

UNLOCKING CULTURAL HERITAGE

**A case study of metadata management at
Europeana**

Katharina W M de Burger

Examensarbete (30 högskolepoäng) i museologi för masterexamen inom ABM-
masterprogrammet vid Lunds universitet.

Handledare: Björn Magnusson Staaf

År: 2024

Title

UNLOCKING CULTURAL HERITAGE: a case study of metadata management at Europeana.

Abstract

This MA thesis introduces Europeana and information science principles to the field of museology. Its aim is to identify challenges in digital cultural heritage metadata management within an international context. To do this, this thesis focuses on specific aspects: the European Data Model, international cooperation, and the metadata management process specific to Europeana.

The study is designed as a case study of the digital cultural heritage portal and database Europeana. The primary method for the collection of empirical material is that of Kusenbach's Go-along. The method is supplemented with unstructured interviews and email interaction, and the analysis of existing documents and websites.

Three theoretical perspectives guide the analysis and discussion of the results of the study. Actor Network Theory and Theory of Infrastructure are particularly relevant for the analysis of the empirical material, and Database as Discourse is applied to the discussion in connection with the topic of power which can be found in earlier literature.

This thesis concludes that common issues and challenges within EDM are related to either faults created through automatic processes, or the absence of data from the onset. Challenges of international cooperation arise due to an unclear understanding of Europeana, expectations and requirements, and inconsistent organisational structures. The metadata management process of Europeana is non-linear and dependent on local structures. Each level has its own challenges. The levels can broadly be divided into: The Provider, the Aggregator, and the Europeana Foundation.

Keywords

Museology, Information science, Digital cultural heritage, Europeana, European Data Model (EDM), Metadata management, data management

Acknowledgements

This thesis would not have been possible without the support and advice from my supervisor Björn Magnusson Staaf, who has offered continuous feedback and encouragement during the writing process.

This thesis would also not have existed if it weren't for the people at Europeana and the Swedish National Heritage Board. My sincerest thanks are offered to these organisations and their employees, for their participation in the study.

Table of Contents

1. Introduction	5
1.1 Problem statement, Aim, and Research Questions	6
1.2 Background	7
1.2.1 Digital Cultural Heritage.....	7
1.2.2 Europeana	9
1.2.3 European Data Model – EDM	11
1.3 Limitations	12
1.4 Outline of this Thesis	14
2. Literature Overview.....	16
2.1 Metadata Research in a Museological Context.....	16
2.2 Knowledge organisation systems and digital cultural heritage.....	18
2.2.1 Knowledge Organisation Systems	18
2.2.2 Digital Cultural Heritage.....	21
2.2.3 Digital cultural heritage portals in research – museum focused, excluding Europeana.	23
2.3 Overview of research about Europeana	25
3. Theory	27
3.1 Actor Network Theory – ANT	28
3.2 Theory of Infrastructure	29
3.3 Database as Discourse.....	31
4. Methodology	33
4.1 Research Design – The Case Study	33
4.1.1 The Go-Along	34
4.1.2 Other methods of data collection	36
4.2 Ethical considerations	38
4.3 Conducting the study.	40
5. Results and Analysis	42
5.1 Actor Network Presentation.....	42
5.1.1 The Aggregation loop	45
5.1.2 The Metadata Network	49
5.2 Theory of Infrastructure – the dimensions of Europeana	62
5.3 Review based on Research Questions.....	65
6. Discussion.....	67
7. Conclusion.....	71
7.1 Suggested future research	72
Bibliography	74
Appendix 1: Alphabetised list of Abbreviations	81
Appendix 2: Index of tables and figures	83
Figures:.....	83
Tables:.....	84

1. Introduction

It is an established fact that cultural heritage institutions face some major challenges when it comes to making the items of which they are stewards of accessible. This holds especially true for the digital sphere, where issues commonly arise when considering how to effectively describe items within their collections, as well as how to share these descriptions across institutions (Skinner 2014, p. 52). One example of an attempt to tackle these challenges is the database Europeana, run by a foundation which shares the same name. This is why this MA thesis takes the form of a case study of the database Europeana, centring around the practices regarding metadata.

Having a background in Library and Information Science, it is not uncommon to come in contact with different kinds of knowledge organisation systems and digitally accessible platforms and databases. Within museology, there is a noted lack of both awareness in regard to access to digital and digitised artefacts, as seen by the few times databases in connection to museums are mentioned during lectures and seminars. There is also a noted absence of museological research about digitally accessible platforms and databases, especially within northern Europe, as no articles on the topic could be found, in the planning stages of this thesis, in the journal *Nordisk Museologi*. The above has led to an interest in introducing a more digital oriented approach to the field. The choice of conducting a field study about, in particular, Europeana follows this same reasoning, with Europeana being selected due to its varied catalogue, and data being supplied through various cultural heritage institutions, including archives, libraries and museums (ALM).

While definitions regarding what Europeana is differ, this thesis operates under the assumption of Europeana as an international digital database that collects and manages digitised cultural heritage. It is run by the Europeana Foundation on assignment of the European Union and its collection is acquired through cooperation with various cultural and memory institutions. As an organisation Europeana shares and promotes Europe's digital cultural heritage, allowing it to be used and enjoyed by everyone for learning, work, or fun. It does so to inspire and inform new perspectives and to encourage conversations about history and culture. Europeana provides access to items from various institutions across Europe, including artworks, books, music, videos, newspapers, and more. It does so through various media formats, including images, text, sound, video, and 3D digital models.

Its access relies on thousands of European providing institutions, however, the foundation does not work with them individually but works through a network of aggregators that collect their items and related data, checks it, enriches it, and links it (Europeana 2024a). The structure and organisation of Europeana lends itself to its exploration as a prime example of international cooperation with data and metadata, providing access to digital cultural heritage.

Lastly, I'd like to forewarn that this MA thesis is lengthy and consist of a high degree of detail. As such it is also heavy in terms of content and technical detail. This is due to it being a patchwork of various disciplines, with some not widely applied within the field of museology. The writing of this thesis, required a large frame of reference. Some of this framework is also necessary for the understanding of the thesis and for fulfilling its aim – which is presented below.

1.1 Problem statement, Aim, and Research Questions

With Europeana's work being rooted in a collaborative effort of a selection of cultural and memory institutions, Europeana also faces many of the same challenges as these institutions, especially when it comes to the digital sphere. The issues in question that are relevant to this MA thesis are difficulties with how to effectively describe items within collections and how to share descriptions across institutions (Skinner, 2014). It is largely these issues that inspire the research aim of this MA thesis, the aim being to contribute to the research field pertaining to the digitalisation of cultural heritage through identifying challenges concerning metadata management in an international context.

The specific research questions (RQs) that guide the exploration of present work processes in an international context regarding (meta)data management are as follows:

- Europeana uses the European Data Model standard for their (meta)data, and aggregators are used to convert/transform local data to EDM. Which issues and challenges can commonly be found in EDM following this process?
- What are the challenges of international cooperation in a digital format in regard to cultural heritage?
- What does the data management process entail in the specific case of Europeana?

The aim and research questions are presented above as the questions themselves are the cause for the structure of the following chapters. The topic of metadata and databases, as the introduction stated, are topics not commonly explored within the realm of museology, and the questions above informed the choice of literature and which fields of research to involve in this thesis. Choices in relation to what knowledge would be needed to understand any empirical materials collected through the study of Europeana, but also in the selection of the type of background that would be needed to provide context and understanding of the topic of Europeana and its significance to the museological field.

1.2 Background

To fully understand the motivations behind this MA thesis, one should understand the basic premises on which it is built. The provided background explains what digital cultural heritage, that which Europeana's database is comprised of, is. It touches upon the discussion between digitised vs. digital within its definition, albeit this is explored in more detail in chapter 2. Further, the background chapter aims to give an overview of the Europeana Foundation and how the Europeana initiative started, as well as a brief introduction to the European Data Model (EDM), the model used for descriptive data. More details regarding EDM can be found in the Results and Analysis chapter, within the sub-chapter 5.1.2 Metadata.

1.2.1 Digital Cultural Heritage

Europeana is one attempt at creating a universally accessible collective access portal to European cultural heritage in the digital sphere. But what is cultural heritage? And more specifically, what is digital cultural heritage?

Cultural heritage can be many things, it comes in a variety of shapes. It can be tangible, e.g. buildings, landscapes, artifacts. It can be intangible, e.g. memories, values, customs, and practices (Nilson & Thorell 2018, p. 9). Definitions can be broader, or narrower. The United Nations Educational, Scientific and Cultural Organisation (UNESCO), which is in many instances viewed as an institute of certain authority regarding the topic, defines cultural heritage as follows:

Cultural heritage includes artefacts, monuments, a group of buildings and sites, museums that have a diversity of values including symbolic, historic, artistic, aesthetic, ethnological or anthropological, scientific and social significance. It includes tangible heritage (movable, immobile and underwater), intangible cultural heritage (ICH) embedded into cultural, and natural heritage artefacts, sites or monuments. The definition excludes ICH related to other

cultural domains such as festivals, celebration etc. It covers industrial heritage and cave paintings.

(UNESCO 2020)

There are also a variety of uses for cultural heritage. It ranges from the purpose of building nations to marketing places, and as such the abstract idea of cultural heritage, of a common history, is used as a tool within the political, cultural, entrepreneurial, and educational. It is used both for emancipation and justifying subjugation.

Cultural heritage refers to a society's use of the past, and contributes to shaping everything from national stereotypes, regional identities, and the reflections on the past. Depending on the region, it differs what cultural heritage is associated with. Within Europe, the first thing that comes to mind tends to be older city centres. In North America it is more likely to be associated with national parks, museums, and galleries. In Oceania it is strongly linked to indigenous identity and landscape (Nilson & Thorell 2018, p. 10).

So, where does digital cultural heritage fall within all of this? The digital part of this means that it can be argued that it falls somewhere between the tangible and the intangible. Digitised or digital cultural heritage can't really be touched or experienced the same way tangible cultural heritage such as a building or artifact might, regardless of whether one should or not. But it is important to note that when this thesis refers to digital cultural heritage, it means both originally created digital content and heritage, as well as digital representations of physical or 'real' cultural heritage. Regardless of the discussed discourse for digital cultural heritage in current and past research. Mostly, due to the Europeana platform providing access to both, and also because whether a record or item is originally in a digital format or whether it is a digital reproduction of a physical item, it exists in data. As such it is saved in bytes to a server, accessible to the web.

The digital items, or representations of items, can be used much the same way that other cultural heritage can. And it is just as difficult to preserve, as data can get corrupted, it degrades with age, and becomes inaccessible if the files aren't updated to match the capabilities of the technology used to access and manage it, as seen in the example above.

The connection between cultural heritage and memory institutions, such as museum, is that of custodianship. Museums, and other heritage institutions/organisations have a kind of institutionalised authority to act as

custodians of the past, particularly in Western societies. Due to this, it can be argued that these institutions and organisations hold a significant part of intellectual capital of the information society. When using emergent digital technologies, there occurs parallel shifts in the organisation and practices of the institutions caring for cultural heritage. If the digital technology use and the institutions have a relationship that is symbiotic in nature, the cultural heritage they care for will appropriate, adapt, incorporate and transform the digital technologies (Cameron & Kenderdine 2007, p. 1).

1.2.2 Europeana

Since the study uses Europeana as its case through which to look for the challenges in international (meta)data management and collaboration in connection to digital cultural heritage, it is important to establish a consensus in regards what Europeana actually is. It is generally referred to as a portal and has been referred to as such since at least 2011, as seen by the press release in January by the European Commission (2011). But the concept of Europeana has shifted and developed over time.

Its history officially starts in 2005, with a letter signed by six heads of state asking European Union (EU) officials to support the development of a European digital Library. The letter was sent by the French president at the time, cosigned by the heads of state of Germany, Spain, Italy, Poland, and Hungary, addressed to the president of the European Commission José Manuel Barosso (Chirac 2005; European Commission 2005, 2014; Reding 2005). A prototype of Europeana went subsequently live in the fourth quarter of 2008. While the idea started with a digital library, museums, audiovisual archives, and galleries joined the library initiative. This meant that at launch, Europeana provided access to roughly 4.5 million digital objects (Europeana Foundation 2024a).

In 2011 the European Commission released the ‘New Renaissance’ report (European Commission 2011; Europeana Foundation 2024a). It is composed by the Reflection Group on digitisation, also known as *Comité des sages*, in connection to their task of providing a set of recommendations for the digitisation, online accessibility and preservation of Europe’s cultural heritage (European Commission 2010). In short, the report “endorses the Digital Agenda's objective of strengthening Europe's digital library Europeana and suggests solutions for making works covered by copyright available online” (European Commission 2011). The report feeds into the European Commission’s strategy *Digital Agenda for Europe* which aims to help

cultural institutions transition towards the digital age (European Commission 2011). The report is especially relevant to the history of Europeana due to some of its key conclusions and recommendations, one such is quoted below.

The Europeana portal should become the central reference point for Europe's online cultural heritage. Member States must ensure that all material digitised with public funding is available on the site, and bring all their public domain masterpieces into Europeana by 2016. Cultural institutions, the European Commission and Member States should actively and widely promote Europeana.

(European Commission 2011)

At this point in time, January 2011, Europeana provided access to more than 15 million digitised books, maps, photographs, film clips, paintings and musical extracts, with most being older works in the public domain (European Commission 2011).

The year 2012 brought with it a large milestone, when Europeana metadata was released under the Creative Commons CC0 1.0 Universal Public Domain Dedication. This means that the metadata itself became freely available for all kinds of use, allowing for digital innovation and creativity (Europeana Foundation 2024a). Europeana released the metadata for more than 20 million cultural objects. It is particularly significant as this means that the released metadata under the CC0 waiver can be used for Linked Open Data (LOD), with the potential to bring together data from libraries, museum and archives, as well as data from sectors such as tourism and broadcasting (Europeana Foundation 2012).

Between the years of 2015 and 2022 Europeana was deemed one of the European Commission's Digital Service Infrastructures (DSI), delivering networked cross-border services (Europeana Foundation 2020, 2024a). As a DSI, Europeana worked to make it easier for institutions to effectively share their collections online, improve the quality of data and content shared with Europeana, and empower cultural heritage institutions in digital transformation. Since 2022, Europeana has been central to the European Commission's *common European data space for cultural heritage*. It is funded under the *Digital Europe programme* (Europeana Foundation 2024a).

The latest public data show that Europeana provides access to an estimated 55 million digital objects (Europeana 2024b) divided over images, text, sound, video and 3D items.

Item type	Amount
Image	31 672 384
Text	24 519 709
Sound	996 587
Video	364 986
3D	4 696

Table 1. Overview of amount of digital items by item type, in descending order based on amount. Copied from: <https://www.europeana.eu/en/about-us> [2024-01-11 11:51:48 CET]

1.2.3 European Data Model – EDM

The European Data Model, or EDM, according to Europeana is “an interoperable framework that allows us to collect, connect and enrich cultural heritage metadata” (Europeana Foundation 2024b).

EDM was developed due to the various European countries and institutes connected to cultural heritage having differing metadata standards. The aim of EDM is, simply put, to make the metadata interoperable. To be able to do this, there was an attempt at identifying the lowest common denominator of the various standards in use, leading to the establishment of the European Semantic Elements (ESE) in 2008 with the launch of the Europeana Prototype. ESE is a flat record structure using the Dublin Core (DC) element sets with some Europeana extensions (Europeana 2013; *Hackweek Part1* 2021).

EDM is not built on any specific community standard. Rather, it adopts as much as possible, an open, cross-domain Semantic Web-based framework, in the attempt to accommodate as large a range of community standards, such as e.g. LIDO¹ (museums), EAD² (archives) and METS³ (libraries). The EDM is modelled as to also allow for data enrichment from third party sources, all the while showing the provenance of all data links connected to the digital object (Europeana 2013). The Europeana Foundation gives this example in their EDM Primer:

¹ LIDO stands for Lightweight Information Describing Objects and is an XML harvesting schema intended for delivering metadata from an organisation’s online collections database to portals of aggregated resources (CIDOC 2024a)

² EAD stands for Encoded Archival Description and is an XML standard for encoding archival finding aids. It is maintained by the Technical Subcommittee for Encoded Archival Standards of the Society of American Archivists in collaboration with the Library of Congress (Library of Congress 2024a)

³ METS stands for Metadata Encoding & Transmission Standard and is a metadata schema for encoding descriptive, administrative and structural metadata regarding objects within a digital library using XML schema language (Library of Congress 2024b)

a digital object from Provider A may be contextually enriched by metadata from Provider B. It may also be enriched by the addition of data from authority files held by Provider C, and a web-based thesaurus offered by Publisher D.

(*Europeana 2013*)

EDM also, in comparison to ESE, is able to accommodate and support more complex objects. One example of this might be a digital book, where individual chapters, illustrations as well as the index can be understood both individually and collectively through EDM (Europeana 2013).

EDM adheres to the modelling principles connected the approach of the Semantic Web, also known as the Web of Data. This entails that there is an absence of an absolute or fixed schema which dictates only one way to represent data. Instead, EDM functions as an anchor, to which other models can be attached, meaning they are at least partly interoperable, while allowing the data to retain most of their original richness. This means that EDM does not necessarily require local practices to change, but it does encourage changes to increase cross-domain applicability of data. This might be done through e.g. the use of publicly accessible vocabularies (Europeana 2013)

1.3 Limitations

For all that this thesis aims to contribute, there are certain limitations that need to be defined. These limitations are related to the scope of this thesis, reflections from the conducted pilot study in preparation for this thesis, prioritisations and specific terminology and their uses.

The applicability of this thesis and its subsequent findings are limited. Not just due to the nature of this thesis, a case study, but also due to its scope. The conducted study focuses solely on the chosen case of Europeana, with the study conducted at their headquarters in the Hague, focusing on their practices and processes related to (meta)data management. This thesis has also attempted to explore the aggregator network through which Europeana is supplied with data, through the example of the Swedish national aggregator SOCH, also known as K-samsök. However, due to the structure of their organisation and data supply, a closer study was not possible, so all information regarding them is based on an informational meeting as well as their website.

This thesis makes use of certain terminology which varies in definition based on which academic field is relevant. One such example is the use of *data management*

or *metadata management*. In the case of this thesis, the term is applied to refer to the collection of practices and processes connected to the flow of data through the organisation. Included practices and processes can be, but are not limited to, quality control, repairing of damaged or broken data and datalinkage, data enrichment, policy making, transference and translation of (meta)data.

Within this thesis, the terms *data* and *metadata* are oftentimes used interchangeably. While this is not standard practice, and there are in most cases and applications of the terminology differences between the two. This thesis has chosen to use these terms interchangeably for the majority of the study, due to Europeana themselves having general tendency to use the term *data* rather than *metadata*.

Terminology is also important to consider from the perspective of the organisation and database studied. The foundation provides access to a glossary through their website Europeana Pro, but even using this glossary, it is at times difficult to use the correct phrases, due to ongoing development and changes. As such, while this thesis attempts to use the correct terminology, mistakes are likely to occur. The terminology for what this thesis in the introduction calls ‘Europeana’ is difficult to nail down. A difficulty which originates from the dynamic changes in its definition and the landscape in which it exists. Europeana can be specified in many different ways: it can be used to refer to the website, the infrastructure, the organisation, and more.

