

# Linked Data: Implementation, Use, and Perceptions across European National Libraries.

Laura Francesca Cagnazzo

This dissertation was submitted in partial fulfilment of requirements for the degree of MSc Information and Library Studies

DEPT. OF COMPUTER AND INFORMATION SCIENCES UNIVERSITY OF STRATHCLYDE

AUGUST 2017

### Declaration

This dissertation is submitted in partial fulfilment of the requirements for the degree of MSc of the University of Strathclyde.

I declare that this dissertation embodies the results of my own work and that it has been composed by myself. Following normal academic conventions, I have made due acknowledgement to the work of others.

I declare that I have sought, and received, ethics approval via the Departmental Ethics Committee as appropriate to my research.

I give permission to the University of Strathclyde, Department of Computer and Information Sciences, to provide copies of the dissertation, at cost, to those who may in the future request a copy of the dissertation for private study or research.

I give permission to the University of Strathclyde, Department of Computer and Information Sciences, to place a copy of the dissertation in a publicly available archive.

Yes [ V ] No [ ]

I declare that the word count for this dissertation (excluding title page, declaration, abstract, acknowledgements, table of contents, list of illustrations, references and appendices is 21,979.

I confirm that I wish this to be assessed as a Type 1 (2) 3 4 5 Dissertation.

Signature:

Date: 16/08/2017

#### Abstract

Since its creation, the digital world has been evolving at exponential rate, presenting both its developers and users with the constant challenge of updating their skills and knowledge in order to support its development, and take advantage of its potential. Alongside with the more popular World Wide Web and Web 2.0, another version of the web has been developing quietly, compared to the spectacular growth of its 'relatives': the Semantic Web, also known as Web 3.0 or Web of Data. If the expression 'Semantic Web' reflects the more general concept, 'linked data' can be defined as the key tool to realise the idea.

This dissertation focused on the investigation of the implementation and use of linked data across national libraries in Europe. Considering the increasing weight that the 'Semantic Web' discussion has been gaining on the international scene over the past decade or so, it was deemed necessary undertaking a study to understand the role that libraries are playing in such context. The selection of the research sample fell on national libraries, as organisations invested with the responsibility of leading and shaping the development of the cultural heritage network of their own country. The aim of the research is to deliver a comprehensive picture of the current state of linked data implementation in Europe, gathering information on reasons, purposes, and uses of this technology, together with perceptions on related benefits and challenges as per the experience of the information professionals involved in the implementing process.

What emerged from this study is a strong need to spread the awareness of the Semantic Web potential within the library environment. Although various institutions have applied linked data to their resources, they are still in the minority, and a considerable lack of expertise has been identified. Some of the projects achieved have showed the advantages that linked data can deliver in terms of augmenting the visibility and discoverability of library data, alongside with overcoming linguistic barriers, and supporting interoperability. However, several issues still remain unsolved, requiring the efforts of further research. Whilst the early implementers bewailed a lack of guidelines and tools to assist their first steps, the situation has started to change, with an increasing number of vendors and developers contributing to the Semantic Web advancement. Nonetheless, libraries clearly expressed the demand for positive and successful cases of linked data use, that can best support and motivate their choice of adhering to the web of data principles.

iii



## Acknowledgements

I would like to express my warmest thanks to Osma Suominen, who first agreed to take part in this research, as interviewee, manifesting his interest and appreciation for this project, and who also suggested some of the latest publications on the topic of linked data implementation in the library sector. I would also like to thank Carla Colombati and Klára Rösslerová for their advice and for addressing me towards helpful contacts.

A big thank you goes to my supervisor, Dr Diane Rasmussen-Pennington, for her guidance, patience, professional and moral support at any step of this work.

Last but not least, I thank my partner, Kenneth Barlee, for the endless supply of cheesecake and Montepulciano d'Abruzzo wine provided throughout the project.

"On two occasions I have been asked, 'Pray, Mr. Babbage, if you put into the machine wrong figures, will the right answers come out?' I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question". (Charles Babbage).

"The best thing to do with your data will be thought of by someone else". (Rufus Pollock).

## **Table of Contents**

Declaratio	nii
Abstract	iii
Acknowle	dgementsv
Table of Fi	iguresx
Table of Table	ablesx
1. Intro	duction1
1.1.	Linked data1
1.2.	Library data1
1.3.	National libraries2
1.4.	Research context
1.5.	Problem statement
1.6.	Methodology4
1.7.	Learning outcomes
1.8.	Dissertation structure
2. Liter	ature Review 6
2.1.	Semantic Web and Linked Data6
2.1.1.	Linked Data principles8
2.1.2.	Linked Data or Linked Open Data?11
2.1.3.	Challenges of the Semantic Web vision12
2.1.4.	Benefits of Semantic Web and Linked Data13
2.2.	Linked Data development: outlining some of the main projects14
2.3.	Linked Data and cultural heritage institutions18
2.3.1.	Why would libraries and cultural institutions need Linked Data?
2.3.2.	Challenges of Linked Data implementation21
2.3.3.	Examples of Linked Data use in the cultural heritage sector

2.4.	The role of national libraries	25
3. Rese	earch Methodology	28
3.1.	Literature review	28
3.2.	Interviews	30
3.3.	Online resources analysis	32
3.4.	Conclusion	33
4. Find	ings and analysis	34
4.1.	Linked Data uses across national libraries	36
4.1.1.	Contributing to the Linked Open Data cloud	36
4.1.2.	Linked Data for bibliographic and authority data	37
4.1.3.	Publishing the national bibliography as Linked Data	39
4.1.4.	Linked Data for digital resources	40
4.1.5.	Thesauri and ontologies	40
4.2.	Implementation process	41
4.3.	Reasons for implementation	43
4.4.	Benefits of Linked Data adoption	45
4.5.	Challenges of Linked Data implementation	46
4.6.	Most popular examples of Linked Data implementation	49
4.7.	Policies	50
4.8.	Collaboration	51
4.9.	Standardisation and interoperability	53
5. Reco	ommendations and Conclusions	56
5.1.	Best practice	56
5.2.	Further research	58
5.3.	Key findings	59
5.4.	Research self-assessment	62

5.5.	Discus	sion	62
References	s		65
Glossary			75
Appendice	S		78
Appendix	<1-	Participant information sheet	78
Appendix	< 2 —	Interview questions	80
Appendix	< 3 —	Interview with the British Library	81
Appendix	<b>‹</b> 4 –	Interview with the National Library of Finland	87
Appendix	< 5 —	Interview with the National Library of Germany	90
Appendix	< 6 –	Interview with the National Library of Latvia	
Appendix	(7 –	Interview with the National Library of Luxembourg	103
Appendix	< 8 –	Interview with the National Library of the Netherlands	106
Appendix	< 9 —	Interview with the National Library of Wales	112
Appendix	< 10 —	Interview with the National Library of Portugal	116
Appendix	< 11 —	Interview with Open Knowledge Greece	117
Appendix	< 12 —	Email interview with the National Library of France	119
Appendix	< 13 —	Email interview with the National Library of Spain	122
Appendix	< 14 —	Email interview with the National Library of Czech Republic	124
Appendix	< 15 —	Email conversations	125

# Table of Figures

Figure 1: Sturtevant, R. (2010) Linked Data (Semantic Web) candies	. 6
Figure 2: Example of graphic representation of an RDF triple	. 9
Figure 3: First version of the LOD cloud diagram published in 2007	14
Figure 4: Latest version of LOD cloud diagram published in February 2017	14
Figure 5: Google Knowledge Graph for Chris Martin, leader of British band Coldplay	16
Figure 6: British Library data model	42

## Table of Tables

Table 1: Overview of research findings – Participating National Libraries	.34
Table 2: Overview of research findings – Non-participating National Libraries	35
Table 3: VIAF and Europeana contributors	36

Introduction

#### 1. Introduction

This research focuses on the challenges and difficulties that European national libraries (NLs) have encountered during the process of creating and implementing linked data (LD) within their resources, and provides an insight of the current situation, with the intent of increasing the awareness of LD potential within the information science sector, and of drawing recommendations for policies and best practice for institutions which perceive the call and the need of embracing this technology, but are still looking for best solutions.

#### 1.1. Linked data

When Tim Berners-Lee invented the web, his intention was to create a virtual environment where anyone could add and share documents. However, he soon realised that his idea had not developed the way he hoped, and he decided to work on a new project, with the aim of having data on the web, and not just documents. "I want to think about a world where everybody has put data on the web... and I'm calling that 'Linked Data'". (TED, 2009).

The main and innovative concept that underlies LD is building relationships. The goal is to provide a seamless experience of navigating a 'web of Trust', where anyone can make his/her own contribution. (Library of Congress, 2012). Cooperation and interoperability are the key benefits brought by LD.

The broader idea to which LD relates to, is the 'Semantic Web' (SW) approach, whose purpose is to communicate the content of web resources in a way that is 'understandable' by computers. (Rasmussen-Pennington, 2016). If machines are enabled to 'comprehend' the meaning of web pages' content, they can build relationships between resources and enrich users' experience, improving discoverability. Nonetheless, a standard way to apply such idea is still to be found. Although the potential behind LD is widely perceived, there is still confusion and reserve on how to benefit from this tool, and the path towards its implementation has revealed several challenges.

#### 1.2. Library data

Cultural institutions have been sharing metadata long before the advent of the Internet. Bibliographic control has always represented one of the major target for libraries, and,

Introduction

ideally, the achievement of a universal bibliographic control. In the path towards its realisation, two major revolutions occurred: the transition from book to card catalogue, and from this to the 'machine readable' version. Standards have been designed *in itinere*, with the intent of enabling interoperability and exchange of records (Tallerås, 2013), such as MARC (*MAchine-Readable Cataloging*), AACR (*Anglo-American Cataloguing Rules*), and the most recent RDA (*Resource Description and Access*). The main trigger for moving forward in the standards development has been the difficulty of achieving the exchange of data on a wide scale, due to the diversity of descriptive standards and schemas adopted, as well as the diversity of languages in use. (Breeding et al., 2016). Interoperability of metadata, essential to improve data exchange and sharing, is still hindered by the semantic gap between different metadata schemas. (Sugimoto et al., 2015). LD offers cultural institutions the means for improving interoperability, alongside with better tools for data management, and for increasing the amount and quality of information. (Byrne & Goddard, 2010).

We are approaching the end of the cataloguing record containing purely library-provided data, and we are moving towards more enriched data, coming from various resources. This requires that library data is structured to be more flexible and widely applicable. (Coyle, 2009). To reuse an expression by Richard Wallis, "We are moving from cataloguing to catalinking". (Wallis, 2013).

#### 1.3. National libraries

The choice of focusing on NLs, besides allowing a manageable sampling of data and research scale, is based on the consideration of the role that such institutions play with respect to preserve and widen access to all knowledge published within their own countries. (Jøsevold, 2016).

Amongst the responsibilities that the International Federation of Library Associations and Institutions (IFLA) attributes to NLs, particularly significant to the purpose of this study are: "the provision of central services (e.g., reference, bibliography, preservation, lending) to users both directly and through other library and information centers; the preservation and promotion of the national cultural heritage". (IFLA, 2017). Developing and disseminating standards for libraries and cultural institutions of the same nationality, and maintaining and making available their collections of national relevance, are two of the key duties of NLs.

Linked Open Data (LOD) provides a way for cultural heritage information to enter the web of data, to radically transform the traditional metadata techniques, and to open their collections to a wider audience, as proved by the initiatives of major organisations such as the Deutsche National Bibliothek (DNB = German NL) and the Bibliothèque National de France (BnF = French NL). (Jones & Seikel, 2016).

#### 1.4. Research context

Considering the swift turnover of technology and the current state of experimentation with LD (which translates into an ongoing ideation and development of new projects), latest research can prove valuable in tracing the present state of implementation of a technology which is still quite recent. Furthermore, while previous studies adopted surveys and literature reviews as main methods to gather information on the use of LD among NLs, or provided the perspective of a single (or few) institution, this work adopts interviews and online resources analysis, with the intent of providing a more comprehensive picture of LD application across Europe. A proof of the existence of a gap in the literature can be deemed the good rate of participation to the study, and the appreciation and interest expressed by the majority of the interviewees, who requested to be informed on the findings of the research, considering this work as a useful instrument to spread awareness of the SW across information professionals.

Albeit some NLs have embraced and applied the LD concept, several institutions still feel intimidated by such tool, since no guidelines have been yet formulated on how to build the systems. Lack of documentation and clear instructions represents a barrier to the LD propagation. (OCLC Research, 2014). Furthermore, the inadequacy of standards represents a primary challenge. If communities providing web content will not agree on implementing the necessary technology, the LD universe will continue to be limited and fragmented.

A study providing detailed descriptions of projects' development and challenges encountered, and reporting the viewpoints of those professionals directly involved in the planning and unrolling of such projects, may represent a powerful means to improve the understanding of LD benefits and identify the key issues that still require a solution.

Introduction

#### 1.5. Problem statement

The questions that this research addresses are:

- What are the benefits of LD implementation as perceived by European NLs that adopted this technology?
- What are the reasons behind organisations' decision of implementing LD? Which are the barriers preventing LD adoption?
- What are the key challenges encountered by European NLs before and during LD implementation?

The investigation focuses on the identification of the reasons behind organisations' decision of adopting or not LD. In addition, the study aims to map and illustrate the results of the LD application by European NLs, since a gap in the existing literature was recognised.

#### 1.6. Methodology

A qualitative approach, in the form of a case study, was the framework considered more appropriate to deliver a breadth of perspective on the subject of LD implementation. With the intent of providing comprehensive answers to the research questions, an integration of data collected through different techniques was operated. First of all, since the results of a preliminary literature review were not satisfactory (although numerous, many resources were too generic and not fit for the purpose of this study), it was decided that interviews would be conducted, to collect reflections and feedbacks from information professionals involved in the LD implementation process. Semi-structured interviews were chosen for their adaptability: a set of questions was prepared, but, in order to allow the flow of the conversation, the order was changed according to need, and questions were modified/omitted. Interviews were carried out via Skype and email. A constant comparative analysis was performed to process data collected through literature review and research interviews. Furthermore, an analysis of online resources was carried out, primarily to gather information from the websites of those institutions which did not take part in the interviews.

Introduction

#### 1.7. Learning outcomes

Considering the present attention to the LD topic, it is of interest to provide an overview of the reasons behind its implementation across European NLs. In addition, the collection of viewpoints of professionals who have been closely involved in the accomplishment of projects based on the use of LD, will contribute to a better grasp of advantages and disadvantages of such method of publishing structured data, and to the formulation of recommendations for policies and best practice. A comprehensive picture of the current state of LD use across Europe will constitute a useful tool for institutions who are considering the adoption of LD within their resources. This research aims to increase the perception of LD potentiality: the opportunity of obtaining augmented and widened visibility and accessibility of library resources to the rest of the world. (Rasmussen-Pennington, 2016).

#### 1.8. Dissertation structure

The main body of this dissertation can be divided into five sections:

- *Introduction*: chapter 1 is intended to prepare the reader to what is coming, offering an overview of this study
- *Literature Review*: chapter 2 presents the results of the literature review performed in order to map the existing sources on the selected topic, and provide the theoretical framework
- Research Methodology: chapter 3 explains the methods applied to the research, for data collection and analysis, providing justification and outline of the main features
- *Findings and Analysis*: chapter 4 reports the findings obtained through the data processing
- Recommendations and Conclusions: chapter 5 offers further reflections on the research topic and findings, formulating recommendations for best practice; a final summary of the study outcomes is outlined.

#### 2. Literature Review

#### 2.1. Semantic Web and Linked Data

The SW was at first presented as a new form of web content that is meaningful to computers, an extension of the existing World Wide Web, in which information is provided with well-defined meaning, with benefit of a better cooperation between humans and machines. Its target is to enhance and bring logic to the traditional web rather than creating a different web. (Berners-Lee et al., 2001; Baker et al., 2011). The first and most important step towards its realisation is to publish data on the web in a machine-readable format. (Berners-Lee, 2000). If SW is the goal, LD is the means to reach it. (Bizer et al., 2009).



Figure 1: Sturtevant, R. (2010) Linked Data (Semantic Web) candies.

The literature shows some uncertainty on the relation between SW and LD, which requires an attempt (at least) of disambiguation. After developing the web, Sir Tim Berners-Lee introduced the idea of an advanced version of it: an extension of the web enabling relationships not only between documents, but also between the elements, the data, within documents. (Tallerås, 2013). It was the frustration of having a web made of exclusively human-readable content that encouraged the development of the SW idea. (Van Hoolan and Verborgh, 2014). Other terms were subsequently adopted, such as 'Giant

global graph' (Berners-Lee, 2007), 'Web 3.0' (Berners-Lee, 2006) and 'Web of data' (Bizer et al., 2009), to express a similar concept.

Beyond a slight ambiguity generated by the use of all these different terms, a common interpretation is that while 'Semantic Web' reflects the overall, high-level vision, the term 'linked data', used for the first time by Berners-Lee in 2006, conveys the practice and methods to concretise such vision. (Heath, 2009; Tallerås, 2013; Calaresu & Shiri, 2015).

LD has also been interpreted as a way of rendering the SW concept more appealing: according to van Hoolan and Verborgh (2014) the elaboration of the LD principles represents a rescaling of the initial ambitious idea (connected to the developments of Artificial Intelligence) to a more realistic and achievable dimension. On the opposite front, Heath (2009) refuses to accept the interpretation according to which LD implies a sort of re-dimensioning of expectations. The importance of links has always been the key component of the SW vision, and there is no application or service that can claim its belonging to the SW unless built following the LD principles. (Heath, 2009). Tallerås (2013) describes the SW as an ambition to establish links between data across various domains, and to make machines capable of operating with these links. The key requirement for this to happen is a minimum level of 'semantic interoperability', enabling machines to 'understand' data.

A different interpretation is offered by Stuart (2011), who argues that the difference between LD and SW is the difference between 4 and 5 stars in Berners-Lee's scheme (illustrated in the next section): data may be part of the SW adhering to RDF standards, but, as for the traditional web, pages are more useful when linked to by external resources. And that is the goal of LD.

The definition of LD provided by its own creator describes LD as "the Semantic Web done right". (Berners-Lee, 2008). From a technical perspective, the term 'linked data' indicates a set of best practises for publishing and connecting structured data on the web. (Bizer et al., 2009). Its main requirements are: data needs to be in a machine-readable format; data's meaning must be explicitly defined; data has to be linked to external datasets. (Volz et al., 2009, in Shiri & Davoodi, 2016). It is important to stress that LD should be interpreted as a continuously evolving set of best practices for the publication of structured data on the web, rather than a specific, well-defined technology. (Van Hoolan and Verborgh, 2014).

"LD aims to achieve the goal of the Semantic Web through the provision of a set of standards, data publishing models and methods that bring consistency, interoperability and shareability to unorganised and unidentifiable data on the web." (Shiri & Davoodi, 2016, p.65). Heath (2009) defines LD as "a means to dismantle data silos". To provide a simple explanation of the LD concept, Heath compares the web of documents to the web of data: as the first was born by linkage of HTML (HyperText Markup Language) documents with hyperlinks, the second one emerged when bits of data started to be linked together with RDF triples.

To summarise, LD is a way of connecting related data across the web using URIs (Uniform Resource Identifiers), HTTP (Hypertext Transfer Protocol), and RDF. (Heath, 2009).

#### 2.1.1. Linked Data principles

The key requirements for LD to be classified as such were clearly stated by Berners-Lee (2006) in his four rules (known as 'Linked Data principles'):

- 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, SPARQL)
- Include links to other URIs, so that users can discover more things.

The first principle requires a unique identification of each concept, which can be achieved associating each notion to a URI. The second principle states that identifiers must be HTTP URIs, that is to say URLs (Uniform Resource Locators) on the web. (Van Hoolan and Verborgh, 2014). HTTP is a communication standard to deliver information on the web. (OCLC, 2012). While URLs represent addresses for documents and other resources on the web, URIs provide a more generic way to identify any entity that exists in the world. (Bizer at el., 2009). URIs, in the LD context, have a disambiguating function, as they are used to uniquely identify resources. The risk related to this element is the lack of stability of URLs, which may determine loss of validity, hence a lost connection. (Van Hoolan and Verborgh, 2014).

The third principle requires the representation of those resources identified by URIs through the adoption of standards, such as RDF, a generic, graph-based data model that can describe all things in the world in the form of statements called 'triples' (see figure 2).

(Bizer et al., 2009; Tallerås, 2013). Statements consist of subject, object and predicate, each of which is represented by URIs. Predicates (or properties) describe the relation between elements. Subject and object of a triple can belong to different datasets: this is how RDF links work. (Bizer et al., 2009). RDF is a straightforward, easily-adaptable data model: connections can be established among any resource, allowing the interaction of heterogeneous data. (Van Hoolan and Verborgh, 2014). RDF is to SW what HTML has been to the traditional web. (Yu, 2007). This data model defines the 'graph database', which is the type of database that builds the SW. In other words, a graph database is the data storage model used by the SW. (Introducing Graph Data, 2017).



Figure 2: Example of graphic representation of an RDF triple. On the left is the URI of the VIAF (Virtual International Authority File) page for J.R.R. Tolkien; on the right is the URI of the page containing the bibliographic description of a printed version of The Lord of the Rings on WorldCat. The two entities are connected by properties drawn from the Dublin Core and DBpedia ontologies.

The first standardised syntax of the SW was based on eXtensible Markup Language (XML), therefore named RDF/XML. Whilst XML focuses on providing added structure to documents, RDF is a tool to express meaning. (Berners-Lee et al., 2001). The main difference between RDF and XML and other models is that RDF is 'schema-neutral': to understand the meaning of an RDF triple, the URI of each resource needs to be followed towards the vocabulary it refers to. (Van Hoolan and Verborgh, 2014). The great advantage of RDF is that it allows the free reuse of subjects, predicates and objects, as their meaning is defined by an external source, and it is not limited to a local dimension proper of the application. Furthermore, RDF supports and encourages change, as new triples can be added at all times, without need of modifying the underlining structure. (Van Hoolan and Verborgh, 2014). Alternative formats have been created in recent years, such as Turtle syntax: in Turtle, each element of a triple is separated, allowing a clearer visualisation of the results. (Van Hoolan and Verborgh, 2014).

The fourth and last principle prescribes that data contains links to other data. (Van Hoolan and Verborgh, 2014). To recapitulate, LD employs HTTP URIs to identify resources; the HTTP protocol as retrieval mechanism; and RDF data model to represent resource description. (Bizer et al., 2009).

Alongside with URIs and RDF, another main component of the SW is ontologies. Defined as 'collections of information', ontologies have the ability of improving retrieval, enabling search engines to find specific concepts, rather than ambiguous keywords. (Berners-Lee et al., 2001). They provide a shared understanding of a domain, gathering together its important notions (expressed in RDF), and establishing their relationships, including hierarchical ones. (Bizer et al., 2009; Antoniou et al., 2012). Within the LD context, ontologies allow the connection between collections, through the sharing of same concepts. (Van Hoolan and Verborgh, 2014).

Amongst the most commonly adopted ontology languages are:

- RDF Schema (RDFS): a general-purpose vocabulary language for describing properties and classes of RDF resources, with a semantics for generalisation hierarchies of such properties and classes (RDFS, 2010; Antoniou et al., 2012);
- Web Ontology Language (OWL): built on RDFS, OWL is a richer vocabulary language for describing properties, classes, and relations between classes. (OWL, 2013; Antoniou et al., 2012). Recommended by W3C, it is probably the most widespread language for creating ontologies nowadays. (Yu, 2007). One of the most popular OWL term is 'sameAs', which allows different datasets to be joined together. (Stuart, 2011).
- Simple Knowledge Organisation System (SKOS): a common data model for sharing and linking knowledge organisation systems via the web. SKOS captures much of this similarity and makes it explicit, to enable data and technology sharing across diverse applications. (SKOS, 2013). According to van Hoolan and Verborgh (2014), SKOS exemplifies the most successful result of the standardisation efforts towards reusing controlled vocabularies, offering a data model to represent such vocabularies on the web.
- Friend of a Friend (FOAF) and Description of a Project (DOAP): ontologies used to build social networks across the web. (Berners-Lee, 2006). FOAF includes the basic terms generally used to describe personal information. (Yu, 2007).

In order to evaluate LD applications, Berners-Lee elaborated the so-called 'five-star scheme':

- ★ Available on the web (in whatever format) but with an open licence, to be
  Open Data
- ★★ Available as machine-readable structured data (e.g. Excel instead of image scan of a table)
- ★★★ As above, plus non-proprietary format (e.g. CSV instead of Excel)
- ★★★★ All the above, plus use of open standards set by W3C (RDF and SPARQL) to identify things
- ★★★★ All the above, plus: link your data to other people's data to provide context.(Berners-Lee, 2006).

Few elements need clarification at this point. Berners-Lee (2006) mentions W3C as the body supporting the development of the SW. The *World Wide Web Consortium* (therefore, W3C) is an international community that creates open standards to support the continuous enhancement of the web. Founded (and currently led) by Berners-Lee in 1994, its goal is to foster compatibility and agreement among organisations, and "lead the World Wide Web to its full potential by developing protocols and guidelines that ensure the long-term growth of the web". (W3C, 2016; W3C, 2017).

One of the standards established by W3C, other essential component of the SW universe, is SPARQL (pronounced "sparkle", recursive acronym for SPARQL Protocol and RDF Query Language), a query language for RDF that allows manipulation and retrieve of data from a triplestore (RDF database). (SPARQL, 2017; Stuart, 2011). Each triplestore should provide an 'endpoint', which is where SPARQL queries can be submitted. The basic principle of SPARQL is trying to find sets of triples that match a given graph pattern. (Antoniou et al., 2012).

#### 2.1.2. Linked Data or Linked Open Data?

A further disambiguation is needed at this point. Although sometimes used as synonyms, 'linked data' and 'linked open data' refer to two different concepts: not all linked data are published under an open license and not all open data are linked. (Heath, n.d.). Berners-

Lee (2006) explains that LD are 'open' when published under a licence that allows reuse (e.g. the *Creative Commons* licences). Nevertheless, LD can be used internally and not necessarily be open.

The tendency to opening data as widely as possible is now spreading across different sectors, such as government organisations and industries, alongside with the cultural environment. (Shiri & Davoodi, 2016). The fact of publishing open data has specific benefits for each sector: for libraries and cultural institutions, it helps providing better information services to users; for international organisations, it brings commercial benefits, taking advantage of the skills and expertise of external developers, who can manipulate and reuse the data; for governmental bodies, it grants transparency, and it benefits of the economic potential of data; for science, it assists reducing the duplication of work, supporting research and progress. (Stuart, 2011).

#### 2.1.3. Challenges of the Semantic Web vision

According to Stuart (2011) the fact that there is no unique or most correct way to publish data on the web does not represent an issue: the essential is that we are making data available in some way. Of a different opinion, Tallerås (2013) argues that one of the main challenges in realising the SW vision is the diversity of metadata adopted in different sectors. The achievement of full interoperability is hindered by the existence of 'semantic heterogeneity' between systems. (Antoniou et al., 2012).

Van Hoolan and Verborgh (2014) claim that the complexity of the SW has slowed its expansion: the expertise that it requires (particularly with respect to publishing data and writing applications) was initially too demanding for the majority of users.

Alongside with more technical challenges, such as those regarding applications architecture, schema mapping, and building user interfaces, Bizer et al. (2009) identify some key issues preventing the wide adoption of LD, on which they invite the research to focus:

- Link maintenance: the web of data continuously changes; the risk of dead links, pointing to URIs no longer maintained, requires a solution
- Licensing: it is essential that for each dataset, conditions attached to the reuse of data are specified

- Trust, quality, and relevance: issue of assessing quality of data and ensuring that data which are most relevant and appropriate to users' needs is identified and made available
- **Privacy**: the risk of exposing sensible data requires a combination of technical and legal measures, as well as awareness of users about what data to provide in each context.

Further challenges detected include: the ideation of structures which are able to cope with data from present but also from the future (issue of foreseeing possible future developments and create adequate long-term solutions); the exposing of data with strings understandable by humans, and not only by machines. (Stuart, 2011). In addition, dealing with the ever-increasing amount of data. (Antoniou et al., 2012).

#### 2.1.4. Benefits of Semantic Web and Linked Data

Even before the elaboration of the LD idea, the key benefit recognised to the SW was its capacity of enabling shared understanding, enhancing interoperability, and promoting synergy among agents on the web. (Berners-Lee et al., 2001). Berners-Lee attributes to the SW the ability to "assist the evolution of the human knowledge as a whole", allowing the linkage of similar concepts, even when created by different organisations, since each entity is identified by a URI. (Berners-Lee et al., 2001, p.38). LD has the potential to overcome barriers and allow interoperability: providing the tools to obtain answers to complex, even multidisciplinary queries, LD can build bridges across fields. (TED, 2009). Machines become capable of 'understanding' the meaning of information, thanks to the adoption of standards such as RDF, URIs and ontologies. The definition of properties and classes through the reference to widely adopted vocabularies offers the basis for meaning and content sharing across domains. (Tallerås, 2013).

From an information retrieval viewpoint, the use of structured data enables search engines to overcome issues such as poor relevancy ranking, inaccuracy, and slowness, thanks to the adoption of RDF as data exposition method, and the connection of data to common vocabularies, which avoids ambiguity. (Spivak, 2009; Byrne & Goddard, 2010). LD creates the opportunity of developing new types of domain-based applications. (Bizer et al., 2009). Furthermore, it suggests a new, more granular way to look at information, made of blocks that come from different sources and can be accessed at different times. (OCLC, 2012).

#### 2.2. Linked Data development: outlining some of the main projects

This section presents some of the most emblematic examples of LD applications, across diverse sectors, leaving on a side, for the time being, the cultural institutions environment. A graphic visualisation of the datasets published in LD format is offered by the *Linking Open Data (LOD) cloud diagram*.



Figure 3: First version of the LOD cloud diagram published in 2007. (Cyganiak & Jentzsch, 2007).



Figure 4: Latest version of LOD cloud diagram published in February 2017. (Abele et al., 2017).

Started in 2007 and supported by the W3C Semantic Web Education and Outreach Group, the *Linking Open Data Project* was generated with the goal of identifying existing datasets available under open licences, converting them into RDF, and publishing them on the web, with links to other data sources. (Bizer et al., 2009). The metadata published in the LOD cloud is collected and curated by contributors to the DataHub, a free data management platform by Open Knowledge International, where organisations can manage and publish collections of their data. (Abele et al., 2017; Datahub, n.d.).

Each node in the diagram represents a distinct dataset published as LD. The arcs indicate links between items in two datasets. Heavier arcs correspond to a greater number of links, and bidirectional arcs indicate that each dataset contains outward links to the other. (Bizer, 2009).

