MASTER'S THESIS IN LIBRARY AND INFORMATION SCIENCE
FACULTY OF LIBRARIANSHIP, INFORMATION, EDUCATION AND IT
2024

Moving towards Linked Open Data in Libraries

A case study from the National Library of Sweden and the National Library of Finland

Sabine Weidinger

UNIVERSITY
OF BORÅS

English title:	Moving towards Linked Open Data in Libraries A case study from the National Library of Sweden and the National Library of Finland
Author(s):	Sabine Weidinger
Completed:	2024
Abstract:	Linked Open Data (LOD) gives libraries the opportunity to open up their data to the web, making it discoverable and reusable. To do this, libraries need to transform their vast amounts of existing data into LOD and move towards creating LOD. Unfortunately, there are many challenges, such as a lack of expertise and mapping problems between formats. This can lead to a feeling of being overwhelmed, which in turn can lead to libraries sticking to traditional ways of cataloguing and presenting their data. It is therefore important that libraries have access to detailed project presentations. Therefore, this master's thesis presents the reasons for adoption, the steps taken, the challenges encountered, and the solutions used for the two cases of the National Library of Sweden and the National Library of Finland. For this case study, a content analysis of existing documents, such as recordings, articles, and reports, was carried out. In addition, interviews were conducted with people who were involved in the projects to gain further insight. The information found was then presented and analysed using the Diffusion of Innovation theory. In a further step, the two cases and other project reports were compared in order to analyse which (organisational) factors may have led to the early adoption by the National Library of Sweden and whether additional challenges were met due to the early adoption. This may also provide some insight into the challenges that may arise in the early adoption of innovations in general.
Keywords:	Linked Open Data, libraries, case study, content analysis, Diffusion of Innovation

Contents

Tables

Figures

1 Introduction

Linked Open Data (LOD) is based on the idea of the Semantic Web, first presented by Berners-Lee et al. (2001). It is based on two key concepts. First, data on the web should be linked in a way that is understandable to both humans and computers, so that both can explore their relationships. Second, data should be open and free, published with a license so that it can be reused by others. This allows information from different sources to be linked and reused, so that data does not need to be repeated.

As most library data is still produced and available in MARC 21 format, it is not fully compatible with the Semantic Web and the Linked Data (LD) principles. Data is (mostly) stored as strings in fields and subfields that are only understood by library software and librarians. In order to open up the library data on the web, it needs to be transformed into a LOD format. This can lead to greater visibility and discoverability outside of libraries. Library data could be integrated into other web resources, which could also lead users back to the library and increase its use. Libraries can also use LOD from other sources and enrich their data to expand the context of their information. Even user-generated data, such as recommendations and summaries, could be included.

As Gaitanou et al. (2022) showed in their literature review on LD in libraries, transformation projects and studies, as well as literature reviews, in the field of libraries, archives, and museums have increased over the last 15 to 20 years. During this time, many challenges have been identified, such as difficulty in choosing between the available tools, vocabularies, and ontologies; a lack of resources and expertise; a steep learning curve; mapping issues; as well as incomplete and erroneous existing data that needs to be cleaned first (Ali & Warraich, 2018; Ullah et al., 2018). This can hinder libraries in the transformation process. It is therefore important to have best practices and step-by-step instructions for the transformation process, as described, for example, by Cole et al. (2013) and Rasmussen Pennington and Cagnazzo (2019).

One library that has already transitioned from a traditional library catalogue using MARC 21 to a LOD catalogue using its own format, which is argued to be aligned with the BIBFRAME 2.0 format, is the National Library of Sweden. The National Library of Sweden is said to be the first national library that has changed its data creation to LOD (Wennerlund & Berggren, 2019). Compared to other libraries, the National Library of Sweden started this project very early, in 2012. It is also interesting that the project took quite a long time, from 2012 until the start of their new cataloguing system in 2018 (Wennerlund & Berggren, 2019). For my master's thesis, I analysed what steps were taken during this project, what challenges were met, and what solutions were used to overcome them. In addition, I analysed the National Library of Finland's LKD and Fennica projects and compared the results of the two cases to analyse which (organisational) factors might have led to the early adoption of LOD, which can be seen as an innovation, by the National Library of Sweden. I further analysed whether the National Library of Sweden met any additional problems due to the early adoption. The thesis is intended to help libraries that

are considering such a transition project or an innovation process in general, as this early adoption may provide an insight into the possible challenges of other early adoptions of innovations or what circumstances may be needed to make early adoptions possible.

2 Problem Formulation

Linked Open Data (LOD) is becoming an increasingly interesting topic in the library community, as it gives libraries the opportunity to open up their data for reuse, which can lead to higher visibility and discoverability (Ullah et al., 2018). It has been shown, for example, in the systematic literature review by Gaitanou et al. (2022), that LOD projects in the library field have increased over the last 15 to 20 years. Several institutions have transformed and published some or all of their data, and some have moved to creating LOD data, but many libraries still create their data in formats such as MARC 21, possibly because their library systems can only work with this format.

Previous projects and research have identified a number of issues and barriers, which will be presented in detail in Chapter 5. However, although the challenges have been identified, how to overcome them has only been described in project reports, and then only rarely. In particular, as far as I am aware, there has been no detailed description of how to move towards creating LOD from scratch. What organisational or other factors might have led some libraries to move towards LOD earlier than others has, to my knowledge, also not been analysed before.

3 Aim and Research Questions

Based on the previously presented lack of in-depth descriptions and information on how to overcome the challenges that arise in real projects (Gaitanou et al., 2022), the aim of this master's thesis is to understand and explain a transformation process of library data in a MARC 21 format into LOD. I will describe the steps taken by the National Library of Sweden to transform the existing library data into LOD, the challenges encountered, and the solutions found to overcome these challenges. I will also present the transition to cataloguing library data in LOD and the transformation back to MARC 21, as many of the libraries that receive their data from LIBRIS XL still work with this format (Wennerlund & Berggren, 2019). This can also give an idea of how LOD and MARC 21 data can co-exist and how MARC 21 can be created from LOD until libraries move away from MARC 21 completely. I will also analyse the more recent projects of the National Library of Finland to see if the different starting points of the projects led to different challenges. In addition, I will compare the two cases to analyse what (organisational) factors might have led to an earlier implementation in the case of the National Library of Sweden.

In order to achieve this, this master's thesis aims to answer the following research questions:

1. What steps have the National Library of Sweden and the National Library of Finland taken to move towards creating LOD?
2. What were the challenges encountered, how were they overcome, and did the National Library of Sweden face additional challenges due to its early adoption?
3. What (organisational) factors may have supported the National Library of Sweden's decision to start the transition to LOD so early?

To answer these questions, I conducted a qualitative content analysis based on existing data and interviews, which will be presented in detail in Chapter 7.

4 Background

Before presenting previously conducted studies and projects, I will give a brief introduction to Linked (Open) Data, the Functional Requirements for Bibliographic Records (FRBR), the Resource Description and Access (RDA), and BIBFRAME.

4.1 Linked (Open) Data

Linked Data (LD) is based on the Semantic Web, which was largely introduced by Tim Berners-Lee. He stated that the Semantic Web is not just about putting data on the web but also about making connections between the data so that both humans and computers can explore the connections (Berners-Lee, 2009).

Berners-Lee (2009, LD section, para. 3) also defined four rules for LD:

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs, so that they can discover more things.

These rules are also the basis for Linked Open Data (LOD), the only difference being that LOD is published under an open license, such as a Creative Commons license. This is done to ensure that the rules for reuse of the data are in place.

Berners-Lee (2009, Is your LOD 5 Star section, para. 3) has also defined a 5-star scheme for L(O)D:

☆	Available on the web (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 (2) plus non-proprietary format (e.g. CSV instead of excel)
☆☆☆☆	All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
☆☆☆☆☆	All the above, plus: Link your data to other people's data to provide context

This 5-star scheme applies to both LD and LOD, except that to get even one star as LOD, the data must be published under an open license.

One of the standards mentioned above that should be used for LOD is the Resource Description Framework (RDF), which is based on the node/edge data model (Mitchell, 2013). This model is also known as a graph model. A graph consists of several nodes that are connected by edges. The edges can either have a direction, in which case the graph is called a *directed graph*, or they can have no direction. With RDF, you can make single statements about a resource or a thing. Several of these statements can be combined to give a description of a resource. Each statement is set up as a triple with a subject, a predicate, and an object. The subject and object are the nodes, and the predicate is the edge, in this case a directed one. The predicate is used to define the relationship between the subject and the object. The subject and the predicate must always be Unique Resource Identifiers (URIs), but the object can also be a literal, for example, for years and page numbers (Wang & Yang, 2018). Multiple triples can therefore be visualised as graphs. To also make the triples understandable to computers, the triples need to be serialised, or, in other words, encoded. This can be done in formats such as XML, JSON, and Turtle. The other standard mentioned above is the SPARQL Protocol and RDF Query Language (SPARQL), which is used to query RDF triples.

As presented above, URIs are a key component of LD and can be used to identify both things and themes on the web. As also explained for the *res* entity in the Library of Congress' Library Reference Model (LRM), this can be anything, "material or physical things and concepts (. …) considered relevant to the bibliographic universe" (Riva et al., 2017, p. 20). Uniform Resource Locators (URLs) are a special type of URIs that can be used to also locate things and concepts on the web.

4.2 FRBR, RDA, and BIBFRAME

The Functional Requirements for Bibliographic Records (FRBR) were based on a study that should present a new framework for what a bibliographic record should contain and was finalised in 1997. The framework describes several groups of entities and their attributes. The first group distinguishes between *work*, *expression*, *manifestation,* and *item.* The second group includes the entities *person* and *corporate body*, those involved in the production of a work. Finally, the third group includes the entities *concepts*, *objects*, *events,* and *places*, which are the subjects of works.

As the distinction in the first group seems to be quite problematic in transformation projects, I will present it in more detail. A *work* is defined as "a distinct intellectual or artistic creation" (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998, p. 16). So, the work does not describe a specific publication, but the intellectual thought behind it. An *expression* is the realisation of the work and can, for example, be in written form, an image, or a performance. So, the original publication of a text would be one expression, and its translation into another language would be a second expression. A *manifestation* is defined as the physical embodiment of an expression. An album released on compact disc by one record studio would be one manifestation, and the release of the same album by the same record studio but on vinyl would be a second manifestation. Lastly, an *item* is a single copy

of a manifestation. So, if a library has several copies of the same book, published in the same year by the same publisher, each of those books would be an item, in this case each with a different barcode for loaning.

Resource Description and Access (RDA) is a set of guidelines for creating data that shall help with the discovery of resources. It is based on the foundations of the Anglo-American Cataloguing Rules (AACR). RDA was designed to be flexible and extensible and has been aligned with the first group of FRBR entities presented above (RDA, 2013). RDA is also aligned with the Functional Requirements for Authority Data (FRAD) and its entities *person*, *family*, *corporate body,* and *place*. The core elements of each group of entities are defined in the framework.

Compared to FRBR, BIBFRAME 2.0, which was developed by the Library of Congress and published in 2016, uses only the following three entities (Library of Congress, 2016):

- A *work* is the conceptual essence of a resource and corresponds to the FRBR entities *work* and *expression*.

- An *instance* is the material embodiment of a work and corresponds to the FRBR entity *manifestation*.

- An *item* is an actual (physical or electronic) copy of an instance and corresponds to the FRBR entity *item*.

BIBFRAME also includes classes such as *agents* (e.g., people, organisations), *subjects* (e.g., topics, places), and *events*. The Library of Congress has also published several tools, such as the bibframe2marc conversion application, to help with the transformation process.

MARC differs from these concepts in that it does not distinguish between the different levels and describes only one resource at a time, with all its information, which would be split up into different levels in FRBR, RDA, and BIBFRAME.

5 Previous Research

Since the emergence of the idea of the Semantic Web and LOD, libraries have undertaken different types of projects to open up their data to the web. Some libraries have transformed and published thesauri and ontologies in LOD, such as the Library of Congress and the National Library of Germany; some have published specific collections or all of their data in LOD, such as the University of Illinois Library; and others have moved to creating their new data in LOD and creating a new framework for managing their metadata, such as the National Library of Sweden.

Most of the published papers on LOD in libraries to date have been based on these projects, mostly from the perspective of those involved in the projects. Some examples are Apenīte and Bojārs (2021), Candela et al. (2018), Cole et al. (2013), Panchyshyn and Park (2015), and Southwick (2015). As these papers are mainly project reports rather than previous research, they will not be included in this chapter. Instead, I will present what has already been analysed, mainly the benefits of LOD and the challenges encountered in projects or experiments.

Over the years, several studies, mainly literature reviews, have analysed what motivates libraries to move towards LOD, in other words, what are the benefits of LOD. One of these benefits is the visibility of library data on the web, which has been presented by Ullah et al. (2018) and Rasmussen Pennington and Cagnazzo (2019). Gaitanou et al. (2022) also mentioned that this increased visibility can lead to better discoverability of the data on the web. As described by Williams (2021), users tend to start their searches in search engines, and if library data appears on the results page, this could lead users back to the libraries, potentially increasing the use of the library.

Another benefit presented by Rasmussen Pennington and Cagnazzo (2019) and Ullah et al. (2018) is the possibility of data enrichment, which could expand the context of the data. Furthermore, library data could also be enriched with user-generated data, such as reviews, which could break down the barrier between curators and consumers, as mentioned by Marden et al. (2013).

Rasmussen Pennington and Cagnazzo (2019) and Ullah et al. (2018) also presented that library data as LOD could be reused by others, which could lead to connecting different resources on the web, leading to the Semantic Web. Regarding reuse, McKenna et al. (2018) also mentioned that others can benefit from using LOD authority files from libraries, as libraries are trusted sources for this type of data, and therefore reusing this data could also increase the credibility of the LOD on other websites or from other institutions.

Ullah et al. (2018) and Alemu (2022) also presented that library data as LOD would improve interoperability with other data, for example, BIBFRAME is said to be compatible with web standards and metadata formats outside the library domain. This is not the case for the most commonly used data in MARC, which is only understood by libraries and library tools (Ullah et al., 2018). Rasmussen Pennington and Cagnazzo (2019) also found that LD can

support multilingualism because entities are identified by URIs, which can include different language versions of a name or concept, rather than strings in a particular language, as also described in Chapter 4.2.

Warraich et al. (2022) noted that LOD is seen as having the potential to provide better services, for example, in search interfaces. Some work has also been done to analyse such LOD portals. For example, Alvite-Díez (2021) carried out an evaluation based on the functionalities and user experience of four LOD portals: the Virtual Polygraph Library Ignacio Larramendi, the Open Data Catalogue of the Miguel de Cervantes Virtual Library, the Bibliographic Data Portal of the National Library of Spain, and the DataBnF of the National Library of France. The author concluded that the visual interfaces minimised the search effort and maximised information discovery. Another study of the functional features in six LD catalogues was conducted by Y.-N. Chen (2023). The author used a checklist of 10 items based on a literature review and described that a *seek or browse* paradigm was present and that contextual browsing based on relationships was possible, as well as refined keyword recommendations. The author also stated that a lack of knowledge of ontologies on the part of the user could lead to problems with SPARQL endpoints.

Another important part of previous research is the presentation of the challenges encountered with regard to LOD. For example, Ullah et al. (2018) identified the need to clean up the library data before transforming the data into LOD. Furthermore, Ali and Warraich (2018) found in their literature review that the existing data, due to incomplete records and data in languages other than English, can lead to technical challenges, for example, when trying to match external sources with the terms in other languages. Warraich et al. (2022) identified in their in-depth interviews that traditional cataloguing standards, such as MARC, caused problems because their structure is not compatible with the relationship structure of LOD. In their exploratory study, Park and Kipp (2019) identified that issues with MARC arise from its fixed fields and machine-readable, rather than human-readable, identifiers. Cole et al. (2013) presented that the MARC fields with their subfields lead to challenges, as the meaning of a subfield with the same name depends on the main field it belongs to and can vary. Lastly, Ali and Warraich (2018) mentioned that the matching or mapping of categories in MARC and other data models is a challenge.

Ullah et al. (2018) found that LOD has a steep learning curve, and Ali and Warraich (2018) and Rasmussen Pennington and Cagnazzo (2019) identified a lack of expertise, lack of knowledge, and lack of technical skills as challenges. Rasmussen Pennington and Cagnazzo (2019) also stated that staff awareness of LOD was low, which was also established in Warraich et al.'s (2022) 18 in-depth interviews with senior library and information science professionals from Pakistan. In contrast, McKenna et al. (2018) observed a high level of knowledge about LD in their online questionnaire, which was answered by 185 information professionals from libraries, archives, and museums. However, the authors mentioned that this could be due to the fact that people interested in LD responded to the survey. Despite this awareness, only 11% of respondents in McKenna et al.'s (2018) survey had actual experience in cataloguing LD.

Ali and Warraich (2018) identified in their systematic literature review that it was challenging for libraries to choose the appropriate LD tool to transform the data, and Rasmussen Pennington and Cagnazzo (2019) found that there was a lack of appropriate tools and infrastructure. McKenna et al. (2018) mentioned that respondents to their questionnaire found the available tools sometimes difficult to use or learn, but respondents also mentioned that the LD tools were moderately useful, but that they needed to be adaptable.

Ali and Warraich (2018) and Rasmussen Pennington and Cagnazzo (2019) also presented that there was a lack of resources and time to start LOD projects in libraries. They also identified that previous projects had found it challenging to choose an ontology, as they either had to find the right one or create their own (Ali & Warraich, 2018).

The lack of collaboration between libraries, LD experts, information retrieval researchers, and users was presented by Ullah et al. (2018) and can also be found in McKenna et al. (2018), who mentioned that the use of different controlled vocabularies and different practices between institutions can be a barrier for LOD projects. Similarly, Park and Kipp (2019) concluded in their exploratory study that libraries often use a mix of schemas and that the use of different schemas slows down the migration from MARC to LD and leads to interoperability issues, for example when libraries use different elements with different names and purposes of different schemas and ontologies to describe the same data. McKenna et al. (2018) also identified that the mapping between different metadata formats and LD models is experienced as a challenge.

Some studies have also experimented with transformations, such as that of Park and Kipp (2019). In their exploratory study, they focused on schemas, ontologies, and data models in 14 libraries. The researchers transformed sample data from MARC 21 records into LD, where available, using the official mappings of the libraries. The authors mentioned that schema maintenance is crucial, as some of the ontologies in their experiments were not maintained and had disappeared. This may also be something that is not easily recognised if the data is only transformed once, whereas the data is generally still created in MARC 21.

