NAISC-L: A Linked Data Interlinking Framework for Libraries, Archives and Museums

A thesis submitted to the
University of Dublin, Trinity College
in fulfilment of the requirements of the degree of
Doctor of Philosophy

Lucy McKenna
ADAPT Centre
School of Computer Science and Statistics
Trinity College Dublin
Ireland

Supervised by Prof. Declan O'Sullivan
Co-supervised by Dr. Christophe Debruyne

November 2020

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Signed: ___________________________ Date: _________________

Lucy McKenna

Acknowledgements

I would like to express my sincere gratitude to Prof. Declan O'Sullivan and Dr. Christophe Debruyne for the invaluable guidance and supervision that they afforded me throughout my PhD.

To my colleagues in the ADAPT Centre, especially Dr. Ademar Crotti Junior, thank you for your advice and encouragement.

To my parents, Catherine and Paul, sisters, Ruth and Katie, extended family and friends – thank you for your unwavering support throughout these years.

Abstract

This thesis presents a framework for **N**ovel **A**uthoritative **I**nterlinking for **S**emantic Web **C**ataloguing in **L**ibraries – or NAISC-L (pronounced noshk-el).

The Semantic Web (SW) is an extension of the current Web where data is given well defined meaning and where the relationships between data are defined in a common machine-readable format (Berners-Lee, Hendler, & Lassila, 2001). Linked Data (LD) describes a set of principles for publishing and engaging with data on the Semantic Web (Berners-Lee, 2006). A LD dataset is structured information encoded using the Resource Description Framework in which resources are identified using HTTP URIs. A LD dataset must contain links to related data (Berners-Lee, 2006) with the view of enhancing the knowledge associated with a specific entity (Papaleo, Pernelle, Saïs & Dumont, 2014).

Within the Library, Archive and Museum (LAM) domain, LD interlinking could be used to enrich data and to improve data discoverability. However, upon reviewing of some of the leading LD projects in LAMs, there was a notable lack of interlinks created for purposes beyond authority control. Interlinking could also be used by LAMs to provide additional information and context for a resource. Literature suggests this lack of interlinking is, in part, due to the technical complexity of available LD tooling. There is also a lack of interlinking frameworks that support the creation of relationship links i.e. links to related resources that are not identical to the entity being interlinked.

Using the results of a LD questionnaire distributed to LAMs, a set of requirements for the development of a LD interlinking framework was defined. The proposed framework, NAISC-L, is comprised of an interlinking process and an accompanying tool. The interlinking process is cyclical in nature and consists of four stages – Entity Selection, Link-Type Selection, Provenance Data Generation, and RDF Graph Generation and Visualisation. The tool consists of a graphical user interface, a knowledge organisation approach which provides a structure for the linking of interlink and provenance data, and a provenance data model.

NAISC-L was evaluated, in terms of its effectiveness, efficiency and user satisfaction, via three usability tests – a think-aloud test, an interlink creation test and a field test. The participants of these evaluations were Information Professionals, including librarians, archivists and metadata cataloguers. Across all experiments, participants achieved a high level of accuracy when interlinking entities, and usability and utility measures indicated positive scores for efficiency and user satisfaction. Overall, NAISC-L was shown to be an effective framework for facilitating the creation of LD interlinks in the LAM domain.

This research has yielded one major contribution, **the design, development and evaluation of the NAISC-L interlink framework**, and two minor contributions. The first minor contribution is **the provision of a report on the current state of LD in the LAM domain,** and the second is **NaiscProv. NaiscProv is an extension of the PROV Ontology, the W3C recommended standard for describing provenance data, which was developed to provide provenance descriptions for LD interlinks.**

Table of Contents

List of Figures

List of Tables

Abbreviations

AAT	Art and Architectural Thesaurus
AIMQ	AIM Quality
AVG	Average
BNB	British National Bibliography
BNF	French National Library (Bibliothèque Nationale de France)
BNE	Spanish National Library (Biblioteca Nacional De España)
CSUQ	Computer System Usability Questionnaire
DNB	German National Library (Deutsche National Bibliotek)
DQ	Data Quality
EAD	Encoded Archival Description
EDM	Europeana Data Model
EES	Effectiveness, Efficiency & Satisfaction
FAST	Faceted Application of Subject Terminology
FRBR	Functional Requirements for Bibliographic Records
GND	Gemeinsame Normdatei
ICT	Interlink Creation Test
IFLA	International Federation of Library Associations and Institutions
IG	Interlinking Guide
ISNI	International Standard Name Identifier
IP	Information Professional
ISO	International Organization for Standardisation
LAM	Library, Archive and Museum
LCSH	Library of Congress Subject Headings
LCNAF	Library of Congress Name Authority File
LD	Linked Data
LDexp	Linked Data Experience
LOC	Library of Congress
LOD	Linked Open Data
LOV	Linked Open Vocabularies
MADS	Metadata Authority Description Schema
MODS	Metadata Object Description Schema
NIG	No Interlinking Guide
NoLDExp	No Linked Data Experience

OWL	Web Ontology Language
PROV-O	Prov Ontology
PSSUQ	Post Study System Usability Questionnaire
RDA	Resource Description and Access
RDAU	Resource Description and Access Unconstrained
RDF	Resource Description Framework
RO	Research Objective
SD	Standard Deviation
SEO	Search Engine Optimisation
SUDOC	French Academic Union Catalogue
SUS	System Usability Scale
SW	Semantic Web
TAT	Think-Aloud Test
TGM	Thesaurus of Graphic Materials
TGN	Thesaurus of Geographic Names
UCD	University College Dublin
UKeiG	UK e-Information Group
ULAN	Union List of Artist Names
UNLV	University of Nevada, Las Vegas
URI	Uniform Resource Identifier
W3C	World Wide Web Consortium
VIAF	Virtual International Authority File
XSLT	Extensible Stylesheet Language Transformations
YSO	General Finnish Ontology

1 Introduction

1.1 Motivation

The Semantic Web (SW) is an extension of the current Web where data is given well defined meaning and where the relationships between data, and not just documents, are defined in a common machine-readable format – creating a Web of Data (Berners-Lee, Hendler, & Lassila, 2001). The efforts of the Semantic Web have led to various standardised technologies for representing, storing, querying, and reasoning over information. Linked Data (LD) describes a set of principles and best practices for publishing, interlinking and engaging with data on the Semantic Web (Berners-Lee, 2006). These principles include the use of HTTP Uniform Resource Identifiers (URIs) for naming resources and for retrieving data using the existing HTTP stack. A LD dataset is structured information encoded using the Resource Description Framework (RDF), the recommended model for representing and exchanging LD (Brickley & Guha, 2014). RDF statements take the form of subject-predicate-object triples, which can be organised in graphs. The subject of a triple must be a resource and that resource may be identified with a URI i.e. a named resource. When a resource does not have a URI, it is called a 'blank node'. The predicate of a triple must be a named resource. The object of a triple can be a named resource, a blank node or a literal. The use of URIs allow both human and computer-based agents to access information about these resources. SPARQL is an RDF query language that allows for the retrieval and manipulation of data stored in RDF format via a SPARQL endpoint.

LD provides a 'protocol' on how to engage with structured data on the Web. Open Data is an initiative for making data freely available on the Web. LD that is published under an open license (e.g., Creative Commons or Open Database Licenses) is known as Linked Open Data (LOD) (Berners-Lee, 2006). However, not all LD is meant to be open, for instance, LD principles can also be applied behind a firewall to facilitate interoperability within an organisation (Denaux, Ren, Villazon-Terrazas, Alexopoulos, Faraotti & Wu, 2017). LOD is classified according to a Five Star rating system, and the requirements for achieving each star can be seen in Figure 1 below. It can be seen that Open Data and LD ‘converge’ when Open Data initiatives avail of LD technologies and principles.

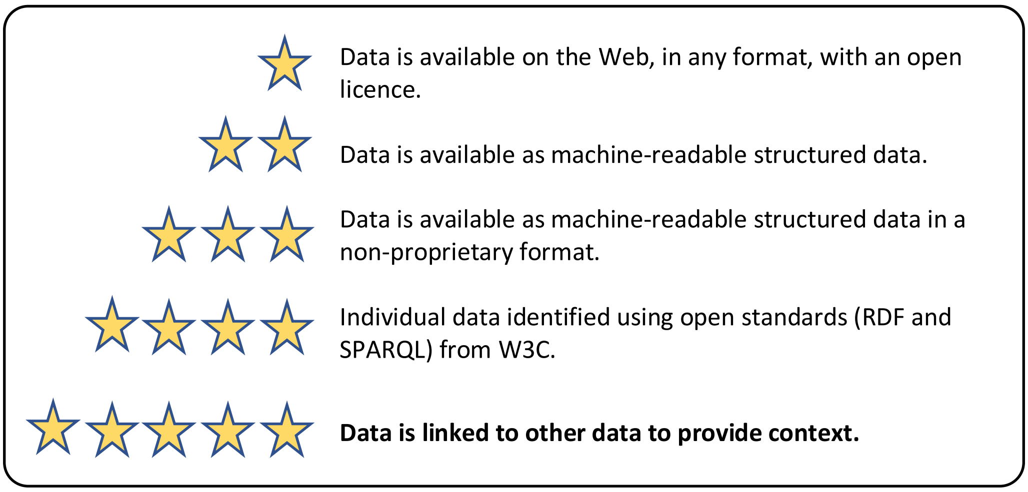


Figure 1: Five Star Linked Open Data (Berners-Lee, 2006)

In order to be considered Five Star, a LD dataset must contain links to related data (Berners-Lee, 2006; Kim & Hausenblas, 2015). The purpose of these LD interlinks is to enhance the knowledge associated with a specific entity or Thing, such as a person, place, concept, work or object (Papaleo, Pernelle, Saïs & Dumont, 2014). These links have the potential to transform the Web into a globally linked and searchable database, rather than a disparate collection of documents (W3C, 2015). This would allow for easier data querying and discovery, as well as the development of novel applications built on top of the Web.

With the Web being one of the first places where people search for information, the Library, Archive and Museum (LAM) domain would greatly benefit from publishing their metadata as LD. LD has the capability to open up and share LAM resources on the Web in ways that were previously restricted by metadata models (Gonzales, 2014). As many of the metadata standards employed by LAMs cannot be processed by Web search engines, a significant amount of relevant content is not visible in Web search results (Guerrini & Possemato, 2016; Pesch & Miller, 2016). Data published as RDF, however, is easily processed by SW search engines (Schilling, 2012) – enhancing data discoverability and visibility. For LAMs, this would not only make it easier for LAM users to find useful information, but it would also provide the opportunity to reach individuals who would not typically use LAM resources (Fons, 2016).

Additionally, as a number of different metadata standards are currently being used across LAMs, data interoperability is extremely challenging (Alemu, Stevens, Ross & Chandler, 2012). Publishing metadata as RDF would allow for seamless sharing and re-use of data across institutions – increasing collaboration and reducing record duplication (Hastings, 2015; Seeman & Goddard, 2015). LD interlinking could also be used by LAMs to aid users in discovering additional information related to their data search and also to provide seamless navigation between internal and external datasets (Alemu et al., 2012; Coyle, 2013; Seeman & Goddard, 2015).

Though the number of LAMs publishing LD is growing, uptake is still relatively slow. This is due to the challenges faced by LAMs when using LD, including:

- Current cataloguing software used by LAMs does not support LD (Cole, Han, Weathers, & Joyner, 2013; Hallo, Luján Mora, & Trujillo, 2014; Mitchell, 2016).
- Steep technical learning curve and complex LD software (Deliot, Wilson, Costabello, & Vandenbussche, 2017; Martin & Clegg, 2012; Smith-Yoshimura, 2018).
- Relatively few projects that demonstrate how LD can benefit LAMs, as well as a lack of implementation guidelines (Hastings, 2015; Mitchell, 2016).
- Financial constraints and a lack of resources (Martin & Clegg, 2012; Smith-Yoshimura, 2018).
- Transforming existing records to RDF (Schilling, 2012) and ontology selection (Smith-Yoshimura, 2018).
- Difficulty establishing interlinks (Smith-Yoshimura, 2018).
- Copyright and intellectual property issues (Schilling, 2012).

There also appears to be an issue whereby LAMs are reluctant to invest time and resources on LD projects without clear signs of success from other institutions and without having a variety of RDF datasets to interlink with (Neish, 2015).

In terms of interlinking, the most common form of interlinks on the SW are **identity links** i.e. links that point to identical entities across datasets (e.g. same as). Upon reviewing prominent existing LAM LD services, see Section 3.2.2, it can be seen that the majority of interlinks are identity links that are used for authority control purposes. Though useful, facilitating the creation of interlinks beyond identity links would be important for data enrichment purposes. Links to related, but not necessarily identical, resources are known as **relationship links** (e.g. similar to, related to, associated with). Relationship links could be used by LAMs to provide additional information and context for a given entity, thus enriching data searches.

As one of the fundamental prerequisites of the SW is the existence of large amounts of meaningfully interlinked resources (Bizer, Heath & Berners-Lee, 2009), it is key that LAMs not only create identity links, but also relationship links. As the full potential of LD interlinking has yet to be realised within the LAM domain, there is a need to explore how Information Professionals (IPs) can be facilitated to create LD interlinks beyond those used for authority control.

1.2 Research Question

The research question investigated in this thesis is:

To what extent can NAISC-L, a domain-specific interlinking framework, facilitate Information Professionals to engage with the process of Linked Data interlinking with effectiveness, efficiency and satisfaction?

The terms used in this research question are defined as follows:

- An *Information Professional* (IP) is a metadata expert working in the LAM domain.
- *Engage* refers to the ability to create, edit and interpret LD interlinks.
- In the context of this research, *Linked Data Interlinking* is the process of creating a link between related or identical entities across LD datasets[1].
- *Effectiveness* is the degree of accuracy as to which users can create LD interlinks.
- *Efficiency* is the time taken to create an interlink.
- S*atisfaction* is the extent to which NAISC-L meets the users' needs and expectations.

[1] In the wider Linked Data context, interlinking can also refer to linking entities within the same dataset.

1.2.1 Research Objectives

In order to address the research question defined above, the following research objectives (RO) were identified:

RO1: Perform a state-of-the-art review of existing LD interlinking frameworks and tools.

RO2: Explore the benefits and challenges of using LD as experienced by IPs.

RO3: Propose a LD interlinking framework for the LAM domain.

RO4: Apply, implement and evaluate the interlinking framework in terms of its effectiveness, efficiency and satisfaction as perceived by IPs.

The proposed interlinking framework is called NAISC-L (pronounced noshk-el) which stands for **N**ovel **A**uthoritative **I**nterlinking for **S**emantic Web **C**ataloguing in **L**ibraries. The NAISC-L framework is comprised of a Linked Data (LD) interlinking process and accompanying tool.

1.2.2 Contributions

The major contribution of this thesis is the proposed interlinking framework – NAISC-L. The minor contributions include a current overview of the state of LD in the library, archive and museum (LAM) domain, and NaiscProv – an extension of the PROV Ontology (Belhajjame, Cheney, Corsar, Garijo, Soiland-Reyes, Zednik & Zhao, 2013) for capturing interlink provenance.

Major

The major contribution of this thesis is the development and demonstration of the interlinking framework, NAISC-L. Unlike existing interlinking frameworks and technologies, NAISC-L was designed specifically with the needs and work processes of the LAM domain in mind. As mentioned, NAISC-L is comprised of an interlinking process and a tool. The interlinking process is cyclical in nature and consists of four stages – Entity Selection, Link-Type Selection, Provenance Data Generation, and RDF Graph Generation and Visualisation. The tool consists of a graphical user interface (GUI), an approach to knowledge organisation, and a data model. The GUI is an instantiation of the interlinking process developed for use by Information Professionals (IPs) in the LAM domain. NAISC-L's knowledge organisation approach is a graphical structure detailing how interlink data is associated with its corresponding provenance data using a series of named graphs. The data model, known as NaiscProv, is used to capture the provenance data for each interlink created.

NAISC-L was developed specifically for the LAM domain with the aim of facilitating IPs to engage with the process of LD interlinking with efficacy and ease. NAISC-L supports the creation of both identity links and relationship links in order to provide a means for LAMs to enrich their LD with a greater variety of interlinks to a broader range of sources. A video demo of NAISC-L can be viewed on this webpage[2], and its code can be found on Gogs[3].

Publications associated with this contribution are:

- **McKenna, L., Debruyne, C., & O'Sullivan, D. (2019). NAISC: An Authoritative Linked Data Interlinking Approach for the Library Domain. In *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)* (pp 11-20). https://doi.org/10.1109/JCDL.2019.00012.**
  This publication describes the NAISC-L framework in detail. Also outlined in this paper are the results of the first phase of user-testing of NAISC-L.
- **McKenna, L. (2017). Engaging librarians in the process of interlinking RDF resources. In *European Semantic Web Conference (ESWC)* (pp. 216-225). https://doi.org/10.1007/978-3-319-58451-5_16.**
  This paper was published as part of a PhD Symposium in which the research question, objectives, motivation and evaluation plan of this thesis were discussed.

Minor

The first minor contribution of this thesis is the provision of a report on the current state of LD in the LAM domain. This report is based on a state-of-the-art review of LD projects in LAMs, as well as the results of a large, international survey of the use of LD in LAMs. The survey identified a number of LD challenges that are being experienced by LAMs. This information could be used

[2] https://www.scss.tcd.ie/~mckennl3/naisc/ accessed 16th August 2020
[3] https://gogs.adaptcentre.ie/mckennl3/NAISC accessed 16th August 2020. Access must be granted prior to viewing NAISC-L code – please email author for access.

by researchers and developers in order to address and provide potential solutions to these issues. The survey also identified the types of systems, datasets and metadata schemas most commonly used by LAMs, information which could also be used in the development of future LAM tooling.

Publications associated with this contribution are:

- **McKenna, L., Debruyne, C., & O'Sullivan, D. (2018). Understanding the Position of Information Professionals with regards to Linked Data: A survey of Libraries, Archives and Museums. In *2018 ACM/IEEE on Joint Conference on Digital Libraries (JCDL)* (pp. 7-16). https://doi.org/10.1145/3197026.3197041.**

  This publication discusses the results of a large-scale international survey of IPs regarding their views on the benefits and challenges of using LD in the library, archive and museum domain.
- **Debattista, J., McKenna, L., & Brennan, R. (2018). Understanding Information Professionals: A Survey on the Quality of Linked Data Sources for Digital Libraries. In *2018 Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE)* (pp. 537-545). https://doi.org/10.1007/978-3-030-02671-4_32.**

  This paper discusses the results of a survey of IPs in which they were asked to select the evaluation criteria they apply when using and searching for external data sources, as well as the common data quality issues they encounter.

The second minor contribution of this thesis is NaiscProv – an extension of the PROV Ontology (PROV-O) which is the W3C-recommended standard for describing provenance data (Belhajjme et al., 2013). NaiscProv was developed in response to a specific problem related to data trustworthiness. During the state-of-the-art and requirements gathering stages of this research, it was noted that the provision of provenance data was an important factor for IPs when making decisions regarding the trustworthiness and authoritativeness of a dataset. As such, the provenance of the interlinking process was captured using NaiscProv with the view that IPs, and others, could use this data in future in order to make authoritative decisions regarding the credibility of the data generated.

Publications associated with this contribution are:

- **McKenna, L., Debruyne, C., & O'Sullivan, D. (2019). Modelling the Provenance of Linked Data Interlinks for the Library Domain. In *Companion Proceedings of the 2019 World Wide Web Conference (WWW)* (pp. 954-958). https://doi.org/10.1145/3308560.3316518.**

  This publication describes the development and use of the NAISC-L Provenance Model as well as the PROV Ontology expansion, NaiscProv.

1.3 Research Overview

This section provides an overview of the research approach taken in the investigation of the thesis, the methods applied to achieve the research objectives, and the strategy for evaluating the research output.

1.3.1 Research Approach

A Design Science and User-Centred Design Approach were applied to the research conducted as part of this thesis.

Design Science

Design Science (DS) is defined as "a research paradigm in which a designer answers questions relevant to human problems via the creation of innovative artefacts, thereby contributing new knowledge to the body of scientific evidence. The designed artefacts are both useful and fundamental in understanding that problem" (Hevner & Chatterjee, 2010; Wieringa, 2014). Thus, knowledge of and a solution to an identified problem are acquired through the process of designing, building and testing an artefact (Hevner, March, Park & Ram, 2004).

DS research must have three identifiable research cycles (Hevner, 2007):

1. The Relevance Cycle: Identification of the contextual environment of the research, the requirements of the artefact, and the ultimate evaluation criteria for the completed artefact.
2. The Rigor Cycle: Application of state-of-the-art research to inform the research processes and to ensure an innovative solution.
3. The Design Cycle: Iterative design, evaluation and refinement of the artefact.

The Design Science Research Methodology (DSRM) provides a process model for conducting research (Hevner et al., 2004; Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007). The steps included in the model are:

1. Problem Identification and Motivation – Define the problem and the importance of finding a solution.
2. Solution Objectives – Define what would be accomplished by the new solution.
3. Design and Development – Create the artefact.
4. Demonstration – Use the artefact to solve the problem.
5. Evaluation – Observe the efficacy and efficiency of the artefact in solving the problem.
6. Communication – Disseminate knowledge.

DS has become a much used and well-established research approach within the Information Systems domain (Goldkuhl, Ågerfalk, & Sjöström, 2017). Due to the applied nature of this thesis, i.e. investigating, proposing and testing a framework for LD interlinking by IPs, DS was an appropriate approach for investigating the research question. In line with this approach, the research question of this thesis was explored in iterative cycles of requirements gathering, artefact design, development and evaluation.

User-Centred Design

User-Centred Design is the process of designing a tool in view of how it will be understood and used by users, thus placing the user in the centre of the design process (Lowdermilk, 2013; Usability First, 2015). The principles that underline user-centred design, as per the International Usability Standard International Organization for Standardisation (ISO) 9241-210 (Travis, 2011), include:

1. The design is based upon an understanding of users, their requirements, their environment and the tasks they complete.
2. Users are involved in all phases of tool design and development.
3. The tool is evaluated by users and refinements are made based on these results.
4. The design and evaluation processes are iterative in nature.
5. Tools are designed with a holistic user experience in mind.
6. The design team includes multidisciplinary skills and perspectives.

Given that the objectives of this thesis were to propose, implement and evaluate a LD interlinking approach for the LAM domain, a user-centred approach was taken in order to ensure that the perspectives of IPs were considered in achieving these aims.

User-Centred Design was combined with DS by including IPs in every DS research cycle. As part of the Relevance Cycle, a questionnaire was distributed to IPs in order to collate a set of requirements for a LD interlinking tool for LAMs (see Section 3.3). As part of the Rigor Cycle, interlinking tools, LD projects in LAMs and research exploring IPs views on LD were reviewed as part of the State-of-the-Art research (see Section 3.1 and Section 3.2). Finally, as part of the Design Cycle, IPs were involved in the evaluation of NAISC-L and their feedback was used to refine future iterations of the framework and tool.

1.3.2 Technical Approach

In order to achieve RO1 and RO2, a state-of-the-art review of LD projects in LAMs and existing interlinking frameworks was initially undertaken. A survey investigating the benefits and challenges of using LD in LAMs, as perceived by IPs, was then conducted. One particular challenge identified was that of LD interlinking, and a set of requirements for facilitating interlinking in the LAM domain were distilled from the results of the survey.

Using these requirements, the first iteration of the NAISC-L framework was developed. The framework underwent four cycles of design and testing. The results of each user-evaluation were used to inform the following design iteration. Through these iterations of development, evaluation and refinement, RO3 and RO4 were achieved.

1.3.3 Evaluation Strategy

The ISO 9241-11:2018 standard defines usability as the “extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” (ISO, 2018). Three separate forms of usability testing were used to evaluate NAISC-L including:

- **Usability Test 1 – Think-Aloud Test:** Think-aloud Tests**,** a widely used method for usability testing, require participants to verbalise their thoughts and actions while interacting with a system (Haak, De Jong & Schellens, 2003). This provides data on the types of difficulties users encounter while using a system, as well as data on what users enjoy about a system. In order to evaluate NAISC-L, fifteen IPs were asked to think-aloud while carrying out a set of six pre-defined interlinking tasks. After completing these tasks, participants took part in a post-test interview that explored their experience of using NAISC-L. They were also asked to complete the Post Study System Usability Questionnaire (PSSUQ) (Lewis, 1992; 2002) – a 19-item instrument used to quantitively assess the usability of a system. The results of the PSSUQ can be viewed in four categories – System Usefulness (SysUse: Items 1-8), Information Quality (InfoQual: Items 9-15), Interface Quality (InterQual: Items 16-18), and Overall (Items 1-19).

- **Usability Test 2 – Online Usability Test**: This user-evaluation was conducted online by 95 IPs and required participants to complete a set of three pre-defined set of tasks using NAISC-L. After finishing these tasks, users were asked to complete the Post Study System Usability Questionnaire (PSSUQ) as well as a data quality questionnaire which was based on the AIMQ methodology (Lee, Strong, Kahn & Wang, 2002).

- **Usability Test 3 – Field Test**: Over one working week, three IPs from the Irish Traditional Music Archive (ITMA) used NAISC-L for a short period each day in order to create a set of thirty interlinks. After completing the field test, the participants took part in a post-test interview which explored their experience of using NAISC-L. They were also asked to complete the Post Study System Usability Questionnaire (PSSUQ) as a quantitative usability and utility measure.

Effectiveness, efficiency and satisfaction (EES) were measured in each experiment using a variety of tools outlined in Table 1 below. A description of these tools and how they were used to evaluate EES is presented in Chapter 5.

Table 1: Experiment Effectiveness, Efficiency and Satisfaction Measures[4]

Measure	Usability Test 1 Think-Aloud Test	Usability Test 2 Online Usability Test	Usability Test 3 Field Test
Effectiveness	Number of Interlinks Completed	Number of Interlinks Completed	
	Interlink Accuracy	Interlink Accuracy	Interlink Accuracy
	PSSUQ: SysUse	PSSUQ: SysUse	CSUQ: SysUse
		Data Quality Questionnaire	
Efficiency	Interlink creation time	Interlink creation time	
	PSSUQ: SysUse	PSSUQ: SysUse	CSUQ: SysUse
Satisfaction	PSSUQ: InterQual	PSSUQ: InterQual	CSUQ: InterQual
	PSSUQ: Overall	PSSUQ: Overall	CSUQ: Overall
	Think-Aloud Test	PSSUQ: InfoQual	Field Test Diary
	Post-Test Interview		Post-Test Interview

1.4 Thesis Overview

The remainder of this thesis is structured as follows:

Chapter 2: Background

This chapter provides useful preliminary information for readers of this thesis. It begins with information about the Semantic Web (SW), LD, interlinking and LD provenance. It then describes the use of Linked Data and data provenance in the library domain.

[4] PSSUQ – Post-Study System Usability Questionnaire
CSUQ – Computer System Usability Questionnaire

Chapter 3: State of the Art

This chapter provides an analysis of existing LD interlinking frameworks and tools, as well as existing LD provenance ontologies and methods. This is followed by a review of LD projects in the library, archive and museum domain.

Chapter 4: The NAISC-L Framework

This chapter first describes the questionnaire which was distributed to LAMs as a means of gathering a set of requirements for LD interlinking. This chapter then describes the NAISC-L framework in detail.

Chapter 5: Evaluation

This chapter describes the methods used to evaluate NAISC-L including a think-aloud observation, an online usability test and a field test.

Chapter 6: Conclusion

This chapter presents the key findings of the research described in this thesis. It discusses to what extent the research question has been answered and the research objectives have been met. Possible directions for future work related to the research in this thesis are also outlined.

2 Background

This chapter presents background information relating to the research of this thesis, particularly LD interlinking and LD provenance. There is an assumption that the reader is familiar with the basic concepts of the SW and the principles of LD, as discussed in Section 1.1. Section 2.1 presents an introduction to LD interlinking, Section 2.2 describes LD provenance and Section 2.3 provides a summary of this chapter.

2.1 Linked Data Interlinking

LD interlinking describes the task of determining whether a named resource (an entity identified by a URI) can be linked to another named resource in order to indicate that they both describe the same thing or that they are related in some capacity (Ferrara, Nikolov & Scharffe, 2011). Interlinks can be created within or across datasets, **however, in the context of this thesis, LD interlinking specifically refers to cross-dataset links**.

As mentioned in Section 1.1, in order to achieve Five Star LOD, a LD dataset must contain links to related data (Berners-Lee, 2006). The purpose of LD interlinks is to provide additional information about an entity in order to improve data discovery (Kim & Hausenblas, 2015). Through the process of interlinking, references can be created from entities in one dataset to entities in another, which in turn may be interlinked to further entities – creating links to a potentially infinite network of data (Heath & Bizer, 2011).

LD interlinks are also known as typed links and the linking property used to describe the relationship between two URIs is known as a link-type (Neubauer, 2017). Identity Links are a specific kind of typed-link where the subject and object URI refer to the same entity (Papaleo, Pernelle, Saïs & Dumont, 2014). Identity links are typically expressed using the `owl:sameAs` property, from the Web Ontology Language[5] (OWL), and the process of creating these links is referred to as *instance matching*. `owl:sameAs` links are the most common type of cross-dataset interlink on the SW (Paris, Hamdi & Cherfi, 2019).

[5] https://www.w3.org/TR/owl-overview/ accessed 27th July 2020

In the absence of a central naming authority on the SW, different datasets often refer to the same entity using different URIs which can then be linked using instance matching (Arioua, Croitoru, Papaleo, Pernelle & Rocher, 2016). Such identity links are often used for the purpose of authority control in LAMs.
Figure 2 below demonstrates the use of instance matching between the James Joyce entity in the French National Library[6] (BNF) to the James Joyce entities in the Virtual International Authority File[7] (VIAF) and DBpedia[8], which both have further interlinks to IdRef[9], the German National Library[10] (DNB) and the Spanish National Library[11] (BNE).

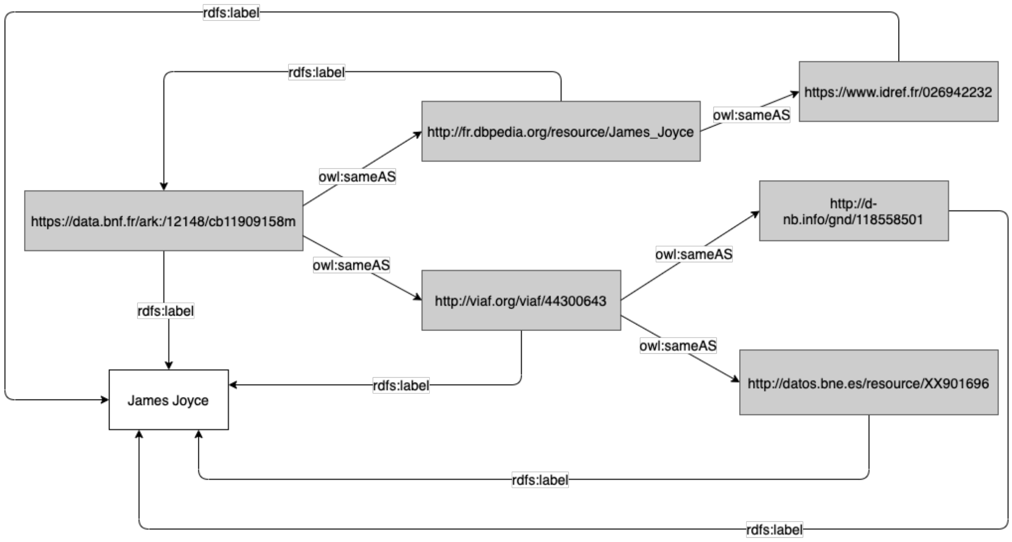


Figure 2: Identity Link Example

The `owl:sameAs` property has strict semantics and should only be used where two things are identical, share the same properties and where all the statements for one entity are true for the other (McGuinness & van Harmelen 2004). Although identity links are the most common type of interlink across LD datasets, the strict semantics of `owl:sameAs` are not always followed leading to the inference of inaccurate data and reducing data quality (Jaffri, Glaser & Millard, 2008; Halpin, Hayes, McCusker, McGuinness & Thompson, 2010; Paris, 2018; Raad, Beek, van Harmelen, Pernelle & Saïs, 2018). These

[6] https://data.bnf.fr/ accessed 9th August 2020
[7] http://viaf.org/ accessed 9th August 2020
[8] https://wiki.dbpedia.org/ accessed 9th August 2020
[9] https://www.idref.fr/autorites.jsp accessed 9th August 2020
[10] https://www.dnb.de/DE/Home/home_node.html accessed 9th August 2020
[11] http://datos.bne.es/inicio.html accessed 9th August 2020

inaccuracies could be reduced by using link-types that express weaker relationships between entities.

Relationship Links are another kind of typed link which are used to point to related entities in other datasets (Heath & Bizer, 2011). Unlike identity links, relationship links do not have to point to exactly the same thing and can be used to provide background knowledge and context for an entity. For example, Figure 3 demonstrates a set of relationship links from the James Joyce entity in the BNF to the James Joyce Wikipedia[12] page and a collection of James Joyce materials held in the Digital Library of University College Dublin[13] (UCD).

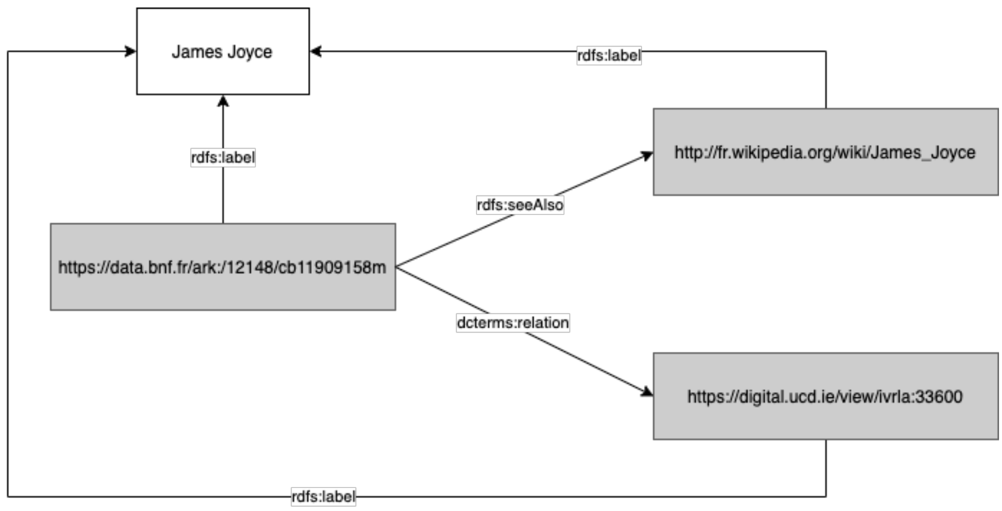


Figure 3: Relationship Link Example

Finally, vocabulary links are used to map the vocabulary of one schema, or ontology, to another. Vocabulary links enable LD applications to understand and integrate data described using different schemas (Heath & Bizer, 2011). Also known as ontology mapping, this form of interlinking is beyond the scope of this thesis.

[12] https://www.wikipedia.org/ accessed 9th August 2020
[13] https://digital.ucd.ie/ accessed 9th August 2020

2.2 Linked Data Provenance

Data provenance is a record describing the origin of a piece of data (Gupta, 2009), and may include information on the date/time, people, institutions, and processes involved in its creation. Provenance statements are a form of contextual metadata which can, in turn, have their own provenance record. Given that any individual can publish to the SW, LD provenance is crucial in establishing the trustworthiness and quality of the data (Dezani-Ciancaglini, Horne & Sassone, 2012). For instance, searching for a specific entity may result in multiple URIs from different sources – provenance information regarding the origin of the data and how it became available would be useful in deciding which URI to follow or to interlink to (Hartig & Zhao, 2010). In the LAM domain, The Open Archival Information System (OAIS) (CCSDS, 2019) and Preservation Metadata: Implementation Strategies (PREMIS) (PREMIS Editorial Committee, 2015), are both widely accepted standards for digital preservation that require the provision of provenance information when archiving digital resources. As such, a NAISC-L Provenance Data Model, described in Section 4.2.2.2, was developed. The provenance ontology and approaches used in the data model are introduced below.

2.2.1 PROV Data Model

The PROV Data Model is a Web-Oriented provenance standard, developed by the W3C Provenance Working Group, for the representation and exchange of provenance information (Belhajjame et al., 2013). The PROV Ontology (PROV-O) is an OWL ontology that maps the PROV Data Model to RDF.

The core classes of the model are the Entities (physical, digital or conceptual objects or Things), the Agents (persons, organisations, software) and the Activities involved the process of creating an Entity – see Figure 4. The properties of the model describe the relationships between these classes.

There are three Agent subclasses, `prov:Person`, `prov:Organization` and `prov:SoftwareAgent`, and three Entity subclasses – `prov:Collection` (a group of entities), `prov:Plan` (a set of actions) and `prov:Bundle`. The Bundle subclass is a named set of provenance descriptions which, as it is an entity, can itself be described thus providing the provenance of the provenance data.

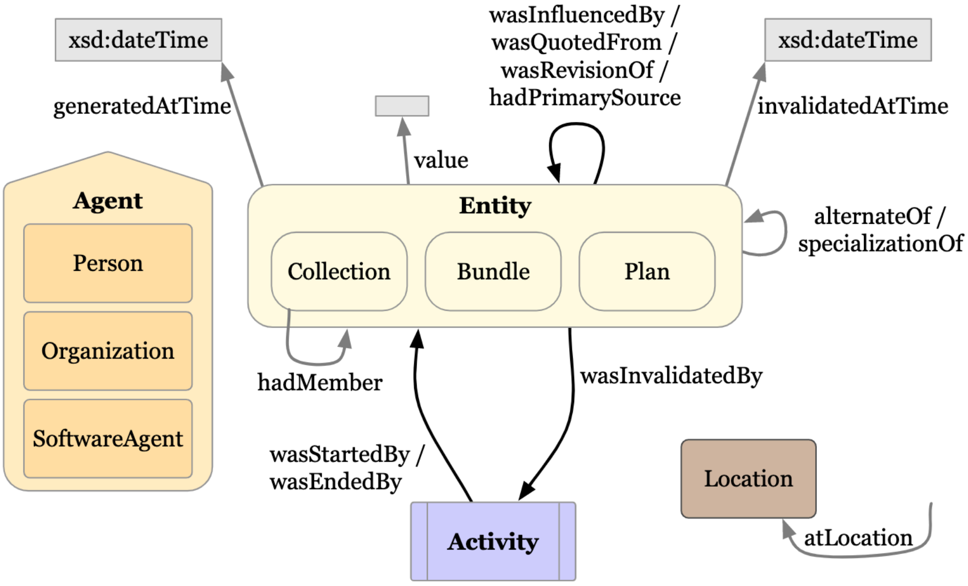


Figure 4: Prov Data Model[14]

As well as PROV-O, there are a number of other ontologies which have been developed for provenance purposes including the Provenance Vocabulary (Hartig & Zhao, 2010), the Open Provenance Model (OPM) (Moreau, Clifford, Freire, Futrelle, Gil, Groth, Kwasnikowska et al., 2011), Provenance Authoring and Versioning ontology (PAV) (Cicaresse, Soiland-Reyes, Belhajjame, Gray, Goble & Clark, 2013), and Provenir (Sahoo & Sheth, 2009). PROV-O was used as part of the NAISC-L Provenance Data Model over these aforementioned vocabularies because it is a W3C-recommended standard and because it can be easily extended for domain-specific purposes, something which was necessary for the provision of interlink provenance. PROV-O was chosen over LAM vocabularies, such as Dublin Core[15] and BIBFRAME[16], as these ontologies use free-text tags for the provision of provenance data. These tags would not allow for interlink-specific provenance and would also limit the kinds of queries which could be run over the data.

[14] https://www.w3.org/TR/prov-o/ accessed 9th August 2020
[15] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/ 10th August 2020
[16] https://www.loc.gov/bibframe/ accessed 10th August 2020

2.2.2 Provenance Approaches

The two approaches used to represent LD provenance in the NAISC-L Provenance Model were RDF Reification (Manola & Miller, 2004) and Named Graphs (Carroll, Bizer, Hayes & Stickler, 2005), both of which are detailed below. Alternative approaches, including the Singleton Property (Nguyen, Bodenreider & Sheth, 2014) and RDF* (Hartig, 2017) are also discussed.

2.2.2.1 RDF Reification

RDF Reification, part of the RDF Standard (Manola & Miller, 2004), is an instance of RDF Statement identified by a URI or declared as a blank node. As such, further statements, or meta-triples, can be about the RDF statement which is why it is a useful approach for the provision of provenance data. The RDF Reification vocabulary consists of the type `rdf:Statement`, which is given a URI, and the properties `rdf:subject`, `rdf:predicate`, and `rdf:object`.

Although reification leads to the addition of four triples for every reified statement, it was deemed useful for the NAISC-L Provenance Data Model as it allowed for the creation of meta-triples describing the origin of LD interlinks. For example, Figure 5 below demonstrates how RDF Reification could be used to describe the provenance of a relationship link from the James Joyce entity in the BNF to the James Joyce Wikipedia page (as seen in Figure 5).

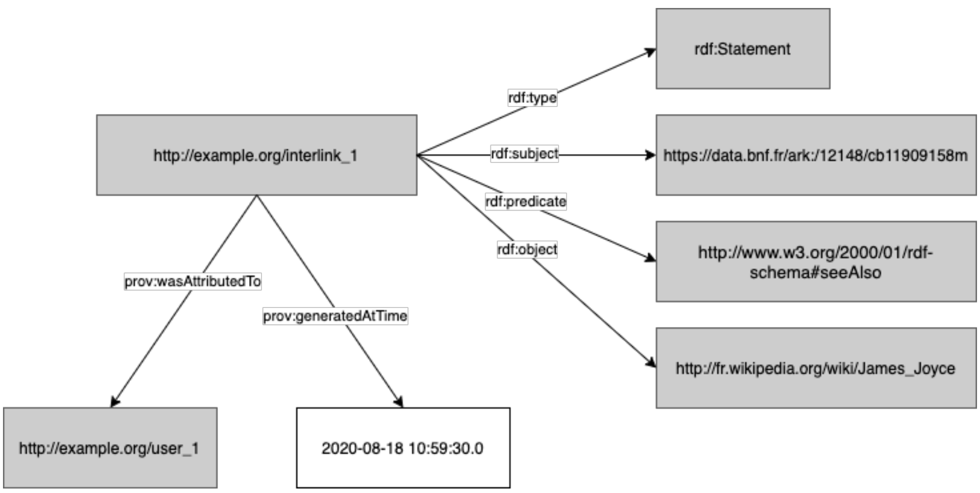


Figure 5: LD Provenance using RDF Reification

2.2.2.2 Named Graphs

A named graph is an RDF sub-graph containing a set of triples that has been assigned a unique name in the form of a URI (Carroll, Bizer, Hayes & Stickler, 2005). These collections of triples can then be published as independent units. Like RDF graphs, a named graph can contain any number of statements. Named graphs are often used in the process of provenance data generation as they allow for the assertion of statements relating to a specific set of triples in a dataset. For example, Figure 6 below portrays a named graph, `http://example.org/named_graph_James_Joyce` (shown as a dashed box), which contains a set of interlinks from the James Joyce entity in the BNF to related entities in Wikipedia and UCD. Outside the named graph, at the bottom of the figure, are two statements that describe the provenance of the named graph using its URI, or name, as the subject.

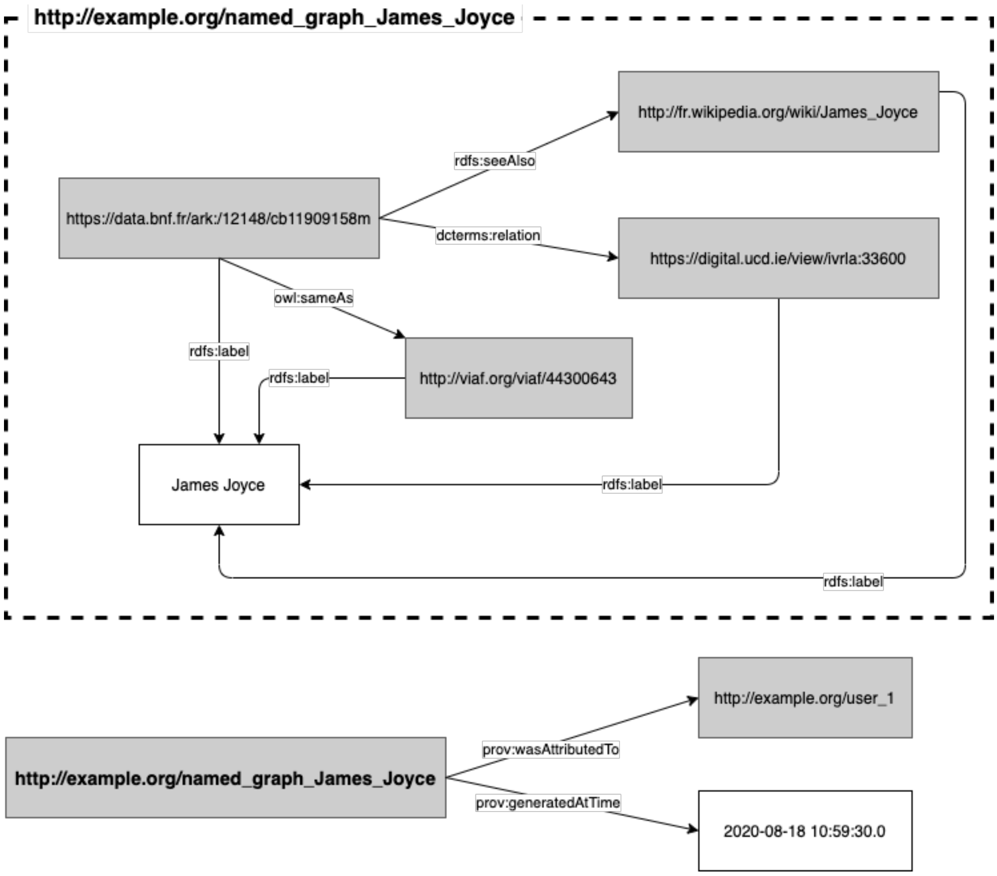


Figure 6: LD Provenance using Named Graphs

Named Graphs are represented as quads, as opposed to triples, whereby the name of the optional graph is added to the triple statements it contains e.g. <subject> <predicate> <object> <graph_name_uri>. When no graph name is provided, the statement is declared in the so-called *default* graph of the RDF dataset. RDF serialisation formats that support named graphs include TriX (Carroll & Stickler, 2004), which provides an XML representation, TriG[17] which is an extension of Turtle[18], and N-Quads[19], which is an extension of N-Triples[20].

2.2.2.3 Singleton Property

The Singleton Property involves the creation of a unique RDF resource that represents a property (Nguyen, Bodenreider & Sheth, 2014). This resource is linked to the property that it represents using `rdf:singletonPropertyOf`. The URI of the resource can be used as the predicate of a triple. Note, this URI can be arbitrarily chosen and does not have to follow a particular pattern. As the singleton property is itself a resource, statements can be added to it to provide provenance data for the triple in which it was used as a property. Figure 7 shows how a singleton property, `rdfs:seeAlso#singleton_1`, could be used to provide the provenance of an interlink from the James Joyce entity in the BNF to the James Joyce Wikipedia page.

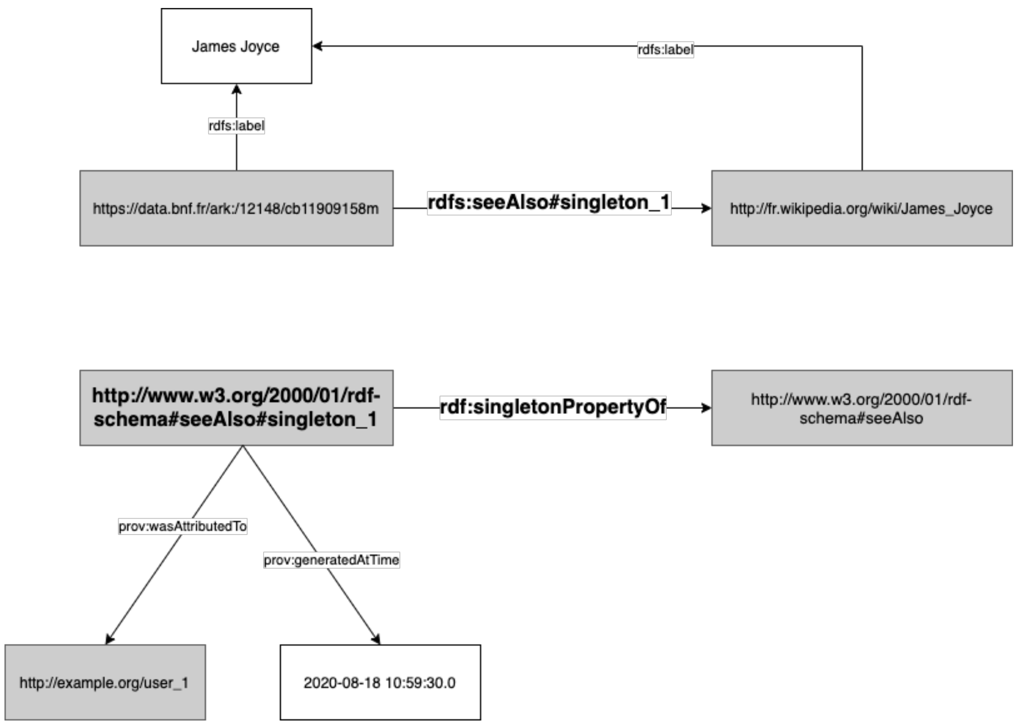


Figure 7: LD Provenance using the Singleton Property

[17] https://www.w3.org/TR/trig/ accessed 10th August 2020
[18] https://www.w3.org/TR/turtle/ accessed 10th August 2020
[19] https://www.w3.org/TR/n-quads/ accessed 10th August 2020
[20] https://www.w3.org/TR/n-triples/ accessed 10th August 2020

Although using the Singleton Property only generates one additional triple per statement, it was not used as part of the NAISC-L Framework as the provenance data is associated with the predicate, or link-type, alone. In the case of NAISC-L, the provenance data needs to be associated with the entire statement as elements of the provenance may refer to the subject or object entity.

2.2.2.4 RDF*

RDF* is an extension of RDF which uses nested triples. Essentially, a meta-triple describing a particular statement can contain this statement as its subject or object i.e. the statement is nested in the meta-triple (Hartig & Thompson, 2014; Hartig, 2017). Figure 8 below demonstrates how RDF* could be used to generate the provenance of a relationship link from the James Joyce entity in the BNF to the James Joyce Wikipedia page. Here the interlink statement is a nested triple, as represented by the orange box, about which two provenance statements, or metadata triples, have been asserted. These particular metadata triples describe by whom and when the interlink was generated.

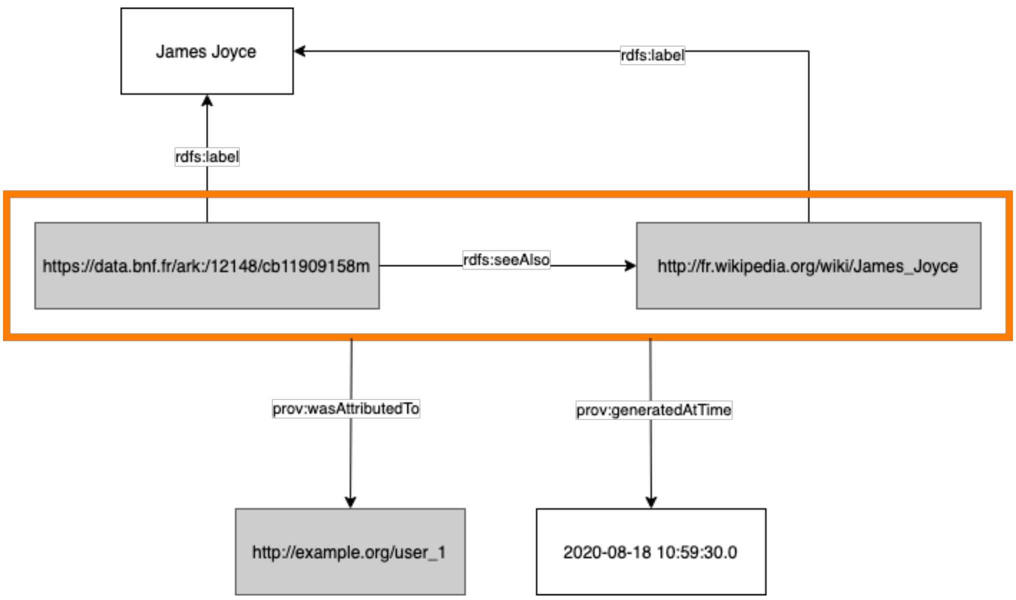


*Figure 8: LD Provenance using RDF**

Using RDF* for the provision of provenance data would not generate any additional triples per statement, providing a very concise representation of meta-triples. RDF* was not implemented as part of the NAISC-L Provenance Model as it requires the extension of RDF, Turtle and SPARQL syntaxes which were not supported by common LD technologies at the time of NAISC-L's design.

However, as RDF* is being supported by a growing number of systems (Hartig, 2019), the NAISC-L Provenance Data Model was reviewed in order to determine how RDF* could be incorporated into future work – see Section 6.3.

2.3 Chapter Summary

This chapter provided an introduction to LD Interlinking and LD provenance, particularly PROV-O, RDF Reification and Named Graphs, all of which are used as part of the NAISC-L Provenance Data Model described in Section 4.2.2.2.

3 State-of-the-Art

In line with the Rigour Cycle of the Design Science Model (Hevner, 2007), the following chapter provides a state-of-the-art review of existing LD interlinking frameworks (Section 3.1), prior research exploring the use of LD in LAMs (Section 3.2.1), and leading LAM LD projects (Section 3.2.2). Also presented is a LD Requirements Questionnaire which explored the benefits and challenges of using LD in LAMs as well as potential solutions to these challenges (Section 3.3).

3.1 Linked Data Interlinking Tools

The tools included in the state-of-the-art-review were discovered by searching Google Scholar[21], ACM Digital Library[22], ScienceDirect[23], Scopus[24], SpringlerLink[25] and IEE Xplore Digital Library[26] using the keywords 'link discovery framework', 'interlinking tool', and 'linked data interlinking'. The aforementioned databases were also directly searched for the eleven LD tools discussed in Nentwig, Hartung, Ngonga Ngomo and Rahm's (2017) survey of link discovery frameworks.

Of the many interlinking tools and frameworks discovered, those included in the review were tools developed for both relationship and identity link discovery, and tools developed exclusively for instance matching[27], i.e. creating `owl:sameAs` links between resources. Tools excluded from the review included those developed solely for ontology mapping or vocabulary alignment, i.e. creating vocabulary links, as this kind of linking is not the focus of this research. As this thesis explores the extent to which a domain-specific interlinking framework can facilitate IPs to engage with LD interlinking, the tools reviewed were further refined to include only those with a GUI as this would play an important role in user-friendliness for the IP population.

[21] https://scholar.google.com/ accessed 14th August 2020
[22] https://dl.acm.org/ accessed 14th August 2020
[23] https://www.sciencedirect.com/ accessed 14th August 2020
[24] https://www.scopus.com/ accessed 14th August 2020
[25] https://link.springer.com/ accessed 14th August 2020
[26] https://www.ieee.org/publications/xplore/ accessed 14th August 2020
[27] Also known as *entity resolution*

AgreementMaker

Although initially developed for ontology matching, the AgreementMaker system can also be used for instance matching (Cruz, Antonelli & Stroe, 2009; Cruz, Stroe, Caimi, Fabiani Pesquita, Couto & Palmonari, 2011). AgreementMaker uses a three-phase process in order to return potential pairs of matching entities. The Lookup Phase requires users to provide a SPARQL endpoint or API for both the source and target datasets. In the Disambiguation Phase, the system uses a variety of algorithms to compute the similarities between the source and target entities via feature comparison, e.g. comparing features such as `rdfs:label`, `rdfs:comment`, `dbpedia:abstract` or `skos:description`. In the final Combination Phase, the similarity values are ranked so that the system can select the best target match for a source entity. AgreementMaker supports instance matching using the `owl:sameAs` and the `skos:exactMatch` link-type. Users can interact with the system via a GUI, see Figure 9, which supports advanced visualisation techniques. The GUI also provides a control panel that users can use to select matching methods and evaluation strategies.

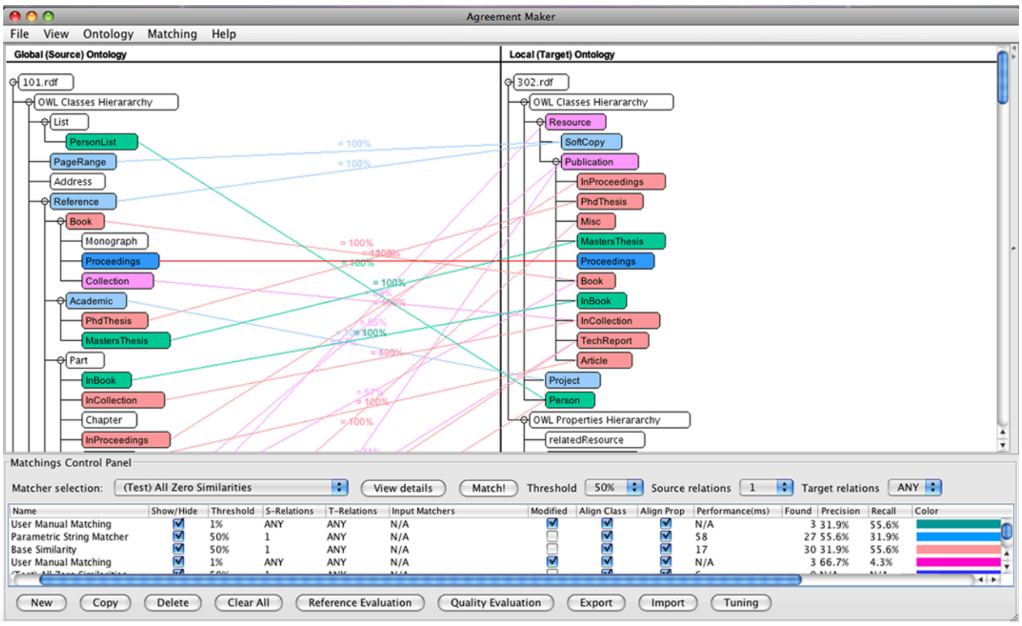


Figure 9: AgreementMaker GUI[28]

[28] Criz, Antonelli, Stroe, Keles & Maduko, 2009

LogMap

Like AgreementMaker, LogMap was initially developed for the purpose of ontology matching but was then extended for instance matching (Jiménez-Ruiz & Grau, 2011; Jiménez-Ruiz, Grau & Horrocks, 2012). LogMap uses lexical indexation, whereby the labels of entities are indexed, in order to find matches between the source and target datasets. The system can then be used to create `owl:sameAs` links between matched pairs. LogMap is open source and can be used from the command line or directly from its Web interface[29] – see Figure 10.

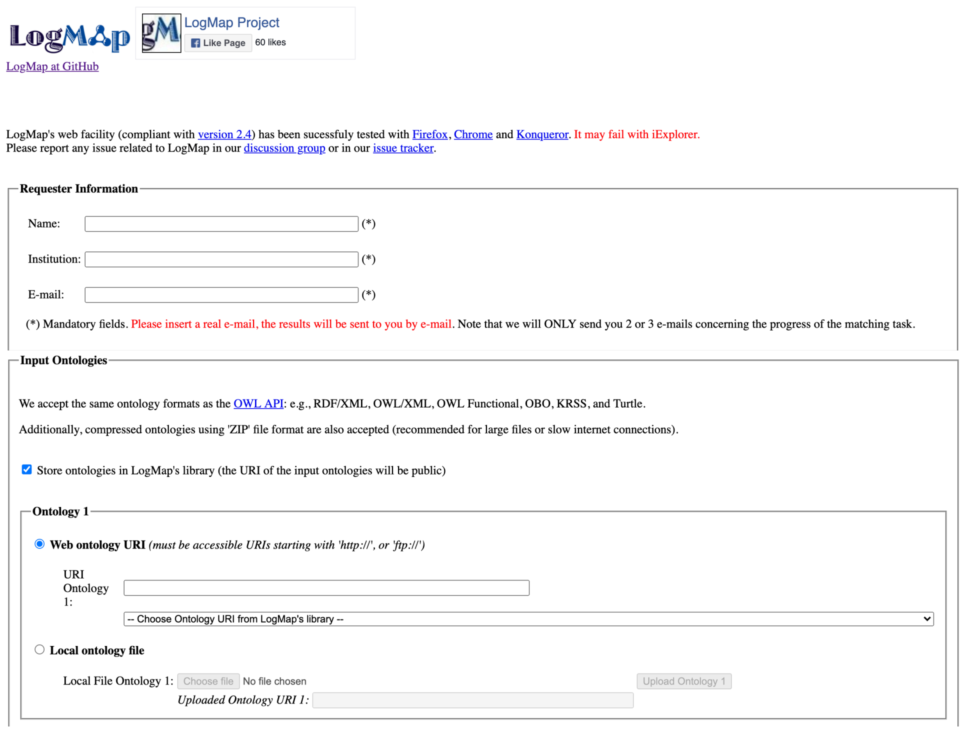


Figure 10: LogMap Web Interface

LinkItUp

LinkItUp (Hoekstra & Groth, 2013) is a link discovery tool that can be used to enrich research output published via the Figshare.com[30] repository. Using the metadata of a Figshare research article, LinkItUp searches for equivalent terms in a number of academic LD datasets including DBpedia, DBLP: Computer Science Bibliography[31], CrossRef[32], ORCID[33], the NIF Registry[34], and DANS

[29] http://krrwebtools.cs.ox.ac.uk/logmap/ accessed 29th July 2020
[30] https://figshare.com/ accessed July 29th 2020
[31] https://dblp.uni-trier.de/ accessed July 29th 2020
[32] https://www.crossref.org/ accessed July 29th 2020
[33] https://orcid.org/ accessed July 29th 2020
[34] https://neuinfo.org/ accessed July 29th 2020

Easy[35]. Results are returned to the user who can then check and manually confirm the discovered suggestions. Confirmed links are added to the article metadata. An RDF representation of the links, using the `owl:sameAs` or the `skos:exactMatch` link-type, can also be downloaded. LinkItUp was accessible via a web dashboard interface[36], see Figure 11, however, it was not functional at the time of writing.

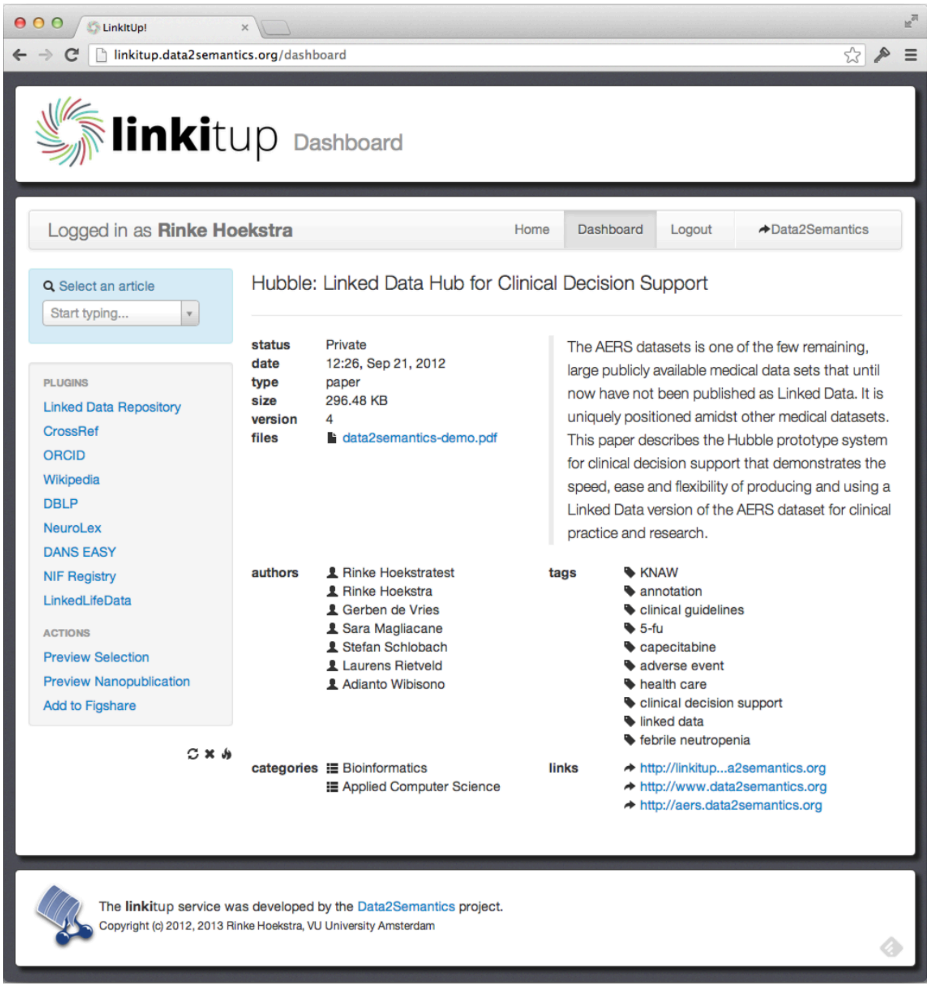


Figure 11: LinkItUp Web Dashboard

SILK

The SILK Link Discovery Framework (Bizer, Volz, Kobilarov & Gaedke, 2009) is a tool for creating relationships between entities from different datasets. Using the SILK – Link Specification Language, users can specify the types of links to be discovered between datasets as well as the conditions entities must fulfil in order to be interlinked. These Link Specifications can be declared manually

---

[35] https://easy.dans.knaw.nl/ui/home accessed July 29th 2020
[36] http://linkitup.data2semantics.org/ inaccessible 27th July 2020

using XML or via the GUI, SILK Workbench. The SILK Workbench[37] is an open source web application that guides users through the link specification creation process – see Figure 12. By comparing the entity properties, SILK can identify possible matched pairs across the specified datasets. The user then manually accepts or rejects the pairs. Following this, the user specifies the link-type to be used to interlink the pairs. In order to create links other than identity links, the datasets must have sufficient information available for the links to be declared and discovered.

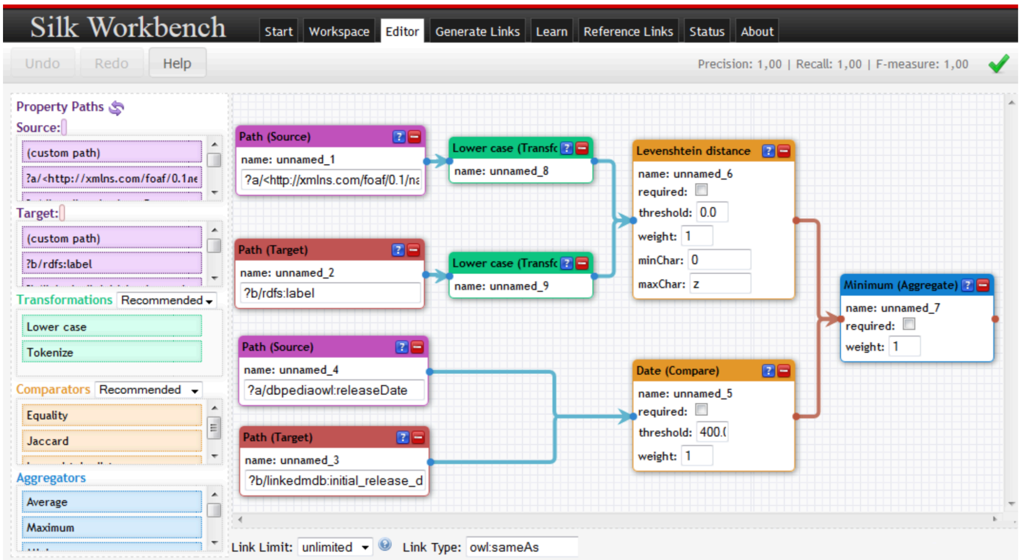


Figure 12: SILK Workbench GUI (www.silkframework.org)

LIMES

The LIMES Link Discovery Framework for Metric Spaces (Ngonga Ngomo & Auer, 2011) uses a series of algorithms and approaches to estimate the similarity of entities from different data sources. LIMES is open source and can be accessed via a web user interface[38], see Figure 13. The GUI assists users in specifying the type of links they wish to create between two datasets and in selecting a machine learning algorithm. After running the algorithm over the datasets, the LIMES GUI presents link candidates to the user who then labels the pairs as either a match or non-match. Following this, the user can specify the link-type to be used to interlink the pairs. Like SILK, in order to use link-types other than those used for instance matching, the datasets must have sufficient information available for the links to be declared and discovered.

---

[37] http://silkframework.org/ accessed 24th July 2020
[38] https://limes.demos.dice-research.org/ accessed 14th July 2020

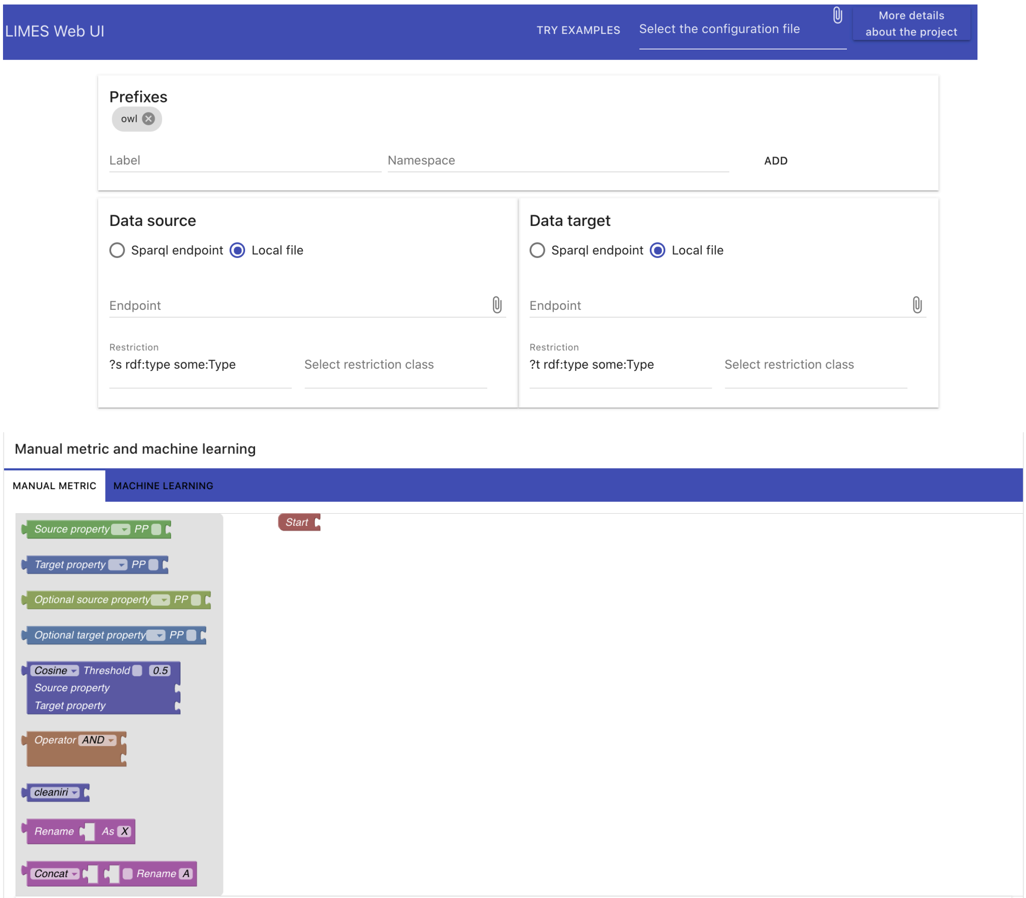


Figure 13: LIMES Web Interface

OpenRefine –RDF Extension

OpenRefine[39], previously GoogleRefine, is an open source tool that can be used to clean, transform and modify unstructured data. It can also be used for data reconciliation, whereby users can match name values, in a local dataset, to equivalent matches in an external database. This process is semi-automated as human judgement is required to confirm matches. OpenRefine provides multiple reconciliation service extensions, some of which are commonly used in LAMs including Wikidata[40], VIAF, DBpedia, the Faceted Application of Subject Terminology[41] (FAST), Getty Vocabularies[42], Europeana[43], and the Library of Congress Subject Heading (LCSH)[44] and Name Authority File (LCNAF)[45].