Observing the cloud, it immediately jumps to the eye the predominance of datasets focusing on life sciences. Scientists are currently amongst the most active LD supporters and developers, having realised the LD potential to uncover an enormous amount of data previously hidden and fragmented across databases, spreadsheet, and documents. Emblematic the case of the *Alzheimer's Disease Neuroimaging Initiative*, born from the drive of a group of corporate and government labs who decided to publish as LD a conspicuous volume of patient information and brain scans. It was demonstrated how a complex scientific query was answered with relevant results if put in the LD databases context, whilst produced irrelevant results on traditional search engines. (Berners-Lee, 2010).

One of the central interlinking hub of the LOD diagram is DBpedia, a main resource in the LD context, generated from the community efforts to extract structured information from Wikipedia and make this information available on the web. (DBpedia, 2017). Originally designed by Chris Bizer, DBpedia is a publicly accessible RDF dataset covering several domains. (TED, 2009). At 2014, DBpedia release consisted of 3 billion RDF triples. The current overall count has recently reached 23 billion triples. (Freudenberg, 2017).

With the key goal of increasing relevance of the search results, Bing, Google, and Yahoo! collaborated to create *Schema.org*, to provide a standard allowing web content creators to mark up their pages with semantic data. This gives the content creator control over how data is described, and reduces the need, for search engines, to analyse and extract entities from web content. (Neish, 2015).

Google adopted LD technologies for the development of the Google Knowledge Graph, which provides information drawn from various sources to millions of users every day. (Rasmussen-Pennington, 2016). It is visually represented by a panel appearing on the right side of the screen when searching for a popular topic or person (see figure 5). The Knowledge Graph is presented by Google as a tool to render information search easier and quicker, obtaining more relevant results: this is achieved thanks to the disambiguation ability of LD and its capacity of 'understanding' information in a way which is closer to the human comprehension than traditional search engines. (Singhal, 2012).



Figure 5: Google Knowledge Graph for Chris Martin, leader of British band Coldplay

Another major company making use of LD is Facebook. The *Open Graph Protocol*, originally created at Facebook in 2010, is used to allow any web page to have the same functionality as any other object on Facebook. (Open Graph Protocol, 2014). Built on RDFa, a W3C recommendation that adds a set of attribute-level extensions for encoding structured data within web documents, the Open Graph Protocol enables site owners to determine how entities are described on the social network. (Heath & Bizer, 2011; RDFa, 2017).

According to the statistics reported by Bizer (2011), in 2009 the majority of data published as LD was geographical data. Amongst the most popular geographical datasets are: GeoNames, present in the LOD cloud diagram since the very first version and linked to by many other resources; LinkedGeoData, which offers a LD version of OpenStreetMap. (LinkedGeoData, 2017).

The yellow of the governmental resources definitely represents an important portion of the diagram. Governments of various countries have embraced the idea of making their

data available under open licenses, such as United States (<u>https://www.data.gov/</u>) and United Kingdom (<u>https://data.gov.uk/</u>). However, not all the government data are LD. (Antoniou et al., 2012). A list of Open Data portals can be found at datacatalogs.org (<u>http://dataportals.org/</u>), currently containing 524 data catalogues, mostly from government organisations. (Neish, 2015).

The world of journalism and media has also been very active in experimenting with the Web 3.0 technologies. In 2009 the *New York Times* started to convert its subject headings index into SW format: the heading list is published in RDF format, and terms are linked to datasets like DBpedia and GeoNames. (Antoniou et al., 2012). The main goal was to maintain and share the extensive newspaper's holdings, and actively encourage use of data via public APIs. (Marden et al., 2013). The newspaper *The Guardian* has also embraced the SW technologies with its Open Platform. (O Crualaoich, 2010). Thomson Reuters has developed *OpenCalais*, a service for annotating news texts with URIs from the LOD cloud referring to places, companies, and people. (Bizer, 2009). Its aim is to improve value, accessibility, and interoperability of any web content. Furthermore, the publisher Springer Nature has developed the *Scigraph*, a LOD platform for the scholarly domain, projected to contain 1.5 to 2 billion triples currently available under a CC BY-NC 4.0 license. (Springer Nature, 2017).

BBC has been a very prolific contributor of the LD landscape. *BBC Things* provides a single reference for the growing collection of entities (persons, places, organisations, and events) mostly used by BBC content. Its data is linked to various open data sources. (BBC Things, 2017). The company has created several ontologies for domains such as wildlife, sport, food and BBC programmes. Amongst the many applications based on LD technology are *BBC Education, BBC Sport,* and *BBC Music.* (Ontologies, 2017).

*Linked Jazz* is a research project investigating the application of LOD technologies to digital cultural heritage resources. (Linked Jazz, n.d.). The building material of this dataset is represented by transcripts of jazz archive interviews, which are first analysed with natural language processing tools, to extract full names (entities) that are then mapped against DBpedia. Users can contribute to creating relationships among interviewees and names mentioned in transcripts, using a LD friendly vocabulary. Relationships are available as RDF triples, an API, and a network visualisation. The result is a wider access to jazz archives' information. (Marden et al., 2013).

This section has highlighted how the web of data principles have been applied across many different sectors. The next segment will focus on the relationship between cultural institutions (libraries in particular) and SW.

#### 2.3. Linked Data and cultural heritage institutions

Access to cultural heritage information presents several challenges, mainly related to the variety of items included in the collections (and their consequential different requirements in terms of metadata, indexing and retrieval methods), and to the different needs and skills of their users. Efforts have been made, over the past few decades, towards the development of techniques and tools to facilitate access and use of digital cultural heritage. (Ruthven & Chowdhury, 2015).

The access key to all collections, physical and digital, is metadata: it is the tool that allows the description, identification, organisation, retrieval, access, use, conservation, delivery, and preservation of all resources. (Sugimoto et al., 2015). Metadata created by cultural institutions is defined by van Hoolan and Verborgh (2014) as 'legacy metadata', since, unlike metadata produced by other organisations, it is supposed to last the long term. Particularly for digital collections, metadata has the potential of enhancing interoperability and linkage between cultural resources: this can be obtained through sharing information on the adopted metadata schemas. (Sugimoto et al., 2015).

Cultural institutions have always faced issues of findability, searchability, and retrievability of their data (Shiri & Davoodi, 2016), and several different solutions have been elaborated throughout the time, each of which has required a disruptive transformation process. (Van Hoolan and Verborgh, 2014). Metadata used by cultural institutions have been created over the years by several agents and for various purposes, with the result of a relevant degree of heterogeneity in the description of the same entities, and consequential lack of interoperability. (Tallerås et al., 2013).

Various bibliographic standards have been created throughout the years in order to enable the exchange of library metadata across institutions. (Tallerås, 2013).

MARC was developed in the 1960s with the intent of allowing the creation and sharing of library materials in a way that computers could understand. (Shiri & Davoodi, 2016). MARC is a flat model that does not easily support the establishment of relationships among entities. Therefore, information professionals have directed their efforts towards the

elaboration of a model able to convert the MARC scheme into an entity-relationships model, with the final goal for libraries to take advantage of the potential of the web, and increase the visibility of their resources. (Shiri & Davoodi, 2016). The other most widespread standard is represented by the AACR (latest version AACR2). Both (MARC and AACR2) developed prior to relational databases and the web, these standards have been increasingly criticised, for several reasons: one for all, the fact that they aim to create bibliographic records for human reading and interpretation, resulting in text strings that are machine-readable, but allowing very limited actions. (Tallerås, 2013).

The recent revision of AACR2 to RDA intended to bring cataloguing practice closer to the SW principles. (Shiri & Davoodi, 2016). More flexible than its precedents, and adaptable to describe a wider range of digital and physical materials, RDA supports users in carrying out the tasks of finding, selecting, identifying, and obtaining information resources, through the entity-relationship model FRBR (*Functional Requirements for Bibliographic Records*). (Shiri & Davoodi, 2016).

#### 2.3.1. Why would libraries and cultural institutions need Linked Data?

The *Library Linked Data Incubator Group* final report revealed that: library data is not integrated with the web and it is often expressed in natural language; library standards only apply to the library community; and the technological development is driven by the expertise of vendors, rather than by information professionals. (Baker et al., 2011). Another main issue of the traditional approach to library metadata, is the isolation of the various records, as MARC proves, being based on the description of a specific edition of a book, rather than on the work entity. (Tallerås, 2013). The Library Linked Data Incubator Group was born under the W3C umbrella, with the mission to improve global interoperability of library data on the web, by promoting discussion and collaboration amongst people involved in LD projects, working in the library sector and beyond. (W3C Incubator Activity, 2010).

Numerous studies have been conducted with the goal of identifying the benefits that the application of the SW principles may convey to the library and cultural heritage environment.

Marden et al. (2013) recognise to LOD the potential of enabling cultural heritage institutions to share their holdings with a wider audience, changing the traditional

relationship between holder, interpreter, and consumer of knowledge. Linked open datasets can facilitate and enrich the navigation experience of users, empowering them to manipulate and attribute their own meaning to cultural heritage collections.

A research investigating the status of LD implementation across digital libraries by Hallo et al. (2016) highlighted some of the key benefits of this technology:

- Improve data visibility
- Allow linkage to other online services
- Improve open data recovery
- Enable interoperability without affecting data source models
- Allow modelling things of interest related to a bibliographic resource such as people, places, events, themes
- Improve credibility of end user resources annotations.

Enhancing data visibility and findability on the web (OCLC, 2012), where most users currently seek information, is a goal libraries should aim for, in order to remain relevant. Augmented visibility may lead to an increase in the use of library data, hence an increase in the number of library users. (Gonzales, 2014, in McKenna, 2017).

Further strengths of LD are related to the support it offers to multilingualism, since each entity is identified by a URI, rather than by a name. (Shiri & Davoodi, 2016). In addition, LOD provides libraries with the opportunity of creating a big pool of data, shareable, extensible, and reusable, with important savings in terms of time and costs of cataloguing efforts (Marden et al., 2013; Ryan et al, 2015), and it allows the progress and innovation of cataloguing systems, as well as the invention of more creative web-applications based on library resources. (OCLC, 2012; Shiri & Davoodi, 2016). It offers to cultural organisations tools for a better data management, allowing to store, share, and reuse data as needed. (Byrne & Goddard, 2010). Furthermore, the process of interlinking RDF resources with those emerging from other cultural heritage institutions and beyond, could lead researchers to a web of related data with the effort of a single information search. (Hastings, 2015, in McKenna, 2017).

Converting data into informative datasets, institutions establish themselves as trusted sources of quality data, supporting the expansion of the SW. (Marden et al., 2013). Ultimately, LOD contributes to the path towards a seamless and unified access to content

of heterogeneous repositories, which is the final aim of the SW concept. (Shiri & Davoodi, 2016).

The survey conducted by OCLC in 2014, and repeated in 2015, investigating LD implementation amongst libraries across the world, revealed some of the main reasons behind the choice of adopting LD, which reflects part of the above identified benefits attributed to LD: exposing data to a larger audience on the web; demonstrate what could be done with datasets as LD; experiment the LD potential after having heard of this technology; enrich user experience; enhance their own data; improve Search Engine Optimisation (SEO). (Smith-Yoshimura, 2015).

#### 2.3.2. Challenges of Linked Data implementation

The previously mentioned OCLC survey throws a light on the several obstacles which currently prevent a wider adoption of LD amongst cultural institutions. The challenges mostly lamented by libraries, according to the study, are:

- Difficulty in understanding SW concepts by staff members
- Inconsistency in legacy data
- Selection of appropriate ontology to express the data
- Establishing the links
- Lack of documentation and advice on how to build the systems. (Smith-Yoshimura, 2015).

In addition, more technical issues were identified by the OCLC survey, preventing the LD consume, such as: mapping of vocabularies; lack of URIs for each element; lack of authority control; matching, disambiguating, and aligning source data and LD resources. (Smith-Yoshimura, 2015). The most alarming information revealed by the study is the fact that most of the participants declared they would not do anything differently, even though aware of the flaws in their services.

Hanneman and Kett (2010) assort the key issues that LD implementation presents in three main classes:

• **Technical challenges**: LD requires infrastructure, that is a means of data storage (triplestore or database), a webserver, and a resolver that interprets incoming

web requests, translates them into queries for the data storage, and returns the results

- Conceptual challenges: data modelling; several ontologies to choose from; URI specification
- Legal challenges: publication rights and licences of LD.

Privacy concerns have also been raised, considering the danger of exposure of sensitive data that an extensive interlinking implies. (Byrne & Goddard, 2010). On a more general level, Hanneman and Kett (2010) highlight the issues of guaranteeing data quality, as well stability of resources, and, above all, the lack of experience reports, and step-by step instructions for building LD services.

To allow a better understanding of some of the primary difficulties presented by LD implementation, it is useful to outline the main steps required to publish library data as LD:

- 1. Modelling a RDF-structure of properties and classes from existing (or specially created) vocabularies
- 2. Converting the existing data from any of the standards to the new RDF-structure
- 3. Linking entities in the new dataset to entities in other LD resources. (Tallerås et al., 2013).

Whilst Tallerås (2013) argues that difficulties can arise at any point of the process, other authors stress how the most challenging phase is the data integration process, mainly due to the lack/inadequacy of tools supporting this step. (Bergman, 2014; Neish, 2015; McKenna, 2017). In particular, the study carried out by McKenna (2017) underlines how the issues encountered at data integration level affect the ability of institutions publishing LD to interlink their datasets to external resources.

Further challenges identified with regards to LD implementation within libraries are:

 Inadequacy of most adopted bibliographic standards for use in the SW: MARC21 standard does not support the use of relationships between bibliographic entities, that instead both FRBR and RDA internalise. (Cole et al., 2013, in McKenna, 2017).

 Lack of agreements on how to provide data; lack of expertise in different areas of data transformation; lack of indicators about the use of LD, once available. (Hallo et al., 2016).

Although technical challenges are a clear obstacle to a LD widespread adoption across libraries, the biggest issue may be recognised in the lack of awareness of LD existence and potential, alongside with the fact that the general tendency leans towards maintaining and improving existing systems, rather than performing a complete change. (Byrne & Goddard, 2010). Despite the discussion around SW and LD has flourished in recent years, there are still only few concrete examples that prove how LD implementation can benefit cultural heritage institutions. (Neish, 2015). The next section offers a selection of the key projects accomplished by cultural institutions in order to adhering to the SW vision.

#### 2.3.3. Examples of Linked Data use in the cultural heritage sector

In 2011 the Library of Congress (LC), following the findings of the report On the record (Library of Congress Working Group on the Future of Bibliographic Control, 2008) highlighting the inadequacy of MARC to the latest programming systems, announced its intention of developing a new bibliographic framework based on LD principles and RDF. (Library of Congress, 2011). The release of Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services (Library of Congress, 2012) signed the beginning of the path leading to BIBFRAME, a dedicated LD model for bibliographic metadata. (Tallerås, 2013). The BIBFRAME initiative has aimed to provide assistance to libraries during their transition from MARC21 formats to LD standards, ensuring the continuity of data exchange, which has allowed, in the recent decades, cataloguing cost savings and resource sharing. (Neish, 2015). The idea at the source of the project requires information professionals to operate a complete mindset change: rather than embedding data in the record, as with all the traditional standards, links are now provided to authoritative sources. (Bartlett, 2013). BIBFRAME, differently than Schema.org, includes FRBR concepts, and reuses the entity-relationships models. (Hallo et al., 2016). Data can be easily shared, annotated, re-used and linked, allowing cooperative cataloguing at a granular level as never before. (Shiri & Davoodi, 2016). A fragment of the BIBFRAME plan is the Library of Congress Linked Open Data Service, enabling both humans and machines to access data values and the controlled vocabularies (promulgated by the LC) that host them. (Library of Congress, n.d.). The service allows access and bulk download of authority

names and controlled vocabularies at no cost, and permits users to link their metadata to the LC data values. (Marden et al., 2013). These initiatives have placed the LC at the forefront in the LD implementation landscape.

The first version of BIBFRAME, released in 2014, was replaced by BIBFRAME 2.0 in 2016: this revisited version took inspiration for improvement from the criticism moved to the former framework by the Cornell University Library, the Harvard Library Innovation Lab, and Stanford University Libraries. (Suominen & Hyvönen, 2017). The Linked Data for Libraries (LD4L) project accomplished by these three institutions between 2014 and 2016, intended to facilitate the retrieval of scholarly information through the creation of a Scholarly Resource Semantic Information Store (SRSIS) model, able to work both within individual institutions and through a network of LOD resources, to capture the value added by librarians, domain experts, and scholars to information resources. (LD4L, 2016). Starting off converting their datasets into BIBFRAME 1.0, the three institutions resolved to replace some BIBFRAME structures with structures from other RDF vocabularies, and resolved to produce the LD4L ontology. (Suominen & Hyvönen, 2017). An ongoing project is Linked Data for Production (LD4P), which sees the same institutions leading the LD4L, together with LC, Columbia, and Princeton Universities, piloting the production of LD for library resources. The goal of the initiative, running over a two-year period (2016-2018), focuses on "developing standards, guidelines, and infrastructure to communally produce metadata as linked open data". (Branan & Futornick, 2017). An extension of BIBFRAME ontology to describe library resources in specialised domains and formats, as well as the elaboration of best practice for ontology extension modelling and for LD production, are some of the objectives of the project. (Branan & Futornick, 2017).

OCLC (Online Computer Library Centre) has been very active on the LD implementation front. The OCLC's LD initiatives include the publication of: VIAF, which combines multiple name authority files into a single name authority service, in order to match and link widelyused authority files and make that information available on the web (VIAF, 2017); a LD version of FAST (Faceted Application of Subject Terminology), derived from the Library of Congress Subject Headings (LCSH); Dewey Web Services, offering part of the Dewey Decimal classification represented as LD; and the release as LD of its catalogue, WorldCat. (OCLC, 2017). Although WorldCat does not fully apply the SW principles, it represents at least an effort in this direction, adopting the Schema.org vocabularies to provide structured data on the web. (Shiri & Davoodi, 2016).

Probably the most famous LD cultural heritage initiative is *Europeana*: launched in 2008, Europeana is an open-access digital library of cultural heritage information, bringing together millions of digital cultural objects from several European institutions. (Ruthven & Chowdhury, 2015). Users are enabled to search across collections of libraries, museums, and archives, breaking the typical institutional silos. (Thorsen & Pattuelli, 2016).

Another project involving the various protagonists of cultural heritage is *Linked Open Data in Libraries, Archives and Museums* (LODLAM): emerging in 2011, this international community gathers information professionals, researchers, and educators interested in, or working with, LOD pertaining to galleries, libraries, archives, and museums. (LODLAM, 2014). The LODLAM community has expanded and evolved through a series of initiatives, actively contributing to the convergence of cultural heritage communities via the implementation of the SW principles. (Thorsen & Pattuelli, 2016).

Focusing on archival data, the LOCAH project (then continued as *Linking Lives*) aimed to publish as LD data from the *Archives Hub*, the JISC (Joint Information Systems Committee) initiative collecting metadata from over 300 archives in the UK. (LOCAH Project, 2014). The mission was to make archival data available to researchers not familiar with archives. (Marden et al., 2013).

Prior to discussing the initiatives of LD implementation carried out by the NLs across Europe, it is deemed necessary to outline the role of such organisations, in particular with regard to the information and cultural landscape of their respective countries.

#### 2.4. The role of national libraries

Sylvestre stresses the authoritative role of NLs, stating that "The service of the nation's libraries cannot be optimized without the leadership of a national library". (Sylvestre, 1987, p.5). Amongst their primary objectives are: the acquisition of a comprehensive collection of the nation's literature, and the publication of the current national bibliography. Furthermore, NLs work as national centre for the international exchange of bibliographic records. (Sylvestre, 1987).

There is no prescriptive definition that can uniquely describe the wide range of organisations falling under the label of 'national libraries'. According to Wainwright (1993) the characteristics of each NL depend strongly on the size, cultural history, economic, and technological development of the country they serve.

NLs are invested with a complex set of tasks and responsibilities, related to "building, preserving and enabling access to collections at a national level, and providing relevant services throughout a nation". (Breeding, 2011, p.21). They invest significant resources in developing technology systems, with noticeable benefit of other libraries within the same nation. They are responsible for establishing standards for record formats, data exchange, and interoperability protocols. The extent to which these organisations are able to provide cataloguing services to other libraries has a remarkable impact on the methods of processing materials adopted by each cultural institution in the country. (Breeding, 2011).

The mission of creating, preserving (for future generations), and making available (for present users) comprehensive collections of national cultural heritage has encountered increasing difficulties since the advent of digitised and born-digital content. (Breeding, 2011). The digital format is becoming the main format, with consequential more demand on library systems and on the infrastructure needed to acquire, use, and preserve this material. In addition, new uses of data and information are being and will be developed and, to create and maintain adequate services, NLs have to take better advantage of the potential and value of metadata, allowing its reuse and linkage. New systems and technologies require investing resources in library staff education and training, so that they are able to participate in the changes. (Sverrisdottir, 2009).

In quality of memory institutions and custodian of the national identity, NLs carry out the utmost important task of building and preserving, for the long term, digital archives. The projects achieved by the Diet Library (Japan), and the LC have proved the impact that the SW and LOD activities promoted by W3C can have on digital archives. (Sugimoto et al., 2015). The conversion of data to LD operated by national institutions implies that 'local' information services are able to reuse them, with beneficial savings in term of time, costs, and further resources. (OCLC, 2012).

There is no doubting the fact that the library and information environment is changing dramatically, and it is easy to foresee that such tendency will continue. NLs have the potential of acting as leaders in the technological development and establishment of standards, since they are usually placed under the government, hence in a better position for having a say in the national policy making. (Hagerlid, 2011).

Taking into account the responsibilities NLs are invested of with regards to leading the advancement of the systems and technology adopted by compatriot institutions, and
preserving the nation's cultural heritage, together with the consideration of their privileged status enabling them to draw from better resources than the rest of the libraries within the country, it should not surprise that some of the early LD implementers have been in fact NLs. (Neish, 2015).

The attention will now focus on outlining the methodology selected and applied to this research.

# 3. Research Methodology

Ritchie & Spencer (2003) affirm that the aim of qualitative research is to provide a deep understanding of a subject through the investigation of people's circumstances, experiences, and perspectives. The human factor has a major impact in qualitative research. In particular, case studies have proved to be the best tool for in-depth and rich analysis of complex phenomena involving a mix of people and processes. (Pickard, 2007). For the above reasons, qualitative research methodology, in the form of case study, was deemed the most appropriate for the purpose of this dissertation, which values individuals' viewpoints as primary source of information. A mixed technique was chosen for data collection: a combination of interviews, literature review, and analysis of online resources. An integration of different methods was judged the best approach to provide a more comprehensive answer to the research questions. This solution derived from the fact that not all institutions contacted participated in the study, determining a need for seeking information about potential projects realised among the literature. Whereas no information on the status of LD implementation at a certain NL could be retrieved through this second phase, an analysis of online resources was carried out, focusing primarily on the examination of the institution's website.

# 3.1. Literature review

The literature review is a complex process of interpretation of documents available from various sources on a specific topic, involving summarisation, analysis, evaluation, and synthesis of the documents. (Onwuegbuzie et al., 2012).

In order to address the research questions, a literature review was conducted, to understand the current state of LD implementation and development across European NLs. The main databases consulted were LISA, LISTA, and Google Scholar. A remarkable number of articles was retrieved using text string such as "linked data AND national libraries". Nonetheless, many of the retrieved sources were too generic, or related to institutions outside Europe. A search log was used to register variants of search phrases adopted and respective results, allowing to refine search criteria, thus enhancing precision and recall. Alerts were set in order to keep up-to-date with the research in the field throughout the data collection phase: they revealed to be a useful tool for discovering many of the 2017 publications included among the references. In addition, mailing lists were essential to provide awareness of an imminent release of the latest version of the LOD cloud (due in July 2017, then postponed to unspecified date), and the most recent count of DBpedia triples.

In order to identify further useful material, trails were followed, starting from resources considered authoritative and more relevant to the research goals. Representativeness rather than comprehensiveness was the main inclusion criteria for the retrieved material. Author's names which appeared with higher frequency were prioritised, and their sources' bibliography offered the starting point for the identification of further resources. A wide range of sources were considered, including books, articles, conference proceedings, webinars, presentations, and YouTube videos. An Excel document was created, to keep track of all the sources considered. The visual display allowed by the spreadsheet facilitated the recording of codes, and, eventually, the identification of the main themes. Each descriptive label was attributed with a different colour, and a legend with all the codes and respective colours was devised. References were recorded manually on a separate document, which was regularly updated with new entries, and checked for consistency.

To provide an adequate overview of the research topic, and increase awareness of the possibilities of LD application, material related to the development of SW and LD across different sectors was sought. This allowed the identification of the modalities of participation to the SW by organisations such as LC, Google, and Facebook.

For the data processing phase, a constant comparative analysis method was applied, comparing each piece of data with other data that was similar or different, formulating concepts on the possible relationships among data, and building categories in which inserting the processed data. (Pickard, 2007). In this type of examination, rather than being preconceived, categories emerge directly from the analysis of data, allowing more flexibility and inclusiveness. This method can be applied to documents and other resource types. (Onwuegbuzie et al., 2012). For these reasons, it was chosen for the present study, in order to analyse data collected via both literature review and interviews. The constant comparative analysis process involves three different phases of coding: an *open coding* stage, in which data are divided into smaller segments that are all given a code or descriptor; an *axial coding* stage, in which codes are grouped into similar categories; and a *selective coding* stage, wherein no new properties or relationships emerge, instead the main categories are identified and the theory is refined. (Onwuegbuzie et al., 2012).

Applied to the literature review of this research, this method followed the steps outlined below:

- Every set of information, after reading, was divided into smaller, meaningful parts, each of which was assigned with a descriptive label (*open coding*);
- Each new chunk of data was compared with previously determined codes, so that similar portions of data were assigned with the same labels (*axial coding*);
- Once all sets of information were labelled, codes were organised in bigger clusters, by similarity, and themes were identified and described, based on each cluster (*selective coding*).

#### 3.2. Interviews

The primary reason for opting for interviews is their property of providing a way of expressing breadth of perspective and individual viewpoints, which is what was sought by this research, whose purpose was to record, analyse, and compare reflections and considerations that may lead to the formulation of best practices.

Kvale's (1996) seven stages of the interview process (defining the key steps of adopting interviews as data collection method) provided the overall framework for interviews' design, structure, and analysis. The starting point of any interview process is the definition of the research aims and learning outcomes, essential in order to select the interview type and elaborate design and planning.

Semi-structured interviews were considered the most appropriate design option, since characterised by a series of predetermined questions (interview guide), but, at the same time, by a flexibility of what and in which order questions are asked, that can vary according to the answers provided by the interviewee. (Robson, 2002). This allows a more dynamic interaction, positively contributing to a more relaxed and comfortable environment. The existence of the interview guide aimed to guarantee that all the main points were hit, while the adaptable structure allowed a more natural conversation flow.

The interview script (see appendix 2) was formulated on the basis of the main topics that emerged from the preliminary literature review, always keeping in close consideration the research questions. The main themes on which the interviews focused were: uses of LD across NLs; reasons for implementing/not implementing LD; policies and technologies adopted; implementation process; identified issues, challenges, and benefits of implementation; best practice and future development.

Following the guidelines provided by Luo & Wildermuth (2009), after the main topics were defined, a list of all the questions that needed to be asked, in relation to each topic, was compiled. In order to avoid defensive and 'forced cognition' reactions from the participants' side, 'lose' questions (offering opt-outs) were preferred. (Becker, 1998).

As suggested by Berg (2001), next to the 'essential' questions, which aim to collect the key information to answer the research questions, other queries were prepared: 'extra' questions, which offered an alternative way to express or explain the essential question; and 'probing' questions, asked to obtain further elaboration of a received answer. Pilot testing was conducted, with the intent of verifying consistency and clarity, improving the script, and providing participants with details on estimated duration upfront.

With respect to the strategy adopted to approach the institutions, the starting point consisted of compiling a full list of European NLs, based on information extracted from *The European Library* website (The European Library, 2013), and respective e-mail addresses to contact. The literature review assisted the process of selecting candidate participants: the professionals involved in LD implementation projects across European NLs were identified through various sources (webinars, articles, mailing lists, blogs on library websites), particularly for those institutions more active and interested in the SW development, such as Sweden, United Kingdom, Germany, France, the Netherlands, Latvia, Finland, and Spain. In case no direct contact was traceable, a generic e-mail address for the institution was used. This was the only option for those libraries that resulted as not having implemented LD, hence lacking of key informants to identify. Social media such as Twitter and LinkedIn were also used as way of connecting with participants at the earliest stage. A participant information sheet, containing information on aims of the research, estimated duration of the interview, and data treatment, was e-mailed to all the candidate participants (see appendix 1).

Nine interviews were carried out via Skype (Finland, Germany, Open Knowledge Greece, Latvia, the Netherlands, British Library, Wales, Luxembourg, Portugal), while three institutions preferred the option of providing email responses, mainly due to technical or time issues (France, Spain, Czech Republic); three more email exchanges (Italy, Bulgaria, Liechtenstein), although not responding the questionnaire comprehensively, offered

useful information for the overall research purposes. Skype conversations were (upon obtaining consent) recorded and notes were taken, as backup. A notepad was organised upfront with sections for the targeted topics, according to the questions, to facilitate the writing process. The breadth, depth, and quality of information conveyed through email interview did not result of same level as the details provided by the Skype interviews. Nevertheless, it was a valuable tool to reach a more comprehensive picture of the state of LD spread across Europe. Prior to processing data collected, a summary of the key findings of each interview was sent to the respective participant to verify that no misunderstanding had occurred, and to seek clarification in case of doubts.

Transcriptions and email responses were coded applying a constant comparative analysis method, similarly to the processing of the data gathered through the literature review.

#### 3.3. Online resources analysis

The meaning of 'online resources analysis' with respect to this study, can be identified in the concept of 'primary Internet research' or 'Internet-mediated research' (IMR), a methodology aiming to the gathering of primary data from the Internet, in order to provide new evidence to answer the research questions. (Hewson, 2008).

A document analysis was applied to the web pages of some of the NLs in Europe. Document analysis has been defined as the examination of static documents available online, "which do not involve interactions between individuals". (Hewson, 2008, p.554). The investigation focused on the websites of those institutions that did not participate in the interview process. Nonetheless, web pages of other libraries which successfully implemented LD were analysed, with the intent of developing an effective method of extracting primary data from other websites.

The examination of web pages in order to establish if and how this technology had been embraced by an institution required the elaboration of a multiphase method. Before consulting the individual institutions' website, the first port of call was to look for information on Datahub (<u>https://datahub.io/</u>), in order to identify datasets created by NLs. The lists of VIAF and Europeana contributors were also considered, to understand which institutions support these LD datasets.

