

# **Digital Archiving**

## **- Contemporary Preservation**

Master's Degree in Computer Science 2012  
Design of Digital Environments

Therese Roksvåg Nilsen

January 30, 2012  
Østfold University College, Norway

This is a final report in the Master's Degree in Computer Science, Design of Digital Environments at Østfold University College, Norway. All use of this thesis must be according to the restrictions set by the University College.

Supervisor: Prof. Børre Ludvigsen.

# Abstract

**Keywords:** Archives, digital preservation, social medias, contemporary, social history,

Tommy Arvidson states in an article from 1999 that one should '*Depict contemporary life now, tomorrow it's gone*'<sup>1</sup> [Arvidson(1999)] This statement is supported by the fact that significant amount of information created in our times is born digital. As such it is ephemeral in nature and many people are not aware of the risk involved by not actively managing this information. A digital personal archive could be a solution to this problem.

This thesis is looking for an answer to the question; *Can a personal digital archive on a national scale be established, and if so by whom?*. To address this question a three-part structure has been devised leading up to a concluding argument.

Throughout the work on this thesis this author have become more aware of the pressing need to organize and preserve personal information in a sustainable manner. However, to ensure that personal material is securely managed we need a centralized trusted service where people can feel safe that the material is being handed by trusted professionals who will secure access to preserve it for generations to come. Hopefully this thesis is the first step in establishing such a service.

---

<sup>1</sup>Authors translation

# Acknowledgements

I would like to express my gratitude to the many people who have helped made this thesis possible.

I would like to thank my supervisor Børre Ludvigsen for making this a focus issue both for me and hopefully for many others. I thank him for giving of his time for guidance and support, and for letting me take a part of field journeys to both Horten and Mo i Rana.

On that note I would like to show gratitude to the employees at the National Library, Mo i Rana, for making us feel welcome, and a special thanks to Jonny Edvardsen and Svein Arne Solbakk for their time and for a wonderful guided tour throughout their facilities and into the mountain.

I would also like to give thanks to Hege Oulie and Jens Gold at Preus fotomuseum in Horten for their time and input, and to Torvill Solberg for helping me with literature search at their library. I show gratitude to Olaf Schjeldrup at UNINETT in Trondheim for giving of his his time and knowledge, and to Jeremy John at British Library for the same, giving me important research information and essential input. I am also grateful to Aslaug Bendiksen at Datatilsynet for information on some of the legal aspects in relating to this thesis.

I would like to thank Østfold University College for the opportunity and support on field journeys, and to the staff at the Computer Science department for help when needed. I would also like to give special thanks to Christian Heide for all the support and help throughout my years as a student at the University College.

I would like to thank my family for all the support throughout my life. For always loving me, helping me and being there for me in every way possible, and for shaping me into the person I am today.

In the end I would like to show my deepest gratitude to my partner Arne Solvang. He has made available his support in a number of ways, and stood by my side always. Without him none of this would be possible.

# Preface

This paper sets out to answer a question which have gained momentum in later years and requires an urgent response, namely;

*Can a personal digital archive on a national scale be established, and if so by whom?*

To answer this question the thesis has been divided into three distinct parts before presenting a summarized concluding argument.

The question of ephemerality tackles issues such as how technological transition contributes to the need for a digital archive of personal material, and how data organization and integrity should be handled to ensure its longevity. The role of commercial actors is then debated and an indication of its fleeting nature is established. This section is concluded with a review of how measures of ephemerality have been enacted in Norway and abroad.

Having established the nature of ephemerality and the challenges it presents the next step is examining how this can be combated through preservation. The preservation chapter investigates a set of preservation strategies and current approaches. It also touches upon relevant legal aspects which could aid or impede preservation efforts.

The two first parts of the thesis have first highlighted the concept of ephemerality and dealt with the consequent need for preservation, the final step is then to tackle the complex issue of trust. Trust is paramount for an archive of this scale and the chapter tackles such issues as who should own the archive, what can be published and how can rights of third parties be protected.

Together these parts have scrutinized the nature of ephemerality, it has reviewed several types of solutions within the field of preservation and it has finally discussed the element of trust in relation to a personal digital archive on a national scale. The findings from these section are all brought together in the final section *A personal national archive*. This section ties back to the initial question

*Can a personal digital archive on a national scale be established, and if so by whom?* The answer is multifaceted and as such the section is divided into several sub sections. The short answer is yes, and it would have to have government involvement and work within the framework set forth by said government. In addition it should be based around several of the best practices identified in this paper.

# Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgements</b>	<b>ii</b>
<b>Preface</b>	<b>v</b>
<b>Contents</b>	<b>vii</b>
<b>List of figures</b>	<b>viii</b>
<b>List of tables</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	2
1.2 Methodology . . . . .	3
1.3 Case history . . . . .	7
1.3.1 Past . . . . .	7
1.3.2 Present . . . . .	9
1.3.3 Future . . . . .	11
<b>2 Ephemerality</b>	<b>13</b>
2.1 Technological transition . . . . .	13
2.1.1 The Domesday Project . . . . .	16
2.2 Data integrity and organisation . . . . .	18
2.2.1 Management strategies . . . . .	21
2.3 The transience of commercial services . . . . .	23