The terminology referring to specific aspects of Europeana are also dynamic. Previously the term *Core Service Platform* was used to refer to the set of services including the aggregation infrastructure, the Europeana website, Europeana APIs, interoperability services and capacity-building efforts, operated as part of the Europeana Digital Service Infrastructure (Europeana 2024b). Although other terminology can also be used. E.g.: *Europeana initiative*, used to capture the collaborative work undertaken by the Europeana Foundation and its consortium partners, along with the Europeana Aggregators’ Forum and the Europeana Network Association (Europeana 2024b); *Data space infrastructure*, the technical infrastructure underlying the common European data space for cultural heritage, consisting of four main digital products, namely, the Europeana website, Europeana Pro website, APIs and aggregation systems (Europeana 2024b). The three terms provided above might be the most encompassing ones relevant to this thesis. *Data space* (DS), as defined in the glossary, denotes a decentralised and standard-based infrastructure to enable trustworthy data sharing between the data space

participants. These data spaces can be both purpose/sector-specific and cross-sectoral. However, due to the confusion in regard to what term denotes what in itself qualifying as a challenge, this thesis does continue to use the more broad and less defined term of *Europeana*, as inspired by the descriptions used in the ‘about’ section of both the main Europeana portal, and Europeana Pro websites.

In preparation for this thesis, a pilot study was conducted to ascertain the usefulness of the chosen method in the study of (meta)data management in databases for cultural heritage institutions. A relevant reflection based on the pilot study to keep in mind for this thesis is that the chosen method, while applicable and useful for gaining insight into processes and practices, to identify challenges, also carries the risk that the findings may only scratch the surface of the questions posed, leaving room for further investigation and research. This is also reflected in the research aim and structure of this thesis, focusing on identifying the challenges in order to contribute to future research.

1.4 Outline of this Thesis

Firstly, for ease of reading it is important to be aware of the existence and content of the appendices. This thesis is heavy both in terms of content and usage of technical terms, due to it being a patchwork of various disciplines and perspectives. As such, a list of all important and prevalently used abbreviations and their meaning can be found in Appendix 1: Alphabetised list of Abbreviations. This thesis makes use of both established abbreviations and abbreviations unique to this thesis, so when unsure, please consult the list.

This thesis aims to provide context and to answer the posed RQs in the following way. In Chapter 2, literature is presented covering some of the relevant perspectives in regard to metadata management, knowledge organisation systems, digital cultural heritage, and digital cultural heritage portals, and provides a quick overview of research about Europeana as a whole. The subchapter regarding knowledge organisation systems is structured slightly differently than the other topics, due to the literature used being mostly educational in nature. This choice has been made as a basic understanding of knowledge organisation systems is necessary to allow for a discussion about challenges to take place.

Chapter 3 covers the theoretical perspectives this thesis relies on and uses to analyse and discuss findings of the conducted field study at the offices of the Europeana Foundation. This thesis mainly relies on three different perspectives, namely: Actor

Network Theory (ANT); Theory of Infrastructure (TI); and lastly, Database as Discourse (DD). ANT and TI are used extensively in analysing the empirical data collected, and DD is more prevalent in discussing aspects of power, a topic introduced in the Chapter 2.

The fourth chapter covers the methodology of the thesis, ranging from the research design and primary method of data collection, the go-along, to which secondary methods have been used in the process to support and supplement findings using the primary method. This chapter also covers the ethical considerations important to this thesis, including a discussion about the use of public records with identifiable information and the use of information originating from introduction meetings. The chapter closes with some notes regarding the way the field study was conducted.

This thesis has chosen to combine and present the results of the conducted study with the analysis thereof, in part due to the qualitative nature of the empirical data, but also in part due to the sheer quantity of collected data and information. As such chapter 5 tackles the topics in two main subchapters, the first using ANT to show how various actors connect and specific cases of challenges, and the second analysing the empirical data using the dimensions of infrastructure proposed by Star & Ruhleder (1996), following the order of the dimensions. As this chapter contains a lot of information, and as such it might be difficult to keep in mind which aspects are concurrent with the RQs posed by this thesis, the chapter concludes with a review following the RQs as a guide.

Chapter 6, the discussion, uses reflections from the Literature Overview, as well as parts of the Background chapter to both contextualise the findings and to discuss further challenges. This chapter also provides some suggested aspects which might lend itself to further research. Finally, this thesis is concluded with __ paragraphs summarizing the main points of chapters 5 and 6 and aims to provide an overview of the answers to the RQs posed by this thesis.

2. Literature Overview

A quick search conducted in the preparation stages of this MA thesis using Lund University Libraries LUBSearch on the term “Europeana”, or “Europeana AND Database” revealed a varied search result. Search results yielded literature published in various languages, and most results appeared to be either conference papers, academic papers, or reports. Some of these will be discussed in chapter 2.3, as there are generally two noticeable trends – one focusing on data, and one on impact. While this is important information for this thesis, literature on metadata research, digital cultural heritage, as well knowledge organisation systems, is also of importance. As such, the choice has been made to approach this chapter in the following way: It starts with a contextualisation of metadata focused research within a museological context, followed by an exploration of information retrieval systems and digital cultural heritage. Then research into other digital cultural heritage portals is presented. The chapter concludes with an overview of research published about Europeana in particular.

2.1 Metadata Research in a Museological Context

While it is not possible to say that there exists no research regarding databases and metadata within a museological context, it is a fair assessment that there exists less research focusing on metadata within museum practices, than e.g. library practices. Much the same can be said regarding metadata and archives. This is important to note, since this study focuses on the digital database Europeana, which deals with a large variety of digitised and digital objects from various institutions within the European Union pertaining to cultural heritage. The objects it contains range from works of literature and text, to photographs and digital representations of objects, as well as digitalised artworks and original digital art. As such, the metadata of which the data records are comprised, follow a variety of standards. It is an established fact that effectively describing the items of which cultural heritage institutions are stewards of, as well as sharing the descriptions across institutions are two major challenges that cultural heritage institutions face (Skinner 2014, p. 52). Two researchers who have looked at the three memory institutions of library, museum, and archive are Mary Elings and Günter Waibel, who did so in 2007. They denoted four key areas of description when it comes to metadata within these institutions, namely:

1. data fields and structure;

2. data content and values;
3. data format, and lastly;
4. data exchange.

Elings & Waibel also exemplify this with a grid structure to show the various standards that are used for each key aspect, in two versions. One denoted using the institution, and one version using the type of material as its denotation (Elings & Waibel 2007). There do exist various standards, but how widely used these standards are, is dependent on the type of institution, and differences exist even between different genres of institutions, e.g. an art museum and an anthropological museum or natural history museum.

The above background knowledge provides a suggested structure for the interpretation and presentation of possible results of the proposed research study. Seeing as there are 4 key areas of description of metadata within cultural heritage institutions, a big task will be to identify the kinds of standards and schemas are used, especially whether multiple such are in use, seeing as Europeana is a digital database which collects digital materials pertaining to and similar materials as the ones cared for by museums, as well as libraries and archives.

Interestingly, when it comes to metadata research conducted in relation to museums and museal practices, some trends can also be divided according to different time periods. Elings and Waibel study (2007) identifying key areas of data description follows a similar trend as seen in e.g. the study of Patel et al. (2005) looking into metadata requirements for a digital museum environments through the exploration of the Augmented Representation of Cultural Objects (ARCO) system. Paul F. Marty's chapter in *Encyclopedia of Library and Information Science* (2009) focusing on Museum informatics also takes more of an identifying approach, looking in broader terms at information science aspects within museums and digital museum catalogues, including metadata and standards, data sharing initiatives, information technologies, digitisation, and digital and online museums. The literature found that investigates metadata in the context of museum published between 2005 and 2010 generally takes a more identifying approach, looking at how and which standards are used, and which requirements museums have of metadata.

More current research focusing on metadata in the context of museum (published anno 2022-2024) instead investigates use cases. Topics explored range from: access to and utilization of information and items (Gibson, Chowdhury & Chowdhury

2024; Zhang & Ren 2024), effects of application of Linked Open Data (LOD) (McKenna, Debruyne & O’Sullivan 2022), accuracy completeness and consistence in metadata in relation to physical objects (Zavalin & Zavalina 2023), and the connection of metadata to museum performance and visitor experience (Philippopoulos et al. 2024). The above does however not exclude more identifying oriented research, as seen by the study by Rahul Pandey and Vinit Kumar (2023) whom conducted a survey to identify which metadata element sets were used for digital art objects in online collections.

2.2 Knowledge organisation systems and digital cultural heritage

2.2.1 Knowledge Organisation Systems

For the purpose of this thesis, literature that is most relevant is educational literature, rather than current research focusing on Knowledge organisation systems and information retrieval systems. Largely, this is due to the focus of this thesis being on identifying the challenges of international data management. As such, a basis of understanding of both database management systems (a form of knowledge organisation system) and information retrieval systems is prudent, to allow results of the conducted study to be effectively discussed and understood. Below follows as such an overview of information provided in particular through the literature written by David Bawden and Lyn Robinson (2012) and G. G. Chowdhury (2010)

Information retrieval systems (IRS) are part of a larger group of what can be called Information technologies (Bawden & Robinson 2012, p. 131). Usually, the concept of information technology (also known as IT) is associated with the wider scope of computers and networks, however, as intended with the original meaning of the concept, information technologies include all tools and machines used to assist in the creation, dissemination and use (read retrieval) of information (Bawden & Robinson 2012, p. 132). Focusing on the digital aspect of that, a fair assumption would be that all information today is digital, has been digital, or may be digital (Bawden & Robinson 2012, p. 132). What can generally be said about those involved with IT, and by extension IRS, is that they are involved in information architecture, a quest of finding the best way of organizing and structuring information spaces, supportive of their user’s needs. Within information architecture, there are some agreed upon pragmatic principles following Dan Brown. **Table 2** showcases these principles.

Principle	Significance
Objects	Treat content as a living thing, with a lifecycle, behaviours, and attributes; recognize different types of content and treat them differently
Choices	Offer meaningful choices to users, keeping the range of choices available focused on a particular task. A greater number of options can make it more difficult for people to reach a decision: people think they like having many options, but they don't (the 'paradox of choice')
Disclosure	Show only enough information to help people understand what kinds of information they will find if they dig deeper. Comes from the general design principle of 'progressive disclosure': people cannot use information they are not yet interested in, or do not understand
Exemplars	Describe the contents of categories by showing examples; it's the simplest and most effective form of explanation
Front and side doors	Assume a majority of users will come to any page or piece of information other than through the home page and prescribed navigation routes; typically, they come via search engine. All pages should tell the visitor where they are, and what else is available; the home page should focus on orienting new users.
Multiple classification	Offer several different classifications for browsing content; allow for users' different mental models, even for quite restricted sets of information.

Focused navigation	Don't mix apples and oranges in a navigation scheme; provide access by different mechanism, for example, topic, timeliness, services, use facet analysis principles.
Growth	Assume the content you have today is a small fraction of what you will have in the future; allow for growth by, for example, having a few main categories and making it easy to create sub-categories.

Table 2. Table containing Brown's (2010) 8 pragmatic principles of information architecture. Table content reproduced from Bawden & Robinson (2012, p. 144). Restructured into two columns for ease of reading.

The principles above are central to information architecture, aspects to keep in mind when designing systems to organise and manage information, as well as when designing systems for retrieval. For this MA thesis, the systems of interest are databases, IRS and possibly digital libraries and repositories. Databases is generally a vague and over-used term, as it is applied as a general term to any collection of digital information (Bawden & Robinson 2012, p. 147). A more restrictive definition are systems that handle structured data, typically in number or short text format, and can be called database management systems (DBMS) or relational database management systems (RDBMS). With the difference between DBMS and RDBMS being that the relative means there is *normalisation*, a process which groups data elements into defined structures without redundancy, resulting in a model dependent on relations between elements (Bawden & Robinson 2012, p. 147). Databases tend to use a structured query language to search through its content. One commonly used in many database systems is SQL (Structured Query Language) (Bawden & Robinson 2012, p. 148).

In comparison, IRS tend to handle less structured data than pure DBMS or RDBMS, however, that does not mean that there is no structured data. IRS is more variable, and ranges from more structured data based, such as with bibliographic databases, or with less or no structured records, such as with web search engines such as Google. Generally, IRS is analysed based on system components. Which components range from analyst to analyst. Bawden & Robinson (2012, p. 148) focus on four, namely: input; indexing; search; and interface. Others focus on other

components. Chowdhury (2010, p. 4) uses the model seen in **Figure 1**, which is more adapted to a library and information science perspective.

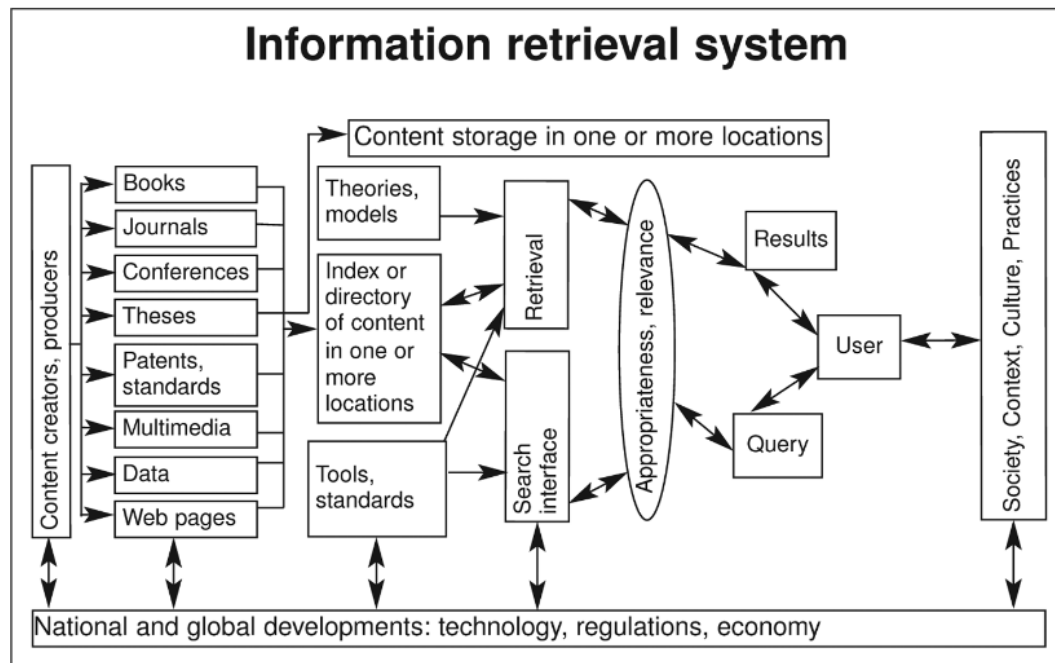


Figure 1. Components of an IRS (reproduced from Chowdhury, 2010; Bawden & Robinson, 2012)

2.2.2 Digital Cultural Heritage

Continuing off the introduction and definition of digital cultural heritage in chapter 1.2.1, current research about digital cultural heritage generally focuses on preservative measures and use-cases of digital cultural heritage. Depending on whether the digital refers to digital in origin or digitised, there are some differences in trends. Generally, research about digital cultural heritage also considers public policy. Below follow some examples from relevant research.

An example looking at originally produced digital content for future reference is the dissertation by Olle Sköld (2018), who draws on disciplines of library and information science as well as archival studies. He does this to search for insights for practical and conceptual support for how to preserve sociocultural aspects of videogames, beyond the code and audiovisual resources of the game itself. The results of the study by Sköld provides a description and analysis of what information objects videogame-related social media can plausibly hold. It also examines the consequences of collecting community-produced social media and

framing it as documentation of sociocultural aspects of videogame, which according to him is a key issue in videogame preservation.

A researcher who has taken a look at Swedish public policy models in regard to cultural heritage from cultural heritage institutions' perspective is Pelle Snickars (2022). Snickars, whose research is data-driven, introduces his article by contextualising, using the Swedish governments researchproposition from 2016, *Kunskap i samverkan – för samhällets utmaningar och stärkt konkurrenskraft* (Prop. 2016/17:50). Most notably he states that the proposition contains nearly nothing about the humanities subjects, and barely anything about ALM. What it apparently does mention is the increased need to cultural heritage institutions to digitise their archives and collections to further data-driven research, increasing research potential as the collections are largely unused due to limited digitisation.

However, it can be argued that research also touches upon a something called *future illiteracy* (Högberg & Holtorf 2021, p. 264). Högberg and Holtorf (2021, p. 264) conducted a study interviewing sixty professionals within the cultural heritage sector of which the results showed a lack of future thinking, a heavy emphasis on the here and now. Preservation is motivated by a need to transmit and safe keep for future generations, but many times there is a lack of shared strategies regarding future heritage management and the future of said heritage. One thing that might be a cause of this is that we do not know if the accumulation of cultural heritage will turn out to be an asset or a burden, whether the preservation will be wanted or unwanted. It might actually be quite naïve to believe the future will be a grateful participant of what is preserved today (Högberg & Holtorf 2021, pp. 264–265). While it might be too early to say anything for sure, some issues of this can already be observed. Older files that are incompatible with newer systems risk not only corruption and corrosion, but also complete loss of access. A real-life example of this was stumbled upon during the reorganisation project of the Swedish Research Institute in Istanbul's library for visiting researchers. The original registration, including localising information, of materials acquired during the donation of the Gunnar Jarring Collection of Central Eurasia, were kept in FileMaker Pro files, dating back to an older version, which cannot run unless converted to the file format concurrent with the newest program version.

2.2.3 Digital cultural heritage portals in research – museum focused, excluding Europeana.

Some examples of local digital cultural heritage portals are the Swedish digital cultural heritage database Kringla, administrated by the Swedish National Heritage Board, and Carlotta, a database system developed for museums and their collections, used and adapted by some Swedish museum institutions.

Kringla is a co-search service which retrieves information from Swedish museums, archives, and registers, consisting of objects in museum collection to buildings of interest. The portal uses two distinct search-inputs, namely free text search and map-search. With map-search results are retrieved based on proximity to a defined geographical area. Search results are able to be filtered after conducted search query. Kringla is managed by the Swedish National Heritage Board (Riksantikvarieämbetet 2017). The search service does not own or contain much of the information it provides access to, instead it makes use of a collection of five Application Programming Interfaces (APIs) to retrieve and present information, although Kringla's "about" page only lists four of them, namely:

- Information about items is retrieved from K-samsök/SOCH
- Spelling from LIBRIS
- Maps from Google Maps
- European cultural heritage data from Europeana.

(Riksantikvarieämbetet 2017)

Mentions of Kringla in research are few, and in the quest of literature in connection with this thesis, only two mentions were found. In the first mention, the search portal was mentioned in the context of conducting a qualitative content analysis of web archive initiatives online presence of which the analysis portion focuses on "voice" (Severson 2017). The second instance the portal was found was in the MA thesis of Anna Rössle (2018), which look at how two Swedish image databases have decided to apply metadata in relation to gender.

Carlotta, as mentioned in the introducing paragraph of this subchapter, is a database system developed for museum collections. The system is owned by the Swedish museum collective called the National Museums of World Culture, which includes the Museum of Ethnography, the Museum of World Culture, The Museum of Far Eastern Antiquities, and the Museum of Mediterranean and Near Eastern Antiquities. The initial concept for Carlotta was to create a flexible system which can be used and adapted to different types of museum collections. Carlotta is based

on CIDOCs⁴ international way of naming fields, adjusted to a Swedish context such as SWETERM in collaboration with CIDOC (Malmö museer 2024). SWETERM is a collection of uniform procedures for term descriptions in Swedish museum database systems, developed by the INSAM (Informationssystem i samverkan vid Svenska museer) project of the 1990s (Sverige 1997, p. 149; Komsell & Melén 2007, p. 30). SWETERM builds upon CIDOCs Conceptual Reference Model (Komsell & Melén 2007, p. 30; CIDOC 2024b; Malmö museer 2024). Carlotta is structured as such that information is registered in various registers. Every museum chooses on an individual level whether to use just one or multiple. After, it is possible to relate different register records. Both client and server parts of Carlotta are JAVA-based (Malmö museer 2024).