The following steps focused on the investigation of the individual websites, with the aid of browser extensions, such as *Semantic Radar*, able to recognise SW structures.

Some of the actions taken to extract information from library websites included:

- Typing 'linked data' into the main search bar of the home page
- Examining pages dedicated to collections, catalogues (records were analysed), digital libraries, national bibliographies
- Examining pages focusing on research/projects undertaken by the institution, alongside with collaboration projects
- Consulting website map
- Scanning weblogs and publications
- Considering Library strategy/policies/standards adopted and services offered

Furthermore, useful information was retrieved, in some cases, through the governmental website collecting openly published datasets within the country.

# 3.4. Conclusion

The three data collections techniques adopted resulted in different types of information gathered. Although both literature review and interviews showed similar, recurring themes, such as reasons behind LD implementation, perceived benefits, and experienced challenges, the research interviews provided much richer data, touching many more themes related to collaboration, viewpoints on best practice and future development, and individual reflections on specific situations, that were hard (if not impossible) to identify through the literature review. Furthermore, the online resources analysis enabled to collect information that the other two methods failed to convey. The combination of data gathered through the different techniques provided answers to the research questions, allowing the achievement of the research objectives.

The next section presents the results of this study, outlining the reasons and modalities of LD adoption amongst European NLs, attempting to shed light on the benefits and challenges encountered during the unrolling of the projects.

# 4. Findings and analysis

Information gathered through interviews, literature review, and analysis of online resources is presented here, with the intent of outlining the status of LD implementation across European NLs. Information accompanied by a reference was gathered through literature review, in particular (but not exclusively) for institutions which have not participated in this research; the lack of a reference indicates that the information instead derives from the interviews. Tables 1 and 2 offer a summary of the research results, indicating the main channels through which data has been collected for each institution. It is necessary to bring to the reader's attention that the Greek participant is not a representative of the Greek NL, but of Open Knowledge Greece, which falls under the umbrella of Open Knowledge International, a global non-profit organisation promoting the value of open data for civil society organisations. (Open Knowledge International, 2017). Open Knowledge Greece, active since 2012, has led the main steps taken by the country with respect to LD implementation, for the NL and the wider network of Greek libraries.

Participating National Libraries				
	Implemented	Intending/Planning to implement	Not implemented	
Skype	Finland Germany Greece (Open Knowledge Greece) Latvia The Netherlands United Kingdom (British Library) Wales	Luxembourg Portugal		
Email	France Spain	Italy	Bulgaria Liechtenstein	

Table 1:	Overview of r	esearch fin	dinas – Partic	ipating Na	tional Libraries
	0.00.0000				

Non-participating National Libraries				
	Implemented	Not implemented	Taking steps towards Semantic Web	
Literature review	Hungary Ireland Sweden		Austria Poland	
Web- resources analysis	Georgia Switzerland Ukraine	Albania (www.bksh.al) Armenia (www.nla.am) Azerbaijan (anl.az/new) Belarus (www.nlb.by) Belgium (www.kbr.be) Bosnia and Herzegovina (www.nub.ba) Croatia (www.nsk.hr) Cyprus (www.cypruslibrary.gov.cy) Denmark (www.kb.dk) Estonia (www.nlib.ee) Iceland (landsbokasafn.is) Kosovo (www.biblioteka-ks.org/#) Lithuania (www.lnb.lt) Macedonia (nubsk.edu.mk) Malta (www.maltalibraries.gov.mt) Moldova (www.bnrm.md) Montenegro (nb-cg.me) Norway (www.nb.no) Romania (www.bibliotecadistato.sm) San Marino (www.bibliotecadistato.sm) Serbia (www.nk.sk) Slovenia (www.nk.sk) Slovenia (www.nk.uni-lj.si) Turkey (www.mkutup.gov.tr)		

 Table 2:
 Overview of research findings – Non-participating National Libraries

The examination of online resources delivered the following information:

- Indications of the work towards LD implementation undertaken by the Swiss NL were found through the opendata.swiss portal (opendata.swiss, 2017), which led to the discovery of a paper by Bättig and Schwer (2016), describing the method adopted by the institution to convert MARC21 records into RDF
- Semantic Radar allowed the identification of RDF structures on the website of the NL of Ukraine, which offers a 'Scientific search for publications in WEB 3.0' option (<u>http://www.nbuv.gov.ua/node/1539/</u>)
- SW elements were detected through Semantic Radar on the website of the Digital Library of the National Parliamentary Library of Georgia (<u>http://dspace.nplg.gov.ge/?locale=en</u>)

# 4.1. Linked Data uses across national libraries

The findings have highlighted the various paths chosen by the European NLs to contribute to the SW, usually depending on specific contexts and choices. This section outlines different ways of participating to the LD universe, alongside with the diverse purposes and uses LD was selected for.

# 4.1.1. Contributing to the Linked Open Data cloud

Providing the institution's data to LD datasets, such as VIAF, Europeana, and Wikidata, was indicated by some of the interviewees as a way of participating to the SW development. Table 3 summarises which NLs contribute to VIAF and to Europeana.

VIAF	Europeana
BL, Czech Republic, Estonia, France,	Austria, Croatia, Finland, France,
Germany, Hungary, Iceland, Ireland,	Greece, Iceland, Latvia, Luxembourg,
Latvia, Luxembourg, the Netherlands,	Norway, Poland, Romania, Scotland,
Norway, Poland, Portugal, Russia,	Serbia, Switzerland, Ukraine.
Scotland, Spain, Sweden, Switzerland,	
Vatican City, Wales.	

Table 3: VIAF and Europeana contributors

Findings and Analysis

Although sharing data with such external resources does not imply an active involvement with LD principles and technical requirements, it is a way of taking part into the LOD cloud, as stated by the NL of Latvia (NLL). Furthermore, for those institutions which currently lack the necessary resources to actuate a full LD implementation, sharing data with the main LD platforms represents the only way to contribute to the web of data expansion, as the NL of Luxembourg affirms in relation to their participation to Europeana. The case of the Polish NL may be considered similar, as in 2012 it provided over a million records to Worldcat. (OCLC, 2012).

A deeper form of involvement with Europeana is represented by the leading role assumed by the Austrian NL in the technical coordination of *Europeana DSI-2*, a project aiming to provide a better customer experience for all Europeana's user groups, and improve metadata quality, which also sees NLs of Latvia and Luxembourg as participants. (Österreichische Nationalbibliothek, 2017).

The NLL suggests two reflections connected with the provision of data to these main LD actors: from one side, the institution affirmed that the decision of passing their authority data to VIAF was considerably delayed due to the dilemma of granting their information for free to an organisation providing paid-for services. In addition, the NLL recognises to Europeana the potential role of incentivising the wider LD adoption across cultural heritage institutions, if they would encourage organisations to publish LOD and integrate it with their data, rather than accepting data in any format and then arrange for their conversion into RDF.

# 4.1.2. Linked Data for bibliographic and authority data

This study has confirmed that LD is most frequently adopted by NLs in order to publish their bibliographic and authority data.

Some organisations have accomplished a LD publication of both bibliographic and authority data, such as: BnF; BNE (Biblioteca Nacional de España = National Library of Spain), with *datos.bne.es* (<u>http://datos.bne.es/inicio.html</u>); DNB; NL of the Netherlands; Open Knowledge Greece; Hungarian NL (NektarWiki, 2011); Swedish NL (Library of Congress, 2012b).

*Data.bnf.fr* (<u>http://data.bnf.fr/</u>), one of the earliest achievements of NLs in the SW context, is a discovery tool built around BnF's collections data: each page, automatically generated, provides an overview of related information merged across different data silos such as their catalogue, archives and manuscripts, digital library (*Gallica*), web archives, and virtual exhibitions.

DNB and Finnish NL both started working with authority data, as they deemed it more reusable and easier to model. The German initiative GND (*Gemeinsame Normdatei* = Integrated Authority File), completed in 2010, targeted the integration of all the authority data previously located in separate databases into a unique database. The underlining principle was to look at everything as entities, as in the FRBR model. (Deutsche National Bibliothek, 2016). In 2011, the DNB developed the first prototype for bibliographic data, and since then the institution has been remodelling and continuously expanding the service. The Finnish NL is currently working towards exposing their bibliographic data as LD.

A group of institutions has worked on a LD version of their library catalogues. LIBRIS, the Swedish Union Catalogue (<u>http://libris.kb.se/</u>), began providing LD in 2008 (Stephens, n.d.), establishing itself as one of the earliest LD projects accomplished across the library environment. The key steps taken towards making the library catalogue available as LD are described by Malmsten (2009), with the aim of providing a step-by-step guide for other institutions:

- Find a way to get the records and relations out of the ILS (Integrated Library System)
- Choose a URL pattern
- Map MARC records to RDF
- Implement content negotiation / record delivery.

In 2010, it was the turn of the Hungarian NL to publish its OPAC as LOD. (NektarWiki, 2011). The NL of the Netherlands has worked on the realisation of a National Catalogue, for all the public libraries of the country, based on semantic technologies. Furthermore, the Welsh NL is at present cooperating with Ex-Libris to achieve a LD exposition of their catalogue. Focusing on authority data, the Irish NL contributed to *Linked Logainm* (<u>https://www.logainm.ie/en/inf/proj-machines</u>), a collaborative project that created a LD version of the bilingual database of Irish place names, *Logainm.ie*. (Grant et al., 2013).

The Portuguese NL and the National Central Library of Florence are planning to make their bibliographic information available as LD. The BNCF's (Biblioteca Nazionale Centrale Firenze) is currently looking at different hypothesis for the publication of bibliographic data, trying to avoid from one side the risk of jumping from MARC silos to LD silos (Suominen and Hyvönen, 2017); from the other, the BNCF believes that on LD it is necessary to follow the mainstream. Since the end of 2014 an experimental project has been run by the Italian institution, with the intent of adopting Wikibase (the software in use by Wikidata) to manage bibliographic data, in order to make data more easily discoverable and ready-to-reuse, and to preserve its quality. (Progetto: GLAM/Biblioteca Nazionale Centrale di Firenze, 2017).

The Polish NL, aiming to prepare the path to LD implementation, has recently introduced the *Descriptors* model, to allow better data segmentation to authority and bibliographic data. The goal of the project has been to transform unstructured data into structured data, facilitating access to the Library catalogue, and establishing links between defined entities in the NL database, laying the foundations for embracing the SW. (Cichoń, 2017).

The Austrian NL has considered the publication as LD of the Austrian Library Network's (OBV's) bibliographic datasets. (Danowski et al., 2013). No evidence of the accomplishment of this or other LD related projects has been detected.

#### 4.1.3. Publishing the national bibliography as Linked Data

As affirmed by the Finnish NL to motivate the priority assigned to the LD publication of their national bibliography, this is a resource which should be the prime responsibility of each NL, likely to be judged (especially from an external viewpoint) as the most interesting database a NL has.

In July 2011 the BL released the British National Bibliography (BNB) as LOD. (Deliot, 2014). The choice of working with the BNB rather than with the whole library catalogue was based on the consideration of its specific scope, being an authoritative source of information about UK publications from 1950s to present. In addition, the BNB's metadata is more consistent and well-maintained than the metadata available throughout the

catalogue. Therefore, the BL believed publishing the BNB as LD a more straightforward process compared to processing the entire catalogue.

2011 also saw the publication as LD of the Swedish national bibliography. (Library of Congress, 2012b). The publication as LD of the Finnish national bibliography was achieved by the Finnish NL with *Fennica*. (http://linkeddata-kk.lib.helsinki.fi/). Currently working to make the complete national bibliography available as LD are the DNB and the NL of the Netherlands.

# 4.1.4. Linked Data for digital resources

NLs that have adopted LD in relation to their digital resources are: BnF, with its digital preservation project; Hungarian NL (NektarWiki, 2011); NLL; NL of the Netherlands.

The NLL has achieved two projects in this context: *Digital Object Management* (DOM) (http://dom.Indb.lv/), and *Rainis and Aspazija* (http://runa.Inb.lv). DOM is mainly an internal system, whose LD export functionality was added when developing a digital repository. The linked digital collection *Rainis and Aspazija* is the most recent development by the NLL in the LD field. It is a pilot project testing how to enrich a digital collection with additional links between objects by annotating named entity references, and exposing this information as LD. The lack of a good tool to annotate the content led to the creation of a new annotation tool. An advantage in this context was the fact that the pilot project could reuse LD published by the DOM system, linking to its URIs, without need of duplicates.

The NL of the Netherlands is currently involved in a national programme for digital heritage, together with other cultural national institutions. The aim is to build a discovery infrastructure for cultural heritage in the Netherlands based on LD principles. However, the project is still at a design phase. At an even earlier stage are the plans of the Portuguese NL for the realisation of a digital platform based on geographic resources, linked to GeoNames and similar datasets.

#### 4.1.5. Thesauri and ontologies

LD have been implemented by some institutions for the publication of thesauri and ontology services. *Finto* (<u>https://finto.fi/en/</u>) signed the start of LD adoption at the Finnish NL. The project benefited of funding from the Ministry of Education and Culture, made

available for developing a thesaurus and ontology service for the cultural heritage and public sectors.

The DNB developed the *GND Ontology*, aiming to solve the issue of name ambiguity in the library sector. The GND ontology aligns with existing vocabularies, such as FOAF and RDA Vocabulary, in order to ensure compatibility. (Deutsche National Bibliothek, 2016). It was adopted by Open Knowledge Greece as tool for the first pilot of converting authority data from MARC to LD.

Looking at Italy, the BNCF has made available the Thesaurus *Nuovo Soggettario* (New Subject Heading) on SKOS. (Biblioteca Nazionale Centrale di Firenze, 2015). The LD publication of the Thesauri was one of the LD initiatives achieved by the NL of the Netherlands in 2015.

As revealed by this study, amongst the challenges of LD implementation, the dilemma of reusing established ontologies versus creating new ones is perceived as a key issue still requiring a solution, as it will be further discussed. At this point, it is deemed useful to indicate some of the existing ontologies mostly adopted (and often adapted) by the European NLs: SKOS, FOAF, RDA, Dublin Core Metadata Element Set, GND, BIBO (Bibliographic Ontology).

# 4.2. Implementation process

Albeit the provision of in-depth reports outlining the technical processes carried out by the European NLs during LD implementation does not strictly fall within the objectives of this research, it was nevertheless judged useful to present some of the steps taken towards LD adoption by the institutions considered, prior to focusing on the implementation reasons, benefits, and challenges identified through the interview process.

As stated by the DNB, the LD implementation process is iterative. Although variations, depending on individual choices, need to be taken into account, a concise summary of the key phases involved is offered by the BNE:

- Intensive data analysis
- Data mining, in order to find rules in their data structure to achieve the passage from a record-driven to an entity-driven model

- Vocabularies selection, and thereafter building
- Web and services design, including establishing links to external datasets.



*Figure 6: British Library data model. Available at: <u>http://dataliberate.com/wp-</u> <u>content/uploads/2012/01/British-Library-Data-Model-v1.01.pdf</u>* 

The phase of converting data from various format to LD has resulted as one of the most challenging for most of the institutions, as it will be discussed further on.

An essential phase of the LD implementation process (the key point without which LD ceases to exist) is represented by establishing connections between a dataset and the external resources forming the LD universe. The data gathered shows the popularity of a group of resources which has been more frequently linked to by the various projects achieved by the NLs, among which are: VIAF, DBpedia, LCSH, ISNI (International Standard Name Identifier), GeoNames, and LinkedGeoData.

Linking to external datasets is considered the priority for the Swedish NL, over any other aspect. External data are treated, by the Swedish institution, as if they were internal and vice versa: great attention is paid to updates within internal datasets as well as to updates in external ones. The ultimate goal for the Swedish institution is to fully replace records with links and relational graphs. (Library of Congress, 2012b).

Findings and Analysis

#### 4.3. Reasons for implementation

Investigating the reasons behind LD implementation, this study highlighted some factors common to a few libraries, whilst others specific to the individual institution, and even to the specific projects carried out.

The fact that LD has been main the subject of conferences and initiatives across the library environment has encouraged in several cases the first steps towards implementation by European NLs. The discussion around the web of data, often presented as the way forward, if not an obliged step for libraries to take in order to stay relevant, has produced an increasing curiosity by institutions such as BL, Welsh NL, BNE, NLL, NL of the Netherlands, and Open Knowledge Greece. In some circumstances, curiosity and will of experimenting prevailed, also considering the fact that other NLs had already adopted this technology. For the Finnish NL, the choice of LD has stemmed from the perceived need of "growing awareness of the time", recognised as a key responsibility of a NL, which is invested in the role of leading the development of the compatriot cultural institutions.

Another very popular factor encouraging the adoption of LD technologies is the goal of augmenting data visibility and discoverability on the web, as declared by France, Spain, Portugal, Germany, the Netherlands, and Italy. In order to obtain a better discoverability of their data through search engines, DNB, Finnish NL, and the NL of the Netherlands have adopted Schema.org as data model, which grants, according to the Finnish NL "the advantage to be generic enough to allow exposing many things, but at the same time specific enough that you can provide many details".

Some of the participants have indicated that the decision of implementing LD has derived from the intention of bringing enriched, open, and reusable information to users, as declared by Luxembourg, Portugal, Germany, and the Netherlands. The data enrichment is enabled by establishing links to external resources, as well as by the intensive phase of preparation required at the beginning of LD implementation. Enhancing the usability of existing datasets was mentioned by Wales, BL, and the Netherlands. Improving existing datasets conveys beneficial effects to the users' end. At this regard, the aim of the National Digital Heritage project, currently undertaken by the Netherlands, aims to facilitate the users' browsing experience, overcoming the fragmentation of information held by different institutions in the country.

LD offers, as expressed by part of the participants' viewpoints, a means for making data available in a new way. The BL judged LD as a valid tool to enhance a long-established resource (BNB). Undeniably the implementation was driven, for the BL, by the goal of achieving better results with their existing data.

Furthermore, a positive factor stressed by several interviewees, which supported the decision of embracing this technology, is the added value brought by LD of rendering data reusable for a wide range of purposes, and to the wider community. LD is deemed by several participants as a powerful tool for reaching out beyond the library environment, to open up data silos and make data available across different sectors, as stated by Sweden, Germany, BL, and France. One of the objectives that the DNB had in mind, while designing their LD projects, was to promote to the wider audience of possible stakeholders that their data was available and free to reuse.

For some institutions, the choice of implementing LD has derived from their intention of adhering to established standards or policies adopted at a higher level. For instance, the BnF has recognised the benefits of using standards recommended by W3C, hence the decision of taking steps towards LD, which belongs to this category. The BL's motivation, instead, was to follow the Government policy encouraging public exposure of data. Similarly, although not having yet reached a full LD implementation, the Scottish NL has embraced the open data movement in order to comply with the national strategy. (National Library of Scotland, 2017).

Among the recorded reasons of LD adoption, a category of factors related to the perceived and proven potential of LD could be identified. The declarations of some of the participants revealed how LD was selected as deemed the best fit for the achievement of their objectives, as in the case of data.bnf.fr, whose goal was to create an entity-driven website that was accessible and visible from search engine results, and provide the underlying data as "freely reusable by, and interoperable with datasets from, stakeholders inside and outside the library sector". An initial risk assessment showed the BnF that LD was to be preferred to relational databases, taking into account the aim in mind. Similarly, the Welsh NL explained that LD was selected as most appropriate way of publishing their collection of shipping records, considering that the nature of the content pertained to people, places, and events, easily expressed through an entity-driven model.

Data integration was the reason explaining the choice of LD technology for the national catalogue of the Netherlands, since they meant to combine their bibliographic records with other resources, in order to provide people searching the catalogue with different kinds of information. For the BNE, the intention was to investigate the multilingual capabilities attributed to LD.

#### 4.4. Benefits of Linked Data adoption

The findings of this study have exposed some of the perceived benefits attributed by institutions to the implementation and use of LD, which were in some cases confirmed by feedbacks and other means of measuring value. It is interesting to notice that some of the perceived benefits overlap with the reasons adduced by participants for LD implementation. In particular, the potential of LD for improving data visibility and discoverability, the opportunity of a wide data reuse and re-purpose, and the LD role in enhancing data quality. With regards to the latter matter, the Netherlands defines publishing LD as a form of "social contract": each organisation is responsible for the data provided, and is invested with the task of ensuring that data stays available in a persistent way, not to disrupt users. According to the institution, LD involves a deep reflection on "what you are doing and if you are doing it the proper way, which is quite useful".

In the majority of cases, the outcomes of the initiatives carried out did not encounter the expected advantages of LD implementation. Reflecting on the matter of data reuse, the DNB declares to have not had the broad uptake they hoped for, especially within the industrial sector. Nonetheless, this is an aspect they are currently working on, in order to improve the library data reuse across different environments. Similarly, the NLW affirms that they had foreseen a higher number of users of their datasets, identifying the main causes of this unsatisfactory result in the lack of outreach and promotion of the service, as well as the in high level of expertise required to realise a good level of LD service.

The NLL implemented LD with the goal of being the authority hub to which other people would refer, but this target was not achieved, as in fact authority data has not been provided by the institution, due to internal misunderstandings. They expected to improve the discoverability of their data, but they are not really able to prove that this objective has been met either. With regards to the benefit of integrating information systems internally, the Latvian institution believes that this goal was reached, although not

primarily thanks to LD, as information integration does not necessarily require LD adoption.

The BL and the NLL have expressed satisfaction with regards to the achievement of their set objective of undertaking LD implementation in order to acquiring a better understanding and competence of LD and SW, which can be considered as a consequence of the general interest raised by the subject across the information science environment.

Several benefits are recognised to LD by the Swedish NL. First of all, its disambiguation ability, since each element is provided with a unique identifier. (Library of Congress, 2012b). A deriving advantage is the support that LD provides to multilingualism, allowing to obtain multilingual data in different formats, as stressed by Greece and Spain. Furthermore, according to the Swedish NL, LD offers a way to get more out of the cataloguing efforts, allowing a considerable workload reduction, since it enables libraries to reuse information already available elsewhere. (Library of Congress, 2012b).

At the boundary between reason adduced and perceived benefit is the target expressed by few of the interviewees of implementing LD in order to achieve an authoritative position as bibliographic/authority data provider, to which other institutions would refer to.

# 4.5. Challenges of Linked Data implementation

The challenges presented here gather different clusters of issues which have hindered, temporarily or definitely, LD implementation across European NLs.

Lack of resources has emerged as a key obstacle for both implementers and nonimplementers. In most cases LD initiatives are developed by a limited dedicated taskforce within the institution. Lack of human resources has in fact represented one of the main barriers preventing LD implementation for Portugal and Luxembourg. Financial restraints as well had a negative effect for many institutions. Such challenging situations are derived from the fact that LD projects are not considered priorities to the overall mission and strategy of the institution.

Another primary challenge that had to be faced by LD adopters, causing several delays to the undertaking of projects, is the lack of knowledge and expertise in this area among information professionals, also denounced as main barrier by non-implementers.

Findings and Analysis

Reflections gathered revealed both the lack of technical skills necessary for initiating LD adoption, alongside with a general deficiency of awareness of the SW and LD principles amongst library staff. Explicative the case of the BL: being amongst the early implementers, they could not take advantage of other examples, having to do most of the work from the scratch. Similarly, as early adopter, BnF had to face issues of lack of practical and technical knowledge of LD requirements, which determined a strong case on whether or not to choose this technology. The development of internal IT skills, enabled through the experience conveyed by previous projects, contributed to a positive outcome and to the accomplishment of data.bnf.fr.

The identification of a knowledge gap carries with it the reflection on the importance of staff training: as stated by the Welsh NL, the acquaintance of the SW principles requires an active participation of developers to conferences, in order to learn about latest developments, as "it is not something that can be learned remotely".

Strategic issues were also detected amongst the main challenges to overcome. Some of the participant institutions have reported the struggle of obtaining the buy-in from senior management in order to get their projects started. As the NLL suggests, high-level management needs to be presented with tangible proofs of LD benefits, which can demonstrate the validity of the investment required. On a different note, the BNE has reported difficulties determined by the lack of a well-defined strategy to shape and clarify steps and objectives of their LD initiative since the start.

A barrier to a wider LD adoption was recognised, on the basis of the perspectives gathered, in a traditional approach commonly assumed by the library professionals: the general tendency is focused on maintaining and improving the systems already in use, rather than opting for a radical change of direction. This was not only the reflection of an external viewpoint (Open Knowledge Greece), but also of NLs' staff members. The Netherlands argued that the 'centralistic' approach shaping the services developed at their institution, for which each project would consist of aggregating data in one place, and then start processing it, has hindered LD implementation at first. LD does not require the aggregation phase, as its advantage is to link to entities in external resources. However, this position proves the challenge related to convincing people to work in a different way than the one they are used to. LD implies a radical change of mindset, which is hard to achieve. Furthermore, Greece stressed how libraries are often unwilling to provide data, as data is

considered a personal resource of the institution, with consequential reluctance of the organisations to open and make their information available.

Besides, challenges have been encountered by participant institutions in relation to rights and licensing issues. For instance, the BL stated that, shortly after the undertaking of the initiative, they soon realised that data would have been hard to reuse, unless published under a CCO licence, with the consequential issue that they could not track which purposes data was downloaded for.

On the same matter, the Netherlands affirms that LD makes sense only in LOD form, that is to say, published under CCO license. The explanation provided is that metadata is the way to render data findable, hence there should not be any barrier to hinder access. The Swedish NL adopted a CCO licence to publish the national bibliography and the authority file, to guarantee that data would be freely reusable. (Library of Congress, 2012b). Although not having implemented LD, the Portuguese NL have also adopted a CCO licence to make their data available.

If publishing data under a permissive license (not requiring registration, feedback, or mention of data provider in case of reuse) is the essential condition to full reusability, it also determines the difficulty, if not impossibility, of tracking who is using LD and for what scope. This factor impedes the measurement and demonstration of the usability, usefulness, and uptake of data, as stated by BL and DNB. The BL has succeeded in finding a partial solution to this problem, thanks to its collaboration with Fujitsu International, which provided the tools to obtain a much clearer picture of their LD users.

A series of obstacles more strictly connected to the technical aspects of the implementation process have been also recognised. What was pointed out by several NLs as a key issue is the lack of examples, and clear step-by-step instructions to guide the implementation phases, alongside with the scarcity of adequate tools and infrastructure to support LD. Furthermore, lack of a uniform standard to transform data into RDF and thus LD, has been identified as a primary challenge by many participants, since this negatively contributes to make the conversion more complex, and hinders interoperability.

Severe difficulties were caused by the various ways of publishing LD. For many NLs this resulted in a long process towards the identification of the best way to present data in the most useful way. A major issue is also represented by the fragmentation of the various

ontologies/vocabularies available, and by the choice of adopting an established ontology or creating a new one. Greece highlighted how, even a few years after the first project was realised, there is still not a specific ontology to describe library data.

There is no doubting that LD requires an elaborate and intensive data processing, as well as an intensive effort in maintaining the data quality with regular and efficient updates: as the BL argues, maintaining the infrastructure over the long term requires considerable resources. One of the most demanding phases of LD implementation for some of the participants has been the formulation and attribution of URIs, alongside with guaranteeing their persistency. The Netherlands declared to have spent over a year formulating a policy to establish how to determine URIs. In addition, the BnF underlined the fact that, being still a niche technology (especially if compared to formats such as JSON and CSV) RDF risks being an obstacle to data reuse, rather than supporting it.

#### 4.6. Most popular examples of Linked Data implementation

Early implementers could not count on many examples from which drawing guidelines, as instead was the case for later LD adopters. Nevertheless, it is interesting to notice that among the most popular models for NLs, with respect to LD implementation, were in fact other NLs. The institutions' names most frequently mentioned include: Swedish NL, BnF, DNB, BNE, and BL, with each of them frequently addressing the others as source of inspiration.

Alongside with carefully looking at the achievements of the Swedish NL, which is described as 'pathfinder' in the area, the BL considered the experience of the LODLAM initiative, and the case of the German project Hbz (Hochschulbibliothekskurs des Landes Nordrhein-Westfalen, <u>https://www.hbz-nrw.de/ueber-uns</u>). The NLL stated to have closely observed the examples offered by VIAF, Europeana, and the LC. The latter is also mentioned as inspiration model by Greece and Portugal.

The NL of the Netherlands declared that sources of important learning outcomes for the LD implementation were for them the results obtained by OCLC and DBpedia. Other cultural heritage institutions working with the SW were also taken into account, such as the Amsterdam Museum. The emphasis lays on the idea of a process of mutual exchange of experience and knowledge, together with drawing lessons from any available context. Furthermore, the Netherlands pointed out the Semantics Conference in Amsterdam

(Semantics Amsterdam, 2017), where different sectors (particularly business) gather, as an opportunity to realise how working with the SW and creating smart LD solutions is becoming a business trend. Adding more sources of inspiration, the BnF recognises the BBC website as an example they looked at, and Wikidata as a current proof of LD usefulness. On a different note, the NLW declared to have considered as a primary model the IIIF (International Image Interoperability Framework), a standard for sharing and working with digitised images, which expresses data in JSON/LD format. IIIF is regarded as a proof of the LOD usefulness by the Welsh institution.

The viewpoint of the BNE has underlined that, although successful results have been obtained, particularly useful and striking case studies able to prove the full LD potential are yet to come.

# 4.7. Policies

None of the respondents declared that their institution holds a written policy covering all the aspects of LD implementation and use. An explanation of this situation can be recognised in the fact that, in most cases, the unrolling of LD initiatives is considered as a low-priority project. Self-explanatory is the approach assumed by the Finnish NL, where a small task force gathers monthly to discuss plans and achievements, drafting policies as they proceed. Nevertheless, some of the NLs stated that LD use is mentioned in the general strategy adopted at higher levels within the organisation.

The Netherlands affirms to have developed a policy for determining URIs, which required a very long elaboration process. The same matter has been tackled by France and Spain. The BNE has identified URI creation, maintenance, and persistence as a key objective guiding their LD policy. Recognising that using URIs associated to entities is a novelty for libraries, Spain affirms that their aim is to create URIs for entities which are able to interact together. The BnF has formulated a set of guidelines shaping their approach (defined 'mixand-match') to LD, implying: reuse of existing classes/properties whenever appropriate and relevant; creation of maintain specific classes/properties whenever needed; use of persistent and dereferenceable URIs.

The general strategy of the NL of the Netherlands explicitly mentions LD as being systematic approach to share their data and make it feasible in the best way, as well as the necessity of adopting Schema.org to obtain better uptake with search engines. Similar

is the case of DNB: their long-term strategy states the goal of building knowledge networks, and refining the LD service is among their strategic priorities for 2017-2020.

