Metadata Quality Assurance for Audiobooks

- An explorative case study on how to measure, identify and solve metadata quality issues

Author: Patrik Carlsson
Abstract

Metadata is essential to how (digital) archives, collections or databases operate. It is the backbone to organise different types of content, make them discoverable and keep the digital records’ authenticity, integrity and meaning over time. For that reason, it is also important to iteratively assess if the metadata is of high quality. Despite its importance, there is an acknowledged lack of research verifying if existing assessment frameworks and methodologies do indeed work and if so how well, especially in fields outside the libraries. Thus, this thesis conducted an exploratory case study and applied already existing frameworks in a new context by evaluating the metadata quality of audiobooks. The Information Continuum Model was used as a way to capture the metadata quality needs of customers/end users who will be searching and listening to audiobooks. Using a mixed methods approach, the results showed that the frameworks can indeed be generalised and adapted to a new context. However, although the frameworks helped measure, identify and find potential solutions to the problems, they could be better adjusted to the context and more metrics and information could be added. Thus, there can be a generalised method to assess metadata quality. But the method needs improvements and to be used by people who understand the data and the processes to reach its full potential.

Keywords: Metadata Quality | Quality Assessment | Information Continuum Model | Audiobooks | Information management
Acknowledgements

First I would like to thank my supervisor, Erica Hellmer, for all her help and guidance during this writing. A huge thank you to S., H. and M. at The Company, for their additional supervision and technical support in preparing the data analysis, as well as all the interviewees for their valuable inputs. I would also like to thank my friends and family who provided me with feedback and support for this work. Last but not least, a huge thank you to my wife, and love of my life, Anna, and my children Frida and Thilo for always making me happy.

Patrik Carlsson
18th May 2023
Table of Contents

Abstract .................................................................................................................. ii
Acknowledgements ............................................................................................... iii
List of terminologies ............................................................................................... vi
1. Introduction ........................................................................................................ 1
   1.1 Outline of the Thesis ..................................................................................... 3
   1.2 Purpose, Aim & Research Questions ........................................................... 3
   1.3 Delimitation ................................................................................................. 4
2. Literature review ................................................................................................. 5
   2.1 A General Overview of the Problem ............................................................ 5
   2.2 Existing Frameworks .................................................................................... 8
   2.3 Metadata Control from Sweden .................................................................. 11
   2.4 The Gap ...................................................................................................... 14
3. Theoretical Framework ....................................................................................... 15
   3.1 What is Quality? .......................................................................................... 15
   3.2 Grand Theory: Critical Theory ................................................................... 18
   3.3 Middle-range Theory: Information Continuum Model ............................... 20
4. Methodology ....................................................................................................... 24
   4.1 Instrumental Case Study .............................................................................. 24
      4.1.1 Strengths and Weaknesses of a Case Study ........................................ 25
      4.1.2 Ethical Role and Bias ........................................................................... 26
   4.2 The Assessment Method ............................................................................. 27
      4.2.1 Measure ............................................................................................... 29
      4.2.2 Identify ................................................................................................ 31
      4.2.3 Solving .................................................................................................. 33
   4.3 Data Collection: Mixed Method .................................................................. 36
      4.3.1 Quantitative Calculations .................................................................... 37
      4.3.2 Qualitative Interviews ........................................................................ 38
5. The Results and Analysis ................................................................................... 41
   5.1 The Quantitative Data .................................................................................. 41
      5.1.1 Implementation ..................................................................................... 41
      5.1.2 Analysis of the Quantitative Data ......................................................... 43
   5.2 The Qualitative Interviews ........................................................................... 48
      5.2.1 Familiarity with the Customers/End Users .......................................... 49
      5.2.2 Opinion about the Assessment Method .............................................. 51
List of terminology

If not stated otherwise, this list’s terms will have the following meaning.

Accuracy: how well metadata portrays the record or resource it represents.\(^1\)

Archivist: here following the Swedish definition, where the role of the records manager (creating or receiving, organising and pluralising information or records, and planning how to store the information or records)\(^2\) and the role of the archivist (acquiring, describing, appraising, preservation and providing access to records and/or management of archival repository)\(^3\) are fused into the same role.

Completeness: all elements of metadata that are needed to provide that an ideal representation of the record or resource is given.\(^4\)

Customer: in this case, it will be used as a synonym for ‘end user (of the streaming service)’. The term ‘customer’ will be used instead of ‘end user’ when discussing another work where the author(s) have used the term customer in the text. This is to prevent confusion when too much interchange between the two terms occurs.

Data: different values representing facts, used in a format suitable to be used and understood by humans and/or machines.\(^5\)

Information: meaningful content, i.e. data put into context, that is communicated to others who find the content relevant and useful.\(^6\)

---


\(^4\) Reiche & Höfﬁg 2013, p. 237.

\(^5\) Nationalencyklopedin, data, 2023, https://www.ne.se/uppslagsverk/encyklopedi/1%C3%A5ng/data [accessed 2023-03-22].

**Metadata:** data describing an information object such as a record, another piece of data or similar.⁷

**Record:** information or data, created or received, and stored on a medium to be used as an extension of (human) memory or to support accountability of activities.⁸

**Recordkeeping:** process for the creation, collection organisation and pluralisation of records.⁹

**Resource:** an information object,¹⁰ which in this case is other than a digital record. For example, an audiobook, a podcast, a video, etc.

**Societal embeddedness:** part of pluralisation, that points at the societal context in which a record exists, and thus highlights the processes that must be put in place to ensure that the record can be shared outside the organisation that created and/or manages it.¹¹

---

⁷ Society of American Archivists 2023b.
1. Introduction

It is said that information is the new gold. But not everything that glitters is gold. Therefore, a crucial part of a sustainable information provision is to ensure that not only that the correct information is found, but also found easily. Because, as the amount of information grows larger, it gets harder to sort out what information is relevant and useful. In his book about how to improve e-government, Eriksson recognised the fact that we still do not know what to do with all these large amounts of information – and what it does to us. Thus, Eriksson recommended the usage of metadata as a tool to help structure, use and validate information. This is also important for website construction, as the content is a key factor for users to consider the website a finished product or not. Furthermore, metadata is the backbone to organise different types of content and making them discoverable. The application of metadata also provides contextualisation to a record or resource, which in turn contributes to the trustworthiness and thus effectiveness of the item. Hence, metadata of high quality is essential for (digital) archives, collections and databases to operate efficiently.

The appliance and control of metadata are also important for another reason in archival theory, namely the records continuum theory, where a record is always in the process of becoming. This means that although the content and structure of a record can be fixed, the context of how and for what it is used is open for change. Thus, in a continuum perspective, as a record is used in different activities, archiving processes are needed to fix this record to the activities with broadening layers of contextual metadata. This is to ensure its accessibility and meaningfulness. This means that even if changes and edits to metadata do not affect the content of a record, they do have an important influence over the record’s authenticity.

For this reason, the archival scientist Foscarini argued that for the modern archivist to maintain the digital records’ authenticity, integrity and meaning over time, the archivists need to focus on the records’ metadata from a “people-centric” and “data-centric” perspective.

---

other words, archivists who operate in the digitalised world and follow the continuum model need to pay attention to the metadata of their records. Because, as the data science scholars Reiche and Höfig noted, as the content may change, information becomes invalid, outgoing links go dead or similar, metadata needs to be (re)evaluated to ensure high quality.\(^{17}\)

But since the term ‘quality’ does not have a final definition, the problem of how to measure, identify and solve quality issues lingers. This is an issue that many managers of digital libraries, repositories and archives have identified when trying to ensure high metadata quality. Different models have been developed to solve these issues but with limited research on how these models work in practice.\(^{18}\) Ergo, there is a lack of verification if these models can be generalised and work in other contexts, other than in which they were created.

The issue becomes more severe when a repository or archive is part of a larger network that aims at interoperability. Because metadata that had high quality in one context in the network can have poor quality in another context in the same network. This mostly happens when the environment aggregates and the metadata is more likely to be used for other purposes than those they were created for. Thus, the task of ensuring sufficient metadata quality level of digital records becomes fundamental, as repositories grow in size, number and diversity,\(^{19}\) and is an important part of generating the “new science of the document” archivists have called for due to the digitalisation.\(^{20}\)

As there is limited verification on how the models and frameworks work in practice to measure, identify and solve metadata quality issues, this thesis will further the research. More concretely, it will look at already existing methods and then apply them in a new context, in a case study. The case is a large European media company. Although not an archive, the location for the case was deliberately chosen as it offered a setting an opportunity to research the topic of metadata quality, which was relevant for archival and information science. More specifically, it offered an opportunity to check metadata quality for audiobooks, which will be the main focus of this thesis: how to ensure high metadata quality for audiobooks, coming from different sources, that later will be public on a streaming platform.

\(^{17}\) Reiche & Höfig 2013, p. 236.


\(^{19}\) Tani, Candela & Castelli 2013, p. 1196.

\(^{20}\) Frendo 2007, p. 158.
1.1 Outline of the Thesis

The rest of this chapter will explain the purpose of this thesis, its aim, research questions and its delimitation. The second chapter, the literature review, will provide a general overview of what has been done to ensure metadata quality, as well as a more detailed explanation of existing frameworks. The reviewed literature will mostly be very international, but one section will also include Swedish literature about metadata quality. The chapter ends with the identified gap in the literature. In chapter three, the theoretical concept of quality will be discussed, before the choices of using Critical theory and the Information Continuum Model as theoretical frameworks are motivated. Chapter four presents the methodology for how the existing frameworks will be assembled to create an automatic method to identify and evaluate metadata quality issues, and a semi-manual method to solve these issues. It will also explain how the verification of the framework’s applicability will be conducted, using a mixed-method approach of cross-tabular comparison and interviews. In chapter five the results of applying the framework on metadata will be presented, and in chapter six the conclusion of the findings will be outlined.

1.2 Purpose, Aim & Research Questions

The purpose of this thesis was to further the research in finding an effective way to assess metadata by measuring, identifying and solving metadata quality issues. This was conducted by applying some already existing models and frameworks in a new context. The aim of this thesis was that the results could contribute to providing a generalisable method to manage metadata quality issues, and thereby solve the identified archive and information science problem of how to get control over the metadata in the digital era. Thus, to serve the purpose and meet the aim, the following research questions were formulated:

1) Do existing frameworks to assess metadata quality issues work in a different context?
2) How does the implementation of the frameworks help to improve metadata quality for internal and external stakeholders?
3) What further adjustments, data governance policies and processes to the frameworks are needed to ensure metadata quality?
1.3 Delimitation

As this thesis was part of the master’s program Sustainable Information Provision with a major in Archives and Information Science it looked at frameworks that did not require programming. Although the challenges and opportunities that digitalisation brings were key themes in the master’s program, the goal was to provide knowledge and understanding of how to manage them and the need for information. Not how to create new digital tools. Therefore, the frameworks and methods that were used and tested were more based on theory, and only required technical support from what one could get from a Microsoft Excel programme or equivalent.

With that being said, this thesis was an explorative research. Using an explorative research design allowed for a deeper understanding of the data, its internal functions and its relations to other factors. It also allowed for an ‘inspection’ of the found data to analyse if it answered the questions one would seek to answer,\(^2\) which correlated with this thesis’ purpose to test how well-existing assessment frameworks worked.

---

2. Literature review

This chapter outlines what has been done to ensure metadata quality. The chapter’s first part will provide a more general overview of how studies are addressing metadata quality issues. The second part will focus on concrete examples of frameworks that have been developed to assess the quality level of metadata, data or information. In section three the Swedish literature on what has been done to ensure metadata quality is presented. At the end of the chapter, the identified gap will be presented. It became clear from the literature that although the importance of metadata is recognised, relatively little research in this field has been done.

2.1 A General Overview of the Problem

A key question within archival and information science is how to appraise records and information. In her article, the archival scientist Foscarini summarised the four appraisal paradigms, showing how appraisal has changed since the end of the 19th century. She finished her article by going a bit deeper into the current paradigm, operating in the digital era. She explained that in the era of digital preservation, it is no longer only agencies and institutions creating records, but also individuals. This means that information comes and is available from all over the world, but we need to be aware of our digital footprints and respect that not everyone wishes to be heard. An appraisal should therefore consider these factors and focus on what the institutions ‘mean’ to the people inhabiting them. In this regard, Foscarini identified metadata as a key part of the appraisal decision process.22

Similarly, by using archival theory, the scholar in digital history Sternfeld recognised the importance of metadata for digital records to preserve the integrity and reliability of their information. Because metadata provides context and can serve as a means to establish the provenance of the record. This in turn gives a record trustworthiness, as it ensures a proper representation of the historical event. But for it to work, archival theory demands a sound metadata infrastructure for digital records.23 This is a view also shared by Frendo, an archivist at the International Law Association, who put focus on how archival theory has tried to find an automated way to create and capture metadata for digital records, to ease the flow of information as well as manage and preserve digital records. But as digitalisation has changed how metadata, and records, are created there needs to be a “new science of the document” as a

---

22 Foscarini 2017.
response to the contemporary challenges. One part of this challenge is indeed how to apply metadata systems that assist in the creation and management of record metadata. Thus, Frendo concluded with a warning that the adaptation of such a metadata system should not be uncritical. Metadata systems need to be able to support the attributes of records for the present and the future.24

A summary of existing quality assessment frameworks was made by Tani, Candela and Castelli, scholars in information science and technology. In their article, they conducted a critical evaluation of different assessment methods that have been developed to provide digital libraries with high-quality metadata. After analysing and summarising the different quality assessment frameworks and approaches to metadata quality issues, they concluded that three general issues needed to be dealt with for future frameworks. These were 1) the need to develop an open and comprehensive framework that allows people from different fields to define their data quality concepts, 2) develop a machine-processable way to effectively indicate genuine quality aspects, and 3) the development of a tool that can be used to augment quality-oriented aspects of the data.25

To get an understanding of how metadata quality was perceived and worked towards in practice, the associate professor and the metadata librarian Park and Tosaka surveyed cataloguing and metadata professionals in the US. Their findings strongly suggested that high-quality metadata was essential to build successful digital repositories, where more than 80% of the 303 responders recognising the importance of metadata quality control (50.8% strongly agreed and 30.4% agreed). They also found out that although there was a strong recognition of the importance of content standards for metadata at over 70% (47.9% strongly agreed and 34.7% agreed), only 27.7% strongly agreed that semantics was important for metadata quality. Park and Tosaka acknowledged that the low level of strong agreement might be due to the unclear definition of ‘metadata semantics’ in their questionnaire. Nonetheless, 68.3% of the responders agreed with the perceived importance of semantics.

Further on, as the responders were asked which criteria they used to measure metadata quality, 76.9% answered accuracy, 74.3% consistency and 65.0% completeness. Park and Tosaka believed that this was due to as resources can change in type and function across collections, it was better for the metadata to be accurate and consistent across the collections, i.e. that the metadata provided the same representation in all collections. Although completeness means that all relevant metadata elements are used to describe a resource, it is

---

25 Tani, Candela & Castelli 2013.
largely defined by the characteristics of the resource or the nature of the collection within which it is stored.

Lastly, as Park and Tosaka asked the responders what they did to solve metadata issues, 69.9% answered staff training and 63.0% answered using metadata creation guidelines. 43.6% answered that they were using some sort of metadata creation tool. 37.6% of the responders answered that they did periodic samples of the metadata quality, and only 21.5% were using emended metadata creation guidelines in their system. This meant that solving metadata was mostly done by humans, following some sort of training or guidelines. Following guidelines was however not without its problems, as Park and Tosaka noted, as many of these guidelines seldom were shared outside the local environment. This, in turn, could cause that many individual guidelines were developed within an organisation, or across organisations sharing data with each other. Interestingly, the responders were mainly looking at data quality during the creation of the metadata, as only 37.6% said they were doing periodic checks on the metadata.