There are three notable mentions of Carlotta in literature. Chronologically, the first mention found in academic literature that was found during literature searches for this MA thesis, is a thesis written by Lina Komsell and Hanna Melén (2007) with the aim to describe and evaluate the usability of Carlotta and one other database system. The second mention was in a research paper by Hassan Taher, Giuseppina Addo, Pille Pruulmann Vengfeldt, Mara Engberg and Åsa Harvard Maare (2022). A paper which explores the possibility of reusing digitised material in a participatory setting. Within the study, Carlotta served as both the origin of the items used for reusing, and the destination of the participant contributions, for a circular remediation process. The third and most recent mention was in a research paper by Admeire da Silva Santos Sundström (2023). The paper focuses on ethical issues within knowledge organisation systems following the implementation of decolonial practices of indigenous collections. Her study makes use of an interview, as well as a content analysis of Carlotta database, due to the identification of problems related to classification and terminology in the database following the Museum of World Culture's implementation of decolonial practices for its indigenous collection.

The literature references above about the two digital cultural heritage portals *Kringla* and *Carlotta* are relevant to this thesis as they serve to illustrate how little research about databases with a museological focus there is. The research that does exist touches more on themes in relation to their usability and how museum practices and attitudes can be gleamed from their online catalogues, or the databases are a tool to access data for other research purposes.

⁴ CIDOC refers to the ICOM International Committee for Documentation. ICOM being the International Council of Museums.

2.3 Overview of research about Europeana

When it comes to literature published about Europeana, its portal, database or organisation, there are generally two different trends. The first noticeable trend is literature focusing on the effects and implications of the existence of Europeana. Common topics include general effects on society, but more specifically power structures, discourse, cultural canon as well as international relations and influence. The second trend is literature focused on data science and research. This trend can also be described as focused on performativity, as studies within this trend tend to look at aspects connected to functions and use. Generally, this research focuses on information models, APIs, and IRSs in broad terms. Some of the research is published independently, by researchers unaffiliated with the foundation of the European commission. Some is published by affiliated individuals or groups. But literature ranges from thesis', reports, conference papers and academic articles, and literature can be found in various languages. The presented literature is mostly in English and Swedish though, with an odd one in Spanish.

Interestingly, even the descriptions of Europeana differ, based on which paper one refers to. At times, Europeana is referred to as a search engine aggregating metadata (Petras, Hill, Stiller & Gäde 2017), at others it is referred to as a portal aggregating metadata (Capurro & Severo 2023). In some earlier research, Europeana is still referred to as a digital library (Valtysson 2012), likely due to its origin and initial design. There are also instances in which Europeana is referred to as a database (Andersson & Othén 2013; Stainforth 2022), or a digitisation project (Stainforth 2016). Sometimes, Europeana is described as having its own collection (Capurro & Severo 2023), and sometimes it seen described more as just a portal or access provider (Petras et al. 2017).

Following the division introduced in the first paragraph, one noticeable research trend in regard to mentions of Europeana, focuses on either the effects or the influence exerted by Europeana – whether this refers to a database, a search engine, or some kind of authority for the digital cultural heritage sector. The effect or influence in question depends on the background and field of the researcher, and topics vary, with the unifying factor being a common focus. Some literature chooses to separate research based on what effect or influence, or based on what aspect of Europeana is focused on, as seen by e.g. Carlotta Capurro and Marta Severo (2023). However, this thesis has chosen to group these together, instead separating research focusing on effect and influence, from more technical oriented research.

Interestingly, some research within this first focus category makes use of terminology similar to ANT, which is used as a theoretical perspective in this MA thesis, albeit without explicitly stating such. It does support the view and study of Europeana as some kind of actor, though. Notably, the equating of Europeana as an actor is done by Capurro & Severo (2023, p. 3) in the context of considering the impact of Europeana being considered as a techno-cultural construct with socio-economic structures, and thus as a socio-political actor. Capurro & Severo (2023, p. 3) even argue that all the different components of Europeana should be considered to fully understand its influence in the cultural heritage sector – essentially reiterating an approach similar to the application of ANT.

Due to the varied nature of the literature, it can be difficult to provide an overview without excessive referencing. Instead, a few examples can be found below, which align with the first focus area. They serve to illustrate ways the focus area is covered through different aspects of Europeana studied.

One study focuses on EDM and the role it plays in governing local and national heritage institutions (Capurro & Plets 2020). Another on digitisation of cultural heritage in relation to sustainable development, and Europeana's influence thereupon (Macri & Cristofaro 2021). A thesis within the focus area discusses object selection for Europeana and how it influences identity (Andersson & Othén 2013). The researcher Elizabeth Stainforth (2016, 2022) has two published papers concerning collective memory, a topic which is also focused on by Bjarki Valtysson (2012).

The second area of focus within literature concerning Europeana tends to lean towards exploring the performativity of the portal and systems. Some of the research is conducted in connection with projects in connection with joining or evaluating having joined the Europeana Aggregation Network, as e.g. the study by Agenjo Xavier and Fransisca Hernandez (2020). Their paper summarises the development of Hispana, the Spanish national aggregator, and analyses the main challenges of cultural heritage information systems, including the effects of the data structure of EDM on European and Spanish memory institutions. Another type of literature focusing on performativity is those focusing on evaluations connected to Europeana (Petras & Stiller 2017; Stiller & Petras 2018).

3. Theory

The theoretical framework for this MA thesis consists of a combination of 3 theoretical perspectives, that each have certain similarities and overlap. Actor Network Theory (ANT) spearheaded by Latour, Callon and Law; the Theory of Infrastructure (TI) in accordance with Star, Bowker, Ruhleder and Edwards; and Database as Discourse (DD) as described by Poster following Foucault's determinations in regard to discourse. Each of these theoretical perspectives will be explored in the following subchapters. A brief overview of their application within the thesis and their connection is provided below.

ANT can be described to function as an overarching way of considering the multitude of factors the results of the study might encounter when exploring the layers of metadata management within an international context. The application of ANT allows for various actors and networks to be identified. What is particularly beneficial is that actors do not necessarily need to be of human nature, and can instead be both inanimate and animate, physical and digital.

TI can then be used to identify which links or parts of the infrastructure that is the database, are most susceptible to faults or issues. According to TI, infrastructure is defined as a network of components that relate to each other (Sandvig, 2013) and follows the set of dimensions as proposed by Star & Ruhleder (1996). In essence, within the context of ANT, TI is used to both identify and deconstruct the components or actors that allow the network to function, and to explore its weaknesses.

Database as Discourse (DD) is used in particular to strengthen the idea that the database explored (Europeana) is a valid and complete actor and/or network. The use of DD allows the database as an entity to be understood as a “discursive production which inscribes positionalities of subjects according to its rules of formation” (Poster, 1995, p. 88).

What DD especially contributes to the study is the idea that the database as such is remote from any authorial presence, while at the same time being authored by many. The database is no one's and everyone's, and at the same time, does belong to someone, a state, an institution, or an organisation. The above leads to DD being applied and used to discuss the database Europeana in relation to the literature

overview, and as such is especially relevant for chapter 6. Discussion. DD is thus relevant to topics up for discussion in relation to cultural canon and power, as well as for some of the challenges presented within the ANT model connected to especially the metadata chart.

3.1 Actor Network Theory – ANT

Bruno Latour (2005) introduces the Actor Network Theory (ANT) in connection to his work *Reassembling the Social: an introduction to Actor-Network-Theory* where he proposes the use of ANT to analyse what he refers to as ‘the social’. The theory has its origins in the 1980s, spearheaded by the researchers Bruno Latour, Michel Callon and John Law (Latour 2005, p. 10; Law 2009, p. 142). The term itself was devised by Callon in 1982 (Law 2009, p. 142), although before that it was at times known as the ‘sociology of translation’ (Callon 1984, p. 197). It is through the application of the sociology of translation to the study of the roles played by e.g. science and technology on, in the case of Callon (1984), structuring power relationship, that ANT developed. So, broadly, it can be said that ANT is often used specifically to study the relationships between science, technology and society (Åsberg, Hultman & Lee 2012, p. 40). However, the idea’s that make up ANT predate the 1980s, and in John Law’s words “[ANT] is itself a network that extends out in time and place, stories of its origins are [...] in part arbitrary” (Law 2009, p. 142).

ANT is an approach, a lens, through which data is viewed and analysed based on relations (Law 2009, p. 141). While ANT is often applied as a theory, there appears to be some dissonance regarding this consideration. Some, like John Law, question whether it is a theory at all. Rather, Law refers to it as an approach. This, due to that theories usually try to explain the why to phenomena. According to Law, ANT is descriptive rather than foundational. It tells *how* rather than *why* relations do or do not assemble (Law 2009, p. 141).

the actor network approach thus describes the enactment of materially and discursively heterogeneous relations that produce and reshuffle all kinds of actors including objects, subjects, human beings, machines, animals, “nature,” ideas, organizations, inequalities, scale and sizes, and geographical arrangements.

(Law 2009. p. 141)

Law goes on to explain how ANT can be viewed as a “toolkit for telling interesting stories about, and interfering in, [...] relations” (Law 2009, p. 142)

Since the MA thesis aim to investigate the challenges faced in metadata management in an international context through the case of Europeana, ANT is indeed intended to be used as a toolkit. The theory or approach, depending on which side of the discussion one places oneself, allows for the mapping of challenges based on a variety of actors as well as how they are inevitably linked together. The allowance of non-human actors within the network lends itself well to exploring aspects of data-management, as it means that Europeana, the database, can be considered both as an actor and a network. Same goes for the Europeana Foundation, the organisation, and the larger Europeana Aggregator Network.

3.2 Theory of Infrastructure

Within the thesis, the Theory of Infrastructure (TI) is used to extend the application of ANT. It can even be claimed that TI is not dissimilar to ANT and fits within the post-humanistic paradigm. Within TI, infrastructure is defined as a network of components that relate to each other (Sandvig 2013), as is introduced in the first paragraphs of the theory chapter. In many ways, TI can be thought of as relatively straight forward. In its simplest form, it builds upon the definition of the word *infrastructure*, explained as the structures which runs underneath structures (Star & Bowker 2010, p. 1), a definition concurring with the etymology of the word. The word *infrastructure* has a French origin, based on the Latin Language. It uses the Latin prefix *infra-*, generally meaning below or underneath, to gain its meaning (Oxford English Dictionary 2023a, 2023b, 2023c). The above can be said to be the base upon which TI has been built and developed. Just like ANT, TI is a theory used much like a lens through which the researcher views the research subject, and guides reflections and conclusions. And similarly, it focuses on the *how* rather than the *why*, or the *what*. TI means to look at something, a process, organisation, structure, collection of things, in the structural sense. It focuses on *how* it works, or doesn't work (Sandvig 2013, p. 90).

So, if simplified infrastructure is a network of components, or parts that allow for a task or function to be performed (Sandvig 2013, pp. 90–95), what is infrastructure within TI? Infrastructures can be described in several ways, of which many descriptions, definitions and theories overlap. Within TI, an infrastructure follows the dimensions proposed by Susan Leigh Star and Karen Ruhleder (1996): *Embeddedness*; *Transparency*; *Reach or Scope*; *Learned as part of a membership*; *Links with conventions of practice*; *Embodiment of standards*; *Built on an installed base*; and *Becomes visible upon breakdown*. It is the culmination and configuration

of these determined dimension which form infrastructures (Star & Ruhleder 1996, p. 113)

Dimension	Meaning
Embeddedness	An infrastructure is hypothetically sunk into other structures, social arrangements, and technologies
Transparency	An infrastructure is transparent to use. It does not need to be reinvented or assembled for each task, instead infrastructure invisibly supports each task
Reach or Scope	An infrastructure reaches beyond a single event or one-site practice
Learned as part of a membership	An infrastructure is familiar and naturalised to members of the community of which it is a part of. Strangers encounter it as something to be learned about
Links with conventions of practices	An infrastructure both shapes and is shaped by conventions of practice of the community
Embodiment of Standards	An infrastructure tends to plug into other infrastructures and tools in a standardized way
Built on an installed base	An infrastructure does not develop from nothing. It inherits both strengths and limitations from earlier iterations and its base or origin
Becomes visible upon breakdown	An infrastructure (the invisible cogs) first becomes visible when it breaks, when something stops working and needs to be repaired or adjusted

Table 3. the dimensions for infrastructure as defined by Star & Ruhleder (1996, p. 113) presented in a table for ease of reading.

The dimensions of infrastructure found in TI, not only define infrastructure, but also essentially supplies a guide. The guide can then be used to identify infrastructures, but also a guide which helps analyse infrastructures. Its application helps identify components, parts, and aspects. In essence, TI allows for the identification of influences, and provides a guide to help identify actors for the actor-network. For components, made up by parts, can also be written as actors, made up by a network of actors, and infrastructures, a connected network of components, a network made up by actors.

Theory of infrastructure is not used to argue the case of Europeana being an infrastructure. As this thesis operates on the assumption that Europeana is a database, it also follows the assumption that all databases are infrastructure, as argued by Francis Hunger (2018, p. 53). The theory is instead used as an analytical tool to bring nuance and context to the empirical data collected through the conducted field study by using the dimensions of infrastructure.

3.3 Database as Discourse

The consideration of databases as discourse, or Database as Discourse (DD) as a theoretical perspective is used to strengthen the position of the database as a valid and complete actor and network, as posed by this thesis. Applying DD allows the database to be considered as an entity in its own right. It also serves as a way of interpreting the challenges and causations of some sections of the analysis of the metadata ANT chart found in chapter 5.1.2.

DD builds upon Foucault's theories and determinations regarding discourse. DD uses an interpretation of Foucault's work to show and prove the complexities of databases, linking it to ideas of language (Poster 1996, pp. 78–82). The researcher Mark Poster (1996), proponent of DD, argues for the consideration of databases as a discourse due to databases affecting the constitutions of the subject. He explains that databases are essentially a form writing and inscribing traces. It extends to the principles of writing as a "différance" (Poster 1996, p. 85), a way of differentiating and distancing, leading to an ultimate realization. Databases are transferable and indefinitely preservable, especially in its electronic and digital form, as a database may last forever everywhere (Poster 1996, p. 85). What differentiates the database from spoken or other written language, is that the database in a sense is remote from any authorial presence while at the same time authored by so many that, according to Poster, it makes a mockery of the very principle of having an author as an authority (Poster 1996, p. 85). The database is no one's and everyone's, while still

belonging to someone, a state, an institution, an organisation. Most importantly, Poster defines the database as “a discourse of pure writing that directly amplifies the power of its owner/user” (Poster 1996, p. 85).

Poster goes on to speak of databases, through Foucault’s notion of discourse, as exteriorities, and looks for their rules of formation as a key to the way databases constitute individuals. He explains that essentially, databases are carefully arranged lists, digitalized to take advantage of the speed of computers. The lists are then partitioned into *fields* for items denoting information about an entry, which is organized in *records*. When databases contain the same field, one can be used to cross-reference the other. This means that these databases, or electronic lists, become additional identities that are constituted for the computer. If a database contains information about human individuals, Poster means that these become additional social identities, and the database in question, a social agent (Poster 1996, pp. 87–88). More broadly, this leads to the comprehension of the database as a “discursive production which inscribes positionalities of subjects according to its rules of formation” (Poster 1996, p. 88).

Some might argue against considering databases as discourses and might even use Foucault’s application of discourse to broader prose containing subjects to deny that databases are or can be considered as a discourse (Poster 1996, p. 88). However, most importantly, databases are grids of specification, something that constitutes one of the three rules of formation for discourses. The grids are systems through which subjects are divided, contrasted, related, and classified. Poster presses that databases are pure grids, organized in “vertical fields and horizontal records and classify objects with a precision that more traditional forms of discourse such as psychology must surely envy” (Poster 1996, p. 88).

Due to the focus of earlier research on the influences of Europeana and the effects of its existence, Database as Discourse is a theory used for the discussion of this thesis in relation to themes found in earlier research. The connection between Europeana and the European Commission leads to questions related to power that cannot be entirely avoided, despite this thesis’ focus on the challenges of metadata management.

4. Methodology

The research design follows that of the case study. As such, it shares many of its limitations. Due to its nature, the case study is generally impervious to generalisation, as findings are limited to the case. However, a case study can serve as a starting point to identify larger underlying issues that are prevalent within a field of research or phenomena. It particularly enables in depth descriptions and enhances understanding of chosen phenomena by providing nuance (Choemprayong & Wildemuth, 2017).

The primary data collection method chosen is that of the go-along (Kusenbach, 2003) or shadowing (Czarniawska, 2014). This method combines an observational method with interview techniques, in a more informal and ‘natural’ environment. This method was chosen following considerations of the posed research questions, which following reflection, are closely related to the work practices and procedures connected to metadata management. The go-along allows for a kind of show and tell, which mitigates some of the traditional limitations connected to observations and interviews. To support and supplement findings of the primary method, secondary methods are used for some of the empirical data collection, including unstructured interview techniques, the analysis of documents and websites, and the taking of screen images. These secondary methods are used when the primary method is not applicable or when something more tangible is needed to illustrate and exemplify points raised through the go-along method.

4.1 Research Design – The Case Study

The case study as such is generally impervious to generalising attempts. Findings are limited to specific cases, hence the name, and for a case study to become fit for generalising many repeated case studies on the same phenomenon should be conducted. However, a case study can serve as a starting point in identifying larger issues that are prevalent within a field of research or phenomena. It enables in depth descriptions and enhances understanding of chosen phenomenon by providing nuances (Choemprayong & Wildemuth 2017, pp. 51–55). As such it cannot be disregarded as a valid method of study, albeit with maybe limited importance to said field. The choice of research subject for the proposed study being Europeana builds on the digital advancements and the increased importance of a collective heritage as well as both interoperability between systems and institutions as well as

transparency towards users and subsequently citizens. It is an environment where these topics and issues are likely to arise, and thus would make a relevant subject of research when exploring said topics, especially within an international context.

4.1.1 The Go-Along

The Go-Along method can also be found as a method under the name shadowing and bears a similarity to participatory observation methods common within ethnographic research. The MA thesis' preference towards the Go-Along as the name of reference for the method is due to the meaning associations caused by the way other terminology is used. Would one e.g. refer to the method through the word shadowing, an impression is created of a silent and over-shadowing observing and might introduce a judging or evaluating element the researcher wishes to avoid. The term go-along instead gives the clear impression of what such a thing might entail, reminiscent of education and teaching, as such the observer is supposedly more of an open mindset and willing to absorb what the observed individual might want to impart on them. While the method is used and applied by different researchers, such as e.g. Czarniawska (2014, 2022), this thesis follows the go-along method as stipulated and described Margarethe Kusenbach (2003).

The Go-Along of Kusenbach (2003) is at times also referred to as a walk-along or ride-along, dependent on the capacity in which the researcher follows along with the informant. She uses the walk-along specifically for following an informant on foot, while the ride-along refers to when the researcher sits with an informant in a/their vehicle. The stipulations of the go-along remain largely unchanged, with the same methods of documentation relevant as well as the mindset. Due to the unpredictability of the conditions of the study, the go-along is considered the most applicable term for the method in regard to this MA thesis. The method itself appears as a fitting way of collecting empirical data in the pursuit of answering the posed research questions due to its description as a combination method, associated with a perceived both ask questions and respond to the data collected (Kusenbach 2003).

