linking and sharing data through URIs on the web. Linked data is used by the public as well as the private sector as a framework for creating information and telecommunication strategies. The three following goals should be taken into account when sharing data online are:

- to increase public understanding and teamwork,
- data and information about the world should be made accessible and available and
- to boost the competence of public and private sectors.

All of these objectives are provided through Linked Data technology. Linked Data is accessible and open. To comply with the Linked Data best practices, a variety of Linked Data publishing tools have been developed. These tools are important as either the tools serve the content of RDF stores as Linked Data on the web or offer Linked Data views over non-RDF data sources. “The tools shield publishers from dealing with technical details such as content negotiation and ensure that data is published according to the Linked Data community best practices [1], [22]”. Some of the most useful tools are: D2R Server, Virtuoso Universal Server, The Talis Platform, etc. [22].

3 The Web of Data

Linked Data is being used by a number of individuals as well as organizations to share their data by grounding it in the web. As a result the global space of data also known as the web of data is created. “The web of data forms a giant global graph consisting of billions of RDF statements from numerous sources covering all sorts of topics, such as geographic locations, people, companies, books, scientific publications, films, music, television and radio programs, genes, proteins, drugs and clinical trials, statistical data, census results, online communities and reviews [1].” The most visible example of adoption and application of the Linked Data principles has been the Linking Open Data project [1]. The project objectives were to identify the available and existing datasets and convert these to RDF by following the Linked Data Principles, before publishing the newly converted RDF’s on the web. The participants in the project were mainly researchers and developers. Over time, the project has grown considerably, and there is involvement of large corporates such as BBC, Thomson Reuters and the Library of Congress. The growth has been obtained by the open nature of the project, where anyone can simply participate and publish data sets according to the Linked Data principles and joining the published data sets with existing data sets on the web. The web of data can be seen as an additional layer that is tightly interwoven with the classic document web and has many of the same properties:

- The web of data is generic and can contain any type of data.
- Anyone can publish data to the web of data.
- The web of data is able to represent disagreement and contra dictionary information about an entity.

Data sets are categorized into the following topical domains: geographic, government, media, libraries, life science, retail and commerce, user-generated content and cross-domain data sets [1].

3.1 Cross-Domain Data

Datasets that firstly appeared on the web of data were not specific to one topic but covered numerous domains. This cross-domain exposure was vital for interlinking domain-specific data into a solid and consistent data space, by avoiding the division of the global space of data into a remote dataset. An example of cross-domain Linked Data is DBpedia [3]. Data is pulled out from public Wikipedia dumps... The parts of the subjects of a Wikipedia article are assigned in an automatic way as a DBpedia URI. Another source of cross-domain Linked Data is *Freebase* [4], which is editable, openly-licensed database in which users and data imports from sources such as Wikipedia and Geonames [5], which have contributed to populate it. Freebase makes RDF descriptions of items in the database, which are linked to items with incoming and outgoing links in DBpedia [3].

3.2 Geographic Data

Geography is another factor that can often connect information from varied topical domains. This is apparent in the Web of Data, where the *Geonames* [5] data set frequently serves as a hub for other data sets that have some geographical component. Geonames is an open-license geographical database that publishes Linked Data about 8 million locations. A second significant data set in this area is Linked Geo Data, a Linked Data conversion of data from the Open Street Map project, which provided information about more than 350 million spatial features. Wherever, the possible locations in Geonames and LinkedGeoData are interlinked with corresponding locations in DBpedia, ensuring that there is a core of interlinked data about geographical locations [3]. Linked Data versions of the *EuroStat* [6], *World Factbook* [7] and US Census data sets are bridging the world of statistics, politics and social geography.