2.4	Measures against ephemerality . . . . .	25
2.4.1	Historical countermeasures . . . . .	25
2.4.2	The National Library . . . . .	27
2.4.3	Practices in other nations . . . . .	30
<b>3</b>	<b>Preservation</b>	<b>33</b>
3.1	Prerequisites . . . . .	33
3.1.1	Strategies . . . . .	35
3.1.2	Sustainability . . . . .	38
3.2	Current approaches . . . . .	40
3.2.1	Digitalization projects . . . . .	41
3.2.2	Social networking platforms . . . . .	46
3.2.3	Web harvesting . . . . .	47
3.2.4	Other projects . . . . .	48
3.3	Open formats and standards . . . . .	52
3.4	Legal aspects . . . . .	53
3.4.1	Review of relevant laws . . . . .	54
3.4.2	Legal impact . . . . .	60
<b>4</b>	<b>Trust</b>	<b>63</b>
4.1	Ownership . . . . .	63
4.1.1	Governmental trust . . . . .	66
4.1.2	Archiving selection . . . . .	70
4.2	Publication . . . . .	71
4.3	Rights of third parties . . . . .	73
<b>5</b>	<b>A personal national archive</b>	<b>77</b>
5.1	Overview . . . . .	78
5.2	Legislation . . . . .	79
5.3	Technology and security . . . . .	80
5.4	Association and access . . . . .	82
5.5	Accumulation . . . . .	83
5.6	Integrity and ethics . . . . .	84
5.7	Roles . . . . .	84



5.8	Contribution to society . . . . .	86
5.9	Conclusion . . . . .	87
	<b>Bibliography</b>	<b>89</b>
<b>A</b>	<b>Design methods</b>	<b>99</b>
A.1	Systematic search . . . . .	99
A.2	Literature searching . . . . .	100
A.3	Classification of design information . . . . .	100
A.4	Specification writing . . . . .	101
<b>B</b>	<b>Interview, Olaf Schjelderup at UNINETT</b>	<b>102</b>

# List of Figures

1.1	Outline of digital preservation . . . . .	3
1.2	Research question . . . . .	4
1.3	Thesis process . . . . .	6
2.1	Dilemma of Modern Media . . . . .	14
2.2	Traditional Preservation Versus Digital Preservation . . . . .	16
2.3	Domesday book . . . . .	17
2.4	Entrance to the archive inside the mountain . . . . .	28
2.5	Example of the archives shelf system . . . . .	28
2.6	NB's digitized archival library . . . . .	29
2.7	NB's 2Pb digitized library backup . . . . .	29
3.1	Digital archiving concerns . . . . .	34
3.2	Digital life-cycle . . . . .	35
3.3	Digitization workflow of the First World War Poetry Digital Archive . . . . .	43
5.1	Model of a personal national archive . . . . .	78
5.2	Model for sustainable preservation . . . . .	81

# List of Tables

4.1 List of responses . . . . .	76
---------------------------------	----



# Chapter 1

## Introduction

The process of active management to ensure accessibility to digital content in the future is called digital preservation. Digital longevity is potentially fleeting because of rapid change in technology leading to technical obsolescence, increasing accumulation of digital information and random approaches to management and organization of digital content. To ensure accessibility in the future actions has to be taken in the present. Unlike paper documents, digital material will not sustain long-term by being neglected. Ongoing active management is required. [Beagrie et al.(2008)Beagrie, Semple, Williams, and Wright]

Professor Børre Ludvigsen at the Østfold University College in Norway started looking at this issue many years ago, and summarized it in the document, *Norske Mediearkiv*. [Ludvigsen(2011)] Ludvigsen feared that the period of time we are in now, with all the technology we have at hand, will end up being less documented than prior times. Ludvigsen's vision is a digital archive for the general public of Norway. An archive for everyone to store the digital material they wish to keep for the future. [Ludvigsen(2011)]

This thesis will look at several aspects related to bringing his vision to life. Initially the general concepts of digital preservation will be outlined along with a background of personal archiving endeavours. This is to establish context so the reader can place a personal archive in relation to previous efforts and evaluate it accordingly. This report will then go into topics such as the concept of ephemerality, preservation, and trust. These segments will take a closer look at the various issues facing digital preservation of personal material, outline the current views and research as well as highlighting potential pitfalls and best practice. In addition to this, various field research with

major Norwegian and international institutions along with corresponding interviews will make up the bulk of the material need to conclude with the final segment of the report, an overview of the peoples archive. This final segment will act as a concluding remark highlighting the recommended courses of action on the basis of the findings from the previous section. The final product will serve as a stepping stone for further research into the topic and hopefully act as a foundation for the eventual implementation of a peoples digital archive. This report will commence by taking a step back, and providing a brief overview of the concepts of preservation, digital preservation and personal information management. Together they form the basis for the next focus of the report which is detailed topics related to personal digital information archiving, and by introducing the underlying concepts, it is hoped that the subsequent discussion on specific topics within personal digital archiving will be easier to follow.

## 1.1 Background

In addition to Ludvigsen, many other scholars and practitioners have taken an interest in the same problem. Some describe the current development in personal archiving as a digital revolution. Jeremy John of the British Library states that a digital revolution is transforming the nature of personal archiving. It is no longer just the rich and famous that can have their personal life documented. International Data Corporation<sup>1</sup> estimated that 70% of material in the digital universe would be created by individuals. Despite the opportunities at hand John is concerned about the sustainability of personal archiving. [John(2009)]

The management of personal information is called Personal Information Management (PIM)<sup>2</sup> and can be looked at as a combination of long-term storage of digital objects, preservation of the objects and enabling access to the objects now and in the future. [Marshall(2007)]

Among the most prominent threats to digital preservation are storage media's short lifespan, technological obsolescence, and slow access to old media. Though the initial selection and installation of software components in a digital repository can go some way to combat this, the storage infrastructure is the crucial component. [Arora(2006)]

In the past information was often contained in formats such as books, maps, photographs and

---

<sup>1</sup><http://www.idc.com/>

<sup>2</sup>[http://en.wikipedia.org/wiki/Personal\\_information\\_management](http://en.wikipedia.org/wiki/Personal_information_management)

recordings which could be interpreted by the naked eye, however today information can take the form of several digital formats and interpretation will require appropriate hardware and software. [Arora(2006)]

Several scholars has at different times suggested that modern society is entering a digital dark age because the lack of coherent strategies to long-term digital archiving. The concern is not just for digital material in the public eye, but also for personal digital belongings. [Marshall(2007)]

The key concepts of digital preservation is outlined in figure 1.1.