Similar to the research by Park and Tosaka, Kuźma and Mościcka, from the Military University of Technology in Warsaw, conducted a literature review to see what scholars found to be the most important criteria when assessing metadata quality. They also tried to find out how archival cartographic documents’ metadata were evaluated in digital libraries, only to find out that there are no studies that have metadata of cartographic studies. However, what Kuźma and Mościcka did find out was that completeness was the most popular criterion to evaluate metadata quality, due to its easiness to count and automate. The second most used criterion was accessibility, and in shared third place were accuracy and consistency/conformance. Their research also showed that, due to the lack of coherence in the field of metadata evaluation, some evaluation concepts had the same names but individual authors understood them differently, or that the same concept had different names. To prevent the same type of metadata from being written in many different ways, Kuźma and Mościcka stressed the need to define common criteria names and definitions. Apart from seeing the need for further research on the quality of metadata, they also concluded that if the quality assessment is to be effective, the measuring needs to be automatic.

Looking at the literature it also became clear that there existed some metadata standards, such as Dublin Core, ISO and MARC, guiding how metadata should be created and managed.

---

26 Park & Tosaka 2010.
27 Park & Tosaka 2010, p. 709.
However, the researchers kept describing these different standards either as insufficient,\textsuperscript{29} difficult to use\textsuperscript{30} or that information managers did not comply with them.\textsuperscript{31} As a result, many researchers and organisations made their own frameworks to assess and deal with metadata quality issues. Some of these frameworks took inspiration from these standards and some were developed independently from these standards.

2.2 Existing Frameworks

To offer a way to evaluate metadata quality based on semantics (the meaning of the words) instead of syntactic (looking at grammatical errors) Margaritopoulos \textit{et al.}, from the University of Macedonia in Greece, proposed a set of logic rules, to ensure metadata provides a true picture as possible of the resource. These logic rules were based on how ‘true’ the metadata stays to the resource in the question of correctness, completeness and relevance. The correctness can be seen syntactically, but also in a semantic way of how true the metadata represents the resource. Completeness here refers to if there is enough metadata to describe the resource. Relevance in this case has to be measured based on the context. Margaritopoulos \textit{et al.} acknowledged that there is a high level of subjectivity in evaluating these three criteria. But from these criteria, they formulated three logic rule categories for quality assessment; 1) rules of inclusion, to ensure that a resource’s metadata field includes values of the same metadata for similar resources; 2) rules of imposition, to make it so the metadata values can be equal to a mathematical or logic expression; and 3) rules of restriction, to set up boundaries within which range the value of the metadata is allowed to be in.\textsuperscript{32}

The information scientists Stvilia \textit{et al.} took a more general approach to the problem. They looked at how to assess information quality (IQ) and then applied it to assess metadata in a case study. The result of their research showed a general framework that established causal connections between the level of information quality and potential problems. Their IQ assessment framework consists of measurable IQ dimensions that are grounded in meaningful


\textsuperscript{30} Tani, Candela & Castelli 2013, p. 1200; Marta Fagerlind & Gunilla Gisselqvist, \textit{Metadata Enligt Dublin Core: Tillämpningar och konsekvenser i de svenska kvalitetsmöjligheterna SAFARI, Svenska miljönätet och Sveröök, Master’s thesis} (Lund: Lund University, 1999), p. 69.


\textsuperscript{32} Margaritopoulos, Margaritopoulos, Mavridis & Manitisaris 2008.
attributes of the entity. A central part of their framework is the taxonomy of identified IQ dimensions. There are 22 dimensions in total and can be divided up into three main categories: intrinsic IQ, relational or conceptual IQ and reputational IQ. The categories within the intrinsic IQ can be valued by internal attributes, or in relation to some set standard. Relational IQ measures the relationships between information and some aspect of its usage. The category’s relevance comes from the fact that the context in which the information entity is in can change or the entity of which the information describes. The dimensions of reputational IQ measure the position of an information entity within a cultural activity or structure. Along with these different dimensions, Stvilia et al. also identified four major reasons why information can vary in quality. These reasons can be due to mapping, changes to the information entity, changes to the underlying entity or condition and context changes.\textsuperscript{33}

Trying to expand the metadata quality discussion outside the library community, the academic and the metadata manager Bruce and Hillmann created a framework that assesses metadata after what they found are the seven most commonly recognised characteristics of metadata quality. These are completeness, accuracy, provenance, conformance to expectations, logical consistency and coherence, timeliness and accessibility. These characteristics are in turn to be evaluated on a three tier-scale, explaining the level of the metadata quality, with the first tier having the lowest quality level, and the third one having the highest level. Bruce and Hillmann estimated that for tier 1 and 2 automated processes can be used to ensure quality. But for tier 3, human intervention is more likely needed.\textsuperscript{34}

Influenced by the characteristics given by Bruce and Hillmann, Reiche and Höfig set up their framework with five characteristics to evaluate metadata quality. The framework by Reiche and Höfig aimed at evaluating public government data, that is documents and proceedings that are freely available and accessible. In their framework, the characteristics of completeness, weighted completeness, accuracy, richness of information and accessibility were used. For every characteristic, Reiche and Höfig provided a function to quantify these qualitative values. The characteristics of provenance, logical consistency and coherence and timeliness were excluded due to their difficulty to quantify in functions, and as Reiche and Höfig valued, they bring little further quality to the users. But for the characteristics chosen, the usage of weighted completeness is due to the drawback of treating every metadata field with the same importance. The value of the weight is, therefore, to be set by the metadata field’s perceived importance. The importance can be formulated, for example, by using support from

\textsuperscript{33} Stvilia, Gasser, Twidale & Smith 2007.
\textsuperscript{34} Bruce & Hillmann 2004, pp. 242–54.
an expert in the area or looking at the results of a survey. Accuracy indicates if the metadata is correctly representing the resource or not. Richness of information indicates if the description is useful for the user. Accessibility is how ‘readable’ metadata is in terms of cognitive accessibility, but also physical and logical acceptability. These metrics, according to Reiche and Höfig, should be able to be applied in general, but they needed more validation to be considered sufficiently viable and reliable.\textsuperscript{35}

Also taking a more technical approach to solving metadata quality issues, Hughes provided with his article an open-source algorithm, designed primarily to provide a metadata assessment infrastructure for the Open Language Archives Community (OLAC). The design is a mixture of Dublin Core best practice guidelines and OLAC domain-specific controlled vocabularies. The algorithm assesses the metadata quality primarily based on three features, 1) if the metadata contains OLAC vocabulary the quality increases, 2) if Dublin Core elements are missing then the quality decreases and 3) by combining these two values the algorithm derives a per metadata record weighted aggregate. Based on these three features, one can derive further metrics to evaluate the metadata and the archives within the OLAC context. The framework by Hughes thus provides infrastructural support to assess metadata. Until there exists a computational agent that can objectively manage the quality issues, Hughes recognised that metadata creation and adjustments are to be carried out mainly by humans.\textsuperscript{36}

Recognising that Hughes’ framework was restricted to only assessing metadata within the Component Metadata Description framework (CMD), Trippel \textit{et al.} set out to make a more flexible framework that assessed the quality based on the underlying schema and the metadata instance. Trippel \textit{et al.} also recognised, just like Margaritopoulos \textit{et al.}, that syntactic issues are now a solved problem, thus identifying the importance of semantics for metadata quality. Therefore, to assess metadata, Trippel \textit{et al.} first defined measurable quality indicators based on Dublin Core and ISOCat. It should be noted that the authors decided to use these two standards for this specific case, emphasising that the framework can use other standards as well if so chosen. These indicators were then calculated by using average values or golden standards of the indicators, to be used to calculate the quality score. Trippel \textit{et al.} could distinguish between three types of applications of the score: 1) to measure the quality at editing time, 2) to evaluate a specific repository, or 3) to give an overall assessment of the quality of the metadata originating from the CMD-providers. Seeing the potential to aid metadata modellers in the creation and assessment of metadata, Trippel \textit{et al.} concluded that their framework looks like a

\textsuperscript{35} Reiche & Höfig 2013.

\textsuperscript{36} Hughes 2004.
promising solution, at first glance. However, the framework described in their text is not a formal solution and needed further testing before they could say how effective their framework is.  

2.3 Metadata Control from Sweden

The literature above is very international but has for the most part Anglo-Saxon influence (mostly from Australia, Canada, the US and Great Britain) in one way or the other. Primarily because they were published in a book or journal originating from one of these countries and/or that the author(s) have an Anglo-Saxon background (originating or writing for an Anglo-Saxon institution). As this thesis is written under a Swedish university, this section will therefore describe what has been done from a Swedish perspective to manage metadata quality.

The importance of metadata for digital records and resources is acknowledged for information management and archives in Sweden. For some general examples, the Swedish National Archives (Riksarkivet) have set standards and regulations for how public agencies shall preserve information and metadata in their original conditions over time, as when they were created.  


provision, and metadata management, therefore, needs to be carried out in an organised manner. However, Samuelsson also identified the difficulty in how to create an adequate metadata management guideline. Due to this difficulty, many organisations were missing adequate guidelines for metadata management. Also, the difficulty in implementing international standards was another obstacle.\(^{41}\)

The struggle to set adequate guidelines was also recognised by the archive consultant and the former archivist Kharkina and Lundell, who briefly discussed how Swedish public museums and the Swedish Performing Arts Agency (Musikverket) were working with metadata. Although these agencies were having individual plans for how to achieve high-quality metadata, they were encountering challenges. Primarily how to structure the metadata, as the agencies were managing it in different ways. Another challenge was how to work with international standards and translate the metadata to English, to reach a broader audience. Kharkina and Lundell concluded that metadata management demands competence and a view of the future, to provide systematic and strategic solutions, that work in practice to preserve cultural heritage and its digitalisation.\(^{42}\)

Further Swedish research about metadata quality has mostly been conducted by bachelor and one-year master students, in different fields. In her Master’s thesis, Odlöw studied how metadata in three multi-media art archives affected users’ knowledge production. Taking a critical perspective, she analysed how the creation of metadata shaped how cultural heritage material is portrayed and how it can be used in the future. She found that the metadata created in the three archives was basic and fact-based. Thus in that sense, the metadata did not direct or affect the knowledge production. However, as the metadata did not give a complete picture of the context, it was not as usable for all users and was directing the knowledge production in that sense instead. To give more context, and make the archives more democratic, Odlöw suggested the concept of sharing the creation of metadata with the users. The usage of so-called ‘crowdsourcing’ will generate more context and accumulate knowledge in the archives. Of course, this shared creation needs to be supervised and follow policies, so that no user is misusing this right, and to guide what to do if misuse would happen.\(^{43}\)


\(^{42}\) Anna Kharkina & Mats Lundell, "Metadata vid arkivering av kulturarv", *Arkiv Information Teknik* no. 2 (2019).

Early research about digital metadata was done by Fagerlind and Gisselqvist. In their thesis, they investigated the implementation of the Dublin Core standard at three Swedish search engines. Their research suggests that for Dublin Core, and the search engines, to properly work they needed a quality-controlled web environment for the metadata. The usage of metadata formats also needed to be easy to apply and follow. For that reason, Fagerlind and Gisselqvist were also sceptical if a standard like Dublin Core could cover all the varying needs different search engines had. In that regard, Björkhem and Lindholm held a different view. Looking at metadata creation for digital libraries, they recommended the mixed usage of Dublin Core and MARC to provide simplicity, flexibility, interoperability and standardisation for metadata creation. Especially if the library was to share its collection with other libraries. For this reason, they also suggested caution when changes and adaptations of the formats were to be done. Björkhem and Lindholm also acknowledged that there might be niche libraries with special needs for their collection and metadata. Therefore, when choosing a standard format, libraries must consider what suits the present but also the future.

Further Swedish research on metadata concerns geographical data. Blomgren and Tuoremaa conducted interviews to investigate what was needed for high-quality metadata for geographic information systems (GIS) used by the Swedish military. As the system contained a vast amount of complex geographical data, edited by many different persons across time, the need for quality metadata was crucial to find specific data. Through the interviews and field trips, Blomgren and Tuoremaa were able to produce a data model tailored to the metadata needs of the military usage of the GIS. Similarly, Bolin and Liljebrand conducted interviews with the fire department, police department and ambulance service (in Swedish grouped under the term ‘blue light actors’), to create a ‘blue light map’ containing common and relevant geo data for all the emergency services. For that purpose, metadata was needed to ensure that the geo data was reliable, if and when the data had been updated, the rights to use the data, etc. Bolin and Liljebrand concluded that a common map for the blue light actors promoted cooperation between the actors, ensured the quality of the map as well as shortened the travel time for the blue light actors to reach the scene. Through cooperation, metadata could be generated based on the blue light actors’ local knowledge of the area in which they operated.

44 Fagerlind & Gisselqvist 1999.
2.4 The Gap

The literature showed that metadata is not only important in archival theory but also in many different fields, be it for military use or preserving cultural heritages, and thus needed to be of high quality in all these areas. To ensure this, many organisations and researchers have been trying to set the metadata creation following some standard or guidelines. However, the studies from Park and Tosaka, Tani, Candela and Castelli, Hughes, Samuelsson, Kharkina and Lundell and Fagerlind and Gisselqvist were all critical to the practicality and effectiveness of these methods. Instead, there have been calls from Tani, Candela and Castelli and Kuzma and Moscicka for an automatic procedure to easily and effectively assess metadata quality. There was also a recognised importance that metadata needed to be assessed even after it has been created, shown by the works of Reiche and Höfig, Odlöw, Bolin and Liljebrand, and Hughes. Different frameworks and methodologies to automate and assess metadata quality post-creation have been developed. Many developers of these frameworks also saw the potential for the frameworks to work together with other frameworks. Research verifying if these existing frameworks can be generalised to other contexts and/or be combined with other frameworks was however limited. Meaning that there was a research gap for if the frameworks do indeed work, and if so how well. This lack of research on assuring metadata quality was also generally addressed in the literature itself, and primarily for fields outside the libraries. This thesis, therefore, set out to test a sample of existing frameworks and see how well they could be applied and used in situations other than the very first trial phases, but also together with other supplementary frameworks.

49 Riksarkivet n.d.
50 Park & Tosaka 2010, p. 709.
51 Tani, Candela & Castelli 2013, p. 1200.
52 Hughes 2004, p. 325.
53 Samuelsson 2015, p. 55.
54 Kharkina & Lundell 2019, p. 33.
55 Fagerlind & Gisselqvist 1999, p. 69.
56 Tani, Candela & Castelli 2013, p. 1203.
57 Kuzma & Moscicka 2020, p. 23.
58 Reiche & Höfig 2013, p. 236.
59 Odlöw 2019, pp. 53–6.
60 Bolin & Liljebrand 2016, p. 31.
62 See for example Park & Tosaka 2010, pp. 697–698; Reiche & Höfig 2013, p. 241; Frendo 2007, p. 158.
3. Theoretical Framework

In this chapter, the concept of quality will be discussed in the first section. How it can be defined, its characteristics and the problems linked to it. In the second section, this thesis explains how critical theory can be used as a tool to solve problems using quality measurements. Lastly, the third section will present the archival theory used as middle-range theory, the Information Continuum Model, that will be used to analyse the metadata quality.

3.1 What is Quality?

Although its recognised importance, there is no universally agreed definition of what metadata quality is.\textsuperscript{64} Meaning that scholars and researchers have picked different elements to focus on when evaluating metadata quality in previous studies. How general or specific these definitions are varies. Thus, although drawing inspiration and using other researchers’ definitions, this thesis has formulated its own definition of metadata quality.