[39] https://openrefine.org/ accessed 16th July 2020
[40] https://www.wikidata.org/wiki/Wikidata:Main_Page accessed 16th July 2020
[41] https://fast.oclc.org/searchfast/ accessed 16th July 2020
[42] https://www.getty.edu/research/tools/vocabularies/ accessed 16th July 2020
[43] https://www.europeana.eu/en accessed 16th July 2020
[44] https://id.loc.gov/ accessed 11th August 2020
[45] https://id.loc.gov/ accessed 11th August 2020

OpenRefine's RDF Extension provides a GUI, see Figure 14, for reconciling data with SPARQL endpoints using the `owl:sameAs` property.

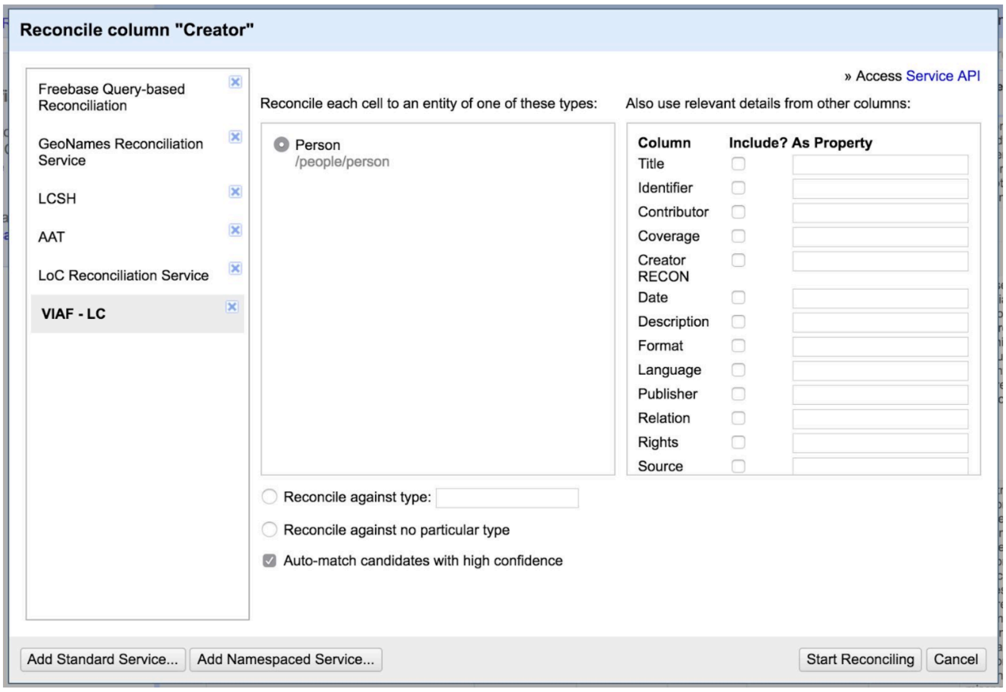


Figure 14: OpenRefine GUI (Dehner, 2017)

3.1.1 Discussion

The above LD interlinking tools have been summarised in Table 2 according to the following criteria:

- Data Input: The type of data formats that can be input into the tool and analysed.
- Supported Link-Types: The types of interlinks that can be created using the tool in order to assess whether the tool can be used to create relationship links, identity links or both.
- Link Generation: Whether the interlinking process is manual, automatic or semi-automatic.
- Integrated Datasets: Whether any external datasets are integrated, or directly accessible, from the tool without having to be imported.
- GUI: How users interact with the tool.
- Domain: Whether the tool was developed for a specific domain.
- Published User-Testing: Whether there is any published research available describing a user evaluation of the tool.
- Interlink Provenance: Whether the tool publishes provenance data for the interlinks generated.

Table 2: Linked Data Interlinking Tools

System	AgreementMaker	LogMap	Linkitup	SILK	LIMES	OpenRefine
Data Input	SPARQL API	RDF	Figshare.com Metadata	RDF SPARQL CSV XML	RDF SPARQL CSV XML	RDF SPARQL CSV XML JSON
Supported Link-Types	owl:sameAs skos:exactMatch	owl:sameAs	owl:sameAs skos:exactMatch	owl:sameAs User-declared	owl:sameAs User-declared	owl:sameAs
Link Generation	Semi-automatic	Semi-automatic	Semi-automatic	Semi-automatic	Semi-automatic	Semi-automatic
Integrated Datasets	None	None	DBpedia DBLP CrossRef ORCID NIF Registry DANS Easy	None	None	Wikidata DBpedia VIAF FAST ORCID LCSH LCNAF Getty *Not an exhaustive list
GUI	GUI	Web Interface	Web Dashboard	Web Workbench	Web Interface	Web Interface
Domain	Unspecified	Unspecified	Academic Research	Unspecified	Unspecified	Libraries Biodiversity Research Other
Published User Testing	No	No	No	No	No	No
Interlink Provenance	No	No	No	No	No	No

It can be seen that the majority of tools were developed solely for instance matching. Only SILK and LIMES allowed for the creation of other types of user-specified interlinks. However, in order to create these links, sufficient information must be available in the dataset. There is an evident need to facilitate the creation of interlinks beyond instance matching. Furthermore, none of the above tools appeared to provide provenance data for the interlinks created. As mentioned in Section 2.2, the provision of LD provenance is crucial in establishing the trustworthiness and quality of the data. Thus, a need for tooling which provides rich data provenance for LD interlinks has also been identified.

In terms of domain specialisation, only OpenRefine had extensions specifically developed for LAMs, although the tool itself was not developed for any particular domain. Additionally, none of the reviewed tools had published user-testing data for their GUIs. As such, there is scope for a LD interlinking framework designed specifically for the LAM domain, and for a GUI that has been tested by the tool's targeted users.

3.2 Linked Data in Libraries, Archives & Museums

Section 3.2.1 below discusses two surveys conducted with IPs which explored the use of LD in LAMs and IPs perceptions of LD. This is followed by Section 3.2.2 which presents a review of some of the leading LD services and projects in LAMs.

3.2.1 Perceptions

OCLC Research[46] conducted an International LD Survey for Implementers in 2014, 2015, and 2018, receiving responses from a combined total of 143 LAMs and research institutions across 23 countries. (Smith-Yoshimura, 2017; Smith-Yoshimura, 2018). Participating institutions, mostly libraries, were actively involved in a LD project or service. The questionnaire gathered data on the types of LD projects being implemented, the data being published and consumed, the rationale for implementing the project, and the barriers encountered.

The chief benefits of implementing LD projects, as reported in the survey, included increased exposure of resources to a wider audience, providing users with richer experiences, demonstrating and testing the capabilities of LD, improving the internal metadata management, and improving search accuracy.

The main barriers in publishing LD included a steep learning curve for implementers, difficulties selecting ontologies, a lack of resources and documentation for building LD services, interlinking issues, and a lack of tooling. The main barriers in consuming LD included challenges with data matching and alignment, difficulty mapping vocabularies, data quality and reliability issues, a lack of "off-the-shelf" tools and a lack of authority control.

It can be seen that, from the perspective of the survey participants, using LD in LAMS has many benefits, both for IPs and LAM users. However, there are a number of barriers in implementing LD, such as a lack of resources and documentation, and a lack of appropriate tooling. These challenges could be a root cause in some of the other reported barriers such as difficulties creating interlinks and issues with data mapping and alignment. It could be argued that

[46] https://www.oclc.org/research/home.html accessed 10th July 2020

tooling designed to target these areas of difficulty could improve IPs' experiences of using LD and in turn improve LD services in LAMs.

While the OCLC study provided a detailed overview of the benefits and challenges of using LD in LAMs, one area that was not explored was IPs level of understanding and interest in implementing LD. However, this perspective was investigated in 2013 via an online survey (LaPolla, 2013) which was distributed to 156 academic cataloguers and library technical-service professionals. The 22-question survey explored the IPs' level of understanding and attitudes towards the SW, as well as their views on its role in the library domain.

Of the 156 responses, 35% of participants rated themselves as "Very Familiar" with the SW and 55% as "Somewhat Familiar". Additionally, the majority of participants were interested in the use of SW technologies within the library domain as they felt it could be useful for library cataloguing. However, issues such as financial constraints, a lack of published best practices, insufficient evidence, and technological complexities were reported as the main barriers to engaging with the SW. These barriers resulted in less than half (42%) of responding institutions actively exploring the implementation of SW catalogues.

Again, the results of this study indicate that, despite the majority of participants being knowledgeable and interested in the SW, there are still many fundamental barriers that prevent IPs in implementing LD in the library domain.

3.2.2 Linked Data Projects & Services

The following section explores the leading LD services and projects emerging from LAMs in order to ascertain the type of projects being implemented, whether interlinking was conducted, and whether LD provenance is available. Where possible, the specific LD tools used in the projects, especially those used for interlinking, are mentioned, however, this information was not always available in publications related to the projects or on the project website.

Library of Congress Linked Data Service

An early adopter of LD was the Library of Congress (LOC) (Summers, Isaac, Redding, & Krech, 2008). The LOC initially made the LCSH available in RDF using the SKOS[47] ontology. The LD service[48] was launched in 2009, exposing approximately 260,000 authority records. LOC has since extended their LD Service to include other controlled vocabularies such as the LCNAF and the Thesaurus of Graphic Materials[49] (TGM), as well as ontologies such as BIBFRAME, Metadata Authority Description Schema – MADS/RDF[50] and PREMIS. Data is available in both MADS and SKOS, and the service provides identity links to controlled vocabularies such as VIAF, Wikidata, FAST, Getty's Union List of Artist Names[51] (ULAN), the General Finnish Ontology[52] (YSO), and RAMEAU[53] (subject indexing language of the BNF), and interlinks to other libraries including the BNF and the DNB. The tooling used to create these interlinks was not specified.

LIBRIS

One of the first large-scale library LD projects was LIBRIS[54]. LIBRIS aimed to convert the six million bibliographic records belonging to the Swedish Union Catalogue, comprising of 175 libraries, to LD. One of the main goals of publishing these records as LD was to share resources beyond the library sector, thus generating increased interest in the libraries' data (Malmsten, 2008, 2009).

The first phase of the project was to transform the records from Machine-Readable Cataloguing[55] (MARC) format, a data structure for bibliographic metadata, into XML. The XML records were then converted to RDF/XML using Extensible Stylesheet Language Transformations (XSLT), and a persistent URI was given to each record. The data was formatted using the KB[56] vocabulary. Interlinks to DBpedia and LCSH were created manually using SPARQL queries

[47] https://www.w3.org/TR/skos-reference/ accessed 30th July 2020
[48] http://id.loc.gov accessed 11th July 2020
[49] https://id.loc.gov/ accessed 10th August 2020
[50] https://www.loc.gov/standards/mads/ accessed 30th July 2020
[51] https://www.getty.edu/research/tools/vocabularies/ulan/ accessed 30th July 2020
[52] https://finto.fi/yso/en/ accessed 29th July 2020
[53] https://rameau.bnf.fr/ accessed 11th July 2020
[54] http://libris.kb.se data accessed 30th July 2020
[55] https://www.loc.gov/marc/bibliographic/ accessed 30th July 2020
[56] https://id.kb.se/vocab/ accessed 14th August 2020

on Sesame Native Store[57], a triple store which has since been succeeded by RDF4J[58]. Data continues to be updated and RDF records can be retrieved from the LIBRIS website via Xsearch[59], a HTTP-based lightweight API.

Europeana Pro

Another large-scale LOD project is Europeana Pro[60], originally data.europeana.eu, which was implemented by Europeana, the European Union's flagship digital library project which collects metadata from approximately 1,500 cultural heritage institutions across Europe (Haslhofer & Isaac, 2011). Via this project, all Europeana datasets have been made available in RDF. The data is represented in the Europeana Data Model[61] (EDM) – a SW inspired metadata framework used as a means of collecting, connecting and enriching the metadata of Europeana data providers (Haslhofer & Isaac, 2011). Using the EDM ensures consistency and interoperability between the datasets of different institutions. Europeana's LD dataset has been manually linked with related resources in Getty's Art and Architectural Thesaurus[62] (AAT), VIAF, GeoNames[63], Wikidata and DBpedia. The tooling used to create these interlinks was not specified. The dataset can be queried using a SPARQL API or a keyword Search API. Some provenance data is provided regarding the provider of the entity metadata.

University of Nevada, Las Vegas

The Library of the University of Nevada, Las Vegas (UNLV) implemented a LD project which involved transforming the 45,000 item records from their Digital Collections department into LD (Lampert & Southwick, 2013; Southwick, 2015). Although no longer active, the aim of the project was to improve the discoverability and interconnectivity of their resources. UNLV also used the project as a means of exploring the feasibility of converting collection records into LD while maintaining their richness and expressivity.

---

[57] https://sourceforge.net/projects/sesame/ 14th August 2020
[58] https://rdf4j.org/ accessed 14th August 2020
[59] http://librishelp.libris.kb.se/help/xsearch_eng.jsp accessed 30th July 2020
[60] https://pro.europeana.eu/page/linked-open-data accessed 30th July 2020
[61] https://pro.europeana.eu/page/edm-documentation accessed 30th July 2020
[62] https://www.getty.edu/research/tools/vocabularies/aat/ accessed 16th July 2020
[63] https://www.geonames.org/ accessed 16th July 2020

UNLV used OpenRefine's RDF Extension in order to transform the metadata records into LD. The data was represented in the EDM, thus the LD is compatible with Europeana. Interlinks were created to well-established controlled vocabularies such as the TGM, LCNAF, FAST, Getty Thesaurus of Geographic Names[64] (TGN). No provenance data was reported to have been provided.

One of the reported challenges of this project included difficulty initiating and planning the project as UNLV found that there was little information available that detailed the experiences and processes of other institutions in implementing LD projects.

British National Bibliography

In July 2011, the British Library released the British National Bibliography (BNB) as LD (Deliot, 2014; Deliot et al., 2017). The BNB LD dataset[65] is an authoritative source of information on UK publications from 1950 to the present and consists of approximately three million records in several languages. The main motivations behind the LD project were to open up the BNB dataset for wider re-use and to allow the British Library to experiment with the SW.

The BNB records were transformed from MARC to RDF in a multi-step process. This process included character set conversion, data normalisation and data matching using the MARC Global[66] tool. Catalogue Bridge, a suite of in-house BL tools, was used to generate local URIs. After these processes, the enhanced MARC records were converted to RDF/XML using XSLT. The RDF/XML was then quality checked using Jena Eyeball[67] and converted to N-Triples. The BNB uses many metadata schemas including the Bibliographic Ontology[68], Bio[69], Dublin Core, FOAF[70], and SKOS. Using CatalogueBridge, the BNB dataset was linked to well-established controlled vocabularies within the LAM domain by matching authorised headings in the BNB records with corresponding URIs in

---

[64] http://www.getty.edu/research/tools/vocabularies/tgn/index.html accessed 16th July 2020
[65] http://bnb.data.bl.uk accessed 18th July 2020
[66] http://www.marcofquality.com/soft/mgfeatures.html accessed 18th July 2020
[67] https://docs.huihoo.com/jena/Eyeball/index.html accessed 30th July 2020
[68] http://bibliontology.com/ accessed 30th July 2020
[69] https://vocab.org/bio/ accessed 30th July 2020
[70] http://xmlns.com/foaf/spec/ accessed 30th July 2020

VIAF and LCSH. Links were also created to WebDewey[71], MARC Countries[72] and MARC Languages Codes[73], Lexvo (now discontinued) and GeoNames by generating URIs automatically from existing record data and also via Crosswalk Matching. Since its initial launch, the BNB dataset has also been linked with the International Standard Name Identifier[74] (ISNI) database and the UK Government Interval Set[75]. The BNB is accessible via a keyword search and a SPARQL endpoint[76]. VoID (Alexander, Cyganiak, Hausenblas & Zhao, 2011), a W3C recommended schema for the description of RDF datasets, was used to provide provenance data for the dataset.

Challenges reported included a steep technical learning curve for staff and difficulty quantifying the service impact of the LD dataset. However, the project has been referenced by many other subsequent LD projects in the library domain, proving it to be quite influential.

The National Polytechnic School of Ecuador

The National Polytechnic School of Ecuador embarked on a LD project whereby bibliographic metadata from the electrical engineering faculty of the school was published as LD (Hallo, Luján Mora & Trujill Mondéjar, 2014). The library records were originally stored in MARC format in a relational database. The Triplify[77] tool, now defunct, was used to generate RDF from the data stored and SILK was then used to interlink the RDF with external LD datasets including Open Library[78], Europeana and LCSH.

Swissbib

Swissbib[79] hosts the meta-catalogue of Swiss University Libraries and the Swiss National Library as LD (Bensman, Prongu, Hellstern, & Kuntschik, 2016). For this project, records were converted from MARC to RDF and then formatted using Dublin Core and the Resource Description and Access Unconstrained

[71] http://dewey.org/ accessed 30th July 2020
[72] http://id.loc.gov/vocabulary/countries accessed 30th July 2020
[73] http://id.loc.gov/vocabulary/languages accessed 30th July 2020
[74] https://isni.org/ accessed 30th July 2020
[75] https://www.epimorphics.com/using-interval-set-uris-in-statistical-data/ – 30th July 2020
[76] https://bnb.data.bl.uk/flint-sparql accessed 30th July 2020
[77] https://www.w3.org/2001/sw/wiki/Triplify accessed 30th July 2020
[78] http://datahub.io/dataset/open-library accessed 30th July 2020
[79] https://data.swissbib.ch/ accessed 30th July 2020

(RDAU) properties element set[80]. The data was modelled as per the Functional Requirements for Bibliographic Records[81] (FRBR) model. Following this, the data was interlinked with VIAF and DBpedia using LIMES. Interlinking was reported to be a challenging and time-consuming process for this project (Bensman et al., 2016).

Data.bnf.fr

The BNF hosts its 'Catalogue Général' as LD via its data.bnf.fr service (Simon, Wenz, Michel, & Di Mascio, 2013). The LD is modelled according to the FRBR model. an entity-relationship model designed for libraries, and formatted in SKOS, Dublin Core, FOAF and RDA Vocabularies[82]. In terms of linking, data.bnf.fr partially relies on available alignments from other projects by interlinking with LOD hubs, like DBpedia and VIAF, in order to link to other LD datasets. The dataset also links with Europeana, GeoNames, WorldCat[83], the DNB, the BNE, the French Academic Union Catalogue[84] (SUDOC), and Agrovoc[85]. The BNF used an in-house string-matching alignment tool in order to generate interlinks.

Datos.bne.es

The BNE has published the majority of its bibliographic and authority records as LD as part of the datos.bne.es[86] project (Vila-Suero & Gómez-Pérez, 2013; Vila-Suero, Villazón-Terrazas & Gómez-Pérez, 2013). The data was transformed from MARC to RDF using MARiMbA[87], a command-line tool designed specifically for the project, and formatted using the BNE ontology[88]. The RDF was modelled as per the FRBR model. MARiMbA was also used to create interlinks to external datasets including VIAF, DBpedia, Libris, the German National Library and SUDOC via property matching.

[80] http://www.rdaregistry.info/Elements/u/ accessed 30th July 2020
[81] https://www.oclc.org/research/activities/frbr.html accessed 30th July 2020
[82] http://www.rdaregistry.info/ accessed 30th July 2020
[83] https://www.worldcat.org/ accessed 30th July 2020
[84] http://www.sudoc.abes.fr/ accessed 30th July 2020
[85] http://aims.fao.org/agrovoc/releases accessed 30th July 2020
[86] http://datos.bne.es/inicio.html accessed 30th July 2020
[87]http://mayor2.dia.fi.upm.es/oeg-upm/index.php/en/technologies/228-marimba/index.html accessed 30th July 2020
[88] http://datos.bne.es/def/ontology.html accessed 30th July 2020

The Linked Data Service of the German National Library

The Linked Data Service of the German National Library[89] (DNB) provides access to its bibliographic and authority data in RDF (Hannemann & Kett, 2010). The library data was transformed to RDF using the Metafacture[90] tool and formatted using RDAU, Dublin Core, FOAF, the GND (Gemeinsame Normdatei) Ontology[91], DNB Metadata Terms and the Agent Relationship Ontology[92]. The data is modelled as per BIBFRAME. The data contains interlinks to German Wikipedia, DBpedia, VIAF, LCSH and RAMEAU. These interlinks were pre-existing in the library's catalogues and transformed to RDF using Metafacture.

CEDAR

An example of an archival LD project is CEDAR[93]. CEDAR is a project in which a collection of Dutch historical census data, archived by the Central Bureau of Statistics and the International Institution of History, was converted to LD (Ashkpour, Meroño-Peñuela, & Mandemakers, 2015; Meroño-Peñuela, Ashkpour, Guéret, & Schlobach, 2017). The census dataset consisted of 6.8 million statistical observations about the demography, labour and housing of Dutch Society from the years 1795 to 1971.

Data was converted to RDF from Excel tables using TabLinker[94], a tool which was developed specifically for the project. Again, using TabLinker, the dataset was manually interlinked to the Historical International Standard Classification of Occupations[95] (HISCO), the ICONCLASS[96] system, and to the Dutch Ships and Sailors datasets[97]. The dataset was also semi-automatically linked to DBpedia, GeoNames and the Amsterdamse Code[98] using existing mappings. Provenance data for the project is provided using PROV. The data was then made available via a SPARQL endpoint.

---

89 https://www.dnb.de/DE/Home/home_node.html accessed 30th July 2020
90 https://github.com/metafacture/metafacture-core/wiki accessed 30th July 2020
91 https://d-nb.info/standards/elementset/gnd accessed 30th July 2020
92 https://d-nb.info/standards/elementset/agrelon accessed 30th July 2020
93 https://www.cedar-project.nl/ accessed 30th July 2020
94 https://github.com/Data2Semantics/TabLinker accessed 30th July 2020
95 https://historyofwork.iisg.nl/ accessed 30th July 2020
96 http://www.iconclass.nl/ accessed 30th July 2020
97 https://old.datahub.io/dataset/dutch-ships-and-sailors accessed 30th July 2020
98 https://www.encyclo.nl/begrip/Amsterdamse_code accessed 10th August 2020

Since transforming the data to RDF, it was reported that users can run more complex queries over the data more efficiently. However, the researchers found the process of harmonising the census data from year to year to be one of the most challenging tasks associated with the project. They also found that there was a lack of SW tools available that were suitable for mapping the size and complexity of the dataset (Meroño-Peñuela, Ashkpour, Rietvel, Hoekstra, & Schlobach, 2012).

Amsterdam Museum

The Amsterdam Museum hosts a variety of cultural heritage objects, such as artwork, clothing, furniture and books, related to Amsterdam and its citizens (de Boer, Wielemaker, van Gent, Hildebrand, Isaac, Van Ossenbruggen & Schreiber, 2012). Prior to the implementation of the LD project[99], the museum had already made the records of its collection publicly available online in XML. The museum then converted these 73,447 XML object records to RDF, following the EDM model using the ClioPatria SW Toolkit[100]. The data was subsequently aligned with the Dutch AAT, ULAN, GeoNames and DBpedia using Amalgame[101], a vocabulary alignment tool. The Amsterdam museum data is available as LOD via Europeana and was previously also available via a SPARQL endpoint.

3.2.3 Discussion

Both LAM surveys discussed in Section 3.2.1 demonstrate that IPs have a favourable view of LD and that they perceive LD to have many benefits for the domain. However, the studies also highlighted many fundamental barriers towards implementing LD that need to be addressed in order for IPs to be able to engage fully with the SW. These barriers include, but are not limited to, the provision of LD resources and documentation for LAMs, as well as the provision of less technologically complex tooling for data mapping, data reconciliation and LD interlinking.

[99] http://semanticweb.cs.vu.nl/lod/am/ accessed 30th July 2020
[100] http://cliopatria.swi-prolog.org/home accessed 30th July 2020
[101] http://semanticweb.cs.vu.nl/amalgame/ accessed 30th July 2020

The LD services discussed in Section 3.2.2 have been summarised in Table 3. Looking at these services, it can be seen that the majority are single-institution initiatives that are interlinked to an average of five external datasets. These datasets are, for the most part, authority files and controlled vocabularies, such as LCSH, AAT, ULAN and VIAF, and datahubs such as DBpedia and Wikidata. Although interlinking with controlled vocabularies is extremely useful and important, this type of linking predates LD. It would be important for LD services to ensure that they also link to datasets beyond those used for authority control to realise the full potential of LD. Linking to large-scale datahubs, such as DBpedia and Wikidata, is also very useful. However, as these datasets do not fall within the LAM domain, it would be important for LAMs to interlink with other LAM datasets, as linking with general datahubs alone may result in a loss of authoritative value given they do not typically use LAM controlled vocabularies.

It was noted that smaller LAMs, such as the Amsterdam Museum, may choose to add their LD to the SW via large-scale aggregators such as Europeana. While this allows for these datasets to be available on the SW, it creates a disconnect between the LAM institution and the data (de Boer et al., 2012). Enabling LAMs to directly interlink with other LD datasets, especially those emerging from other LAM institutions, could help maintain the authoritative value of the data and also allow smaller institutions to have their own node in the Linked Open Data Cloud[102].

Only four of the projects, the LOC, the National Polytechnic School of Ecuador, the BNF and the BNE, linked with other library datasets, and only one project, CEDAR, linked with specialised datasets. In order to provide richer knowledge discovery for LAM users, LAMs should prioritise interlinking with other LAMs. Also, only three services, Europeana, CEDAR and the BNB, appeared to provide LD provenance information. In order to encourage other institutions to interlink with their data, LAMs need to provide provenance data.

[102] http://lod-cloud.net accessed 30th July 2020

Finally, the majority of interlinks created by the projects were `owl:sameAs`, `skos:exactMatch` and `skos:closeMatch`. While some `rdfs:seeAlso` links were noted, there remains vast potential for LAMs to create interlinks that provide additional information and context for a given entity by using other kinds of link-types.

Overall, Section 3.1 of this review indicated that there are limited interlinking tools available that are capable of creating links beyond instance matching. Additionally, none of the tools reviewed generate provenance data for the interlinks. The survey results, discussed in Section 3.2.1, indicate that IPs find interlinking to be a barrier when implementing LD projects and that they find LD tooling to be technologically complex. The LD services reviewed in Section 3.2.2 had interlinks to an average of five other datasets, with the majority of these interlinks being instance matches. Additionally, only three services appeared to provide provenance data.

Overall, there appears to be a need for a user-friendly LD interlinking framework that facilitates the creation of relationship links and that provides provenance data for the links generated. In order to collate a more detailed set requirements for such a tool, a LD Requirements Questionnaire, discussed in the following section, was distributed to LAMs.

Table 3: LD Services & Projects in LAMs

Institution	Library of Congress	Swedish Libraries	Europeana	University of Nevada, Las Vegas	British National Library	National Polytechnic School of Ecuador	Swiss University Lbraries and National Library	French National Library	National Library of Spain	German National Library	The Dutch Central Bureau of Statistics and the International Institution of History	Amsterdam Museum
Service Name	ID.LOC.GOV – Linked Data Service	Libris	Europeana Pro	-	British National Bibliography bnb.data.bl.uk	-	Swissbib	data.bnf.fr	datos.bne.es	Linked Data Service of the German National Library	CEDAR	-
Year Launched	2009	2008/2009	2011/2012	2013	2014	2014	2016	2013	2011	2010	2012/2013	2010/2011
Active LD Service	Yes - Open Access	Yes - Open Access	Yes - Open Access	No	Yes - Open Access	No	Yes - Open Access	Yes - Open Access	Yes - Open Access	Yes - Open Access	Yes - Open Access	Yes via Europeana
Aim	Transform the LCSH to LD	Transform the Swedish Union Catalogue to LD	Transform collections to RDF	Transform Digital Collection to LD	Transform the British National Bibliography to LD	Transform bibliographic metadata from the electrical engineering faculty to LD	Transform the Swiss University Libraries' and the Swiss National Library catalogues to LD	Transform the Catalogue Général to LD	Transform the library catalogue to LD	Transform bibliographic and authority data to LD	Transform Dutch historical census data to LD	Transform collection records to LD
RDF Converter	XSLT	XSLT	XSLT	OpenRefine	XSLT	Triplify	Metafacture	CubicWeb framework	MARiMbA	Metafacture	TabLinker	ClioPatria SW Toolkit
Interlinking	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Interlinking Tool/Technologies	Not specified	SPARQL	Not specified	OpenRefine	Catalogue Bridge	SILK	LIMES	In-house tooling	MARiMbA	Metafacture	TabLinker	Amalgame
Interlinked Datasets	BnF RAMEAU DNB VIAF YSO Wikidata FAST ULAN	DBpedia LSCH	AAT VIAF GeoNames Wikidata DBpedia	TGM LCNAF FAST TGN	VIAF LCSH Dewey.info MARC Country & Language Codes GeoNames ISNI UK Gonvernment Interval Set	OpenLibrary Europeana LCSH	VIAF DBpedia	DBpedia VIAF Europeana GeoNames WorldCat DNB SUDOC Agrovoc	VIAF DBpedia Libris DNB SUDOC	German Wikipedia DBpedia VIAF LCSH RAMEAU	HISCO* Iconclass* Dutch Ships and Sailors* DBpedia GeoNames Amsterdamse Code	Dutch AAT ULAN GeoNames DBpedia
Access Point	Keyword search	Keyword search	SPARL API Search API	SPARQL endpoint	Keyword Search SPARQL endpoint	SPARQL endpoint	Keyword search	Keyword search SPARQL Endpoint	Keyword search	Keyword search	SPARQL endpoint	Europeana SPARQL endpoint
LD Provenance	No	No	Yes	No	Yes	No	No	No	No	No	Yes	No

3.3 Linked Data Requirements Questionnaire

Section 3.1 and Section 3.2 above indicated that there is a need for a LD interlinking framework that facilitates the creation of relationship links and that is less technologically complex than existing LD tooling. In order to gain a more defined and detailed understanding of this issue, a LD Requirements Questionnaire was distributed to LAMs. The online questionnaire was developed using Qualtrics[103] survey software and consisted of 50 questions – a copy can be found in Appendix 1. The questionnaire was distributed to IPs by sharing the link via LAM related conferences, organisations and contacting IPs directly. The questionnaire was completed by 185 participants.

The questionnaire was divided into six parts:

1. Cataloguing Experience
2. Usability of Cataloguing Tools
3. Knowledge and Views on Linked Data
4. Linked Data Project Experience
5. Usability of Linked Data Tools
6. Linked Data for Information Professionals

Questions in Part 1 and Part 2 of the questionnaire explored participants' cataloguing experience and gathered information on their perceived usability of current cataloguing tooling. Part 3 and Part 4 of the questionnaire measured the participants' self-perceived knowledge of LD and explored their experiences using LD Finally, Part 5 and Part 6 of the questionnaire gathered information on the participants' perceived usability of current LD tooling and explored solutions for common LD challenges in LAMs.

Of the 50 questions, most were multiple choice, however, some open-ended questions were also included. Some questions were based on the results of the LD studies discussed in Section 3.2.1. Included in Part 2 and Part 5 of the questionnaire was the Computer System Usability Questionnaire (CSUQ) (Lewis, 1995), a software usability and utility measure which is detailed in Section 5.2.3 of this thesis. The CSUQ consists of 19 statements about which the user rates agreement on a seven-point scale from Strongly Agree (1) to

[103] https://www.qualtrics.com/uk/ accessed 13th July 2020

Strongly Disagree (7) – thus lower scores indicate fewer usability issues. However, for the purpose of this questionnaire a subset of nine statements, focusing on system efficiency, effectiveness, learnability and ease of use, were used.

3.3.1 Participants

The participants of this questionnaire were primarily IPs with experience working in the LAM domain. Also recruited were Researchers and Academics with experience in the LAM domain. IPs were encouraged to participate regardless of whether they had any prior experience working with LD. This was done in an attempt to recruit a broad range of participants, rather than just IPs who are highly experienced in LD. That being said, it is possible that many participants who completed the survey already had a prior interest in the SW and LD in order to be motivated to do so.

Non-probabilistic sampling methods were used to recruit participants (Daniel, 2011). This was done by directly contacting LAMs with a request to distribute the questionnaire link amongst staff and members, sharing the questionnaire link on Twitter[104], and by directly contacting known IPs and researchers. Snowball sampling was also used whereby participants were asked to forward the questionnaire link to other potential participants.

Prior to completing the questionnaire, participants were supplied with an Information Sheet and a Consent Form (see Appendix 2). Only participants who provided consent participated in the questionnaire.

The questionnaire received 230 responses of which 185 were deemed suitable for analysis. The responses removed from the study were eliminated on the basis that participants did not confirm that they had experience working as an IP or a researcher in the LAM domain, or due to question incompleteness.

Participants were from 20 different countries, with the majority coming from Ireland (28%), the USA (23%) and the UK (20%). Within Ireland and the UK,

[104] https://twitter.com/lucymckenna01/status/918911344962875392 accessed 11st Nov 2020

the types of institutions where the questionnaire was shared included large reputable university libraries, national libraries, archives, digital repositories and LAM related associations. In the US, the questionnaire was shared within a large reputable library research centre and on Twitter by an international IP association.

Participants represented a variety of LAMs and research institutions including Academic Libraries (56%), Research Institutions (7%), Public Libraries (7%), Special Libraries (6%), Archives (6%), National Libraries (5%), Museums (4%), and Special Archives (1%). 88% of participants had more than 4 years of experience working in the LAM or research domain, with the majority of participants (61%) having 10 or more years of experience. As the majority of participants came from an Academic Library setting, it is important to note that the results of the survey may not be generalisable across all LAM domains.

3.3.2 Results

Part 1 – Cataloguing Experience

74% (N = 132) of participants reported that they were currently involved in metadata cataloguing at their workplace. When asked what metadata formats were applied, this subgroup of participants reported a total of 41 formats with the most commonly cited being MARC 21 (73%), Dublin Core (43%), Encoded Archival Description[105] (EAD) (20%), MARC XML (17%), and the Metadata Object Description Schema[106] (MODS) (16%).

With the majority of participants coming from the library domain, unsurprisingly the most frequently reported metadata format used was MARC 21. As MARC 21 does not inherently allow for linking, the standard is incompatible with LD. Converting MARC 21 to RDF, though possible, is extremely challenging (Cole, Han, Weathers, & Joyner, 2013). Although MARC is still the most commonly used library metadata model, there seems to be a consensus that it is no longer the most effective means of encoding library metadata and that LD may be the way forward (Kroeger, 2013; Sprochi, 2016). However, the future of the bibliographic record still remains unclear.

[105] https://www.loc.gov/ead/ accessed 11th August 2020
[106] https://www.loc.gov/standards/mods/ accessed 11th August 2020

Part 2 – Usability of Cataloguing Tools

94% (N = 173) of participants reported having previous experience using a cataloguing tool. Of these participants, the most commonly reported tools can be seen in Table 4. Also included is the number of participants who indicated whether they initially required the support of a technical person to use the cataloguing tool. The CSUQ scores for these tools can be found in Table 5.

Table 4: Commonly Used LAM Cataloguing Tools

Tool	Technical Assistance Required			No. of Participants	Percentage of Participants (N = 173)
	Yes	No	Unsure		
Aleph	20	20	5	45	26%
Omeka	2	33	6	41	24%
DSpace	13	20	7	40	23%
Sierra	13	23	3	39	20%
Koha	5	18	2	25	18%
Fedora	12	8	7	27	16%
Voyager	8	12	4	24	14%
Filemaker	3	15	0	18	11%
Alma	9	20	9	20	11%
Millenium	3	9	4	16	10%

It can be seen that only Fedora[107] had more participants requiring assistance than not, and only Aleph[108] had an equal number across both groups. Interestingly, both Fedora and Aleph had the highest scores in the CSUQ signifying more usability issues for these tools. However, all tools had overall scores less than four indicating generally positive responses to the CSUQ statements. Higher scores were noted for statements regarding the tools' interface, suggesting increased usability issues in this area.

Having a knowledge of the types of systems IPs regularly use as well as what they like and dislike about the systems is useful for designing future tooling for LAMs.

[107] https://duraspace.org/fedora/ accessed 31st July 2020
[108] https://www.exlibrisgroup.com/products/aleph-integrated-library-system/ – 31st July 2020

Table 5: Cataloguing Tool CSUQ Scores

Tool	It was simple to use this tool	I could effectively complete my work using this tool	I was able to complete my work quickly using this tool	It was easy to learn how to use this tool	Whenever I make a mistake using the tool, I recovered easily and quickly	The organisation of information on the tool's screen was clear	The interface of this tool was pleasant	This tool had all the functions and capabilities I expected it to have	Overall, I was satisfied with this tool	Score
Aleph	4	2	3	4	4	4	4	3	3	**3.44**
Omeka	3	2	3	2	3	3	3	4	3	**2.88**
DSpace	3	2	3	3	3	4	4	4	3	**3.11**
Sierra	3	3	3	3	3	4	4	4	3	**3.22**
Koha	3	2	3	3	3	3	3	3	3	**2.88**
Fedora	5	3	3	4	4	4	4	3	3	**3.66**
Voyager	3	2	2	3	3	4	4	3	3	**2.88**
Filemaker	3	2	3	3	3	3	3	3	3	**2.88**
Alma	4	2	3	3	3	4	4	3	3	**3.11**
Millennium	3	2	2	2	2	4	4	3	3	**2.66**
Average	**3.4**	**2.2**	**2.8**	**3**	**3.1**	**3.7**	**3.7**	**3.3**	**3**	**3.07**

Part 3 – Knowledge and Views on LD

Of the 185 participants, 29% (N = 54) had prior experience working on a LD project or service – going forward, this group will be referred to as the LDExp group. 71% (N = 131) of participants had no prior LD experience and will be referred to as the NoLDExp group.

Participants were asked to rate their prior knowledge of the SW and LD as either Extremely Knowledgeable (EK), Very Knowledgeable (VK), Moderately Knowledgeable (MK), Slightly Knowledgeable (SK) or Not at all Knowledgeable (NK), the results of which can be seen in Table 6.

Table 6: LD Questionnaire – Participant SW & LD Knowledge Ratings

Topic	Rating	Number of Participants	Percentage of Participants N = 185
SW	EK	11	6%
	VK	26	14%
	MK	70	38%
	SK	48	26%
	NK	30	16%
LD	EK	11	6%
	VK	37	20%
	MK	72	39%
	SK	46	25%
	NK	19	10%

Results show that the vast majority participants had some prior knowledge of the SW (84%) and LD (90%), with the majority of participants rating themselves as at least Moderately Knowledgeable for both the SW (58%) and LD (65%).

Looking at the LDExp group versus the NoLDExp group:

- 95% of the LDExp group rated their knowledge of the SW as Moderately Knowledgeable or above, compared with 44% of the NoLDExp group.
- 95% of the LDExp group rated their knowledge of LD as Moderately Knowledgeable or above, compared with 54% of the NoLDExp group.
- 16% of participants had no prior knowledge of the SW and 10% had no prior knowledge of LD. All of these participants were part of the NoLDExp group.

These high knowledge ratings could allow for the results of this study to be treated with increased confidence. However, it is important to bear in mind that participants were asked to rate their own level of knowledge of the SW and LD, without having to demonstrate this knowledge, thus running the risk of participants being more or less knowledgeable than they rated themselves to be.

Participants were asked whether they thought there were benefits to publishing LAM data as LD. Of those who responded 'Yes' (N = 150), the most commonly reported benefits of publishing LAM metadata as LD were:

- Expose data to a larger audience (89%)
- Improve data accessibility (82%)
- Easier metadata sharing (81%)
- More efficient data searches (73%)
- Increased metadata openness (71%)
- Improved authority control on the SW (73%)
- Create a research environment (55%)
- Improved search engine optimisation (SEO) (51%).

Participants were also asked whether they thought LAMs faced barriers to publishing LD. Of those who responded 'Yes' (N = 124), the most commonly reported barriers were:

- Integrating LD publication into the cataloguing workflow (77%)
- Cleaning data (67%)
- Time consuming (66%)
- Steep learning curve (63%)
- Copyright issues (52%)
- Difficulty using LD tools (52%)
- Inadequate LD tools available (50%)
- Difficulty establishing links (43%)
- Difficulty using SPARQL endpoints (42%)
- Insufficient number of controlled vocabularies available as LD (41%).

Following this, participants were asked whether there were benefits for LAMs in consuming LD. Of those who responded 'Yes' (N = 154), the most frequently reported benefits of consuming LD were:

- Improved data discovery (85%)
- Interlinking across institutions (81%)
- Enriched bibliographic metadata (79%)
- Interlinking across datasets (75%)
- Harmonise data from multiple sources (73%)
- More efficient data searches (70%)
- Improved metadata quality (68%)
- Automated authority control (53%)
- Reduced time spent cataloguing (52%).

Again, participants were asked whether they thought LAMs faced barriers to consuming LD. Of those who responded 'Yes' (N = 119), the most mentioned barriers were:

- Difficulty ingesting into the catalogue (75%)
- Time consuming processes (71%)
- Issues with dataset reliability (58%)
- Difficulty using LD tools (55%)
- Authority control issues (53%)
- Lack of LD tools available (51%)
- Issues with data re-usability (49%)
- Difficulty establishing interlinks (45%)
- Use of unstable URIs (45%).

Other concerns that were frequently mentioned included the cost, both financial and timewise, of publishing and consuming LD. This included the time and cost of training current staff in using LD and hiring new IT staff. Participants indicated that in order to invest time and finances into LD, more useful examples of its application would need to be seen.

Part 4 – LD Project Experience

Of the 185 participants, 29% (N = 54) stated that they had previously been directly involved in the implementation of a LD project or service. Of this subgroup, 51% indicated that the project involved both the consumption and publication of LD, 41% indicated the project only involved LD publication, and 8% reported that the project involved only the consumption of LD. Evidently, more LAMs are publishing LD than are consuming it. It is possible that this lack of interlinking and integration is a result of the challenges reported in Part 3 of the questionnaire.

With regards LD consumption, results indicated that the most frequently used datasets were the AAT (55%), DBpedia (52%), VIAF (52%), Wikidata (26%), GeoNames (38%), Getty Thesaurus of Geographic Names (TGN) (35%), LCNAF (29%), WorldCat.org (29%), and Europeana (26%).

In relation to LD publication, participants indicated that the most common types of data published were bibliographic data (55%), digital collection data (55%), authority files (35%), controlled vocabularies (32%), ontologies (32%) and holdings data (22%). Additionally, the most commonly reported ontologies used for LD publication were Schema.org[109] (50%), SKOS (45%), DCTerms[110] (43%), DCE[111] (37%), FOAF (33%) and local vocabularies (31%).

Part 5 – Usability of LD Tools

The most commonly reported LD tools that the LDExp group (N = 54) had experience using can be seen in Table 7. Also included here is the number of participants who indicated whether or not they initially required technical support in order to use the tool. Participants also completed the CSUQ for these tools – see Table 8.

[109] https://schema.org/ accessed August 3rd 2020
[110] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/ accessed July 31st 2020
[111] https://www.dublincore.org/specifications/dublin-core/dces/ accessed July 31st 2020

Table 7: Commonly Used LD Tools

Tool	Technical Assistance Required			Number of Participants	Percentage of Participants (N = 54)
	Yes	**No**	**Unsure**		
OpenRefine	2	12	1	15	28%
Protégé	3	6	4	9	20%
Fedora	6	2	2	10	18%
Apache Fuseki	1	2	3	6	15%
RDF Refine	1	3	2	6	11%
Virtuoso	1	4	0	5	11%
Pubby	1	3	0	4	11%
Blazegraph	1	3	0	4	11%

Again, Fedora was the only tool where more participants stated that they required assistance than those who did not. Fedora also had one of the highest CSUQ scores, however, both Virtuoso[112] and BlazeGraph[113] received higher scores, indicating more perceived usability issues for these tools. That said, all tools had overall scores less than four indicating generally positive responses to the CSUQ statements.

Knowing the types of LD tools IPs use and the types of interfaces they rate most positively is useful for designing a LD tool specifically for LAMs.

[112] https://virtuoso.openlinksw.com/ accessed July 31st 2020
[113] https://blazegraph.com/ accessed July 31st 2020

Table 8: LD Tool CSUQ Scores

Tool	It was simple to use this tool	I could effectively complete my work using this tool	I was able to complete my work quickly using this tool	It was easy to learn how to use this tool	Whenever I make a mistake using the tool, I recovered easily and quickly	The organisation of information on the tool's screen was clear	The interface of this tool was pleasant	This tool had all the functions and capabilities I expected it to have	Overall, I was satisfied with this tool	Score
OpenRefine	4	2	3	4	4	4	4	3	3	**3.44**
Protégé	3	2	3	2	3	3	3	4	3	**2.88**
Fedora	3	2	3	3	3	4	4	4	3	**3.11**
Apache Fuseki	3	3	3	3	3	4	4	4	3	**3.22**
RDF Refine	3	2	3	3	3	3	3	3	3	**2.88**
Virtuoso	5	3	3	4	4	4	4	3	3	**3.66**
Pubby	3	2	2	3	3	4	4	3	3	**2.88**
Blazegraph	3	2	3	3	3	3	3	3	3	**2.88**
Average	**3.37**	**2.25**	**2.87**	**3.12**	**3.25**	**3.63**	**3.63**	**3.37**	**3**	**3.12**

Part 6 – LD for IPs

When asked to provide their thoughts on the idea of developing LD tooling specifically for the LAM domain, participants reported that such tools would:

- Enable more LAMs to become part of LD community.
- Allow LAMs to fully engage in the LD ecosystem.
- Make LD creation and usage more accessible for IPs.
- Enable LD to be incorporated into cataloguing workflows.
- Make it easier for IPs to understand the benefits of LD.
- Help reduce the technological barrier.
- Be more likely to be used in LAMs.

Participants indicated that such tooling should be:

- Integrable with library management systems and cataloguing workflows.
- LAM standards compliant
- Considerate of IPs point of view.
- Attuned to IPs working environment.
- Usable without having to understand complex LD technicalities or requiring the help of an IT professional.
- User friendly, low-tech interface.
- Available in the public domain.

However, some concerns were raised which included:

- Bespoke tools may limit how LAMs could interact with communities not using these tools, potentially limiting the use of their work.
- Whether tools would be able to interact with closed vendor systems.
- If too bespoke it may be difficult to adapt tooling to the individual needs of specialised teams.
- Workflows and data processes differ across institutions.
- It was also highlighted that there are already existing LD Tools that LAMs could use. While this is true, based on the LD challenges reported by the participants, these tools are difficult to use.

When asked to rate the usefulness of a LD Interlinking Tool for IPs, see Table 9, the vast majority of participants (77%) indicated that they thought it to be a useful idea. When asked to explain why they rated the idea as useful, see Table

10, the most commonly cited reasons included that interlinking resources provides tremendous value to LAMs in terms of enriching data and improving resource discoverability, and a bespoke tool could help overcome the technical knowledge gap of IPs who find interlinking to be one of the most challenging areas of creating LD. However, concerns regarding how adaptable such a tool would be to the individual needs of institutions were raised, as well as a need for useful case studies if time and finances were to be invested in using such tooling.

Table 9: Usefulness of a LD Tool for LAMs

Rating	**Number of Participants**	**Percentage of Participants (N = 185)**
Extremely Useful	74	40%
Moderately Useful	55	30%
Slightly Useful	13	7%
Neither Useful nor Useless	35	19%
Slightly Useless	0	0%
Moderately Useless	2	1%
Extremely Useless	6	3%

Table 10: LD Questionnaire – LD Tool Usefulness Quotes

Rating	Participant Comment
Extremely Useful	Reconciliation is one of LOD's greatest problems
Moderately Useful	Needs to be a means for IPs to establish connections as they create/interact with data
	More institutions would get involved if there were less barriers to entry
	Overcome the technical knowledge gap of content experts that create metadata
	Create potential for new interdisciplinary research
	Reduce the need for IPs to learn to use different interfaces, search strategies and vocabularies
	A dedicated approach would mean that more IPs would "buy into" using the system
	Time saving
Slightly Useful	A tool with awareness of the sources that IPs trust/prefer will be more efficient
	Potentially useful within particular LAMs but may have limited use within sectors that do similar work but are not 'cultural heritage' institutions.
	More valuable if there was a way to automate interlinking across collection silos
Neither Useful nor Useless	Need useful case study to be convinced
	Unsure if IPs will take kindly to doing more or having roles changed
	Until there is more evidence that LD is the future of information systems, there may not be a big buy in from the financially strapped heritage sector
Useless	* The participants who rated the tool as slightly, moderately or extremely useless did not provide feedback.

Participants then listed the functions they thought a LD interlinking tool should have and the most commonly stated were:

- Data enrichment (61%)
- Awareness of common data sources (58%)
- Automatic linking to controlled vocabularies (54%)
- Configurable to the institution's workflow (54%)
- Ability to integrate LD datasets into the catalogue (52%)
- Data cleaning (49%)

- Link validation (48%)
- Link discovery (46%)
- Review data source quality (43%)
- Vocabulary alignment/reconciliation (43%)
- Automatically link to ontologies (42%)
- Remove the need for understanding LD technicalities (40%)
- Create controlled vocabularies in SKOS (35%).

Also documented were the LD datasets that participants would find most useful to interlink to which included; Workdcat.org (72%), id.loc.gov (59%), NAF (45%), ORCID (44%), British National Bibliography (BNB) (43%), VIAF (41%), AAT (40%), Europeana (40%), The British Museum's Semantic Web Collection (34%), and GeoNames (33%).

Finally, participants were asked what quality criteria (Zaveri, Rula, Maurino, Pietrobon, Lehmann & Auer, 2016) they apply when using, or searching for, external data sources – see Table 11. Participants indicated that Trustworthiness (66%) was by far the most important criteria, followed by Interoperability (51%) and Licensing (49%). Having a knowledge of these criteria would be useful in aiding IPs in determining datasets to integrate and to interlink with.

Table 11: Dataset Quality Criteria

Criteria	Number of Participants	Percentage of Participants (N = 185)
Trustworthiness	122	66%
Interoperability	94	51%
Licensing	91	49%
Completeness	76	41%
Understandability	63	40%
Provenance	72	39%
Timeliness	70	38%
Syntactic Validity	67	36%
Availability	59	32%
Conciseness	43	23%
Versatility	20	11%
None/Unsure	11	6%
Other	17	9%

3.3.3 Discussion

The findings of the survey indicate that IPs consider the primary benefits of LD publication and consumption to include:

- Cross-institutional linking and integration resulting in additional context for data interpretation and improved cataloguing efficiency.
- Improved data discoverability and accessibility.
- Enriched metadata and improved authority control.

With regards to the challenges of using LD, the overarching barriers for LAMs in publishing and consuming LD reported by the participants were:

- Interlinking and Integration Issues: difficulty selecting appropriate ontologies and link-types, and difficulty with entity resolution and vocabulary mapping.
- LD Tooling Issues: inadequate for the requirements of the library domain, technologically complex and difficulty integrating into cataloguing workflows.
- Resource Quality Issues: LD datasets and URIs maintenance issues, lack of LD guidelines and use cases for LAMs, and difficulty creating URIs.

Interestingly, these benefits and challenges confirm those reported in the LD surveys reviewed in Section 3.2.1. However, the questionnaire data allowed for a deeper investigation into these challenges, and also provided data on potential solutions.

A more in-depth exploration of the 'Interlinking and Integration' issue indicated that IPs are experiencing difficulties in selecting appropriate link-types when interlinking and also in deciding which datasets to interlink to and integrate with. A cause of these difficulties may be a lack of provenance data, as some of the aforementioned decision-making issues could be somewhat alleviated if IPs had access to data describing the origin of the metadata.

With regards to the 'LD Tooling Issues', participants specifically mentioned that tools are often challenging to learn and to use, inadequate for use in LAMs, and difficult to integrate into cataloguing workflows.

Finally, in relation to ‘Resource Quality Issues’, participants highlighted concerns with the reliability of available LD resources – another issue that may be resolved with the provision of provenance data.

Overall, the results of this questionnaire provided a detailed overview of the challenges IPs face when using LD, as well as more in-depth exploration of the issues faced when interlinking. In response to these challenges, the vast majority of participants (77%) agreed a LD interlinking tool, designed specifically for IPs, could be useful. The most commonly cited reasons for this being that a bespoke tool could help overcome the technical knowledge gap of IPs, make LD more accessible to IPs, increase the number of LAMs using LD, and create new research opportunities. Participants also provided a detailed list of functions that a bespoke interlinking tool for LAMs should have.

3.4 Chapter Summary

This chapter presented a review of six LD interlinking frameworks, two surveys exploring the use of LD in LAMs, and twelve LAM LD services. Based on this review, a LD Requirements Questionnaire was developed in order to further explore the issues identified. The results of the questionnaire provided a strong basis and justification for the development of NAISC-L, with a number of user requirements being distilled from its results. These requirements are discussed in Section 4.1.

4 The NAISC-L Framework

This chapter details NAISC-L, a framework for **N**ovel **A**uthoritative **I**nterlinking for **S**emantic Web **C**ataloguing in **L**ibraries. In link with the Relevance Cycle of the Design Science Approach, Section 4.1 presents a set of user requirements for NAISC-L which were distilled from the State-of-the-Art Review and LD Requirements Questionnaire presented in Chapter 3. Following this the NAISC-L Framework and all of its components are discussed in detail in Section 4.2. This chapter is then summarised in Section 4.3.

4.1 Linked Data Interlinking Tool Requirements

In line with the Relevance Cycle of the Design Science Model (Hevner, 2007), a set of user requirements for the development of a LD interlinking framework for LAMs were distilled from the State-of-the-Art Review and the results of the LD Requirements Questionnaire.

R1. **The interlinking framework should facilitate the creation of identity and relationship links.** As highlighted in Section 2.1, existing interlinking tools primarily facilitate the creation of identity links. However, research indicates that the `owl:sameAs` property is being used erroneously to create links between entities which do not refer to exactly the same thing (Ding, Shinavier, Finin & McGuinness, 2010; Heiko, 2014; De Melo, 2013). Facilitating the creation of relationship links may reduce these errors and also enrich the type of knowledge associated with linked entities.

R2. **The interlinking framework should be designed with the needs and expertise of IPs in mind.** In the LD Requirements Questionnaire, IPs reported that current LD tooling is often inadequate for use in LAMs and technologically complex. Participants indicated that designing a tool for the LAM domain would make LD more accessible to LAMs, and that a bespoke LD interlinking tool would be useful. Participants suggested variety of features that a LD interlinking tool for LAMs should have, the majority of which were incorporated in the design of NAISC-L. These included that the tool should be adaptable to LAM workflows, hide complex LD technicalities, have a user-friendly interface, have an awareness of common LAM data sources, enabling linking to LAM controlled vocabularies, provide data quality scores, and incorporate link-types from LAM ontologies. General-purpose LD tools do not provide these LAM specific features, as such, a bespoke tool is necessary to meet the needs of IPs.

Some features, suggested by the IPs who completed the LD Requirements Questionnaire, which were not included in the design of NAISC-L included vocabulary alignment, SKOS vocabulary creation, automatic linking to ontologies, and data cleaning. Many of these features were related to the creation of vocabulary links and were thus beyond the scope of this thesis. Features such as data cleaning, which is not specifically related to the task of LD interlinking, were also not included in the design of NAISC-L.

R3. **The framework should provide provenance data for the interlinks generated.** As discussed in Section 2.2, given anyone can publish to the SW, when interlinking resources, data provenance is needed to establish the trustworthiness and quality of the data (Dezani-Ciancaglini, Horne & Sassone, 2012). Of the tools reviewed in Section 3.1, none published LD provenance for the interlinks generated. Of the LD services reviewed in Section 3.2.2, only two provided LD provenance for the published resources. In the LD Requirements Questionnaire above, participants reported having difficulty in deciding which datasets to interlink with as well as concerns with the quality of LD resources – both issues that could be lessened if IPs had access to provenance data for the resources.

4.2 NAISC-L Framework

With the above requirements in mind, the NAISC-L Framework was developed. In line with the Design Cycle of the Design Science Model, NAISC-L was iteratively designed and refined based on the results of three user evaluations described in Section 5. Figure 15 displays the role of NAISC in the architecture of a LD application (Heath & Bizer, 2011). As mentioned, the framework consists of an interlinking process and a tool. The tool includes an approach to knowledge organisation, a provenance data model and a GUI. Each component of the framework will be discussed in turn in the following sections.

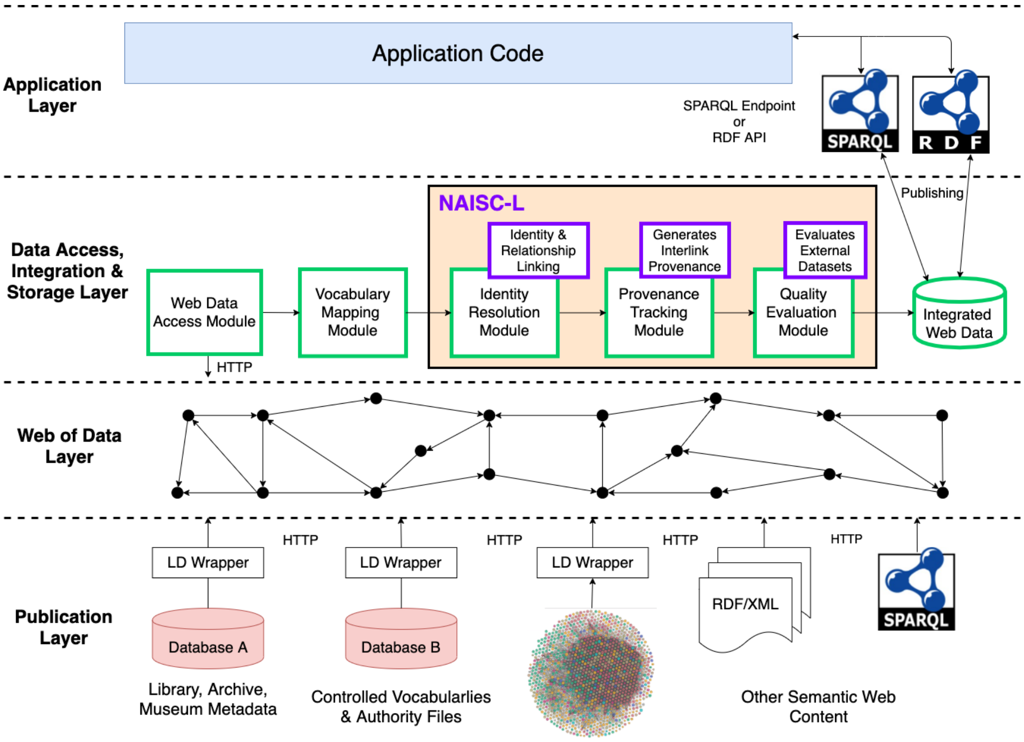


Figure 15: Role of NAISC-L in the Architecture of a LD Application

4.2.1 NAISC-L Interlinking Process

The NAISC-L interlinking process consists of four cyclical steps, as seen in Figure 16 below. These steps are Entity Selection, Link-Type Selection, Provenance Data, and RDF Graph Generation and Visualisation.

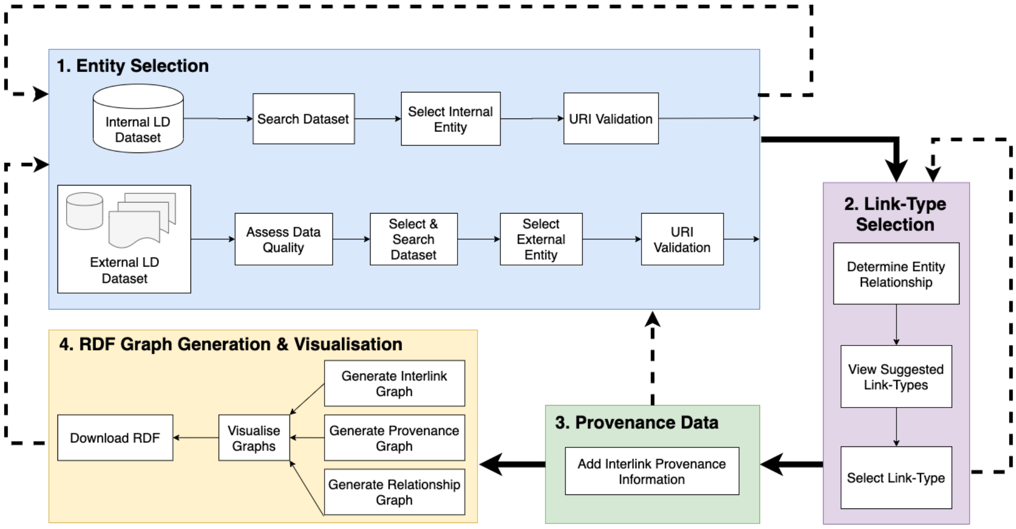


Figure 16: NAISC-L Interlinking Process

The interlinking process draws upon prior work related to ontology mapping workflows (Falconer & Storey, 2007; Euzenat & Shvaiko, 2013; Debruyne, Walshe & O'Sullivan, 2015). These workflows are used for ontology matching and alignment, resulting in vocabulary links. Although the NAISC-L Interlinking Process can be mapped onto these workflows, in order to meet NAISC-L's user-requirements, certain distinctions were specified. The ontology mapping workflows do not stipulate whether the interlinking approach is manual or (semi-)automatic, and also the links are chosen after correspondences are found. The NAISC-L Interlinking Process, however, uses manual matching and alignment of entities by guiding users through the process of selecting a link-type. Provenance also plays a more prominent role in the NAISC-L Interlinking Process in response to R3 above, which relates to establishing the authoritativeness of the data.

Step 1 and Step 2 of the interlinking process are manual. Existing interlinking mappings and algorithms successfully facilitate semi-automatic instance matching, however, as mentioned, there is a need to facilitate the creation of relationship links. Relationship linking requires a knowledge of the entities being interlinked, a knowledge of the context in which the interlink is being created and a knowledge of the purpose of the interlink. Few datasets would have sufficient information available for such relationship links to be declared and discovered automatically. As such, a manual approach was used for the design of NAISC-L.

By placing the Entity Selection step first, followed by the Link-Type Selection step, users have a holistic view of the entities to be linked and are able to make an informed decision of the most appropriate link-type that should be used to create a meaningful interlink. A justification is then manually added, as part of the Provenance Data step as, again, this data is contextual and unique for each interlink. The provenance data, the RDF graphs, and the visualisations are then generated automatically.

Each step of the interlinking process is described in greater detail below.

Step 1 – Entity Selection

This step involves selecting an entity, from an internal dataset, from which to create an outward link. The URI of the selected entity is then validated by NAISC-L. Note, NAISC-L presumes that a LD dataset already exists within a given institution. The user then chooses an external dataset to interlink with. Dataset quality ratings are provided to aid in the decision-making process. From this external dataset, the user selects an entity to interlink with. Again, the URI of the selected entity is validated. This step can be repeated so as to create a linkset – a collection of interlinked entitles.

The external dataset quality rating is based on three quality metrics – Trustworthiness, Interoperability and Licensing. These metrics were chosen as they were the top three quality criteria used to evaluate external data sources, as selected by participants of the LD Requirements Questionnaire (see Section 3.3). The aim of providing this score was to assist users in selecting high-quality resources to interlink with. The datasets included in the rankings were the most commonly consumed LD resources mentioned in the LD Requirements Questionnaire, as well as the LD services discussed in the State-of-the-Art Review (see Section 3.2.2).

Users also have the option of defining entities as per the FRBR entity-relationship model. FRBR is a library reference model developed by the International Federation of Library Associations and Institutions (IFLA) that aims to provide a logical structure to bibliographic information (Riva, La Boeuf & Žumer, 2016). In FRBR, entities are divided into three groups. A Group 1 entity, or Endeavour, is a work, expression, manifestation, or item which is the product of an intellectual or artistic undertaking. A Group 2 entity, or Responsible Entity, is a person, family or corporate body responsible for the creation, dissemination or custodianship of an Endeavour. Group 3 entities serve as Subjects of an Endeavour and can be a concept, object, event, place or any Group 1 or Group 2 entity. FRBR was incorporated in NAISC-L to aid users in selecting an appropriate link-type as some properties are specific to the category of entity being described. Additionally, the model was used in three of the LD services reviewed in Section 3.2.2.

Step 2 – Link-Type Selection

Step 2 takes the user through the process of creating an interlink between an Internal Entity and a Related Entity. This is done in two stages which are presented to the user via an Interlinking Guide.

Interlinking Guide

The first stage of the Interlinking Guide requires the user to determine the kind of relationship that exists between the entity pair. To do this, the user selects one of six natural language Relationship Terms. Each Relationship Term is defined and the user should select the definition that most accurately describes the connection between the entity pair (see Figure 34, Section 4.2.2.3). The terms were inspired by the types of identity and similarity links identified by Halpin et al. (2010) in their analysis of `owl:sameAs` statements on the SW.

The Relationship Terms are:

- **Is Identical To:** In this case, the Internal Entity and the Related Entity are exactly the same i.e. the URIs represent the same thing/person/place/item. The properties ascribed to both entities are also exactly the same. In other words, these entities can be substituted for each other and all statements made about the entities remain true in either context. Defining the relationship between two entities as Identical indicates that all properties ascribed to, and statements made about, the Related Entity are also true for the Internal Entity.

- **Is Identical In Certain Contexts To:** In this case, the Internal Entity and the Related Entity are considered to be the same only in a specific context. Here context could refer to a given time, place or perspective. For instance, from January 20, 2009 to January 20, 2017, the statement 'Barack Obama is the same as the President of the United States' would be true. However, this statement would not be correct in the context of a different timeframe. Another example is, in the context of Twitter, it would be appropriate to refer to a person using their Twitter handle. However, it would likely be inappropriate to use this handle as a substitute for a person's name in a place outside of the social media platform.

- **Is Almost Identical To:** In this case, the Internal Entity and the Related Entity represent the same thing/person/place/item, however, all of the properties ascribed to, or the statements made about, the Related Entity may not be true for the Internal Entity. Therefore, even though the URIs represent the same thing, the entities cannot be substituted for one another. For example, different datasets may provide slightly different longitude and latitude coordinates for the same place. Thus, although referring to the same place, the properties of one entity are different to the other meaning the entities are not exactly the same or interchangeable.

- **Is Similar To:** In this case, the Internal Entity and the Related Entity do not represent the same thing/person/place/item, however, they do represent something very similar and have many, but not all, properties in common. For example, the prime version of the painting Virgin of the Rocks by Leonardo da Vinci, which is held in the Louvre, is similar to the later version of the painting of the same name held in the National Gallery, London.

- **Is Associated With:** In this case, the Internal Entity and the Related Entity are not identical and share little or no properties in common. However, the entities are closely aligned in some fashion and can be associated with each other in certain contexts. This type of link would be useful in creating pointers to external entities that may be of interest to those researching the internal entity.

- **Is Different To:** When two URIs represent distinctly different entities but these entities may be easily confused for one another. This could be used to state that two people or items, with the same name, are in fact two distinct entities.

The second stage of the Interlinking Guide is to select a link-type to connect the Internal Entity and the Related Entity pair. The link-types presented to the user are narrowed down depending on the Relationship Term selected. The suggested link-types are taken from vocabularies commonly used in LAMs, as identified in the LD Requirements Survey, Section 3.3.

Step 3 – Provenance Data

Provenance data describing by whom, where, when and how an interlink was created is automatically generated by NAISC-L. With regards to 'why' an interlink was created, this justification data is supplied by the user after selecting a link-type. This justification can include, but is not limited to, a description of the relationship between the entities, the purpose of the interlink, the interlink context, and the rationale behind the chosen link-type. The data is structured as per the NAISC-L Provenance Data Model, described below in Section 4.2.2.2.

Step 4 – RDF Graph Generation and Visualisation

NAISC-L data is stored in a relational database (RDB) and is uplifted to RDF using R2RML, a W3C Recommendation used to express mappings from RDBs to RDF (Das, Sundara, & Cyganiak, 2012). NAISC-L's Knowledge Organisation, detailed in Section 4.2.2.1, consists of three named graphs – an interlink graph, a provenance graph and a relationship graph. The data for each graph is uplifted to RDF using a separate R2RML mapping (see Appendix 17). These mappings were created using the JUMA mapping tool (Crotti, Debruyne

& O'Sullivan, 2017, 2018). The graphs can be viewed and downloaded in TriG, NQuads and TriX format. Interactive visualisations of the graphs are also generated using GoJS[114] software. Samples of the RDF graph and visualisation output can be seen in Figures 37, 38 and 39 in Section 4.2.2.3.

4.2.2 NAISC-L Tool

As mentioned previously, the NAISC-L Tool consists of an approach to knowledge organisation, a provenance data model and a GUI, all of which are detailed in the following sections.

4.2.2.1 NAISC-L Knowledge Organisation

NAISC-L's Knowledge Organisation, Figure 17, comprises of three named graphs – an interlink graph, a provenance graph and a relationship graph. NAISC-L's Knowledge Organisation was designed to facilitate simple and efficient querying of interlinks created using NAISC-L, as these links would be the most utilised component of NAISC-L's output.