3.3 Media Data

The potential of Linked Data has been exploited by one of the most recognized media companies is British Broadcasting Corporation (BBC). The BBC has adopted the Linked Data principles and technologies for publishing and managing their workflows content. The BBC is known for running various radio stations and television channels that have different management systems. Following earlier experiments with publishing their catalogue of programs as RDF, the BBC during 2008 released two large sites that combine publication of Linked Data and conventional web pages. The first of these, <http://www.bbc.co.uk/programmes> provides a URI for RDF description of every episode of emission TV or radio programs broadcasted across the BBC's various channels [9]. The second of these sites, <http://www.bbc.co.uk/music>, publishes Linked Data about every artist whose music has been played on BBC radio stations, including incoming links from the specific program episode during which it was broadcasted. This music data is interlinked with DBpedia, and it receives incoming links from a range of music-related Linked Data sources. BBC has launched the site <http://www.bbc.co.uk/wildlifefinder> [9], which presents itself to users as a conventional web site with extensive information about animal species, behaviors and habitats.

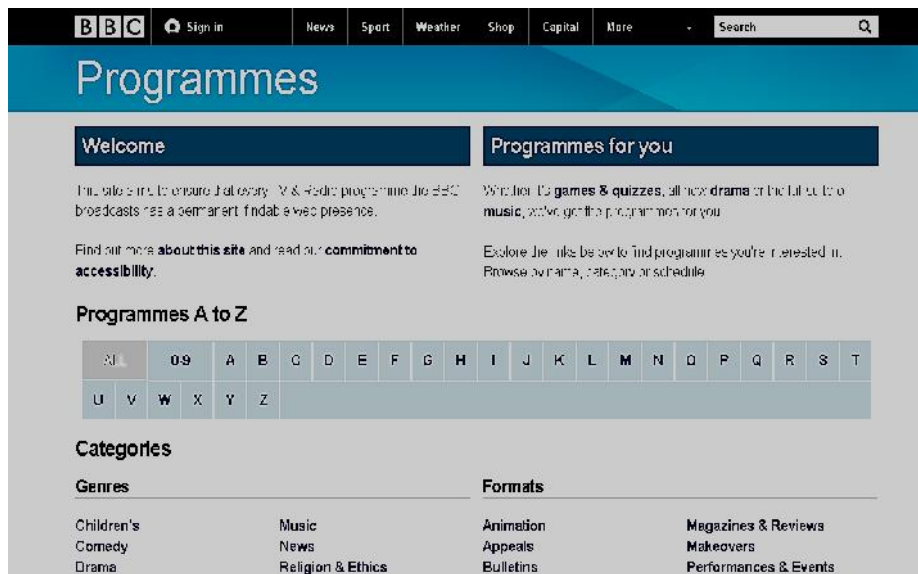


Fig. 1. The web site of BBC Programs [8]



Fig. 2. The web site of BBC Wildlife Finder [9]

3.4 Government Data

Both the government and the public sector organizations generate rich data ranging from crime statistics, economic performance to records of elected representatives and reports on school performance. In the recent years, there has been a need to increase government transparency. This has been achieved as more data is available on the web for the public to access it easily. Government websites such as data.gov.uk and data.gov have exploited many benefits of Linked Data best practices. The potential of Linked Data has increased the volumes of data shared by the government on the web significantly. Furthermore, the members of the public have exploited the benefits too; as they are able to work,

analyze, find new insights and develop tools to represent the findings to others from the data easily accessed on the web [1].

3.5 Education and Libraries

The Linked Data also supports libraries by producing high quality structured data. The intention is to integrate and interlink the content of library catalogs on a global scale according topics, location or historical period, etc. It enables interlinking of library catalogs with third part information, such as picture and video archives, or knowledge bases like DBpedia and by relying on web standards it makes library data easily accessible. American Library of Congress and the German National Library of Economics, LIBRIS and the Swedish National Union Catalogue are good examples of libraries that are supported by Linked Data [1], [22].

3.6 Life Science Data

Linked Data will significantly support the Life Sciences by enabling the connection of various datasets that are used by researchers in this field. Bio2RDF project has interlinked several used data sets, including UniProt (The Universal Protein Resource), KEGG (The Kyoto Encyclopedia of Genes and Genomes), CAS (The Chemical Abstracts Service), PubMed, and the Gene Ontology. The W3C Linking OpenDrug Data [10] is trying to enable interlink of openly-licensed data about drugs and clinical trials between the pharmaceutical companies like Eli Lilly, AstraZeneca, and Johnson & Johnson [1].