Niu (2018) conducted a content analysis of 76 publications on LD projects in 53 libraries, based on the Diffusion of Innovation theory. The study showed that the adoption of the LD innovation was a decentralised process in libraries, originating from different lead adopters, and that it involved a high degree of reinvention, for example, by adapting BIBFRAME to the needs of the libraries. Niu (2019) conducted a similar study on the adoption of research data management services in 44 libraries, which showed that early adopters had to do a lot of research and created tools, templates, and training materials from which later adopters could benefit. Niu (2020) then compared the two studies to identify patterns in the diffusion and adoption of innovations in libraries, as this will need to happen on a regular basis. Niu presented that both diffusion processes were decentralised with a high degree of reinvention. The author also presented differences, such as the driving factors for adoption – standardisation for LD and data management policies for research data management services. Niu concluded that collaboration and multi-disciplinary groups are important to

share knowledge about innovations, that early adopters face high costs and a steep learning curve, and that different types of innovations can pose different challenges.

The Diffusion of Innovation theory has been used in several research papers in the field of library and information studies to analyse the adoption of innovations. The theory has, for example, been used to study the adoption of library automation software by Moloney (1997) for the Kogarah Municipal Library, by Ponelis and Adoma (2018) for the use of open source systems in Ugandan university libraries, and by Olson (2010) to analyse the decision-making process for library management systems. Neo and Calvert (2012) analysed the diffusion and adoption of Facebook in nine public libraries in New Zealand, including motivational factors and barriers to adoption. Raynard (2017) conducted a study to analyse factors and barriers to e-book use in academic libraries in order to develop marketing strategies to promote the adoption of e-book reading. Furthermore, Lund et al. (2020) analysed the perceptions of artificial intelligence by academic library staff based on the categories of adopters defined in the Diffusion of Innovation theory.

As can be seen, most of the previous research to date has been literature reviews and surveys of challenges and benefits. Some research has been done on user experience in catalogues and some on data transformations outside of actual projects. Most of the papers found were project descriptions with an insider's view, some short and some very detailed, such as that by Southwick (2015). As Rasmussen Pennington and Cagnazzo (2019) pointed out, such detailed descriptions of implementation projects can help other libraries starting their own projects.

6 Theory: Diffusion of Innovation

New ways of doing things can be seen as innovations. So LOD was an innovation and may still be seen as such by many in the library field. As presented before, not many libraries have moved on to creating LOD, and many have only done one-time experiments. The question arises, why the National Library of Sweden made this change so early and what (organisational) factors might have helped. To answer these questions, we can look at the Diffusion of Innovation theory. Rogers (2003, p. 5) defined the diffusion of innovation as ”the process in which an innovation is communicated through certain channels over time among the members of a social system.” An innovation in this sense does not have to be something really new, it just has to be perceived as new by one person or a group of people.

The innovation decision can be divided into three types. The optional decision is the one made by an individual, although the decision can be influenced by others. The collective decision is made by consensus in the group. Lastly, the authority decision is made by a small number of people who have power, status, or expertise. The last one is most often found in organisations. Relevant activities leading up to the decision can also take place before the diffusion process begins, for example, when special funding is given to a department that will make it easier to implement the innovation. (Rogers, 2003)

Rogers (2003) defined five stages in the innovation decision process. In the knowledge phase, an individual gains an understanding of the innovation. Knowledge may be awareness of the existence of the innovation, knowledge of how to use it, or knowledge of the underlying principles of the innovation. In the persuasion phase, individuals form an attitude towards the innovation, which may be positive or negative. In this phase, individuals seek information about the innovation. Sometimes the innovation may also be preventive, so one that is not necessary because of an actual need but to avoid an undesirable event in the future. In the decision phase, an individual engages in activities that lead to a decision about whether or not to implement the innovation. The implementation phase is when the innovation is put into use. This phase may also include a re-invention of the innovation, which means that the innovation is modified to meet the needs of the user. And in the confirmation phase, it is evaluated whether the decision was the right one and whether the innovation should continue to be used.

Rogers (2003) identified five attributes that mainly lead to a faster or slower adoption. The first is the relative advantage, which is the degree to which it is perceived to be better than the current option. In previous studies, this has been found to be one of the strongest variables for adoption. The second is compatibility with existing values and needs. The more compatible it is with existing values, the less uncertain it is. The third variable is the complexity of the innovation, so how difficult it is to use or implement. The fourth is trialability, which is the degree to which you can experiment with it with

limited resources. And the last is the observability, which is how visible the results are.

While studying at Iowa State University, Rogers (2003) also defined five groups of adopters: innovators, early adopters, early majority, late majority, and laggards. Innovators are venturesome, have large networks, and are the first to adopt innovations. Previous research has identified several factors that can help with adopting innovations earlier, such as significant financial resources and technical knowledge. Innovators also need to be able to cope with a high degree of uncertainty about the innovation and must be willing to live with setbacks. They are also often gatekeepers, helping to spread the innovation. Early adopters are often more integrated into their local systems and are often seen as role models in their community. The early majority are those who adopt the innovation before the majority and may consult for quite some time before adopting it. The late majority adopts after the average, often out of economic necessity or peer pressure. Lastly, the laggards are very traditional, they look rather to the past and often have fewer resources, so they will also try to be certain about an innovation before adopting it.

The diffusion of innovation was initially studied only for individuals. However, since the 1980s, there has been a growing interest in the innovation process within organisations. This process can be more complex because it involves more people, some of whom are for the innovation and some of whom are against it. Previous research has found that larger organisations seem to be more innovative, as do organisations with a less formalised structure. Previous research has also identified that concentration on a few individuals for the decision making has a negative effect on innovativeness. Interconnectedness between individuals in an organisation and the flow of information showed a positive influence on innovativeness, as did uncommitted resources. (Rogers, 2003)

Rogers (2003) divided the innovation process in organisations into five stages, the first two being part of the initiation and the last three being about the implementation. The first stage is the agenda-setting, where organisational problems are defined and/or a need for an innovation is perceived. At this stage, the needs and problems are prioritised, and the search for a useful solution begins. Many organisations also use opportunistic surveillance to look for innovations that could benefit the organisation. Therefore, sometimes the innovation process is started because of knowledge about it. In the matching phase, problems are matched with the innovation, and benefits and possible problems are described. In the redefining or restructuring phase, innovations are re-invented to fit the needs of the organisation. Innovations can lead to uncertainty in organisations, especially when they represent a paradigm shift that leads to a new way of carrying out a task. In the clarifying phase, innovations are put into practice. And in the routinizing phase, the innovations become part of the regular work and are no longer perceived as innovations.

7 Methods

In this chapter, I will present the methods I have used for data collection and data analysis. I will also present the cases that form the basis of this master's thesis.

7.1 Case Study

A qualitative case study approach was used for this master's thesis. This means that specific cases were investigated using different methods to provide an example of how libraries implement and adopt LOD, how existing data can be transformed into LOD, and what challenges were faced and how they were overcome in specific projects. This may not be generalizable to every type of library, but it may provide a useful example of how it can be done and what needs to be considered before starting an innovation project.

The first case was based on the National Library of Sweden and their cataloguing tool, Libris XL, in which new data is created as LOD. As presented in Chapter 1, the National Library of Sweden was one of the first libraries to start its LOD project in 2012, making this case an interesting example of an innovator. The National Library of Sweden did not just transform their data, they also decided to create their new data in LOD. They started doing this in 2018 and stated that they were the first national library to create all their data as LOD (Wennerlund & Berggren, 2019).

To see if there might be differences depending on the time of implementation, I also analysed the National Library of Finland's Fennica and LKD projects. The LKD project started in 2022 and is still ongoing at the time of writing.

7.1.1 Case: National Library of Sweden/Kungliga Biblioteket

The National Library of Sweden collects and gives access to almost everything that is published in Sweden and foreign material that has a Swedish connection (Kungliga biblioteket, 2024a). They also have the task of promoting the quality of research and collaboration in the public library sector in Sweden. (Kulturdepartementet, 2022)

Libris is Sweden's union catalogue that was initiated in the 1970s and is used by Swedish university libraries, special libraries, public libraries, and the National Library of Sweden. Libris is maintained and developed by the National Library of Sweden. A lot of data is imported from vendors and libraries, edited and enriched in Libris, and finally exported to the local library systems. This means that publications only have to be catalogued once, and libraries only have to add their holdings. Libris also includes Sweden's national bibliography and authority files[1]. The project of moving Libris to LOD began

[1] Authority files are controlled vocabularies for people that include the authoritative (preferred) form of the name, as well as other forms of the name so that, for example, the author of a book can be uniquely identified.

in 2012, and the new version of the catalogue, Libris XL, was launched in 2018. (Wennerlund & Berggren, 2019)

7.1.2 Case: National Library of Finland/Kansalliskirjasto/ Nationalbiblioteket

The National Library of Finland also collects literature that is published in Finland or that is related to Finland. They also ensure that their work and services reflect Finland's diversity and multilingualism. (National Library of Finland, 2020)

The National Library of Finland's strategic choices and development areas for 2021–2030 state that the library will "develop the description of information resources in accordance with the principles of open, linked data." (National Library of Finland, 2020, p. 4). They started their first LOD project in 2016 by converting their national bibliography Fennica into LOD (Suominen, 2023).

In 2022, they started their Linked Library Data (LKD) project with the plan to build their own data model for LOD based on BIBFRAME 2.1.0 and the official RDA to prepare for the transformation of their data into LOD and the future creation of LOD. The project was also originally connected to the platform change of Melinda, the Finnish union catalogue. (National Library of Finland, 2024)

During the interview, I got the information that the system change of the Finnish union catalogue Melinda, in which about 50 libraries are included, will be postponed, at the time of writing, probably until 2028. The interviewees mentioned that the completion of the data model itself is very important, as Axiell's Quria[2] system is already used by one library in Finland and will be implemented in several more in 2024.

7.2 Methods for Data Collection

Several methods of data collection were used for this case study. One of these was the use of existing documents. I searched databases as well as project homepages and Youtube channels for data on the LOD projects. Some texts were also recommended by people I had contacted for interviews. The videos were transcribed with help of the online tool *revoldiv.com* before I proofread them.

For the National Library of Sweden, I analysed 21 recordings of, for example, conferences, seven presentation slides, four articles, five reports, and one press release. For the National Library of Finland, I analysed six recordings, six presentation slides, three articles, and three websites. All the analysed data can be found in Appendix A: Analysed Data from the National Library of Sweden/Kungliga Biblioteket and Appendix B: Analysed Data from the National Library of Finland/Kansalliskirjasto/ Nationalbiblioteket, including the codes I assigned to each document, which I use to refer to the documents in the following chapters. I would like to point out that it is not surprising that

[2] The system is said to rely on FRBR and RDA.

there was less data for the National Library of Finland, as it includes newer projects, so there was less opportunity to attend as many conferences. The National Library of Sweden has also posted the recordings from their *Libris användardag* (Libris user day) on their Youtube channel.

Lastly, I want to mention that it is important to keep in mind that the existing data does not represent objective information, as it was created for a specific purpose and audience, for example, to present the project at a conference for other LOD specialists (Bryman, 2016).

The next step was to interview people who had been involved in the projects to gain some further insight. The interviewees were selected on the basis of people I had found in connection with the projects or who had been recommended to me. As a result, I was not able to get an insight into all the different departments involved in or affected by the project, for example, the cataloguing departments. For the National Library of Sweden, I conducted one 90-minute interview with someone involved in the development of the data model and the editor, and one 30-minute interview with someone involved in the organisation of the project. For the National Library of Finland, I conducted one 60-minute interview with two people involved in the development of the data model. The interviews were conducted via Zoom and recorded so that I could focus on the interview without being distracted by taking notes. Before the interviews were conducted, the interviewees signed a consent form for the recording, as this included personal data such as their voices. The interviews were also transcribed with the help of *revoldiv.com* and proofread by me. I started the interviews by asking the interviewees to talk freely about the projects, focusing on the steps, challenges, and solutions. I then asked follow-up questions about things they had said or questions that were still open. This gave me the opportunity to get answers to my research questions and added information from the interviewees that I had not thought of before. In preparation for the interviews, an interview guide, see Appendix C: Interview Guide, was prepared with a list of questions. This guide included the questions that I wanted answered and also served as a checklist to keep track of whether the questions had been answered by the interviewees during the initial project presentation or if they still needed to be asked.

7.3 Methods for Data Analysis

I conducted a qualitative content analysis to extract information about the reasons that led to the projects; the steps taken to convert the existing data into LOD, to set up a new cataloguing tool, and to transform the LOD created into MARC 21 for use by the libraries. I also looked for information about the challenges encountered and how they were overcome.

The online programme *QCAmap* was used for the content analysis. Documents and transcriptions were uploaded either as .txt or Word documents. The research questions were set up in the programme with the type of *Inductive Category Formation,* as the codes emerged from the data during the analysis, which is typical for the qualitative approach (Zhang & Wildemuth, 2017). In addition, the unit of analysis is typically a theme for this type of analysis (Zhang & Wildemuth, 2017). As this could not be selected in the programme, I

used *Single word,* as one word could be an indicator for a theme. Lastly, I defined the C*ontext Unit*, which is the largest text component that can be in one category (Mayring, 2022), in this case a paragraph.

I then started to code all the material. The themes emerged during this process based on my interpretations. As I had also identified some possible themes during the literature review and some based on the Diffusion of Innovation theory, I kept these in mind during the coding process, such as the challenges or factors for early adoptions of innovations. In the best case scenario, several coders would have done this before using the *Coder Agreements* feature in *QCAmap* to achieve more objective themes. This might also have resulted in more passages being mapped, as another person might have identified different challenges in the data than me. Unfortunately, since I have written this thesis alone, this was not the case.

After all texts were coded, the codes were validated to ensure that the coding was consistent (Zhang & Wildemuth, 2017). I then downloaded spreadsheets for each research question for further analysis. I then summarised and analysed the reasons, steps, challenges, and solutions for each case to provide an overview of the transition process. In the next step, I compared the results of the National Library of Sweden with the results of the National Library of Finland, as well as some other projects and previous research, to analyse whether the National Library of Sweden met additional challenges as an innovator and what factors might have led to the early adoption.

8 Findings, Analysis, and Interpretation

In this chapter, I will present and interpret the findings of the content analysis of the data from Appendix A: Analysed Data from the National Library of Sweden/Kungliga Biblioteket and Appendix B: Analysed Data from the National Library of Finland/Kansalliskirjasto/ Nationalbiblioteket. First, I will present the themes that emerged from the qualitative content analysis. The extracted information will then be presented in a summarised form, first for the National Library of Sweden and then for the National Library of Finland. During the presentation, I will also analyse and interpret the data according to the Diffusion of Innovation theory.

8.1 Themes

I analysed all the texts and transcribed recordings, as well as all the transcribed interviews with *QCAmap*, as described in *Chapter 7.3.* This subchapter presents the themes that emerged from the data based on the four questions I had set up in the programme.

The following themes emerged for the question: *What led to the project?*[3]. One theme was about the *constraints* of the used library system. Another theme had to do with the *existing data* and the problems caused by, for example, changing cataloguing rules. I also found themes that had been identified in previous research, such as *discoverability*, *visibility*, *(re-)usability* and *interoperability*. I also found data on how LOD could help with *multilingualism* and how it could be used to enrich *(enrichment)* the library data with data from other sources. The analysis also included data on *change*s in the library community and the web in general. Lastly, *organisational* factors, such as policies, were identified.

For the question, *What steps were taken?,* I found the theme of *research and collaboration,* which includes steps taken before creating their own data models as well as collaboration with others. The theme *requirements and tests* includes information about the necessary requirements the libraries had for the data model or the editor, as well as early tests. Information on how the team was set up and who was involved in the projects is part of the theme *team*. The *data clean up* theme includes any information about how or when existing data will be cleaned up during or after the projects. Another theme is *data model,* which covers all the steps that have been taken to set up their data models. The theme *vocabularies* contains information about which vocabularies are (supposed to be) used in the LOD data. The *mapping* theme contains all the information about how the MARC 21 data has been mapped to the libraries' LOD data models. Another theme is *transformation,* which includes information about how the transformation has been carried out or is planned to be carried out in the future. All information about how works have been extracted from the data, or how this is planned, is included in the theme *work extraction*. I also found three subthemes for the *Libris XL* theme, which is the Swedish union catalogue. One subtheme is about the *research* that went into creating the editor. The second subtheme is called *editor* and is about how the

[3] The themes are shown in italics.

editor was developed, and the third is *launch and further development*, which is about everything related to the launch of the editor and the development since then. Information about *open source* availability was another theme. Data on how the LOD can be accessed and how it is licensed is part of the theme *license and access*. Data on how information about the project has been communicated inside and outside the libraries is included in the theme *communication*. The *future steps* theme includes all data that presents possible future or already planned developments, and the theme *checklist* includes all steps presented at a conference for such developments.

For the question, *What challenges were encountered?,* the theme *existing data* includes all information about erroneous or missing data that led to challenges. The *MARC* theme includes all the challenges that arose due to the MARC data format. The *data model and mapping* theme includes all information found about challenges related to the data model, either because of the National Library of Finland's specific use or because of the non-availability of a library specific data model at the beginning of the National Library of Sweden's project. Challenges in linking data are part of the theme *links*. The *legal issues* theme includes information on what needs to be considered when publishing the data under an open license. In addition, the *transformation* theme covers all issues related to the actual transformation of the data, and the *work extraction* theme covers all the issues that have arisen when trying to extract works from the data. The theme *Libris X*L includes all issues that have arisen during the development of the editor. And challenges due to the new way of *cataloguing* are another theme. Problems that arose, for example, as a result of organisational changes were included in the theme *organisational*. And lastly, the theme of *communication and* p*ersonal challenges* includes all the data on how people were affected by the change and the challenges that arose because of the way the projects were communicated.

Finally, for the question *What solutions were found?*, I found the following themes that presented solutions to the challenges related to *Libris XL*, *existing data*, *data model*, *personal issues,* and *work extraction*.

8.2 National Library of Sweden/Kungliga Biblioteket

In this chapter, I will present and analyse the findings of the content analysis for the project of the National Library of Sweden based on the data that can be found in Appendix A: Analysed Data from the National Library of Sweden/Kungliga Biblioteket. I will start with the reasons for starting the project before presenting the steps taken in the project, the challenges encountered, and their solutions.