The NLL, although not possessing a policy on LD, as LD is not contemplated amongst the highest priorities, reported of a reference to implementing and developing LD, and becoming a hub for authority data, in their general strategy. Publishing data in LD format is also mentioned as one of the tasks of the Latvian Institute of Bibliography, which falls under the umbrella of the NL. In addition, the NLL has developed guidelines to briefly outline what LD is, and how to put it into practise.

Since the BL does not have an infrastructure for LD and hosts their data on external platforms, an actual policy on the matter is not perceived as needed. The BNB project began as an experimental initiative of one of the departments of the Library, the Collection Metadata service. Therefore, there was no requirement for a policy regulating LD; but very clear licensing policies, for managing all metadata formats, are in act. Similar is the case of the Portuguese NL: although the institution does not have a LD policy, due to the fact that no definitive decision nor plans for implementation have been made, a strategy of "open and linked data" is adhered to.

#### 4.8. Collaboration

The question directed to understand the general perception on the relationship between collaboration and LD development has received a wide variety of responses.

The BL highlighted how collaboration can assume different forms. In their case, the relationship with TALIS responded to the need of integrating a gap of knowledge, experience, and technical infrastructure. On the other hand, the relationship with TSO is of a commercial rather than collaborative nature, since it is a paid-for service; but still essential, as TSO provides the BL of an infrastructure they would not otherwise have. The collaboration with Fujitsu International has been beneficial in order to understand who the users of their LD service are. In this case, it was a mutual exchange, as the BL provided a source of rich data useful for Fujitsu International to test their software.

The DNB believes that collaboration is essential at several levels, in particular at content and technical level. The DNB's experience in the matter has seen the collaborative efforts shared with other German speaking libraries publishing LD (such as the Swiss NL), with the goal of consolidating the various data models adopted. At present, this cooperation has

reached the completion of the second iteration of recommendations on how to model bibliographic data in RDF across Germany. The invitation of the German NL is towards a more pragmatic cooperation on issues such as data modelling, how to implement new technologies, the ways of describing data, and how to achieve interoperability at content level. Another observation made by the DNB concerns the diversity of the community interested in LD within the library sector, which makes the dialogue between the parts crucial: the call is for forums where issues, ideas, and solutions can be discussed.

DBpedia has been named as a valid partner by both Germany and the Netherlands. DNB is working to integrate library data into DBpedia to a greater scale in order to do text mining, and also automated categorisation of articles, which is the research project DBpedia is currently dealing with. The NL of the Netherlands benefited of the knowledge of a DBpedia expert during the unrolling of their project. A close cooperation with DBpedia was also established by the Swedish NL, which was also involved with the Swedish Cultural Heritage Board and the Swedish National Archive in the organisation and delivery of a series of workshops aiming to publish data as LD, as well as improving existing datasets, creating interfaces overlying LD datasets, and establishing links to external resources. (Lindström & Malmsten, 2015).

The key role played by conferences as collaborative occasion, and source of useful learning, was underlined by several respondents. Finland, Germany, and the Netherlands referred to SWIB (Semantic Web in Bibliotheken, <u>http://swib.org/</u>) as a good opportunity for information professionals working with LD to exchange experiences and knowledge.

For the NL of Luxembourg collaboration is essential to keep up-to-date with the latest developments, and, more importantly, it represents for them the only way to participate in the SW advancement, since lack of resources has prevented a full LD implementation. The main achievement of the Luxembourgish institution in the SW direction has been the project based on establishing mutual links between their catalogue and the Dictionary of Luxembourgish Authors (*Autorenlexikon*), produced by the National Centre for Literature. A pilot was run between 2012 and 2013, to investigate the possibility to create links between the authors in ALEPH (bibliographic/authority records in the Library catalogue) and their corresponding authors in *Autorenlexikon*. (Popistasu, n.d.). Cooperating with Europeana, they are able to contribute to the LOD cloud with their data, even if not directly. Similarly, the NL of Portugal declares that collaboration is the only way they may

be able, in the foreseeable future, to actively participate in the LD development, as they currently lack of the necessary resources.

The NLL stresses the need for a more active collaboration among institutions within the same country, suggesting that LD implementation should be part of a vision at national level, "almost part of the development strategy of the country information system". The NLW mentions two collaborations that benefited them for providing tools and expertise: the relationship with BBC RES, essential source of knowledge at the start of their experience with LD, to the point that they would not have probably completed the first project without their consultancy; and their link with IIIF, which is instead "more a collaborative process of getting in the right direction".

An added advantage of collaboration is identified by the Netherlands in its role of demonstrating the value of LD: the more organisations cooperate towards the LD development, the easier would be to bring resources together in a relevant way for users.

The Finnish NL declares its satisfaction with respect to the current status of collaboration: "The community is finally trying to get together and look at these matters together". The Finnish institute regards at the event organised by LD4P and LD4L-Labs (occurred in April 2017 at Stanford University) as the first opportunity for professionals to meet and discuss the future of bibliographic LD, perhaps signing the beginning of a deeper cooperation amongst institutions.

BNE states that better cooperation will be enabled by a wider number of institutions sharing their data, assigning to main connection hubs such as VIAF, Wikidata, and Geonames, the key role of gateways between datasets. Wider participation, alongside with enhanced interoperability, allows institutions to specialise and focus on a specific subject, hence supporting further development.

# 4.9. Standardisation and interoperability

As for the theme of collaboration, the topic of standardisation and interoperability has generated a diversified range of responses.

The BL argued that LD requires adherence to the SW standards, but this is not necessarily advantageous for institutions with a long metadata tradition, since it forces intensive data processing in order to prepare data for conversion. According to the BL, whilst

standardisation within the library community was understood in the past on the basis of everyone doing the same thing, following restrictive sets of guidelines, this is no longer sustainable. LD offers an alternative, since it allows "identification of entities with an opaque URI, around which you can cluster any number of different labels, to suit the needs of different communities, but all referring back to the same entity".

Adoption of common standards should be a primary goal for the information and library environment, according to Greece, which adds what a powerful tool LD is, in terms of reaching interoperability at international level, beyond linguistic barriers.

An interesting and comprehensive answer on the standardisation subject was provided by the NLL. First of all, the interviewee clarified that, if referring to the different serialisations of RDF (such as RDF/XML or Turtle) then it does not matter which serialisation is chosen, since they are basically equivalent; and libraries who deal with LD would be normally able to convert between them and consume any of them. Content negotiation comes useful in this context, as it allows users to obtain data in different formats. The advice of the NLL is to provide representations of data in multiple formats, to ensure that whoever has preferences for one or the other can still understand the data. Furthermore, the Latvian institution raises the issue of ontologies and vocabularies adoption, admitting that, although sustaining the reuse of existing ones, in practice, it is often easier to develop a new ontology that better fits the institution's data; an issue also highlighted by NLW and BL. In most cases, even when using standard ontologies, an adaptation process is still required. Whatever the case is, what would be helpful for data users is to provide mapping from the specific vocabulary used to common vocabularies. The topic of mapping is also touched on by the DNB, with reference to the MACS initiative (Multilingual Access to Subjects, http://www.dnb.de/EN/Wir/Kooperation/MACS/macs.html), aiming to understand if mapping of subject headings in different languages is achievable. The project was started about ten years ago by OCLC, BL, BnF, and Swiss NL, to map subject headings in German, French, and English. The DNB is currently working to publish the complete MACS dataset.

The importance of adopting standards is well-recognised by implementers as well as by non-implementers. A consideration coming from the NL of Luxembourg, which was also formulated by the Swedish NL few years ago (Library of Congress, 2012b), is the utmost importance for institutions of small countries to adopt international standards and follow what the bigger institutions are doing, to avoid the risk of being left adrift.

The Czech NL declares that they have always been aware of the importance of the adoption of standards to guarantee data exchange and interoperability. Their work in this direction includes the publication of the national bibliography, sharing their records with both Czech and foreign languages institutions, and staying up-to-date with the international standards development. The Portuguese NL has recently replaced local metadata schema with international standards.

A general tendency across libraries to move from XML standards towards LOD, encouraged by the promise that the conversion will allow better interoperability, has been noticed by the Welsh NL. The related reflection is that converting XML into LOD does not necessarily guarantee interoperability, unless people agree on the standards.

The viewpoint of the Netherlands is that standards adoption, in the LD context, is the way both humans and machines can 'understand' the information provided. Although discussion can be fruitful, the institution identifies the threat of focusing the debate around the 'perfect model', rather than looking for pragmatic solutions to shape services for the users' benefit. The advantage that the Netherlands attributes to LD is that it allows to adopt various ways of modelling data, and bring them together without necessarily having to choose a unique way. However, in general, the need for a better consent, at least across the library sector, on which way to choose towards LD implementation, is recognised. The Finnish NL foresees that "if the current trend continues, if everyone keeps using different ways to create LD, the data published as LD will not be well used, as it is very difficult to combine, or do a comparative analysis of, different databases".

BNE attributes to LD the "potential of allowing interoperability far beyond what has been achieved" so far, offering the advantage, also identified by the Netherlands, of guaranteeing full interoperability even when adopting different models and vocabularies. The attention, with LD, has moved from content to structure level.

The reflection offered by the BnF is that, while adopting the core grammar (RDF, RDFS, OWL, SKOS, SPARQL) is crucial in building interoperability across different datasets, each implementer is then free to adapt this with a choice of a specific data model based on various vocabularies. "The greatest challenge in this is to have a simple and easily understandable data model that makes easier data consumption, and to maintain its classes and properties over time as reused vocabularies are maintained and other vocabularies are created and adopted by the community."

# 5. Recommendations and Conclusions

This section presents the recommendations collected through interviews and literature review, with the intent of supporting institutions which are contemplating the future adoption of LD as a possibility, as well as those who are in need of maintaining existing datasets. Furthermore, suggestions for research directions are outlined. A summary of the key findings emerged in this study, together with a brief self-assessment of the research, will conclude this work.

# 5.1. Best practice

In order to provide advice to new implementers, it was deemed appropriate to interrogate, on the matter of best practice, institutions who have already experienced working with LD. Below is a summary of the responses collected from the interviewees on the topic of recommendations for making the first steps towards LD adoption:

- Make use cases and carefully evaluate if LD is the right technology for your scope
- Start working with LD, even on something small, in order to learn
- Take advantage of the increasing resources available to support implementation
- Look at examples offered by successful projects
- Get in touch with LD implementers, through conferences or other means, to get a feeling of best practices
- Seek for expert developers to carry out the implementation outside the institution, if necessary
- Focus on data specific to your institution (e.g. national bibliographies for NLs)
- Ensure to have a community of stakeholders wider than just the library community
- Consider URI syntax: reflect on how you want to identify your data and keep it available permanently
- Reuse data, whenever possible (e.g. reuse national authority files, if already published elsewhere)
- Collaborate with local universities and benefit from their expertise in matters such as ontologies modelling

- Ensure you can count on professionals who understand both the technical and the content sides
- Adopt an entity-based approach to data
- Design a careful roadmap, a detailed strategy, before acting
- Use existing vocabularies, whenever possible

According to the Netherlands, LD adoption across the library and information sector would highly benefit of a clear message coming from the higher sphere, possibly at government level, rather than being the choice of an individual institution. In this sense, the cases of BL and Scottish NL adhering to governmental policies, are emblematic. Recognising the need for LD at national level, elaborating a strategy to be cohered to by all cultural heritage institutions, would represent a powerful way for supporting the SW development. Therefore, the need for a clear message and an incentive encouraging libraries towards the SW has been identified.

Further recommendations, gathered by the OCLC surveys previously referred to, include (Smith-Yoshimura, 2015):

- Focus on goals, rather than technical matters
- Pick a problem you can solve
- Consider legal issues from the start
- Develop a good understanding of LD structure, available ontologies, and your own data
- Strive for long-term data reconciliation and consolidation
- Involve your institution/community

Promoting awareness of LD has emerged as a primary necessity in order to favour a wider adoption of such technology. This study has demonstrated the usefulness of mailing lists, webinars, and conferences in order to stay up-to-date with latest development, and as source of advice and best practice.

Recommendations should also address potential risks for implementers. One of the issues underlined by this research is related to the persistency of URIs: maintaining solid and permanent URIs should be a priority for cultural institutions, considering their key function in the LD context. (Van Hoolan and Verborgh, 2014). Likewise, great attention should be dedicated to maintaining metadata and metadata schemas as logical instance, in order to keep their semantic consistency. (Sugimoto et al., 2015). In order to create a high quality service, institutions should aim to provide their datasets with resolvable URIs, a SPARQL end-point, and the opportunity of bulk downloads. (Suominen & Hyvönen, 2017).

The identified lack of awareness and expertise denounces a strong need for promoting training across library staff. (Byrne & Goddard, 2010). In addition, the technical difficulties determined by the variety of data models, should encourage libraries to work together to find as much common ground as possible, rather than elaborating new solutions. (Suominen & Hyvönen, 2017). This matter will be expanded in the next section focusing on further research. Suominen & Hyvönen (2017) encourage libraries to have a close dialogue with their users, to investigate and better address their needs, and to reach a practical understanding of how they want to produce LD. Furthermore, the authors express a consideration on the necessity of a more open and transparent collaboration between institutions.

The study has shown that the cases of well-defined policies regarding LD services are a minority. The *Library Linked Data Incubator Group Final Report*, in its final section dedicated to recommendations, which gathers most of the advice previously outlined, invites institutions to develop policies for managing RDF vocabularies and their URIs. (W3C Incubator Group, 2011). Among the report's advice, as summarised by Hallo et al. (2016), figure the need for: finding a way of obtaining feedback on data usage; tools for visualisation, mapping refinement, and data analysis; a better discussion about rights of LOD; community participation in cataloguing and quality control of published data; improved visibility through adhering to initiatives such as Schema.org.

To summarise, as argued by the Swedish NL, "Linked data is about contributing what you can: you work on what is unique to you and you make it available". (Library of Congress, 2012b).

# 5.2. Further research

The completion of this study has revealed some key aspects on which further efforts and research are required in order to facilitate the future participation of libraries in the LD and SW conversation.

The digital humanities have the potential to lead the debate on the LD matters. In order to do so, they should closely collaborate with engineers to ensure and ease the access to cultural long-lasting values, rather than opting for an opportunistic behaviour aiming to

the publication of their own datasets in the LOD cloud. (Van Hoolan and Verborgh, 2014). Besides, the Netherlands has identified the need of a better communication between developers and system providers from one side, and library professionals on the other, as the first often state that they do not offer to libraries LD technologies since it is not what libraries demand, while the latter lament that their intention of implementing LD is hindered by lack of tools made available by the vendors.

Recognising the importance of adopting LD to the achievement of a web in which all data is connected to each other, van Hoolan and Verborgh (2014) claim the necessity of finding solutions towards the automation of the linking process, which results time-consuming and too demanding if carried out manually. If SW standards do not fully satisfy the needs and features of library data, then libraries have the ability and the opportunity to expand existing standards to make them more relevant and useful. (Shiri & Davoodi, 2016).

What has emerged from this study as the key priority on which future research should focus is the urge for libraries to agree on a common model, in order to reduce the complexity of data integration. (Svensson, 2013). At present, there is no one unique model fit for all different data kept by libraries. Cooperation between cultural institutions, and between the information environment and potential data consumers from various sectors, is essential to find a common pattern. (Svensson, 2013). The achievement of metadata interoperability will depend on the ability of cultural institutions to agree on standards. (Breeding et al., 2016). The tool for succeeding will be the elaboration of a holistic and integrated strategy guiding the design of search and discovery systems, with the intent of providing seamless and inclusive access to data and information. (Shiri & Davoodi, 2016).

#### 5.3. Key findings

Different to other international standards (such as RDA, which became enforced upon an official date) the LOD implementation and use across libraries has been an uneven and gradual process, currently still at its early stages. (Frederick, 2017).

The potential role played by libraries in easing the access to the web of data is seen by Stuart (2011) as natural evolution of their task of facilitating the access to the web of documents, with the only difference of the SW's requirement of a closer engagement with

the latest technology development. The author's conclusion is quite drastic: libraries failure to embrace the web of data will determine the end of librarianship.

The message that most clearly emerged from this research is the need for information professionals to cooperate towards the identification of solutions most apt to fix those issues that still hinder the wide participation of libraries to the SW. Discussing each other's results, confronting projects and achievements, finding ways to overcome obstacles that may be tailored to different circumstances, rather than built ad hoc for individual cases, are best practices that institutions should embrace. Most of the participants claimed the necessity of increasing awareness of LD potential; goal that can be achieved showing the positive results obtained, collecting feedbacks, looking at examples of successful LD applications across sectors, and drawing learning outcomes. In a time of budget cuts and limited resources, there is no wonder if librarians are struggling to receive approval for projects presenting no guarantee of success and advantage. Implementing LD for the only reason of following a trend is likely not to be a sufficient motivation, especially for smaller institutions with limited resources. It is essential to find a way to measure the value that LD can convey and spread this awareness. Positive feedbacks and concrete demonstration of LD potential would be the most powerful tool to support the SW cause. We have seen, from the declarations of some of the interviewees, that NLs recognise themselves responsible for guiding the technological advancement of their compatriot institutions, being able to benefit from richer resources, and from a more influential and close position to government bodies. In a world that sees an exponential growth of digital items, it is utmost important that NLs take action to preserve the national cultural heritage for the long term, adopting the most advanced and effective tools. LD can offer a means to increase visibility and discoverability of library data, often buried into the deep web. At the same time, LD can alleviate the dramatically increasing workload of cataloguers, offering a way to benefit from information drawn from authoritative sources, rather than having to generate records already available elsewhere. The Swedish NL's vision about the future of library metadata sees the end of the record, as intended so far, and its replacement with relationships and links among resources. (Library of Congress, 2012b). LD has the potential of uniquely identifying entities, with a disambiguating function that allows the overcoming of linguistic barriers. Nonetheless, the findings have revealed that many institutions struggle to put into action plans focusing on the application of the SW principles. Even some of the institutions that can count on most advanced systems found

themselves facing several issues in their path towards LD implementation, one of the primary ones being lack of knowledge and clear step-by-step guidelines on how to proceed.

The recorded viewpoints permitted to highlight an obstacle which is, at present, far to be solved, and that may potentially sign the end of the information science sector's participation to the SW discourse: the lack of agreement on matters such as data modelling and ontologies adoption. Libraries may miss the opportunity of taking advantage of LD if failing to find a common implementation path, with the result of passing from facing the problem of 'MARC silos' to having to deal with the newly created 'LD silos'. (Suominen & Hyvönen, 2017).

We are still some distance from the world of linked data that Tim Berners-Lee, James Hendler, and Ora Lassila envisaged fifteen years ago when they first proposed a Semantic Web. [...] For the time being, we must content ourselves with more modest ambitions for linked open data. First among these is the agreement on standards, for without agreed-upon standards, linking becomes much more problematic and labour-intensive. (Jones & Seikel, 2016, p.ix).

Another aspect to consider is that, while the discussion across the library environment has mainly focused on using LD for representing bibliographic and authority data, an improved understanding of the wide range of the possible LD applications should be sought.

With LD, libraries have the chance of taking the leadership as providers of seamless access to information sources, not only for their users, but also for the wider web community. (Shiri & Davoodi, 2016). This requires commitment, but first of all, collaboration and disposition to share achievements and converge efforts towards the elaboration and design of (ideally) universally applicable standards and solutions. Contexts such as conferences and forums can assist the discussion in order to better address the still unsolved key issues identified in this context. The dialogue needs to involve not exclusively library professionals, but also vendors and system developers, to enable an improved understanding of each other's goals and priorities, and establish a more fruitful cooperation; users, and the wider community of stakeholders, should also be consulted, to ensure that LD services are tailored to best support their needs.

#### 5.4. Research self-assessment

The methodology selected for carrying out this study succeeded in answering the research questions stated in the introductory chapter. In particular, interviews resulted as the most useful way for gathering participants' viewpoints related to their experiences of LD implementation, allowing the identification of encountered challenges and barriers to overcome, alongside with perceived and proved benefits. Nonetheless, the research has presented some challenges.

The phase of participants' recruitment required a long process of identification of the most apt people to contact, through literature review, website analysis, and third parties. In most cases, first attempts to establish contact proved unsuccessful, and required followup. Not all the libraries replied to the invitation, as foreseen.

Resources which could have been extremely helpful in understanding which of the European NLs have implemented LD were found to be out of date (therefore of limited use), in particular the surveys related to the contributors of the LOD cloud, and of Datahub. This determined the need for an online analysis, an intensive and time demanding phase, which did not provide guaranteed outcomes, although the browser extensions employed revealed to be useful in spotting LD users not otherwise identified. An obstacle slowing this phase was related to linguistic barriers, as in most cases the English version of various websites was available for a limited number of webpages, or sometimes non-existing, with the consequence of requiring the assistance of Google translating tools. In addition, technical disruptions were often experienced whilst browsing several library websites.

Overall, taking into account pitfalls and limits, the combination of the three different techniques employed has allowed the delivery of a satisfactory outcome.

#### 5.5. Discussion

The interview process allowed the collection of rich data that, once processed, has conveyed information not previously identified through the literature review. In particular, as hoped at the phase of methodology selection, interviews turned out to be a good tool for gathering individual perceptions and viewpoints.

The study has showed that, despite the achievements of several European NLs, LD is still considered a niche technology, and does not have a wide presence within the information
sector. Approximately a third of the NLs in Europe has made efforts towards contributing to the SW, or is considering to do so. Nonetheless, within each organisation, a very restricted number of staff members is usually in charge of planning and carrying out LD implementation, in most cases weighted as side, non-primary projects. With few exceptions, such as the DNB case, a long process to obtain the senior management buy-in has delayed the initialisation of LD-related projects.

In general, a striking element emerging from this study is that each case of LD implementation is a case on its own, meaning that LD has been adopted in different ways, for various purposes, making use of different tools. With regards to the purposes, NLs have turned to LD technologies to publish their authority/bibliographic data and national bibliographies, as well as building applications for digital resources.

Amongst the adduced reasons for implementation are the goal of augmenting data visibility and discoverability on the web, and making data available for reuse, in a way that is understandable to all users, within and beyond the library environment. It is interesting noticing how several institutions have stated that the idea of embracing the SW was in the first place encouraged by the popularity of this subject across the library and information sector: that is to say, LD being presented as the future for library data management was the trigger which pushed many institutions to experiment with it, and learn about it.

However, some institutions have manifested disappointment about the results achieved and the path towards them, as well as reservations about the truthfulness of promising visions of the LD application to the library context. On this note, it is interesting to report the doubts expressed by the NLL on the real potential of LD, which has not yet been proved in a satisfactory way. In fact, several challenges have been identified by the various institutions, which have hindered the adoption of LD. Lack of human and financial resources, together with lack of expertise and clear guidelines on how to proceed, resulted among the primary obstacles. Issues related to rights and licensing were mentioned by BL and NLL. It is worth underlining that the only 'external' (to the library and information sector) viewpoint recorded, that is the Open Knowledge Greece experience, reported a common approach amongst library professionals to consider data as a 'property' of the institution who has collected it, with the consequential tendency to keep it close rather than opening data up. Although a general turn towards opening data has been registered in recent years, there are reasons that explain libraries' reticence: for instance, the BL states that converting the BNB to LD has entailed the end of the income previously

63

generated by subscription fees required to access the resource. Besides, a significant issue, agreed by many institutions, derives from the fact that making data available under a CCO license (not implying registration nor the citation of data source), it is extremely difficult, with current tools, to track who is downloading the data and for which purpose, which can understandably represent cause of concern for data providers.

With regards to viewpoints of the benefits of LD, many of the answers overlapped with the reasons underlying LD implementation. Comprehensibly, what are the perceived and proven values of this technology, have encouraged the undertaking of the many projects accomplished by the European NLs.

What can best support the LD cause is positive feedbacks from users, as well as clear demonstration of LD value by implementers, delivered through use cases, as the NLL suggests. Collaboration is recognised by most of the participants as a powerful instrument towards LD development, and a better participation of libraries to the SW conversation. Conferences in particular are valued as an effective means for spreading awareness of the latest achievements in the field, and a great opportunity for mutual learning. The DNB laments the scarcity of forums where to discuss technical issues, which could facilitate the initiation of LD projects for new implementers. Albeit the situation seems to have slightly improved in recent years, in terms of resources available to support institutions taking the first steps towards LD implementation, unsolved issued have been identified which still require solution, such as the dilemma between adopting existing ontologies or building new ones, as mentioned by BL and Wales.

# References

Abele, A., McCrae, J.P., Buitelaar, P., Jentzsch, A. and Cyganiak, R. (2017) *Linking Open Data cloud diagram 2017*. Available at: <u>http://lod-cloud.net/</u> (Accessed: 10 August 2017).

Antoniou, G., Groth, P., van Harmelen, F. and Hoekstra, R. (2012) *A Semantic Web primer.* 3<sup>rd</sup> edn. Cambridge, Massachusetts: MIT Press.

Aste, M., Mataloni, M.C. and Martinelli, L. (2015) 'Linked Data: Il mondo di internet e il ruolo delle biblioteche, degli archivi e dei musei'. *Digitalia*, pp.65-73.

Baker, T., Bermès, E., Coyle, K., Dunsire, G., Isaac, A., Murray, P., Panzer, P., Schneider, J., Singer, R., Summers, E., Waiters, W., Young, J. and Zheng, M. (2011) *Library Linked Data Incubator Group final report, W3C.* Available at: http://www.w3.org/2005/Incubator/IId/XGR-IId-20111025/ (Accessed: 10 August 2017).

Bättig, Y. and Schwery, P. (2016) *Linked Open Data in der Schweizerischen Nationalbibliothek*: *Modellierung der Titeldaten*. Available at: <u>http://ead.nb.admin.ch/ogd/2016-03-21\_Modellierung-Titeldaten.pdf</u> (Accessed: 10 August 2017).

BBC (2017) *Ontologies*. Available at: <u>http://www.bbc.co.uk/ontologies</u> (Accessed: 10 August 2017).

BBC (2017) *BBC Things*. Available at: <u>http://www.bbc.co.uk/things/</u> (Accessed: 10 August 2017).

Becker, H.S. (1998) *Tricks of the trade: How to think about your research while you're doing it.* Chicago: University of Chicago Press.

Berg, B.L. (2001) *Qualitative research methods for the social sciences*. Boston: Allyn and Bacon.

Bergman, M. (2014) A decade in the trenches of the Semantic Web. Available at: <u>http://www.mkbergman.com/1771/a-decade-in-the-trenches-of-the-semantic-web/</u> (Accessed: 10 August 2017).

Berners-Lee, T. (2000) *Weaving the web: The past, present and future of the World Wide Web by its Inventor*. London: Texere.

Berners-Lee, T. (2006) *Linked Data*. Available at: <u>http://www.w3.org/DesignIssues/LinkedData.html</u> (Accessed: 10 August 2017).

Berners-Lee, T. (2008) *The Linked Open Data movement, W3C*. Available at: <u>http://www.w3.org/2008/Talks/0617-lod-tbl/#(3)</u> (Accessed: 10 August 2017).

Berners-Lee, T. (2010) 'Long live the web'. *Scientific American*, December. Available at:

http://www.cs.uu.nl/docs/vakken/b1iuw/materiaal/lit/Long%20Live%20the%20Web%20

-%20A%20Call%20for%20Continued%20Open%20Standards%20and%20Neutrality.pdf (Accessed: 10 August 2017).

Berners-Lee, T., Hendler, J. and Lassila, O. (2001) 'The Semantic Web'. *Scientific American*, 284(5), pp.29-37. Available at:

https://pdfs.semanticscholar.org/566c/1c6bd366b4c9e07fc37eb372771690d5ba31.pdf (Accessed: 10 August 2017).

Biblioteca Nazionale Centrale di Firenze (2015) *Nuovo Soggettario.* Available at: <u>http://thes.bncf.firenze.sbn.it/thes-dati.htm</u> (Accessed: 12 August 2017).

Bizer, C. (2009) 'The emerging web of Linked Data'. *The Semantic Web*, September-October, pp.87-92.

Bizer, C. (2011) 'Evolving the web into a global data space'. *28th British National Conference on Databases (BNCOD2011)*, Manchester, UK. 12 July 2011. Available at: <u>https://pdfs.semanticscholar.org/0022/125472ff6705e5e68bc421be4e633c7de8bd.pdf</u> (Accessed: 10 August 2017).

Bizer, C., Heath, T. and Berners-Lee, T. (2009) 'Linked Data: The story so far'. International Journal on Semantic Web and Information Systems, 5(3), pp.1-22.

Branan, B. and Futornick, M. (2017) *Linked Data for Production (LD4P)*. Available at: <u>https://wiki.duraspace.org/pages/viewpage.action?pageId=74515029</u> (Accessed 07 August 2017).

Breeding, M. (2011) 'The special challenges for national libraries'. *Computers in Libraries*, 31(5), pp.21-24.

Breeding, M., Kroeger, A. and Sandy, H.M. (2016) 'Sharing metadata across discovery systems', in Spiteri, L. (ed.) *Managing metadata in web-scale discovery systems*. London: Facet Publishing, pp.17-55.

Brindley, L. (2002) 'The role of national libraries in the twenty-first century'. *Bodleian Library Record*, 17(6), pp.464-481.

British Library (n.d.) *The British National Bibliography*. Available at: <u>http://www.bl.uk/bibliographic/natbib.html</u> (Accessed: 10 August 2017).

Byrne, G. and Goddard, L. (2010) 'The strongest link: Libraries and Linked Data'. *D-Lib Magazine*, 16(11/12). Available at: http://www.dlib.org/dlib/november10/byrne/11byrne.html (Accessed: 10 August 2017).

Calaresu, M. and Shiri, A. (2015) 'Understanding Sematic Web: A conceptual model'. *Library Review*, 64(1/2), pp.82-100.

Cichoń, M. (2017) National Library of Poland: Descriptors model as an attempt of opening library data for reuse. Available at: <u>http://library.ifla.org/1617/</u> (Accessed: 10 August 2017).

Cole, T.W., Han, M.J., Weathers, W.F. and Joyner, E. (2013) 'Library MARC records into Linked Open Data: Challenges and opportunities'. *Journal of Library Metadata* 13(2-3), pp.163-196.

Coyle, K. (2009) 'Metadata mix & match'. *Information Standards Quarterly*, 21(1), pp. 9-11. Available at: <u>http://www.kcoyle.net/isqv21no1.pdf</u> (Accessed: 10 August 2017).

Danowski, P., Goldfarb, D., Schaffner, V., Seidler, W. (2013) *Linked (Open) Data: Bibliographische daten im Semantic Web. Bericht der ag linked data an die verbundvollversammlung*. Available at:

http://eprints.rclis.org/20981/1/Patrick%20Danowski%2C%20Doron%20Goldfarb%2C%2 OVerena%20Schaffner%2C%20Wolfram%20Seidler\_Linked%20%28Open%29%20Data\_Bi bliographische%20Daten%20im%20Semantic%20Web.pdf (Accessed: 12 July 2017).

