A DIGITAL DARK NOW?
Digital information loss at three archives in Sweden

Anna-Maria Underhill & Arrick Underhill

Master's thesis (30 credits) in Archival Science within the Master's Program of Archival Science, Library- and Information Science and Museum Studies at Lund University
Supervisor: Lars Björk
Year: 2016

Title
A digital dark now? Digital information loss at three archives in Sweden

Abstract
The purpose of this study is to examine digital information loss at three Swedish archives. Digital preservation is a complex issue and something most archival institutions struggle with today. While there is merit in focussing on successes in this struggle, doing so to the exclusion of failures runs the risk of creating a blind spot for existing problems. We hope that this study will contribute to an open and productive conversation about digital preservation and how it can be improved.

This study is based on interviews at three different kinds of archives, each of which have a different function in society. Limiting the study to a small number of archives allows for a more in-depth examination of how and why digital information loss occurs. The interviews concern the struggles dealt with by these archives and how they handle their digital information. We use a broad definition of digital information in order to look at parts as well as whole digital objects and their metadata. We also examine digital internal work documents, as these may serve as a contextual support for an archive's collections. The results are analyzed from several perspectives including the transition between the Records Lifecycle Model and the Records Continuum Model, an ontological understanding of digital information, the SPOT model for risk assessment and the OAIS Reference Model.

Some of the conclusions in this study echo previous research, such as the need to prioritize organizational issues. Others conclusions are more representative of the current state of digital preservation at these archives. One such conclusion is that there is a delicate balancing act involved between setting up systems for successful future digital preservation while managing existing digital collections which may not have been preserved correctly. Another conclusion we arrived at is that some institutions are unable to undertake a more proactive form of digital preservation, as the nature of the materials they preserve precludes such interference from archivists. We also discovered that when discussing digital preservation, the tendency remains to think of digitized material first rather than born digital information. Another finding of this study is that information loss may involve only a part of the information. Losses of metadata or the connections between information may be more common than the loss of entire digital objects. Finally, one question has followed this study from the beginning to the end: How can you know that you have lost something you never knew existed?

Keywords
Archive, information, digital preservation, digital archives, electronic archives, e-archive, future archives, electronic records, internet archive, long-term digital preservation, digital dark age, information loss, risk assessment

TABLE OF CONTENTS

1 Introduction 5
1.1 Motivation 5
1.2 Intent and research questions 7
1.3 Disposition 8
2 Background and terminology 10
2.1 Digital information 10
2.2 Digital preservation 13
2.3 Digital preservation strategies 15
2.4 Information loss and a digital dark age 16
2.5 Current projects within digital preservation in Sweden 18
2.6 Swedish legislation concerning public records 19
3 Previous research: Digital preservation issues 20
3.1 More than just technology 20
3.2 Limited funding 22
3.3 A changing profession 23
3.4 Digital public records 25
3.5 Risk assessment 26
3.6 Information loss 27
4 Theory 29
4.1 The Records Continuum Model 29
4.2 The ontology of digital phenomena 31
4.3 The SPOT model for risk assessment 33
4.4 The OAIS Reference Model 36
5 Method 40
5.1 Procedure and selection 40
5.2 Interviews 41
5.3 Results and analysis 43
5.4 Limitations 43
5.5 Ethical considerations 44
6 Results 45
6.1 Archive A 45
6.1.1 Digital holdings and information flow 45
6.1.2 Responsibilities for digital information 47
6.1.3 What is an e-archive? 48
6.1.4 An example of information loss: Issues with format migration 49
6.1.5 Strategies for digital preservation management 49
6.1.6 Cloud services, copyright and personal data integrity laws 50
6.2 Archive B 51
6.2.1 Digital holdings and information flow 51
6.2.2 Responsibilities for digital information 53
6.2.3 What is a record in the digital world? 54
6.2.4 Information loss: Analog vs digital 54
6.2.5 Strategies for digital preservation management 55
6.2.6 Cloud services 56
6.3 Archive C 56

6.3.1 Digital holdings and information flow 56
6.3.2 Responsibilities for digital information 58
6.3.3 What is digital preservation? 59
6.3.4 Information loss: Effects on the digital object 59
6.3.5 Strategies for digital preservation management 60
6.3.6 Intellectual property rights 61
7 Analysis 62
7.1 Archive A 62
7.1.1 The Records Continuum Model 62
7.1.2 The ontology of digital phenomena 63
7.1.3 The SPOT model for risk assessment 64
7.1.4 The OAIS Reference Model 65
7.2 Archive B 68
7.2.1 The Records Continuum Model 68
7.2.2 The ontology of digital phenomena 69
7.2.3 The SPOT model for risk assessment 69
7.2.4 The OAIS Reference Model 70
7.3 Archive C 72
7.3.1 The Records Continuum Model 72
7.3.2 The ontology of digital phenomena 73
7.3.3 The SPOT model for risk assessment 74
7.3.4 The OAIS Reference Model 75
7.4 Discussion 77
8 Concluding remarks 82
9 References 85
10 Division of labor 92

1 Introduction

1.1 Motivation

Within a short period of time digital technology and the internet have radically shifted the ways in which we live our lives. These changes have affected the nature and location of public discourse, what we work with, how we communicate, our experience of culture, how we consume the news and many other things. Records of our day to day existence are comprised largely of binary code on our smartphones, computers, and at servers in remote locations. Our photographs, correspondence, money, degrees, music, political discussions, applications, police reports and medical records often exist exclusively in digital form. The hectic evolution of this technology and its spread into every corner of our lives has made it apparent that the preservation of digital information is both immensely important and an incredible challenge.

Cultural heritage institutions such as libraries, archives and museums (LAMs) are institutions in society which are responsible for the preservation of information over long periods of time. As these institutions handle an ever increasing amount of digital material in preservation and workflows, LAMs have a relatively new set of preservation concerns to deal with. The cycle of technological obsolescence and the struggles it entails for these institutions are a cause for concern.

The documents preserved by archives possess a wide range of functions and originate from a diverse group of sources. These might range from state administration, which is the foundation for the principle of legal security and a functional democracy, to historical research and cultural collections. Due to the fact that archives preserve documents which are unique – and often fundamental to societal function – it is especially important to focus on these institutions in making sure that their digital preservation practices will provide the best possible chance for their digital holdings to survive well into the future.

Analog material which has been digitized, whether this has been performed in order to increase its accessibility or improve chances for preservation, is not free from the complications inherent to the preservation of born digital material. On the contrary, some digitization projects have proven to have lifespans considerably shorter than the analog material they were meant to be a copy of. One example of this is the 1980s Domesday Book digitization project. That project helped to bring the problem of digital preservation into general awareness by its failure in preservation terms. A particularly egregious example, this costly and highly publicized digitization project was unreadable already within two decades of its creation due to obsolete format and media (Corrado & Moulaison 2014, p. 8).

Many different projects and organizations are engaged with the challenges posed by digital preservation. Some of these are securely established, funded by the government, and regulated by law. Others led a more precarious existence taking up the responsibilities for cultural material which falls through the cracks. In regards to the governmental level, in 2011 the Swedish government stated that it was their goal to become the best in the world when it came to utilizing the possibilities of digitization. This goal was formulated in a agenda titled "IT in the service of humanity: A digital agenda for Sweden"[1] (Näringsdepartementet 2011). With this goal in mind, the governmental cultural sector has formulated a national strategy for the period 2012-2015, which they call "Digital cultural heritage"[2] (Kulturdepartementet 2011a). This strategy presents guidelines for how cultural institutions should work with questions of digitization and encourages each agency to formulate a plan for digitization and digital accessibility. The Swedish National Archives (Riksarkivet) have been the coordinator for this project, creating a coordination secretariat to oversee the work called Digisam. Among other things, the secretariat has organized working groups where agencies, institutions, and associations throughout the entire cultural sector can meet and discuss these challenges. Digisam has recently received an extended mandate (Kulturdepartementet 2011b; 2015).

While the issue of digital preservation strategies in Sweden is currently being taken up by the Digisam project, at least as regards governmental cultural institutions, Digisam is fundamentally oriented towards the future, and is not as concerned with digital information loss which may have already occurred. This study focus on potential and actual digital information loss which has taken place in the past or may take place in the future.

The looming threat of a future 'digital dark age' – the idea that much of our history may be permanently lost due to a failure to preserve digital information – is something many are aware of. The term was first mentioned by Terry Kuny (1997) in an article titled "A digital dark ages? Challenges in the of electronic prevention information". The term refers to a problem intrinsic to digital information: that aging formats, software, and hardware will make the information unreadable if there is a failure to constantly maintain it. This problem with the preservation of digital information has been described as a "civilizational issue" (Brand 1999, p. 46). The struggles of long-term digital preservation have recently entered into the public discourse via several blog posts and speeches by Vint Cerf, Vice President of Google (Anderson 2015, pp. 20-21). It should nonetheless be pointed out that the problems have been known for a much longer period of time (Quisbert 2008, p. 1). While the term 'digital dark age' may sound exaggerated, the risks and consequences it describes are not. This paper aims to contribute to an understanding of the circumstances and causes which are currently relevant to digital information loss in Sweden.

[1] Authors' translation, original title "IT i människans tjänst: En digital agenda för Sverige".
[2] Authors' translation, original title "Digit@lt kulturarv".

1.2 Intent and research questions

There are many different kinds of archives and the regulations governing them vary from country to country. We are interested in examining how digital information loss occurs at organizations in Sweden. For this study we have chosen three archives which differ from one another in the kind of materials they preserve and the rules which they are regulated by. The intent of this study is to aid in understanding digital information loss and the challenges it poses to these three archival institutions.

Research questions:

- What digital information has been lost or is at risk of being lost?
- Which conditions have led to the loss of this material, or its at-risk status?

In everyday language, the terms 'data' and 'information' are often used interchangeably. They can also be defined according to the degree they have been processed. From this perspective, data serve as the building blocks of information, they are uninterpreted and acontextual. Information, on the other hand means interpreted, contextualized data. Both of these can be communicated. Information becomes meaningful together with knowledge (The Consultative Committee for Space Data Systems (CCSDS) 2012, ch. 2, pp. 3-5). When we write about information loss in this paper, we do not exclusively mean information. We include data loss since it may be the case that only a part of the information missing.

There are many kinds of digital information loss, not all of which will be taken up in this paper. To clarify the concept, some of the ways digital information loss can occur include:

- Loss of information due to obsolete formats, software or hardware
- Loss of information due to physical damage to digital storage media
- Loss of information due to failure to backup
- Loss of information due to improper ingest procedures
- Loss of information due to sabotage
- Loss of data due to repetitive copying
- Loss of data due to lossy data compression
- Loss of data due to damage of bit stream in digital media (bit rot)
- Loss of metadata or improperly entered metadata

(Gladney 2007, p. 10; Rosenthal 2011, not paginated; Runardotter, Quisbert, Nilsson, Hägerfors & Mirijamdotter 2006, p. 26).

The term 'loss', in this context, does not necessarily mean that the information no longer exists. It may be forgotten in the archives or stored on obsolete storage media, but due to budgetary constraints or other considerations, recovering access would take an inordinate amount of effort. Therefore the term 'loss' can be considered to encompass things which are effectively unsearchable, unreadable, or just plain missing. This aspect of the term 'loss' is crucial when making the distinction between preservation and storage. While storage only requires that the information is on physical storage media, preservation entails that it is also possible to read, access and

understand the information. Stored material may in some instances be considered "lost" from a digital preservation perspective. An example of this would be digital photographs stored in an obsolete format. The data is still present, but there may be no software available to open it. This means that aspects of digital information can be lost in a way that is fundamentally different from analog material.

This study is focused on the unintended loss of digital information. Culling, as defined by the Swedish National Archives, is an intentional action (Riksarkivet 1999, p. 5). This paper does not look at the culling of records. What exactly constitutes a record is a complex issue. We have decided to adopt a broad understanding of records:

> Archives are thus made up of material (documents, photos, audio and video recordings, etc.) that constitutes cultural heritage, organizational memory, evidence, etc. and is often referred to as records.
>
> *Mari Runardotter 2007, pp. 10-11*

We have chosen a holistic perspective of archives' digital preservation in this study. This means that we look at digital information in the archives' collections together with work documents and other information in the environment, since these may contain important contextual information for understanding material in the collections. In those cases where this is within the public sector, the records are sometimes a matter of "public record", which in a Swedish context has a specific definition which is given in section 2.6.

This study hopes to provide an opportunity for cultural heritage institutions to gain a better understanding of their own digital preservation efforts. By bringing information loss into focus, the importance of sufficient support for digital preservation can be more effectively emphasized. This study aims to contribute to raising awareness of the issues at stake to decision makers. Decisions about digital preservation must be made grounded in facts and experience rather than relying solely on hypothetical situations. As archives are a fundamental part of the nation's historical, intellectual, administrative, and democratic health, loss of digitally preserved archival information may be seen as endangering that health. Lastly, this research may also be of interest for historians studying the growing ubiquity of the digital world in everyday life, how the increasing dominance of digital information affects historical research, and the effects of these issues have on archives and other cultural heritage institutions.

1.3 Disposition

The thesis is structured as follows: first a background is provided in order to give the reader a more in-depth understanding for the problems of digital preservation. Next there will be a brief review of some of the previous research concerning this topic in order to locate the study within its proper context. After this, two sections will follow explaining the study's theoretical perspectives and method. These will be followed by a presentation of the study's results, which are taken from interviews, and their analysis from the selected interpretive frameworks. The analysis section will

conclude with a discussion on a more general level, which also contextualizes the study within previous research. The thesis concludes with a summary and a discussion of the concluding remarks.

2 Background and terminology

This section will provide a more in-depth description of the problems of digital preservation and why information loss occurs. It will also provide a short presentation of the work being done with digital preservation in Sweden, and to a more limited degree, internationally. The study's key terminology is presented and defined here.

2.1 Digital information

Within digital preservation, it is common to refer to archives which preserve digital material as "electronic archives", which contain "electronic records". The line between analog and digital material can be somewhat unclear when discussing electronic records. Electronic media can be analog or digital: while old compact cassette tapes are electronic, they are not digital. There are of course plenty of magnetic tape mediums which *are* digital. So, to avoid the confusion, we use the term digital rather than electronic unless something within digital preservation is specifically termed "electronic", such as an electronic archive (e-archive), or electronic services (e-services).

Digital refers to information represented by discrete symbols rather than the continuous representations of analog. As a further clarification: when we talk about digital data or information here we are talking about data or information composed of binary code. Therefore, when we talk about digital information, data, or media in this paper we are talking about electronic data written in binary code (Wikipedia 2016a). As another example, the information on a VHS is not digital since it is not written in binary code. It is instead non-binary, analog information encoded by varying the magnetic tension on a strip. A CD, on the other hand, is digital since the information is composed of binary code which is read optically. Binary code is a system of numbers based on ones and zeros. (InterPARES 2 Project 2008, p. 9).

The definition of data and information vary depending on the origin of the definition. One way of framing data and information is according to the Data, Information, Knowledge, Wisdom hierarchy which treats each tier as being more highly processed and contextualized than the one which precedes it. From this perspective, data and information refer to different levels of complexity. In this sense, data could be referred to as acontextual recorded symbols, or values and information would be data in a context which is interpretable by a user who can be either a human or a computer (Rowley 2006, p. 167). Data and information can be differentiated by intent. While data can be seen as the atomic level of information which cannot be broken down any further, information is meant to communicate something. Digital objects serve as yet another level of complexity, and refer to a complete entity or unit. These may be

temporary collections of information originating from different sources. A medical record is an example of this, as it usually consists of different documents compiled together from different systems and different sources, depending upon what information the user has requested (InterPARES 2 Project 2008, pp. 16, 26).

Digital information requires an information carrier, but its identity should not be confused with that carrier (Yeo 2007, pp. 328-329). In the analog world, with a document such as a letter the content, context, and structure are all collected together in one place (Cook 2007[1994], pp. 426-427). While some analog information can be examined with the naked eye, others require special apparatuses for playback. In many cases, digital information is independent of its carrier and vice versa – for example, same information (a song) can be produced whether it is recorded on a CD or as a .flac file. Information carriers also have specific playback requirements: a CD reader is needed to playback a CD, while software and hardware compatible with .flac are needed to play back the .flac file.

Digital information can be structured in different ways. Both hardware and software are needed to read it: hardware deals with the physical aspects and machine-readability side of information while software interprets and presents it for the user. On a computer, the underlying software which allows all other software programs to be run is called the operating system, or OS (Borghoff, Rödig, Scheffczyk & Schmitz 2003, p. 8). Operating systems are structured in different ways. This is one source of the incompatibility issues between common operating systems such as Windows, Mac, and Linux. The programs running on the OS also have a part in reading information. Regardless of what OS one has installed, one will be unable to read a video file in a particular format if they lack the appropriate program. Since digital information is interpreted by software, some researchers mean that one cannot really separate the information from the software which it is dependent on (Allison, Currall, Moss & Stuart 2005, pp. 368-369).

Formats specify how information is organized in a file (InterPARES 2 Project 2008, pp. 16, 22). An example of this could be word processing documents which are saved as .doc or .pages files. Both of these are word processing file formats but the information in them is structured differently, which may cause problems if one attempts to open them in the wrong program. Moreover, the formats themselves exist in different versions, .doc from Microsoft Word has nearly a thirty year history behind it (Wikipedia 2016b). Older .doc files may be incompatible with the newer versions of the Word program due to these changes.

Formats vary in who is allowed to view their specifications and what kind of license agreement governs the format's usage. There are several kinds of formats including open source, unpublished, proprietary and free. Open source formats are those whose specifications are published and accessible to anyone. The specifications of unpublished formats are considered trade secrets and are not viewable by just anyone. Proprietary formats may be published but their access and use is governed through restrictions and license agreements. Free simply means that the format is free to use by anyone (InterPARES 2 Project 2008, pp. 33, 37; Wikipedia 2016c). Open source is generally preferred within long-term preservation since it provides some measure of future security. When someone wishes to access a file 15 years from now, the hope is

that they will have a better chance of finding the specifications of the format, which will allow them to find a way to view the contents of the file (Corrado & Moulaison 2014, p. 69).

Within a LAM institution, information is often divided according to whether it is born digital or digitized. Digitized means that a digital representation has been created from material which was originally analog. A photograph taken with a digital camera which is then saved to a computer never existed in analog format. Born digital material such as this can be defined as "[...] items created and managed in digital form." (Erway 2010, p.1). Some researchers take the perspective that digitized documents themselves are born digital because they always result in the creation of a new document in digital form. Those responsible for the creation of this new digital object have made choices in regards to which features should be preserved. Different choices would have resulted in a different document (Owens 2012). Digitize usually means "to change (information or pictures) to digital form" (Merriam-Webster 2015). In Swedish, digitization may also refer to the ongoing permeation of the digital world in our daily life, when the word 'digitalisering' is used. According to the Digitization Commission (Digitaliseringskommissionen) it can be used in two ways:

> *Digitization of information* is the process where an analog original is transformed into digital information. This means that the information can become structured, searchable, and made available via digital channels.
>
> *Societal digitization* is the profound process of transformation on the societal and individual level which gradually makes it more difficult to separate the digital world from other parts of our lives. This means that individuals and organizations can communicate and exchange information with other people, organizations and their environment in a completely new way. Digitalization and the use of IT-based solutions can contribute to increased access and effectivity for businesses and public administrations.[3]
>
> *Digitaliseringskommissionen 2014, pp. 28-29*

When the concept is used in this study, we mean the digitization of analog information, unless otherwise specified. The creation of digital search aids and the transformation of analog material into a digital format are both included in this concept. The consequences of digitization of analog materials have been discussed by many authors (among others Björk 2015; Dahlgren & Snickars 2009).

With digital information in archives and recordkeeping a distinction can be made between the original (master copy) and the copy accessed by users, which is sometimes called a derivative file. In the digital world, master and copy designates a relationship between files. The master is often placed in a separate location in order to protect it against alteration or loss. The access version a user interacts is frequently a version which is separate from the master. While there need not necessarily exist any appreciable difference between the two versions, the designation of one of them

[3] Authors' translation.

as the master helps to ensure authenticity, which is especially important when dealing with public records (Federal Agencies Digitization Guidelines Initiative 2014).

The question of what is to be designated a copy, and what is the original, is not entirely a simple one. The designation is an aid in managing an archive containing digital material. The International Association of Sound and Audiovisual Archives (IASA) writes in their recommendations that sound files in audio archives are often copies of the original recordings, but that these should be treated as the originals, despite the fact that they are copies. However, it can still be said that one file is being designated as the "original" in this case (IASA 2005, not paginated).

2.2 Digital preservation

Digital preservation presents unique challenges which must be identified in order to be handled successfully. Adapting to these challenges has been a difficult task for the archiving world. The Digital Preservation Coalition – a UK-based, non-profit organization – was set up in 2001 to encourage digital preservation best practices (Digital Preservation Coalition 2016). They point out six main differences between digital and paper-based materials:

- Digital information is hardware and software dependent
- Technological obsolescence
- Information carriers are subject to decay with dramatic consequences
- Loss of integrity (difficulty to maintain authenticity)
- There is no passive preservation (a digital resource which is not selected for active preservation treatment at an early stage will very likely be lost or become unusable in the near future)
- Preservation concerns need to be addressed at all stages of a digital object's lifespan

(Digital Preservation Coalition 2008, pp. 32-33).

When discussing digital preservation, it is important to clarify that storage is not the same thing as preservation. Digisam clarifies this by saying that storing information simply means that it exists somewhere, such as on a hard drive, while preservation entails making sure that it is possible to read and understand the information now and in the future (Digisam 2016a).

There is a massive amount of digital information being created daily, which consists of everything from private information such as photographs or family videos to public information like government records and academic work. This information also includes media which were formerly printed – such as magazines and newspapers which today are predominantly digital. Archives are one of the institutions working to save this information. The transition to mostly digital material has caused many headaches: while in the past one could preserve information simply by storing it in a room with the appropriate environmental conditions and safety controls, this is no longer an option in a world where information is predominantly digital. Digital information does not survive hundreds of years simply by being stored

on a shelf. The material needs to be kept readable and understandable in the face of quick and dramatic technological shifts. The survival of information is dependent upon the maintenance of its infrastructure and migrating it to contemporary formats. While the preservation of context is crucial in digital information, it is not unique to the digital realm – the same can be said of analog records. The interpretation of analog public records also requires knowledge about how the organizations which created them worked.

Digital material may be required to be saved for varying periods of time; in some cases decades, in others only a few months. It is important to specify the kind of preservation that is intended in order to specify what requirements are important to meet these goals. One definition of short, medium, and long-term preservation is:

- Short-term preservation – solutions that are used for a short time, 5 years maximum.
- Medium-term preservation – solutions that are used during a system's lifetime, 10 years maximum.
- Long-term preservation – solutions that are used after the originating system's lifetime, number of years unspecified.

Digital Cultural Heritage Roadmap for Preservation Project (DCH-RP) 2014, p. 49

There is no general consensus as to what constitutes long-term preservation. Other definitions of long-term describe it as a period of 50 years (Borghoff et al. 2003, p. VIII). We use the definitions provided by the DCH-RP in this text.

Archives can also be divided up into local, intermediary, and final archives. A local archive is the collection of information that an organization actively uses, such as information in the active organizational system, records in employee's offices, or elsewhere on the premises of the organization. A local archive often lacks search capability and has no direct preservation functions. In an intermediary archive, information is collected which is no longer frequently accessed and which need not exist physically near the organization which created it. Furthermore, an intermediary archive usually has search functions and the goal of preservation. The information stored at local and intermediary archives is usually owned by the organization which created it. The information is then handed over to an archival institution, called a final archive. Ownership rights are then transferred over to the archival institution, who is then considered to be responsible for the information. The final archive has search capability preserves this information (Aspenfjäll 2013, pp. 10-11).

There also exist another type of archive called dark archives. These can simply be defined as archives where information is preserved but is not accessible in most cases. Dark archives are often used in order to separate the original master copies of a file from the copies that users actually access. These dark archives are generally only accessed when new material is being placed in them, and are otherwise protected in order to maintain the authenticity of the originals by placing them in an environment that is as tamper and error proof as possible (Corrado & Moulaison 2014, p. 87).