4 Consuming Linked Data

All data that are published on the web, according to the Linked Data principles becomes part of a single, global data space. This section gives an overview of deployed Linked Data applications. As the availability of Linked Data is a relatively recent phenomenon, the presented applications are mostly applications and prototypes of the first generation that will likely undergo significant evolution as lessons are learned from their development and deployment. Nevertheless, they already give an indication of what will be possible in the future as well as of the architectural patterns that are emerging in the field of Linked Data. Significant volumes of Linked Data have been published on the web. As a consequence, numerous efforts are done to researching and building applications that exploit web of data. For now these efforts can be broadly classified into three categories: Linked Data browsers, Linked Data search engines and domain-specific Linked Data applications [1].

3.7 Linked Data Browsers

Linked Data browsers allow users to navigate by following RDF links. This is similar to how users would navigate between HTML pages when using traditional web browsers. The tabulator browser [11], for example, allows the user to traverse the web of data and expose pieces of it in a controlled fashion in outline mode to discover and highlight a pattern of interest, and then query for any other similar patterns in the data web. The results of the query form a table that can then be analyzed with various conventional data presentation methods, such as faceted browsing, maps, timelines, and so on. The tabulator is an RDF browser which is designed for users' interested in the semantic web and give them a meaning to access and interact with the entire web of data [21]. Tabulator and Marbles are among the data browsers which track the provenance of data while merging data about the same thing from different sources. The fig. 3 shows the Marbles Linked Data browser displaying data about Tim Berners-Lee which has been merged from different sources [12].

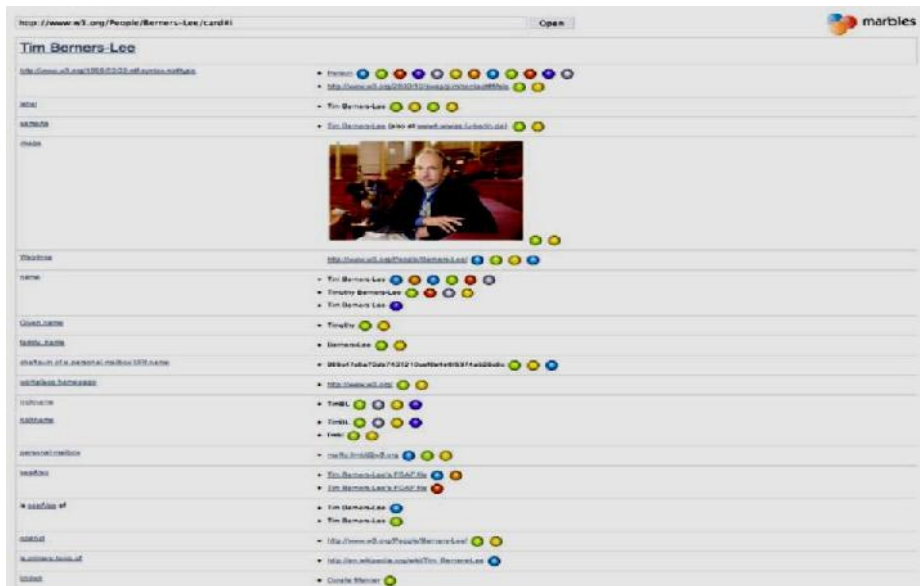


Fig. 3. Marbles browser displaying data about Tim Berners-Lee from different data source

3.8 Linked Data Search Engines

A vast amount of search engines have been developed that examine Linked Data from the web by following RDF links and provide query abilities over collected data. Data is gathered from different data sources. For example search engines, such as Sig.ma [13], Falcons [14], and SWSE [15] provide keyword-based search services oriented towards human users and follow a similar interaction paradigm as existing market leaders, such as Google and Yahoo. Linked Data search engine offer a richer interaction ability to the users which exploit the underlying structure of the data. The figure below displays Sigma search engine. The user is presented with a search box into which they can enter keywords related to the item or topic in which they are interested, and the application returns a list of results that may be relevant to the query. However, rather than providing links from search results through to the source documents in which the queried keywords are mentioned, both SWSE and Falcons provide a more detailed interface to the user. Falcons provide users the option of searching for objects, concepts and documents. Linked data search engines can be divided into two categories: human-oriented search engines and application-oriented indexes [13].