Datahub (n.d.) *About*. Available at: <u>https://old.datahub.io/about</u> (Accessed: 10 August 2017).

DBpedia (2017) *About*. Available at: <u>http://wiki.dbpedia.org/about</u> (Accessed: 26 June 2017).

Deliot, C. (2014) *Publishing the British National Bibliography as Linked Open Data*. Available at: <u>http://www.bl.uk/bibliographic/pdfs/publishing\_bnb\_as\_lod.pdf</u> (Accessed: 26 June 2017).

Deutsche National Bibliothek (2014) MACS: Multilingual Access to Subjects. Available at: <u>http://www.dnb.de/EN/Wir/Kooperation/MACS/macs.html</u> (Accessed: 10 August 2017).

Deutsche National Bibliothek (2016) *Integrated authority file*. Available at: <u>http://www.dnb.de/EN/Standardisierung/GND/gnd\_node.html</u> (Accessed: 10 August 2017).

Deutsche National Bibliothek (2016) *Linked Data service of the German National Library*. Available at:

http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkeddata\_node.html (Accessed: 17 July 2017).

Dunsire, G. and Willer, M. (2011) 'Standard library metadata models and structures for the Semantic Web'. *Library Hi Tech News*, 28(3), pp.1-12.

Finto (n.d.) Available at: https://finto.fi/en/ (Accessed: 10 August 2017).

*GND ontology* (2017) Available at: <u>http://d-nb.info/standards/elementset/gnd</u> (Accessed: 10 August 2017).

Frederick, D.E. (2017) 'Disruption or revolution? The reinvention of cataloguing (data deluge column)'. *Library Hi Tech News*. Available at: <u>https://doi.org/10.1108/LHTN-07-2017-0051</u> (Accessed: 10 August 2017).

Freudenberg, M. (2017) 'ANN: DBpedia version 2016-10 released', listhub.w3.org, 4 July. Available email: <u>public-lod@w3.org</u>.

Gonzales, B.M. (2014) 'Linking libraries to the web: Linked Data and the future of the bibliographic record'. *Information Technology and Libraries*, 33(4), pp.10-22.

Grant, R., Lopes, N., Ó Carragáin, E. and Ryan, C. (2013) Using the Linked Logainm dataset. Available at:

http://apps.dri.ie/locationLODer/docs/using\_linked\_logainm\_en.pdf (Accessed: 10 August 2017).

Hagerlid, J. (2011) 'The role of the national library as a catalyst for an open access agenda: The experience in Sweden'. *Interlending & Document Supply*, 39(2), pp.115-118.

Hallo, M., Luján-Mora, S., Maté, A. and Trujillo, J. (2016) 'Current state of Linked Data in digital libraries'. *Journal of Information Science*, 42(2), pp.117-127.

Hannemann, J. and Kett, J. (2010) *Linked Data for libraries*. Available at: <u>http://www.ifla.org/past-wlic/2010/149-hannemann-en.pdf</u> (Accessed: 10 August 2017).

Hastings, R. (2015) 'Linked Data in libraries: Status and future direction'. *Computers in Libraries*, 35, pp.12-16.

*Hbz* (2016) Available at: <u>https://www.hbz-nrw.de/ueber-uns</u> (Accessed: 10 August 2017).

Heath, T. (n.d.) *Linked Data: Connect distributed data across the web: Frequently asked questions.* Available at: <u>http://linkeddata.org/faq</u> (Accessed: 19 June 2017).

Heath, T. (2009) 'Linked Data? Web of Data? Semantic Web? WTF?' *Tom Heath's Displacement Activities*, 2 March. Available at:

http://tomheath.com/blog/2009/03/linked-data-web-of-data-semantic-web-wtf/ (Accessed: 21 June 2017).

Hewson, C. (2008) 'Internet-Mediated Research as an emergent method and its potential role in facilitating mixed methods research', in Hesse-Biber, S.N. and Leavy, P. *Handbook of emergent methods*. New York: The Guilford Press, pp.543-570.

IFLA (2017) National Libraries section. Available at: <u>https://www.ifla.org/national-libraries</u> (Accessed: 10 June 2017).

Jones, E. and Seikel, M. (2016) *Linked Data for cultural heritage*. London: Facet Publishing.

Jøsevold, R. (2016) 'A national library for the 21st century – Knowledge and cultural heritage online'. *Alexandria: The Journal of National and International Library and Information Issues*, 26 (1), pp.5-14.

Kvale, S. (1996) Interviews: An introduction to qualitative interviewing. London: Sage.

Latvian National Library (n.d.) *About the collection*. Available at: <u>http://runa.lnb.lv</u> (Accessed: 10 August 2017).

Latvian National Library (n.d.) *DOM Access*. Available at: <u>http://dom.lndb.lv/</u> (Accessed: 10 August 2017).

LD4L (2016) LDL4 2014. Available at: <u>https://www.ld4l.org/ld4l-2014/overview</u> (Accessed: 07 July 2017).

Library of Congress Working Group on the Future of Bibliographic Control (2008) On the record. Available at: <u>https://www.loc.gov/bibliographic-future/news/lcwg-</u> <u>ontherecord-jan08-final.pdf</u> (Accessed: 10 August 2017).

Library of Congress (2011) A Bibliographic Framework for the digital age. Available at: <u>https://www.loc.gov/bibframe/news/framework-103111.html</u> (Accessed: 10 August 2017).

Library of Congress (2012) *Bibliographic Framework as a Web of Data: Linked Data model and supporting services.* Available at: <u>https://www.loc.gov/bibframe/pdf/marcld-report-11-21-2012.pdf</u> (Accessed: 10 August 2017).

Library of Congress (2012) *Linked data at the National Library of Sweden*. Available at:

https://www.youtube.com/watch?v=eY2eAKS\_OrU&index=15&list=PLPaeTNUMteonaX1 5ayHDkhKI6WE2u92Hn (Accessed 12 July 2017).

Library of Congress (n.d.) *Library of Congress Linked Data service: About.* Available at: <u>http://id.loc.gov/about/</u> (Accessed: 08 July 2017).

Lindström, N. and Malmsten, M. (2015) 'Building interfaces on a networked graph' in Cervone, F.H. and Svensson, L.G. (eds.) *Linked Data and user interaction. IFLA Publications, volume 162.* Berlin/Boston: De Gruyter.

LinkedDataTools.com (2017) *Introducing graph data*. Available at: <u>http://www.linkeddatatools.com/introducing-rdf</u> (Accessed: 22 June 2017).

*LinkedGeoData.org* (2017). Available at: <u>http://linkedgeodata.org/About</u> (Accessed: 27 June 2017).

Linked Jazz (n.d.) Available at: https://linkedjazz.org/ (Accessed: 27 June 2017).

*Linked Logainm.ie* (2017) Available at: <u>https://www.logainm.ie/en/inf/proj-</u> <u>machines</u> (Accessed: 10 August 2017).

LOCAH project (2014) Available at: <u>http://locah.archiveshub.ac.uk/about/</u> (Accessed: 07 July 2017).

LODLAM (2014) *About*. Available at: <u>http://lodlam.net/about/</u> (Accessed 07 July 2017).

Luo, L. and Wildemuth, B.M. (2009) 'Semistructured interviews', in Wildemuth, B.M. *Applications of social research methods to questions in information and library science*. Westport, Conn.: Libraries Unlimited, pp.232-241.

McKenna, L. (2017) 'Engaging librarians in the process of interlinking RDF resources'. *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10250, pp.216-225.

Malmsten, M. (2009) *Exposing library data as Linked Data*. Available at: <u>https://pdfs.semanticscholar.org/2e07/91d88a65cb2517e284c2bfca02b7c6660f30.pdf?</u> <u>ga=2.76240145.545710788.1502378221-2111383068.1502378221</u> (Accessed: 10 August 2017).

Marden, J., Li-Madeo, L., Edelstein, J. and Whysel, N. (2013) 'Linked Open Data for cultural heritage: Evolution of an information technology', *31st ACM International Conference on Design of Communication*. Greenville, North Carolina, USA. 30 September - 01 October 2013. Available at: <u>https://core.ac.uk/download/pdf/27297984.pdf</u> (Accessed: 04 June 2017).

National Library of Scotland (2017) *Data*. Available at: <u>https://data.nls.uk/</u> (Accessed: 10 August 2017).

Neish, P. (2015) 'Linked Data: What is it and why should you care?'. *The Australian Library Journal*, 64(1), pp.1-10.

NektarWiki (2011) National Széchényi Library (national library of Hungary) on the Semantic Web. Available at: <u>http://nektar.oszk.hu/wiki/Semantic\_web</u> (Accessed: 10 August 2017).

Daithi O Crualaoich (2010) 'Linked Data at the Guardian' *Open Platform blog*, 18 October. Available at: <u>https://www.theguardian.com/open-platform/blog/linked-data-open-platform</u> (Accessed: 26 June 17).

OCLC (2012) Linked Data for libraries. Available at: https://www.youtube.com/watch?v=fWfEYcnk8Z8&t=10s&index=14&list=PLPaeTNUMte onaX15ayHDkhKI6WE2u92Hn (Accessed: 17 June 2017).

OCLC (2012) National Library of Poland to add 1.3 million records to WorldCat. Available at: <u>http://www.oclc.org/en/news/releases/2012/201269.html</u> (Accessed: 10 August 2017).

OCLC (2017) *WorldCat: Data strategy*. Available at: <u>https://www.oclc.org/en/worldcat/data-strategy.html</u> (Accessed: 07 July 2017).

OCLC Research (2014) *Linked Data survey results* 4–Why and what institutions are publishing (Updated). Available at: <u>http://hangingtogether.org/?p=4167</u> (Accessed: 04 May 2017).

OCLC Research (2017) International Linked Data survey. Available at: <u>http://www.oclc.org/research/themes/data-science/linkeddata.html</u> (Accessed: 10 August 2017).

Onwuegbuzie, A.J., Leech, N.L. and Collins, K.M. (2012) 'Qualitative analysis techniques for the review of the literature'. *The Qualitative Report*, 17(28), pp.1-28. Available at: <u>http://nsuworks.nova.edu/tqr/vol17/iss28/2</u> (Accessed: 21 May 2017).

Opendata.swiss (2017) *Bibliographic data from Swiss libraries, archives and other institutions.* Available at: <u>https://opendata.swiss/en/dataset/bibliographic-data-from-swiss-libraries-archives-and-other-institutions</u> (Accessed: 10 August 2017).

Open Knowledge International (n.d.) *About*. Available at: <u>https://okfn.org/about/</u> (Accessed: 10 August 2017).

Österreichische Nationalbibliothek (2017) *Europeana DSI-2*. Available at: <u>https://www.onb.ac.at/en/research/research-activities-at-the-austrian-national-library/europeana-dsi-2/</u> (Accessed: 25 July 2017).

OWL (2013) Available at: https://www.w3.org/OWL/ (Accessed: 22 June 2017).

Pickard, A.J. (2007) Research methods in information. London: Facet Publishing.

Popistasu, R. (n.d.) *Linked Data: From ALEPH/PRIMO to Autorenlexikon and back*. Available at: <u>http://www.slideserve.com/violet/roxana-popista-u-it-staff-roxana-maria-popistasu-bnl-etat-lu</u> (Accessed: 25 July 2017).

Progetto: GLAM/Biblioteca Nazionale Centrale di Firenze (2017) Available at: https://it.wikipedia.org/wiki/Progetto:GLAM/Biblioteca\_Nazionale\_Centrale\_di\_Firenze# Wikibib (Accessed: 18 July 2017).

Public Library of Veroia Linked Data project (2013) Available at: http://libver.math.auth.gr/drupal/?q=node/17 (Accessed: 10 August 2017).

Rasmussen-Pennington, D. (2016) 'Demystifying Linked Data: Are you ready for what's next?' *CILIP Update*, July-August, pp.34-36.

*RDFa* (2017) Available at: <u>https://en.wikipedia.org/wiki/RDFa</u> (Accessed: 27 June 2017).

*RDFS* (2010) Available at: <u>https://www.w3.org/2001/sw/wiki/RDFS</u> (Accessed: 22 June 2017).

Ritchie, J., and Lewis, J. (2003) *Qualitative research practice: A guide for social science students and researchers.* London: Sage.

Robson, C. (2002) *Real world research: A resource for social scientists and practitioner- researchers.* 2<sup>nd</sup> ed. Oxford: Blackwell.

Ruthven, I. and Chowdhury, G.G. (2015) *Cultural heritage information: Access and management*. London: Facet Publishing.

Ryan, C., Grant, R., Ó Carragáin, E., Collins, S., Decker, S., and Lopes, N. (2015) 'Linked Data authority records for Irish place names'. *International Journal of Digital Libraries*, 15, pp.73-85.

Semantics Amsterdam 2017 (2017) Available at: <u>https://2017.semantics.cc/</u> (Accessed: 10 August 2017).

Shiri, A. and Davoodi, D. (2016) 'Managing Linked Open Data across discovery systems', in Spiteri, L. (ed.) *Managing metadata in web-scale discovery systems*. London: Facet Publishing, pp.57-90.

Shotton, D. (2013) 'Libraries and Linked Data #6: Why publish library catalogues as open linked data?' *Semantic Publishing*, 1 March. Available at: <u>https://semanticpublishing.wordpress.com/2013/03/01/lld6-catalogues-and-linked-data/</u> (Accessed: 29 June 2017).

Singhal, A. (2012) 'Introducing the Knowledge Graph: Things, not strings' *Google: Official blog*, 16 May. Available at:

https://googleblog.blogspot.co.uk/2012/05/introducing-knowledge-graph-thingsnot.html (Accessed: 26 June 2017).

SKOS (2013) Available at: <u>https://www.w3.org/2001/sw/wiki/SKOS</u> (Accessed: 22 June 2017.

Smith-Yoshimura, K. (2015) *Linked Data implementations—Who, what and why?* Available at: <u>http://www.oclc.org/research/themes/data-science/linkeddata.html</u> (Accessed: 04 March 2107).

Söderbäck, A. (n.d.) *Why libraries should embrace Linked Data*. Available at: <u>https://code4lib.org/files/LIBRIS\_code4lib.pdf</u> (Accessed: 11 July 2017).

SPARQL (2017) Available at: <u>https://www.w3.org/2001/sw/wiki/SPARQL</u> (Accessed: 22 June 2017).

Spivak, N. (2010) 'Eliminating the need for search: Help engines'. *Nova Spivack,* January 16. Available at: <u>http://www.novaspivack.com/uncategorized/eliminating-the-need-to-search</u> (Accessed: 25 June 2017).

Springer Nature (2017) *Springer Nature Scigraph*. Available at: <u>http://www.springernature.com/gp/researchers/scigraph?countryChanged=true</u> (Accessed: 25 June 2017).

Stephens, A. (2016) 'Functions, tasks and roles of national libraries in the 21st century'. *Alexandria: The Journal of National and International Library and Information Issues*, 26(2), pp.145-198.

Stuart, D. (2011) *Facilitating access to the web of data: A guide for librarians*. London: Facet Publishing. Strauss, A. L. and Corbin, J. M. (1998) *Basics of qualitative research: Techniques* and procedures for developing grounded theory. London: Sage

Sugimoto, S., Nagamori, M., Mihara, T. and Honma, T. (2015) 'Metadata in cultural contexts: From manga to digital archives in a Linked Open Data environment', in Ruthven, I. and Chowdhury, G.G. *Cultural heritage information: Access and management*. London: Facet Publishing, pp.89-112.

Sturtevant, R. (2010) *Linked Data (Semantic Web) candies.* Available at: <u>https://www.flickr.com/photos/reedsturtevant/4288406152/in/photostream/</u> (Accessed: 10 August 2017).

Suominen, O. and Hyvönen, N. (2017) 'From MARC silos to Linked Data silos?'. *O-bib*, 4(2). Available at: <u>https://www.o-bib.de/article/view/2017H2S1-13/5883</u> (Accessed: 10 August 2017).

Svensson (2013) 'Evolution of bibliographic data exchange'. *Information Standards Quarterly*, 25(4), pp.5-13. Available at:

http://www.niso.org/apps/group\_public/download.php/11931/FE\_Svensson\_bibliograp hic\_models\_isqv25no4.pdf (Accessed: 10 August 2017).

Sverrisdottir, I.S. (2009) 'The next ten years in national libraries'. Alexandria: The Journal of National and International Library and Information Issues, 21(2), pp.56-70.

SWIB17: Semantic Web in Libraries (2017). Available at: <u>http://swib.org/</u> (Accessed: 10 August 2017).

Sylvestre, G. (1987) 'Guidelines for national libraries'. Paris: General Information Programme and UNISIST (United Nations Educational, Scientific and Cultural Organisations). Available at:

http://unesdoc.unesco.org/images/0007/000761/076173eb.pdf (Accessed: 15 June 2017).

Tallerås, K. (2013) 'From many records to one graph: Heterogeneity conflicts in the linked data restructuring cycle'. *Information Research*, 18 (3). Available at: <u>http://www.informationr.net/ir/18-3/colis/paperC18.html#.WUEwnZB95dg</u> (Accessed: 16 June 2017).

Tallerås, K., Massey, D., Dahl, J.H.B.D. and Pharo, N. (2013) 'Ordo ad chaos -Linking Norwegian black metal' in *Libraries, black metal and corporate finance: current research in Nordic library and information science*, 23, pp.138-152. Available at: <u>https://www.diva-portal.org/smash/get/diva2:883968/FULLTEXT01.pdf</u> (Accessed: 17 July 2017).

TED (2009) *Tim Berners-Lee: The next web of open, linked data*. Available at: <u>https://www.youtube.com/watch?v=OM6XIICm\_qo&t=6s</u> (Accessed: 04 May 2017).

The Bibliographic Ontology (2009) Available at: <u>http://bibliontology.com/#</u> (Accessed: 10 August 2017).

The European Library (2013) *Discover contributors*. Available at: <u>http://www.theeuropeanlibrary.org/tel4/discover/contributors</u> (Accessed: 10 August 2017).

*The Open Graph protocol* (2014) Available at: <u>http://ogp.me/</u> (Accessed: 27 June 2017).

Thomson Reuters (n.d.) *About Open Calais*. Available at: <u>http://www.opencalais.com/about-open-calais/</u> (Accessed: 27 June 2017).

Thorsen, H.K. and Pattuelli, C. (2016) 'Linked Open Data and the cultural heritage landscape', in Jones, E. and Seikel, M. (eds.) *Linked Data for cultural heritage*. London: Facet Publishing.

Van Hoolan, S. and Verborgh, R. (2014) *Linked Data for libraries, archives and museums: How to clean, link and publish your metadata*. London: Facet Publishing.

VIAF: Virtual International Authority File (2017) Available at: <u>https://viaf.org/</u> (Accessed 07 July 2017).

Volz, J., Bizer, C., Gaedke, M. and Kobilarov, G. (2009) 'Discovering and maintaining links on the Web of Data', in *Proceedings of the 8th International Semantic Web Conference*, Chantilly, VA. October 25-29, 2009, pp.650-665.

W3C (2016) *W3C Mission*. Available at: <u>http://www.w3.org/Consortium/mission</u> (Accessed: 22 June 2017).

W3C Incubator Activity (2010) *W3C Library Linked Data Incubator Group*. Available at: <u>http://www.w3.org/2005/Incubator/Ild/</u> (Accessed: 07 July 2017).

W3C Incubator Group (2011) *Library Linked Data Incubator Group final report*. Available at: <u>http://www.w3.org/2005/Incubator/IId/XGR-IId-20111025/</u> (Accessed: 18 May 2017).

W3C (2013) *W3C Semantic Web activity*. Available at: <u>http://www.w3.org/2001/sw/</u> (Accessed 25 June 2017).

Wainwright, E. (1993) 'The national library in an electronic age: Dinosaur or catalyst?'. *Alexandria: The Journal of National and International Library and Information Issues*, 5(2), pp.111-118.

Wallis, R. (2013) 'Linked Data for libraries: Great progress, but what is the benefit?' *Semantic Web in Libraries (SWIB13) Conference*, Hamburg, Germany. 26 November 2013. Available at: <u>https://www.slideshare.net/rjw/linked-data-for-libraries-great-progress-but-what-is-the-benefit</u> (Accessed 07 July 2017).

World Wide Web Consortium (2017) Available at: https://en.wikipedia.org/wiki/World\_Wide\_Web\_Consortium (Accessed: 22 June 2017).

Yu, L. (2007) *Introduction to the Semantic Web and Semantic Web services*. Boca Raton, FL: Chapman & Hall/CRC.

# Glossary

AACR	Anglo-American Cataloguing Rules
AACR2	Anglo-American Cataloguing Rules, 2 <sup>nd</sup> edition
ΑΡΙ	Application Programming Interface
BIBFRAME	Bibliographic Framework
BIBO	Bibliographic Ontology
BL	British Library
BNB	British National Bibliography
BNCF	Biblioteca Nazionale Centrale Firenze = National Central Library of Florence
BNE	Biblioteca Nacional de España = National Library of Spain
BnF	Bibliothèque National de France = National Library of France
СС	Creative Commons
CC0	"no copyright reserved" option in the Creative Commons toolkit
CSV	Comma-separated values
DNB	Deutsche National Bibliothek = National Library of Germany
DOAP	Description of a Project
DOM	Digital Object Management
FAST	Faceted Application of Subject Terminology
FOAF	Friend of a Friend
FRBF	Functional Requirements for Bibliographic Records
GND	Gemeinsame Normdatei = Integrated Authority File
HTML	HyperText Markup Language
НТТР	Hypertext Transfer Protocol

- IFLA International Federation of Library Associations and Institutions
- IIIF International Image Interoperability Framework
- ISNI International Standard Name Identifier
- JISC Joint Information Systems Committee
- JSON JavaScript Object Notation
- LC Library of Congress
- LCSH Library of Congress Subject Headings
- LD Linked Data
- LD4L Linked Data for Libraries
- LD4P Linked Data for Production
- LOD Linked Open Data
- LODLAM Linked Open Data in Libraries, Archives and Museums
- MACS Multilingual Access to Subjects
- MARC MAchine-Readable Cataloging
- NL National Library
- NLL National Library of Latvia
- NLW National Library of Wales
- OCLC Online Computer Library Center
- OWL Web Ontology Language
- RDA Resource Description and Access
- RDF Resource Description Framework
- RDFa Resource Description Framework in Attributes
- RDFS RDF Schema
- SKOS Simple Knowledge Organisation System

- SPARQL SPARQL Protocol and RDF Query Language
- SRSIS Scholarly Resource Semantic Information Store
- SW Semantic Web
- SWIB Semantic Web in Bibliotheken
- TSO The Stationery Office
- URI Uniform Resource Identifiers
- URL Uniform Resource Locators
- VIAF Virtual International Authority File
- W3C World Wide Web Consortium
- XML eXtensible Markup Language

# Appendices

# Appendix 1 – Participant information sheet

Name of department: Computer & Information Sciences

**Title of the study:** Linked Data: Implementation, Use and Perceptions across European National Libraries.

### Introduction

My name is Laura Cagnazzo and I am a postgraduate student at the University of Strathclyde. I would like to invite you to take part in my final project, investigating the adoption and use of linked data across European national libraries.

**Researcher**: Ms Laura Cagnazzo; email address: <u>laura.cagnazzo.2016@uni.strath.ac.uk</u> **University of Strathclyde**: 16 Richmond St, Glasgow G1 1XQ; telephone: +44 (0) 141 552 4400

### What is the purpose of this investigation?

This research focuses on the challenges and difficulties that European national libraries have encountered during the process of implementing linked data within their systems. It aims to provide an insight of the current situation, with the intent of increasing the awareness of the linked data potential within the information science sector, and potentially drawing recommendations for those institutions, which perceive the call and the need of embracing this new method, but are still looking for best solutions.

#### Do you have to take part?

In order to gain a comprehensive picture of the current state of linked data implementation and use in the information and library sector, with specific focus on European national libraries, I will interview those professionals directly involved in the realisation of linked data-related projects. In addition, I will gather viewpoints of those institutions that have not implemented linked data, with the intent of understanding the reasons preventing such decision. Participation is entirely voluntary. If you do not wish to take part, you do not have to give a reason and you will not be contacted again. Similarly, if you do agree to participate you are free to withdraw at any time during the project if you change your mind.

### What will I do in the project?

Your contribution will consist of an interview, lasting approximately 30 minutes. There will be only one single interview with myself. We will arrange an appointment which is convenient to you. I am happy to conduct the interview via Skype or email (in this case, I will send you the set of questions and you will be able to reply in your own time). Interviews will be held over the next couple of weeks, preferably within the end of June (adjustments can be made if dates are not suitable). I will be glad to send you a summary of the findings, upon completion of the study, if you are interested.

#### Why have you been invited to take part?

Participants to this study are selected on the basis of the following criteria:

- Professionals working at European national libraries directly/closely involved in the implementation of linked data within library system
- Professionals working at European national libraries in roles focusing on metadata standards (who are more likely to be involved in future linked data implementation)
- Professionals working at European national libraries particularly interested in the linked data development within information science.

Your experience and reflection are deemed particularly valuable in order to achieve the expected learning outcomes of this research.

### What are the potential risks to you in taking part?

At the best of my knowledge, no potential risks are involved. Only a small portion of your time is required!

### What happens to the information in the project?

All your personal data will be anonymised and treated strictly confidentially. All data will be stored in secure manner and duly disposed of within one year after the completion of the study. If required, anonymity can be guaranteed avoiding the association between you and your organisation and/or nationality.

The University of Strathclyde is registered with the Information Commissioner's Office who implements the Data Protection Act 1998. All personal data on participants will be processed in accordance with the provisions of the Data Protection Act 1998.

#### What happens next?

If you are happy to take part in this project, you will be asked to sign a consent form to confirm your availability. Once I have received the form back, I will contact you so we can arrange the details of the interview in a way and at a time that is convenient for you.

If you decide not to participate, no further action is needed. I would like to seize the opportunity to thank you for the time you took to read this far.

In case the results of the study may be considered for publication, further information will be provided prior to take any further action.

Thank you for reading this information – please ask any questions if you are unsure about what is written here.

This investigation was granted ethical approval by the Departmental Ethics Committee.

If you have any questions/concerns, during or after the investigation, or wish to contact an independent person to whom any questions may be directed or further information may be sought from, please contact:

### My supervisor:

Dr Diane Pennington Lecturer in Information Science Email: <u>diane.pennington@strath.ac.uk</u> Telephone: +44 (0) 141 548 3900

### **Departmental Ethics Committee:**

University of Strathclyde Department of Computer and Information SciencesLivingstone Tower 26 Richmond Street Glasgow G1 1XH Email: <u>enquiries@cis.strath.ac.uk</u> Telephone: +44 (0) 141 548 3189

# Appendix 2 – Interview questions

1. Have you implemented LD in any of your library resources?

If 'no':

2. Can you explain your organisation's concerns preventing you from opting for an implementation of LD? **Extra:** 2.1. What barriers were identified that discouraged or impeded the adoption of this technology?

3. Have you reflected on benefits that LD could bring to your institution?

4. What impact, if any, could the examples offered by successful projects accomplished by other institutions have on your future choices? **Extra:** 4.1. Have you identified any application of LD that, in your opinion, has proved particularly successful and beneficial to library resources? **Probing:** 4.2. Can you provide examples?

5. Have you ever reflected on the role that the adoption of agreed standards plays with respect to creating uniform systems and allowing interoperability among datasets? **Extra**: 5.1. Have you ever considered what the relationship between LD spread and further development and standardisation is?

6. Is there any plan in place at your institution for a future change of direction? Probing:6.1. Can you provide more details?

7. Would you like to add any further comment/reflection? **Extra:** Your thoughts on future development on LD within the information and library sector or in general?

If 'yes':

1.1. What were the reasons behind your choice of implementing LD?

2. Have you identified the key barriers that delayed or could potentially have prevented LD implementation at your institution? **Probing:** 2.1. Can you provide more details on this?

3. Can you describe the main features of the project realised? **Extra:** 3.1. What were the main steps taken? 3.2. How much of the process could benefit of automated or ready-to-use technologies?

4. What example have you looked at, if any? **Extra:** 4.1. Do you reckon a particularly successful project within the information sector has proved the usefulness of LD? 4.2. How about projects beyond the Information & Library sector?

5. What is the policy adopted at your institution with respect to LD implementation and use, if any?

6. Have you ever reflected on the impact that the adoption of agreed standards has with regards to creating uniform systems and allowing interoperability among various datasets? **Extra**: 6.1. What would you say is the relationship between LD spread and further development and the adoption of standards?

7. Has collaboration with other institutions helped the actuation of your project? How do you believe it could support the development of LD within the information and library context? **Extra:** 7.1. How do you believe collaboration between cultural institutions could support the development of LD?

8. What were the foreseen benefits of implementing LD? Did reality meet the expected outcomes? **Extra:** 8.1. What were your expectations prior to starting the implementation? 8.2. What do you think went well?

9. How about the challenges? Foreseen and actually encountered ones? **Extra:** 9.1. Have experienced issues related to training, implementation cost, time management, resources required by the process? 9.2. What would you say went wrong?

10. Would you like to add any further comment/reflection? **Extra:** 10.1. Do you have any suggestions for those institutions who are looking into implementing LD or reflections on future LD development within the ILS sector?

# Appendix 3 – Interview with the British Library

#### Q.: Have you implemented LD in any of your library resources? How?

A.: Before we start, I wanted to clarify one thing. When you ask if we have implemented LD in our library resources. What we have done is to make one product available as LOD, which is our British National Bibliography (BNB), but it is not like having a LD library system. In our Library Management System we are using MARC21, and we are processing the data to convert into RDF. BNB is a relatively small subset of our data, we are not making our entire catalogue available as LOD. We chose BNB because it has got a clearly defined scope, and its metadata is more consistent than the metadata available in our catalogue,

some of which date back to the 19<sup>th</sup> century. We thought that processing the BNB would be relatively a more straightforward task than processing the entire British Library catalogue. That was just a clarification for the first question.

#### Q.: What were the reasons behind your choice of implementing LD?

A.: There were many reasons for implementing LD. First of all, we wanted to experiment with LD technologies. There was a lot of information coming out from the Joint Steering Committee for Development of RDA, from this and other committees, that LD was going to be the technology of the future. So we wanted to see what the technology would involve and what the infrastructure would be like, that would be necessary to support it. That was one motivation.

Another motivation was that Government policy at the time was strongly encouraging public bodies to show a good example to the rest of the community by making its data publicly available as openly as possible. And we saw LD as one mechanism for doing that. We thought that by making that available we may encourage people to use it in ways that we had not considered before. And we also thought that we might get some benefits by doing so.

A third motivation is that the BNB has been produced since the 1950s, and it went through transitions between different technologies: first, printed technologies in the 1950s-70s, then it was transferred to CD-ROM during the 1980s-1990s and then it became an online resource. So we saw LD as another way to make this product available.