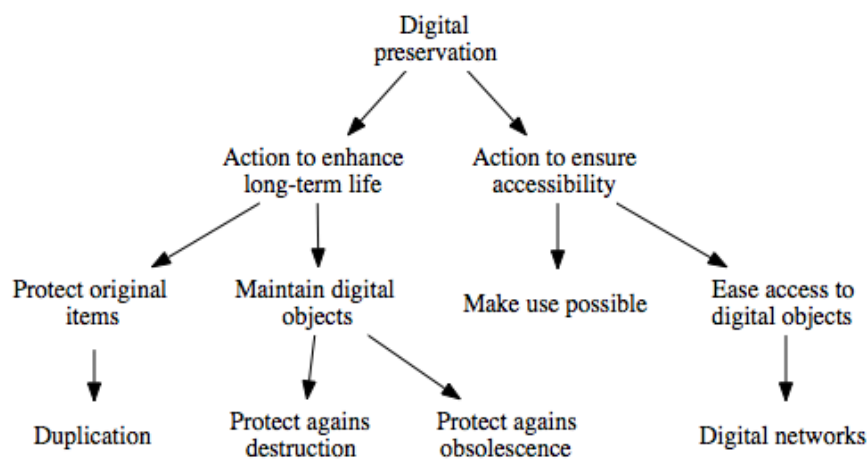


Figure 1.1: Outline of digital preservation based on [Arora(2006)]

## 1.2 Methodology

This thesis is based on the proposal *Norske Mediearkiv* by Børre Ludvigsen at Østfold University College. In this project the focus has been to identify works, projects and publications on the subject of personal archiving and the importance of personal contemporary preservation. The thesis have been established by design methods and fieldwork mainly based on published literature. The design methods used have not been followed to the letter, but have guided and supported the process throughout the timeframe. The mentioned methods can be seen in appendix A.

In figure 1.2 the research question and subsequent breakdown of the thesis is illustrated leading up to a concluding section segment.

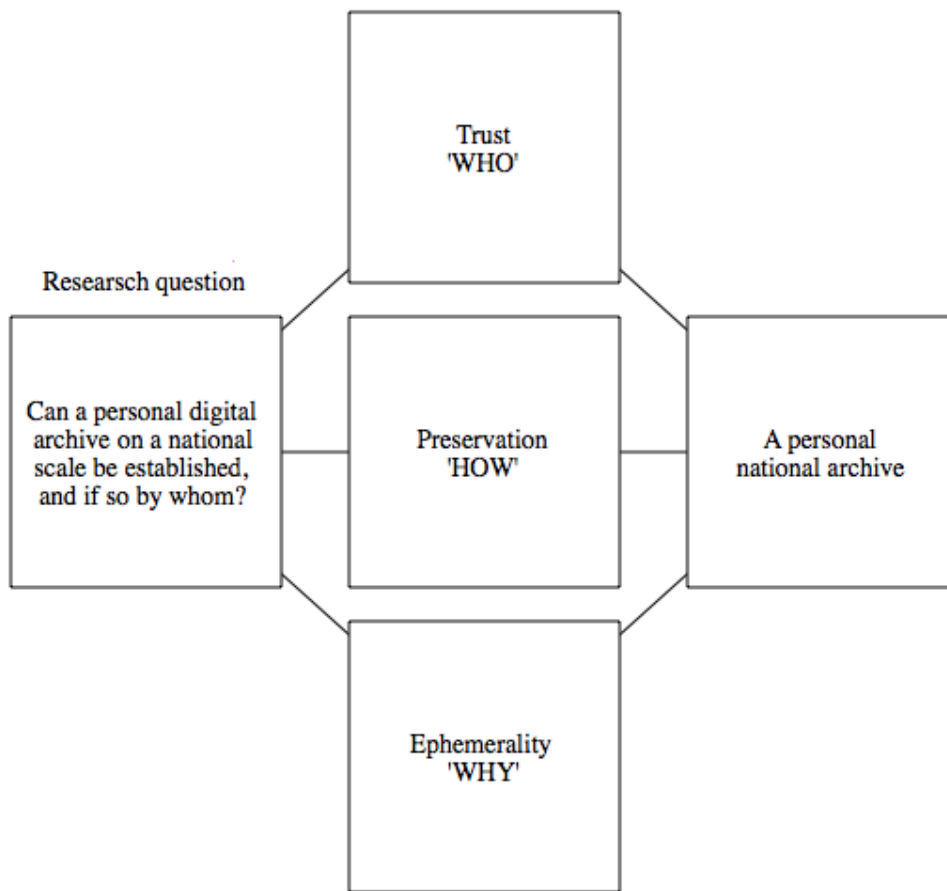


Figure 1.2: Research question



Fieldwork that have been carried out include visits to libraries, museums and archives, as well as visits and meetings with institutions that potentially could be facilitators for *Norske Mediearkiv*. Field trips include;

- British Library<sup>3</sup>, London. *For literature and meeting with Jeremy John.*
- Preus museum<sup>4</sup>, Horten. *Meeting representatives with Børre Ludvigsen and for literature*
- National Portrait Gallery<sup>5</sup>, London. *Inspiration*
- John Soane museum<sup>6</sup>, London. *Inspiration*
- UNINETT<sup>7</sup>, Trondheim. *Meeting for input, potential facilitator*
- The Data Inspectorate<sup>8</sup>, Oslo. *Meeting with representatives, potential facilitator*
- The National Library<sup>9</sup>, Mo i Rana *Meeting representatives with Børre Ludvigsen, guided tour of the facilities, potential facilitator*
- National Archives<sup>10</sup> in Kew, London. *Inspiration, Domesday, archive, Catalog day*

As mentioned has the main focus in this thesis been published literature. As such there will not be given an ordinary literature review as the thesis it self contains review of the literature in each section.