In their article, Stvilia et al. recognised that information quality (IQ) is to be identified as “fitness for use” within one or more task(s) or context(s). They also recognised that IQ can be identified as “meeting or exceeding customer expectations” or “satisfying the needs and preferences of its users”, as the need for information and the tasks/actions needed to be performed are both shaping each other. IQ is therefore highly contextual. Stvilia et al., therefore, highlighted the need to consider the context, but also the societal views, when one is to assess IQ.\textsuperscript{65}

By using the purpose of metadata: to adequately and correctly describe a resource so that an end user can get a true picture of the resource even without accessing it, Margaritopoulos et al. identified some important characteristics for metadata quality, using a court case as an analogy. During a court process, it is necessary to hear a witness’ testimony of the facts of the case. Before the witness gives the testimony, they must swear that they will ‘tell the truth, the whole truth and nothing but the truth’. The testimony then provided will be assessed for its quality, based on its distance to the oath. In other words, does the testimony contain all the true facts (i.e. the ‘truth’ and correctness of the testimony), all possible aspects of the facts (i.e. the ‘whole truth’ and completeness of the testimony) and is it relevant to the case under examination (i.e. the ‘nothing but the truth’ and relevance of the testimony)?\textsuperscript{66}

\textsuperscript{64} Tani, Candela & Castelli 2013, p. 1194.
\textsuperscript{65} Stvilia, Gasser, Twidale & Smith 2007, pp. 1721–2.
\textsuperscript{66} Margaritopoulos, Margaritopoulos, Mavridis & Manitsaris 2008, p. 106.
Margaritopoulos et al. described that correctness can be divided into two levels. The first and lower level is objective and inspects if metadata follows syntactical and grammatical rules and standards. If there are misspelt words, inconsistent formatting, inappropriate values in the metadata field, etc., then quality is missing from an objective and syntactical perspective. The second and higher level of correctness is instead more subjective and focuses on semantic rightness: is the representation true and absent of any deception? Completeness of metadata refers to if it can provide a sufficient description of the record. Here it is important to decide which metadata fields are important and which are not, to provide a sufficient description. Lastly, even if the describing metadata record is complete and correct, the values in the metadata fields might not be relevant for the context of use. Although relevance can seem to be similar to correctness and/or completeness, the discriminating factor here is to look at the context of usage.\textsuperscript{67} For example, a metadata value describing an audiobook can be correct when it describes which format it was recorded in but be irrelevant as one would be more interested in if one can open the format at hand, or be complete in describing the narrator in the original language but be irrelevant after it has been translated into another language. Notably, one can see that there exists two levels of metadata quality: one lower and objective level that is easy to solve, and one upper and subjective level where the quality is to be decided and evaluated based on the context.

Another important feature of metadata quality, as Bruce and Hillmann highlighted, is that the question of the quality is passed down from creator to aggregator to user. The user in turn trusts that someone in the chain of the information provision has been paying attention to what they see as quality.\textsuperscript{68} In other words, that someone has ensured quality for the end user. Apart from that, Bruce and Hillmann’s 3-tier framework (introduced in section 2.2) also provides a good way to categorise the quality level of metadata. Metadata in the first tier is identified using the following indicators:

- The ability to validate the syntax and syntactic binding (the schema).
- Be used by the appropriate namespace declaration.
- Containing an administrative “wrapper” to identify it uniquely.

The quality can be improved further when the elements of tier two are applied:

\textsuperscript{67} Margaritopoulos, Margaritopoulos, Mavridis & Manitsaris 2008, pp. 106–8.
\textsuperscript{68} Bruce & Hillmann 2004, p. 242.
• Using controlled vocabularies, linked to publicly available sources of terms.
• If available, elements defined by a specific community, to be discoverable by the community.
• A full complement of general elements, free of assumptions about who will use the metadata, for general discoverability.
• Provenance information about the methodology used for creating the metadata.

For the highest level of metadata quality, the indicators for tier three are:

• An expression of metadata intentions, based on an explicit, documented application profile and registered in conformance to a general metadata standard.
• Metadata from a trusted source, known to regularly update the metadata and include a controlled vocabulary.
• Full provenance information.69

The quality concept from Stvilia et al.70 and the feature of quality being “passed down the line”, recognised by Bruce and Hillmann,71 resonates with general philosophical definitions of what constitutes as quality in business. In short, three experts, William Edwards Deming, Philip Crosby and Joseph Juran, have constituted three definitions of quality. Deming took a more statistical approach and said that the quality of a product reflects the quality of the process. In other words, Deming’s definition does not include the customers’ needs from the product. The customers’ needs are instead included in Crosby’s definition as well as Juran’s. According to Crosby, quality comes from conformance to customers’ requirements or needs. For Juran on the other hand, the definition of quality stems from the fitness of a product for its intended use. Juran’s definition thus incorporates the customers’ need the most.72

It becomes clear that the definition of metadata quality, just like information quality or the quality of a product (or service), must take the expectations and requirements of the end users at heart. This of course leads back to the issue stated in the Introduction, as well as above in this chapter: the loose definition of “fitness for use and satisfying the needs and preferences of

---

71 Bruce & Hillmann 2004.
its users". The needs and preferences of the end-users are open up for subjective speculation. Quality thus becomes like beauty, to be evaluated in the eyes of the beholder. Or, as Bruce and Hillmann so delicately put it: "Like pornography, metadata quality is difficult to define. We know it when we see it, but conveying the full bundle of assumptions and experience that allow us to identify it is a different matter". But even though a more narrow and precise definition of metadata quality is in the realm of subjectivity, is there a way to use theory to make it more objective?

3.2 Grand Theory: Critical Theory

This thesis is interpretive in nature. That means that its ontological view is that there exists no single reality. Instead, there are multiple complex realities, each individually observed and embedded in its context. Epistemologically, this means that all entities are in a constant state of simultaneous shaping. All knowledge we acquire is the result of the interaction between the researcher and the subject, bound to the time and context of when the interaction happened. The interpretive focus on context resonates with the theoretical concepts of quality addressed earlier in this and the previous chapter. This thesis will therefore use critical theory as a grand theory, as it engages with examining current thoughts and social structures.

The usage of critical theory can give researchers new perspectives, and prevent them from getting stuck with seemingly neutral descriptions of what exists. It can also help prevent the reproduction of institutionalised power relations taken for granted. Critical theory, therefore, aims to raise awareness of the political character of social phenomenon and encourages critical reflection over what one might take for granted, both what one observes and over oneself. Using critical theory thus enables the possibility to see the end-users’ perspective when defining the parameters that shall provide metadata quality. It can in other words provide a way to incorporate the needs and expectations of the end-users, as Stvilia et al., Bruce and Hillmann, Crosby and Juran saw as necessary for quality.

Another reason for using critical theory is the current archival paradigm. The so-called community paradigm, or the questioning paradigm, identified by John Ridener, is heavily

73 Stvilia, Gasser, Twidale & Smith 2007, p. 1721.
74 Bruce & Hillmann 2004, p. 238.
76 Pickard 2013, p. 11.
80 Mitra 2016, p. 82.
influenced by critical theory. Primarily by using postmodernism, archivists in the current paradigm are trying to appraise digital records so they resonate with the public. The archivists’ goal is to understand the relationship between representation and reality, history and memory, place and identity as well as archival practice and social needs. The current paradigm also acknowledges that appraisal occurs based on personal criteria rather than a universal and objective rule. The personal criteria are also recognised as actions that are socially grounded in time and space. Thus, the questioning paradigm, by using critical theory, questions assumptions that have previously been taken for granted for the archivist to gain an understanding of the dynamic of the records and information, and thus become culturally engaged with creators and the public.\footnote{John Ridener, From Polders to Postmodernism: A Concise History of Archival Theory. (Duluth, Minnesota: Litwin Books, 2009), pp. 101–40.}

Again, the usage of critical theory opens up the possibility to look beyond one’s perception of quality, and possibly also indoctrinated and taken-for-granted concepts of quality. Instead, it can provide a way to, not only find and understand the customers’ needs, but also a way to update one’s understanding of the needs. This is a crucial part of ensuring the metadata quality, because, as Kużma and Mościcka stated “Understanding contemporary needs is the basis for understanding why /.../ resources should be described in a special way.”\footnote{Kużma & Mościcka 2020, p. 23.} Ensuring high metadata quality is therefore not a one-time task but rather iterative, as needs, expectations and also the space in which the metadata exists can change. Something Reiche and Höfig also acknowledged.\footnote{Reiche & Höfig 2013, p. 236.}

The advantages and suitability of using critical theory as a guide to ensure metadata quality has thus been identified. But the above-mentioned question of ‘if a theory can help make the subjective definition of quality more objective’ remains unanswered.

Building on the court case-metadata analogy from Margaritopoulos et al., we can identify a missing piece for how the verdict of the assessed metadata quality goes. Indeed, a testimony in a court of law shall hold the truth (be correct), the whole truth (be complete) and nothing but the truth (be relevant), just like metadata describing a resource.\footnote{Margaritopoulos, Margaritopoulos, Mavridis & Manitsaris 2008, p. 106.} However, the extent of a verdict made in a court of law is to be set by, indeed, the law. Although there are many different types of governments and methods on how to set laws, the starting point is always subjective (‘this should be illegal/legal’). But once set, the laws can operate as an objective standpoint (‘this is illegal/legal’). A philosophical example of this is the deontology of Kant’s categorical
imperative, where rules are unconditional or absolute to determine if an action is good or bad.\textsuperscript{85} Or, to put it in more relevant terms, the laws and rules outlying how the management of the data and information is to be done shall be found in the company’s data governance. The data governance rules that will guide the management of metadata quality can in turn be based on the criteria and contemporary understanding of the customers’ needs identified by using critical theory. In other words, objectivity can be ‘implemented’ in the process by setting rules for how the management of metadata quality should be. More objectivity can also be considered to be added, as these rules are based on the requirements of the end users. However, it might be necessary to mention that these laws and rules will, just like any government’s laws and rules and following critical theory, not be static. Instead, they must be updated and adjusted to fit the contemporary situation.

\textbf{3.3 Middle-range Theory: Information Continuum Model}

As critical theory is a so-called grand theory, it only provides an abstract explanation of how the world works. Therefore, a smaller, so-called middle-range theory, is needed to give a more detailed, but limited, explanation and understanding of a specific situation or phenomenon.\textsuperscript{86} This in turn shall aid this research in preventing the results to be too abstract or vague.\textsuperscript{87} In this thesis, archival theory will be used as middle-range theory.

The archival theorist Eastwood explained that the purpose of theory in any field is to give explanations for general facts and to contemplate the nature of a subject. In an applied discipline, like archival science, theory should relate to method and practice to decide what to do when. Archival theory, Eastwood explained, is needed for the archivist to analyse the purpose and functions of the archive. The role of the archivist is not only to save everything that is historically relevant in the archive, but also to know the nature of the record they manage, and for that theory is needed.\textsuperscript{88} As Eastwood stated: “theory dictates the social agenda of archivists, who stand as protectors of evidence to ensure that social relations may be pursued on objective grounds.”\textsuperscript{89} In other words, through their work, archivists give trustworthiness, accountability and authenticity to the information found in the archive. But this work needs to be guided by theory to set methods and practices. If not, the information can ‘go bad’, become

\textsuperscript{85} The Editors of Encyclopaedia Britannica, "categorical imperative", in Encyclopedia Britannica (2023).
\textsuperscript{87} Alvesson & Sköldberg 2017, p. 208.
\textsuperscript{89} Eastwood 1994, p. 129.
corrupted or falsely interpreted, for example when the context of a record’s creation is missing.90

The purpose of archival theory explained by Eastwood resonates with the previously mentioned criteria for metadata quality, that is to make sure that the metadata gives a true and relevant representation of the resource.91 To understand contemporary needs, archival theory through the Information Continuum Model works as an effective theoretical tool. Based on the Records Continuum Model, which is intended to be used by archivists to map and analyse the uses of records over time, the Information Continuum Model was developed to aid librarians analyse the representation, recalling and dissemination of information.92 Even though the Information Continuum Model is intended for librarians and not archivists it is still very much useful for this thesis. Because, as Eastwood made clear, the grounds of theory most suit the perspectives and purposes of the applied discipline. For that reason, the knowledge for the theory can be drawn from other fields and disciplines, if they are relevant.93 Thus, it is relevant as metadata is used to give a true representation and aid in finding (recalling) and thus reusing (disseminating) information resources.

The Information Continuum Model consists of four dimensions: create, capture, organise and pluralise, through four perspectives: action/structure, technology, storage/memory and categorisation (see figure 3.1 below). Action/structure explains in which structures our actions take place and how we influence the structures. Technology concerns the different tools used to manage records. Storage/memory is where the information is located. Categorisation concerns how to classify and the levels of classing information.94 The dimension ‘create’ equates to the act of when a record or a resource is created. Capture is when the created record or resource is brought into a system. Organise concerns how the record or resource shall be stored and classified in the given system. Lastly, pluralisation is how information is taken beyond the organisation’s context into forms of societal totalities.95 To better understand pluralisation, Frings-Hessami recommended thinking of the societal embeddedness in where a record or resource is created, i.e. the sociocultural context, technology used, laws and

95 Upward 2000, p. 122.
customs. The fourth dimension thus allows adherence to the contextual and societal views, important aspects for assessing information quality highlighted by Stvilia et al.

![Information Continuum Model by Frank Upward](image)

**Figure 3.1: The Information Continuum Model by Frank Upward.**

To illustrate an example of the process of how information moves through the dimensions, when this thesis is being written it is in the create-dimension. Once finished it will be submitted to the university’s database, and thus be captured as a master’s thesis. It will later be organised in the database based on the subject, keywords, the faculty it was part of etc. Lastly, the thesis will be pluralised when (hopefully) this thesis will be used to set a generalisable method and view on how to manage metadata quality issues.

When applying the model, it usually goes as the mentioned example where one starts from the inner dimension, create, and works outwards through capture and organise until pluralise. From pluralise, it is possible to start over again at create, as a record or resource is shared and reused. But it is also possible to read the model inwards. In that case, one would first look at the societal embeddedness one is located in. What laws, norms and technology affect us and in which sociocultural context are we? Then one looks at how these factors influence the

---

96 Frings-Hessami 2021.
98 Upward 2000, p. 130.
organising and management of systems. In the capturing part, one would look at which systems one would like to store the record or resource, based on preferences, but influenced by the organisation and pluralisation dimensions. Lastly, the creation will then adhere to the requirements from the three other dimensions.\textsuperscript{99}

To illustrate this process with the same example, the process of writing this thesis needed to meet the laws, rules and customs of how to conduct research at a university (for example, the research has to be ethical). But as it is a master’s thesis, there were also academic expectations for this thesis to meet (for it to pass and get a high end-note as well). Based on this societal embeddedness in the fourth dimension, the university in turn set up certain procedures for how they would organise this thesis. For example, using the e-service Ouriginal to inspect this work for plagiarism, and peer review to check the thesis’ credibility. The capture dimension had to provide a system that adhered to the pluralisation and organisation dimensions. In this case, it was the e-service DiVA, which served as a way not only to store the thesis but also how to make it accessible and spread its scientific information. Lastly, the dimension of creation was affected by all of these other dimensions, how the research was done, the citation of sources, what was written and how it was written (as it was to be published and inspected using digital tools, the thesis was written using Microsoft Word, another digital tool).

This process, of working from the outside inward, is very useful as it provides a way to analyse information a priori before it is created. This is a way to make sure that the information will stay relevant and usable for as many stakeholders as possible for as long as they need them.\textsuperscript{100} Using the Information Continuum Model thus made it possible to analyse the metadata, before it reached the end-users, to see if it was of relevance and was of quality for them.