Authenticity can be a major issue for digital records and is important to their evidentiality. Insofar as such records serve as proof of business conduct, all of the properties of the record involved in content, context, and structure are of the utmost importance (Borglund 2008, p. 15). There are two main facets of authenticity in the digital archives: validating and maintaining the integrity of the original file, and the thoroughness with which metadata and contextual information were recorded when the digital object was first produced (Yeo 2010, pp. 4-5). There is always an element of trust when it comes to authenticity. A user has to be able to trust that the document provided by the archive has not been manipulated in any way which affects its authenticity, which can be difficult to verify. In order to choose which aspects of digital material to save, archivists must understand why something is being preserved: though archives and libraries are seldom able to know to what end something will be used, it must be assumed that users will limit their usage to something within the spirit of the original intention (Thibodeau 2002, p. 14-15). Ensuring the authenticity of records has been an issue of primary concern for projects such as the InterPARES project, which has developed plans to protect the authenticity of digital records during all periods of their life cycle (Cook 2013, pp. 100-101).

2.3 Digital preservation strategies

Digital preservation strategies can be characterized by the issues they prioritize. Certain digital preservation strategies prioritize protection against technical issues such as hardware, software, and format obsolescence while others are primarily concerned with funding and organizational issues. While any combination of these issues are important at one time or another, different organizations have different needs depending on their size, where they find themselves in the preservation process, and what kind of support they receive.

One such handbook has been developed for digital preservation with support from the European Commission. This is the DCH-RP, which lists four different kinds of strategies for digital preservation which they describe as techno-centric, analytic, incremental, and durable digital object. The first strategy is concerned with maintaining the original hardware and software in order to keep digital information readable. The analytical strategy aims to predict which data will be in demand in the future and either ensure its future usability, or attempt to recover the data at a later date when it is needed. The incremental strategy utilizes emulation and migration, and the last strategy of durable digital objects seeks to adapt the formats of current data so that they will be self-contained within an environment, ensuring their readability in the future (DCH-RP 2014, pp. 24-25).

While there are many different strategies for the preservation of digital material, migration is currently the most commonly used method (DCH-RP 2014, p. 25). The term migration can have different meanings within digital preservation. One of these meanings is to migrate a digital object whose format is obsolete (or risks becoming obsolete) by transferring it over to a newer format. Another meaning is the transfer of files from one information carrier to another (InterPARES 2 Project 2008, pp. 31, 48).

Emulation is another option for digital preservation. It targets the operating environment that the information exists in rather than the file. Emulation preserves

the material in its existing format, but recreates the functionality which existed before its hardware/software/format became obsolete (Kristiansson 2002, pp. 207-208). An example of this would be an older computer game which was programmed for Windows 95. In order to run this game on a current-day computer, the emulation option would involve reproducing the Windows 95 operating environment.

Emulation and migration have been treated as opposing strategies in the past, but some claim that this is not really the case, and that they should in fact complement one another (Anderson, Delve, & Pinchbeck 2010, pp. 115-116). The emulation itself will eventually require migration. Emulation can become too complicated to be viable in the long run, especially when dealing with a complex network of relationships and all of the specifications which they entail (Thibodeau 2002, p. 19-20).

One choice that many preservation institutions have made in order to cut costs is to utilize cloud computing services. Cloud computing allows information infrastructure to be used when it is needed. Clients pay a cloud service provider for this service. The advantages of cloud computing for the client are that they need not own and maintain the hardware and software they use, nor must they have the technical knowledge needed to maintain it. Cloud computing has become a popular choice internationally for institutions in need of cheap computing power. Several countries have so-called "Cloud First" policies, mandating that cloud computing options be considered before other IT solutions, unless the latter are more inexpensive (Kundra 2011, pp. 1-2, 9).

While a full discussion of cloud services in an archival context would be out of place here, a brief description follows. Cloud services provide access to software, platforms, or infrastructure such as storage and hardware. A client can purchase more of a service from a cloud service provider as needed, and the service can even be controlled to meter usage, thereby controlling spending and ensuring that the institution pays the same amount each month. Furthermore, there are three flavors of cloud storage: public, private, and hybrid. Public cloud storage usually involves large scale infrastructure and can be accessed via the internet. An example of private cloud storage would be an organization's own network which may only be accessible locally. Hybrid cloud storage is a mixture of these two previous kinds (Beagrie, Charlesworth & Miller 2014, pp. 6-7, 13-15).

The usage of cloud services are not unproblematic in an archival context. In Sweden, cloud services are less often chosen due to perceived lack of security, complex legal issues, and long-term sustainability. The Swedish Civil Contingencies Agency (Myndigheten för samhällsskydd och beredskap) has provided guidelines to organizations for how they should handle alternatives such as cloud services in connection with the public procurement of IT-related services. When the information contains sensitive personal information, organizations are advised to avoid cloud computing (Myndigheten för samhällsskydd och beredskap 2013, p. 22).

2.4 Information loss and a digital dark age

The digital dark age refers to a scenario where a large portion of our future historical record will be lost if steps are not taken immediately to preserve the digital information we have now. The term 'digital dark age' references the so-called

historical Dark Ages, a time from which we have relatively few written records, compared to the periods before or after. There are many causes for concern in relation to a digital dark age, but format, software, and hardware obsolescence, together with technical issues like bit rot and the fact that digital preservation requires a constant stream of funding are among the most common (Kuny 1997, p. 5). Though it has a melodramatic ring to it, the term may nevertheless be accurate. It is easy enough to imagine digital information being preserved 10 or 20 years into the future, but preservation on the scale of centuries seems questionable. The digital dark ages could become reality in many ways. If we lack the ability to read and display obsolete file formats, or the required hardware needed to access older storage media does not exist, then we will lose such information. Information which was never targeted for preservation in the first place will also be lost, as will anything saved on storage media which has decayed (Rosenthal 2011, not paginated).

The problem of the digital dark age was recently brought into the public consciousness when Vint Cerf, Vice President of Google, discussed it in articles and speeches. Cerf intended to promote a preservation approach that preserved digital objects in their original context. The discussion triggered a wide variety of reactions from the preservation community, ranging from outright denial and indignation to acknowledgement of the problem (Anderson 2015, pp. 20-23). Regardless of the response, the significance of this was that the threat of a digital dark age entered into a broader public discourse.

Sometimes digital preservation fails to preserve that which it intends to save. When digital material becomes damaged or missing, or its recovery would require an unreasonable reallocation of resources, it can be termed information loss. Strategic plans for digital preservation are, in part, meant to prevent the loss of access to digital material. Some of the reasons for information loss include damage to the physical information carrier, obsolete file formats, inability to access a reader for older information carriers such as floppy disks, broken links between information and its metadata, problems with licensing and copyright agreements, and intentional or unintentional damage to files (Gladney 2007, p. 10).

Obsolescence occurs when formats, hardware, or software become outdated and are no longer supported. When new operating systems are released, older file types and programs eventually become unusable. An example could be text documents on a Commodore 64. This becomes a problem because, as technology changes rapidly, it makes it difficult to preserve digital objects for more than a few years at a time (Corrado & Moulaison 2014, pp. 227, 229, 231). One consequence of this is that it takes a lot of money for archiving institutions to maintain their digital holdings and keep them current. Obsolescence is currently one of the greatest threats to successful digital preservation. If a file cannot be read, then it is nearly the same thing as a document having been destroyed. If obsolescence has progressed too far, large amounts of archival resources may need to be reallocated to recover a collection – not to mention the fact that it must also be maintained afterwards. It has been claimed that obsolescence is causing the greatest period of information loss ever witnessed (Brand 1999, p. 46).

2.5 Current projects within digital preservation in Sweden

There are many different efforts currently aimed at improving digital preservation in Sweden. In this section we mention some of the efforts which affect organizations in the public sector and to some extent organizations in the private sector. This involves a shared e-archive for governmental agencies, harmonization of data, and a more centralized management of digital cultural heritage.

The Swedish government is working to develop a unified e-archive for all governmental organizations via a collaboration between the authorities at the National Government Service Centre (Statens servicecenter) and the Swedish National Archives. The general director for the National Government Service Centre, Thomas Pålsson, explains that there has been a wide range of quality in digital preservation efforts up to the present day. He claims that under the current circumstances, the legal requirements placed on this information risk going unfulfilled. Without terming it as such, he describes fears which could serve equally well as the definition of a digital dark age: that information, or the access to it will be lost due to insufficient storage techniques and a lack of standards. Pålsson describes the up and coming electronic archive as an intermediary archive, which will ensure the standardization of the information passing through it before being delivered to the Swedish National Archives. The national electronic archive will be introduced during the period from 2017-2018 (Statens servicecenter 2015; undated).

The eARD project, which ended in 2014, was meant to improve the handling of digital information within Swedish governmental administration. The project was led by the Swedish National Archives with the participation of national, regional and local authorities. Part of the project was meant to develop specifications for how metadata was structured within electronic archives and electronic registers which are shared by administrations, called FGS (FörvaltningsGemensamma Specifikationer). These specifications are intended to improve the interorganizational transfer of information and reduce the difficulty of the work involved in locating records for private persons and civil servants (Riksarkivet undated a, undated b).

In the spring of 2016 Digisam released a report of its activities during the period 2011-2015. This report presents their evaluations and suggestions which Digisam has arrived at in their work. Generally, it can be said that Digisam advocates a more centralized handling of digital cultural heritage information through collaborations, collective storage solutions and a centralized structure for the digitization of analog material. It suggests that infrastructure solutions be carried out in collaboration with the Swedish University Computer Network (SUNSET). Another suggestion is that the Center for the Conversion of Media (Mediakonverteringscentrum) be developed into a national resource for large scale digitization. Digisam proposes the introduction of mobile digitization units, where digitization can be carried out locally at organizations, rather than transporting fragile cultural heritage material. This is intended to aid in maintaining a high level of digital preservation competencies for digitization, which Digisam claims is difficult in today's circumstances of project-based initiatives and geographically dispersed competencies. Furthermore, in response to the demand from institutions, a legal support for organizations is suggested which can provide advice on legal issues with digital cultural heritage information. In order to develop a more centralized way of working with digital

cultural information, Digisam suggests that an organization should be introduced which builds on Digisam's work and which serves as a center for continued collaborative work (Digisam 2016b, primarily pp. 8-9, 14, 31, ch. 13, 51-53, 61, 70).

2.6 Swedish legislation concerning public records

There are four constitutional laws in Sweden. One of these is the Freedom of the Press Act (Tryckfrihetsförordningen, SFS 1949:105). Within this constitutional law the Principle of Free Access to Public Records (Offentlighetsprincipen, SFS 1949:105, Ch. 2) is defined, which legislates governmental transparency and the right of the public to view public records. The Principle of Free Access to Public Records itself provides a definition as to what constitutes a 'record' (here called 'document') within public administration, and what makes such a record a 'public record':

> Document is understood to mean any written or pictorial matter or recording which may be read, listened to, or otherwise comprehended only using technical aids. A document is official if it is held by a public authority, and if it can be deemed under Article 6 or 7 to have been received or drawn up by such an authority.
>
> *SFS 1949:105, Ch. 2 § 3*

The definition of a 'record' (called 'document' above) in a Swedish context is independent of whatever media it takes the form of. The same law applies to digital as well as analog information. A record becomes public in Sweden as soon as it is received or dispatched. A record can be considered free or classified (SFS 1949:105, Ch. 2). While one can go further into the definition of what is considered to be a government agency and where the boundary between work material and a public record is, such a discussion is not within the scope of this paper.

As noted previously, this thesis uses the term 'record' outside of the framework of 'public record', as defined in section 1.2. When a record – in the meaning of a public record – arises, we specify it as such.

3 Previous research: Digital preservation issues

This section sketches out the research context relevant to our study. Though this study is primarily grounded within archival science, we also look at other areas of research concerned with digital preservation. It should be noted here that archival science as an independent field of research is a relatively new development and retains its origins in an interdisciplinary approach (White & Gilliland 2010, p. 233). Digital preservation itself is an even newer field (Quisbert 2008, p. 1). The research discussed here extends into other disciplines such as library, data, and information science, as well as information and media technology. Both Swedish and international research will be discussed, though we will only take up the Swedish perspective when dealing with digital public records, due to the country-specific nature of the laws governing them (Runardotter 2007, p. 5).

Hugo Quisbert describes long-term digital preservation as a new research area where models, methods, and tools are being developed and tested to ensure that digital information will be readable in the future. Quisbert structures research concerning digital preservation according to a progression between four phases: from hands-on approaches, to becoming more involved with the software used in preservation (and attempting to make some of those processes automatic), to management and organizational issues, and finally a fourth and as yet undetermined phase which Quisbert claims is characterized differently by different researchers. He also claims that this research has shifted away from hands-on work towards a more theory-based approach, and from a focus on techniques to an increasing emphasis on questions of organization (Quisbert 2008, pp. iii, 1, 10-11).

3.1 More than just technology

In this section we examine literature concerning the technical and organizational aspects of digital preservation issues. They are concerned with when and where digital preservation begins and what kinds of qualities are desirable in a system for long-term digital preservation.

In her article “Information technology challenges for long-term preservation of electronic information”, Viveca Asproth (2005) writes about the massive quantities of information which are now being created digitally and explains that the increasing availability of e-services will cause this quantity to grow even more. The future she describes is the present we now find ourselves in, eleven years after the article was written. Asproth discusses four areas that involve challenges for digital preservation. These include technical challenges, legal issues, organizational challenges, and context and metadata (Asproth 2005, pp. 30-32).

With technical challenges Asproth means things such as obsolete file formats, hardware and software, together with gaps in technical knowledge and problems with security. Concerning legal issues, she has in mind problems such as authenticity, provenance, and the integrity of digital materials. Asproth sees organizational challenges as being the most serious of all these issues. She is concerned with the uncertainty surrounding just who is responsible for digital information, and the lack of knowledge concerning what happens when no one actively takes care of digital holdings. Asproth also sees that cooperation and communication between archivists and IT personnel is another area in need of improvement. Accordingly, part of the solution is for archivists to improve their IT competencies. Asproth writes that despite the organizational nature of these digital preservation challenges – which possess both a strategic and structural character – the greatest focus hitherto has been on solving technical problems. She questions if it might not be the case in the future that we will be unable to understand the context in which digital information has been created. It is these challenges which she has in mind when she writes about problems involving context and metadata. Asproth means that the difficulties in understanding a records context are greater when dealing with digital materials than they are with their analog counterparts (Asproth 2005, pp. 30-32).

She also points out that technical developments have transformed archival practice, due in part to the fact that content and context exist in and of themselves in the digital realm, while they only exist together in the analog world. She claims that this affects what one can identify as a record (Asproth 2005, p. 29).

> In paperbound records the content and the context exist together, while in electronic documents they are in fact separated and stored independently of one another. Therefore documents do not exist as such, except in the moment of capturing. Information is stored in bits and pieces, and can only, with help of software, be connected and transformed to a document as required. An example is a digital map, which consists of numerous objects organized in different layers. The map, when it appears, is merely a view constructed at that moment for some specific purpose.
>
> *Asproth 2005, p. 29*

Asproth states that there are many problems which must be solved in order to reach the goal of secure and effective long term digital preservation, claiming that, within archival practice and information technology, a collaborative effort is needed to create these solutions (Asproth 2005, p. 28).

In his dissertation "Design for recordkeeping: Areas of improvement", Erik Borglund (2008) researches the design of information systems where archival information is created and managed. Borglund places a special emphasis on the importance of proactivity when it comes to the preservation of digital records. This proactivity must occur on several levels regarding the digital record, the system where the record resides, the organization which manages the record, and the user themselves (Borglund 2008, p. iii).

Along with the creation of the digital record, its requirements for preservation must already be fulfilled. If this is not the case, the digital records risk being preserved with

to a substandard degree of quality. In this context, Borglund also discusses the organizational aspect of digital preservation. Like Asproth (2005, p. 32), Borglund criticizes how technical aspects have been excessively focussed on in the past. He means that previously, research has had a reactive and technical orientation. With this he means that it has tried to solve the problems of long-term digital preservation by treating information which has already been created. One aspect of what Borglund examines is what organizations define as a record, arriving at the conclusion that there is a discrepancy between the concept of 'record' within archival theory and how it is understood out in the archiving world. He states that organizations must identify what is a record within their specific organization, so that this can be utilized in the design of information systems which manage information (Borglund 2008, pp. 17-18, 37-38). Keeping both Asproth's and Borglund's critique in mind, we have chosen to see the challenges with digital information loss from a perspective which includes technical matters without being limited to them.

While Borglund focusses on the organizational aspects concerning the construction of information systems, Hugo Quisbert (2008) focusses on the technical. In his dissertation, Quisbert examines which framework and specific demands should be placed on the information systems which are used for long-term digital preservation. The study posits three challenges for digital preservation: technological obsolescence, an ever-increasing amount of information produced, and preservation competencies. Since the first two things are impossible to affect, Quisbert claims that the solution is to focus on competence within long-term digital preservation work (Quisbert 2008, pp. 1-4, 7).

3.2 Limited funding

There are several studies which attempt to provide a picture of what digital preservation currently looks like in practice. Some of these are primarily concerned with funding strategies, while others are concerned with enacting the best possible digital preservation in the face of extremely limited funding.

The first of these studies was carried out by Nancy Maron, Kirby Smith and Matthew Loy (2009). The authors look at funding strategies for digital resources. They examine supplementary funding strategies at various institutions with established digital preservation programs and attempt to determine what successful strategies look like. In the study they also examine the results of a previous report that investigated various sources of income for archives containing a wide variety of different materials. Concluding that a focus beyond the basic, fundamental revenue is necessary in order to survive budget cuts, Maron, Smith and Loy advocate a more flexible and dynamic approach to funding strategies which should also further the interests of the archive. The basic support for those institutions who depend upon a host institution must already be stable however, in order to make this dynamic approach possible (Maron, Smith & Loy 2009, p. 28).

The attempts of smaller institutions to meet an unfunded government mandate which required these institutions to make their data available to the public at no cost was the topic of a study titled "From theory to action: "Good enough" digital preservation for under-resourced cultural heritage institutions". Written by Schumacher et al. (2015) it

examines how several smaller universities went through the process of analyzing and vetting tools for digital preservation. The institutions collaborated in creating their own analytical chart called Tool Grid, which they used to compare digital preservation tools such as Archivematica, Preservica, MetaArchive, Duracloud, Curator's Workbench and Internet Archive. The most important result of this study was that, for smaller institutions, digital preservation is sometimes only achievable in increments, and that such an option is preferable to the idea of an either/or proposition: a comprehensive digital preservation program or no digital preservation at all. As smaller and medium sized institutions are often not in control of the policies which guide them, they must make do and try to preserve as well as they can. The study concluded that while there was no realistic chance of enacting these plans for some of the institutions, every step forward would, at a minimum, reduce their future workload (Schumacher et al. 2015, pp. 6-9, 15).

3.3 A changing profession

The profession of archivist has undergone rapid changes during the last few decades. The following studies deal with the relationship between the archivist's professional role and the growing importance of the digital world. They are also concerned with the role of the archivist in teaching others what material should be preserved.

Terry Cook has been a prominent voice in the dialogue over the changing roles of archives and archivists. In "Electronic records, paper minds", Cook (2007[1994]) reflects over the archival profession in the postmodern world. He claims that the archiving profession must transform itself from a passive, custodial role where one describes the content of records into an active role where greater focus is laid on the description of context. Cook sees two reasons for this: the first being that an ever greater amount of information is being created, which requires archivists to take an oversight role. The other reason is that the digital world is radically different from the analog one. Cook is critical towards the profession's attempts to solve the problems of electronic records with concepts, terminology and practice which have their origin in an analog understanding. He claims that the changes in the profession are more than a simple adaptation to the new technical reality: they constitute a paradigm shift. If one considers a paper document, content, context, and structure are bound together in one place. People can take part in all of them with the naked eye. Cook claims that this is not the case for digital information, where these aspects are separate. Here the archivist's job is to understand and describe these things and how they fit together. Furthermore, he points out that in the digital world, it is often a large group of people, and sometimes several organizations, who are involved in the creation of information. Cook uses a geographical information system as an example – it gathers information originating from many different sources. Cook claims that in this type of information there is no document in the meaning usually understood by archivists and information managers – instead, a number of different ways of seeing information exist, depending upon what a user is asking of the system (Cook 2007[1994], pp. 402-403, 413, 422-426). In a later article, Cook (2013) claims that there have been four archival paradigms since the middle of the 1800s: premodern, modern, postmodern, and contemporary. He claims that we are in the middle of yet another paradigm shift since the writing of "Electronic records, paper minds". In the contemporary paradigm, archivists have more of a supporting role in empowering communities. The records

are stored with the community, and the archivist's role is to support them, especially regarding digital records (Cook 2013, p. 116).

In the Swedish context, Mari Runardotter (2007) has researched the effects of digital reality on the archival profession in her licentiate thesis "Information technology, archives and archivists: An interacting trinity for long-term digital preservation". In her thesis Runardotter examines the archivist's situation, where professional responsibilities include preserving digital information in the long-term. Runardotter comes to the conclusion that an archivist's work responsibilities become unwieldy due to a lack of understanding amongst other professions. Archiving concerns are not well understood amongst these other groups, and such questions have a low priority and status among them. Due to a lack of understanding from these collaborators, archivists are forced to continually re-educate those around them. Furthermore, Runardotter states that archivists themselves suffer from a lack of IT competence, while IT personnel suffer from a lack of archiving knowledge. Cooperation and communication between the groups is often sub-par. Due to the common presupposition that archivists work only with paper records, they are shut out from digital preservation work. Since archivists are not often involved in the development of new information systems, they lack the ability to affect the system's suitability for long-term digital preservation, simultaneously as they are left alone to take care of such systems after they are phased out. Runardotter states that many organizations today lack strategies for long-term digital preservation. She also claims that the digital format has led to public records not being recognized as such, causing a failure to adhere to the legal standards which govern public records (Runardotter 2007, pp. 51-52, 70-71).

During the same year, Runardotter wrote an article together with Anita Mirijamdotter and Christina Mörtberg called "Being an archivist in our times: Trying to manage long-term digital preservation" (2007). Here, the authors have carried out a smaller study in order to gain a more in-depth understanding of archivists who have digital as well as analog records to manage. During the course of their research they interviewed an archivist at a university and made observations. Like Asproth (2005, p. 32) and Borglund (2008, pp. 17-18) Runardotter, Mirijamdotter and Mörtberg claim that a disproportionate amount of focus has been placed on the technical aspects of digital preservation and that they wish to offer a different perspective in their article. The authors claim that the technical development has created a situation where archivists seldom have a background in history, and have seen their profession begin to resemble that of librarians and other information specialists. The archivist interviewed in the study expressed frustration over having to teach co-workers how and what to archive. The archivist expressed that archiving is the collective struggle of all of those who produce information, not something that the archivist solves all by themselves when the record is no longer current. This reiterates the problem where archives and archiving are given low priority, and that organizations do not necessarily see the value of preservation work. The archivist in the study has encouraged the university to develop a strategy for long-term digital preservation that would improve cooperation with the IT division, but these efforts were ultimately futile. This lack of cooperation is something the archivist sees as highly problematic – digital information must be managed even before it is produced. The archivist expressed that there were no clear guidelines regarding when digital information

should be registered and how it should be saved – such as when one deals with web pages. Due to the uncertainty with digital preservation, the archivist made use of what she called "double security", which means that the digital information is also printed out on paper (Runardotter, Mirijamdotter & Mörtberg 2007, pp. 48, 54-56).

The insufficient level of cooperation between archivists and IT personnel is discussed by Karen Anderson, Göran Samuelsson and Marie Morner Jansson (2011) in "Benchmarking information management practice and competence in Swedish organizations". The authors set out to examine the information cultures of organizations. They claim that the deficient cooperation seen between archivists and IT-personnel has led to information systems which are less appropriate for long-term digital preservation than they otherwise would be. Anderson, Samuelsson and Jansson state that while many organizations take advantage of the technical possibilities, they lack the competency to properly secure the information for the needs of today and for the future. Moreover, when important information is lost it is a blow to the organization's daily needs. It also makes following legal requirements more difficult and hampers future users of the information. Even these authors claim that there is a lack of IT competencies with archivists, while at the same time archivists' competencies aren't used to the full advantage of the organizations they work for (Anderson, Samuelsson & Jansson 2011, pp. 32-34).