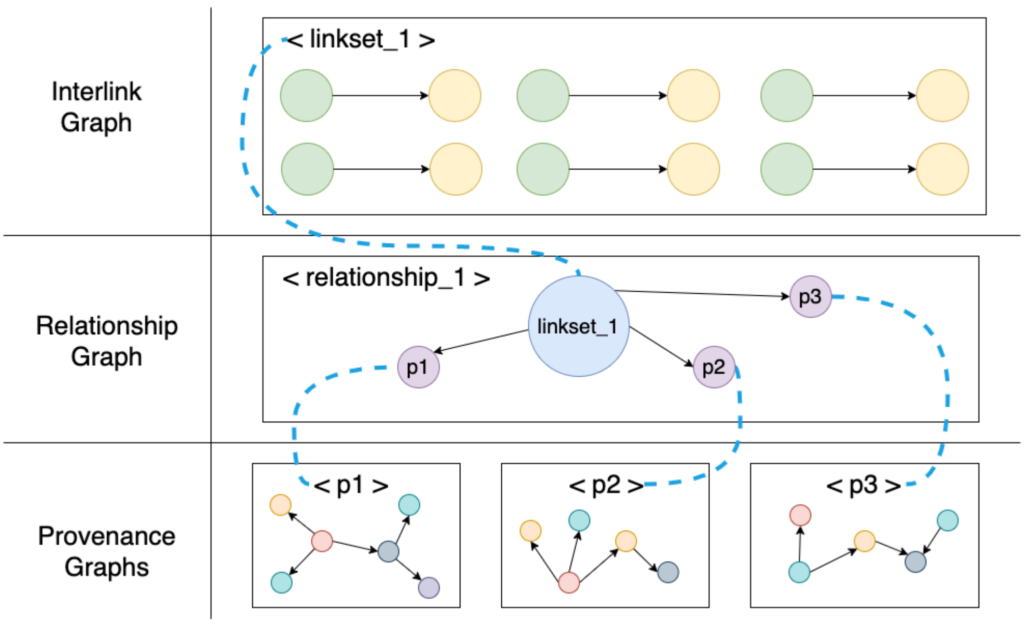


Figure 17: NAISC-L Knowledge Organisation

[114] https://www.nwoods.com/products/gojs/index.html accessed August 3rd 2020

1. **Interlink Graph:** This is a named graph containing a collection of interlinks known as a linkset. When changes are made to the linkset in NAISC-L, these changes are reflected in the Interlink Graph once an interlinking session is complete. Interlinking sessions are controlled by NAISC-L users and are completed when users actively update the Interlink Graph with the additions, deletions or revisions they have made to the linkset. A linkset has only one named graph that contains all of its active interlinks. An active interlink refers to an established interlink that has not been deleted or revised. Thus, in the context of NAISC-L, a linkset and an Interlink Graph are interchangeable. This design allows for simple and efficient querying of the interlinks.

2. **Provenance Graph:** This is a named graph, in the form of a `prov:Bundle`, that contains the provenance data of the links in an Interlink Graph. Additionally, as a `prov:Bundle` is itself an entity, the provenance of the Provenance Graph is also captured. Multiple provenance graphs can be associated with one Interlink Graph, as a new provenance graph is created for every interlinking session. A Provenance Graph contains the origin data of the interlinks created during an interlinking session, as well as the origin data for the linkset itself. It also provides a history of the interlink deletion and revision activities that occurred during an interlinking session. These descriptions are provided using RDF Reification. Although using RDF Reification results in more complex provenance queries, structuring the graphs as such simplifies the querying of interlinks.

3. **Relationship Graph:** This is a named graph containing a set of statements linking an Interlink Graph with its Provenance Graphs using the property `prov:has_Provenance`. This property, which is part of PROV-AQ: Provenance Access and Query[115] (Moreau, Hartig, Simmhan, Myers, Lebo, Belhajjame, Miles & Soiland-Reyes, 2013), specifies how to obtain a provenance record associated with a resource.

---

[115] https://www.w3.org/TR/prov-aq/ accessed 18th August 2020

Separating the data across the three graphs simplifies some of the queries that users can formulate and run over the data, while still allowing for queries that span across graphs, as facilitated by the relationship layer. The Interlink Graph allows a user to view and query a linkset. Should the user wish to review the provenance of a linkset or an interlink, the Relationship Graph can be used to direct the user to the associated Provenance Graphs. The user can then search for the provenance of a particular interlink within the Provenance Graphs, as interlinks are described using RDF Reification. The user can also query interlink revisions and deletions within the Provenance Graphs. Sample SPARQL queries can be found in Figure 43 in Section 4.2.2.3.

Other designs that were considered for NAISC-L's Knowledge Organisation included creating an Interlink Graph for every interlinking session and then linking this Interlink Graph to a Provenance Graph (also created for each interlinking session). This would generate multiple Interlink Graphs for a single linkset. This structure was rejected as it would result in interlinks, that have been revised or deleted from the linkset, remaining in the Interlink Graphs. Thus, querying interlinks would be more complex as users would have to query the status of an interlink before using it. Additionally, users would have to query across multiple named graphs when exploring the interlinks.

Another design that was considered involved creating a named graph for each interlink created, again resulting in multiple Interlink Graphs per linkset. These Interlink Graphs would be linked to a Provenance Graph created for the interlinking session. Thus, a Provenance Graph could be linked to multiple Interlink Graphs. In this design, the Provenance Graph would use the named graphs, instead of RDF Reification, to make statements about statements. Although querying the provenance information would be somewhat simplified using this method, querying the interlinks would be more complex, as it would involve querying multiple graphs. As such, this design was also rejected.

4.2.2.2 NAISC-L Provenance Data Model

A set of requirements for the design of the provenance data model were distilled from the results of the LD Requirements Questionnaire discussed in Section 3.3. These requirements state that the provenance data should:

- Allow for different levels of granularity i.e. viewing the provenance at different levels, for example, at the link level or the linkset level.
- Keep track of revisions made to the interlinks.
- Link to sources used in the dataset.
- Link to people, organisations, and groups that contributed to the dataset.
- Allow for the justification of the sources used to create a link.
- Allow for the justification of the type of link created between resources.

Further requirements for the provenance model were established using a set of ontological competency questions (Gruninger & Fox, 1995; Bezerra, Freitas & Santana, 2013), see Table 12. The competency questions were inspired by common requirements for data provenance on the SW (Groth, Cheney & Miles, 2012) and in LAMs (Li & Sugimoto, 2014).

Table 12: Interlink Provenance Competency Questions

Who created the interlink?	Who revised the interlink?
How was the interlink created?	How was the interlink revised?
Why was the interlink created?	Why was the interlink revised?
Where was the interlink created?	Who created the linkset?
When was the interlink created?	When was the linkset created?
Why was the interlink created?	Who created the provenance data?
When was the interlink revised?	When was the provenance data created?

The PROV Ontology (PROV-O), described in Section 2.2.1, was used as the foundation of the provenance data model. Existing PROV-O classes, sub-classes and properties were used to describe by whom, where, when and how interlinks were created. PROV-O was extended in order to describe why an interlink was created, and to provide additional details on how it was created. This extension, called NaiscProv, includes the addition of interlink specific subclasses and properties – see Figure 18. Specifically, a new Entity subclass, `naiscProv:Interlink`, was created in order to identify the reified statements as interlinks. Using the URI of the reified interlink, the user can then query the provenance of the link. The property, `naiscProv:hasJustification`, was added in order to provide 'why' provenance for a specific interlink. Additionally, two Activity subclasses were declared in order to describe how interlinks were created and deleted.

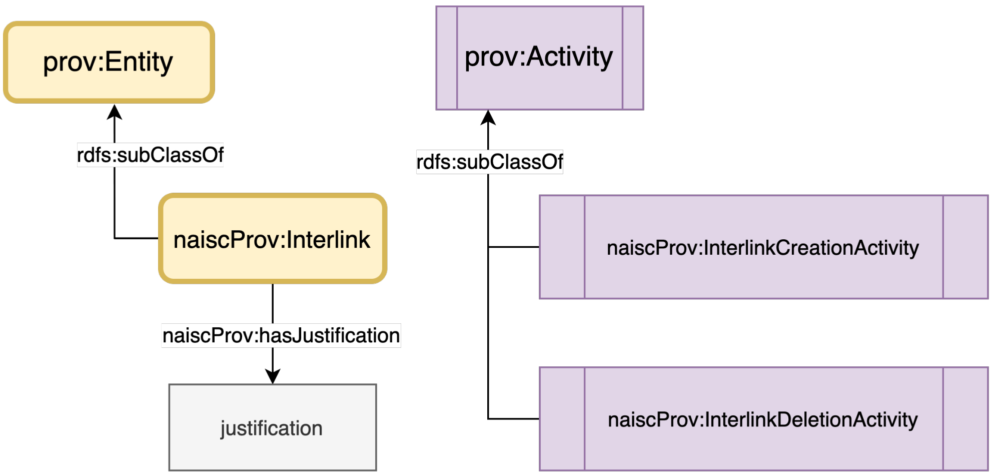


Figure 18: PROVO-O Extension – NaiscProv

In addition to PROV-O, The FOAF ontology (Brikley & Miller, 2014), a vocabulary of people-related terms, was used in the Provenance Model to provide the names of agents. FOAF was chosen as it is a highly used ontology on the SW, with 401 ontologies linking to it[116]. Also, Schema.org[117], a structured vocabulary for use on the web, was used to provide additional creator details such as their occupation. Again, this ontology was chosen as it is a well-used schema, with 77 ontologies linking to it[118].

Example 1, see Figure 19, presents a snippet of a provenance graph, `ex:ProvenanceGraph_1`, which displays how the provenance model is used to describe the creation of an interlink. In the graph it can be seen that a new interlink, `ex:Interlink_A`, was created as a result of a `nasicProv:InterlinkCreationActivity`. The interlink is an entity described using RDF Reification and assigned a URI. The URI is given properties identifying the entity as a `naiscProv:Interlink` and capturing

[116] https://lov.linkeddata.es/dataset/lov/vocabs accessed 16th August 2020
[117] https://schema.org/ accessed 10th August 2020
[118] https://lov.linkeddata.es/dataset/lov/vocabs/schema accessed 16th August 2020

the rationale of the interlink using the `naiscProv:hasJustification` property. Other provenance information included in the graph are the Agents responsible for creating the interlink, such as the software (`prov:SoftwareAgent`), the person (`prov:Person`), and the institution (`prov:actedOnBehalfOf`).

The interlink is part of a linkset, `ex:Linkset_1`, which is a specialised entity called a `prov:Collection`. The linkset is the interlink graph in which the interlink is contained. As the linkset is an entity, statements can be made describing its provenance, for example, when it was created and by whom.

The provenance graph, `ex:ProvenanceGraph_1`, is a specialised entity called a `prov:Bundle`, identified by a URI. As it is an entity, statements can be made about the provenance graph, allowing for the provenance of the provenance to be captured.

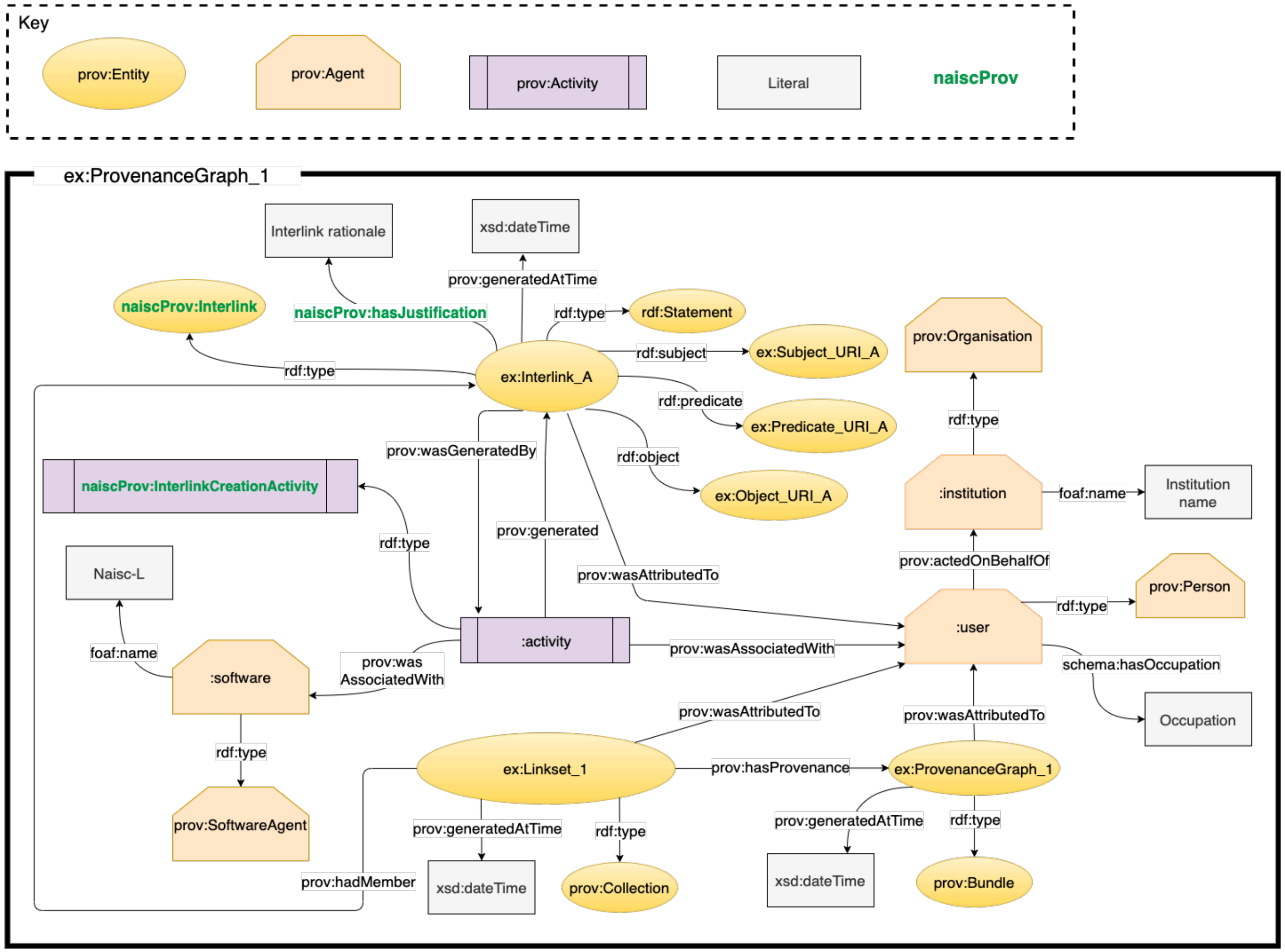


Figure 19: Provenance for the Creation of an Interlink

Example 2, see Figure 20, presents a snippet of how the provenance data model is used to represent an interlink revision. Note that the name of this provenance graph, `ex:ProvenanceGraph_2`, is different to the name of the provenance graph in the Example 1. This is because the interlink revision occurred as part of a different interlinking session, thus the provenance data was generated in a new graph.

In the graph it can be seen that a new interlink, `ex:Interlink_B`, was created as a result of a `nasicProv:InterlinkCreationActivity`. The new interlink, `ex:Interlink_B`, is part of the same linkset used in Example 1, `ex:Linkset_1`. Note that `ex:Interlink_B` is a revision of `ex:Interlink_A`, the interlink which was created in Example 1, which was invalidated as a result of a `naiscProv:InterlinkDeletionActivity`. As can be seen from this example, when an interlink is revised, the old interlink is invalidated and a new revised interlink is created. This new interlink points to the old interlink using `prov:wasRevisionOf`. This method was employed as it allows for the tracking of all changes to an interlink over time and the stores the preceding versions of an interlink for posterity. As the provenance data for `ex:Interlink_A` is held in the `ex:ProvenanceGraph_1`, a link to the graph is provided using `prov:has_provenance`. This allows users to trace the history of an interlink from graph to graph.

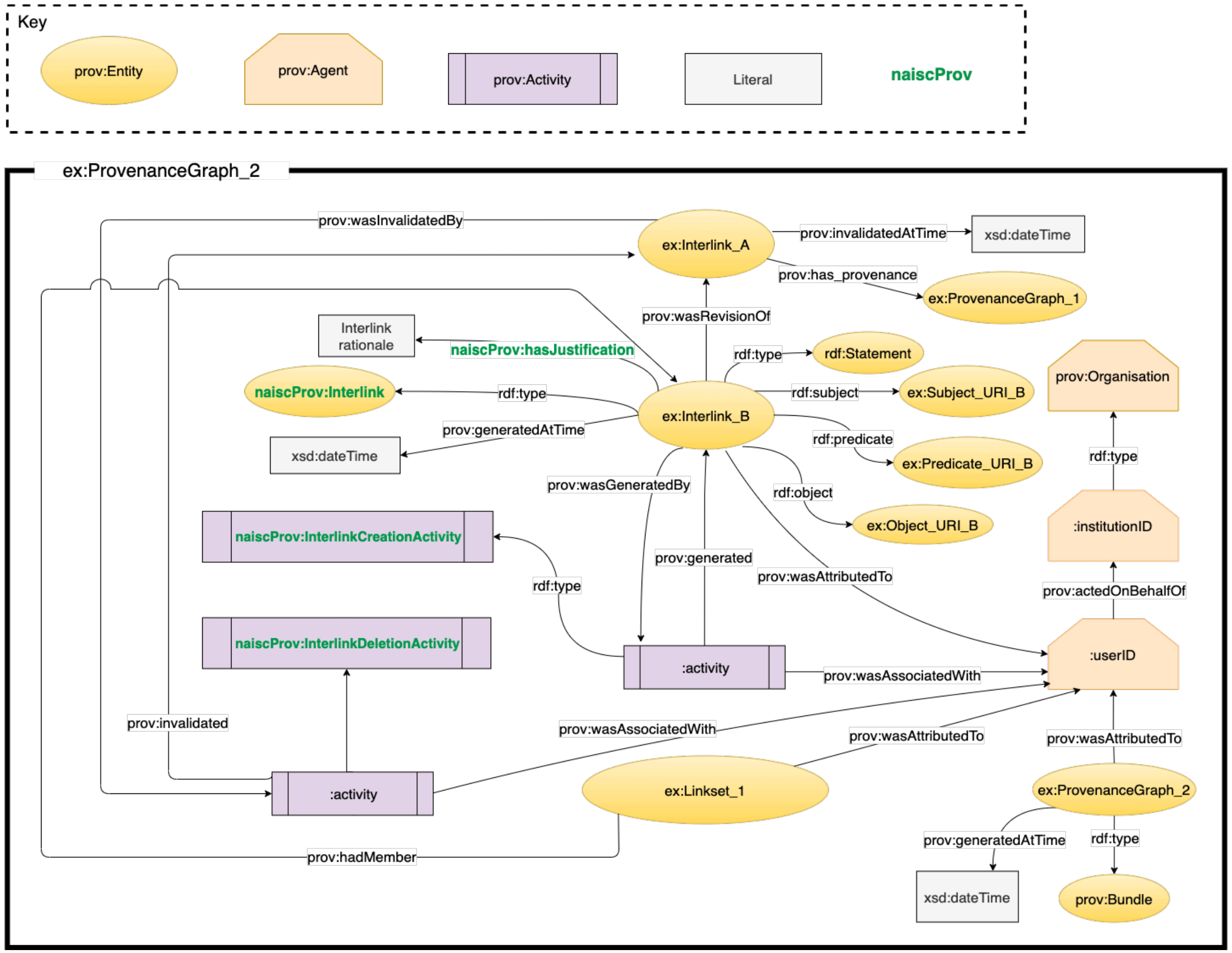


Figure 20: Provenance for the Revision of an Interlink

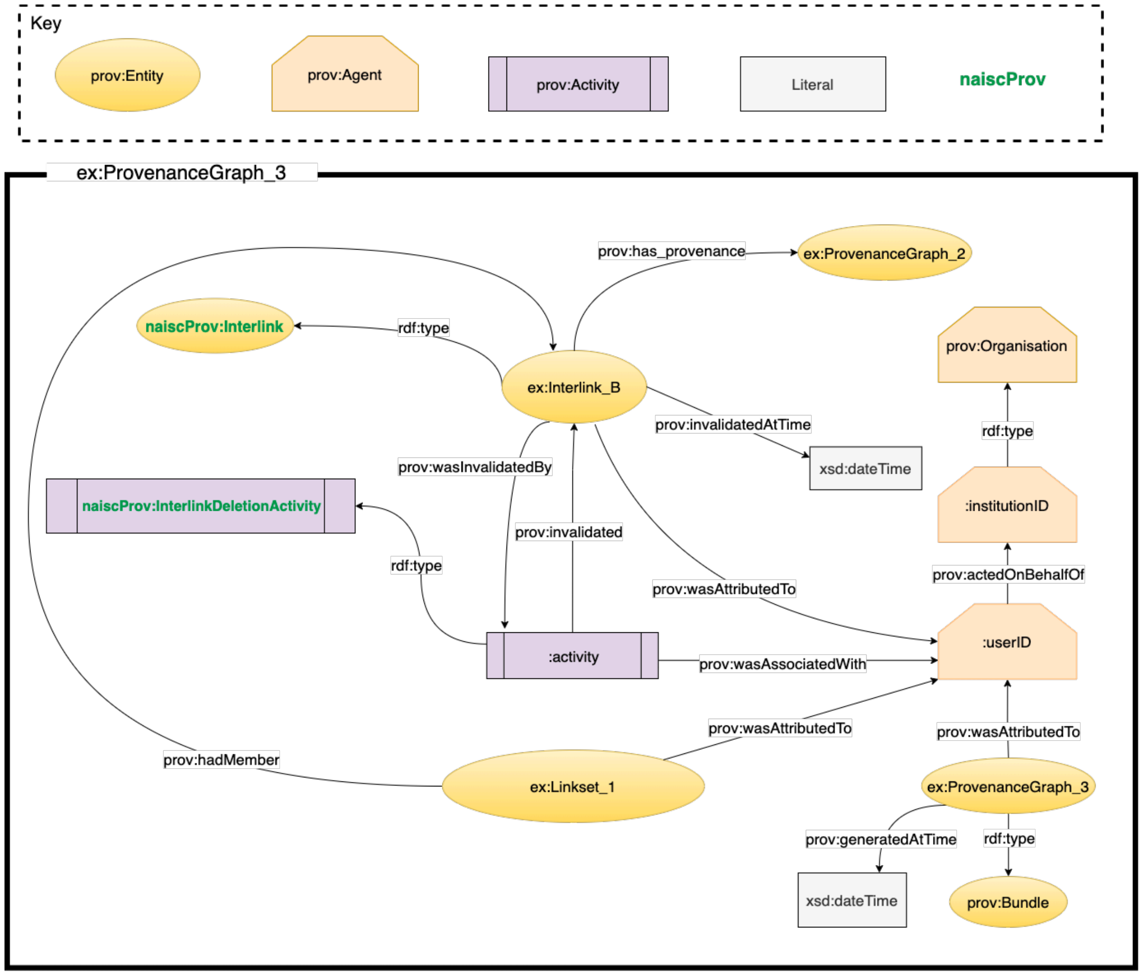


Figure 21: Provenance for the Deletion of an Interlink

Example 3, see Figure 21, presents a snippet of how the provenance data model is used to represent an interlink deletion. Note that the name of the provenance graph, `ex:ProvenanceGraph_3`, is different from the name of the provenance graphs in the previous examples. This is because the interlink deletion occurred as part of a different interlinking session, thus the provenance data was generated in a new graph.

In this graph, `ex:Interlink_B` had been invalidated as a result of a `naiscProv:InterlinkDeletionActivity`. As the provenance data for `ex:Interlink_B` is held in the `ex:ProvenanceGraph_2`, a link to the graph is provided using `prov:has_provenance`. This allows users to track the history of an interlink from graph to graph.

The three provenance graphs `ex:ProvenanceGraph_1`, `ex:ProvenanceGraph_2` and `ex:ProvenanceGraph_3`, used in the examples above, are linked to the Interlink Graph, `Linkset_1`, using the property `prov:has_provenance.` These triples are contained in a relationship graph, `ex:RelationshipGraph_1` – see Figure 22. Every named graph, or linkset, has one relationship graph. Note, the names of the provenance graphs have been presented in numerical order for this example only. Note, NAISC-L's knowledge organisation does not rely numerical order, instead, users should avail of the provenance information, such as the generation date of the provenance graph, to deduce the order of versions.

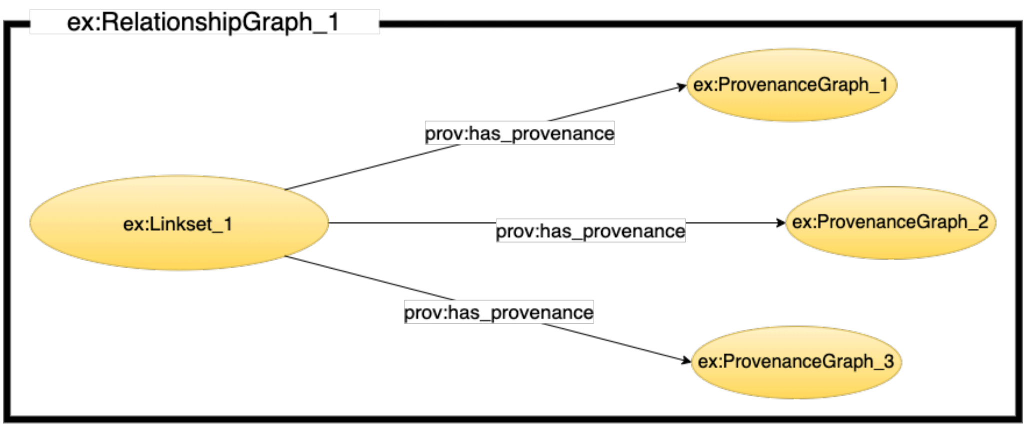


Figure 22: Relationship Graph

4.2.2.3 NAISC-L Graphical User Interface

The final component of the NAISC-L tool is the GUI. In line with the Design Cycle of the Design Science Model, the GUI was iteratively designed, testing and refined based on the results of the user evaluations discussed in Section 5. NAISC-L was built using Apache Tapestry[119], a component-oriented framework for creating web applications in Java, Bootstrap[120] CSS library, and a MySQL[121] database. Other, previously mentioned, technologies also used as part of the framework included R2RML (RDB to RDF mapping) and GoJS (data visualisation). A video demo of NAISC-L can be viewed on this webpage[122], and its code can be found on Gogs[123].

NAISC-L Mock-up

The first stage of the GUI design process involved designing an interactive mock-up of NAISC-L using PowerPoint[124]. This mock-up was used to evaluate the initial design ideas for the GUI – see Figure 23 and 24. The mock-up was reviewed by three IPs and their feedback was used to guide the development of NAISC-L Iteration 1.

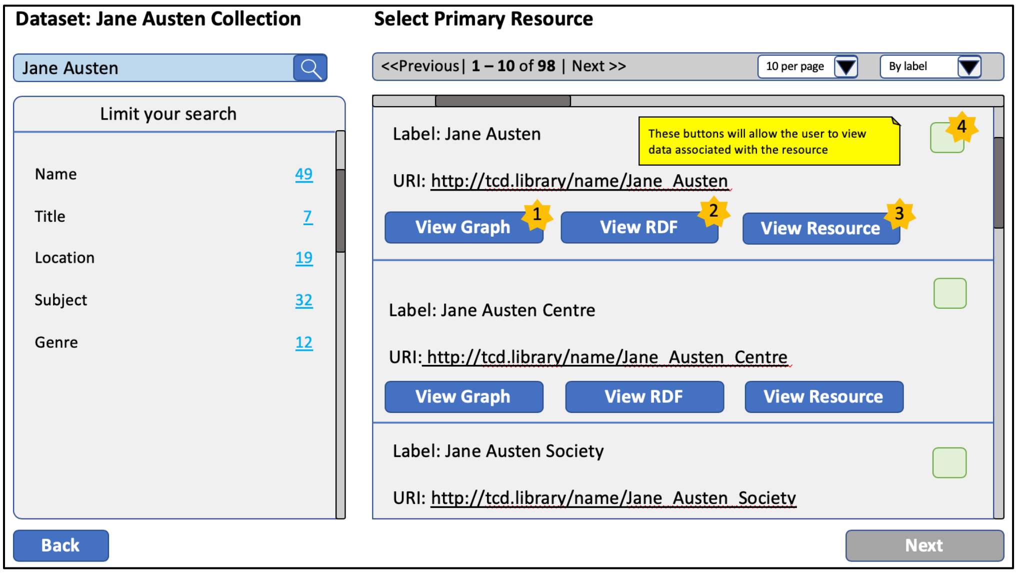


Figure 23: Mock-up for Internal Entity Selection

---

[119] https://tapestry.apache.org/ accessed 16th August 2020
[120] https://getbootstrap.com/ accessed 26th August 2020
[121] https://www.mysql.com/ accessed 16th August 2020
[122] https://www.scss.tcd.ie/~mckennl3/naisc/ accessed 16th August 2020
[123] https://gogs.adaptcentre.ie/mckennl3/NAISC accessed 16th August 2020. Access must be granted prior to viewing NAISC-L code – please email author for access.
[124] https://www.microsoft.com/ accessed 3rd August 2020

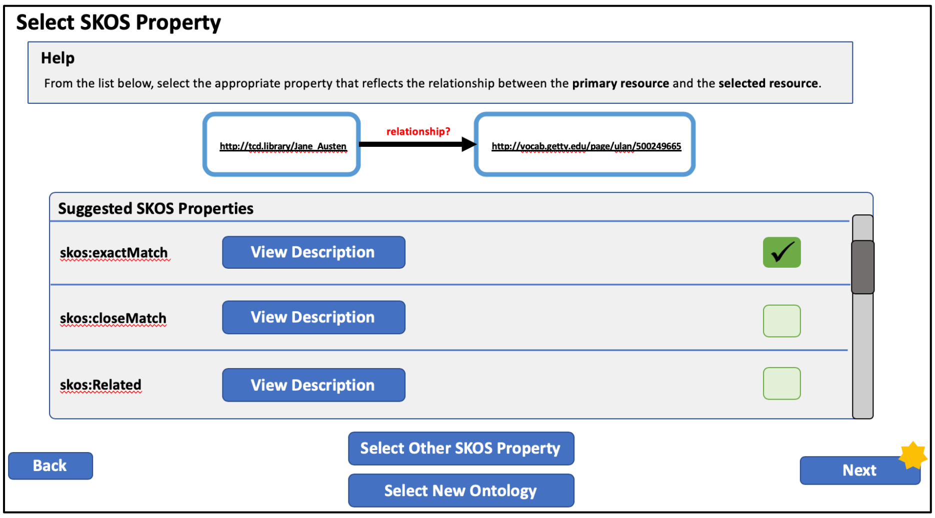


Figure 24: Mock-up for Link-Type Selection

NAISC-L Prototypes

The first prototype developed was NAISC-L Iteration 1. This prototype was then evaluated via Usability Test 1 (see Section 5.3) and the results were used to refine the tool, leading to NAISC-L Iteration 2.

Both NAISC-L Iteration 1 and Iteration 2 followed the interlinking process as described in Section 4.2.1. However, in Iteration 1, the terminology used to describe the process was slightly different. Where the terms Internal and Related Entity are used in the current version of NAISC-L, Primary and Secondary Resources were used in Iteration 1 – as can be seen in Figure 25 which shows the homepage of NAISC-L Iteration 1. This terminology was changed due to the fact that in LAMs a 'resource' has multiple meanings and users found this terminology to be confusing at times.

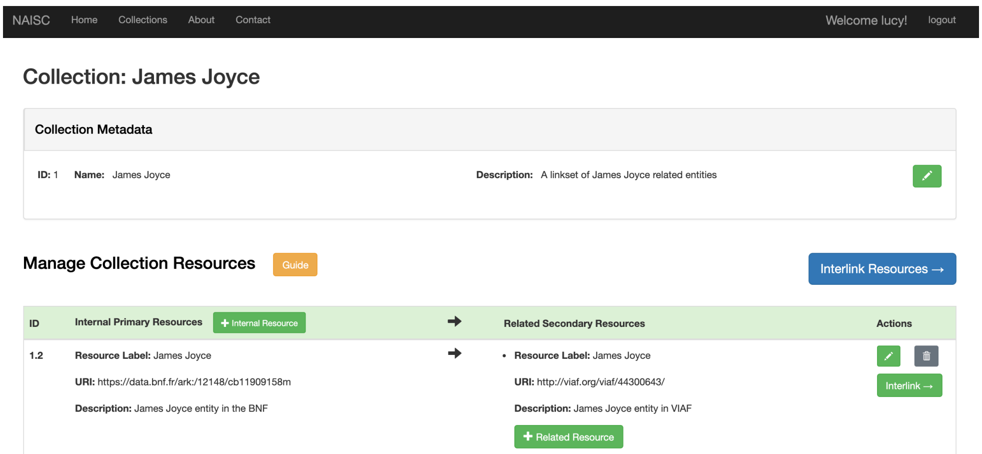


Figure 25: NAISC-L Iteration 1 – Homepage

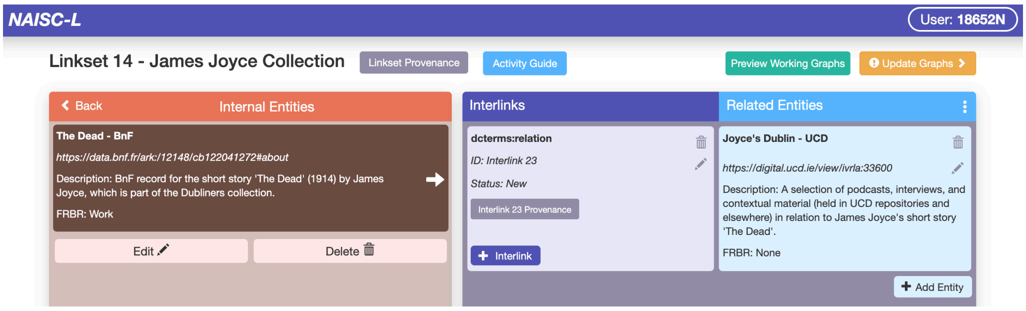


Figure 26: NAISC-L Iteration 2 – Homepage

Figure 26, which shows the homepage of NAISC-L Iteration 2, differed significantly to Figure 25, showing the homepage of Iteration 1. Firstly, as a result of user-feedback, the different components of each interlink were more clearly presented using colour coding and arrows. Secondly, Iteration 2 provided textual provenance information for the linkset and each interlink in a modal, see Figure 27, whereas this information was only available at the end of an interlinking session in Iteration 1. Thirdly, in order to aid in the interlinking process, Iteration 2 allowed users to view the interlink graph, both in RDF and via a visualisation, in a modal while they were creating links, see Figure 28. Again, this was only available at the end of an interlinking session in Iteration 1. These changes were made as a result of user feedback from Usability Test 1.

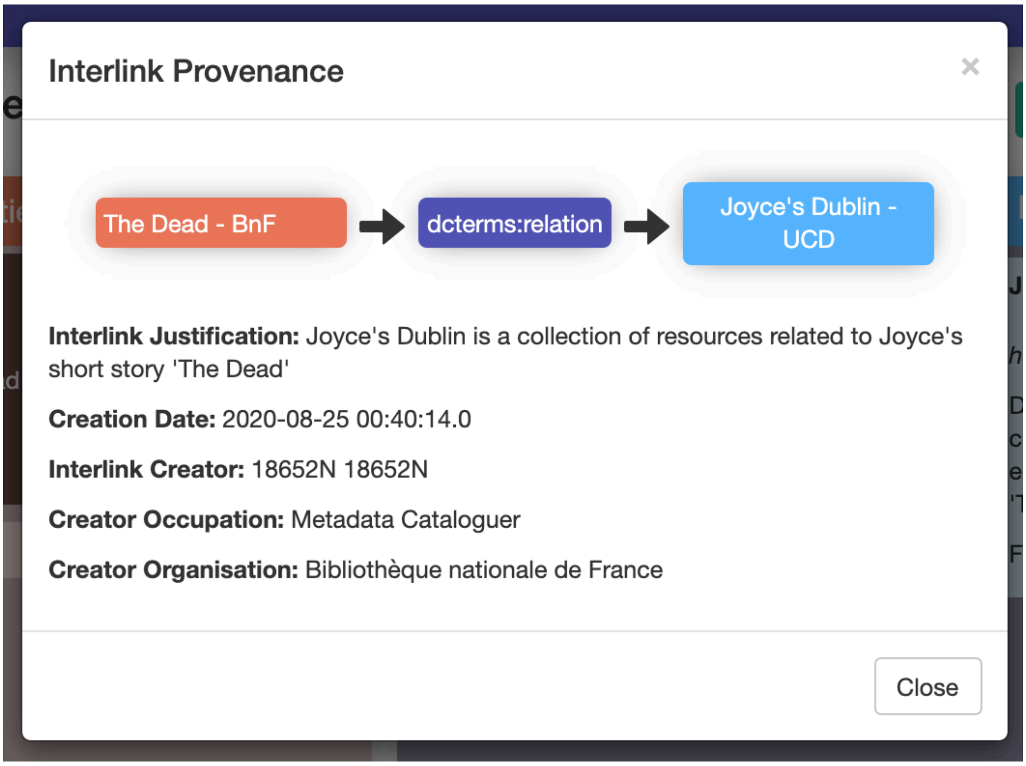


Figure 27: NAISC-L Iteration 2 – Interlink Provenance Modal

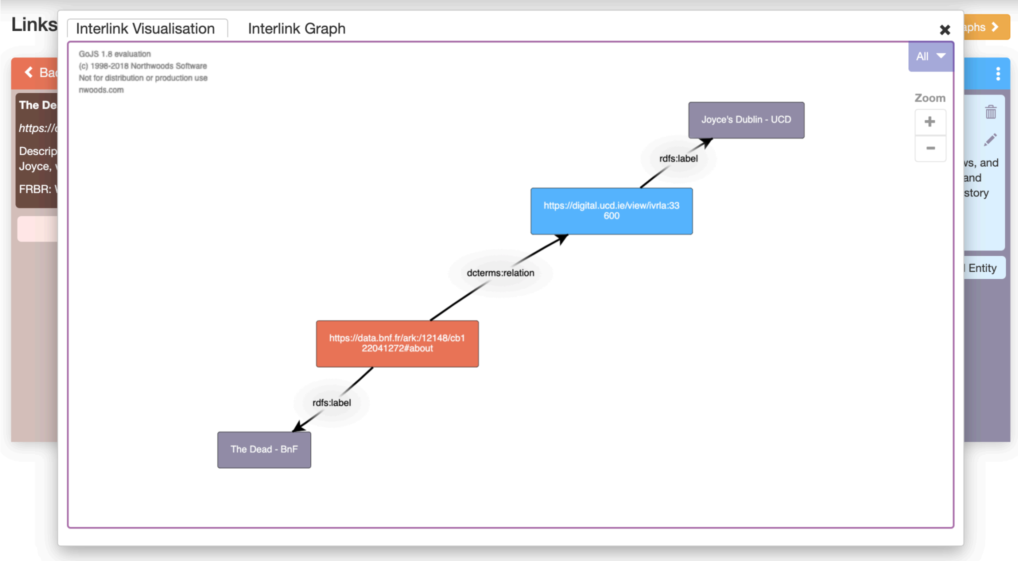


Figure 28: NAISC-L Iteration 2 – Interlink Graph Modal

In NAISC-L Iteration 1, each stage of the interlinking process was presented on separate pages, see Figures 29, 30 and 33. Whereas in NAISC-L Iteration 2, the entire process was presented on one page by using modals (popup child windows) – see Figures 31, 32 34, 35 and 36. This was done in order to provide increased efficiency and to reduce the navigation errors observed in the evaluation of NAISC-L Iteration 1. In addition, NAISC-L Iteration 1 used SemFacet[125] (Grau, Kharlamov, Marciuska, Zheleznyakov & Arenas, 2016), a semantic facet-based search tool, in order to search for Primary Resources/Internal Entities. SemFacet was not incorporated into Iteration 2 as users found it more intuitive to access and search the LD Web Service.

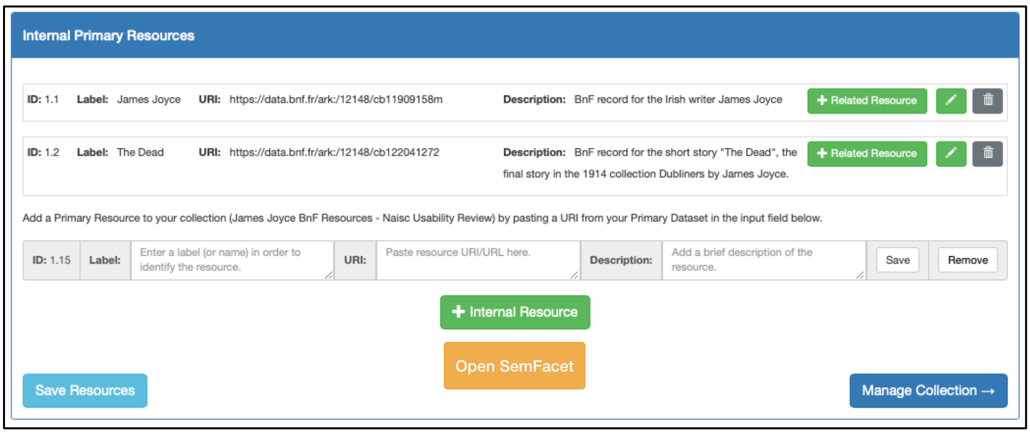


Figure 29: NAISC-L Iteration 1 – Add Primary Resource

[125] https://www.cs.ox.ac.uk/isg/tools/SemFacet/ accessed 3rd August 2020

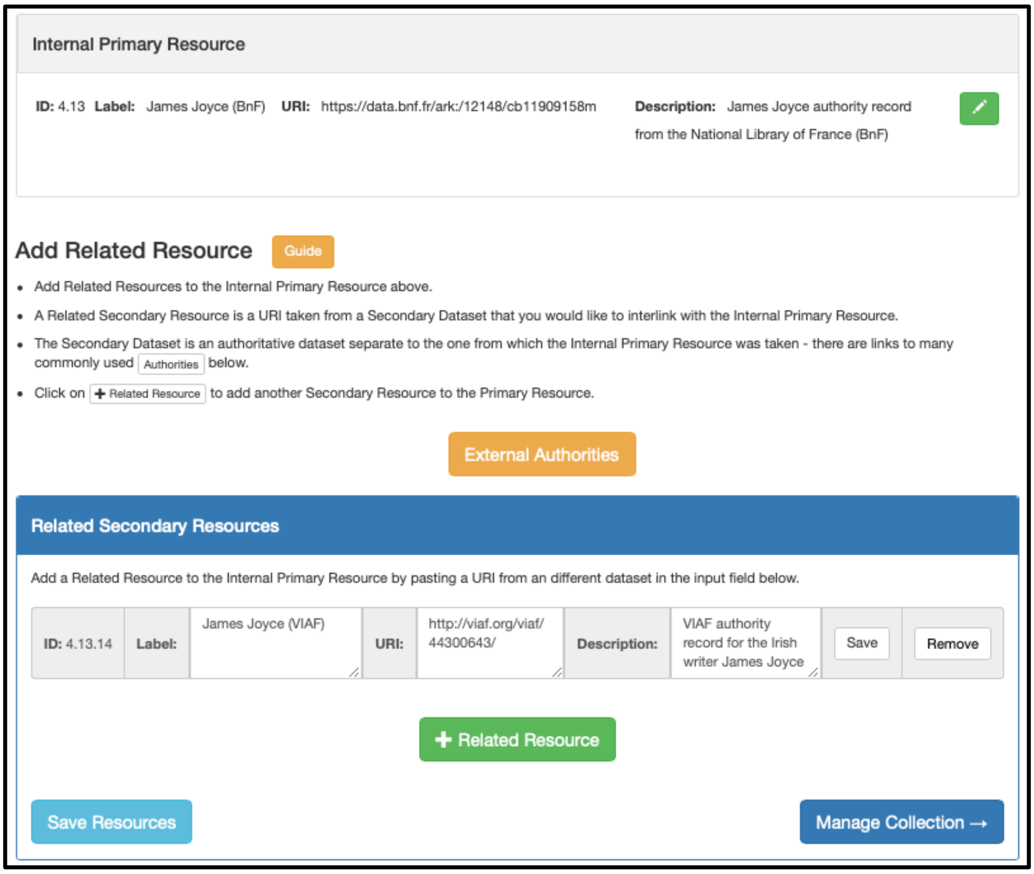


Figure 30: NAISC-L Iteration 1 – Add Secondary Resource

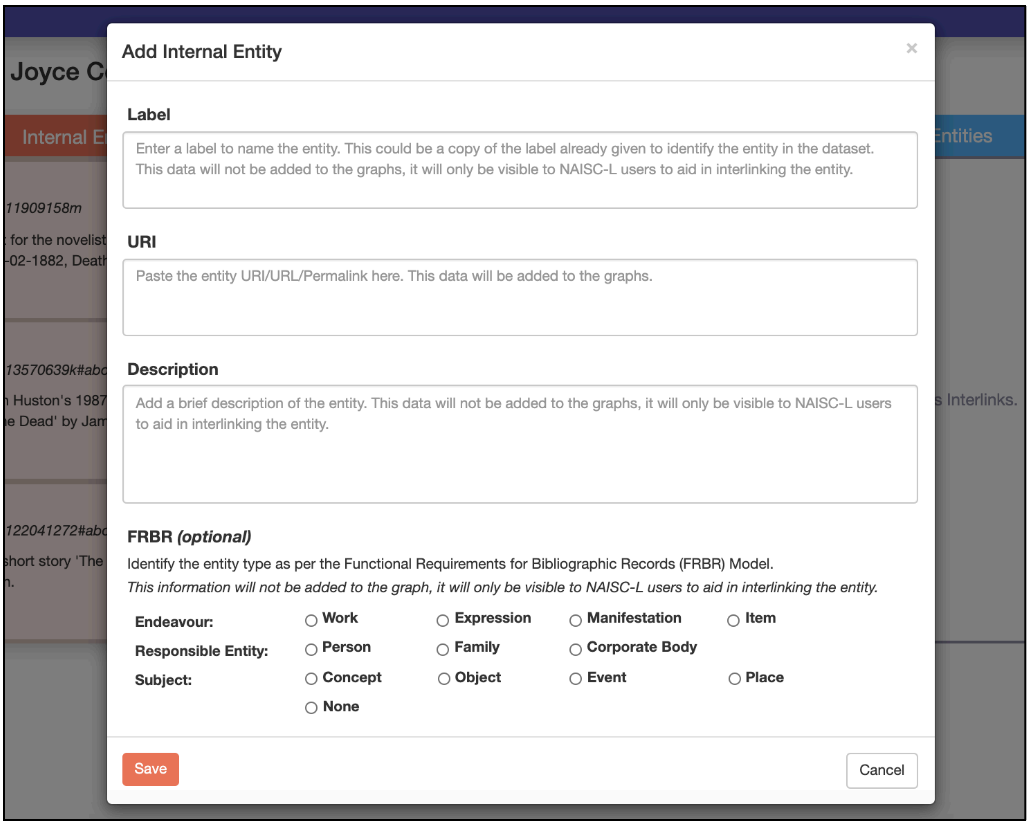


Figure 31: NAISC-L Iteration 2 – Add Internal Entity

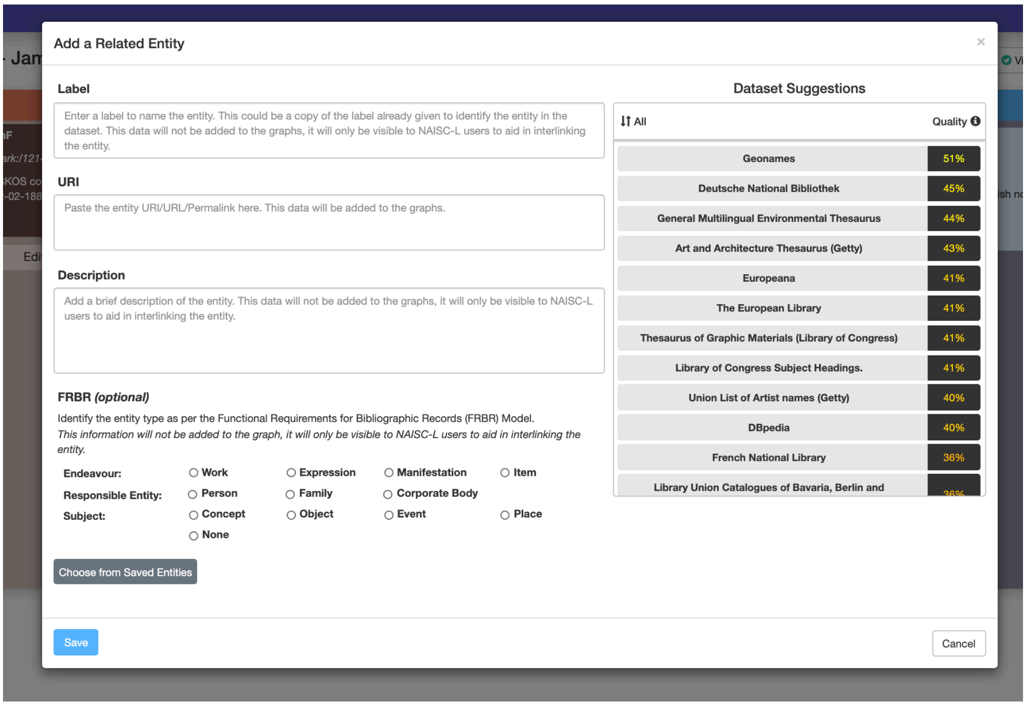


Figure 32: NAISC-L Iteration 2 – Add Related Entity

Another difference between iterations was that only NAISC-L Iteration 2 provided data quality scores for External Datasets, see Figure 32 as this feature was not yet developed for Iteration 1. Similarly, only Iteration 2 used the FRBR model to aid in the interlinking process.

The Interlinking Guide was available in both iterations. The Relationships Terms used in Iteration 1 were refined and altered for Iteration 2 based on user-feedback regarding terminology and providing more detailed definitions. The Interlinking Guide was presented on a single page in Iteration 1 – see Figure 33. However, the guide was presented in a series of modals in Iteration 2 so as to guide users through each stage separately in order to reduce errors – see Figures 34, 35 and 36. An additional feature in Iteration 2 facilitated users to pull link-types directly from Linked Open Vocabularies[126] (LOV) if their preferred link-type was not listed – see Figure 37.

[126] https://lov.linkeddata.es/dataset/lov/ accessed 3rd August 2020

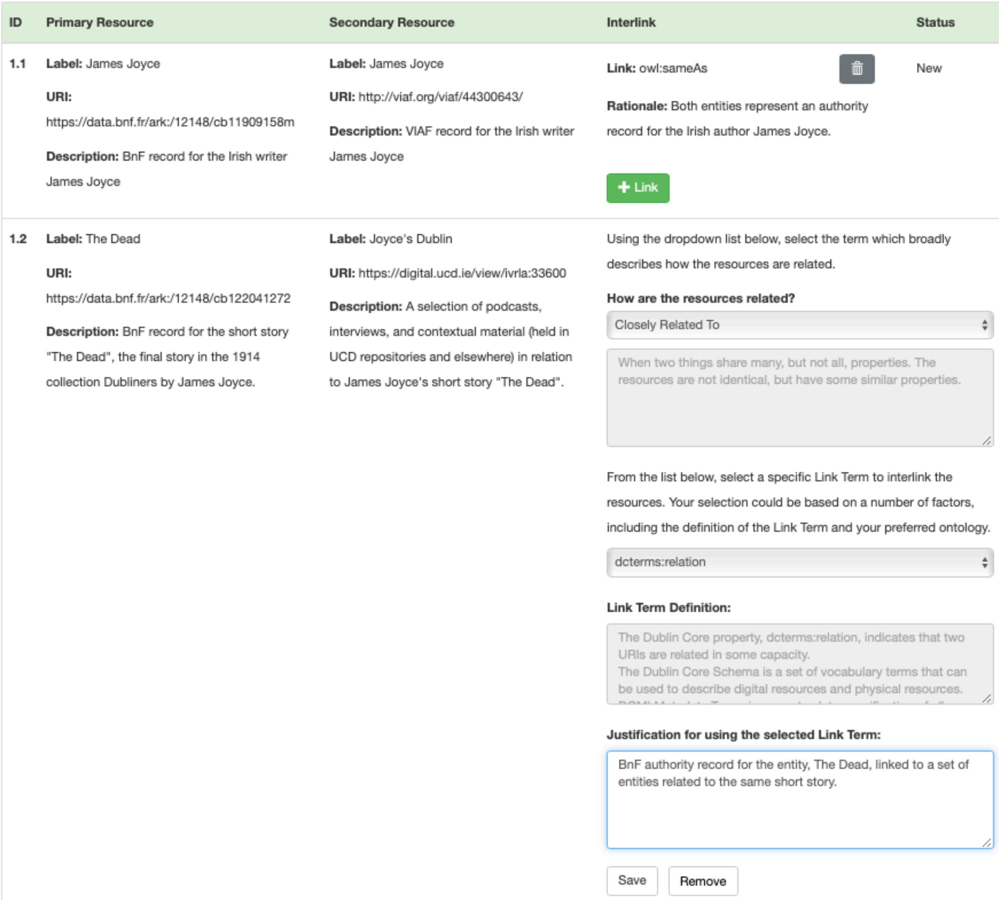


Figure 33: NAISC-L Iteration 1 – Interlinking Guide

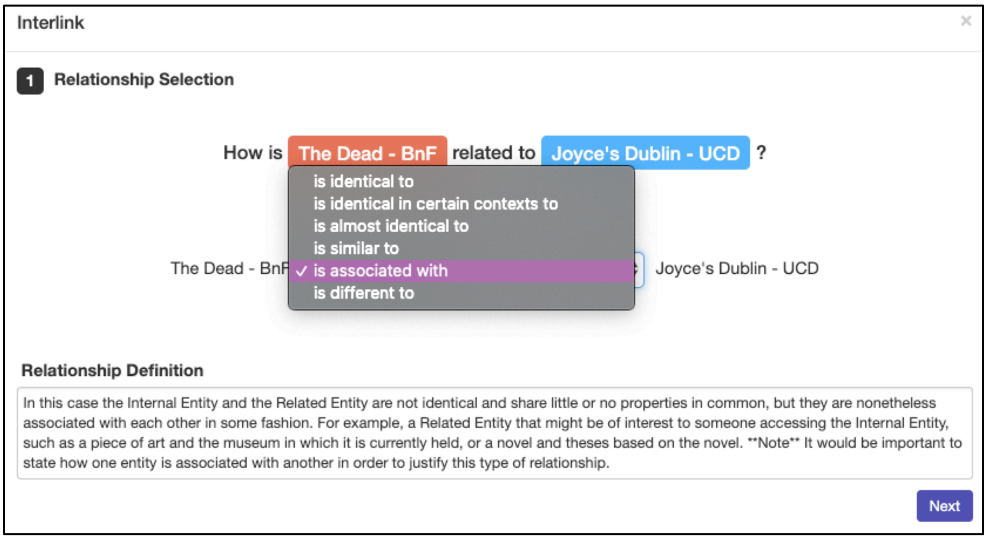


Figure 34: NAISC-L Iteration 2 – Interlinking Guide – Select Relationship

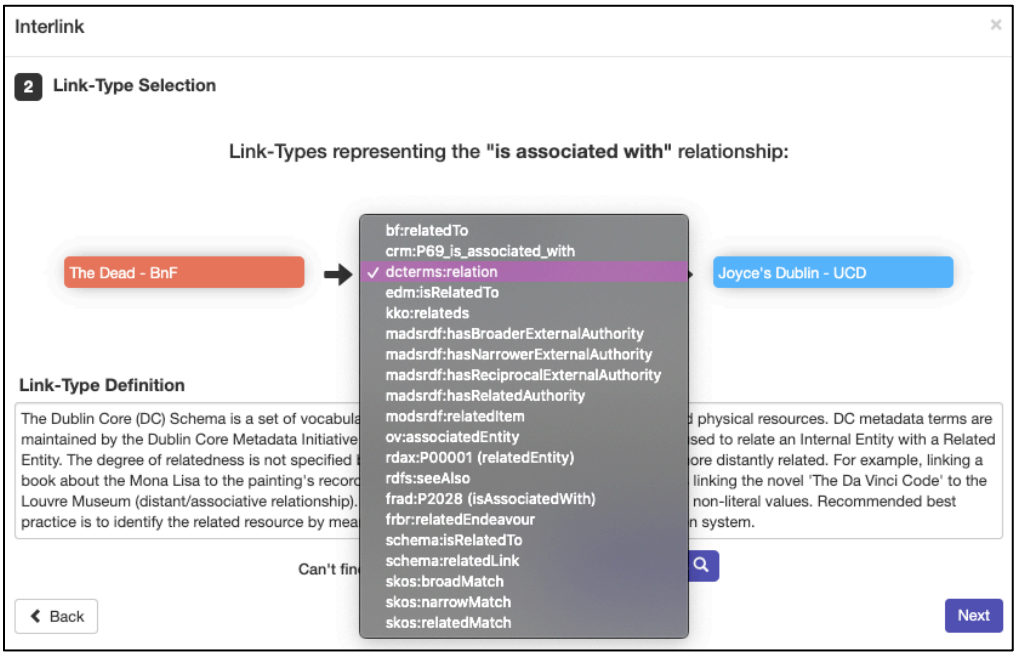


Figure 35: NAISC-L Iteration 2 – Interlinking Guide - Link-Type Selection

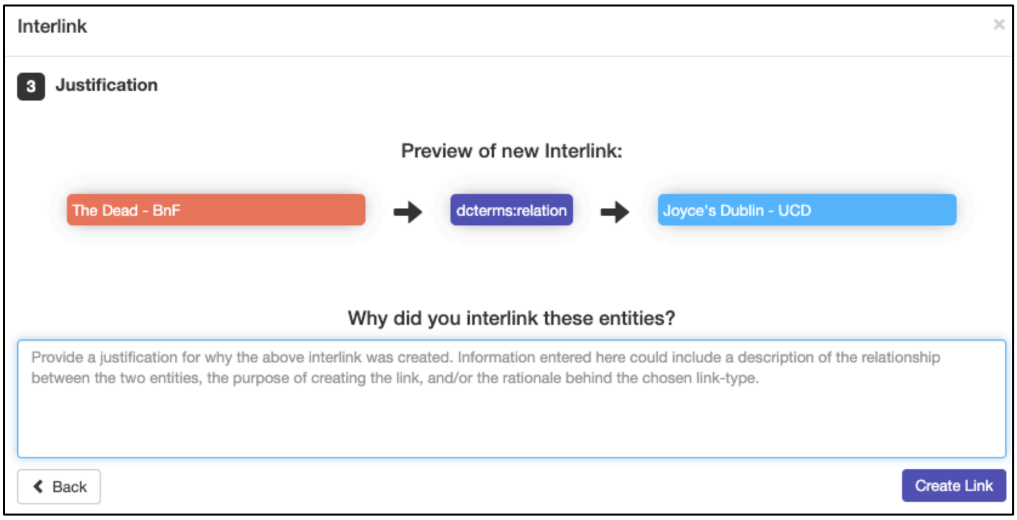


Figure 36: NAISC-L Iteration 2 – Interlinking Guide - Justification

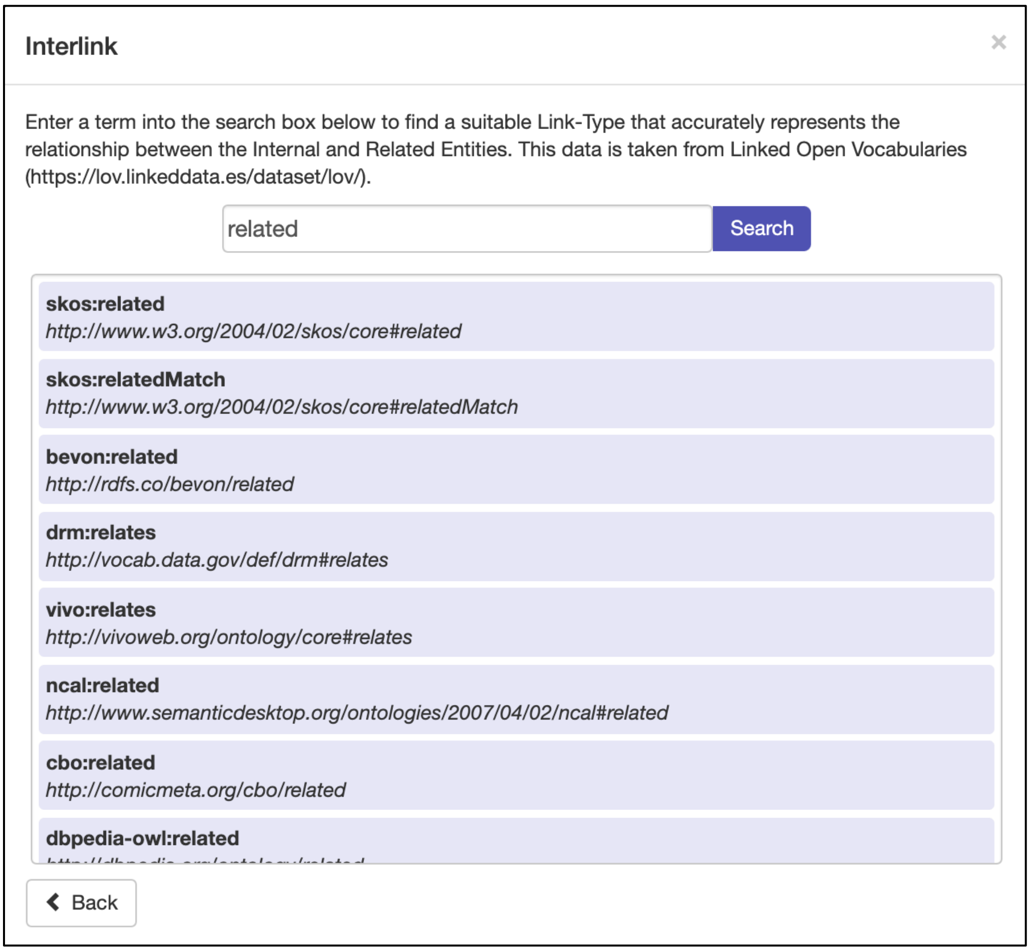


Figure 37: NAISC-L Iteration 2 – LOV Link-Type Search

In both iterations, once a user completes an interlinking session, they can view the interlink and provenance RDF graphs and visualisations. As mentioned, the visualisations are generated using GoJS software. The interlink visualisations were similar across both iterations except that NAISC-L Iteration 2 used colour coding in order aid user comprehension of the graph – Internal Entities were represented in orange and Related Entities in blue – see Figure 38. This colour coding was used throughout the interlinking process and was added as a result of user feedback.

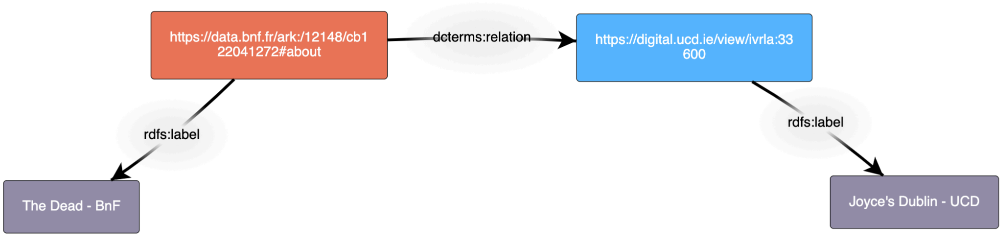


Figure 38: NAISC-L Iteration 2 – Interlink Graph Visualisation

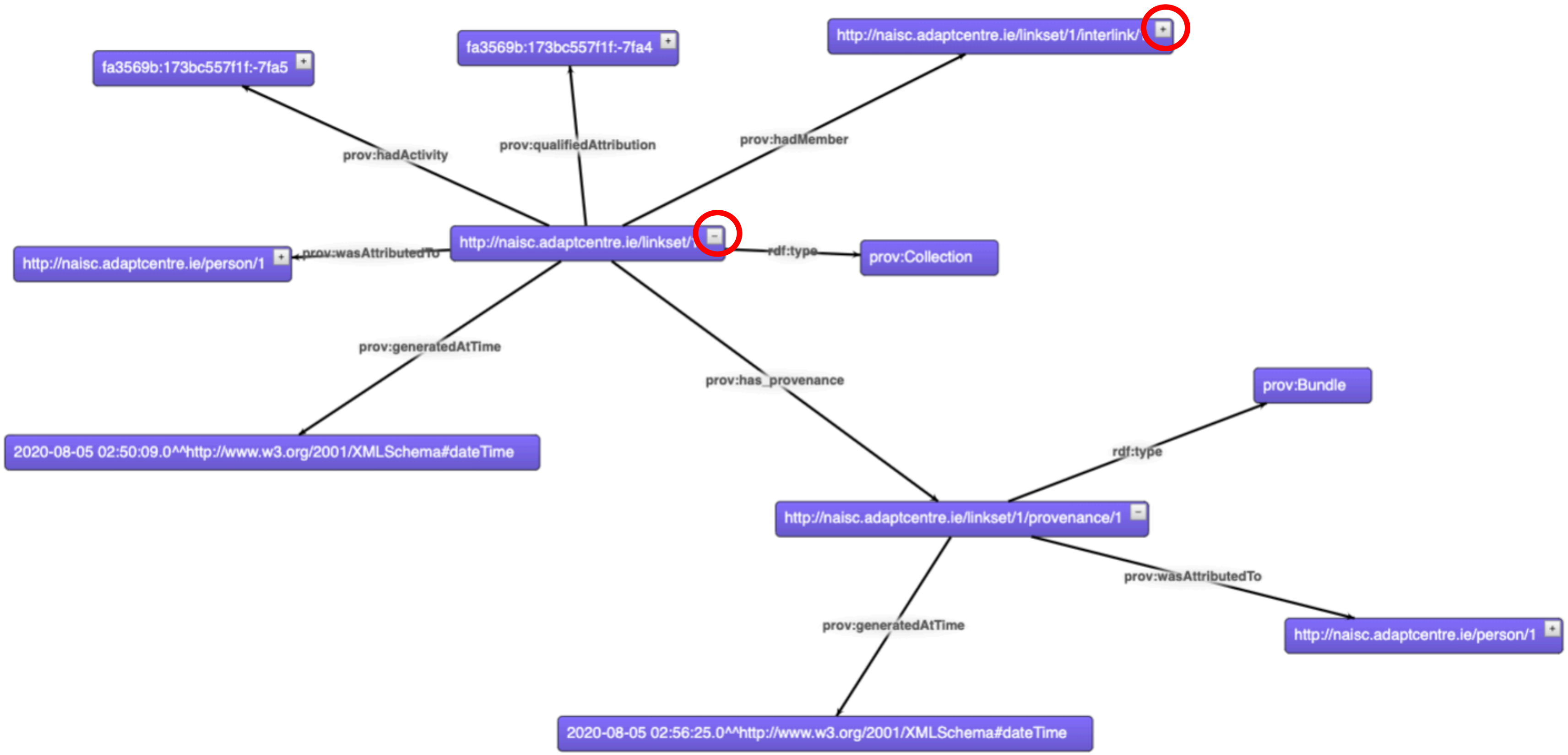


Figure 39: NAISC-L Iteration 2 – Provenance Graph Visualisation

With regards to the provenance visualisation, as the graphs are quite large, users are able to expand and retract the nodes in order to view the data at different levels of detail and granularity – see Figure 39 (sample expansion points circled in red). As well as making the graph easier to navigate, this was also one of the provenance requirements distilled from the LD Requirements Questionnaire.