# *Q.:* Have you identified the key barriers that delayed or could potentially have prevented LD implementation at your institution?

A.: Probably the main initial barrier was rights. Historically we have sold the BNB as a printed product, available on subscription. As MARC data, it was a product that we sold as MARC exchange files to other institutions. So, from the Library perspective, in a time of economic restraint, there was an economic issue here. Our perception of LD was that we would have to adopt a very permissive licence for it, which would actually give the data away for free. We had to take that to the Library Board, to receive approval. And what they eventually approved was that we would make the data available under a CCO licence. We discovered quite quickly that making LD available under any more restrictive licence would make it effectively useless to potential users, because they wouldn't be able to redistribute the data without intervention. This would be the case even if we made the data available under a CC-BY licence. Therefore, that was a major issue that we had to resolve even before we started the project.

Another issue was that we didn't have any expertise in LOD. None of us was a developer, nor an IT specialist. We were all basically cataloguers. So we realised that we couldn't do it our own and that we needed advice, so we partnered initially with TALIS, who, at the time, were interested in LOD and were doing quite a lot of work in this area. Subsequently we have contracted with TSO, as we do not have an infrastructure to support LD at the

moment. For example, we do not have a triplestore. We are simply 'renting' space, and procuring expertise, from TALIS first and now TSO. These were two barriers.

The data we are processing is in MARC21 format. So we had to convert the data into RDF/XML and then into triples, which was quite a complex process. First of all, we had to model the data, we had to decide how we wanted to present it. We had taken the decision that we were not really aiming at the library market in this case, that we won't see this as an alternative to the MARC service. We were seeing this as something different in which other communities might be interested. So we decided to present the data in a slightly different way than we would have done for the library community. Therefore, the modelling took us quite a bit of time. Then building/testing the conversion was quite intensive. There is more information on our website about the project.

The other issue that was identified while we were going through the process, was a lack of infrastructure, a lack of tools. Things had to be done pretty much from the scratch, there was not a lot that you could just take off the shelf.

We also suffered from the fact of being relatively early implementers. There was not the kind of consistency or expertise that we were used to in the MARC environment. There were not actually consistent opinions on how to best do things, and that was particularly true about the modelling. Some of these issues, I think, have not been completely solved yet. There are still questions on the modelling of real world objects and data about real world aspects.

#### Q.: Is there any plan for future project involving LD at the British Library?

A.: I think it really depends on how technology develops. I believe LD has to really prove itself, in the sense of the benefits that it can bring. At the moment we are using it in a very limited way, essentially making something available for people to consume. We are not really consumers ourselves. We see a lot of potential, but we do not really see the kind of infrastructure emerging at the moment that, we would be comfortable abandoning what we are doing now. We are currently in a holding pattern, I would say.

#### Q.: What example have you looked at, if any?

A: There were not a huge number of examples to look at. We looked at what the National Library of Sweden had done: they are pathfinders in the area. We were also looking at what had been done with LOCAH, a project looking at archive data. Other examples were the HBZ, the German project. In addition, TALIS was doing some other work of their own, so we were consulting with them during the development phase.

# *Q.:* What is the policy adopted at your institution with respect to LD implementation and use, if any?

A.: There is no institutional policy on LD, driven by the fact that we have not got an infrastructure for LD, so there is no need for a policy. We created this service as an initiative of the Collection Metadata service, as something experimental. It was not the initiative of the Library itself. We did not call up on the IT or any other department for any

other resource. Therefore, we do not have anything that you could call a policy. We have very clear licensing policies instead, not only for LD, but for all of our metadata.

# *Q.:* Have you ever reflected on the impact that the adoption of agreed standards has with regards to creating uniform systems and allowing interoperability among various datasets?

Obviously you need an agreed understanding at some level in order to do LD. You need to follow RDF and other Semantic Web standards. However, on the other hand, for institutions with a long tradition of metadata, that does not necessarily help a great deal, because it challenges you to get your metadata into a condition where it can become LD. For us it was largely the issue with identification of entities, the fact that historically entities have not been uniquely identified. Although we have identifiers for persons and corporate bodies for some of our metadata, we do not have it across our metadata. We do not have consistent identifiers for subjects for our metadata, nor for works or even for products. In some case you cannot really rely on ISBNs. The key thing on which LD is going to depend on is identification. And identification is actually difficult. I believe, in some sense, that LD will allow us to work in a completely different way than the way we worked in the past. In the past standardisation within the library community was understood on the basis of everyone basically doing the same thing, to a very restrictive set of guidelines. With the goal of, for example, creating a heading for William Shakespeare in the same way as the Library of Congress. This is really not sustainable and it does not really help when you try to internationalise, because each institution does not necessarily follow the same rules. With LD what you are looking for is identification of those entities with an opaque URI, around which you can cluster any number of different labels, to suit the needs of different communities, but all referring back to the same entity. And that is what I see as the big strength of LD, that would allow us to internationalise things in a way that has not been possible before, by using a common URI to identify things, but having labels for that URI locally definable.

# Q.: Has collaboration with other institutions helped the actuation of your project? How do you believe it could support the development of LD within the information and library context?

There are many different ways to collaborate. With TALIS we collaborated because they had the knowledge, the experience and the technical infrastructure that we did not have. I would not define the relationship with TSO collaborative, but commercial, as we pay for a support service. However, we do not have an infrastructure, so that is one way in which an institution can collaborate. We have also worked, over the last year or two, with Fujitsu International. They were interested in analysing how our LD were being used and they developed a dashboard to analyse the LD usage. This collaboration helped us to fill the gap of knowledge deriving from the fact of not knowing who was using our LD. Fujitsu could benefit of a rich source of data to test their software, and we were benefitting of their technical expertise and toolkit. We were actually able to get a much clearer picture of what was being done with our data. And that has subsequently helped us to build a much clearer reporting dashboard with TSO: we are now able to see who is using LD, where they come

from, how often they come back. We still do not know what they are doing with it, but at least we know who they are.

There is surely an issue for cultural institutions, for the whole sector working on LD infrastructure, about how we are going to maintain it over time. RDA cataloguing standard has got a registry of all the classes and properties that they use and the vocabularies they use. And that is freely available data at the moment. However, maintaining the infrastructure is not free. How is the community going to maintain that over the time? This may be an area where there could be collaboration between institutions, but it is still at a very early stage.

# *Q.:* What were the foreseen benefits of implementing LD? Did reality meet the expected outcomes?

What we expected to get out of the project was an understanding of the technology to learn some lessons. And I think that was achieved, as we learned a lot. And we were able to make the BNB available in a completely different way with LD. Within the relatively narrow terms of what we set out to do, that was successful. I do not think that we went into the project with a huge number of expectations about LD. We went to learn about LOD.

### Q.: Have you received any feedback?

A.: I do not think we have received any feedback from staff members. There has been some interest from the IT department and the British Library Labs. The external feedback was generally positive. People have welcomed this initiative, since we were open about this kind of experiment. We have been clear about the lessons learned and what we could have done better. We received positive feedbacks from users. We have a regular user community; people are downloading and using our data. The initial perception that the usage was falling off is not actually the case, as the usage is constant and relatively high level. However, it became clear that there was a group of researchers who were interested in our data, but for whom RDF was a s big a barrier as MARC. There was a need for a more accessible format. This led us to create the "Researcher Format" in which we make tranches of our metadata available as csv. Datasets are created to complement exhibitions or commemorations and are available from our downloads page.

I think that, if we compare our service with *data.bnf*, surely they were able to provide a more attractive service, but they have based that on much more resource investment in data, and they had access to developers, which we had not. Their service is still based on MARC data converted into RDF with LD applications. The methodology is quite similar to ours, but they have a much more integrated process. In their case, it is much more of a core service to the library, while for us it is more a peripheral service.

We have put our data on data.gov.uk and we have finally received five stars from them, which was what we aimed at.

Q.: How about the challenges? Foreseen and actually encountered ones?

A.: One issue was that we did not know what people were doing with our LD. With many of our data we were able to get people to obtain a licence, so we were able to see what they were using it for. So there was a transactional approach to that, which allowed us to monitor who was using the data. We could also often work out what the use of the data was, by the nature of the data. In addition, we could do surveys, ask them what they were using it for. With LD we had no idea if anyone was even using it. All the statistics we were getting back were not particularly informative and (wrongly) implied that the usage was actually declining over time.

What disappointed us was the relative lack of sources of data that we could link to, that were of any sort of value. I think this is the issue. If you are adopting a particular data model to make your data available, you are putting constraints on the data that may affect how data is reused in the future. Our definition of things can be very different from somebody else's definition, which actually made it quite difficult to align those properties, or attributes. You cannot say that two resources or properties are 'same as' if they are not.

We were really keen to point to other resources of information or to reuse other people's data, but that has only been possible to a limited extent, using VIAF, Geonames, Lexvo for example. Beyond that, we were not able to go.

We also thought that there would be the opportunity of working with other institutions. We have talked to other institutions about potential join up between our datasets; e.g. museums have got the objects and we have the books about them. That may be a practical and feasible way of linking, but it has not really taken off since everyone has many other things to do.

*Q.:* Would you like to add any further comment/reflection? Do you have any suggestions for those institutions who are looking into implementing LD or reflections on future LD development within the ILS sector?

A.: There are now many more resources about LD than there used to be, on how to get a better grip of LD concepts, and I would encourage people to have a look at them. For example, there is a LD resources explorer developed by Dublin Core and other partners, LD4PE (http://explore.dublincore.net/about/). There are starting to be sources of training. It is important to look at examples. Institutions starting now have access to many more resources, also library related. However, I believe there are still unresolved issues. For instance, with regards to ontologies: there are still discussions about whether you should develop your own ontology or reuse an existing one. We chose the second approach, so we used Dublin Core ontology and other existing ones as much as we could. This is related to issues of domain (DNS) control and how much trust we can have in ontologies being available online reliably. I believe there are still many unresolved issues, e.g. with respect to several data models (e.g. EDM, BIBFRAME, ours, etc.). Our BNB only contains 4 million records; the data model we use for it may not be adaptable to our whole catalogue. I believe that for institutions starting working with LD now it would be easier than it was for us, but still not plain sailing.

# Appendix 4 – Interview with the National Library of Finland

# *Q.:* Have you implemented LD in any of your library resources? Can you describe the main features of the projects realised?

A.: We mostly use LD for authority data, but we are currently working towards exposing our bibliographic data as LD. It still a work in progress which has not been announced yet. I was tasked with publishing bibliographic LD 1-2 years ago, and it soon became my main project. However, it is still only one person working full time with LD, not every member of the staff deals with it. It is more about preparing for a future scenario and experimenting the potential of making our data available as LD. Our dedicated task force to LD development decided to adopt Schema.org as data model to publish our data: it offers the advantage to be generic enough to allow exposing many things, but at the same time specific enough that you can provide many details. In addition, it is not governed by a single entity, but it is regulated by a community. I am sure there are still things to improve in that community, but there is a good premise that Schema.org may become a shared model that others could use.

Finto project was the thing that broke the ice and opened the way to LD adoption, since the National Library had a burst of funding available from the Ministry of Education and Culture to create this sort of vocabulary/ontology/authority service which is not only for libraries, but also for museums, archives and the public sector. This signed the beginning of our work with LD, as the National Library had finally the resources to hire people with expertise in LD.

Fennica is our national bibliography, so it was the asset that was first in line to be published as LD, since, from an external perspective, it is our most interesting database. The first step for me was to get a feeling of what type of information was in there and then I started converting the data using the different converting tools available. And eventually I built this sort of pipeline that stiches together different tools and it uses Schema.org in the end. I have recently reached the point where I have full conversion, first to BIBFRAME, then to Schema, and then publishing data in a SPARQL endpoint and data dumps, besides some documentation.

#### Q.: What were the reasons behind your choice of implementing LD?

A.: It is something that has been brewing for a very long time. We have started working with LD, here at the National Library of Finland, approximately four years ago. The Library acquired awareness that other institutions around the world were starting to publish their data as LD about a decade ago, but at that time there weren't enough resources to start a project in that direction, and there was uncertainty on the correct way of approaching it. There is not a single, specific reason behind the choice of adopting LD. It is more about growing awareness of the time - which is something a national library should be doing - and gradually allocating more resources to it.

*Q.:* Have you identified the key barriers that delayed or could potentially have prevented LD implementation at your institution?

A.: The fact that there are so many different ways of publishing LD can surely represent a barrier to its adoption. Especially the early developers took various approaches. It didn't seem fruitful to us doing the same and starting the whole process from the scratch. Therefore, we agreed on a data model to reuse, rather than creating a new one. We have been waiting for the Library of Congress – considering their role as guardians of MARC and their authority in the field - to provide guidelines on how to do the transition to LD. They started to do so with BIBFRAME, but they haven't been very good at communicating about it; and the reason behind their lack of communication is likely to be that they do not have enough resources to do so. For long time I have been waiting for the Library of Congress. Eventually, it did not materialise, at least not in a complete way. When I started to work on bibliographic data, I decided to do a review of all the different approaches, to try and understand what is the best practice, if there is one.

# *Q.:* What is the policy adopted at your institution with respect to LD implementation and use, if any?

A.: We have a small task force that works with LD at the National Library of Finland. We have meetings once a month to discuss our plans and achievements, but we do not have a specific policy in place. We draft our policies as we go.

### Q.: What example have you looked at, if any?

A.: We looked at the work accomplished by several national libraries. In particular, the early LD developers, such as the national libraries of France, Spain, England and Germany. Sweden has also been quite active, but they are not putting enough effort in communicating their achievements. They are currently building a whole new library system, that uses LD internally: a very ambitious project, based on the adaptation of BIBFRAME 2.0. They actually started creating their own model, but they ended with something very similar to BIBFRAME 2.0.

# *Q.:* What would you say is the relationship between LD spread and further development and the adoption of standards?

A.: I am worried that, if the current trend continues, if everyone keeps using different ways to create LD, the data published as LD will not be well used, as it is very difficult to combine, or do a comparative analysis of different databases. For example, it would be interesting to see which work that is in our national bibliography is also in the Swedish national bibliography, but you can't do that at the moment, since these resources are structured in a different way, and you need to dig quite deep to be able to compare information.

# *Q.:* How do you believe collaboration between cultural institutions could support the development of LD?

A.: It is definitely starting to happen more than it used to, I am quite happy about the current status of collaboration on LD, although ideally it should have taken place earlier. There are events such as the SWIB (Semantic Web in Libraries) Conference, which places collaboration at its heart: various projects are presented, people meet each other. There

will be a workshop on BIBFRAME this coming autumn in Frankfurt. The community is finally trying to get together and look at these matters together. The event organised by LD4P and LD4L-Labs, that occurred last April at Stanford University, represented the first opportunity for professionals to meet and discuss the future of bibliographic LD. It may be the start of a deeper collaboration.

## Q.: What were the foreseen benefits of implementing LD?

A.: We have more experience on the authority side, as we have been doing it for much longer, but I think one advantage of LD implementation is that it makes data more accessible, also to institutions outside the library sector. If you publish a MARC record, nobody is going to care, because nobody understands them, except for librarians. With LD it is different. We built APIs for instance, which are not exactly LD, but they make the data available in a more convenient form. Another advantage of LD is enabling to analyse our data very efficiently, which guarantees a higher quality of data. Even non-technical people can reuse and tweak SPARQL queries written by others (usually shared on a wiki) even if they don't know how to write a whole new query from scratch.

It is a very efficient way to analyse our data, and I am hoping to do the same with bibliographic data. Previously we had our data in MARC records and they were somehow locked in the system and it is not easy to perform an analysis on that. You can look at an individual record, but search facilities are very limited. LD enables more sophisticated queries. There is also the linking aspect. You can do this with traditional systems too, but with LD is somehow natural linking your own data to other people's data. Therefore, we have started to do that, we have linked the YSO (General Finnish Ontology) to the LCSH (Library of Congress Subject Headings) and the Finnish place name registry, for locations. Establishing the links is a way to find problems with your own data and to enrich it. We are also planning to link to Wikidata, so that we can take advantage of the data published in Wikidata and push some of our data in there.

### Q.: How about the challenges encountered?

A.: It requires a lot of work and it is not always clear which are the benefits of the implementation compared to the resources you have available. Since we had the resources, it seemed like a good idea working with LD. There are not many people working at the Library who understand how to use LD, so training is surely a main issue. When we first started the Finto project, we were running workshops to teach people how to write a SPARQL query and events to increase the awareness of LD, but we do not do that very often anymore. It is still a sort of niche thing to do, a side project for many institutions. Many people at the National Library heard of LD, but they don't know how to work with them, they are focused on other tasks. And I do not expect this to change very quickly.

*Q.:* Do you have any suggestions for those institutions who are looking into implementing LD?

A.: I hope that we will converge on the data model to use to create LD. For anyone who wants to start doing this, it would be best to get in touch with those who have already

been working with LD, through conferences for example, in order to get a feeling of best practices before making any decision that can confine your resources into a little world, if you do things differently.

# Appendix 5 – Interview with the National Library of Germany

Our LD service is at the core of our business, so I am happy to give information about it.

# Q.: Can you describe the main features of the LD projects realised?

A.: We started with the authority data because we thought they are more reusable, and who is working on cultural heritage projects is interested in people, places, topics, events, and so on. Also because authority data is much easier to model than bibliographic data. That was our first prototype in 2010. In 2011, if I remember correctly, we did the first prototype for the bibliographic data, and since then we have been remodelling and continuously expanding the service.

We have been developing our services iteratively. Our first step was to model the authority data using the common vocabularies, RDA Group 2, RDA relators and some home-grown properties for persons and corporate bodies, SKOS and DCTerms for subject headings in the first iteration. While we were doing that, we had another project running, called Gemeinsame Normdatei (GND), *Integrated Authority File* in English, which involved the colocation and integration of all the authority data we had into one single database. We used to have separate databases for authority data: one for people (PND), one for corporate bodies (GKD), one for subject headings (SWD), and one for creative works, especially music (EST). We had a project integrating all these four into one common file, getting rid of different cataloguing rules. We consolidated the view, looking at everything as entities: for instance, if a person is a subject and also an author, it is still seen as the same person. That, of course, forced us to remodel the authority data once the data migration was done. After that, we built our own RDF vocabulary, the GND Ontology. That was one core part of the first four years of the service.

The other core part was then the modelling and publication of the bibliographic data. We first looked at all different models that were out there, in particular looking very closely at FRBR. Then we realised that we did not have sufficient good enough quality data to supply all four entities of the FRBR model. Therefore, we ended up with a very flat representation of the bibliographic data for the first iteration. We then started a cooperation with the other German speaking libraries who publish LD, for instance the Swiss National Library, the Bavarian State Library (BSB), and the German regional library networks. We wanted to see if we could start consolidating our different data models, because everyone was doing it in its own way, which made impossible data integration through LD. We started a working group and we have now finished the second iteration of our recommendations on how to model bibliographic data in RDF in Germany. We are currently trying to implement the second round of recommendations, which should be ready by September. What we have done in parallel to this, is also to complete the data set we publish, to include not only printed and electronic material, but also musical sheets and recordings,

which I believe will be ready in September too. Once that is done, we have set an important milestone, that is making the complete national bibliography available as LD. It is now more matter of polishing the corners and make sure that everything runs smoothly, perhaps look at customers' feedbacks, to see if there is anything we need to change. Then we can say that we have all range of LD service up and running, and after that we need to see where we go from there.

The next important step we will be looking at is the use of Schema.org through the catalogue, to have a better uptake with search engines, and also to improve the data set descriptions. Currently, we're adopting again Schema.org and also DCAT vocabulary to make better data set description. Use of VoiD is also on the agenda. The goal is to obtain better data integration with other data portals, like the EU data portal, and the German Government data portal for public sector data. This will probably keep us busy till 2018.

### Q.: Is there any plan for future projects?

A.: Not really new plans. We are working on continuous development. We have initiated cooperation with other German libraries, to do LD agreeing on a common model, particularly for the bibliographic data. What we are also doing right now, is to increase the reuse of the authority data within the larger German cultural heritage community. We are now looking at metadata set description for search engines, and also the markup of our catalogue with Schema.org. Another research topic is how to represent and use application profiles when doing http content negotiation.

#### Q.: What were the reasons behind your choice of implementing LD?

A.: It is a long story. It started off as a research project. When I first heard about Semantic Web, in 2002, I could not really make sense of it. It was Artificial Intelligence 2.0. I thought: if it didn't work in the 1960s, why should it work now? I read the article by Berners-Lee, Lassila, and Hendler, but it did not make sense to me. Then, in 2003, I started at the National Library of Germany, and I realised that, the way data is curated here, with bibliographic data always linked to authority data, and the bibliographic data being linked internally as well, people with places and so on, it was somehow reflecting the data modelling idea of the Semantic Web. So the next question was: "why don't libraries do Semantic Web, since their data is predestined for it?". After about five years we got some people together looking more deeply into it, and we thought that this idea would work fairly well with the data we had. Also, in 2008, the National Library of Sweden had already put out a prototype, showing that is feasible and makes sense to publish library data as LD. So we decided to contribute to the LD space with the things we had, and that would make sense as well, and that is how it all started.

# *Q.:* Did you consider the LD potential of augmenting the visibility of library resources?

A.: Yes, we did. We wanted to reach out to other communities that do not usually care about library data. Let them know what our data is and that it is free to reuse. That is also the reason why we use a very liberal license: all is published under a CCO licence, RDF data

since the beginning, while MARC data was initially being sold for a fee, but no longer. In addition, we decided not to require registration, to encourage reuse. And that paid out.

# *Q.:* Have you identified the key barriers that delayed or could potentially have prevented LD implementation at your institution?

A.: We didn't really have any. We had the buy-in from senior management since the beginning. We even managed to get it so far that the LD service is on par with the MARC21 service. We also provide them on equal footing, seen as equally important. We were very lucky to obtain the buy-in from the senior management since the beginning. An obstacle was, of course, finding staff that would be willing to look into it and maintain the service on a day-to-day basis. This was quite foreseeable though: staffing is always one of the main obstacle when you start a new project.

We have not had many problems implementing the service. We also had good buy-in from our developers, who dealt with the conversion, and we established a good relationship with the digital service department, who does most of the customer contact.

So, we had problems at the beginning, at the research project phase, but then when we went into production we did not; and we also managed to reuse the workload we already had produced for the MARC21, which solved quite many issues.

### Q.: What example have you looked at, if any?

A.: Our main inspiration was the Swedish National Library, who did the first implementation. Then, we have been in close contact with the national libraries of Spain and France, and of course with the British Library, being them the main other LD implementers in Europe. And it is interesting to see that we all ended up with different data models for how to do the data description. And I believe that, for the foreseeable future, we will have to leave it like this.

*Q.:* What is the policy adopted at your institution with respect to LD implementation and use, if any?

A.: Building knowledge networks is part of our long-term strategy as stated in our "Strategic Compass". Refining the LD service is part of our strategic priorities for 2017-2020, cf. "Strategic Priorities 2017-2020".

*Q.:* Have you ever reflected on the impact that the adoption of agreed standards has with regards to creating uniform systems and allowing interoperability among various datasets?

A.: I am part of the W3C *Data Exchange Working Group* (DXWG), which has been recently formed. One of the topics we will be looking at is the use of application profiles as their own but also in the LD context. To describe them in a machine-readable fashion, and to use HTTP to denote which profile the data addresses, and also to do content negotiation on profiles.

I think we should be aiming to cooperate to see how we can make data integration easier, finding common modelling patterns perhaps. We should have a closer look at how to deal with authority data, which may mean that we all agree to use VIAF for persons, instead of our authority files. When it comes to subject headings, I do not think that there will be a common subject heading authority file for all languages, but instead we will have to look at the results for instance of the MACS project, whether there is intellectual mapping or maybe ultimate intellectual mapping of subject headings in different languages, and see how we can get that more closely integrated. MACS was a project started about ten years ago from OCLC, British Library, National Library of France and National Library of Switzerland to map subject headings in German, French, and English. Per term basis, saying that you have a triple or quadruple of subject heading, that means the same thing in all different languages. The database was closed last year, and we are finding new ways of distributing the management of the data to increase the number of mappings. They added a new vocabulary a year and a half ago, the Nuovo Soggettario, but there are not many mappings available. We intend to publish the complete MACS data set at intervals. It would be static data dumps, consisting of all topics in the database being linked to each other using SKOS relations.

SHACL (Shapes Constraints Language) and ShEx (Shape Expressions) are fairly new technologies that can be used to describe RDF data in a similar way you would describe XML through XML Schema. So you can use that for data description, and also data validation. What we are working at in DXWG is how to describe application profiles in a format-independent manner and how to link that description to format-specific schemas. That way you would first describe how the data is organized and then go on to say this is how you would do it in XML (as an XML schema), this is for JSON (or JSON-LD) and this is how to do it in RDF (e. g. which classes and properties to use).

# *Q.:* How do you believe collaboration between cultural institutions could support the development of LD?

A.: It is very necessary on several levels. On the technical level, how to implement the protocols, how to implement new technologies, querying methods, from SPARQL to Triple Pattern Fragments. We need collaboration on developing ways to describe the data, and modelling the data. What I have seen so far is not much collaboration really. People are talking to each other and showing what they are doing. My impression is that in the recent years there has been much experimentation, and proofs of concepts in experimental projects, with regards to LD production, e. g. to see how BIBFRAME can be implemented. I am curious to see the findings of the *Linked Data for Production* (LD4P) project, which is using BIBFRAME as data exchange format.

In addition, I believe we need more discussion about the different parts of the library data ecosystem, being: data models, cataloguing rules, and exchange formats, which are currently very much intertwined. And I think we should aim at having as much separation between them as possible and to figure out which part of the ecosystem is responsible for what. For instance, minimising the impact of cataloguing rules on the exchange formats. The difficulty derives from the fact that the community interested in LD within the library

sector, is very diverse: there are people who tend more to look at the technical aspect; and people who tend to look more at the content side of it. This is good from one side, since we need experts for both parts, but the issue is that sometimes they do not really get together. And it is hard to find a forum where people can discuss these things. With IFLA there is the *Linked Data Special Interest Group*, that tends to be focused on advocacy, telling libraries what LD is and why it is important, but there is no discussion on data models, or how to achieve interoperability at content level. What we miss is a forum to discuss these issues. A very good conference is SWIB (Semantic Web in Bibliotheken) and I believe there will come more in the next five years. Much depends on the results of the LD4P, and on how much data you can get published using IFLA LRM in RDF as data model. And it may well be that we end up with different ways of doing this, as there is UNIMARC versus MARC21, that are compatible on the syntactic, but not on the semantic level, at least not the whole parts. If we look at MARC 21, there are so many ways to fill fields using the same background data. I think interoperability will keep being a problem for library data interchange. What would be interesting to see is the data uptake of the community outside libraries, what they think of the different ways of publishing library data, because we do not want to do this only for library users: one of the core mission related to LD at the German National Library is to reach out at communities who have not looked much to library data so far, because they could not deal with MARC21.

One interesting partner is DBpedia: we are interested in integrating library data into DBpedia to a greater scale to do text mining, and also automated categorisation of articles, which is the research project they are dealing with right now. We're also working with Wikidata to see how that integrates with library authority files and how we can improve data reuse and interchange.

# *Q.:* What were the expectations before starting implementation and what were the actual outcomes?

A.: People are much more interested in the authority data than bibliographic data. This is something that had expected and has proven itself true, since authority data is much more reusable. However, we have not had such a broad uptake as we hoped for. The people who reuse data are mainly from cultural heritage institutions. We have not seen much uptake through industry, but we are working on it. Particularly because they are starting to realise the benefit of reusing authority data from libraries and other cultural heritage institutions in their own workflows, especially the publishing industry. We have seen uptake through the research sector.

What we have also learned is that there is a lot of interest across the library community on what we are doing, but many institutions find it hard to get started. When speaking to other libraries who publish LD, I realise that the technical approach to creating the data differs very much between the various libraries. It depends also on what library system you use, and your cataloguing method, and on which means the library has to access the data in the central bibliographic system. We do not know all the customers that use our data in RDF, simply because we do not require feedback or to be mentioned as data provider. So we depend on an anecdotic evidence of data reuse in many places.

Our Identifier concept is entity-based, i.e. the http-URI (e.g. <u>http://d-nb.info/gnd/36221-9</u> identifies the Entity University of Strathclyde, just as <u>http://d-nb.info/gnd/1130185230</u> identifies the person Michael Rodgers working there. Those entity URIs are formatindependent and we can then serve metadata about those entities in several formats, e. g. MARC 21, MARC-XML and several RDF serialisations.

*Q.*: Do you have any suggestions for those institutions who are looking at implementing LD?

A.: The first is to consider the URI syntax, how you identify the things you have.

The second is to reuse data, whenever possible, for instance reusing national authority files, if someone else has already published them.

The third is to focus on data that is specific to their organisation or institution. For instance, national libraries should focus on publishing the national bibliography as LD; academic libraries may want to consider special collections they have that may be of interest. It is about looking at what entities each organisation has, how to address them, which means considering the URI concept or URI schema to use, and what would be the greatest addition, content wise, to the LD cloud, something that nobody else can publish.

In addition, I suggest to talk to other libraries who have already published LD, to see which data model would suit best, and reflect if that data model does actually fit with the data of the institution, which is a fairly long process, but necessary to optimise data reuse.

Another recommendation for those who want to get started with LD is to ensure to have a community of stakeholders larger than just the library community, and to be able to count on people who understand both the technical side, including the interaction over the HTTP and all that, and the bibliographic content side, because these people have often different views of the data. Actual LD is part of the Semantic Web, so it questions what the data means. Cataloguers and bibliographers know that much better than the technicians usually do. It is also good to try and have a group of people in the middle, who understand enough of both sides.

It helps sometimes to collaborate with local universities for instance, if they have people who are experts in ontology modelling, when it comes to RDF data.

Another important aspect is to try and find an entity-based approach to the data, so you speak of people, places, things, and publications; and the important question is how are those linked together, and those things can be described in different formats for different audiences, but they are still about the same thing.

# Appendix 6 – Interview with the National Library of Latvia

Before you start with the questions, could you please give me some context? Is there some particular point that you would like to prove with your research?

[outline of research provided]

It sounds very interesting and useful not only for those organisation who are about to start, but also for those, like the National Library of Latvia, who have done something, but still have questions on what the best practices are, on the possible LD uses, and how to best implement LD, if you should do it at all.

# Q.: How have you implemented LD at the National Library of Latvia?

[Details on projects were received by email]

A.: These is how LD is implemented in the National Library of Latvia:

1) authority data is exposed as LD via VIAF: it does not involve NLL implementing LD technology by itself but rather just sending MARC records to VIAF. Yet, as a result NLL's authority data is a part of VIAF "cloud" and available as Linked Data. VIAF and ISNI IDs are added back to relevant NLL's MARC records (maintained in ExLibris ALEPH). We have also experimented with integrating authority data locally (from various NLL's systems), plus making it available as LD, internally but that has not gone further than experiments. The main issue is that the IT department sees the value in aggregating authority data but (1) don't see / understand the value of LD; and (2) have a different concept of authority data than what library people normally understand by authority records.