8.2.1 What led to the project?

Several of the reasons presented in Chapter 5 were also found in the analysed data from the National Library of Sweden, for example, it was presented at a conference that library data is not understandable outside the library community and that it is therefore necessary to describe the data in a more widely understandable way, such as LD (B2). Three documents also included

information about the benefits of visibility and (re-)usability and that the reuse could generate more interest in the work of libraries (B10, B13, C4). One interviewee mentioned that it would be possible to use the created LOD in search engines' knowledge graphs to show where the book can be borrowed or that data on Swedish authors could be used in Wikidata (A1). I also found an interesting previous experiment with a Google search for *The Great Gatsby* that was presented at a conference (B10). The presenters showed that in order to find a link that had anything to do with a library, they had to go to page 50 of the results. They mentioned that this was partly due to the fact that there had been a movie with the same name, but it still showed that library data was quite invisible on the web (B10). The presenters mentioned that there is a lot of library data available and that it would be great to see which library has the book you are looking for and how many copies of it there are (B10).

In addition, several documents included information about the possibility of data enrichment, as it would be possible to connect data from different libraries and other sources (B13, C7, D1, E2, E5). It was also mentioned in C7 that "to actually reap the benefits of linked data (…) we must also use it as an integral part of our systems, not only expose it", which is a strong reason to start the development of the new editor. This would also affect the systems, as the difference between internal and external data sets would disappear and the systems would have to work with protocols that could be used with external sources (C7). All the factors presented so far can be seen as relative advantage of the innovation that make LOD a better option than the currently used MARC format.

The data also showed that changing cataloguing rules and cataloguing practices have resulted in inconsistent data and that changing the data model could improve data management (D1, D4). The resources that need to be catalogued are also becoming more diverse, for example, with e-resources, and MARC is not well suited for these types of resources (D4). The previously used system architecture, Voyager, was also quite old and was not designed to handle the large amounts of data required when working with new data types, which also led to performance problems (B7). This again highlighted the need for a new editor.

One interviewee mentioned that they wanted to make the new cataloguing interface more intuitive, more flexible, more accessible, and easier to use (A1). The interviewee also mentioned that it was not clear at the beginning that LD should be used in the new system, but that it became clear early on that what they wanted could not be done with the MARC format, which is primarily machine-readable (A1, A2). The decision to build the system in-house was partly made because of the risk of a lock-in effect to vendor infrastructure, as the data is often tightly linked to the vendor's software (B13). In addition, one presenter mentioned that when vendor software is used, it becomes increasingly difficult to adapt it to the local needs of the libraries (E20). When asked at a conference why the National Library of Sweden had not asked Exlibris, the vendor of the Voyager system, if they could adapt the software, the presenter replied that they felt "most vendors to be fairly reactive and [that they] don't take so many risks because they don't have a real economical incentive right now" (B5). One interviewee also mentioned that at the time

they decided to change the system, there was no system available for purchase that could work with LOD, and that they felt that they could develop it themselves as they had already developed import and export routines and the search interface (A1). This also shows that they had fulfilled a factor that is said to help with early adoptions, namely knowledge.

Another reason for the project seems to have been changes in the community itself (A1, B5, B13). For example, it was mentioned that at the time the Library of Congress had experimented with publishing its subject headings and that in general more and more vocabularies were being published in LOD, such as Friend of a Friend (FOAF) and Dublin Core (DC) (A1). LOD and the Semantic Web were becoming more important in general development, and the National Library of Sweden wanted their data to speak the same language as the data from other web services (D4). A presenter at a conference also mentioned that "things will not work the same, and libraries need to understand and take charge of how things will work" (B5).

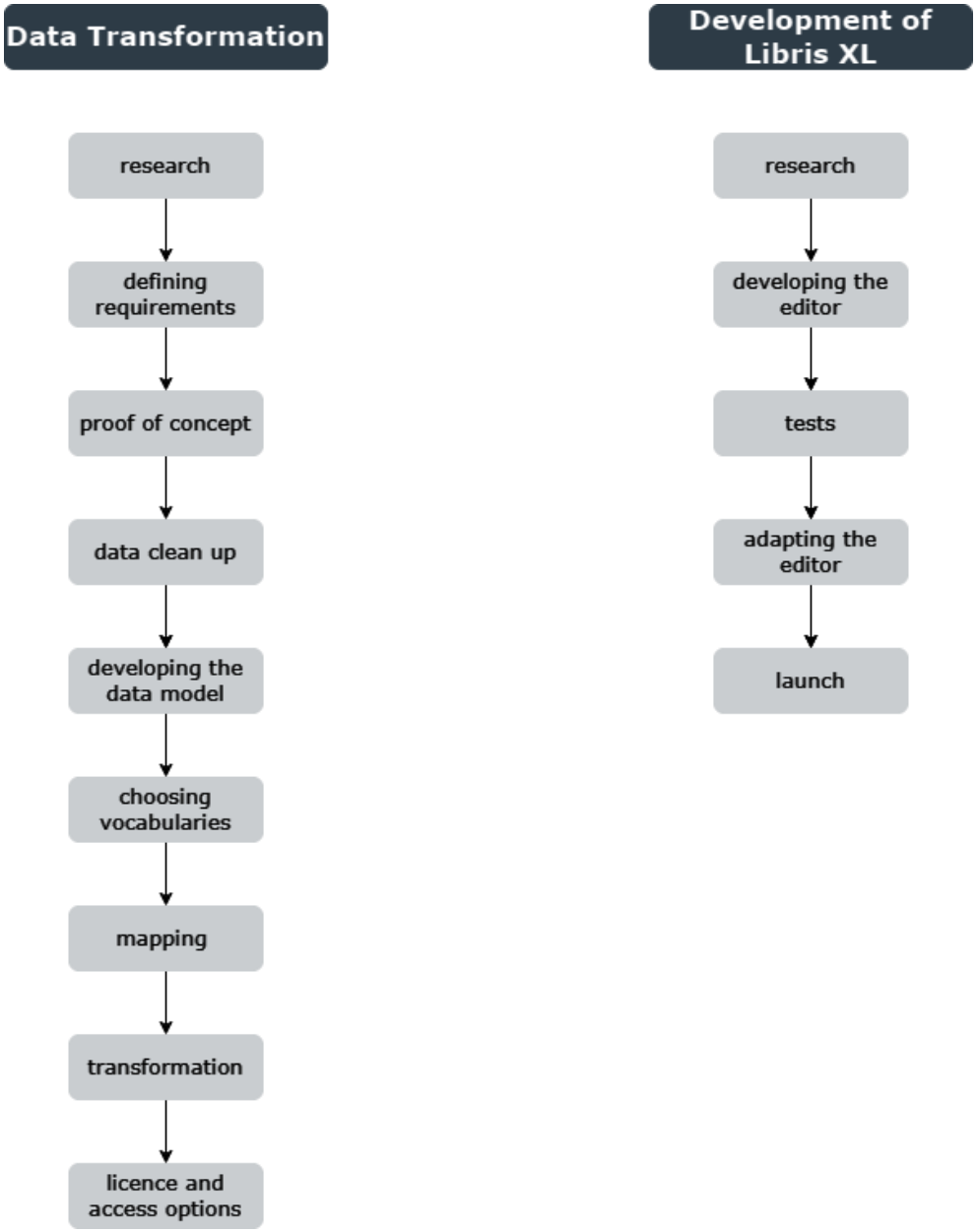
As Libris had originally been independent and only later became part of the National Library of Sweden and was still an independent department at the start of the development, they were able to make the decision to change the system and move to LOD largely on their own (A1). In the Diffusion of Innovation theory, uncommitted resources have been shown to have a positive effect on innovativeness. In the case of the National Library of Sweden, the resources do not seem to have been uncommitted, but they were committed to the task of developing the union catalogue, so for a part of the project team, this was part of their work. Another factor that seems to have helped was that the Libris department had previously worked on LOD projects, for example, by transforming a subset of MARC data and publishing it as LOD to show that library data could be transformed into LOD (A1). This can again be seen as fulfilling one of the factors presented by Rogers (2003) for innovators, as the National Library of Sweden already had technical knowledge of what they wanted to do. It also shows that the trialability of the innovation could be high, as this has also been presented in previous research.

In general, it can be said that the decision to go ahead with this project was an authority one, as a small number of people, in this case with expertise, were able to make the decision. I would also argue that the decision was based on opportunistic surveillance, as staff had already looked at LOD before the need for implementation and a problem – the need for a new editor – had arisen. In addition, this would have been part of the agenda-setting stage based on the Diffusion of Innovation theory.

8.2.2 Steps taken

In this chapter, I will present all the steps taken by the National Library of Sweden in their LOD project. An overview of the steps can also be found in *Figure 1*.

Figure 1. Flowchart of the steps taken by the National Library of Sweden



8.2.2.1 Team

Before going into detail about the steps that were taken, I want to present some information about who was involved in the project team. At the start of the project, the team consisted of about 10 developers and project managers, but it was mentioned that the number fluctuated as, for example, some people changed jobs (A1). The minimum number of people involved at any one time was three to four (A1). One interviewee also stated that the developers were heavily involved throughout the project and were mostly, or at least until the re-organisation, involved in the project full-time, while the rest of the project team still had to deal with the day-to-day work alongside the project (A1).

Cataloguers and metadata specialists from the National Library of Sweden were also involved in the process (A1, A2, B18). As the project progressed, cataloguers from outside the National Library of Sweden became more

involved, for example, in the development of profiles for resources such as sheet music, where their expertise was needed (A2). Other people involved were UX designers and communication and support staff (C5). One interviewee also mentioned that there was a "very broad experience of people working, knowing the Marc format very, very well. And also, the developers who were (…) super into the semantic web and how to use different formats and what kind of formats could talk to each other" (A2).

8.2.2.2 Research and Collaboration

The first research for the project was carried out in 2009, before the project itself started in 2012, at the same time as the development of BIBFRAME had started (B7). As the National Library of Sweden was one of the first to do this, there was little to no previous experience to draw on (C2). There were projects that were doing things like converting data, such as the National Library of Germany, but none that were preparing to create the data in LOD at the time (B7, B8). The team followed the progress of the Library of Congress and the projects of the LD4 community[4] (A1, B7, B8). They also looked at systems on the market but found that there was no system at the time that could meet all their requirements (A1, B7).

During and after the project, the National Library of Sweden exchanged thoughts and experiences about the innovation at various conferences, such as the ELAG (European Library Automation Group)[5], the SWIB (Semantic Web in Libraries) conference[6] and the BIBFRAME conference in Europe (A1). This also shows that the National Library of Sweden has spread the innovation and can be seen as a gatekeeper, which is supposed to be a characteristic of innovators.

8.2.2.3 Defining Requirements, Data Analysis, and Proof of Concept

It was also important to define the requirements for the new system. It was, for example, found that the system would still need to support MARC 21 for many years to come, as the local library systems are still dependent on MARC, so in order to work with the data they need to get it in the MARC format, which also applies to the National Library of Sweden's search engine and import routines from vendors (A1, B20, C2, D4). The National Library of Sweden also needed to build a system that could work with other authorities' systems that would receive data from the library, such as the statistiska centralbyrån (B7). As mentioned above, the National Library of Sweden also wanted an interface that was simple and intuitive and that could work with different and new types of media, such as digital and electronic media, moving images, and sound (A1, A2, B5). Lastly, they wanted to build a system that was based on web standards and open source so that others could use it to build their own applications (B8, D4).

[4] The LD4 community aims to advance library and archival practices and focuses on linking data on the Web. They received grants from the Mellon Foundation and conducted different projects on, for example, how to implement LOD in libraries.
[5] This conference focused on LOD, among other topics.
[6] This conference focuses entirely on the Semantic Web.

The National Library of Sweden had also analysed how much MARC 21 was used in their system and how much it was being adapted to their local needs (A1). In 2012, at the beginning of the project, the National Library of Sweden carried out a proof of concept where they transformed some MARC 21 data into LD (A1). As one interviewee noted, this was quite far away from the later transitions, partly because there were no examples from BIBFRAME yet (A1). This part of the project can be seen as part of the matching phase, as the National Library of Sweden had to figure out how best to go forward with the editor.

8.2.2.4 Data Clean-Up

Prior to the transformation, the National Library of Sweden did some cleaning up of the MARC data (A1), but it was not clear from the analysed data what exactly had been done. It was also mentioned in one interview that they were still working on normalising the data and that at the time of writing they were in the process of normalising electronic resources, as the information that it is an e-resource is entered into the data in four to five different ways, and they are trying to harmonise this (A1). In general, it got clear from the interview that they felt that it would have been good to have had more time to go through the data and analyse it with their future use cases in mind (A1). One interviewee also mentioned that it would have been better "to have a few more years to (…) normalise the data before [they] went live" (A1), but also said that "you can keep doing that forever (…) [and] we really just have to take the plunge" (A1).

8.2.2.5 Data Model

An important decision for the move to LOD was the data model. As mentioned before, BIBFRAME had not been published when the National Library of Sweden started the project, so the National Library of Sweden had to work on their own data model that would fit their library data (A1, B21). The National Library of Sweden did what had been done in previous LOD projects and tests, they looked at the available vocabularies and came to the conclusion that they needed to define terms and have an in-house vocabulary[7] in Swedish as all the available ones were in English, and they needed the labels and definitions in Swedish for cataloguing (A1). As they adapted the available models to their needs, this can be seen as part of the redefining phase.

When the first version of BIBFRAME was published in 2014, the National Library of Sweden reviewed it, but realised that it was still very much based on MARC (A1). The project team also felt that LOD was not implemented well enough for their purposes and decided to continue working on their own model (A1). When BIBFRAME 2.0 came out in 2016, they reviewed it again and found that their own model was not far off and aligned it with BIBFRAME 2.0, feeling that it could be a community model and might become the dominant vocabulary for library data (A1, B12, B21, D4, E4). In general, they used many of the BIBFRAME properties and classes, but also included terms from schema.org, SKOS[8] and RDA (B2).

[7] They called this vocabulary *KBV application vocabulary.*
[8] Simple Knowledge Organisation System

The analysed data also showed that the National Library of Sweden decided to use RDF, which is presented in more detail in Chapter 4.1, because they considered it to be the most flexible description model (B5, B11).

8.2.2.6 Vocabularies

When using LOD, it is important to define which vocabularies will be used in the data. It could be found in the data that the National Library of Sweden uses the National Library of Sweden's base vocabulary, which can be found at *https://id.kb.se/vocab/* (A1, C5). The vocabulary includes terms in Swedish and English and has been aligned with other common vocabularies such as Dublin Core (DC) and schema.org (A1). SKOS is used for the subject headings and concepts, and schema.org for persons and organisations (A1, B21). They used the OWL equivalent property and sub-property classes and mapped them to BIBFRAME (A1, B21). They also kept track of how these terms fit with other vocabularies to improve future interoperability (B21). They also considered what other vocabularies they could use, for example, for geographical data, as it makes sense to use a vocabulary created by experts and to use the same vocabulary as others (B5, B7). The analysed data showed that it is important to have a way of knowing when the used data has changed so that you can react to the change (B20), but no way of doing this was described in the analysed data.

During the analysis, I found out that the transformation does not necessarily mean that there will already be links for everything. When asked during the interviews what was already linked, the answer was people and the Swedish subject heading (A1). For places, they use country code links from the LC[9], which they have translated into Swedish and geographical area codes, but they have not yet included a more detailed way of linking places (A1). One interviewee mentioned that they had discussed different sources for geographic headings, such as using GeoNames, but had realised that this would not work, because of GeoNames' license which would make it impossible for the National Library of Sweden to use a CC0 license for their data (A1).

They also considered Wikidata, but questioned whether it could be trusted as anyone could edit it (A1). They also considered using their own instance of Wikidata but decided that it "would kind of defeat the whole purpose of decentralising" (A1). They had previously done a test with Wikidata, linking to agents using a triangulation of the name and dates (birth and death, where applicable), in which they were able to link 50 000 of their 200 000 agents in Libris but found that some famous ones were not linked, for example, because of white spaces in the name (A1). Something similar with links to DBpedia was presented at a conference, where information from DBpedia, such as an image and some more information about the author, was integrated into a results page (B7). Regarding the unique identifiers for the works themselves, it was mentioned at a conference that they did not have a problem with this as Sweden has a central union catalogue and all other libraries use this data, so they could just use these identifiers (B20).

[9] Library of Congress

8.2.2.7 Mapping

An important step in transforming MARC data into LOD is to set up the mapping between the MARC fields and the entities in the LOD model. The National Library of Sweden has therefore mapped MARC to RDF, which can be found at *https://id.kb.se/marcframe/* (C4). The analysed data showed that the mapping took quite some time, partly due to the challenges posed by the MARC structure (A1, B20), which will be presented in more detail in Chapter 8.2.3.2. A further complication was that the LOD had to be transformed back into MARC 21 (A1), which will be presented in more detail in Chapter 8.2.3.5.

When asked if some information from MARC 21 was not included in the transformation from LOD, one interviewee replied that some things that were not included were already on their way to becoming obsolete or were transported through other fields (A1). But the interviewee also mentioned that there is little information in MARC that is not useful, and that this is partly due to what is needed in the Libris web search (A1). The interviewee also mentioned that as long as this web search is used, the National Library of Sweden will be careful not to simplify the MARC data (A1).

8.2.2.8 Transformation

Technically, the National Library of Sweden decided to use a PostgreSQL database[10] and an Elasticsearch[11] index (A1, B21). They also decided to use JSON-LD for serialisation, which was becoming popular at the time (A1, B20, D4). In addition, someone on the development team had been working with the W3C standard and already had some knowledge of JSON-LD (A1), which again shows the existing knowledge that is helpful when deciding to implement an innovation. The actual transformation was executed using a configuration file in JSON that structures all the fields (A1). The logic of the configuration file was written in Java and includes some pre- and post-processing steps (A1). One interviewee also mentioned that they needed to include a MARC vocabulary layer in the transformation, the so-called Plan B, for things they were not sure they wanted or needed but did not want to throw away because they might still need them (A1). The full transformation can be found on GitHub, with the main transformation in the *marcframe.json*, the *marcframe-hold.json* for the holdings, and the *marcframe-auth.json* for the inclusion of agents and subject headings (A1). All existing data was transformed in 2018 during the editor change (D4).

8.2.2.9 Work Extraction

In the first transformation, the National Library of Sweden did not differentiate between BIBFRAME works and instances because from the MARC records, they would have created as many works as instances, resulting in twice as many entities (A1). At the time of writing, the National Library of Sweden has already done some work on extracting works from the data. To make this easier, they have done some data clean up, so that they could normalise and deduplicate the data (A1). They made simple queries using the main title

[10] PostgreSQL is an open source relational database management system.
[11] Elasticsearch is a search engine developed in Java and based on the open source Apache Lucene search engine software library.