In addition to the visualisations, users can also view and download the RDF graph in TriG, NQuads or TriX format. Since NAISC-L's knowledge organisation relies on named graphs, the aforementioned RDF serialization formats are more appropriate than TURTLE, N-Triples and RDF/XML. Sample Interlink, Relationship and Provenance RDF graphs (in TriG) can be seen in Figures 40, 41 and 42.

```
<http://naisc.adaptcentre.ie/linkset/13> {
    <https://data.bnf.fr/ark:/12148/cb11909158m#about>
            owl:sameAs  <http://viaf.org/viaf/44300643/> .

    <https://data.bnf.fr/ark:/12148/cb122041272#about>
            dcterms:relation  <https://digital.ucd.ie/view/ivrla:33600> .
}
```

Figure 40: Interlink RDF Graph

```
<http://naisc.adaptcentre.ie/linkset/13/relationshipGraph> {
    <http://naisc.adaptcentre.ie/linkset/13>
            prov:has_provenance  <http://naisc.adaptcentre.ie/linkset/13/provenance/22> , <http://naisc.adaptcentre.ie/linkset/13/provenance/21> .
}
```

Figure 41: Relationship RDF Graph

```
<http://naisc.adaptcentre.ie/linkset/13/provenance/22> {
    _:b0    a              prov:Attribution ;
            prov:agent    <http://naisc.adaptcentre.ie/person/13> ;
            prov:hadRole  "Created the linkset" .

    <http://naisc.adaptcentre.ie/linkset/13/interlink/20>
            a                       naiscProv:Interlink ;
            prov:hadRevision        <http://naisc.adaptcentre.ie/linkset/13/interlink/22> ;
            prov:has_provenance     <http://naisc.adaptcentre.ie/linkset/13/provenance/21> ;
            prov:invalidatedAtTime  "2020-08-19 03:28:02.0"^^xsd:dateTime ;
            prov:wasInvalidatedBy   [ a                     naiscProv:InterlinkDeletionActivity ;
                                      prov:invalidated       <http://naisc.adaptcentre.ie/linkset/13/interlink/20> ;
                                      prov:wasAssociatedWith  _:b1 , <http://naisc.adaptcentre.ie/person/13>
                                    ] .

    <http://naisc.adaptcentre.ie/linkset/13>
            a                          prov:Collection ;
            prov:generatedAtTime       "2020-08-19 03:21:27.0"^^xsd:dateTime ;
            prov:hadMember             <http://naisc.adaptcentre.ie/linkset/13/interlink/22> , <http://naisc.adaptcentre.ie/linkset/13/interlink/20> ;
            prov:qualifiedAttribution  _:b0 ;
            prov:wasAttributedTo       <http://naisc.adaptcentre.ie/person/13> .

    _:b2    a          prov:Organization ;
            foaf:name  "Bibliothèque nationale de France" .

    <http://naisc.adaptcentre.ie/person/13>
            a                    prov:Person , foaf:Person , schema:Person ;
            prov:actedOnBehalfOf  _:b2 ;
            foaf:familyName       "19627K" ;
            foaf:givenName        "19627K" ;
            schema:hasOccupation  "Metadata Cataloguer" .

    <http://naisc.adaptcentre.ie/linkset/13/interlink/22>
            a                           naiscProv:Interlink , rdf:Statement ;
            rdf:object                  <http://viaf.org/viaf/44300643/> ;
            rdf:predicate               owl:sameAs ;
            rdf:subject                 <https://data.bnf.fr/ark:/12148/cb11909158m#about> ;
            naiscProv:hasJustification  "Matching identical James Joyce entities for authority control" ;
            prov:generatedAtTime        "2020-08-19 03:25:14.0"^^xsd:dateTime ;
            prov:wasAttributedTo        <http://naisc.adaptcentre.ie/person/13> ;
            prov:wasGeneratedBy         [ a                      naiscProv:InterlinkCreationActivity ;
                                          prov:generated          <http://naisc.adaptcentre.ie/linkset/13/interlink/22> ;
                                          prov:wasAssociatedWith  _:b1 , <http://naisc.adaptcentre.ie/person/13>
                                        ] ;
            prov:wasRevisionOf          <http://naisc.adaptcentre.ie/linkset/13/interlink/20> .

    _:b1    a          prov:SoftwareAgent , foaf:Agent ;
            foaf:name  "NAISC" .

    <http://naisc.adaptcentre.ie/linkset/13/provenance/22>
            a                     prov:Bundle ;
            prov:generatedAtTime  "2020-08-19 03:28:02.0"^^xsd:dateTime ;
            prov:wasAttributedTo  <http://naisc.adaptcentre.ie/person/13> .
}

<http://naisc.adaptcentre.ie/linkset/13/provenance/21> {
    <http://naisc.adaptcentre.ie/linkset/13/interlink/20>
            a                           naiscProv:Interlink , rdf:Statement ;
            rdf:object                  <http://viaf.org/viaf/44300643/> ;
            rdf:predicate               owl:sameAs ;
            rdf:subject                 <https://data.bnf.fr/ark:/12148/cb11909158m#about> ;
            naiscProv:hasJustification  "Matching identical entities for authority control" ;
            prov:generatedAtTime        "2020-08-19 03:22:36.0"^^xsd:dateTime ;
            prov:wasAttributedTo        <http://naisc.adaptcentre.ie/person/13> ;
            prov:wasGeneratedBy         [ a                      naiscProv:InterlinkCreationActivity ;
                                          prov:generated          <http://naisc.adaptcentre.ie/linkset/13/interlink/20> ;
                                          prov:wasAssociatedWith  _:b1 , <http://naisc.adaptcentre.ie/person/13>
                                        ] .

    <http://naisc.adaptcentre.ie/linkset/13>
            a                          prov:Collection ;
            prov:generatedAtTime       "2020-08-19 03:21:27.0"^^xsd:dateTime ;
            prov:hadMember             <http://naisc.adaptcentre.ie/linkset/13/interlink/20> , <http://naisc.adaptcentre.ie/linkset/13/interlink/21> ;
            prov:qualifiedAttribution  _:b0 ;
            prov:wasAttributedTo       <http://naisc.adaptcentre.ie/person/13> .

    _:b0    a              prov:Attribution ;
            prov:agent    <http://naisc.adaptcentre.ie/person/13> ;
            prov:hadRole  "Created the linkset" .

    <http://naisc.adaptcentre.ie/linkset/13/interlink/21>
            a                           naiscProv:Interlink , rdf:Statement ;
            rdf:object                  <https://digital.ucd.ie/view/ivrla:33600> ;
            rdf:predicate               dcterms:relation ;
            rdf:subject                 <https://data.bnf.fr/ark:/12148/cb122041272#about> ;
            naiscProv:hasJustification  "Joyce's Dublin is a collection of materials related to Joyce's short story 'The Dead'" ;
            prov:generatedAtTime        "2020-08-19 03:23:16.0"^^xsd:dateTime ;
            prov:wasAttributedTo        <http://naisc.adaptcentre.ie/person/13> ;
            prov:wasGeneratedBy         [ a                      naiscProv:InterlinkCreationActivity ;
                                          prov:generated          <http://naisc.adaptcentre.ie/linkset/13/interlink/21> ;
                                          prov:wasAssociatedWith  _:b1 , <http://naisc.adaptcentre.ie/person/13>
                                        ] .

    _:b2    a          prov:Organization ;
            foaf:name  "Bibliothèque nationale de France" .

    <http://naisc.adaptcentre.ie/linkset/13/provenance/21>
            a                     prov:Bundle ;
            prov:generatedAtTime  "2020-08-19 03:24:27.0"^^xsd:dateTime ;
            prov:wasAttributedTo  <http://naisc.adaptcentre.ie/person/13> .

    <http://naisc.adaptcentre.ie/person/13>
            a                    prov:Person , foaf:Person , schema:Person ;
            prov:actedOnBehalfOf  _:b2 ;
            foaf:familyName       "19627K" ;
            foaf:givenName        "19627K" ;
            schema:hasOccupation  "Metadata Cataloguer" .

    _:b1    a          prov:SoftwareAgent , foaf:Agent ;
            foaf:name  "NAISC" .
}
```

Figure 42: Provenance RDF Graphs

Figure 43 demonstrates two SPARQL queries used to explore the provenance of the interlinks generated in the Interlink Graph above. The first query runs over the Relationship Graph in order to retrieve the Provenance Graphs associated with the Interlink Graph. The second query runs over the Provenance Graph in order to explore the provenance of a reified interlink statement.

```
SELECT ?provenanceGraph

FROM NAMED <http://naisc.adaptcentre.ie/linkset/13/relationshipGraph>

WHERE
(
   <http://naisc.adaptcentre.ie/linkset/13>
   naiscProv:hasProvenance ?provenanceGraph
)
```

```
SELECT ?justification
FROM NAMED <http://naisc.adaptcentre.ie/linkset/13/provenance/21>
FROM NAMED <http://naisc.adaptcentre.ie/linkset/13/provenance/22>

WHERE
(
  ?i rdf:type rdf:Statement ;
     rdf:subject <http://data.bnf.fr/ark:/12148/cb1222041272#about> ;
     rdf:predicate dcterms:relation ;
     rdf:object <https://digital.ucd.ie/view/ivrla:33600> ;
     naiscProv:hasJustification ?justification .
)
```

Figure 43: SPARQL Query

Overall, it can be seen that the NAISC-L GUI evolved significantly from one iteration to the next as a result of having IPs evaluate NAISC-L and then incorporating their feedback into the design process.

4.3 Chapter Summary

A set of user requirements for the creation of a LD interlinking tool, derived from the State-of-the-Art Review and the LD Requirements Questionnaire, were outlined at the beginning of this chapter. How these three requirements were met is discussed below:

R1. **The interlinking framework should allow for the creation of interlinks that express both identity links and relationship links.** This requirement was met through the design and use of the Interlinking Guide which assists users in creating a interlinks that express relationships beyond identity links. One limitation of NAISC-L, however, is that automated identity linking was not incorporated into the framework, thus all aspects of the interlinking process are manual.

R2. **The interlinking framework should be designed with the needs and expertise of IPs in mind.** This requirement was achieved by incorporating commonly used LAM data sources and vocabularies into NAISC-L. Additionally, the framework was evaluated by IPs (see Chapter 5) and their feedback was incorporated into every aspect of NAISC-L design.

R3. **The interlinking framework should publish provenance data for the interlinks generated.** NAISC-L generates rich provenance data for the interlinks and linksets that it generates. Additionally, the provenance data model was developed with the requirements of LAMs in mind.

Overall, NAISC-L has been designed to meet the requirements distilled as being important from the State-of-the-Art review and from the LD questionnaire. The following chapter presents the usability tests that were performed to evaluate the framework.

5 Evaluation

This chapter describes and presents the findings of three experiments undertaken to evaluate NAISC-L. A brief summary of each experiment is given in Section 5.1. This is followed by a description of the user evaluation instruments used in the experiments in Section 5.2. Sections 5.3 presents Usability Test 1 which was used to evaluate the first iteration of NAISC-L. This is followed by Section 5.4 which describes Usability Test 2 which was used to evaluate the second iteration of NAISC-L. Finally, Section 5.5 discusses the Field Test experiment. Overall conclusions arising from the three experiments are provided in Section 5.6 This is followed by a summary of the Evaluation chapter in Section 5.7.

5.1 Experiment Summaries

Usability Test 1

Usability Test 1 was used to evaluate the first iteration of NAISC-L. The experiment consisted of a Think-Aloud Test (van den Haak, De Jong & Schellens, 2003) in which participants, 15 IPs, were tasked with the creation of a set of interlinks using the NAISC-L framework. A post-test interview and a usability questionnaire were used to gather further data on the utility and usability of NAISC-L.

Usability Test 2

Usability Test 2 was used to evaluate the second iteration of NAISC-L. The experiment consisted of 96 IPs, from a variety of LAM backgrounds and with some prior knowledge of LD, remotely completing three interlinking tasks using NAISC-L. Participants then completed the PSSUQ and a data quality questionnaire in order to provide feedback on their user experience.

Usability Test 3

A field test was conducted in order to evaluate the second iteration of NAISC-L in a real information environment. The field test involved three IPs, working in a music archive, who were asked to use NAISC-L over a period of one week and to maintain a diary of their user-experience. This was then followed by a post-test interview and the CSUQ.

5.2 Evaluation Instruments

5.2.1 Pre-Test Questionnaire

A pre-test questionnaire was developed in order to ascertain how participants rated their knowledge of the SW, LD, RDF, URIs and ontologies prior to partaking in an experiment. Participants were asked to rate their knowledge of these topics on a five-point Likert scale ranging from 'Not at all Knowledgeable' to 'Extremely Knowledgeable'. The questionnaire also inquired whether participants had ever been directly involved in the implementation of a LD project or service, and if so, the kinds of LD activities that they gained experience in. The pre-test questionnaire was used as part of all usability tests. The Pre-Test Questionnaire can be found in Appendix 3.

5.2.2 Post-Test Interview

Semi-structured interviews were conducted as part of Usability Test 1 and the Field Test in order to gain an insight into the participants' experience of using NAISC-L. During a semi-structured interview, the participant is asked several predefined questions, however, the interviewer can deviate from these in order to further probe a participant's response (Ritter & Winterbottom, 2017). This interview format was chosen as the structured portion ensures that the required information is gathered, and the unstructured portion allows for any unforeseen responses to be investigated. The questions asked as part of the interviews were:

1. What is your overall impression of the tool?
2. What worked well?
3. What challenges did you encounter?
4. Are there any functions you would like to add or remove?
5. What is your impression of the process for selecting link-types in order to link internal and external entities?
6. What is your impression of the provenance data stored for the links and interlinking session?
7. Do you think this tool could be useful for the LAM domain?

5.2.3 PSSUQ & CSSUQ

The Post-Study System Usability Questionnaire (PSSUQ) (Lewis, 1992, 2002) is used to measure system usability at the end of a scenario-based user-study. Unlike other usability questionnaires, such as the System Usability Scale (SUS) (Brooke, 1996), the PSSUQ also provides a score for system utility – an aspect deemed necessary in the evaluation of NAISC-L.

The PSSUQ consists of 19 positive items about which the user rates agreement on a seven-point Likert scale from Strongly Agree (1) to Strongly Disagree (7), allowing for more nuanced responses in contrast to those which use five-point scales such as the SUS. The scale also has a Not Applicable (N/A) option. Responses can be calculated to provide an overall usability score as well as scores for three subscales including:

1. System Usefulness – Items 1-8 (SysUse).
2. Information Quality – Items 9-15 (InfoQual).
3. Interface Quality – Items 16-18 (InterQual).
4. Overall – Items 1-19.

It is important to note that lower PSSUQ scores indicate a more positive user perception of the questionnaire items. The PSSUQ was used as part of Usability Test 1 and Usability Test 2. The PSSUQ can be found in Appendix 4.

The Computer System Usability Questionnaire (CSUQ) (Lewis, 1995) is used for measuring system usability and utility as part of a survey or during field research. The CSUQ is almost identical to the PSSUQ except for some small differences in wording. As the PSSUQ is designed to be given directly after completion of a set of scenario-based tasks, its items are worded in the past tense (e.g. ‘I felt comfortable using this system’). However, as the CSUQ does not have to be completed directly after an interaction with a system, its items are worded in the present tense (e.g. ‘I feel comfortable using this system’). Additionally, where the PSSUQ uses the phrase ‘tasks and scenarios’, (e.g. ‘I was able to efficiently complete the tasks and scenarios using this system’), the CSUQ uses the word ‘work’ (e.g. ‘I am able to efficiently complete my work using this system’). Like the PSSUQ, lower CSUQ scores indicate a more positive perception of the questionnaire items. The CSUQ was used as part of the Field Test. The CSUQ can be found in Appendix 5.

5.2.4 Thematic Analysis

"Thematic analysis is a method for identifying, analysing, and reporting patterns within data", (Braun & Clarke, 2006, p. 79). It involves the systematic break down of data derived from qualitative research into codes, or categories, and discovering common themes by analysing and combining them. It is a method often used for the analysis of user-study data (Rosala, 2019).

An inductive approach, or 'bottom-up' approach, was followed in order to identify patterns in the data. This data-driven method involved creating themes as they emerged from the transcripts, as opposed to a 'top-down' approach whereby data is fit into pre-existing coding frameworks (Braun & Clarke, 2006). Key themes were then selected based on whether they captured something significant concerning the research hypothesis of the experiment.

A six-step thematic analysis process (Nowell, Norris, White & Moules, 2017) was used in order to analyse and discover themes in the qualitative data gathered as part of Usability Test 1 and the Field Test. These steps involved:

1. Data Familiarisation – this achieved by transcribing and re-reading the data.
2. Tagging individual observations and quotations with appropriate codes – this was achieved using N-Vivo 12[127] qualitative data analysis software.
3. Identifying patterns and themes in the codes – this was achieved by grouping related and similar codes.
4. Reviewing Themes – this was achieved by reviewing the codes to ensure they supported their assigned theme.
5. Defining and naming themes – this was achieved by refining and describing each theme.
6. Reporting on themes – this was achieved through the provision of extract examples and data visualisations.

[127] https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home accessed 12th August 2020

5.2.5 Data Quality Questionnaire

Data quality (DQ) is defined as the fitness for use of data for given application or use-case, and it is often measured according to a set of dimensions such as accessibility, trustworthiness and completeness (Zaveri et al., 2012). A modified version of the AIM Quality (AIMQ) questionnaire (Lee et al., 2002), as used in order for participants to evaluate the quality of the data they created using NAISC-L during Usability Test 2.

The AIMQ questionnaire consists of 65 statements (see Appendix 6) regarding DQ about which the user rates their level of agreement on a scale of 0 (disagree) to 10 (agree). The AIMQ measures DQ according to 14 quality dimensions: Appropriate Amount, Believability, Completeness, Concise Representation, Consistent Representation, Ease of Operation, Free of Error, Interpretability, Objectivity, Relevancy, Reputation, Security, Timeliness, and Understandability. In terms of scoring, higher ratings indicate a more positive perception of the statements (note that scores for negative statements are reversed).

For the purpose of this research, a subset of 25 statements (see Appendix 7) was used to evaluate the DQ of NAISC-L output. It was decided to modify the questionnaire in order to reduce its completion time. The selected statements were spread across 13 of the 14 dimensions evaluated by the AIMQ. The Security dimension was not included in the modified questionnaire as this information was not deducible from NAISC-L's output. The statements were chosen based on their ability to evaluate DQ in terms of LAM requirements. These requirements were based on the quality criteria used by IPs when evaluating external data sources. These criteria were established as part of the LD Requirements Questionnaire – see Section 3.3.

5.2.6 Cronbach's Alpha

Cronbach's alpha is used to indicate the internal consistency, or reliability, of questionnaires made up of Likert-type scale items (Cronbach, 1951; Tavakol & Denneck, 2011). Alpha values above 0.70 are typically considered to be an acceptable standard of reliability, however, the standard of reliability should be increased depending on the importance of the decision being made based on the test (Nunnally, 1978; Cho & Kim, 2015). Although the questionnaires adopted in this thesis have been widely used, Cronbach's alpha was applied to estimate the reliability of these instruments as part of each experiment.

5.3 Usability Test 1

In line with the DS approach, Usability Test 1 was conducted in order to evaluate the first iteration of NAISC-L, as described in Section 4.2.2.3. The focus of this user experiment was to evaluate the usability and utility of NAISC-L, and to gather detailed feedback on how NAISC-L could be improved as part of the next iteration. The experiment consisted of a Think-Aloud Test, an interview and the PSSUQ. Usability Test 1 was completed by 15 IPs, from varying backgrounds, who had some knowledge of LD.

5.3.1 Think-Aloud Test

Think-Aloud Tests (TATs) are a widely used method for evaluating the usability of software, GUIs, and websites (van den Haak et al., 2003). During a TAT, participants are asked to verbalise their thoughts and actions while carrying out a number of scenario-based tasks, thus providing data on the types of difficulties they encounter and highlighting the areas of a system that require further improvement (Becker & Yannotta. 2013; van den Haak et al., 2003).

The scenario, or context, of the TAT used to evaluate NAISC-L was that of a cataloguer creating interlinks from entities, pertaining to the Irish author James Joyce, located in the French National Library (BNF) LD dataset, to related entities found in other LD datasets. The BNF was used as part of the scenario as it has a well-established LD dataset and because the participants, who were all based in Ireland, were not likely to be overly familiar with its contents. James Joyce was used as the focus of the tasks as, given that all participants were working in Ireland, they were likely to be somewhat familiar with the author and his works, thus allowing them to create meaningful interlinks despite using an unfamiliar dataset.

As TATs typically have six to eight tasks (Andrews, Burleson, Dunks, Elmore, Lambert, Oppegaard, Pohland et al., 2012), six scenario-based tasks, representative of activities that users might carry out on NAISC-L, were developed. The six tasks of the TAT were:

1. Creating a New Linkset: This involved creating a linkset in which interlinks, from the BNF dataset, could be created and stored.
2. Adding an Internal Entity: This involved searching for a specific entity, the person James Joyce, in the BNF dataset using SemFacet[128] a semantic search tool, and then adding the URI to the linkset in NAISC-L.
3. Adding a Related Entity: This involved searching the VIAF dataset for the person James Joyce and adding the URI of the entity to the linkset.
4. Selecting an Appropriate Link-Type: This task required participants to select an appropriate link-type that described the relationship between six pairs of entities in order to create six interlinks.
5. Reviewing the RDF graph and Visualisation: This task required participants to review the RDF output for the interlinks in various RDF serialisation formats (Turtle, RDF-XML and N-Triple), and via a graphical visualisation. This 'review' task involved the participant discussing the extent to which they understood the RDF output.
6. Reviewing a Sample Provenance Visualisation: This task required participants to review a sample RDF visualisation of the provenance data for one of the interlinks created. The 'review' involved the participant discussing the extent to which they understood the data presented in the graphical visualisation.

Observations made while participants completed the tasks were documented, all comments were audio-recorded, and the time taken to complete each task was noted. Activity on the GUI was also screen-recorded – these recordings were used to clarify statements made by the participant when the audio data was transcribed and analysed at a later date. Although participants could ask questions during the test, in keeping with TAT guidelines, assistance was not provided unless necessary, and any help needed to solve a task was documented.

[128] https://www.cs.ox.ac.uk/isg/tools/SemFacet/

A more detailed description of the TAT process, scenario and tasks can be found in Appendix 8.

5.3.2 Hypothesis

This hypothesis being investigated as part of this experiment is stated as follows:

Hypothesis 1.1 (H1.1): Using the NAISC-L Framework to create LD interlinks yields high task performance with sufficient usability for IPs.

Task performance above 83%, for both the number of interlinks completed and interlink semantic accuracy, was considered to be high as a score of 83% indicated that participants completed an average of 5 out of 6 interlinks accurately. Usability was measured using the PSSUQ which is described in Section 5.2.3. In this experiment, *'sufficient usability'* was considered to be scores strictly lower than a neutral score of 4 (on a scale from 1 to 7). As stated previously, lower PSSUQ values indicate a better perception of a system.

5.3.3 Methodology

In order to test **H1.1**, a user experiment, comprising of a Pre-Test Questionnaire, a TAT, a Post-Test Interview and the PSSUQ, was conducted. This experiment was carried out individually with each participant and there was no time limit. The TAT and interview data were analysed using Thematic Analysis. Usability and utility were further evaluated using the PSSUQ questionnaire.

The experiment was structured as follows:

1. **Informed Consent:** Here the experiment was explained in detail to the participant both verbally and in writing. Participants who provided written consent to complete the TAT proceeded with the experiment (see Appendix 9 for the Participant Information Sheet and Consent Form).
2. **Pre-Test Questionnaire:** Participants rated their knowledge of the SW, LD, RDF, URIs and ontologies as part of the pre-test questionnaire. The questionnaire is described in more detail in Section 5.2.1
3. **Think-Aloud Test**: Participants were given the TAT scenario and tasks as described in Section 5.3.1.

4. **Post-Test Interview**: The post-test interview consisted of seven questions, outlined in Section 5.2.2, which explored the participants' experience of using NAISC-L.
5. **Post-Test Questionnaire:** After completing the TAT, the participants were asked to fill out the PSSUQ, as described in Section 5.2.3.

5.3.4 Participants

Research suggests that usability tests of 15 participants uncover an average of 97.05% of usability problems (Faulkner, 2003) – thus 15 IPs were recruited to complete Usability Test 1. Non-probabilistic sampling methods were used, whereby LAMs were contacted directly with a request for participants. LAMs known to conduct research and those with functioning LD projects were contacted. IPs with a known interest in LD were also contacted. All LAMs contacted for the experiment were located in Dublin, Ireland.

The participants in Usability Test 1 were 15 IPs who had some prior knowledge of LD. The participants' self-perceived rating of their knowledge of the SW, LD, RDF, URIs and ontologies can be seen in Table 13 below. With the exception of Participant 15, all participants rated themselves as knowledgeable for each of the five topics, with the majority considering themselves Moderately Knowledgeable in all areas. Participant 15, however, rated themselves as Slightly Knowledgeable for LD only. Four of the participants indicated that they had previous experience implementing a LD project. Overall, it can be seen that all participants had some prior awareness and knowledge of LD, but that none rated themselves as Very or Extremely Knowledgeable in the area, suggesting that no participant considered themselves to be an expert user.

Of the 15 participants, seven worked in academic libraries across four third-level institutions, three worked in a national library, two worked in a museum, two worked in a music archive, and one worked in a government library.

Table 13: Usability Test 1 – Participant Knowledge Ratings

Participant	Institution	Topic / Rating	Not at all Knowledgeable	Slightly Knowledgeable	Moderately Knowledgeable	Very Knowledgeable	Extremely Knowledgeable	Previous work on a LD Project?
		Semantic Web		●				
		Linked Data		●				
1	Academic Library	**RDF**		●				No
		URIs		●				
		Ontologies		●				
		Semantic Web			●			
		Linked Data			●			
2	Government Library	**RDF**		●				Yes
		URIs		●				
		Ontologies		●				
		Semantic Web			●			
		Linked Data			●			
3	National Library	**RDF**			●			Yes
		URIs				●		
		Ontologies			●			
		Semantic Web			●			
		Linked Data			●			
4	National Library	**RDF**			●			No
		URIs			●			
		Ontologies			●			
		Semantic Web			●			
		Linked Data			●			
5	National Library	**RDF**			●			No
		URIs		●				
		Ontologies		●				
		Semantic Web		●				
		Linked Data			●			
6	Academic Library	**RDF**		●				No
		URIs			●			
		Ontologies		●				
		Semantic Web			●			
		Linked Data			●			
7	Academic Library	**RDF**		●				No
		URIs			●			
		Ontologies			●			
		Semantic Web			●			
		Linked Data			●			
8	Academic Library	**RDF**			●			Yes
		URIs			●			
		Ontologies				●		
		Semantic Web			●			
		Linked Data			●			
9	Academic Library	**RDF**			●			No
		URIs			●			
		Ontologies		●				
		Semantic Web			●			
		Linked Data			●			
10	Museum	**RDF**			●			No
		URIs				●		
		Ontologies			●			
		Semantic Web			●			
		Linked Data			●			
11	Academic Library	**RDF**			●			No
		URIs			●			
		Ontologies				●		
		Semantic Web			●			
		Linked Data			●			
12	Academic Library	**RDF**			●			No
		URIs			●			
		Ontologies			●			
		Semantic Web			●			
		Linked Data			●			
13	Archive	**RDF**			●			Yes
		URIs			●			
		Ontologies			●			
		Semantic Web			●			
		Linked Data			●			
14	Archive	**RDF**		●				No
		URIs		●				
		Ontologies			●			
		Semantic Web	●					
		Linked Data		●				
15	Museum	**RDF**	●					No
		URIs	●					
		Ontologies	●					

5.3.5 Results

5.3.5.1 Think-Aloud Test

The results of the TAT tasks have been summarised in Table 14 below. Here the degree to which the participant was able to complete the task is documented, as well as the time (in minutes) that it took participants to complete each task, and whether they required assistance.

Tasks 1, 2, 3, 5 and 6 had a mean completeness score of 100%, and Task 4 (interlinking activities) had a mean completeness score of 95.55%. As mentioned previously, for the purpose of this experiment, task performance above 83% was considered to be high. Given that the mean task performance across all six tasks was 99%, high task performance was achieved overall.

At an individual level, task performance of 100% was achieved by all participants across all tasks, except for Task 4 where Participant 13 and 15 achieved 66.66% as they completed only four of the six interlinking activities. It is worth noting that this was due to time constraints rather than an inability to complete the activities.

The average time it took participants to complete the TAT was 29.87 minutes. The task which took the longest time for each participant to complete was Task 4 with an average of 19.6 minutes. However, this was expected given that the task required the participant to complete six separate interlinking activities, resulting in an average time of 3.27 minutes per activity. A box plot of the time taken for each task can be seen in Figure 44. Here, outliers can be noted for Task 1 and Task 2.

Assistance was required by some participants in Task 2 (40%), Task 3 (26.67%) and Task 4 (46.67%). The assistance required for Tasks 2 and Task 3 was primarily navigational i.e. the participant was unsure which button to click in order to proceed with the task. Assistance given in Task 4 was primarily to point out which of the two entities displayed on the GUI needed to be interlinked, and to highlight that clicking on a Relationship Term or Link-type would provide its definition

Table 14: Think-Aloud Test – Task Results

Participant	Task 1			Task 2			Task 3		
	Complete (%)	Time (mins)	Assistance Required	Complete (%)	Time (mins)	Assistance Required	Complete (%)	Time (mins)	Assistance Required
1	100%	1.77	No	100%	3.63	Yes	100%	2.43	No
2	100%	1.38	No	100%	3.75	No	100%	3.92	No
3	100%	0.87	No	100%	1.83	No	100%	1.70	No
4	100%	3.77	No	100%	4.00	No	100%	5.57	Yes
5	100%	1.50	No	100%	2.83	No	100%	3.63	N
6	100%	1.07	No	100%	2.48	No	100%	4.13	Yes
7	100%	0.75	No	100%	4.10	Yes	100%	3.82	No
8	100%	1.63	No	100%	4.00	No	100%	3.23	No
9	100%	0.87	No	100%	4.10	No	100%	4.55	No
10	100%	1.15	No	100%	3.30	Yes	100%	2.93	Yes
11	100%	3.90	No	100%	4.82	No	100%	4.13	No
12	100%	2.28	No	100%	3.97	No	100%	3.12	No
13	100%	1.60	No	100%	6.40	Yes	100%	4.10	No
14	100%	0.57	No	100%	3.35	Yes	100%	2.57	No
15	100%	1.38	No	100%	6.90	Yes	100%	4.58	Yes
Mean	**100%**	**1.63**	**0%**	**100%**	**3.96**	**40%**	**100%**	**3.63**	**26.67%**
Standard Deviation		**0.96**			**1.27**			**0.95**	

Participant	Task 4			Task 5			Task 6			Total	
	Complete (%)	Time (mins)	Assistance Required	Complete (%)	Time (mins)	Assistance Required	Complete (%)	Time (mins)	Assistance Required	Complete (%)	Time (mins)
1	100%	13.37	Yes	100%	2.68	No	100%	2.68	No	100%	23.95
2	100%	12.13	No	100%	2.75	No	100%	2.75	No	100%	24.07
3	100%	8.28	No	100%	2.02	No	100%	2.02	No	100%	14.88
4	100%	33.57	No	100%	2.45	No	100%	2.45	No	100%	49.71
5	100%	22.25	Yes	100%	2.80	No	100%	2.80	No	100%	33.13
6	100%	14.20	No	100%	1.63	No	100%	1.63	No	100%	23.57
7	100%	18.30	No	100%	0.85	No	100%	0.85	No	100%	28.02
8	100%	28.00	No	100%	1.59	No	100%	1.59	No	100%	38.69
9	100%	33.00	Yes	100%	4.13	No	100%	4.13	No	100%	37.10
10	100%	23.12	Yes	100%	1.22	No	100%	1.22	No	100%	20.33
11	100%	11.28	Yes	100%	3.88	No	100%	3.88	No	100%	45.68
12	100%	28.75	No	100%	0.08	No	100%	0.08	No	100%	27.75
13	66.66%	18.05	Yes	100%	1.03	No	100%	1.03	No	94.44%	28.57
14	100%	15.17	No	100%	1.05	No	100%	1.05	No	100%	22.36
15	66.66%	14.60	Yes	100%	1.53	No	100%	1.53	No	94.44%	30.29
Mean	**95.55%**	**19.60**	**46.67%**	**100%**	**1.98**	**0%**	**100%**	**1.98**	**0%**	**99.25%**	**29.87**
Standard Deviation		7.82			1.09			1.09			9.22

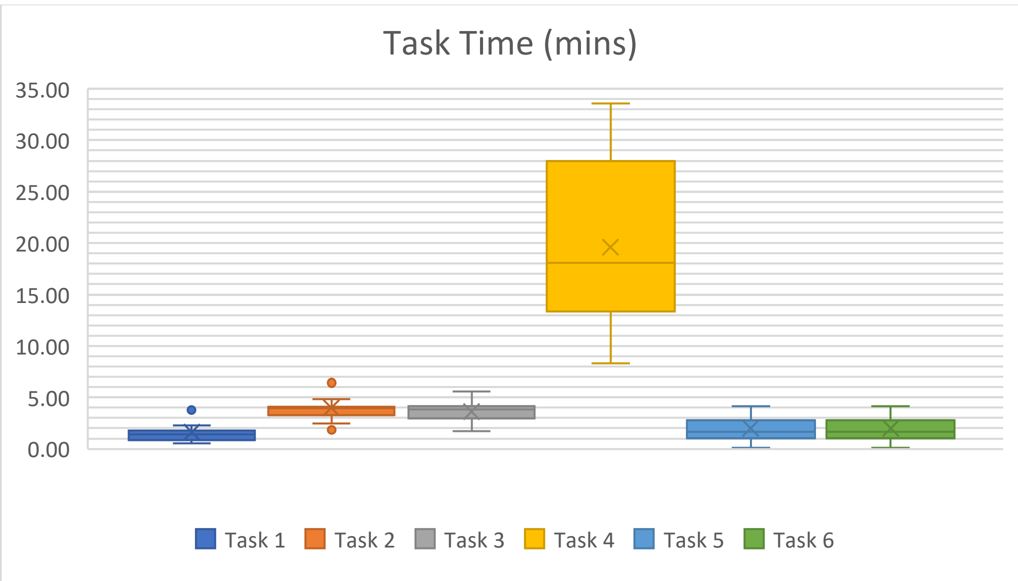


Figure 44: Think Aloud Test – Task Time Box Plot

As mentioned, Task 4 consisted of six interlinking activities which required participants to select an appropriate link-type to:

1. Link the BNF entity for James Joyce to the VIAF entity for James Joyce.
2. Link the BNF entity for the short story 'The Dead' to the entity for 'Joyce's Dublin', a collection of media related to the short story 'The Dead', located in the digital library of University College Dublin (UCD).
3. Link the BNF entity for 'Ulysses' to the entity of an audiobook of Ulysses in WorldCat.
4. Link the BNF entity for 'Ulysses' to the entity for a performance of Ulysses in the Abbey Theatre Archive[129].
5. Link the BNF entity for the film 'The Dead' to the entity for the film 'The Dead' in the Irish Film and TV Research Online archive[130].
6. Link the BNF entity for the film 'The Dead' to the entity of a newspaper article about the film in The New York Times archive[131].

The goal of Task 4 was to evaluate the usability and utility of NAISC-L's link-type selection process. This involved evaluating how successful participants were in choosing a reasonable and semantically accurate link-type to represent the relationship between each pair of entities. For the purpose of this research, a reasonable Link-type was considered to be a predicate that, according to its ontological definition, could be used to meaningfully link the given entities.

Table 15 below outlines the link-types selected by each participant when completing Task 4 of the TAT. Link-types marked in green are considered reasonable or accurate, and link-types in red are considered inaccurate. The overall accuracy for each activity is given in the final row of the table. Task 4 had an average accuracy score of 91.12% across all interlinking activities indicating that high task performance (over 83%) was achieved overall. Looking at the accuracy of the individual activities, all but Interlinking Activity 3 had an average accuracy score above 83%.

[129] https://www.abbeytheatre.ie/about/archive/ accessed August 12th 2020
[130] https://www.tcd.ie/irishfilm/ accessed August 12th 2020
[131] https://www.nytimes.com accessed August 12th 2020

Although the relationship between each pair of entities could be meaningfully defined in multiples ways, the level of agreement between participants for link-type selection was noted to be as follows:

- Activity 1 – 66.67% of participants selected `owl:sameAs`.
- Activity 2 – 53.34% of participants selected `dcterms:relation`[132].
- Activity 3 – 40% of participants selected `dcterms:relation` and 33.34% selected `ov:similarTo`[133].
- Activity 4 – 40% of participants selected `dcterms:relation` and 40% selected `ov:similarTo`.
- Activity 5 – 76.93% of participants who completed the activity selected `owl:sameAs`.
- Activity 6 – 46.15% of participants who completed the activity selected `ov:similarTo` and 38.46% selected `dcterms:relation`.

[132] Dublin Core Metadata Terms – https://www.dublincore.org/specifications/dublin-core/dcmi-terms/ accessed 12th August 2020
[133] Open Vocab – https://vocab.org/open/ accessed 12th August 2020

Table 15: Think-Aloud Test – Interlink Accuracy

Participant	Activity 1	Activity 2	Activity 3	Activity 4	Activity 5	Activity 6	Semantic Accuracy
1	owl:sameAs	rdfs:seeAlso	rdfs:seeAlso	rdfs:seeAlso	owl:sameAs	rdfs:seeAlso	100%
2	owl:sameAs	dcterms:relation	dcterms:relation	dcterms:relation	owl:sameAs	dcterms:relation	100%
3	owl:sameAs	rdfs:seeAlso	ov:similarTo	sio:isRepresentedBy	owl:sameAs	rdfs:seeAlso	83.33%
4	ov:similarTo	ov:similarTo	ov:similarTo	ov:similarTo	owl:sameAs	dcterms:relation	100%
5	rdfs:seeAlso	rdfs:seeAlso	ov:similarTo	ov:similarTo	owl:sameAs	dcterms:relation	100%
6	owl:sameAs	dcterms:relation	sio:represents	ov:similarTo	rdfs:seeAlso	rdfs:seeAlso	83.33%
7	owl:sameAs	dcterms:relation	dcterms:relation	dcterms:relation	dcterms:relation	rdfs:seeAlso	100%
8	owl:sameAs	dcterms:relation	dcterms:relation	dcterms:relation	owl:sameAs	dcterms:relation	100%
9	ov:similarTo	dcterms:relation	dcterms:relation	dcterms:relation	dcterms:relation	rdfs:seeAlso	100%
10	owl:sameAs	owl:sameAs	owl:sameAs	owl:sameAs	owl:sameAs	dcterms:relation	50%
11	ov:similarTo	dcterms:relation	ov:similarTo	dcterms:relation	owl:sameAs	dcterms:relation	100%
12	owl:sameAs	dcterms:relation	ov:similarTo	umbel:isLike	owl:sameAs	owl:sameAs	83.33%
13	dcterms:relation	rdfs:seeAlso	dcterms:relation	rdfs:seeAlso			100%**
14	owl:sameAs	rdfs:seeAlso	sio:represents	ov:similarTo	owl:sameAs	owl:sameAs	66.67%
15	owl:sameAs	dcterms:relation	dcterms:relation	ov:similarTo			100%**
Accuracy	100%	93.33%	80%	86.67%	100%*	86.67%*	91.12%

* Note that for Activity 5 and Activity 6 accuracy is based on 13, rather than 15, link-types.

** Note that for Participant 13 and 14, accuracy is based on 4, rather than 6, activities.

5.3.5.2 Thematic Analysis: Think Aloud Test & Interview

Three rounds of thematic analysis, as visualised in Figure 45, were conducted on the TAT and post-test interview data (see Appendix 10 for a sample transcript). The initial round of coding identified 125 separate tags – see Figure 46 for a coding snippet. A second round of coding was then conducted in order to categorise and merge similar tags resulting in the generation of 42 codes. Of these 42 codes, 33 were grouped into five themes. These 33 codes were selected based on the total number of references to the code in the data, the number of participants associated with the code, and the relevance of the code in relation to the experiment hypothesis (**H1.1**) and the research question of this thesis. Nine codes were excluded from the themes based on having less than three references in the transcripts and/or being unrelated to the experiment hypothesis. These unused codes can be found in Appendix 11.

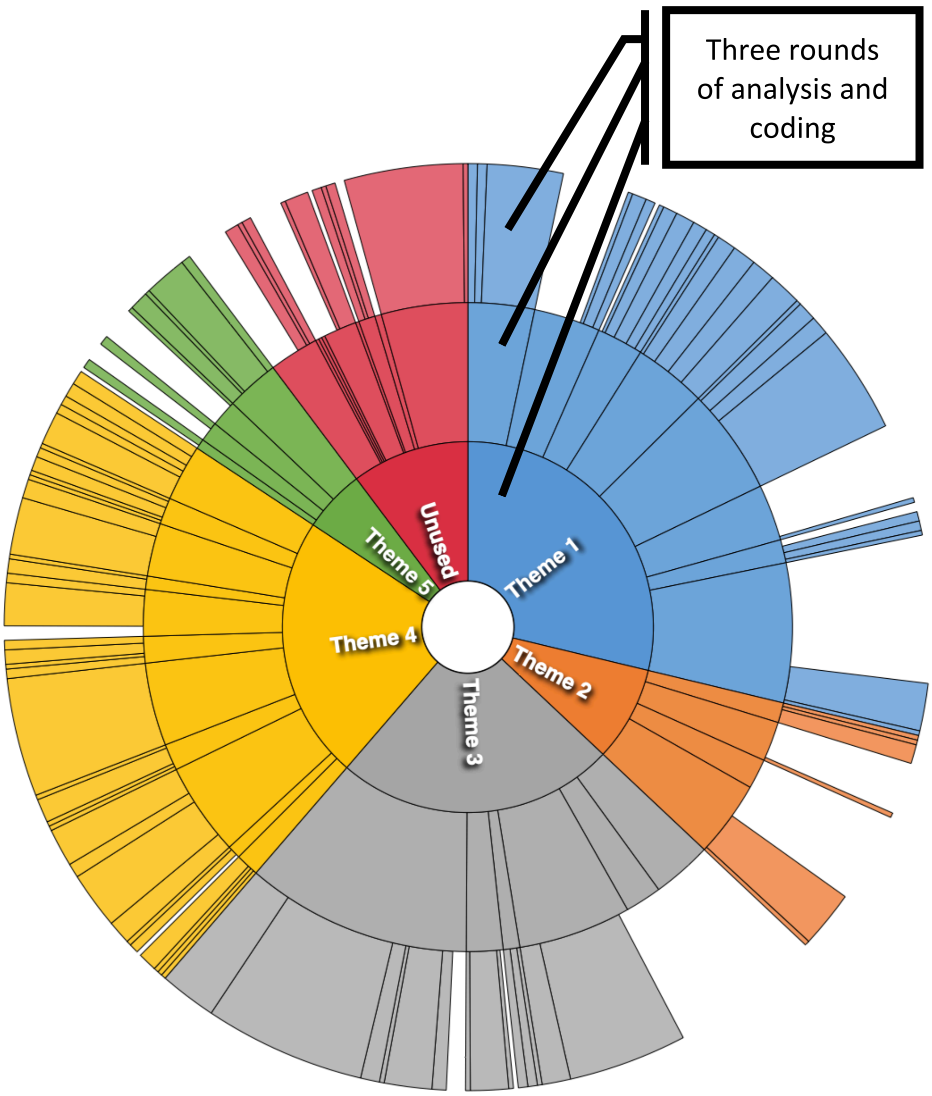
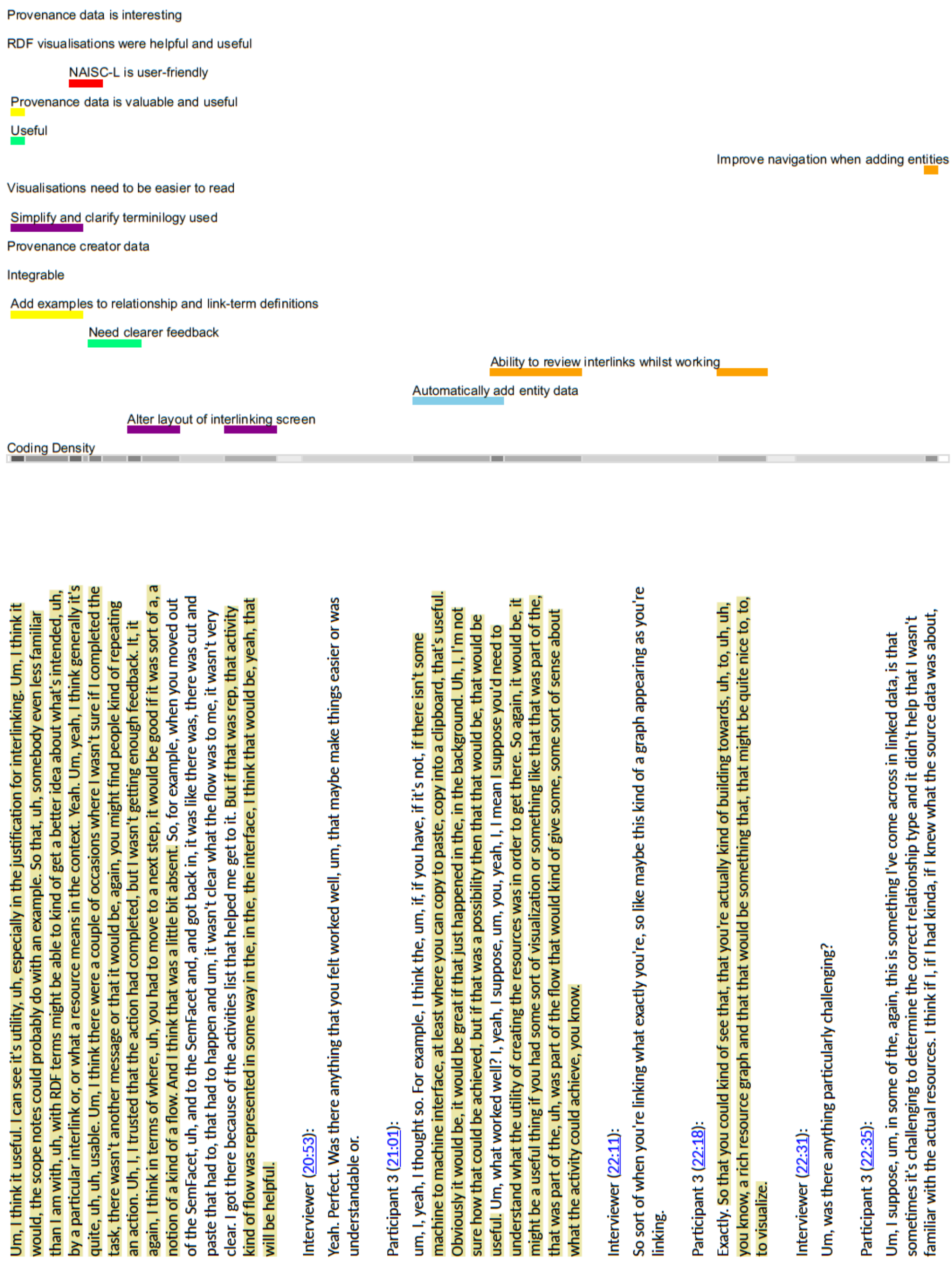


Figure 45: Usability Test 1 – Transcript Coding Hierarchy



Um, I think it useful. I can see it's utility, uh, especially in the justification for interlinking. Um, I think it would, the scope notes could probably do with an example. So that, uh, somebody even less familiar than I am with, uh, with RDF terms might be able to kind of get a better idea about what's intended, uh, by a particular interlink or, or what a resource means in the context. Yeah. Um, yeah, I think generally it's quite, uh, uh, usable. Um, I think there were a couple of occasions where I wasn't sure if I completed the task, there wasn't another message or that it would be, again, you might find people kind of repeating an action. Uh, I, I trusted that the action had completed, but I wasn't getting enough feedback. It, it again, I think in terms of where, uh, you had to move to a next step, it would be good if it was sort of a, a notion of a kind of a flow. And I think that was a little bit absent. So, for example, when you moved out of the SemFacet, uh, and to the SemFacet and, and got back in, it was like there was, there was cut and paste that had to, that had to happen and um, it wasn't clear what the flow was to me, it wasn't very clear. I got there because of the activities list that helped me get to it. But if that was rep, that activity kind of flow was represented in some way in the, in the, the interface, I think that would be, yeah, that will be helpful.

Interviewer (20:53):

Yeah. Perfect. Was there anything that you felt worked well, um, that maybe make things easier or was understandable or.

Participant 3 (21:01):

um, I, yeah, I thought so. For example, I think the, um, if, if you have, if it's not, if there isn't some machine to machine interface, at least where you can copy to paste, copy into a clipboard, that's useful. Obviously it would be, it would be great if that just happened in the, in the background. Uh, I, I'm not sure how that could be achieved, but if that was a possibility then that that would be, that would be useful. Um, what worked well? I, yeah, I suppose, um, you, yeah, I, I mean I suppose you'd need to understand what the utility of creating the resources was in order to get there. So again, it would be, it might be a useful thing if you had some sort of visualization or something like that that was part of the, that was part of the, uh, was part of the flow that would kind of give some, some sort of sense about what the activity could achieve, you know.

Interviewer (22:11):

So sort of when you're linking what exactly you're, so like maybe this kind of a graph appearing as you're linking.

Participant 3 (22:18):

Exactly. So that you could kind of see that, that you're actually kind of building towards, uh, to, uh, uh, you know, a rich resource graph and that that would be something that, that might be quite nice to, to, to visualize.

Interviewer (22:31):

Um, was there anything particularly challenging?

Participant 3 (22:35):

Um, I suppose, um, in some of the, again, this is something I've come across in linked data, is that sometimes it's challenging to determine the correct relationship type and it didn't help that I wasn't familiar with the actual resources. I think if I, if I had kinda, if I knew what the source data was about,

Figure 46: Usability Test 1 – Coding Snippet

The five themes distilled from the data include:

1. NAISC-L Framework Usability and Utility – 8 codes
2. Provenance Data Usability and Utility – 4 codes
3. Enrich Descriptions and Definitions – 6 codes
4. GUI Requirements – 11 codes
5. Automation – 4 codes

Theme 1 and Theme 2 relate to the usability and utility of the interlinking process and the provenance data. Codes for these themes indicate that participants had a positive experience using NAISC-L, that they found it intuitive, useful and user-friendly. Participants also found the provenance data to be useful and thorough.

Several new requirements for NAISC-L were identified throughout the experiment. Codes from Theme 3, Theme 4 and Theme 5 of the thematic analysis include a number of modifications and additions to NAISC-L which were suggested by participants. The codes in Theme 3 primarily suggest clarifying and simplifying the terminology used in NAISC-L and enriching descriptive text with examples. Codes in Theme 4 relate to new requirements identified for the GUI such as improving navigation, making visualisations easier to read, and adding a function that would review interlinks visually while they are being created. Theme 5 focuses on participants' feedback regarding adding automated processes to NAISC-L as a way of saving time. Codes in this theme indicate that participants would prefer that certain functions in NAISC-L were automated such as auto-populating entity fields and justification fields with data, as well as providing link-type suggestions.

All themes and codes distilled from the transcript data are detailed in Table 16 below. This table also includes the code descriptions, the number of references to the code in the transcripts (NR), the number of participants who referred to the code (NP), and supporting participant (P) quotes. Three supporting quotes were chosen randomly for each code in order to demonstrate how the codes and themes were derived from the transcript data.

Table 16: Usability Test 1 – Thematic Analysis

Theme	Code	Description	NR	NP	Selection of Quotes
Theme 1: NAISC-L Framework Usability and Utility	Useful	Participants thought that NAISC-L was useful for creating interlinks, and that the interlinks themselves would be useful to the LAM domain.	42	14	"The important point is that this is very useful", P. 7 "I think there's a real opportunity for something like that in this kind of environment", P. 10 "I could actually see it working with the digital collections very well. Because that's, that's a tool that we're building ourselves. So, I think, um, I think I could definitely see it working there because it does bring, it brings, really brings added value I think to the whole process of discovery if you can make these links happen. Yeah. So, there's definite value.", P. 11
	RDF visualisations were helpful and useful	Participants found the visualisations helpful for understanding the interlinks.	32	12	"I think, I love the visualisation at the end. For me that really would help me try to figure it out", P. 4 "It's interesting to see it visualised like that", P. 5 "it's obviously uh, very clear and good to see, to review what you've done and the structure that you've created", P. 6 "I found this really useful at the end to see the output", P.9
	Quicker and easier to use over time	Participants stated that NAISC-L became easier to use over time.	20	9	"I think maybe the more I'd be, the more I'd use it, the more familiar and quickly I'd be able to go through it", P. 4 "It seemed fairly clear and nicely presented and easy to, uh, to create, especially when you've done a couple and kind of get into the swing of it", P. 6 "once I did a few then I got the hang of it", P. 11
	Clear GUI	The layout of the GUI was clear and neat	20	9	"My overall impression is that the screens are very clean. That navigating around, finding the button that you're looking for is really clear, there's no issue there", P. 7 "the interface is really good and it actually really, compares really well to the kind of metadata tools we would use here in terms of that just the, the support it gives for, for why you're doing this like, you know, so yeah", P. 9 "It's all very easy to use, I have to say, I didn't really find anything particularly unpleasant about using it. It was very clear and thorough. So, I wasn't searching around for any buttons or trying to find anything for a while. No it's, it's, it's great", P. 14

Theme 1: NAISC-L Framework Usability and Utility	Intuitive and straightforward	Participants found the flow of the interlinking process to be intuitive.	18	10	“So, I thought that was actually quite good. It's quite intuitive.”, P. 8 “I found it very user friendly. Very intuitive.”, P. 12 “That flows quite well actually”, P. 14
	Relationship and link-type definitions were useful and helpful	Participants found the Relationship and link-type definitions useful when creating interlinks.	16	8	“And also I did really like the, the embedded descriptions of what I was using, so if I chose a Dublin Core or an OWL, uh, option, what that meant because I think that's, for me as a beginner kind of in actually implementing it. It's, like I would know Dublin Core but it wouldn't know OWL. And, yeah, it's just nice to actually have it in front of you”, P. 9 “I did find the definitions helpful, you know, I don't use OWL or any of those other ontologies, so certainly for me that was a learning curve. It's very useful to have the definitions in situ”, P. 11 “I liked, I liked the inbuilt explanations, I liked the dropdown“, P. 12
	User-friendly	Participants found NAISC-L to be easy to use.	14	10	“I think um, the idea of having an interface to make it more usable is often a big barrier to it being adopted. So, I can see, I think this would be really useful one day.”, P. 9 “It seemed pretty easy to use”, P. 10 “Um, certainly just to do the few, I found it very user friendly. Very intuitive.“, P. 12
	Suitable for non-expert LD users	Participants found NAISC-L to be suitable for IPs.	7	5	“I think it's, um, I think it's pretty intuitive. You know, especially for someone like me who was very, very basic knowledge.”, P. 2 “if you are building it for a cataloguer then, you know, it's, it's, I think you're probably going in the right direction”, P. 8 “And particularly with Linked Data, I could see this being useful for students or for researchers who aren't, as I say, often it's part of their projects that they may not have time to learn XML or learn whatever it is, this would be really nice way of illustrating the use of it, even for someone like me who has really limited knowledge of it.”, P. 9

Theme 2: Provenance Data Usability and Utility	Valuable and useful	Participants indicated that the provenance would be useful when making decisions regarding data authority and trustworthiness.	23	10	"I like the fact that you've included the bit for the justification for your decision because that also then flags to other users or other cataloguers that yeah, actually she's provided the evidence there", P 1. "I would definitely be looking at that first to see if I really trusted the resource", P. 2 "Um, it's important. Um, I think if you want authoritative data you need good sources of information. So, I think it's really important that it's there.", P. 11
	Interesting	Participants found the provenance data to be interesting.	12	7	"I think it's really interesting because it's not something I've come across that often", P. 2 "I do like having provenance with an, with who created it, it has a date", P.8 "Yeah, no, I quite, quite like this one. I actually quite liked seeing this. It was nice, it was very, very, I found it very interesting to see, uh, to see the associations", P. 10
	Thorough	Participants indicated that the data captured in the provenance graph was sufficient.	9	6	"It looks comprehensive to me", P. 1 "I don't know why you'd want any more than that. I think it's really good. It's really good information to have in there", P.2 "I don't think there was an excess of information around the resources themselves and the information describing the links between them, that all, I'm happy with what's there, I can't see anything missing', P. 7
	Creator Data	Participants indicated that the data captured for the creator should not be recorded at the level of the individual but that more information regarding their department and area of expertise could be provided.	6	3	"I questioned, um, the importance of the information about the, the record creator, that coming, that coming up there, it feels maybe like that's, um, above and beyond what's required", P. 7 "Um, I think I'd, I'd like something a little bit more granular than University College Dublin just because if you have projects within new UCD working on things and they're publishing their kind of, their records", P. 8 "potentially you might use, um, uh, you might use a friend of a friend data about the actual author so that you could kind of get some provenance about the provenance, yeah. You know, sort of like what, you know, it's, so if it's coming from, um, if it's a Joyce Scholar, you know, um, yeah, yeah, I'd feel, I, I'd feel I might trust that justification in, in a more, um detailed way", P. 3

Theme 3: Enrich Descriptions and Definitions	Simplify and clarify terminology	Participants found some of the vocabulary to be too complex.	63	15	“Yeah, and you do have to be very careful about language as well because certain words that we do use as librarians that can be used in various terms. Like the word collection for example, can be used in lots of different ways”, P. 4 “I suppose the, well certainly the descriptions, um, for the relationship end of things wasn't clear. But the link term description, again as I said, was clearer. And I suppose there were at least two occasions when I kind of struggled to pick the link relationship that was appropriate”, P. 7 “I don't actually understand these terms so it's a little difficult”, P.10
	Clarify the content required in data-fields	Participants were at times unsure what data to enter into a field.	34	10	“So, what are you looking for it in this box?”, P. 1 “So, with the label, I'm not sure now to what extent I need to identify that it's VIAF here”, P. 12 “I think the fact that it's free text sometimes is, um, I feel like you could just put anything in and if you do a search on free texts, you could just get lots of gobbledygook, you know? It would be good if they, it kind of forced you to, to enter in something more specific”, P. 15
	Add FRBR[134] entity relationships	Participants indicted that FRBR entity relationships would assist in interlinking decision making.	18	6	“I, what I'm thinking of is work, expression, manifestation”, P. 5 “This is the kind of semantics that we would rack our heads over regularly is, what is our record actually about? The thing itself or the concept of the thing or a representation of it”, P. 8 “So, here is again another work and here is, okay, an expression of that work in FRBR terms”, P. 12
	Add examples to relationship and link-type descriptions	Participants stated that interlink examples would help them to select an appropriate link-type.	11	4	“Um, I think it would, the scope notes could probably do with an example. So that, uh, somebody even less familiar than I am with, uh, with RDF terms might be able to kind of get a better idea about what's intended, uh, by a particular interlink or, or what a resource means in the context”, P. 3 “So, I think a little more descriptive, maybe some examples. If you could add a couple of examples to the descriptions, that would be quite helpful I think”, P. 10 “I think in order to, unless you're dealing with this on an, on a kind of a daily basis, unless you're kind of living and breathing this, it's, it is, it can be hard to, um, get, to get that, you know, without sort of concrete yeah, concrete examples”, P. 13

[134] FRBR (Functional Requirements for Bibliographic Records) – https://www.oclc.org/research/activities/frbr.html

Theme 3: Enrich Descriptions and Definitions	Clarify the purpose an interlink	Participants were at times unsure of the purpose of an interlink.	11	6	"And the other thing which I haven't mentioned so far, is the user, the end user. Who are your end users? What is it adding to them? What value is it adding to them? How are they going to use it? The different, what different types of users do you have? Um, you know, for the academic researcher, this might be all brilliant, but for the local history person who just wants an image of James Joyce, this is overkill for them", P.8 "It's just a thing with Linked Data, for me, to get my head around is why you are doing it, like why are you relating these two sources from the same, that relate to the same person", P. 9 "And what's the purpose of, of the interlink? Is it just to describe the relationship between the two?", P. 14
	Clarify the level of visibility for data entered	It was not always clear to participants what data would be published or added to the RDF graph.	5	4	"So, the interlink has appeared now with, the link metadata description and the rationale. Okay. Does that rationale appear to the public?", P.7 "And is this then, this justification, is this for your purposes to see what our decision-making process was, or would this be something that you would see as being used for anything?", P. 9 "Everything I put here, would that be publicly visible?", P. 11
Theme 4: GUI Requirements	Alter layout of interlinking screen	It was sometimes unclear which entities needed to be interlinked.	28	12	"I thought those were the ones who were linking, but no, we're linking this to that", P. 1 "So, by pair of resources, do you mean this and this?", P. 14 "Am I trying to link the, the books with the author?", P. 15
	Improve navigation when adding entities	Participants were sometimes unsure of what button to click in order to add an entity.	25	11	"So how do I get to VIAF dataset", P. 6 "I'm going to go to the edit button on the right-hand side. The little editor's pen. I can't see where I'm going to add the URI", P. 7 "So, I'm not sure what to do with the URI that I've just found", P.10
	Visualisations need to be easier to read	Natural language labels need to be added to the visualisations.	17	9	"I'd like to see the labels so I can read it better", P. 12 "I can imagine for a very complex records, um, yeah the natural language would be great to be at to have", P.14 "This is quite hard to read though", P. 15

Theme 4: GUI Requirements	Need clearer feedback	It was unclear whether data had saved or copied.	16	12	"The page popped there, but I'm not sure if that meant that it actually did save that. So, I'm hoping that it did", P. 2 "I trusted that the action had completed, but I wasn't getting enough feedback", P. 3 "Not sure if that copied", P. 6
	Integrable	Participants indicated that, in order to be usable, it would be important that NAISC-L be integrated into current cataloguing software	14	9	"I think in anything like this, it would probably need to be integrated with the cataloguing tool", P. 3 "the question you always ask then is, how could this be integrated with a cataloguer's workflow within the context of an academic research library, em, using a library management system", P. 7 "And I think the other thing that would be interesting would be to see if this could be developed as a module that can be buried in other cataloguing systems", P. 10
	Edit and rank external dataset list	Participants sometimes did not notice the external datasets. Participants also indicated an interest in adding more datasets and ranking them.	8	6	"I'm looking at the screen, I'm not seeing that open in front of me, so I am going to use Google to get access to that data set", P. 7 "Yeah, I think cultural heritage, Irish cultural heritage, open data sets would be really, really useful', P. 5 "Are they ranked? Do you have them ranked?", P. 8
	Incorporate dataset search within NAISC-L	Participants did not like switching between tabs	8	4	"Click directly into the search for each resource", P. 8 "Bury SemFacet search within your own tool", P. 10 "the way tool works now is that you have to open the other repository, you have to then know that repository. Know where to find the URI. Know how to search it and so on. So again, if it's technically possible, it would be nearly, or it would be better if you didn't have to go to the repository", P. 12
	Personalisable	Participants indicated that NAISC-L would need to be personalised for different institutions.	8	3	"You could probably have a free text where you could type in from your standard terminologies perhaps", P. 1 "I mean, an organisation will probably at the start, want to set it up with whatever their preferences.", P. 8 "And then along those lines, um, I can imagine that if you had the tool for a specific context or a specific institution or a specific collection type, that the chosen linking vocabularies will be very close to what you actually need", P. 12

Theme 4: GUI Requirements	Ability to review interlinks whilst working	Participants stated that viewing the RDF visualisations whilst creating interlinks would be useful.	7	6	"So that you could kind of see that, that you're actually kind of building towards, uh, to, uh, uh, you know, a rich resource graph and that that would be something that, that might be quite nice to, to, to visualize.", P. 3 "I'm just thinking what would be useful from a library point of view when you're kind of adding further links, is to be able to review your approach for similar items and the wider kind of set of links, which I don't know, you may well be able to do with that", P. 6 "Um, and I'm wondering, is there any way that you could actually see that [the visualisation], uh, uh, um, during step four?", P. 13
	Add more link-types	More link-types needed in order to express different kinds of interlinks.	4	2	"Again, I would pick relation, but again, I'm screaming for something more concrete", P. 8 That was one of the things I was struggling with, um, is to find something where the, the word of the link, the link-type was meaningful to me and as specific as possible.", P. 12
	Interactive visualisations	Make the RDF visualisations more interactive by adding a zoom function and clickable URIs.	4	2	"So, is there no way to zoom in and zoom out of this?", P. 13 "Um, there was just a, the, the only real thing I would say is the clickable links. That would be handy just to see, so you can go and view the resource and sort of remind yourself of your justifications for your choices.", P. 14
Theme 5: Automation	Time consuming process	Participants found NAISC-L the process to take a lot of time but that this may improve with increased experienced and when using real data.	16	8	"I wonder if you were adding, just say you had 300 items to add links to, and maybe this is early days as well, what, what, what people, what would they, how would they feel about going through having to do one by one", P. 4 "I mean, in a real world, I would sort of see the materials so I'd know what I'm dealing with", P. 11 "And then the other thing is that, you know, for sort of, um, new-fangled technology, it's very manual, you know and that's, I don't see that as a drawback of the tool, but maybe that's just the reality of creating those links", P. 12
	Automatically suggest link-types	Automatically select appropriate link-types/predicates.	8	5	"Definitely the more, the more automation the better", P. 4 "I think probably in an ideal world, you'd do it automatically", P. 5 "But I mean if you're constantly having to declare links should, should, isn't the idea that in some way those links would be generated automatically or intuitively", P. 11

Theme 5: Automation	Automatically add entity data	Automatically populate data fields.	6	4	“Also, yeah. I don't know, uh, if there if there was anything, if you could get something from the original resource in terms of its description or additional metadata that's associated with that original term, or the primary or the secondary term”, P. 3 “It won't add in these automatically, no? I need to add these descriptions?”, P. 14 “Can you automate it so you can import directly?”, P. 15
	Automate provenance justification	Provide a list of pre-written interlink justifications.	3	2	“you could possibly come up with, you know, again, you know, fixed descriptions”, P. 7 “So, you could also come up with a list of justification, of like generic standard justifications over time if you decided, okay look, here's the types of things that we're always saying. So, let's have a list of them”, P. 8

5.3.5.3 PSSUQ Scores

As mentioned in Section 5.2.3, PSSUQ items are scored from 1 to 7 with lower scores indicating more positive perceptions. Also, for the purpose of this experiment, sufficient usability was considered to be scores strictly lower than 4. The PSSUQ scores for each participant and the mean scores for each item can be found in Table 17 below. It can be seen that the mean score for each item is below 4, indicating that participants were generally in agreement with the PSSUQ items and that sufficient usability was achieved for all items. The mean score for each item, except for Item 9, was lower than 3 signifying that participants' perceptions were mostly positive. Item 9, which reads, *"The system gave error messages that clearly told me how to fix problems",* had a mean score of 3.64 signifying more mixed perceptions of this item. It is worthy of note that four participants choose N/A for this item which may have impacted results, and others verbally commented that they disagreed with the items as they did not face any errors whilst completing the TAT tasks. Upon reflection, a task which purposefully led participants to encounter an error could have been included so that they could experience error messaging.

The mean SysUse, InfoQual and InterQual subscale scores, as well as the mean Overall score, can also be found in Table 17. All mean scores were less than 3 indicating only mild usability and utility issues overall. Again, sufficient usability was achieved as the mean score for each scale was lower than 4.

Table 17: Usability Test 1 – PSSUQ Scores

Participant	1. Overall, I am satisfied with how easy it is to use this system	2. It was simple to use this system	3. I could effectively complete the tasks and scenarios using this system	4. I was able to complete the tasks and scenarios quickly using this system	5. I was able to efficiently complete the tasks and scenarios using this system	6. I felt comfortable using this system	7. It was easy to learn to use this system	8. I believe I could become productive quickly using this system	9. The system gave error messages that clearly told me how to fix problems	10. Whenever I made a mistake using the system, I could recover easily and quickly	11. The information (such as on-line help, on-screen messages, and other documentation) provided with this system was clear
1	4	4	4	3	3	5	3	2	2	2	4
2	3	3	3	4	3	4	2	2	5	4	4
3	3	2	3	3	3	3	2	2	4	3	3
4	3	3	3	2	2	3	3	2	N/A	N/A	4
5	2	3	3	3	2	3	3	2	4	4	2
6	2	3	3	2	2	3	2	2	N/A	N/A	3
7	2	2	3	4	3	3	2	2	6	3	4
8	2	2	2	4	2	2	2	2	4	2	2
9	2	2	1	1	1	1	1	1	2	2	1
10	3	3	1	2	1	4	3	2	N/A	3	2
11	3	3	3	4	3	4	2	4	5	4	3
12	1	1	1	1	1	1	1	3	N/A	1	1
13	4	4	5	3	3	2	1	1	3	3	2
14	1	1	1	2	2	1	1	1	1	1	1
15	3	4	2	4	5	5	3	1	4	2	4
Mean	**2.53**	**2.67**	**2.53**	**2.80**	**2.40**	**2.93**	**2.07**	**1.93**	**3.64**	**2.62**	**2.67**
Standard Deviation	**0.88**	**0.94**	**1.15**	**1.05**	**1.02**	**1.29**	**0.77**	**0.77**	**1.43**	**1.00**	**1.14**

Participant	12. It was easy to find the information I needed	13. The information provided for the system was easy to understand	14. The information was effective in helping me complete the tasks and scenarios	15. The organisation of information on the system screens was clear	16. The interface of this system was pleasant	17. I liked using the interface of this system	18. This system has all the functions and capabilities I expect it to have	19. Overall, I am satisfied with this system	SysUse (1-8)	InfoQual (9-15)	InterQual (16-18)	Overall (1-19)
1	4	4	4	5	2	2	2	2	3.50	3.57	2.00	3.21
2	4	3	4	4	1	1	2	2	3.00	4.00	1.33	3.05
3	3	2	2	3	2	2	2	2	2.63	2.86	2.00	2.58
4	3	4	2	2	2	2	N/A	3	2.63	2.14	1.33	2.26
5	2	2	2	2	1	1	1	2	2.63	2.57	1.00	2.32
6	3	3	3	2	2	2	4	3	2.38	2.00	2.67	2.32
7	3	N/A	N/A	2	2	2	3	2	2.63	2.57	2.33	2.53
8	2	2	2	3	2	2	3	2	2.25	2.43	2.33	2.32
9	1	1	1	1	2	1	1	1	1.25	1.29	1.33	1.26
10	3	4	3	2	2	2	3	2	2.38	2.43	2.33	2.37
11	3	3	3	3	4	4	5	4	3.25	3.43	4.33	3.53
12	1	1	1	1	1	1	5	1	1.25	0.86	2.33	1.26
13	3	4	3	3	1	1	2	3	2.88	3.00	1.33	2.68
14	2	2	2	1	1	1	2	1	1.25	1.43	1.33	1.32
15	5	5	4	5	3	4	N/A	N/A	3.38	4.14	2.33	3.32
Mean	**2.80**	**2.86**	**2.57**	**2.60**	**1.87**	**1.87**	**2.69**	**2.14**	**2.48**	**2.58**	**2.02**	**2.42**
Standard Deviation	**1.05**	**1.19**	**0.98**	**1.25**	**0.81**	**0.96**	**1.26**	**0.83**	**0.71**	**0.93**	**0.80**	**0.69**

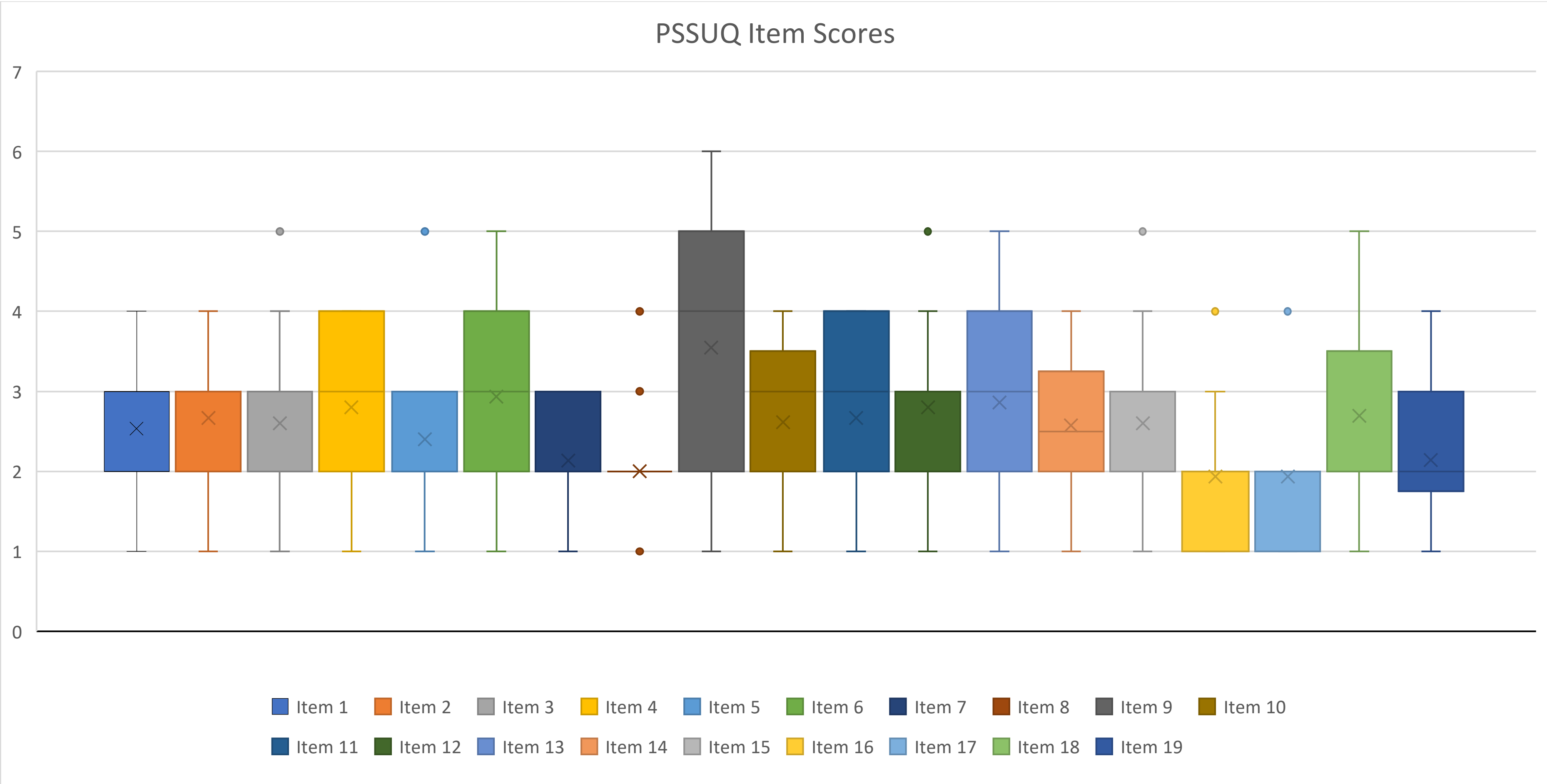


Figure 47: Usability Test 1 – PSSUQ Scores Box Plot

Figure 47 above presents a box plot for each item in the PSSUQ. Outliers can be seen in Items 3, 5, 8, 12, 15, 16, and 17. However, given that sufficient usability (scores less than 4) was achieved for all items, it was decided not to exclude any data from the analysis of this experiment.