3.4 Digital public records

In her dissertation "'The emperor's new clothes' Recordkeeping in a new context", Maria Kallberg (2013) examines how Swedish municipalities attempt to live up to the demands of managing digital public records, how technical advancements have affected the professions of archivists and registrars, and how information capture functions together with archives legislation. She investigates the awareness of how to properly carry out recordkeeping duties in the world of public agencies, where records are predominantly digital.

Regarding the information which exists in information systems, Kallberg comes to the conclusion that collaboration is necessary between researchers in the fields of information systems and archives in order to ensure that the role of archives are fulfilled in the future. She writes that the task of identifying records needs to be adapted to function across a wide spectrum of digital systems. Referencing an earlier study from 2011 concerning Swedish work with businesses which attempted to improve their preservation efforts, Kallberg wrote that business systems were being phased out without any long-term solution for preserving them. Since they had no proper strategy for handling digital records, the approach taken by these agencies towards long term preservation often consisted of a paper-based strategy. Kallberg writes that despite the lack of proactive work and strategies, an awareness of the problems with the preservation of electronic records exists. On the other hand, funding, knowledge, and cooperation between professions is often insufficient and needs to be improved upon (Kallberg 2013, pp. 36-37, 47, 102, 123).

3.5 Risk assessment

We have encountered several studies during our research that deal with threat studies, risk analysis, and digital preservation. Some of these studies involved determining the efficacy of a particular risk assessment method, while other sought to create a typology of threats.

The first paper discussed here is titled "Risk Assessment: Using a risk based approach to prioritise handheld digital information". This study was an inspiration for our own, especially with regards to its questions and methodology. Written in 2008 by Rory McLeod on behalf of the British Library's Digital Library Program, the study applies the LIFE risk assessment method to the library's digital holdings, attempting to map out where threats are realized in the archival process. Though it relied on questionnaires rather than interviews, the questions were quite similar to our own. One conclusion of the report was that organizational issues currently take precedence over technical ones in digital preservation. This, however, does not exclude the danger of technical threats: McLeod concluded that the decay of physical media such as CDs and DVDs was among the greatest threat to the library's holdings at that time (McLeod 2008, not paginated).

A similar study was carried out much earlier by Cornell University in the United States, on behalf of the Council on Libraries and Information Resources, titled "Risk management of digital information: A file format investigation" (Lawrence, Kehoe, Rieger, Walters & Kenny 2000). This study is useful to gain a historical perspective on how threats to digital preservation have been handled in the past. The core of the report analyzes format migration according to risk assessment methodology. Finding insufficient support in their own discipline at that time, the authors turned to computer science research. The study considers different conversion tools and compares their actual output with a loss-free output calculated by the authors in order to check for subtle errors. The study also considered different kinds of formats and found that the most significant problem (at the time) was finding the specifications for proprietary formats such as TIFF (Lawrence et al. 2000, pp. 4, 12).

Another study dealing with threats to digital preservation, but without a risk assessment model, was concerned with how to identify and categorize these threats. In a paper written by Justin Littman (2007), the study details the threats to a digital preservation project which was then under development. The project, carried out by the National Endowment for the Humanities and the Library of Congress, involved a review of the digitization of newspapers from the early 20th century which had become public domain. In contrast to the other studies, this was intended only to identify the threats, not to recommend corrective measures or strategies to handle them. In the concluding remarks, Littman notes that the greatest threat faced during the development of the project was actually operator error, which occurred when one of the developers accidentally deleted a large cache of files. The project also encountered other failures involving software and hardware (Littman 2007, not paginated).

3.6 Information loss

The risks for information loss are usually raised while discussing digital preservation in a general context. The focus has often been on finding solutions to prevent the development of information loss. In this section we look at studies which deal primarily with the information loss which has already occurred.

In a Swedish context, Erik Borglund and Karen Anderson (2011) are among those who have contributed to the research concerning information loss. In "L.O.S.T records: The consequence of inadequate recordkeeping strategies", they utilize a case study to discuss information loss which is set to occur due to a strategic mistake. Borglund and Anderson mean that the research that has occurred up until the present day has looked at either electronic records management systems or at long-term preservation, but studies have seldom been performed where one looks at the records which are being produced now and their future availability. In their study, Borglund and Anderson investigate how recordkeeping functioned within the extensive railway infrastructure construction project Ådalsbanan. The project, run by the Swedish Rail Administration, entailed the construction of a 180 km railway in northern Sweden, between Sundsvall and Långsele at a cost of 6.6 billion Swedish kronor. The documents which were produced within the project were judged to be viable for up to 100 years, because this information would constantly be in use for things such as future repairs (Borglund & Anderson 2011, pp. 272-273).

A long-term strategy for recordkeeping was created for the Ådalsbanan project which specifically targeted information in the form of documents. This meant that information not defined as a document fell outside of this strategy. This untargeted information was stored in a database and in a business information system. Having never been designated for preservation in the first place, the information will eventually be lost. Borglund and Anderson claim that something which is a document is easier to handle as an aggregated "unit", while other information which perhaps has less clearly defined boundaries is more difficult to identify (Borglund & Anderson 2011, pp. 279-280). Like Cook (2007[1994], pp. 410-411), Borglund and Anderson point out that a system which is in use must be structured for digital preservation from the beginning, and that migration must be planned before the system is put into use. Digital records, the authors write, demand a proactive approach. The authors attempt to show this in a discussion of the Records Continuum Model (RCM). They claim that if the organization had performed a process-oriented mapping of when and where the information was created, the organization would likely not have put its information at such high risk. The authors also claim that a process mapping would have provided the document with a context; which would have been decisive in preserving the authenticity of the records, as well as keeping them understandable. Without a context it is difficult not only to understand the information, but also to judge its authenticity (Borglund & Anderson 2011, pp. 279-281). As the authors write:

> Without the provenance and the contextual links between records, records cannot be demonstrated to be authentic and reliable, evidentiality is lost and the use of the records for knowledge and understanding about what has happened will be difficult.
>
> *Borglund & Anderson 2011, p. 279*

Taking inspiration from this study, we make use of RCM in order to better understand the digital preservation efforts of archives (discuss in section 4.1). The study has also contributed to our decision to take a more holistic approach in our work.

In an international context, we looked at a study by Joy Perrin, Heidi Winkler, and Le Yang (2015). In "Digital preservation challenges with an ETD collection: A case study at Texas Tech University", The authors examine digital information loss which occurred in a database containing electronic theses and dissertations (ETD) during the years 2005-2012. In the study they take up four specific cases of information loss encountered by the organization within their ETD database. These cases involved loss of metadata, database corruption and automatic overwrites of important information. In 2011 the library chose to focus on long-term digital preservation and secured funds to do so. This time, they were able to follow proper digital preservation procedures including the creation of a backup in separate geographical location and the introduction of a daily automatic control to prevent unwanted overwrites (Perrin, Winkler & Yang 2015, pp. 99-102).

Insufficiently developed strategies and the total lack of strategies led to information loss in both of the two studies just discussed. What should be mentioned in this context is that the Swedish National Archives conducted a 2009 survey which showed 97 % of the 322 agencies who took part in the survey stated that they archived digital material – while at the same time only 5 % answered that they had an overview and well-documented strategies for preservation (Riksarkivet 2011, p. 79). Digisam carried out a study between 2013-2014 which showed a similar set of problems. Seven cultural agencies took part in this study. All of them had cultural heritage material in digital format, but none of the agencies had a documented strategy for long-term digital preservation (Digisam 2014, pp. 11, 13). These reports can be seen as an indication that there is much work to be done in this area.

4 Theory

This section presents the interpretive frameworks which are used in the analysis of the study's results. Four interpretive frameworks will be used: a model which shows how technology has changed our understanding of a record's lifecycle and how it has affects archival practice, an ontological understanding of digital information, a risk assessment model and a model to help understand the challenges and digital information loss from a bird's eye perspective. These have been chosen in order to highlight research questions from multiple, complementary directions. While it can be debated that we are in a gray zone between theory and models, we feel that these interpretive frameworks are the most appropriate tools to understand the results of our study.

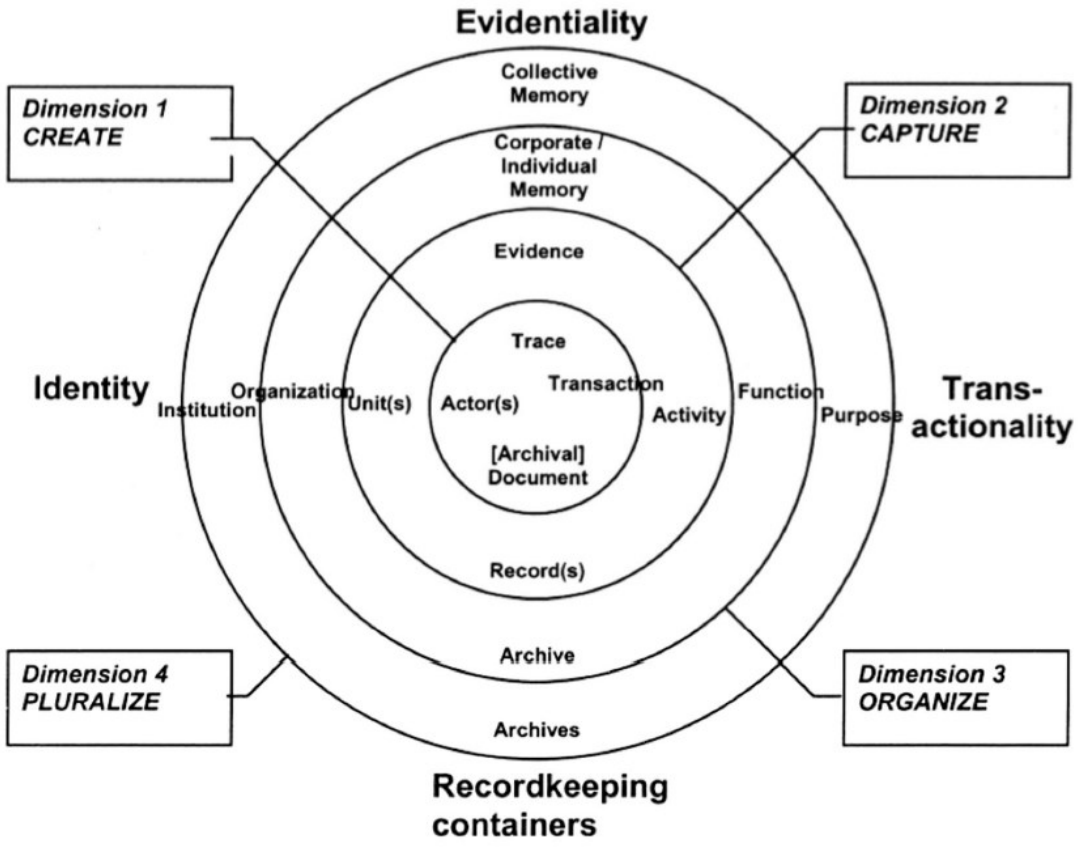
4.1 The Records Continuum Model

The Records Continuum Model (RCM) (see Fig. 1) referred to here was developed by Frank Upward, who published two papers describing the model in the 1990s (Upward 1996, 1997). Its creation was a response to perceived flaws in the formerly dominant Records Lifecycle Model, or RLM (Quisbert 2008, p. 23). RLM was a model with neatly delineated boundaries and responsibilities, which divided up the treatment of records management and archives administration and the respective responsibilities of each. As the name implies, it viewed the record as proceeding through developmental stages, from its birth to its death. In her article titled "An integrated approach to records management", An Xiaomi has characterized the transition from the RLM to the RCM as a shift in the conception of archival practice from reactive to proactive. She identifies the RCM's focus on integration and collaboration as one of its most beneficial aspects. In RCM, the record is within the domain of the continuum before it is even produced, and the responsibilities for it are shared both by archivists and those who produce them (Xiaomi 2003, p. 27). Upward motivated his proposal for the continuum as being better suited to serve the creators of records and their users (Quisbert 2008, p. 23). He also explained that there was a need for a model to accommodate the then new paradigm of preservation which was emerging with electronic recordkeeping (Upward 2000, p. 1). This required the iteration of a postmodern paradigm which was both more fluid and allowed for the interplay of multiple aspects – rather than the discrete descriptions provided by RLM. These advantages grow when dealing with digital material, since the RLM's focus on the custodianship of physical objects and repositories is poorly suited to material whose most important aspects involve its reproducibility rather than physicality (Xiaomi 2003, pp. 25-28).

Upward's model provides a graphical depiction of actions which occur in

recordkeeping throughout space and time. In general, the model is intended to serve as a "conceptual tool" to understand the processes which occur in recordkeeping (Upward 1997, p. 4). Through the dimensions and coordinates of RCM, more variables can be combined when examining best practices for archives management.

Fig. 1. Records Continuum Model.



Source: Upward 2005, p. 203. © Frank Upward (used with permission)

Upward's model visualizes the axes and dimensions which records move through, helping to locate them within the spectrum of usage and preservation. The model includes time, the function of the record and actions taken on the record.

Upward explains that the model can be used by creating a "dimensional analysis" by connecting the different coordinates in the model. These different combinations show where a record is at in the dimensions of the RCM. Different dimensional analyses can be undertaken at different points in time during a record's existence. These may be useful to show which needs are relevant in using the records at the point in its existence (1996, p. 7).

In Upward's model, the first dimension, document creation, is the point from which a document is set in motion and enters into the realm of recorded information. The second dimension, records capture, is the point where this information is incorporated into a framework, structuring its usage so that it is both consistent and coherent. This dimension emphasizes relationships linking to the network of related documents (Xiaomi 2003, p. 25). The third dimension, the organization of corporate and personal memory, configures the framework around the document so that those approaching it from a different knowledge structure will also be able to understand it. Upward describes this as the "organization of memory" (Upward 1997, p. 1). Xiaomi's article describes it as incorporating different families of records under one organization. The fourth dimension, pluralization of collective memory, is when the information radiates outward from behind its organizational context and into social contexts. The aspect of service is emphasized here, towards those institutions whose documents are preserved by the archive, as well as towards users (Xiaomi 2003, p. 25).

The axes are presented by Upward as coordinates which can combine with one another dimensionally (Upward 1996, p. 6). The first of these is the transactional axis. It can be defined by the kind of actions which can be performed on or with documents, and is concerned with accountability. The second axis is that of identity. It can be thought of as what work group the document belongs to and what organization it falls under. Here, records are grouped and tagged with metadata. The third axis is that of evidence. (Xiaomi 2003, pp. 25). This axis is about the materialized by-products of past events, or traces as Upward calls them (Upward 1996, p. 6). The fourth and final axis is recordkeeping. This is concerned with the "vehicles for storage" which hold records (Upward 1996, p. 6; Xiaomi 2003, p. 25).

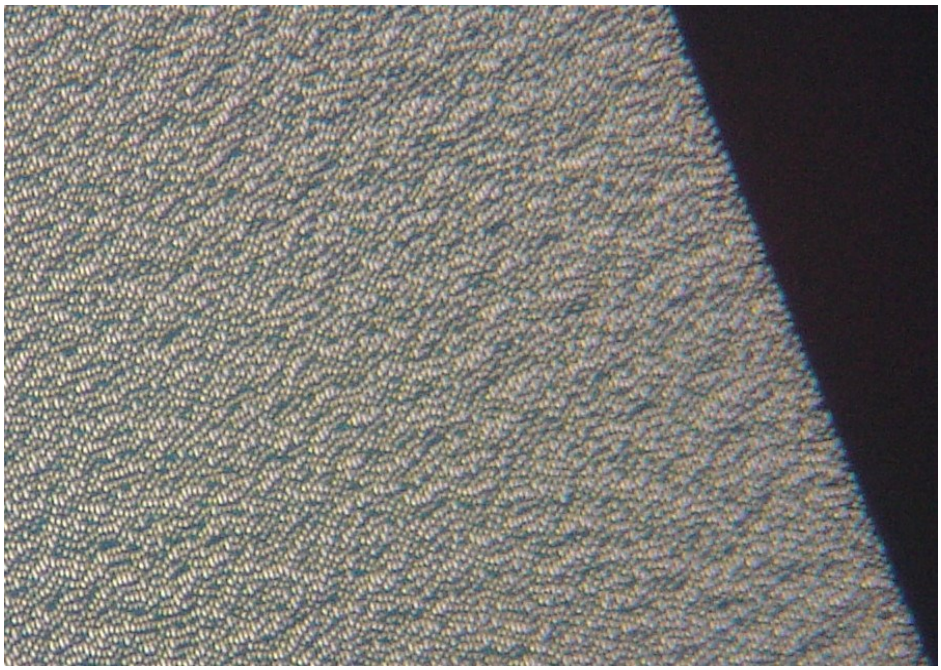
The RCM is also used by Kallberg (2013), Borglund and Anderson (2011) and Borglund (2008) as a theory to understand information loss and change in archival practice. We will use the RCM together with an analysis of how the shift away from RLM has taken place out in real-world archival practice. This will serve as a framework for interpretation of the digital preservation work that the organizations perform and to discuss the challenges of information loss.

4.2 The ontology of digital phenomena

In "Overview of technological approaches to digital preservation and challenges in the coming years", Kenneth Thibodeau (2002) describes an ontology of digital information which may be helpful in understanding digital preservation. He claims that digital information has three aspects: physical, logical, and conceptual (Thibodeau 2002, p. 6).

Thibodeau explains that, in order for digital information to exist, the binary code must be inscribed onto a physical information carrier, such as a hard drive on a computer. Without inscription the digital code cannot exist. This is what Thibodeau means by the physical aspect of digital information (Thibodeau 2002, pp. 6-7). In colloquial speech, the digital realm is often described as being non-physical, but digital information is physical, even if it may seem otherwise. Binary code inscribed on optical storage media takes the topographical surface of pits and land (see Fig. 2).

Fig. 2. Image of the physical pits and lands on a CD-ROM.



Source: Kennerly, J. (2007). Licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license.

Thibodeau points out that there are many different ways to inscribe a digital object, which depend among other things on which kind of media is being used. In order to be able to read these physical bits, they must have an internal logic which can be interpreted by software. This is what Thibodeau calls the logical aspect. The logical aspect of digital information can be structured in different ways, with the result that not all formats can be read by all software. An example of this is when a user cannot open a document on a computer because they lack the appropriate software (Thibodeau 2002, pp. 7-8).

Another aspect of digital information concerns when it is recognized and understood by a user. This, according to Thibodeau, is its conceptual aspect. If we imagine a digital photograph whose code is inscribed on a CD-ROM, it follows a logic which must be interpreted by software. However, a user will still be unable to take part of the information if it has not been presented in a way which makes the information available to the user's perception. Though this can occur via other human-computer interfaces, it is most commonly done via a screen. Thibodeau means that the recipient of information need not necessarily be a person – it can even be another computer. The conceptual aspect is described by Thibodeau as "the object we deal with in the real world". This can be understood as information which carries meaning (Thibodeau 2002, pp. 8-10).

In order to take part of a digital object, the physical bits must first be localized and then interpreted by a logic to be presented conceptually. Thibodeau concludes that the preservation of digital information is foremost a matter of preserving the possibility to reproduce the conceptual information. If it is no longer possible to recreate the information on a conceptual level, then one has not successfully preserved the information. In order to take part of the conceptual information, one also needs the physical and logical information. Furthermore, Thibodeau means that the preservation of digital information can seldom be accomplished without altering the information in one way or another. With time the hardware fails, requiring the information to be moved from one information carrier to another. Eventually the software also becomes obsolete, so the information must be moved over to a new format which can be read. However, this need not affect the conceptual aspect (Thibodeau 2002, pp. 12-13).

When working with the preservation of physical objects – a book for example – changes will have an effect on preservation and authenticity. This is not the case in digital preservation, according to Thibodeau. He claims that digital preservation is actually contingent upon these changes, due to hardware which fails over time, obsolete formats and technical developments. In order to judge which changes can be made without affecting preservation and authenticity, Thibodeau claims that the intention of the preservation must be identified. Through identification of what the 'essential respects' are, one knows which changes can be made. By essential respects, Thibodeau means those qualities and relationships which cannot be changed without altering the defining characteristics of the information. Take an older digital text document for example, one that was originally displayed as green text on a black background, and is now reproduced as a modern PDF with black text on a white background. One could claim that the essential respects have been preserved, if it is the information in the text which has been identified as essential for preservation, and not the visual environment. Thibodeau summarizes that "(one can only preserve the ability to reproduce the document." In order to find the ideal way to preserve a certain kind of digital material, he claims that one needs to identify the three aspects of the information and what the essential respects are (Thibodeau 2002, pp. 12-14).
This ontological understanding of digital information will be used in order to analyze the challenges struggled with by archives regarding digital information loss, and to understand which aspects risk being or already have been lost. We will also consider how responsibility is allotted for the aspects of digital objects.

4.3 The SPOT model for risk assessment

> Ultimately, digital preservation is an exercise in risk management.
>
> *Corrado & Moulaison 2014, p. 68*

The Simple Property-Oriented Threat (SPOT) model for risk assessment is intended to function as a diagnostic tool for those involved in digital preservation tasks. The model was developed in response to the authors' quest for a simple, effective diagnostic tool which could be scaled to handle different sizes of repositories involved in various phases of digital preservation. It has been developed by Sally Vermaaten, Brian Lavoie, and Priscilla Caplan (2012) and was meant to be an

alternative to other threat assessment models such as Drambora, which covers a large variety of technical and organizational threats but requires a large investment of time to carry out (Vermaaten, Lavoie & Caplan 2012, p. 11). As a more user-friendly alternative to such models, SPOT can be utilized in a broader range of scenarios.

The authors divide threats towards digital preservation into two categories: threats against content and threats against organization. They specify that their model deals only with the first kind of threat: threats towards content. This means that a complementary model would be necessary in order to handle the organizational threats to digital preservation. While the SPOT model alone cannot be used to analyze threats to digital preservation at a repository, it has the potential to reduce the number of factors which require analysis down to a manageable level. An institution can take up a different kind of analysis for the organizational issues and then collocate the results. The authors encourage the development of a similarly simplified model for the identification of organizational risks (Vermaaten, Lavoie & Caplan 2012, pp. 1-2, 12).

The SPOT model identifies key properties of the content of digital preservation and the kinds of threats each one might face. It is meant to identify risks within an institution's own digital preservation strategy, or to help create a set of criteria to determine whether a digital preservation solution offered by a third party is suitable. SPOT can be used to identify threats which have already been realized or used for ongoing audits of a system (Vermaaten, Lavoie & Caplan 2012, p. 2).

The model is grounded in a review of the literature surrounding risk assessment and digital preservation. This review was specifically focussed on models meant to aid in the analysis of threats to digital preservation. The authors carry out a comparison between other threat models, categorizing them according to four qualities: "conceptual clarity, appropriate detail and consistent granularity, comprehensiveness, and simplicity". As the authors did not find any of the reviewed models to contain all four of these qualities, they decided to construct their own (Vermaaten, Lavoie & Caplan 2012, p. 2).

The authors claim that there are six essential properties for digital preservation, which must preserve the following properties:

- Availability
- Identity
- Persistence
- Renderability
- Understandability
- Authenticity

The threats against these properties are described as being outcome-oriented. By this, the authors mean that they do not attempt to locate the origin of the threat in their model, but instead deal only with the threat itself. It is up to the repository to decide how to deal with the source of the threat (Vermaaten, Lavoie & Caplan 2012, p. 7).

What follows is a list of the essential properties and the threats attributed to them:

Availability, the first of the properties in the list, may be the most fundamental. It is concerned with whether the material has even been identified for long term preservation. The threats that endanger this property are equally basic: the information may have degraded to the point that it can no longer be restored, it may never have been preserved at all (intentionally or unintentionally), it may be missing, or it may only have been partially preserved (Vermaaten, Lavoie & Caplan 2012, p. 8).