At its core, the go-along attempts to combine the strengths of primarily two methods, namely: ethnographic observations and interviews. Two tools or methods which are widely used to study everyday lived experiences (Kusenbach 2003, p. 458). The go-along is especially conducive to explore two key aspects of lived experience, namely: “the constitutive role and the transcendent meaning of the

physical environment, or place” (Kusenbach 2003, p. 458). The go-along not only aims to combine the two methods for their strengths, but also negate certain weaknesses found in the application participant observations and interviews. During non-participatory observations, while the actions of an individual in their natural environment might be observed, a participant in the study generally does not narrate or comment upon what is going through their mind while performing tasks or actions, or when responding to situations within their environment. As such, there is a limit to what kind of information and empirical data can be obtained through non-participatory observations (Kusenbach 2003, p. 459). With interview methods there is a contrasting issue, especially in the case of the traditional sit-down interviews. During an interview, whether unstructured, semi-structured or fully structured, an individual might try to explain actions, practices, and experiences, but due to the format, the individual cannot show these to the researcher. The traditional sit-down interview might thus provide a glimpse of something, but in actuality, the format keeps a participant from engaging in their natural behaviour and activities that their ‘natural’ environment would have allowed them to. Which could lead to difficulties grasping exactly what a participant might speak about or refer to. Additionally, the data that is collected during an interview process depends on whether the participants are both able and willing to divulge information the researcher is interested in, or that is relevant to the study. So, with both traditional observation studies and interview studies, important aspects might remain invisible or unnoticed, and if noticed, unintelligible. Especially when it comes to experience and practices in everyday life (Kusenbach 2003, p. 459).

The most basic structure for the go-along, or one could call it its outlining structure, is as follows: fieldworkers or the researcher accompany individual informants on their ‘natural’ outings and actively explore their subjects (the informants) stream of experiences and practices as they move through and interact with their physical and social environment. This is aided by the asking of questions, listening to, and observing the informant. In essence, the go-along is a modest, systemic, and outcome-oriented version of, what is essentially, hanging out, with an informant (Kusenbach 2003, p. 463). What makes the go-along unique compared to other methods is the allowance of simultaneously linking observed patterns with the subjects’ experiences and interpretation (Kusenbach 2003, p. 463).

Methods for data collection within the go-along method vary. Most of it depends on what the subject or informant is comfortable with as well as the researcher’s

preferences, but viable methods range from audio-recording, the jotting down of key phrases and happenings, photography, and sometimes noting things down after the fact either in a natural lull of activity or speech, or after a concluded shorter go-along (Kusenbach 2003, p. 465). In Kusenbach's description of the go-along, she describes it as important to at least try to give informants as little direction as possible, even regarding what she would like her informants to speak about. When prodded for instructions, she would request them to comment on whatever came to mind while looking at and moving through places, as well as share what they usually experienced during routine actions. At times she would also point out hard to overlook features to find out the informants' thoughts about it (Kusenbach 2003, p. 465).

In the study conducted for this MA thesis, the go-along is the primary method for the collection of empirical data. The chosen method of documentation is that of jottings, an extensive field diary detailing containing the practices observed as well as both questions and answers, as well as reflections annotated. Furthermore, the diary and connected folders are supplemented by screen images to aid understanding of challenges encountered and observed at the Europeana Foundation Offices in the Hague, the Netherlands. Due to the nature of the research questions and the focus of the study, the go-along is applied as a guide, but the choice was made to add a short introduction and a little directive to the informants encountered during the fieldwork, to create a more comfortable experience for the informant. The informants are directed to think of their encounter and time with the researcher as a teaching opportunity, or an introductory or a new employee, to treat the researcher in a way not dissimilar to how the informant would introduce the work, processes, and environment to a new employee. This was added following a conducted pilot study, as it notably created a more comfortable and clear experience for the informant, and encouraged the informant to speak comfortably of issues that arise during specific work processes. Further allowances are made in regard to guiding an informant, due to the unpredictability of the location and availability of informant, as well as time-constraints.

4.1.2 Other methods of data collection

Due to the limits of the go-along and the large amount of empirical data a theoretical approach and analysis using ANT needs, there were several other methods implemented for the collection of empirical data. Mostly to supplement and contextualize the data collected using the primary method, the go-along. This means

that all methods mentioned and elaborated below are secondary, or even tertiary, methods.

4.1.2.1 Unstructured interview

As part of the networking and planning stages of the thesis, informational meetings were held with representatives of relevant organisations. In the case of the thesis, this meant digital meetings using Google Meet and Microsoft Teams to meet with one representative of Europeana, who would act as both informant and point of contact, and one representative of SOCH, or K-Samsök, at separate instances. These meetings were largely unstructured with both the researcher and the informant's exchanging information, in regard to the organisations and the project. These meetings were annotated using jottings by the researcher, and some of what was said during these meetings form the basis of further material acquisition. Both representatives made use of organisation-specific presentations and showed some materials from internal sources which are non-replicable. Furthermore, due to limited availability of informants, some sessions that took place during the go-along field study were more similar to unstructured interviews, with aspects of show-and-tell. These are, however, included in the go-along materials, due to precisely the more naturally occurring show-and-tell moments.

4.1.2.2 Documents and Websites

A unique opportunity with working with public institutions as cases for researching specific phenomena or fields, is that commonly, a lot of information and material is accessible via public platforms. This is limited to completed reports, research, initiatives, and information of use to either the public or other institutions, yet enough is available to be used to supplement and enrich empirical data collected through other methods, such as the go-along this thesis makes use of. Europeana, as an example, has both the official database/portal website through which it provides access to digitised objects, but it also has the Europeana Pro website, and the Europeana Knowledgebase, aimed at cultural institutions, both current data providers, and potential providers and collaborators, as well as aggregators. Through these websites, access is also available to documents with tailored information, manuals, guides, and training courses, facilitating access and contribution as well as fostering a deeper understanding of the initiative and its work.

When openly accessible websites or documents are used for empirical data, or as supplement to the go-along study, the information will be reported through the same

reference system as the other chapters of the thesis, and the websites and documents can be found reported in the bibliography when used in other chapters.

4.2 Ethical considerations

Good research ethics are based on transparency. The idea is that as a researcher, one has a responsibility towards the public, your research subject as well as the research community to speak truth about one's research, to consciously evaluate and report one's standpoint, to openly explain one's methods and results. One is responsible to keep good order in one's research, through documentation. A researcher should aim to conduct research without causing harm, injury or distress to humans, animals, or the environment, and one should be fair in one's judgement of other's research. This all is according to the Swedish Research Council (Svenska Vetenskapsrådet) (Hughes Tidlund & Von Unge 2022, pp. 49–51). Since the conducted research study makes use of a qualitative research method, and as such falls within a category of methods and studies where one meets humans and follows their everyday existence, the point of how a researcher should aim to conduct research without causing harm is especially focused on. When contacting individuals, it might be difficult for them to understand or see the consequences of participating in research. This brings to mind the issue of anonymity and confidentiality, which is often discussed when speaking of methods like interviews and observations, which the primary method of the go-along combines. Depending on the focus of the study it might be probable that anonymity to an extent can be ensured. Depending on the topic of the research, being able to identify those that contributed can be harmful to that individual. As such, it is important as a researcher to provide as much information as possible regarding what it means to contribute to the proposed research, so that the individual might make an informed decision. To facilitate this, one might make use of a form of consent, which in an easy to grasp way establishes the studies purpose, method, possible consequences as well as information about how details about the participating individuals will be treated. The form also needs to inform the individual about how participation is one their own terms and can be terminated at will without the need for an explanation. In a way, such a form functions as a type of contract between researcher and contributor (Hughes Tidlund & Von Unge 2022, pp. 51–52).

With the conducted research study, the aim is to keep the informants anonymous. The researcher has followed multiple specialists within the same company to gain as much insight as possible regarding the practices conducted with the active work on metadata management in an international context and more generally the upkeep

of information in and access to the database. Their experiences of occurring issues are central to the study. However, the focus is precisely on what they encounter in their work with metadata. The study does not focus on opinions or incriminating or identifying information. The results are in no way dependent on the identity of the informants, and as such the aim is that all names and identifying information is kept out from both field annotations and other documentation.

The method of the go-along is also not without other risks, not just because anonymity can be hard to completely ensure when conducted within an organization. The presence of the researcher can impact and affect the behaviour of the informant, and might impact their performance, which can lead to negative consequences. Results collected from the method can have unforeseen consequences, as shown by Czarniawska (2022, p. 166), where one of her informants became the victim of a hasty restructuring within the company following their participation in Czarniawska's shadowing research. As the researcher, one might also inadvertently share observations and questions which might affect procedures and practices, which might also influence the results compiled after a concluded go-along. As such, it is important to stay aware of these possibilities as the researcher and to try to explain these possibilities to an extent within e.g. the consent form.

Lastly, an ethical consideration that is important to reflect on is the dilemma of the conversation. Something that both unstructured interviews, semi-structured interviews and the go-along have in common. When participating in a study, the informant might feel inclined that they are supposed to answer all questions. Prolonged time spent together with the interviewer might also mean that the informant or research participant might divulge information during the session or conversation which they might not truly feel comfortable with sharing (Hughes Tidlund & Von Unge 2022, pp. 51–53). This dilemma can also be described as the difficulty of knowing when something is on or off the record, something that might be even more prevalent in the go-along method used in the study, where the researcher has spent a total of 9 workdays at the office and spent hours with an informant. It is mostly a question of the distance between the researcher and the research (Hughes Tidlund & Von Unge 2022, pp. 51–54). While these things might be difficult to avoid, the researcher might mitigate some of these risks through repeatedly bringing up the topic of consent, as well as ask for confirmation whether something might be allowed to be included in a jotting. The boundary between on and off record might also be explained at various intervals, and more detailed in the

beginning of the observation period. With the proposed research study this might entail informing the informant about that all information or conversation shared not relating to the practices applicable to the work conducted with metadata is not relevant to the study and shall thus not be recorded. This also means most if not all conversations not pertaining to the topic will be disregarded. While this might be applicable in theory, at times the reality clashes with the ideas one might have beforehand. As such, continuous reflection is necessary.

4.3 Conducting the study.

As can be expected when doing field work, the theory of the research design rarely completely matches up with the actual conducting of the study. Field conditions heavily influence and impact what can be done and how the researcher can move around. As stated above, the work at the Europeana office in the Hague functions on a largely hybrid model. The busy days, when most attend the office, tend to be Tuesdays and Thursdays, with Tuesdays being the most meeting heavy. Some employees work entirely remotely. Others depend on various factors and requirements. The Go-Along, in the way that Kusenbach describes it, is largely a method applied to the physical space, with the researcher moving alongside their informants in real time, in a physical space, encountering ‘natural’ behaviour and processes.

In the case of the research and collection of empirical data for this MA thesis, certain liberties have been taken due to the structure of the ‘field’. So as not to negatively impact the workday of employees, I was generally present in one of the free offices and operating on an open-door concept, where when employees had time they could simply drop by to discuss aspects of their work, could book standing meetings or invite me along on whatever they were working on. When not engaged with an individual, the time available was spent on structuring and transcribing annotations from interactions, and the analysis of empirical material using the tools provided by the theoretical perspectives of this thesis. The hybrid working model of Europeana also meant that some interactions had to be planned in advance, and conducted digitally, whether through emails or digital meetings through Google Meet. Some interactions were also followed by email interaction due to the ease of sharing documents and other additional materials in relation to the moment of interaction. Being an external visitor to the offices, I had limited access to the internal digital eco-systems, and as such often had to be invited into the system through shared screen functions in digital meetings, screenshots from active

processes and presentation PowerPoints from earlier conferences, to name just a few.

The theoretical framework of the thesis, especially ANT, has been a constant lens through which empirical data was contextualised and placed in relation to other notes. As such, the models used to present a collection of the challenges, found in chapter 5. Results and Analysis, were part of the ongoing process and worked on in the moments in between interaction moments at the offices in the Hague. The chosen theoretical framework in combination with the informants active reasoning make it difficult to separate a clear result from the interpretation of results, as both the empirical data collection throughout the field study, as well as the processing and interpreting thereof relies on deductions of observed and encountered challenges and issues.

5. Results and Analysis

The thesis has chosen to combine the result presentation and analysis, as just a clean presentation of the field notes and other methods would not bring value to the discussion. Rather, it is through the application of analysis that field notes and other collected materials can contribute value. The results and analysis chapter opens with a presentation of results using ANT, aided by created charts of important actors, processes and connections. Within the ANT subchapter, special focus is allotted for what is dubbed the Aggregation loop, and the branch of Metadata. Of special interest in these headings are examples from or about particular aggregators that supply data to the Europeana database. After this presentation, the chapter continues with a subchapter using TI to analyse the collected empirical data in relation to the eight dimensions of TI.

5.1 Actor Network Presentation

During the field study conducted at the Europeana Foundation Headquarters at the Royal Library in the Hague, the notes lent themselves to the separation of challenges posed. Truly, Europeana can, due to its organisational structure, be called a network of networks, in true actor network fashion. However, due to the constraints of the MA thesis, an in-depth exploration of the network of networks, is not possible. Rather, a broad overview is presented, in relation to the challenges discussed. This overview is represented in **Figure 2**. In particular the branches concerning the Aggregator Network and Metadata are focused on. These descriptions can overlap at times and connect in relation to the bigger issues, as seen in the model with reoccurring mentions. Some branches will be explored while covering the ones explicitly mentioned, in part since they in one way or another allow the processes to happen, which is not adequately presented in **Figure 2**, due to limitations of mapping in relation to keeping the model both uncluttered and accessible.

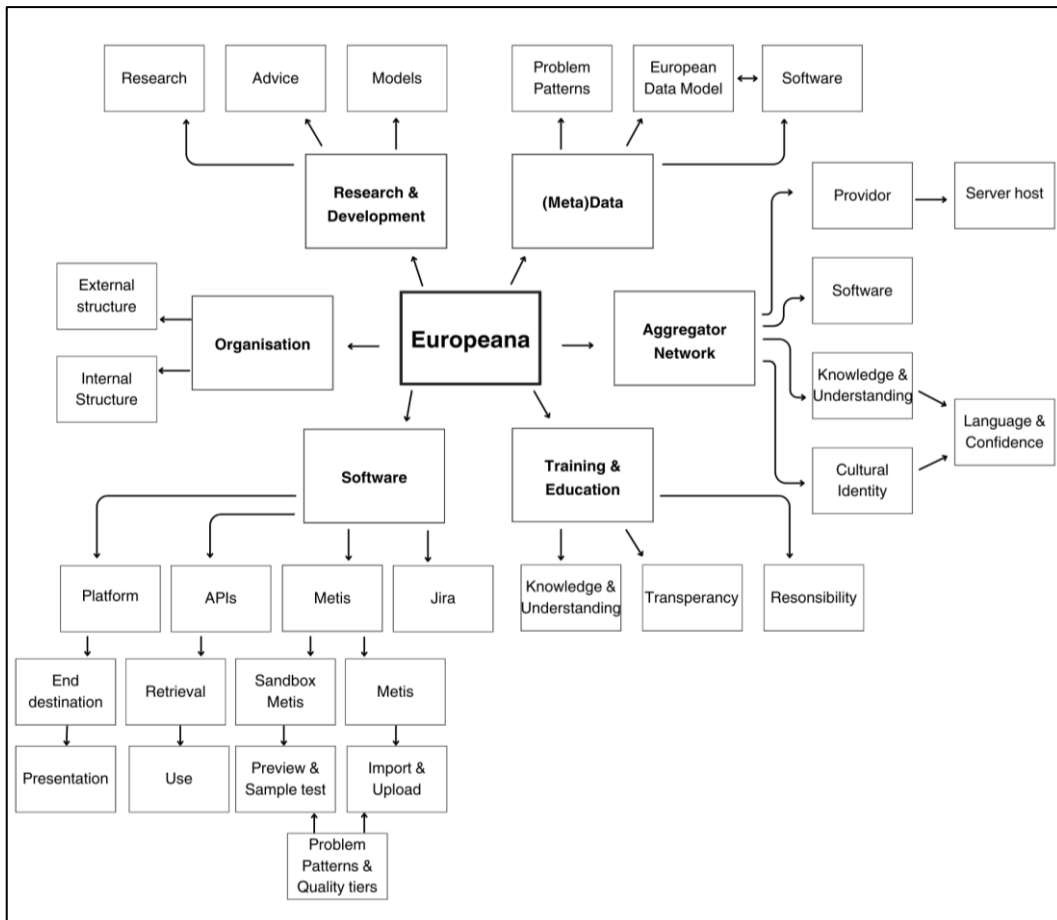


Figure 2. An ANT model of Europeana. The model is based on jottings from the conducted field study at the Europeana Foundation office in the Hague, the Netherlands. Branches are named following the identification of broader and narrower themes in connection to processes, cases and challenges mentioned and encountered during the conducting of the study. The arrows denote the smaller identified challenges or themes in connection to the larger theme blocks. The Model was created keeping in mind the ANT perspective through which empirical data is contextualised and placed in relation to itself.

The ANT model presented in **Figure 2** shows an extremely simplified actor network, with actors and networks named by theme or common denominator as encountered during the conducted field study using primarily the method of the go-along. During the researchers two week stay at the Europeana Foundation office in the Hague, certain challenges, themes, and occurrences stood out, aiding in the creation of the model presented in **Figure 2**. Below follows an overview of some of the relations, actors and networks mentioned in the map. An overview of the organisation, and in particular the Europeana Foundation can be found in **Figure 3**.