Figure 48 below presents a box plot of the overall and subscale scores for all participants. One outlier, Participant 11, can be seen in the results of InterQual subscale. Again, as the overall InterQual score was below the targeted value of 4, it was decided not to exclude any data from the analysis of this experiment.

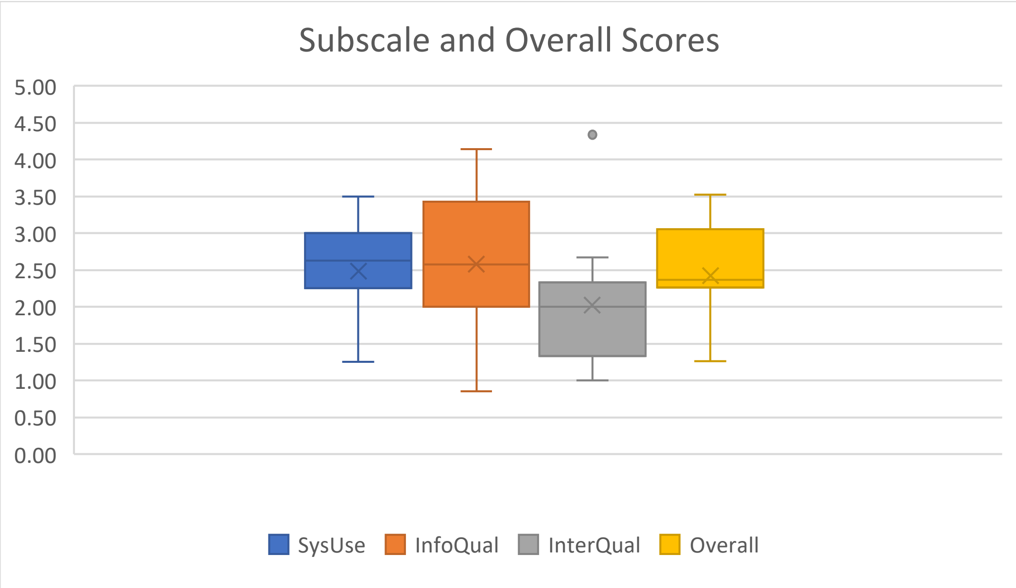


Figure 48: Usability Test 1 – PSSUQ Subscales Box Plot

Reliability

As stated in Section 5.2.6, Cronbach's alpha indicates the internal consistency of questionnaire items. The alpha index for the PSSUQ questionnaire applied in this experiment was 0.92, indicating a high level of internal consistency and reliability in the responses.

5.3.6 Discussion

The hypothesis (**H1.1**) being investigated as part of Usability Test 1 was whether *'using the NAISC-L Framework to create LD interlinks yields high task performance with sufficient usability for IPs'*. Here *'high task performance'* was achieved if 83% of interlinks were completed with 83% accuracy, and *'sufficient usability'* was achieved if PSSUQ scores were lower than a neutral score of 4 (on a scale from 1 to 7).

In the TAT, the mean number of interlinks created and the mean accuracy score for the interlinking task was above 83%. This indicates a high task performance for the creation of interlinks using the NAISC-L framework. The mean score for each item of the PSSUQ, as well as for the SysUse, InfoQual, InterQual and Overall scores, was lower than 4 indicating that sufficient usability was achieved for IPs when using NAISC-L.

Overall, the experiment indicated that **IPs can use NAISC-L for the creation of LD interlinks with high performance and sufficient usability – confirming the hypothesis (H1.1) of this experiment**.

The research question of this thesis focuses on facilitating IPs to engage with the process of Linked Data interlinking with effectiveness, efficiency and satisfaction. Effectiveness is considered to be the degree to which users can accurately complete LD interlinks. Efficiency refers to the time taken to create an interlink, and satisfaction is the extent to which NAISC-L meets a user's needs and expectations. Table 18 below outlines how effectiveness, efficiency and satisfaction (EES) were measured in this experiment.

Table 18: Usability Test 1 – EES Measures

Area	Measure	Results
Effectiveness	TAT Task 4: No. of Interlinks Created	Task 4 had a mean completeness score of 95.55% as all but two of the fifteen participants created six interlinks.
	TAT Task 4: Accuracy of the Interlinks	The mean accuracy score was 91.12%. 10 of the 15 participants had an accuracy score of 100%.
	PSSUQ: SysUse	The SysUse portion of the PSSUQ includes items which assess system efficacy. The mean SysUse score was 2.48, indicating that participants had mostly positive perceptions of these items.

Efficiency	TAT Task 4: Time taken to create 6 interlinks	The mean time to create six interlinks was 19.6 minutes resulting in an average of 3.27 minutes per interlink. There were no outliers for this task. The inter-quartile range was 14.63 and the standard deviation was large at 7.82 minutes.
	PSSUQ: SysUse	The SysUse portion of the PSSUQ includes items which assess system efficiency. The mean SysUse score was 2.48, indicating that participants had mostly positive perceptions of these items.
	Thematic Analysis	A code in Theme 5 indicated that participants found NAISC-L to be '*time consuming'* as the process was quite manual. However, participants also indicated that the experience would be different with increased familiarity.
Satisfaction	PSSUQ: InterQual & Overall	The InterQual portion of the PSSUQ investigates whether a system met the expectations of a user. The mean InterQual score was 2.02, indicating that participants had mostly positive perceptions of these items. The Overall PSSUQ score includes items which specifically measure user satisfaction. The mean Overall score was 2.42, indicating that users had a mostly positive experience of NAISC-L usability and utility.
	Thematic Analysis	A number of codes emerged from the data which indicate that the participants were satisfied with NAISC-L. These codes form part of Theme 1, which relates to the usability and utility of NAISC-L, and include: • Useful • RDF visualisations were helpful and useful • Quicker and easier to use over time • Clear GUI • Intuitive and straightforward • Relationship and link-type definitions were useful and helpful • User-friendly • Suitable for non-expert LD users

From Table 18 it can be seen that participants considered NAISC-L to be effective and satisfactory. Feedback on the efficiency of NAISC-L was somewhat mixed as the SysUse portion of the PSSUQ had positive results, however, the thematic analysis revealed that participants found the process to be time-consuming. Overall, participants had a positive reaction to the NAISC-L interlinking process, finding it usable, useful and user-friendly. Some suggestions were made, such as increased automation and changes to the GUI, in order to make the tool more efficient and to increase usability, which will be incorporated into future iterations of NAISC-L.

5.3.7 Experiment Summary

Section 5.4 presented the results of Usability Test 1 which evaluated the performance of participants in creating interlinks as well as the usability of NAISC-L. The performance of participants was found to be high and sufficient usability was achieved. Measures of effectiveness and satisfaction were also high but measures of efficiency were found to be mixed. However, participants indicated that efficiency might improve with familiarity, the use of real data, and increased automation. New user requirements were also identified, primarily for the GUI. In line with the DS approach, these results were used to inform the second iteration of NAISC-L. A discussion of the results of Usability Test 1 in relation to the results of Usability Test 2 and the Field Test can be found in Section 5.6.

5.4 Usability Test 2

In line with the DS approach, Usability Test 2 was conducted in order to evaluate the second iteration of NAISC-L, as described in Section 4.2.2.3. The focus of this experiment was to evaluate the usability and utility of NAISC-L, and to evaluate the quality of the data created using NAISC-L. Usability Test 2 consisted of an Interlink Creation Test, the PSSUQ and a data quality (DQ) questionnaire. The experiment was completed online by 96 IPs from varying LAM backgrounds.

5.4.1 Interlink Creation Test

The Interlink Creation Test (ICT) required participants to create three interlinks using NAISC-L. The same entities were provided for all users, as such, the primary task was to select an appropriate link-type in order to interlink them. The scenario, or context, of the ICT was the same as the TAT described in Section 5.3.1 i.e. a cataloguer creating interlinks from entities, pertaining to the Irish author James Joyce, in the French National Library (BNF) to related entities found in other LD datasets. The ICT was followed by the PSSUQ (as described in Section 5.2.3) and a DQ questionnaire (as described in Section 5.2.5). All components of this test were conducted online.

Prior to completing the ICT, participants were randomly split into four user groups. Participants were presented with a different version of NAISC-L depending on their user group. Versions either included or excluded the Interlinking Guide and also either included or excluded the Provenance RDF Graph and Visualisation portions of NAISC-L (see Table 19).

Table 19: ICT – NAISC-L Versions

NAISC-L Version	**A**	**B**	**C**	**D**
Interlinking Guide	Yes	No	Yes	No
Provenance Output RDF Graph and Visualisation	Yes	Yes	No	No

This versioning was done in order to compare participants' user experience, interlinking accuracy, interlink completion, and DQ perception depending on the level of guidance and provenance information they were presented with. Users without the interlinking guide selected a link-type via Linked Open Vocabularies. Users without the provenance output did not have access to the provenance graph or visualisation. However, all participants were able to view the provenance data during the interlinking process.

5.4.2 Hypotheses

The hypotheses being investigated as part of this experiment are as follows:

Hypothesis 2.1 (H2.1): Using the NAISC-L Framework to create LD interlinks yields high task performance with sufficient usability and sufficient data quality for IPs.

Hypothesis 2.2 (H2.2): The number of interlinks completed is higher for participants who used the NAISC-L Interlinking Guide when compared to participants who did not use the Interlinking Guide.

Hypothesis 2.3 (H2.3): Interlink accuracy is higher for participants who used the NAISC-L Interlinking Guide when compared to participants who did not use the Interlinking Guide.

Hypothesis 2.4 (H2.4): PSSUQ scores are better for participants who used the NAISC-L Interlinking Guide when compared to participants who did not use the Interlinking Guide.

Hypothesis 2.5 (H2.5): Data Quality perceptions are better for participants who had access to the Interlink Provenance Output RDF Graph and Visualisation when compared to participants who did not have access to the provenance output.

Task performance and usability were evaluated via the ICT. Task performance above 66%, for both the number of interlinks completed and interlink accuracy, was considered to be high as a score of 66% indicated that participants completed an average of 2 out of 3 interlinks accurately. Usability was measured using the PSSUQ (described in Section 5.2.3) and DQ was evaluated using a data quality (DQ) questionnaire (described in Section 5.2.5). In this experiment, *'sufficient usability'* was considered to be scores strictly lower than a neutral score of 4 (on a scale from 1 to 7). As stated previously, lower PSSUQ values indicate more positive perceptions of a system. *'Sufficient data quality'* was considered to be scores above 5, as higher values in the DQ questionnaire indicate more favourable perceptions.

5.4.3 Methodology

In order to test the experiment hypotheses, an online user experiment, comprising of a Pre-Test Questionnaire, an ICT, the PSSUQ and a DQ Questionnaire was conducted. This experiment was carried out individually by the participants online with no time limit. Qualtrics survey software was used in order to conduct and distribute this user experiment. Participants completed the Pre-Test Questionnaire, PSSUQ and DQ Questionnaire via the survey. The survey also included a link to the online versions of NAISC-L, thus allowing participants to complete the ICT.

The experiment was structured as follows:

1. **Informed Consent:** Here the experiment was explained in detail to the participant in writing. Participants who provided consent to complete the usability test proceeded with the experiment (see Appendix 12 for the Participant Information Sheet and Consent Form).
2. **Pre-Test Questionnaire:** Participants rated their knowledge of the SW, LD, RDF, URIs and ontologies as part of the pre-test questionnaire. The questionnaire is described in more detail in Section 5.3.1
3. **Interlink Creation Test**: Participants were given a scenario and three interlinking tasks to be complete via one of four version of NAISC-L as described in Section 5.5.2.
4. **Post-Test Questionnaires:** After completing the ICT, the participants were asked to fill out the PSSUQ and a DQ Questionnaire as described in Sections 5.2.3 and 5.2.5 respectively.

Statistical Tests

A variety of statistical tests, defined below, were used to analyse the experiment data. Note that, across these tests, the null-hypothesis (H_0) is accepted when the p-value is greater than the chosen alpha level. The null hypothesis is rejected and the alternative hypothesis (H_1) is accepted when the p-value is less than the chosen alpha level. The standard alpha level (α = .05) was applied to the tests conducted in this thesis. P-values were rounded to three decimal points. The statistical tests used to analyse the data of this experiment include:

- **Levene Test of Homogeneity of Variances:** The Levene test (Levene, 1960) is used to assess the equality of variances for a variable across two or more groups. The null-hypothesis of this test is that the population variances are equal, or homogenous. The alternative-hypothesis of this test is that the population variances are not homogenous. Homogeneity of the data is required for some statistical tests used in this thesis.

- **Shapiro-Wilk Test:** The Shapiro-Wilk Test (Shapiro & Wilk, 1965) is used to assess the normality of data. The null-hypothesis of this test is that the population is normally distributed. The alternative-hypothesis of this test is that the population is not normally distributed. The normality of the data is required for some of the statistical tests used in this thesis.

- **Mann-Whitney U Test:** The Mann-Whitney U Test (Mann & Whitney, 1947) is a nonparametric test used to determine whether two independent samples derive from the same population by assessing whether there is a statistically significant difference between the medians of the two samples. The Mann-Whitney U Test does not assume data normality. The null hypothesis of this test is that the population medians are equal, indicating no statistically significant difference between the samples. The alternative hypothesis of this test is that the population medians are not equal, indicating a statistically significant difference between the samples.

- **Independent-Samples T-Test:** The Independent Samples T-Test (Student, 1908) is a parametric test used to compare the means of two independent groups in order to determine whether there is a statistically significant difference between their means. The test assumes homogeneity of the variances and normality of the data, however, it can be used on non-normalised data where the number of participants is greater than 20. The null hypothesis of this test is that the population means are equal, indicating no statistically significant difference between the groups. The alternative hypothesis of this test is that the population means are not equal, indicating a statistically significant difference between the groups.

- **Kruskal-Wallis Test:** The Kruskal-Wallis Test (Kruskal & Wallis, 1952) is a rank-based nonparametric test used to determine whether there is a statistically significant difference between the medians of two or more independent groups. The Kruskal-Wallis Test does not assume data normality. The null hypothesis of this test is that the population medians are equal, indicating no statistically significant difference between groups. The alternative hypothesis of this test is that the population medians are not all equal, indicating a statistically significant difference between groups.

- **One-Way ANOVA:** The one-way ANOVA (Fisher, 1919) is a parametric test used to determine whether there is a statistically significant difference between the means of two or more independent groups. The one-way ANOVA assumes homogeneity of the variances and normality of the data, however, the test can be used on non-normalised data where the number of participants is greater than 20. The null hypothesis of this test is that the population means are equal, indicating no statistically significant difference between the groups. The alternative hypothesis of this test is that the population means are not all equal, indicating a statistically significant difference between the groups. **Tukey's Test** (Tukey, 1949) was used to confirm where differences occurred by comparing all possible pairs of means.

- **Spearman Rank-Order Correlation Coefficient:** Spearman's correlation (Spearman, 1904) is a nonparametric test used to measure the strength of a monotonic relationship between variables. The null hypothesis of this test states that there is no correlation between the variables. The alternative hypothesis states that there is a correlation between the variables. The Spearman test is less sensitive to outliers than the Pearson correlation test.

- **Pearson Correlation Coefficient:** Pearson's correlation (Pearson, 1895) is a parametric test used to investigate how strongly two continuous variables are linearly related. A relationship is considered to be linear when a change in one variable is associated with a proportionate change another. The null hypothesis of this test states that there is no correlation between the variables. The alternative hypothesis states that there is a correlation between the variables. The Pearson test assumes normality of the data, however, as $N \geq 20$, the test was suitable for use on non-normalised data.

5.4.4 Participants

Non-probabilistic sampling methods were used to recruit the participants for this study whereby LAMs were contacted directly with a description of the research and a link to the survey to distribute to staff. Specifically, LAMs known to conduct research and those with functioning LD projects were contacted. IPs with a known interest in LD were also directly contacted as well as IPs who previously stated that they would like to be contacted for future NAISC-L research after completing a LD survey conducted during the requirements gathering process as described in Section 3.3. Snowball sampling was also used whereby the link to the survey was forwarded by participants to other suitable participants.

Overall, the survey was fully completed by 97 participants, however, one of these participants was a self-described independent consultant in Linked and Structured Metadata in libraries and on the Web. As this person was not, and had never been, an IP, their data was excluded from the survey. A further 139 surveys were excluded from data analysis on the basis that they were incomplete. It is suspected that this high number is due to the experiment taking between 20 to 30 minutes to complete.

The 96 participants included in the survey were all IPs and were dived into the four versions of NAISC-L as outlined in Table 20. Although participants were evenly distributed amongst the four NAISC-L versions via the survey, there is a slight discrepancy (+/-4) between the groups due to some incomplete responses.

Table 20: ICT – Participants per NAISC-L Version

NAISC-L Version	A	B	C	D
No. of Participants (N)	23	25	22	26

As part of the experiment the participants rated their knowledge of the SW, LD, RDF, URIs and ontologies as Not Knowledgeable (NK), Slightly Knowledgeable (SK), Moderately Knowledgeable (MK), Very Knowledgeable (VK), or Extremely Knowledgeable (EK). These results can be seen in Table 21 below. Note an inclusion criterion for the experiment was that participants must have at least a slight knowledge of LD.

Table 21: Usability Test 2 – Participant Knowledge Ratings

Version	N	Rating	SW	LD	RDF	URIs	SPARQL	Ontologies
A	23	NK	1	0	1	1	4	1
		SK	5	5	9	5	9	6
		MK	9	8	4	7	5	10
		VK	7	9	8	8	4	5
		EK	1	1	1	2	1	1
B	25	NK	0	0	0	1	6	0
		SK	4	4	5	0	5	6
		MK	12	10	10	14	6	12
		VK	5	7	6	5	4	4
		EK	4	4	4	5	4	3
C	22	NK	0	0	2	1	4	2
		SK	4	5	6	4	12	4
		MK	9	7	8	10	2	9
		VK	6	6	3	4	2	4
		EK	3	4	3	3	2	3
D	26	NK	0	0	2	0	8	2
		SK	5	3	8	3	5	4
		MK	10	10	6	10	9	7
		VK	7	9	6	6	3	8
		EK	4	4	4	7	1	5

On analysis of the LD Knowledge scores across the four groups, the Levene Test of Homogeneity of Variances found that the data was homogenous (see Table 22). The Shapiro-Wilk Test was then conducted in order to assess the normality of the data. The results indicated that the data was not normally distributed for each of the four groups – see Table 22.

Table 22: LD Knowledge – Levene & Shapiro-Wilk Tests

Group	**Levene Test ($\alpha = .05$)**	**Shapiro-Wilk Test ($\alpha = .05$)**
A	p = 0.649 Accept H_0 – the variances are equal based on median	p = .004, Accept H_1 – the data is not normally distributed
B		p = .008, Accept H_1 – the data is not normally distributed
C		p = .012, Accept H_1 – the data is not normally distributed
D		p = .007, Accept H_1 – the data is not normally distributed

In order to assess whether the groups are comparable based on prior LD knowledge both the Kruskal-Wallis Test and a One-Way ANOVA were conducted, the results of which can be found in Table 23. It can be seen that H_0 is accepted for both tests, thus **it is assumed that the prior LD knowledge ratings are not significantly different between the groups.** As such, groups were balanced in terms of participants' prior LD knowledge.

Table 23: LD Knowledge – Kruskal-Wallis and ANOVA Tests

Topic	**Kruskal-Wallis Test ($\alpha = .05$)**	**One-Way ANOVA ($\alpha = .05$)**
LD Knowledge	$\chi 2(2) = .914$ p = .822 Accept H_0 -medians are equal	$F(3,92) = .357$ p = .784 Accept H_0 – means are equal

5.4.5 Results

5.4.5.1 Interlink Creation Test

The results of the ICT are discussed below in terms of the number of interlinks created, interlink accuracy, and interlink creation time.

Interlinks Completed

Table 24 provides the average number of interlinks completed across the four groups. Participants had to create three links, hence possible values for the average are within the range $[0,3] \subseteq \mathcal{R}$. High task performance was achieved across all groups as the average number of interlinks created was above 66%.

Table 24: ICT – No. of Interlinks Completed

Interlinks Completed	Group A	Group B	Group C	Group D
Average	2.70	2.76	2.64	2.46
Percentage	90%	92%	88%	82%

On analysis of the completeness scores for the four groups, the Levene Test of Homogeneity of Variances found that the data was homogenous and the Shapiro-Wilk Test indicated that the data was not normally distributed for all groups – see Table 25.

Table 25: Interlink Completeness – Levene & Shapiro-Wilk Tests

<table>
<tr><th>Group</th><th>Levene Test
(α = .05)</th><th>Shapiro-Wilk Test
(α = .05)</th></tr>
<tr><td>A</td><td rowspan="4">p = 0.538
Accept H_0, variances are equal based on median</td><td>p = .000, Accept H_1 – the data is not normally distributed</td></tr>
<tr><td>B</td><td>p = .000, Accept H_1 – the data is not normally distributed</td></tr>
<tr><td>C</td><td>p = .000, Accept H_1 – the data is not normally distributed</td></tr>
<tr><td>D</td><td>p = .000, Accept H_1 – the data is not normally distributed</td></tr>
</table>

It was found that **there was no significant difference between the average number of interlinks completed across the four groups** as determined by a Kruskal-Wallis Test and a one-way ANOVA – see Table 26. This suggests that participants were able to complete a similar number of interlinks regardless of the version of NAISC-L which they used.

Table 26: Interlinks Completeness – Kruskal-Wallis & ANOVA Tests

Case	**Kruskal-Wallis Test (α = .05)**	**One-Way ANOVA (α = .05)**
Interlink Completeness	$\chi 2(2) = 1.512$ p = .680 Accept H_0, medians are equal	F(3,92) = .728 p = .538 Accept H_0, means are equal

As part of investigating **H2.2** of this experiment, the interlink completeness scores from Group A and Group C were combined to form an Interlinking Guide (IG) Group (N = 45) and the completeness scores from Group B and Group D were combined to form a Non-Interlinking Guide (NIG) Group (N = 51). A Mann-Whitney U Test and an Independent-Samples T-Test were performed between the IG and NIG groups – see Table 27.

Table 27: Completeness – Mann-Whitney U Test & Independent T-Test

Case	**Mann-Whitney U Test (α = .05)**	**Independent-Samples T-Test (α = .05)**
Interlink Completeness	U = 1106.5 p = .664 Accept H_0, medians are equal	t = .379 p = .706 Accept H_0, means are equal

Results suggest that there is no statistically significant difference in the number of interlinks completed between the IG group than the NIG group. As such **H2.2** is rejected as **the number of interlinks completed does not appear to be higher for participants who used the NAISC-L Interlinking Guide when compared to participants who did not use the Interlinking Guide**.

Interlink Accuracy

Table 28 below provides the average interlink accuracy scores across the four groups. These scores are based on how successful participants were in choosing a reasonable and semantically accurate link-type to represent the relationship between each pair of entities. For the purpose of this research, a reasonable link-type was considered to be a predicate that, according to its ontological definition, could be used to meaningfully link the given entities. Again, the average score lies within the range $[0,3] \subseteq \mathcal{R}$.

Table 28: ICT – Interlink Accuracy

Accuracy	Group A	Group B	Group C	Group D
Average Score	2.13	1.12	2.09	0.81
Average Percentage	71.00%	37.33%	69.66%	27.00%

High task performance was achieved for Group A and Group C only as both had an average accuracy score above 66%. Group A and Group C conducted the ICT using a version of NAISC-L which included the Interlinking Guide.

The Levene Test of Homogeneity of Variances found that the data was homogenous and the Shapiro-Wilk Test indicated that the data was not normally distributed across the four groups – see Table 29.

Table 29: Interlink Accuracy – Levene & Shapiro-Wilk Tests

Group	Levene Test ($\alpha = .05$)	Shapiro-Wilk Test ($\alpha = .05$)
A	$p = .737$, Accept H_0, variances are equal based on median	$p = .000$, Accept H_1 – the data is not normally distributed
B		$p = .003$, Accept H_1 – the data is not normally distributed
C		$p = .000$, Accept H_1 – the data is not normally distributed
D		$p = .000$, Accept H_1 – the data is not normally distributed

There was a statistically significant difference between the accuracy scores of the groups as determined by the Kruskal-Wallis H Test and a one-way ANOVA – see Table 30.

Table 30: Interlink Accuracy – Kruskal-Wallis & ANOVA Tests

Case	**Kruskal-Wallis Test** (α = .05)	**One-Way ANOVA** (α = .05)
Interlink Accuracy	$\chi2(2)$ = 26.822 p = .000 Accept H_1,, medians are not equal	F(3,92) = 12.138 p = .000 Accept H_1,, means are not equal

A Tukey posthoc test revealed that interlink accuracy was statistically significantly higher for Group A (2.13 ± .81 interlinks, p = .002) and Group C (2.09 ± 1.19 interlinks, p = .004) when compared to Group B (1.12 ± .81 interlinks). The Tukey posthoc test also revealed that interlink accuracy was statistically significantly higher for Group A (2.13 ± .81 interlinks, p = .000) and Group C (2.09 ± 1.19 interlinks, p = .000) when compared to Group D (0.80 ± .89 interlinks). There was no statistically significant difference between Group B and Group D (p = .645) or Group A and Group C (p = .999). This suggests that **interlink accuracy was significantly higher for participants who used a version of NAISC-L that included the Interlinking Guide (Group A and Group C), when compared to those who did not have access to the Interlinking Guide (Group B and Group D).**

In order to explore **H2.3**, the accuracy scores from Group A and Group C were combined to form an Interlinking Guide (IG) Group (N = 45) and the accuracy scores from Group B and Group D were combined to form a Non-Interlinking Guide (NIG) Group (N = 51). A Mann-Whitney U Test and an Independent-Samples T-Test were performed between the IG and NIG groups – see Table 31.

Table 31: Accuracy – Mann Whitney U Test & Independent T-Test

Case	**Mann-Whitney U Test** (α = .05)	**Independent-Samples T-Test** (α = .05)
Interlink Completeness	U = 482 p = .000 Accept H_1, medians are not equal	t = 5.937 p = .000 Accept H_1, means are not equal

These results suggest that interlink accuracy is statistically significantly higher for the IG group than the NIG group according to the Mann-Whitney U test and the Independent-Samples T-Test. This supports **H2.3** in that **interlink accuracy is higher for participants who used the NAISC-L Interlinking Guide when compared to participants who did not use the Interlinking Guide.**

Time

The average time it took participants from each group to complete the ICT can be seen in Table 32 below.

Table 32: ICT – Average Time

	Group A	Group B	Group C	Group D
Average Time 3 Interlinks (hh:mm:ss)	00:16:19	00:13:26	00:14:59	00:12:45
Average Time 1 Interlink (hh:mm:ss)	0:05:26	0:04:29	0:05:00	0:04:15

The Levene Test of Homogeneity of Variances found that the data was homogenous, and the Shapiro-Wilk Test indicated that the data was not normally distributed – see Table 33.

Table 33: ICT Time – Levene & Shapiro-Wilk Tests

Group	Levene Test (α = .05)	Shapiro-Wilk Test (α = .05)
A	p = .786, Accept H_0, variances are equal based on median	p = .000, Accept H_1 – the data is not normally distributed
B		p = .008, Accept H_1 – the data is not normally distributed
C		p = .000, Accept H_1 – the data is not normally distributed
D		p = .000, Accept H_1 – the data is not normally distributed

It was found that **there was no significant difference between the time taken to complete the test across the four groups** as determined by a Kruskal-Wallis H Test and a one-way ANOVA – see Table 34.

Table 34: ICT Time – Kruskal-Wallis & ANOVA Tests

Case	Kruskal-Wallis Test (α = .05)	One-Way ANOVA (α = .05)
ICT Time	$\chi 2(2) = 1.615$ p = .656 Accept H_0, the medians are equal	F(3,92) = .458 p = .712 Accept H_0, the means are equal

Although the time difference between groups was not significant, Group B and Group D had lower average times than Group A and Group C. It is possible that this slight increase in time for Group A and Group C was due to the participants using the Interlinking Guide. In order to investigate this further, the participant data was once again divided into the IG and NIG groups and a Mann-Whitney U Test and an Independent-Samples T-Test was conducted – see Table 35. Results suggest that **there is no significant difference between the IG and NIG groups for the time taken to complete the ICT.**

Table 35: Time – Mann-Whitney U Test & Independent-Samples T-Test

Case	Mann-Whitney U Test (α = .05)	Independent-Samples T-Test (α = .05)
ICT Time	U = 1021.5 p = .355 Accept H_0, medians are equal	t = 1.097 p = .284 Accept H_0, means are equal

Overall, although the Interlinking Guide may increase the time taken to complete an interlink, this increase is not statistically significant. However, the effect of the guide on improving interlink accuracy is significant. The positive impact on interlink accuracy of the Interlinking Guide on interlink accuracy could thus outweigh the extra time necessary in creating links.

5.4.5.2 PSSUQ

The PSSUQ was used to evaluate the usability of each version of NAISC-L. Table 36 shows the average (AVG) scores and standard deviation (SD) per group. See Appendix 13 for detailed PSSUQ results.

It can be seen that sufficient usability was achieved in all areas for each group as all average scores were less than 4. The groups with the lowest average scores were Group A and Group C, both of which included the Interlinking Guide.

Levene's Test of Homogeneity of Variances and the Shapiro-Wilk Normality Test were applied to the PSSUQ data. As can be seen in Table 37, the p-values indicate that the PSSUQ data is homogenous and normalised across all groups.

Table 36: Usability Test 2 – PSSUQ Scores

PSSUQ	A		B		C		D	
	AVG	SD	AVG	SD	AVG	SD	AVG	SD
SysUse	3.05	1.14	3.27	1.37	2.84	1.12	3.59	1.46
InfoQual	2.93	1.18	3.73	1.35	3.08	1.29	3.73	1.22
InterQual	2.86	1.08	2.99	1.51	2.7	1.13	3.38	1.59
Overall	2.98	1.07	3.42	1.3	2.92	1.13	3.62	1.32

Table 37: PSSUQ – Levene & Shapiro-Wilk Tests

PSSUQ	A		B		C		D	
	Levene	Shapiro-Wilk	Levene	Shapiro-Wilk	Levene	Shapiro-Wilk	Levene	Shapiro-Wilk
SysUse	0.932	0.555	0.986	0.171	0.874	0.406	0.449	0.625
InfoQual		0.338		0.149		0.248		0.613
InterQual		0.366		0.139		0.05		0.111
Overall		0.305		0.06		0.468		0.528

Figures 49, 50, 51 and 52 show boxplots of the average PSSUQ scores for Group A, Group B, Group C and Group D respectively.

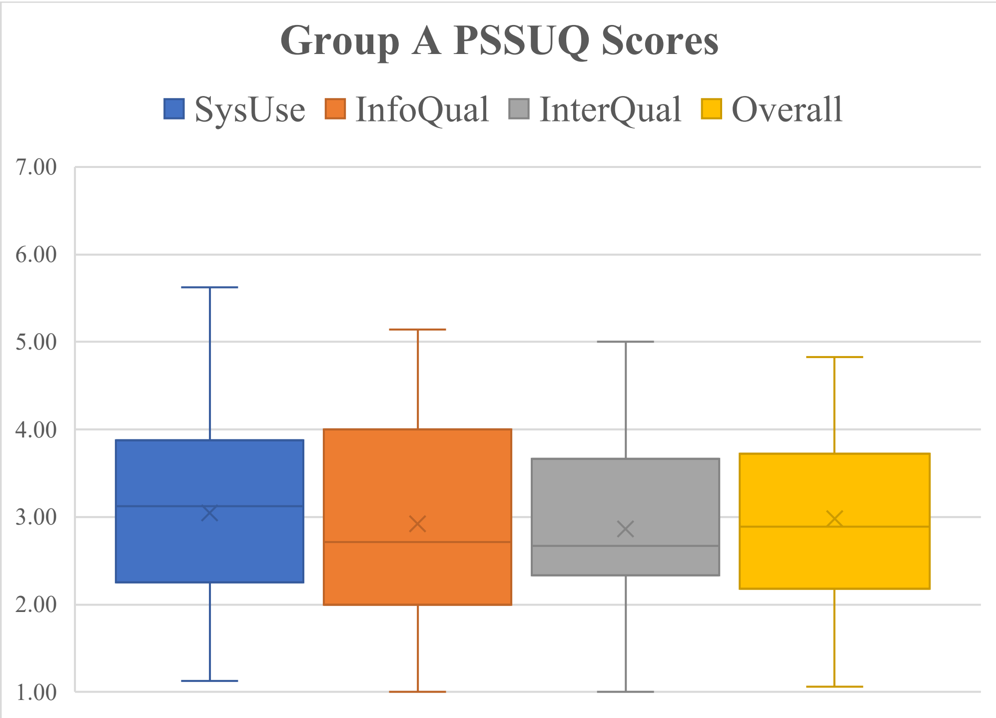


Figure 49: Group A PSSUQ Score Boxplot

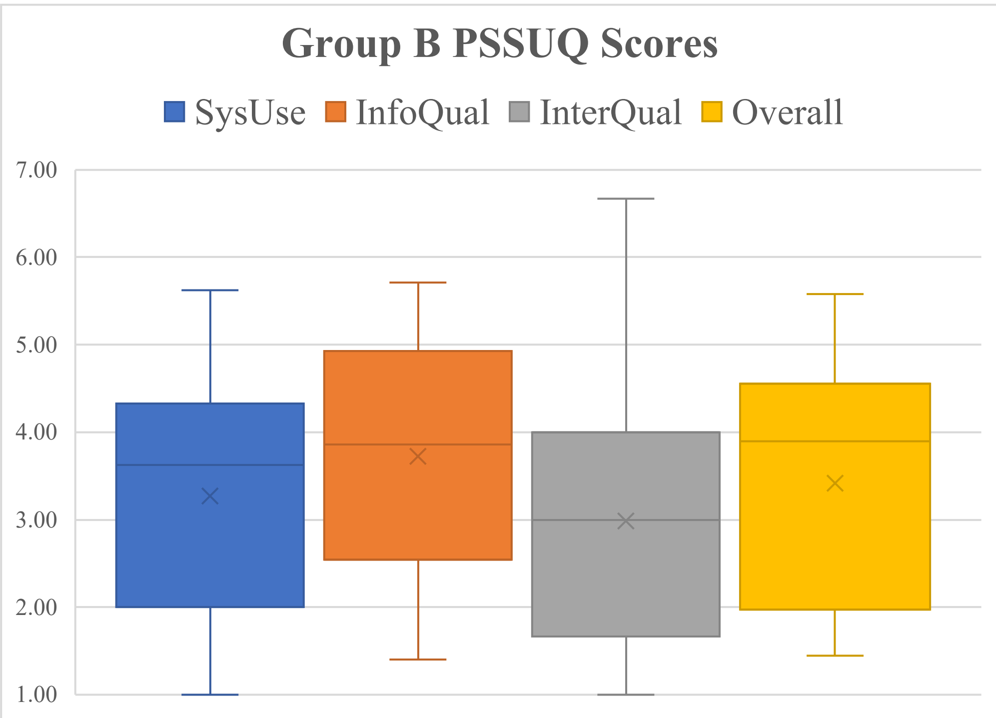


Figure 50: Group B PSSUQ Score Boxplot

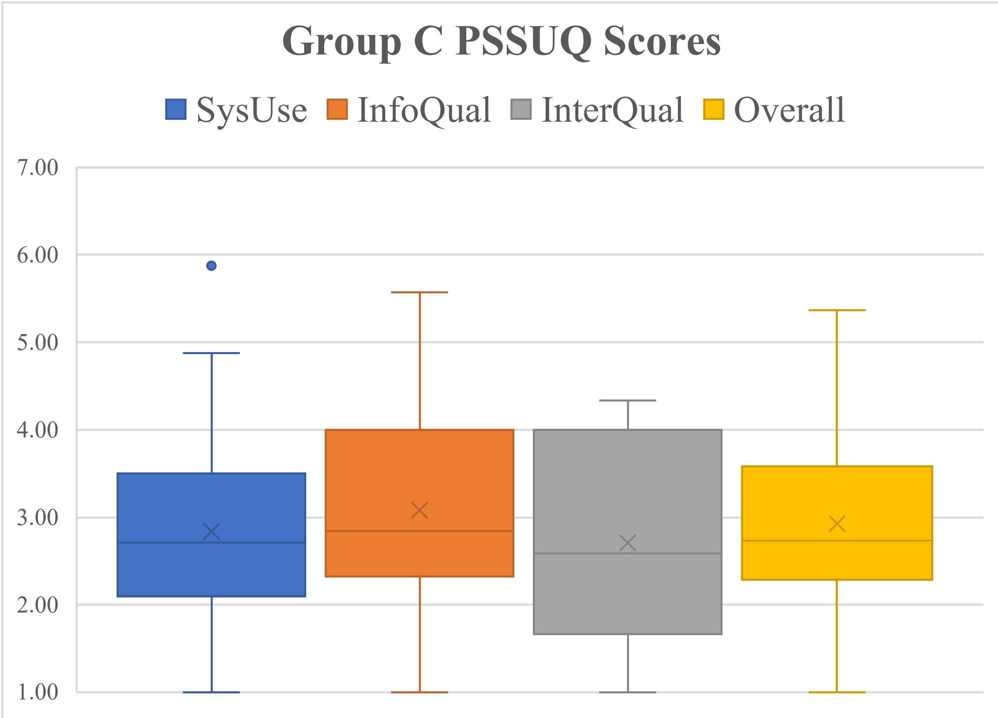


Figure 51: Group C PSSUQ Score Boxplot

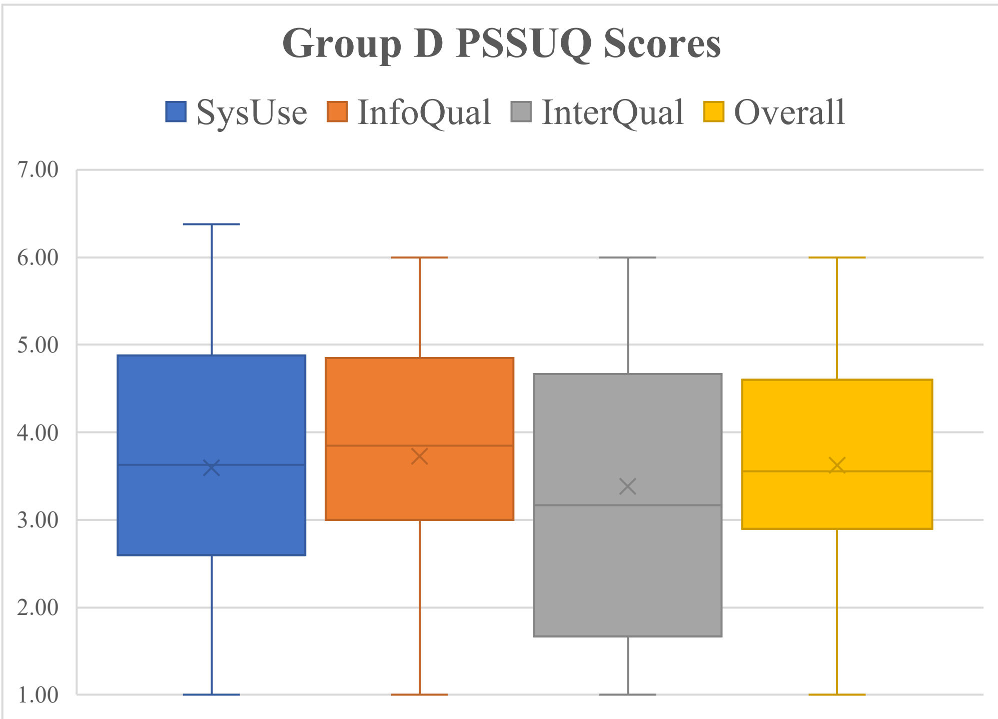


Figure 52: Group D PSSUQ Score Boxplot

As can be seen in Figure 51, the SysUse score for Group C has a data point that could be considered an outlier, however, this point was not removed from the data as it did not affect the normality of the distribution according to the Shapiro-Wilk test.

Both the Kruskal-Wallis Test and a One-Way ANOVA were used to compare groups for every aspect of the PSSUQ with all data points – see Table 38. Results indicate that **there is no statistically significant difference between the PSSUQ scores of each group**.

Table 38: PSSUQ – Kruskal-Wallis & ANOVA Tests

PSSUQ	**Kruskal-Wallis Test (α = .05)**	**One-Way ANOVA (α = .05)**
SysUse	$\chi 2(2) = 4.033$ p = .258 Accept H_0, medians are equal	F(3,92) = 1.443 p = .235 Accept H_0, means are equal
InfoQual	$\chi 2(2) = 7.516$ p = .057 Accept H_0, medians are equal	F(3,92) = 2.599 p = .057 Accept H_0, means are equal
InterQual	$\chi 2(2) = 2.418$ p = .490 Accept H_0, medians are equal	F(3,92) = 1.076 p = .363 Accept H_0, means are equal
Overall	$\chi 2(2) = 5.124$ p = .163 Accept H_0, medians are equal	F(3,92) = 1.788 p = .155 Accept H_0, means are equal

In order to investigate **H2.4**, the PSSUQ scores from Group A and Group C were combined to form an IG Group (N = 45) and the PSSUQ scores from Group B and Group D were combined to form an NIG Group (N = 51). A Mann-Whitney U Test and an Independent-Samples T-Test were then performed between these groups – see Table 39.

Table 39: PSSUQ – Mann-Whitney U Test & Independent-Samples T-Test

PSSUQ	Mann-Whitney U Test (α = .05)	Independent-Samples T-Test (α = .05)
SysUse	U = 904.5 p = .074 Accept H_0, medians are equal	t = -1.821 p = .072 Accept H_0, means are equal
InfoQual	U = 779.5 p = .007 Accept H_1, medians are not equal	t = -2.791 p = .006 Accept H_1., means are not equal
InterQual	U = 981.0 p = .219 Accept H_0, medians are equal	t = -1.432 p = .155 Accept H_0, means are equal
Overall	U = 849.5 p = .029 Accept H_1, medians are not equal	t = -2.253 p = .027 Accept H_1, means are not equal

Results indicate that both **the InfoQual and Overall PSSUQ scores are statistically significantly lower for the IG group than the NIG group**. This suggests that participants had significantly better perceptions of system information quality and overall system usability and utility for versions of NAISC-L that included the Interlinking Guide. As such, participants who used the Interlinking Guide had a better user-experience of NAISC-L when compared to participants who did not use the Interlinking Guide thus supporting **H2.4**.

5.4.5.3 Data Quality Questionnaire

The DQ questionnaire was used to evaluate the perceived DQ of each version of NAISC-L. Table 40 shows the average scores and standard deviation (SD) per group. It can be seen that sufficient data quality was achieved for all groups as all had average scores greater than 5. See Appendix 14 for detailed DQ results.

Table 40: DQ Questionnaire – Overall Scores

DQ	Overall Scores			
Group	**A**	**B**	**C**	**D**
	6.17	7.53	7.40	5.48
	5.84	8.58	6.08	10.00
	5.76	6.67	7.58	5.48
	6.96	5.64	7.00	8.00
	5.00	7.58	3.47	9.04
	6.29	5.76	5.44	7.20
	4.96	5.60	4.92	7.56
	6.96	6.80	6.29	8.96
	6.24	6.36	5.80	6.00
	9.17	6.56	6.72	6.08
	6.75	7.92	4.61	6.96
	8.88	5.64	8.72	6.80
	7.38	6.90	7.12	5.36
	5.56	4.64	5.12	6.28
	7.96	6.76	9.52	6.00
	5.48	4.70	6.36	6.91
	8.35	7.00	4.92	6.19
	5.80	6.48	7.78	9.47
	7.95	5.92	5.13	5.00
	5.72	5.00	7.75	4.69
	8.32	7.88	4.72	7.13
	6.59	4.48	10.00	7.48
	6.52	5.00		8.04
		7.00		6.48
		7.14		6.04
				6.39
Average	**6.72**	**6.38**	**6.48**	**6.89**
SD	**1.19**	**1.09**	**1.63**	**1.35**

The Levene Test found that the data was homogenous, and the Shapiro-Wilk Test indicated that the data was normally distributed – see Table 41.

Table 41: DQ Questionnaire – Levene & Shapiro-Wilk Test

Group	Levene Test ($\alpha = .05$)	Shapiro-Wilk Test ($\alpha = .05$)
A	p = .328, Accept H_0, variances are equal based on median	p = .197, Accept H_0, the data is normally distributed
B		p = .626, Accept H_0, the data is normally distributed
C		p = .711, Accept H_0, the data is normally distributed
D		p = .265, Accept H_0, the data is normally distributed

The Kruskal-Wallis H Test and a One-Way ANOVA were conducted in order to compare the DQ scores between groups – see Table 42.

Table 42: DQ Questionnaire – Kruskal-Wallis & ANOVA Tests

Case	Kruskal-Wallis Test ($\alpha = .05$)	One-Way ANOVA ($\alpha = .05$)
DQ	$\chi 2(2) = 1.680$ p = .641 Accept H_0, medians are equal	F(3,92) = .731 p = .2536 Accept H_0, means are equal

It was found that there was **no significant difference between the DQ scores across the four groups** as determined by the Kruskal-Wallis H Test and the One-Way ANOVA. This suggests that participants perceived a similar level of DQ regardless of the version of NAISC-L which they used to complete the ICT.

In order to investigate **H2.5**, the DQ scores for Group A and Group B were combined to form a Provenance Output (PO) Group (N = 48) and the DQ scores from Group C and Group D were combined to form a No Provenance Output (NPO) (N = 48) Group. A Mann-Whitney U Test and an Independent-Samples T-Test were performed between the PO and NPO groups – see Table 43.

Table 43: DQ – Mann Whitney U Test & Independent-Samples T-Test

Case	Mann-Whitney U Test (α = .05)	Independent-Samples T-Test (α = .05)
DQ	U = 1062.5 p = .754 Accept H_0, medians are equal	t = -.509 p = .612 Accept H_0, means are equal

Results indicate that the **DQ scores were not statistically significantly different between the PO and NPO groups** according to the Mann-Whitney U and the Independent-Samples T-Test. This suggests that **viewing the provenance output did not significantly improve participants' perceptions of NAISC-L data quality when compared to participants who did not have access to the provenance output,** thus, **H2.5** is rejected. It is worthy of note that provenance data was available to all users during the interlinking process but only available to Group A and Group B at the data output stage.

5.4.5.4 Correlations

Table 44 displays the correlations, using the Pearson and the Spearman correlation tests, between interlink accuracy and LD knowledge.

Table 44: Correlation between Interlink Accuracy and LD Knowledge

Interlink Accuracy & LD Knowledge	Pearson (α = .05)		Spearman (α = .05)	
	Coefficient	**p-value**	**Coefficient**	**p-value**
A	-.373	.079	-.374	.079
B	.329	.109	.367	.071
C	.234	.294	.213	.341
D	.133	.517	.197	.336

None of the correlations between accuracy and LD knowledge were found to be statistically significant indicating that a participant's prior level of self-perceived LD knowledge did not significantly influence interlink accuracy. This suggests that participants of all LD knowledge levels were able to create interlinks with the same degree of accuracy.

Table 45 displays the correlations, using the Pearson and the Spearman correlation tests, between interlink completeness and LD knowledge.

Table 45: Correlation between Interlink Completeness & LD Knowledge

Interlink Completeness & LD Knowledge	**Pearson (α = .05)**		**Spearman (α = .05)**	
	Coefficient	**p-value**	**Coefficient**	**p-value**
A	-.462	.026*	-.518	.011*
B	-.244	.240	-.265	.200
C	.302	.172	.308	.163
D	-.218	.285	-.232	.254

Only the correlation between the number of interlinks completed and LD knowledge for Group A was found to be statistically significant (scores marked with a *). This suggests that a participant's prior level of self-perceived LD knowledge influenced the number of interlinks created for Group A only.

Tables 46, 47, 48 and 49 show the correlation coefficients between the DQ questionnaire and the PSSUQ for Group A, Group B, Group C and Group D respectively. For Group A, there is a statistically significant correlation between DQ and the PSSUQ across all scores (scores marked with a *). The negative numbers indicate that as DQ scores increase, PSSUQ scores decrease (i.e. the usability improves with better DQ scores, as smaller values in the PSSUQ indicate positively perceived usability). There is also a statistically significant correlation between DQ and InterQual scores for Group C and Group D (scores marked with a *). Again the negative numbers indicate that as DQ scores increase, PSSUQ scores decrease.

Table 46: Group A – Correlation between DQ & PSSUQ

Group A				
DQ & PSSUQ	Pearson (α = .05)		Spearman (α = .05)	
	Coefficient	p-value	Coefficient	p-value
SysUse	-.463	.026*	-.501	.015*
InfoQual	-.607	.002*	-.597	.003*
InterQual	-.556	.006*	-.644	.001*
Overall	-.579	.004	-.632	.001*

Table 47: Group B – Correlation between DQ & PSSUQ

Group B				
DQ & PSSUQ	Pearson (α = .05)		Spearman (α = .05)	
	Coefficient	p-value	Coefficient	p-value
SysUse	-.135	.519	-.079	.706
InfoQual	-.055	.795	-.040	.848
InterQual	-.264	.203	-.223	.283
Overall	-.137	.514	-.082	.697

Table 48: Group C – Correlation between DQ & PSSUQ

Group C				
DQ & PSSUQ	Pearson (α = .05)		Spearman (α = .05)	
	Coefficient	p-value	Coefficient	p-value
SysUse	-.318	.149	-.293	.186
InfoQual	-.333	.131	-.381	.080
InterQual	-.437	.042	-.524	.012*
Overall	-.325	.140	-.360	.099

Table 49: Group D – Correlation between DQ & PSSUQ

Group D				
DQ & PSSUQ	Pearson (α = .05)		Spearman (α = .05)	
	Coefficient	p-value	Coefficient	p-value
SysUse	-.292	.148	-.200	.328
InfoQual	-.364	.067	-.160	.435
InterQual	-.456	.019*	-.368	.064
Overall	-.378	.057	-.266	.189

5.4.5.5 Reliability

Table 50 shows Cronbach's alpha indexes (see Section 5.2.6) for the PSSUQ and DQ questionnaire. These results suggest that the findings of these instruments can be reliably considered as scores are greater than 0.70.

Group	PSSUQ	DQ Questionnaire
A	0.947	0.927
B	0.964	0.923
C	0.963	0.960
D	0.949	0.952

5.4.6 Discussion

Five hypotheses were investigated as part of Experiment 2. The first was **H2.1** which stated that *'using the NAISC-L Framework to create LD interlinks yields high task performance with sufficient usability and sufficient DQ for IPs'*. Here *'high task performance'* was achieved if 66% of interlinks were completed with 66% semantic accuracy, *'sufficient usability'* was achieved if PSSUQ scores were lower than 4 (on a scale from 1 to 7) and *'sufficient DQ'* was achieved for DQ questionnaire scores above 5.

In the ICT, the mean number of interlinks completed and the mean interlink semantic accuracy score was above 66%. This indicates that high task performance was achieved for the creation of interlinks using the NAISC-L framework regardless of which version participants used. Similarly the average SysUse, InfoQual, InterQual and Overall PSSUQ scores across all groups was less than 4, indicating that sufficient usability was achieved for all groups using the NAISC-L framework. Finally, the average overall DQ score for each group was above 5, indicating that sufficient DQ was achieved for all groups using the NAISC-L framework.

Overall, the experiment indicated that **IPs can use NAISC-L for the creation of LD interlinks with high performance, sufficient usability and sufficient DQ – confirming H2.1 of this experiment**. Interestingly, **H2.1** was accepted for all versions of NAISC-L.

H2.2 of this experiment investigated whether 'the number of interlinks completed *is higher for participants who used the NAISC-L Interlinking Guide when compared to participants who did not used the Interlinking Guide'*. There was no statistically significant difference between the number of interlinks created across each of the four groups. Similarly, there was also no significant difference between the number of interlinks created by IG group (Group A + Group C) and the NIG group (Group B + Group D).

Additionally, there was no correlation found between perceived LD knowledge and the number of interlinks completed for Group B, Group C and Group D. This indicates that these participants were able to create a similar number of interlinks regardless of prior LD knowledge. However, there was a correlation found between the number of interlinks created and perceived LD knowledge for Group A. This is despite the fact that there was no statistically significant difference found between the perceived LD Knowledge ratings of the groups.

Overall, the experiment indicated that **the number of interlinks completed is not higher for participants who used the NAISC-L Interlinking Guide when compared to participants who did not use the Interlinking Guide – leading to the rejection of H2.2**.

H2.3 of this experiment investigated whether '*interlink semantic accuracy is higher for participants who used the NAISC-L Interlinking Guide when compared to participants who did not use the Interlinking Guide'*. On analysis of the data, it was found that there was a statistically significant difference between the accuracy scores of each the four groups. Namely, interlink accuracy was statistically significantly higher for Group A and Group C when compared to Group B and Group D respectively. Note that both Group A and Group C used the Interlinking Guide whereas Group B and D did not. When the scores of the groups were combined to form an IG group (Group A + Group C) and an NIG group (Group B + Group D), the accuracy scores of the IG group were found to be statistically significantly higher than those of the NIG group.

In addition, no correlation between perceived LD knowledge and interlinking accuracy was found. This indicates that participants were able to create interlinks with similar levels of accuracy regardless of prior LD knowledge.

Overall, the experiment indicated that **interlink accuracy is higher for participants who used the NAISC-L Interlinking Guide when compared to participants who did not use the Interlinking Guide – leading to the confirmation of H2.3**.

H2.4 of this experiment investigated whether '*PSSUQ scores are better for participants who used the NAISC-L Interlinking Guide when compared to participants who did not use the Interlinking Guide'*. On comparison of the SysUse, InfoQual, InterQual and Overall PSSUQ scores between all groups, it was found that there was no statistically significant difference between the scores. However, when the PSSUQ scores of the groups were combined to form an IG group (Group A + Group C) and an NIG group (Group B + Group D), it was found that the InfoQual and Overall scores were statistically significantly lower for the IG group than the NIG group. As lower scores in the PSSUQ indicate more favourable perceptions, it can be concluded that participants who used the Interlinking Guide perceived better overall usability and utility for NAISC-L when compared to participants who did not use the Interlinking Guide.

Overall, results indicated that **InfoQual and Overall PSSUQ scores are better for participants who used the NAISC-L Interlinking Guide when compared to participants who did not use the Interlinking Guide – confirming H2.4**.

H2.5 of this experiment investigated whether '*DQ perceptions are better for participants who had access to the Interlink Provenance Output RDF Graph and Visualisation when compared to participants who did not have access to the provenance output*'. On comparison of the DQ scores between all groups, it was found that there was no statistically significant difference between them. Similarly, when the DQ scores of the groups were combined to form a PO group (Group A + Group B) and an NPO group (Group C + Group D), it was again found that there was no statistically significant difference between the groups.

Overall, the experiment indicated that **DQ scores are not better for participants who had access to the Interlink Provenance Output RDF Graph and Visualisation when compared to participants who did not have access to the provenance output – leading to the rejection of H2.5**.

The research question of this thesis focuses on facilitating IPs to engage with the process of Linked Data interlinking with effectiveness, efficiency and satisfaction. Effectiveness is considered to be the degree to which users can accurately complete LD interlinks. Efficiency refers to the time taken to create an interlink, and satisfaction is the extent to which NAISC-L meets a user's needs and expectations. Table 50 below outlines how effectiveness, efficiency and satisfaction (EES) were evaluated in this experiment.

Table 50: Usability Test 2 – EES Measures

Area	Measure	Results
Effectiveness	ICT: No. of Interlinks Created	The average completeness score per group was above 66% indicating that participants, on average, completed at least 2 out of 3 interlinks.
	ICT: Semantic Accuracy of the Interlinks	The average accuracy score for Group A and Group C was above 66% indicating that participants, on average, completed 2 out of 3 interlinks accurately. The average accuracy for the IG Group was significantly higher than the NIG Group. This indicates that the Interlinking Guide was effective in improving interlink accuracy.
	PSSUQ: SysUse	The SysUse portion of the PSSUQ includes items that evaluate efficacy. The mean SysUse score per group was less than 4, indicating that participants had mostly positive perceptions of the items.
Efficiency	ICT: Time to create 3 interlinks	The average time to create 3 interlinks was between 12 and 16 minutes, resulting in an average of 4 to 5 minutes per interlink. Group A and Group C had higher average times than Group B and Group D, however, this difference was not statistically significant. It is possible that this slight increase in time for Group A and Group C was due to the participants using the Interlinking Guide.
	PSSUQ: SysUse	The SysUse portion of the PSSUQ includes items that assess system efficiency. The mean SysUse score per group was less than 4, indicating that participants had mostly positive perceptions of these items.
Satisfaction	PSSUQ: InterQual & Overall	The InterQual portion of the PSSUQ evaluates whether a system meets user expectations. The mean InterQual score per group was less than 4, indicating that participants had mostly positive perceptions of these items. The Overall PSSUQ score includes items that specifically measure usersatisfaction. The mean Overall score per group was less than 4, indicating that users had a mostly positive experience of NAISC-L usability and utility.
	DQ	The overall mean DQ score per group was greater than 5 indicating that participants had mostly positive perceptions of these items.

From Table 50 it can be seen that participants considered NAISC-L to be effective, efficient and satisfactory. Overall, participants across all groups had a positive response to NAISC-L, as indicated by the PSSUQ and DQ questionnaire results. Notably, participants who used NAISC-L Version A, the version which included both the Interlink Guide and the provenance output, had high semantic accuracy and high interlink completeness scores. This group also had better PSSUQ scores when compared to versions of NAISC-L which did not include the Interlink Guide. The presence or absence of the provenance output did not seem to have a significant impact on perceptions of DQ. However, as all participants had access to the provenance information during the interlinking process, it is possible that this influenced the DQ scores.

5.4.7 Experiment Summary

Section 5.4 presented the results of Usability Test 2 which evaluated the performance of participants in creating interlinks as well as the usability of NAISC-L and the DQ of its output. The performance of participants was found to be high, and sufficient usability and DQ were achieved. Results also indicated that, regardless of prior LD knowledge, participants who had access to the Interlinking Guide had significantly higher accuracy than those without access. Similarly, participants who had access to the Interlinking Guide had better perceptions of NAISC-L's usability and utility. A discussion of the results of Usability Test 2 in relation to the results of Usability Test 2 and the Field Test can be found in Section 5.6.

5.5 Usability Test 3

A Field Test was conducted in order to evaluate the second iteration of NAISC-L, as described in Section 4.2.2.3, in a real information environment using data held in the institution. The experiment was completed by three IPs working in a music archive and consisted of a Field Test, a post-test interview and the CSUQ.

5.5.1 Field Test

Field Tests are research activities conducted in the user's context (Farrell, 2016). This approach was chosen as testing under realistic conditions can capture information and reveal issues that may not arise in an artificial environment. The method used for this Field Test was a diary study whereby participants maintained a log in which they documented comments on their experience of using NAISC-L in real-time. This was then followed up with a post-test interview and CSUQ in order to gain further insight into the users' experience.

For this experiment, NAISC-L was evaluated in the context of the Irish Traditional Music Archive[135] (ITMA). ITMA holds a vast collection of materials relating to Irish traditional music, songs and dance. ITMA was recently involved in the LITMUS[136] (Linked Irish Traditional Music) project which focused on the development of the first LD framework tailored to the needs of Irish traditional song, instrumental music, and dance. The project included the development of the LITMUS ontology to represent contemporary and historical Irish traditional music practice, documentation and performance, as well LD pilot project. This project involved using 20 years of TG4 Gradam Ceoil[137] (Irish traditional music awards) performance data in order to create a LD dataset that demonstrated the use of the LITMUS ontology and vocabularies.

Over one working week, three IPs at ITMA used NAISC-L for a short period each day in order to create a set of interlinks. These interlinks connected some of the musicians and bands referenced in TG4 Gradam Ceoil LD dataset to related entities in VIAF, the OCLC-hosted name authority service. The aim of

[135] https://www.itma.ie/ accessed July 20th 2020
[136] https://www.itma.ie/litmus/info accessed July 20th 2020
[137] https://www.tg4.ie/en/other-brands/gradam-ceoil/about-gradam-ceoil/ – July 20th 2020

these interlinks was to provide authoritative information for specific individuals or groups, as well as to link to other LAMs that contributed to a VIAF record.

5.5.2 Hypothesis

The hypothesis for this experiment is similar to hypothesis of Usability Test 1 (**H1.1**) (see Section 5.3.2). However, unlike **H1.1**, this hypothesis is specifically concerned with measuring IPs ability to use NAISC-L to create interlinks in a LAM setting using real data held in the institution's database. The hypothesis is stated as follows:

Hypothesis 3.1 (H3.1): Using the NAISC-L Framework, in a LAM context, to create LD interlinks from an institution's dataset yields high accuracy with sufficient usability for IPs.

Accuracy and usability were evaluated via a Field Test which is described in Section 5.5.1. For the purpose of this research, interlink accuracy above 75% was considered to be high. Usability was also measured using the CSUQ which is described in Section 5.2.3. In this experiment, *'sufficient usability'* was considered to be scores strictly lower than a neutral score of 4 (on a scale from 1 to 7). As stated previously, lower CSUQ values indicate a better perception of a system.

5.5.3 Methodology

In order to test **H3.1**, a user experiment, comprising of a Pre-Test Questionnaire, a Field Test, a Post-Test Interview and the CSUQ, was conducted. Interviews were carried out individually with each participant and the resulting data was analysed using Thematic Analysis. Usability and utility were evaluated using the CSUQ questionnaire.

The experiment was structured as follows:

1. **Informed Consent:** Here the experiment was explained in detail to the participant both verbally and in writing. Participants who provided written consent to complete the Field Test proceeded with the experiment (see Appendix 15 for the Participant Information Sheet and Consent Form).

2. **Pre-Test Questionnaire:** Participants rated their knowledge of the SW, LD, RDF, URIs and ontologies as part of the pre-test questionnaire. The questionnaire is described in more detail in Section 5.2.1
3. **NAISC-L Introductory Session:** A 30-minute introductory session was conducted at ITMA prior to the initiation of the Field Test in which a brief introduction to NAISC-L was given.
4. **Field Test:** Participants at ITMA used NAISC-L for a period of one week in order to create interlinks from the TG4 Gradam Ceoil LD dataset to related entities in VIAF.
5. **Post-Test Interview**: The post-test interview consisted of seven questions, outlined in Section 5.2.2, which explored the participants' experience of using NAISC-L.
6. **Post-Test Questionnaire:** After completing the interview, the participants were asked to fill out the CSUQ, as described in Section 5.2.3.

5.5.4 Participants

Non-probabilistic sampling methods were used, whereby LAMs, known to have a LD dataset, were contacted directly with a request to conduct a Field Test of NAISC-L. ITMA responded to the request and three IPs working at the archive volunteered to participate in the Field Test. All participants had some prior knowledge of LD. The participants' self-perceived rating of their knowledge of the SW, LD, RDF, URIs and ontologies can be seen in Table 51 below. All participants rated themselves as knowledgeable for each of the five topics and all considered themselves Moderately Knowledgeable for LD. Additionally, one participant indicated that they had previous experience implementing a LD project. Overall, it can be seen that all participants had some prior awareness and knowledge of LD, but that none rated themselves as Extremely Knowledgeable in the area, suggesting that no participant considered themselves to be an expert user.

It was previously mentioned in Usability Test 1 that 15 participants are often recruited for user testing as they discover the majority of system issues (Faulkner, 2003). However, trialling and testing a system over a prolonged period, in this case one week, requires high engagement and involvement from participants, thus the number recruited was lower. Nonetheless, research indicates that three participants can discover approximately 65% of issues (Virzi, 1992; Nielsen & Landaur, 1993), including the majority of the most significant problems (Krug, 2013).

Table 51: Field Test Participant Knowledge Ratings

Participant	Institution	Topic / Rating	Not at all Knowledgeable	Slightly Knowledgeable	Moderately Knowledgeable	Very Knowledgeable	Extremely Knowledgeable	Previous work on a LD Project?
1	Archive	Semantic Web			•			Yes
		Linked Data			•			
		RDF			•			
		URIs			•			
		Ontologies			•			
2	Archive	Semantic Web		•				No
		Linked Data			•			
		RDF		•				
		URIs			•			
		Ontologies		•				
3	Archive	Semantic Web			•			No
		Linked Data			•			
		RDF			•			
		URIs					•	
		Ontologies				•		

5.5.5 Results

5.5.5.1 Field Test

As mentioned, for the duration of the Field Test the participants used NAISC-L to create interlinks from performers in the TG4 Gradam Ceoil dataset to name authority records for the same person or group in VIAF and, where there was no entry in VIAF, in other external datasets. One aim of the Field Test was to evaluate how successful participants were in choosing a reasonable and semantically accurate link-type to represent the relationship between each pair of entities. For the purpose of this research, a reasonable link-type was considered to be a predicate that, according to its ontological definition, could be used to meaningfully link the given entities.

A total of 34 interlinks were created by the participants over the course of a week. These interlinks were all `owl:sameAs` links from ITMA to VIAF (27 interlinks), Library of Congress (LOC) (5 interlinks), the German National Library (DNB) (1 interlink), French National Library (BNF) (1 interlink). It is worthy of note that the participants of the field-test consciously decided to use NAISC-L to specifically create only links of type `owl:sameAs` as this was a real task that they wished to perform on the TG4 dataset that they were not able to complete previously due to a lack of appropriate tooling. Despite using the same link-type throughout, participants nevertheless gained a full experience of the interlinking process.

Table 52 below provides details on the number of interlinks created per person as well as the accuracy of the interlinks. All participants had an accuracy score of 100% meaning that high accuracy (over 75%) was achieved. Although it was decided by the participants to create only `owl:sameAs` interlinks, measuring the accuracy is still useful as accuracy is dependent on both the selected link-type and the chosen external entity. In this case, the external entity of each interlink was verified to be identical to its internal entity.

Table 52: Field Test – No. of Interlinks & Interlink Accuracy

Participant	No. of Interlinks	Accuracy
1	9	100%
2	14	100%
3	11	100%
Total	**34**	**100%**

5.5.5.2 Thematic Analysis: Field Test Diary & Interview

Three rounds of thematic analysis, as visualised in Figure 53 below, were conducted on the Field Test diaries and post-test interview data (see Appendix 16 for a sample transcript). The initial round of coding identified 52 separate tags – see Figure 54 for a coding snippet. A second round of coding was then conducted in order to categorise and merge similar tags resulting in the generation of 30 codes. As with Usability Test 1 (see Section 5.3), an inductive, or 'bottom-up,' approach was followed in order to identify patterns in the data. Nevertheless, it was found that the 30 codes distilled from the Field Test data could be grouped according to the themes which emerged from Usability test 1.

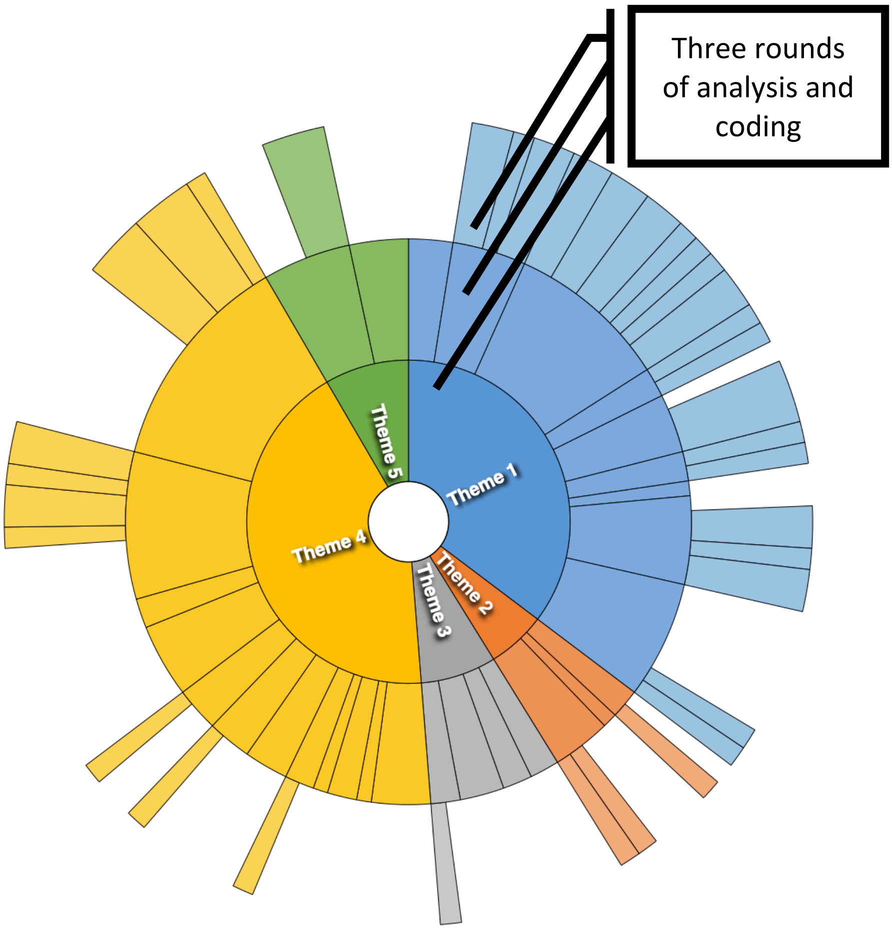
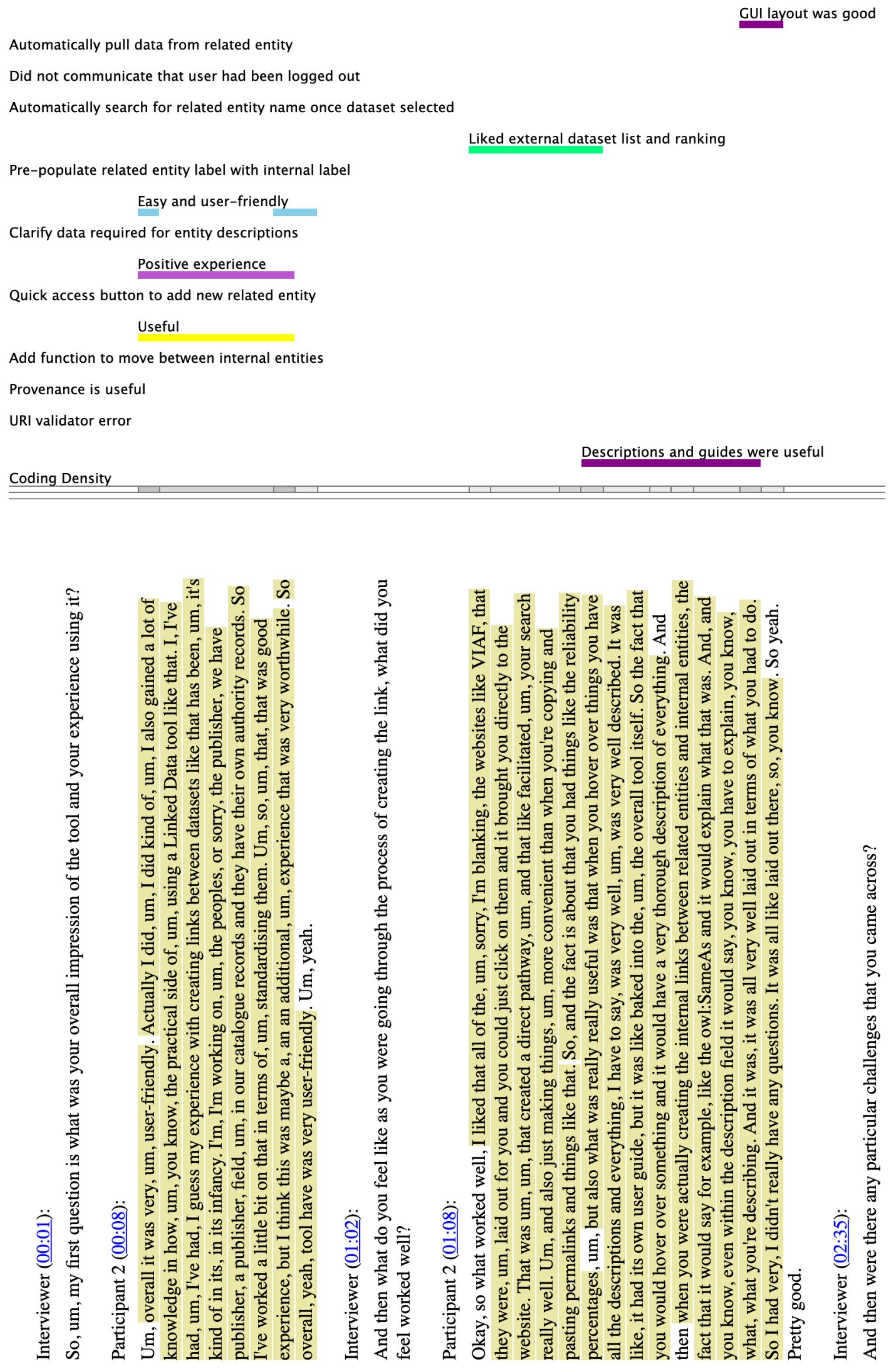


Figure 53: Field Test Coding Hierarchy


GUI layout was good

Automatically pull data from related entity

Did not communicate that user had been logged out

Automatically search for related entity name once dataset selected

Liked external dataset list and ranking

Pre-populate related entity label with internal label

Easy and user-friendly

Clarify data required for entity descriptions

Positive experience

Quick access button to add new related entity

Useful

Add function to move between internal entities

Provenance is useful

URI validator error

Descriptions and guides were useful

Coding Density

Interviewer (00:01):

So, um, my first question is what was your overall impression of the tool and your experience using it?