Fig. 4. Sigma search engine [13]

Swoogle, Sindice and Watson also known as application oriented indexes aid the Linked Data applications through providing APIs to find RDF documents on the Web that reference certain keywords or URI. There is a variety of Linked Data applications that are applicable to the needs of a specified user group. An example of Linked Data applications could be found on the data.gov and data.gov.uk websites where the applications provide transparency by making the government data visible to the public members. An example of the Linked Data government application is the *US Global Foreign Aid Mashup* [16] shown in fig. 5. The application demonstrates the spending figures from the United States Agency for International Development (USAID) and other government departments. An application of Linked Data that helps educators to create and manage lists of learning resources (e.g., books, journal articles, web pages) is *Talis Aspire*. Talis Aspire is written in PHP and users access the application through a Web interface. It is used by students and lecturers on daily basis to handle, store and access Linked Data and it has been deployed in many universities around the United Kingdom [17].

In the recent years, a variety of services have been developed for providing domain with specific functionality by linking up data from a variety of Linked Data sources in comparison to the Linked Data search engines and browsers that offer a generic functionality [1], [22].

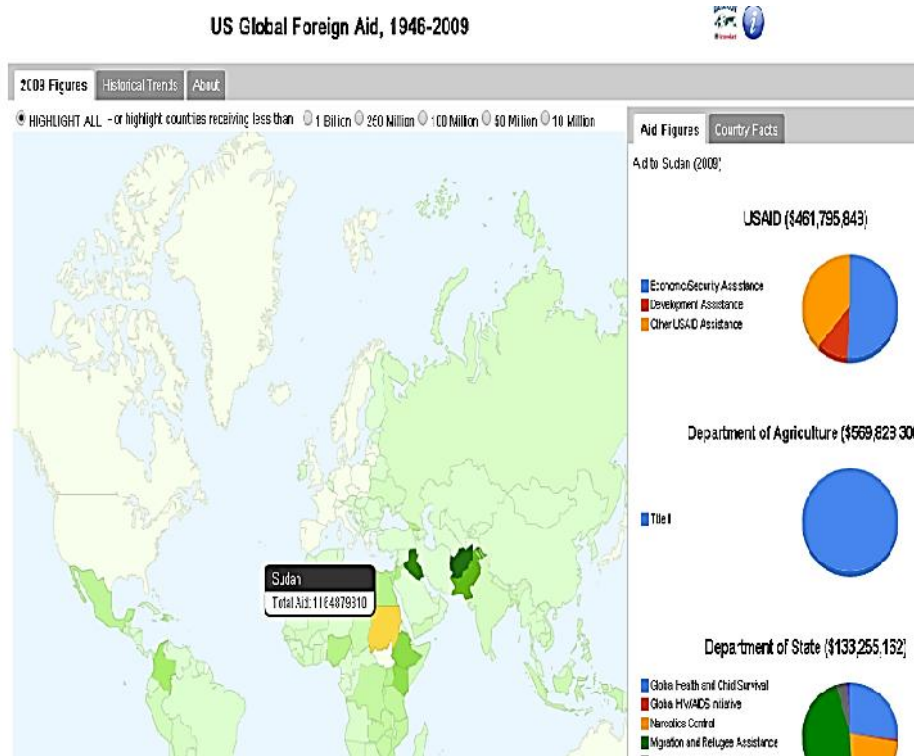


Fig. 5. US Global Foreign Aid combining data from different departments of the US

The screenshot shows the University of Plymouth website interface. The main content area displays the book 'Financial accounting and reporting' by Elliott, Barry, Elliott, Jamie. The book details include:

- Item type: Book
- Author(s): Elliott, Barry, Elliott, Jamie.
- Publisher: Financial Times/Prentice Hall
- Edition: 12th ed.
- ISBN 10: 0273712314
- ISBN 13: 9780273712312
- Date: 2008.