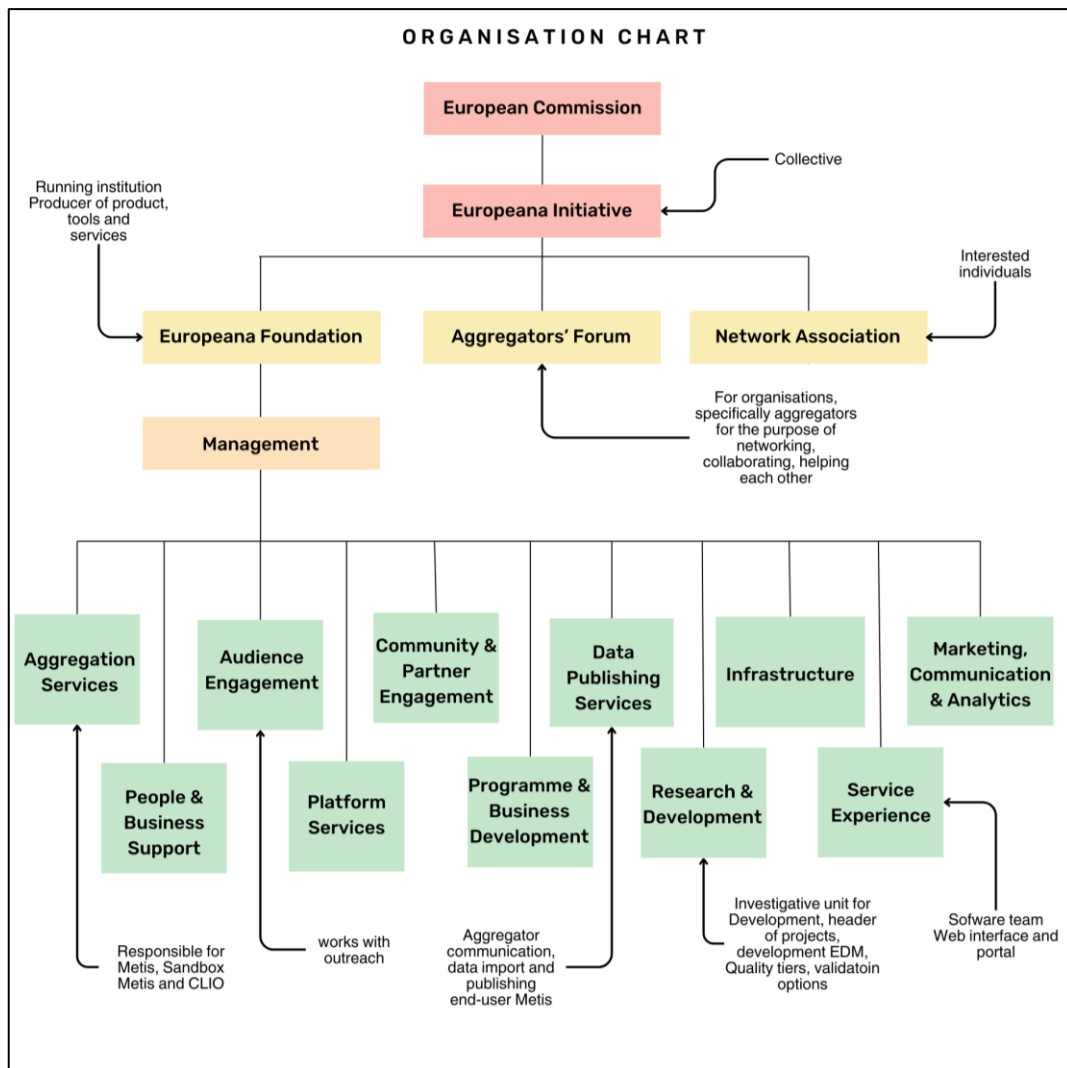


Figure 3. A chart of the organisational structure of Europeana and the Europeana Initiative. Colour coded to denote levels in the organisational structure. Annotated with clarifications for ease of interpretations. The chart was made during the conducted field study to support the ANT models presented in other figures in the hopes of clarifying where in the organisation challenges are noted.

5.1.1 The Aggregation loop

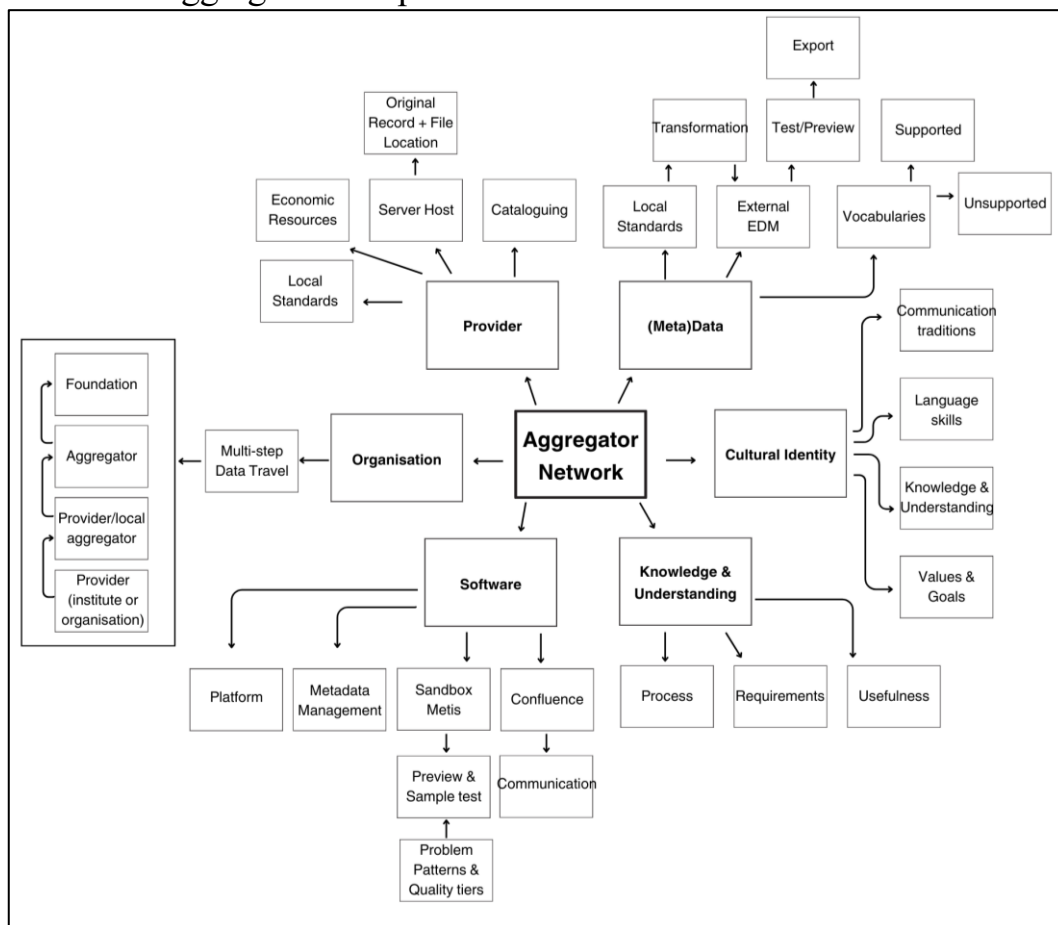


Figure 4. A model of the aggregation network node identified from the model found in figure 2. The model is based on jottings from the conducted field study at the Europeana Foundation office in the Hague, the Netherlands. Branches are named following the identification of broader and narrower themes in connection to processes, cases and challenges mentioned and encountered during the conducting of the study. The arrows denote directions of steps in flows and processes, or simply some of the smaller identified challenges or themes in larger theme blocks. The Model was created keeping in mind the ANT perspective through which empirical data is contextualised and placed in relation to itself.

Within specifically the aggregation loop, as the section of the actor network concerned with the aggregation, there are in general 6 core areas of challenges, as seen in **Figure 4**. Just like in the broader chart of **Figure 2**, there are certain areas that overlap, and influence each other, as is expected using ANT. Below follows a general description and analysis of the core areas of challenges, followed by some specific examples of encountered or described situations where challenges presented themselves during the conducted field study and adjacent informational meetings.

One such core area is that which this thesis would like to call the *data flow*. This references the travelled path the data follows from record creation (locally) to record availability (through Europeana). And includes challenges posed by local organisational and decentralised structures. The flow through which data moves is extremely dependent on local organisational structures and is heavily case dependent. Some countries have a national aggregator, acting as a buffer between local providing institutions and Europeana. Some countries don't have a national aggregator, and dependent on the content of the providing cultural heritage institution's collection, they might be matched with a theme-based aggregator within the network. Some countries have local aggregators as an in-between step before the data reaches the national aggregator. Either way, generally the data flow is a multi-step process, and issues with the data, regardless of what the issue is, can occur at any one of these stages.

Certain types of challenges depend on certain limitations. These can be resource based, or they can be knowledge based. Common resource challenges include:

- Software limitations
- Server limitations

Knowledge-based challenges are a bit more difficult to generalise, although general trends noticed by the staff at the Europeana Foundations, knowledge-based challenges present themselves in a lack of understanding of the general information process, a lack of awareness of quality requirements, as well as not knowing which type of data is 'useful', or which type of data is compatible with the standards set by Europeana.

Beside challenges and aspects connected to the actual flow of data, or technical and resource-based challenges, there also stands to argue in regard to other influences on the network. One such large influence is culture and cultural identity. Culture influences all individuals and organisations, it informs and forms human interaction, organisational structures, and interactions, as well as user needs and query formulation. Something which is hinted at in **Figure 1**.

5.1.1.1 Case 1: The Spanish aggregation system

The Spanish aggregation system was used in many examples put forth by informants during the conducted field study at the Europeana Foundation offices. The Spanish aggregation system is largely decentralised, but falls under the

ministry of culture, as does the Spanish national aggregator. However, two regions/provinces have separate systems and aggregators, namely, Catalonia and the Basque region. These two regions' aggregators are directly connected with Europeana without the intermediary national aggregator. This follows these regions general trends, as both also have their own National Library Catalogue. Outside of the named two regions, the system relies on local providers, providing their data to local aggregators or repositories (or multiple depending on the region and specific set-up), which then supply data to the national aggregator. Generally, all or a significant majority of providing organisations and institutions within the Spanish system are dependent on technological providers to run infrastructure, mappings to EDM as well as software.

The choice of outsourcing brings with it various considerations, some of which were discussed with an informant during the field study at Europeana. This discussion revealed that when looking whether something can feasibly be done inhouse, considerations surround which requirements such a thing would have. More specifically, aspects surround the skillsets required, and subsequent cost and investment into information infrastructure for handling the outsources task within the institution instead. Choosing to outsource allows an institute to not have to worry about such things. Outsourcing does come with some risks and limitations though. Through outsourcing, it can be debated whether the institute in question loses the authority over their collection and its management. If changes need to be made to data records, a company can charge by the record. If a software company decides to change their operating model, or change the way the software functions and operates, changes what services they offer, an institute would (depending on the market of such things) have to go along with the changes. Outsourcing does, however, mean that an institution does not have to divert their focus from their main objective and scope of operation. It allows the time of employees to be spent on other important aspects of the institution.

During the conducted field study, there were a few other examples mentioned and discussed related to the challenges and development possibilities in connection to the Spanish aggregation system. Two such examples are presented below, namely the example of the Virtual Library of Malaga and a new aggregator from the Basque region. Both examples are concerning aggregators directly connected to Europeana, without the Spanish national aggregator as an in-between step.

The Virtual Library of Malaga provides access to 5 different local providing institutions/organisations. An issue encountered with this aggregator is that access to the digital objects is only given through the virtual library portal, and as such, the institution listed on the data sets is that of the Virtual library of Malaga, and not the actual providing institution. The issue itself was only discovered following the request to access information on the usage of the materials provided through the aggregator, based on the institute of item origin. So, due to the way the data is structured, the requested information cannot be provided.

Another recent example of an issue, and subsequent solution, was encountered with the new aggregator from the Basque region. Materials from their collection in the local language were listed as under the language Vasco, an inaccurate and dated name for the language now known as Euskara. The issue was discovered by the aggregator following a data excerpt test using the Europeana service known as Metis Sandbox, which allows providers to test a sample of their data and see whether any issues jump out over a larger portion of the data set they aim to make available. The issue was then flagged to DPS team and taken to the software teams for adjustment allowing the correct language name to be displayed.

5.1.1.2 Case 2: The Swedish National Aggregator (SOCH)

An issue discovered in the early stages of this study, already hinted at during the exploration phase, is the fact that ‘outsiders’ have a varying knowledge base in regard to the processes and requirements posed by Europeana. During an informational meeting at the start of the project with the Swedish national aggregator SOCH the informant mentioned that they were unsure when their next data upload to the Europeana server would take place following a recent issue encountered at their latest upload attempt. The data set upload appeared to have been rejected following an issue with the licensing listed on various records, and to the knowledge of the informant, SOCH was unsure what the issue was about or how to resolve it, leading to a need to communicate and cooperate with Europeana to resolve the possible issue. This type of communication is, however, not something those at SOCH find difficult, as one gets feedback quickly.

The Swedish aggregation process occurs mostly automatic, with providing institutions mapping their data to the data schema and structure of SOCH, which is then automatically mapped to EDM at a later stage. The data within SOCH is continuously updated to match their providing institutions. The process of data export to Europeana is in comparison done manually, where SOCH needs to create a request or work-order to update their datasets in Europeana. The manual nature

of the tasks related to the provision of data to Europeana means that at times this task competes with SOCH's other responsibilities.

5.1.1.3 Case 3: *The support of Ukraine*

In some cases, one can observe more crossover between actors than others. One such case is the situation following the Russian offensive in the Ukraine. During the field study, an informant mentioned it as a sidenote to a discussion about the processes involved in accrediting new aggregators, the addition of new data sets, data requirements and offered support. Following the military offensive attacking the Ukraine, there was a scramble to make available as much digital Ukrainian cultural heritage lest physical items are destroyed. To support this drive to save cultural heritage for the future, there was a working group set up within the Europeana Initiative, fittingly named *Supporting the digital cultural heritage of Ukraine Working Group*. The main aim of the working group is formally to “support existing initiatives and develop new, concerted actions that respond to the current needs of the Ukrainian digital heritage sector” (Europeana 2024c, 2024d).

As part of Europeana's initiatives to help with the preservation of Ukrainian cultural heritage there is also available an informational page via the Europeana Pro portal to the initiative *Laptops for Ukraine*, an initiative run by the European Commission, the Ukrainian Ministry of Digital Transformation and Digital Europe (Europeana 2024e). The laptops for Ukraine initiative “aims to collect and deliver laptops, smartphones and other equipment for schools, hospitals and public administration in Ukraine's most affected war regions” (Europeana 2024e) and the initiative was extended in June 2023 to also “help cultural heritage institutions to digitally document, scan and photograph documents and objects at risk of being lost due to the war”(Europeana 2024e). The page on the Europeana Pro portal it is also explicitly stated that “The Europeana Foundation applauds this initiative and is proud to support it. [The Europeana Foundation calls] on [their] large pan-European network of cultural heritage institutions to donate and take their part in preserving Ukraine's cultural heritage through digitalisation” (Europeana 2024e).

5.1.2 The Metadata Network

With the theme of this MA thesis being metadata management, it stands to reason that one important node of the ANT analysis is that following the processes of the actual data in circulation. Below, in **Figure 5**, a simplified map denoting various steps and phases can be found, as well as some dependencies.

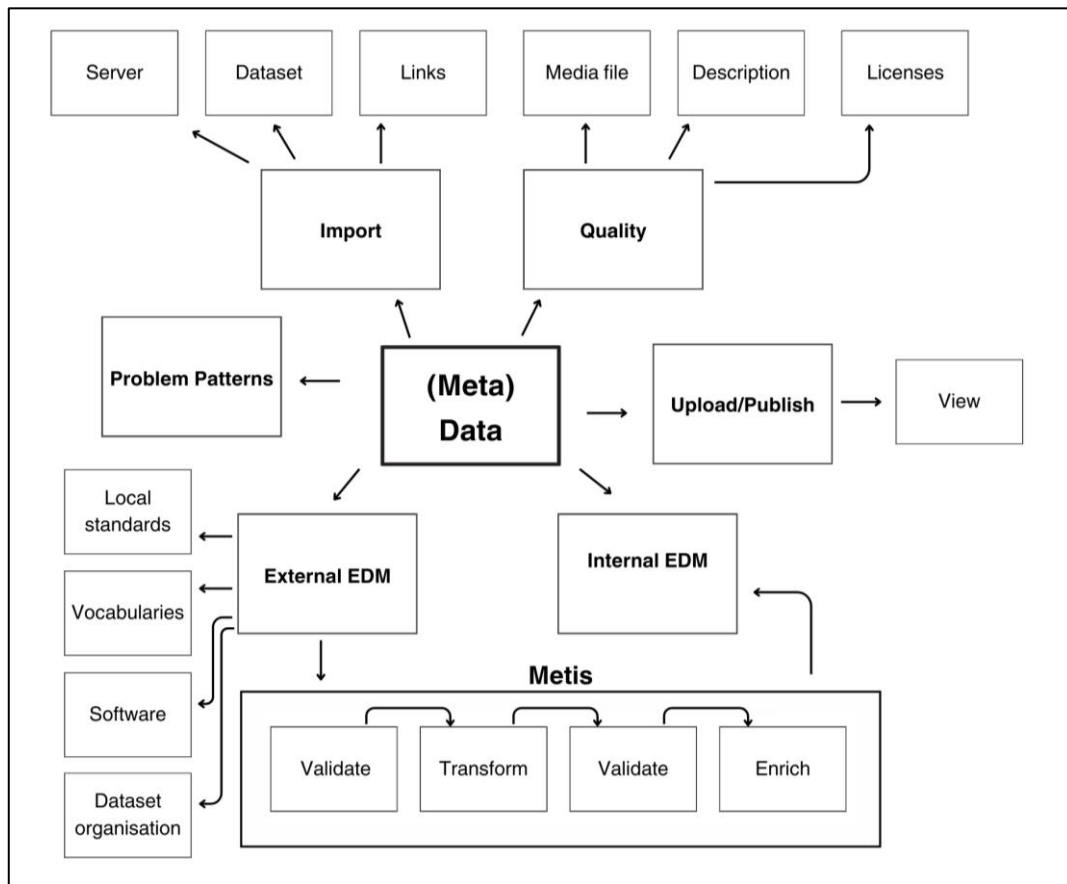


Figure 5. A model of the (Meta)Data node identified from the model found in figure 2. The model is based on jottings, software insights and documents obtained from the conducted field study at the Europeana Foundation office in the Hague, the Netherlands. Branches are named following the identification of broader and narrower themes in connection to processes, cases and challenges mentioned and encountered during the conducting of the study. The arrows denote directions of steps in flows and processes, or simply some of the smaller identified challenges or themes in larger theme blocks. The Model was created keeping in mind the ANT perspective through which empirical data is contextualised and placed in relation to itself.

A central part of the metadata management is the metadata schema, which is accepted and developed by Europeana, also known as European Data Model (EDM). While EDM is briefly introduced in chapter 1.2 Background, there is a bit more to it, as can be found below. Following the EDM description and more detailed insight, this chapter goes on to explain the process involved in the import of datasets, focusing on the process of going from the provided original records in the External EDM schema to the Internal EDM schema and publishing to the

Europeana portal. Finally, before providing specific examples mentioned or encountered during the field study, this chapter also provides an overview and analysis of the work with Problem patterns.

To review what has been mentioned in chapter 1.2.3 European Data model - EDM, the aim for the creation of EDM was to create an interoperable framework for collecting, connecting and enriching cultural heritage metadata (Europeana Foundation 2024b). EDM is based on ESE and borrows from various metadata standards and schemas in an attempt to accommodate a large range of cultural heritage institutions standards, and is modelled to allow for data enrichment from third party sources (Europeana 2013; *Hackweek Part1* 2021). In their own EDM primer, Europeana describes how EDM adheres to the modelling principles connected to the approach of the Semantic Web. This entails an absence of absolute or fixed schemas which dictates only a singular way to represent data. In their own words, EDM functions as an anchor to which other models can be attached, which means that EDM does not require local practices to change even though it does encourage it to increase cross-domain applicability (European Commission. Directorate General for Communications Networks, Content and Technology. 2024).

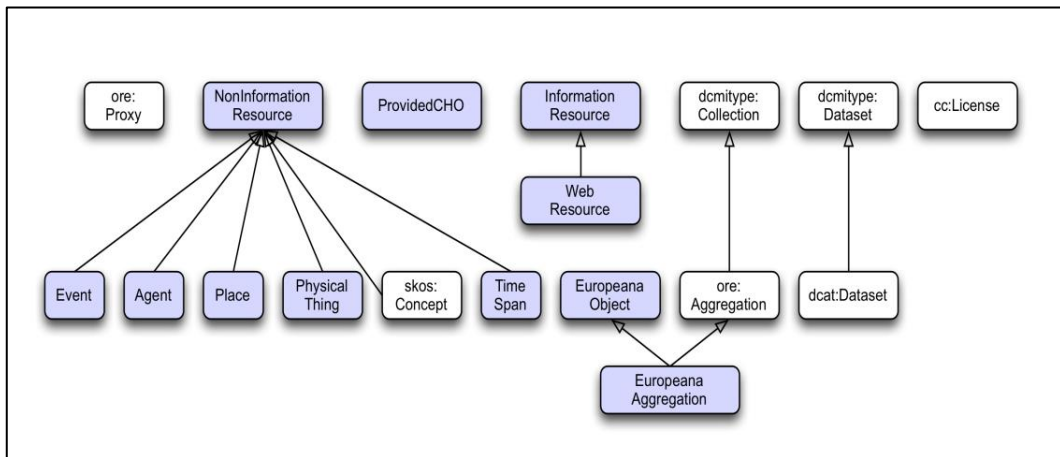


Figure 6. The EDM class hierarchy and names. The classes introduced by EDM are shown in light blue. The classes in white are re-used from other schemas, with the schema indicated before the colon. The model originates from: [EDM Definition v5.2.8 102017.pdf \(europeana.eu\)](#), p. 7

Due to the nature of EDM as an integration medium for collecting, connecting, and enriching the data and descriptions provided by Europeana’s content providers, it is essentially impossible to account for all elements that are contained within EDM.