\textsuperscript{100} Frings-Hessami 2021, p. 151.
4. Methodology

This section will first describe the type and characteristics of the case that will be used in this study. The section afterwards will illustrate the methodology that will be used to measure, identify and solve metadata quality issues. The section will have sub-sections for each part of the process explaining how each phase was conducted in detail. The last section of this chapter will present the mixed methods used to analyse the quality of the metadata. It will first be a quantitative data collection, looking at the results from applying the methods to measure, identify and solve the audiobook metadata issues. Following, there will be a qualitative data collection through interviews with relevant stakeholders, giving their input on how the assessment method could potentially aid them in their work to ensure high metadata quality.

4.1 Instrumental Case Study

A case study was conducted to determine if the assessment method worked. As Halperin and Heath stated, using a case study is a way to find out if a theory can be applied and ‘work’ in another context other than for the context it was developed in.\textsuperscript{101} Of course, there are many ways how to conduct case studies, depending on time, scope, amount of cases and what is to be studied. For the many different ways of how to do case studies, one can in general identify three types of case studies of what they investigate. These three types are: intrinsic, instrumental and collective. In short, intrinsic case studies aim to gain more knowledge of the case itself. Instrumental case studies are for investigating a certain phenomenon or theory, where the case itself becomes the setting or vehicle for the investigation. Collective case studies use more than one case to investigate a certain phenomenon or theory.\textsuperscript{102} As this thesis set out to test existing frameworks, it was therefore an instrumental case study.

When doing instrumental case studies, the researcher finds an issue or concern to focus on, in this case, metadata quality issues, and then selects a bounded case to illustrate the issue. Although the focus was on the issue and not the case, it is still important to describe the case where the research takes place,\textsuperscript{103} as well as the boundaries due to limitations in time and resources.\textsuperscript{104} The time scope for this thesis was 20 weeks, the time granted for writing. As

\textsuperscript{101} Halperin & Heath 2017, p. 215.
\textsuperscript{102} Pickard 2013, p. 102.
\textsuperscript{104} Pickard 2013, p. 103.
mentioned in the introduction, the thesis used data and resources from a company (from now on called ‘The Company’ as they wished to remain anonymous) which allowed me to conduct my research, meaning that this was a single-N case study. Although not an archive, The Company was deliberately chosen as it was a large media company located in Europe, which partly worked with developing an online streaming platform. For that reason, The Company had access to multimedia content, such as videos, music and audiobooks. They were at the time of writing facing issues with the metadata quality of their resources coming from other companies further up the pipeline. These issues needed to be solved before the resources, with their metadata, were passed on further down the pipeline, and went online and became public. In other words, The Company offered an opportunity to research the topic of metadata quality, which was relevant for archival and information science, not only in theory but also in practice.

To narrow down the scope of the metadata that would be analysed, I asked The Company what type of resource and what type of metadata they had the most issues with. They responded that the resources with the most pressing issues were audiobooks. Thus this thesis set out to look at the audiobooks’ metadata quality defined from the end users’ perspective (as explained in section 3.1). The types of metadata that The Company identified to be the most critical ones, in descending order from most important to least, were: title, author, description, language, speaker, genre, length, part of series and release date. Thus, the thesis also limited itself by looking at these metadata fields.

4.1.1 Strengths and Weaknesses of a Case Study

Conducting a single-N case study comes with some disadvantages. There is the risk of selection bias, where the result is not accurate and instead misleading. The risk is higher when the study tries to investigate a broader phenomenon and make some sort of generalisation.\(^\text{105}\) There might indeed be a high level of selection bias in this case. The choice to approach The Company as a case study was partly due to connections I had with some employees, as I have been working with them earlier. This of course raises some other bias issues, which will be discussed in subsection 4.1.2 below. However, the reason why I wanted to use The Company as my case was due to its broad portfolio in data products and data projects. The Company could provide me with the rich and detailed insights that Pickard highlighted as important criteria for choosing a site.\(^\text{106}\)

\(^{106}\) Pickard 2013, p. 104.
Another disadvantage concerns the ability to generalise. When doing a small-N, and particularly a single-N, it becomes hard to generalise the findings. However, as this thesis is building on the findings of previous research and applying them in a different case, the level of generalisation will take, and give, support to the previous findings. In other words, although this is a single-N case study it is part of a cluster of other case studies investigating the same phenomenon. Thus, the possibility for generalisation of the results from this thesis must be evaluated together with the results from the other research covering metadata quality. As Halperin and Heath mentioned, one way to have confidence that the results from a small-N study can be robustly generalised is to replicate the study but with a new case,¹⁰⁷ which this thesis did.

Some strengths of conducting a single-N case study is that it allowed going deeper into detail than what a large-N would allow.¹⁰⁸ Using a single-N also allowed for more time and effort spent on approaching the informants and ‘social actors’ inhabiting the case at its heart with their own experiences and knowledge.¹⁰⁹ This was of benefit as the research design of this thesis was a mixed method. Another strength was that by using a single-N case study, the internal validity is much higher than compared with a large-N study.¹¹⁰

4.1.2 Ethical Role and Bias

There is a need to be transparent to explain my ethical role and potential bias in this research. To start the transparency, the answer to the question ‘For whom shall this research be useful?¹¹¹ are archivists and the academy. Meaning this thesis meant to further the research in archival and information science by executing the purpose of this thesis and finding an effective and easy method to assess metadata quality, that can be generalised. The results of this thesis are primarily directed to archivists trying to gain control of the records’ metadata in the archive, but should also be useful for others managing different resources in any repository.

It is also important to go further into why I decided to do this research by The Company. As mentioned above, I had contacts there from before I began writing this thesis which made it easier to approach The Company. The Company’s portfolio of data products was well-suited to find a research area that would be relevant to this thesis. It should also be noted that when I

¹⁰⁸ Creswell & Poth 2018, p. 102.
approached them I had other ideas for the research topic. But as I was pitching the ideas to The Company, they had the suggestion to look into their metadata issues, as it seemed relevant to my studies and was something that they needed help with. In other words, they helped me to find a limited area and relevant issue in the field of archival and information science to study. Thus, the level of selection bias goes as far as I had contacts at The Company before I approached them. If I had not, I would have tried to approach some other, but similar, organisation with my ideas. Then the research would be different in location and research area.

The last point for my ethical role in this research concerns anonymity. As mentioned, The Company has been anonymised in this work. This has a drawback as it makes it near impossible to verify the results from this thesis.\textsuperscript{112} However, it was seen as necessary as a safety measure to respect the clause in my contract with The Company, which stated that this research had to be careful with The Company’s data. If not anonymised, competitors could for example get an advantageous inside view of how The Company worked. This thesis should not be used as a secret intel for the competitors, but as a way to find a generalisable method to assess metadata quality. This also ties into the next reason for anonymising The Company, which was that The Company had no objections to the publication of this thesis, as long as they remained anonymous. In other words, it was seen as a necessary step to conduct and publish the research, without having to break any ethical rules or have any conflict of interest.\textsuperscript{113}

4.2 The Assessment Method

The assessment method to measure, identify and solve the metadata quality issues, was divided into three individual steps. I have visualised the process in Figure 4.1 below. After the metadata was received, the first step was to measure the quality level, i.e. quantifying the quality by putting a number on it. For this purpose, the framework by Reiche and Höfig\textsuperscript{114} was used, as they recognised that their framework needed more validation to be considered sufficient and reliable. The choice was also influenced by the fact that it was not dependent on any criticised, external standard, like Dublin Core. Instead, the framework consists of mathematical equations that can, without much effort, be applied and used either through a tool like Microsoft Excel or programmed in code. This can therefore allow people from other fields to define their quality needs in the equations while also automating the process, which was called for by Tani, Candela

\textsuperscript{112} Vetenskapsrådet 2017, p. 41.
\textsuperscript{113} Vetenskapsrådet 2017, p. 46.
\textsuperscript{114} Reiche & Höfig 2013.
and Castelli. The framework by Reiche and Höfig also provided a way to calculate the most important factors of metadata quality identified by Park and Tosaka (accuracy, consistency and completeness) and Kužma and Mościcka (completeness, accessibility, accuracy and consistency/conformance). It also conformed to the three logic rules set by Margaritopoulos et al. and it drew inspiration from Bruce and Hillmann. It was therefore seen as the best framework, without the weaknesses of following complicated standards other frameworks have, while still encompassing the most needs called for by other authors and scholars.

![Figure 4.1: Visual representation of the quality proofing process](image)

The next step was to identify the metadata quality level. This step was separated from the step of measure, because even though there was a number to the value, what the number meant and what its impact was still needed to be defined. In other words, there needed to be a process of giving the ordinal data meaning. This process of identification had to be drawn from the theory of what quality is, explained in section 3.1 above. Thus, the articles from Bruce and Hillmann and Margaritopoulos et al. were used as they provided a theoretical framework to classify the metadata quality level based on its character and sort it on a three-level scale, which also indicated if the issue was objective or subjective in nature.

Knowing the character of the metadata quality would in turn be important for the process of solving the issues. As the quality was to be perceived by the end users, the Information Continuum Model was chosen as a support to get the end users' perspective along with the framework by Stvilia et al. The reason for using the framework by Stvilia et al. was partly due to its problem-solution focus, but also because they already applied the framework to two sets of metadata. The chances for the framework to work was therefore higher, as the thesis brings accumulated knowledge for if the framework works in other contexts.

---

115 Tani, Candela & Castelli 2013, p. 1203.
119 Bruce & Hillmann 2004.
120 Bruce & Hillmann 2004.
121 Margaritopoulos, Margaritopoulos, Mavridis & Manitsaris 2008.
4.2.1 Measure

To measure the level of metadata quality, the quality metrics by Reiche and Höffig\textsuperscript{123} were used. They used mathematical formulas to quantify qualitative values. The values were Completeness, Weighted Completeness, Accuracy, Richness of information and Accessibility. Although Accessibility is an important factor for metadata quality, in Reiche and Höffig’s article they saw it more in concern with the trouble of using governmental jargon.\textsuperscript{124} As the audiobooks varied in their target audience (children, adults, education level, interests, etc.) it would have required a separate formula for each target group. Therefore, due to the limited scope of this thesis, accessibility was not used.

The formulas used to calculate each value are shown in Table 4.1 below and originate from the work by Reiche and Höffig. Note that Richness of Information (RoI) appears two times. This is because Reiche and Höffig provided two formulas for the richness, one for when a field contains only one or few numerical or vocabulary values and one for free text.\textsuperscript{125}

To calculate the values, Microsoft Excel was used with sheets containing the metadata for the resources. For detailed explanations and examples of how Excel was used to calculate each value, see the Appendix under the section ‘Excel formulas’.

\textsuperscript{123} Reiche & Höffig 2013.
\textsuperscript{124} Reiche & Höffig 2013, p. 238.
\textsuperscript{125} Reiche & Höffig 2013.
<table>
<thead>
<tr>
<th>Metrics</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completeness</td>
<td>$\frac{\sum_{i=1}^{n} [field_i \neq null]}{n}$</td>
</tr>
<tr>
<td>Weighted completeness</td>
<td>$\frac{\sum_{i=1}^{n} w_i [field_i \neq null]}{\sum_{i=1}^{n} w_i}$</td>
</tr>
<tr>
<td>Accuracy</td>
<td>$1 - \frac{\sqrt{\sum_{i=1}^{n} d(field_i)^2 + 0.0001}}{\sum_{i=1}^{n} d(field_i) + 0.0001}$</td>
</tr>
<tr>
<td>Richness of information (numerical and vocabulary)</td>
<td>$\frac{\sum_{i=1}^{n} - \log P(Value_i)}{n}$</td>
</tr>
<tr>
<td>Richness of information (free text)</td>
<td>$\frac{\sum_{i=1}^{n} tf(word_i) \cdot \log(\frac{m}{df(word_i)})}{n}$</td>
</tr>
</tbody>
</table>

Table 4.1 Metrics and formulas

The value for the Completeness of an audiobook is based on the ratio of all non-empty metadata fields (field; $\neq$ null) to the total number of metadata fields ($n$) that were used to describe the audiobook. Weighted Completeness was calculated by summing up all non-empty fields (field; $\neq$ null) multiplied by their respective weights ($w_i$). This sum was then divided by the sum of the total weights ($w_i$).

In the formula for Accuracy, the difference ($d$) between internal and external data values was calculated. In this case, the internal metadata provided by The Company (from now on called the Given dataset) was compared with metadata from external sources describing the same audiobooks (from now on called the Comparing dataset). An explanation of how the external metadata was collected is given below in section 5.1.1. The smaller the difference the greater the accuracy. Depending on the nature of the metadata value, the difference could be found by using a Boolean calculation (for vocabulary values) or subtraction of one numerical value with the other. The sum was then used to calculate the Euclidian distance ($\sqrt{\sum_{i=1}^{n} d(field_i)^2}$) divided by the sum of the differences ($\sum_{i=1}^{n} d(field_i)$). To prevent errors in the calculations if both sums were equal to zero, both the numerator and denominator had 0.0001 added to them.

The Richness of information for numerical and vocabulary values calculated the informativeness based on how unique the value ($Value_i$) was. The probability ($P$) indicated how often a certain value appeared in the dataset, and was multiplied with $-\log$. The product for each
\(-\log P(\text{value})\) was summed up and then divided by the total amount of numerical and vocabulary values \((n)\) used in the dataset, to get the richness of information for numerical and vocabulary values.

Calculating the Richness of information for free text a Term Frequency-Inverse Document Frequency (tf-idf) function was used. The term frequency (tf) returned how often the selected keyword \((\text{word}_i)\) occurred in the data set. The sum of the keyword frequencies was multiplied by the logarithm \((\log)\) of the total amount of audiobooks \((m)\) divided by the number of audiobooks which had the word \((\text{df(word))}\) in the description. The sum of the tf-idf calculation was at the end divided by the used amount of keywords \((n)\).\(^{126}\)

In summary, these calculations are evaluating different aspects of the metadata. The Completeness evaluates if metadata is presented or not. Weighted Completeness tells if the most relevant or important metadata is present or not. In other words, these two calculations are checking the existence of the required elements, in this case, the metadata. Whereas the calculations for the RoI and Accuracy are evaluating the content of the metadata.

4.2.2 Identify

After measuring the metadata quality, it was necessary to identify the quality level by putting it in a framework for classification. This would serve as a guideline on how crucial the issue was and what the following step(s) would be. The identification schema was based on the three-tier classification by Bruce and Hillmann\(^{127}\) along with the classification from Margaritopoulos \(\textit{et al.}\)\(^{128}\) The classification table to identify the level of metadata quality was structured in the following way:

<table>
<thead>
<tr>
<th>Tier level</th>
<th>Objective or subjective</th>
<th>Characteristics</th>
<th>Examples</th>
</tr>
</thead>
</table>
| 1          | Objective               | • All fields filled in  
• Values are error-free  
• Values are appropriate                                                      | • When all fields are filled in, Completeness should have the value 1.  
• No errors or abnormalities in the Excel calculations.  
• Input in fields follow rules set in Excel. |
| 2          | Subjective              | • Controlled vocabulary  
• Description containing a variety of elements  
• Metadata is given free from assumption.                                              | • Fields follow predefined conditions (e.g. set terms for genres)  
• When Richness of information has a high value, there should be a variety of elements in the description.  
• The metadata should be general and independent from any particular community. |
| 3          | Subjective              | • Vocabulary used by the target audience  
• The metadata is similar/updated throughout repositories  
• Metadata fits the context                                                               | • For relevant vocabulary, the Richness of information for free text should have a high value.  
• High level of Accuracy  
• For example, a suspense thriller should have a high age restriction                  |

Table 4.2 Identifying the issue

Note that the scale went similarly to the scale from Bruce and Hillmann, meaning that tier level 1 was the lowest and 3 was the highest. One way would be to translate the meaning of each tier into 1 = sufficient, 2 = good and 3 = excellent. Of course, there were instances where the metadata quality could not even be identified to be classified as tier 1. These were the
audiobooks which needed the most urgent improvement in metadata quality, as the level was insufficient.