Participant 2 (00:08):

Um, overall it was very, um, user-friendly. Actually I did, um, I did kind of, um, I also gained a lot of knowledge in how, um, you know, the practical side of, um, using a Linked Data tool like that. I, I've had, um, I've had, I guess my experience with creating links between datasets like that has been, um, it's kind of in its, in its infancy. I'm, I'm working on, um, the peoples, or sorry, the publisher, we have publisher, a publisher, field, um, in our catalogue records and they have their own authority records. So I've worked a little bit on that in terms of, um, standardising them. Um, so, um, that, that was good experience, but I think this was maybe a, an an additional, um, experience that was very worthwhile. So overall, yeah, tool have was very user-friendly. Um, yeah.

Interviewer (01:02):

And then what do you feel like as you were going through the process of creating the link, what did you feel worked well?

Participant 2 (01:08):

Okay, so what worked well, I liked that all of the, um, sorry, I'm blanking, the websites like VIAF, that they were, um, laid out for you and you could just click on them and it brought you directly to the website. That was um, um, that created a direct pathway, um, and that like facilitated, um, your search really well. Um, and also just making things, um, more convenient than when you're copying and pasting permalinks and things like that. So, and the fact is about that you had things like the reliability percentages, um, but also what was really really useful was that when you hover over things you have all the descriptions and everything, I have to say, was very well, um, was very well described. It was like, it had its own user guide, but it was like baked into the, um, the overall tool itself. So the fact that you would hover over something and it would have a very thorough description of everything. And then when you were actually creating the internal links between related entities and internal entities, the fact that it would say for example, like the owl:SameAs and it would explain what that was. And, and you know, even within the description field it would say, you know, you have to explain, you know, what, what you're describing. And it was, it was all very well laid out in terms of what you had to do. So I had very, I didn't really have any questions. It was all like laid out there, so, you know. So yeah. Pretty good.

Interviewer (02:35):

And then were there any particular challenges that you came across?

Figure 54: Field Test – Coding Snippet from Participant 2

The five themes that emerged from the data include:

1. NAISC-L Framework Usability and Utility – 9 codes
2. Provenance Data Usability and Utility – 3 codes
3. Enrich Descriptions and Definitions – 4 codes
4. GUI Requirements – 12 codes
5. Automation – 2 codes

Theme 1 and Theme 2 relate to the usability and utility of the interlinking process and the provenance data. Codes for Theme 1 indicate that participants found NAISC-L to be useful, user-friendly and straightforward. Participants also found the provenance data to be useful and that it "adds gravitas" (Participant 3), or authority, to the interlinks.

A number of new requirements for NAISC-L were distilled from the experiment data. The codes in Theme 3 indicate a need to simplify the link-type definitions used in NASIC-L and to provide more precise descriptions for the data that should be entered into entity description and justification fields. Codes in Theme 4 highlight new requirements for the GUI such as fixing a URI validator error, adding copy buttons to entity labels, pre-populating related entity data fields with data from the internal entity, and pre-populating the justification field with data from the related entity description. Similar to Usability Test 1, Theme 5 relates to suggestions automating some of the NAISC-L processes in order to reduce the time it takes to create an interlink. Suggestions included automatically searching a dataset for a related entity and, once an entity is selected, auto-populating the appropriate data fields.

All themes and codes distilled from the data are detailed in Table 53 below. This table also includes the code descriptions, the number of references to the code in the transcripts (NR), the number of participants who referred to the code (NP), and supporting participant (P) quotes. Two quotes were chosen randomly for each code in order to demonstrate how the codes and themes were derived from the interview and field test diary data

Table 53: Field Test Themes & Codes

Theme	Code	Description	NR	NP	Selection of Quotes
Theme 1: NAISC-L Framework Usability and Utility	Easy and User-friendly	Participants stated that NAISC-L was easy to use, user-friendly, intuitive and straightforward.	11	3	"Overall it was very, um, user-friendly", P. 2 "Very easy to use", P.3
	Useful	Participants stated that NAISC-L is useful, has scope and worthwhile using.	8	3	"I could see huge opportunities that we could do things with it", P.1 "It would be very, very useful. I found it very, very useful", P.2
	Positive user-experience	Participants indicated that they had a positive experience when using NAISC-L.	6	2	"It was a very positive impression", P.1 "Actually I did, um, I did kind of, um, I also gained a lot of knowledge in how, um, you know, the practical side of, um, using a Linked Data tool like that… but I think this was maybe a, an, an additional, um, experience that was very worthwhile", P.2
	Descriptions and guides were useful	The guides, pop-ups, relationship term and interlink term descriptions were useful.	5	1	"What was really, really useful was that when you hover over things you have all the descriptions and everything, I have to say, was very well, um, was very well described. It was like, it had its own user guide, but it was like baked into the, um, the overall tool itself", P. 2
	Straightforward interlinking process	Participants stated that the process of creating an interlinking was quick and straightforward.	4	2	"When you picked your first and then you got the different options then depending on what you would pick the first time made it, made it very kind of straight forward. Because that's really, that's a really difficult thing to get your, for people to get their head around so that was, um, uh, that was, uh, displayed really well", P.1 "I found that straightforward enough because I just stuck to the same method every time", P. 3

Theme 1: NAISC-L Framework Usability and Utility	Better at using over time	Participants indicated that they became better at using NAISC-L over time	3	2	“Once you get used to the tool, that was fine”, P.1 “I suppose at the beginning you just have to take your time to get the feelers and figure out how it, how it works. But once you got into a rhythm, you know, then it, it, it worked fine”, P.3
	Good GUI Layout	Participants stated that the interlinking process was displayed well and that the interface was well laid out.	2	2	“Displayed really well”, P. 1 “And it was, it was all very well laid out in terms of what you had to do. So I had very, I didn't really have any questions. It was all like laid out there, so, you know”, P.2
	Liked external dataset list and dataset ranking	The list of externals datasets was convenient and the dataset ranking worked well.	2	1	“I liked that all of the, um, sorry, I'm blanking, the websites like VIAF, that they were, um, laid out for you and you could just click on them and it brought you directly to the website. That was um, um, that created a direct pathway, um, and that like facilitated, um, your search really well. Um, and also just making things, um, more convenient than when you're copying and pasting permalinks and things like that… and the fact is about that you had things like the reliability percentages”, P.2
	Liked visualisations	A participant liked viewing the interlinks via the visualisation	1	1	“Being able to create the links and then show them graphically. That was really good.”, P.1
Theme 2: Provenance Data Usability and Utility	Provenance is useful	Participants indicated that they found the provenance data and justifications useful	4	3	“The justifications for making them would, would be useful, I think”, P. 1 “Because sometimes it is useful because even, I know this isn't the exact same thing, but even in VIAF, it will tell you where things, where records have been created, where, where publications associated with the performer have been created. And I think that is very useful”, P. 2
	Justification adds gravitas	The justification strengthened the provenance data.	2	1	“I suppose it is because, um, I suppose it gives a bit of gravitas, you know, it was, it was, you know, I think, um, you know, that this person is professional working in an organisation who knows her stuff and then that sort of, to me, gives, gives that relationship that you've created more gravitas, more meaning.”, P.3

Theme 2: Provenance Data Usability and Utility	Personal information not necessary	Adding a creator name may not be useful when publishing provenance data outside of an institution.	1	1	"Whether you'd need the personal information there, I don't know. Whether you can have the personal information because of GDPR as well.", P.1
Theme 3: Enrich Descriptions and Definitions	Clarify data required for justifications	Participants were unsure of the difference between the entity description and interlink justification	3	1	"I suppose sometimes I wondered what was I really supposed to write in here?...I just, I thought well maybe I need to say something like this is the same person because I know that his sound recordings by the same name are also in ITMA. So is it that we have to be that prescriptive or, yeah, so sometimes I was a bit unsure about, even though know you've got, you, you, you tell me what I should write in there. Yeah, I am on sort of unsure, yeah, what, what I should say", P.3
	Add dataset acronyms	Add acronyms to list of datasets.	2	1	"Oh yeah, just the beginning I was going where is VIAF"", P.3
	Clarify data required for entity Descriptions	Participants were unsure of the purpose of the entity description.	2	2	"I was wondering, um, I was wondering, I guess what the protocol was for descriptions", P.2 "I don't know if I was using the description right", P. 3
	Simplify language	Some jargon used descriptions.	2	1	"Some of the text that was about, the about text and stuff. Um, some of it was very jargony and it really, it was very, it's not plain English. And I found some of it difficult to read now to be honest. So, I think that could probably be written, rewritten", P. 1
Theme 4: GUI Requirements	URI validator error	Participants experienced errors with the URI validator where some valid URIs were rejected".	15	3	"There was a couple of times that happened, when we cut and pasted from VIAF and it happened to [name] as well. Um, we just got an error to say, please enter a valid URI", P.1 "But when you sort of copy and pasted your VIAF link in it said, oh, it's not a valid link, but it obviously was, you click out of it, click back. So that was just a small bit annoying", P. 3

Theme 4: GUI Requirements	Streamline process of adding a related entity and an interlink	It was suggested that participants could be guided through adding a related entity using dropdown menus and that it should lead directly to the interlinking process.	10	2	"I think if you were doing it in a more heightened way for the, the, you could do it in steps where it would bring you through as opposed to you choosing it. Like, which one of these do you want? You know, sort of the, this is my entity, this is a person. Do you want to use these? In the same way that you had the other ones you know. Do you want to choose a, you know, these are the ones that have people in them, which one of these do you want to use?... it would suggest datasets.", P.1 "You know, sometimes I sort of forgot that I wasn't finished with the process, then I have to go in and do the interlink", P.3
	Quick access button to add new related entity	Add a button to add a related entity on the main screen.	5	3	"Everyone clicks on the picture, they don't click on the three dots", P.1 "One thing that I did notice when I first logged in was, you know, the add new option, the add new, I clicked on that straight away. And what you're meant to do is, you're meant to click the three dots", P.2
	Function to copy entity label	Add a function that could copy the label of an internal entity or a related entity.	4	2	"Copy the name, copy it to your clipboard. So, then you can just paste it into the browser", P.1 "If you had like a little copy button there and then you just copied straight into the search box here", P. 3
	Indicate that an internal entity has been interlinked	Add an indication that allows users to differentiate between interlinked and non-interlinked internal entities from the list view.	3	2	"So, you know what, like even just a number to say, five links created or six things created so that when you're going through a big, long list and you go, who have I done and who have I not done?", P.1 "I suppose there's nothing here to say you've been in, you've made a link. Maybe if it changed colour or something?", P.3
	Pre-populate the justification field with the related entity description	In order to save time and avoid repetition participants suggested pre-populating the justification field with the related entity description.	3	1	"There's the, the justification was in two different places, but in a way it was the same information. Um, so that'd be another thing if you, if you were just to find it once, that you could carry that justification through and maybe be able to edit it afterwards if needs be.", P.1

Theme 4: GUI Requirements	Pre-populate the related entity label with the internal entity label.	In order to save time and avoid repetition participants suggested pre-populating the related entity label field with the data from the internal entity label field.	3	2	“Oh, another little time saving thing that might be a good idea. Um, again, it's very, very small, but when you're going into your, you're creating your related entity, um, if, if there was an option that you could just copy the label from the internal entity and so you just, there could be an option where you just go, um, copy internal entity label or create a new one because it could be a related entity or it could be identical. So you know, you could just say, oh, copy that and then it automatically pastes into the label field”, P.2 “Can the label be pre-populated the person's name?”, P.3
	Click name of dataset to access	Participants indicated that they often clicked on the dataset name expecting it to take them to the Manage Linkset screen	2	2	“The manage, you know, when you've got the, at the very front, when you're going into it, it says manage dataset. I'm constantly clicking on the name of the dataset to get into it, as opposed to going to the manage button, you know?”, P.1 “So, um, I suppose it's something just very simple but I just thought you should be able to just click there”, P.3
	Login Errors	Need to add an indictor on screen when a user has been automatically logged out of NAISC-L. Errors logging in to NAISC.	2	2	“In just in terms of logging on and getting usernames. It was a little bit of errors and stuff coming up”, P. 1 “I think at one point, it wasn't a challenge necessarily, it just automatically logged me out I think at some point, um, so I got a bit confused cause I was trying to work on, I was working through the dataset and um, it wasn't, um, communicating with me. I was kind of like, oh, what is, this is very strange. Like it wasn't recognising permalinks. and then I looked and um I saw that it had logged me out.”, P.2
	Save related entity button was below the fold	Change the late of the related entity screen so that the save button is more visible	2	1	“But a couple of times I clicked the choose button first because the save is actually under the fold. So, you don't see it on the screen”, P.1
	Function to move between internal entities	Add a function that allows a user to move between one entity and the next without having to return to the main screen	1	1	“Um, also another thing, as you're moving through data, as you're moving through the dataset, um, one thing I noticed, um, which could be useful, um, there's arrow buttons on this side. The arrow button is pointing to the related entities. I thought it also could be handy if the arrow buttons moved between internal entities. So, you click and then it would go to the next internal entity, if that makes sense”, P.2

Theme 4: GUI Requirements	Dataset description pop-ups	A participant noticed that dataset description pop-ups do not disappear automatically if clicked	1	1	"So when you click on the dataset, so say when you have the list of datasets on the right hand side and you click on say VIAF and a popup comes up to show, this is what VIAF is and you go to another dataset and that pops up over the one that's there. And especially, it's especially annoying with VIAF because VIAF is so long. So, you can't really, you can read it, but it's, sometimes you don't even notice that it's there", P.1
Theme 5: Automation	Automatically pull data for related entity	Automatically add related entity data from the dataset	6	2	"Yeah even just be like, you can like just say, you know, pool or a pull, um, information from this specific permalink", P.2 "I don't know if there's any way of sort of being able to identify that's something is a URI and automatically pick it up or something?", P.3
	Automatically search for related entity once dataset selected	Participants stated that it would be useful if an automatic search was conducted rather than having to type in the search terms	4	3	"I don't know if it is possible, but once you have something copied and you pick VIAF that it takes it from your clipboard and puts it in. Um, so that you're, like once you choose the interlink, it goes off and actually does that search for you in a way as opposed to you having to, uh, um, to, to manually put it back in again", P.1 "I thought maybe it would be a good idea is if you click on VIAF and it automatically would search the name through the database for you instead of you having to go in and then manually type out the name again", P.2

5.5.5.3 CSUQ

As mentioned, the CSUQ items are scored from 1 to 7 with lower scores indicating more positive perceptions. For the purpose of this experiment, sufficient usability was considered to be scores lower than 4. The CSUQ scores for each participant and the mean scores for each item can be found in Table 54 below. It can be seen that the mean score for each item, except for Item 9, is below 4 indicating that participants were generally in agreement with the CSUQ items and that sufficient usability was achieved for almost all items. However, Item 9, which reads, *"The system gave error messages that clearly told me how to fix problems",* had a mean score of 5 signifying more negative perceptions of this item. The reason behind this negative perception was distilled from the interview data, analysed in Section 5.6.6.2 above, where participants indicated that there was an intermittent error with the URI validator which they were unable to resolve and found "annoying" (Participant 2).

The mean SysUse, InfoQual and InterQual subscale scores, as well as the mean Overall score, can also be found in Table 54. The mean scores for the SysUse and InterQual were less than 3, indicating mostly positive perceptions of their items. The mean InfoQual and Overall scores were 3.19 and 3.22 respectively, indicating some mixed responses to items. All mean scores were less than 4, signifying that sufficient usability was achieved for the Field Test and suggesting only mild usability and utility issues overall.

Reliability

As stated in Section 5.2.6, Cronbach's alpha indicates the internal consistency of questionnaire items. The alpha index for the CSUQ questionnaire applied in this experiment was 0.95, indicating a high level of internal consistency and reliability in the responses.

Table 54: Field Test – CSUQ Scores

Participant	1. Overall, I am satisfied with how easy it is to use this system	2. It is simple to use this system	3. I can effectively complete my work using this system	4. I am able to complete my work quickly using this system	5. I am able to efficiently complete my work using this system	6. I feel comfortable using this system	7. It is easy to learn to use this system	8. I believe I became productive quickly using this system	9. The system gives error messages that clearly tell me how to fix problems	10. Whenever I make a mistake using the system, I recover easily and quickly	11. The information (such as on-line help, on-screen messages, and other documentation) provided with this system is clear
1	2	3	3	4	3	3	2	2	5	4	5
2	1	2	2	3	3	1	2	2	5	2	3
3	2	2	2	2	2	2	2	1	5	4	3
Mean	1.67	2.33	2.33	3.00	2.67	2.00	2.00	1.67	5.00	3.33	3.67
Standard Deviation	0.47	0.47	0.47	0.82	0.47	0.82	0.00	0.47	0.00	0.94	0.94

Participant	12. It is easy to find the information I need	13. The information provided for the system is easy to understand	14. The information is effective in helping me complete my work	15. The organisation of information on the system screens is clear	16. The interface of this system is pleasant	17. I like using the interface of this system	18. This system has all the functions and capabilities I expect it to have	19. Overall, I am satisfied with this system	SysUse (1-8)	InfoQual (9-15)	InterQual (16-18)	Overall (1-19)
1	4	4	3	4	3	2	4	2	2.75	4.14	3.00	4.13
2	2	3	1	1	1	1	1	1	2.00	2.43	1.00	2.47
3	3	2	2	2	2	2	4	2	1.88	3.00	2.67	3.07
Mean	3.00	3.00	2.00	2.33	2.00	1.67	3.00	1.67	2.21	3.19	2.22	3.22
Standard Deviation	0.82	0.82	0.82	1.25	0.82	0.47	1.41	0.47	0.39	0.71	0.87	0.69

5.5.6 Discussion

The hypothesis (**H3.1**) being investigated as part of the Field Test was whether 'using the NAISC-L Framework, in a LAM context, to create LD interlinks from an institution's dataset yields high accuracy with sufficient usability for IPs'. Here 'high accuracy' was considered to be interlinks with over 75% accuracy, and 'sufficient usability' to be CSUQ scores strictly lower than a neutral score of 4 (on a scale from 1 to 7).

In the Field Test, the mean accuracy score for all participants was above 75%. This indicates high accuracy for the creation of interlinks using the NAISC-L framework. The mean SysUse, InfoQual, InterQual subscale scores and the mean Overall score, were all lower than 4 indicating that sufficient usability was achieved for IPs when using NAISC-L.

Overall, the experiment indicated that IPs, in a LAM context, can use NAISC-L for the creation of LD interlinks with high accuracy and sufficient usability – confirming the hypothesis (**H3.1**) of this experiment.

The research question of this thesis focuses on facilitating IPs to engage with the process of Linked Data interlinking with effectiveness, efficiency and satisfaction. Effectiveness is considered to be the degree to which users can accurately complete LD interlinks. Efficiency refers to the time taken to create an interlink, and satisfaction is the extent to which NAISC-L meets a user's needs and expectations. Table 55 below specifically outlines how effectiveness, efficiency and satisfaction (EES) were measured in this experiment.

Table 55: Field Test EES Measures

Area	Measure	Results
Effectiveness	Semantic Accuracy of interlinks	All interlinks created were 100% accurate as the `owl:sameAs` property was used to link identical persons across datasets.
	CSUQ: SysUse	The SysUse portion of the CSUQ includes items that address system efficacy. The mean SysUse score was 2.21, indicating that participants had mostly positive perceptions of the items.
Efficiency	CSUQ: SysUse	The SysUse portion of the CSUQ includes items that address system efficiency. The mean SysUse score was 2.21, indicating that participants had mostly positive perceptions of the items.
Satisfaction	CSUQ: InterQual & Overall	The InterQual portion of the CSUQ investigates whether a system met the expectations of a user. The mean InterQual score was 2.22 indicating that participants had mostly positive perceptions of these items. The Overall CSUQ score includes items which specifically measure user satisfaction. The mean Overall score was less than 4, indicating that users had a mostly positive experience of NAISC-L usability and utility.
	Thematic Analysis	A number of codes, indicate that the participants were satisfied with NAISC-L, emerged from the data These codes form part of Theme 1, which relates to the usability and utility of NAISC-L, and include: • Easy and user-friendly • Useful • Positive experience • Descriptions and guides were useful • Straightforward interlinking process • Better at using over time • Good GUI layout • Liked external dataset list and dataset ranking • Liked the visualisations

Overall, participants had a positive reaction to the flow of the NAISC-L framework and stated that it was useful and user-friendly. This suggests that NAISC-L is both effective and satisfactory. The new requirements distilled from the data were primarily suggestions for automating certain steps and adding extra functions to the GUI in order to make the interlinking process more time efficient.

5.5.7 Experiment Summary

Section 5.5 presented the results of a Field Test of NAISC-L in a music archive environment. The IPs were able to use NAISC-L as part of their cataloguing workflow and, even though the IPs had little to no prior experience with LD, all were able to successfully create interlinks. High interlink accuracy and sufficient usability were achieved overall. The data also indicated that participants found NAISC-L to be effective and satisfactory. Regarding efficiency, although no negative feedback was noted, the new user requirements were primarily suggestions for speeding up the interlinking process. A discussion of the results of Field Test in relation to the results of Usability Test 1 and Usability Test 2 can be found in Section 5.6.

5.6 NAISC-L Evaluation Conclusion

The research question of this thesis is, *'To what extent can NAISC-L, a domain-specific interlinking framework, facilitate IPs to engage with the process of LD interlinking with effectiveness, efficiency and satisfaction?'*.

The following section will discuss NAISC-L effectiveness, efficiency and user satisfaction across all three experiments. Note for Usability Test 2, unless otherwise stated, the version of NAISC-L being considered is Version A which included all components of the NAISC-L Framework.

5.6.1 NAISC-L Effectiveness

The effectiveness of NAISC-L was measured in terms of the mean number of interlinks completed and the semantic accuracy of the interlinks. It also took into account the mean SysUse PSSUQ score. Across all experiments NAISC-L was found to have high completeness and high accuracy. In Usability Test 1, participants accurately completed an average of 5 out of 6 interlinks. Similarly, in Usability Test 2, participants accurately completed 2 out of 3 interlinks. Participants achieved 100% accuracy in Usability Test 3. In addition, the results of Experiment 2 indicated that prior LD knowledge did not significantly affect accuracy scores, suggesting that non-expert LD users were able to use NAISC-L with the same level of success as expert LD users.

The mean SysUse scores were below 4 across all three experiments indicating that participants had mostly positive perceptions towards PSSUQ/CSUQ items which measured system effectiveness. Specifically, the average SysUse scores for Usability Test 1, Usability Test 2 (NAISC-L Version A) and Usability Test 3 were 2.48, 3.05 and 2.21 respectively. These scores suggest only mild issues in terms of NAISC-L effectiveness.

Overall these results indicate that IPs, with varying levels of LD knowledge, could effectively use the NAISC-L interlinking process and tool to create LD interlinks.

5.6.2 NAISC-L Efficiency

The SysUse portion of the PSSUQ/CSUQ was used to measure efficiency across all three experiments and the time taken to complete a set of interlinks was used to measure efficiency in Usability Test 1 and Usability Test 2.

The mean SysUse scores were below 4 across all three experiments indicating mostly positive perceptions of PSSUQ/CSUQ items which measured system efficiency. Although the SysUse scores suggest that participants did not experience significant issues in terms of efficiency, the Thematic Analysis of Usability Test 1 and Usability Test 3 indicates that participants found the process to be time-consuming. Participants suggested automating certain functions, such as automatically adding internal and related entity data, automatically searching for related entities, and automatically selecting link-types, as a means of saving time. While increased automation would indeed likely improve efficiency, it would be important to ensure a balance between automatic and manual processes as increased automation has the potential to detract from the contextually rich interlinks created manually by domain-expert IPs.

In Usability Test 1 the average time to create an interlink was 3.27 minutes and in Usability Test 2 the average time taken to create an interlink was 5.45 minutes. Users had to create six interlinks in Usability Test 1 and three interlinks in Usability Test 2. It is possible that the average time per interlink was lower in Usability Test 1 as participants had the opportunity to create more links, thus gaining more experience using the tool, which may have led to their interlinking speed increasing over time – a factor which was mentioned in the Thematic Analysis of the experiment.

Overall, the results indicate that IPs could efficiently use the NAISC-L to create LD interlinks.

5.6.3 NAISC-L Satisfaction

Across all three experiments the InterQual and Overall PSSUQ/CSUQ scores were used the measure satisfaction. Data from the Thematic Analysis was used to measure satisfaction in Usability Test 1 and Usability Test 3, and DQ questionnaire scores were used to measure satisfaction in Usability Test 2.

The mean InterQual and Overall scores were less than four across all three experiments indicating mostly positive perceptions of PSSUQ/CSUQ items which measured system satisfaction. Specifically, the average Overall scores for Usability Test 1, Usability Test 2 (NAISC-L Version A) and Usability Test 3 were 2.42, 2.98 and 3.22 respectively. These scores suggest only mild issues in terms of NAISC-L usability and utility, indicating that participants had a satisfying user experience overall.

In Usability Test 1 and Usability Test 3, the Thematic Analysis revealed that participants considered the NAISC-L interlinking process to be useful, straightforward and intuitive. The Thematic Analysis also revealed that participants found the NAISC-L tool to be user-friendly, clear and suitable for non-expert LD users.

In Usability Test 2, the average overall data quality score (NAISC-L Version A) was greater than 5 indicating mostly positive perceptions of quality NAISC-L data. This suggests that participants were satisfied with the quality of NAISC-L's output.

Overall, the results indicate that IPs found that NAISC-L satisfactorily met their requirements and expectations.

5.7 Chapter Summary

This chapter presented and discussed the results of three experiments used to evaluate NAISC-L. The first experiment consisted of a Think-Aloud Test which was used to evaluate the first iteration of NAISC-L. The second experiment consisted of an Online Interlink Creation Test which was used to evaluate the second iteration of NAISC-L. Finally, the third experiment was a Field Test which was also used to evaluate the second iteration of NAISC-L but in the context of a LAM, in this case a music archive. The results of these experiments indicated that NAISC-L can be used by IPs to successfully engage with the process of LD interlinking with sufficient effectiveness, efficiency and satisfaction.

6 Conclusion

This chapter draws conclusions from the research presented throughout this thesis. Section 6.1 discusses the extent to which the research objectives of this thesis, outlined in Chapter 1, have been achieved. The contributions of the research are revisited in Section 6.2. Section 6.3 presents potential future work that could be undertaken based on the findings of this thesis. Final remarks are presented in Section 6.4.

6.1 Research Objectives

The extent to which the objectives, posed to address the research question of this thesis, were achieved is analysed in this section.

The research question of this thesis was:

To what extent can NAISC-L, a domain-specific interlinking framework, facilitate Information Professionals to engage with the process of Linked Data interlinking with effectiveness, efficiency and satisfaction?

6.1.1 RO1

The first research objective was to *perform a state-of-the-art review of existing LD interlinking frameworks and tools*. This objective was achieved through the analysis of existing LD frameworks and LAM LD services conducted in Chapter 3. The review firstly evaluated current LD Interlinking Frameworks and Tooling. The characteristics examined included the link-types supported by the framework, whether the framework was designed for the LAM domain, whether the framework provides provenance data, and whether usability testing was conducted. The review indicated that frameworks primarily support the creation of identity links with the exception of two tools, SILK and LIMES, which allow users to specify their preferred link-type. No framework was designed specifically for the LAM domain, however, OpenRefine does have extensions that are library specific. None of the frameworks generate interlink provenance data and none have published usability testing research.

Also analysed as part of the State-of-the-Art Review were LAM LD services. The characteristics examined included whether interlinking was conducted, the datasets linked to, and whether an LD provenance was generated. The LD services reviewed were interlinked to an average of five other datasets, with the majority of these being controlled vocabularies and authority files. The types of interlinks created were primarily `owl:sameAs` statements and only two services appeared to provide provenance data.

The State-of-the-Art Review indicated a need for a framework that facilitates the creation of relationship links and that provides provenance data. A set of requirements for such an interlinking framework were identified from the review and presented in Chapter 4.

6.1.2 RO2

The second research objective was to *explore the benefits and challenges of using LD as experienced by IPs.* This was achieved through the execution of a LD Requirements Questionnaire which was completed by 185 IPs from a variety of backgrounds. The findings of the survey indicated that the primary benefits of LD, as perceived by IPs, included improved data discoverability and accessibility, enriched metadata and exposing data to a larger audience. With regard to the challenges of using LD, the primary barriers reported by participants were difficulty interlinking and integrating data, technologically complex tooling inadequate for LAMs, and resource quality issues. Like the State-of-the-Art Review, the results of the questionnaire were used in formulating a set of requirements for an interlinking tool for LAMs.

6.1.3 RO3

The third research objective was to *propose a LD interlinking framework for the LAM domain.* This research object was achieved by proposing the **N**ovel **A**uthoritative **I**nterlinking for **S**emantic Web **C**ataloguing in **L**ibraries (NAISC-L) Framework, described in Chapter 4. NAISC-L supports the user requirements distilled from the State-of-the-Art Review and the LD Requirements Questionnaire by guiding users through the creation of identity links and relationship links. It was also designed with the LAM domain in mind and provides quality scores and links to commonly used LAM data sources. NAISC-L also supports link-types from frequently used LAM ontologies. Finally, the framework generates detailed provenance data for the interlinks created. The interlink and provenance data is presented via interactive visualisations so as to aid user comprehension.

6.1.4 RO4

The fourth research objective was to *apply, implement and evaluate the interlinking framework in terms of its effectiveness, efficiency and satisfaction as perceived by IPs*. This objective was achieved by applying the NAISC-L interlinking process to a GUI which allowed for IPs to interact with the framework. NAISC-L was evaluated by IPs through a series of user experiments, including a Think-Aloud Test, an Online Usability Test and a Field Test, the results of which are discussed in Chapter 5. Across all experiments, users achieved a high level of accuracy when linking entities suggesting that NAISC-L is effective for the creation of interlinks. Similarly, the mean PSSUQ/CSUQ SysUse scores were less than four across all experiments, suggesting mostly positive response to statements measuring system efficiency. Users reported being able to create interlinks efficiently, however, automating certain functions, such as adding entity data and selecting link-types, was suggested as a means of saving time. Finally, the mean InterQual and Overall PSSUQ/CSUQ scores were less than 4 across all experiments indicating mostly positive responses to statements that measure user satisfaction. Users also reported that they found the NAISC-L interlinking process to be useful, straightforward and intuitive, and the GUI to be user friendly, clear and suitable for non-expert LD users. These results suggest that NAISC-L satisfactorily met the requirements and expectations of IPs.

6.2 Contributions

This section briefly revisits the contributions of this thesis, which were initially presented in Chapter 1.

The major contribution of this thesis is the development and demonstration of the interlinking framework, NAISC-L. NAISC-L advances the state-of-the-art by presenting an interlinking framework, that facilitates the creation of relationship links and identity links, which is accessible via a GUI that was designed to reduce LD technicalities and to support the needs of IPs. The experiments used to evaluate NAISC-L considered the framework's effectiveness, efficiency and user satisfaction, and the results indicated that these three measures were achieved. These user-experiments also advance the state-of-the-art as they are the first to evaluate the usability of LD interlinking tooling for LAMs. Table 56 overleaf demonstrates how NAISC-L compares to the LD interlinking tools discussed in Section 3. It can be seen that NAISC-L advances the capabilities of existing frameworks by facilitating the creation of relationship links, integrating multiple LAM datasets, and generating interlink provenance.

It is envisaged that the NAISC-L Framework will have an impact on the adoption of LD in LAMs by facilitating IPs to create LD interlinks with greater ease and efficacy than existing LD tooling allows. The full potential of LD interlinking has yet to be realised within the LAM domain due to a notable lack of interlinks created for purposes outside of authority control. NAISC-L, however, facilitates the creation of relationship links to authoritative data sources. LAM metadata that has been enriched with relationship links would improve data discovery and allow users to navigate seamlessly between internal and external datasets.

As NAISC-L was specifically designed for LAMs and tested by IPs, it could be used as a model for the development of other LD tooling for LAMs. NAISC-L could also be easily adapted for use in other domains. It has been shown that the NAISC-L Interlinking Process and GUI can be successfully and easily used by those who are not LD experts or computer scientists. By replacing LAM specific content, such as controlled vocabularies, authority files, with domain-specific content from other areas, it is likely that NAISC-L could be used effectively in other fields.

Table 56: NAISC-L vs Existing Linked Data Interlinking Tools

System		AgreementMaker	LogMap	Linkitup	SILK	LIMES	OpenRefine	NAISC-L
Data Input		SPARQL API	RDF	Figshare.com Metadata	RDF SPARQL CSV XML	RDF SPARQL CSV XML	RDF SPARQL CSV XML JSON	RDF Manual Text Entry
Supported Link-Types	**Supports Identity Links**	✓	✓	✓	✓	✓	✓	✓
	Identity Link-Types	owl:sameAs skos:exactMatch	owl:sameAs	owl:sameAs skos:exactMatch	owl:sameAs User-declared*	owl:sameAs User-declared*	owl:sameAs	owl:sameAs skos:exactMatch ov:commonItem ov:commonManifestation ov:commonExpression ov:commonWork ov:commonEndeavour madsrdf:hasExactExternalAuthority frad:P2019 frad:P2037 User-declared*
	Suports Relationship Links	×	×	×	✓	✓	×	✓
	Relationship Link-Types	–	–	–	User-declared*	User-declared*	–	madsrdf:hasCloseExternalAuthority madsrdf:hasRelatedAuthority skos:closeMatch skos:relatedMatch skos:broadMatch skos:narrowMatch skos:relatedMatch schema:sameAs schema:isSimilarTo schema:isRelatedTo schema:relatedLink bf:relatedTo dcterms:relation edm:isSimilarTo edm:isRelatedTo frbr:relatedEndeavour modsrdf:relatedItem ov:similarTo ov:associatedEntity crm:P69_is_associated_with frad:P2028 rdfs:seeAlso User-declared*
Link Generation		Semi-automatic	Semi-automatic	Semi-automatic	Semi-automatic	Semi-automatic	Semi-automatic	Manual
Integrated Datasets		None	None	DBpedia DBLP CrossRef ORCID NIF Registry DANS Easy	None	None	Wikidata DBpedia VIAF FAST ORCID LCSH LCNAF Getty + Not an exhaustive list	AAT BNB BNF BNE DBpedia DNB Europeana FAST Food & Agriculture GeoNames LCNAF LCSH TGM TGN ULAN VIAF + Not an exhaustive list
GUI		GUI	Web Interface	Web Dashboard	Web Workbench	Web Interface	Web Interface	Web Interface
Domain		Unspecified	Unspecified	Academic Research	Unspecified	Unspecified	Libraries Biodiversity Research Other	Libraries Archives Museums
Published User Testing		No	No	No	No	No	No	Yes
Interlink Provenance		No	No	No	No	No	No	Yes

* Users can define and employ any link-type

The first minor contribution of this thesis is the provision of a report on the current state of LD in the LAM domain. This report is based a review of LD services in LAMs, as well as the results of the LD Requirements Questionnaire. The questionnaire highlighted a number of challenges that are being experienced by IPs when using LD. These results could be used to inspire future research exploring potential solutions to the other LD barriers being experienced by LAMs.

The second minor contribution of thesis is NaiscProv – an extension of PROV-O designed for describing the provenance of an interlink. NaiscProv was developed in response to a specific problem related to data trustworthiness, identified in the State-of-the-Art Review and the LD Requirements Questionnaire.

LAMs are typically well-established and trusted sources of information, as such, LD generated by IPs is likely to be treated with increased credibility over data generated by non-authoritative sources. However, current LD tooling and LAM LD services have a notable lack of provenance data. NAISC-L generates provenance data for the interlinks it creates, providing users with a knowledge of the data's origin, allowing them to make informed decisions regarding its quality and authoritativeness. This would likely lead to an increased use of LD published by LAMs, including an increase in the number of external data sources pointing to the data. NaiscProv could be also used as a model for future PROV-O extensions which capture the provenance of other LD activities.

6.2.1 Uptake

As stated in Chapter 1, this research is already having impact within the research community with publications in well-known venues such as the 2019 and 2018 ACM/IEEE on Joint Conference on Digital Libraries, the 2019 World Wide Web Conference and the Extended Semantic Web Conference 2017. NAISC-L was also presented at the Semantic Web in Libraries Conference 2018 and 2019, and the LIBER Conference 2019.

Invitations were extended from Information Today Europe[138], EuropeanaTech Insight[139], and eLucidate[140], the online journal of the UK eInformation Group (UKeiG), to provide an article on NAISC-L for their respective publications. An invitation was also extended from UKeiG to present on NAISC-L at the group's annual general meeting. Finally, a Library Software Company has expressed an interest in incorporating NAISC-L into a commercialised Library Management System.

[138] https://www.infotoday.eu/ accessed 7th August 2020
[139] https://pro.europeana.eu/page/europeanatech-insight accessed 7th August 2020
[140] https://elucidate-ukeig.org.uk/index.php/elucidate accessed 7th August 2020

6.3 Future Work

NAISC-L Interlinking Process

The results of the Linked Data Requirements Questionnaire, see Section 3.3, indicated that IPs would also find it useful to create vocabulary links using NAISC-L. This could be achieved by extending the interlinking process to include ontology mapping. One possible method of achieving this would be to fully integrate the Simple Knowledge Organization System (SKOS) data model (Miles & Bechhofer, 2009) with NAISC-L. SKOS can be used in order to define, develop and link ontologies on the Semantic Web. Using SKOS as part of NAISC-L would allow users to create links between concepts across different vocabularies. This process could be supported by developing a Vocabulary Linking Guide similar to the Interlinking Guide described in Section 4.

NAISC-L Graphical User Interface

During Usability Test 1 (see Section 5.3) and Usability Test 3 (see Section 5.5) participants suggested automating certain features of the NAISC-L GUI in order to improve interlinking efficiency. One such feature is the population of Internal and Related Entity data fields. At present, in order to add an entity, users need to open the webpage of their chosen dataset, search for an entity, and then copy and paste the entity data into the appropriate text fields in the NAISC-L GUI. In order to make this process more efficient, rather than accessing datasets from separate tabs or windows, users will be able to conduct a keyword search of individual integrated datasets directly in the NAISC-L GUI. Once the user selects an entity, its data will automatically populate the appropriate text fields. This will reduce the time taken to search for and add entities to NAISC-L.

Another GUI feature that could be automated is the External Dataset selection process. Currently, when adding a Related Entity, users are presented with a list of over sixteen External Datasets that they can search through. This list of datasets can be filtered according to their topic and function in order to aid users in selecting an appropriate dataset. However, this decision-making process could be made more efficient by suggesting a smaller selection of datasets to user. The list of potential datasets will be refined based on the internal entity's FRBR Model definition (see Section 4.2.1). This will be achieved by suggesting datasets that contain similar types of entities, for example, FRBR Responsible

Entities are commonly linked with name authority datasets such as the Library of Congress Named Authority File[141] (LCNAF), as such only name authorities would be suggested in this case. The list of suggested datasets will then be further refined according to their quality rating, with only the top-rated datasets being presented to the user. However, users will still be able to explore the full list of integrated datasets if required. Finally, once a user selects an External Dataset, the label of the Internal Entity will be used to automatically search the selected dataset for a Related Entity. Overall, these changes will make the process of searching for and adding a Related Entity more efficient.

The final feature that would benefit from increased automation is the Interlinking Guide (see Section 4.2.1), which supports users in selecting a link-type. Currently, users are presented with a list of link-types based on the selected Relationship Term describing the association between an Internal and External Entity. The number of suggested link-types will be simplified using the FRBR definitions of both the Internal and External Entities, as certain link-types can only be used with specific FRBR forms. The list will be further refined according to the ontologies used in the Internal and External Datasets, as certain link-types are only compatible with specific ontologies. Overall, this will decrease the number of link-types presented to the user, making it easier and more efficient to interlink entities, while also reducing the potential for errors.

NAISC-L Provenance Model

As discussed in Section 2.2.2.4, RDF* is an approach that can be used to create meta-triples describing the provenance of an RDF statement. Despite the fact that there are many benefits to using RDF*, when compared to RDF Reification, the approach was not incorporated into the NAISC-L Provenance Model because, at the time of design, RDF* and its extensions were not supported by many LD technologies. However, since then, LD technologies are increasingly providing support for RDF* (Hartig, 2019). As such, incorporating RDF* into the NAISC-L Provenance Model was explored.

[141] http://id.loc.gov/authorities/names retrieved 12th November 2020

One benefit of using RDF* to generate LD provenance is that it does not add any additional triples to the graph – unlike RDF Reification which adds four triples per statement. This is due to the use of nested triples – described in Section 2.2.2.4. Another benefit is the SPARQL* extension which allows users to query meta-triples in a more simplified manner using fewer queries.

Figure 55 displays how RDF* could be used to replace RDF Reification in the NAISC-L Provenance Model. It can be seen that there are fewer triples in the graph, when compared to Figure 19 in Section 4.2.2.2. Note that the interlink justification, generation date/time and the creator of the interlink are directly linked to the nested triple. This differs to the current provenance model whereby this data is linked to the interlink URI that is generated as part of the RDF Reification process. As such, in order to retrieve the provenance of an interlink, the user must first query the reified statement to retrieve its URI and, following this, query its provenance. However, using RDF* and SPARQL*, the user can find this information directly using the nested triple.

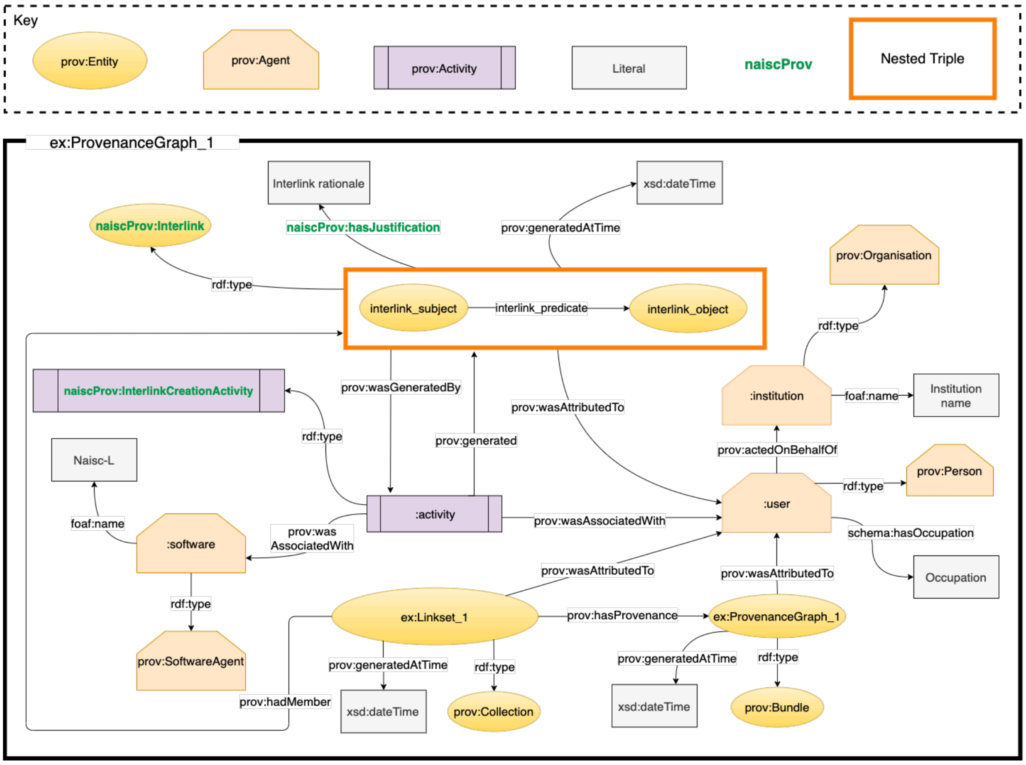


Figure 55: Using RDF in the NAISC-L Provenance Model*

Figure 56 demonstrates how a nested triple is portrayed in RDF Turtle syntax using the Turtle* extension which captures the concept of a nested triple by enclosing it using the strings '<<' and '>>'.

Figure 56: Turtle Output*

```
<<:The_Dead_BNF_URI rdfs:seeAlso :Joyces_Dublin_UCD_URI>>
      naiscProv:hasJustification "Joyce resources related to The Dead" .
      prov:wasAttributedTo :Librarian#1
      prov:generatedAtTime "2020-08-12 13:28"
```

Figure 57 demonstrates a SPARQL* Query of the data in the Figure 56. As can be seen, this query is shorter and more simplified when compared to a similar query using RDF Reification as presented in Figure 43.

Figure 57: SPARQL Query*

```
SELECT ?justification WHERE (
<< :The_Dead_BNF_URI rdfs:seeAlso :Joyces_Dublin_UCD_URI >>
   naiscProv:hasJustification ?justification
)
```

Overall, RDF* provides a more efficient approach to generating and querying LD provenance. RDF* could be applied to future iterations of NAISC-L to improve its efficiency.

6.4 Final Remarks

It is hoped that NAISC-L can be of benefit to LAMs that have already established a LD service. By integrating the framework into their LD creation process, NAISC-L can facilitate the creation of an increased variety of links to a larger number of external data sources.

It is also hoped that NAISC-L would benefit the research community. Researchers can employ NAISC-L in their interlinking processes, use the findings presented in this thesis in their research, and apply their expertise to contribute to the framework and its implementations.

References

Alemu, G., Stevens, B., Ross, P. and Chandler, J. (2012). Linked Data for libraries: Benefits of a conceptual shift from library-specific record structures to RDF-based data models, *New Library World, 113(11/12),* 549-570.

Alexander, K., Cyganiak, R., Hausenblas, M. & Zhao, J. (2011, March 03). *Describing linked datasets with the VoID vocabulary*. Retrieved from http://www.w3.org/TR/void/.

Andrews, C., Burleson, D., Dunks, K., Elmore, K., Lambert, C. S., Oppegaard, B., Pohlan, E. E., Saad, D., Scharer, J. S., Wery, R. L., Wesley, M. & Zobel, G. (2012). A new method in user-centered design: Collaborative prototype design process (CPDP). *Journal of Technical Writing and Communication, 42(2)*, 123-142.

Arenas, M., Grau, B. C., Kharlamov, E., Marciuška, Š., & Zheleznyakov, D. (2016). Faceted search over RDF-based knowledge graphs. *Journal of Web Semantics, 37*, 55-74.

Arioua A., Croitoru M., Papaleo L., Pernelle N., Rocher S. (2016). On the explanation of sameAs statements using argumentation. In: Schockaert S., Senellart P. (eds) Scalable Uncertainty Management. SUM 2016. *Lecture Notes in Computer Science, vol 9858*, 51-66. Springer, Cham.

Ashkpour, A, Meronõ-Peñuela, A, & Mandemakers, C.A. (2015). The aggregate Dutch historical censuses. *Historical Methods, 48*, 230–245.

Becker, D. A., & Yannotta, L. (2013). Modeling a library web site redesign process: Developing a user-centered web site through usability testing. *Information Technology and Libraries, 32(1)*, 6-22.

Belhajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S., & Zhao, J. (2013, April 30). *PROV-O: The PROV ontology*. Retrieved from https://www.w3.org/TR/prov-o/.

Bensman, F., Prongu, N., Hellstern, M., & Kuntschik, P. (2016). *Swissbib goes linked data*. Paper presented at Semantic Web in Libraries (SWIB) 2016.

Berners-Lee, T. (2006, July 27). *Linked data*. Retrieved from www.w3.org/DesignIssues/LinkedData.html.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American, 284(5)*, 34-43.

Bezerra, C., Freitas, F., & Santana, F. (2013). Evaluating ontologies with competency questions. *IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) 2013,* 284-285.

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data: The story so far. *International Journal on Semantic Web and Information Systems (IJSWIS) 2009, 5(3),* 1-22.

Bizer, C., Volz, J., Kobilarov, G., & Gaedke, M. (2009). Silk – a link discovery framework for the web of data. In *CEUR Workshop Proceedings, 538*, 1-6.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3(2)*, 77-101.

Brickley, D. & Guha, R.V. (2014, February 25). *RDF Schema 1.1*. Retrieved from https://www.w3.org/TR/rdf-schema/.

Brickley, D. & Miller, L. (2014, January 14). *FOAF Vocabulary Specification 0.99*. Retrieved from http://xmlns.com/foaf/spec/.

Brooke, J. (1996). SUS: a 'quick and dirty' usability scale. In P.W.Jordan, B. Thomas, B.A. Weerdmeester, & I.L. McClelland (Eds*.) Usability Evaluation in Industry* (pp. 189-194). London: Taylor and Francis.

Carroll, J. J., Bizer, C., Hayes, P., & Stickler, P. (2005). Named graphs. *Journal of Web Semantics, 3(4)*, 247-267.

Carroll, J. J. & Stickler, P. (2004, May 13). *TriX: RDF Triples in XML*. Retrieved from https://www.hpl.hp.com/techreports/2004/HPL-2004-56.pdf.

Cho, E., & Kim, S. (2015). Cronbach's coefficient alpha: Well known but poorly understood. *Organizational Research Methods, 18(2),* 207-230.

Ciccarese, P., Soiland-Reyes, S., Belhajjame, K., Gray, A. J., Goble, C. & Clark, T. (2013). PAV ontology: provenance, authoring and versioning. *Journal of Biomedical Semantics, 4(1)*, 37.

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.

Consultative Committee for Space Data Systems (CCSDS) (2019, September). *Reference model for an open archival information system (OAIS)*. Retrieved from http://www.oais.info/.

Coyle, K. (2013). Library linked data: An evolution. *Italian Journal of Library & Information Science, 4(1),* 53.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16(3)*, 297-334.

Crotti Junior, A., Debruyne, C., & O'Sullivan, D. (2017). Juma: An editor that uses a block metaphor to facilitate the creation and editing of R2RML mappings. In *European Semantic Web Conference (ESWC)* (pp. 87-92). Springer, Cham.

Crotti Junior, A., Debruyne, C., & O'Sullivan, D. (2018). An editor that uses a block metaphor for representing semantic mappings in linked data. In *European Semantic Web Conference (ESWC)* (pp. 28-33). Springer, Cham.

Cruz, I. F., Antonelli, F. P., & Stroe, C. (2009). AgreementMaker: efficient matching for large real-world schemas and ontologies. *Proceedings of the VLDB Endowment, 2(2),* 1586-1589.

Cruz, I. F., Stroe, C., Caimi, F., Fabiani, A., Pesquita, C., Couto, F. M., & Palmonari, M. (2011). Using AgreementMaker to align ontologies for OAEI 2011. In *ISWC international workshop on ontology matching (OM), 814*, 114-121.

Das, S., Sundara, S. & Cyganiak, R. (2012, September 27). R2RML: RDB to RDF Mapping Langauge. Retrieved from https://www.w3.org/TR/r2rml/.

De Boer, V., Wielemaker, J., Van Gent, J., Hildebrand, M., Isaac, A., Van Ossenbruggen, J., & Schreiber, G. (2012, May). Supporting linked data production for cultural heritage institutes: the Amsterdam museum case study. In *Extended Semantic Web Conference (ESWC)* (pp. 733-747). Springer, Berlin, Heidelberg.

De Melo, G. (2013). Not quite the same: Identity constraints for the web of linked data. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*.

Debattista, J., McKenna, L., & Brennan, R. (2018). Understanding information professionals: A survey on the quality of linked data sources for digital libraries. In *2018 Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE)*.

Debruyne, C., Walshe, B., & O'Sullivan, D. (2015, December). Towards a project centric metadata model and lifecycle for ontology mapping governance. In *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services* (pp. 1-10).

Deliot, C. (2014). Publishing the British National Bibliography as linked open data. *Catalogue & Index, 174*, 13-18.

Deliot, C., Wilson, N., Costabello, L., & Vandenbussche, P. Y. (2017, November). The British National Bibliography: who uses our linked data?. In *International Conference on Dublin Core and Metadata Applications* (pp. 24-33).

Denaux R., Ren Y., Villazon-Terrazas B., Alexopoulos P., Faraotti A., Wu H. (2017). Knowledge architecture for organisations. In: Pan J., Vetere G., Gomez-Perez J., Wu H. (eds) *Exploiting Linked Data and Knowledge Graphs in Large Organisations*. Springer, Cham.

Dezani-Ciancaglini, M., Horne, R., & Sassone, V. (2012). Tracing where and who provenance in linked data: A calculus. *Theoretical Computer Science, 464*, 113-129.

Ding, L., Shinavier, J., Finin, T., & McGuinness, D. L. (2010). owl: sameAs and Linked Data: An empirical study. In *Proceedings of the Second Web Science Conference.*

Euzenat, J., & Shvaiko, P. (2013). *Ontology matching* (2nd ed.). Springer, Berlin, Heidelberg.

Farrell, S. (2016, October 23). *Field Studies*. Retrieved from Nielsen Norman Group, https://www.nngroup.com/articles/field-studies/.

Faulkner, L. (2003). Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behaviour Research Methods, Instruments, & Computers, 35(3)*, 379-383.

Falconer, S. M., & Storey, M. A. (2007). A cognitive support framework for ontology mapping. In *The Semantic Web* (pp. 114-127). Springer, Berlin, Heidelberg.

Ferrara, A., Nikolov, A., & Scharffe, F. (2011). Data linking for the semantic web. *International Journal on Semantic Web and Information Systems (IJSWIS), 7(3),* 46-76.

Fisher, R. (1919). XV.—The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh, 52*(2), 399-433.

Fons, T. (2016). Making Libraries Visible on the Web. *Library Journal, 141*(13), 44-46.

Goldkuhl, G., Ågerfalk, P., & Sjöström, J. (2017). A design science approach to information systems education. In *International Conference on Design Science Research in Information System and Technology* (pp. 383-397). Springer, Cham.

Gonzales, B. M. (2014). Linking Libraries to the Web: Linked Data and the Future of the Bibliographic Record. *Information Technology & Libraries, 33*(4), 10-22.

Groth, P., Gil, Y., Cheney, J., & Miles, S. (2012). Requirements for provenance on the web. *International Journal of Digital Curation, 7*(1), 39-56.

Grüninger, M., & Fox, M. S. (1995). Methodology for the design and evaluation of ontologies. In *Workshop on Basic Ontological Issues in Knowledge Sharing*. International Joint Conference on Artificial Intelligence.

Guerrini, M., & Possemato, T. (2016). From Record Management to Data Management: RDA and New Application Models BIBFRAME, RIMMF, and OliSuite/WeCat. *Cataloging & Classification Quarterly, 54*(3), 179-199.

Hallo, M., Luján Mora, S., & Trujillo, J. C. (2014). Transforming library catalogs into Linked Data. *Paper presented at the Proceedings of the 7th International Conference of Education, Research and Innovation.*

Halpin, H., Hayes, P. J., McCusker, J. P., McGuinness, D. L., & Thompson, H. S. (2010). *When owl: sameas isn't the same: An analysis of identity in linked data.* In International Semantic Web Conference (pp. 305-320). Springer, Berlin, Heidelberg.

Hannemann J. & Kett J. (2010). Linked data for libraries. In *Proceedings of the World Library and Information Congress, 76th IFLA General Conference and Assembly* (pp. 1–11). Gothenburg.

Hartig, O. (2017). RDF* and SPARQL*: An alternative approach to annotate statements in RDF. In *International Semantic Web Conference* (Posters, Demos & Industry Tracks).

Hartig, O. (2019). *Position Statement: The RDF* and SPARQL* Approach to Annotate Statements in RDF and to Reconcile RDF and Property Graphs*. Retrieved from blog.liu.se/olafhartig/2019/01/10/position-statement-rdf-star-and-sparql-star/.

Hartig, O., & Thompson, B. (2014). Foundations of an Alternative Approach to Reification in RDF. *Retrieved from https://arxiv.org/abs/1406.3399.*

Hartig, O., & Zhao, J. (2010). Publishing and consuming provenance metadata on the web of linked data. In *International Provenance and Annotation Workshop* (pp. 78-90). Springer, Berlin, Heidelberg.

Hastings, R. (2015). Linked Data in Libraries: Status and Future Direction. *Computers in Libraries, 35*(9), 12-16.

Haslhofer, B., & Isaac, A. (2011). data.europeana.eu: The europeana linked open data pilot. In *International Conference on Dublin Core and Metadata Applications* (pp. 94-104).

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian Journal of Information Systems, 19*(2), 87-92.

Hevner, A., & Chatterjee, S. (2010). Design science research in information systems. In *Design Research in Information Systems* (pp. 9-22). Springer, Boston, MA.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 75-105.

Hoekstra, R., & Groth, P. (2013). Linkitup: link discovery for research data. In *2013 AAAI Fall Symposium Series,* (pp. 28-35).

International Organization for Standardisation (2018). *Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts (ISO 9241-11)*. Retrieved from https://www.iso.org/standard/63500.html.

Jaffri, A., Glaser, H. & Millard, I.C. (2008). URI disambiguation in the context of linked data. In *1st Workshop on Linked Data on the Web (LDOW).*

Jiménez-Ruiz, E., & Grau, B. C. (2011). LogMap: Logic-based and scalable ontology matching. In *International Semantic Web Conference* (pp. 273-288). Springer, Berlin, Heidelberg.

Jiménez-Ruiz, E., Grau, B. C., & Horrocks, I. (2013). LogMap and LogMapLt results for OAEI 2012. *Ontology Matching*, 152.

Kroeger, A. (2013). The road to BIBFRAME: the evolution of the idea of bibliographic transition into a post-MARC Future. *Cataloging & Classification Quarterly, 51*(8*),* 873-890.

Krug, S. (2014). *Don't make me think, revisited: a common sense approach to Web usability.* (3rd ed.). San Francisco, California: New Riders, Peachpit, Pearson Education.

Kruskal, W.H. & Wallis W.A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association, 47*, 583-621.

Lampert, C. K., & Southwick, S. B. (2013). Leading to linking: Introducing linked data to academic library digital collections. *Journal of Library Metadata, 13(2-3),* 230-253.

LaPolla, F. (2013). Perceptions of librarians regarding Semantic Web and Linked Data technologies. *Journal of Library Metadata, 13*(2-3*),* 114-140.

Li, C., & Sugimoto, S. (2014, October). Provenance description of metadata using Prov with Premis for long-term use of metadata. In *International Conference on Dublin Core and Metadata Applications* (pp. 147-156).

Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: a methodology for information quality assessment*. Information & Management, 40(2),* 133-146.

Levene H. (1960) Robust tests for equality of variances. In *Contributions to Probability and Statistics – Essays in Honor of Harold Hotelling*, eds I. Olkin, S.G. Ghurye, W. Hoeffding,W.G. Madow and H.B. Mann (pp. 278-292). Stanford University Press, Stanford, California.

Lewis, J. R. (1992). Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. In *Proceedings of the Human Factors Society Annual Meeting 36(16),* 1259-1260. Sage CA: Los Angeles, CA: SAGE Publications.

Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction, 7(1)*, 57-78.

Lewis, J. R. (2002). Psychometric evaluation of the PSSUQ using data from five years of usability studies. *International Journal of Human-Computer Interaction, 14(3-4),* 463-488.

Lowdermilk, T. (2013). *User-centered design: a developer's guide to building user-friendly applications.* CA, USA: O'Reilly Media, Inc.

Malmsten, M. (2008). *Making a library catalogue part of the semantic web.* Paper presented at the International Conference on Dublin Core and Metadata Applications-Metadata for Semantic and Social Applications 22-26 September 2008, Berlin (DC-2008).

Malmsten, M. (2009). Exposing library data as linked data. *Paper presented at the IFLA Satellite Preconference* sponsored by the Information Technology Section Emerging Trends in Technology: Libraries between Web 2.0, Semantic Web and Search Technology, Florence.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics,* 50-60.

Manola F. & Miller, E. (2004, February 10). *RDF Primer*. Retrieved fromhttps://www.w3.org/TR/rdf-primer/.

Martin, C., & Clegg, C. (2012). A Socio-Technical Systems Analysis of the Potential Barriers to Adoption of Open Data Approaches by UK Public *Bodies.* In the *4th ACM International Conference on Web Science*.

McGuinness, D. L. & van Harmelen (2004, February 10). *OWL web ontology language*. Retrieved from https://www.w3.org/TR/owl-features/.

McKenna, L. (2017). Engaging librarians in the process of interlinking RDF resources. In *European Semantic Web Conference (ESWC).*

McKenna, L., Debruyne, C., & O'Sullivan, D. (2018). Understanding the Position of Information Professionals with regards to Linked Data: A survey of Libraries, Archives and Museums. In *2018 ACM/IEEE on Joint Conference on Digital Libraries*.

McKenna, L., Debruyne, C., & O'Sullivan, D. (2019). Modelling the Provenance of Linked Data Interlinks for the Library Domain. In *Companion Proceedings of the 2019 World Wide Web Conference*.

McKenna, L., Debruyne, C., & O'Sullivan, D. (2019). NAISC: An Authoritative Linked Data Interlinking Approach for the Library Domain. In *2019 ACM/IEEE Joint Conference on Digital Libraries*.

Meroño-Peñuela, A., Ashkpour, A., Guéret, C., & Schlobach, S. (2017). CEDAR: the Dutch historical censuses as linked open data. *Semantic Web, 8*(2), 297-310.

Meroño-Peñuela, A., Ashkpour, A., L. Rietveld, Hoekstra, R. & Schlobach, S. (2012). Linked humanities data: The next frontier? A Case-study in Historical Census Data. In *Proceedings of the 2nd International Workshop on Linked Science (LISC 2012) - Tackling Big Data.* CEUR Workshop Proceedings. Retrieved from research.vu.nl/en/publications/linked-humanities-data-the-next-frontier.

Miles, A. & Bechhofer, D. (2009). SKOS Simple Knowledge Organisation System Reference. *W3C Recommendation 18 August 2009*. Retrieved from https://www.w3.org/TR/skos-reference/.

Mitchell, E. T. (2016). Library linked data: Early activity and development. *Library Technology Reports, 52*(1), 5-33.

Moreau, L., Clifford, B., Freire, J., Futrelle, J., & Gil, Y. (2011). The Open Provenance Model core specification (v1.1). *Future Generation Computer Systems 27(6),* 743–756.

Moreau, L., Hartig, O., Simmhan, Y., Myers, J., Lebo, T., Belhajjame, K., Miles, S., & Soiland-Reyes, S. (2013, April 30). *PROV-AQ: Provenance access and query*. Retrieved from w3.org/TR/prov-aq/.

Nielsen, J., & Landauer, T. (1993). A mathematical model of the finding of usability problems. In *Human Factors in Computing Systems* (*CHI)*.

Neish, P. (2015). Linked data: what is it and why should you care? *The Australian Library Journal, 64*(1), 3-10.

Nentwig, M., Hartung, M., Ngonga Ngomo, A. C., & Rahm, E. (2017). A survey of current link discovery frameworks. *Semantic Web, 8*(3), 419-436.

Neubauer, G. (2017). Visualization of typed links in Linked Data. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare, 70*(2), 179-199.

Ngomo, A. C. N., & Auer, S. (2011). Limes – A time-efficient approach for large-scale link discovery on the web of data. In *Proceedings of IJCAI.*

Nguyen, V., Bodenreider, O., & Sheth, A. (2014). Don't like RDF reification? Making statements about statements using singleton property. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 759-770).

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*(1), *1-13*.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Paris, P. H. (2018). Assessing the Quality of owl: sameAs Links. In *European Semantic Web Conference* (pp. 304-313). Springer, Cham.

Paris, P. H., Hamdi, F., & Cherfi, S. S. S. (2019). Interlinking RDF-based datasets: A structure-based approach. *Procedia Computer Science, 159*, 162-171.

Papaleo, L., Pernelle, N., Saïs, F., & Dumont, C. (2014). Logical detection of invalid sameas statements in RDF data. In *International Conference on Knowledge Engineering and Knowledge Management* (pp. 373-384). Springer, Cham.

Paulheim, H. (2014). Identifying wrong links between datasets by multi-dimensional outlier detection. In *WoDOOM*, pp. 27-38.

Pearson, K. (1895). Notes on the history of correlation. *Proceedings of the Royal Society of London, 58,* 240-242.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems, 24(3),* 45-77.

Pesch, O., & Miller, E. (2016). Using BIBFRAME and library linked data to solve real problems: An interview with Eric Miller of Zepheira. *The Serials Librarian, 71(1),* 1-8.

PREMIS Editorial Committee (2015). *PREMIS data dictionary for preservation metadata.* Retrieved from https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf.

Raad, J., Beek, W., Van Harmelen, F., Pernelle, N., & Saïs, F. (2018, October). Detecting erroneous identity links on the web using network metrics. In *International Semantic Web Conference* (pp. 391-407). Springer, Cham.

Ritter, M., & Winterbottom, C. (2017). *UX for the Web: Build websites for user experience and usability.* Birmingham, UK: Packt Publishing Ltd.

Riva, P., Le Boeuf, P., & Žumer, M. (2016). *FRBR – Library reference model*. Retrieved from http://www.ifla.org/files/assets/cataloguing/frbr-lrm/frbr-lrm_20160225.pdf

Rosala, M. (2019). *How to Analyze Qualitative Data from UX Research: Thematic Analysis.* Retrieved from https://www.nngroup.com/articles/thematic-analysis/.

Sahoo, S. S., Sheth, A. (2009). Provenir ontology: Towards a framework for escience provenance management. In *Microsoft eScience Workshop*. Retrieved from https://corescholar.libraries.wright.edu/knoesis/80.

Schilling, V. (2012). *Transforming library metadata into linked library data.* Association for Library Collections & Technical Services (ALCTS). American Library Association. Retrieved from http://www.ala.org/alcts/resources/org/cat/research/linked-data.

Seeman, D., & Goddard, L. (2015). Preparing the way: creating future compatible cataloging data in a transitional environment. *Cataloging & Classification Quarterly, 53(3-4),* 331-340.

Shapiro S.S. & Wilk M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*, 591-611.

Simon, A., Wenz, R., Michel, V., & Di Mascio, A. (2013, May). Publishing bibliographic records on the web of data: opportunities for the BnF (French National Library). In *Extended Semantic Web Conference (ESWC)* (pp. 563-577). Springer, Berlin, Heidelberg.

Smith-Yoshimura, K. (2017). Analysis of International Linked Data Survey for Implementers. *D-Lib Magazine, 22(7/8),* 141–167. http://doi.org/10.1045/july2016-smith-yoshimura.

Smith-Yoshimura, K. (2018). Analysis of 2018 international linked data survey for implementers. *Code4Lib Journal*, *(42).* Retrieved from https://journal.code4lib.org/articles/13867.

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.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15*, 72-101.

Sprochi, A. (2016). Where Are We Headed? Resource Description and Access, Bibliographic Framework, and the Functional Requirements for Bibliographic Records Library Reference Model. *International Information & Library Review, 48*(2), 129-136.

Student (Gosset, W.S) (1908). The probable error of a mean. *Biometrika, 6*(1), 1–25.

Summers, E., Isaac, A., Redding, C., & Krech, D. (2008). LCSH, SKOS and linked data. In *The International Conference on Dublin Core and Metadata Applications-Metadata for Semantic and Social Applications,* 22-26 September, Berlin.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2,* 53.

Travis, D. (2011). *ISO 13407 is dead. Long live ISO 9241-210!.* Retrieved from https://www.userfocus.co.uk/articles/iso-13407-is-dead.html.

Tukey, J. (1949). Comparing individual means in the analysis of variance. *Biometrics, 5*(2), 99-114.

Usability First (2015). Introduction to User-Centered Design. Retrieved from usabilityfirst.com/about-usability/introduction-to-user-centered-design/.

Van Den Haak, M., De Jong, M., & Schellens, P. (2003). Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. *Behaviour & information technology, 22(5)*, 339-351.

Vila-Suero, D., & Gómez-Pérez, A. (2013). datos.bne.es and MARiMbA: an insight into library linked data. *Library Hi Tech, 31(4),* 575-601.

Vila-Suero, D., Villazón-Terrazas, B., & Gómez-Pérez, A. (2013). datos.bne.es: A library linked dataset. *Semantic Web, 4(3),* 307-313.

Virzi, R. A. (1992). Refining the test phase of usability evaluation: How many subjects is enough?. *Human factors, 34(4),* 457-468.