This is mostly due to the fact that EDM as this integration medium, can be said to include any element found in the original description. This forms an open set of data elements, meaning that the set can be extended as new requirements and providers join the information space. There is, however, a well-identified set of elements that EDM uses to carry out its tasks, dividable into two main categories. Namely: The elements introduced by EDM, and the elements re-used from other namespaces. (Europeana 2017, p. 8) The most common re-used elements come from the following major namespaces (Europeana 2017, p. 8):

- The Resource Description Framework (RDF) and the RDF Schema (RDFS) namespaces
- The OAI Object Reuse and Exchange (ORE) namespace
- The Simple Knowledge Organization System (SKOS) namespace
- The Dublin Core (DC) namespaces for properties from the element, terms and types namespaces
- The W3C Data Catalog Vocabulary (DCAT) namespace
- The Creative Commons (CC) namespace
- The SIOC Services Ontology Module namespace

Interestingly, during the field study it was found that Europeana actually uses 2 versions of EDM, but providing institutions only need one of them. The data accepted by Europeana needs to be formatted according to the EDM External Schema. During the data import it is then transformed using the EDM Internal Schema, which is the metadata format used when providing access to digital items through the Europeana portal and publishing the records.

For the purpose of identifying challenges the above knowledge is needed, as well as a few specific properties and their requirements. See **Table 4** below:

EDM Property or tag	Description and Comment	Obligation/ Requirement
edm:IsShownBy	An unambiguous URL reference to the digital object on the provider's web site in the best available resolution/quality. It will lead users to the digital object on the	Only allowed to occur once Either isShownAt OR isShownBy is Mandatory

	provider's website where they can view or play it.	
edm:IsShownAt	An unambiguous URL reference to the digital object on the provider's web site in its full information context. It will lead users to the digital object displayed on the provider's web site in its full information context.	Only allowed to occur once Either isShownAt OR isShownBy is Mandatory
dc:title	A name given to the resource. Typically, a Title will be a name by which the resource is formally known	Mandatory to supply either dc:description OR dc:title Use language tags (xml:lang)
dc:description	An account of the resource	Mandatory to supply either dc:description OR dc:title
xml:lang	Attribute tag defining language of property content	Must be valid according to ISO639-1 OR ISO639-2 OR ISO639-3
dc:type	The nature or genre of the resource. Type includes terms describing general categories, functions, genres, or aggregation levels for content. The type of the original analog or born digital object as recorded by the content	

	<p>holder: typically containing values such as photograph, painting, sculpture etc. which, ideally, will have been taken from a controlled vocabulary.</p> <p>Providers are recommended to map the values entered in this property to the five material types used in Europeana: TEXT, IMAGE, SOUND, VIDEO and 3D but to keep the original local values in this property.</p>	
edm:type	The Europeana material type of the resource	<p>All digital objects in Europeana must be classified as one of the five Europeana types (in upper case): TEXT, IMAGE, SOUND, VIDEO or 3D</p> <p>Mandatory</p>
dc:subject	<p>The topic of the resource</p> <p>This is the subject of the original analog or born digital object.</p> <p>Recommended to use a term from a controlled vocabulary</p>	Repeat the property for multiple subject term

Table 4. List of excerpts of 'important' EDM properties or tags and their requirements. Sources: [EDM Definition v5.2.8 102017.pdf \(europeana.eu\)](https://www.europeana.eu/edm/definition/v5.2.8/102017.pdf) and <https://docs.google.com/spreadsheets/d/1atZr1w-h9AdWwWSBYLCck6fAdJSxrCNP56QRLNY1jLg/edit#gid=1273976273>

A central part of the data chart and actor network is the process of data import/export. When Europeana receives a data set from a providing institution through an aggregator, the data records provided are in EDM, and more specifically, external EDM. In general, the data set is then processed using the Metis software developed by a team within the Europeana foundation and an external collaborator where the external EDM is first validated with the schema, and then it is transformed into internal EDM, validated against that schema, and enriched. These are the steps shown in the data ANT model in **Figure 5**. In reality, the entire process is a bit more complicated, and includes a few more steps.

WORKFLOW STEP	PROCESSED / RECORDS	START TIME	END TIME	ACTIONS
Pr Preview	0 / 0	—	—	
M Process Media	0 / 0	—	—	
E Enrich	101765 30 / 228703	12/03/2024 19:40	—	Report
N Normalise	228703 / 228703	12/03/2024 16:46	12/03/2024 19:40	
Vi Validate (EDM internal)	228703 / 228703	12/03/2024 16:01	12/03/2024 16:46	
T Transform	228703 / 228703	12/03/2024 15:15	12/03/2024 16:01	
Ve Validate (EDM external)	228703 127 / 228830	12/03/2024 12:37	12/03/2024 15:15	Report
Ho Import OAI-PMH Full Processing	228830 / 228830	12/03/2024 10:17	12/03/2024 12:37	

Figure 7. Screenshot showing a generic example of a dataset being run through Metis, shared by informant.

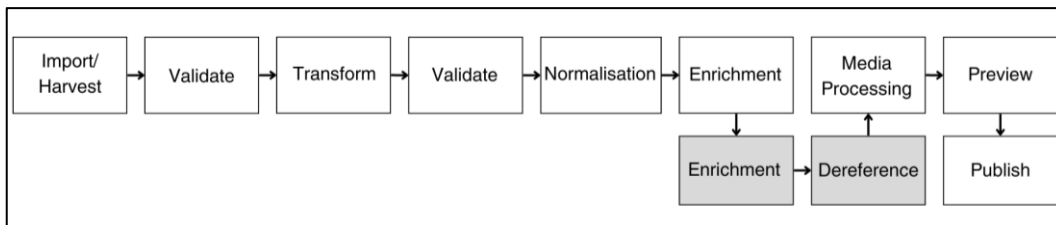


Figure 8. Simplified model of import/publish process of the Metis software, in order. The arrows signify data and process direction, and the grey blocks denote underordered processes included in a process step.

The first step in Metis is known as harvesting or import, as it is shown as in **Figure 7** and **Figure 8**. The import step is done through either OAI-PHM or HTTP files. This step is applicable whether it concerns an entirely new dataset or updates to an existing dataset, as the process can be done incrementally through OAI-PMH, which allows for the selection to process only updated or new items/records.

After import, the dataset is run through an automatic validator run through Metis, which checks whether the data is compliant with the EDM External Schema. When

discrepancies are found, the records are noted, as can be seen in the example in **Figure 7**. The system creates a report, which then can be used as a reference when checking for issues, and in considering possible solutions or causes for occurrences.

The next step in the Metis process is known as transformation, which is a workflow step. A workflow step means that the system actively works with and changes something about the provided record. Metis ‘cleans’ the important data and actively changes the schema and elements from EDM external to EDM internal. This process entails the creation of the Europeana Aggregation class, adds the Europeana Proxy class and `edm:country` and `edm:language` fields are overwritten by information from the Dataset information form. It is also here that Europeana IDs are generated. Internally, these two versions are sometimes referred to as EDM flavours. There are two versions, or flavours, to distinguish between information originating from the provider, which is kept in its original state as much as possible, and the information Europeana and Metis adds to a record. After the external EDM records are transformed into internal EDM records, they are again validated, this time by comparison with the EDM internal schema.

Following transformation, the data goes through normalisation, during which unnecessary spaces are removed and HTML markup tags are cleaned from text values. Duplicate statements are removed, and language identifiers are normalised to ISO standards (namely ISO 639-X). Dates are normalised and encoding errors in available media URIs are fixed.

The next step in the process is generally referred to as enrichment, which is done in two-fold. This step generally enriches records through addition of context and dereferences records through resolving LOD vocabularies. Context is added from literal values, such as places, agents, concepts, and timestamps. It equivalences relations and adds resolved entities to the record as context. Enrichment computes to Europeana completeness, an indicator for richness of data, and determines the year(s) of the records. Dereferencing is instead done through the examination of values of certain fields for references to known external sources, such as Wikidata or Geonames, or to the Europeana entity collection. The sources are mapped to the EDM format using XSLTs if the source is external.

The final active process part of the dataset import through Metis is the media processing step. Certain fields in the record contain media links and file information. These links and the accompanying file information are collected

during this step. During media processing file resources are downloaded through the system to create technical metadata, such as general properties like media-type and file size as well as type-specific properties and to add these to the record. This step demands a lot from the providing server, and if the server is too slow issues might occur and files might not be able to be accessed. There is an option to manually adjust the speed of media file retrieval to adjust to the providing server's capacity, in a process called throttling. Upon retrieval of the file and technical data, the file is deleted.

The final step before dataset upload is to preview. It is here that Content- and Metadata tiers are determined and assigned, and the record is saved to MongoDB, a database designed to preserve the record and all its components and relations. When the dataset is published the records are saved to a different database called Solr, which is designed for searching, and the final version of the records are uploaded to definitive storage so they can be accessed by Europeana's API services.

Analysing the import and publication process as a whole reveals a few challenges, observations which are supported by various conversations had during the duration of the field study. A general truth is that issues can occur at any stage of the process. The causation of these issues falls in one or multiple of the following categories: Lack of knowledge or understanding, use of incompatible vocabularies and formats, system and software limitations and software bugs.

The reports created in Metis, and Metis Sandbox are partly created by errors identified when comparing with metadata schemas but are also informed by established and documented Problem Patterns. The work on identifying these patterns in supplied metadata in EDM has been ongoing since at least 2014, however, quite sporadically. Despite the sporadic focus through the year, issues in supplied metadata have been tracked quite extensively, originally in what is internally known as a "monster" document, filled with both noted issues and improvement suggestions. Currently, there are about 43 identified patterns, of which 5 are actively implemented in Metis Sandbox, allowing for aggregators to identify issues in their supplied data in an early stage, and possibly implement changes.

While the document and process of noted patterns is called Problem Patterns, an informant mentioned that the name is quite inaccurate. Not all noted patterns are truly problems, errors, or faults. Some noted patterns are more about data and record

quality, and recommendations to make data more accessible than it would be otherwise.

Some noted patterns and documented issues can, however, not be amended, changed, or be improved. In part due to the organisational structure, as the Europeana Foundation cannot require changes to be made by aggregators and providers. Some things cannot be changed due to e.g. local legislation, or limited resources. Instead, where possible, compromises are offered, and larger issues are prioritised over smaller ones.

As was mentioned when explaining the preview step in the import/publication process, Metis assigns quality tiers on provided datasets. Quality is hard to measure and to define, but what it means in this case is that Metis checks the data against established criteria, and the more data matches the criteria, the higher the quality tier which is assigned to the data record.

To conclude this chapter and to provide some examples to illustrate discussed challenges, case descriptions can be found below. The titles are indicative of the challenges encountered or mentioned during the conducted field study at the Europeana Foundation office in The Hague, the Netherlands.

5.1.2.1 Broken Links

When clicking on an item on the Europeana portal, and a black screen is shown where an item should be, it is generally indicative that the link providing either the preview, or the link to the actual digital item, is somehow broken. Sometimes the issue is purely limited to the preview window, but not always. The issue of broken links is connected to the EDM properties `<edm:isShownBy>` and `<edm:isShownAt>`.

The ‘break’ in links can happen in two different ways. The first is that the link was non-function from the beginning or might have been in the wrong field. Cases of this tend to be found before publishing as it is an issue that Metis scans and reports and is usually found in the Media processing stage. The second, and more challenging way a link can break, is due to *link-rot*. Link-rot essentially means that a media file link to the providers server (where the digital file is kept) stops being accessible after the record has been published. Link-rot can happen anytime, even nearly a year or more after upload. Occurrences of link-rot were originally quite hard to check, due to it having to be done manually, and as such many occurrences

were never found. This led to the development of CLIO – Checking Link In Operation, by the Aggregation Systems Department.

5.1.2.2 Incompatible vocabulary

The use of unsupported or incompatible vocabularies for subject terminology and field content can cause various issues for the accessibility and retrievability of records. One example of which is that of a local controlled vocabulary being used, which is not supported by EDM. Generally, content derived from locally controlled vocabularies have the content replaced by a URL, even in the view of the record post-publishing.

As of the writing of this MA thesis, Europeana supports and keep up to date with 26 different controlled vocabularies, yet many institutions choose to create local vocabularies for contextual data. Vocabularies which in turn also need to be kept up to date, mapped and documented. Due to the local contextual vocabularies generally not being interoperable with EDM, the use of different vocabularies creates a doubling of work, for all involved.

5.1.2.3 Enrichment

Some issues with the metadata are machine created. During the enrichment process in Metis, the system deals with reference ambiguity, which leads to Metis possibly incorrectly enriching the record with linked contextual data. A past persistent issue was that records of items with unknown or uncertain authors got automatically enriched with Keith Wharton Whitman. This occurs due to Whitman in his Wikidata record being tagged with “also known as: the anonymous”, leading to Whitman incorrectly being tagged as the creator of a large number of items in the enrichment process. An informant explained that this issue is due to encountering a literal value rather than an entity URI⁵. Metis connects the literal value to an entity in Europeana’s entity connection, which in this particular example thus leads to Whitman being linked as the creator in the metadata. To fix the issue, since the system cannot recognise that this is faulty, the term “anonymous” is excluded from the machine automatic enrichment, but the issue cannot be fixed and can thus still be found in older records.

Another interesting example of an enrichment challenge is that of *Inde*. An informant shared a study published specifically on this example. The study explains the issue as follows:

⁵ URI stands for Uniform Resource Identifier.

When a user in Europeana the single access point to European cultural heritage, searches for poison in the collections provided by Swiss institutions, she will find photographs from India and Indian movie covers. The relevance of the retrieved documents to the query is not comprehensible. A deeper investigation reveals that retrieved objects were automatically enriched with the term poison and its multilingual equivalents. In Latvian *poison* means *Inde* which is the same keyword the French-speaking domain expert gave the objects to describe its content: India.

(Olensky, Stiller & Dröge 2012)

The issue, like Olensky, Stiller & Dröge (2012) point out, is due to ambiguities. More specifically, synonyms, homonyms, and cross-lingual ambiguities. These ambiguities can both cause challenges in retrieving the relevant materials based on a search query, but can also lead to software incorrectly enriching records.

5.1.2.4 Persistent Identifiers

A current topic being investigated by Europeana, in part due to a interdepartmental working group but also within the larger Europeana Aggregation Forum, is that on of persistent identifiers for metadata records and digital items. The theme was also brought up during the European Aggregation Forum's online General Assembly which occurred during the conducted field study.

A persistent identifier, sometimes abbreviated to PID, is a “globally unique and long-lasting reference to potentially any sort of digital or non-digital entity, providing the information required to reliably identify, verify, locate and access it” (Europeana 2024f). Essentially, PIDs “ensure that the digital entity is ‘set in stone’ and can always be findable through that identifier” (Europeana 2024f). The absence of a PID can thus lead to difficulties in finding, accessing, and locating digital entities, including data records about said entity. Currently, without PID, when identifying information of data records are changed in some way, the record can get ‘lost’, meaning that it becomes harder to find, or that the connection to other relevant records becomes unavailable. On the Europeana Pro portal page *help us to make cultural heritage data more persistent*, one of the most common issues in relation to the absence of PIDs are broken links, which in the case of Europeana can affect “all forms of (re)use, [and] also negatively impact the traffic towards the institution’s website and its [Search Engine Optimisation] ranking amongst search engines” (Europeana 2024f).

The work on introducing and encouraging the use of PIDS is ongoing within the Europeana Initiative, and the updates given during the Europeana Aggregators Forum General Assembly were the implementation plan for persistence and uniqueness of identifiers in the data space is being refined. Further, a report on the

practices of aggregator towards PIDs and a report on the practices of cultural institutions towards PIDs are completed, but not yet available on the Europeana Pro Model, and I was not able to take a look at the mentioned reports during the timeframe of the field study. The data space annual report 2022-2023 did however identify three challenges to investigate to allow for better support and handling of PIDS, namely:

- 1) stability of object identifiers provided by data partners, 2) stability of identifiers originated by the aggregation into the Europeana infrastructure, 3) stability of identifiers shared with (or reused by) our audiences

(European Commission. Directorate General for Communications Networks, Content and Technology. 2024)

The report also states that about 13% of data records published on the Europeana website “contains a possible PID”(European Commission. Directorate General for Communications Networks, Content and Technology. 2024). This data is based on the identification of data matching one of three candidate PID schemes in use that are known to be supported by reliable PID policies: ARK⁶, HANDLE⁷ and DOI⁸.

5.1.2.5 Missing data – limitations of applicability/access

The challenge of missing data is broad and non-specific. During the field study there were various cases and examples mentioned which might fall within this category, with at times the aspect of missing data could be more or less obvious. The causes of missing data, and the effects of missing data are many. Some causes are due to issues with automatic mapping and thus software coding and design. Others might be due to a lack of awareness of specific metadata requirements and an uncertainty regarding which metadata elements to use. Sometimes missing data in EDM records is due to choosing the wrong data element. Other variables can be the general absence of information about an item and the way metadata is managed and created by an items institute of origin.

When data is missing, depending on the stage of the process in which it is discovered, might lead to the supplementing of data by e.g. an aggregator, as seen in the case of a dataset provided through the aggregator Deutsche Digitale Bibliothek. The items in the dataset had not originally been given titles by the providing institution, prompting the aggregator to use inventory numbers instead of titles, with the aim of making the titles unique. Upon import to the Metis software,

⁶ ARK stands for Archival Resource Key.

⁷ HANDLE is a general-purpose global name service, and the Handle system manages handles, which are unique names for digital objects and other internet resources (Lannom, Boesch & Sun 2003).

⁸ DOI stands for Digital Object identifier.

the titles were then flagged as unrecognisable. Having no title would be a larger issue though, so the titles created by the Deutsche Digitale Bibliothek were the better option.

5.2 Theory of Infrastructure – the dimensions of Europeana

Applying the theory of infrastructure to the materials collected during the field study and the preparation there off, results in some interesting conclusions. The analysis follows the dimensions as stipulated by Star and Ruhleder (1996) and presented in **Table 3**. These dimensions of infrastructure, interestingly can be applied both on a larger scale, as in Europeana as a whole, or on a smaller scale, focusing on EDM. This chapter aims to give examples and to discuss, at least to an extent, both.

When Star and Ruhleder (1996) talk about *embeddedness*, they suppose that hypothetically, an infrastructure is sunk into other structures, social arrangements and technology. While this dimension is hard to prove or apply, on a superficial level one can argue that Europeana is a portal, sunken within the entirety of the European investment in cultural heritage and identity building. It is embedded in the European commission's cultural policies and attempts at competing with the commercial sector and safeguarding access to cultural heritage for the future. Various services provided by European can also be found in other technologies and services. One example being how the Swedish national aggregator retrieves data regarding European cultural heritage from Europeana, for their access portal Kringla.