4.2.3 Solving

Due to the limited resources of this thesis, the part of solving the quality issues was limited to what the potential solution(s) could be, not actual problem-solving. Nonetheless, the process of solving the quality issues must be based on its scope and nature. Therefore, after the issues were identified, they were compared with the Information Quality (IQ) metrics presented by Stvilia et al., who also mapped what the sources of IQ problems could be, to find potential solutions. The four sources of IQ problems are mapping, changes to the information entity, changes to the underlying entity and context changes.

Mapping problems occur when the information is incomplete, ambiguous, inaccurate, inconsistent or redundant in its description of a state, event or entity. IQ problems can also arise if the information entity (in this case, the metadata) changes, or the underlying entity it represents (in this case, the audiobooks). Changes in context occur when the information object has moved from one culture or sociotechnical structure to another.

These four sources can then affect any of the 22 IQ dimensions that Stvilia et al. identified. These 22 dimensions can be split up into three categories: intrinsic, relational or contextual and reputational. Intrinsic dimensions assess the information by internal attributes or by following some standard. Relational or contextual dimensions tell the relationship between the information and the context in which it is used. Reputational dimensions measure where an information object is located within a cultural or activity structure.

Based on the nature of what metadata does as information, Stvilia et al. used the 18 dimensions from the intrinsic and relational or contextual categories to identify IQ problems. These will also be used in this thesis and are shown in Table 4.3 below. It should be noted that their definitions of the dimensions have been ‘translated’ to this context, thus explaining the dimensions of the metadata.  

\[\text{129 Stvilia, Gasser, Twidale \\& Smith 2007, pp. 1722–9.}\]
<table>
<thead>
<tr>
<th>Type</th>
<th>Dimension</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrinsic</td>
<td>Accuracy/validity</td>
<td>Legitimacy or validity of the information following some reference source.</td>
</tr>
<tr>
<td></td>
<td>Semantic Consistency</td>
<td>The extent of using the same value (vocabulary control) and elements.</td>
</tr>
<tr>
<td></td>
<td>Informativeness/Redundancy</td>
<td>How informative the metadata is.</td>
</tr>
<tr>
<td></td>
<td>Precision/Completeness</td>
<td>The quality of a model to generate information following a general-purpose ontology.</td>
</tr>
<tr>
<td></td>
<td>Currency</td>
<td>The age of the metadata</td>
</tr>
<tr>
<td></td>
<td>Naturalness</td>
<td>The extent to which the model or schema and content of a resource are expressed by conventional, typified terms and forms according to some general-purpose reference source.</td>
</tr>
<tr>
<td></td>
<td>Cohesiveness</td>
<td>The extent of the content is focused on one topic.</td>
</tr>
<tr>
<td></td>
<td>Complexity</td>
<td>The extent of cognitive complexity of the metadata is measured by some index.</td>
</tr>
<tr>
<td></td>
<td>Structural consistency</td>
<td>The extent of similar attributes or elements is used for the metadata to provide the same structure, format and precision.</td>
</tr>
</tbody>
</table>

*Table 4.3 IQ dimensions*
<table>
<thead>
<tr>
<th>Type</th>
<th>Dimension</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relational/Contextual</td>
<td>Accuracy</td>
<td>How well the information represents the object, event, process or phenomenon.</td>
</tr>
<tr>
<td></td>
<td>Semantic Consistency</td>
<td>The extent of using the same value (vocabulary control) and elements, suggested by an external standard.</td>
</tr>
<tr>
<td></td>
<td>Informativeness/Redundancy</td>
<td>The degree to which the information is new or informative.</td>
</tr>
<tr>
<td></td>
<td>Precision/Completeness</td>
<td>The extent the information object matches the precision and completeness needed in the context.</td>
</tr>
<tr>
<td></td>
<td>Volatility</td>
<td>The amount of time the information remains valid in the context of a particular activity.</td>
</tr>
<tr>
<td></td>
<td>Naturalness</td>
<td>The degree of the model and content of the metadata are semantically close to the resources.</td>
</tr>
<tr>
<td></td>
<td>Complexity</td>
<td>Degree of cognitive complexity of the metadata concerning a particular activity.</td>
</tr>
<tr>
<td></td>
<td>Relevance</td>
<td>The extent the metadata applies to a given activity.</td>
</tr>
<tr>
<td></td>
<td>Structural consistency</td>
<td>The extent of similar attributes or elements is used for the metadata to provide the same structure, format and precision, suggested by an external standard.</td>
</tr>
</tbody>
</table>

Table 4.3 IQ dimensions (cont.)

Using the framework and IQ definitions From Stvilia et al. not only started answering the question of ‘how to think’ when solving the issue, but also if it should be a manual or automatic task. Because we live in a digitalised world, it is recognised that manual techniques are not
enough to manage the ever-expanding volume and diversity of metadata. The need to automate the process has been called for by many researchers. Although not time-efficient, it is often recommended in the literature that there needs to be a human in the loop to ensure high metadata quality. As Foscarini and Oliver wrote:

The notably missing piece of the puzzle is the ‘human’ component of every human activity system, where people carry out their purposeful activities through the information they create. People, processes, and structures are all part of human activity systems and are inextricably linked to one another.

In other words, to stay efficient, the process needed to be automated, but there was also a need for human intervention. This can at first seem like a contradicting situation, but as mentioned in the identification the metadata quality can be divided into different levels, and if the problem is objective or subjective in nature.

4.3 Data Collection: Mixed Method

To test the effectiveness of already existing frameworks and how well they adapt to new contexts, the analysed data needed to be some sort of indicator of how well the frameworks worked. For this reason, a mixed method was chosen to combine quantitative data with qualitative data. The assessment method was analysed quantitatively, using statistical calculations. The results from the qualitative analysis were later used to formulate interview questions for a qualitative analysis explaining how well the assessment method produced valuable information and helped to solve metadata quality issues.

More concretely, the first two steps of the assessment method, the measuring and identification, used quantitative methods to collect the data, as they were numerically calculated. After quantitative and numerical values for the metadata quality levels were calculated, the data for the last step, solving, was collected by analysing the results from the

---

131 See for example Reiche & Höfig 2013; Trippel, Broeder, Durco & Oehren 2014; Gilliland & McKemmish 2012.
measuring and identification. Data collection for if the assessment method did indeed bring any valuable information to stakeholders was done through interviews. The interview questions were designed to investigate how the relevant stakeholders at The Company incorporated the theoretical definition of quality used in this thesis (see section 3.1), their opinions of the assessment method as it was, and what they thought was missing or could be improved with the method.

Using a mixed method came of course with the issue of needing to provide both the trustworthiness of qualitative research and the rigour of quantitative research. Meaning that the result should have validity/credibility, be generalisable, have synchronic reliability and be objective. One of the strengths of mixed methods was that there was no set design of how to use the quantitative and qualitative methods, but a variety of combinations, where one method could dominate the other.\(^{134}\) This was of benefit as the first research question of this thesis (do existing frameworks to assess metadata quality issues work in a different context?) is quantitative in nature and was answered through the implementation of the assessment method. The second research question (how does the implementation of the frameworks help to improve metadata quality for internal and external stakeholders?) in turn is more qualitative and was answered through the interviews, and the third one (What further adjustments, data governance policies and processes to the frameworks are needed to ensure metadata quality?) drew from the results of both these methods.

4.3.1 Quantitative Calculations

The calculations of the quality metrics expressed in section 4.2.1 did not only serve as a way to quantify qualitative values. They also served as a way to gather information about the quality of the metadata for the entire dataset, by calculating the range, average, median and standard deviation.

The range was used to see how far away the resource with the highest quality score differs from the one with the lowest. The average and median scores were calculated to be used as reference points if the data was skewed. The standard deviation was also calculated to see how the scores disperse around the average. Also, depending on where the average was located, using the standard deviation would help tell the urgency to improve the metadata quality. For

\(^{134}\) Pickard 2013, pp. 10–1, 18–23.
example, a low average and a wide standard deviation would have indicated that there were many flaws in the metadata quality.\footnote{Pickard 2013, pp. 284–93.}

Similarly, the identification from section 4.2.2 was calculated based on the results from the measurement calculations. Thus, it provided a summarised score of how good the individual, as well as the collective metadata, was. These two results were then analysed in an internal cross-tabular comparison, to show the general data but could provide a way to find links between different variables. Which was valuable for trying to identify where the metadata quality issue was and what to do against it. Meaning it served to find out the last step of the assessment method, solving, by analysing the results from measuring and identifying through the IQ dimensions from Table 4.3 to see what the potential solutions could be. But there was also an external cross-tabular comparison with the results from Reiche and Höfig.\footnote{Reiche & Höfig 2013.} The external cross-tabular analysis worked as a way to test reliability, by analysing if the different results were significantly similar.\footnote{Pickard 2013, pp. 22–3, 286–8.} It thus helped to show if the assessment method was indeed applicable in other contexts and thus indicated if the already existing frameworks are generalisable.

4.3.2 Qualitative Interviews

The choice of using interviews was partly due to the recommendation by Creswell and Poth to provide an in-depth understanding of the case that was needed for a qualitative study.\footnote{Creswell & Poth 2018, p. 98.} It was also chosen as it allowed a deeper investigation with the key informants and ‘social actors’ that had more knowledge and experience of the field, whom Pickard highlighted as valuable data sources.\footnote{Pickard 2013, pp. 104–6.} Indeed, the interviews were conducted with different key stakeholders within The Company at different positions, but who still had connections to the end users through their roles. The choice of interviewing internal stakeholders, and not the end-users, was due to deemed difficulty with reaching out to them. Instead, by using qualitative interviews with the internal key informants, the thesis could extract their judgment of what worked, what did not work and why they, through their roles, could represent the perspectives of the end-users.\footnote{Michael Quinn Patton, Qualitative research & evaluation methods: integrating theory and practice, Fourth edition ed. (Thousand Oaks, California: SAGE Publications, Inc, 2015).}
As the thesis tried to have answers for a specific topic it made sense to have semi-structured interviews with open-ended questions. This was beneficial for two reasons. First, the semi-structured interviews gave control over the research as all the interviewees had to answer the same questions. But as the questions were open-ended, it had the second benefit of allowing the interviewees to give more detailed and deeper answers, explaining if the metadata met the necessary quality for the end-users and why. This was key as the user perspective was the determining factor of metadata quality. For this reason, it was ill-suited to use a closed, fixed-response structure for the interviews. Because, even if all the interviewees would still answer the same questions, there would have been the risk of missing pieces of the end-users’ perspectives, as the answer options would be limited, and maybe even poorly formulated. Using an unstructured interview would also have been ill-suited, as these types of interviews aim to get a holistic understanding of a topic. It would therefore be difficult to get a clear, specific answer whether the metadata was of quality for the users or not, which this thesis needed to answer.\footnote{Pickard 2013, pp. 196–203.}

The questions asked were based on the quantitative calculations done in the first two steps of the assessment model, as well as the Information Continuum Model. By reading the model inwards, it allowed to use the cultural and sociocultural context as a reference point when analysing and judging the expected quality of the metadata. It was therefore used to make an a priori judgment if the metadata was of high quality for the customers before the metadata would be sent further down the pipeline.\footnote{Frings-Hessami, Sarker, Oliver & Misita 2020, pp. 142–8.} This of course needed to be verified, which was done by asking the interviewees if they agreed with the a priori analysis of the metadata quality.

After the interviews had been conducted they were transcribed to start the analysis. When transcribed, the process of data reduction began, which set out to discard redundancies and to find the parts of the responses that were of interest to this research. This process was also used to make the data more manageable and easier to organise. The data reduction began with a general read-through of the transcriptions, to get a general impression. Then the transcripts were reread while looking for similarities, differences, patterns and thematic connections which served in coding the data. The codes were based on the theoretical definition of quality, the Information Continuum Model and the frameworks in the assessment method. This would not only serve as a way to analyse the interviews and answer the research questions for this thesis but also to keep continuity and better connect the findings to previous literature. The analysis
thus consisted of bringing these codes into a web of meaning, by iteratively cross-checking the responses with each other, to serve as a foundation for the conclusions.\textsuperscript{143}

\textsuperscript{143} Halperin & Heath 2017, pp. 304–7.
5. The Results and Analysis

This chapter will explain how the research was conducted. As it was split up into two phases, one quantitative and one qualitative, there will be two main sections discussing each part separately with one closing section for discussion. The quantitative and qualitative sections also have a respective separate analyses.

5.1 The Quantitative Data

The analysed dataset was a sample of 101 randomised audiobooks with their metadata. Using a randomised dataset with each audiobook having the same probability of being selected allowed it to be representative of the whole population, although it is relatively small. Thus, it would allow generalising the findings of the data giving the results a stronger level of internal validity.\(^{144}\)

The samples were also to include the metadata fields identified as most important (title, author, description, language, speaker, genre, length, part of series and release date). Except for the field ‘part of series’ all other fields were included, along with a few other metadata fields. One of these other fields was ‘language prediction’, which was the result of a model to predict the spoken language in any audiobook. As The Company developed this model to serve as a substitute for when the field ‘language’ was empty, I decided to include this field in my research as well and give it the same weight as ‘language’.

These samples were analysed by using Microsoft Excel and the formulas used are explained in the Appendix. First, the calculation of measuring the metadata quality was conducted, as explained in section 4.2.1 above, and thereafter the quality level was identified, as explained in section 4.2.2.

5.1.1 Implementation

In measuring the quality of the metadata, the weights used for calculating the weighted completeness were based on the importance of the respective metadata category. In other words, as there were nine prioritised metadata categories given, the category of ‘title’ got a weight of 9, ‘author’ a weight of 8, ‘description’ 7 and so on until ‘release date’ which got a weight of 1 as it was the least important. Any further metadata category that might have been included in

the dataset got a weight of 0, as they were seen as irrelevant for end users (as it mostly was metadata for internal use) and thus this research.

For calculating the Richness of Information (RoI) for numerical and vocabulary values, only the genres were used. This was because it was seen as the only type of such metadata that could provide a clearer picture of what the content of the audiobook was. Due to the nature of the calculation, how unique the information is, it would not make sense to measure the RoI of names. To give an example, between a famous author appearing several times in the dataset and an unknown author appearing once, the calculation would value the unknown author higher than the famous one. Meaning the more famous the author is, the less information is extracted from the name. Thus excluding the author and speaker. It also wouldn’t make sense to calculate how often a certain length or how often a release date appeared and titles could be misleading (never judge a book based on its cover). Lastly, the category of language only explained if the listener would be able to understand the content or not. Therefore, the field of genres was seen as the only field that could have, or lack, an RoI in this context. Thus, the RoI for numerical and vocabulary values was used solely for the RoI from the genres.

Furthermore, as one of the requirements to reach tier 2 was that the vocabulary was controlled, a predefined list of genres to use served to control the vocabulary. In other words, the calculation was checking if the audiobooks had genres that fitted the predefined list. If they had one or more genres from the list, the audiobook then got the sum of the defined values per respective genre. Any genre not included in the list did not generate any value and thus no RoI.

For the calculation of RoI free text, the descriptions of the audiobooks were used. Due to the nature of the calculation, and due to the limits of Excel, three words or collocations were chosen to be used to measure the RoI. They were chosen manually, based on which words might give the clearest picture of what the audiobook was about to evaluate their RoI. After the RoI for the genres and descriptions were calculated, they were summed up, giving the total RoI for each audiobook, which was used for the analysis.