Identity is a property based on difference. This difference singles one digital object from others and therefore makes it possible to locate. The property can be contextual: requiring, for example, contextual information to differentiate between two objects with the same name. Failures to preserve the proper metadata, incorrectly linked metadata, or inaccessible metadata are the threats to this property (Vermaaten, Lavoie & Caplan 2012, p. 8).

Persistence is concerned with how well the digital object maintains its integrity over time. One threat to persistence is anything environmental which may cause premature deterioration to the storage media where the information is located. Media obsolescence and/or non-availability of the necessary hardware are two other threats relevant to this property. Intentional or accidental destruction of information also be a factor in a SPOT analysis of threats to persistence (Vermaaten, Lavoie & Caplan 2012, pp. 9-10).

Renderability is concerned with the representation and interpretation of digital objects for an end user. In order for a digital object to be renderable, its key aspects must be able to be reproduced so that the object will function as intended. These are called 'significant characteristics' in the model and must be identified and preserved. If the necessary hardware or software is unavailable, this is a threat to renderability. Lack of information about the format of the digital object, or an inability to determine the relevant details concerning the original rendering environment are also considered threats by the SPOT model. Another threat to the property is a failure to preserve the features which enable users to interact with the object (Vermaaten, Lavoie & Caplan 2012, p. 9). The authors of the model point out that the preservation of digital information is fundamentally concerned with the preservation of an object's 'significant characteristics' (Vermaaten, Lavoie & Caplan 2012, p. 9).

Understandability is mostly concerned with what other kind of information outside the primary object may be necessary in order to understand it. This includes all kinds of contextual information that may contribute to understanding. Threats to this property include an insufficient amount of support material, or a failure to maintain such supplementary information. The authors point out that such information can generally only be made understandable for the information's designated community of users. This means that one must identify who the potential users are and then save the supplementary information needed by these groups. A risk is that one fails to identify the needs of certain groups or misidentifies the designated community (Vermaaten, Lavoie & Caplan 2012, p. 10).

Authenticity entails the ability to verify that a digital object is what it claims to be. Threats to this include a failure to include metadata relevant to authenticity, which may not have been included at the object's creation or the beginning of the preservation process, intentional manipulation of such information, and undocumented changes to a digital object made during the archive's custodianship (Vermaaten, Lavoie & Caplan 2012, p. 10).

We believe that the concept of 'significant characteristics', which is named in the description of the property Renderability, may have much in common with Thibodeau's concept of 'essential respects' (see section 4.2). The authors of the SPOT model clarify, like Thibodeau (2002, pp. 12-14) that the digital preservation need not require a strict preservation of an object's bit stream, which can be changed without affecting an object's 'significant characteristics'/'essential respects'. As an example, a PDF document is not appreciably changed from the user's perspective when it is converted to a PDF/A. When such a conversion is made, the content of the document is not affected in any appreciable way for the user, but its prospects for successful preservation have been improved (Vermaaten, Lavoie & Caplan 2012, p. 9).

We will use this model in order to identify risks concerning content within the institutions' digital preservation efforts and to improve our understanding of the risks they themselves identified. The authors of the model state that it can be useful in mapping the threats against the OAIS Reference Model. The OAIS model is a commonly used to map out the organization of digital preservation (Vermaaten, Lavoie & Caplan 2012, p. 15). Following this advice, we will utilize SPOT together with the OAIS model in order to interpret our results (see section 4.4).

4.4 The OAIS Reference Model

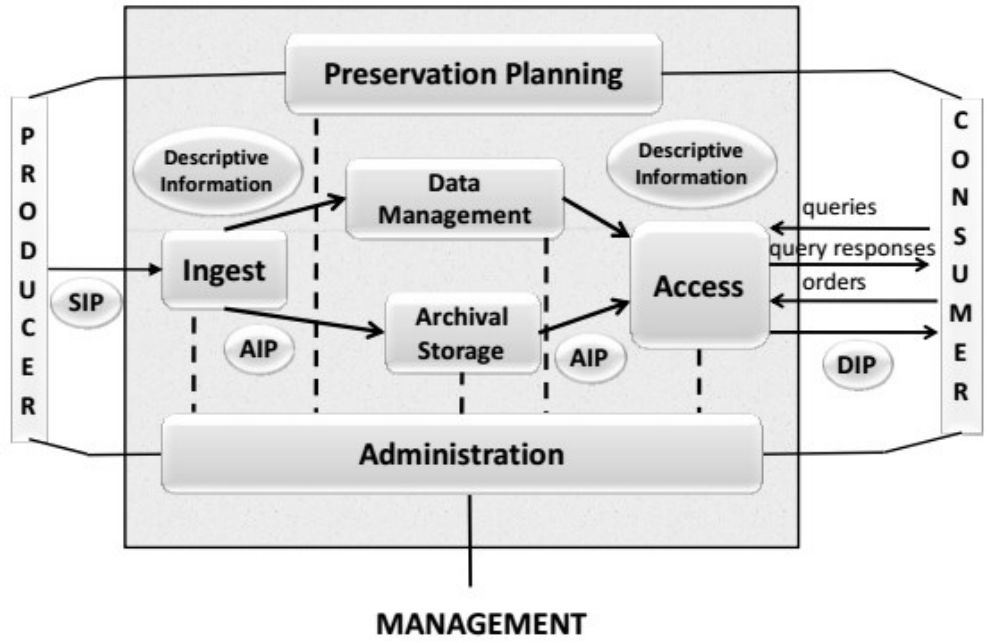
The Open Archival Information System (OAIS) Reference Model was originally developed to support the long-term preservation of information (see Fig. 3). It was developed due to the need to protect information threatened by technological obsolescence. At its foundation, OAIS is a model for an archive: it describes a framework, archival functions and provides well-defined concepts necessary for long-term preservation (CCSDS 2012, ch. 1, p. 1). The Consultative Committee for Space Data Systems (CCSDS), the author of the 2012 version of the OAIS model, defines the OAIS as:

> An OAIS is an Archive, consisting of an organization, which may be part of a larger organization, of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community.
>
> *CCSDS 2012, ch. 1, p. 1*

The OAIS model is an abstraction: it is not limited to actualization in any one form. It can be instantiated in different kinds of physical archives, and with different hardware and software (CCSDS 2012, ch. 1, p. 2). As noted earlier, the model is commonly used when talking about the organization around digital preservation (Vermaaten, Lavoie & Caplan 2012, p. 15). The discussion of OAIS in this study will be

constrained to fit the space available: we only take up its most basic aspects together with a few others that we feel are a particularly good match for analyzing the situations we encountered.

Fig. 3. The OAIS Reference Model.



Source: CCSDS 2012, ch. 4, p. 1.

The four major components of the OAIS system can interact in many different ways.

1) **Producer** - something or someone which creates the information to be preserved. Among other things, this can include persons, companies, and other archives. In OAIS terminology, submissions are made by producers.

2) **Manager** - person or persons who set policy for an archive.

3) **Archive** - an actualized organization preserving information that serves a designated group of users.

4) **Consumer** - someone or something who interacts with an OAIS to access its preservation content. In OAIS terminology, material is distributed by the archive to a consumer.

(CCSDS 2012, ch. 1, pp. 9-14; ch. 2, pp. 2-3).

OAIS also deals with Information Packages, which are what the archive preserves. These Information Packages can be made up of two kinds of information, the inclusion of both of which is optional:

1) **Content Information** - the original target of preservation, in our study the digital object and any additional material necessary to interpret it.

2) **Preservation Description Information** (PDI) - refers back to the Content Information and is necessary so that contextual and identity information about it is preserved.

(CCSDS 2012, ch. 2, pp. 5-6).

The OAIS model makes it a point to differentiate between Information Packages that the archive preserves and Information Packages that the archive receives or distributes. This is because they may be structured in entirely different ways and because one may not contain the same components as the other. There are three varieties of Information Package:

1) **Submission Information Package** (SIP) - what the archive receives. The form and content of these are generally decided upon together by the producer and the archive.

2) **Dissemination Information Package** (DIP) - what the archive sends out.

3) **Archival Information Package** (AIP) - what the archive preserves

(CCSDS 2012, ch. 2, pp. 7-8).

The most commonly reproduced diagram from the OAIS model shows the functional entities of an archive and the information flow between them (see Fig. 3).

OAIS systems contain six functional entities:

1) **The Ingest Functional Entity** - this entity receives SIPs from producers or from the administrative workings of the archive itself and readies them for preservation.

2) **The Archival Storage Functional Entity** - concerned with AIPs and deals with storing them, maintaining them, and retrieving them.

3) **The Data Management Functional Entity** - concerned with describing AIPs and administrative information, making sure they are properly grouped and also issues of successful database functionality.

4) **The Access Functional Entity** - concerned with making sure that all aspects of the consumer's finding and using information in the OAIS function as they should. This includes requests and delivery together with access restrictions.

5) **The Administration Functional Entity** - concerned with the general function of the archive, this entity deals with the relationships between producers, consumers, submissions, and standards. It also includes dealing with questions

of archive policy. It is also concerned with overseeing the archive's infrastructure and issues of migration.

6) **The Preservation Planning Functional Entity** - primarily directed towards the functioning and maintenance of the OAIS itself, ensuring that it can fulfill its primary functions in the long term. It keeps an eye on technological obsolescence and manages the measures to counteract it, keeping its holdings understandable for its community.

(CCSDS 2012, ch. 4, p. 1-3).

In this paper we use the OAIS model to map out the workings of the archives we interviewed. The explicitness of the model may help to clarify where things go awry in the digital preservation process. Since the SPOT risk assessment model only addresses technical issues, we will use the OAIS model to complement this and deal with the organizational aspects of digital preservation. The OAIS model will also aid us in creating a visualization which shows where information loss occurs.

5 Method

In this section we will present the method and structure of our study. This will include the choices we have made, the material we have gathered, and how the results will be analyzed. The section concludes with a discussion of the study's limitations and ethical considerations.

5.1 Procedure and selection

In order to be able to examine the occurrence of digital information loss or the risk of digital information loss, we have decided to perform interviews at three different Swedish archives. Each of these archives differ from one another in the kind of material they preserve. Two of them are involved with maintaining public records. These interviews will help us to obtain information directly from archival institutions, who have firsthand experience of the issues. We have determined that the scope suitable for this research is three archival institutions. There are many different kinds of archives, each of which face different challenges. Due to the constraints of this study, the inclusion of more than three archives would necessitate a shallower analysis.

In order to gain access to the information we wish to analyze we will conduct interviews with archivists and IT-personnel at the chosen archives. Ideally, we will conduct interviews with those employees who have a long work history at the selected institution, since they will be more likely to be in the know about digital information loss in the workplace. The interviews will be conducted in Swedish and take place at the archives.

The answers provided by these informants will be anonymized in order to prevent the identification of their workplace or themselves. The archives will be kept as anonymous as possible: it is not of interest to name them since this study is not about the archives themselves, but rather digital information loss. Loss of information can also be a sensitive question, which is another reason to keep the archives anonymous. The archives will be referred to as: Archive A, Archive B, and Archive C. Anonymity will be assured by allowing the institutions the opportunity to read a proof of the results and provide their input into how successful they feel their anonymity has been maintained.

Archive A and B are both part of a host organization and have more than 10 employees. Archive C is autonomous and has less than 10 employees. What is an archive and what is part of a host organization can be a difficult line to draw. In this study we have chosen to define 'archive' as an organization that has an archival

function as one its main tasks. Both Archive A and B are separate from the rest of the organization, having their own name and structure. At the same time they are following the decisions made at different parts of the organization. We chose to define them as an archival organisation and will not include interviews with people in the host organization.

All the informants at Archive A will be given names that begin with an A, all the informant at Archive B will be given names with a B and Archive C's names will begin with C. Our hope is that this will make it easier to follow the study's results and analysis. The names used will be: Abby, Albert, Bob, Charles and Cathy. The names are not gender specific. The variation in the number of persons interviewed at the archives is based upon to what degree our questions were answered in the initial interview.

5.2 Interviews

In order to immerse ourselves in the challenges of digital information loss, we first needed to sketch an outline of what the institutions' digital preservation work looked like. The interviews occurred between January and April of 2016 and were semi-structured, containing 14 core questions. As the archives face different challenges, and their circumstances differ greatly from one another, it can be expected that they will bring up different aspects of digital preservation which are not included by the others. Other questions will be added as needed, based on the answers provided by the archives. We consider the inconsistencies which arise during the interview process to be unproblematic since the intention is to show three different archives and their experiences of digital information loss or the risk thereof. While we do not intend to make a direct comparison between the archives, we will compare them insofar as it helps to highlight different aspects of digital information loss. The difference which arises during the interviews will be understood as an expression of how digital preservation works at each of the three archives.

The interview questions presented below are intended to help us see the big picture of an institution's involvement with digital preservation, as well as to provide us with an understanding of digital information loss at that particular institution.

Interview questions:

1) What kind of digital material do you archive?
2) Where is the digital material stored in your collections? (Floppy disks, CD-ROMs, DVDs, USB sticks, external hard drives, computers, servers, cloud services? An OAIS-based system for long term digital preservation?)
3) Do you have digital catalogues for your material?
4) What is the oldest digital material your institution has (that you are aware of)?
5) What kind of format is your digital material saved in?
6) Do you currently have a plan for digital preservation such as guidelines, strategies, or best practices that you follow?
7) Do you have a specific plan at your workplace for how digital material will be preserved?

8) Have you ever been made aware of digital information loss? (For example, loss when information was migrated, damaged search paths, cloud or other services that went out of business, physical media that isn't possible to read any longer? Is there any material that you are aware of which is generally considered abandoned / too difficult to recover?)
9) As far as you know, has your institution ever experienced a cyber-attack?
10) How long does your institution intend for your digital resources to remain useable? (Is there a specific time-frame?)
11) Do you have any strategies to keep pace with technological development?
12) What kind of metadata do you save with your digital objects: keywords, format, hardware, and/or system information? (Are there criteria somewhere for what kind of metadata your institution saves?)
13) Are there any copyright, patent or other intellectual rights that might affect the ability for you to preserve your digital objects, i.e. in connection to the migration of digital material?
14) How do users take part of your digital information? (Do complications arise with access, issues of copyright infringement or the dissemination of copyrighted digital information?)

The interview questions are formulated in order to indirectly access the information we want to learn about by breaking down the big questions about digital preservation. We believe that only asking the organizations about what kind of information loss they have experienced will not provide us with a clear understanding of the situation. This is because digital information loss can be understood in different ways and also because there are different understandings of what constitutes digital preservation. As previous research shows, it can be difficult to clearly identify what constitutes a record in a digital context. This inability to clearly identify a record makes it more difficult to answer the question of one has experienced information loss. Therefore, we have chosen to break down the larger questions into smaller, more concrete ones concerned with digital information loss.

As described in section 1.2 we also include data when we write about information, and with records we mean the information that is intended to be preserved by an archive. Other information which is also of use will be taken up in the discussion. Another reason for this is that Swedish archival science does not maintain the separation between recordkeeping and archives which exists in other countries (Kallberg 2013, p. 109). In summary, while we will be focussing on long-term digital preservation at these archives, we will not be excluding relevant information from short or medium-term preservation. This means that the study will also be able to deal with work documents and other material. As the research carried out by Borglund and Anderson (2011) shows, a holistic approach is more informative when it comes to information loss, since this can occur before an organization manages to identify the digital objects which should be preserved in the long-term. This means that we examine the digital information at an organization that is meant to be preserved, together with some digital information which has yet to be targeted for preservation. Two of the archives maintain public records, and their experiences can tell us something about the situation concerning the preservation of digital public records in Sweden today.

5.3 Results and analysis

In the results section we will review the most important aspects of the interviews. The results emphasize the information from the interviews that is pertinent to an understanding of information loss. The archives will be presented one at a time. The analysis section will utilize the interpretational frameworks we have chosen in order to analyze each archive in turn, and it will conclude by discussing them together. We will use the Records Continuum Model to analyze how archives have transitioned to proactive digital preservation work. Afterwards we will use Thibodeau's ontological understanding of digital information to look at which aspects have been lost or risk being lost. The third part of the analysis considers the technical aspects of information loss using the SPOT risk assessment model, and the fourth section examines what are primarily organizational issues using the OAIS Reference Model.

5.4 Limitations

This study will only include examples from three archival institutions where digital information is preserved and digital information loss has occurred or risks occurring. The results will be limited to focus on information loss and its risk at these specific institutions. Moreover, the examples, consequences, questions and struggles of these three archives also tell us in general about the situation with digital preservation in Sweden today. The laws governing public records differ from country to country, which also make this study country-specific. However, the challenges faced may likely be found in other countries.

A potential limitation of the study is the truthfulness of the informants and the risk that they may be unaware of digital information loss at their institution. An institution may have several personnel dealing with digital preservation, so we need to identify who is most capable of providing us with the most comprehensive set of answers. It could be the case that the employee who has been most involved in digital preservation has recently been replaced. If no succession plans were in place, this might mean that the new employee is less informed about digital preservation issues of the past at their workplace.

In the interviews, there will be people with different levels of knowledge who can provide different insights into their organizations and technology. There could be other examples of digital data loss which do not come up during these interviews. It is beyond the scope of this study and would likely be impossible to map out all the digital information loss at the three archives. Instead, it is our goal to sketch out problem areas and show how information loss is actualized.

A potential weakness of this study is the question of how can one know they have lost something they never knew they had. Knowledge of systems and digital holdings is not always passed along with personnel changes. If there is digital information which remains improperly registered, the knowledge of it might not be transmitted, and could be seen as lost – but without the knowledge that it is lost.

5.5 Ethical considerations

The main ethical consideration affecting this research is the question of anonymity. We hope that anonymity will assist us in getting the answers we need from archivists and IT-personnel in order to conduct our study. We chose to make the institutions anonymous in order to avoid the risk of anyone being labelled a whistleblower, and to avoid pointing out any particular institution. There are unresolved digital preservation issues throughout the archiving sector. These questions do not have simple answers. We hope that the informants will be as forthcoming as possible with their answers. In order to ensure their anonymity, we have provided the opportunity for the interviewees to read the results section before it is finalized.

6 Results

The results from each archive are presented in this section below. Each archive is discussed in its own section and are presented in alphabetical order. The structure for each archive is similar, but varies according to which questions were emphasized during the interviews. This study is meant to highlight the struggles with digital information loss at the three archives, and any comparisons between the archives are made in order to discuss the larger digital preservation issues.

6.1 Archive A

6.1.1 Digital holdings and information flow

Abby, one of the two staff members interviewed at Archive A, estimates that the archive currently maintains around 700 GB worth of digital material which is kept in a e-archive and on local servers. Due to the fact that public records in Sweden are official as soon as they are produced, Archive A is legally required to take care of the analog as well as the digital public records which fall under their custodianship. Such digital records may take the form of any kind of media. This is true both for Archive A and the host organization to which it belongs. As Archive A serves in an archiving capacity for this host organization, it is the archive's assignment to preserve the digital public records produced by it. However, the archive is currently unable to accept submissions from the host organization in digital form. Archive A also preserves archives of private individuals, businesses and associations, which we refer to as private archives.

Abby explains that the Archive A is currently undertaking efforts to make digital submissions of public records possible. She expects that there will be more variety to these submissions than just the digital versions of paper documents; they may consist of programs, databases, and audiovisual material. Regardless of the kind of digital material, its preservation may be required by law.

While the archive is not currently capable of accepting digital submissions, there are exceptions to this: a few smaller digital record submissions have been made to the archive. These include public records kept in information systems which have been phased out by newer versions. The lack of a preservation strategy has placed the information in these systems at such a high risk that the archive decided to take emergency action to preserve it. When the archive has saved such systems, it is the information contained in them rather than the systems themselves which have been

saved. This information is then stored in an open-source coded structure, and preserved in an information system modelled according to OAIS. Archive A calls this OAIS-based system an "e-arkiv", or electronic archive. Abby explained that while it is possible to save the information in these kinds of older systems, such interventions demand a great deal of resources.

The use of open-source based software, which in this instance has been utilized in order to preserve the information in these old systems, is a strategy the archive intends to pursue as much as possible in the future. Archive A also tries to follow sustainable format choices for long term digital preservation. These formats currently include XML for text, PDF/A for many kinds of documents, and TIFF or JPEG for images. The archive has also undertaken digitization projects in order to facilitate better access for its users. These projects are intended to increase searchability, improve accessibility, and improve the overall chances of successful long term preservation.

During the beginning of the first interview we were only able to learn about specific digital projects. This led us to indicate our interest in hearing about all the digital material held by the archive by asking the question: "What material does the archive have in binary code which you want to preserve over time?". This led to a discussion which yielded new information about local work documents, registers stored by an information system outside of the archive's control, and digital public records produced at the archive. We also learned that digital information on CD-Rs and other loose media might be stored in archive boxes in the private archives. Abby does not work with this material, and could not confirm this. We therefore decided to interview another staff member, Albert, who was in a better position to tell us about deliveries to the private archives.

During the interview with Albert described the process of delivery of private archive collections, which originate from private persons, businesses and associations. He explained that it was not uncommon for digital material such as CD-ROMs, DVDs, and USB sticks to arrive with such deliveries of what were otherwise analog material. He also told us that there are no routines to follow when dealing with such material. He explained that private archives are not a high priority, and that the majority of resources are allocated towards maintaining public records. Therefore, it can take some time before the private archives are packed up, and even longer before they are registered. When these archives are registered, this is usually done by interns. If a CD or similar information carrier is discovered, it may be registered – but rarely if ever are the contents of such media themselves registered. Whatever data exists on such storage media stays with the archive boxes in the physical archive.

Tracking the digital information at Archive A, we mapped out who it is produced by, where it is stored and who is responsible for maintaining it:

1) <u>Digital public records created within the archive</u>
   Producer: Archive A
   Stored: On servers at the host organization, outside Archive A's e-archive
   Maintained: By the host organization

2) Digital public records which have been saved through emergency actions
   Producer: The host organization
   Stored: On servers at the host organization, inside Archive A's e-archive modelled after OAIS
   Maintained: By the host organization

3) Analog public records digitized by the archive
   Producer: The host organization (original), Archive A (digital copy)
   Stored: Locally on the Archive A's servers
   Maintained: By Archive A

4) Digital material in the analog private archives
   Producer: Private persons, businesses and associations
   Stored: On different information carriers in analog private archives
   Maintained: No, only stored

5) Analog material in private archives digitized by the archive
   Producer: Private persons, businesses and associations (original), Archive A (digital copy)
   Stored: Locally on the Archive A's servers
   Maintained: By Archive A

Archive A utilizes an information system which houses some of its public records. The responsibility for this system and its servers is not under Archive A's control and has been excluded from this study.

6.1.2 Responsibilities for digital information

We learned of two instances where the responsibility for digital preservation was shared between the archive and its host organization. While the host organization was responsible for the servers and backup in both of these cases, it was the archive who was responsible for keeping the information readable. The archive maintains its own servers for those digital materials which do not require higher security measures. Abby expressed concern over the lack of transparency and communication with the host organization, and was particularly concerned with how it takes care of it servers. She was uncertain about the quality of their backup, and what kind of hardware they used. Abby explained that she is working on improving contact with the IT division at the host organization.

There are several persons involved in digital preservation at Archive A, including digital archivists and the personnel responsible for oversight. Abby stated that the people dealing with the preservation of digital material at Archive A are relatively few compared to the number of staff members who produce it. Maintenance and security issues for digital preservation are managed by the IT department of the host organization. Due to the issues with security and compliance with personal data integrity laws, most of the sensitive information is stored at this level, ensuring that the people who have the most knowledge about security are in control of it.

The various kinds of digital information created by the archive itself were discussed during the interview with Abby: video material, sound recordings, and other material which constitute public records. There is currently no way of knowing how this material has been maintained over the years. Abby explained that part of the problem was a lack of knowledge concerning this kind of information, and she clarified that the identifying material as a public record or knowing how to take care of it were not always crystal clear issues. She also stated that it was not really possible to say to what degree data loss has occurred in these records, and that the solution was a more active preservation strategy.