---

<sup>3</sup><http://www.bl.uk/>

<sup>4</sup><http://www.preusmuseum.no/norsk/index.php>

<sup>5</sup><http://www.npg.org.uk/>

<sup>6</sup><http://www.soane.org/>

<sup>7</sup><http://www.uninett.no/>

<sup>8</sup>Datatilsynet, <http://datatilsynet.no/>

<sup>9</sup>Nasjonalbiblioteket, <http://www.nb.no/>

<sup>10</sup><http://www.nationalarchives.gov.uk/>

Figur 1.3 gives an overview of the process.

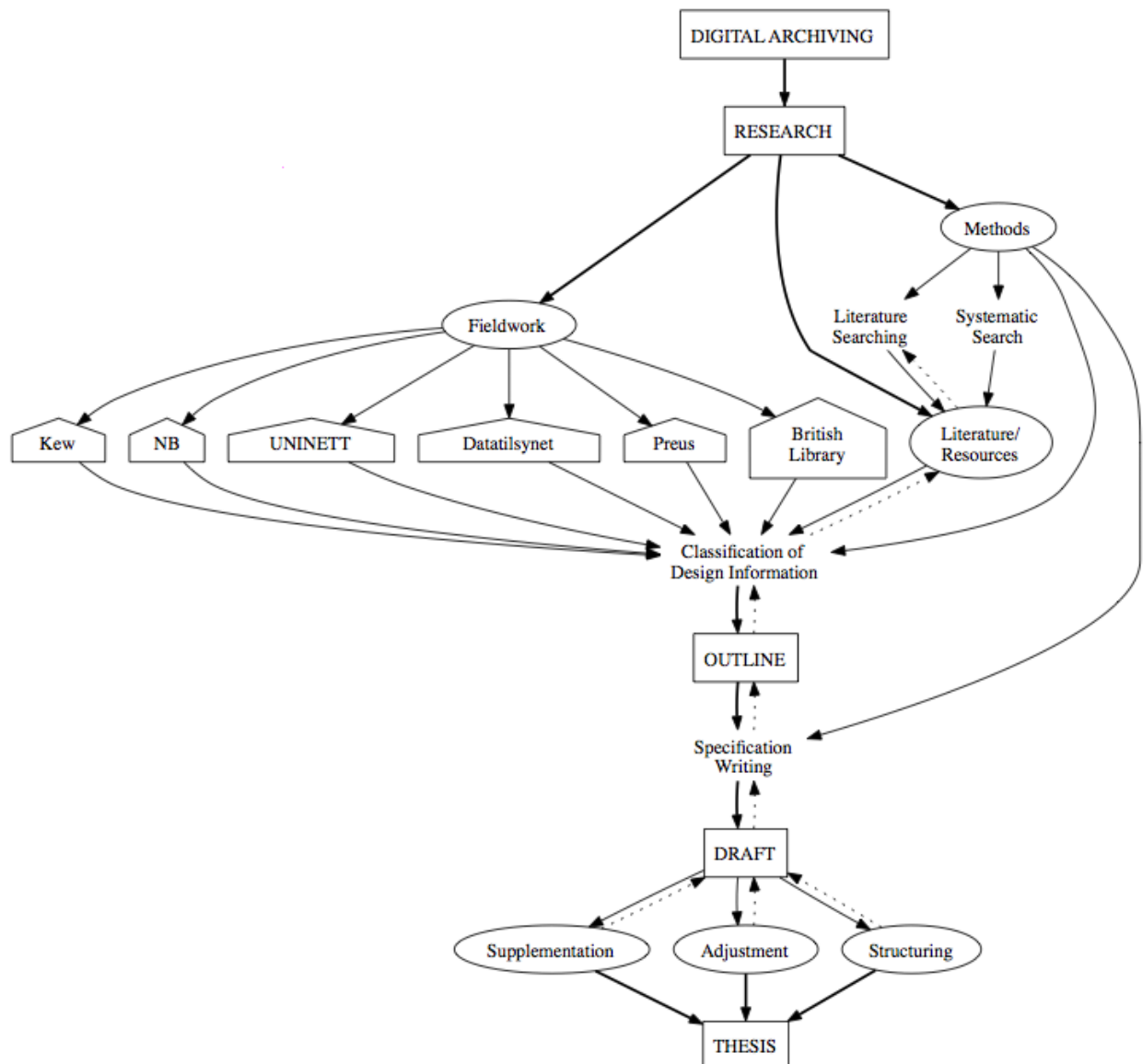


Figure 1.3: Thesis process

## 1.3 Case history

This section will provide a brief overview of how personal archiving have been pursued in the past, present and what the future may hold. It is meant to provide a reference point for this project, and illustrate how individuals have approached personal archiving throughout the ages.

### 1.3.1 Past

Throughout history man has documented his existence with images, illustrations and text all as new inventions has come to be. This section will look at how private individuals in the somewhat distant and recent past have documented parts of their lives through text and how private photos laid the basis for photo historical research in Norway.