W3C, (2015). *Linked data*. Retrieved from w3.org/standards/semanticweb/data

Wieringa, R. J. (2014). *Design science methodology for information systems and software engineering*. Berlin Heidelberg: Springer-Verlag.

Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., & Auer, S. (2016). Quality assessment for linked data: A survey. *Semantic Web, 7(1),* 63-93.

Appendices

Appendix 1 – Linked Data Requirements Questionnaire

Do you have experience in one or more of the following areas?

- ☐ Working in a cultural heritage institution (e.g. library, archive, museum, gallery)
- ☐ Cataloguing library or cultural heritage metadata
- ☐ Researcher in the area of library science
- ☐ Researcher for a cultural heritage project
- ☐ Researcher in the area of Linked Data
- ☐ ⊗None of the above

What is your current occupation?

____

What type of institution do you currently work in e.g. academic library, public library, archive, museum?

____

Overall, how long have you worked in the library, archive, cultural heritage and/or research domain?

- ○ Less than 1 year
- ○ 1-3 years
- ○ 4-6 years
- ○ 7-9 years
- ○ 10 years or more

Are you involved in the metadata cataloguing process in your current workplace?

- ○ Yes
- ○ No

What metadata formats do you apply when cataloguing? Please select as many as appropriate.

- ☐ AudioMD and VideoMD
- ☐ BIBFRAME
- ☐ Dublin Core
- ☐ Darwin Core
- ☐ EAD (Encoded Archival Description)
- ☐ Extended Date/Time Format (EDTF)
- ☐ ISO 19115 Geographic Information
- ☐ LIDO – Lightweight Information Describing Objects
- ☐ MADS (Metadata Authority Description Standard)
- ☐ MADS-RDF
- ☐ MARC 21
- ☐ MARCXML
- ☐ METS (Metadata Encoding & Transmission Standard)
- ☐ MIX (NISO Metadata for Images in XML)
- ☐ MODS (Metadata Object Description Standard)
- ☐ MODS-RDF
- ☐ PB Core (Public Broadcasting Core Metadata)
- ☐ TEI (Text Encoding Initiative)
- ☐ TextMD (Technical Metadata for Text)
- ☐ UNIMARC
- ☐ VRA Core

☐ Other(s) – please list:

☐ ⊗Unsure/None

Throughout your career, what cataloguing tools/software have you gained experience in using?

☐ Aleph

☐ CALM

☐ Capita Prism (TALIS Prism)

☐ Dspace

☐ Fedora

☐ Koha

☐ Sierra

☐ Evergreeen

☐ Filemaker

☐ Lucidea

☐ Mandarin

☐ Omeka

☐ SobekCM

☐ Other – please state:

☐ ⊗None/Unsure

Did you require the support of a technical person to be able to use any of the tool(s) you selected previously?

	Yes	Maybe/Unsure	No
Aleph	○	○	○
CALM	○	○	○
Capita Prism (TALIS Prism)	○	○	○
Dspace	○	○	○
Fedora	○	○	○
Koha	○	○	○
Sierra	○	○	○
Evergreeen	○	○	○
Filemaker	○	○	○
Lucidea	○	○	○
Mandarin	○	○	○
Omeka	○	○	○
SobekCM	○	○	○
Other – please state:	○	○	○
Other – please state:	○	○	○
⊗None/Unsure	○	○	○

The following questions/statements will give you an opportunity to express your satisfaction with the Linked Data tools that you have experience in using.

Think about the tasks that you completed when using the tool while you answer these statements.

Please read each statement and indicate how strongly you agree or disagree with the statement by selecting a number on the scale.

It was simple to use this tool.

I could effectively complete my work using this tool.

I was able to complete my work quickly using this tool.

It was easy to learn how to use this tool.

Whenever I made a mistake using the tool, I recovered easily and quickly.

The organisation of information on the tool's screen was clear.

The interface of the tool was pleasant.

The tool had all the functions and capabilities I expected it to have.

Overall, I was satisfied with this tool.

	Strongly agree 1	2	3	4	5	6	Strongly disagree 7
Aleph	○	○	○	○	○	○	○
CALM	○	○	○	○	○	○	○
Capita Prism (TALIS Prism)	○	○	○	○	○	○	○
Dspace	○	○	○	○	○	○	○
Fedora	○	○	○	○	○	○	○
Koha	○	○	○	○	○	○	○
Sierra	○	○	○	○	○	○	○
Evergreeen	○	○	○	○	○	○	○

Filemaker	○	○	○	○	○	○	○
Lucidea	○	○	○	○	○	○	○
Mandarin	○	○	○	○	○	○	○
Omeka	○	○	○	○	○	○	○
SobekCM	○	○	○	○	○	○	○
Other – please state:	○	○	○	○	○	○	○
⊗None/Unsure	○	○	○	○	○	○	○

This section of the questionnaire is going to explore your knowledge of the Semantic Web and Linked Data, as well as your opinions on the usefulness of Linked Data in the library and cultural heritage domain.

How would you rate your current knowledge of the following topics:

	Extremely knowledgeable	Very knowledgeable	Moderately knowledgeable	Slightly knowledgeable	Not knowledgeable at all
Semantic Web	○	○	○	○	○
Linked Data	○	○	○	○	○
RDF	○	○	○	○	○
URIs	○	○	○	○	○
SPARQL	○	○	○	○	○
Ontologies	○	○	○	○	○

Do you think that **publishing** library and other cultural heritage metadata as Linked Data has the potential to **add value to the Semantic Web**?

- ◯ Yes
- ◯ Unsure
- ◯ No

In what way do you think that publishing library and other cultural heritage metadata as Linked Data could add value to the Semantic Web? Please select as many as appropriate.

- ☐ Create a research environment
- ☐ Easier metadata sharing
- ☐ Expose data to a larger audience
- ☐ Improve authority control on the Semantic Web
- ☐ Improve data accessibility for other institutions
- ☐ Improve Search Engine Optimisation (SEO)
- ☐ Increased metadata openness
- ☐ More efficient data searches
- ☐ Other(s), please list:

____

Do you think libraries and other cultural heritage institutions face barriers to **publishing** Linked Data?

- ◯ Yes
- ◯ Unsure
- ◯ No

What are these barriers? Please select as many as appropriate.

- [ ] Copyright issues
- [ ] Difficulty cleaning data
- [ ] Difficulty incorporating Linked Data publication into current workflow
- [ ] Difficulty establishing links
- [ ] Difficulty using Linked Data software/tools
- [ ] Difficulty using SPARQL endpoints
- [ ] Inadequate Linked Data software/tools available
- [ ] Insufficient availability of controlled vocabularies in Linked Data format
- [ ] Insufficient availability of useful ontologies
- [ ] Lack of useful case studies
- [ ] Steep learning curve
- [ ] Time consuming
- [ ] Other(s), please list: ______________________

Do you think that **consuming** (using) Linked Data resources has the potential to **benefit the library and cultural heritage domains**?

- ○ Yes
- ○ Unsure
- ○ No

What benefits do you think **consuming** Linked Data offers the library and cultural heritage domains? Please select as many as appropriate.

- ☐ Automated authority control
- ☐ Enriched bibliographic metadata
- ☐ Harmonising data from multiple sources
- ☐ Improved metadata quality
- ☐ Improved data discovery
- ☐ Interlinking across datasets
- ☐ Interlinking across institutions
- ☐ More efficient data searches
- ☐ Reduction in time spent creating metadata
- ☐ Other(s), please list:

Do you think libraries and other cultural heritage institutions face barriers to consuming Linked Data?

- ○ Yes
- ○ Unsure
- ○ No

What are these barriers? Please select as many as appropriate.

- ☐ Authority control issues
- ☐ Data format volatility issues
- ☐ Dataset reusability issues
- ☐ Dataset reliability issues
- ☐ Difficulty creating controlled vocabularies

- [ ] Difficulty establishing interlinks
- [ ] Difficulty incorporating Linked Data into cataloguing systems
- [ ] Difficulty using Linked Data software/tools
- [ ] Insufficient useful case studies
- [ ] Lack of suitable Linked Data software/tools available
- [ ] Lack of useful resources available as Linked Data
- [ ] Large size of RDF dumps
- [ ] Slow dataset updates
- [ ] Time consuming processes
- [ ] Unstable endpoints
- [ ] Use of unstable URIs
- [ ] Other(s), please list:

____

Have you ever been directly involved in the implementation of a Linked data project/service?

○ Yes

○ No

How was Linked Data used in this project/service?

○ Consumed Linked Data

○ Published Linked Data

○ Consumed and Published Linked Data

What Linked Data datasets were consumed as part of the project/service?

- ☐ AAT – Art and Architecture Thesaurus
- ☐ AGROVAC
- ☐ American Numismatic Society's Thesaurus of Numismatic Concepts
- ☐ The British Museum's Semantic Web Collection
- ☐ British National Bibliography (British Library)
- ☐ CIDOC-CRM
- ☐ DBpedia
- ☐ Dewey Web Services
- ☐ DPLA – Digital Public Library of America
- ☐ Drug Encyclopedia
- ☐ Europeana
- ☐ The European Library (TEL)
- ☐ EuroVoc – Multilingual Thesaurus of the European Union
- ☐ FAST – Faceted Application of Subject Terminology
- ☐ GEMET – General Multilingual Environmental Thesaurus
- ☐ GeoNames
- ☐ ISNI – International Standard Name Identifier
- ☐ Library of Congress Linked Data Services – id.loc.gov
- ☐ NAF – Library of Congress / NACO Authority File
- ☐ ORCID
- ☐ Research Libraries UK Lined Open Data

- ☐ Smithsonian Libraries' Books Online
- ☐ TGM – Thesaurus of Graphic Materials
- ☐ TGN – Getty Thesaurus of Geographic Names
- ☐ ULAN – Union List of Artist Names
- ☐ University College Dublin's Digital Library
- ☐ VIAF – Virtual International Authority File
- ☐ Wikidata
- ☐ WorldCat.org
- ☐ WorldCat.org Works
- ☐ Others – please list:

- ☐ ⊗Unsure/None

What type of data was published as Linked Data? Please select as many as appropriate.

- ☐ Authority files
- ☐ Bibliographic data
- ☐ Controlled Vocabularies
- ☐ Digital collections
- ☐ Encoded Archival Descriptions
- ☐ Geographic data
- ☐ Holdings data
- ☐ Museum object data

- [ ] Ontologies
- [ ] Statistical Data
- [ ] Other(s), please list:

____

What RDF Vocabularies and Ontologies were used for the project/service?

- [ ] ARCH – Archival Collection Ontology
- [ ] BIBO- The Bibliographic Ontology
- [ ] BIO – A Vocabulary for Biographical Information
- [ ] BF – BIBFRAME Vocabulary
- [ ] BLT – British Library Terms RDF Schema
- [ ] CIDOC-CRM
- [ ] CERIF semantic vocabularies
- [ ] DAT – Data Catalogue Vocabulary
- [ ] Dcterms – DCMI Metadata Terms
- [ ] DCE – Dublin Core Metadata Element Set
- [ ] EAC-CPF Description Ontology for Linked Archival Data
- [ ] ELM – Europeana Data Model vocabulary
- [ ] The Event Ontology
- [ ] FABIO – FRBR-aligned Bibliographic Ontology
- [ ] FOAF – Friend of a Friend
- [ ] FRBR – Expression of Core FRBR Concepts in RDF
- [ ] GEO – WGS84 Geo Positioning

- [ ] ISBD – ISBD elements
- [ ] LIB – Library Extension of Schema.org
- [ ] Local Vocabulary
- [ ] MADS – Metadata Authority Description Schema
- [ ] MAP – DPLA Metadata Application Profile
- [ ] MO – Music Ontology
- [ ] MODS – Metadata Object Description Schema
- [ ] Nomisma Ontology
- [ ] ORE – The OAI ORE Terms Vocabulary
- [ ] ORG – Care Organisation Ontology
- [ ] OWL
- [ ] Radatana
- [ ] RDA – Local Vocabulary VOCABS
- [ ] RDAG2 – RDA Group 2 Elements
- [ ] RDF Schema
- [ ] REV – Review Vocabulary
- [ ] schema – Schema.org vocabulary
- [ ] SIOC
- [ ] SKOS – Simple Knowledge Organisation System
- [ ] VIVO – VIVO Core Ontology
- [ ] viaf.org/ontology

- [ ] Other(s) – please list:

- [ ] ⊗Unsure/None

Did you **personally gain experience** using any Linked Data Tools whilst working on the project? If so, which ones?

- [ ] AgreementMaker
- [ ] Apache Fuseki
- [ ] ARC2
- [ ] CODI
- [ ] D2R Server
- [ ] Fedora Commons
- [ ] KnoFuss
- [ ] LIMES
- [ ] LogMap
- [ ] OAI2LOD Server
- [ ] OpenRefine (GoogleRefine)
- [ ] Protege
- [ ] Pubby
- [ ] RDF Refine
- [ ] RiMOM
- [ ] RuleMiner
- [ ] SAIM

- ☐ SERIMI
- ☐ SILK
- ☐ IOC Exporters
- ☐ SLINT
- ☐ SparqPlug
- ☐ Talis Platform
- ☐ Triplify
- ☐ Vapour Validation
- ☐ Virtuoso Universal Server
- ☐ Zhisi.Links
- ☐ Other – please state:

____

- ☐ ⊗None/Unsure

Did you require the support of a technical person to be able to use any of the tools?	Yes	Maybe/Unsure	No
AgreementMaker	○	○	○
Apache Fuseki	○	○	○
ARC2	○	○	○
CODI	○	○	○
D2R Server	○	○	○
Fedora Commons	○	○	○

KnoFuss	○	○	○
LIMES	○	○	○
LogMap	○	○	○
OAI2LOD Server	○	○	○
OpenRefine (GoogleRefine)	○	○	○
Protege	○	○	○
Pubby	○	○	○
RDF Refine	○	○	○
RiMOM	○	○	○
RuleMiner	○	○	○
SAIM	○	○	○
SERIMI	○	○	○
SILK	○	○	○
IOC Exporters	○	○	○
SLINT	○	○	○
SparqPlug	○	○	○
Talis Platform	○	○	○
Triplify	○	○	○
Vapour Validation	○	○	○
Virtuoso Universal Server	○	○	○
Zhisi.Links	○	○	○
Other – please state:	○	○	○

⊗None/Unsure	○	○	○

The following questions/statements will give you an opportunity to express your satisfaction with the Linked Data tools that you have experience in using.

Please read each statement and indicate how strongly you agree or disagree with the statement by selecting a number on the scale.

It was simple to use this tool.

I could effectively complete my work using this tool.

I was able to complete my work quickly using this tool.

It was easy to learn how to use this tool.

Whenever I made a mistake using the tool, I recovered easily and quickly.

The organisation of information on the tool's screen was clear.

The interface of the tool was pleasant.

The tool had all the functions and capabilities I expected it to have.

Overall, I was satisfied with this tool.

	Strongly agree 1	2	3	4	5	6	Strongly disagree 7
AgreementMaker	○	○	○	○	○	○	○
Apache Fuseki	○	○	○	○	○	○	○
ARC2	○	○	○	○	○	○	○
CODI	○	○	○	○	○	○	○
D2R Server	○	○	○	○	○	○	○
Fedora Commons	○	○	○	○	○	○	○
KnoFuss	○	○	○	○	○	○	○
LIMES	○	○	○	○	○	○	○

LogMap	○	○	○	○	○	○	○
OAI2LOD Server	○	○	○	○	○	○	○
OpenRefine (GoogleRefine)	○	○	○	○	○	○	○
Protege	○	○	○	○	○	○	○
Pubby	○	○	○	○	○	○	○
RDF Refine	○	○	○	○	○	○	○
RiMOM	○	○	○	○	○	○	○
RuleMiner	○	○	○	○	○	○	○
SAIM	○	○	○	○	○	○	○
SERIMI	○	○	○	○	○	○	○
SILK	○	○	○	○	○	○	○
IOC Exporters	○	○	○	○	○	○	○
SLINT	○	○	○	○	○	○	○
SparqPlug	○	○	○	○	○	○	○
Talis Platform	○	○	○	○	○	○	○
Triplify	○	○	○	○	○	○	○
Vapour Validation	○	○	○	○	○	○	○
Virtuoso Universal Server	○	○	○	○	○	○	○
Zhisi.Links	○	○	○	○	○	○	○
Other – please state:	○	○	○	○	○	○	○
⊗None/Unsure	○	○	○	○	○	○	○

This section of the questionnaire will explore your thoughts on the development of a Linked Data tool for the library, archive and cultural heritage domain.

If a Linked Data tool was developed specifically for librarians, archivists and related professionals – do you think it would be more important for the tool to enable the:

- ◯ Publication of Linked Data
- ◯ Consumption of Linked Data
- ◯ Both publication and consumption of Linked Data
- ◯ Neither/None
- ◯ I do not think these professionals require bespoke Linked Data tools

Why do you think this?

____
____
____
____
____

How useful would you consider a Linked Data **interlinking tool** for librarians, archivists and related professionals to be?

- ◯ Extremely **useless**
- ◯ Moderately useless
- ◯ Slightly useless
- ◯ Neither useful nor useless
- ◯ Slightly useful
- ◯ Moderately useful
- ◯ Extremely **useful**

Why do you think this?

____
____
____
____
____

In your opinion, what functions would be important for the tool to have? What would you like to be able to do with such a tool? Please select as many as appropriate.

- ☐ Awareness of common library, archive and cultural heritage data sources
- ☐ Create controlled vocabularies in SKOS (Simple Knowledge Organising System)
- ☐ Configurable to the institution's workflow
- ☐ Data cleaning
- ☐ Data enriching
- ☐ Integrate Linked Data datasets into the catalogue
- ☐ Link discovery
- ☐ Link to related controlled vocabularies automatically
- ☐ Link to related ontologies automatically
- ☐ Review the quality of an external data source before use
- ☐ Remove the need for understanding Linked Data technicalities
- ☐ Validate/verify generated links
- ☐ Vocabulary alignment/reconciliation
- ☐ Others(s), please list:

____

- ☐ ⊗Unsure/None

If an Interlinking Tool for libraries and cultural heritage institutions did exist Linked Data datasets would you find most useful to interlink with? Please select as many as appropriate.

- ☐ AAT – Art and Architecture Thesaurus
- ☐ AGROVAC
- ☐ American Numismatic Society's Thesaurus of Numismatic Concepts
- ☐ The British Museum's Semantic Web Collection
- ☐ British National Bibliography (British Library)
- ☐ CONA – Cultural Objects Name Authority
- ☐ DBpedia
- ☐ Dewey Web Services
- ☐ DPLA – Digital Public Library of America
- ☐ Drug Encyclopedia
- ☐ Europeana
- ☐ The European Library (TEL)
- ☐ EuroVoc – Multilingual Thesaurus of the European Union
- ☐ FAST – Faceted Application of Subject Terminology
- ☐ GEMET – General Multilingual Environmental Thesaurus
- ☐ GeoNames
- ☐ ISNI – International Standard Name Identifier
- ☐ Library of Congress Lined Data Services – id.loc.gov
- ☐ NAF – Library of Congress / NACO Authority File
- ☐ ORCID

- [ ] Research Libraries UK Lined Open Data
- [ ] Smithsonian Libraries' Books Online
- [ ] TGM – Thesaurus of Graphic Materials
- [ ] TGN – Getty Thesaurus of Geographic Names
- [ ] ULAN – Union List of Artist Names
- [ ] University College Dublin's Digital Library
- [ ] VIAF – Virtual International Authority File
- [ ] Wikidata
- [ ] WorldCat.org
- [ ] WorldCat.org Works
- [ ] Other(s) – please list:

- [ ] ⊗None/Unsure

When completing different metadata tasks, what evaluation criteria do you apply when using, or searching for, external data sources? Please select as many as appropriate.

- [ ] Availability of the external source (e.g. SPARQL endpoint is accessible)
- [ ] Licensing issues (e.g. Can I use this external source freely?)
- [ ] Syntactic validity (e.g. Are dates in the correct format, correct spelling?)
- [ ] Conciseness (e.g. Is there any redundancy within the external source?)
- [ ] Completeness (e.g. Do all external metadata fields have values?)

- [ ] Trustworthiness (e.g. Can this provider be trusted that all data is correct?)
- [ ] Understandability (e.g. Are all records in the external source labelled and ready for human consumption?)
- [ ] Timeliness (e.g. Are all records up to date?)
- [ ] Provenance (e.g. Does the external source provide provenance/origin information on the data?)
- [ ] Interoperability (e.g. Does the external source use well-known standard schemas to represent the data?)
- [ ] Versatility (e.g. Is the data available in different languages?)
- [ ] Other, please list:

____

- [ ] ⊗None/Unsure

Can you give an example of a data quality issue or concern that you experience frequently?

____
____
____
____
____

Do you have any additional thoughts or feedback regarding the topics discussed during this questionnaire?

____
____
____
____
____

Appendix 2 – Requirements Questionnaire Information Sheet & Consent Form

Participant Information Sheet

Who is conducting the research?

The research is conducted by a PhD student from Trinity College Dublin.

What is the aim of the questionnaire?

The aim of this questionnaire is to gather information on librarians' current cataloguing processes and cataloguing interfaces, as well as their knowledge, use and views on linked data for libraries.

Why was I asked to participate?

You have been asked to participate in this research as you have been identified as someone who has experience working as an Information Professional.

What will the participation in research involve?

If you agree to participate in the research, you will be asked to complete a questionnaire in which you will be asked to answer a series of questions regarding the topics outlined above. The majority of these questions will require you to select an answer from a number of options, and some will require you to write your own answer. It should take approximately 30 minutes to complete.

What happens to the information I provide?

The data gathered from this questionnaire will be used as part of the above research project. All information you provide will be treated with full confidentiality and, if published, will not be identifiable as yours.

What if I change my mind about participating in this research?

Your participation in this questionnaire is completely voluntary. You are free to withdraw from the research at any time without any penalty. Should you choose to withdraw from the research, all information you have provided will be deleted.

Any questions? Contact Lucy McKenna at lucy.mckenna@adaptcentre.ie.

Consent Form

Researcher: Lucy McKenna (lucy.mckenna@adaptcentre.ie)

Background: The aim of this research is to gather information on librarians' current cataloguing processes and cataloguing interfaces, as well as their knowledge, use and views on linked data for libraries.

Procedure: If you agree to participate in the research, you will be asked to complete a questionnaire in which you will be asked to answer a series of questions regarding the topics outlined above. It should take approximately 30 minutes to complete the questionnaire.

Publication: The data gathered from this questionnaire will be used as part of the researcher's PhD thesis and may be also be presented at academic conferences. All information you provide will be treated with full confidentiality and, if published, will not be identifiable as yours.

Declaration:

- I am 18 years or older and am competent to provide consent.
- I have read a document providing information about this research. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction.
- I agree that my data is used for scientific purposes and I have no objection that my data is published in scientific publications in a way that does not reveal my identity.
- I understand that I may stop electronic recordings at any time, and that I may at any time, even subsequent to my participation have such recordings destroyed.
- I understand that no recordings will be replayed in any public forum or made available to any audience other than the current researcher.
- I freely and voluntarily agree to be part of this research study, without prejudice to my legal and ethical rights.
- I understand that I may refuse to answer any question and that I may withdraw at any time without penalty.
- I understand that my participation is fully anonymous and that no personal details about me will be recorded.
- I have received a copy of this agreement.

Signature: ____________________________ Date: _______________

Appendix 3- Pre-Test Questionnaire

How would you rate your knowledge of the following topics? (tick the appropriate box):

	Extremely Knowledgeable	Very Knowledgeable	Moderately Knowledgeable	Slightly Knowledgeable	Not at all Knowledgeable
Semantic Web					
Linked Data					
RDF					
URIs					
Ontologies					

Have you ever been directly involved in the implementation of a Linked Data project or service?

☐ Yes ☐ No

If yes, what kinds of activities did you gain experience in?

☐ Creating URIs ☐ Creating Ontologies

☐ Creating and Publishing RDF data ☐ Integrating RDF data

☐ Interlinking internal RDF resources to external RDF resources

☐ Other, please describe:

Appendix 4 – PSSUQ

Item	1 Strongly Agree	2	3	4	5	6	7 Strongly Disagree	N/A
Overall, I am satisfied with how easy it is to use this system								
Comment:								
It was simple to use this system								
Comment:								
I could effectively complete the tasks and scenarios using this system								
Comment:								
I was able to complete the tasks and scenarios quickly using this system								
Comment:								
I was able to efficiently complete the tasks and scenarios using this system								
Comment:								
I felt comfortable using this system								
Comment:								
It was easy to learn to use this system								
Comment:								
I believe I could become productive quickly using this system								

Comment:								
The system gave error messages that clearly told me how to fix problems								
Comment:								
Whenever I made a mistake using the system, I could recover easily and quickly								
Comment:								
The information (such as on-line help, on-screen messages, and other documentation) provided with this system was clear								
Comment:								
It was easy to find the information I needed								
Comment:								
The information provided for the system was easy to understand								
Comment:								
The information was effective in helping me complete the tasks and scenarios								
Comment:								
The organization of information on the system screens was clear								
Comment:								

The interface of this system was pleasant							
Comment:							
I liked using the interface of this system							
Comment:							
This system has all the functions and capabilities I expect it to have							
Comment:							
Overall, I am satisfied with this system							
Comment:							
Overall, I am satisfied with how easy it is to use this system							
Comment:							

Appendix 5 – CSUQ

Item	1 Strongly Agree	2	3	4	5	6	7 Strongly Disagree	N/A
Overall, I am satisfied with how easy it is to use this system								
Comment:								
It is simple to use this system								
Comment:								
I can effectively complete my work using this system								
Comment:								
I am able to complete my work quickly using this system								
Comment:								
I am able to efficiently complete my work using this system								
Comment:								
I feel comfortable using this system								
Comment:								
It is easy to learn to use this system								
Comment:								
I believe I became productive quickly using this system								
Comment:								

The system gives error messages that clearly tell me how to fix problems								
Comment:								
Whenever I make a mistake using the system, I recover easily and quickly								
Comment:								
The information (such as on-line help, on-screen messages, and other documentation) provided with this system is clear								
Comment:								
It is easy to find the information I need								
Comment:								
The information provided for the system is easy to understand								
Comment:								
The information is effective in helping me complete my work								
Comment:								
The organisation of information on the system screens is clear								
Comment:								
The interface of this system is pleasant								
Comment:								

I like using the interface of this system								
Comment:								
This system has all the functions and capabilities I expect it to have								
Comment:								
Overall, I am satisfied with this system								
Comment:								

Appendix 6 – AIMQ Questionnaire (Lee et al, 2001)

All items are measured on a 0 to 10 scale where 0 is not at all and 10 is completely. Items labels with‘‘(R)’’are reverse coded.

Accessibility. (4 items, Cronbach’s Alpha¼.92)
This information is easily retrievable.
This information is easily accessible.
This information is easily obtainable.
This information is quickly accessible when needed.

Appropriate Amount. (4 items, Cronbach’s Alpha¼.76)
This information is of sufficient volume for our needs.
The amount of information does not match our needs. (R)
The amount of information is not sufficient for our needs. (R)
The amount of information is neither too much nor too little.

Believability. (4 items, Cronbach’s Alpha¼.89)
This information is believable.
This information is of doubtful credibility. (R)
This information is trustworthy.
This information is credible.

Completeness. (6 items, Cronbach’s Alpha¼.87)
This information includes all necessary values.
This information is incomplete. (R)
This information is complete.
This information is sufficiently complete for our needs.
This information covers the needs of our tasks.
This information has sufficient breadth and depth for our task.

Concise Representation. (4 items, Cronbach’sAlpha¼.88)
This information is formatted compactly.
This information is presented concisely.
This information is presented in a compact form.
The representation of this information is compact and concise.

Consistent Representation. (4 items, Cronbach’sAlpha¼.83)
This information is consistently presented in the same format.
This information is not presented consistently. (R)
This information is presented consistently.
This information is represented in a consistent format.

Ease of Operation. (5 items, Cronbach’s Alpha¼.85)
This information is easy to manipulate to meet our needs.
This information is easy to aggregate.
This information is difficult to manipulate to meet our needs. (R)
This information is difficult to aggregate. (R)
This information is easy to combine with other information.

Free of Error. (4 items, Cronbach's Alpha¼.91)
This information is correct.
This information is incorrect. (R)
This information is accurate.
This information is reliable.

Interpretability. (5 items, Cronbach's Alpha¼.77)
It is easy to interpret what this information means.
This information is difficult to interpret. (R)
It is difficult to interpret the coded information. (R)
This information is easily interpretable.
The measurement units for this information are clear.

Objectivity. (4 items, Cronbach's Alpha¼.72)
This information was objectively collected.
This information is based on facts.
This information is objective.
This information presents an impartial view.

Relevancy. (4 items, Cronbach's Alpha¼.94)
This information is useful to our work.
This information is relevant to our work.
This information is appropriate for our work.
This information is applicable to our work.

Reputation. (4 items, Cronbach's Alpha¼.85)
This information has a poor reputation for quality. (R)
This information has a good reputation.
This information has a reputation for quality.
This information comes from good sources.

Security. (4 items, Cronbach's Alpha¼.81)
This information is protected against unauthorized access.
This information is not protected with adequate security. (R)
Access to this information is sufficiently restricted.
This information can only be accessed by people who should see it.

Timeliness. (5 items, Cronbach's Alpha¼.88)
This information is sufficiently current for our work.
This information is not sufficiently timely. (R)
This information is not sufficiently current for our work. (R)
This information is sufficiently timely.
This information is sufficiently up-to-date for our work.

Understandability. (4 items, Cronbach's Alpha¼.90)
This information is easy to understand.
The meaning of this information is difficult to understand. (R)
This information is easy to comprehend.
The meaning of this information is easy to understand

Appendix 7 – Data Quality Questionnaire

All items are measured on a 0 to 10 scale where 0 is not at all and 10 is completely. Items labels with‘‘(R)’’are reverse coded.

Appropriate Amount.

This information is of sufficient volume for our needs.

The amount of information does not match our needs. (R)

Believability.

This information is believable.

This information is of doubtful credibility. (R)

This information is trustworthy.

Completeness.

This information includes all necessary values.

This information is incomplete. (R)

This information is sufficiently complete for our needs.

Concise Representation.

This information is presented concisely.

This information is presented in a compact form.

Consistent Representation.

This information is presented consistently.

Ease of Operation.

This information is easy to manipulate to meet our needs.

This information is easy to combine with other information.

Free of Error.

This information is incorrect. (R)

This information is accurate.

This information is reliable.

Interpretability.

It is easy to interpret what this information means.

Objectivity.

This information is objective.

Relevancy.

This information is useful to our work.

This information is relevant to our work.

Reputation.

This information comes from good sources.

Timeliness.

This information is not sufficiently timely. (R)

This information is sufficiently up-to-date for our work.

Understandability.

The meaning of this information is difficult to understand. (R)

This information is easy to comprehend.

Appendix 8 – Think-Aloud Test Protocol

This document provides a detailed plan of usability testing protocol for the evaluation of the NAISC-L Framework.

Goals

1. Create a project/collection
2. Add primary resources to a collection
3. Add secondary resources to a primary resource
4. Link primary and secondary resources
5. Add provenance data
6. View output

Equipment

- Usability test activities for the participant
- Consent form
- Information sheet
- Pens
- Researcher Observation forms
- Internet access
- Laptop with screen recorder
- Audio recorder (& charger)
- Stopwatch
- RDF Dataset

Plan

1. **Welcome the participant and explain the aims and objectives of the usability test.**

   Thank you for agreeing to take part in this user review. The purpose of this study is to review the RDF Interlinking Framework that I have developed as part of my PhD research.

2. **Provide the participant with an information sheet, consent form and discuss the contents of the documents.**

   Participation in this study is entirely voluntary and you are free to withdraw from the study at any stage. Quotes and results from your review may be used my documentation reporting on the study, however your identity will remain anonymous. Additionally, with your consent, this test will be audio recorded.

3. **Explain the usability test process to the participant**

   The aim of this study is to review the usefulness of the interlinking framework as well as how easy or difficult the interface is to use. The review will comprise of you completing a set of activities using the framework, completing a short pre and post-test questionnaire, and finally answering a few questions on your experience using the framework. The review should take no longer than 40 minutes. Any questions?

4. **Pre-Test Questionnaire**

5. **Think Aloud Test**

   As mentioned previously, during the test you will be asked to complete a set of activities. These activities will be completed using the think-aloud method. In other words, you will be asked to verbalise your actions, thoughts and questions while you are completing each activity. This is how I will collect data about the framework. I may ask you questions or give you verbal prompts to continue thinking aloud during the review process if required.

I want to make it clear that we are testing the framework, not you. There are no wrong or right ways to complete these activities, just complete them in the way that makes the most sense to you.
Also, I want to know exactly what you think about the website, so please do not worry that you are going to cause offence if you report any difficulties or negatives. By using this information, I will be able to improve the framework

As mentioned before, as you complete the activities, I am going to ask you to think aloud and to verbalise what going through your mind as much as possible.

If you have questions during the activities, feel free ask them. I may not be able to answer them right away since I am interested in how you complete each activity by yourself without assistance, but I will try to answer any questions you have once the activities are complete.

6. Demonstration Activity

- I will now demonstrate to you how to think aloud while completing a simple activity.

7. Scenario

- Read scenario aloud

8. Tasks

- Prompting questions or instructions may be given if the participant is silent for a prolonged period of time or if the participant is not following the task instructions.
- If a participant is unable to complete a task, assistance may be required. If this occur:
    - Ask what the user would do if the researcher was not present
    - Take note of how you influenced/ affected the participant
- Finish the test once all tasks are completed.

Participant Task Sheet

Scenario:

You are a metadata cataloguer working for the France National Library (Bibliotheque nationale de France). You have created a set of metadata records for a collection of recently donated items related to the Irish writer and poet James Joyce. Following this you have published these records in RDF format.

In order to produce 5 Star Linked Open Data, you plan to link the objects, concepts and terms found in your RDF dataset to related resources found in other institutions and external datasets. This will provide richer context to the data for future users. To create these links, you plan to use the NAISC Framework. NAISC stands for Novel Authoritative Interlinking of Schema and Concepts, and is an Interlinking Framework designed with the needs and expertise of Information Professionals in mind.

Tasks:

1. You intend to create a number of interlinks that will connect the resources in your BnF dataset to related resources found in authoritative external datasets. To begin this process, you first need to create an Interlink Collection which will contain the links that you create. As your dataset is related to James Joyce, the title of the Interlink Collection can reflect this as can the Description. You should use your participant number (__) as the Creator name.

2. The BnF dataset contains a resource/record for the writer James Joyce. You plan to interlink this resource to a related resource in an external dataset. In order to do this, you must first add an Internal Resource to your collection. This Internal Resource is the Uniform Resource Identifier (URI) that represents the resource/record for James Joyce in the BnF dataset that you are working with. You can search the BnF dataset for the URI using the SemFacet tool.

3. Now you need a to add a resource that you would like to interlink with the Internal Resource you just added to your interlink collection. To do this, you

must add a Related Resource to the Internal Resource. In this case, the Related Resource is going to be the VIAF (Virtual International Authority File) record for James Joyce.

4. Using the data in the Interlink Collection create a link between each pair of resources in the collection. An interlink should describe the relationship between the two resources.

5. Now view the interlink graph and the interlink RDF output.

6. View the provenance data describing the interlinks you just created.

Evaluator Observation Sheet

Task 1 <Create Collection>

<table>
<tr><td>Goal/Output:</td><td colspan="2">Create a new collection</td></tr>
<tr><td>Inputs:</td><td>- Collection Name
- Collection Description
- Creator Name</td><td>Organisation Name

Date</td></tr>
<tr><td>Assumptions:</td><td colspan="2">- User already has a 4 Star Linked Dataset
- Collection does not already exist</td></tr>
<tr><td>Steps:</td><td colspan="2">- Enter information in form
- Click Create button</td></tr>
<tr><td>Time for expert:</td><td colspan="2"></td></tr>
<tr><td>Instructions for user:</td><td colspan="2">As mentioned, you intend to create a number of interlinks that will link the resources in your dataset to related resources found in authoritative external datasets. To begin this process, you first need to create an Interlink Collection which will contain the links that you create. As your dataset is related to James Joyce, the title of the Interlink Collection can reflect this. You should use your participant number (Participant __) as the Creator name.</td></tr>
<tr><td>Time for User</td><td colspan="2"></td></tr>
<tr><td>Notes:</td><td colspan="2"></td></tr>
</table>

Task 2 <Add Primary Resource>

Goal/Output:	Add a primary resource to a collection	
Inputs:	- Resource URI - Resource Description	
Assumptions:	- RDF Dataset - RDF Dataset is loaded and prepared for searching in SemFacet - Resources named in instructions exist in dataset	
Steps:	- Click on Edit Collection - Click on Add Primary Resource - Click on Open SemFacet - Search for resource	- Click on NAISC tab - Enter URI - Enter Description - Click Save
Time for expert:		
Instructions for user:	*Your dataset contains a resource/record for the writer James Joyce. You plan to interlink this resource to a related resource in an external dataset. In order to do this you must first add the Uniform Resource Identifier (URI), representing the resource for James Joyce in your dataset, to the Interlink Collection. You can search for the URI using the SemFacet tool.*	
Time for User		
Notes:		

Task 3 <Add Secondary Resource>

Goal/Output:	Add a Secondary Resource to a Primary Resource	
Inputs:	- Resource URI - Resource Description	
Assumptions:	- Related resource exists in authority	
Steps:	- Click on Manage Collection - Click on Add Secondary Resource - Click on Authority Link - Search Authority for resource	- Click on NAISC tab - Enter URI - Enter Description - Click Save
Time for expert:		
Instructions for user:	*Following this you must search an external authoritative dataset for a related resource that you can interlink with your record for James Joyce. Search the VIAF (Virtual International Authority File) dataset for a record for James Joyce and add the URI to the Interlink Collection so that it can be linked to the first resource URI that you added.*	
Time for User		
Notes:		

Task 4 <Interlink>

Goal/Output:	Interlink the Primary and Secondary Resources.
Inputs:	- Select ontology - Select predicate - Add justification
Assumptions:	- Relationships between the resources are relatively self-explanatory
Steps:	- Click on Manage Collection - Click on Start Interlinking Session - Determine relationship - Select predicate - Click on Preview Interlinks - Click on End Interlinking Session - Click on Publish Interlinks
Time for expert:	
Instructions for user:	• *Using the data in the Interlink Collection, named Test Collection, that has been previously added prior to this review, create a link between each pair of resources in the collection. An interlink should describe the relationship between the two resources.* • *Following this, review and publish the new interlinks.*
Time for User	
Notes	

Task5 <View and Publish Interlinks>

<table>
<tr><td>Goal/Output:</td><td colspan="2">Publish Interlinks</td></tr>
<tr><td>Inputs:</td><td colspan="2">- Select ontology
- Select predicate
- Add justification</td></tr>
<tr><td>Assumptions:</td><td colspan="2">- Relationships between the resources are relatively self-explanatory</td></tr>
<tr><td>Steps:</td><td>- Click on Manage Collection
- Click on Start Interlinking Session
- Determine relationship
- Select predicate</td><td>- Click on Preview Interlinks
- Click on End Interlinking Session
- Click on Publish Interlinks</td></tr>
<tr><td>Time for expert:</td><td colspan="2"></td></tr>
<tr><td>Instructions for user:</td><td colspan="2">Now review and publish only the newly created interlinks.</td></tr>
<tr><td>Time for User:</td><td colspan="2"></td></tr>
<tr><td>Notes:</td><td colspan="2"></td></tr>
</table>

Task 6 <Provenance>

Goal/Output:	Add provenance data	
Inputs:		
Assumptions:	- Automatically completed based on data entered during session	
Steps:	- Click on Review Provenance Data - Publish Provenance Data	
Time for expert:		
Instructions for user:	*Review the provenance data describing the interlinks you just created.*	
Time for User		
Notes:		

9. Post-Test Interview

1. What is your overall impression of the tool?
2. What worked well?
3. What challenges did you encounter?
4. Are there any functions you would like to add or remove?
5. What is your impression of the process for selecting link-types in order to link internal and external entities?
6. What is your impression of the provenance data stored for the links and interlinking session?
7. Do you think this tool could be useful for the LAM domain?

10. PSSUQ

11. Debrief and Wrap-up

Thank you for participating in thus user review, I really appreciate you giving us your time. Are there any questions you would like to ask before we wrap up?

Appendix 9 – Think-Aloud Test Information Sheet & Consent Form

Participant Information Sheet

Who is conducting the research?
The research is conducted by a PhD student from the ADAPT Centre, a research centre based in Trinity College Dublin.
PhD Student: Lucy McKenna, ADAPT Centre, Trinity College Dublin
Academic Supervisor: Prof. Declan O'Sullivan, ADAPT Centre, TCD
Co-Supervisor: Dr. Christophe Debruyne, ADAPT Centre, TCD

What is the aim of the user review?
The aim of user review is to gather information on the usefulness and ease of use of an RDF Interlinking Framework that had been developed for Information Professionals.

Why was I asked to participate?
You have been asked to participate in this research as you have been identified as someone who is familiar with the Semantic Web and Linked Data, and who is also an Information Professional.

What will the participation in research involve?
You will also be asked to complete a number of activities using the Interlinking Framework whilst being observed by the researcher. Once finished, you will be asked to provide feedback on your experience using the tool in the form of a questionnaire and brief interview. This should all take approximately 40 minutes of your time. With your permission, the review process will be audio-recorded.

What happens to the information I provide?
The data gathered from this questionnaire will be used as part of the researcher's PhD thesis and may be also be presented at academic conferences. All information you provide will be treated with full confidentiality and, if published, will be anonymised. Audio recordings will not be made available to anyone other than the researcher and will not be replayed in any public forum or presentation of the research.

What if I change my mind about participating in this research?
Your participation in this research is completely voluntary. You are free to withdraw from the research at any time.

Any questions?
Please feel free to contact the researcher, Lucy McKenna, via email at lucy.mckenna@adaptcentre.ie.

Consent Form

Researcher: Lucy McKenna (lucy.mckenna@adaptcentre.ie)

Background: The aim of this research is to test an RDF Interlinking Framework that has been developed for Information Professionals.

Procedure: If you agree to participate in the research you will be asked to complete a review of the Interlinking Framework which will involve completing a number of activities using the tool and providing your feedback on your experience. It should take approximately 40 minutes to complete the review and feedback.

Publication: The data gathered from this questionnaire will be used as part of the researcher's PhD thesis and may be also be presented at academic conferences. All information you provide will be treated with full confidentiality and, if published, will not be identifiable as yours.

Declaration:

- I am 18 years or older and am competent to provide consent.
- I have read a document providing information about this research. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction.
- I agree that my data is used for scientific purposes and I have no objection that my data is published in scientific publications in a way that does not reveal my identity.
- I understand that I may stop electronic recordings at any time, and that I may at any time, even subsequent to my participation have such recordings destroyed.
- I understand that no recordings will be replayed in any public forum or made available to any audience other than the current researcher.
- I freely and voluntarily agree to be part of this research study, without prejudice to my legal and ethical rights.
- I understand that I may refuse to answer any question and that I may withdraw at any time.
- I understand that my participation is fully anonymous and that no personal details about me will be recorded.
- I have received a copy of this agreement.

Participant's Initials: _______________________ Date: ________________

Appendix 10 – Sample Think-Aloud Test & Post-Test Interview Transcript

This appendix contains the Think-Aloud Test and Post-Test Interview transcript of Participant 3 from Usability Test 1 as an illustration of the experiment process.

Think-Aloud Test – Participant 3

Participant 3 (00:23):
Should I read through all the activities or can I do these one by one?

Interviewer (00:27):
Oh no yeah, one by one is fine. So we'll just start with the first one there.

Participant 3 (00:30):
Okay. You intend to create a number of interlinks that will connect the resources in your BNF dataset to related resources in an authoritative external dataset. You will first need to create an interlink collection, which will contain the links you create. As your dataset is related to James Joyce, the title of interlink collection as can the description. You can use your participant number as creator name. And I'm number three. Okay. Okay. So I'm going to, call this the James Joyce collection and uh, it's going to be a collection of resources relating to the Irish author James Joyce. Okay. And I'm the creator so I presume that I'm going to put that into the creator box. Uh, my organization I going to find, see if I can find it in the drop down and there it is and the creation date, it's auto-filled there with today's date and I'm happy with that. I presume I can get a different date if I need to, yeah. Okay. And now create, very good, I can see it up there and now I'm seeing a list of the, this new resource and uh, the option to create another one. But I'm going to use the manage collection option I think to move on to the next activity, I presume that completes the first activity.

Interviewer (02:07):
Yeah, perfect.

Participant 3 (02:07):
Your dataset contains a resource record for the writer James Joyce, your plan to interlink this resource to a related resource in an external dataset. In order to do this, you must add an internal resource to your collection. This internal resource is the uniform resource identifier that represents the resource for James Joyce in the BNF Dataset that you're working with. You can search the Dataset for the URI using the SemFacet tool. Okay. So I presume I do that by going to manage collection. Okay. I'm going to go into manage collection and I want to try to interlink this resource. Okay. And you plan to interlink this resource to a related resource in an external. Add internal resource. So I'm looking for add internal resource. So plus internal resource. Okay. Use the

SemFacet tool to search the primary dataset for your URI so I'm going to do something like open SemFacet and I'm looking for.

Interviewer (03:12):

Sorry, I'm going to help you here just because it's not always very clear.

Participant 3 (03:14):

SemFacet search. Okay, very good and in SemFacet I'm looking for James Joyce. And I'll search. And I can see that there's a James Joyce looks like the right kind of entity. It's somebody with a, there's a date of birth and death. I presume I, so I can get a copy of the URL. Okay, I presume that copied onto the clipboard? So I'm going to make that presumption and go back in and do Add Internal Resource. And I'm going to paste, I hope. Okay. I'm hoping that that's the right thing. I didn't have any confirmation that that's, but I presume that that's something that was just copied in there. And I'm going to use the label James Joyce and description, Uh, Irish author. And save. Okay. And I presume that's saved too because it's coming back and it's no longer in a form field. Very good, okay. Following this, you must add a related resource that you would like to link to the BNF record for James Joyce to your collection search, the VIAF Virtual International Authority File dataset for a record for James Joyce and add the URI to the interlink collection. So that can be linked to the first resource that I added. So I can do that by, let's see, related resource. So I'm looking for something like related resource. Related resource. Okay, that's fine. And I'm going to search the external authorities and I want VIAF. Okay. Click on that link. And again I'm looking for James Joyce. Okay. So I'm presuming that that's the same James Joyce, it's the same date of birth. There's a second possibility, but I'll, this one seems to be more a richer resource, seems to be greater linkage. Good. So I am presuming now that I need to copy the URL for this resource myself, so I'll do that from the search bar. I don't see any option to paste it in from the, to capture to the clipboard other than doing like that. And then I'll Add Related Resource. Paste that in here. Also call this James Joyce. And to distinguish that. Using the data in the interlink collection, create a link between each pair of resources in the connection. An interlink should describe the relationship between the two resources.

Interviewer (07:09):

So I might just ask you to click on home.

Participant 3 (07:13):

Okay.

Interviewer (07:13):

Because I just prepared, a different collection earlier that has a lot more interlinks. You just click on that one there.

Participant 3 (07:21):

Okay.

Interviewer (07:22):

And yeah, so now you can just go into this collection.

Participant 3 (07:26):

You go into this manage collection. Okay. Manage collection resources. Guide. Do I need a guide? Well, let's see what I'm trying to do. Use the data in the interlink collection. Create and link between each pair of resources in the connection. Okay. Each pair of resources. So this, I'm presuming that the pair is on the left and on the right. Uh, so here's a resource. That's one from VIAF and one from the BNF and I'm going to interlink these two by clicking on Interlink. Um, so clicking on link. I'm guessing that these are identical. They're both authority records for the same person. So I'm going to call them Same As. MMM. Same As. Okay. I think they're going to be Same As. Link term definition. Okay. Two is exactly the same thing. I think that's fine. I'm happy with that. Justification using it for provenance. Okay. Okay. So, uh, label The Dead and then we've got Joyce's Dublin. Okay. So now I'm going to try and link the Dead and a MMM resource in UCD. MMM. I think these are related to? I think it's a See Also. So I'm going to try and link these two. MMM.

Interviewer (15:47):

Perfect.

Participant 3 (15:48):

Okay. Um, save links I presume? Okay. I think that's saved. Very good. And now view and download the interlinks, okay. So I presume I am viewing the interlinks now. Now, I would like to, no, View Interlinks, like that. Okay. Okay. I'm getting an RDF graph. Okay. I only seem to have one resource triple in there, so I'm presuming that's fine. Okay. I'm going to download and I'll take it RDF XML and that seems to download fine. And. I'm not sure if that's the same file I download. So maybe, maybe the, the download name isn't as clear what, what resource or collection and I'm downloading from in the file name that downloads. Uh, now view and download? So I've downloaded the interlinks. It did download, but I just need to identify which is the last download file. View the provenance data describing the interlinks you just created. Okay. View provenance data.

Interviewer (17:52):

There you go.

Participant 3 (17:53):

Okay. Very good. Okay, very good. In the actual. Okay, so now I'm seeing another graph and I can see that the justification, um, fields have been filled out with the information that I've put in and that seems good. Okay. So there I can see that I've created a nice RDF resource. Hopefully helpful to somebody else.

Interviewer (18:42):

Perfect.

Participant 3 (18:42):

Very good

Post-Test Interview – Participant 3

Interviewer (18:43):

Great, so that's everything. Um, so I might just ask you a couple of questions.

Participant 3 (18:51):

Sure.

Interviewer (18:52):

So first, what is your overall impression of the tool?

Participant 3 (18:57):

Um, I think it useful. I can see it's utility, uh, especially in the justification for interlinking. Um, I think it would, the scope notes could probably do with an example. So that, uh, somebody even less familiar than I am with, uh, with RDF terms might be able to kind of get a better idea about what's intended, uh, by a particular interlink or, or what a resource means in the context. Yeah. Um, yeah, I think generally it's quite, uh, uh, usable. Um, I think there were a couple of occasions where I wasn't sure if I completed the task, there wasn't another message or that it would be, again, you might find people kind of repeating an action. Uh, I, I trusted that the action had completed, but I wasn't getting enough feedback. It, it again, I think in terms of where, uh, you had to move to a next step, it would be good if it was sort of a, a notion of a kind of a flow. And I think that was a little bit absent. So, for example, when you moved out of the SemFacet, uh, and to the SemFacet and, and got back in, it was like there was, there was cut and paste that had to, that had to happen and um, it wasn't clear what the flow was to me, it wasn't very clear. I got there because of the activities list that helped me get to it. But if that was rep, that activity kind of flow was represented in some way in the, in the, the interface, I think that would be, yeah, that will be helpful.

Interviewer (20:53):

Yeah. Perfect. Was there anything that you felt worked well, um, that maybe make things easier or was understandable or.

Participant 3 (21:01):

um, I, yeah, I thought so. For example, I think the, um, if, if you have, if it's not, if there isn't some machine to machine interface, at least where you can copy to paste, copy into a clipboard, that's useful. Obviously it would be, it would be great if that just happened in the, in the background. Uh, I, I'm not sure how that could be achieved, but if that was a possibility then that that would be, that would be useful. Um, what worked well? I, yeah, I suppose, um, you, yeah, I, I mean I suppose you'd need to understand what the utility of creating the resources was in order to get there. So again, it would be, it might be a useful thing if you had some sort of visualization or something like that that was part of the, that was part of the, uh, was part of the flow that would kind of give some, some sort of sense about what the activity could achieve, you know.

Interviewer (22:11):

So sort of when you're linking what exactly you're, so like maybe this kind of a graph appearing as you're linking.

Participant 3 (22:18):

Exactly. So that you could kind of see that, that you're actually kind of building towards, uh, to, uh, uh, you know, a rich resource graph and that that would be something that, that might be quite nice to, to, to visualize.

Interviewer (22:31):

Um, was there anything particularly challenging?

Participant 3 (22:35):

Um, I suppose, um, in some of the, again, this is something I've come across in linked data, is that sometimes it's challenging to determine the correct relationship type and it didn't help that I wasn't familiar with the actual resources. I think if I, if I had kinda, if I knew what the source data was about, what it genuinely represents, that I'd be have been quicker to kind of decide on the, on the, the, the relationship type.

Interviewer (23:09):

Um, is there any functions, are there any functions that you think should be added to the tool or any sort of data fields that you feel are missing or.

Participant 3 (23:19):

Well, as I said, I think, I think in the scope notes it would be useful to, to include examples. Um, I think it would be obviously helpful if, um, you didn't have to cut and paste URLs, but again, that could be a big technical challenge and mightn't be achievable in the, in, in a prototype. So I think that would be a nice addition.

Interviewer (23:46):

Yeah. Yeah, for sure. Um, and then what was your impression of the process for selecting the linking terms? How did you find that?

Participant 3 (23:56):

Um, it wasn't too bad. I think there might've been, it might've been useful to kind of have that as in a, in its own, uh, maybe a modal or something like that where you could, where you can kind of again where maybe a scope is there. Uh, something about, you know, about why you might use this relationship terms. So I think that maybe taking that into a modal, might be, might be helpful where you're kind of, that's just the activity that you're doing there and it's a one, two, three type step.

Interviewer (24:33):

Right. So for each link that you're creating, a little explanation as to why you should use, why you would use.

Participant 3 (24:40):

What you might use. Yes. Yeah. Obviously those, um, yeah, I, I think that would be good. And if that was the main focus, I think having it over on the right hand column, you're, it's a sort of a, I think the focus has to be on, on what's going on there. Yeah. Also, yeah. I don't know, uh, if there if there was anything, if you could get something from the original resource in terms of its description or additional metadata that's associated with that original term, or the primary or the secondary term, which would help you determine? Because one of the things I did find is that I wasn't 100% sure what the resources were. I wasn't familiar with them because it was a, a canned example and maybe if I had gone, but maybe if I had gone to the original resource had actually examined it in, in more detail I'd know that immediately, but because I was taken out of the context of where it coming from, I was kind of guessing. Yeah. Yeah.

Interviewer (25:46):

Um, and then what's your impression of the provenance data? Do you feel like it's enough or not enough information?

Participant 3 (25:56):

No, no, I, I think it's good. Um, I think it would be good to, uh, if you could, um, obviously the, the, uh, the label has been overwritten by the thing. So maybe that, that's a little bit, uh, that that's a minor quibble and probably easily addressed. Um.

Interviewer (26:22):

Do you feel if you were looking at someone else's links and you had this information available, would that be sufficient for you to trust their links or would, would there be something else that you'd like to see before you'd kind of ingest their data?

Participant 3 (26:36):

Um, so, uh, potentially you might use, um, uh, you might use a friend of a friend data about the actual author so that you could kind of get some provenance about the provenance, yeah. You know, sort of like what, you know, it's, so if it's coming from, um, if it's a Joyce Scholar, you know, um, yeah, yeah, I'd feel, I, I'd feel I might trust that justification in, in a more, um detailed way.

Interviewer (27:11):

Yeah. So something more than just saying had role cataloguer, it actually might be they're a cataloguer and they're also the James Joyce scholar.

Participant 3 (27:19):

Yeah, exactly. Yeah. So I mean that, that would certainly be an interesting thing to, to, to, to um, include.

Interviewer (27:27):

Perfect. And, um, uh, do you think this tool could be useful for interlinking internal and external resources, RDF resources overall? Or how do you feel about it?

Participant 3 (27:39):

Well, in the, in the context of the library here? Yes, definitely. Um, we, uh, I think in anything like this, it would probably need to be integrated with the cataloguing tool. Like it would, it would be much easier to, um, include linked resources if, if that was done at the time of cataloguing. Rather than subsequently it, you know, I think that that would be a, that would definitely be something that would be useful if it was built into the cataloguing tool.

Interviewer (28:15):

Cool. Great.

Participant 3 (28:16):

I hope that's helpful.

Interviewer (28:18):

Yeah. I just had a quick post-test questionnaire. I might give you that? Okay. Um, and be as brutal as you like because whatever feedback you give me is going to help me.

Participant 3 (28:31):

Okay.

Appendix 11 – Usability Test 1 – Unused Codes

	Code	Description	No. of References	No. of Participants
Unused Codes	Unfamiliarity with Mac	Some participants had difficulty copying, pasting and scrolling due to not being accustomed to the Mac laptop on which the experiment was performed.	26	9
	Concerns and uncertainties around LD	Participants expressed that creating LD requires a lot of resources which may be difficulty to acquire and/or justify.	15	5
	Need to establish an institutional style	Institutions would need to set an interlinking convention in order to ensure interlinking consistency.	10	5
	Standardisation in Museums	Museums have issues regarding data standardisation and use traditional cataloguing methods.	5	2
	Unclear on the level of a resource record	Participant was unclear whether an entity represented a collection or an individual resource.	2	1
	Date Format	US date format was used.	1	1
	LD is useful	Participant expressed that they thought LD was useful for LAMs.	1	1
	Name format for SemFacet	Participant was unsure how to format a name in order to conduct a search in SemFacet.	1	1
	Spellcheck	Participant found spellcheck for data fields useful.	1	1

Appendix 12 – Usability Test 2 Information Sheet & Consent Form

Participant Information Sheet

Who is conducting the research?

The research is conducted by a PhD student from the ADAPT Centre, TCD

PhD Student: Lucy McKenna, ADAPT Centre, Trinity College Dublin

Academic Supervisor: Prof. Declan O'Sullivan, ADAPT Centre, TCD

Assistant Supervisor: Dr. Christophe Debruyne, ADAPT Centre, TCD

What is the aim of the research?

The aim of this research is to gather information on the usability and usefulness of a linked data interlinking tool called NAISC-L.

What is NAISC-L?

NAISC-L (pronounced noshk-el) stands for **N**ovel **A**uthoritative **I**nterlinking for **S**emantic Web **C**ataloguing in **L**ibraries. NAISC is also the Gaelic word for links. NAISC-L is an interlinking model and tool which was developed specifically for the library domain for the creation of linked data interlinks between related internal and external library resources.

Why was I asked to participate?

You have been asked to participate in this research as you have been identified as someone who has experience working as a librarian/library assistant/cataloguer/metadata expert.

What will the participation in research involve?

If you agree to participate in the research you will be asked to complete a number of tasks using the interlinking tool. Following this you will be asked to provide feedback on your experience. It should take approximately 30 minutes to complete the evaluation.

What happens to the information I provide?

The data gathered from this questionnaire will be used as part of the above research project. All information you provide will be treated with full confidentiality and, if published, will not be identifiable as yours.

What if I change my mind about participating in this research?

Your participation in this questionnaire is completely voluntary. You are free to withdraw from the research at any time without any penalty. Should you chose to withdraw from the research, all information you have provided will be deleted.

Any questions?

Please feel free to contact the researcher, Lucy McKenna, at lucy.mckenna@adaptcentre.ie.

Consent Form

Publication: The data gathered from this user evaluation will be used as part of the researcher's PhD thesis and may be also be presented at academic conferences. All information you provide will be treated with full confidentiality and, if published, will not be identifiable as yours.

Declaration:

- I am 18 years or older and am competent to provide consent.
- I have read a document providing information about this research.
- I agree that my data is used for scientific purposes and I have no objection that my data is published in scientific publications in a way that does not reveal my identity.
- I freely and voluntarily agree to be part of this research study, without prejudice to my legal and ethical rights.
- I understand that I may refuse to answer any question and that I may withdraw at any time without penalty.
- I understand that my participation is fully anonymous and that no personal details about me will be recorded.

I consent to participate in this research ☐

I do not consent to participate in this research ☐

Appendix 13 – Usability Test 2 PSSUQ Scores

	Group A - PSSUQ																			SysUse (1-8)	InfoQual (9-15)	InterQual (16-18)	Overall (1-19)
	Overall, I am satisfied with how easy it is to use NAISC-L	It was simple to use NAISC-L	I could effectively complete the tasks using NAISC-L	I was able to complete the tasks and scenarios quickly using NAISC-L	I was able to efficiently complete the tasks using NAISC-L	I felt comfortable using NAISC-L	It was easy to learn to use NAISC-L	I believe I could become productive quickly using NAISC-L	NAISC-L gave error messages that clearly told me how to fix problems	Whenever I made a mistake using NAISC-L, I could recover easily and quickly	The information (such as on-line help, on-screen messages, and other documentation) provided with NAISC-L was clear	It was easy to find the information I needed	The information provided for NAISC-L was easy to understand	The information was effective in helping me complete the tasks	The organisation of information on NAISC-L screens was clear	The interface of NAISC-L was pleasant	I liked using the interface of NAISC-L	NAISC-L has all the functions and capabilities I expect it to have	Overall, I am satisfied with NAISC-L				
	4	3	2	2	3	2	3	3	1		2	2	2	1	4	3	4	7	4	2.75	2.00	4.67	2.89
	4	4	3	5	6	6	4	7	7	5	4	4	4	4	6	3	4	4	4	4.88	4.86	3.67	4.63
	7	7	7	7	1	5	5	6	2	2	2	5	5	5	4	5	5		4	5.63	3.57	5.00	4.67
	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2.00	2.00	2.00	2.00
	5	4	4	6	5	6	5	5	5	5	5	5	5	4	5	4	4	4	4	5.00	4.86	4.00	4.74
	3	4	1	2	2	3	4	1			3	3	3	1	3	3	2	3	3	2.50	2.60	2.67	2.59
	3	2	2	2	2	3	3	2			2	3	2	2	2	4	3	4	2	2.38	2.20	3.67	2.53
	2	2	2	3	3	2	2	2			3	2	3	3	1	1	1	2	2	2.25	2.40	1.33	2.12
	3	2	3	3	3	4	3	4	4	4	3	3	4	3	3	4	4	3	4	3.13	3.43	3.67	3.37
	3	3	5	3	3	3	4	1	3		1	1	1	1	2	1	1	1	3	3.13	1.50	1.00	2.22
	2	2	2	5	2	2	2	4			3	3	3	3	2	2	2	5	2	2.63	2.80	3.00	2.71
	2	2	2	2	2	2	2	2			3	3	2	2	1	3	2	3	2	2.00	2.20	2.67	2.18
	3	3	3	3	3	4	3	3	2	1	2	5	4	4	1	1	2	4	4	3.13	2.71	2.33	2.89
	4	2	3	3	3	5	3	2	4	3	4	5	4	4	5	2	4	3	3	3.13	4.14	3.00	3.47
	2	2	2	2	2	3	3	2			2	3	2	2	2	3	3	2	2	2.25	2.20	2.67	2.29
	4	6		5	5	5	4	5	7	6	6	5	6	3	3	4	4		4	4.25	5.14	4.00	4.82
	2	2	3	2	4	2	1	1		2	2	1	2	1	3	2	3	3	1	2.13	1.83	2.67	2.06
	2	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1.13	1.14	1.00	1.11
	3	3	3	4	4	4	3	3		3	4	3	3	2	2	3	3	4	3	3.38	2.83	3.33	3.17
	5	5	3	4	4	5	3	3	6	4	4	5	6	3	3	3	3	4	4	4.00	4.43	3.33	4.05
	1	1	1	1	1	1	1	2				1	1	1	1	1	1	1	1	1.13	1.00	1.00	1.06
	4	3	4	5	5	4	3	3	3		4	3	6	5	3	2	3	3	4	3.88	4.00	2.67	3.72
	4	5	2	3	3	5	3	3	3	3	4	4	4	3	3	2	3		3	3.50	3.43	2.50	3.33
Average	**3.22**	**3.04**	**2.73**	**3.26**	**3.00**	**3.43**	**2.91**	**2.91**	**3.64**	**3.15**	**3.00**	**3.13**	**3.26**	**2.61**	**2.70**	**2.57**	**2.78**	**3.15**	**2.87**	**3.05**	**2.93**	**2.86**	**2.98**
SD	**1.317**	**1.517**	**1.355**	**1.566**	**1.351**	**1.499**	**1.1**	**1.586**	**1.875**	**1.511**	**1.243**	**1.393**	**1.566**	**1.277**	**1.365**	**1.135**	**1.14**	**1.424**	**1.076**	**1.1401**	**1.18049**	**1.081969**	**1.067316**

	Group B - PSSUQ																			SysUse (1-8)	InfoQual (9-15)	InterQual (16-18)	Overall (1-19)
	Overall, I am satisfied with how easy it is to use NAISC-L	It was simple to use NAISC-L	I could effectively complete the tasks using NAISC-L	I was able to complete the tasks and scenarios quickly using NAISC-L	I was able to efficiently complete the tasks using NAISC-L	I felt comfortable using NAISC-L	It was easy to learn to use NAISC-L	I believe I could become productive quickly using NAISC-L	NAISC-L gave error messages that clearly told me how to fix problems	Whenever I made a mistake using NAISC-L, I could recover easily and quickly	The information (such as on-line help, on-screen messages, and other documentation) provided with NAISC-L was clear	It was easy to find the information I needed	The information provided for NAISC-L was easy to understand	The information was effective in helping me complete the tasks	The organisation of information on NAISC-L screens was clear	The interface of NAISC-L was pleasant	I liked using the interface of NAISC-L	NAISC-L has all the functions and capabilities I expect it to have	Overall, I am satisfied with NAISC-L				
	4	5	4	5	5	3	4	3	6	5	6	5	6	5	6	3	/	/	6	4.13	5.57	3.00	4.76
	5	5	2	5	5	5	4	5	6	2	5	/	6	7	4	4	4	6	5	4.50	5.00	4.67	4.72
	4	5	4	5	4	5	4	2	4	3	5	5	5	5	2	2	2	4	4	4.13	4.14	2.67	3.89
	2	2	1	2	2	2	1	1	/	2	2	2	2	2	1	1	1	3	2	1.63	1.83	1.67	1.72
	3	2	2	2	2	2	1	2	1	1	3	3	3	2	2	3	2	4	2	2.00	2.14	3.00	2.33
	4	/	4	3	4	3	5	3	5	4	5	5	5	4	4	3	3	5	5	3.71	4.57	3.67	4.11
	1	1	1	1	1	1	1	1	3	/	3	4	3	2	3	2	2	2	1	1.00	3.00	2.00	1.83
	5	/	5	/	4	/	4	4	4	5	4	5	4	4	4	4	4	4	4	4.40	4.29	4.00	4.25
	2	2	2	2	2	2	2	2	4	3	2	2	2	2	1	1	1	1	1	2.00	2.29	1.00	1.89
	4	4	7	7	7	7	6	3	5	7	6	7	6	7	2	5	4	6	6	5.63	5.71	5.00	5.58
	1	1	2	2	2	1	1	1	/	/	2	2	2	2	1	1	1	1	2	1.38	1.80	1.00	1.47
	1	1	1	1	1	3	1	1	1	1	3	3	2	2	1	1	1	/	1	1.25	1.86	1.00	1.44
	2	2	2	2	2	2	2	2	/	/	1	1	2	2	1	1	1	1	1	2.00	1.40	1.00	1.59
	5	5	4	4	4	4	4	4	7	4	5	5	4	4	5	4	4	4	4	4.25	4.86	4.00	4.42
	4	4	3	3	4	3	5	3	4	5	5	5	4	4	4	4	3	5	4	3.63	4.43	4.00	4.00
	5	4	5	5	5	7	4	6	7	7	5	5	5	5	6	5	5	6	5	5.13	5.71	5.33	5.37
	3	3	3	3	3	3	3	3	/	/	5	5	5	5	5	3	3	4	4	3.00	5.00	3.33	3.71
	1	1	1	1	1	3	2	1	2	3	4	3	3	3	3	1	1	3	2	1.38	3.00	1.67	2.05
	3	2	2	5	3	2	3	4	/	/	3	3	3	3	2	1	1	3	3	3.00	2.80	1.67	2.71
	5	4	6	5	5	2	2	4	2	1	5	6	6	4	3	2	2	5	6	4.13	3.86	3.00	3.95
	2	2	2	3	3	3	3	2	/	/	4	4	3	3	3	3	3	4	2	2.50	3.40	3.33	2.88
	4	4	1	6	5	5	1	6	3	4	6	6	4	3	5	7	7	6	6	4.00	4.43	6.67	4.68
	5	5	4	5	5	5	4	4	/	/	2	3	4	4	4	4	4	/	4	4.63	3.40	4.00	4.13
	3	3	3	3	3	3	3	3	/	/	3	3	3	3	3	3	3	3	3	3.00	3.00	3.00	3.00
	5	/	7	7	7	7	3	2	7	/	7	7	5	7	1	1	1	/	5	5.43	5.67	1.00	4.94
Average	**3.32**	**3.05**	**3.12**	**3.63**	**3.56**	**3.46**	**2.92**	**2.88**	**4.18**	**3.56**	**4.04**	**4.13**	**3.88**	**3.76**	**3.04**	**2.76**	**2.63**	**3.81**	**3.52**	**3.27**	**3.73**	**2.99**	**3.42**
SD	**1.434**	**1.461**	**1.796**	**1.798**	**1.675**	**1.755**	**1.44**	**1.451**	**1.947**	**1.87**	**1.536**	**1.589**	**1.366**	**1.582**	**1.587**	**1.582**	**1.55**	**1.592**	**1.7**	**1.371914**	**1.345166**	**1.506962**	**1.300159**