Below the book details is a 'My Notes' section with a text area containing the note: 'Chapters 2 and 3 are useful for the 3rd seminar - more as an example of a poor argument than a good one!'.

Fig. 6. The application of Talis Aspire [17]

DBpedia Mobile is a Linked Data application which helps tourists to explore a city. The application runs on an iPhone or other smartphone. It is a location-oriented DBpedia client application for mobile device and its components are a map view, the Marbles Linked Data Browser and a GPS-enabled

launcher application. It accesses web data and makes possible for users to publish their current location, pictures and reviews to the web as Linked Data, to enable them to be used by other applications [1], [18]. Depending on the current GPS position of a mobile device, DBpedia Mobile provides a map containing information about nearby locations from the DBpedia dataset. According to the map, users can explore background information about locations. They can navigate into DBpedia and other datasets which are interlinked [18]. On a mobile device, the map view can be dragged on the touch screen, or it can be zoomed using the provided controls. It can be viewed in displays with at least 320x240 pixels resolution. At the moment, only Opera Mobile 8 is working, because most of other mobile browsers don't have the proper Document Object Model (DOM) level 1 and 2 support to host the underlying Google Maps API. It works on the iPhone, but at the moment is not being used, as iPhone specifics haven't been implemented yet. When you click on a resource it shows up a summary view of the selected item generated by the Marbles Linked Data Browser. This view has a short text describing the resource and sometimes an image. The application may also be used on desktop computers. The desktop view can be used to explore DBpedia resources all over the world. The application successfully works with Firefox 2, Internet Explorer 7, Safari 3 and Opera 9. DBpedia Mobile makes use of these components: DBpedia datasets (labels, geo-coordinates and categories), Google Maps API, Map tiles coming by OpenStreetMap, etc. [1], [18].

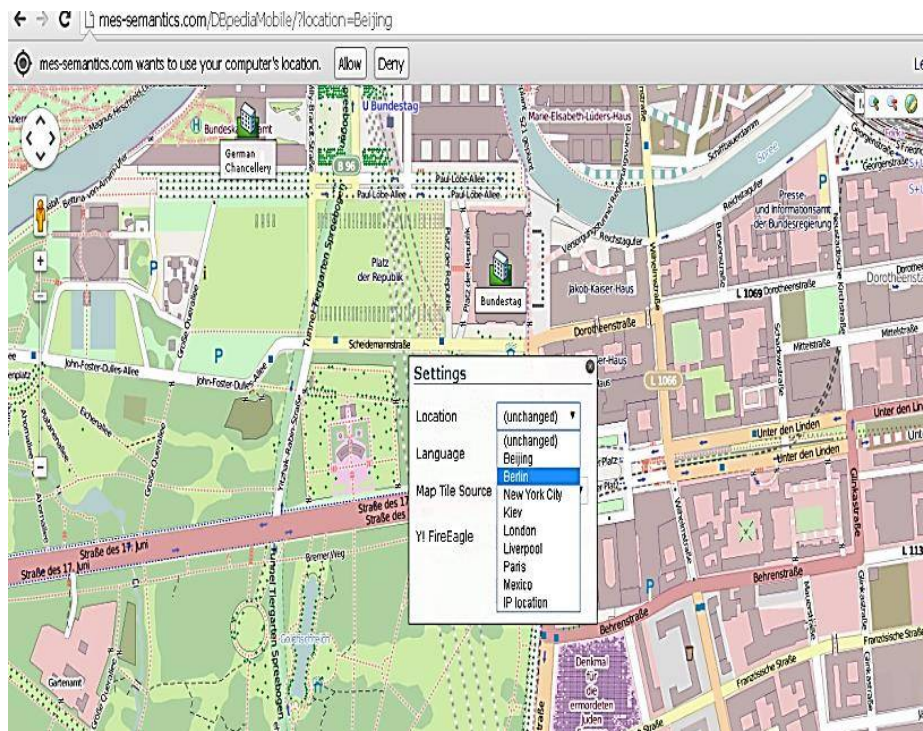


Fig. 7. A view of DBpedia Mobile's Map [18]