(kbv:hasTitle[12] and bf:mainTitle[13]) and primary contributor (bf:contribution) (C6). Those with the same main title and author were then merged into clusters before some similar clusters were further merged (B2). The next step was for the National Library of Sweden to check that the instances in the clusters were actually the same (B2). In a second attempt, they used all titles (bf:mainTitle, bf:subtitle, bf:hasPart, bf:partNumber, bf:partName, marc:parallelTitle, marc:equalTitle) and removed punctuation to achieve larger clusters because they had to match exactly (B2). They also ignored generic subtitles, such as *novel* (B2). Again, there were some challenges due to missing data, which will be explained in Chapter 8.2.3.6. They decided not to include subject headings in the clusters, as each instance probably would not have gotten exactly the same subject headings (B2). The data could then also be normalised using information from other instances in the same cluster (B2).

At the end of the project in 2020, this was still only a proof of concept (A1). However, by December 2023, the National Library of Sweden had completed a first round of work extractions with a subset of Swedish modern fiction, excluding series and collections. They extracted 50 106 works, which were then linked to 135 000 instances (A1). The work with the most instances had 65 (B2). This made it possible to reduce the duplication of data and reuse the created works as links (A1). It should also be mentioned that workshops were held with other Nordic national libraries to discuss the possibilities of extracting works in 2019 (B19).

8.2.2.10 License and Access Possibilities

The LOD is licensed under the Creative Commons CC0 license, as are the National Bibliography and the Authority Files (A1, D2). The only exception is MARC field 664, which is a non-public note that is already filtered out of the data (D2).

The National Library of Sweden's LOD can be accessed in several ways. One is the read and write API that uses the JSON-LD via the Elasticsearch index, which will also be the basis for the next web search and can also be used by others to build their own services (A1, B20). Another option is the SPARQL endpoint, which can be found at *https://libris.kb.se/sparql* and is also used by the National Library of Sweden for their own analysis (A1).

8.2.2.11 Libris XL

After presenting the development of the data model and the transformation, as well as the access possibilities, I will now present and analyse the development of the editor.

8.2.2.11.1 Research and Tests

For the editor, the UX designers conducted contextual interviews and observations with cataloguers from the National Library of Sweden and other libraries at their workplaces from 2009 onwards (B6). It turned out that being

[12] *kbv* stands for the base vocabulary of the National Library of Sweden.
[13] *bf* stands for BIBFRAME.

at their workplaces was helpful, as the UX designers could observe whether the cataloguers kept notes such as post-its nearby or whether they sometimes asked their colleagues for something (B6). The UX designers used the data to create three personas with different levels of experience and expertise (B6, B7). The first persona was described as someone who catalogues full-time and works with other specialists (B6, B7). The second persona was defined as a part-time cataloguer who also works at the information desk (B6, B7). And the third persona was described as someone in a public library who catalogues only a few hours per week (B6, B7). The observations and feedback were also used to find out what help materials should be produced (C5). A team from the National Library of Sweden created training videos[14], documentation, help texts, and additional support was offered (B14, C5, D4).

Several reference groups were created, for example, for special libraries such as music libraries, which would have different needs from the other libraries (A1, A2, B16). It can also be mentioned that the expertise of the cataloguers was used for the editor design, the mapping, and the transformation process, as not all libraries needed the same fields for their systems (A1, B12). The National Library of Sweden also carried out tests with the other libraries to make sure that the system would work for the libraries' workflows (A1). This would indicate interconnectedness between the different departments and libraries, which is described in the Diffusion of Innovation theory as a helping factor for innovativeness.

In May 2018, a test version of the editor was released for cataloguers to evaluate (A2, B12, B15, D4). There were test sessions, workshops, and demos of the system organised by the National Library of Sweden (A2). A part of the team even went on a tour to present the interface and format and held four workshops in the spring before the launch, open to any library that wished to participate (A2). As a result of these tests, a number of adjustments were made (B5), which will be presented in more detail in Chapter 8.2.3.7.

8.2.2.11.2 Editor

Compared to the previous system, cataloguing in the new Libris XL is more about selecting entities rather than entering information from the source (B20). Information about an entity is also catalogued in the same place rather than in different fields, as in MARC 21, for example, comments about the language are entered next to the language (B5). Links are displayed as tags and can be clicked on (B5). Authorised input is clearly marked with a special icon, and strings are clearly displayed as they look like simple plain text and have a dashed line around them (B5). One change from before is that fields now have labels in natural language instead of codes (B5). It was mentioned that these labels are connected in the background to what they are, so it does not really matter what cataloguing rules are used (B7). The team also created cataloguing profiles for different types of data to be used (A1). The editor also includes help text so that cataloguers do not have to go to another page to check something (B14).

[14] The tutorials can be found on the National Library of Sweden's *YouTube page.*

New for libraries is the ability to make batch exports of the data they need (B12). In addition, data can be exported automatically and continuously, as libraries can set up how often they want to receive the data (B12). Individual records can also be retrieved via the Libris web search (B12). And it happens instantly, not overnight, as before (B12).

8.2.2.11.3 Launch and further development

After long discussions about the stability of the system and the best, or rather, least problematic, time for the libraries to move to the new system, the editor was launched on June 4th, 2018 (A2, B18, D4). The system was down for one week, during which neither the old nor the new editor could be used, but the search engine was still accessible (D4). All functionalities, such as the import and export jobs, were deployed, and the catalogue was indexed for searching in the search engine (D4). The first cataloguing in the new system took place on June 8th, 2018 (D4). The import routines did not work in the beginning but were restarted on June 25th, 2018 (D4).

The launch also started the clarifying phase, in which the innovation is used widely in the organisation, or in this case, in all the libraries that are part of the union catalogue. It is also important to communicate about the innovation so that people know how to work with it and to reduce uncertainty about the innovation. How this was done in the project is described in more detail in Chapter 8.2.2.13.

It could be found in the data that after the launch of the editor, libraries had high expectations for new releases with additional features (D4). A support team was therefore set up to deal with problems and initially met daily to deal with all incoming issues (D4). The project team analysed, documented, categorised, and prioritised the feedback (D4). In the end, not all wishes could be implemented because there were sometimes conflicting ideas, and the team had to work out how to reconcile them (A1).

The team had already been working on new features, templates, and bug fixes and had been releasing new versions of the editor since September 2018, initially on a monthly basis (D4). In 2019, for example, the team worked on a filter for fiction that could be exported with the system and also on expanding the links in the data (B19). In 2020, it was decided that the project should end because the primary goal had been achieved and people could work with the system (A1).

This can be identified as the routinizing phase, in which the innovation becomes part of everyday work and is supposed to no longer be seen as an innovation. On the basis of the analysed data, it cannot be said for sure whether this is actually the case for all the people involved, so this would need to be analysed by interviewing or surveying more people. Rogers (2003) also mentioned that re-inventing the innovation can lead to a sense of making the innovation their own, so it would be possible that further adaptations could lead to greater acceptance of the editor.

8.2.2.12 Open Source

One of the principles of the National Library of Sweden is to work transparently, so they have made their source code for the editor and the transformation available on *GitHub* (A1, B4, B20, E4). This has been done since the beginning of the project, even before the project was finished (B20). It should also be mentioned that the code has already been reviewed and tested by (at least) one library, the National Library of Finland (B7). Again, this can be seen as a way of spreading the innovation, as it may make the implementation easier for other libraries as well as furthering knowledge through discussion.

8.2.2.13 Communication

I would also like to present how the National Library of Sweden communicated their project. The data showed that they used social media and their blog to keep the libraries informed (D4).

When asked what advice the interviewees would give for such projects, one interviewee mentioned that "communication is super important, and trying to be as transparent and (…) [to make] sure that people have the right expectations" (A1), as the innovation probably will not have all possible features in the first release (A1). The other interviewee mentioned that "communication is the key" (A2) and that "you have to communicate with the ones who are involved in any sense (…), those who are involved in the project of course, but also the ones who are more far away from the project, who [are] depending on the outcome of the project" (A2). It was also mentioned that it was assumed that the team had done this well, in particular communicating extensively before and after the launch (A2). The National Library of Sweden even had a sub-project on how to prepare the libraries for the change (A2). It would be interesting to know if people in the other libraries also felt that the communication had been done well, but no information on this could be found in the analysed data.

8.2.2.14 Future Steps

I was also able to find some information in the data about (possible) future steps. One interviewee mentioned that they were working on including more links, for example, for the MARC fields 760 and 770, as well as geographic data (A1). The interviewee also stated that they are interested in mapping the Swedish subject headings to the Library of Congress' subject headings, and that they intend to integrate the archive and the audio-visual database into Libris XL (A1). It was also found that they would like to extract more works, possibly starting with fiction in languages other than Swedish, for example, translations (A1).

One interviewee mentioned that the wider adoption of LOD in libraries is hindered by the small number of systems on the market that can work with LD (A1). The interviewee further explained that the National Library of Sweden had considered making Libris XL an implementation-only product but that they did not have the resources to do so at the time of writing (A1). It could be argued that this could reduce the complexity of implementing LOD, as other libraries would not need the same technical knowledge. This would also

strengthen the National Library of Sweden's position as a gatekeeper for the innovation.

A new web search was under development at the time of writing (A1, B17). The first beta version was released on May 29th, 2024, and can be found at *https://beta.libris-qa.kb.se/* (Kungliga biblioteket, 2024b). One interviewee mentioned that the new web search will initially co-exist with the current web search (A1). It can be noted that the availability of such a new search interface may increase the observability of the results of the innovation, as the advantage of LOD will be visible.

8.2.3 Challenges and solutions

In this chapter, I will present and analyse the challenges that were encountered in the Libris XL project and the solutions that I found in the analysed data.

8.2.3.1 Existing Data

It was found that one of the most difficult parts of the transformation was the existing data (A1). An example of this was missing data, partly due to changes in the used cataloguing rules that, for example, had not required all the fields at the time of creation (A1, D1, D4). One interviewee mentioned that in order to get more out of this data, it would have been necessary to go back to the resources, which libraries generally do not have the human resources to do (A1). Therefore, the National Library of Sweden decided to work with what they had (A1). It was also mentioned in the interview as advice for other libraries that it is very important for the transformation to clean up messy data (A1).

8.2.3.2 MARC

The data showed that there were a number of problems with transforming the data that is based on the MARC format. It was mentioned that a general problem with MARC is that it uses codes, and if you do not know MARC well, these codes will not mean anything to you (A1). So, sometimes people might have added things that were already included in some codes (A1). There were also issues with new data types that did not fit into MARC, so, some fields were added (A1). In general, fields have also been used in different ways over time, making automatic transformation more difficult (A1). One interviewee also mentioned that the data in the union catalogue had not been normalised over time (A1).

The data also showed that MARC is not always clear, as information about entities is often divided into different fields, and it takes extensive knowledge of MARC to understand what the differences are (B6). Another challenge is that MARC fields sometimes contain information about different entities, such as field 245, which can contain information about the title, author, and form, among other things (B6, C4). There was also information about problems arising because a different indicator could mean that the input in a MARC field belonged to a different class in the data model, making transformations more complicated than simply mapping the fields to an entity (A1). Another problem with MARC was repetitions, as these were not always clear and sometimes not properly keyed (A1).

In addition, different libraries needed different types of MARC data in their local systems (A1). So, the National Library of Sweden had to do a lot of analysis, development, and testing, which was quite labour intensive and not originally estimated (A1).

8.2.3.3 Data Model and Mapping

It could be found that the mapping between different data models, for example, between BIBFRAME and schema.org, was complicated because they have different granularities (B21, D4). Another problem that arose during the workflow tests with various libraries was that some libraries used fields that the National Library of Sweden had not intended to include (A1). One interviewee mentioned that it was important for the team to design the editor in such a way that all libraries could continue their work, and therefore they adapted the mapping (A1).

8.2.3.4 Links

There were also some challenges with linking. There was, for example, one issue with ISBN numbers because several ISBNs were found in one record, but only one should be linked, and the question arose as to how to decide which one (A1).

Another problem came up with display notes, which are comments about the link (A1). One interviewee mentioned that if you would just link, you would lose the comments (A1). One possible solution that was presented was that you would say this is the relationship, and this relationship has this note. Another option would be to use annotations from RDF-star, which will be part of the next version of RDF (A1). And one interviewee noted that they intend to use the latter version but that it has not been implemented yet (A1). The interviewee also mentioned that it would not always be necessary to include the note, for example, if it just says what the link points to (A1).

In general, the National Library of Sweden is still working on linking to more things (A1). This includes relations with entities within the system, where they need to figure out what can be linked, and links outside their own system, such as the Library of Congress (B4). The use of data from different sources also led to the question of responsibility, as the library, for example, did not have responsibility for what was written in Wikidata but wondered if it was their responsibility if it was used in their data (A1). This also seems to be the reason why they have not included such links (yet).

8.2.3.5 Transformation back to MARC

It was mentioned in both interviews that the transformation back to MARC 21 was one of the biggest challenges, as this also affected the transformation from MARC 21 to BIBFRAME (A1, A2). Even though they had had an idea of how to model the data in RDF, they were not able to stray too far from MARC because they would not have been able to transform the LOD back to MARC (A1, B2). In general, the National Library of Sweden wanted to try to reconcile data that was fragmented across different MARC fields, such as information about the publication from fields 008, 260, and 264 (A1). However, one interviewee mentioned that it would have been difficult to know what needed

to go back from BIBFRAME to the MARC fields (A1). The data also showed that there was little to no experience to draw upon for the back-transformation into MARC 21, as, for example, the Library of Congress had not been working on anything like this, so the National Library of Sweden had to figure this out for themselves (B14). One interviewee mentioned that this was one of the reasons why the project had taken longer than they had originally anticipated (A1). This can be seen as a challenge that innovators will face, as they are mostly on their own in the development process.

8.2.3.6 Work Extraction

A major challenge that could be identified in the data is the extraction of works. It was mentioned by one presenter that when there is ambiguous information, humans can often easily determine what is what, but that this is not the case for machines, which need specific and precise information (B1). It was also mentioned that machines can generally work with ambiguity, for example, in machine learning, but that they need a lot of raw data to learn patterns from it (B1). One interviewee mentioned that one of the challenges in extracting works is that there is no division of this in the MARC records (A1). So, if you took the information from the MARC records and divided it into instances and works, you would end up with as many instances as works, and you could not link them because you would have a lot of duplicate entities of the same work (A1). When trying to cluster the duplicates, the denormalised data was again a challenge (A1). For example, if one of the MARC records was missing some information, for example, if the illustrator was not included in the 700 field, it would look different, but it could still have been the same work (A1). And that could be cleaned up to some extent, for example, by getting the information from the 245 field (A1). This again led to the importance of data clean up and normalisation, as it also helps with clustering the works (B1).

A question that had also arisen in this context is how to define a work, as there are different views on this, for example, the differences between RDA and BIBFRAME presented in Chapter 4.2 (B4). The National Library of Sweden mentioned that it is therefore sometimes difficult to decide whether something is a new work, especially in the case of children's picture books and comics (B2).

8.2.3.7 Libris XL

There were several challenges experienced with the interface when trying to make it coherent and easy to use (A1). One interviewee mentioned that it is one thing to present data, but another to make it clear what can be edited (A1). Originally, the input fields were only visible as such when clicked on, to support scanning through the data that had already been imported (B5). But this was not intuitive enough and led to confusion about what could be modified in the editor (B5). This was changed to include borders after the National Library of Sweden had received the feedback (B5). Another thing that was changed was the visibility of what information was linked by showing it as tags (B5).

Another issue came up when it was not possible to link to a person (B16). Now there is something called *snippet*, which makes it possible to select that a

person is a local entity (B16). Also, things that are not linked are shown with a dashed line around them, so it is easy to see what is linked and what is not (B7).

There were also problems with the ergonomics of the interface (B16). Cataloguers felt that they had to make too many clicks, especially if they were cataloguing all day (B16, B18, C5). An example of this was adding a contributor, which initially required 15 to 20 clicks (B16). Changes have been made, for example, it is now possible to cut and paste a whole section into another one, for example, if a person is entered as the primary contributor but is not and this needs to be changed (B16).

They had originally used a two-column model for the interface, but this caused problems, especially with long lists (B7). They therefore changed it to a three-column model, where one column is for the label, one is for editing, and one is for interactions, like help texts (B7).

After testing the beta version, several libraries also had concerns about missing functionalities, such as the duplicate handling they had in Voyager (B18, D4). I asked one of the interviewees about this via email and was told that this had been implemented in one of the first releases. The interviewee also mentioned that they had implemented a function to move all holdings from one description to another. After the launch, some libraries also lacked templates and profiles for their resource types (D4). They either found workarounds or paused cataloguing for certain resource types until these templates were available (D4).

8.2.3.8 Cataloguing

As the new Libris XL interface was “a radically different experience from the previous MARC 21 tool“ (D4) and the way in which data was entered was also different from the previous one, this created challenges for the cataloguers (D4). Instead of typing in the information, they had to scroll or search and click on the links (D4). This shows that the implementation of the innovation has led to a paradigm shift in which the task is carried out in a new way. It also shows that LOD is not compatible with existing systems because it changes the way cataloguing is done. This may be one of the reasons why LOD has not been used more widely in libraries.

In general, I could also find information that the impact on cataloguing or the time it took to get back to their previous productivity seemed to have been influenced by the types of resources and/or how many types of resources the library had to catalogue (D4). I also found a statement in the data that the new cataloguing is something that everyone will get used to as they work with it over time (B15), just as everyone got used to cataloguing in MARC.

One interviewee mentioned that “I think also the ones who had prepared in a good way, they also started to work quite fast also in the new setup” (A2). It was also mentioned that the cataloguers at the National Library of Sweden had “prepared very well for the transition. So, they had a plan for the possibility that it would be a catastrophe, and the catastrophe was not there” (A2). It was also noted that the change may have been easier for newer cataloguers because

they were not as used to MARC as very experienced MARC cataloguers (A2). In general, it seems to have been difficult for the cataloguers because they had the same workload but needed more time at first because they were not used to the system (A2). However, it seems that by the end of 2018, the cataloguers at the National Library of Sweden were almost back to their normal productivity (D4).

Unfortunately, I did not have any data from the cataloguers to analyse, so I cannot say what challenges they faced and how satisfied they are with the system. It would also be interesting to see if they are already benefiting from the new system, which would make it possible to observe the results of the implementation of the innovation.