Both analog public records and analog private archives which have been digitized by the archive are maintained solely by Archive A. The archive is responsible for both hardware, backup server and readability of this material. Digital material in the analog archives, listed under number 4, is not actively taken care of.

In general, Abby felt that the greatest challenge for digital preservation at her archive involved organizational issues rather than technical ones. One example of this we encountered was that while some staff members believed a certain kind of storage media to be unreadable, others knew this not to be the case, and that there exist inexpensive hardware solutions which can access the data

6.1.3 What is an e-archive?

While some digital information was stored in what Abby called an e-archive, other digital information was "just" on servers. When we asked Abby to define what she meant by 'e-archive', she explained that it is a system which includes an organization set up around it, with rules, plans for culling records, and backup. Abby explained that the e-archive at Archive A is an OAIS-based system with its own interface and serves as both a way to collect and to preserve information.

Abby stated that it was sometimes unclear what specific properties and functions were required in order to constitute an e-archive. She took up the example of another system maintained by the archive, which only houses one specific collection. While this system also has its own interface together with an organization around it regarding backup and rules, making it sound very similar to an e-archive – it does not fall under what Archive A calls an 'e-archive'. Abby continued to say that when one talks about e-archives, people mostly think about the OAIS-model. However, she said that this model is very abstract and does not say much, if anything, about the practical aspects. Abby posed the rhetorical question of whether the OAIS model is best seen as a model for an information system or if it should instead be understood as a model for an organization which works with digital preservation. Abby speculated that the distinction between an e-archive and the digital information outside of it would likely become less rigid in the future: with time it will all just fall under the umbrella term 'archive'.

One difference Abby did see between e-archive and the digital information kept outside of it was the attention paid to security and controlled access. This is the reason why the e-archive is not maintained by Archive A, but by the host organization, who has the responsibility for security and restricting access to the appropriate users. While talking about security we asked Abby if she had any

knowledge of cyber-attacks against the archive. She answered that there had been DoS[4] attacks against the host organization, but nothing seems to have been directed specifically at the archives themselves.

6.1.4 An example of information loss: Issues with format migration

One concrete example of data loss we were told about at the archive involved a PDF/A converter tool, which was used to convert documents from their original formats to PDF/A. Most of the files were converted without errors, but there were some which were less successful.

We have separated these into three types:

1) The file was saved as a locked or protected PDF which did not allow for conversion.
2) The file was in an unusual format that the conversion program was unable to recognize.
3) The file was originally saved with an incorrect file extension and could neither be opened nor converted.

The archivist can test the error in the third type and eventually figure out the correct file extension, but such guesswork is time consuming. Abby states that files of the first and second types will be possible to read so long as there is a program to read them. An example of a format which has caused difficulties is Word 97, which has not always converted successfully to PDF/A. Abby also stated that no further migration will be possible with the files of the first and second type. Furthermore, she mentioned that a different conversion program could potentially be able to read a greater number of formats, and that this may prevent such material from being lost. The files of the third file type are already considered lost and cannot be opened without guessing the correct file extension, renaming them, and testing for functionality. Abby speculated that altering formats might potentially be seen as having an effect on their authenticity. Obviously this is a non-issue if the alternative is that they remain unreadable. Abby reiterated here that she believes the problem of outdated formats to be less significant than the need for proper management and organization within digital preservation. In this context, Abby added that she believes there to be an excessive amount of trust placed in technical solutions, of which format standards like PDF/A are an example.

Abby also mentioned that other parts of the host organization have experienced information loss. As Archive A is responsible for overseeing these materials, Abby has taken an active role in helping to prevent such losses from occurring in the future.

6.1.5 Strategies for digital preservation management

When we discussed succession planning, Abby mentioned that an insufficient amount of time had been allowed for this process when she took the position. The information

[4] DoS or Denial-of-Service attacks are intentional disruptions of service which flood it with request and block legitimate users from accessing it (Wikipedia 2016d).

system she inherited from her predecessor was an ad-hoc solution adapted to budgetary and other constraints which existed at the time. After her predecessor moved on to her new job, Abby maintained informal contact with them, learning as much as possible about the existing solutions. This reliance on knowledge that resides only with individuals rather than in written plans was characterized by Abby as a risk for the archive.

It came to light that the archive does not have a general strategic plan for digital preservation. While there were such work plans in the past, these have become outdated. Abby stated that the institution's current strategy is to figure out which programs are in most need of care and focus on them first. However, Abby said that it would still be for the best if such information was written down. There are currently not enough personnel to carry out all of the necessary efforts for digital preservation at Archive A, according to Abby. When we asked Abby if she thought the digital material at her archive would be readable in 400 years, she stated that the material at Archive A would remain readable as long as it was maintained.

Finances were also a concern for Abby, who said that project-based funding had been the norm up until now, and that a shift in perspectives towards digital preservation would be necessary in order to raise its profile within Archive A. She was concerned that a lack of sufficient resources could lead to digital information loss in the future. This was discussed in the context of digital preservation's need for constant funding. In response to a question concerning whether or not the archive might need to be more restrictive regarding what digital material it takes in, Abby answered affirmatively. The concern was that the archive risked taking in a greater amount of digital material than it had the resources to preserve. Albert mentioned something similar to this sentiment, stating that while archival registration had been carried out with an extreme attention to detail in the past, the focus today had shifted more towards a strategy of collection en masse. Archive A does not see how it could handle a reduction in resources for digital preservation without digital information loss ensuing as a result. Apart from financial resources, another cause for concern was the effects of insufficient planning for digital preservation. As an example, Abby pointed out that if the host organization's information systems had been written in open source from the beginning, the archive could instead be focusing on current digital preservation issues rather than rewriting old code.

6.1.6 Cloud services, copyright and personal data integrity laws

Abby stated that the archive does not use cloud services due to the legal risks for sensitive information. The regulations for personal data integrity preclude the usage of cloud services for storing potentially sensitive material. She mentioned that other, similar archives have used these kinds of services, and have received criticism due to concern that information can end up in the wrong hands. Abby also expressed concern that if a cloud service goes bankrupt the information may be lost, or in the very least, expensive to get back.

Regarding the question of copyright and how it may hinder preservation and access, Abby suspected that such issues must exist, but was uncertain as to what they were. This uncertainty was due to a lack of documentation. Information pertaining to the rights of certain materials seemed to only be known to certain employees rather than

being written down. No overview of the various copyright agreements exists. There may exist conflicting interpretations of agreements regarding ownership of different aspects of information between system suppliers and the archive.

When asked about how users take part in the archive's digital material, the answer was: as of now, they don't. This was seen as a possibility in the near future for some digital holdings but the archive preferred to err on the side of caution when it came to adherence to the personal data integrity law. This meant that allowing access outside the archive to many digital materials was unlikely. It was also unclear if the archive could digitize and send out certain materials in PDF form, even if a user were to request them. The archive cannot ensure the security of an email from both the sender and the receiver side, and therefore preferred to use paper copies as a way to comply with the law. Concerning the other digital material, the only certain way to follow the law was to only allow local access to it at the archive.

6.2 Archive B

6.2.1 Digital holdings and information flow

Archive B is responsible for the maintenance of public records in much the same way as Archive A. They are legally required to ensure that their material is maintained over time and kept accessible. Another similarity between the two archives is that Archive B also accepts deliveries of private material. When Archive B takes in records in digital form, these are taken care of by the IT division at the host organization. These digital records include text documents, images, audio, video, and IT systems. The majority of material dealt with by Archive B is analog, but digital material is occasionally delivered as well.

The IT division's digital preservation system is based on the OAIS model. Bob, who we interviewed at Archive B, stressed the importance of registering digital and analog materials together when they belong to the same archive, even though the analog material goes to Archive B and the digital material goes to the IT division. He explained that there had previously been problems where such materials had been ingested via these respective routes, yet had not been linked to one another. This meant that they were not necessarily registered together, which made it difficult to locate all the material belonging to an archive. This also resulted in a situation where neither division was aware of the other's holdings – holdings which really belonged to one and the same archive. This problem has now been solved by creating a single consistent path for both kinds of material to enter the organisation.

When asked about how much material the IT division maintained in digital holdings, Bob responded that the organization as a whole has taken in approximately 2,5 TB's worth. This amount is not larger due to the fact that it is mostly made up of text. The organization also produces its own digital information, which predominantly consists of digitized documents in the form of high quality TIFF files. This material takes up a much greater amount of space: approximately 1 Petabyte, or twice that amount if the backup is included.

Regarding the digital public records produced by the archive itself, Bob explained that these are not stored digitally. In this situation, Archive B has chosen to print out copies of the records it produces – such as emails – in order to preserve them. The archive is currently working on a solution to manage this material digitally.

We also discussed other situations where digital storage media is taken in with analog material. This digital material may be part of the public record or it may originate from private archives. One of the first things that Archive B does when dealing with this material is to determine whether it is simply a digital copy of already existing analog material, or if it is unique. If it is something unique – which also warrants digital preservation – Bob explained that it would be sent to the IT division.

As an example of such an instance where the archive received a mixed delivery of analog and digital material, Bob told us of a case where the archive received two CD-ROMs. One of these contained information which was identical to the analog material, and the other contained new information. In this case, Bob said that the archive chose to discard the first CD, and print out the information contained in the second rather than send it to the IT division. When we asked what his reasons were for printing out this information, Bob answered by saying that paper records are considered to be more trustworthy as evidence in a legal setting than their digital counterparts. He also stated that CD-ROMs should only be treated as media for transport, not for storage. Such media had occasionally found a way into the archives and remained hidden amongst analog material. In years past, the existence of such digital storage media was usually registered, but this registration was not always carried out consistently. Bob explained that there are now routines in place in order to ensure that this material is handled properly rather than hidden away in an archive box.

Tracking the digital information at Archive A, we mapped out who it is produced by, where it is stored and who is responsible for maintaining it:

1) Analog public records digitized by the archive
   Producer: External (original), Archive B (digital copy)
   Stored: At the IT division in an OAIS-modelled system
   Maintained: By the IT division of the host organization

2) Digital public records mixed in with a delivery of analog *public records*
   Producer: External
   Stored: At the IT division in an OAIS-modelled system, printed out on paper in certain cases, occasionally hidden in archival boxes
   Maintained: By the IT division of the host organization, by Archive B if in paper form, or for the ones in boxes: if they are not found, then they cannot be maintained but only stored

3) Digital public records produced locally
   Producer: Archive B
   Stored: Not kept in digital format but printed out on paper

Maintained: No current intention to save such material digitally, though this will change in the future

4) Digital material in the *private archives* mixed in with a delivery of analog material
   Producer: Private persons, businesses and associations
   Stored: At the IT division in an OAIS-modelled system, occasionally hidden in archival boxes
   Maintained: By the IT division of the host organization, or for the ones in boxes: if they are not found, then they cannot be maintained only stored

6.2.2 Responsibilities for digital information

The IT division is responsible for nearly all of the information in digital form. While it may occasionally occur that digital information inadvertently arrives with an analog delivery to Archive B, it is either moved to its correct place or printed out when it is found. Archive B is responsible for making sure that such material finds its rightful place.

When a user requests material which contains information which is sensitive for security or integrity, Archive B is responsible for granting or denying access. If the material is in digital form, and is cleared for access, then a request for the information is sent over to the IT division. It is entirely up to the IT division to handle backup, readability, migration, and similar issues. The division of labor can be summarized by saying that Archive B is responsible for access to the information, while the technical aspects are managed by the IT division. The organization saves their digital material in specified formats which rarely change, and Bob explained to us that format changes are not a particularly pressing issue for the IT division.

Archive B and its host organization have worked together with a security organization to certify that their holdings are secured against cyberattacks. This organization advised the host organization that such attacks generally come from the inside, originating with disgruntled employees. Attacks from the outside are less common. Therefore, Archive B and the host organization primarily defends itself against these kinds of problems by storing their digital holdings on different servers, with different levels of access for employees. This strategy of having a structure with multiple layers and controlled access is Archive B's main defense against cyber-attacks. There is also a dark archive which is not connected to the primary archives – the files linked to by the registry are copies on a separate server. These copies are the results seen by users when they perform a search. Bob said that security is a constant issue which can never be solved, and which requires a constant struggle due to ever changing technology.

Bob also explained that each of the information systems used within Archive B and its host organization have a system administrator assigned to them. Part of the reason for this is to prevent outdated systems containing sensitive material from being forgotten and disposed of improperly, which could mean an uncontrolled exposure of their contents.

6.2.3 What is a record in the digital world?

When we asked Bob what exactly was preserved from an information system, his answer echoed that of Archive A's. He explained that it is not the system itself which is preserved, but the information contained in it. We also discussed the issue of what actually constitutes a record in an information system, given that there often exist temporary combinations of information, depending upon what one is requesting from the system. Bob pointed out that this was something which needed to be identified before the system is put into use. It is the final records themselves, and not drafts or other extraneous information, which are to be preserved, said Bob. He continued by saying that if one does not define in advance which parts are records to be preserved, then a great amount of extraneous work has already been created: the work of sorting through all of this information after the fact. Bob stated that it is also necessary to identify which records are scheduled for culling. When we asked if he believed that this process – of defining in advance what constitutes a record in an IT system – functioned as it should out amongst the organizations who deliver material to Archive B, Bob expressed his doubts.

While discussing the digital public records produced by the archive which were printed out on paper, we asked Bob if he felt that this worked sufficiently well. He said that it did, but mentioned that it was important to keep new and old personnel alike informed as to what exactly constitutes a public record, especially where digital records were concerned. He also felt that this was a subject which should be emphasized more and stated that the level of familiarity with this knowledge has a significant effect on the material itself. He also emphasized that an over-reliance on a few go-getters to educate coworkers about what constitutes a public record was unsustainable in the long run, and that education concerning these practices needed to be a continual process.

While not everything related to public records is required to be saved, Bob said that he believed there was a tendency to save more when it came to digital material, since it only takes up a little space on a hard drive. He explained that this was problematic because it results in a glut of less-useful information which obscures the important information.

6.2.4 Information loss: Analog vs digital

When the question concerning digital information loss came up, Bob made several comparisons with losses in analog material. He stated that digital information loss was a serious issue and that preventative measures needed to be taken, but that it was also important to keep in mind that losses occur even in analog material. He made a comparison between the costs of maintaining analog and digital materials, and said that both required constant maintenance. While paper records require environmental controls together with rent for storage space, the maintenance of digital records incurs other costs such as maintaining servers, hardware, and the rest of the infrastructure. Bob also pointed out that the cost of storing digital material has declined steadily over the years.

Work is currently being done to harmonize the structure of metadata and digital information submissions which are delivered to the host organization. Bob sees this

as something that will decrease the risk for digital information loss in the future. When this process is completed, digital deliveries will be easier and it will be possible to automate aspects of the migration process This will also assist in reducing the cost of such deliveries.

When we asked Bob the same question we had asked Archive A – if he thought his archive's digital information would be readable in 400 years – he answered: probably so. However, he also made a comparison to written material from the 1600s and said that few people understand Latin, much less the laws, rules, and general functions of the administrative systems which produced the records from that time. Here he emphasized the importance of describing systems when the material is digital, explaining that information is difficult or impossible to understand outside of its context. In a digital context, this may be what determines whether the information saved is of any value to those who inherit our records 400 years from now. This, according to Bob, is why one of the most important abilities for archivists is to be able to describe information systems and the processes within administrative services.

6.2.5 Strategies for digital preservation management

Archive B and its host organization have both strategic and work plans for their digital holdings. They have focussed on finding more general solutions for future digital preservation. Bob claims that Archive B and the host organization will soon be able to reap the benefits of this work. Up until this point, the work of digital preservation has been undertaken to the best of everyone's abilities. Once the long-term solutions are implemented, one will then save that which can still be saved, said Bob. There is, however, a risk that the information in older systems, which have not yet been taken in by the archive, and which have remained in disuse for a long period of time, will be lost. Attempting to handle these older systems while simultaneously trying to keep pace with the new material being created would be a "Sisyphean task", Bob explained. Therefore, he felt that the best strategy was one which focussed on the future.

Bob felt that there were several obstacles which could potentially hinder digital preservation on a general level in Sweden. He perceived one of these to be an inconsistent level of knowledge amongst archivists. Another obstacle he identified was poor communication between IT divisions and archiving staff. In Bob's opinion, digital preservation was often viewed within IT divisions as simply being a question of extra backup – their understanding of the problems of long term digital preservation did not often go further than this. However, he said that this situation had begun to change, as staff have now had the experience of dealing with systems which are old enough to become unsupported. It becomes an expensive task to figure out what to do with them. Generally, when this occurs, the IT division contacts the archivists to determine what course of action to take. There is a risk, however, that the archivist has insufficient knowledge to handle such questions, which in the end makes it difficult to communicate the problem clearly to decision makers.

The greatest challenge is oversight, according to Bob. He expressed concern that this may not be functioning as well with digital information as it has with analog. If the oversight from the host organization does not work as it should, then digital information loss will occur. This will be the case when systems are no longer used or

maintained, and there is no one left who knows what it contains or how it was created. Bob stated that he was not in a position to say how many systems may have become lost throughout the years.

Regarding finances, Bob said that the archive had been forced to make personnel cuts in order to have enough funding for digital preservation. He did not otherwise express worry over the economic situation.

6.2.6 Cloud services

When we asked about cloud services, Bob stated that these were definitely not an option due to security issues. He doubted that the kind of rigorous, controlled access required by the archive could be maintained in a cloud environment.

6.3 Archive C

6.3.1 Digital holdings and information flow

Archive C focusses on housing one specific kind of cultural heritage. They differ from both Archive A and B in that there is no legal requirement for them to preserve their material. While the earliest objects in their collection are analog, almost everything they receive today is in digital form. Nearly the entire spectrum of hardware and formats utilized over the years by submitters is represented at this archive.

When receiving new digital material, the archive first determines whether or not it can be read. Mistakes are occasionally made, such as when files are saved incorrectly. In these cases the archive contacts the producer to resend the information. This can be considered their first line of defense against digital information loss. When Archive C receives a submission which consists of files, these are saved on a server, while other kinds of digital submissions are kept on their original information carrier. Due to the nature of the material, the archive does not feel it has the ability to affect which form it is delivered in. It often occurs that material submitted to the archive is in a proprietary format, which was a concern for Charles, who we interviewed at Archive C. Charles expressed concern about what would happen if the license agreement were to expire, or if the business responsible for the format were to go bankrupt. When the archive receives a file submission the archive keeps this in its original format. Those who submit material to the archive are expected to maintain their own copy, which the archive calls the “master”. The primary copy maintained by the archive is designated the “submaster”, and another copy is made as a backup. This backup copy is made in one of the formats or information carriers which have been predetermined to be suitable by the archive. When someone wants to use the archive’s material, they are provided with yet another copy of the submaster.

The archive maintains a database which contains all of the material which has been submitted as individual files, together with metadata for the whole collection. The database, which is in a proprietary format, is stored on the archive’s server. The archive also has a mirror of this database maintained by a cloud service, run by

Company Z. While the database is only accessible at the archive itself, a limited catalog of the material is available online.

The archive sees problems with the long term usage of their current database, since it is based on a commercial application rather than an open-source one. The archive is currently receiving assistance in the development of a new open-source database, which they hope to begin working with in 2016. Since the current database has become excessively complex over the years, the personnel have difficulty understanding all its functions in their entirety. The new database is modelled after one utilized by a similar organization, but adapted by Archive C to suit their own needs. Charles stated that he was uncertain whether the database was situated within a structure based on the OAIS model.

A portion of the ingested analog material has been digitized. Some of this, together with born digital material, has been copied and is preserved at another institution (Institution X), as well at Archive C. The archive was uncertain just how much digital material they preserved at their own location, and planned to perform an inventory to find out.

A user who wishes to make use of Archive C's material often makes a request for a specific format which will work on the hardware and software they have available. The archive fulfils these requests to the best of their abilities, but is generally restricted to certain formats. As the material is in many cases intended to be presented in a specific context, and sometimes even on certain hardware, this is an important issue. The archive does its best to communicate this to its users. When the archive sends out a copy that has been requested, the copy does not count as a part of the material the archive needs to save or preserve over time.

Internal work material produced by the archives is usually in digital format, though some of this was originally analog and was later digitized. In contrast to Archives A and B, these are not public records and there exists no legal requirement to preserve them, except for some of the records related to bookkeeping. Work materials are saved in formats which are proprietary, and are treated as living, changing documents. These formats include word, tiff, and jpeg. The files are kept in such a way that they can be changed as needed, and can be divided into two kinds: supplementary material concerning the archive's collection, and internal working documents, such as license agreements and receipts. The first group is stored on the server but is not present in the database, though the archive intends to change this situation. Putting this information into the database would necessitate a reallocation of resources in the future.

Apart from the submitted material and internal work documents, the archive also has a collection of research copies which vary in quality. While most of these are copies of material which exists in the archive's holdings, some of them constitute unique material. This collection, which consists of a mixture of analog and digital material, is kept separate from the general collection. Cathy, who we also interviewed at the archive, explained that she has encountered material in this collection which can no longer be read.

Tracking the digital information at Archive A, we mapped out who it is produced by, where it is stored and who is responsible for maintaining it:

1) Deliveries of digital material to the archive
   Producer: External
   Stored: Submaster in original form. If it is an individual file a copy is stored on the server, otherwise copied to the same storage media as submaster. Everything stored at Archive C.
   Maintained: By Archive C

2) Analog deliveries digitized by the archive
   Producer: External (original), Archive C, occasionally other institutions (digital copy)
   Stored: On storage media or on server
   Maintained: By Archive C

3) Digital-born work material
   Producer: Archive C
   Stored: On local computers, server, and different information carriers. Some material located in a cloud service
   Maintained: By Archive C, Company Z (material located in cloud service)

4) Analog work material which has been digitized
   Producer: Archive C (original), Archive C (digital copy)
   Stored: Uncertain
   Maintained: By Archive C

5) Collection of research quality copies in digital format
   Producer: External (original), Archive C, occasionally other institutions (digital copy)
   Stored: Stored apart from other material at Archive C
   Maintained: No

6.3.2 Responsibilities for digital information

The archive is solely responsible for the digital information stored there, excluding the backup maintained by Company Z. Cathy and Charles explained that the resources for the long-term preservation of their collection simply do not exist. While they do everything in their ability to maintain their holdings, this is on a day to day basis: looking towards the future is not possible. Up until a few years ago a collaboration existed between the archive and Institution X. This institution made lossless copies of the material, which they also preserved. However, the media which allowed for easy lossless copying is being phased out. Moreover, Institution X has discontinued the practice of using this kind of media, and now makes copies as digital files instead. These copies are not lossless. Institution X came to the conclusion that it was not within their mission to save a full quality copy of the original, and chose instead to preserve copies which were of a lower quality than those kept by Archive C. The quality of these copies was insufficient from Archive C's perspective. When Archive C realized the change in the quality of copies maintained by Institution X,

they felt obligated to discontinue the collaboration. Cathy and Charles both said that the lower quality copies did not constitute preservation as they defined it.

Archive C cannot, according to Cathy and Charles, be considered to be "preserving" the digital part of their collection: instead, they "store" it to the best of their ability. Cathy explained that there is no institution which has both the resources and the assignment to preserve the material maintained by Archive C. Since all of their holdings are in one physical location, the archive expressed hope that Institution X can at least make space for some copies, as a backup in case a disaster, such as a fire, were to occur at Archive C. This would at least reduce the risk by storing the material at two geographically separate locations. Not even the metadata is secure, says Charles: it is located in a database which is not coded in open-source. Charles and Cathy state that while they wish to work proactively, this is impossible for them under the current circumstances.

6.3.3 What is digital preservation?

When we asked Cathy and Charles what digital preservation meant to them, Charles answered that it meant to be able to reproduce the information at or sufficiently near the original quality. Cathy explained that she does not believe the archive currently works with digital preservation, but is instead working towards it. Cathy also said that she felt the concept digitization was used to mean different things. She explained that when one considers the migration of digital material or the digitization of analog material, there are many choices to be made which have significant consequences for the information. Dependent upon which choices are made, such as file format, the effects can be great. Therefore, according to Cathy, preservation can mean different things to different people when talking about digital preservation.