The second dimension Star and Ruhleder (1996) speak of, *transparency*, state that infrastructures are transparent to use. This means that infrastructures do not need to be reinvented or assembled for each task, and instead the infrastructure invisibly supports each task. While this might hold true in theory as repeated tasks do not need for the systems in use to be reinvented or reassembled. However, dependent on the needs of various actors, aspects of the infrastructure might end up being redesigned for the chance of increased efficiency or to support new developments. These redesigns are not indicative of a lack of transparency though, rather they are a part of the continued adaptation of an infrastructure to better support its functions. These adaptations then make it so the infrastructure increases its ability to support each task, and even new tasks as they appear.

The dimension of *reach or scope* might be the easiest to note and describe. The dimension states that infrastructure reaches beyond a single event or one-site practice (Star & Ruhleder 1996). The international nature of Europeana means that the infrastructure of Europeana automatically reaches beyond a single event or single location. Various institutes and aggregators work with the systems and standards part of Europeana in their efforts to provide access to digital or digitised cultural heritage. Aspects of Europeana are also used by various institutions to in turn enrich and contextualise their own records and items, as seen in the example of SOCH in chapter 5.1.1.2.

Learned as part of a membership is a dimension which denotes that an infrastructure is familiar and naturalised to members of the community of which it is a part of, and that strangers encounter it as something to be learned about (Star & Ruhleder 1996). TI does not explicitly explain what constitutes a member and what constitutes a stranger, which leaves this dimension, and subsequent analysis quite nuanced. The statement itself though, that the Europeana infrastructure, is learned as part of a membership, is undoubtedly true. Those that deeply involved with the everyday workings of Europeana, are both familiar with the various aspects of the infrastructure and interact with it in a naturalised way. Depending on level of involvement, this familiarity and natural applicability varies. The findings, conversations and observations of the conducted field study leads to the idea of a scale rather than binary consideration of membership.

During the field study, an informant referred to the fact that Europeana had at times been a bit of a “black box”. A statement which can be related back to the experience of learning about Europeana. Attempts are being made to illuminate parts of this black box, but limited funding and time means that there is currently only a beta training platform with courses. Something which has also reinforced the idea of a membership, as analysed in this dimension, is at times the presentation of Europeana and its efforts have been presented at seemingly random conferences in relevant geographical areas. These conferences contributed to a sense of “you were there, or you weren’t” as another informant described, which becomes even more challenging considering how information about procedures, methods, and relevant/needed skills is spread out over various websites.

An infrastructure *links with conventions of practices*. As such, it both shapes and is shaped by conventions of practices (Star & Ruhleder 1996). Europeana consists of a mixture of practices from the entirety of the cultural heritage sector and the IT

sectors. At the same time, the requirements and limitations of the various aspects of the Europeana infrastructure limits accepted contributions as e.g. non-complicit data records cannot be imported nor be made accessible through the portal. As such it shapes practices of contributing communities. This can be seen both through the literature overview and its focus on influence, as well as the ANT analysis provided in chapter 5.1.

Embodiment of Standards denotes that infrastructures tend to plug into other infrastructures and tools in a standardised way (Star & Ruhleder 1996). At its core, this dimension can be likened to the one above, and the one below. Like in an actor network, at its core, an infrastructure does not stand alone. The interconnectedness, the connection of standards and practices through traceable pathways, is what allows interoperability, allows an infrastructure to support whatever it needs to support. In the case of Europeana we see the creation of the metadata model EDM, but also the fact that EDM support a collection of controlled (and thus in a sense standardised) vocabularies, such as Wikidata, Geonames, LOD and more.

An infrastructure is *built upon an installed base*. It does not develop from nothing and inherits both strength and limitations from earlier iterations and its base or origin (Star & Ruhleder 1996) An example of this, and challenges posed by it, is e.g. how Europeana still, 16 years after its first launched prototype, is largely unknown outside the cultural heritage sector, and even within it at times. This dimension can be seen in the development of Europeana going from the idea of a digital library to a portal providing access to European digital cultural heritage. Or in how EDM developed from ESE and still contains properties and elements identified in the creation and documentation of ESE.

The eighth dimension, an infrastructure *becomes visible upon breakdown*, refers to how infrastructures first become visible when it breaks down, how when something stops working and needs to be repaired or adjust the invisible cogs of which an infrastructure consists of come out into the open. This dimension can be seen in Europeana's own efforts at documenting and resolving common issues within the application of EDM as seen in their Problem Pattern work. Another example embodying this dimension is how it didn't become clear there was a lack of knowledge in regard to requirements in license marking in the case of SOCH and their mostly automated process of data mapping for scheduled data exports. The failed automatic enrichment and dereferencing by the Metis software is another example.

5.3 Review based on Research Questions

Considering how much empirical data as well as analytical reflections and positioning is contained within chapter 5, the following paragraphs provide a short review of the major points made above. These points serve as a reference for chapter 6 for ease of understanding, as well as to take a step back to reflect on the general points the analysis of the empirical data make in relation to the RQs posed at the beginning of this MA thesis. As such this review follows the order of the three RQs.

The first RQ is related to the use of EDM and aims to identify the challenges following the common processes involved in the conversion of local data to EDM. The RQ being: which issues and challenges can commonly be found in EDM following the conversion of local data to EDM? The empirical data introduced two variants of EDM, external and internal, which were added to the equation, so the empirical data collected aimed at answering this question instead introduced a focus on the broader processes involved in exporting and importing data, as well as Europeana's ongoing internal work of identifying problem patterns. The empirical data found that issues can occur at any stage in the metadata processes, and at any level. Some issues are more common at the local level. Some occur solely during the process of translating external EDM to internal EDM. Some issues are pre-existing, and a part of the original metadata. Other issues occur after the data has already been published and made accessible through the Europeana Portal. Examples of issues mentioned or observed during the conducted field study are:

- Broken Links
- Incompatible Vocabularies
- Issues with Enrichment
- Lack of Persistent identifiers
- Missing Data

The second RQ is about identifying the complexities and challenges of international cooperation in a digital format. More specifically, the second RQ is: what are the challenges of international cooperation in a digital format in regard to cultural heritage? Based on the empirical data, this MA thesis has identified the following challenges:

- Unclear understanding of what Europeana is and the expectations on being involved in the cooperative.
- Inconsistent organisational structures, resources, and competence on the local level.
- Difficulties finding or accessing information.

- Different communication practices formed by cultural identities.

The third RQ asks what the data management process entails at Europeana. The original formulation of the RQ being: what does the data management process entail in the specific case of Europeana? The answer to this is difficult to summarise due to the process not being entirely linear. An attempt can be found below, organised into the different organisational levels and in somewhat order of processes.

1. Provider level:

- a) Items are digitised (unless already digital), saved to a server and catalogued in a local metadata standard.
- b) Local metadata standards are mapped to EDM or to a different national standard mapped to EDM.

2. Aggregator level:

- a) Data arrives at Aggregator and is transformed into external EDM.
- b) Aggregator checks sample of dataset in Metis Sandbox to discover early issues.
- c) If there are issues, issues are hopefully resolved, which is done in collaboration with the providing institutions.

3. Europeana Foundation:

- a) Imports external EDM from aggregators and runs it through Metis.
- b) Metis validates, transforms into internal EDM, validates, normalises, enriches the data records, and processes media. Reports regarding issues are created concurrently.
- c) Depending on issues, contact is made with aggregators to resolve and discuss.
- d) Metadata is uploaded to database through which portal APIs can retrieve records based on queries.
- e) APIs are used to check published records every once in a while, to discover any post-publication issues.

6. Discussion

A common denominator when touching upon the topic of Europeana, is a difficulty defining what Europeana truly is. Research about, or mentioning, Europeana, offers various definitions and descriptions, but there appears to be a lack of consensus. Depending on the aspect focused on, the definition shifts. The involvement of Europeana, and the phrases used in their own online channels to describe their history and future, can even be discussed to have contributed to the confusion and vagueness surrounding Europeana. A common theme noted in chapter 5 is aptly referred to as ‘lack of knowledge/understanding’ and can be applied to various aspects and actors within the international context of Europeana, whether speaking about the providing institutions and aggregators, or about a wider user-centred perspective.

Returning to the topic of transparency, chapter 5 spoke of transparency in terms of dimensions of infrastructure. Transparency can also be discussed in the context of the words dictionary definition. According to the Oxford English Dictionary, *transparency* is a noun derived from the adjective *transparent* (Oxford English Dictionary 2023d), which in a figurative sense means “frank, open, candid, [and] ingenuous” (Oxford English Dictionary 2024a) as well as “easily seen through, recognized, understood, or detected; manifest, evident, obvious, clear” (Oxford English Dictionary 2024b). While a lot of information is available to the public, and to providing institutions, in regard to the various structures, processes and standards employed by Europeana, there can be said to be some challenges in fulfilling the second cited definition – most notably the requirement of “easily”. A challenge of which the Europeana foundation is both aware and attempting to tackle. An informant explained how Europeana at times had been somewhat of a “black box” when speaking of the efforts made in improving transparency. The available information, while published, is at the moment of writing still spread out over various webpages and kept up to date to varying degrees and with synchronization.

As a ‘stranger’ (**Table 3**), the entrance to information about Europeana is found through the webpages europeana.eu and pro.europeana.eu. Some of the hyperlinks on informative pages of the Pro portal might direct you to a Confluence page hosting what is dubbed the Europeana Knowledgebase. Confluence is a wiki⁹

⁹ A *wiki* is a type of webpage designed in a way which allows its content to be edited by anyone who accesses it, and uses a simplified mark-up language (Oxford English Dictionary 2023e)

development and management software (Atlassian 2024) and knowledge base software (Europeana 2024g). According to its own description, the Europeana Knowledge Base collects “all [their] guidelines and documentation in one place” (Europeana 2024g). Depending on where one falls on the mentioned ‘scale of membership’, one might find oneself here following a similar path, or one might find it using a direct access link through a different route. Once an individual has found their way to the Europeana Knowledge Base, principles of transparency are more applicable.

Despite the aforementioned vagueness and diffusion surrounding consensus about what Europeana is, considering Europeana as a database allows the application of the theory Database as Discourse to the discussion regarding power, effect, and influence. Topics that have previously been covered independently in previous research, whether focusing on aspects of Europeana acting in a governing role in local and national heritage institutions (Capurro & Plets 2020), the consideration of Europeana as a techno-cultural construct with socio-economic structures and thus socio-political implications (Capurro & Severo 2023), or literature focusing on Europeana’s influence on digitisation of cultural heritage in relation to sustainable development (Macrì & Cristofaro 2021). Topics which are also seen in discussions involving Europeana and the context of identity and collective memory (Valtysson 2012; Andersson & Othén 2013; Stainforth 2016, 2022).

The following central themes of DD are particularly relevant to the discussion about power:

- Databases affect constitutions of their subjects.
- Databases make a mockery of the principle of having an author as an authority while still belonging to an entity.
- A database is a discourse of pure writing that amplifies the power of its owner/user.
- A database is a discursive production which inscribes positionalities of subjects according to its rules of formation.

The first central theme of DD, the fact that databases affect constitutions of their subjects, connects with themes of power and influence observed in earlier research. For the requirements posed by the metadata schema, the software and the organisation shape and give meaning to items within the collection of records. The vocabularies supported by EDM limit which descriptions can be given to entities. This first central theme, and its meaning, can also be seen in the similarities of the

third and fourth, as a database is a discourse of pure writing that amplifies the power of its owner/user and a database is a discursive production which inscribes positionalities of subjects according to its rules of formation.

The second central theme, that which argues that databases make a mockery of the principle of having an author as an authority while still belonging to an entity, can be seen in how many individuals and organisations are involved in its processes at any one time. Whether one speaks of the employees at the foundation, the entire network of aggregators, the influence of the European Commission, the Aggregator's forum, or the Network Association. Actors seen in the representation of the organisational structure in **Figure 3**. This theme can also be seen in the continued development of software and APIs, the inclusion of more vocabularies to allow for even greater chance at interoperability.

While it is briefly touched upon in the discussion of effects as guided by the theory DD, there are certain areas worth focusing on when considering the metadata aspect of this MA thesis and the focus of challenges. Elings & Waibel (2007) denoted four key areas of description when they investigated descriptive standards and metadata sharing between ALM institutions:

1. Data fields and structures
2. Data content and values
3. Data format
4. Data exchange

Below, these are discussed focusing on EDM.

Seeing as EDM is not a static schema, and instead a model used for combining relevant metadata properties, as well as containing its own properties, it might be a tad more difficult to present or speak of EDM in the same structure as proposed by Elings & Waibel. EDM also supports multiple standards in regard to data content and values. The model does however use a clearly denoted data format, using XML (Extensive Mark-up Language) to format its data fields and content. When external EDM is exported to Europeana from providing aggregators, it is only accepted through OAI-PMH or HTTP.

This thesis has at times alluded to the concept of collections, and whether Europeana constitutes as having a collection of its own. This is not a point of investigation this thesis has concerned itself with, however, it does deserve to be commented on. The result and analysis chapter 5.1.2 focussing on metadata speaks

of the causes of the issue of broken links, for example. More specifically, the issue of link rot, which is caused by changings of location or identifier of an item. An issue which exists due to the fact that Europeana does not have the item on a server of their own, instead linking to the providing institutions server location for the digitised item. In this sense, one can argue that Europeana does not have its own collection. This thesis would implore one to consider the following in this discussion: Europeana does, upon receiving data in the format of external EDM, translate, enrich, and dereference the records and save these to their own database. As such, the enriched records make up a collection of sorts, with data created through the import process to the Europeana Portal.

7. Conclusion

To conclude, this thesis has taken a look at challenges in the management of digital cultural heritage metadata within an international context through the conducting of a case study of the digital cultural heritage portal Europeana. This MA thesis has identified these challenges through the guidance of three specific research questions, namely:

- Europeana uses the European Data Model standard for their (meta)data, and aggregators are used to convert/transform local data to EDM. Which issues and challenges can commonly be found in EDM following this process?
- What are the challenges of international cooperation in a digital format in regard to cultural heritage?
- What does the data management process entail in the specific case of Europeana?

Through the application of Actor Network Theory and Theory of Infrastructure on the empirical material collected through the conducted field study using the go-along and accompanying methods, this thesis has found the following:

- Common issues and challenges found in EDM are related to:
 - Broken Links
 - Incompatible Vocabularies
 - Faulty Enrichment
 - Lack of Persistent identifiers
 - Missing Data
- Challenges of international cooperation in a digital format in regard to cultural heritage are:
 - Unclear understanding of what Europeana is and the expectations on being involved in the cooperative.
 - Inconsistent organisational structures, resources, and competence on the local level.
 - Difficulties finding or accessing information.
 - Different communication practices formed by cultural identities.
- The data management process isn't linear although can be divided into three general levels, with some local variations:
 - The Provider level where items are digitised, saved to a server and catalogued, and where local metadata standards might be mapped to EDM or a different national standard mapped to EDM.

- The Aggregator level where data is transformed into external EDM, datasets are checked for issues, and any possible issues hopefully resolved in collaboration with the providing institutions.
- The Europeana Foundation, which imports external EDM through Metis and validates the data and transforms it into internal EDM, enriches the data before it is uploaded to the database. This database is accessible for portal APIs to retrieve records based on queries, and APIs can check published records to discover any post-publication issues.

The relevant empirical material to this question also showed how issues, errors and challenges can occur at any of these levels, and that many times the process jumps between steps based on a case-by-case basis.

The contextualisation of the results in chapter 6. Discussion, where the material was discussed in relation to the literature overview and the perspective of Database as Discourse, notes how the difficulty observed in clearly defining Europeana can be considered one of its challenges. The discussion highlights the difficulties posed by the ways in which Europeana can still at times be considered to be a type of “black box” and connects to the analysis through the discussion of access to information about procedures and requirements of Europeana and the concept of being ‘a stranger’ vs. a ‘member’.

7.1 Suggested future research

The discussion and analysis of the collected empirical material, as well as reflections following the development of the background and literature overview of this thesis leads to some interesting perspectives which are worth investigating in the future. As such, the final sub-chapter of this MA thesis proposes various topics which could be investigated next.

As is stated in chapter 1 and the introduction of this MA thesis, there is a noted absence of museological research about digitally accessible platforms and databases. This absence is especially noticeable in regard to northern Europe following a search in the journal *Nordisk Museologi*. The focus on Europeana in this MA thesis as such aims to introduce similar research to the field of museology, encouraging to merge ideas from information science with museology.

This MA thesis identifies various challenges in connection to metadata management in an international context regarding digital cultural heritage. Challenges which could contribute, if further investigated, to making cultural heritage more accessible. The limitations of this thesis means that some encountered topics and questions went unexplored, as such there are a few suggestions pertaining to future research which can be found below.

Topics of interest surrounding Europeana in museological research could include:

- The ways in which the records classify as a collection in a digital context and how thematic excerpts in blogpost constitute as exhibits.
- Reasons behind sharing data in a large-scale initiative such as Europeana – what is the aim with sharing data, or in other words – what does a museum hope to accomplish through joining a digital cultural heritage portal?
- The effects of intent on metadata quality and item selection

Topics of interest for future research in regard to metadata within museology

- Investigating different openly available digital museum catalogues and digital cultural heritage portals
- Which metadata standards are used in museums? Can be limited to selected geographical zones. Which needs and requirements influence choice?
- Processes involved in metadata sharing in the context of artefacts on loan.

Topics of interest in a broader ALM context

- What is Europeana? The common definitions and assumptions.
- Collaborations in the digital sphere between ALM institutions – challenges and benefits

Bibliography

- Andersson, S. & Othén, A. (2013). *Ting och tendens : makt, representation och identitet i arbetet med Europeana*.
- Åsberg, C., Hultman, M. & Lee, F. (2012). Möt den posthumanistiska utmaningen. In: Åsberg, C., Hultman, M., & Lee, F. (eds.) *Posthumanistiska nyckeltexter*. Lund: Studentlitteratur, pp. 29–46.
- Atlassian (2024). *Confluence: The Ultimate Wiki Software*. Atlassian.
<https://www.atlassian.com/software/confluence/use-cases/wiki> [2024-04-18].
- Bawden, D. & Robinson, L. (2012). *An introduction to information science*. London: Facet.
- Brown, D. (2010). Eight Principles of Information Architecture: Bulletin of the American Society for Information Science & Technology. *Bulletin of the American Society for Information Science & Technology*, 36(6), pp. 30–34, doi:10.1002/bult.2010.1720360609.
- Callon, M. (1984). Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St Brieuc Bay. *Sociological Review*, 32(S1), pp. 196–233, doi:10.1111/j.1467-954X.1984.tb00113.x.
- Cameron, F. & Kenderdine, S. (2007). Introduction. In: Cameron, F. & Kenderdine, S. (eds.) *Theorizing digital cultural heritage: a critical discourse*. Cambridge, Mass: MIT Press, pp. 1–15.
- Capurro, C. & Plets, G. (2020). Europeana, EDM, and the Europeanisation of Cultural Heritage Institutions: Digital Culture & Society. *Digital Culture & Society*, 6(2), pp. 163–189, doi:10.14361/dcs-2020-0209.
- Capurro, C. & Severo, M. (2023). Mapping European Digital Heritage Politics: An Empirical Study of Europeana as a Web-based Network. *Heritage & Society*, 0(0), pp. 1–21, doi:10.1080/2159032X.2023.2266801.
- Chirac, J. (2005). letter_barroso2005_5978.pdf.
- Choemprayong, S. & Wildemuth, B. M. (2017). Case Studies. In: Wildemuth, B. M. (ed.) *Applications of social research methods to questions in information and library science*. Santa Barbara, California: Libraries Unlimited, pp. 51–59.
- Chowdhury, G. G. (2010). *Introduction to modern information retrieval*. 3rd ed. London: Facet.