2) Digital object management (DOM) system - when developing a new, custom system (a digital repository) we added LD export functionality to it. Digital object metadata is published as LD and available using content negotiation: <u>http://dom.lndb.lv/</u>

Notes:

- an internal SPARQL endpoint is also available but is not used much in practice

- a major issue in implementing LD was figuring out how to model this information using existing RDF vocabs. this was not solved 100% so the native XML representation (also available via content negotiation) can be richer than the relevant RDF data.

- DOM is mainly an internal system and there has been debate whether to keep the LD interface public or not

3) Linked Digital Collection "Rainis and Aspazija" http://runa.lnb.lv

This is our most recent development. It is a LD pilot project - a Linked Digital Collection - where we experimented with enriching a digital collection with additional links between objects (by annotating named entity references, see further info in the paper) and exposing this information as LD.

This is where LD published by the DOM system come in handy as the pilot project does not have to fully duplicate what is already published in the other system - it can just refer to its URIs.

Authority data – our internal experiment was integrating authority data, but the developers had somehow a different idea of what authority data is.

The other project, the Digital Object Management (DOM) system represents a good example, where we outsourced an information system, and part of the work and part of the money was set aside for the LD interface. It was built in good quality and it works. Although, now there is an issue: there are some bugs, but these are not being given priority to be fixed, because money is limited, the IT department doesn't prioritise this, as they do not see the value of LD, who is using it and what LD is needed for.

In terms of the 'Raisin and Aspazija' project, the linked digital collection, an important learning point was that, we could not find a good tool for annotating this text content with mentions of ISNI entities. It is basically a process when you enrich text with links to some records about entities, like people or places. Consequently, we had to create our own annotation tool.

### Q.: What were the reasons behind your choice of implementing LD?

A.: Basically we were seeing LD as the future of library and information systems. Or, at least, this is the way LD has been appearing in conferences and papers during the years. We see that others are using it and, we may not be at the forefront, but we want to see the value for ourselves. To gain competence in LD. I guess that's also why the Library hired me, as I am the resident Semantic Web expert. Maybe it feels a bit silly that we are doing it because others are doing it, but if the big players with much more experience and resources see a value in it, then there must be something.

# *Q.:* Have you identified the key barriers that delayed or could potentially have prevented LD implementation at your institution?

A.: Somehow getting LD adopted in practice I feel like it is full of challenges for us. The main challenge is limited resources, both financially, but especially in people, in competencies of developers. That's true particularly for this last project, 'Raisin and Aspazija', because it was developed in house without any additional financing. So people who are already busy, cut out some time to devote to this important project, but not important enough for Ministries to give additional money. A consequence of this, is that we could only relay on developers that the Library already had. My role at the Library is more that of an informed customer for the IT department, rather than a programmer myself, so I know what needs to be done, but I am not the one developing systems for the end user. That is the IT department job, and they are the ones who will need to integrate the system supports and so on. And the role of developers is to maintain what we have and support systems that are procured elsewhere, like Ex-Libris and such. That means that you are basically asking people, whose competency is maintaining the systems, to do a quite cutting edge pilot project, like research and development. And it is painful when you

try to explain them what you need, why you need LD. Well, we got somewhere in the end, but it took a long time, the project was delayed. So people, combined with limited resources. If you have the resources, you can hire people with a good expertise; but libraries often have tight budgets, which do not allow that.

In addition, there was the issue of administrative delays. Mostly it occurs when we need to do something which involves giving our data to some other party. For example, with VIAF. I cannot even claim in this case that we published LD, since the only thing we had to do, was taking our data and giving it to OCLC for integrating it into VIAF. But still, that was the first time that our data appeared as LD. In the end, to the consumer it doesn't really matter if we publish LD or if we supply the publisher. That is a good way to publish LD: provide it to someone who publishes the data. I think the project was delayed by a year just to take this single decision: 'shall we do this?' And this because OCLC is an organisation providing some of its services upon payment of fees. So the question was: if OCLC provides paid-for services, why should we give OCLC our data for free? In the end, after a year or so, the project was approved and we started being part of VIAF, which is good.

The last challenge is related to understanding and proving the value of LD. If you do not have a way to show the management what are the benefits of LD, then why should they approve the project?

*Q.:* What is the policy adopted at your institution with respect to LD implementation and use, if any?

A.: LD doesn't seem to be the highest, or high-enough priority. It is enough to send people to conferences, and do some pilot projects, but... maybe the management is getting tired of LD and wants to finally get value out of it. However, implementing LD is mentioned in the Library strategy, the fact of becoming a hub for authority data, including linked authority data. So there is a line, in a high-level policy, mentioning LD as something to develop. And it is also included as one of the tasks of my department/structure at the Library, the Institute of Bibliography. In addition, there are guidelines, probably developed 4-5 years ago, for LD, just setting out the context, the strategies, what is LD and how to do it. So, there are some policies and some documents. It is OK to have documents, but you still need to have the practical buy-in and the willingness to implement those documents.

#### Q.: What example have you looked at, if any?

A.: We looked at many examples. You explore anything you can find. I have also given some LD training to the libraries, and, as a part of that, I have also collected some examples. Some of the major ones: VIAF, Library of Congress, Europeana. Then the data provided by the libraries, the good examples of Library LD, such as: the national libraries of France, Germany, Sweden (with LIBRIS), Finland and Spain. I am not sure how much of Europeana's data has been used. That goes for all of them and for our digital collection project: 'we have published LD, but what is next?' I guess that is the issue with open data in general, when you publish something for an unspecified wide range of users.
*Q.:* Have you ever reflected on the impact that the adoption of agreed standards has with regards to creating uniform systems and allowing interoperability among various datasets?

A.: I guess it is essential to adopt the same standards, your question is self-answering in some ways! It is like asking if common language is good for interoperability: of course it is!

*Q.:* Do you think the fact of using different ways of expressing LD is an issue? Considering, for example, all the types of serialisation available.

A.: if you mean the different serialisation of RDF, then it doesn't matter at all, unless you have some consumer who can only understand some of them; because they are basically equivalent, and libraries who work with LD would be normally able to convert between them and consume any of them. In terms of RDF/XML versus Turtle and such. That is not a problem.

On this note, content negotiation is useful, where you can ask resources for different representations/formats of data. So, in case you are worried about people being able to consume different versions, or in general you want to provide representations of data in multiple formats, that ensures that whoever has preferences for one or the other can understand the data.

Like we do in 'Raisin and Aspazija', where you can get the data in RDF/XML or Turtle. In DOM there is only one RDF representation, which is XML, but you can also get data in a completely different format, which is not really RDF, but it is almost a raw representation of what is in the database, so it might be even more precise or more complete than the RDF version. That is one thing.

Then, there are different ways in which people publish information, either using a combination of existing vocabularies or ontologies, one or more ontologies that are standards, or they use their own ontology. I think in Norway they use their own ontology, but they use LD internally. National Library of Spain uses its own ontology. At the National Library of Latvia is considered better to use standards, use existing resources. And there is no reason not to do that, if your case is standard, when you can use given examples and basically copy from it and use it for your data. However, sometimes it is easier to develop your system and your data model if you develop your own ontology. Then it is the question of how to map it to the existing ones. Modelling data in RDF is a lot of work, so I understand why many develop their own ontology, which is precisely matched to their needs. Considering the challenges that the RDF data modelling involves, should you do it rather than doing your own thing?

*Q.:* Don't you think that, if anyone does his/her own thing this is going to hinder future developments?

A.: There is a theoretical 'nice to have' answer to this question, and a practical one. The theoretical answer is that we have this nice LD web where applications automatically process the data and, in some limited way, understand the data, because they use common vocabularies. In that sense you should say that you should use the same

vocabularies and ontologies, because others would understand the data. However, I think that, in practice, every set of data, provided by different sources, even those that use the same ontologies, they use them in different combinations, there are some differences. Which means, if you are building an information system in practice, unless it is something very generic, which just shows you this RDF data, then you would be probably still handling data from each system separately, with some custom processing; you would still have to do something which is specific to that data set. If that is the case, then it does not matter much if they use existing vocabularies, but in their own special way, or if they use their own special vocabulary. Nevertheless, it would help if they provide mapping from their specific vocabulary to common vocabularies, just as a courtesy.

Q.: Has collaboration with other institutions helped the actuation of your project? How do you believe it could support the development of LD within the information and library context?

A.: What we mostly did was internal. The National Library is quite a large organisation as such, so we were mostly working by ourselves, of course if we don't count the contractors who were doing this work for us, the DOM project. However, it is good and inspiring to go to conferences and collaborate with people. It is not really collaboration with other institutions as such: it is more like a collaboration with the community, being part of the community. And that way you discuss and learn. If I have to mention one organisation, it would be Europeana, since they are good specialists and a core part of the community. Also Library of Congress; and the collaboration with VIAF gave great results. Otherwise we mostly do things ourselves. What could be done to improve the development of LD would be a collaboration among different organisations within the same countries. If there was some kind of umbrella organisation, of which libraries, museums, archives are part of, which moves forward the development of LD, that has a vision for it at national level. For that, I guess, you would need buy-in at government level, or higher level anyways. It would be almost part of the development strategy of the country information system.

# *Q.:* What were the foreseen benefits of implementing LD? Did reality meet the expected outcomes?

A.: One of the foreseen advantages was to be the authority hub to which other people would reference, to have a LD site, a hub, where you have identifiers that others can use. I don't think we have achieved that, because the authority data is missing. You can become authority hub if you provide information about the objects that other organisations/people want to refer to. In our case, the VIAF part is the only one where we have a hub, but it is actually VIAF's hub, not ours.

Another foreseen advantage was better discoverability: provide structured information of our objects/records, so that they are better discovered. Did we achieve this? Yes and no. I am not sure that we used LD to really improve the discovery of our bibliographic information, apart for the VIAF part. And in this, an important step forward would be Europeana. Others at the Library are already working on this, we are already providing data to Europeana. It would really help if they would ask institutions to provide their data

as LD. In our case, with the DOM project, we could have given our data to them, but it is not really part of the workflow. Also, as I mentioned when talking about all different formats, not all RDF combinations are the same: we would still need to map our information to Europeana. But that way we could provide them richer information and our data would be better discovered.

Then there was the advantage of integrating our information systems internally. I think LD initiative helped in this, but not directly. Staff members of the IT department do not use LD themselves, their competence is not in LD; but doing the pilot project gave them some understanding of how they could use this LD approach. The trouble for LD is that the integration of information systems doesn't really require LD. LD is just one of the ways in which things can be done. I think the IT department is more likely to use existing and more limited APIs. However, what they found valuable in the LD approach, is to have URIs for objects, unique identifiers. So they do not understand the RDF part, but they like the idea of using URIs and the fact that you can enrich data with external resources.

And finally there was a foreseen advantage of raising our competence and experience with LD. That was achieved and is still being achieved, since we are learning more about LD and that helps. We are now exploring how to move to BIBFRAME. We already know something or quite a lot.

#### Q.: How about the challenges? Foreseen and actually encountered ones?

A.: In the DOM project, and also in the linked digital collection, the problem was that quite a lot of time was spent to discuss/understand how to represent the data in RDF. Which brought us to build a set of classes and properties, in our case from different ontologies, such as FOAF, Dublin Core and Schema.org. How to put them together to represent what we wanted to express was a problem. In this case we were looking to express for example annotated correspondence, letters, and that was not something for which we could take existing examples and just copy them. Modelling data in RDF is a lot of work. And sometimes you then end up not figuring out how to express something. Like in the case of the XML data provided for DOM, it may happen that, even if the RDF version is more interoperable, since others can look at your data and kind of understand it, because it uses existing technologies, the XML data is more complete and we could have published it without creating the RDF version and probably saving lots of time. We could have created our own ontology, almost copying what it is in XML. So the question is: 'should you spend time on RDF modelling?'

Challenges are mostly competency and financing, and problems of developing a system using limited internal resources, with limited competence and maybe interest in it. Which means that you cannot do much with it. I wish I had the challenge of training, because that would mean I had someone to train! I am training librarians, but they are not the people developing the systems. So we need more people who can develop LD systems.

But the main challenge, apart from the above, is understanding what the benefits of LD are and providing motivation to the management and the IT department. Showing them why we need LD. Because you can otherwise do things in a different way, using proprietary

APIs and still integrate information systems. So why choosing LD? We need an answer to that.

# *Q.:* Do you have any suggestions for those institutions who are looking into implementing LD?

A.: To get started with LD you ideally need to have the necessary resources, to be able to pull it off, and have a management or high level buy-in. So that your effort is supported both financially and in terms of decisions. If you don't have these, then it is difficult, and if you don't have enough competencies on the development side, then it is probably better to team up with someone who can do the development for you.

And, as I mentioned it already with regards to Europeana, and this is something that would be useful not only for fresh starters, but for also data consumers and for those who are already active with LD, it would be great if they provided incentives for others to publish LOD and integrate it with their data. Because, as long as the old way works, to provide them data, then there is not much incentive to do something new, let's say for the IT departments. However, if there was a stronger message, a stronger need, then you can get some better value off it. Again, it goes back to 'how do you show the value of LD?'. How do you incentive new players and existing players to publish LD?

I wanted to ask you what you learned from this study.

[brief outline of learning outcomes at current state of research provided]

I would question your sentence "Benefits of LD are well-known", because we assume that they are, we almost take this like a truth in itself; but are they? When you need to explain the benefits of LD to someone who has not bought this, who doesn't know about it and runs a library and you need to prove that it is worth spending money and time on this, what would you say, apart the fact that anyone else is doing it?

[suggesting positive feedback from users may be a measure of value]

OK, feedback from users and success stories from existing collaborations and projects, where you maybe know the users directly, that is indeed the best way to measure and understand what is the value of providing this. The problem with providing open data is that, beyond the fact that they say open data is the new oil, institutions may say: "now we can sell data for some income, which we will lose if we open our data and it will not be supplemented". How can you show them that you can obtain value from LD that is at least the same of the money they would lose? If for open data you say "some unknown user will use your data in an unknown way", then it is not appealing.

So, existing usage or board stories of how others are using it and getting value out of it, that is what is necessary and would be useful for some organisations thinking about starting publishing LD. It would really help anyone to have this, a set of good successful stories, examples of how people can get value out of it.

I would suggest you to contact the Oslo Public Library, as they are really serious about using LD. They are building a library system based entirely on LD: they want to use LD as a

core technology, rather than as a side functionality. It would be really interesting to understand the reasons of such decisions, the foreseen benefits of doing this.

## Appendix 7 – Interview with the National Library of Luxembourg

#### Q.: Have you implemented LD in any of the library resources?

A.: Yes and no. It is a bit more complicated, in the sense that we did some small LD project, we worked together with another institution from Luxembourg, the National Centre for Literature (CNL). They created the Dictionary of Luxembourgish Authors, at first only in printed version, but they have had an online version since a couple of years. The National Library is managing the online catalogue and discovery interface not only of the National Library but of the entire network of Luxembourgish libraries, including university, public and special libraries. We decided to link our catalogue to the Dictionary of Luxembourgish Authors, and the nice thing was that the CNL liked the project so much that they decided to link their data back to us, to our catalogue, so that people could check in which library a specific book was available. In that sense it was LD, but the reason I say it was not exactly, is that we did not have the resources at the time to use the LD technologies, we did not use triplestores etc.

Since neither in our authority file nor in their (CNL) Dictionary there were not enough identifiers that we could use to properly link the resources, we had to find a different way, so we actually used bibliographic information: for example, for a certain author, we didn't want to use only name, DOB and DOD, but we also extracted information from books, articles, things that the author created, to better identify him/her. That was the first step and unfortunately we have not managed to move to the second step, because in the meanwhile we also started exporting our data to VIAF. The idea was to incorporate all the VIAF identifiers and ISNIs in our authority files. They (CNL) also wanted to import all these identifiers in their Dictionary so that we could then properly link based on either ISNI or VIAF identifiers.

#### Q.: What were the reasons behind your choice of implementing LD?

A.: The idea was that we could bring more information to our users. We wanted to link the authors in our catalogue to the authors that figure in the Dictionary of Luxembourgish Authors, because there is more biographical information in there, sometimes photos.

We aim to continue on the same path that we have started years ago, with author information, because, as National Library we have a mission to create an authority file for people and organisations and what we did, linking our catalogue to the Dictionary, was just a small step. Our idea is to also link it to the big LD resources, like VIAF, DBpedia, Wikidata etc.

While I was before the only person working on LD implementation, now there are at least other two people from the metadata team that are also interested and we are thinking that we might do other projects once we will move to the new building. It is also because from the metadata side, they are now looking at RDA, so they think that once we adopt LD it is going to be easier for them to do things from a content point of view, more than from a technical one.

#### Q.: What were the barriers that hindered LD implementation?

A.: For us it was mainly about lack of human resources. Our team is quite small and we didn't have enough people to invest in such project, since we had not done anything like that before. It would have been the issue of learning how to do it properly. Also, another problem would have been the time to implement the whole infrastructure. So lack of human resources and time were the main barriers to a full implementation. However, we are still interested in the idea of working with LD. One major barrier at the moment is that we are going to move to a new building next year: this is a huge project that it's taking many resources, including the IT team and other member of staff that might otherwise be interested in working with LD.

*Q.:* What is the policy adopted at your institution with respect to LD implementation and use, if any?

A.: There is no policy, at least not for the time being.

#### Q.: What example have you looked at, if any?

A.: One of the first example we looked at, which was always an inspiration for us, and we would like to imitate one day, was the National Library of France, with their data.bnf. As far as I know, they were one of the first national libraries with such a huge project, which included many resources. Just like them, we have a union catalogue, for the whole network of libraries. We also have digitised content, digital-born content and we would like one day to publish all this data on a platform such as data.bnf, where all these resources are not only linked together, but also linked to external data sets, like VIAF, DBpedia, or similar. Other projects we looked at were also the one realised by the National Library of Spain. The British Library started a bit later, but we were also looking at what they were doing.

*Q.:* Have you ever reflected on the impact that the adoption of agreed standards has with regards to creating uniform systems and allowing interoperability among various datasets?

A.: We believe that's the way to go! That is why for all our projects, even for digitisation projects, we are usually trying to look what the standards are, or at least, look at what the other libraries are doing. If there is not a widespread standard, we still aim to create a product that other people can use afterwards, from our resources. Of course, that means that you have a lot of work to do to transform your data, to be able to use the existing standards, if you didn't use them since the beginning. It's a bit of a problem for us. However, we are aware that is the way to go and especially for small institutions, like we are, if you do not use international standards, if you do not try to go with what the bigger libraries are doing, then you are just going to be lost.

# *Q.:* How do you believe collaboration between cultural institutions could support the development of LD?

A.: I think it is very important to collaborate. That is why our project was developed with the cooperation of another institution, at a local level. We are aware that collaboration is really important, that is why we are trying, if not to work together, as now we do not have the necessary resources, at least to look at what other institutions are doing. If we can get involved, we try to get involved. For this reason, for instance, the National Library of Luxembourg is sending its data to Europeana: if we cannot implement something here for the time being, at least we can be part of the Europeana LD platform. I am personally involved in the Linked Open Data Special Interest Working Group and I am Deputy Coordinator within the IGeLU (International Group of Ex-Libris Users). We want to be aware of what our colleagues are doing and collaborate on a different level if we do not have the resources to publish our data in triplestores and use a SPARQL interface.

#### Q.: What were the foreseen benefits of implementing LD?

A.: We decided to adopt LD because we thought that it was going to be an added service for our users. And that was indeed the case. There were a lot of people who liked that we were implementing this, on both our side and the CNL side. It was better than expected, because at the beginning we thought that there might have been few people noticing it, but many people actually noticed and appreciated the change.

#### Q.: How about the challenges?

A.: First challenge was trying to obtain the approval of our management team. I was the one proposing the project at first, so it came from the IT team and initially it was not easy to explain to the directors what the benefits would be. They thought that it would have been lots of work and maybe not much out of it. That is why we did a pilot project, right at the beginning and, I have to say, from a technical point of view the results were quite disappointing, since I did not get many links between the resources; but what helped us was to make a mock-up interface of how it would look like if we went with it as a normal project. And seeing how it might look afterwards was an immediate success with our management team and they gave us permission to go ahead with the project. It was really helpful.

The other challenge was that I did not have enough time, because I was in charge of the LMS, so this was a side project. That is why I could not go ahead and do the full LD implementation as I had wished, that is creating a triplestore and go from there. I had to see how much time I had available and how I could get the results with the available resources. When I started the project I was the only person working with LD, nobody else did.

# Q.: Do you have any suggestions for those institutions who are looking into implementing LD?

A.: For us what was really important was to start small. I remember taking part in a SWIB Conference, four years ago. The National Library of Spain was doing a presentation about

their project and they were saying that it is really nice that more and more institutions were starting to use LD, but they would really recommend to start small and not try to go full scale, change the data you have and use ontologies directly, because that might be too big of a project, you may not complete it and then the management would be against the idea of implementing LD. On the other end, nowadays it is easier to get the approval by the management sphere, because you see many more examples of LD projects. An example that I used with my team and the directors was that even Google has on its search page linked resources and that users find this really interesting. Of course we are not Google! We do not have the resources to do something like that, but we can start small, because we also have lots of data and we can try to link it to resources that our users may find interesting.

#### Q.: Would you like to add any further comment/reflection?

A.: I would be interested in seeing the results of your research, because we are still not doing as much as I would like us to do with LD, but I am really trying to encourage our library to do more. Also, we are a much smaller library compared to the national libraries of France or Germany, but it helps if I can go to my director and say "look, we can start doing things like this".

## Appendix 8 – Interview with the National Library of the Netherlands

I have been following the developments of the other national libraries for years now. It was initially quite a small stable group of people working with LD. Only recently it looks like it is becoming a bigger group of libraries doing things in this direction.

*Q.:* How have you implemented LD in your library resources? Can you describe the main features of the project realised?

A.: I was committed to a project focused on building the digital library infrastructure for the public libraries in the Netherlands. That was a separate project that then became part of the tasks of the National Library, but it used to be in the hands of a separate organisation. We built the National Catalogue based on semantic technology and LD. That was my first and largest experience with LD at the National Library. That project is now part of the National Library tasks. Next to this project we also started publishing all bibliographic data as LD. We published our Thesauri, about two years ago, as LD. In addition, we are now very close to publishing our first draft of the National Bibliography as LD. It will be hopefully live in a couple of months. It is a big step for us. So, these are the three main projects involving LD implementation.

We are also involved in a national programme for digital heritage, together with all the national institutions, such as the National Archive and others. We are building a discovery infrastructure for cultural heritage in the Netherlands which is based on LD principles. However, it is still in a design phase.

For the National Catalogue, we started considering the users' requirements, which should be the starting point of any project. You should consider what people really need and what you can offer them doing things in a certain way. Working with LD, data modelling and conversion are two big things of course. We started building the platform at the same time that the Library of Congress started talking about BIBFRAME. We took a long time thinking whether to adopt the BIBFRAME approach or try to implement FRBR, which now, through the new RDA version, is much more accessible than it was five years ago. We then decided to trust our judgment and we did our own thing inspired by the BIBFRAME model.

Another interesting aspect of our project, was creating a chapter for DBpedia, because we really wanted to connect to DBpedia, but, especially for the public libraries, there was a linguistic barrier. There was no Dutch version of DBpedia at that time. So we went to Leipzig, to the University which is one of the founders of DBpedia, and we started to talk about creating a Dutch chapter. It is still available, but it is a spin-off of our project.

#### Q.: What were the reasons behind your choice of implementing LD?

A.: There are different reasons for the different projects. For our National Catalogue we looked at LD from the data integration perspective, because we wanted to combine bibliographic records with other kind of resources, in order to provide people searching the catalogue with different kind of information. So we built a platform that is able to aggregate different kind of resources and all data gets translated to LD, harmonised and integrated in that way. So one of the reasons was data integration.

Another reason: we saw LD as a way to enrich our data with other resources, like DBpedia and other similar resources available.

In addition, to improve discoverability of our data, for instance by using Schema.org, so that the main search engines, like Google, can really understand what we are talking about.

For the publishing of the Thesauri and the National Bibliography, we really think it is important for us to be the authority on bibliographic data in the Netherlands, so that people can use our data in the most convenient way. And we thought that publishing data as LD was the best way so that people could really access our data in an open and reusable way.

For me it is much more a principle thing: it is our responsibility to provide our data as LD, because it is the best way to obtain data integration with data of other organisations and it is the best way to make your information available, findable to other people. So, I think it is a kind of obligation to do it the right way. And I believe that, if we all approach it like this, results will show. However, we are not there yet.

For a long time, we developed special services for specific groups, audiences. That led to an enormous complexity of infrastructure, with different tools. We are now trying to harmonise all the resources. We are trying to do things so that we publish our data in the most usable way for our current users, but also our future users or new groups that want to use our data. What we are now trying to do with the National Digital Heritage project is looking from the users' perspective. It is currently very hard in the Netherlands to get a comprehensive picture of all information that is available on a certain topic. I think if you go back to your information and you publish it in a very usable way adopting the standards and aligning to reference network, such as thesauri and classification systems; if you really do it properly and you publish your data, it will be a lot easier to surf for the user in the end.

It is also a very important way for the institution of approaching your own information, in order to stay relevant; you should really think about doing these things. In general people think about LD that it is a publication format, a set approach, but it is much more than just another publication format, it is a technology for data integration.

*Q.:* Have you identified the key barriers that delayed or could potentially have prevented LD implementation at your institution?

A.: There are many barriers to LD implementation. The key barriers are:

- Lack of knowledge. It is something we are really struggling with. To give you an example: for our National Bibliography we have set a LD environment with the usual ways, triplestores, SPARQL endpoint and all the rest. But our IT department said "these are all new things for us and we do not know how to support them". So we ended up finding an external hosting party to support that environment. I think that's still a big problem. My son is in his third year of Computer Science and he still has not heard of Semantic Web and LD. I think we have a general problem in that area.

- Lack of usable tools. There are lots of tools around, but many tools are in experimental phase and often they never grow up to be really usable. I must say that we started this project in 2011 and if I look back, in the meantime quite a good number of tools have been put on the market, really sophisticated things and you can get good support on them; but it is still a challenge, if you really want to choose to do things the LD way.

- Licensing. LD really only makes sense if you do things as LOD. There were many discussions on licences. Part of our metadata is still not CCO available, so we had to make deals with companies such as OCLC. That's quite hard. And it is something we really need to get out from. Metadata is the way to make our data findable, so there shouldn't be barriers there. That's my personal way of looking at it. Europeana did a great job requiring all CCO data and now this awareness is spreading.

- Doing LD also has to do with organising things in a different way than the one people are used to. One thing I like about LD, if you do things by the book, it should be sufficient to organise your own thing and publish it and then data will be integrated. So you do not need to get all the parts together and bring them into a central system and then do lots of interesting things with them. However, especially in my organisation, people still think in a centralistic way, so if they build a service for users, they aggregate all data in their system and then they start working with it. That's the change we are trying to bring in our organisation, in order to say "OK, we have different groups producing interesting pieces of metadata and content and they should be sufficient to concentrate on that and then have an infrastructure that brings all the parts together, using LD in that way. There is a thing going on called 'resource-centric' approach against a 'repository-centric' approach and that is very basic for LD.

In general, especially with the Catalogue project, we experienced that we were building this for the various libraries in the Netherlands and they were saying "Yes, nice platform, but it should behave like the one we had before". So they were asking for a catalogue that behaved like the old one. It is really a pity when you have so much more sophisticated intelligence and people just request the things they are used to. That's the general problem with innovation and doing things differently. People have to mind-switch and that's not easy. So it is a barrier.

There are still several opinions about LD and the many possible ways to approach it. It would be nice if, when we started the project, there were guidelines, some guidance or general ideas.

*Q.:* What is the policy adopted at your institution with respect to LD implementation and use, if any?

A.: We only have a policy for determining URIs. It is not really a document on the use of LD. However, our general policy explicitly mentions LD as being strategic approach to share our data and make our data feasible in the best way. The adoption of Schema.org to ensure that even Google understands what kind of information we have. If you want to do things the right way, then you need to adopt tools like Schema.org and publish data in an open way, and also use well-known vocabularies.

#### Q.: What example have you looked at, if any?

A.: If you look at what OCLC has done with data modelling, you can surely learn a lot from that.

I have particularly been following the national libraries, what they have been doing. The British Library was already doing great things when we were working at our project. Also, the National Library of Sweden. For us, in particular, the German National Library was interesting, as they use the same cataloguing platform as we do.

We learned a lot from DBpedia, when we were creating the Dutch chapter. It is a very important resource in the LOD cloud and they have an enormous experience in working with LD. We actually had one of their PhD students working part-time on our project, just to transfer knowledge, and that was great.

In Netherlands we also have a number of other cultural institutions that are doing things with LD as well. We did a lot of exchange with them. The Amsterdam Museum together with the Free University in Amsterdam was one of organizations doing interesting projects with semantic technologies. It's, as I said, learning from each other.

If you look beyond the library sector, there is a very interesting semantic conference in Amsterdam (Semantics Conference), where you see all different kind of environments, especially business, working on semantics and LD. It is interesting to see the developments. It is really becoming a business working with LD and creating smart solutions with it.

*Q.:* Have you ever reflected on the impact that the adoption of agreed standards has with regards to creating uniform systems and allowing interoperability among various datasets?

A.: Especially in the library community, we keep talking about the perfect model, but what we should be really talking about is the pragmatic way to really get things done for the users' benefit. I think we are still a long way from that.

Data integration is a very important aspect of LD. You automatically have to adopt standards, or at least use the ones that are available and do things in a way that people can understand what you are doing and machines can 'understand' too.

*Q.:* Do you think that the fact that there are so many different standards you can choose from to implement LD is an issue?

A.: It is a difficult problem. If you look at BIBFRAME, they decided to develop a special BIBFRAME vocabulary, which is alright if you look at BIBFRAME being a replacement of MARC, as library exchange format; but if you want to use BIBFRAME to communicate to other environments beyond libraries, it is hard. For example, if you look at a book description in Wikipedia, it is easy and straightforward for everybody; if you look a bibliographical description of a book then things become very complicated (works, manifestations, expressions etc).

What LD facilitates is that you can model your data in different ways at the same time. For the catalogue, we had a description model for user interface using all specifics of bibliographical description; but we also supported, with the same API, Schema.org, in order to present our information in Schema.org vocabulary to Google. In LD you are free to adopt several ways of modelling and you can bring them together, you do not really have to choose to do it in a way rather than in another way. However, in general, there should be more consent about the way you want to do things, at least in the library sector.