6.3.4 Information loss: Effects on the digital object

Archive C was formerly able to make use of a technology which enabled lossless copying, but this option is no longer commonly available. This was made possible via information carriers which were once the standard for the material received by the archive, during a period which began in the 1990s. Charles saw this as a paradigm of lossless copying which had now passed. The new technology which replaces it creates constraints which affect certain properties of the copy and may change how the material can be accessed or presented. Furthermore, Charles pointed out that some of the technology used today causes a small amount of information loss already during the actual production phase.

Cathy described several examples where she had seen how different items in the collection were very sensitive to which settings were used during migration or digitization. While this was barely noticeable for certain kinds of material, the difference was clear with others. Therefore, every object requires individual attention, which makes it impossible for Archive C to migrate or digitize their collection en masse. The result of this work is heavily dependent on the knowledge of individual staff, their degree of feeling for the material and their ability to keep that knowledge current with the ever-changing technology. Cathy stated that this was, in her opinion, one of the greatest challenges concerning the preservation of the material collected by Archive C. Its success is dependent upon knowledgeable, motivated and dedicated

individuals. Resources must also exist to support these individuals: without them, the quality of digital preservation will suffer. Cathy then told us of another organization which had built up an extensive knowledge base within this field. Due to a political shift towards the far right, this organization found itself spontaneously defunded by their government. Today it survives only in a limited form. Cathy expressed concern over the loss of knowledge that occurred in this situation. Archive C and the organization had previously maintained a crucial exchange. Cathy described knowledge as a "perishable good", and expressed worry that the future would see a disproportionate focus on technical quality rather than on a professional level of knowledge.

Many of the materials at Archive C are intended to be experienced in a certain context, on a certain kind of hardware. If one takes part in the information on another kind of hardware than that which it was intended for, then it may be possible that this experience deviates too far from the original intent. This may present the user with a less authentic experience. Charles said that while the user wants digital files, the information may have been planned to be presented in a different way.

Archive C struggles with the pace of technological shifts. The producers of the material stored at this archive utilize many different kinds of equipment and formats. Charles said that the fast pace of technological developments in the digital world constitute one of the greatest challenges. From the archive's perspective, it means that they must be ready to deal with many different kinds of formats, hardware, and software. It is impractical for the archive to own all of the necessary equipment. Charles explained that technological development was both friend and foe: ostensibly, the technology becomes simpler but the all the more often a situation arises where a digital object is bound to a specific program. This was a reason to focus on open source solutions, said Charles.

Generally, Charles and Cathy could not recall any instances of digital information loss in their collection. On the other hand, some of the digital information carriers at the archive are becoming antiquated, and their decay threatens the information stored on them. According to Charles, there have only been a few small issues with the archive's work material, but this is more often due to human error than to technical problems. The problem of compatibility has become less of a concern over the years in Cathy's and Charles' experience. Within the collection of research copies, some of which are unique, loss has been experienced. The archive is not focussed on preserving this material, though they realize its potential value. The information loss which affects the quality of digital materials, occurring during migration or digitization of the material is a constant challenge for Archive C.

6.3.5 Strategies for digital preservation management

Archive C currently has no strategic preservation plans in place. Cathy explained that while she is aware of the existence of plans for digital preservation, it would be meaningless to attempt them. The resources required to follow such plans are nonexistent at Archive C. She explains that, in order to solve the problem efficiently, it must be solved on a higher level, and tied into the general preservation of cultural heritage. There is no reason for every institution to reinvent the wheel. Ideally, the

material would be taken care of by the larger organisations which have the resources and knowledge to take care of digital preservation in the long term.

There is currently no organization whose assignment it is to preserve the material at Archive C, at least not in a way Cathy and Charles consider to truly count as preservation. Cathy explained that Digisam's work has helped with the discussion regarding digital preservation and brought problematic aspects to the forefront. While Digisam's task has been to develop solutions for governmental collections, even private organizations have been invited into their discussions. Cathy pointed out that even if Digisam were to develop a solution for private organizations such as Archive C, the help will arrive too late. A solution needs to be found soon, rather than in the future for Archive C to avoid information loss.

Due to a lack of resources, Archive C does not have the ability to work with long term digital preservation, and is worried about how they will continue to uphold a sufficient level of quality work in their underfunded state. The resources are such that their personnel have difficulty finding time to work with one another and maintaining good communication. Cathy said that is a challenge to run the archive and clarified: "A reduction of resources does not even exist in our imagination; in such a case I do not see how we could continue to run this place". She continued by explaining that there are many continuous, unavoidable expenses, and that their rent will be increased soon. The archive has no specific time frame in mind for preserving their material. Asking whether the information maintained by Archive C would be readable in 400 years seemed unrealistic, so we asked if they believed that their material would be readable in 30 years instead. Charles answered: we are working to keep it readable today. When asked if they could continue to work at their current level of funding, Charles and Cathy answered that it will be impossible to handle the digital information under such circumstances for a longer period of time.

6.3.6 Intellectual property rights

Archive C has not encountered any obstacles related to intellectual property laws in their digital preservation. On the other hand, Cathy has encountered a confusion about digital material – an analog mode of thinking – which focusses on the physical information carrier. For example, if three different institutions have three versions of the same material, a tendency exists to think that one of these copies is the "original", and somehow different. However, they are all copies, explained Cathy, and what is important is not the physical object but the content contained therein. Which copy is unimportant, except in choosing the one with the least information loss.

7 Analysis

This section will analyze the results of our study using the four interpretive frameworks described in section 4. These will be used to understand which aspects of the digital information preserved by these archives risk being lost or have already been lost, as well as which circumstances have led to the development of this situation. This will help us to gain a wider understanding of those challenges faced by these three archives regarding actual and potential digital information loss.

Each archive will be analyzed in turn. The results will first be discussed vis a vis the transition between the Records Lifecycle Model and the Records Continuum Model. Afterwards, we will consider how the responsibility for certain aspects of records were divided using Thibodeau's ontological understanding of digital information. Next, we will use the SPOT model to analyze which properties of the records are threatened. The organizational risks will be considered in the last section, using the OAIS Reference Model to understand how the organization works with digital preservation and to locate where the problems show up in the flow of records. After these individual analyses, all three archives will be discussed together.

7.1 Archive A

Archive A has information in digital format which is meant to be preserved over time. This includes the material it receives from the host organization and individual archives, as well as the information that the archive itself produces in the form of its own documents or via the digitization of analog materials.

7.1.1 The Records Continuum Model

It is not currently possible to for Archive A to accept deliveries of digital public records. Despite this, some deliveries have been made as an emergency preservation measure. This measure targeted public records which were stored in outdated information systems and which were in danger of being lost completely. The situation developed in part due to a lack of a proactive strategy concerning the preservation of information systems. If the producers of this information, who are a part of the host organization, had enacted a preservation strategy for these information systems, such a situation could have been avoided. Another factor in the development of this problem involved the treatment of digital material according to analog practices. Digital information was maintained as though it were analog, resulting in a situation where preservation measures were looked into after the information was no longer used. This was also a problem with the CDs and other digital information carriers stored in the private archives: without taking measures to preserve them as digital

information, their contents will eventually be lost. Archive A is now working proactively to enable digital deliveries for the future. Information systems will still be phased out during this developmental phase, which could potentially cause a recurrence of a similar crisis situation. Considered from Upward's perspective, Archive A is wedged between two paradigms of preservation: simultaneously as the archive is trying to transition to proactive work, it must deal with the fact that newly old digital material needs to be taken care before the new proactive measures can be implemented. This mode of management is similar to the concept of a record as it is understood according to the Record Lifecycle Model and not in the Records Continuum Model. If the information systems received by Archive A had been handled proactively, the system would probably have been coded in open source from the beginning. Now, the archive has instead been forced to spend time on the recoding of information systems, which Abby viewed as a diversion of resources that could otherwise have been better placed elsewhere.

The digital information targeted by the emergency preservation action was mostly saved. The information loss which did occur was associated with the migration of those documents which were attached to various transactions from their original format to PDF/A. A part of the problem was that the conversion software was unable to read all of the formats which were present within the material. This is one example of how the challenges with digital preservation can be of a purely technical nature. Another stumbling block for the archive in the migration of these records was that some of them were saved incorrectly from the beginning, which is human error rather than a technical problem. This problem manifested itself in two ways: either the document was saved a locked or protected PDF, or the file was saved with an incorrect file extension. The original intent of saving the PDF as protected or locked may have been to prevent changes and protect its authenticity. However, this has led to a situation where the document can only be read so long as supported software exists which can read the format, since it is not possible to migrate such material. This could be seen as an example of how the desire to uphold authenticity can clash with digital preservation in the long run. Such information will be lost over time.

7.1.2 The ontology of digital phenomena

When we discussed those documents which had been given incorrect file extensions, Abby took up the question of whether or not guessing the correct extension and correcting it could possibly be seen as affecting the authenticity of the record. After all, the producer had chosen to save the information in this way. Abby made the case that it would be unacceptable to make similar changes to an analog document. Interpreting the situation from Thibodeau's point of view, changing the filename does not affect its authenticity. Thibodeau explains that certain changes are unavoidable in the preservation of digital information and may be necessary in order to preserve the information over time. He explains that the only thing that can be preserved is "(t)he ability to reproduce the document." (Thibodeau 2002, p. 13). According to this theory, it is the 'essential respects' which govern what changes are acceptable. In this case, one could say that the essential respects include the readability of a document's contents, but not a file extension. Those documents which have been saved in such a way can be understood as though their physical binary code has been preserved, but a distortion of their logical aspect has blocked access to the conceptual aspect of the information. Guessing the way to these documents' readability should be a way to

repair the logical aspect, allowing one to get at the conceptual aspect. What can be said here is that human error – in this case, using an incorrect file extension – has made extra work necessary in order to re enable the readability of the documents. Furthermore, such work can only be performed if resources allow for it. In all likelihood the information will simply be lost, despite the fact that it exists in a purely physical sense.

The work undertaken to preserve the information in the e-archive is divided between Archive A and the IT division of the host organization. While the IT division has the responsibility for the hardware and backups, (i.e. the information's physical aspects), Archive A has the responsibility to ensure that the information is readable – maintaining its logical aspect. Due to the fact that users cannot currently access the material outside of the archive, it is also up to Archive A to offer a point of access, allowing users to take part of the information. In other words, this means that they must also maintain the conceptual aspect of the information. Archive A has the responsibility for the physical, logical as well as the conceptual aspects of the information outside of the e-archive. These divisions of responsibility are in some ways problematic. During the interview Abby expressed concern over the physical aspect of the information at the IT division. Her efforts to improve contact between the IT division and the archivists can be seen as an attempt to counteract the problems taken up by Bob during the interview with Archive B.

Concerning the CD-ROMs and other digital information carriers in archive boxes, these can be considered to be stored at the present time. While the binary code on the information carriers – their physical aspect – is preserved, the logical and conceptual aspects are not, meaning that these aspects will eventually be lost. This also means that the physical aspect will eventually be lost due to the inevitable decay of information carriers such as CD-ROMs. If action is taken before this occurs, it will likely be possible to recreate the logical and conceptual aspects, provided that the contextual information of the digital object – such as the formats typical for the time, or which software could read these kinds of formats – is known. This type of digital detective work demands a great deal of resources and it is unlikely that the archive will be able to make these kinds of interventions.

7.1.3 The SPOT model for risk assessment

Applying the SPOT analysis to Archive A, it becomes even more apparent that most of the threats to its digital preservation were of an organizational rather than technical nature. Nevertheless, we found threats to every property on the list. Rather than ordering these threats according to their severity, we list them in the order they appear in the SPOT analysis as presented earlier in the text.

There are four threats listed in the SPOT model which endanger the property of availability. The first of these is concerned with digital objects that have become unusable over time. Waiting until information systems are unable to be used any longer, and then taking emergency actions to preserve them, can be seen as a manifestation of this threat. An example of the second threat to availability – a failure to identify something for preservation – can be seen in the CD-ROMs and other digital material in private archives. This material is not maintained and is not always registered. The same threat may be realized if digital public records are never

identified as such. Though the cause of this is rooted in an organizational issue, the SPOT model is only concerned with the outcome. Copyright issues with certain materials may be a threat to availability at Archive A, but this remains unconfirmed. The property of identity was threatened by a failure to capture sufficient metadata, specifically concerning the correct file extensions.

There were some threats which we found difficult to place within the SPOT model. Two of these included threats to the property of persistence, and both of them involved migration issues. Migration was prevented in the first case by files which had been saved as locked PDFs, and in the second, there existed no conversion programs for certain formats. Within the threats listed by the SPOT-model, we found that Archive A lacked the equipment to read certain materials. This included the absence of a needed floppy disk reader, and other readers may soon be necessary for the assorted information carriers in archive boxes. Renderability was another property which was threatened by the incorrect file extensions – it is not possible, without trial and error, to discover what hardware and software are necessary to display them correctly. The same problem can even be seen as a threat to authenticity. Understandability was not threatened at Archive A, and the only potential threat to authenticity we could see was concerned yet again with the file extensions. Authenticity is threatened by this if the alterations made to correct file extensions are not properly recorded.

7.1.4 The OAIS Reference Model

In order to clarify the information losses and risks at Archive A, we have mapped them out onto the OAIS model. This allows us to create an overview over the archive's handling of digital information and helps us to pinpoint where problems occur according to the OAIS model's functional entities, components and information packages. We begin by rigidly outlining the relevant issues according to the OAIS structure, and follow this up by discussing some of these problems in depth.

Functional entities

Ingest - We learned about two problems that could be located within the Ingest functional entity at Archive A. The first of these was that the archive cannot currently accept digital materials. The second problem was the lack of proper routines for handling the digital material which arrives with private archives.

Data management - The lack of communication between the e-archivists and the IT division can be seen as a hindrance to the Data Management functional entity.

Access - The fact that users are currently unable access most of the archive's digital material, except for those materials which are accessible at the archive itself, affects the functional entity Access. (see also: Consumer component)

Administration - There were four problems with the Administration functional entity at Archive A. The first of these involved controlling how obsolete systems were handled (see also: Component "Producer"). The second problem was that the producers of digital material at Archive A were disproportional to the number of people responsible for the preservation of digital material. The third problem involved

the lack of strategies and plans concerning questions of archive policy and the functioning and maintenance of the OAIS itself (see also: Preservation Planning). The fourth problem was the lack of digital preservation competencies amongst coworkers (see also: Producer).

Preservation planning - The two problems involving the Preservation Planning functional entity at Archive A included the lack of written strategies for digital material and the inability of the converters to work with all of the existing formats.

Components

Producer - The main issue with the Producer component at Archive A was that information was saved incorrectly from the beginning, which makes later preservation difficult or impossible.

Manager - The Manager component is not fulfilling its responsibility of ensuring that strategies and plans exist for digital preservation at Archive A

Consumer - Users cannot currently access digital material, which affects the Consumer component. Some of the digital can be accessed at the archives, however.

Information packages

Submission Information Package (SIP) – The archive has difficulty in making sure that Submission Information Packages are identified for delivery and ensuring that they are saved correctly.

While mapping out the flow of digital information at Archive A, we discovered that some digital information exists at the archive which is not actively maintained. Instead, this information could be understood as simply being stored. The information we are referring to here is the CD-ROMs and other digital information carriers which were included in some deliveries to the analog private archives. This is a problem with the OAIS functional entity Ingest. That this information is only stored can be understood as the direct consequence of Archive A's priorities. Albert explained that the highest priority is placed on public records, since the archive is legally obliged to preserve these, while private material is not covered by such laws and is therefore less important.

Another reason for information loss, as Albert stated, is the lack of routines and strategies for handling this material. This issue falls under the functional entities Administration and Preservation Planning, as well as the component Manager. Albert explained that it can take a long time before the material is gone through, and even longer before it is registered – and that this work often falls on interns who may lack the knowledge of how to carry out such work properly. It is probably the case that when archives arrive with such digital information carriers that their existence is not always registered – and it is even less likely that their contents are registered. This

information loss could also be attributed to is the lack of collaboration between archivists who handle analog material and those who handle digital material.

The probable information loss in the digital material within private archives can be understood as the consequence of organizational issues related to the functional entities Administration and Preservation Planning together with the components Producer and Manager. Abby expressed that while she thought that the number of staff dealing with digital preservation needed to be increased, she also thought that digital preservation competencies among the general staff could be improved. As Abby sees it, the current level of funding for digital preservation at Archive A is at its limit.

It may be the case that the situation at Archive A appears as it does due to a lack of strategic plans for the preservation of digital material. The fact that such plans are not written down makes the problems less visible. Another challenge for Archive A is that the funding for digital preservation has, up until now, been project based. This is a situation Abby hopes to change.

An increase in the employees' competencies with digital preservation would reduce the risk that this information – some of which counts as public records – is not properly identified as such. This problem is an issue of oversight and falls under the functional entity Administration, the component Producer, and concerns Submission Information Packages. Abby stated that it is not possible to know how much digital material has been created without being identified as a public record. She took up the example of video or sound recordings, saying that such material may be less likely to have been identified as a public record by the personnel who produced it. It may be useful to point out a difference between analog and digital material here: that analog materials which have been lost may still be found again, while this is less likely for digital material. If we imagine two cases where an employee produces a video – in the first instance on their smartphone, in the second on VHS – the physical existence of the VHS tape makes it easier to identify as information worth saving, while the video on the smartphone is experienced as abstract and disposable. The physical existence of the latter is located within all the other information and functionality of a smartphone. One the other hand, information loss will still be the result if the VHS is not identified as a public record or is taken home by one of the employees. As previous research has shown, digital information is less apt to be identified as a public record than information located on media which has a more obvious physical presence.

Another organizational question which affects digital preservation at Archive A is concerned with succession planning. This involves issues that can be categorized under the functional entities of Administration and Preservation Planning, together with the component Manager. When Abby took over her predecessor's position, approximately three work hours were set aside for the transition, which Abby described as inadequate. If it had not been the case that Abby and her predecessor had the possibility to maintain contact, this lack of sufficient succession would likely have led to information loss. However, if better routines had existed for the documentation of work concerning digital preservation, the risks caused by this succession would have been much smaller. In any case, it is not always possible to make a smooth

transition between employees, such as in cases where an employee suddenly must leave (Corrado & Moulaison 2014, p. 69).

The consequences of the inadequate documentation become clear when considering material which is strongly controlled by copyright/agreements/licensing. When we discussed with Abby whether there exist any copyright issues which limit the ability to reproduce and disseminate material digitally, she answered that it was probable that such issues exist. However, she said that this information has not been written down, and is impossible to learn about outside of directly asking one of the employees who happen to have knowledge of it.

During the interview it came to light that Archive A has a semantic division between the digital information contained in their e-archive, and the digital information that is outside of this e-archive. Abby stated that electronic archives are often defined by whether they are constructed according to the OAIS model or not. She also pointed out that one of the archive's digital collections has an organization surrounding it which could be seen as similar to the OAIS model, and could therefore also be seen as an e-archive. Abby believes that the division between analog and digital material which exists today will disappear with time, and that one will refer to all the material preserved by the archive as simply "archives". The lack of clarity which exists in the definition of what counts as an e-archive and what doesn't is likely to have a negative effect on preservation.

7.2 Archive B

All the digital information preserved by Archive B is gathered into one OAIS modelled information system which is maintained by the IT division of their host organization. Deliveries of digital material including deliveries by Archive B to themselves are also maintained by the IT division. Such material includes digitizations made by the archive itself. The digital public records produced by Archive B are not preserved digitally but rather printed out on paper. Since all of the digital information which is meant to be preserved exists within the same system and is managed in the same place, the risk for data loss is lower.

7.2.1 The Records Continuum Model

The archive has focussed on ensuring that digital preservation works smoothly in the future. Such efforts have been directed at solving problems with digital preservation on a more general level, so that one can preserve proactively. One aspect of this work has involved the harmonization of information. This can be seen as an implementation of the Record Continuum Model and an attempt to adapt to the new preservation paradigm. The digital information which now risks being lost is that which was created before the archive had managed to transition to the new preservation style. This concerns information systems which have fallen out of use amongst those organizations which produce public records, who may have even been forgotten that they ever used such a system. Bob explained that it was no longer possible to fight on both fronts; one had to focus on making future digital information safe rather than fretting about what may have already been lost. Since Bob does not work directly with these questions, he did not have insight into how many

undelivered obsolete information systems may have been lost in the transition between these two preservation paradigms.

One consideration for making the transition to a preservation model based on the Records Continuum Model is that what constitutes a record in the digital world must be defined in advance. This means defining which part of the information is actually is a record. When we asked Bob if he believed that this was taken care of out amongst the organizations who produce information received by the archive, he answered that it probably did not. This suggests that when the information is ready for preservation, one will need to go through large amounts of information in order to try and understand what it is which should be saved. Even if there are processes which have been developed in order to preserve proactively, this does not mean that everything functions as it should. It can be seen as an expression for the set of problems which Runardotter (2007, p. 71) and Runardotter, Mirijamdotter och Mörtberg (2007, p. 54) pointed out nine years ago: that public records in digital format are less often recognized as public records.

In their study, an archivist expresses frustration of continuously having have to teach co-workers what records are and how they are to be handled. Bob's answers can be seen as an expression of that the same set of challenges still exist today. The need to continually educate producers as to what is a record and which records are public is something that Bob sees as a never-ending challenge, and shows how work with digital preservation is a continuous and constant process.

7.2.2 The ontology of digital phenomena

Despite the fact that Archive B now has a secure system for digital information, they choose to print out certain information on paper. Bob said that authenticity was the reason behind this. Printing out such information entails a loss of the physical, logical, and to a certain extent even the conceptual aspect. That which Archive B has identified as the essential respects can be assumed to be the contents of the text itself. The visual characteristics may also be considered to be insignificant as well. From this, one can draw the conclusion that the risk of information being manipulated in the digital world is so problematic that, for something to function as evidence in a legal setting it is better for it to be printed out. Paradoxically, this information which is printed out on paper, like the example taken up by Bob, originates with a digital file.

In the case of Archive B, the archive is responsible primarily for the analog material. All of the digital preservation questions reside with the IT division. However, access to this material is the responsibility of the archive. From Thibodeau's perspective, this could be understood that the IT division is responsible for the physical and logical aspects of the material, while Archive B is responsible for the conceptual, in this case to offer a point of access where the visitor can take part in the information.

7.2.3 The SPOT model for risk assessment

We found only a few threats to digital preservation at Archive B that could be elaborated on by the SPOT model. Several of these threats were concerned with problems which had occurred in the past, and which do not reflect the current digital

preservation scenario at Archive B. One example of this older material may involve older systems which have not yet been delivered to the archive and which may be stored or forgotten at the organization where they were last used. This threat to availability can also be seen with material which was not ingested properly and which may still remain in archive boxes. The possibility that digital public records were not identified as such and thus never preserved constituted yet another threat to availability at Archive B. Identity was threatened by the previous existence of two disconnected ingest routes: one for digital and the other for analog material. This property threatened the material which belonged to the same archive, but was not linked together. The CD-ROMs which are left in archive boxes were also a threat to persistence, as these media will eventually decay, taking their information with them if they are not found and preserved. Finding playback devices for obsolete media will also become more difficult as time goes on. We found no threats to renderability at Archive B. Concerning understandability, potential threats included the archivist's ability to identify the needs of their users and what contextual information should be preserved. Authenticity may be threatened at Archive B due to its "print it out" strategy for internally produced digital public records. Printing these records removes access to their metadata, and potentially disconnects them from other contextual information as well. Although authenticity can be assured after printing, this can no longer be verified for the documents previous existence as a digital document.

7.2.4 The OAIS Reference Model

In order to clarify the information losses and risks at Archive B, we have mapped them out onto the OAIS model. This allows us to create an overview over the archive's handling of digital information and helps us to pinpoint where problems occur according to the OAIS model's functional entities, components and information packages. We outline the issues according to the OAIS structure, and analyze some of the issues in a more in-depth discussion afterwards.