4 Scenarios (Case Studies)

4.1 Bibliographic Data

Bibliographic data cluster includes use cases about bibliographic records. This cluster is based on the following submitted use cases and scenarios:

- Use Case AGRIS ¹

AGRIS (International Information System for the Agricultural Sciences and Technology) database has been joining bibliographic references, such as research papers, studies and theses, each consisting of metadata such as conferences, researchers, publishers, institutions, etc. The AGRIS database publishes bibliographic records from word-wide agricultural libraries and documentation centers. The AGRIS indexer makes use of new incoming records for searching a web index for related resources. It makes use of the title element, the combined author/subject elements and the conference and journal elements of the record [21].

- Bibliographic Network ² Use Case

The International Federation of Library Associations and Institutions (IFLA) re-examined the bibliographic data to create a framework that would generate a clear, precisely stated, and commonly shared understanding of what the bibliographic record can provide information about. Also people expect the record do the answer to the user needs. Also, IFLA wants to recommend a basic level of functionality and basic data requirements for records produced by national bibliographic agencies. Linked Data techniques would make possible for these data, the concepts and relationships between them to be interpreted as an information graph. Web standards can be used to satisfy the user's discovery needs. Data needed for this case consists of:

- Title
- Sub-title
- Author/Editor/Composer/etc.
- Publisher
- Date of publication
- Language
- Type of material [21].

Community Information Service ³ Use Case

Academic organizations of different sizes (research groups, university departments, scholarly societies, special interest groups, etc.) have a strong interest in keeping awareness and qualitative information in their domain. Their aim is to publish this information to the broader academic community and to the general public. A significant part of this information is bibliographic metadata which can be found from library resources, especially information about books and articles published in a particular field. The aim is to make a library catalog and other publisher-generated bibliographic metadata freely available to community data curators so it can be easily selected by author, affiliation, subject, etc.

Curator of a community information service filters data from input sources to determine that books, articles, photographs and videos were published lately [21].

4.2 Authority Data

¹ http://www.w3.org/2005/Incubator/1ld/wiki/Use_Case_AGRIS

² http://www.w3.org/2005/Incubator/1ld/wiki/Use_Case_Bibliographic_Network

³

http://www.w3.org/2005/Incubator/1ld/wiki/Use_Case_Community_Information_Service

- Use Case AuthorClaim ⁴

The AuthorClaim registration service aim is to link scholars with the records about the works that they have written, as documented in a bibliographic database. The application makes available the identification of authors. In the application scenario, document metadata records are classified by subject experts. Each expert takes a binary decision whether a document is part of a category or not. The resulted document collection creates an issue of a subject report. Linked Data can be used to further generalization of the basic application and provide the re-use of the application's results [21].

- Use Case Authority Data Enrichment ⁵

Authority control is the practical work of creating and maintaining authority data for bibliographic entities. Authority control makes possible catalogers to disambiguate resources with similar or identical characteristics as well as collocating resources that logically belong together. Linked Data could provide the re-use of external data sets by linking instead of copying and merging. A librarian aim is to enrich a particular authority data record happening in his institutional data set (i.e. a person) with additional information. Firstly, the librarian has to explore for equivalent authoritative entities in the data set of one or more remote data providers. Next, the librarian has to identify an equivalent entity in the search result. In case of a successful identification, both representations have to be joined and the librarian has to make a decision what characteristics and/or relationships are important for the enrichment of the local representation [21].

- Use Case Linked Data Service of the German National Library ⁶

In Germany, authority data is put together and kept in a collaborative way. This data, as well as the German National Library's bibliographic data, is part of many libraries and other cultural heritage institutions. Linked Data gives a suitable framework for publishing relevant data at the German National Library and make a link of it to other data sources of interest. Users that have interest can gain access of the data either through the existing search interface of the library's web portal or by downloading the data in the form of a database [21].