8.2.3.9 Organisational

There were also some issues that could be described as organisational. One of these was the fact that Libris was not originally part of the National Library of Sweden, and when it was integrated, it was still a separate department with developers and people with cataloguing knowledge but no one who was cataloguing (A1, A2). One interviewee mentioned that some people felt that Libris XL was not a product of the National Library of Sweden but of Libris (A2). The interviewee also mentioned that "cataloguers at the National Library who were so heavily involved in the process, they still felt as vulnerable as all the other librarians out in the country " (A2). It was also noted that this did not initially change even after the re-organisation, when Libris staff were integrated into several departments at the National Library of Sweden (A2). This could be seen as a lack of interconnectedness, which could also lead to lower innovativeness.

The re-organisation of the National Library of Sweden also created resource challenges as most people were split up into different departments (A1). This meant that people involved in the project now had to get resources for the project from different heads of department, and suddenly there were fewer people working on the project (A1). One interviewee suggested that if there is a re-organisation in the company during such a project, it should be considered "if the project needs a restart in some way" (A1).

It was also mentioned in an interview that it would be helpful for the project team to have someone who can talk to the stakeholders, as it takes up a lot of the developers' time if they have to talk to the reference groups at the same time as they are supposed to be developing, as this naturally leads to longer development times (A1).

8.2.3.10 Communication and Personal Challenges

It was also mentioned that the project was difficult for some people, especially as it was not clear where they were going (A2). The libraries had heard that something was going to change quite early in the project, but they did not really know what was coming and how it would affect their work for a long time (A2). So it was mentioned that they might have talked about the changes too early (A2).

Another challenge that was mentioned was the psychological impact on people of moving from MARC, which they had worked with for a long time and were very knowledgeable about, to something completely new (A1). People may have felt that something they were good at, something they felt safe with, was being taken away, and they needed to work with something they were not yet good at (A1). But it was also mentioned that experienced cataloguers became more accepting of the new editor over time as they got used to it (B5).

The data also showed that new collaborations had been formed in the Libris network as a result of the development and that the new system had advanced equality between junior and senior cataloguers (D4). Therefore, it seems that interconnectedness improved through the project, which could be useful in future innovation projects, as this is said to have a positive effect on innovativeness.

8.3 National Library of Finland/Kansalliskirjasto/ Nationalbiblioteket

In this chapter, I will present and analyse the findings for the National Library of Finland's LOD projects based on the data that can be found in Appendix B: Analysed Data from the National Library of Finland/Kansalliskirjasto/ Nationalbiblioteket.

8.3.1 What led to the project?

Like the National Library of Sweden, the National Library of Finland had general reasons for moving towards LOD. One of these was the (re-)usability of the data, as, for example, institutions outside the library field would be able to reuse the data (K1, K2). It could also be found in the data that this would make library data more valuable and visible internationally (I5, J1).

It was also found that cataloguers could benefit from existing LOD, for example, by reusing annotations from works in a new edition (I5). Libraries could also enrich their data, for example, by linking to research data (I5). Enrichment in general could also present the data in new contexts, which could also increase the value of the library's data (J1, K1). The data also showed that interoperability between different types of library data and interoperability with other data in the world were reasons for the change (H6, I5). This again shows the relative advantage of LOD for libraries.

It was noted in one presentation that work on data transformation could help to identify where there are problems with the quality of existing data and help to improve it (H6). It was also mentioned by one interviewee that "one of the biggest selling points of linked data has been (…), that we can (…) actually support multilanguage or multilingual approach to data through it" (G1), which is very important as Finland has two national languages.

It was also found in the data that they needed to move to LOD in order to be "better prepared when the MARC format is eventually replaced by newer metadata models" (J3). Another reason given was the new possibilities for discovery interfaces and the use of linked works for better recommendation

services (I5). It was also found in one slide that LOD could be used "as the basis for AI Open metadata" (I5) to make "more transparent AI systems" (I5) as well as to enhance the possibilities of semi-automated methods (I5).

On an organisational level, one of the reasons for the move to LOD was the so-called "Metadata Vision" that was published in 2020 (G1). It was a kind of discussion paper that presented how the National Library of Finland saw the future of metadata in the cultural heritage sector in Finland (G1). This included the need to move to LOD and to have centralised repositories in order to provide better services to end users (G1). The public libraries also launched a project called "Towards the Metadata Vision (2022–2023)", which explored what LD could bring to public libraries and was based on interviews with library staff from Finland and Sweden (I5). This would also indicate that the decision was an authority one, as the metadata visions were most likely at least finalised by a small number of people.

One of the things that probably helped with the decision were the projects that had already been carried out. Examples of this are the Kilda[15] project and the Fennica[16] project (G1). This again shows that they had the necessary knowledge, which helps in the early adoption of innovations.

The LKD project was originally also connected to the platform change of Melinda, which is Finland's national metadata repository (G1, K1, K2). It was found that the system should promote open services and open science and be open source if possible (J2). It can also be argued that the decision was based on opportunistic surveillance, as staff had been testing and looking at LOD before the need for a new system arose. Based on the Diffusion of Innovation theory, this would also be part of the agenda setting stage.

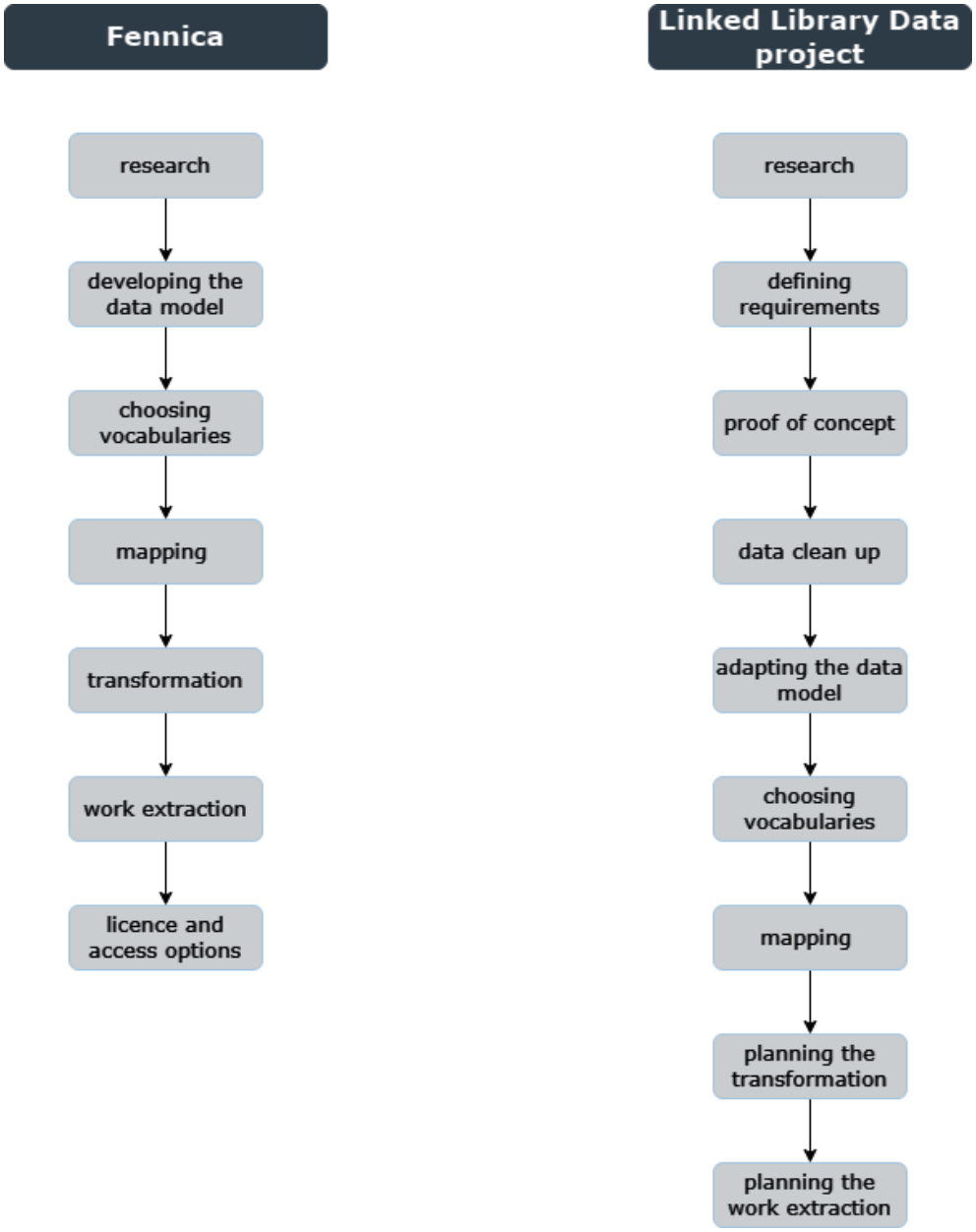
It should be noted that one interviewee mentioned that they are in a kind of second wave of implementing LOD in libraries and that they feel that there is still a lot to be solved, but that it also gives them the opportunity to influence this (G1). Therefore, it can be argued that they are early adopters based on the Diffusion of Innovation theory, as they did not start as early as the National Library of Sweden, but they had conducted their first projects from 2016 onwards and had stated their move towards creating LOD in their "Metadata Vision".

8.3.2 Steps taken

In this section, I will present the steps that the National Library of Finland has taken towards LOD, as shown in *Figure 2*.

[15] The Kilda project aimed to support the transition to LOD. It included, among other things, the translation of RDA, the implementation of RDA, and the publication of the National Metadata Thesaurus as LOD.
[16] The Fennica project was about publishing the National Bibliography of Finland, Fennica, as LOD.

Figure 2. Flow chart of the steps taken by the National Library of Finland



8.3.2.1 Team

The National Library of Finland has two main divisions, one more focused on library work and the other more focused on network services (G1). These divisions are divided into several units, and the LKD project involves people from these different units (G1). For example, there are people from the interoperability services, from Melinda, and from Finna (G1). The team consists of a project manager, a systems specialist, and two information specialists working full time on the project, one person working 50% on the LKD project and 50% on the Melinda project, and one RDA specialist working 30% on the project (G1, H1, I1, I5).

In addition, there are several expert groups specialising in different aspects of cataloguing, for example, for music (G1, K1). These groups include members from the National Library as well as staff from different library sectors (G1).

For this project, a new group has been created for LOD, which includes members from all these cataloguing expert groups (G1). The people involved are again from different library sectors and from libraries with different systems to get different points of view (G1). It was mentioned that the work with these groups will be intensive during 2024 and that the group for LOD will continue after the end of the project (G1).

When asked if anything had changed in the way they worked together, one interviewee answered that they "fused together" (G1) and that they had learned to understand each other's work better (G1). Therefore, it can be said that the interconnectedness that helps with innovativeness in organisations has already improved in the National Library of Finland, which may help when moving to a new cataloguing system. One interviewee also mentioned that there is an assumption that the end result will be better if you work together and everyone understands where everyone's coming from (G1).

8.3.2.2 Research and Collaboration

Prior to the start of the project in autumn 2020, the National Library of Finland started to look at the market (H6). They analysed two systems, Libris XL and Folio[17] in more depth (G1). From spring 2022 onwards, they also conducted a deeper analysis of Libris XL and even had a local test instance (I1, I5). They created GitHub pull requests and tested cataloguing with different materials in Libris XL (I5). They already had some preliminary ideas about how to work with component parts, such as music and book articles (I1). The team experienced some performance problems but already had ideas on how to solve them (I1). The team from the National Library of Finland also had discussions with the team from the National Library of Sweden (I5). Overall, the data showed that the team found Libris XL to be "easy to use even without in depth knowledge of the format or RDA" (I1) and that it "enables different approaches and implementation methods" (I1). The National Library of Finland concluded that the implementation in Finland would require Libris XL to be made more universal and that it would be a lot of work to implement it with the local systems, but that this would be the case with any new system (I1).

The analysis of Libris XL has also led the National Library of Finland to have a closer cooperation with the National Library of Sweden (G1). In addition, the National Library of Finland is also part of the BIBFRAME interoperability group BIG, which is looking at interlinguality and what the smallest common denominator version of BIBFRAME might be. They are also part of Share VDE[18] and one person from the National Library of Finland is also on the organising committee of the BIBFRAME workshop in Europe (G1, H1).

One interviewee mentioned that, to some extent, they are on their own, as other organisations do not follow RDA as strictly as they do (G1). One exception is the Washington Libraries, who use RDA strictly but do not use BIBFRAME (G1). The National Library of Finland also reviewed what the Library of

[17] Folio is an open source library service platform developed by a collaboration of libraries, developers, and vendors.
[18] Share-VDE is a library-led initiative to bring together bibliographic catalogues and authority files from different libraries in a shared discovery environment based on LD.

Congress was doing (H1). In general, it was mentioned that it is important to learn from what others have done (H1, H6).

8.3.2.3 Defining Requirements and Tests

One requirement for the National Library of Finland was to preserve the characteristics of Finnish cataloguing, for example, by using aggregates for music (G1). One interviewee also mentioned that interoperability with what others have done and cooperation were important to them (G1). In addition, the interviewee mentioned that they would still have to convert their data back to MARC 21 for a while as library systems change slowly (G1).

The National Library of Finland carried out some tests with different types of records from the Union Catalogue Melinda (H1). They converted them using the Library of Congress' MARC to BIBFRAME converter and analysed how this differed from what they wanted based on their national cataloguing conventions and the nature of the data (H1). As a second option, they started from the RDA as a kind of guiding light (H1). These results were used to help them make decisions about their own transformations and pre-processing steps (H1). And through both approaches, they hoped to find a harmonious model (H1). This again can be seen as part of the matching phase in which they need to figure out how to implement the innovation, which is also the phase in which the development of the new editor is likely to be at the time of writing.

8.3.2.4 Data Clean-Up

The National Library of Finland had decided that the transformation would involve some preliminary data clean up, as the data comes from different databases and will include quite a few duplicate records (G1). One interviewee also mentioned that the National Library of Finland is constantly making corrections to their data in Melinda, so it is an ongoing process (G1). One interviewee also mentioned that they expect this to lead to a better starting point for the transformations and that this should help when moving to native BIBFRAME in the new Melinda system (G1).

8.3.2.5 Data Model

In order to create their own BIBFRAME model, the National Library of Finland started by comparing their MARC data with the BIBFRAME data they had created, as mentioned in Chapter 8.3.2.3 (H1). They also compared it with the Libris XL and Share VDE data models (K1). The model was based on BIBFRAME 2.1.0 and the official RDA in 2022, with a division between work and expression (G1, H4, K2). The BIBFRAME model has been extended where necessary, but they have worked on maintaining interoperability with the original BIBFRAME model, which they call the default BIBFRAME (K2).

One thing that was important for their data model was that they needed to be able to catalogue component parts, especially for music (G1). So each song on an album is a separate work based on RDA (G1). In RDA, this is done with aggregates, and they want to integrate this into their BIBFRAME model (G1). They also want to integrate miscellaneous additions, such as age limits, which are not generally supported in BIBFRAME (G1). This can be seen as the redefining phase, in which the team adapts the data model to their needs.

In comparison, the Fennica data model was largely based on schema.org with its bibliographic extensions, and the OCLC WorldCat LOD model was used as a reference (K3). It can be noted that schema.org was developed by the major search engine companies and was therefore not rooted in the library field.

8.3.2.6 RDA and Mapping

RDA was implemented in all union catalogue libraries in Finland in 2016 and is also used by Finnish museums and archives for describing agents (J2). This means that RDA was originally implemented in the MARC 21 format (J2). One interviewee mentioned that there was no official mapping between RDA and BIBFRAME yet (G1). There was a mapping from the University of Washington Libraries, but it was several years old and RDA had evolved since then, so the National Library of Finland decided to do their own mapping (G1).

The mapping for the LKD project has been done between the official RDA and BIBFRAME to ensure that all key RDA elements can be used and interoperability is ensured (H1, H2, I1). They have also mapped RDA to the Finnish version of BIBFRAME for their own transformations (G1). It was mentioned that one of the most difficult parts was to divide the BIBFRAME work into work and expression (H2, I5, I6), which I will present in more detail in Chapter 8.3.3.2 (I5). In general, the BF (BIBFRAME) work was divided into BFFI (BIBFRAME FINLAND) work and BFFI expression (G1). One interviewee explained that some of the subclasses of BF work are therefore included in both BFFI work and BFFI expression, and some are only included in one of them (G1). An example would be *language*, which will be in the BFFI expression, and *intended audience*, which will be in the BFFI work (G1). It was also mentioned that they think that it should generally be easy to merge the BFFI work and the BFFI expressions back together to get the BF work (G1). They also mentioned that they have not tried this yet but that they did not think it would break interoperability (G1).

The first raw version of the National Library of Finland's data model was published as version 0.1.0 in January 2023 (I5). It included the nationally mandatory elements for narrow and broad description levels according to RDA (I5). The second version, 0.2.0, was published in June 2023 with some fixes and better translations of the terms into Finnish (I5). At the time of writing, the current version is 0.5.0 and can be found at *http://schema.finto.fi/lkd/*.

8.3.2.7 Vocabularies

The National Library of Finland uses a controlled vocabulary based heavily on RDA with some national additions (G1). All their controlled LOD vocabularies and thesauri are part of their thesaurus and ontology service, Finto, which can be accessed at *https://finto.fi/en/* (G1, I5). They have the YSO, which is the general Finnish ontology with all the subject headings (G1, I1, I5). Here they have included links to Wikidata and the Library of Congress' subject headings (G1, I4). For places, they use YSO places, which also has links to Wikidata and to the National Geographic Survey, which is the official Finnish place data (G1, I1). This means that all coordinates shown in YSO places come from the National Geographic Survey (G1). They have also set up different centuries for time (G1). FINAF is the ontology for agents, including many different name

versions and links to Finna, which contains the published material of agents, as well as links to the ISNIs and ORCIDs, if they are scientists (G1, I1, I5). The National Library of Finland also has a metadata thesaurus and some specialised vocabularies, for example, for instruments (G1). As Finland has two official languages, Finnish and Swedish, all data in the vocabularies is in these two languages (G1). In addition, a large part of the data in the ontologies is also translated into Northern Sami and English (G1).

Some of the links to the vocabularies were already included in the Melinda MARC records, and some are supposed to be included during the data clean up or the pre-processing steps (G1). One interviewee mentioned that there is a working group that creates authority records for works with the most expressions and manifestations (G1). At the time of writing, they had created 400 of these records and were beginning to include them in the MARC records as a pilot project (G1).