#### 1.3.1.1 Photography

In Norway the basis for photo historical research was laid by Ragna Sollied a librarian at Bergen public library. Private book collections occasionally included old photo albums. Sollied recognized their historical value and retained the albums that normally would have been thrown away. She stored the photographs in shoeboxes, categorizing them by photographer. In 1967 she published the book *Eldre bergenske fotografer*, making the foundation for many other publications and further research. [Larsen and Lien(2007)]

#### 1.3.1.2 Diaries

Diaries and personal letters details the contemporary in a fresh, intimate, plain-spoken and lively way. Giving the reader a window to the past, showing not just the times gone by, but also the similarities in todays hopes and worries, and a clear picture of who the writer was. [Stowe(2002)] This section will give a brief description of three diaries. One written in the 17 century, one from the middle of the last century, and one written at the end of the last decade.

**Samuel Pepys, 1660-69** During 1660-69 Samuel Pepys kept a diary recording his own life and the world around him. He wrote his diary in shorthand in a time of censorship and his writings

were never intended for the public eye. [Pepys(2003)] It is however available in most book stores today.

Pepys wrote by hand on paper and his writings were published almost 200 years later illustrating the durability of these analog journals [Tomalin(2003)]. In contrast research done by the British Library in the project *Digital Lives* shows nearly 30% of participants have experienced serious computerized data loss at home. [John et al.(2010)John, Rowlands, Williams, and Dean]

**August Andersen, 1940-45** In April 1940 the editor of the newspaper *Farsund Avis* urged the people in the area to write about what was happening.

*'Everyone should write a journal. We are experiencing something big. But unfortunately our memories only last for a short time. Let's spend a few minutes each day to secure the memories and experiences on paper. In a few years they will be invaluable'*<sup>11</sup>

[Einar Holmer Hoven (1940), cited in Kåre Rudjord Preface [Andressen(1986)]]

August Andersen started writing letters to his brother in America and he kept writing a letter every week throughout the war. The stories are not all historically accurate. His source material was what he heard from the people he met, illegal radio from London and German propaganda, and he had no opportunity to verify what he was told. Despite this his letters provide a valuable insight into the occupation from the uncensored viewpoint of a common man. August died in 1946, and his manuscript of letters were published 40 years later. [Kåre Rudjord Preface [Andressen(1986)]]

To read the thoughts of a man on the sidelines of a war one has to get the book from the depot of the National Library, and because of the Legal Deposit Act it will be available forever.

**Regine Stokke, 2008-09** In August of 2008 a young girl from Kristiansund, Norway got the tragic news of Acute myeloid leukemia. She started blogging about her condition on November 4 2008 and kept on through out her illness. She died on December 3 2009, 18 years of age. In March of 2010 her blog, a selection of comments and pictures, and contributions from her family and friends was published as a book. [Stokke(2010)]

Regines book differs from the other diaries highlighted in this thesis. Regine wrote her blog post for people to read. Like the others she is documenting a condition and what is happening around

---

<sup>11</sup>Authors translation

her, but the other diaries were written in secret. Samuel Pepys wrote in code in a time of censorship. August Andreassen wrote about and during time of war, violence and censorship, and all though Andreassen was consciously writing to document what was going on during the war he was hiding his letters. But Regine Stokke were posting her writings on the Internet for all to see, and people saw, in 2009 her blog was one of the most read in Norway. Still would it have been here for all to see in 10 years if it had not made the transition from blogosphere to publication?

### 1.3.2 Present

The emerging problem of information overload, which is only heightened by the digital environment, have prompted individuals to look for new methods of managing their personal information in a more efficient manner. [Kim(2010)] This section will examine some of the prevailing trends within personal archiving today, as well as a few examples which will be discussed further in later chapters.

#### 1.3.2.1 Social networks

A growing trend is to have remote services acting as hubs for personal digital collections. These services can provide security, management tools, and allow users to publish their collections with friends or even the general public. [Beagrie(2005)]

The increased desire and openness to share personal information and communicate with the outside world have lead to a rise in software tailored for the creation and publication of blogs or digital photographs. These items or communiqués are then shared through a third party provider such as Flickr<sup>12</sup>. Such services offers various options for categorizing, searching, and publishing digital information. [Beagrie(2005)]

The increased adaptation of such activities could suggest that the general public would be open to continue publishing personal information also for the long-term, and the familiarity with categorizing images, grouping them, and publishing material regularly may go some way in making people aware of the basic concepts of archiving.

Another interesting development in the social networking sphere is the establishment of services

---

<sup>12</sup><http://www.flickr.com/>

such as Ourmedia<sup>13</sup>. This service allows individuals to publish and store music, video and other personal media free of charge of an indefinite amount of time, the only caveat is that the information is open to the general public. [Beagrie(2005)]

### 1.3.2.2 Digitalising

A common approach for moving analogue archives into the digital realm is to embark on a digitalization project. There are numerous examples of high profile projects, often conducted by notable institutions, such as a nations national archives or library. The concept of digitalization will be discussed throughout this thesis, as such only one example is highlighted here to illustrate how digitalization can be conducted and for which purpose.

**The First World War Poetry Digital Archive** The First World War Poetry Digital Archive<sup>14</sup> is an example of a project which have successfully managed to digitize a vast amount of analog material from a specific era and subject matter. It has processed over 4500 items and made these accessible to the general public. It is interesting to note that the public have also contributed directly, adding over 6500 items to the archive through a digital submission tool. [Lindsay(2009)]

The aim of this project was to provide an online portal for a collection of digitized material which were previously largely inaccessible in its analog form. The subject matter was World War One poetry and supporting material. [Lindsay(2009)]

### 1.3.2.3 Lifelogging

Another recent trend is that of life logging, attempting to capture nearly all aspects of an individuals life through the use of technology. A prime example of this approach is the MyLifeBits project.