- CIDOC (2024a). *What is LIDO?* ICOM CIDOC.
<https://cidoc.mini.icom.museum/working-groups/lido/lido-overview/about-lido/what-is-lido/> [2024-04-19].
- CIDOC (2024b). *Home | CIDOC CRM*. <https://cidoc-crm.org/> [2024-05-15].
- Czarniawska, B. (2014). Observations on the move: Shadowing. In: *Social Science Research: From Field to Desk*. 1 Oliver's Yard, 55 City Road London EC1Y 1SP: SAGE Publications Ltd.
- Czarniawska, B. (2022). Skuggning i fältarbete. In: Ahrne, Göran & Svensson, P. (eds.) *Handbok i kvalitativa metoder*. Stockholm : Liber, pp. 153–166.
- Elings, M. W. & Waibel, G. (2007). Metadata for all: Descriptive standards and metadata sharing across libraries, archives and museums. *First Monday*, doi:10.5210/fm.v12i3.1628.
- European Commission (2005). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - i2010 : digital libraries {SEC(2005) 1194} {SEC(2005) 1195}*.
- European Commission (2010). *Boosting cultural heritage online: the European Commission sets up a Reflection Group on digitisation*. European Commission.
https://ec.europa.eu/commission/presscorner/detail/en/ip_10_456 [2024-02-20].
- European Commission (2011). *Digital Agenda: 'Comité des Sages' calls for a 'New Renaissance' by bringing Europe's cultural heritage online*. European Commission.
https://ec.europa.eu/commission/presscorner/detail/en/ip_11_17 [2024-02-20].
- European Commission (2014). *Timeline of digitisation and online accessibility of cultural heritage | Shaping Europe's digital future*. <https://digital-strategy.ec.europa.eu/en/library/timeline-digitisation-and-online-accessibility-cultural-heritage> [2024-04-11].
- European Commission. Directorate General for Communications Networks, Content and Technology. (2024). *Common European data space for cultural heritage: annual report 2022/2023*. LU: Publications Office.
- Europeana (2013). Europeana Data Model Primer.
- Europeana (2017). Definition of the Europeana Data Model v.5.2.8.
- Europeana (2024a). *About*. Europeana. <https://www.europeana.eu/en/about-us> [2024-01-11].

- Europeana (2024b). *Glossary of Terms*. Europeana PRO.
<https://pro.europeana.eu/page/glossary> [2024-03-11].
- Europeana (2024c). *Europeana Aggregators' Forum Activity Report 2022*.
 Europeana PRO. <https://pro.europeana.eu/post/europeana-aggregators-forum-activity-report-2022> [2024-05-16].
- Europeana (2024d). *Supporting the digital cultural heritage of Ukraine Working Group*. Europeana PRO. <https://pro.europeana.eu/project/supporting-the-digital-cultural-heritage-of-ukraine-working-group> [2024-05-16].
- Europeana (2024e). *Laptops for Ukraine*. Europeana PRO.
<https://pro.europeana.eu/page/laptops-for-ukraine> [2024-05-16].
- Europeana (2024f). *Help us to make cultural heritage data more persistent*.
 Europeana PRO. <https://pro.europeana.eu/post/help-us-to-make-cultural-heritage-data-more-persistent> [2024-05-16].
- Europeana (2024g). *Europeana Knowledge Base - Confluence*.
<https://europeana.atlassian.net/wiki/spaces/EF/overview?homepageId=330367111> [2024-04-18].
- Europeana Foundation (2012). *Europeana opens up full dataset for re-use*.
 Europeana PRO. <https://pro.europeana.eu/post/europeana-opens-up-full-dataset-for-re-use> [2024-02-20].
- Europeana Foundation (2020). *Europeana DSI*. Europeana PRO.
https://pro.europeana.eu/page/europeana-dsi?utm_campaign=internal_link&utm_medium=Digital%20service%20in%20frastructures&utm_source=about-us%2Fmission [2024-02-20].
- Europeana Foundation (2024a). *Our Mission*. Europeana PRO.
<https://pro.europeana.eu/about-us/mission> [2024-02-01].
- Europeana Foundation (2024b). *Metadata*. Europeana PRO.
<https://pro.europeana.eu/share-your-data/metadata> [2024-02-13].
- Gibson, R. C., Chowdhury, S. & Chowdhury, G. (2024). User versus institutional perspectives of metadata and searching: an investigation of online access to cultural heritage content during the COVID-19 pandemic. *International Journal on Digital Libraries*, 25(1), pp. 105–121, doi:10.1007/s00799-023-00385-y.
- Hackweek Part I* (2021).
- Högberg, A. & Holtorf, C. (2021). Final Reflections. In: Holtorf, C. & Högberg, A. (eds.) *Cultural Heritage and the future*. London ; New York: Routledge, Taylor & Francis Group, pp. 264–269.
- Hughes Tidlund, I. & Von Unge, E. (2022). Etik och kvalitativa metoder. In: Silow Kallenberg, K., Nystrand von Unge, E., & Wiklund Moreira, L.

- (eds.) *Etnologiskt fältarbete: nya fält och former*. Lund: Studentlitteratur, pp. 47–65.
- Hunger, F. (2018). Epistemic Harvest. *A Peer-Reviewed Journal About*, 7(1), doi:10.7146/aprja.v7i1.115064.
- Komsell, L. & Melén, H. (2007). *Museum på Webben - en undersökning om användbarhet och åtkomst*.
- Kusenbach, M. (2003). Street Phenomenology: The Go-Along as Ethnographic Research Tool. *Ethnography*, 4(3), pp. 455–485, doi:10.1177/146613810343007.
- Lannom, L., Boesch, L. C. B. P. & Sun, S. (2003). *Handle System Overview*. Internet Engineering Task Force. Request for Comments No. RFC 3650.
- Latour, B. (2005). *Reassembling the social: an introduction to actor-network-theory*. Oxford: Oxford University Press, UK.
- Law, J. (2009). Actor Network Theory and Material Semiotics. In: Turner, B. S. (ed.) *The new Blackwell companion to social theory*. Chichester: Wiley-Blackwell, pp. 141–158.
- Library of Congress (2024a). *EAD: Encoded Archival Description (EAD Official Site, Library of Congress)*. <https://www.loc.gov/ead/> [2024-04-19].
- Library of Congress (2024b). *Metadata Encoding and Transmission Standard (METS) Official Web Site | Library of Congress*. <https://www.loc.gov/standards/mets/> [2024-04-19].
- Macri, E. & Cristofaro, C. L. (2021). The Digitalisation of Cultural Heritage for Sustainable Development: The Impact of Europeana. In: Demartini, P., Marchegiani, L., Marchiori, M., & Schiuma, G. (eds.) *Cultural Initiatives for Sustainable Development: Management, Participation and Entrepreneurship in the Cultural and Creative Sector*. Cham: Springer International Publishing, pp. 373–400.
- Malmö museer (2024). *Carlotta - Om Carlotta*. <https://carlotta.malmo.se/carlotta-mmus/web> [2024-04-10].
- Marty, P. F. (2009). Museum Informatics. In: Bates, M. J. & Maack, M. N. (eds.) *Encyclopedia of Library and Information Sciences, Third Edition*. CRC Press, pp. 3717–3725.
- McKenna, L., Debruyne, C. & O’Sullivan, D. (2022). Using Linked Data to create provenance-rich metadata interlinks: the design and evaluation of the NAISC-L interlinking framework for libraries, archives and museums. *AI & SOCIETY*, 37(3), pp. 921–947, doi:10.1007/s00146-021-01373-z.

- Nilson, T. & Thorell, K. (2018). Introduction. In: Nilson, T. & Thorell, K. (eds.) *Cultural heritage preservation: the past, the present and the future*. Halmstad: Bulls Graphics AB, pp. 8–21.
- Olensky, M., Stiller, J. & Dröge, E. (2012). Poisonous India or the Importance of a Semantic and Multilingual Enrichment Strategy. In: Doderer, J. M., Palomo-Duarte, M., & Karampiperis, P. (eds.) *Metadata and Semantics Research*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 252–263.
- Oxford English Dictionary (2023a). infrastructure, n.
- Oxford English Dictionary (2023b). infrastructure, n., Etymology.
- Oxford English Dictionary (2023c). infra-, prefix.
- Oxford English Dictionary (2023d). transparency, n., sense 1.a.
- Oxford English Dictionary (2023e). wiki, n.
- Oxford English Dictionary (2024a). transparent, adj., sense 2.a.
- Oxford English Dictionary (2024b). transparent, adj., sense 2.b.
- Pandey, R. & Kumar, V. (2023). A Survey of the Metadata Element Sets Used for Digital Art Objects in the Online Collections of the Museums of India: DESIDOC Journal of Library & Information Technology. *DESIDOC Journal of Library & Information Technology*, 43(4), pp. 226–233, doi:10.14429/djlit.43.04.19228.
- Patel, M., White, M., Mourkoussis, N., Walczak, K., Wojciechowski, R. & Chmielewski, J. (2005). Metadata requirements for digital museum environments. *International Journal on Digital Libraries*, 5(3), pp. 179–192, doi:10.1007/s00799-004-0104-x.
- Petras, V., Hill, T., Stiller, J. & Gäde, M. (2017). Europeana - a search engine for digitised cultural heritage material: Datenbank-Spektrum. *Datenbank-Spektrum*, 17(1), pp. 41–46, doi:10.1007/s13222-016-0238-1.
- Petras, V. & Stiller, J. (2017). A Decade of Evaluating Europeana - Constructs, Contexts, Methods & Criteria. In: Kamps, J., Tsakonas, G., Manolopoulos, Y., Iliadis, L., & Karydis, I. (eds.) *Research and Advanced Technology for Digital Libraries*. Cham: Springer International Publishing, pp. 233–245.
- Philippopoulos, P. I., Drivas, I. C., Tselikas, N. D., Koutrakis, K. N., Melidi, E. & Kouis, D. (2024). A Holistic Approach for Enhancing Museum Performance and Visitor Experience: Sensors (14248220). *Sensors (14248220)*, 24(3), p. 966, doi:10.3390/s24030966.
- Poster, M. (1996). *The second media age*. Cambridge, MA: Polity Press/Blackwell Publishers Inc.
- Reding, V. (2005). The role of libraries in the information society.

- Riksantikvarieämbetet (2017). *Om Kringla*. <https://www.raa.se/hitta-information/kringla/om-kringla/> [2024-04-10].
- Rössle, A. (2018). *Är en bild värd mer än tusen ord? : - En studie i metadataanvändning i Kringlas och Stockholms digitala stadsmuseums bilddatabaser*.
- Sandvig, C. (2013). The Internet As Infrastructure. In: Dutton, W. H. (ed.) *The Oxford handbook of Internet studies*. Oxford: Oxford University Press, pp. 86–106.
- Severson, P. (2017). Searching for voice as democratic values in contemporary Swedish web archive initiatives. Presented at the Digital Democracy Critical Perspectives in the Age of Big Data.
- Skinner, J. (2014). Metadata in Archival and Cultural Heritage Settings: A Review of the Literature. *Journal of Library Metadata*, 14(1), pp. 52–68, doi:10.1080/19386389.2014.891892.
- Sköld, O. (2018). *Documenting Videogame Communities : A Study of Community Production of Information in Social-Media Environments and its Implications for Videogame Preservation*. Department of ALM.
- Snickars, P. (2022). Tema kulturarv: Temamodeller om ABM i svenska statliga offentliga utredningar 1945–89. *Nordisk Museologi*, 34(2), pp. 73–96, doi:10.5617/nm.10072.
- Stainforth, E. (2016). From museum to memory institution: the politics of European culture online. *Museum and Society*, 14(2), pp. 323–337, doi:10.29311/mas.v14i2.646.
- Stainforth, E. (2022). Collective memory or the right to be forgotten? Cultures of digital memory and forgetting in the European Union: *Memory Studies*. *Memory Studies*, 15(2), pp. 257–270, doi:10.1177/17506980211044707.
- Star, S. L. & Bowker, G. C. (2010). How to Infrastructure. In: Lievrouw, L. A. & Livingstone, S. (eds.) *Handbook of New Media: Social Shaping and Social Consequences of ICTs, Updated Student Edition*. 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd.
- Star, S. L. & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *INFORMATION SYSTEMS RESEARCH*, 7(1).
- Stiller, J. & Petras, V. (2018). Learning from Digital Library Evaluations: The Europeana Case. *ABI Technik*, 38(1), pp. 37–45, doi:10.1515/abitech-2018-0006.
- Sundstrom, A. da S. S. (2023). Ethical Issues from Decolonial Practices in Knowledge Organization: The Case of Indigenous Collections in

- Varldskulturmuseet: *Knowledge organization*, 50(6), pp. 439–451,
doi:10.5771/0943-7444-2023-6-439.
- Sverige (1997). *IT i kulturens tjänst: slutbetänkande*. Stockholm: Fritze.
- Taher, H., Addo, G., Pruulmann-Vengerfeldt, P., Engberg, M. & Harvard Maare, Å. (2022). Reuse and Appropriation: : Remediating Digital Museum Collections and Digital Tools for a Participatory Culture in Transition. *Baltic Screen Media Review*, 10(1), pp. 122–138.
- UNESCO (2020). *Cultural heritage*. UNESCO.
<https://uis.unesco.org/node/3079731> [2024-02-05].
- Valtysson, B. (2012). Europeana: The digital construction of Europe’s collective memory: Information, Communication & Society. *Information, Communication & Society*, 15(2), pp. 151–170,
doi:10.1080/1369118X.2011.586433.
- Xavier, A. & Hernandez, F. (2020). OAI-PMH and Linked Open Data in the context of Hispana and Europeana. Some historical reflections: *JLIS.it: Italian Journal of Library, Archives & Information Science*. *JLIS.it: Italian Journal of Library, Archives & Information Science*, 11(1), pp. 1–16,
doi:10.4403/jlis.it-12573.
- Zavalin, V. & Zavalina, O. L. (2023). Exploration of Accuracy, Completeness and Consistency in Metadata for Physical Objects in Museum Collections. In: Sserwanga, I., Goulding, A., Moulaison-Sandy, H., Du, J. T., Soares, A. L., Hessami, V., & Frank, R. D. (eds.) *Information for a Better World: Normality, Virtuality, Physicality, Inclusivity*. Cham: Springer Nature Switzerland, pp. 83–90.
- Zhang, J. & Ren, T. (2024). A conceptual model for ancient Chinese ceramics based on metadata and ontology: A case study of collections in the Nankai University Museum. *Journal of Cultural Heritage*, 66, pp. 20–36,
doi:10.1016/j.culher.2023.10.012.

Appendix 1: Alphabetised list of Abbreviations

Abbreviation	Definition
ALM	Archives, Libraries, and Museums
ANT	Actor Network Theory
API	Application Programming Interface
CC	Creative Commons
CIDOC	ICOM International Committee for Documentation
CLIO	Checking Link In Operation
DBMS	Database Management System
DC	Dublin Core
DD	Database as Discourse
DPS	
DS	Data space
DSI	Digital Service Infrastructure
EDM	European Data Model
ESE	European Semantic Elements
EU	European Union
HTML	HyperText Markup Language
HTTP	Hypertext Transport (or Transfer) Protocol
ICH	Intangible Cultural Heritage
IRS	Information Retrieval Systems
ISO	International Standards Organisation
IT	Information Technology
KOS	Knowledge Organisation System
LOD	Linked Open Data
LUB	Lund University Libraries
MA	Master of Arts (degree)
OAI-PMH	Open Archives Initiative – Protocol for Metadata Harvesting
RDBMS	Relational Database Management System
RQ	Research question
SOCH	Swedish Open Cultural Heritage
SQL	Structured Query Language

TI	Theory of Infrastructure
UNESCO	United Nations Educational, Scientific and Cultural Organisation
URL	Uniform Resource Locator
XSLT	eXtensible Stylesheet Language Transformations
XML	eXtensive Markup Language

Appendix 2: Index of tables and figures

Figures:

- Figure 1.** Components of an IRS (reproduced from Chowdhury, 2010; Bawden & Robinson, 2012).....21
- Figure 2.** An ANT model of Europeana. The model is based on jottings from the conducted field study at the Europeana Foundation office in the Hague, the Netherlands. Branches are named following the identification of broader and narrower themes in connection to processes, cases and challenges mentioned and encountered during the conducting of the study. The arrows denote the smaller identified challenges or themes in connection to the larger theme blocks. The Model was created keeping in mind the ANT perspective through which empirical data is contextualised and placed in relation to itself.....43
- Figure 3.** A chart of the organisational structure of Europeana and the Europeana Initiative. Colour coded to denote levels in the organisational structure. Annotated with clarifications for ease of interpretations. The chart was made during the conducted field study to support the ANT models presented in other figures in the hopes of clarifying where in the organisation challenges are noted.44
- Figure 4.** A model of the aggregation network node identified from the model found in figure 2. The model is based on jottings from the conducted field study at the Europeana Foundation office in the Hague, the Netherlands. Branches are named following the identification of broader and narrower themes in connection to processes, cases and challenges mentioned and encountered during the conducting of the study. The arrows denote directions of steps in flows and processes, or simply some of the smaller identified challenges or themes in larger theme blocks. The Model was created keeping in mind the ANT perspective through which empirical data is contextualised and placed in relation to itself.45
- Figure 5.** A model of the (Meta)Data node identified from the model found in figure 2. The model is based on jottings, software insights and documents obtained from the conducted field study at the Europeana Foundation office in the Hague, the Netherlands. Branches are named following the identification of broader and narrower themes in connection to processes, cases and challenges mentioned and encountered during the conducting of the study. The arrows denote directions of steps in flows

and processes, or simply some of the smaller identified challenges or themes in larger theme blocks. The Model was created keeping in mind the ANT perspective through which empirical data is contextualised and placed in relation to itself.....50

- Figure 6.** The EDM class hierarchy and names. The classes introduced by EDM are shown in light blue. The classes in white are re-used from other schemas, with the schema indicated before the colon. The model originates from: EDM_Definition_v5.2.8_102017.pdf (europeana.eu), p. 751
- Figure 7.** Screenshot showing a generic example of a dataset being run through Metis, shared by informant.55
- Figure 8.** Simplified model of import/publish process of the Metis software, in order. The arrows signify data and process direction, and the grey blocks denote underordered processes included in a process step.55

Tables:

- Table 1.** Overview of amount of digital items by item type, in descending order based on amount. Copied from: <https://www.europeana.eu/en/about-us> [2024-01-11 11:51:48 CET] 11
- Table 2.** Table containing Brown’s (2010) 8 pragmatic principles of information architecture. Table content reproduced from Bawden & Robinson (2012, p. 144). Restructured into two columns for ease of reading.20
- Table 3.** the dimensions for infrastructure as defined by Star & Ruhleder (1996, p. 113) presented in a table for ease of reading.30
- Table 4.** List of excerpts of 'important' EDM properties or tags and their requirements. Sources: EDM_Definition_v5.2.8_102017.pdf (europeana.eu) and <https://docs.google.com/spreadsheets/d/1atZr1w-h9AdWwWSBYLCCk6fAdJSxrCNP56QRLNY1jLg/edit#gid=1273976273>54