8.3.2.8 Transformation

The Fennica project used a batch process to transform the data (H5). Several tools were analysed, but it had become clear that "none of them seemed just right for our purposes (…) but there were some interesting approaches and tools" (H6), so a pipeline was created to use the different tools sequentially (H4, H5, H6, J3). They started with a data dump from the ALEPH database, which at that time was the integrated library system of the National Library of Finland (H5, J3). The one million records were then split into smaller batches for parallel processing (H5, J3). The data was first cleaned and transformed into MARCXML using the Catmandu toolkit (H6, I4, J3). The Library of Congress' converter marc2bibframe2 was then used to transform the data into BIBFRAME expressed as RDF (H6, I4, J3). A SPARQL CONSTRUCT query was then used to convert the BIBFRAME data into schema.org (H6, I4, J3). Another SPARQL query was then used to include links to external datasets, such as SKOS and YSO (H6, I4, J3). Lastly, the data was serialised in HDT (Header, Dictionary, Triples) (H6). The whole transformation took about six hours and generated about 40 million RDF triples from 1 million MARC records (H6, K3).

The LKD transformation is set up similarly. The National Library of Finland had considered making their own version of the Library of Congress' converter but decided against it because it is actively developed and written in a programming language that is difficult to modify (G1). The LKD pipeline will include some preliminary fixes, including some data clean up (G1). Then the Library of Congress' converter will be used to obtain the conventional BIBFRAME RDF (G1). In the post-processing, the data will be transformed into the Finnish BIBFRAME (G1).

Like the National Library of Sweden, the National Library of Finland needs to convert the data back to MARC 21 for the union catalogue Melinda and the other libraries (G1). One interviewee also mentioned that this will result in a conversion loss and that they "hope to sort of identify the loss that we are experiencing and then decide whether it's something that we can live with or whether it needs to be dealt with" (G1).

As the National Library of Finland and the Finnish library sector have not yet started to use the data model, it can be argued that they have not yet reached the clarifying phase based on the Diffusion of Innovation theory.

8.3.2.9 Work Extraction

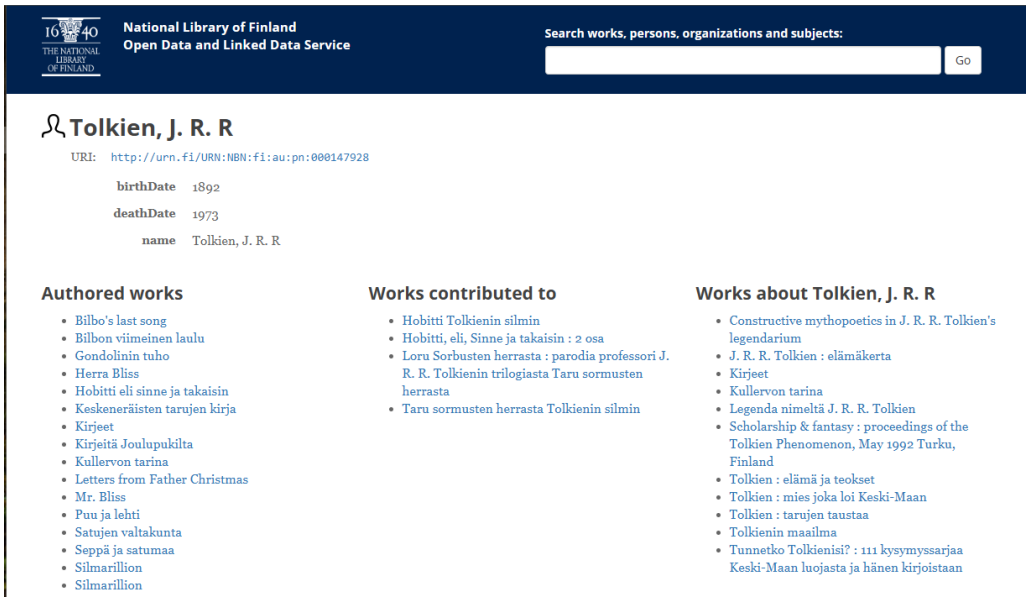
A part of the work extraction will be done through the transformation based on the division of entities as presented in Chapter 8.3.2.6. It was also mentioned by one interviewee that this will lead to many duplicates (G1). The interviewee also mentioned that this will also make it necessary to cluster the duplicates to get to the real BFFI works, expressions, and manifestations, and that this should only be necessary once (G1). When asked if the knowledge about extracting works from the Fennica project was helpful, they replied that some of it was, at least the insight into how difficult it is to cluster works (G1).

8.3.2.10 License and Access Possibilities

As open data is promoted in Finland based on the national policy (J2), the Fennica project has been published under a CC0 license so that it can be used in other applications or, for example, visualisations (J2, J3). All the code for the Fennica project is available on the GitHub page *https://github.com/NatLibFi/bib-rdf-pipeline* (I3).

As the LKD project is not finished yet, I can only present how the LOD from the Fennica project can be accessed. There is a SPARQL endpoint at *http://data.nationallibrary.fi/bib/sparql* and the possibility to download data dumps at *http://data.nationallibrary.fi/download/* (H6, J3, K3). There is also a website (*data.nationallibrary.fi*) where you can browse or search (H4, H6, J3). For example, you can open the page for an author and see the works the person has written as well as works about that person (H4, H6), as can be seen in *Figure 3*. The page also includes links to Finna, the search engine, so you can find more information about where the resources can be borrowed (H4, H6). Data can also be exported in a variety of formats, including RDF/XML, Turtle, N triples, JSON-LD, and HDT (H4, I3).

Figure 3. Screenshot from https://data.nationallibrary.fi



8.3.2.11 Checklist

Finally, I would like to present a checklist for LOD projects that was presented at a conference (H2). The first point that was presented is that BIBFRAME needs to be translated if you are working in a non-English environment, and there is no translation into your language available yet (H2). It was also mentioned that you need to consider how much needs to be translated because cataloguers interact with the data model through the interface of the cataloguing system and the cataloguing rules, and this also means that the data model may not be visible to them, therefore probably not everything needs to be translated (H2). Next, you need to figure out whether everything you need is in BIBFRAME or if you need additions, and if so, whether anything can be reused or if you need to create it yourself (H2). You also need to decide which namespaces will be used (H2). Then you need to think about how you are going to transform your data and whether you need to transform it back to MARC 21 and, if so, how much conversion loss you are prepared to accept (H2). Another issue is the extraction and clustering of works (H2). The presenter mentioned that it is important to remember that this task is "far from trivial" (H2). Another point presented was the choice of controlled vocabularies that will be used and the question of how to deal with authorities such as persons, places, and events (H2). Finally, the presenter mentioned that it is important to consider your library system and your customers in all of this (H2).

8.3.3 Challenges and solutions

I will now present and analyse the challenges that were met by the National Library of Finland in their projects and the solutions they found for them.

8.3.3.1 Existing Data

One problem with the data was the traditional way of handling authority data, with strings and preferred forms that can and do change, leading to many duplicates in the database (H4). This also means that it is not always clear whether a person with the same name is the same person in both cases (H4). In addition, not all authors may be included in authority files (H4, H6). It was also found in the data that it is important to be careful with authority data, as you do not want to link an author to a book that has not been written by that person (H6). Another problem found was errors or missing information in the MARC records (I4).

8.3.3.2 Data Model and Mapping

One interviewee mentioned that the person had not expected how long it would take to familiarise themselves with the BIBFRAME data model (G1). This also shows that LOD as an innovation has a high level of complexity, which may be a reason for the lack of adoption.

Some problems arose because of the different ways in which annotations are used (J1). Another problem arose because of the differences between RDA and BIBFRAME, which made the adoption of the data model more complex (G1, H2). The National Library of Finland ended up using mostly broad matches from RDA to BIBFRAME because BIBFRAME is sometimes more fine-grained, as properties are also tied to specific subclasses of work or instance,

whereas RDA works with more generalised relationships (H2). It could be found in the data that the mapping was not straightforward, as MARC records contain information about many different entities (J3).

It was also mentioned that mapping between data models can be very difficult as there are many different options, such as *rdfs:subClassOf* or *owl:equivalentClass,* and *owl:sameAs* (H2). There are also options in SKOS such as *exactMatch*, *closeMatch*, *narrowMatch,* and *broadMatch* (H2).

8.3.3.3 Links

It could be found in the data that there were problems with links, for example, because the links from the vocabularies were not yet included in the MARC records because there might be several subfields in a field that would have different terms from the vocabulary (G1). It was presented that some records of the Library of Finland are in WorldCat but that they could not use their identifiers because they did not know their OCLC numbers (I4). Originally, authority records for people were not linked to the Virtual International Authority Files (VIAF) and International Standard Name Identifiers (ISNI) (I4), but this has already been changed.

8.3.3.4 Legal issues

The data also included legal issues that may arise from contracts with commercial partners that may restrict the sharing of metadata (J2). Another challenge is the ownership of the data and the possibility of openness, as personal information may be embedded in the metadata (J2).

8.3.3.5 Transformation (back to MARC)

The National Library of Finland has not done the transformation yet, but it was mentioned by both interviewees that they are dreading it because they think it might be more complex than they think (G1).

It was also mentioned that the transformation between MARC 21 and BIBFRAME is not trivial and that they will probably have to transform the data back to MARC at least until the end of the decade (H2, J1). One interviewee also mentioned that the transformation back to MARC means that some things are just going to be taken into BIBFRAME because they are needed later and that the LOD would be simpler if this would not be necessary (G1). This adds to the complexity of the innovation and makes the back transformation a barrier to adoption. It could also be found that a lossless transformation back is probably impossible and that it is therefore necessary to decide what and how much loss can be accepted (J1).

8.3.3.6 Work Extraction

Regarding the extraction of works, the FRBR Work-Set Algorithm seemed to have difficulties when trying to extract works for the Fennica project, as missing metadata was an issue (I1). It was therefore mentioned that it is necessary to check the data after using the algorithm (H4).

A general problem presented for the extraction of works is that there will always be duplicates at the beginning (H6). And missing data will again be a

challenge when trying to cluster the works with the help of algorithms, as some data that may be needed for the merge is not included in all the records or is expressed in different ways (J1, H6). One possible solution presented in the interview would be the creation of work authority files with stable identifiers, which has been started as a pilot as mentioned in Chapter 8.3.2.7 but may need to be adapted after the transformation (G1).

9 Discussion

In this chapter, I will first compare the two cases and other projects before presenting the answers to the five research questions presented in Chapter 3.

9.1 Comparison of the two cases and previous literature

In order to find out if there were specific factors that led to the early adoption by the National Library of Sweden, as well as if the National Library of Sweden faced any additional challenges due to their early adoption, I will now compare the reasons for starting the projects and the challenges faced in the two cases, the National Library of Sweden and the National Library of Finland, as well as some projects and previous literature that I had found during my literature search in the databases *Emerald*; *Library & Information Science Abstracts LISA*; *Library, Information science & Technology Abstracts LISTA; Scopus* and *Web of Science* for Chapter 5.

9.1.1 What led to the projects?

I will begin by comparing the reasons for carrying out LOD projects. An overview of the factors can be found in *Table 1*.

Table 1. Comparison of reasons for implementation

Factors for the implementation	National Library of Sweden	National Library of Finland	Other projects
Change in the community	x	x	
Constraints of MARC	x	x	x
Catalogue enrichment	x	x	x
Improving the existing data	x	x	
Improving visibility of the data	x	x	x
Multilingualism		x	x
Organisational factors, e.g. new strategy		x	x
Previous projects	x	x	
(Re-)usability of the data	x	x	x
Usability outside the library field	x	x	x

One reason that was found in the data from the National Library of Sweden was the change in the community, for example, because several LOD vocabularies had been published. It could also be found that they knew that things were going to change, and they wanted to take charge of that. Something similar was mentioned by staff at the National Library of Finland, who mentioned that they wanted to build their data model now because they wanted to be prepared when MARC would be replaced.

Another reason presented by Williams (2021) is that MARC does not fit the requirements anymore. For example, it lacks interoperability with data formats outside the library sector and is understood only by library software and librarians. This could also be found in the data from the National Library of Sweden, where it was mentioned that MARC is not suitable for newer resources such as e-books. A very important reason for the National Library of Finland was also that LOD supports multilingualism, which is very important as Finland has two national languages. This benefit has also been presented previously by Rasmussen Pennington and Cagnazzo (2019).

Rasmussen Pennington and Cagnazzo (2019) and Ullah et al. (2018), as well as the National Library of Sweden and the National Library of Finland, presented that the change to LOD could improve the visibility of library data on the web. Furthermore, the National Library of Sweden and Gaitanou et al. (2022) presented that the change to LOD would make the data more usable outside the library field. This (re-)usability of library data has also been presented as a reason for LOD projects by Rasmussen Pennington and Cagnazzo (2019), and Tharani (2015). This reason could also be found in the data from the National Library of Sweden and the National Library of Finland. It could also be found in the data from the National Library of Finland that cataloguers could benefit from this, as they could, for example, reuse data from works in their expressions. Another important factor is the possibility of enriching library data with data from other sources, expanding the context of the data. This could be found in the data of both cases, as well as in Rasmussen Pennington and Cagnazzo (2019) and Ullah et al. (2018).

The data from the National Library of Sweden also showed that the data model change could be used to improve the data, which was inconsistent due to thirty years of changing cataloguing rules and practices. Something similar, namely that the change could help to identify problems in the quality of the data, could also be found in the data from the National Library of Finland. The data from the National Library of Finland further showed that LOD could be used for better discovery systems and recommendation services. The first was also found in Gaitanou et al. (2022) and Rasmussen Pennington and Cagnazzo (2019). In addition, the data from the National Library of Finland also presented the possibility of using the data for transparent AI systems.

In both analysed cases, the move towards LOD was connected to a desired editor change. The National Library of Sweden wanted to make the editor more intuitive and flexible and needed a new one, as their previous editor, Voyager, did not meet their requirements anymore. The National Library of Finland still intends to make this change but has postponed it until probably 2028.

Vincent et al. (2018) stated that one of the reasons for publishing the national bibliography in LOD was a new strategy of the National Library of Scotland. Something similar was also found in the data of the National Library of Finland with its own "Metadata Vision", as well as that of the Finnish public libraries. Furthermore, the data for the National Library of Finland showed that a national policy for open data was another reason for the project. I had not found anything similar for the National Library of Sweden, but it was mentioned that the department for Libris was able to decide on the change itself, as they were responsible for the editor.

Based on the analysis of both cases, it also seems that previous LOD projects and experiments were helpful in starting these large projects, as they already had some experience. The data from the National Library of Sweden also showed that the team had knowledge of setting up import and export routines and developing the search interface. In addition, it could be found in the data from the National Library of Sweden that none of the available systems met all their requirements and that developing the system themselves would give them more flexibility than working with a vendor. Therefore, it seems that the existing knowledge in both LOD and system development has been a strong reason for starting the project so early. It can also be assumed that knowledge is a very important factor in starting innovation projects in general.

Based on the Diffusion of Innovation theory, larger organisations also tend to be more innovative, which would be the case for the National Library of Sweden and the National Library of Finland, as the decision in these cases also affected a larger group of libraries because of the union catalogues. Many other large libraries have not made this change, so this may not have the strongest impact on such decisions.

9.1.2 Challenges and Solutions

Now, I will compare the challenges that were met in the different projects, as also presented in *Table 2*.

Table 2. Comparison of challenges

Challenges	National Library of Sweden	National Library of Finland	Other projects
Choice of vocabularies	x	x	x
Data model			
Interface development	x		
Lack of expertise			x
Lack of resources			x
Lack of staff awareness			x
Legal issues		x	x

Links	x		x
Mapping	x	x	x
MARC format	x	x	
Missing and erroneous data	x	x	x
Organisational challenges	x		
Personal challenges	x		
Work extraction	x	x	

One challenge that has been found in previous literature, such as Rasmussen Pennington and Cagnazzo (2019), namely the lack of staff awareness of LD in libraries, could not be found in the data from the National Library of Sweden and the National Library of Finland. This could be due to the fact that I mainly analysed information given by the project team and a lot from conference presentations. Therefore, I have not gotten any information on how aware, for example, the cataloguers or people outside the national libraries were. Furthermore, a lack of expertise and resources was mentioned as an issue by Ali and Warraich (2018), Gaitanou et al. (2022), and Rasmussen Pennington and Cagnazzo (2019), as well as a steep learning curve. It seems that in both presented cases, there was already some knowledge that helped to start the projects. However, one interviewee from the National Library of Finland also mentioned that the person had not expected that it would take so long to familiarise themselves with BIBFRAME. There was no mention of the learning curve in the data from the National Library of Sweden. It would have been interesting to know more about this, but unfortunately, I did not ask about it in the interviews.

Ali and Warraich (2018) and Gaitanou et al. (2022) presented that the choice of vocabularies, ontologies, and tools is an issue in LOD projects. This was also found to some extent in the projects of the National Library of Sweden and the National Library of Finland. For example, the National Library of Sweden looked at what they could use to link their places in a more granular way. Another example was the search for a transformation tool in the National Library of Finland's Fennica project, in which they ended up using a pipeline of different tools as no tool was able to do everything. McKenna et al. (2018) and Rasmussen Pennington and Cagnazzo (2019) also stated that the use of different schemas and vocabularies can lead to interoperability issues and slow down the adoption of LOD. In the data from the National Library of Sweden, several challenges were found in relation to linking, such as when no authority file was available for specific people. Jin (2021) also mentioned a problem with choosing a link, for example, for their works, where they ended up using the OCLC numbers.

A really important issue identified by the National Library of Sweden and the National Library of Finland is missing and erroneous data, as well as different ways of entering data. An example given by the National Library of Finland is that authorities were originally written as strings with preferred names, which

often changed over time, leading to issues in identifying the correct person and making it difficult to cluster works. This was also presented by Ali and Warraich (2018), who mentioned that data in languages other than English could make the work more difficult. It was also presented by Jin (2021) and Ullah et al. (2018) that data clean up is a very important and necessary step for transformation projects.

In the case of the National Library of Sweden, several challenges were identified due to the way in which data is included in MARC, such as information about entities that is divided into separate fields in MARC. Another one is the fact that one MARC field can contain information about multiple entities. The necessary transformation back to MARC could also be identified as a major challenge in both analysed cases, as well as information that, without this, they would have been able to create simpler LOD.

Regarding the data model, it could be found that the National Library of Sweden had not intended to include all the data they ended up with because some libraries needed some fields for their systems. Something similar was found in the data from the National Library of Finland, where it was mentioned that the creation of the data model was made more complicated by the integration with all the local systems. The National Library of Finland also met a challenge due to the differences between BIBFRAME and RDA. Both Ali and Warraich (2018) and Gaitanou et al. (2022) mentioned mapping issues, which could also be found in the data from the two cases. The National Library of Sweden had issues due to different granularities between the schemas, and the National Library of Finland met challenges due to the different options for mapping the data.