For Accuracy, external data from the publishing companies were used as reference points in the Comparing dataset. When there was no data about the audiobooks at the websites from the publishing companies, the websites Goodreads, Booklooker, Audiolibrix and/or Amazon were used in the following order. Goodreads and Booklooker are two platforms that are in large part driven by the readers. Audiolibrix on the other hand is a retailer of audiobooks. They were chosen as they are indeed specialising in audiobooks, and thus have most of the relevant metadata to make a comparison. Amazon was chosen as a last effort to find some data to
compare, due to Amazon’s vast product range. For these reasons, Goodreads and Amazon were both used, even though Goodreads is owned by Amazon.

When calculating the accuracy of the title and language, a Boolean function was used. If the title and language in the given dataset were the same as the reference dataset, then the function would give the value 0. If they were not the same then the function would give the value 1. This as the function by Reiche and Höfig is explained as measuring the distance of difference between the two datasets. Thus 0 indicated no difference between the metadata values and that they were the same, and 1 indicated that there was indeed a difference. The difference in duration and release date was calculated by subtracting the values from the two datasets from each other. Thus, if they were the same, the result would have been 0 and still indicating no difference. Accuracy for the genres and descriptions was calculated by subtracting the RoI values from the two datasets. To keep the difference in RoI clean, the calculation of the RoI in the comparing dataset used the same list of genres and keywords used for the RoI in the given dataset.

To identify the tier, or level, of the metadata quality, an IF-AND function was used in Excel. It evaluated the measurements from the calculations and ranked the audiobooks accordingly. In this case, for an audiobook to reach tier 3, the highest level, it had to meet the requirements for tier 2 and 1 as well. In this case, to reach tier 3, accuracy had to be lower than 0.2, total RoI higher than 0.03, weighted completeness higher than 0.86 and completeness higher than 0.9. The requirements were formulated after a general first impression of the metadata and the first results of calculating the metadata level for each metric. Meaning they were my formulations and they can be formulated differently in other contexts. Tier 2 only looked at RoI weighted completeness and completeness, but with the same values, and Tier 1 looked only at weighted completeness and completeness, also with the same values. If any audiobook wouldn’t reach tier 1, the text ‘Improve’ would appear, indicating that the audiobook acutely needed an improvement in its metadata.

5.1.2 Analysis of the Quantitative Data

Similar to Reiche and Höfig’s results, this thesis looked at the minimum and maximum values, as well as the average and the median for each metric. However, this thesis also included the standard deviation and how many audiobooks within each metric were indicated to be at a critical level. For completeness and weighted completeness, the critical level was set below 0.5
(50%). For all the RoI it was below 0.01 and Accuracy above 0.2. Thus the following results can be seen in the table below.

<table>
<thead>
<tr>
<th></th>
<th>Completeness</th>
<th>Weighted Completeness</th>
<th>RoI (Genres)</th>
<th>RoI (Description)</th>
<th>Total RoI</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>0.833</td>
<td>0.705</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Max</td>
<td>1.000</td>
<td>1.000</td>
<td>0.038</td>
<td>0.090</td>
<td>0.099</td>
<td>0.557</td>
</tr>
<tr>
<td>Average</td>
<td>0.897</td>
<td>0.825</td>
<td>0.009</td>
<td>0.022</td>
<td>0.031</td>
<td>0.194</td>
</tr>
<tr>
<td>Median</td>
<td>0.917</td>
<td>0.864</td>
<td>0.008</td>
<td>0.021</td>
<td>0.030</td>
<td>0.090</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.039</td>
<td>0.073</td>
<td>0.006</td>
<td>0.020</td>
<td>0.021</td>
<td>0.198</td>
</tr>
<tr>
<td>Critical %</td>
<td>0.00%</td>
<td>0.00%</td>
<td>69.31%</td>
<td>29.70%</td>
<td>19.80%</td>
<td>40.59%</td>
</tr>
</tbody>
</table>

**Table 5.1 Results of the metrics**

The formulations for the critical levels were my own. For future research and usage of this framework, the critical levels can be defined differently.

Calculating the number of critical cases can provide a good explanation of where to begin when improving the metadata quality in general. However, Table 5.1 shows only a summary of the calculations. It doesn’t tell the full quality of the metadata for the audiobooks. For that, Table 5.2 provided a better overview.

<table>
<thead>
<tr>
<th></th>
<th>Amount</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adequate level</td>
<td>72</td>
<td>71.29%</td>
</tr>
<tr>
<td>Tier 3</td>
<td>30</td>
<td>29.70%</td>
</tr>
<tr>
<td>Tier 2</td>
<td>21</td>
<td>20.79%</td>
</tr>
<tr>
<td>Tier 1</td>
<td>21</td>
<td>20.79%</td>
</tr>
<tr>
<td>Improve</td>
<td>29</td>
<td>28.71%</td>
</tr>
</tbody>
</table>

**Table 5.2. Tier categorisation**

The data had similarities to the results of the three quality assessments made by Reiche and Höfig.\(^{145}\) Both the maximum values for Completeness and Weighted Completeness had a maximum value of ‘1’ and a high minimum value, along with the average and median being high as well and close to each other. The RoI was low, just like the results from Reiche and Höfig, but lower than the median and average from Reiche and Höfig’s calculations. Their RoI calculations were in the range of 0.2 to 0.4 and the lowest maximum was calculated to be 0.69.

\(^{145}\) Reiche & Höfig 2013, p. 241.
This can be due to their calculations for free text might have calculated how often every word in the description occurred, instead of a few selected keywords as in this calculation. However, this difference might not be of too much importance as Reiche and Höfig noted that the record with the highest score (of 0.82.) was a description consisting only of two words. They saw this as critical for even if the description was easy to comprehend, it probably was not meaningful due to its shortness and therefore not of quality.¹⁴⁶

The Accuracy was where the results differ the most. Although Reiche and Höfig also received ‘0’ as their minimum value for the three datasets they calculated, the maximum for all of them was also ‘1’. The average and median for all three datasets were also all around 0.5, meaning that the accuracy varied quite equally from 0, fully accurate, to 1, not accurate at all. Meaning that even though this research used external sources for measuring the Accuracy it got better results when the reference record was available. Although, using external reference data might explain why the median and average in this research are so different from each other, along with the high standard deviation.

This data along with the results from the tier categorisation indicated that there was potential to increase the metadata quality. Even if more than 70% of the audiobooks were deemed to have an adequate level of metadata quality, almost 60% of them were in Tier 1 or 2, meaning that they could be improved. Focusing on the genres would have been a good starting point, as this was the RoI calculation with the most audiobooks deemed at a critical level. The high amount of audiobooks having a critical level in the RoI Genre can be that out of the 101 samples, 66 of them only had one genre assigned to them, and 14 had none.

Another interesting realisation that emerged when comparing Tables 5.1 and 5.2 with each other was the amount counted as critical in the RoI Description and the amount identified as needing improvement. These were 30 and 29 respectively. Indeed, all audiobooks that were identified as needing improvement had the value ‘0’ for their RoI Description, meaning that they did not have any description at all. (The 30th audiobook with a critical level of RoI Description had a value of 0.00828, which probably was due to the description length of 322 words.) Thus, empty description fields seemed to be a solid indicator to find audiobooks with generally low metadata quality.

To get a clearer picture of where Accuracy could be improved, the instances where the given and compared datasets were completely identical (giving the value ‘0’) for each metadata field were calculated. The results are shown below in Table 5.3. Although not a perfect

---

summarisation, as some accuracy values were Boolean and others were subtractions, it showed where improvements could be made. For example, the low amount for Speaker and Language was not due to contradicting data, but rather that in the given dataset, these fields were mostly empty. Similarly, in the given dataset, there were 32 empty fields for the Author. Meaning that the accuracy could be improved ‘simply’ by adding the missing metadata to the dataset.

<table>
<thead>
<tr>
<th>Metadata field</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>75</td>
</tr>
<tr>
<td>Autor</td>
<td>55</td>
</tr>
<tr>
<td>Description</td>
<td>21</td>
</tr>
<tr>
<td>Language</td>
<td>17</td>
</tr>
<tr>
<td>Language Prediction</td>
<td>87</td>
</tr>
<tr>
<td>Speaker</td>
<td>19</td>
</tr>
<tr>
<td>Genre</td>
<td>13</td>
</tr>
<tr>
<td>Length</td>
<td>51</td>
</tr>
<tr>
<td>Release Date</td>
<td>38</td>
</tr>
</tbody>
</table>

*Table 5.3 Amount of accurate metadata fields (having the value ‘0’)*

With that being said, it is important to mention that the given dataset did include data for the Author and Speaker. However, in the given dataset, they were referred to as an artist or author ID, which was a numerical value. These values had to be converted into their real names before a comparison could be done. However, not every name could be converted correctly, meaning that the amount for accurate Author and Speaker should be seen as a minimum, rather than as an end value. In other words, the level of Accuracy can be increased by improving or ‘cleaning’ the metadata for comparison.

In the field of Language, Accuracy could be improved very quickly. The calculation of Language Prediction in this case was done by comparing the predicted language in the given dataset with the language described in the comparing dataset. This was done as no Language Prediction could be found on any other website, thus the best reference point was the described language. The Language prediction proved very accurate when comparing it with the described language in the comparing dataset. Indeed, the reason why Language Prediction did not have the value of 101 can be due to the 14 cases when comparing audiobooks had empty Language fields. Thus, the Language field in the given dataset could use the Language Prediction fields, as they were accurate and true.
Moving on, the differences in Title between the given and the comparing dataset were mostly due to whether the title included the name of the series it was part of, and then how it was structured. For example, in the given dataset the Title could read [name of the episode] - [name of the series] whereas in the comparing dataset, it could be the other way around: [name of the series] – [name of the episode]. Therefore, the differences were mostly due to structure, not that there were two different titles.

For Length, 51 audiobooks had an exact match between the two datasets. Another 42 audiobooks had a difference of only one to five minutes between the two datasets. Similarly, even if the Description and Genres had very low exact matches, many audiobooks had an Accuracy level equal to or lower than 0.02. There were 55 audiobooks with such a level of Accuracy for the Description and 94 for the Genres. Meaning that even if most audiobooks were not 100% accurate in these three categories, they still could be seen as very accurate.

Last but not least the Release Date, where the difference counted in days could be very off. On one hand, there were indeed 38 audiobooks where the Release Dates were fully accurate. On the other hand, 23 audiobooks were off with more than one year, with the highest one with 115 years. However, that meant that 40 audiobooks were off with a maximum of one year.

So what did this mean for solving the issues? Using the IQ dimensions from Stvilia et al., the potential problems and suitable solutions were evaluated. As the Accuracy was calculated by comparing external data as a reference source, there might have been some issues with the relational/contextual Accuracy. Because the measuring of Accuracy in this case was rather ‘how well does the given metadata correspond with external metadata?’ instead of ‘how well does the given metadata reflect the content of the resource?’ Of course, the metadata was mostly extracted directly from the publishers’ websites. But it was not always that the publishers had all the relevant metadata for this study. In a few instances, they didn’t even display the audiobook on their website. In those cases, metadata for comparison was extracted from the websites Goodreads and Booklooker. When even these websites couldn’t provide any information, the last resorts were Audiolibrix and Amazon, who, similar to The Company, rather relied on the metadata they got from the audiobook publishers than creating their own metadata. Meaning although not identified as a problem, there might be uncertainty if the representation is indeed corresponding with the resources, which, as Stvilia et al. would suggest, would call for checking the stability of the resource, and its provenance.

---

Further on, the level of RoI suggested that the IQ dimension of intrinsic Informativeness/Redundancy can be improved. But since it was to a large part accurate with the external metadata, it must be at least informative enough for the end-users, although not new. There were also issues with the Relational/Contextual Precision/Completeness. One issue was the cases where the Weighted Completeness was deemed too low, and the tier evaluation labelled the audiobook with ‘Improve’. The other issue was that most of the Audiobooks had at least one metadata field empty and thus were not 100% complete. The intrinsic Precision/Completeness could also be improved further by applying a full-fledged ontology to, for example, the description.

By using the IQ dimensions, the analysis indicated that there were some mapping-related IQ problems, primarily as most audiobooks had incomplete metadata (mainly from missing Language), but also that the RoI could rise and be less redundant and inconsistent due to conversion errors. To solve these issues Stvilia et al. would suggest investigating how well the metadata represents the audiobook and/or looking at the provenance, mediation and upkeep of the metadata.\footnote{Stvilia, Gasser, Twidale & Smith 2007, pp. 1723–4.}

5.2 The Qualitative Interviews

To get feedback if the assessment framework had any value in giving clarity about the quality issues, six interviews with internal stakeholders in The Company were conducted. The stakeholders were introduced to me by The Company after I explained to them what kind of interviews I wanted to conduct, what questions I wanted to ask and what results I was looking for. To keep The Company anonymous, the interviewees were also anonymised to their roles they had in The Company. The following table summarises their roles and in which department they worked.
<table>
<thead>
<tr>
<th>Department</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Strategy</td>
<td>Manager of Data Governance</td>
</tr>
<tr>
<td>Team Audiobooks</td>
<td>Product Owner</td>
</tr>
<tr>
<td>Team Quality Engineering</td>
<td>Data Product Manager</td>
</tr>
<tr>
<td>Team Metadata</td>
<td>Data Product Manager</td>
</tr>
<tr>
<td>Team Metadata</td>
<td>Data Scientist</td>
</tr>
<tr>
<td>Team Recommendation</td>
<td>Data Scientist</td>
</tr>
</tbody>
</table>

*Table 5.3 The interviewees*

The interviewees will thus be referred in the text to the roles from the department they worked in, e.g. “the Product Owner from Team Audiobooks stated…”, to maintain clarity of who said what. To further strengthen their anonymity, they will also be addressed using the pronouns they/them. Their answers are presented in the following three sub-sections, which are divided into the nature of the questions that were asked.

5.2.1 Familiarity with the Customers/End Users

This section looks at what the interviewees answered on how well they understood the end-users’ needs and expectations of metadata quality and how they could incorporate these needs and expectations in their work when solving the metadata quality issues. This was primarily answered through questions 2-7, which are found in the Appendix under ‘Interview Questions’. The questions were based on the theoretical definition of quality, explained in section 3.1, as well as the societal embeddedness from the Information Continuum Model.

The level of familiarity with the end users varied across the different roles. But they all had some familiarity with their expectations and needs for the metadata. For example, the Manager of Data Governance from Data Strategy had a rather general and business-focused perspective. This was due to the role being a rather internal facing one, thus getting knowledge about end-users through the teams they were serving with how to manage the data. Similarly, the Data Product Manager from Team Quality Engineering identified their position as possibly the furthest away from the end users. However, when working with data that would reach the end users, the Data Product Manager from Team Quality Engineering still had to evaluate all the data from the end-users’ perspective, thinking about what they would feel with the data at hand. One way to get this perspective was by using the streaming platform, in other words being an end user oneself. This was a similar case with the Product Owner from Team Audiobooks, who
also used customer interviews, questionnaires, research and looking at reviews for the streaming platform, to get valuable insight into the end users' needs and expectations. The Data Product Manager and Data Scientist from Team Metadata got insight from the end-users partly by working closely with other teams further down the pipeline who provided them with feedback and complaints from the end users. Listening to, finding out and solving metadata issues was a high priority for Team Metadata, the data scientist said. Another method, the Data Product Manager from Team Metadata mentioned was together with the other teams to set up the requirements for the metadata, make use cases and track the live performance on the streaming platform. The Data Scientist from Team Recommendation used anonymised usage data to help provide recommendations for what content the end users would be interested to consume.