Functional entities

Ingest - The problems with the Ingest functional entity at Archive B included two issues which occurred in the past and have now been resolved. The first of these issues involved CDs which were improperly dealt with when they arrived at the archive and the other concerned the fact that there were two different routes for ingesting digital and analog materials, which were not always linked to one another.

Administration - There were three issues with the Administration functional entity at Archive B. The first of these was on the Producer's side, in that obsolete systems were not always properly monitored. The second of these involved the fact that digital public records produced by the archive were printed out rather than stored digitally (see also; component Producer). The third problem was an uneven level of digital preservation competencies amongst producers at the archive and at the organizations who deliver material to the archive. This meant that digital public records were not always identified as such or saved properly (See also: Producer)

Data management - The problem described earlier in the Ingest functional entity also falls under Data Management: in the past there existed two different routes for

ingest of digital and analog material, resulting in a disconnection between the analog and digital material belonging to the same archive.

Preservation planning - There were two issues with the Preservation Planning functional entity at Archive B. The first of these involved dealing with digital material which had not yet arrived at the archives (see also: Producer). The second of these involved the printing out of digital public records produced by the archive (See also: Administration).

Components

Producer - There were two problems with the Producer component at Archive B. These involved the ability of producers to identify public records in the digital realm, and the fact that digital public records produced by the archive are not preserved digitally.

Manager - The issue of oversight for Archive B in the digital realm was an issue for the Manager component. This involved ensuring that employees identify and properly take care of digital public records.

Information packages

Archival Information Package (AIP) - The AIPs at Archive B were affected by the fact that digital public records were by printing out on paper rather than preserved in their original form.

Preservation Description Information (PDI) - The loss of connection between digital and analog in the past affected the PDI of the information package.

While in the past there have been cases where digital material has arrived with analog deliveries and has been placed in archive boxes, there are now routines for handling this issue, which falls under the OAIS functional entity Ingest. The digital information which has found its way into archive boxes yet remains undiscovered risks being lost. Regarding just how consistently the digital information has been registered, Bob answered that it had not always been as consistent as it is today.

Another past risk which has already been corrected is the fact that deliveries to the archive previously had two routes of ingest, one for digital material and one for analog. This was a problem for both Ingest and Data Management, and negatively affected the Package Description Information aspect of the Information Packages. This meant that the connection between the analog and digital material was sometimes lost. While no information was directly lost, the connection may have been, resulting in the loss of contextual information.

Archive B does not experience funding as a threat to digital preservation at their institution. They have fully-developed strategies and work plans. The challenge that

Archive B judged to be the greatest was the question of oversight, which concerns the functional entities Administration and Preservation Planning, together with the components Manager and Producer. There is a concern that those who produce digital public records may not be knowledgeable enough to identify them and save them in the proper way. Older systems containing public records may not have been identified for preservation and may incur information loss before their arrival at Archive B.

Regarding the digital public records which Archive B has chosen to print out – such as emails – there may be some risk of information loss. This is an issue for the functional entities Administration and Preservation Planning, the component Producer, and the Information Packages SIP and AIP. Printing out what was originally a digital public record means that the transition between the submitted information package and its archived form runs the risk of losing descriptive information (PDI) from the original submitted information package. It also runs the risk of losing any contextual information such as links to other documents which it may have contained in its digital form.

7.3 Archive C

Unlike the first two archives, Archive C does not preserve digital public records. Due to funding issues, they can be said to store their material rather than preserve it. It should also be noted that Archive C does not control the format or information carrier that their producers deliver to them. Archive C has generally higher requirements for the preservation of their material, a factor which in and of itself complicates digital preservation at this archive.

7.3.1 The Records Continuum Model

Archive C is engaged in a constant struggle to preserve information as unchanged as possible while technology is making this more difficult. In the past a ubiquitous technology existed for making lossless copies. This has been gradually phased out and the material currently preserved with this technology will gradually be lost if it is not migrated. Problematically, they do not have the resources to handle the material themselves, and high demands are placed on the technology and knowledge simultaneously as there exists no agency which has the assignment and resources to truly preserve this material.

The material Archive C intends to preserve is especially vulnerable to technological development. The producers of this material quickly adapt to and make use of new technology, which has a strong effect on Archive C due to the fact that the material the archive receives is saved in its original condition. In contrast to Archives A and B, Archive C does not have the possibility to affect the producers of their material nor do they see that they have the ability to affect how material is delivered to them in. This can be interpreted as Archive C not having the possibility to work proactively like Archives A and B can. They cannot implement a records continuum handling of the digital material. They are forced into a preservation structured modelled after analog preservation, which in practice entails a constant struggle for the archive. The submaster which they preserve is judged, in other words, as something that they do not wish to change the logical aspect of.

Archive C does not follow any preservation plans, since they do not have the means to enact them. The organization has unique material, which at the time of this writing is not preserved but stored, until a solution becomes available. Archive C sees that it needs a more general solution and that a government organization should have the explicit assignment to preserve this material, which the archive was working towards. While the work with Digisam is intended to be for governmental organizations, non-governmental organizations have also been invited into the discussions. Cathy points out that even if Digisam were to eventually offer digital preservation even for non-governmental organizations, this is not a solution because it is so far off in the future that one must try to find other solutions.

The material can judged to be at risks for reasons other than the passage of time and inactive preservation. Much of the material only exists at the archive, if something were to happen at this location the material would be lost. The producers save their materials in the master versions, but it is not certain that these will be readable over time. It may also be the case that copies do not get saved if the person dies. Even if these are saved physically, it does not mean that they will be preserved.

7.3.2 The ontology of digital phenomena

Archive C no longer considers its digital material to be preserved. While the archive previously saw itself as working with long term digital preservation, this has not been the case since 2010, when their collaboration with Institution X ended. This collaboration ended because Institution X ceased to maintain lossless copies, insisting instead on lesser quality copies. Moreover, it was not the larger institution's role to maintain such material at such a high quality. This can be understood via Thibodeau's ontology of digital information: the 'essential respects' identified by Cathy and Charles differ from those 'essential respects' which the other organization deems necessary for preservation. Preservation can take different forms. For Cathy and Charles, this means that preservation is only possible if the information is reproduced in a quality sufficiently near the original. There should be no noticeable degradation in the quality.

The high requirements for preserving that which is identified by Archive C as 'essential respects' can themselves be seen as a threat towards the preservation of their collection. In order to preserve the material in its original quality, a high level of familiarity with complicated technology is necessary. The same problem exists with the digitization of their analog material. This means that only an organization which can maintain those levels of quality can preserve the material in the opinion of Archive C. Not preserving those 'essential respects' which Archive C has identified is the same thing as information loss. It is not a question of total information loss, but information loss to a degree that it is no longer a question of preservation. One can say that the loss of knowledge means an information loss for Archive C, since the acceptable modalities migration and digitization is no longer possible. This is why the closure of another institution that Archive C had an exchange with which was seen as such a blow .Cathy stated that there were few who understood these problems.

There are three types of information loss here. One type is total information loss, in the form of damaged hardware which potentially could occur due to a disaster like a

fire at the archives or the deterioration of physical media. This would cause information loss by destroying the physical aspect of the information. A partial information loss is another type, occurring via unsatisfactory copies, where the conceptual aspect is damaged. A third type would be through a technological development that makes it impossible to participate in recreating the conceptual aspect as it was intended. At the present time, there is no information loss of the material that the archive stores and has the intention to store/preserve. However, the risk that it will be lost is great. The reason that this is the case is that they themselves don't have the ability to preserve the information's 'essential respects'.

The archive does not present the conceptual aspect of their material – this takes place with the user instead. However, the archive attempts to offer advice and assistance so that their material's conceptual aspects are expressed correctly. If a certain material is meant to be presented or engaged with in a certain way, Archive C tries to ensure that the user does so. The material may need to be displayed on a certain kind of hardware. With time it will become more and more difficult for the archive to maintain this level of fidelity. The hardware which is necessary to take part in the conceptual object in its proper form will become more and more difficult to locate and maintain. Archive C gives an example of other hardware which is used in order to take part in the conceptual aspect and they explain that it is often unsatisfying, and communicates the conceptual object in a distorted fashion. The users asked for the material in different logical and physical compositions, so that it would work with the playback devices available at their institution, a situation which was not always to Archive C's satisfaction.

7.3.3 The SPOT model for risk assessment

Archive C presents a difficult situation to analyze, as they currently consider themselves as only capable of storing – not preserving – their material. Nevertheless, a SPOT analysis can be useful in showing the numerous ways in which their material is threatened. The SPOT analysis for Archive C shows threats to every property listed, with the exception of authenticity. The threats to availability at Archive C begin with the entire collection being seen as not selected for preservation. Another problem related to availability is that many materials are in proprietary formats. Another issue with availability at Archive C is that the technology used in making lossless copies is no longer widely used. This can be seen as making it more difficult to preserve an important aspect of their digital objects. The property of identity is threatened at Archive C by the current database, which is not coded in open-source, creating a risk for the connections between the digital objects and their metadata risk to be lost during a transition to a new system. Persistence was another large problem for Archive C. Since they are unable to affect what format/media they receive their material in, many kinds of media obsolescence are developing simultaneously. Moreover, the materials on outdated information carriers will require migration, which will divert already scarce resources. Lastly, the equipment required for playback runs the gamut of technology developed over several decades. Eventually, the correct playback equipment will become unavailable. There were two threats to renderability at Archive C. Due to the fact that each and every digital object requires specific attention to its technical settings, Archive C has no ability to batch migrate its material. Secondly, the organization Archive C had previously collaborated with did not identify the same characteristics as being essential for preservation that Archive C

did. Lastly, the property of understandability could be threatened if relevant information about formats, playback requirements, and other technical aspects were crucial aspects which needed preservation themselves. This does not seem to be a problem at this time however.

7.3.4 The OAIS Reference Model

It is problematic to map out Archive C according to the OAIS model since the archive considers itself as storing rather than preserving its material. Nevertheless, this is done here in order to clarify where problems exist. This section begins with an outline of the issues within the OAIS model, which is followed by an in-depth discussion of some of the problems.

Functional entities

Administration - The fact that there is no organization which has been assigned the responsibility for this collection in a way that preserves the essential respects of the material is a problem with the Administration functional entity.

Access - There is a problem with the Access functional entity at Archive C due to the difficulty of balancing the authenticity of DIPs with user friendliness. Addressing concerns of accessibility must not lead to excessive divergence from the original intention of how the digital object should be displayed

Preservation planning - The Preservation Planning functional entity can be considered absent at Archive C, due to a lack of means.

Data management - The fact that the old database is in a proprietary format is an issue with the Data Management functional entity, though this problem is in the process of being corrected.

Components

Consumer - The Consumer component is affected by the difficulty of taking part in collection due to a conflict between making materials user-friendly and ensuring authenticity in their presentation.

Producer - There is a problem with the Producer component where the archive cannot affect the format the material is delivered in. This results in a proliferation of formats at Archive C. This variety of formats causes an exponential increase in the vulnerability of the collection to technological obsolescence.

Information packages

Preservation Description Information (PDI) - There was a lack of Preservation Description Information for the materials in the research collection at Archive C.

Dissemination Information Package (DIP) - There is difficulty in maintaining the balance of authenticity when dealing with DIPs in a way that is both user friendly and does not diverge from the original intent of the digital object.
Archival Information Package (AIP) - There were two problems with the Archival Information Packages at Archive C: they were saved in an excessive number of formats, and their preservation was vulnerable to which settings they were saved with.

Archive C does not follow any preservation plans, since they do not have the means to do so. This problem involves the functional entities Administration and Preservation Planning. The organization has unique material which is currently stored rather than preserved. Archive C sees that it needs a more general solution and that a government organization should have the explicit assignment to preserve this material, which Cathy said the archive was working towards. They have been invited to discussions with Digisam, which is primarily meant to be for governmental organizations. Cathy points out that even if Digisam were to offer digital preservation assistance for non-governmental organizations, this is too far in the future to offer a viable solution.

One instance of potential information loss at Archive C involves dissemination information packages, and concerns the Producer component and the Access functional entity. It is difficult to maintain the balance of authenticity when dealing with DIPs in a way which is both user friendly and does not diverge from the original intention of how the digital object should be displayed. The archive must negotiate with consumers to provide them with a digital object which is both usable and faithful to the original.

As mentioned earlier, Archive C cannot affect which formats SIPs arrive in. This issue involves the functional entities Ingest and Preservation Planning, the information package types AIPs and SIPs, and the Producer component. The prolific variety of formats dealt with by Archive C together with high degree of fidelity required by the digital objects constitutes a threat to the preservation of these materials. The collection at Archive C requires a great deal of specialized knowledge and funding if it is to be properly preserved.

There is an issue with the Database Management Functional entity due to the fact that the old database is written in a proprietary code. This is a threat to the contextual information and connections between digital objects.

The material at Archive C can be judged to be at risk for yet another reason: much of the material only exists at the archive, and if something were to happen at this location the material would be lost. This concerns the Archival Storage functional entity, and the components Producer and Archive. The producers save their materials in master versions, but it is not certain that these will be readable over time. It may also be the case that these are not saved in the event of the producer's death. Even if these are physically saved, it does not mean that they will be preserved.

7.4 Discussion

In this section we will discuss our analysis of digital information loss and the conditions which have led to or increased the risk of it at the three archives. The discussion will take place against the larger backdrop of previous research. Our study has shown that the three archives in this study struggle with a wide range of challenges concerning digital information loss. What constitutes an overwhelming challenge for one archive may be a non-existent problem for another. The study also shows that while digital preservation problems are often due to a mixture of technical and organizational issues, the informants we interviewed were primarily concerned with organizational issues.

Digital preservation may seem to be dominated by technological concerns at first glance. Previous research, such as Asproth (2005, p. 32) och Borglund (2008, pp. 17-18) claim that a disproportionate amount of attention has been paid to technical challenges. The results of our study confirm that organizational issues were the foremost concern for the three archives. Corrado & Moulaison claim that an archive which lacks a sufficiently well-developed organizational structure for digital preservation also lacks digital preservation itself (Corrado & Moulaison 2014, ch. 1). A proactive approach to digital preservation such as that emphasized in the Records Continuum Model has been deemed essential by researchers such as Borglund and Anderson (2011 pp. 272, 280) and Cook (1997, p. 28). That digital preservation out in the real world sometimes appears to continue to follow a Records Lifecycle approach rather than adapting to the new paradigm and taking on a proactive approach has been noted by Kallberg (2013 p. v). Whether or not the archives in our study chose to undertake a proactive approach was largely determined by their ability to do so rather than their willingness. The paths taken by Archive A and Archive B towards the new preservation paradigm are quite different. While Archive B has been able to focus on the future and implement an understanding of records based on the Record Continuum Model, Archive A has felt forced to save information in older systems which risks being lost. Unable to implement a full digital preservation strategy at the present, Archive A is coping with its responsibilities as best it can in the face of limited funding. That such an incremental approach to digital preservation is preferable to none at all was noted earlier in Schumacher et al. (2014, p. 15). On the other hand, the transition to a proactive digital preservation in and of itself requires a prioritization of certain material for Archive B. Bob characterized the attempt to save older information simultaneously as one works to introduce proactive preservation a "Sisyphean task". He explained that one will simply have to try and save that which can still be saved after the new way of working is put into place. This situation makes it clear that the transition from a post-custodial to a proactive preservation brings about a direct confrontation with digital information loss. While a vast quantity of digital information has been produced over the last few decades, it is only now that organizations are coming to grips with how to preserve it all. An interesting result in the study is that Archive C does not have the capability to undertake proactive preservation. The nature of the material received by Archive C leaves it unable to affect many of the material's aspects. The copy (submaster) which Archive C preserves is always saved in its original format. According to Charles, those who have produced the information saved by Archive C have utilized both new and old formats. This results in a vast number of format types for the archive to contend with, and is one of the greatest challenges for the archive. In comparison, Bob at Archive B

experiences file formats as a non-issue. For Archive A, organizational issues were the primary concern. This shows that the same digital preservation issues impact the three archives differently and to different degrees. Archive B's strategy differs from that of Archive A's, the latter of which is treading water while dealing with issues from the past and simultaneously preparing the foundation for proactive digital preservation. Both archives deal with information that is inadequately preserved before it is delivered. The predominance of organizational issues should not be interpreted as a lack of technical issues. Archive C, like McLeod (2008, not paginated) faced both organizational and technical threats to their digital preservation efforts. One example of this is that both archives A and C have been required to reallocate resources to the recoding of databases which were not written in open source.

The decision to examine parts and aspects of information rather than restricting ourselves to whole digital objects has given us access to a much more comprehensive set of results. This has helped us to show the big picture of what digital information loss looks like at these institutions. We were able to see potential and actual information loss by looking at parts of digital objects, whole digital objects, work documents and metadata. This also allowed us to discover a particular type of information loss at Archive C. Cathy and Charles stated that they would be hesitant to describe their information as "preserved" any longer if its migration or digitization meant a loss of quality. For Archive C, the information loses some of its 'essential respects' in Thibodeau's sense, if it no longer exist in the same high quality. This brings up the problem of just what can be said to qualify as preservation for a particular digital object at a particular archive. While digitization is often treated as though it had one, monolithic meaning, there is a multiplicity inherent in digitization, which has been discussed by Owens (2012). Even a failure to include certain kinds of metadata may disqualify a digital object from being considered to be preserved (Corrado & Moulaison 2014, pp. 69-70). In order to carry out migration and digitization processes at Archive C, a great amount of knowledge and a feeling for the material is required. This problem forces the archive to migrate each digital object individually rather than utilize batch migration. That knowledge is a perishable good which must be sustained was mentioned by Cathy during the interview, and is also mentioned in Digisam's report. The proposals set up by Digisam in their report have also argued for the need to sustain competency levels over time (Digisam 2016b, p. 31). However, Archive C risks losing the portion of their collection saved on one kind of information carrier, which is deteriorating rapidly. This material must be migrated. That technical limitations could lead to information loss was also evident at Archive A. Some of the losses they dealt with were due to software limitations: the software they used to migrate text documents to PDF/A could not recognize all of the formats in the collection that was targeted for migration. Information loss in connection to format migration is something that both Archive A and C have experienced. Further complicating the issue at Archive A, some material cannot be migrated because it was saved with an incorrect file extension, a problem which originated with the producers of the material.

Funding was a threat to digital preservation at both archives A and C, but not at B. The financial situation at Archive C is so severe that it places their entire enterprise at risk. However, an increase in the availability of financial resources would not in and of itself solve the problems at Archive C. Cathy stated that a large-scale and long-

term solution must be found, by placing the collection under the care of a larger and more stable institution with the means to preserve it. Archive C explains that the situation has developed to the point where they have not even been able to adopt any sort of digital preservation strategy. They simply lack the capability to follow and enact such a strategy. If this situation does not change, Archive C will soon experience information loss. Since Archive C is a private archive, it falls outside the framework of an initiative such as Digisam. Even Archive A lacks strategies for digital preservation, which is one factor in its concern for organizational issues. Abby explained that more personnel who are involved with the preservation of digital information are needed and that the current personnel were in need of improved digital preservation competencies. The lack of knowledge, resources, and strategies has led to a situation where information in digital format has found its way into archive boxes amongst analog material. This reminds us of Kallberg's claim that many professionals are aware of the problems with digital preservation but that funding, knowledge, and cooperation are frequently insufficient (Kallberg 2013, pp. 102, 123). It is possible that information loss has already occurred at Archive A in the private archives. Even if one would take measures to save this information, it is not likely that everything could be found, since it may never have been registered. Eventually such undocumented information will be lost due to obsolete hardware or software. The situation also highlights the importance of distinguishing "storage" from "preservation" when it comes to digital information and is an example of how an analog understanding of records is directly damaging for information in digital format. At Archive B, there has been a problem with the same type of situation and even there, in those cases where digital information has not been registered, information loss will occur if it has not already.

Both archives A and B are producers of public records, some of which are digital. Since previous research (Runardotter 2007, p. 71) shows that digital official records are less likely to be identified as such than their analog counterparts, we chose to discuss this question during the interviews. Abby at Archive A confirmed that this was probably an issue: it is likely the case that digital public records are produced but not recognized by their producer as public records. Therefore, these records remain unpreserved. However, she pointed out that it is impossible to know to what degree this has occurred – how can one know that they have lost something they never knew they had? The question highlights a quandary in digital information loss: in order to be able to make a statement concerning which losses have been experienced, one has to have a knowledge of which information one ought to have. When this information is never even registered such an assessment becomes impossible. In this specific case – with public records – it could be claimed that information loss in the digital realm does not differentiate itself from the loss of analog materials. However, previous research shows that it is more common for digital information to go unrecognized as a public record, and it follows that losses are greater within digitally produced information. This means that some digital public records risk never being preserved and will be lost sooner or later due to technological and organizational issues. An analog public record which goes missing may eventually be rediscovered. A digital public record which is not recognized as such, and is stored on an employee's computer, will probably be lost when the person leaves their job or goes into retirement. Bob was of the opinion that the difference between information loss in analog format compared to digital was not especially great. However, the above

scenario could be said to indicate the opposite. If the production of digital information goes undocumented, it can be difficult to discover losses, since it is not possible to judge the condition of digital information without accessing it. Analog information leaves an empty physical space when it is gone while digital information is not obviously missing in the same way. A loss of metadata or other contextual information is equally difficult to detect. Furthermore, digital information is not bound to its information carrier, even though its existence presupposes an information carrier of some sort.

In their research, Borglund and Anderson take up the issue that digital information does not always take the form of clearly designated entities, which potentially makes it more difficult to identify for preservation. In the Ådalsbanan project, Borglund and Anderson discovered that information which did not clearly take the form of documents fell outside the preservation strategy for the project and was not preserved (Borglund & Anderson 2011, pp. 280-281). The problem of identifying what constitutes a record in a system is something which was discussed with Bob. Bob meant that this identification does not function perfectly out in the organizations who produce public records, and like the archivist interviewed in Runardotter, Mirijamdotter and Mörtberg (2007, p. 54), he claimed that perpetual education of the producers of archival information was necessary. A truly proactive approach to digital preservation is not possible in such circumstances. This is relevant to what Bob saw as one of the greatest challenges: oversight. He emphasized that oversight over the handling of digital public records must function as well as it does with their analog counterparts. Bob was concerned that this was not currently the case. This also echoes Quisbert's claim that since we cannot affect other threats to digital preservation such as the amount of digital information produced or technological obsolescence, competency must be focussed on (Quisbert 2008, pp. 3-4). It is interesting to look at Archive C in this context, where oversight is not an option.

Another problem discussed at archives A and B was a lack of satisfactory communication between archivists and IT personnel, which has been written about in the research for at least a decade (Anderson, Samuelsson & Jansson 2011, pp. 32-43; Asproth 2005, p. 31; Runardotter 2007, pp. 77-78; Kallberg 2013, p. 123). Though the interview with Bob shows that this problem remains today, he also stated that the situation is changing. At Archive A the issue is still relevant. The lack of communication between the two departments could be seen as the result of how the responsibility for digital information is divided. If the host organization for Archive A had taken a more proactive approach, the systems would not have reached the point where emergency action was necessary to preserve them nor would Archive A have been forced to allocate resources towards recoding them in open source.

Thibodeau's description of the ontology of digital objects has enabled us to see how the responsibility for various aspects of digital information has been divided up at the three archives. The results show that the division of responsibilities takes different forms at the three organizations. While Archive A is responsible for the logical and conceptual aspects (concerning information which resides in their e-archive), Archive B is only responsible for the conceptual aspects. Archive B currently prints out the digital public records it produces. The need for context in understanding information and whether this should be preserved as paper or digital documents has been taken up

by Asproth (Asproth 2005, p. 32). Archive C on the other hand is responsible for the physical and logical aspects but not the conceptual ones. In the case of Archive C, the responsibility for the conceptual aspects falls on the user. The digital object accessed by users at Archive C is a display copy and not the sub master. This can be compared with Archive B where the user accesses a display copy, rather than the one preserved in the dark archives.