Something that was only found in the data from the National Library of Sweden and the National Library of Finland was the complexity of extracting works. First, works have to be created from the data, which leads to many duplicates. Then these have to be clustered into the actual works, which is quite complicated, for example, because of the denormalised data.

Warraich et al. (2022) also mentioned legal challenges, which were also found in the data from the National Library of Finland, due to contracts that libraries have with commercial partners that may limit the openness of the data as well as personal information that may be included in the data.

One challenge that was only found in the data from the National Library of Sweden was the challenge of creating an easy and intuitive interface. They had to make several changes because, for example, it was not clear to users what data could be edited. In the beginning, some steps also took a very long time, which was changed to make the interface more ergonomic. Another challenge only found for the National Library of Sweden was a problem with exporting the data to the libraries' systems, partly due to the different MARC data required in their systems. The fact that these challenges were only identified for the National Library of Sweden can be explained by the fact that no other analysed case or project report had involved a library that had created an interface to create its data in LOD, transformed it back to MARC, and had to export it to different systems.

The National Library of Sweden also encountered issues due to the re-organisation, as developers from the project team were divided into different departments and resources had to be obtained from these departments. The data from the National Library of Sweden also included information about personal challenges, for example, the uncertainty of what was coming that was experienced by cataloguers and the other libraries. The psychological effect that such a data model and interface change have on people was also mentioned in one interview. Both of these challenges were also presented in the Diffusion of Innovation theory when innovations are implemented.

9.2 Answers to the Research Questions

In this chapter, I will present the answers to the research questions, starting with the steps taken in the analysed projects, which are presented in more detail in Chapters 8.2.2 and 8.3.2. The first step taken by the National Library of Sweden was to research what data models and systems were available. As no data model met all their needs, they built their own and created their own vocabulary in Swedish. The National Library of Sweden also decided to develop their own editor, as no existing system met all their requirements. Next, a proof of concept was carried out by transforming some MARC 21 data into LD.

Regarding the data model, the National Library of Sweden started to create their own in 2012. They looked at BIBFRAME 1.0 when it was published but decided that it was not what they wanted to use and continued to work on their own model. In 2016, they examined BIBFRAME 2.0 and decided to align their model with BIBFRAME 2.0 as it was similar to theirs and they felt that BIBFRAME could become a community model. The National Library of Sweden then mapped their MARC to RDF, which took quite a long time as they always had to keep the back transformation to MARC 21 and the needs of the other libraries in mind.

The National Library of Sweden decided to use JSON-LD to serialise their LOD and executed the transformation with a configuration file in JSON with a Java logic and included pre- and post-processing steps. At the time of writing, the LOD includes links to people and subject headings from the National Library of Sweden's vocabularies, as well as some geographical area codes. In 2023, the National Library of Sweden extracted and clustered BIBFRAME works for Swedish fiction. They intend to continue this process in the future, probably starting with fiction in other languages. The data is licensed under Creative Commons CC0 and can be accessed via a read and write API and a SPARQL endpoint.

The National Library of Sweden also developed the editor with the help of UX designers and cataloguers within the National Library of Sweden and in the libraries that receive their data from Libris XL. A beta version of the editor was available for testing and feedback in the spring before the launch. The National Library of Sweden also worked with specialists for the data types to prepare different profiles for various kinds of data and included help texts in the editor. A part of the project team also prepared documentation and video tutorials. The source code as well as the code for the transformation are available on GitHub.

The National Library of Finland's first step was to examine the market. They looked closely at the National Library of Sweden's Libris XL editor and even had a local test instance. In addition, they did some initial tests, transforming their data into BIBFRAME. As they intended to divide the BIBFRAME work into work and expression, they needed to split the entities for their mapping. The first version of the data model was published in January 2023, and at the time of writing, the current version is 0.5.0.

The National Library of Finland will use links to their own controlled vocabularies, which are in Finnish and Swedish and partly in Northern Sami and English. The transformation shall be executed with a pipeline that will first do some data clean up before transforming it with the Library of Congress' converter into the conventional BIBFRAME and lastly converting it into the Finnish BIBFRAME model. As the project has not been finished at the time of writing, the transformation has not been carried out yet.

Next, I will present the (additional) challenges that the National Library of Sweden faced and how they overcame them, which is described in more detail in Chapter 8.2.3. One problem that arose because of the different needs of the libraries is that some data from MARC fields that the National Library of Sweden had not intended to include had to be included, and therefore the mapping had been adjusted accordingly. The analysed data showed that a large part of the mapping and transformation problems came from the necessary transformations back to MARC. In addition, the structure of MARC made the process more complicated because some entities are distributed across multiple fields, and some fields contain information about multiple entities. Another challenge for the transformation was the existing data, which included mistakes or missed information. In addition, some fields had been used differently over time, for example, due to different cataloguing rules. The National Library of Sweden did some data clean up, but it was mentioned that they had not been able to do as much as they would have liked.

There have been some challenges with linking people, as the preferred names of contributors have changed over time, making it difficult to know if the person is actually the person the authority file is about. It could also be found in the data from the National Library of Sweden that linking to data from other sources is difficult because you are not responsible for that data, but you are responsible for what is included in your data. It was also stated that it would be necessary to, for example, get a notification if something changes in the linked data, but no solution for this was found in the analysed data.

The extraction of works was and is a major challenge, as this process initially produces many of the same works from the MARC records. The clustering of these works is made more complicated by the problems with the existing data, as presented before. One (partial) solution described by the National Library of Sweden was to clean up the data by, for example, getting the illustrator from field 245 and adding it to field 700 if it was missing there, so that the work could be merged with the others.

There were also several challenges with the design of the editor. The National Library of Sweden had originally intended to make it easy to browse through

the data, but after testing, they received feedback that the interface made it difficult to know what could be edited. The National Library of Sweden also needed to make some things clearer, such as what data is linked and what is just a string. In addition, people missed features they were used to in the previous editor. Some of these features, such as one to handle duplicates, have been implemented since. There was also feedback on the ergonomics, because in the beginning it took 15 to 20 clicks to add some things, and this has been adapted. To know how intuitive the editor is now, it would be necessary to interview or survey cataloguers who work with it.

Personal challenges could also be found in the analysed data, as cataloguers had to get used to the new way of creating metadata, which was very different from what they were used to and good at. The fact that they had heard about some upcoming changes early on in the project, but then nothing happened for a long time, may have added to the unease of the affected people. To get a more detailed view of this, interviews or a survey with these people would be necessary.

The re-organisation of the National Library of Sweden and the division of members of the Libris department into different departments and the need to obtain resources from these departments made it difficult to work on the project as there were suddenly fewer people involved. The fact that Libris had been its own department for so long and was not originally part of the National Library of Sweden also seems to have led to tensions and a feeling that Libris XL was not a product of the National Library of Sweden.

Regarding issues that may have arisen due to the early adoption of the innovation and the presented comparison with the challenges met by the National Library of Finland and other projects presented in Chapter 9.1, it can be said that the implementation time was probably longer because they had little to nothing to look at when working on their own data model and editor. The data model took the National Library of Sweden at least 4 years from the start of the project to the alignment with BIBFRAME in 2016, while the National Library of Finland shall be finished in 2024 after 2 years of development. As the move to a new editor by the National Library of Finland is postponed until around 2028, it cannot be said that this has taken the National Library of Sweden longer, but it is also difficult to compare, as I do not know how many people are involved in the National Library of Finland's Melinda project, so this may be due to the available resources. To make a full comparison, it would also be necessary to analyse the second case further after they have launched their new editor. It can be noted that a distinction between innovators and early adopters based on the Diffusion of Innovation theory does not seem to be particularly helpful in the analysed cases, as both still share most of the same problems.

Lastly, I will present the factors that supported the National Library of Sweden's decision to start the project so early. One reason that could be found in the data for the early adoption was that the National Library of Sweden needed a new system because their previous system, Voyager, did not meet their requirements any longer. Voyager did not work well with newer media types and had become quite slow as it was not designed to handle the huge

amounts of data that were now being imported. It is interesting to note that the National Library of Finland's project was also originally linked to the planned system change, so this seems to have been a driving force for the start of such projects. The National Library of Sweden did not decide to use LOD at the start, but they quickly realised that what they wanted could not be done with MARC, so they decided that the next step would be to use LOD.

The fact that someone in their department had already experimented with transforming data into LOD can also be assumed to have helped with the decision. They also had experience in developing the search interface themselves, so it seemed possible to develop the editor themselves. Both show that the National Library of Sweden had the necessary knowledge, which is a beneficial factor for the early adoption of innovations. In general, it can be said that the National Library of Sweden, or the Libris department, felt that things were changing, including the availability of LOD vocabularies, and felt that libraries needed to take charge of them.

9.3 Limitations of this Master's Thesis

Due to the available resources, only four people could be interviewed for this thesis. This gives only a limited view of the projects and does not include the group highly affected by the change, the cataloguers. Furthermore, no people from libraries outside the National Library of Sweden were included, so nothing can be said about how it felt for them to wait for the big change that the National Library of Sweden had presented or how it felt to have to make that change without having a say in it. I was only able to analyse data that was available online and that I had found during my research, so some information may not have been included in the analysis. This also applies to project reports from other libraries, as not everything will have been discoverable in the databases I used or have been in languages I understand. The latter also applies to all recordings of the National Library of Finland in Finnish.

10 Concluding Discussion

As presented in the previous research in Chapter 5, there are many reasons that may have led to the slower adoption of LOD in the library field. It has also become clear through the analysis in Chapter 8 and Chapter 9.1 that the attribute of relative advantage, based on the Diffusion of Innovation theory, has not been enough to outweigh other attributes such as the complexity of the innovation. The analysis of the two cases has also shown that the fact that most systems are still working with MARC 21 makes the transition more difficult because moving towards LOD still requires a back transformation, which makes the mapping between MARC and LOD, in these cases, BIBFRAME, more difficult than it would be without it.

Based on the different projects, it can be said that one of the most important steps in LOD seems to be data clean up, as missing and erroneous data complicates both the transformation and the extraction of works. Therefore, one piece of advice that can be given is to plan enough time to do data clean up. Particularly for union catalogues, it is also important to identify the needs of all participating libraries, as their needs may differ from those of the project library and from each other. It is also important to involve cataloguers and metadata specialists to ensure that all special cases are considered.

Two factors seem to have been important in the National Library of Sweden's early adoption of LOD. Firstly, the knowledge they already had about LOD and interface development at the start of the project. And secondly, the fact that they needed a system change and decided that LOD was the way forward. That a necessary system change can be a driving force was also found in the data of the National Library of Finland, as their LKD project was also linked to the intended editor change. One interviewee also mentioned that libraries do not change their systems too often and that when they do, they need to decide whether to move to LOD now or wait for the next change (A1).

For future research, I suggest that it would be important to interview people from different departments, including UX experts and cataloguers, to get a broader view of how the innovation project felt to them and what problems they might have encountered, as these (mostly) could not be identified in the analysed data. It would also be interesting to compare the information provided by the cataloguers within the project libraries, in this case the national libraries, and the libraries that receive their data from the union catalogues.

It would further be interesting to conduct follow-up interviews with staff from the National Library of Finland after they have moved to their new system to see how their development and implementation of the editor might differ from that of the National Library of Sweden.

In general, it would also be interesting to look at the transition in different types of libraries, as now only national libraries with union catalogues have been analysed as cases. One possibility would be the Deichman Library in Oslo, Norway, which moved to creating LOD in 2016.

I think that this thesis can be seen as a basis for comparison of future innovation implementation projects and that it can help libraries in planning their future projects, as some major challenges have been presented that need to be considered in the planning phase, both specifically for LOD projects and for innovation implementations in general. Based on this case study, previous knowledge seems to be an important factor in starting large innovation projects, as both analysed cases had previously carried out smaller LOD projects. In addition, innovators and early adopters need to be aware that they will be mostly on their own at the beginning of such projects and that they will have to figure things out for the most part on their own, which may lead to a longer implementation time.

References

Alemu, G. (2022). *The future of enriched, linked, open and filtered metadata: Making sense of IFLA, LRM, RDA, Linked Data and BIBFRAME*. Facet.

Ali, I., & Warraich, N. F. (2018). Linked data initiatives in libraries and information centres: A systematic review. *The Electronic Library*, *36*(5), 925–937. https://doi.org/10.1108/EL-04-2018-0075

Alvite-Díez, M.-L. (2021). Linked open data portals: Functionalities and user experience in semantic catalogues. *Online Information Review*, *45*(5), 946–963. https://doi.org/10.1108/OIR-07-2020-0295

Apenīte, M., & Bojārs, U. (2021). National Library of Latvia Subject Headings as Linked Open Data. In R. Verborgh (Ed.), *Lecture Notes in Computer Science: v.12739, The semantic web: ESWC 2021 satellite events, virtual event, June 6-10, 2021, revised selected papers* (pp. 33–37). Springer. https://doi.org/10.1007/978-3-030-80418-3_6

Berners-Lee, T. (2009, June 18). *Linked Data - Design Issues*. https://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, *284*(5), 34–43. https://doi.org/10.1038/scientificamerican0501-34

Bryman, A. (2016). *Social research methods* (Fifth edition). Oxford University Press.

Candela, G., Escobar, P., Carrasco, R. C., & Marco-Such, M. (2018). Migration of a library catalogue into RDA linked open data. *Semantic Web*, *9*(4), 481–491. https://doi.org/10.3233/SW-170274

Chen, Y.-N. (2023). An investigation of linked data catalogue features in libraries, archives, and museums: a checklist approach. *The Electronic Library*, *41*(5), 700–721. https://doi.org/10.1108/EL-03-2023-0070

Cole, T. W., Han, M.-J., Weathers, W. F., & Joyner, E. (2013). Library Marc Records Into Linked Open Data: Challenges and Opportunities. *Journal of Library Metadata*, *13*(2-3), 163–196. https://doi.org/10.1080/19386389.2013.826074

Gaitanou, P., Andreou, I., Sicilia, M.-A., & Garoufallou, E. (2022). Linked data for libraries: Creating a global knowledge space, a systematic literature review. *Journal of Information Science*, *50*(1), 204-244. https://doi.org/10.1177/01655515221084645

IFLA Study Group on the Functional Requirements for Bibliographic Records. (1998). *Functional Requirements for Bibliographic Records: Final Report*. K. G. Saur. https://doi.org/10.1515/9783110962451

Jin, Q. (2021). Enhanced Discovery with Linked Open Data for Library Digital Collections. *Technical Services Quarterly*, *38*(1), 17–32. https://doi.org/10.1080/07317131.2020.1854575

Kulturdepartementet. (2022, April 22). *Strategi för ett starkt biblioteksväsende 2022-2025*. https://www.regeringen.se/globalassets/regeringen/bilder/kulturdepartementet/kultur/biblioteksstrategi_k-dep_webb_220420.pdf

Kungliga biblioteket. (2024a). *Färdplan 2024-2026*. https://urn.kb.se/resolve?urn=urn:nbn:se:kb:publ-724
Kungliga biblioteket. (2024b, May 29). *Välkommen med på en utvecklingsresa!* https://www.kb.se/5.33d8b23218fbd1bd5edb0.html
Library of Congress. (2016, April 21). *Overview of the BIBFRAME 2.0 Model (BIBFRAME - Bibliographic Framework Initiative, Library of Congress)*. https://www.loc.gov/bibframe/docs/bibframe2-model.html
Lund, B., Omame, I., Tijani, S., & Agbaji, D. (2020). Perceptions toward Artificial Intelligence among Academic Library Employees and Alignment with the Diffusion of Innovations' Adopter Categories. *College & Research Libraries*, 865. https://doi.org/10.5860/crl.81.5.865
Marden, J., Li-Madeo, C., Whysel, N., & Edelstein, J. (2013). Linked open data for cultural heritage. In M. J. Albers (Ed.), *SIGDOC '13 : proceedings of the 31st ACM International Conference on Design of Communication* (pp. 107–112). ACM. https://doi.org/10.1145/2507065.2507103
Mayring, P. (2022). *Qualitative Content Analysis: A Step-by-Step Guide*. Sage.
McKenna, L., Debruyne, C., & O'Sullivan, D. (2018). Understanding the Position of Information Professionals with regards to Linked Data. In J. Chen (Ed.), *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries* (pp. 7–16). ACM. https://doi.org/10.1145/3197026.3197041
Mitchell, E. T. (2013). Building Blocks of Linked Open Data in Libraries. *Library Technology Reports*, *49*(5), 11-25.
Moloney, P. (1997). Kogarah Municipal Library: library automation and innovation diffusion—a study. *The Australian Library Journal*, *46*(2), 154–160. https://doi.org/10.1080/00049670.1997.10755794
National Library of Finland. (2020). *Strategic Plan of the National Library of Finland 2021-2023*. https://urn.fi/URN:NBN:fi-fe2020042822711
National Library of Finland. (2024). *Linked Library Data project (LKD) - Linkitetty kirjastodata -projekti*. https://www.kiwi.fi/pages/viewpage.action?pageId=298647995
Neo, E., & Calvert, P. J. (2012). Facebook and the diffusion of innovation in New Zealand public libraries. *Journal of Librarianship and Information Science*, *44*(4), 227–237. https://doi.org/10.1177/0961000611435038
Niu, J. (2018). Comparing the diffusion and adoption of linked data and research data management services among libraries. *Information Research*, *25*(2), paper 855. http://www.informationr.net/ir/25-2/paper855.html (Archived by the Internet Archive at https://bit.ly/2WOJqrJ)
Niu, J. (2019). Diffusion and adoption of research data management services. *Global Knowledge, Memory and Communication*, *69*(3), 117–133. https://doi.org/10.1108/GKMC-05-2019-0057
Niu, J. (2020). Diffusion and adoption of Linked Data among libraries. *Proceedings of the Association for Information Science and Technology*, *57*(1), Article e216. https://doi.org/10.1002/pra2.216
Olson, N. (2010). *Taken for Granted: The Construction of Order in the Process of Library Management System Decision Making* [Dissertation]. Högskolan i Borås, Institutionen Biblioteks- och informationsvetenskap / Bibliotekshögskolan. urn:nbn:se:hb:diva-3553

Panchyshyn, R. S., & Park, A. L. (2015). Resource Description and Access (RDA) Database Enrichment: The Path to a Hybridized Catalog. *Cataloging & Classification Quarterly*, *53*(2), 214–233. https://doi.org/10.1080/01639374.2014.946574

Park, H., & Kipp, M. (2019). Library Linked Data Models: Library Data in the Semantic Web. *Cataloging & Classification Quarterly*, *57*(5), 261–277. https://doi.org/10.1080/01639374.2019.1641171

Ponelis, S. R., & Adoma, P. (2018). Diffusion of open source integrated library systems in academic libraries in Africa. *Library Management*, *39*(6-7), 430–448. https://doi.org/10.1108/LM-05-2017-0052

Rasmussen Pennington, D., & Cagnazzo, L. (2019). Connecting the silos. *Journal of Documentation*, *75*(3), 643–666. https://doi.org/10.1108/JD-07-2018-0117

Raynard, M. (2017). Understanding Academic E-books Through the Diffusion of Innovations Theory as a Basis for Developing Effective Marketing and Educational Strategies. *The Journal of Academic Librarianship*, *43*(1), 82–86. https://doi.org/10.1016/j.acalib.2016.08.011

RDA. (2013). *Resource, Description and Access: 2013 Revision*. Facet. https://doi.org/10.29085/9781856047159

Riva, P., Le Bœuf, P., & Žumer, M. (2017). *IFLA Library Reference Model: A Conceptual Model for Bibliographic Information*. https://repository.ifla.org/bitstream/123456789/40/1/ifla-lrm-august-2017_rev201712.pdf

Rogers, E. M. (2003). *Diffusion of Innovations* (Fifth edition). Free Press.