**MyLifeBits** Microsoft's MyLifeBits<sup>15</sup> project illustrates that digital capture and storage is now at a stage where one can feasible capture nearly all areas of an individual's life digitally. Through a

---

<sup>13</sup><http://www.archive.org/details/ourmedia>

<sup>14</sup><http://www.oucs.ox.ac.uk/ww1lit/>

<sup>15</sup><http://research.microsoft.com/en-us/projects/mylifebits/>

continuous digital capture device, digitization of analog materials and archiving of born digital<sup>16</sup> content the project have been able to archive a significant portion of the alpha user's life. [Beagrie(2005)]

#### 1.3.2.4 Personal archiving

**Paradigm** Another research project focused on personal digital archiving is the PARADIGM<sup>17</sup> project sponsored by the JISC<sup>18</sup>. This project sets out to define a best-practice for ensuring long-term access to politicians personal digital archives. [Beagrie(2005)]

### 1.3.3 Future

It has to be noted that given the rapid technological developments in the field, the future of digital archiving is hard to predict. The examples below highlights potential outcomes suggested by experts in the field.

Jeremy John envisions a future where digital objects can be classified and tagged to certain aspects of a personals life upon its initial creation. For example events such as starting school, changing jobs, communications with certain persons could all be automatically categorized and as such be woven into the right section of a personal archive. [John(2008)]

Personal digital archiving research is in its infancy and the focus so far has mainly been on capturing the archives of famous individuals or at an institutional level. This is slowly changing and Cox (2009) highlights the potential of "citizen archivists" where people grow accustomed to administrate and manage personal digital collections through the adaptation of software specially designed for these tasks. The importance of having regular people acting as archivists of their own personal digital collections is considered crucial if the information is to be preserved long-term. [Cox(2009)]

**LPWS** Lifetime Personal Web Space is an ambitious vision consisting of a dedicated webspace where the user can store his or hers entire life's worth of digital information. It is proposed to

---

<sup>16</sup><http://en.wikipedia.org/wiki/Born-digital>

<sup>17</sup><http://www.paradigm.ac.uk/>

<sup>18</sup><http://www.jisc.ac.uk/>

function as one central hub with several interconnections which allows access to external entities. The system will store information that was important in the users past in a searchable archive. As the information age it is envisioned that older files will be accessible in a basic version while newer additions will take advantage of current technology. [Cohn and Hibbitts(2004)]

The user could choose to make content private or public, or even select its audience from its network. It should also be possible to connect to other users and thus form a network with friends and colleagues. It is also proposed that when a user dies the record could be stored for posterity as a record of that users life and contributions to society. [Cohn and Hibbitts(2004)]



## Chapter 2

# Ephemerality

Ephemeral is defined as '*short-lived*' or '*transitory*'<sup>1</sup>. As such a sustainable archive should aim be anything but ephemeral. This chapter will examine the inherent ephemerality of digital media and what has been done to combat it. It has been divided into the sections of technological transition, data integrity and organization, the transience of commercial services and concludes with an overview of measures against ephemerality.

### 2.1 Technological transition

One of the prevalent problems facing the establishment of a digital personal archive (on both a personal and institutional level) is in many ways connected with the move from paper to bits. This entails how one increasingly accumulate information in digital form, and how this presents new challenges, as well as opportunities, which were not present in the analog age.

Arora (2006) proposes that preservation is one of the primary weakness of the digital resources. Unlike the preservation of analog material which could be stored relatively easily and needed little to no continued attention, digital material require regular updates to combat obsolescence of software or hardware. [Arora(2006)] In other words whereas benign neglect could often be a sustainable (if unintended) means of storing analog material, in a digital environment such practice is but assured to put all the material in jeopardy over time. [Marshall(2008a)] In the past it was possible and likely

---

<sup>1</sup><http://dictionary.reference.com/browse/ephemeral>

quite common to gather personal information in a box or other means of physical storage and place it in a safe place, where bearing any major incidents this material would be available for generations to come. In the case of digital media even the same process of storing the digital files in a physical location would likely be insufficient due to the physical failure of the hardware (ie computer, hard drive, dvd's etc) or the obsolescence of the chosen storage format itself. [JISC(2009a)]

Figure 2.1 provides an illustration of the relationship between storage space and life expectancy throughout history. As can be seen we have the dilemma where storage capacity is growing exponentially while life expectancy is diminishing.

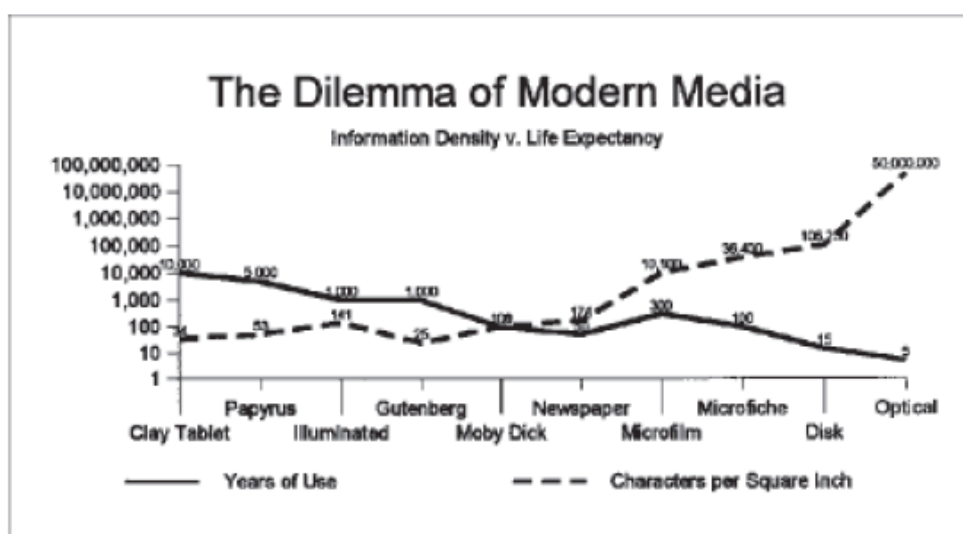


Figure 2.1: The Dilemma of Modern Media [Arora(2006)]