Through their roles and positions, the interviewees could identify different needs and expectations of the metadata for the end-users. The Product Owner for Team Audiobook, the Data Scientist from Team Metadata and the Data Product Manager from Team Metadata said that the expectations and the most important aspects of the metadata are dependent on the context. For example, a Harry Potter fan will have higher expectations and react differently if a book from the series was wrongly described, than someone just waiting for the audiobook to be published. But in general, the metadata needed to provide enough information about the content of the audiobook, just at one glance. Then one could look at metadata such as title, part of series, author and narrator. In the case where audiobooks were parts of series, the Product Owner of Team Audiobook stressed the need for consistency. The Data Scientist from Team Metadata identified the need to have complete sets of data, mentioning instances of the title being cut off after a certain character length in the metadata they had received metadata from companies further up the pipeline. For this reason, the Data Product Manager from Team Metadata stressed the need for metadata integrity and timeliness. Having a Richness of information was seen as important by the Data Scientist from Team Recommendation, as metadata with rich information helps make the recommendation system better and more accurate. The Data Product Manager from Team Quality Engineering also saw Richness of information as important as it enhances the positive experience of the streaming platform. But more importantly, they argued, was to have Accuracy, as that would be the most basic need, and then completeness before Richness of information.
5.2.2 Opinion about the Assessment Method

This section focuses on the interviewees’ answers to questions 8 and 9, to find out what they thought about the assessment method. These questions were based on the results of the quantitative analysis. Thus, the aim of questions 8 and 9 were to ask the interviewees if the assessment method provided any new and valuable information about the metadata quality and its issues.

All the interviewees found the assessment method to be of value in varying degrees in its ability to bring clarity to what the metadata quality issues were. One clear example of this was the response from the Data Product Manager from Team Metadata who, from their perspective, saw the method as giving a clearer understanding of the quality issues, but recognised that other stakeholders might not, due to unfamiliarity with the data and metrics. Such was the case for the Manager of Data Governance from Data Strategy, who although didn’t fully know the data, was very much interested in how to use the new information can be used to help solve problems. They could see the potential to generalise the assessment method, with the only question of how it can be applied and adapted for other datasets. Similarly, the Product Owner from Team Audiobooks found some of the information from the frameworks good but was unsure of how to act on it. They wanted more information about what each metric meant, the tier scale and how everything was calculated. In other words, the information could be more precise through a description of the processes. This would be valuable, for example, to see if a field was indeed containing the correct information and not just any information. But even though the Product Owner from Team Audiobooks wanted more details about the measuring and identification of the metadata quality, they could see the data at hand being useful and relevant for other teams and people.

One of these people was the Data Scientist from Team Metadata, who thought the assessment method was a really helpful tool to directly see what the problems were. Using the tier scale, one could easily filter out certain audiobooks and locate issues at different quality levels. They also saw how frameworks could be used to test the results from different data models and see why they were not working or not as expected. The Data Product Manager from Team Quality Engineering also thought the framework provided new information and a clearer picture of what and where the issues were. Another person was the Data Scientist from Team Recommendation, who thought the framework could be used as a template to identify which rows might have an issue, what the issue was and if it really was an issue. This would have helped get an understanding of what raw data actually meant, and not reach false conclusions.
based on misleading data. For this reason, they would have liked to have more metadata fields analysed in the method.

5.2.3 How to Move Forward

Although the previous section was to find out what the interviewees thought about the assessment method, it was also important to find out what they thought would be the following steps afterwards. For that reason, the interviewees were asked questions 10-12. These questions were based on critical theory, meaning the aim of these questions was to uncover what worked well with the assessment method, what needed to be improved and why. The questions would also answer what the interviewees thought were the best ways to solve the metadata quality issues.

What became clear from the interviews was that even if the assessment method did provide information and clarity about what the metadata quality issues were, more information and clarity would have been appreciated. For example, the Data Product Manager from Team Metadata wanted more justification for why the metrics used in this research were chosen above other metrics. But mostly it was requested that the framework would have had more metrics to measure the metadata quality. The Manager of Data Governance from Data strategy and the Data Scientist from Team Recommendation both wanted metrics to check consistency to prevent duplicates in the data. The Manager of Data Governance from Data Strategy also wanted to have metrics to check the timeliness of the metadata. Both these two persons, as well as the Data Product Manager from Team Quality Engineering, mentioned that the framework could have been improved if relevant stakeholders would have participated in giving insights and business inputs when the metrics were decided and formulated. For example, configuring the weights to fit the context even better was something the Data Scientist from Team Metadata was very interested in, as what one looks for in the metadata might vary if the content is for children, adults or a general audience. Nonetheless, all four of them found the assessment method useful, where the Manager of Data Governance from Data Strategy could see it being used to generate a backlog on how to allocate, align and prioritise resources to solve the metadata quality issues. This idea was shared by the Product Owner from Team Audiobook and the Data Product Manager from Team Quality Engineering. The data from the framework would also be useful to see how things were being improved but also changed as new audiobooks would arrive and change the catalogue.
To solve the issues, the Data Scientist from Team Metadata argued it should be done automatically, due to the enormous amount of data they were working with. Everyone else was rather for a semi-manual solution process. According to the Manager of Data Governance from Data Strategy, the work to solve the issues would probably be manual, as no clear solution emerged from the data. But they could see the potential possibility to automate the solving process but for that one would need to know the data better and also know its provenance. The Product Owner from Team Audiobook could also see the possibility in the future to automate the problem-solving but didn’t think that the technology was ready just yet. The Data Product Manager from Team Quality Engineering thought that some parts of the problem-solving could be automated, but also perform qualitative quality checks done by humans, to ensure that the metadata is indeed correct and makes sense to a human. The idea of quality checks was shared by the Data Scientist from Team Metadata and the Product Owner from Team Audiobook. The latter also included that the results from the quality checks should be presented and shared in workshops with relevant stakeholders, which was one idea of how to keep finding solutions to future problems. Using workshops could also be a way to find solutions for how to improve metadata protected by copyrights and other legal factors, an issue identified both by the Product Owner from Team Audiobook and the Data Scientist from Team Recommendation.

Translating what the interviewees said about the issues to the IQ dimensions, the issue with timeliness identified by the Manager of Data Governance from Data strategy can be seen as an issue with the intrinsic Currency, the age of the metadata. The problem with duplicates is an intrinsic Informativeness/Redundancy, and the changes in format or schema intrinsic Naturalness. The solution given by the Manager of Data Governance is similar to the general suggestions from Stvilia et al. on how to solve these problems: one should look into the provenance of the metadata of how it came to be where it is now. The issues identified by the Product Owner from Team Audiobook and Data Product Manager from Team Quality Engineering were intrinsic Informativeness/Redundancy and relational/contextual Accuracy, as end-users needed to know the content of the audiobook at one glance, by making sense of the metadata. The issue of metadata being cut, identified by the Data Scientist from Team Metadata, was a relational/contextual precision/completeness issue. For these issues, looking into the mapping would have been a good starting point for solving the problems.
5.3 Discussion

The quantitative analysis showed that the framework from Reiche and Höfig\textsuperscript{150} could be applied in a new context and to new data. To a large extent, the results from this thesis were similar to the three analyses Reiche and Höfig conducted on governmental public data. Therefore, it is an indication that their framework is generalisable. The quantitative analysis also showed that the combined frameworks from Bruce and Hillmann\textsuperscript{151} and Margaritopoulos \textit{et al.}\textsuperscript{152} could be used to identify the metadata quality level. The IQ framework from Stvilia \textit{et al.}\textsuperscript{153} could also be applied to the metadata to distinguish the IQ problems and their potential solutions. Thus, the four different frameworks have shown a potential to be generalised to other datasets and contexts to measure, identify and solve metadata quality issues. Be it as own individual frameworks or combined as in this thesis.

The qualitative analysis showed that even if the interviewees had different roles in The Company and distance from the end-users, in some way or another other they could incorporate the end-users’ values and expectations of the metadata. Seeing it through the Information Continuum Model, the interviewees at The Company were working first inwards and then outwards. This was demonstrated as they were adhering to the expectations of the end-users as well as paying attention to the legal boundaries in their work. The interviewees were, directly or indirectly, looking at the societal embeddedness (the fourth dimension) where they as employees and as a company were working in. This of course affected how they organised (the third dimension) and used metadata depending on how they were going to work with the metadata and type of audiobook. When they later got (captured, the second dimension) new metadata from the other companies up the pipeline, they had to structure, organise and prepare the new data accordingly before it was pluralised and shared with the public to use. From there the cycle started again, between the stages of pluralisation (how information was taken beyond the organisation’s context into forms of societal totalities) and capturing. But as the starting point was at the pluralisation and then going inwards on the Information Continuum Model, they ensured that the metadata would be of quality for the end-users.

Because working in this way, of paying attention to the societal embeddedness, they could satisfy some of the important requirements to meet the definition of quality. First of all, it gave them an understanding of the contemporary needs of the end users for the metadata, which was

\textsuperscript{150} Reiche & Höfig 2013.
\textsuperscript{151} Bruce & Hillmann 2004.
\textsuperscript{152} Margaritopoulos, Margaritopoulos, Mavridis & Manitsaris 2008.
\textsuperscript{153} Stvilia, Gasser, Twidale & Smith 2007.
stressed by Kużma and Mościcka. But it also showed that they were following the important aspect pointed out by Bruce and Hillmann, that they were paying attention to what the end users saw as quality through the information provision. Meaning that the assessment method was not only useful to help establish metadata quality for the internal stakeholders, but also the external ones, the end-users.

Further on, it became clear that cooperation between teams and stakeholders is needed to bring the assessment method to its full potential, by determining what aspects of the metadata are the most important and relevant. Depending on the role and position, the knowledge and usage of the metadata could vary. This became clear as more IQ problems could be identified from the interviewees’ responses, than from the quantitative analysis of the assessment framework. Therefore, one important condition for using the assessment method to its fullest potential is knowledge and understanding of the metadata and the processes of managing the metadata. This condition resonates with what Eastwood stated, information management and archiving need to be guided by theory to set methods and practices. This will prevent the metadata to lose its informatic value, becoming corrupted or being falsely interpreted. It also resonates with the conclusion from Kharkin and Lundell that working with metadata demands competence and looking forward to the future. In other words, just applying the assessment method gives limited value. To extract the full value one needs to understand how the metadata and repository work, be it audiobook metadata in a digital library or a record in an archive.

---

154 Kużma & Mościcka 2020, p. 23.
157 Kharkin & Lundell 2019, p. 35.
6. Conclusion

It is said that information is the new gold. But not everything that glitters is gold. For that reason, the importance and usability of metadata are widely recognised. Apart from making resources and records discoverable, enabling differentiating gold from fool’s gold, metadata provides contextualisation, trustworthiness, effectiveness, accessiblility and meaningfulness to the resource or record it describes. Be it for complex geographical data used by the military\textsuperscript{158} or an open language community.\textsuperscript{159} Scholars like Samuelsson have thus realised the importance of high-quality metadata not only for separate resources or records but also for the entirety of the repository or archive. However, he also recognised the difficulty in creating and managing metadata,\textsuperscript{160} another realisation shared by other scholars. Three of these scholars are Tani, Candela and Castelli who recognised three general issues that needed to be dealt with for future frameworks. These were 1) the need to develop an open and comprehensive framework that allows people from different fields to define their data quality concepts, 2) to develop a machine-processable way to effectively indicate genuine quality aspects, and 3) the development of a tool that can be used to augment quality-oriented aspects of the data.\textsuperscript{161}

The purpose of this thesis was to meet the above-mentioned issues by furthering the research in finding an effective and easy way to assess metadata by measuring, identifying and solving metadata quality issues. The aim was for the results to contribute to providing a generalisable method to manage metadata quality issues. Thereby it would solve the identified archive and information science problem of how to get control over the metadata in the digital era and be beneficial to archivists maintaining digital archives and managers of digital repositories. Thus, to serve the purpose and meet the aim, the following research questions were formulated:

1) Do existing frameworks to assess metadata quality issues work in a different context?
2) How does the implementation of the frameworks help improve metadata quality for internal and external stakeholders?
3) What further adjustments, data governance policies and processes to the frameworks are needed to ensure metadata quality?

\textsuperscript{158} Blomgren & Tuoremaa 2004.
\textsuperscript{159} Hughes 2004.
\textsuperscript{160} Samuelsson 2015, pp. 45, 55.
\textsuperscript{161} Tani, Candela & Castelli 2013, p. 1203.
The quantitative results showed that the quantifying framework by Reiche and Höfig can be applied in another context. The results were also positive for combining the classification frameworks from Bruce and Hillmann and Margaritopoulos et al. The combination could indeed provide a way to identify the scope of metadata quality issues. Similarly, it was possible to analyse the quantitative data using the IQ dimensions from Stvilia et al. Thus, the first research question has been answered, with the frameworks applying to a different context, although some adjustments and adaptations were needed. This of course was expected from the theory, as nothing should be taken for granted and one must understand the contemporary needs.

A further sign of the frameworks being applicable in new contexts was also how the measuring and identification could be done using Microsoft Excel. Of course, one could use more advanced tools, like statistical programs, to measure and identify the metadata quality. But as it was possible to use Microsoft Excel to do the analysis, the methods should be easier to apply in other contexts and situations, as the technological tool to assess the given metadata is very much standardised.

The second question was answered by the interviewees. The assessment method did provide a way to spot where and what potential metadata quality issues could be. This could be used in helping prioritise and allocate resources to solve the issues internally, thus making it easier for the internal stakeholders to improve the quality of the metadata for the end users. Be it in the content of the metadata or how metadata was applied to improve the user experience in finding audiobooks to consume.

The answer to the last research question did emerge during the preparation for the quantitative analysis but was also pointed out by the Manager of Data Governance from Data Strategy. Namely, this assessment method should not be conducted in isolation, but for the metrics to be effective and reflect real-world situations, it needed business input. This was similar to the points made by Eastwood, that disciplines might rely on knowledge from other disciplines to know the nature of how things act,162 and Iacovino that the archiving discipline and other disciplines should get a mutual understanding of each other’s methods.163

Therefore, there is a high probability that the frameworks are generalisable as they were successfully implemented on a new dataset in a new context. The feedback from the interviewees also confirmed the ability to generalise the frameworks. However, their feedback also clearly showed that the frameworks can be improved for them to reach their full potential.

---

One could therefore say that the search for a good assessment method for metadata quality is completed. Instead, the focus should be on how we can further improve the methods we have in order to reach their full potential.

Thus, when applying the assessment method correctly, archivists should be able to get control over the records’ metadata. It will be helpful to maintain the digital records’ authenticity, integrity and meaning over time, which was called by Foscarini. Applying the assessment method correctly would also mean using some sort of theoretical framework as a guideline for how to use it. Be it the Information Continuum Model, or as the interviewees did, relying on feedback from the end users to find the ways how the metadata should be managed. As long as the warning from Frendo is heard, that the assessment method should not be uncritically adapted, the method should be helpful when facing contemporary challenges and creating the “new science of the document” to keep the records relevant for the present and the future.

6.1 Limitations

Similar to the research from Reiche and Höfig and Stvilia et al. this research was conducted at a very small scale. In this case, it was a single-N case study and not a small-N case study as Reiche and Höfig and Stvilia et al. did. So although this thesis helped to verify the possibility to generalise the used frameworks, more case studies would be needed before one can truly say that they are generalisable.

6.2 Recommendations for Further Research

The framework from Stvilia et al. did show how to potentially solve the metadata quality problems. However, due to the time limit of this thesis, it was not fully used in the sense of actually executing the solutions. Further research could therefore be to apply their framework and do a comparison before and after to see how much the metadata quality has improved.