*Q.:* How do you believe collaboration between cultural institutions could support the development of LD?

A.: We need to share the things we are doing and learn from them. I must say that the reason I am really happy with conferences like SWIB, it is that it is really useful in order to understand practical aspects of doing LD. I think it is really important to exchange information and learn from each other.

I think collaboration is essential from a learning point of view, but also in order to demonstrate the value of LD, because if we are all working on LD then it will be a lot easier to bring things together in a relevant way for the average users.

*Q.:* What were the encountered benefits of implementing LD?

A.: What I have noticed is that it really forces to think in another way to your data and I think that's an interesting aspect. Publishing LD is sometimes called a 'social contract'. So, if you publish your data you should do it in a persistent way, so that people can continue using it. You have to make sure that data stays available, which is something that is not always easy, if you consider problems with web pages. But you also have to open your data and show what it is like, so there is also the quality aspect. You need to reflect on what you are doing and if you are doing it the proper way. Other people's eyes will force you to do it the right way; and that is a large benefit, because you are responsible of your data. Sometimes policies are written and then forgotten. The passage to LD means that you'll have to re-think all the things you are doing and that's quite useful.

If you look at the big publishers, like Springer Nature, they are adopting LD technology in order to get better quality at lower costs. There is a business advantage too, which is interesting.

#### Q.: How about the challenges presented by LD implementation?

A.: In general people think that it is a lot of overhead. If you want to do LD the proper way then, in a certain way, you have to do more things than you really need to, like organising proper identifiers, which is quite hard. Each LD resource is identified by a URI which is unique and resolvable; but in order to come up with these URIs it takes time. We spent one year and a half writing a policy for that. And we are still discussing with other 20 people to have it approved. This only to organise a very basic aspect of LD. Also, because it is new people, they need to learn new things and work with different tools than the ones they are used to. In general, it looks like a lot of trouble in the end.

It was also quite hard to get an idea of which tools were really usable and fit for our purpose. We spent a lot of time on this. We had a learning project, to learn about tools and how to approach things. What we learned during the project is that LD is more promise that reality. So we ended up building a semantic platform which is based on combining a knowledge graph with index search technology, because we could not get it working completely as semantic platform.

Another important part of the project was designing the user interface, in order to really get the value of LD. That was quite a hard project, since our users required our old catalogue interface, so we had to do experiments with all kind of different approaches. Initially we thought that combining our resources with other LD resources would be very easy to do. In practice it is quite hard, because of the scarce availability of relevant LD. Also, because of the many different ways to bring them in a usable way to users. With LD is very easy to deliver a huge amount of data, but you want to provide the users with relevant data. In addition, there is a challenge for retrieval systems. It is very easy with LD to generate lots of information, links, but in order to make the right selection, you really need to have high quality data. That was something that somehow disappointed me: the fact that it needs lots of work, it needs many different things, probably it firstly need to do work on your own data, to make data richer and create more connections, more relevant ones.

In general, I think the challenge is the poor availability of relevant resources and I want to believe that it is a matter of time for this to change. A big problem with LD is to understand what is the right way to present your data so that it can be reused by a large group of people.

Q.: Would you like to add any further comment/reflection? Do you have any suggestions for those institutions who are looking into implementing LD or reflections on future LD development within the ILS sector?

A.: It is very important that the suppliers of library technology make a serious deal out of it. I have been talking to suppliers for digital heritage and they say that they do not work with LD, because their clients do not ask for LD support, so they do not build it. On the other side, if you talk to institutions, they will say that they would love to adopt LD but suppliers do not provide the tools. There is a miscommunication problem. I think what they are doing in the UK, with data.gov.uk, is very important. It should be the way of doing things, and not the choice of an individual institution. That is what we are trying to do with our national strategy for digital heritage, which includes also libraries. We believe that LD is one of the fundamentals of data integration, and working together, and bringing all our different information together. So we are trying to promote this in a national set of requirements for all the systems in our environment. Therefore, if you want to be a supplier for museums or libraries in the Netherlands, you need to adopt the LD principles. Government and other initiatives at national level could help us and make things much easier. But, of course, it will be a long-term thing.

On the other end, if you really want to do it the practical way and you just start working with LD, you can achieve a lot too. So it is really important to do both things: go, work with it and learn about it and then work together and look at doing things in the right way. That is my approach.

## Appendix 9 – Interview with the National Library of Wales

Q.: Have you implemented LD in any of your library resources?

A.: Yes, we have implemented LOD in the following projects:

We started a number of years ago looking at shipping records - a project where we had a volunteer in the Library transcribing some physical shipping records into Excel. We needed a way to store and make available the data and due to the content being about people, places and events, LOD seemed the best option.

In 2015 we entered the LODLAM competition, looking at what data Wikidata and Wikipedia held about NLW (National Library of Wales) images. LODLAM is a collection of libraries, museums, galleries, and archives, and the LODLAM LOD Conference is held roughly every two years.

More recently we have been looking at implementing the IIIF (International Image Interoperability Framework) standard, based on LOD, to make NLW digitised images available. We worked with LOD Annotations for transcription projects. We developed a Crowdsourcing system. In addition, we have a Wikidata scholar, who is looking to enhance the data on Wikidata for our Landscapes collection. We are also working with our catalogue supplier Ex-Libris to expose our catalogue data as LOD. We are sharing out data with Europeana using a EDM format.

Generally, we have been storing lots of our data as XML, and more recently we have been moving to converting our XML data to LD, and it tends to be the case for all the projects we have worked on. Instead, with the IIIF we may be start looking at storing data natively in LOD format; but it is still quite the beginning.

We mostly developed the LD conversion in house. With IIIF we used an off-the-shelf image server, but wrote conversion programs to transform XML into LD. Using an explicit LD standard is a lot easier than having to make your own choices about metadata fields. For the shipping records project we had to create a kind of standard ourselves, and that takes much more research and work than reusing an existing one.

#### Q.: What were the reasons behind your choice of implementing LD?

A.: It is mostly due to interest. We kept going to conferences and hearing about LD. We wanted to try to see what this was about. With our first project, involving the shipping records, there were not many choices. We could have decided to do that just in a database. However, to try and make it available to researchers, we had to create some sort of API on the database. So we thought that by releasing it as LOD, users could have just downloaded LD and process it themselves, and we would have not needed to support it that much.

With IIIF, it was mostly looking at the standards, which was really useful, because it had many other features, like standards for digitised images that had existing viewers, and enabling to share content. It was actually something in the background, rather than a reason to choose it, but it was really powerful.

Another reason is related to the intention of enhancing existing datasets, reflecting on how we could take our datasets and match them with data which are held elsewhere, such as Wikipedia and Europeana, and try to enhance our data.

# Q.: Have you identified the key barriers that delayed or could potentially have prevented LD implementation at your institution?

A.: One of the main barrier was the lack of top down understanding and drive. I am not sure the senior management understands the benefits of LD and how important it is going to become over the coming years.

Also, it requires staff training, particularly developers being able to travel to conferences, and being able to learn about latest developments. And that was easier in the past than it has been more recently. It is not something that you can learn remotely.

*Q.:* What is the policy adopted at your institution with respect to LD implementation and use, if any?

A.: We do not have an official policy, and so far it has mostly been a matter of interest rather than upper level management pushing towards LD. All we currently do with images follows IIIF standard, but it is not written down anywhere. There is no official policy to say that we will release all of our projects in LOD. I do not think there is any mention of LD in our general strategy either.

### Q.: What example have you looked at, if any?

A.: I found that the IIIF example proves the usefulness of LOD: it is a standard for sharing and working with images, particularly digitised images, but it can be used with other types of images. It is in JSON/LD format, which means that you can display your digitised box of newspaper through a shared collection of viewers. In the past we used to develop a viewer for different projects, while now we can just use an off-the-shelf viewer for all of our content. Other people can reuse our content in their systems. So, I think this has been a very successful example.

### Q.: Have you looked at any of the national libraries projects?

A.: At the first IIIF standard meeting, in 2013, saw a group of libraries participating which were already involved in it, including British Library, Oxford, Stanford, Princeton, Harvard, BNF. And the actual meeting was at the National Library of Denmark. So there were already several big institutions involved.

Q.: Have you ever reflected on the impact that the adoption of agreed standards has with regards to creating uniform systems and allowing interoperability among various datasets?

A.: It appears that many library XML standards that we have been using for a long time are going towards LOD, with the promise that it will allow better interoperability. We have definitely seen that with IIIF, but with a standard like METS (Library of Congress format) which is more important internally, to hold our structure and preservation information, we will have to wait and see if the move to LD assists interoperability. I think that moving form XML to LOD does not necessarily guarantee interoperability, unless people agree on the standards. However, it does feel that things are slowly changing, away from XML to more LOD standards.

Q.: Has collaboration with other institutions helped the actuation of your project? How do you believe it could support the development of LD within the information and library context?

A.: IIIF is a very useful community, which has two meetings every year and it is led by UK use cases. So, whenever we go to these meetings everyone brings different use cases, which are discussed, and they decide what is common between everybody, and then they start mapping that to LOD format. It has been very positive and useful seeing what other people are doing. For example, we learned, bringing our case to one of those meetings,

how to structure a scroll in IIIF. And then we came back to the Library and we were able to find scrolls which we previously did not know what to do with, and have now a way forward to make them available.

With the shipping records, we worked with the BBC RES (Research and Education Space) project, a LOD project looking at gathering LOD about education. And they were really helpful, they gave us lots of advice about what relationship to use. As it was our first project, I do not think we would have gone anywhere without their help. We had questions and they answered them. Instead with IIIF is more a collaborative process of getting in the right direction.

*Q.:* What were the foreseen benefits of implementing LD? Did reality meet the expected outcomes?

A.: Making our data more available and more out there for people to be able to see it and use it for their purposes. With the shipping records, I really hoped that we would release the datasets and get many people to reuse it, but that has not quite happened, for a number of reasons, I think. Probably the fact that we have not advertised widely that we have got this really rich dataset available; also I think that it requires a certain level of ability to work with LD. You have to target certain people who are really interested and keen to put in the effort. It is really interesting to see what the Wikidata scholar has done: all the Wikidata is stored as LOD, but since it is stored all in one place, there is so much tools they can use within it. They are able to visualise all the landscape drawings and put all of them on a map, and do other things with LD because it is all in one place. That is something we have not been able to do with our own release of LD. Although with IIIF we hope someone will be able to develop the tools in the future.

#### Q.: How about the challenges met?

A.: Looking at LOD and trying to choose amongst various different relationships was quite intimidating. Understanding where to start from and where to start looking at these things. And it was very difficult to know where to go for advice. I think that was our biggest challenge. The BBC RES was good; we went for a meeting down to London.

There are other several issues: looking for vocabularies, finding a vocabulary that suits your data. Sometimes you have to adapt your vocabulary from existing ones. And there are so many out there, it is hard to find the one that suits best.

If you can find a standard which does what you want to do, like we have found IIIF, is good. We are now looking at what to do with WWWI, the list of soldiers who died in WWWI from Wales. There was already a big ontology out there created by another project. So we were able to use that off-the-shelf, with some modifications. With the shipping records, instead, we could not find anything similar, so it was a lot more difficult.

*Q.:* Do you have any suggestions for those institutions who are looking into implementing LD?

A.: I remember the National Library of Scotland came to visit us and gave us a talk about LOD, and they suggested to start with something. It does not have to be big, it can be a manual, a record, but just to start with something, otherwise you will never learn what the opportunities are. I think that is a very good advice.

## Appendix 10 – Interview with the National Library of Portugal

#### Q.: Have you implemented LD at the National Library of Portugal?

A.: We have not yet implemented it, but we intend to. It is in our strategy. We have a strategy of open and linked data: every metadata is open and we have a website with all the information. It is all open and available under CCO licence. However, it is not ready for the Semantic Web yet, as we do not have the resources to do it, but we are searching for help in that area.

#### Q.: What would be the reasons for planning to implement LD?

A.: We want to publish our data as LD because we want to be more visible on the web, more visible to search engines. Also we want our data to be reusable by other players, and we want them to be linked and enriched by other sources.

*Q.:* What have the obstacles been so far, which have prevented the implementation?

A.: We have very few human resources. We do not have the knowledge to do it. On the library side, we try to stay up-to-date, reading everything that we can, and understanding the concepts. Nevertheless, we need the technological help to do it. So we tried the support of some technological partners, such as the software *Aliada*, from Spain; but we are still searching, we have not made any decision yet. We hope that participating in some European project could help us, because we do not have the knowledge and the ability to go alone.

#### Q.: What are the examples you are looking at?

A.: We are following all the national libraries examples very closely, especially France, Spain, British Library, and also Library of Congress BIBFRAME initiative. In addition, we are involved in the RDA transition process, not implementing yet, just the map of the workgroups; but we are following all these initiatives.

*Q.:* Could you provide some details of the project involving LD that you have in mind? How and for which purpose you want to use LD?

A.: Our authority data is already mapped to VIAF, but we would also like to transform our bibliographic data. In addition, we would like to have links to other vocabularies, in particular to geographic names in what concerns digital objects, because we have a digital library and we are planning to build a platform with geographic resources, and it would be

important to have links to GeoNames and similar vocabularies, that could make our resources more available.

### Q.: Do you have a policy on LD implementation?

A.: No, we do not. We are currently exploring and processing, but we have not made any kind of decision or option. We are at a very early stage.

Q.: What do you believe is the role of standards with regards to interoperability?

A.: We think it is crucial, very important. We follow all the standards we can! For bibliographic data it is the norm: we have UNIMARC. And for what concerns our digital objects, we have just implemented a new metadata model, and we are applying standards to our old metadata schemas. This is because two years ago we had some local schema, but we replaced them with international standards.

Q.: What do you think is the role of collaboration in the context of LD adoption?

A.: It is the only way for us to do anything in this field, because we had a hard time in our public finances, with very severe budget cuts. Therefore, we cannot afford to buy all that is necessary for the implementation. We do not have the internal resources to do it.

*Q.:* What do you recognise are the biggest obstacles preventing LD implementation?

A.: Lack of staff and lack of budget.

Q.: Which do you think are the benefits of implementing LD?

A.: Augmenting visibility of our data, as well as enriching our data. In addition, being able to offer new services. LD would create the possibility of new ways of displaying our data, and also of making the data available beyond the library community.

*Q.: Would you like any comment? What do you think about the future development of LD within the library environment?* 

A.: I think we should work more in order to understand how data is being used, because there is a lot of triples accessible and downloadable, but we do not know... There are few practical examples of LD applications. It would be useful to know how LD is applied by other institutions.

## Appendix 11 – Interview with Open Knowledge Greece

Firstly, I would like to state that our project is not officially appointed by the National Library of Greece. Instead I represent Open Knowledge Greece. We are working on a project with open data, to show what can be done with LD and Library data. So my opinions do not represent the official response of the National Library of Greece, but of Open Knowledge Greece.

*Q.: Can you provide some details of the projects achieved by your organisation in Greece?* 

A.: A first pilot of transforming data of authority files from MARC format to LD used the ontology developed by the German National Library. It was also the controlled vocabulary for Greek Wikipedia. We are trying to persuade libraries in Greece to use this core catalogue in order to develop their own authority files. Core authority files are provided by the National Library of Greece. At first, we started with smaller libraries in Greece, public libraries. We had the full bibliographic catalogue and the authority files. So we made the first tries to convert bibliographic data from MARC format into LD.

We chose BIBO as data model for bibliographic records. Then we adopted the ontology of the Deutsche Nationalbibliothek for the authority files. After the initial conversion, we also developed an application which uses these data sets. We established links with different bibliographic data sets. We started in 2012 and we made an application for using the authority files to provide users information with data from the library but also data from DBpedia. We also used other data sets that were available at the time, similar to DBpedia. So we had to use the same conversion method and technique already in use at the National Library of Greece to transform their authority catalogue into RDF. We then developed authority files for different data sets and this was used as authority control for Wikidata and Wikipedia.

#### Q.: Why do you believe libraries should implement LD?

A.: It is the trend that libraries are going to use LD someday, I believe in almost all of their functions. Libraries are the best candidate for blowing LD technologies in their infrastructures. Even though it is not very easy to get library staff to get involved in all this movement towards using LD functionalities.

#### Q.: Have you ever reflected on the barriers that hinder LD use across libraries?

A.: Mostly, it is people from libraries that would like to keep doing their job in the traditional way. Also, they do not want to provide their data. They are difficult to persuade, since they think that data belongs to them and they do not want to open it. They believe that data is somehow a resource of the institution, so you cannot make any application, for profit or non-profit, without them being involved in any single step of the development. People do not know how this will change. In my opinion, they do not want to leave the way they are conducting their jobs.

#### Q.: What example have you looked at, if any?

A.: 2012 was still early years of LD adoption for libraries. The state-of-the-art at that time were the projects of British Library, and the national libraries of France, Spain and Germany. I think by that time the Library of Congress had not yet realised a LD project, but definitely their example influenced our work. In fact, we were trying to implement similar features into our project.

*Q.:* Is there any policy in place at your organisation with regards to LD implementation and use?

A.: There is no policy, no official document available.

#### Q.: What do you believe are the benefits that LD can bring to libraries?

A.: Multilingualism is a big gain of LD adoption. LD gives a very easy way to obtain multilingual data in different formats. Also, you can reuse data from different data sets, data that was originally used for a different purpose: so you can re-purpose data and implement new application for data. In addition, I think many infrastructures for libraries can have improvements for workloads, as they can reuse information that is already available somewhere else. They do not have to re-implement authority files, they can have a core authority file and that can be reused. We are currently trying to obtain feedbacks from the libraries in Greece, but it is a bit too early.

*Q.:* What would you say is the relationship between LD spread and further development and the adoption of standards?

A.: Interoperability is a big topic now. If we consider Europe, there are different countries with different languages and cultures. LD is a big plus for this. Common standard is something that somehow, someday we will have to adopt in all functionalities of information theory and information technologies.

*Q.:* How do you believe collaboration between cultural institutions could support the development of LD?

A.: We have recently joined Europeana Tech Group, which gathers different institutions across Europe. In addition, there is a movement that wants collaboration amongst libraries in Greece. We are doing workshops for libraries and librarians.

*Q.:* What would you recognise as the main challenges presented by LD implementation?

A.: The fragmentation of the different ontologies used today is a big challenge. Even after 5/6 years after the first project was realised, there is still not a concrete and specific ontology to describe library data. Across projects you can see the adoption of different ontologies. However, there is no drive towards interoperability. There is not a good standard on how to transform data into RDF and LD.

Q.: Would you like to add any further comment/reflection?

A.: Data technology is helping a lot in the area of authority files and I think there is more to come in the next few years.

At last year KohaCon there was a big conversation about using LD in Koha. As a good practice, they showed that librarians can at least have LD as data quilt in the records, so that it will be ready to inter-link to different data sets.

## Appendix 12 – Email interview with the National Library of France

Q.: Have you implemented LD in your library system?

A.: Yes

### Q.: What were the reasons behind your choice of implementing LD?

A.: Linked data was used for digital preservation and for discovery. As preservation mainly used the technology (RDF and SPARQL) for its database, but not the other aspects of the linked data philosophy (entity driven web interface, reusable open data), I will mainly focus on the discovery tool (data.bnf.fr).

As to preservation, the choice of RDF technologies vs. a search engine or relational database was the result of a risk analysis. The use of W3C standards and the flexibility of the model were the two main reasons behind that choice.

As to data.bnf.fr, linked data was considered the best suited means to our goals. Our goals were: make an entity driven web site that was accessible and visible from search engine results; make the underlying data freely reusable by, and interoperable with datasets from, stakeholders inside and outside the library sector (archives, museums, research and other stakeholders from the world of cultural heritage). Linked data technologies seemed the best fit for all those goals.

# *Q.:* Have you identified the key barriers that delayed or could potentially have prevented LD implementation at your institution?

A.: As early adopters, we had classical scalability and internal skills issues. We had knowledge of linked data standards, but had no practical or technical knowledge of linked data technologies at the time of choosing them, at a time when they faced scalability issues. This forced us to have an extremely articulate argumentation to see why it was worth investing effort in such technologies. Having developed internal IT skills on an up and running preservation system helped the adoption of similar technologies for data.bnf.fr.

#### Q.: Can you describe the main features of the project realised?

A.: Preservation system (SPAR): the database of the system is recording all the metadata required to preserve our digital assets. Part of this data act as the brain of the system, telling him how to behave on a particular asset.

Data.bnf.fr: entity driven website built around BnF's discovery data. All the knowledge about BnF collection data is structured around an entity driven website, where each page is automatically generated and provides a conspectus of the things around and about it that are available at BnF, merged across different data silos (catalogue, archives and manuscripts, digital library, web archives, virtual exhibitions...). The 3 main aims of the website were to be visible on the web and particularly on search engines, to be legible and understandable (linked, organised pages on the browser side; linked, organised data on the data reuser side); to be reusable (all underlying data technically available and legally reusable through an Etalab license, which is close to CC-BY)

Q.: What example have you looked at, if any? Do you reckon a particularly successful project within the information sector has proved the usefulness of LD? How about projects beyond the Information & Library sector?

Inside the library sector, LIBRIS inspired us; outside, we were inspired by the BBC website and Dbpedia. Nowadays, Wikidata is a good example of the usefulness of LD.

*Q.:* What are the policies and systems adopted at your institution with respect to LD implementation and use?

A.: For data model and vocabularies used, we have a mix-and-match approach: reuse classes/properties whenever appropriate and relevant; create and maintain specific classes/properties whenever needed; use persistent URIs as much as possible; have dereferenceable URIs everywhere.

*Q.:* Have you ever reflected on the impact that the adoption of agreed standards has with regards to creating uniform systems and allowing interoperability among various datasets? What would you say is the relationship between LD spread and further development and the adoption of standards?

The core grammar (RDF, RDFS, OWL, SKOS, SPARQL) is seminal in building interoperability across different lined-data-driven datasets, but each dataset refines this core logic with a specific data model made of various vocabularies where each implementer is absolutely free of his/her own choices. The greatest challenge in this is to have a simple and easily understandable data model that makes easier data consumption, and to maintain its classes and properties over time as reused vocabularies are maintained and other vocabularies are created and adopted by the community.

*Q.:* How do you believe collaboration between cultural institutions could support the development of LD?

A.: LD should be a means, not an end. Collective, entity management tools across institutions should be a good use case for data curation, in the back office, and entity discovery, in the front office, and LD can be a good means to that end – which requires a strict and consistent provenance management policy to work across different data producers.

*Q.:* What were the foreseen advantages of implementing LD? Did reality meet the expected outcomes?

A.: The foreseen advantages where increased reuse, across domains; building a unified data model across internal silos; increased discoverability and visibility from search engines.

Unforeseen advantages were the most visible entities on the web were rare, niche entities, for which we were competitive.

#### Q.: How about the challenges? Foreseen and actually encountered ones?

A.: Unforeseen issues: reuse is still impeded by the RDF data model. Everybody wants JSON and CSV, not RDF, which remains a niche technology in the IT sector.

The main technical challenge was regular and efficient data updates, as data.bnf.fr is built on an export from pre-existing data silos with their own data update policies, and is a massive website where all information is linked and can be the target, or the source, of a link.

Allowing for data model improvement without jeopardizing existing reuses proved quite of a challenge.

Skills and human resources were, and still are, the greatest challenge: In this innovation sector, people on the library and IT side are very mobile, and the skills required remain very specific. This means a lot of training and documentation is required to make the project sustainable.

*Q.:* Would you like to add any further comment/reflection? Do you have any suggestions that you would like to pass to institutions who are looking into implementing LD or reflections on future LD development within the ILS sector?

Make use cases, then ask yourself what technology is best suited for you. LD may, or may not, be the answer, depending on the use case and your own resources.

## Appendix 13 – Email interview with the National Library of Spain

Q.: What were the reasons behind your choice of implementing LD?

A.: Experimentation; improve resources visibility in the web; investigate multilingual capabilities; open the door to data reusing (both as consumers and publishers).

*Q.:* Have you identified the key barriers that delayed or could potentially have prevented LD implementation at your institution?

A.: At the very beginning, lack of knowledge or expertise; in other stages (always looming, really) lack of funding, lack of human resources; at the beginning, lack of a clear definition and purpose, a clear strategy is badly needed in this kind of project.

Q.: Can you provide more details on this?

A.: The turning point for us was to find a focus, a project combining data wrangling but also real and tangible targets; being available to show to our managers a product (such as datos.bne.es) was pivotal to make LD one of our key projects for the future.

Q.: Can you describe the main features of the project realised?

A.: Data modelling, based on FRBR reference model; extract and connect useful entities (such as Work or Agent), and profit from them, building useful insights into our data. Extra: 3.1. What were the main steps taken? Exhaustive data analysis, finding strengths and weaknesses; data mining, finding rules in our data structure to jump from a record-driven model to an entity-driven model; vocabulary selection (thereafter was vocabulary building); web and services design.

*Q.:* How much of the process could benefit of automated or ready-to-use technologies?

A.: The data analysis and RDF conversion was carried out by an in-house tool; all other services (data repository, data access, web indexing) was all based on cutting edge free software.

Q.: What example have you looked at, if any?

A.: Mainly data.bnf.fr.

Q.: Do you reckon a particularly successful project within the information sector has proved the usefulness of LD? How about projects beyond the Information & Library sector?

A.: I think that great really useful and impressive real, in production, case studies have yet to come.

*Q.:* What are the policies and systems adopted at your institution with respect to LD implementation and use?

A.: URI creation, maintenance and persistence. Use of Open Source technologies; all the technologies (both at the data level and the web level) use the same JSON objects; content negotiations.

#### Q.: Can you provide examples/more details?

A.: URIs for entities are relatively new for libraries, mostly to access bibliographic resources. Our view is the creation of URIs for entities (Agents, Works, Places...) interacting together.

Q.: Have you ever reflected on the impact that the adoption of agreed standards has with regards to creating uniform systems and allowing interoperability among various datasets?

A.: LD has the potential of allowing interoperability far beyond what has been achieved until now. Two (or many) can be describes using different models and vocabs, and become fully interoperable. The standards have moved to the data structure level, rather than in content.

Q.: Has collaboration with other institutions helped the actuation of your project? How do you believe it could support the development of LD within the information and library context?

A.: The more institutions share their data, the more possibilities for cooperation. Institutions will focus in their respective expertise or interest fields, relying on third parties to get extra data. Great connection hubs (such as VIAF, ISNI, Geonames, Wikidata) will be essential as gateways between datasets.

*Q.:* How do you believe collaboration between cultural institutions could support the development of LD?

A.: LD interoperability allows participating institutions specialise in what they do best or what they know more about.

*Q.:* What were the foreseen advantages of implementing LD? Did reality meet the expected outcomes?

A.: One of the outcomes of our project has been to visualise how it can evolve, that is to say, to realise the potential of LD as a way of working with data, to have a fresh look at our data, revise how we create it.

### Q.: How about the challenges? Foreseen and actually encountered ones?

A.: To find a solid strategy to continuing development. It's really important to take the next steps quite cautiously, with a vision; the main danger is wasting the scarce resources available.

#### Q.: What would you say went wrong?

A.: We lacked at the beginning of a strategy, at several levels, that lead us to a dead-end, forcing us to think it over what we really wanted and how to get it. For future LD developers, the main advice is design a roadmap before even putting to work.

# Appendix 14 – Email interview with the National Library of Czech Republic

*Q.: Have you implemented (or are you planning to) LD in any of your library resources?* 

A.: No.

*Q.: What barriers were identified that discouraged or impeded the adoption of this technology?* 

A.: The NL CR databases consist of data of different quality, part of records was catalogued according to AACR2 or RDA (1996 -), but most of data were result of retrospective conversion of printed catalogue cards (up to 2/3). We plan to improve quality of data before we make steps to LD. So, instead of testing LD we decided to strengthen authority work, to receive reliable results later.

Q.: Have you reflected on benefits that LD could bring to your institution?

A.: Yes, we can see the benefits for users, but on the other hand we can imagine growth of work connected with authority data.

*Q.:* What impact, if any, could the examples offered by successful projects accomplished by other institutions have on your future choices?

A.: Positive examples could help the LD to come to the fore. That can bear on better financial support etc.

*Q.:* Have you ever reflected on the role that the adoption of agreed standards plays with respect to creating uniform systems and allowing interoperability among datasets?

A.: Yes, we are a producer of a national bibliography, we share our records with both Czech and foreign libraries, we always take into account standardisation.

*Q.:* Have you ever considered what the relationship between LD spread and further development and standardisation is?

A.: Yes, we observe development of international standards.

## Appendix 15 – Email conversations

# Email response (original text and translation) provided by the National Central Library of Florence.

L'impegno sui Linked data della BNCF ha prodotto a oggi la disponibilità del Thesaurus del Nuovo soggettario in SKOS <u>http://thes.bncf.firenze.sbn.it/thes-dati.htm</u>

Sul versante dati bibliografici stiamo lavorando a varie ipotesi. Da un lato vogliamo evitare il rischio dei silos (<u>https://www.asist.org/events/webinars/from-marc-silos-to-linked-data-silos</u>), dall'altro pensiamo che sui linked data occorre seguire il mainstream <u>https://it.wikipedia.org/wiki/Progetto:GLAM/Biblioteca Nazionale Centrale di Firenze#</u><u>Wikibib</u>

Anche la BNCF ritiene che i metadati creati dalle biblioteche debbano essere "del web" e non solo "nel web". Migliorare la ricerca dell'utente e potenziare le capacità di controllo bibliografico delle biblioteche.

The work invested on LD by the BNCF (Biblioteca Nazionale Centrale Firenze= National Central Library of Florence) has produced the availability of the Thesaurus Nuovo Soggettario (New Subject Heading) on SKOS <u>http://thes.bncf.firenze.sbn.it/thes-dati.htm</u>

For what regards bibliographic data, we are working at various hypothesis. On one side we want to avoid the risk of silos (<u>https://www.asist.org/events/webinars/from-marc-silos-to-linked-data-silos</u>); on the other side, we believe that on LD it is necessary to follow the mainstream.

<u>https://it.wikipedia.org/wiki/Progetto:GLAM/Biblioteca\_Nazionale\_Centrale\_di\_Firenze#</u> <u>Wikibib</u>

The BNCF believes that library metadata should be "of the web" and not only "on the web". Improving users' search experience, and enhancing the bibliographic control capacity of libraries.

### Email response provided by the National Library of Liechtenstein

Dear Ms Cagnazzo,

the National Library of Liechtenstein hasn't yet implemented linked data due to lack of resources.

I agree, that this information is used in your research.

### Email response provided by the National Library of Bulgaria

Dear Laura,

The National Library of Bulgaria has not implemented linked data and does not have plans of implementing it for the foreseeable future.