Southwick, S. B. (2015). A Guide for Transforming Digital Collections Metadata into Linked Data Using Open Source Technologies. *Journal of Library Metadata*, *15*(1), 1–35. https://doi.org/10.1080/19386389.2015.1007009

Suominen, O. (2023, November). Finnish National Bibliography Fennica as Linked Data. *Europeana Pro*. https://pro.europeana.eu/page/issue-9-swib

Tharani, K. (2015). Linked Data in Libraries: A Case Study of Harvesting and Sharing Bibliographic Metadata with BIBFRAME. *Information Technology and Libraries*, *34*(1), 5–19. https://doi.org/10.6017/ital.v34i1.5664

Ullah, I., Khusro, S., Ullah, A., & Naeem, M. (2018). An Overview of the Current State of Linked and Open Data in Cataloging. *Information Technology and Libraries*, *37*(4), 47–80. https://doi.org/10.6017/ital.v37i4.10432

Vincent, H., Cunnea, P., & Pretto, A. de (2018, August 29). *From Scottish Bibliographies Online to National Bibliography of Scotland: Reinventing a National Bibliography for the 21st Century*. IFLA WLIC 2018 – Kuala Lumpur, Malaysia – Transform Libraries, Transform Societies. https://library.ifla.org/id/eprint/2275/

Wang, Y., & Yang, S. Q. (2018). Linked Data Technologies and What Libraries Have Accomplished So Far. *International Journal of Librarianship*, *3*(1), 3. https://doi.org/10.23974/ijol.2018.vol3.1.62

Warraich, N. F., Rasool, T., & Rorissa, A. (2022). Challenges and prospects of linked data technology: A qualitative study of Pakistani LIS

professionals' insights. *Digital Library Perspectives*, *38*(2), 175–188. https://doi.org/10.1108/DLP-02-2021-0012

Wennerlund, B., & Berggren, A. (2019). *Leaving Comfort Behind: A National Union Catalogue Transition to Linked Data*. IFLA WLIC 2019 - Athens, Greece - Libraries: dialogue for change. https://library.ifla.org/id/eprint/2745/

Williams, J. (2021). From Silo to the Web: Library Cataloging Data in an Open Environment. In J. Williams (Ed.), *Technical Services in the 21st Century* (Vol. 42, pp. 175–191). Emerald Publishing Limited. https://doi.org/10.1108/S0732-067120210000042013

Zhang, Y., & Wildemuth, B. M. (2017). Qualitative Analysis of Content. In B. M. Wildemuth (Ed.), *Applications of social research methods to questions in information and library science* (Second edition, pp. 318–329). Libraries Unlimited.

Appendix

Appendix A: Analysed Data from the National Library of Sweden/Kungliga Biblioteket

Interviews (A)

A1 – Interview, 90 minutes, 08.03.2024

A2 – Interview, 25 minutes, 11.03.2024

Recordings (B)

B1 – Casalini libri. (2022, September 29). *BFWE 2022 - Niklas Lindström & Fredrik Klingwall, Ambition and Ambiguity* [Video]. Youtube. https://youtu.be/goQuMvCJnlc?si=BqC80vaLQhuQpZQB

B2 – Casalini libri. (2023a, October 6). *BFWE 2023 - Andreas Andersson, Clustering, extracting and linking bibliographical work entities* [Video]. Youtube. https://youtu.be/hGAo_D_Y_Fg?si=xbeMEWaiSVKvzLv9

B3 – Casalini libri. (2023b, October 9). *BFWE 2023 - Roundtable, Exchanging experiences* [Video]. Youtube. https://youtu.be/o8-5gj44dXg?si=RNuQIb4SHuv9GNaC

B4 – Dublin Core Metadata Initiative. (2021, December 27). *Panel 2 - Bibframe Implementation* [Video]. Youtube. https://youtu.be/ecSIdJr9fJ0?si=FH9hsNDvLy9CWbRa

B5 – Elag Conference. (2014, June 25). *elag2014 1106 3 The LIBRIS upgrade* [Video]. Youtube. https://youtu.be/ZR8xm01sSzM?si=hJa8qENSejdlIHww

B6 – Kungliga biblioteket. (2014, February 6). *Libris XL och nytt katalogiseringsverktyg. Projekt vid Kungl. biblioteket* [Video]. Youtube. https://youtu.be/AYGQktFlodg?si=HkWPtWbo4iH0TTlr

B7 – Kungliga biblioteket. (2014, December 4). *Att hänga samman - Libris XL och nya katalogiseringsverktyget* [Video]. Youtube. https://youtu.be/ELiNTtzLIvM?si=amMpYe8zG7ewPGf2

B8 – Kungliga biblioteket. (2014, December 11). *LIBRIS XL, Maria Hedenström* [Video]. Youtube. https://youtu.be/kn9yTSqFI6A?si=J2K96VslKvDinTqV

B9 – Kungliga biblioteket. (2015, September 23). *ELAG 2015 I What Are We Describing?* [Video]. Youtube. https://youtu.be/HjCAgVQWzaE?si=AT8HEhe5wme4lol6

B10 – Kungliga biblioteket. (2016, June 23). *Metadata utanför biblioteken* [Video]. Youtube. https://youtu.be/Ho-SLIrk6hE?si=dwIoeaSwYVdzyF06

B11 – Kungliga biblioteket. (2017, November 29). *Format och vokabulär - Niklas Lindström, Kungliga biblioteket* [Video]. Youtube. https://youtu.be/9IFE4J9dIOs?si=RR2aB36j43hVwREU

B12 – Kungliga biblioteket. (2017, November 29). *Import, export och de lokala bibliotekssystemen - Charlotta Simonsson, Kungliga biblioteket* [Video]. Youtube. https://youtu.be/N37-cgLnWgU?si=0zBNVMFNpaaJf-b5

B13 – Kungliga biblioteket. (2017, November 29). *Kort om sammanlänkad information - Peter Krantz, Kungliga biblioteket* [Video]. Youtube. https://youtu.be/oSFWIAwu3bs?si=1HRXV2wR-R3xAzGP

B14 – Kungliga biblioteket. (2017, November 29). *Libris tar sikte på synlighet - Anna Berggren, Kungliga biblioteket* [Video]. Youtube. https://youtu.be/-3wo0gLjGkg?si=FgAqxIaKzP_JF2Xk

B15 – Kungliga biblioteket. (2017, November 29). *Presentation av användningstesterna - Erik Hellström, Kungliga biblioteket* [Video]. Youtube. https://youtu.be/FTpXrDipTuo?si=x_YdMJmW8Ay19XkO

B16 – Kungliga biblioteket. (2018, December 20). *Det praktiska arbetet i Libris katalogisering* [Video]. Youtube. https://youtu.be/_Y67izImwxI?si=UOtNWOxYxGwLjCXa

B17 – Kungliga biblioteket. (2018, December 20). *Libris svarar* [Video]. Youtube. https://youtu.be/-3geYlxdeZY?si=cKeIBQRMVffBDomF

B18 – Kungliga biblioteket. (2018, December 20). *Sex månader med nya Libris* [Video]. Youtube. https://youtu.be/N26nglSxhDk?si=uM_IXVt2NIuQ3CsR

B19 – Kungliga biblioteket. (2019, November 26). *Bodil Wennerlund - Libris - Vad händer nu?* [Video]. Youtube. https://youtu.be/o36r86jAMAY?si=B9zaelTWiXtrzojV

B20 – Elag Conference. (12.06.20213). *ELAG2013: D3-2/9 Triple Bypass -- Open MARC Surgery* [Video]. Youtube. https://youtu.be/0o6ZgkAV5Pw?si=fVbG5nfj5PL1Upwn

B21 – SWIB. (2019, December 2). *Target vocabulary maps* [Video]. Youtube. https://youtu.be/A_1BIDAlbeI?si=m8lbrFTIsHrgJltr

Slides (C)

C1 – Klingwall, F. (2019). *Extensions for past and future relationships* [Slides]. https://www.kb.se/download/18.d0e4d5b16cd18f600eaf3/1569323774643/Panel%20-%20Extensions%20for%20past%20and%20future%20relationships.pdf

C2 – Lindström, N. (n.d.). *BIBFRAME in production: Libris XL. Linked Data: The Swedish Union Catalogue* [Slides]. https://www.casalini.it/EBW2018/web_content/2018/presentations/Lindstrom_2.pdf

C3 – Lindström, N., & Klingwall, F. (n.d.). *On Ambition and Ambiguity* [Slides]. https://www.bfwe.eu/attachments/bfwe-22-lindstroem-klingwall.pdf

C4 – Malmsten, M. (n.d.). *Exposing library data as linked data* [Slides]. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=2e0791d88a65cb2517e284c2bfca02b7c6660f30

C5 – Klingwall, F. (n.d.) *Working with BIBFRAME at the National Library of Sweden* [Slides]. https://www.kb.se/download/18.d0e4d5b16cd18f600eaca/1569309498978/Working%20with%20BIBFRAME%20at%20the%20National%20Library%20of%20Sweden.pdf

C6 – Andersson, A. (2023). *Clustering, Extracting and Linking Bibliographic Work Entities* [Slides]. https://www.bfwe.eu/attachments/bfwe23-andersson.pdf

C7 – Malmsten, M., Sköld, M., & Berggren, A. (2013). Triple bypass [Slides]. Slideshare. https://de.slideshare.net/geckomarma/presentation-elag-2013?from_search=0

Articles (D)

D1 – Malmsten, M. (2008). Making a Library Catalogue Part of the Semantic Web. In J. Greenberg & W. Klas (Eds.), *Metadata for semantic and social applications: Proceedings of international conference on Dublin Core and metadata applications, 22-26 September 2008 : DC 2008 : Berlin, Germany* (146-152). Dublin Core Metadata Initiative.

D2 – Malmsten, M. (2013). Cataloguing in the open : the disintegration and distribution of the record. *JLIS : Italian Journal of Library, Archives and Information Science = Rivista Italiana Di Biblioteconomia, Archivistica E Scienza Dell'informazione*, 4 (1), 2013, 417–424. https://doi.org/10.4403/jlis.it-5512

D3 – Malmsten, M. (2017). A Step Towards a Distributed Model for Bibliographic Data in Sweden. *Bibliothek Forschung Und Praxis*, 41(1), 40–44. https://doi.org/10.1515/bfp-2017-0013

D4 – Wennerlund, B., & Berggren, A. (2019, September 26). Leaving Comfort Behind: a National Union Catalogue Transition to Linked Data. https://library.ifla.org/id/eprint/2745/1/s15-2019-wennerlund-en.pd

Reports (E)

E1 – Kungliga biblioteket. (2013*). Annual report to CENL (and CDNL) 2012*. https://www.cenl.org/wp-content/uploads/2019/07/Sweden_Annual_Report_2012-2013.pdf

E2 – Kungliga biblioteket. (2015). *Annual Report to CENL and CDNL 2014*. https://www.cenl.org/wp-content/uploads/2019/07/Rapport-till-Cenl-2014_end_20150513.pdf

E3 – Kungliga biblioteket. (2016). *Annual Report 2015 to CENL and CDNL*. https://www.cenl.org/wp-content/uploads/2019/07/Annual-Report-2015-to-CENL-from-National-Library-of-Sweden_6.7-2016-675_final.pdf

E4 – Kungliga biblioteket. (2017). *Annual Report 2016 to CENL and CDNL*. https://www.cenl.org/wp-content/uploads/2019/07/Annual-report-2016-The-National-Library-of-Sweden.pdf

E5 – Kungliga biblioteket. (2018). *Annual Report 2017 to CENL and CDNL*. https://www.cenl.org/wp-content/uploads/2019/08/Report-from-National-Library-of-Sweden_2017.pdf

Press Release (F)

F1 – Kungliga biblioteket. (2018b, July 6). *KB becomes the first national library to fully transition to linked data*. https://www.mynewsdesk.com/se/kungliga_biblioteket/pressreleases/kb-becomes-the-first-national-library-to-fully-transition-to-linked-data-2573975

Appendix B: Analysed Data from the National Library of Finland/Kansalliskirjasto/ Nationalbiblioteket

Interviews (G)

G1 – Interview with two interviewees, 60 minutes, 08.03.2024

Recordings (H)

H1 – Casalini libri. (2022, September 29). *BFWE 2022 - Matias Frosterus, Harmonious marriage of BIBFRAME and RDA - the Finnish approach* [Video]. Youtube. https://youtu.be/HlJl5kd866I?si=evf79W8-t3XatVpS

H2 – Casalini libri. (2023, October 9). *Adapting the BIBFRAME data model to our needs* [Video]. Youtube. https://youtu.be/_OOZKv-vQn4?si=CCY1crKscSFh4rb_

H3 – Kungliga biblioteket. (2017, November 29). *Länkad kulturarvsdata i Finland - Jessica Parland von Essen* [Video]. Youtube. https://youtu.be/VrZzmzHCtio?si=JBu12LyZ2ZOnjpDt

H4 – Kungliga biblioteket. (2018, December 20). *Finlands nationalbibliografi som länkade data* [Video]. Youtube. https://youtu.be/LPiuZyY_VTI?si=BuXmQPr4-mKnvP9x

H5 – SWIB. (2016, December 21). *From MARC silos to Linked Data silos* [Video]. Youtube. https://www.youtube.com/watch?v=xp-PRtdF5_U

H6 – SWIB. (2017, December 19). *Finnish National Bibliography Fennica as Linked Data* [Video]. Youtube. https://youtu.be/sLMxALIQKmQ?si=yppUQio5WESQ-FVJ

Slides (I)

I1 – Lager, L., & Frosterus, M. (n.d.). *The Curious Case of Finnish Metadata, Libris XL, and the Libris Data Model* [Slides]. https://www.kb.se/download/18.2dd50321183eebadf405aac/1670567494690/4%20Librisdagen%202022%20Melinda-LKD.pdf

I2 – Suominen, O. (2017). *Finnish National Bibliography Fennica as Linked Data* [Slides]. https://swib.org/swib17/slides/suominen_fennica.pdf

I3 – Suominen, O. (2018, December 5). *Finlands nationalbibliografi som länkad data* [Slides]. Speaker Deck. https://speakerdeck.com/osma/finlands-nationalbibliografi-som-lankad-data

I4 – Suominen, O., & Hyvönen, N. (2016). *From MARC silos to Linked Data silos?* [Slides]. https://swib.org/swib16/slides/suominen_silos.pdf

I5 – Frosterus, M., & Saarikko, J. (2023, October 6). *The road to the future of Finnish library: Finnish linked library data project* [Slides]. https://www.kiwi.fi/display/LKD/LKD-

projektin+tuotokset?preview=/273090685/376406093/Seoul-2023-10-06-Saarikko-Frosterus-update2.pdf

I6 - Frosterus, M. (2023). *Adapting the BIBFRAME data model to our needs: challenges encountered and lessons learned* [Slides]. https://www.bfwe.eu/attachments/bfwe23-frosterus.pdf

Articles (J)

J1 – Frosterus, M. (2022). The Garden of Finnish Linked Data Springing to Life. *IFLA Metadata Newsletter*, 8(1). https://repository.ifla.org/bitstream/123456789/1976/1/IFLA_Metadata_Newsletter_June_2022.pdf

J2 – Hormia-Poutanen, K., Kivinen, M.-L., Hyvönen, N., Virtanen, M., & Suominen, O. (2016). Renewing the Library System Environment in Finland – What We Have and What We Need. *Bibliothek Forschung Und Praxis*, 40(3). https://doi.org/10.1515/bfp-2016-0066

J3 – Suominen, O. (2023, November). Finnish National Bibliography Fennica as Linked Data. *Europeana Pro*. https://pro.europeana.eu/page/issue-9-swib

Websites (K)

K1 – National Library of Finland. (2024). *Action plan of the Linked Library Data Project for 2022 - Linkitetty kirjastodata*. https://www.kiwi.fi/display/LKD/Action+plans+of+the+Linked+Library+Data+Project#:~:text=The%20project%20will%20last%20until,database%20conversion%20will%20be%20piloted.

K2 – National Library of Finland. (2024b). *Linked Library Data project (LKD) - Linkitetty kirjastodata -projekti*. https://www.kiwi.fi/pages/viewpage.action?pageId=298647995

K3 – Suominen, O. (2019, April 5). *Fennica RDF data model - Data catalog*. https://www.kiwi.fi/display/Datacatalog/Fennica+RDF+data+model

Appendix C: Interview Guide

- When and why was the decision made to start the LOD project?
    - Who was involved in this decision?
- Which departments and how many people were involved in the project?
- Have you been in contact with other libraries about LOD?
- What steps have been taken to transform the existing data into LOD?
    - Did you conduct any data clean up?
    - What technologies did you use for the transformation?
    - What data model did you use?
    - What vocabularies and ontologies did you use?
    - How did the mapping process go?
    - What data was linked during the transformation process?
    - Did you enrich the data further, e.g., with DBpedia?
    - How, if at all, have you extracted works from the data?
- What challenges arose during implementation?
- What solutions were used to overcome the challenges?
- How can people access the data?
- What future steps are planned?
- Finishing up, what is one thing you would advise other libraries to do when starting their LOD project?

Only concerning the National Library of Sweden:

- How was the new cataloguing system set up?
    - Were the cataloguers involved in the process, and if so, how?
    - What challenges arose?
    - Which solutions were used to overcome the challenges?
    - Have you received feedback from the cataloguers? Are they satisfied?
- Do you think some of the challenges wouldn't have arisen if you hadn't started the project so early?