The challenge and need of moving from an analog to a digital world is further highlighted by concrete actions by established institutions. For instance in 2000 The British Library coined the term '*eManuscripts*' for the digital equivalent of letters, diaries, notebooks, family photographs and films. In addition the institution started to collect these digital entities as part of its overall holdings. [John(2008)]

On the other side of the Atlantic the Library of Congress<sup>2</sup>, in 2000, appropriated \$100 million for digital preservation initiatives, where the aim was to ensure that '*born digital*' information would be accessible for both current and future generations. While the learnings from the case example

<sup>2</sup><http://www.loc.gov/index.html>

of the Library of Congress will be discussed in further detail at a subsequent level of this report, its actions illustrates that as early as the year 2000 major institutions started to take note of the shift from analog to digital accumulation of personal information, and actively set out to address this. [Library of Congress(2011)]

Despite the apparent identification of the challenges of moving from analog to digital form as well as the subsequent actions of some of the leading archival institutions, we are still far removed from fully addressing the challenges connected to the shift from a analog to a digital environment. There are numerous examples from our recent past of how the move from paper to bits have jeopardized significant amounts of information. One such example is that of the UNIVAC<sup>3</sup> computer which were used to store data from the 1960 Census. Once this material was handed over to the National Archives in the 1970s the UNIVACs had become obsolete and the material could only be retrieved at a significant cost. In addition early data from space missions and satellite images were not retrievable and is likely lost forever. [Barksdale and Berman(2007)] These are just some of the examples that highlight what could be the worst outcome if the challenges of collecting and preserving digital material is not addressed adequately.

While Marshall (2008) warns of the potential of a coming digital dark age, where digital information is lost forever [Marshall(2008b)], and JISC argues that there is currently no digital media which is as reliable or durable as some of its analog counterparts, we should not ignore the fact that digital information also offers significant benefits and that avoiding the challenges is not an option, they must be addressed. [JISC(2009a)] In fact digital media will likely lead to a more diverse set of information, whereas in the analog past the recording of daily life was a privilege only afforded to the rich. [John(2009)] In addition the ease of which digital information can be created and stored allows insight into new aspects of human life. [Kim(2010)] Through continual preservation and active management one can assure that 'born digital' content survives between several generations of both men and hardware. [Library of Congress(2011)] This paper will continue to highlight these challenges in more detail and offer suggestions as to how they can be addressed in a national digital archive.

Figure 2.2 illustrates how digital material requires frequent management, unlike its traditional counterparts.

---

<sup>3</sup><http://en.wikipedia.org/wiki/UNIVAC.I>

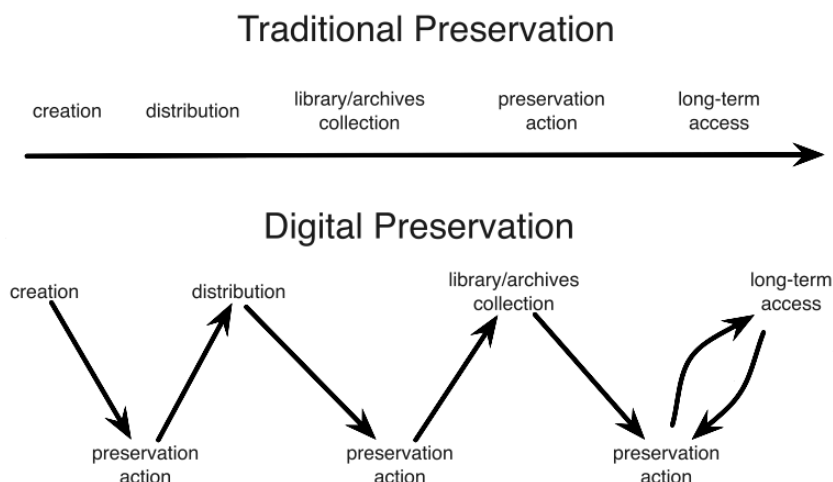


Figure 2.2: Traditional Preservation Versus Digital Preservation [Library of Congress(2011)]

### 2.1.1 The Domesday Project

One example of the difference in durability between analog and digital material is the story of the Domesday project. The efforts to collect data from contemporary society may have been 900 years apart, but the longevity of the two projects presents an intriguing set of questions regarding the challenges of moving from an analog world into the digital realm.

**The original Domesday project** The original Domesday project was conducted in 1086 and is regarded as the first great work of a bureaucratic state. King William I of England commissioned an overview of the land after 20 years of rule. [Finney(1986)]

The book was hand-written in black and red ink, on sheep-skin parchment, most likely by one monk and is still available at the UK National Archives in Kew. [The Domesday Book Online(2011)] Figure 2.3 shows the current state of the book during a field visit to the National Archive.