	Group C - PSSUQ																			SysUse (1-8)	InfoQual (9-15)	InterQual (16-18)	Overall (1-19)
	Overall, I am satisfied with how easy it is to use NAISC-L	It was simple to use NAISC-L	I could effectively complete the tasks using NAISC-L	I was able to complete the tasks and scenarios quickly using NAISC-L	I was able to efficiently complete the tasks using NAISC-L	I felt comfortable using NAISC-L	It was easy to learn to use NAISC-L	I believe I could become productive quickly using NAISC-L	NAISC-L gave error messages that clearly told me how to fix problems	Whenever I made a mistake using NAISC-L, I could recover easily and quickly	The information (such as on-line help, on-screen messages, and other documentation) provided with NAISC-L was clear	It was easy to find the information I needed	The information provided for NAISC-L was easy to understand	The information was effective in helping me complete the tasks	The organisation of information on NAISC-L screens was clear	The interface of NAISC-L was pleasant	I liked using the interface of NAISC-L	NAISC-L has all the functions and capabilities I expect it to have	Overall, I am satisfied with NAISC-L				
	2	2	2	2	3	3	2	2	2	3	1	1	1	2	1	1	1	2	2	2.25	1.57	1.33	1.84
	2	2	2	2	2	2	2	3	3	3	3	3	3	3	2	2	2	3	2	2.13	2.86	2.33	2.42
	1	1	1	1	1	1	1	1			1	1	1	1	1	1	1	1	1	1.00	1.00	1.00	1.00
	3	3	4	5	4	3	3	3	2		2	4	4	3	6	2	2		3	3.50	3.50	2.00	3.29
	3	4		4	4	2	2	3		2	3	4	4	4	3	3	2	5	3	3.14	3.33	3.33	3.24
	5	5	5	6	6	5	4	3	7	6	6	6	5	5	4	4	4	5	6	4.88	5.57	4.33	5.11
	1	2	2	2	2	2		1	4	3	2	3	2	2	3	3	4	5	2	1.71	2.71	4.00	2.50
	2	2	2	3	3	4	3	3			3	3	3	2	3	4	4	4	4	2.75	2.80	4.00	3.06
	2	2	2	2	2	2	2	2	1	1	2	7	2	2	1	1	1	1	1	2.00	2.29	1.00	1.89
	1	1	7	3	3	2	1	2	6	7	2	3	2	2	1	1	1	3	3	2.50	3.29	1.67	2.68
	2	2	6	1	2	1	1	2			4	3	4	4	1	3	3	2	2	2.13	3.20	2.67	2.53
	1	1		1	1	1	2	2		1	1		1	1	1	1	1	1	1	1.29	1.00	1.00	1.13
	3	3	4	3	4	4	2	2	2	2	2	3	4	3	3	3	2	2	2	3.13	2.71	2.33	2.79
	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4.00	4.00	4.00	4.00
	3	2	1	2	1	1	2	1			1	2	1	1	2	2	2	1	1	1.63	1.00	1.67	1.53
	3	3	4	3	3	4	4	4	5	5	4	4	4	4	4	3	3	3	4	3.50	4.29	3.00	3.74
	3	3	2	4	3	5	3	1			4	4	3	4	5	2	5	5	4	3.00	4.00	4.00	3.53
	3		2		2	3	3	3			3	3	3	3	2	2	2	2	2	2.67	2.80	2.00	2.53
	3		4		2		3				2		4		1		2	3	3	3.00	2.33	2.50	2.70
	3	3	2	3	3	3	3	1	2		3	3	3	3	3	3	3		3	2.63	2.83	3.00	2.76
	6	6	7	6	6	5	6	5	5	5	6	6	5	5	5	4	4	5	5	5.88	5.29	4.33	5.37
	6	2	5	7	3								5	7	4	4	4			3.83	5.33	4.00	4.70
Average	**2.82**	**2.65**	**3.40**	**3.20**	**2.91**	**2.85**	**2.65**	**2.40**	**3.58**	**3.50**	**2.81**	**3.53**	**3.09**	**3.10**	**2.73**	**2.52**	**2.59**	**3.00**	**2.76**	**2.84**	**3.08**	**2.70**	**2.92**
SD	**1.402**	**1.276**	**1.828**	**1.691**	**1.345**	**1.352**	**1.195**	**1.114**	**1.801**	**1.848**	**1.435**	**1.5**	**1.311**	**1.477**	**1.513**	**1.096**	**1.231**	**1.487**	**1.342**	**1.120086**	**1.289323**	**1.134772**	**1.13477**

	Group D - PSSUQ																			SysUse (1-8)	InfoQual (9-15)	InterQual (16-18)	Overall (1-19)
	Overall, I am satisfied with how easy it is to use NAISC-L	It was simple to use NAISC-L	I could effectively complete the tasks using NAISC-L	I was able to complete the tasks and scenarios quickly using NAISC-L	I was able to efficiently complete the tasks using NAISC-L	I felt comfortable using NAISC-L	It was easy to learn to use NAISC-L	I believe I could become productive quickly using NAISC-L	NAISC-L gave error messages that clearly told me how to fix problems	Whenever I made a mistake using NAISC-L, I could recover easily and quickly	The information (such as on-line help, on-screen messages, and other documentation) provided with NAISC-L was clear	It was easy to find the information I needed	The information provided for NAISC-L was easy to understand	The information was effective in helping me complete the tasks	The organisation of information on NAISC-L screens was clear	The interface of NAISC-L was pleasant	I liked using the interface of NAISC-L	NAISC-L has all the functions and capabilities I expect it to have	Overall, I am satisfied with NAISC-L				
	4	4	4	5	4	5	2	2			3	3	3	3	3	3	3	3	3	3.75	3.00	3.00	3.35
	1	1	1	1	1	1	1	1	4	1	1	1	1	1	1	1	1	1	1	1.00	1.43	1.00	1.16
	5	5	5	3	3	4	3	4	6		2	3	4	4	4	3	3	4	4	4.00	3.83	3.33	3.83
	7	7	3	4	7	7	4	7		2	5	7	5	7	4	4	7	7	7	5.75	5.00	6.00	5.61
	3	5	4	6	2	2	3	3			6	5	4	3	3	1	1	1	2	3.50	4.20	1.00	3.18
	4	5	5	5	4	6	5	5	6	6	3	4	2	2	5	5	4		5	4.88	4.00	4.50	4.50
	5	5	5	5	5	5	5	4		7	3	5	3	7	5	4	6	5	5	4.88	5.00	5.00	4.94
	3	3	3	4	4	5	5	1	7	3	4	4	4	4	1	4	2	3	4	3.50	3.86	3.00	3.58
	3	4	3	3	4	3	3	2	6	5	4	5	3	3	3	3	2	4	3	3.13	4.14	3.00	3.47
	4	4	4	4	4	4	4	4	4	3	4	4	4	4	3	3	3	5	4	4.00	3.71	3.67	3.84
	6	1	3	2	3	4	2	5				4	5	5	2	2	5	7	5	3.25	4.00	4.67	3.81
	5	2	2	2	2		2	4		4	6	6	1	4	2	2	2	5	5	2.71	3.83	3.00	3.29
	4	2	4	4	4	6	3	3	3	3	4	4	4	4	4	3	4		4	3.75	3.71	3.50	3.72
	6	6	6	6	6	6	6	4	6	5	5	5	5	5	7	6	6	6	6	5.75	5.43	6.00	5.68
	4	4	4	4	4	5	4	2		2	5	6	6	7	4	2	4		5	3.88	5.00	3.00	4.24
	2	2	3	4	4	3	3	2			2	2	2	2	2	1	1	2	2	2.88	2.00	1.33	2.29
	3	1	2	1	2	3	3	1	1	3	1	3	3	3	3	1	1	3	2	2.00	2.43	1.67	2.11
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.00	1.00	1.00	1.00
	6	6	6	6	6	6	6	6			6	6	6	6	6	6	6	6	6	6.00	6.00	6.00	6.00
	4	1	4	1	1	1	1	4	4	4	4	4	4	4	5	6	7	4	4	2.13	4.14	5.67	3.53
	2	1	2	4	4	2	1	2	7	3	2	2	2	2	3	2	2	1	2	2.25	3.00	1.67	2.42
	1	2	1	1	1	2	2	2	1	4	2	3	2	2	1	1	1	3	2	1.50	2.14	1.67	1.79
	3	4	5	6	6	6	6	4	6	6	5	4	4	5	6	5	4	4	4	5.00	5.14	4.33	4.89
	4	5	3	3	5	5	3	2	3	3	3	4	3	4	2	2	2	4	3	3.75	3.14	2.67	3.32
	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4	3	3	2.88	3.00	3.67	3.05
	6	7	7	7	7	6	6	5			4	4	4	6	6	4	4	6	6	6.38	4.80	4.67	5.59
Average	**3.81**	**3.46**	**3.58**	**3.65**	**3.73**	**4.04**	**3.35**	**3.19**	**3.78**	**3.24**	**3.52**	**3.92**	**3.38**	**3.88**	**3.42**	**3.04**	**3.31**	**3.83**	**3.77**	**3.59**	**3.73**	**3.38**	**3.62**
SD	**1.618**	**1.946**	**1.549**	**1.775**	**1.766**	**1.8**	**1.616**	**1.594**	**2.046**	**1.6**	**1.526**	**1.466**	**1.389**	**1.717**	**1.691**	**1.629**	**1.917**	**1.833**	**1.601**	**1.455378**	**1.219425**	**1.585032**	**1.315961**

Appendix 14 – Usability Test 2 DQ Scores

	Group A - DQ																									
	The data is of sufficient volume for my organisation's needs	The amount of data does NOT match my organisation's needs	The data is believable	The data is of doubtful credibility	The data is trustworthy	The data includes all necessary values	The data is INcomplete	The data is sufficiently complete for my organisation's need	The data is presented concisely	The data is presented in a compact form	The data is presented consistently	This information is easy to manipulate to meet my organisation's needs	The data is easy to combine with other data	The data is INcorrect	The data is accurate	The data is reliable	It is easy to interpret what the data means	The data is objective	The data is useful to my organisation's work	The data is relevant to my organisation's work	The data comes from good sources	The data is NOT sufficiently timely	The data is sufficiently up-to-date for my organisation's work	The data is easy to comprehend	The meaning of the data is difficult to understand	Overall
	0	10	10	10	9	5	5		7	10	5	5	5	10	10	5	2	5	5	2	8	5		4	5	6.17
	0	1	7	8	7	5	5	2	8	8	7	2	7	8	7	8	6	7	6	3	7	7	7	7	6	5.84
	6	5	6	4	6	6	7	4	6	7	7	5	5	7	6	6	7	3	6	6	7	7	5	7	3	5.76
	0	10	8	8	8	0	10	8	8	8	8	0	0	10	8	8	8	8	8	8	8	10	7	7	8	6.96
	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5.00
	7	2	6	7	7	7	8	7	7	7	7	6		8	7	6	6	5	5	5	6	6	7	7	5	6.29
	1	1	5	5	5	2	0	1	5	2	5	5	8	7	8	5	10	8	3	5	8	5	5	8	7	4.96
	7	6	8	2	8	7	6	7	9	9	9	7	8	6	8	8	7	8	6	6	8	5	5	8	6	6.96
	5	5	7	8	7	7	5	8	4	4	7	6	4	9	8	8	4	8	6	7	8	7	7	4	3	6.24
	10	8	9	9	9	9	10	10	10	10	10		10	10	10	8	9	1	10	10	10	8	10	10	10	9.17
	0	10	10	7	8	5	5		4	4	10		7	10	10	10	7	0			10	5		7	6	6.75
	8	9	9	10	9	9	9	9	10	10	10	9	10	10	10	10	7	5	8	8	10	9	9	7	8	8.88
	8	8	8	9	8	8	8	8	8	8	8		4	8	8	7	7	8	7	7	8	1	9	6	8	7.38
	0	10	8	5	5	5	5		8	8	8		7	5	5	5	4				6			3	3	5.56
	8	8	9	9	8	8	8	8	7	7	8	8	8	9	8	8	8	6	8	8	8	8	8	8	8	7.96
	5	5	8	5	5	5	5	5	5	5	5	5	5	5	5	5	3	5	5	5	5	5	5	5	5	5.48
	7	9	9	10	9	9	9	7	8	7	9		7	9	8	8	8	8		8	9	10	8	8	8	8.35
	5	5	5	5	5	5	5	5	10	10	10	5	5	5	5	10	5	5	5	5	5	5	5	5	5	5.80
	0	10	10	10	10	5	5		7	7	7		5	10	10	10	9	9			10			9	8	7.95
	5	5	7	7	7	5	5	5	7	7	7	5	5	7	7	7	5	5	5	5	5	5	5	5	5	5.72
	5	5	10	10	10	10	10	5	9	9	10	5	5	10	10	10	10	10	5	5	10	10	5	10	10	8.32
	0	10	7	7	7	5	5		8	8	8	8			5	7	4	5	8	8	7	7	7	7	7	6.59
	6	7	7	8	7	7	8	5	7	7	7	5	5	8	7	7	5	3	5	5	8	8	8	6	7	6.52
Average	**4.26**	**6.70**	**7.74**	**7.30**	**7.35**	**6.04**	**6.43**	**6.06**	**7.26**	**7.26**	**7.70**	**5.35**	**5.95**	**8.00**	**7.61**	**7.43**	**6.35**	**5.77**	**6.11**	**6.05**	**7.65**	**6.57**	**6.68**	**6.65**	**6.35**	**6.67**
SD	**3.26**	**2.87**	**1.59**	**2.22**	**1.58**	**2.26**	**2.34**	**2.30**	**1.75**	**2.07**	**1.65**	**2.06**	**2.19**	**1.83**	**1.79**	**1.74**	**2.14**	**2.49**	**1.62**	**1.88**	**1.68**	**2.17**	**1.62**	**1.81**	**1.97**	**1.19**

	Group B - DQ																									
	The data is of sufficient volume for my organisation's needs	The amount of data does NOT match my organisation's needs	The data is believable	The data is of doubtful credibility	The data is trustworthy	The data includes all necessary values	The data is INcomplete	The data is sufficiently complete for my organisation's need	The data is presented concisely	The data is presented in a compact form	The data is presented consistently	This information is easy to manipulate to meet my organisation's needs	The data is easy to combine with other data	The data is INcorrect	The data is accurate	The data is reliable	It is easy to interpret what the data means	The data is objective	The data is useful to my organisation's work	The data is relevant to my organisation's work	The data comes from good sources	The data is NOT sufficiently timely	The data is sufficiently up-to-date for my organisation's work	The data is easy to comprehend	The meaning of the data is difficult to understand	
																										Overall
	10	5				10	8	9	5	6	6		5		6		8		5	5	9			8	8	**7.53**
	10	10	10	10	10	10	10	10	3	3	0	10	10	10	10	10	10		10	10		10	10	10	10	**8.58**
	3	3	7	7	7	3	8	7	4	6	8	8	10	9	8	8	4	5	7	7	7		8	8	8	**6.67**
	5	5	7	10	5	5	5	5	8	8	8	5	5	5	5	5	5	5	5	5	5	5	5	5	5	**5.64**
	8	9	8	9	8	5	5	5	8	8	8	5		9	8	8	8	5	8	8	8	9	8	8	9	**7.58**
	0	0	5	5	5	8	0	0	9	9	9	5	9	8	5	5	8	8	1	8	8	5	8	8	8	**5.76**
	5	5	10	10	10	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	**5.60**
	5	9	7	3	7	7	7	5	5	6	6	5	8	8	8	7	7	7	5	7	8	10	8	7	8	**6.80**
	5	5	6	5	6	6	8	5	6	6	8	8	9	8	7	7	7	6	5	5	7	5	5	6	8	**6.36**
	6	8	6	8	6	6	8	6	6	7	7	3	2	9	6	6	6	6	6	6	8	9	8	7	8	**6.56**
	5	5	10	10	10	5	5	5	9	9	9	5	10	5	10	10	8	10	10	10	10	5	9	9	5	**7.92**
	5	5	8	8	2	5	5	5	8	8	8	5	5	5	5	7	5	7	5	5	5	5	5	5	5	**5.64**
	5	5	7	9	7	7	9	5	7	6	7	5		9	8	8	7	7	5		7			7	8	**6.90**
	5	5	5	5	6	6	4	5	5	5	5	5	5	5	5	5	6	5	0	0	6	5	0	6	7	**4.64**
	5	5	8	9	8	8	9	5	7	7	7	7	7	5	7	7	7	7	5	7	8	5	5	8	6	**6.76**
	3	2	9				10	5	6	7	5	2	2	10	5	6	5	4	3	3	6	5	5	3	2	**4.70**
			6	10	6				7	7	7	7	7	10	7	7	7	3	7	7	7			7	7	**7.00**
	3	1	9	8	9	2	1	3	6	2	10	3	10	8	9	9	3	9	9	9	9	9	6	7	8	**6.48**
	5	5	10	6	6	2	1	3	9	7	7	7	5	7	7	7	7	5	6	5	10	5	5	9	2	**5.92**
	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	**5.00**
	7	9	8	10	8	8	9	7	8	8	8	5	6	10	8	8	8	8	8	8	8	5	8	8	9	**7.88**
	5	5	5	5	5	5	5	5	5	6	6	5	5	5	5	5	1	5	4	4	5	5	5	1	0	**4.48**
			5	5	5	5	5		6	6	5		5	5	5	5	6	5			5	5		6	6	**5.00**
			5	5	5					8	8	5	9				9				5	9		8	8	**7.00**
	5	5	10	10	5	5	5	5	10	10	10				10	10	0	10	10	10	10	5	5	5	2	**7.14**
Average	**5.23**	**5.27**	**7.33**	**7.48**	**6.57**	**5.82**	**5.96**	**5.23**	**6.54**	**6.60**	**6.88**	**5.45**	**6.26**	**7.27**	**6.83**	**6.96**	**6.08**	**6.23**	**5.83**	**6.32**	**7.13**	**6.24**	**6.15**	**6.64**	**6.28**	**6.34**
SD	**2.13**	**2.47**	**1.84**	**2.26**	**1.95**	**2.08**	**2.76**	**1.93**	**1.76**	**1.77**	**2.05**	**1.75**	**2.44**	**2.03**	**1.75**	**1.68**	**2.28**	**1.83**	**2.55**	**2.40**	**1.72**	**1.97**	**2.17**	**1.96**	**2.52**	**1.09**

	Group C - DQ																									
	The data is of sufficient volume for my organisation's needs	The amount of data does NOT match my organisation's needs	The data is believable	The data is of doubtful credibility	The data is trustworthy	The data includes all necessary values	The data is INcomplete	The data is sufficiently complete for my organisation's need	The data is presented concisely	The data is presented in a compact form	The data is presented consistently	This information is easy to manipulate to meet my organisation's needs	The data is easy to combine with other data	The data is INcorrect	The data is accurate	The data is reliable	It is easy to interpret what the data means	The data is objective	The data is useful to my organisation's work	The data is relevant to my organisation's work	The data comes from good sources	The data is NOT sufficiently timely	The data is sufficiently up-to-date for my organisation's work	The data is easy to comprehend	The meaning of the data is difficult to understand	Overall
	7	5	7	10	10	10	10	7	7	7	7	7	6	10	8	8	7	8	7	7	9	4	5	7	5	7.40
	5	5	5	5	6	6	10	6	6	6	6	6	6	10	6	6	6	6	6	6		3	7	7	5	6.08
	6	6	10	9	10	9	3	3	9	9	9	9	9	9	10	10	3	3	9	9	9		8	8	3	7.58
	5	5	8	7	8	5	9	5	9	9	9			9	5	8	8	4	5	5	8			8	8	7.00
	5	5	0	3	3	0									0	2	3	0	5	10	9	5	5	2	2	3.47
	5	5	5	10	8	5	5	5	6	6	5	5	5	5	7	7	1	5	5	5	9	5	5	4	3	5.44
	0	0	7	8	5	2	2	0	7	6	6	5	7	8	5	5	7	5	2	1	5	5	10	6	9	4.92
	5	5	7	8	7	5	5	5	7	7	7	5	5	8	7	7	8	5	5	5	8	5	5	10		6.29
	9	1	9	9	9	9	8	7	9	9	9	8	9	9	0	0	0	0	0	0	0	10	8	9	4	5.80
	8	8	8	8	8	3	3	8	9	9	9	0	8	0	0	0	9	4	10	10	9	8	9	10	10	6.72
				3	5	5	5		5	5	5		5	5	5	5	5	5			5	5		5	5	4.61
	5	2	10	10	10	10	10	10	9	10	10	9	9	10	10	10	10	7	8	8	9	8	8	8	8	8.72
	7	8	8	9	8	5	9	7	7	7	7	7	7	9	5	7	5	5	7	7	5	9	7	7	9	7.12
	5	5	6	5	5	5	5	5	5	6	5	5	5	5	5	5	5	5	5	5	6	5	5	5	5	5.12
	9	7	10	10	10	10	10	10	9	9	10	9	9	10	9	10	10	10	10	10	10	10	9	9	9	9.52
	4	4	10	10	10	4	5	4	6	6	5	4	6	10	10	8	7	5	6	6	8	5	4	6	6	6.36
	5	5	4	4	4	5	5	4	6	6	6	7	7	6	4	4	6	2	5	5	5	5	5	4	4	4.92
			10	5	10	5	5		10	10	10			10	10	10	10	4			7	5	5	7	7	7.78
	7	5		4	8		0		10		9		5		10		6			6	5	0	5		2	5.13
			5	5	5	5	5		10	10	10	10	10	10	10	10	10	5			5	5	5	10	10	7.75
	5	5	6	5	5	4	5	4	6	6	6	6	6	5	5	5	3	5	3	3	5	5	4	3	3	4.72
			10	10	10	10	10										10				10					10.00
Average	**5.67**	**4.78**	**7.25**	**7.14**	**7.45**	**5.81**	**6.14**	**5.63**	**7.60**	**7.53**	**7.50**	**6.38**	**6.89**	**7.79**	**6.24**	**6.35**	**6.32**	**4.65**	**5.76**	**6.00**	**6.95**	**5.63**	**6.26**	**6.75**	**5.85**	**6.41**
SD	**2.00**	**2.02**	**2.57**	**2.51**	**2.27**	**2.75**	**2.93**	**2.47**	**1.69**	**1.67**	**1.88**	**2.39**	**1.66**	**2.67**	**3.29**	**3.07**	**2.93**	**2.26**	**2.58**	**2.77**	**2.42**	**2.37**	**1.83**	**2.30**	**2.63**	**1.63**

	Group D - DQ																									
	The data is of sufficient volume for my organisation's needs	The amount of data does NOT match my organisation's needs	The data is believable	The data is of doubtful credibility	The data is trustworthy	The data includes all necessary values	The data is INcomplete	The data is sufficiently complete for my organisation's need	The data is presented concisely	The data is presented in a compact form	The data is presented consistently	This information is easy to manipulate to meet my organisation's needs	The data is easy to combine with other data	The data is INcorrect	The data is accurate	The data is reliable	It is easy to interpret what the data means	The data is objective	The data is useful to my organisation's work	The data is relevant to my organisation's work	The data comes from good sources	The data is NOT sufficiently timely	The data is sufficiently up-to-date for my organisation's work	The data is easy to comprehend	The meaning of the data is difficult to understand	
																										Overall
	0	10	8	8	8	0	10	0	7	7	7	0	0	10	7	7	7	7	0	0	9	8	0	8	9	5.48
	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10.00
	6	5	6	7	5	5	5	6	5	6	6	6	6	5	5	5	6	5	5	5	7	5	5	5	5	5.48
	8	10	10	10	7	5	10	8	10	10	10	8	7	10	8	8	8	0	7	8		10		10	10	8.00
	8	7	10	10	10	8	9	10	10	10	10	10	5	9	8	10	10	10	6	6	10	10	10	10	10	9.04
	5	10	8	10	5	5	5	10	8	10	10	5	8	10	5	7	3	5	10	10	9	10	4	5	3	7.20
	5	5				0		5	10	10	10	10	10				10	10	8	8	5	2	8	10	10	7.56
	10	10	10	10	10	5	5	5	10	10	5	9	10	10	10	5	10	10	10	10	10	10	10	10	10	8.96
	3	4	6	7	6	6	7	4	6	7	7	4	7	7	6	6	7	7	7	7	3	7	6	6	7	6.00
	7	7	7	8	7	5	5	7	4	4	5	5	5	5	7	7	3	5	9	9	9	5	7	5	5	6.08
	0	0	10	10	10	0	10	0	10	10	10	0	10	10	10	10	10	5	0	10	9	5	5	10	10	6.96
	5	5	10	10	10	5	5	5	5	8	8	7	7	10	10	10	10	5	5	5	5	5	5	10	0	6.80
	5	5	10	8	7	5	3	3	5	5	1	7	5	5	5	5	3	5	5	5	8	5	5	7	7	5.36
	7	8	7	8	7	5	5	6	7	7	6	5	6	5	6	6	7	5	6	6	8	5	6	7	6	6.28
	5	5	8	7	7	5	7								5	5	7	5	6	6	6	7	7	7	3	6.00
			8	8	8	9	5	5	8	8	10	5	6	8	7	7	5	5	7	7	8	5	5	7	8	6.91
			7	5	5	5	5		8	8	8			5	5	5	7	5			5			8	8	6.19
			10	10	10	5	5		10	10	10			10	10	10	10		10	10	10	10	10	10	10	9.47
	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5.00
				10	0				2	2	2			10	10	10	0	5			10			0	0	4.69
			7	8	8	8	7		8	8	8		9		5	7	7	3			5			8	8	7.13
	6	4	8	9	8	7	9	7	9	9	9	9	7	9	7	7	7	7	7	7	5	8	6	8	8	7.48
	8	10	10	10	10	5	5	10	10	10	10	5	8	10	10	10	8	5	8	1	9	5	5	10	9	8.04
	5	5	8	9	5	9	3	5	9	8	8	5	5	5	7	7	8	5	5	5	10	5	5	8	8	6.48
	5	5	7	7	7	5	5	5	7	7	7	7	5	5	7	7	7	7	5	5	7	5	5	7	5	6.04
	1	9	10	10	9	3	5	1	6	5			5	10	10	10	10	5	5	5	8	5	5	5	5	6.39
Average	**5.43**	**6.62**	**8.33**	**8.56**	**7.36**	**5.20**	**6.25**	**5.32**	**7.56**	**7.76**	**7.58**	**6.10**	**6.64**	**7.96**	**7.40**	**7.44**	**7.12**	**5.84**	**6.35**	**6.52**	**7.60**	**6.61**	**6.09**	**7.54**	**6.88**	**6.88**
SD	**2.70**	**2.72**	**1.57**	**1.55**	**2.35**	**2.51**	**2.18**	**2.95**	**2.26**	**2.21**	**2.55**	**2.79**	**2.33**	**2.27**	**2.00**	**1.94**	**2.64**	**2.26**	**2.65**	**2.65**	**2.10**	**2.34**	**2.33**	**2.39**	**2.95**	**1.35**

Appendix 15 – Field Test Information Sheet & Consent Form

Participant Information Sheet

Who is conducting the research?

PhD Student: Lucy McKenna, ADAPT Centre, Trinity College Dublin
Academic Supervisor: Prof. Declan O'Sullivan, ADAPT Centre, TCD
Assistant Supervisor: Dr. Christophe Debruyne, ADAPT Centre, TCD

What is the aim of the research?

The aim of this research is to gather information on the usability and usefulness of a linked data interlinking tool called NAISC-L-L.

What is NAISC-L?

NAISC-L (pronounced noshk-el) stands for **N**ovel **A**uthoritative **I**nterlinking for **S**emantic Web **C**ataloguing in **L**ibraries. NAISC-L is also the Gaelic word for links. NAISC-L is an interlinking model and tool which was developed specifically for the library domain for the creation of linked data interlinks between related internal and external library resources.

Why was I asked to participate?

You have been asked to participate in this research as you have been identified as someone who has experience working as a librarian/library assistant/cataloguer/metadata expert.

What will the participation in research involve?

If you agree to participate in the research you will be asked to use a linked data interlinking tool for 8 working days and to maintain a feedback log documenting your experience using the tool. Following the field test period, you will be asked to participate in a post-test interview which will investigate your experience of using the tool. With your permission, this interview will be recorded and should take approximately 30 minutes of your time.

What happens to the information I provide?

The data gathered from this evaluation will be used as part of the above research project. All information you provide will be treated with full confidentiality and, if published, will not be identifiable as yours. Audio recordings will not be made available to anyone other than the researcher and will not be replayed in any public forum or presentation of the research.

What if I change my mind about participating in this research?

Your participation in this research is completely voluntary. You are free to withdraw from the research at any time without any penalty. Should you chose to withdraw from the research, all information you have provided will be deleted.

Any questions? Please contact the researcher at lucy.mckenna@adaptcentre.ie.

Consent Form – Field Test

Researcher: Lucy McKenna (lucy.mckenna@adaptcentre.ie)

Background: The aim of this research is to gather information on the usability and usefulness of a linked data interlinking tool developed for the library domain.

Procedure: If you agree to participate in the research you will be asked to complete a post-test interview which will take approximately 20-30 minutes of your time, and to complete a brief usability questionnaire.

Publication: The data gathered from this questionnaire will be used as part of the researcher's PhD thesis and may be also be presented at academic conferences. All information you provide will be treated with full confidentiality and, if published, will not be identifiable as yours.

Declaration:

- I am 18 years or older and am competent to provide consent.
- I have read a document providing information about this research. I have had the opportunity to ask questions and all my questions have been answered.
- I agree that my data is used for scientific purposes and that any summary interview content or direct quotations from the interview, that are made available through academic publication or other

academic outlets, will be anonymised so that I cannot be identified.

- I understand that care will be taken to ensure that other information in the interview that could identify me, such as the names of third parties will not be revealed and will be anonymized if used in any research publications.
- I understand that if I make illicit activities known, these will be reported to appropriate authorities.
- I understand that I may stop electronic recordings at any time, and that I may at any time, even subsequent to my participation, have such recordings destroyed.
- I understand that no recordings will be replayed in any public forum or made available to any audience other than the current researcher.
- I freely and voluntarily agree to be part of this research study, without prejudice to my legal and ethical rights.
- I understand that I may refuse to answer any question and that I may withdraw at any time without penalty.

Participant's Initials: ___________________ Date: _________________

Appendix 16 – Field Test Interview Sample Transcript

This appendix contains the Field Test Interview transcript of Participant 1 from Usability Test 3 as an illustration of the experiment process.

Participant 1

Interviewer (00:01):

The first question I'm going to ask is what was your overall impression of the tool and the experience of using it?

Participant 1 (00:08):

Um, I thought it was very good. I thought it was really useful. I could see huge opportunities that we could do things with it. Um, a very good way of creating, of doing what it's supposed to do, creating the interlinks between things. Um, there was, uh, yeah, it was, it was, uh, it was a very positive impression.

Interviewer (00:30):

And what did you feel worked well? What was like the parts that you like, liked the most that you felt.

Participant 1 (00:34):

Um, being able to create the links and then show them graphically. That was really good. Um, the, once you had the interlinks created, you know, when you got the two sets on either side, the bit in between, I thought that was really good. Um, giving you all the different options. Everything I had was Owl:sameAs because that is what we were doing, but, um, there was, I could see that there was huge scope there. And the way that, once you picked your first one, I can't remember the terminology now. When you picked your first and then you got the different options then depending on what you would pick the first time made it, made it very kind of straight forward. Because that's really, that's a really difficult thing to get your, for people to get their head around so that was, um, uh, that was, uh, displayed really well.

Interviewer (01:21):

Okay. Um, were there any particular challenges that you experienced?

Participant 1 (01:26):

Um, there was a bits in, in, in just in terms of logging on and getting usernames. It was a little bit of errors and stuff coming up and you'd go back in now and again, it's probably just to do with the web server. Um, uh, the, um, there was a couple, there was a couple of little things probably will come up as we go through, um, the, the very first one, when you don't have any links there and you go to add new, everybody is clicking on the little picture of the add new button. Everyone click on the picture, they don't click on the three dots. Um, the other one then is when you're, when you have the, when you, you know, you got the FRBR entities down the bottom and you have choosen something and then underneath that is your save button. But a couple of times I clicked the choose button first because the save is actually under the fold. So you don't see it on the screen. On my screen anyway you've got to scroll down to see it. Um, so I was clicking there, the choose one, and that actually throws you out in a way. And you got to start it again, so little things like that, what were were, and that's just kind of, once you get used to the tool, that was fine. But um, when you knew when you were, these things were but straight off the bat, they were kind of just, just annoying more than anything.

Interviewer (02:43):

Yeah. Um, and then if there was one significant change you could make to the tool, what?

Participant 1 (02:48):

Copy. Copy this piece of text. Copy the name, copy it to your clipboard. So then you can just paste it into the browser. Um, uh, cause a lot of times, it happened a few times. I opened the name, opened VIAF and by the time I had gone through all the clicks to get to VIAF, um, I'd forgotten the name, right. And then you've got, you've gotta go back in again.

Interviewer (03:12):

So yeah, that's a good one. Um, um, any other functions you would like to add other than the copy?

Participant 1 (03:21):

Um, if it was possible, I don't know if it is possible, but once you have something copied and you pick VIAF that it takes it from your clipboard and puts it in. Um, so that you're, like once you choose the Interlink, it goes off and actually does that search for you in a way as opposed to you having to, uh, um, to, to manually put it back in again, it's just streamlining the whole thing.

Interviewer (03:47):

Yeah, that'd be good. Um, and then you, you touched on this already, but what was your impression of the, like the three steps for creating the link? So like choosing the, the term and then the property and then kind of creating the provenance?

Participant 1 (04:02):

Um, yeah, it it's the, uh, if you hadn't shown me how to do it, it would have taken a while to figure it out, just the way it's displayed. In the fact that it's one, two, three. Yeah. Um, so there's, um, and I can completely see that it's displayed in that way, but you, you may, it's not, it's not necessarily intuitive, you know, um. And then when you, yeah, cause you've, you've got to choose your, your entity and then, then go through all your different options of it. Um, yeah. Yeah. It's, it's, it's, it's a, it's really hard one to, to, to figure out. If you, try to get it all onto the one screen, that was, that was the way to do it. Put them all into the middle but if you were doing it in a, kind of a more, um, it was a very kind of simplistic design in a way and it was, it was straightforward and basic. Um, I think if you were doing it in a more heightened way for the, the, you could do it in steps where it would bring you through as opposed to you choosing it. Like, which one of these do you want? You know, sort of the, this is my entity, this is a person. Do you want to use these? In the same way that you had the other ones you know. Do you want to choose a, you know, these are the ones that have people in them, which one of these do you want to use? And then from there, sort of going through different windows, if you wanted to separate them out, but that's kind of down the line, I suppose, in looking at the functionality straight off in this one.

Interviewer (05:32):

So do you mean if, um, if you were linking a person to a person that it would suggest properties specifically for people?

Participant 1 (05:39):

Well, it would suggest datasets. It would suggest datasets because, you know, you had geographical terms, you had temporal terms all together. So, but that's kind of bringing you through, through it in steps so it only gives you what it is that, what it is that you need. Um, would be a way of, if you going to be doing it linearly, as opposed to doing it graphically on one side of the screen, to the other.

Interviewer (06:06):

Right. Yeah. And then, um, see the steps when you're choosing the term. So when you're choosing, like is identical to, and then you choose Owl:sameAs, um, how did you feel about those three steps specifically?

Participant 1 (06:19):

Um, Oh yeah. That, um, that I thought I, that I thought was very good. Yeah. And then once you had the two entities together, I thought that actually was, that was the easy bit, you know.

Interviewer (06:27):

Okay, so it was kind of getting the entities.

Participant 1 (06:30):

Getting the entities in place, um, was kind of the, the, the messier bit, I suppose. And then once you, once you had it in place, it was fairly straightforward as to what you needed to do there.

Interviewer (06:41):

So those three steps you felt worked quite well.

Participant 1 (06:44):

Yes.

Interviewer (06:44):

That's good. Um, and then, um, did you get to take a look at the provenance data for the links that you made.

Participant 1 (06:53):

Um, where, uh, where they came from?

Interviewer (06:57):

Yes.

Participant 1 (06:58):

Yeah. I had a look at that. I did look into that and then realised what was going in there, you know, so, because I would always going in the ju, the definitions, saying, this is the same person. Um, and so after I had looked at it, I think if I'd go back, that was kind of just the last one I did, and I would have been, but I would have been a bit more, a bit more, um, a bit clearer about what I was saying and why.

Interviewer (07:20):

Okay. Yeah. And, um, do you feel like that provenance data is sort of a useful element to have in the tool? Like do you think it would be of use to you going down the line that you'd have this data on who, when, where, why it was created?

Participant 1 (07:37):

Possibly so, yeah. I suppose if you're splitting datasets up between people and I can look back at it and go [name], why did you choose that person to that person, as opposed to saying, you know, somebody else, or I don't think, I think this is wrong, at least you're able to go back to the person who did it and say, you know, um, just even knowing who created the link, why they created it. Yeah. There's the, the justification was in two different places, but in a way it was the same information. Um, so that'd be another thing if you, if you were just to find it once, that you could carry that justification through and maybe be able to edit it afterwards if needs be. But, um, because we found ourselves, found myself just repeating or saying this is the same person and because I'd already justified it somewhere else, I wasn't justifying it again in the tool. But the provenance, that's, that's where the justification really needs to be as well within the Provenance data. So maybe just a copy of what's in the other place.

Interviewer (08:31):

Yeah. So like, where do you mean sort of where the description is when you're maybe choosing the entity and you end up kind of saying the same thing twice?

Participant 1 (08:38):

Yeah. So even if you copied it across and give you the option to edit it.

Interviewer (08:42):

Yeah. That would make more sense.

Participant 1 (08:43):

It would make more sense. Yeah. Um.

Interviewer (08:48):

And then, say if you were looking at someone else's dataset and they had provenance information, like, like the provenance information that you were creating there, do you think that would be useful for you in kind of judging the dataset as to whether you'd use it or not, or?

Participant 1 (09:01):

Um, if you can see justifications for things, yes. I think you would. Whether you'd need the personal information there, I don't know. Whether you can have the personal information because of GDPR as well. But the justifications for making them would, would be useful, I think. Yeah. Yeah.

Interviewer (09:18):

Um, and then just sort of the last question, overall, do you think the tool would be useful for interlinking internal and external resources?

Participant 1 (09:25):

Absolutely. Absolutely. Yeah. Yeah.

Interviewer (09:30):

And you could see it being used by yourselves hopefully going forward maybe or?

Participant 1 (09:33):

I can do. Yeah. Yeah. I can, I can, like we're only just dipping our toe into the water here now, but, um, I can, I could see it being very useful down the line. Yeah. If we, if we managed to get ourselves in a position where we have a dataset to be looking at, you know?

Interviewer (09:49):

Yeah, yeah, yeah. Well, those are all my questions. Was there anything you wanted to touch on that you wrote?

Participant 1 (09:54):

I'll just go through them now, let me see, um, some of the text that was about, the about text and stuff. Um, some of it was very jargony and it really, it was very, it's not plain English. And I found some of it difficult to read now to be honest. So I think that could probably be written, rewritten. I said, a bugbear of mine, Gaelic is, instead of Irish, because Gaelic is referring, mostly refers to Scots Gaelic.

Interviewer (10:21):

Okay. So I wrote that somewhere did I?

Participant 1 (10:23):

Yes. Um, um, let me see now, just have looked here to see. Yeah, this is me just proofreading. Um, yeah, that was, clicked on the add new image instead of the thing. So when you click on the dataset, so say when you have the list of datasets on the right hand side and you click on say VIAF and a popup comes up to show, this is what VIAF is and you go to another dataset and that pops up over the one that's there. And especially, it's especially annoying with VIAF because VIAF is so long. So you can't really, you can read it, but it's, sometimes you don't even notice that it's there.

Interviewer (11:02):

Okay.

Participant 1 (11:04):

Um, uh, yeah, that's that one. Save under the fold, cut and paste from VIAF. Oh yeah. There was a couple of times that happened, when we cut and pasted from VIAF and it happened to [name] as well. Um, we just got an error to say, please enter a valid URI.

Interviewer (11:27):

Yeah. Um, that's so I have a URI validator in the tool and every now and then it just doesn't accept a real URI. So it's something that like, that's a URI validator that I took from Java and it should work. But it doesn't. So it's something that definitely has to be looked at.

Participant 1 (11:46):

But you know, you just kept clicking on it and eventually it passed.

Interviewer (11:49):

That's even stranger.

Participant 1 (11:53):

Mmm. Okay. Oh yeah. And that's the other thing, when you get the error and then you'd make a change to the URI. So like if you add the hash whatever to it, it doesn't, there's no change in the screen to say that it's something else, that it's still an error screen, the just stays as it is.

Interviewer (12:10):

Right. Yeah.

Participant 1 (12:13):

Yeah. And again, that's all coming from your, the Java. Yeah. So on the internal entities screen, there's no indication there to say that you've done this one, but there are interlinks on this one.

Interviewer (12:25):

Okay.

Participant 1 (12:25):

So, you know what, like even just a number to say, five links created or six things created so that when you're going through a big, long list and you go, who have I done and who have I not done? And this person had a link and this person, like now, even now going back through it, you had to kind of go to where, where was I, did I get something for this person?

Interviewer (12:41):

Yeah. That would make sense.

Participant 1 (12:42):

Good. Um, yeah. Didn't allow me to create, there was a link for one of them. The first time it didn't allow me to do it the second time it did again, just the server kind of things. Yeah. That's the copy. Found one person in two different VIAF entries. Two different, the same person in two different entities. Um, so I would just put them into as two links. So made an interesting graph. Same as, same as. The data description, constantly clicking the name of the dataset. Yeah. I've been doing that as well. The manage, you know, when you've got the, at the very front, when you're going into it, it says manage dataset. I'm constantly clicking on the name of the dataset to get into it, as opposed to going to the manage button, you know?

Interviewer (13:35):

Yeah. So something better there. Yeah.

Participant 1 (13:38):

Or even both?

Interviewer (13:40):

That it would kind of highlight the whole.

Participant 1 (13:42):

Highlight the whole thing. Yeah. Yeah. Yeah. So these are more like I'm coming at this very much from a, uh, a web, because I've worked in backends of websites. I know how things work in the background. That's probably what I've, I'm coming at it from.

Interviewer (13:55):

But still makes sense like to have the usability kind of stuff, because if that kind of gets in the way of everything else then, so it's still going to cause other problems. No, that's great. Um, that's annoying about the URI validator because I thought I had that sorted, but obviously not, but anyway at least you got them in there in the end. Okie dokes. Is there anything else that you'd like to add?

Participant 1 (14:17):

Um, no, I think, I think that was, that was, I think you've kind of covered everything that I had had, that I had wanted to say anyway. Yeah.

Appendix 17 – R2RML Mappings

Interlink Graph Mapping

```
# Mapping created using Juma editor.
 @prefix rr: <http://www.w3.org/ns/r2rml#> .
 @prefix rrf: <http://kdeg.scss.tcd.ie/ns/rrf#> .
 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

  <#TriplesMap1>
   rr:logicalTable [
     rr:sqlQuery """select l.id as link_id, d.id as linkset_id, p.primaryurl, s.secondaryurl,
concat(pr.ontologyURI, pr.predicateShort) predicate from Dataset d join PrimaryResource p on
d.id = p.dataset_id join Link l on l.primaryresource_id = p.id join SecondaryResource s on
s.link_id = l.id join Predicate pr on pr.link_id = l.id where (d.id = {DATASET_ID} and
pr.linkstatus = 'added to graph')""";
   ];

   rr:subjectMap [
      rr:column "primaryurl";
                    rr:graphMap [ rr:template "http://naisc.adaptcentre.ie/linkset/{linkset_id}" ; ]
    ];

   rr:predicateObjectMap [
     rr:predicateMap [
       rr:column "predicate";
       rr:termType rr:IRI;
     ];

rr:objectMap [
       rr:column "secondaryurl";
       rr:termType rr:IRI;
      ];
   ];
   .
```

Relationship Graph Mapping

```
# Mapping created using Juma editor.
@prefix rr: <http://www.w3.org/ns/r2rml#> .
@prefix rrf: <http://kdeg.scss.tcd.ie/ns/rrf#> .
@prefix naisc: <http://naisc.adaptcentre.ie> .
@prefix prov: <http://www.w3.org/ns/prov#> .

 <#TriplesMap1>
 rr:logicalTable [
   rr:sqlQuery """select d.id as linkset_id, pb.id as provBundle_id from Dataset d join
ProvBundle pb on d.id = pb.dataset_id where d.id = {DATASET_ID}""";
 ];

  rr:subjectMap [
     rr:template "http://naisc.adaptcentre.ie/linkset/{linkset_id}";
              rr:graphMap [ rr:template
"http://naisc.adaptcentre.ie/linkset/{linkset_id}/relationshipGraph" ; ]
  ];

 rr:predicateObjectMap [
   rr:predicateMap [
     rr:constant prov:has_provenance;
     rr:termType rr:IRI;
   ];

    rr:objectMap [
     rr:template "http://naisc.adaptcentre.ie/linkset/{linkset_id}/provenance/{provBundle_id}";
    ];
 ];
 .
```

Provenance Graph Mapping

```
# Mapping created using Juma editor.
  @prefix rr: <http://www.w3.org/ns/r2rml#> .
  @prefix rrf: <http://kdeg.scss.tcd.ie/ns/rrf#> .
  @prefix naisc: <http://naisc.adaptcentre.ie/> .
  @prefix foaf: <http://xmlns.com/foaf/0.1/> .
  @prefix prov: <http://www.w3.org/ns/prov#> .
  @prefix naiscProv: <http://naisc.adaptcentre.ie/naiscProv#> .
  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
  @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
  @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

  <#TriplesMap22>
  rr:logicalTable [
    rr:sqlQuery """select d.id as linkset_id, d.date as linkset_creationDate, d.creatorid as
linkset_creatorId, pb.id as provBundle_id, pb.creatorid as provBundle_creatorId, pb.date as
provBundle_creationDate, pr.id as predicate_id from Dataset d join ProvBundle pb on d.id =
pb.dataset_id join PrimaryResource p on d.id = p.dataset_id join Link l on
l.primaryresource_id = p.id join SecondaryResource s on s.link_id = l.id join Predicate pr on
pr.link_id = l.id join `User` person on person.`id` = d.creatorid where (d.id = {DATASET_ID}
and (pr.linkstatus = 'published' or pr.linkstatus = 'deleted'))""";
  ];

   rr:subjectMap [
      rr:template "http://naisc.adaptcentre.ie/linkset/{linkset_id}/provenance/{provBundle_id}";
      rr:class prov:Bundle;
            rr:graphMap [ rr:template
"http://naisc.adaptcentre.ie/linkset/{linkset_id}/provenance/{provBundle_id}" ; ]
   ];

  rr:predicateObjectMap [
    rr:predicateMap [
      rr:constant prov:generatedAtTime;
      rr:termType rr:IRI;
    ];

 rr:objectMap [
      rr:column "provBundle_creationDate";
      rr:termType rr:Literal;
      rr:datatype xsd:dateTime;
     ];
```

```
  ];
  rr:predicateObjectMap [
    rr:predicateMap [
      rr:constant prov:wasAttributedTo;
      rr:termType rr:IRI;
    ];

rr:objectMap [
      rr:parentTriplesMap <#TriplesMap2>;
      rr:joinCondition [
       rr:child "provBundle_creatorId";
       rr:parent "person_id";
     ];
    ];
  ];
  .

  <#TriplesMap1>
  rr:logicalTable [
    rr:sqlQuery """select d.id as linkset_id, d.date as linkset_creationDate, d.creatorid as
linkset_creatorId, pb.id as provBundle_id, pb.creatorid as provBundle_creatorId, pb.date as
provBundle_creationDate, pr.id as predicate_id from Dataset d join ProvBundle pb on d.id =
pb.dataset_id join PrimaryResource p on d.id = p.dataset_id join Link l on
l.primaryresource_id = p.id join SecondaryResource s on s.link_id = l.id join Predicate pr on
pr.link_id = l.id join `User` person on person.`id` = d.creatorid where (d.id = {DATASET_ID}
and (pr.linkstatus = 'added to graph' or pr.linkstatus = 'deleted'))""";
  ];

   rr:subjectMap [
      rr:template "http://naisc.adaptcentre.ie/linkset/{linkset_id}";
      rr:class prov:Collection;
            rr:graphMap [ rr:template
"http://naisc.adaptcentre.ie/linkset/{linkset_id}/provenance/{provBundle_id}" ; ]
   ];

  rr:predicateObjectMap [
    rr:predicateMap [
      rr:constant prov:generatedAtTime;
      rr:termType rr:IRI;
    ];
```

```
   rr:objectMap [
    rr:column "linkset_creationDate";
    rr:termType rr:Literal;
    rr:datatype xsd:dateTime;
   ];
];
rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant prov:wasAttributedTo;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:parentTriplesMap <#TriplesMap2>;
    rr:joinCondition [
     rr:child "linkset_creatorId";
     rr:parent "person_id";
   ];
  ];
];
rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant prov:qualifiedAttribution;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:parentTriplesMap <#TriplesMap4>;
    rr:joinCondition [
     rr:child "linkset_creatorId";
     rr:parent "person_id";
   ];
  ];
];
rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant prov:hadMember;
    rr:termType rr:IRI;
  ];
```

```
   rr:objectMap [
    rr:parentTriplesMap <#TriplesMap6>;
    rr:joinCondition [
     rr:child "predicate_id";
     rr:parent "predicate_id";
              ];
                   rr:joinCondition [
                    rr:child "provBundle_id";
                    rr:parent "predicate_creationProvBundle";
   ];
  ];
];
rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant prov:hadMember;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:parentTriplesMap <#TriplesMap14>;
    rr:joinCondition [
     rr:child "predicate_id";
     rr:parent "predicate_id";
              ];
                   rr:joinCondition [
                    rr:child "provBundle_id";
                    rr:parent "predicate_delProvBundle";
   ];
  ];
];
.
```

```
<#TriplesMap10>
  rr:logicalTable [
    rr:sqlQuery """select d.id as linkset_id, d.date as linkset_creationDate, d.creatorid as
linkset_creatorId, pb.id as provBundle_id, pb.creatorid as provBundle_creatorId, pb.date as
provBundle_creationDate, pr.id as predicate_id from Dataset d join ProvBundle pb on d.id =
pb.dataset_id join PrimaryResource p on d.id = p.dataset_id join Link l on
l.primaryresource_id = p.id join SecondaryResource s on s.link_id = l.id join Predicate pr on
pr.link_id = l.id join `User` person on person.`id` = d.creatorid where (d.id = {DATASET_ID}
and (pr.linkstatus = 'added to graph' or pr.linkstatus = 'deleted'))""";
  ];

   rr:subjectMap [
      rr:template "http://naisc.adaptcentre.ie/naisc";
      rr:termType rr:BlankNode;
      rr:class prov:SoftwareAgent;
      rr:class foaf:Agent;
            rr:graphMap [ rr:template
"http://naisc.adaptcentre.ie/linkset/{linkset_id}/provenance/{provBundle_id}" ; ]
   ];

  rr:predicateObjectMap [
    rr:predicateMap [
      rr:constant foaf:name;
      rr:termType rr:IRI;
    ];

     rr:objectMap [
      rr:constant "NAISC";
     ];
  ];
  .
```

```
  @prefix naisc: <http://naisc.adaptcentre.ie/> .
  @prefix foaf: <http://xmlns.com/foaf/0.1/> .
  @prefix prov: <http://www.w3.org/ns/prov#> .
  @prefix naiscProv: <http://naisc.adaptcentre.ie/naiscProv#> .
  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
  @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
  @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
  @prefix schema.org: <https://schema.org/> .

 <#TriplesMap2>
 rr:logicalTable [
   rr:sqlQuery """select d.id as linkset_id, d.creatorid as linkset_creatorId, pb.id as
provBundle_id, person.id as person_id, person.firstname as person_firstName, person.lastname
as person_lastName, person.userOrganisation as person_organisation,
person.userOrganisationRole as person_role from Dataset d join ProvBundle pb on d.id =
pb.dataset_id join PrimaryResource p on d.id = p.dataset_id join Link l on
l.primaryresource_id = p.id join SecondaryResource s on s.link_id = l.id join Predicate pr on
pr.link_id = l.id join `User` person on person.`id` = d.creatorid where (d.id = {DATASET_ID}
and (pr.linkstatus = 'added to graph' or pr.linkstatus = 'deleted'))""";
 ];
  rr:subjectMap [
    rr:template "http://naisc.adaptcentre.ie/person/{person_id}";
    rr:class prov:Person;
          rr:class foaf:Person;
    rr:class schema.org:Person;
          rr:graphMap [ rr:template
"http://naisc.adaptcentre.ie/linkset/{linkset_id}/provenance/{provBundle_id}" ; ]
  ];

 rr:predicateObjectMap [
   rr:predicateMap [
    rr:constant foaf:givenName;
    rr:termType rr:IRI;
   ];

    rr:objectMap [
    rr:column "person_firstName";
    ];
 ];
```

```
rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant foaf:familyName;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:column "person_lastName";
   ];
];
rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant schema.org:hasOccupation;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:column "person_role";
   ];
];
rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant prov:actedOnBehalfOf;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:parentTriplesMap <#TriplesMap3>;
    rr:joinCondition [
     rr:child "person_id";
     rr:parent "person_id";
   ];
  ];
];
.
```

```
 <#TriplesMap3>
 rr:logicalTable [
   rr:sqlQuery """select d.id as linkset_id, d.creatorid as linkset_creatorId, pb.id as
provBundle_id, person.id as person_id, person.firstname as person_firstName, person.lastname
as person_lastName, person.userOrganisation as person_organisation,
person.userOrganisationRole as person_role from Dataset d join ProvBundle pb on d.id =
pb.dataset_id join PrimaryResource p on d.id = p.dataset_id join Link l on
l.primaryresource_id = p.id join SecondaryResource s on s.link_id = l.id join Predicate pr on
pr.link_id = l.id join `User` person on person.`id` = d.creatorid where (d.id = {DATASET_ID}
and (pr.linkstatus = 'added to graph' or pr.linkstatus = 'deleted'))""";
 ];

  rr:subjectMap [
     rr:template "http://naisc.adaptcentre.ie/organization/{person_organisation}";
     rr:termType rr:BlankNode;
     rr:class prov:Organization;
           rr:graphMap [ rr:template
"http://naisc.adaptcentre.ie/linkset/{linkset_id}/provenance/{provBundle_id}" ; ]
  ];

 rr:predicateObjectMap [
   rr:predicateMap [
     rr:constant foaf:name;
     rr:termType rr:IRI;
   ];

    rr:objectMap [
     rr:column "person_organisation";
    ];
 ];
 .
```

```
<#TriplesMap4>
rr:logicalTable [
  rr:sqlQuery """select d.id as linkset_id, d.creatorid as linkset_creatorId, pb.id as
provBundle_id, person.id as person_id, person.firstname as person_firstName, person.lastname
as person_lastName, person.userOrganisation as person_organisation,
person.userOrganisationRole as person_role from Dataset d join ProvBundle pb on d.id =
pb.dataset_id join PrimaryResource p on d.id = p.dataset_id join Link l on
l.primaryresource_id = p.id join SecondaryResource s on s.link_id = l.id join Predicate pr on
pr.link_id = l.id join `User` person on person.`id` = d.creatorid where (d.id = {DATASET_ID}
and (pr.linkstatus = 'added to graph' or pr.linkstatus = 'deleted'))""";
];

 rr:subjectMap [
   rr:template "http://naisc.adaptcentre.ie/person/{person_id}/role/createdTheLinkset";
   rr:termType rr:BlankNode;
   rr:class prov:Attribution;
         rr:graphMap [ rr:template
"http://naisc.adaptcentre.ie/linkset/{linkset_id}/provenance/{provBundle_id}" ; ]
 ];

rr:predicateObjectMap [
  rr:predicateMap [
   rr:constant prov:hadRole;
   rr:termType rr:IRI;
  ];

   rr:objectMap [
   rr:constant "Created the linkset";
   ];
];
rr:predicateObjectMap [
  rr:predicateMap [
   rr:constant prov:agent;
   rr:termType rr:IRI;
  ];

   rr:objectMap [
   rr:parentTriplesMap <#TriplesMap2>;
   rr:joinCondition [
    rr:child "person_id";
    rr:parent "person_id";
   ];
```

```
  ];
 ];
 .

  @prefix naisc: <http://naisc.adaptcentre.ie/> .
  @prefix foaf: <http://xmlns.com/foaf/0.1/> .
  @prefix prov: <http://www.w3.org/ns/prov#> .
  @prefix naiscProv: <http://naisc.adaptcentre.ie/naiscProv#> .
  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
  @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
  @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

 <#TriplesMap5>
 rr:logicalTable [
  rr:sqlQuery """select d.id as linkset_id, pb.id as provBundle_id, person.id as person_id, p.id
as primary_id, p.primaryurl, p.primaryurllabel, s.id as secondary_id, s.secondaryurl,
s.secondaryurllabel, pr.id as predicate_id, concat (pr.ontologyURI, pr.predicateShort)
predicate, pr.description as predicateJustification, pr.date as predicate_date, pr.replaced,
pr.creationProvBundleId as predicate_creationProvBundle, l.id as link_id from Dataset d join
PrimaryResource p on d.id = p.dataset_id join Link l on l.primaryresource_id = p.id join
SecondaryResource s on s.link_id = l.id join Predicate pr on pr.link_id = l.id join `User` person
on person.`id` = pr.creatorid join ProvBundle pb on pb.id = pr.creationProvBundleId where
(d.id = {DATASET_ID} and (pr.linkstatus = 'added to graph' or pr.linkstatus = 'deleted'))""";
 ];

  rr:subjectMap [
    rr:template
"http://naisc.adaptcentre.ie/linkset/{linkset_id}/interlinkCreationActivity/interlink/{predicate_i
d}/{predicate_creationProvBundle}";
    rr:termType rr:BlankNode;
    rr:class naiscProv:InterlinkCreationActivity;
          rr:graphMap [ rr:template
"http://naisc.adaptcentre.ie/linkset/{linkset_id}/provenance/{provBundle_id}" ; ]
  ];

 rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant prov:generated;
    rr:termType rr:IRI;
  ];
```

```
   rr:objectMap [
    rr:parentTriplesMap <#TriplesMap6>;
    rr:joinCondition [
     rr:child "predicate_id";
     rr:parent "predicate_id";
   ];
  ];
];
rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant prov:wasAssociatedWith;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:parentTriplesMap <#TriplesMap10>;
    rr:joinCondition [
     rr:child "linkset_id";
     rr:parent "linkset_id";
   ];
  ];
];
rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant prov:wasAssociatedWith;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:parentTriplesMap <#TriplesMap7>;
    rr:joinCondition [
     rr:child "person_id";
     rr:parent "person_id";
   ];
  ];
];
.
```

```
<#TriplesMap6>
rr:logicalTable [
  rr:sqlQuery """select d.id as linkset_id, pb.id as provBundle_id, person.id as person_id, p.id
as primary_id, p.primaryurl, p.primaryurllabel, s.id as secondary_id, s.secondaryurl,
s.secondaryurllabel, pr.id as predicate_id, concat (pr.ontologyURI, pr.predicateShort)
predicate, pr.description as predicateJustification, pr.date as predicate_date, pr.replaced,
pr.creationProvBundleId as predicate_creationProvBundle, l.id as link_id from Dataset d join
PrimaryResource p on d.id = p.dataset_id join Link l on l.primaryresource_id = p.id join
SecondaryResource s on s.link_id = l.id join Predicate pr on pr.link_id = l.id join `User` person
on person.`id` = pr.creatorid join ProvBundle pb on pb.id = pr.creationProvBundleId where
(d.id = {DATASET_ID} and (pr.linkstatus = 'added to graph' or pr.linkstatus = 'deleted'))""";
];

 rr:subjectMap [
   rr:template "http://naisc.adaptcentre.ie/linkset/{linkset_id}/interlink/{predicate_id}";
   rr:class naiscProv:Interlink;
   rr:class rdf:Statement;
        rr:graphMap [ rr:template
"http://naisc.adaptcentre.ie/linkset/{linkset_id}/provenance/{provBundle_id}" ; ]
 ];

rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant rdf:subject;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:column "primaryurl";
    rr:termType rr:IRI;
   ];
];
rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant rdf:predicate;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:column "predicate";
    rr:termType rr:IRI;
   ];
```

```
];
rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant rdf:object;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:column "secondaryurl";
    rr:termType rr:IRI;
   ];
];
rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant naiscProv:hasJustification;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:column "predicateJustification";
   ];
];
rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant prov:wasRevisionOf;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:template "http://naisc.adaptcentre.ie/linkset/{linkset_id}/interlink/{replaced}";
    rr:termType rr:IRI;
   ];
];
rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant prov:generatedAtTime;
    rr:termType rr:IRI;
  ];

 rr:objectMap [
    rr:column "predicate_date";
    rr:termType rr:Literal;
```

```
    rr:datatype xsd:dateTime;
   ];
 ];
 rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant prov:wasAttributedTo;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:parentTriplesMap <#TriplesMap7>;
    rr:joinCondition [
     rr:child "person_id";
     rr:parent "person_id";
   ];
  ];
 ];
 rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant prov:wasGeneratedBy;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:parentTriplesMap <#TriplesMap5>;
    rr:joinCondition [
     rr:child "predicate_id";
     rr:parent "predicate_id";
   ];
  ];
 ];
 .
```

```
  @prefix naisc: <http://naisc.adaptcentre.ie/> .
  @prefix foaf: <http://xmlns.com/foaf/0.1/> .
  @prefix prov: <http://www.w3.org/ns/prov#> .
  @prefix naiscProv: <http://naisc.adaptcentre.ie/naiscProv#> .
  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
  @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
  @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

 <#TriplesMap7>
 rr:logicalTable [
  rr:sqlQuery """select d.id as linkset_id, pb.id as provBundle_id, person.id as person_id,
person.firstname as person_firstName, person.lastname as person_lastName,
person.userOrganisation as person_organisation, person.userOrganisationRole as person_role
from Dataset d join ProvBundle pb on d.id = pb.dataset_id join PrimaryResource p on d.id =
p.dataset_id join Link l on l.primaryresource_id = p.id join SecondaryResource s on s.link_id =
l.id join Predicate pr on pr.link_id = l.id join `User` person on person.`id` = pr.creatorid where
(d.id = {DATASET_ID} and (pr.linkstatus = 'added to graph' or pr.linkstatus = 'deleted'))""";
 ];

  rr:subjectMap [
    rr:template "http://naisc.adaptcentre.ie/person/{person_id}";
    rr:class prov:Person;
           rr:class foaf:Person;
    rr:class schema.org:Person;
           rr:graphMap [ rr:template
"http://naisc.adaptcentre.ie/linkset/{linkset_id}/provenance/{provBundle_id}" ; ]
  ];

 rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant foaf:givenName;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:column "person_firstName";
   ];
 ];
 rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant foaf:familyName;
    rr:termType rr:IRI;
```

```
  ];

   rr:objectMap [
    rr:column "person_lastName";
   ];
];
rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant schema.org:hasOccupation;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:column "person_role";
   ];
];
rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant prov:actedOnBehalfOf;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:parentTriplesMap <#TriplesMap8>;
    rr:joinCondition [
     rr:child "person_id";
     rr:parent "person_id";
   ];
  ];
];
.
```

```
 <#TriplesMap8>
 rr:logicalTable [
   rr:sqlQuery """select d.id as linkset_id, pb.id as provBundle_id, person.id as person_id,
person.firstname as person_firstName, person.lastname as person_lastName,
person.userOrganisation as person_organisation, person.userOrganisationRole as person_role
from Dataset d join ProvBundle pb on d.id = pb.dataset_id join PrimaryResource p on d.id =
p.dataset_id join Link l on l.primaryresource_id = p.id join SecondaryResource s on s.link_id =
l.id join Predicate pr on pr.link_id = l.id join `User` person on person.`id` = pr.creatorid where
(d.id = {DATASET_ID} and (pr.linkstatus = 'added to graph' or pr.linkstatus = 'deleted'))""";
 ];

  rr:subjectMap [
     rr:template "http://naisc.adaptcentre.ie/organization/{person_organisation}";
     rr:termType rr:BlankNode;
     rr:class prov:Organization;
            rr:graphMap [ rr:template
"http://naisc.adaptcentre.ie/linkset/{linkset_id}/provenance/{provBundle_id}" ; ]
  ];

 rr:predicateObjectMap [
   rr:predicateMap [
     rr:constant foaf:name;
     rr:termType rr:IRI;
   ];

    rr:objectMap [
     rr:column "person_organisation";
    ];
 ];
 .
```

```
@prefix naisc: <http://naisc.adaptcentre.ie/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix naiscProv: <http://naisc.adaptcentre.ie/naiscProv#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<#TriplesMap13>
rr:logicalTable [
  rr:sqlQuery """select d.id as linkset_id, pb.id as provBundle_id, person.id as person_id,
p.primaryurl, p.primaryurllabel, s.secondaryurl, s.secondaryurllabel, pr.id as predicate_id,
concat (pr.ontologyURI, pr.predicateShort) predicate, pr.description as predicateJustification,
pr.deletionDate as predicate_deletionDate, pr.replacedBy, pr.deletionProvBundleId as
predicate_delProvBundle from Dataset d join PrimaryResource p on d.id = p.dataset_id join
Link l on l.primaryresource_id = p.id join SecondaryResource s on s.link_id = l.id join
Predicate pr on pr.link_id = l.id join `User` person on person.`id` = pr.deletedBy join
ProvBundle pb on pb.id = pr.deletionProvBundleId where (d.id = {DATASET_ID} and
pr.linkstatus = 'deleted')""";
];

 rr:subjectMap [
   rr:template
"http://naisc.adaptcentre.ie/linkset/{linkset_id}/interlinkDeletionActivity/interlink/{predicate_i
d}/{predicate_delProvBundle}";
   rr:termType rr:BlankNode;
   rr:class naiscProv:InterlinkDeletionActivity;
         rr:graphMap [ rr:template
"http://naisc.adaptcentre.ie/linkset/{linkset_id}/provenance/{provBundle_id}" ; ]
 ];

rr:predicateObjectMap [
  rr:predicateMap [
   rr:constant prov:invalidated;
   rr:termType rr:IRI;
  ];

   rr:objectMap [
   rr:parentTriplesMap <#TriplesMap14>;
   rr:joinCondition [
    rr:child "predicate_id";
    rr:parent "predicate_id";
```

```
    ];
   ];
  ];
  rr:predicateObjectMap [
   rr:predicateMap [
     rr:constant prov:wasAssociatedWith;
     rr:termType rr:IRI;
   ];

    rr:objectMap [
     rr:parentTriplesMap <#TriplesMap10>;
     rr:joinCondition [
      rr:child "linkset_id";
      rr:parent "linkset_id";
    ];
   ];
  ];
  rr:predicateObjectMap [
   rr:predicateMap [
     rr:constant prov:wasAssociatedWith;
     rr:termType rr:IRI;
   ];

    rr:objectMap [
     rr:parentTriplesMap <#TriplesMap17>;
     rr:joinCondition [
      rr:child "person_id";
      rr:parent "person_id";
    ];
   ];
  ];
  .
```

```
<#TriplesMap14>
rr:logicalTable [
  rr:sqlQuery """select d.id as linkset_id, pb.id as provBundle_id, person.id as person_id,
p.primaryurl, p.primaryurllabel, s.secondaryurl, s.secondaryurllabel, pr.id as predicate_id,
concat (pr.ontologyURI, pr.predicateShort) predicate, pr.description as predicateJustification,
pr.deletionDate as predicate_deletionDate, pr.replacedBy, pr.creationProvBundleId as
predicate_creationProvBundle, pr.deletionProvBundleId as predicate_delProvBundle from
Dataset d join PrimaryResource p on d.id = p.dataset_id join Link l on l.primaryresource_id =
p.id join SecondaryResource s on s.link_id = l.id join Predicate pr on pr.link_id = l.id join
`User` person on person.`id` = pr.deletedBy join ProvBundle pb on pb.id =
pr.deletionProvBundleId where (d.id = {DATASET_ID} and pr.linkstatus = 'deleted')""";
 ];

  rr:subjectMap [
    rr:template "http://naisc.adaptcentre.ie/linkset/{linkset_id}/interlink/{predicate_id}";
    rr:class naiscProv:Interlink;
            rr:graphMap [ rr:template
"http://naisc.adaptcentre.ie/linkset/{linkset_id}/provenance/{provBundle_id}" ; ]
  ];

 rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant prov:hadRevision;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:template "http://naisc.adaptcentre.ie/linkset/{linkset_id}/interlink/{replacedBy}";
    rr:termType rr:IRI;
   ];
 ];
 rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant prov:invalidatedAtTime;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:column "predicate_deletionDate";
    rr:termType rr:Literal;
    rr:datatype xsd:dateTime;
   ];
```

```
 ];
 rr:predicateObjectMap [
  rr:predicateMap [
   rr:constant prov:wasInvalidatedBy;
   rr:termType rr:IRI;
  ];

   rr:objectMap [
   rr:parentTriplesMap <#TriplesMap13>;
   rr:joinCondition [
    rr:child "predicate_id";
    rr:parent "predicate_id";
   ];
  ];
 ];
 rr:predicateObjectMap [
   rr:predicateMap [
    rr:constant prov:has_provenance;
    rr:termType rr:IRI;
   ];

   rr:objectMap [
    rr:template
"http://naisc.adaptcentre.ie/linkset/{linkset_id}/provenance/{predicate_creationProvBundle}";
    rr:termType rr:IRI;
   ];
  ];
 .
```

```
   @prefix naisc: <http://naisc.adaptcentre.ie/> .
   @prefix foaf: <http://xmlns.com/foaf/0.1/> .
   @prefix prov: <http://www.w3.org/ns/prov#> .
   @prefix naiscProv: <http://naisc.adaptcentre.ie/naiscProv#> .
   @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
   @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
   @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

  <#TriplesMap17>
  rr:logicalTable [
   rr:sqlQuery """select d.id as linkset_id, pb.id as provBundle_id, person.id as person_id,
person.firstname as person_firstName, person.lastname as person_lastName,
person.userOrganisation as person_organisation, person.userOrganisationRole as person_role
from Dataset d join ProvBundle pb on d.id = pb.dataset_id join PrimaryResource p on d.id =
p.dataset_id join Link l on l.primaryresource_id = p.id join SecondaryResource s on s.link_id =
l.id join Predicate pr on pr.link_id = l.id join `User` person on person.`id` = pr.deletedBy where
(d.id = {DATASET_ID} and pr.linkstatus = 'deleted')""";
  ];

   rr:subjectMap [
     rr:template "http://naisc.adaptcentre.ie/person/{person_id}";
     rr:class prov:Person;
           rr:class foaf:Person;
     rr:class schema.org:Person;
           rr:graphMap [ rr:template
"http://naisc.adaptcentre.ie/linkset/{linkset_id}/provenance/{provBundle_id}" ; ]
   ];

  rr:predicateObjectMap [
   rr:predicateMap [
     rr:constant foaf:givenName;
     rr:termType rr:IRI;
   ];

    rr:objectMap [
     rr:column "person_firstName";
    ];
  ];
  rr:predicateObjectMap [
   rr:predicateMap [
     rr:constant foaf:familyName;
     rr:termType rr:IRI;
```

```
  ];

   rr:objectMap [
    rr:column "person_lastName";
   ];
 ];
 rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant schema.org:hasOccupation;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:column "person_role";
   ];
 ];
 rr:predicateObjectMap [
  rr:predicateMap [
    rr:constant prov:actedOnBehalfOf;
    rr:termType rr:IRI;
  ];

   rr:objectMap [
    rr:parentTriplesMap <#TriplesMap18>;
    rr:joinCondition [
     rr:child "person_id";
     rr:parent "person_id";
   ];
  ];
 ];
 .
```

```
<#TriplesMap18>
rr:logicalTable [
  rr:sqlQuery """select d.id as linkset_id, pb.id as provBundle_id, person.id as person_id,
person.firstname as person_firstName, person.lastname as person_lastName,
person.userOrganisation as person_organisation, person.userOrganisationRole as person_role
from Dataset d join ProvBundle pb on d.id = pb.dataset_id join PrimaryResource p on d.id =
p.dataset_id join Link l on l.primaryresource_id = p.id join SecondaryResource s on s.link_id =
l.id join Predicate pr on pr.link_id = l.id join `User` person on person.`id` = pr.deletedBy where
(d.id = {DATASET_ID} and pr.linkstatus = 'deleted')""";
];

 rr:subjectMap [
   rr:template "http://naisc.adaptcentre.ie/organization/{person_organisation}";
   rr:termType rr:BlankNode;
   rr:class prov:Organization;
        rr:graphMap [ rr:template
"http://naisc.adaptcentre.ie/linkset/{linkset_id}/provenance/{provBundle_id}" ; ]
 ];

rr:predicateObjectMap [
  rr:predicateMap [
   rr:constant foaf:name;
   rr:termType rr:IRI;
  ];

  rr:objectMap [
   rr:column "person_organisation";
  ];
];
.
```