5 Conclusion

Linked Data are essential to connect semantic data on the web. Principles of linked data and their practices are adopted by increasing numbers of data providers, which have resulted in the creation of global data space on the web, containing billions of RDF triples. As the web has made revolution in publishing and consuming of data, and linked data also has great potential in how we access and utilize by them. Linked data provide evolution of the web in global data, enabling applications operating on top of an unlimited set of data sources, through standardized access mechanisms. It is an important step in the evolutionary "leadership" web. Techniques offered through linked data provide overview of the data, special view for specific cases, such as in space and time which look like "appropriate area" between an interface and specific application data at that interface. Finally, we can say that the future of the web development and online data will be carried out continuously, in large measure through Linked Open Data.

References

⁴ http://www.w3.org/2005/Incubator/lld/wiki/Use_Case_AuthorClaim

⁵

http://www.w3.org/2005/Incubator/lld/wiki/Use_Case_Authority_Data_Enrichment

⁶http://www.w3.org/2005/Incubator/lld/wiki/Use_Case_Linked_Data_Service_of_the_German_National_Library

1. W3C (Word Wide Web Consortium) Official Site: Cluster BibData – Library Linked Data, http://www.w3.org/2005/Incubator/1ld/wiki/Cluster_BibData
2. DBpedia Official Site, <http://dbpedia.org>
3. Freebase Official Site, <http://www.freebase.com>
4. GeoNames Official Site, <http://www.geonames.org>
5. Datahub, <http://ckan.net/package?q=eurostat&groups=lodcloud>
6. FU Berlin Official Site: World Factbook, <http://www4.wiwiss.fu-berlin.de/factbook>, <http://datahub.io/dataset/world-factbook-fu-berlin>
7. BBC Official Site: BBC Programmes, <http://www.bbc.co.uk/programmes>
8. BBC Official Site: Wildlife Finder, <http://www.bbc.co.uk/wildlifefinder>
9. W3C Official Site: HCLSIG/LODD, <http://www.w3.org/wiki/HCLSIG/LODD>
10. W3C Official Site: Tabulator: Generic Data Browser, <http://www.w3.org/2005/ajar/tab>
11. Marbles Official Site: Marbles Linked Data Engine, <http://marbles.sourceforge.net/>
12. Sig. ma Official Site: Sig.ma EE – Semantic Information Mashup Enterprise Edition, <http://sig.ma/>
13. IWS Official Site: Falcons, <http://iws.seu.edu.cn/services/falcons/documentsearch/>
14. SWSE Official Site: SWSE – Semantic Web Search Engine, <http://www.swse.org/>
15. Data-Gov Official Site: US Global Foreign Aid, <http://data.gov.tw.rpi.edu/demo/USForeignAid/demo-1554.html>
16. W3C Official Site: A Linked Open Data Resource List Management Tool for Undergraduate Students, <http://www.w3.org/2001/sw/sweo/public/UseCases/Talis/>
17. DBpedia Official Site: DBpedia Mobile, <http://wiki.dbpedia.org/DBpediaMobile>
18. Bechhofer S., Ainsworth J., Bhagat J., Buchan I., Couch P.: Why Linked Data is Not Enough for Scientists, Institutes and Research Centres, University of Southampton, UK (2011)
19. Hartig O., Zhao J.: Publishing and Consuming Provenance Metadata on the Web of Linked Data, IPAW (2010)
20. Berners-Lee T., Chen Y., Chilton L., Connolly D., Dhanaraj R.: Tabulator: Exploring and Analyzing Linked Data on the Semantic Web, Massachusetts Institute of Technology, USA (2006)
21. Edelstein J., Galla L., Li-Madeo C., Marden J., Rhonemus A., Whysel N.: Linked Open Data for Cultural Heritage: Evolution of an Information Technology, <http://www.whysel.com/papers/LIS670-Linked-Open-Data-for-Cultural-Heritage.pdf> (2013)
22. Bizer C., Heath T., Berners-Lee T.: Linked Data – The Story So Far <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf> (2009)