Though it falls outside the intent of this study, an interesting question arose concerning terminology. During the interviews it became clear that there are several terms which were understood differently by different informants. When we asked about which information in digital format was preserved by the archive, several of the informants began by only referring to material which had undergone digitization from analog to digital. The focus on digitization over born digital information has also been noted in the previous research (Rhodes & Neascu 2009, p. 42). One informant used the word digitized to mean not only information which had undergone digitization but also information migrated between digital formats. Another term which created confusion was the concept of "e-archive" at Archive A, something which has been taken up in the results and analysis. At Archive C the concept "digital preservation" led to another discussion, which has also been taken up earlier in this text. A final term which can be mentioned is "information loss": when this was discussed during the interviews, the informants initially only thought of the loss of entire digital objects.

This study shows that it is beneficial to examine digital preservation in terms of information loss, which is the consequence of inadequate digital preservation. Digital information demands constant work and attention: the physical infrastructure must be maintained and updated, formats must be migrated, information security needs to be seen to, and employees' competencies must be refreshed. A continuous flow of resources is required in order for organizations to carry out this work. A reduction of resources is a direct threat which will sooner or later lead to information loss.

In conclusion it can be stated that potential and actual information losses occur in many different ways. What is entirely unproblematic for one archive might constitute a great challenge for another. Examples of this include changes in format. At Archive B, these pose no threat, as this archive has the ability to affect the formats information is produced in. The situation is dramatically different at Archive C, where format changes are one of the absolute greatest challenges they have to overcome. Another example is that Archive B does not experience funding as a limitation in their preservation work, while Archive C is so hindered by it that they no longer feel that they truly preserve their material. Digital preservation work is continual and cannot be solved by technical solutions alone. Technological obsolescence is relentless, and personnel's knowledge must be kept up to date. Constant vigilance is also required to maintain IT security. Cloud services were not seen as an option at any of the archives which handled information categorized as sensitive due to legal or personal integrity issues. The development of new technology brings new risks along with it. What could be said here is that there is not one solution for digital information loss but many, and that these solutions will require constant readaptation.

8 Concluding remarks

In this section we will summarize the intent, research questions, methodology, results and analysis of our study. We will emphasize the most important aspects of the analysis and show how the interpretive frameworks allowed us to understand information loss at the three archives. The section concludes with suggestions for future research.

We have examined digital information loss at three archives in Sweden. These archives differ from one another in their function, preservation objectives and the rules governing them. They have been selected in order to access a spectrum of experience. This study has been carried out in the hope that it will aid in understanding digital information loss and the challenges it poses to these three archival institutions. It has sought the answers to two questions. The first of these asks what digital information has been lost or risks being lost. The second examines which circumstances have led to such losses or increased the risk of them. We feel it has been beneficial to our study to focus directly on information loss rather than digital preservation in general. Since the lack of an active digital preservation leads to information loss sooner or later, it is interesting to see what this loss actually looks like. The study's results have been interpreted from four frameworks. These interpretative frameworks helped us to understand the preservation paradigm of the archives, different aspects of digital information, what technical risks their digital information faced, and where in the organization information loss had occurred or was likely to occur.

Archives are one of the social institutions whose job it is to preserve information over a long period of time. With this objective in mind, we feel that it is beneficial to examine these institutions' challenges with digital information loss. Interviews were the best method available to us and allowed us to answer our research questions. The interview questions were meant to provide us with a more thorough understanding of the archives' digital preservation. These questions were meant to be more specific than the research questions themselves yet still within the same conceptual field. The questions were formulated as they were in order to reduce the complexity of the problem for the informants and clarify what we meant by digital information loss. It has been beneficial for us to use a broad definition of digital information, as it is not always entire digital objects which go missing, but parts, such as data or metadata. This definition also led us to examine how organizations manage their internal work documents.

Our results section has presented what we have deemed to be the most important results from the interviews. Each archive has been dealt with individually. This section presented the information flow at the organizations and how responsibility

was divided for digital information. Afterwards we discussed the central questions which were discussed at each archive: what constitutes an e-archive, how information loss is differentiated between analog and digital information, and what digital preservation entails. We also presented the digital preservation strategies of the archives, and factors that impact the dissemination of information.

The results were then analyzed from the four interpretive frameworks which made it possible for us to identify and discuss actual and potential information loss. The transition from the Records Lifecycle Model (RLM) to the Records Continuum Model (RCM) made it possible for us to see that Archive A has been forced to handle information according to the RLM. Archive B has focused more on the implementation of the RCM. An interesting result is that Archive B was well aware that the actual transition from RLM to RCM can mean information loss, since an organization cannot try to save older material simultaneously as it prepares the way for the implementation of RCM. The informant at Archive B clarified this would be a "Sisyphean task". Another interesting result is that Archive C cannot implement a digital preservation system modelled after RCM, due to the fact that the nature of their material means that they cannot control the format and information carrier of the digital material they receive.

Thibodeau's ontological theory enabled us to understand how the responsibility for digital information was shared at the different organizations. The theory also made it easier for us to see when parts, rather than whole digital objects were missing. This meant that we could identify a kind of information loss at Archive C which would have otherwise gone unnoticed. The material preserved by Archive C can be extremely sensitive to which settings are used in digitization as well as format migrations. An informant at Archive C explained that different settings can dramatically affect the end result when material is digitized. When an inappropriate format is chosen, the effect on the information can be so great that this can be considered information loss.

The authors of the SPOT model only intend for it to highlight risks with technical issues. We combined SPOT with the OAIS model in order to look at organizational as well as technical issues within digital information loss. This provided us with a satisfactory overview of the problems. Looking at how digital information flows through an organization we could show that there have been problems with ingest at archives A and B. An example would be that digital information has been handled as though it were analog, resulting in CD-ROMs simply stored in archival boxes, rather than being preserved.

Earlier research shows that public records are less likely to be identified as such within the digital realm. This makes it difficult to say where information loss has actually occurred. If a digital public record is not identified for preservation in the first place, it will eventually become unreadable or lost. However, if these records are never registered, how does one know that such losses have occurred?

This study has not set out to make a direct comparison between the three archives. Instead, it is meant to highlight the different challenges faced by the archives

concerning digital information loss. Something which is an insurmountable problem for one archive is a non-existent problem for another.

A summary of the actual and potential information loss we found at the three archives included:

- Loss of parts or whole digital objects during migration
- Loss of the connections between analog and digital information belonging to the same archive
- Loss of information due to it having been saved in an incorrect format
- Loss of data in connection with technological changes
- Loss of digital information when stored together with analog
- Loss of information due to obsolete hardware
- Loss of metadata due to databases written in code that is not open source

The reasons behind such actual and potential information loss were:

- Human error during the production of information
- An analog understanding and treatment of digital information
- A lack of organizational structure and strategies for digital preservation
- Lack of resources
- Technological limitations
- Lack of competencies amongst staff who produce digital information

While we focussed on the information which has been identified for preservation, material which falls outside the scope of such systems is often relevant. It may serve as contextual support information for other digital objects. Were an archive to lose some of their work documents, it would involve information loss for the rest of their collection. The understanding of the collection over a longer period of time is dependent on such descriptive documents.

Suggestions for future research might include a quantitative investigation into the readability of digital information at archives, which would provide insight into what portion of holdings in digital format are still readable. Some studies on this topic have already been carried out in Sweden, such as Ina-Maria Janssons (2012) Master's thesis "A lost cultural heritage? Digital private archives – problems, solutions and the future."[5]. Further investigation could also be carried out into the semantic division between e-archives and other material in digital format at archives, which could examine how such a division is understood by the archives personnel.

[5] Authors' translation, original title "Ett förlorat kulturarv? Digitala personarkiv – problem, lösningar och framtid".

9 References

Allison, A., Currall, J., Moss, M. & Stuart S. (2005). Digital identity matters. *Journal of the American society for information science and technology,* 56(4), pp. 364–372.

Anderson, D. (2015). Historical reflections: The digital dark age. *Communications of the ACM*, 58(12), pp. 20-23.

Anderson, D., Delve, J. & Pinchbeck, D. (2010). Toward a workable emulation-based preservation strategy: Rationale and technical metadata. *New review of information networking*, 15(2), pp. 110-131.

Anderson, K., Samuelsson, G. & Jansson, M. M. (2011). Benchmarking information management practice and competence in Swedish organizations. In *Proceedings of the 5th European conference on information management and evaluation.* Como, Italy 8-9 September 2011, pp. 28-36.

Aspenfjäll, J. (2013). *Förstudie Statens servicecenter e-arkiv och e-diarium.* eBuilder. http://www.ltu.se/cms_fs/1.109114!/file/Forstudierapport_SSC.pdf

Asproth, V. (2005). Information technology challenges for long-term preservation of electronic information. *International journal of public information systems*, vol. 2005:1, pp. 27-37.

Beagrie, L., Charlesworth, A. & Miller, P. (2014). *The National Archives guidance on cloud storage and digital preservation: How cloud storage can address the needs of public archives in the UK*. First edition. Kew: The National Archives. https://www.nationalarchives.gov.uk/documents/archives/cloud-storage-guidance.pdf

Björk, L. (2015). *How reproductive is a reproduction?* Doctoral thesis. University of Borås. Borås: University of Borås. urn:nbn:se:hb:diva-881

Borghoff, U. M., Rödig, P., Scheffczyk, J. & Schmitz, L. (2003). *Long-term preservation of digital documents: Principles and practices*. Berlin: Springer.

Borglund, E. (2008). *Design for recordkeeping: Areas of improvement*. Doctoral thesis. Mid Sweden University. Sundsvall: Mid Sweden University. urn:nbn:se:miun:diva-204

Borglund, E. & Anderson, K. (2011). L.O.S.T records: The consequence of inadequate recordkeeping strategies. In Barzdins, J. & Kirikova, M. (red.) *Databases*

and information systems VI: Selected paper from the ninth international Baltic conference (DB&IS 2010). Amsterdam: IOS Press, pp. 271-282.

Brand, S. (1999). Escaping the digital dark age. *Library journal*, 124(2), pp. 46-49.

Consultative Committee for Space Data Systems (CCSDS) (2012). *CCSDS Recommended practice for an OAIS Reference Model*. Washington: CCSDS Secretariat.
http://public.ccsds.org/publications/archive/650x0m2.pdf

Cook, T. (1997). What is past is prologue: a history of archival ideas since 1898, and the future paradigm shift. *Archivaria*, 43, pp. 17-63.

Cook, T. (2007[1994]). Electronic records, paper minds: The revolution in information management and archives in the post-custodial and post-modernist era. *Archives & social studies: A journal of interdisciplinary research,* 22(2), pp. 300-328.

Cook, T. (2013). Evidence, memory, identity, and community: Four shifting archival paradigms. *Archival science*, 13(2), pp. 95-120.

Corrado, E. M. & Moulaison, H. L. (2014). *Digital preservation for libraries, archives, and museums*. Lanham: Rowman & Littlefield.

Dahlgren, A. & Snickars, P. (red.) (2009). *I bildarkivet: Om fotografi och digitaliseringens effekter*. Stockholm: Kungl. Biblioteket.

Digital Cultural Heritage Roadmap for Preservation (DCH-RP) (2014). *A roadmap for preservation of digital cultural heritage content*. http://www.dch-rp.eu/

Digisam (2014). *Digitalt bevarande vid kulturarvsinstitutioner: Nulägesanalys och framtida behov*. Stockholm: Digisam.
http://www.digisam.se/images/docs/rapporter/Digitalt%20bevarande%20vid%20kulturarvsinstitutioner_nulagesanalys%20och%20framtida%20behov.pdf

Digisam (2016a). *Vad menas med långsiktigt digitalt bevarande?* http://www.digisam.se/index.php/digitalisera/bevara [28 May 2016]

Digisam (2016b). *Ett digitalare kulturarv: Digisams verksamhet 2011-2015*. Stockholm: Digisam.
http://ettdigitalarekulturarv.digisam.se/Ett_digitalare_kulturarv_Digisam2011-2015.pdf

Digital Preservation Coalition (2008). *Digital preservation handbook.*

Digital Preservation Coalition (2016). http://www.dpconline.org/about [29 April 2016]

Digitaliseringskommissionen (2014). *En digital agenda i människans tjänst: En ljusnande framtid kan bli vår* (SOU 2014:13). Stockholm: Näringsdepartementet.

Erway, R. (2010). Defining "Born Digital". OCLC Research. https://www.oclc.org/resources/research/activities/hiddencollections/borndigital.pdf

Federal Agencies Digitization Guidelines Initiative (2014). Archival master file. http://www.digitizationguidelines.gov/term.php?term=archivalmasterfile [7 April 2016]

Gladney, H. M. (2007). *Preserving digital information*. Berlin: Springer.

International Association of Sound and Audiovisual Archives (IASA) Technical Committee (2005). *The safeguarding of the audio heritage: Ethics principles and preservation strategy*. Version 3. http://www.iasa-web.org/sites/default/files/downloads/publications/TC03_English.pdf

InterPARES 2 Project (2008). The InterPARES 2 project glossary. In *International research on permanent authentic records in electronic systems (InterPARES) 2: Experiential, interactive and dynamic records*. Duranti, L. & Preston, R. (eds.). Padova: Associazione Nazionale Archivistica Italiana. http://www.interpares.org/display_file.cfm?doc=ip2_book_glossary.pdf

Jansson, I.-M. (2012). *Ett förlorat kulturarv?: Digitala personarkiv – problem, lösningar och framtid*. Master's thesis. Institutionen för ABM. Uppsala universitet. Uppsala: Univ. urn:nbn:se:uu:diva-175831

Kallberg, M. (2013). '*The emperor's new clothes' Recordkeeping in a new context*. Doctoral thesis. Mid Sweden University. Sundsvall: Univ. urn:nbn:se:miun:diva-20217

Kennerly, J. (2007). [Fotografi] https://commons.wikimedia.org/wiki/File:CDRomPits.jpg [1 June 2016]

Kristiansson, G. (2002). Långsiktigt bevarande av digital arkivinformation. Attachment 2 in Arkivutredningen. *Arkiv för alla - nu och i framtiden: Betänkande* (SOU 2002:78). Stockholm: Kulturdepartementet.

Kulturdepartementet (2011a). *Digit@lt kulturarv: Nationell strategi för arbetet med att digitalisera, digitalt bevara och digitalt tillgängliggöra kulturarvsmaterial och kulturarvsinformation 2012 – 2015*. Stockholm: Kulturdepartementet. http://www.regeringen.se/contentassets/8ad272b4832140acae16f36fc82c8c6d/digitlt-kulturarv---nationell-strategi-for-arbetet-med-att-digitalisera-digitalt-bevara-och-digitalt-tillgangliggora-kulturarvsmaterial-och-kulturarvsinformation-ku11.015

Kulturdepartementet (2011b). *Uppdrag till Riskarkivet att inrätta ett samordningssekretariat för digitalisering, digitalt bevarande och digital förmedling av kulturarvet*. Regeringsbeslut 5. Stockholm: Kulturdepartementet.

Kulturdepartementet (2015). *Regleringsbrev för budgetåret 2016 avseende Riksarkivet.* Regeringsbeslut 13. Stockholm: Kulturdepartementet.

Kundra, V. (2011). *Federal cloud computing strategy.* Washington: The White House.
https://www.dhs.gov/sites/default/files/publications/digital-strategy/federal-cloud-computing-strategy.pdf

Kuny, T. (1997). A digital dark ages? Challenges in the of electronic prevention information. In *63rd International Federation of Library Associations and Institutions (IFLA) council and general conference*. Copenhagen, Denmark 31 August-5 September 1997.

Lawrence, G. W., Kehoe, W. R., Rieger, O. Y., Walters, W. H. & Kenney, A. R. (2000). *Risk management of digital information: A file format investigation*. Washington DC: Council on Library and Information Resources.
http://www.clir.org/PUBS/reports/pub93/pub93.pdf

Littman, J. (2007). Actualized preservation threats: Practical lessons from chronicling America. *D-Lib magazine*, 13(7).

Maron, N. L., Smith, K. K., & Loy, M. (2009). *Sustaining digital resources: An on-the-ground view of projects today*. Itaca case studies in sustainability.
http://www.sr.ithaka.org/wp-content/mig/reports/4.17.2.pdf

McLeod, J. (2008). *Risk Assessment: Using a risk based approach to prioritise handheld digital information.* The British Library.
http://www.bl.uk/ipres2008/presentations_day1/20_McLeod.pdf

Merriam-Webster Dictionary (2015). Digitization. http://www.merriam-webster.com/dictionary/digitization [15 April 2016].

Myndigheten för samhällsskydd och beredskap (2013). *Vägledning: Informationssäkerhet i upphandling*. Karlstad: Myndigheten för samhällsskydd och beredskap. https://www.msb.se/RibData/Filer/pdf/26589.pdf

Näringsdepartementet (2011). *It i människans tjänst: En digital agenda för Sverige*.
http://www.regeringen.se/contentassets/6136dab3982543bea4adc18420087a03/it-i-manniskans-tjanst---en-digital-agenda-for-sverige-n2011.12

Owens, T. (2012). All digital objects are born digital objects. *The Signal: Digital Preservation* [blog], May 15. http://blogs.loc.gov/digitalpreservation/2012/05/all-digital-objects-are-born-digital-objects/ [1 April 2016]

Perrin, J., Winkler, H. & Yang, L. (2015). Digital preservation challenges with an ETD collection: A case study at Texas Tech University. *Journal of academic librarianship*, 41(1), pp. 98-104.

Quisberg, H. (2008). *On long-term digital preservation informations systems: A framework and characteristics for development*. Doctoral thesis. Luleå Unversity of Technology. Luleå: Univ. http://epubl.ltu.se/1402-1544/2008/77/LTU-DT-0877-SE.pdf

Rhodes, S. & Neacsu, D. (2009). Preserving and ensuring long-term access to digitally born legal information. *Information & communications technology law*, 18(1), pp. 39-74.

Riksarkivet (1999). *Om gallring: Från utredning till beslut*. Solna: Riksarkivet. http://riksarkivet.se/Media/pdf-filer/gallring_webb.pdf

Riksarkivet (2011). *Förstudie om e-registrering och e-arkivering: Underlagsmaterial*. Stockholm: Riksarkivet. https://riksarkivet.se/Media/pdf-filer/Projekt/Forstudie_e-arkiv_e-diarium_Bilagor.pdf

Riksarkivet (undated a). *Projektet e-arkiv och e-diarium eARD* [fact sheet]. http://riksarkivet.se/Media/ pdf-filer/Projekt/Infoblad%20eARD.pdf

Riksarkivet (undated b). *Projektet e-arkiv och e-diarium, eARD gör det enkelare att följa sina ärenden* [fact sheet]. https://riksarkivet.se/Media/pdf-filer/Projekt/eARD_informationstext.pdf

Rosenthal, D. (2011). Are we facing a "digital dark age?". *DSRS's Blog* [blog], 8 February 2011. http://blog.dshr.org/2011/02/are-we-facing-digital-dark-age.html [9 January 2016]

Rowley, J. (2006). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of information science*, 33(2), pp. 163–180. http://jis.sagepub.com/cgi/content/abstract/33/2/163

Runardotter, M. (2007). *Information technology, archives and archivists: An interacting trinity for long-term digital preservation.* Licentiate thesis. Luleå Unversity of Technology. Luleå: Univ. http://pure.ltu.se/portal/files/375637/LTU-LIC-0708-SE.pdf

Runardotter, M., Mirijamdotter, A. & Mörtberg, C. (2007). Being an archivist in our times: Trying to manage long-term digital preservation. *International journal of public information systems*, vol. 2007:2, pp. 47-61.

Runardotter, M., Quisbert, H., Nilsson, J., Hägerfors A. & Mirijamdotter, A. (2006). The information life cycle: Issues in long-term digital preservation. *Arkiv, samhälle och forskning*, vol. 2006:1, pp. 17-29.

Schumacher, J., Thomas, L., Vandecreek, D., Erdman, S., Hancks, J., Haykal, A., Miner, M., Prud'homme, P.-A. & Spalenka, D. (2014). *From theory to action: "Good enough" digital preservation for under-resourced cultural heritage institutions.* Washington, DC: Institute of Museum and Library Services. http://powrr-wiki.lib.niu.edu/images/a/a5/FromTheoryToAction_POWRR_WhitePaper.pdf

SFS 1949:105. *Tryckfrihetsförordning*. Stockholm: Justitiedepartementet. English version: The International Relations and Security Network (2008). *The constitution of the kingdom of Sweden*. http://www.parliament.am/library/sahmanadrutyunner/Sweden.pdf

Statens servicecenter (2015). *Klartecken för statligt e-arkiv*. Press release 8 May 2015. http://www.statenssc.se/Documents/Klartecken%20f%C3%B6r%20statligt%20e-arkiv%2020150508.pdf

Statens servicecenter (undated). *Ny e-arkivtjänst bidrar till en effektiv e-förvaltning*. http://www.statenssc.se/VaraTjanster/Sidor/Ingen%20menyrubrik/E-arkivintervju-med-Thomas-P%C3%A5lsson.aspx [25 April 2016]

Thibodeau, K. (2002). Overview of technological approaches to digital preservation and challenges in coming years. In *The state of digital preservation: An international perspective*. Conference proceedings. Washington (DC), United States of America 24-25 April 2002, pp. 4-31. http://www.clir.org/pubs/abstract//reports/pub107

Upward, F. (1996). Structuring the records continuum. Part 1: Post custodial principles and properties. *Archives and manuscripts*, 24(2), pp. 268-285.

Upward, F. (1997). Structuring the records continuum. Part 2: Structuration theory and recordkeeping. *Archives and manuscripts*, 25(1), pp. 10-35.

Upward, F. (2000). Modelling the continuum as paradigm shift in recordkeeping and archiving processes, and beyond: A personal reflection. *Records management journal*, 10(3), pp. 115-139.

Upward, F. (2005). The records continuum. In McKemmish, S. (red.). *Archives: Recordkeeping in society*. Wagga Wagga: Charles Sturt University.

Vermaaten, S., Lavoie, B. & Caplan, P. (2012). Identifying threats to successful digital preservation: The SPOT model for risk assessment. *D-Lib magazine*, vol. 18, nr 9/10. http://www.dlib.org/dlib/september12/vermaaten/09vermaaten.html

White, K. L & Gilliland, A. J. (2010). Promoting reflexivity and inclusivity in archival education, research, and practice. *Library quarterly*, 80(3), pp. 231-248.

Wikipedia (2016a). Digital data. https://en.wikipedia.org/wiki/Digital_data [30 April 2016]

Wikipedia (2016b). Microsoft word. https://en.wikipedia.org/wiki/Microsoft_Word [30 April 2016]

Wikipedia (2016c). File format. https://en.wikipedia.org/wiki/File_format [30 April 2016]

Wikipedia (2016d). Denial of service.
https://sv.wikipedia.org/wiki/Denial_of_Service [25 May 2016]

Xiaomi, A. (2003). An integrated approach to records management. *Information management journal*, 37(4), pp. 24-30.

Yeo, G. (2007). Concepts of record (1): Evidence, information, and persistent representations. *The American archivist*, 70(2), pp. 315-343.

Yeo, G. (2010). Nothing is the same as something else: Significant properties and notions of identity and originality. *Archival science*, 10, pp. 85–116.

10 Division of labor

We originally planned to carry out our own individual studies on two different research topics. During our initial research we discovered a significant degree of overlap between our topics. This led us to the decision to combine them into one study. We have developed several areas together such as the identification of the research problems, intent, and research questions. While Arrick put more focus on the international perspective in the previous research, Anna-Maria has focussed more on the Swedish side. Together, we identified a list of common search words in Swedish and English. We read through the key literature and discussed it amongst ourselves. While Arrick placed more emphasis on interpretive frameworks, Anna-Maria spent more time with the analysis of the results. Both of us participated in the interviews. Anna-Maria posed the questions while Arrick transcribed and highlighted important sections. The first stage of the results and analysis were created together in the form of notes organized by topic. Anna-Maria then structured these into a first draft, which Arrick developed into a polished text. When the text was nearly complete, we both read it aloud together and made the necessary changes. The background, previous research and method were handled in a similar fashion.