Another recommendation would be to find and try ways to quantify other metrics other than the ones given by Reiche and Höfig, as they were requested by the stakeholders. Apart from the metrics of timeliness and consistency identified in the interviews, there are also the

\footnote{Foscarini 2017, p. 126.}
\footnote{Frendo 2007, p. 165.}
\footnote{Frendo 2007, p. 158.}
metrics of provenance, conformance to expectations, logical consistency and coherence identified by Bruce and Hillmann.\textsuperscript{167}

What became clear from the interviews was that this thesis could have had other results if the metrics for the assessment method were defined together with the stakeholders. In this thesis, they were set by myself based on my own best judgement. If I had approached the stakeholders earlier, asking them what they see as important metrics, and introduced them earlier to the assessment method, they might have given other answers about what they missed from the data. Therefore, one further research area could be to apply the assessment method together with the relevant stakeholders and see how they experience the method from start to finish, and with continuous feedback.

One last area to look for further research would be to find some guidelines on how to set the weights in the best way possible. This as they are a given factor in the framework by Reiche and Höfig, and as that was something all the interviewees mentioned that it needed to be configured. But when one should apply the framework to new datasets, describing new resources or records in new contexts, it can be useful to have a rule of thumb on how the weights should be set.

Bibliography


Foscarini, Fiorella & Oliver, Gillian, "Introducing the information culture framework as a component of the digital curator’s toolkit", *Framing the Digital Curation Curriculum Conference (DigCurV 2013), Proceedings of the Framing the Digital Curation Curriculum Conference (DigCurV 2013)* (Florence: Rheinisch-Westfälische Technische Hochschule Aachen, 2013).


Ridener, John, From Polders to Postmodernity: A Concise History of Archival Theory. (Duluth, Minnesota: Litwin Books, 2009).


Appendix

Excel formulas

Completeness

Using the formula =COUNTA(A2:C2)/COUNTA(A$1:C$1) in cell 2D will give the percentage of completeness, for the title ‘Lord of the Rings’ and can easily be filled in cells below to find the level of completeness for other items

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Title</td>
<td>Genre</td>
<td>Publisher</td>
</tr>
<tr>
<td>2</td>
<td>Lord of the Rings</td>
<td>Fantasy</td>
<td>0,666666667</td>
</tr>
<tr>
<td>3</td>
<td>Harry Potter</td>
<td>0,666666667</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Star Wars</td>
<td>Sci-Fi</td>
<td>Audiobooks Ltd.</td>
</tr>
</tbody>
</table>

Weighted Completeness

Using the formula =COUNTA(A4)*G$3 in cell G4, excel sees if the metadata is given and multiplies it with its given weight. This formula can later be filled in the rest of the cells. To find the value for Weighted Completeness, the given values for each resource are summed up (for example in cell J4 using the formula =SUM(G4:I4), and then divided with the sum of the weights, found in cell J3 (using the =SUM(G3:I3). The formula used in cell D4, under Weighted Completeness, is, therefore, =J4/J$3, which can easily be filled in the cells below.

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Title</td>
<td>Genre</td>
<td>Publisher</td>
<td>Weighted Completeness</td>
<td>Metadata value</td>
<td>Metadata weight</td>
<td>Title</td>
<td>Genre</td>
<td>Publisher</td>
</tr>
<tr>
<td>4</td>
<td>Lord of the Rings</td>
<td>Fantasy</td>
<td>Audiobooks Ltd.</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Harry Potter</td>
<td>0,666666667</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Star Wars</td>
<td>Sci-Fi</td>
<td>Audiobooks Ltd.</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Richness of information (numerical and vocabulary)

To calculate the richness some separate steps are needed to be taken. The first step is to find out how many instances of each metadata value appear. Using genres as an example, the first formula used is =LEN(A5)-LEN(SUBSTITUTE(A5,””))+1 in cell 5B, to find out how many different genres are given for a resource. The formula is then filled in the cells below, giving the number of genres per resource. Then the total sum of used genres is calculated in cell 5C using =SUM(B:B) (note that the sum is furthest to the right in case more metadata values would be added later on).
The next step is to calculate the uniqueness of each genre. For that, we make a separate calculation for how often each genre appears in the dataset. Using \( =\text{SUM}((\text{LEN}(\text{A}:\text{A}) - \text{LEN}((\text{SUBSTITUTE} (\text{A}:\text{A}, \text{C}15, \text{""}))))/\text{LEN(C}15) \) Excel looks at how often the genre given in cell C15 appears in the list given above. The frequency, or probability, that the given genre appears in this dataset is calculated in cell E15 using the formula \( =\text{D}15/\text{C}5 \) (amount/total). Lastly, the uniqueness of the genre is given by simply giving \( =-\text{LOG(E}15) \). These three formulas are then easily filled in the cells below as well.

For step three you need to find the sum of the uniqueness of each resource. By inserting the formula \( =\text{IF(ISTRUE(SEARCH(F3,A5)),F4,0)} \) in cell F5, it will search if the given genre listed in cell F3 is present in cell A5. If the genre is present it will return the value of uniqueness available in cell F4. The formula is then filled in the cell range F5:F19. The values for each resource are then summed in column E using a simple \( =\text{SUM} \) formula.

The last step is to calculate the Richness of Information. This calculation is simply dividing the sum of the value of uniqueness by the total amount of metadata values (genres) that is used in the dataset. Therefore in cell K5, the formula is \( =\text{E}5/\text{E}3 \).
Richness of Information (free text)

This step is to calculate the value of a free text based on words or collocations located in the text. For that reason, there needs to be a way to insert the words or collocations one wants to evaluate. In this example, there are two input fields for the words or collocations per resource. For the resource description in cell C2, the input fields are D2 and E2. Note in the image below that there are two columns labelled “description” and that the words and collocations are all written in lowercase letters. This is because the formula used in column I to count the occurrence of a word in a free text =SUM(LEN()-LEN(SUBSTITUTE((),))/LEN()) is case sensitive. Thus, the words or collocations in columns D and E need to be lowercase, and the descriptions in column F are converted into lowercase using the formula =LOWER(). More concretely, the formula in I2 is =SUM(LEN(F2)-LEN(SUBSTITUTE(F2,D2,””))/LEN(D2), to find the occurrence of Word 1 in cell F2, which has the formula =LOWER(C2).

Column H calculates the number of words each description has. To do this, cell H2 has the formula =(LEN(F2)-LEN(SUBSTITUTE(F2,””))+1. Column J calculates how many different descriptions Word 1 appears in. Using the formula =SUMPRODUCT( --ISTEXT(SEARCHD2:F:F)) in cell J2 thus counts how often the word ‘story’ appears in all the given descriptions. In this example, ‘story’ appears in total in two descriptions.

The next step are the first to calculate the Richness of Information. As the formula is a TF-IDF (Term Frequency-Inverse Document Frequency) calculation, we first need to calculate the TF for Word 1 and then its IDF before we multiply the two values with each other. The TF is the occurrences of the word in the text divided by the total number of words in the text. Thus,
the formula in K2 is =I2/H2. IDF is the logarithm of the total number of texts divided by the number of texts containing the word. This gives the formula =LOG(A$2/J2) in cell L2. Cell M2 has the formula =K2*L2 which is the value of TF-IDF.

The formulas can then be reused to calculate the TF-IDF for Word 2, as seen in the image below. It can also be reused further if more words or collocations are desired. In the end, the sum of all the individual TF-IDF needs to be calculated. In cell S2 this is done simply by using =M2+R2.

The last step required to obtain the Richness of Information is to divide the sum of all the TF-IDF by the number of different words or collocations used. In this case, two different words or collocations were used per description. Thus, the formula in cell G2, seen below, is =S2/2.

Accuracy

The calculation for Accuracy should be done last. This is due to it partly uses the values from the Richness of Information (RoI) as a value in the calculation. For this calculation, two datasets are needed, as they will be compared with each other. One should be the ‘given’ dataset that you mainly are working with. The other one is the ‘compare’ dataset. What also is important is that the resources are lined up in the same order, as the top resource in the list ‘Given’ is compared with the top resource in the list ‘Compare’. The second top resource in ‘Given’ is compared with the second top resource from ‘Compare’ and so forth.
To keep the example simple, the lists ‘Given’ and ‘Compare’ are on the same Excel sheet (see image below), along with the calculation for accuracy. Reiche and Hösig explained in their article that depending on what the value is to be compared, a Boolean value can be useful. That is the case when the values for Title and Publisher have been compared. For example, in cell G5 the formula used is =IF(ISTEXT(SEARCH(A5,A12)),0,1). If the title in the list ‘Given’ is the same as the title in ‘Compare’ the formula gives the value 0 as in zero distance, meaning that they are the same. If they are not the same the value given is 1, to indicate that there is a distance between them. A similar Boolean calculation occurs in J5.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Given</td>
<td>Title</td>
<td>Description</td>
<td>Genre</td>
<td>Publisher</td>
<td>Duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Description</td>
<td>Genre</td>
<td>Publisher</td>
<td>Duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lord of the Rings</td>
<td>New</td>
<td>novel</td>
<td>Fantasy</td>
<td>Audiobooks Ltd.</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harry Potter</td>
<td>A</td>
<td>story</td>
<td>about</td>
<td>Fantasy</td>
<td>Audiobooks Ltd.</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Star Wars</td>
<td>Adventures</td>
<td>Sci-Fi</td>
<td>Sci-Fi</td>
<td>Audiobooks Ltd.</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Star Trek</td>
<td>#DIV/0!</td>
<td>Sci-Fi</td>
<td>Stories Inc.</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>Title</td>
<td>Description</td>
<td>Genre</td>
<td>Publisher</td>
<td>Duration</td>
<td>Euclidian</td>
<td>Sum</td>
<td>Accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Description</td>
<td>Genre</td>
<td>Publisher</td>
<td>Duration</td>
<td>Euclidian</td>
<td>Sum</td>
<td>Accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.000</td>
<td>0.000</td>
<td>0.281</td>
<td>1</td>
<td>15.332</td>
<td>26.838</td>
<td>16</td>
<td>0.694</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.000</td>
<td>0.000</td>
<td>0.281</td>
<td>1</td>
<td>20.012</td>
<td>30.012</td>
<td>16</td>
<td>0.686</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.000</td>
<td>0.000</td>
<td>0.281</td>
<td>1</td>
<td>25.032</td>
<td>35.032</td>
<td>16</td>
<td>0.678</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For calculating the distance between the Description and the Genre, the ROI is used for each value respectively. Simply, put it takes the already calculated Richness of Information from each dataset and subtracts them from each other, giving the difference. For example, cell I4 contains the value of the calculated ROI of the genre ‘Fantasy’ in cell C4 minus the calculated ROI of the genre ‘Drama’ in cell C11.

Do note that this requires that the calculations for ROI have to be done on the ‘Compare’ dataset as well. The calculation for the ROI for free text should also use the same keywords or collocations as as for the ‘Given’ dataset.

Similarly, if there is another comparison between numerical values, as in column E containing the duration of each audiobook, the difference is simply calculated using subtraction between the value from ‘Given’ and the value from ‘Compare’. For example, in cell K4 the calculation is E4-E11. Do note that this becomes 50 – 75 which equals -25. But to avoid having some positive and some negative values, in all cells where a subtraction occurs (in this case, range H3:H16 and K3:K6) the difference is given in absolute value, using the =ABS() formula. Thus the formula in cell K4 is =ABS(E4-E11).

The next step before being able to calculate the accuracy is to find the Euclidian distance between all these values. The Euclidian distance is found by taking the square root from the sum of each distance squared. Therefore, the formula in cell L3 is =SQRT(SUMSQ(G3:K3).
This is then divided by the sum of the differences, which can be found in M3 using the formula =SUM(G3:K3). However, as the formula is to calculate the difference, with 0 meaning no difference, the formula in N4 is =1-((L4+0.0001)/(M4+0.0001)), to avoid a situation of 0/0, which would have resulted in ‘error’, as would have happened in cell N5. Therefore, although not given in the original formula from Reiche and Höfig, the additional 0.0001 makes such a small difference to the addition but still makes it possible to use the formula under these given terms for calculating the accuracy.

**Evaluating the Tiers**

Since the idea is to identify the metadata quality based on its values and sorting it into a tier explaining the level of quality, the following formula has been formulated:

=IF(AND(S3<=0.2,R3>0.03,N3>=0.9,O3>=0.86),"3",IF(AND(R3>0.03,N3>=0.9,O3>=0.86),"2",IF(AND(N3>=0.9,O3>=0.86),"1","Improve"))). Meaning when Completeness, Weighted Completeness, Roll are high and Accuracy low, it will be sorted as a Tier 3 with the label ‘3’ quality. If only the Completeness, weighted Completeness and Roll are high then it is a Tier 2 with the label ‘2’, and if only Completeness and Weighted Completeness are high enough then it will be classed as a Tier 1 using the label ‘1’. If none of these tiers is met, the text “Improve” will appear to show which audiobooks critically need to be improved.

**Summarising the values**

To get a proper overview, the values need to be summarised. In this case, the summarisation shows the range of the values (the maximum and minimum), the average, median, standard deviation, how many cases are deemed critical and how many they are and their proportion. A second, separate summarisation explains how many audiobooks are in which tier, and how many need to be improved.
The calculations are simple, MIN(range), for the minimum value, MAX(range), for the maximum, AVERAGE(range), MEDIAN(range), and STDEV.P(range) to get the standard deviation for the population. To calculate the critical values, COUNTIF was used on the range with a given value, deemed to be seen as a critical border. For the two calculations for completeness, the value 0.5 (50%) is used in this example, making the formula in cell X7 be =COUNTIF(N2:N102,"<0.5"). For the ROI calculations, 0.01 was used thus the formula =COUNTIF(R2:R102,"<0.01") is used on cell AA7. The critical level for Accuracy was set at 0.2 giving the formula =COUNTIF(S2:S102,">0.2") making values above 0.2 as critical. The results are then divided with the total amount of audiobooks in row 8, by using the formula =Y7/COUNTA($A:$A).

The summarisation of the amount of each tier is using a simple COUNTIF function. Cell W11 looks in the range of how many audiobooks have been labelled with “3”, cell W12 labelled with “2” and W13 for “1”. Cell W14 sum all these up, to get a perspective on how many audiobooks have at least a sufficient level of metadata quality. This can also be compared with W15 showing how many are labelled and that they need to be improved.

**Interview Questions**

1. What is your role in the company, and in which department do you work?

2. In your own words, how close are you to the customers/users?

3. How do you incorporate customer/user values in your work?
4. How often do you need to take the customers'/users' perspective into your work?

5. How do you ensure that you have a customer/user perspective in your work?

6. Based on your experience, what are the most important aspects of the metadata to be of quality for the customers/users (e.g. completeness, accuracy or richness of information)?

7. What would you say are the customer/user needs and expectations of the metadata? What is important for them?

8. Are any of the following conclusions surprising to you, and do you find any of the conclusions critical? Why or why not?:
   - About 30% of the audiobooks are in critical need to have their metadata improved.
   - The data indicates there needs to be more focus on adding genres or making the genres more informative.
   - The data indicate that lack of description means general poor metadata quality.
   - Most audiobooks need metadata about the language spoken to be complete.
   - About 30% of the Audiobooks have an excellent level of metadata quality.
   - Except for the Accuracy, the median and average are very close to each other, and the standard deviation is relatively low.

9. Does the data give you a clear(er) picture of what the data quality issues are? Why, why not?

10. Do you feel this method gives a good idea of where and how to start solving the issues to produce metadata of high quality? If yes, please elaborate.

11. Do you feel that the solutions/improvements can be done manually or automatically? What should the human in the loop pay attention to?

12. Do you think something is missing in this method, some missing data or something that could have been clearer (e.g. the weights, other metadata values or other calculations)?