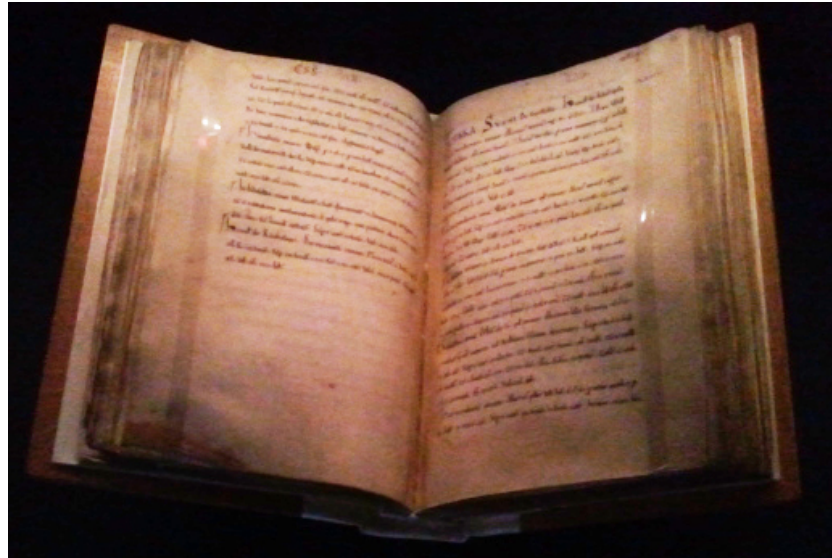


Figure 2.3: Domesday book at display in National Archives in Kew, London. Photo: TRN

**The Domesday project 900th anniversary** In 1986 it was 900 years since the Domesday project and the BBC wanted to highlight this occasion with a Domesday project of the 20th century. In 1983 Peter Armstrong a BBC Television producer started a project establishing a database of how the British look at Britain in 1986, a snapshot of modern England. [The Domesday Book Online(2011)] It was a large scale interactive project using microcomputers to engage schools around Britain to each produce a database from their area. [Finney(1986)]

Approximately a million people around the UK contributed [The Domesday Book Online(2011)] and the material was stored away in various locations on four tapes. These tapes had not being played for about 17 years when engineers at the BBC Archives wanted to copy them. Three tapes were flawless, but the fourth showed signs of deterioration and fine scratching. [Finney(1986)]

The material was also compiled and sold on computer disks. The chosen disks were only compatible with very expensive computers and subsequently not many disk were sold. Today only a few remain, in various conditions. [The Domesday Book Online(2011)]

The Domesday project book from 1086 has not change much since it was written and it's accessibil-ity is about the same as it was 1000 years ago, but technological advancement and copyright issues have meant that the second project is only partly available after great costs were incurred to restore it. [Misund and Høiberg(2001)]

These two projects provide an interesting illustration of the complexity and challenges of obtaining, storing and presenting digital information on a grand scale while at the same time ensuring its longevity. While the original Domesday project have survived the test of time it has not been available to a great number of people, whereas the second Domesday project had ambitions to not only capture contemporary life in Britain but also share its findings with as many people as possible. It provides a clear example of the benefits of digital information, (scope, accessibility, diversity) while at the same time illustrating some of its main challenges (longevity, technological obsolescence and copyright issues). These are all important things to consider when continuing work on a national personal archive. It can be argued that one should aim for the longevity of the original while pursuing the ambitions of the second Domesday project.

## 2.2 Data integrity and organisation

Digital materials are at an increasing risk of being lost due to factors such as processing and continual migration, evolving and incompatible operating systems, a myriad of file formats, ad hoc backup and organization of material as it accumulates. This section will review some of the leading causes where data integrity and organizations contribute to the ephemerality of the digital content.

The physical aspect of digital material can act as a risk factor to data integrity. One of the biggest differences between paper and digital content is machine dependency. Digital content can only be accessed using machines. The appropriate hardware and software has to be available to access the material. Furthermore, digital storage medias are highly fragile, and magnetic and optical deterioration is a constant threat because of physical exposure, such as heat, humidity and dust. Climatic conditions can be harmful but may not leave any physical signs of damages. The damage can lead to corruption of the digital content but it may not be identifiable which fraction has been corrupted. It is therefore crucial that measures for digital preservation is taken at an early stage. [Arora(2006)]

As new formats have grown to become common it often emerges that the life span of most digital devices is discouragingly short, even more so that initially anticipated. People store their life on computers, iPads, mobile phones, DVDs, memory cards, memory sticks and so on, but format and hardware changes combined with the risk of damages to hardware and viruses can result in dramatic loss. 5 1/4 –inch floppy disks, highly pervasive a decade ago, is today as good as obsolete, and a lot of information is lost forever. [Barksdale and Berman(2007)]

Arora (2006) identifies the biggest threat to digital content as techno-obsolence. Three to five years is the typical timeframe for migration because of continuous change and technological development. [Arora(2006)] To summarize it seems apparent that the physical nature of digital material poses a higher risk of data loss than its analog counterparts. Whereas disastrous events, theft or simple misplacement could be causes of losing analog material, digital data could be exposed to these factors as well as the ones mentioned above. As such it can be said that the physical nature of digital content contributes to its ephemerality.

It is common to store data on numerous digital devices and services. In many cases archiving is a side-effect of this ad hoc style of information management but digital content cannot sustain for long under this passive strategy. One often assumes that digital content can be stored and recovered when needed, and that the crucial element is to have back-up of some sort. [Marshall(2008b)] But ones and zeros will degrade over time. All archiving must frequently be refreshed into new media, and with digital content this frequency is fairly high, and ideally the refreshing should take advantage of improvements in technology. [Fraser and Schewe(2009)]

There are many examples that suggest a lot of digital information has already been lost, some may be possible to recover but this may constitute substantial costs. [Arora(2006)]

Arora (2006) specifies three examples in his overview:

- The 1960 Census was stored on Univac paper tapes that only could be read by a UNIVAC type II-A tape drive and ended up being obsolete by the mid-seventies. They ended up being saved by one of two machines capable of reading the tapes and resulted in great costs.
- NASA<sup>4</sup>/NSF<sup>5</sup>/NOAA<sup>6</sup> saved TOVS<sup>7</sup>/AVHRR<sup>8</sup> documentation of global warming over a 20-year period.
- In mid-1980s the New York State Archives attempted to preserve tapes containing data from the Land Use and Natural Resources Inventory Project (LUNR). The tapes were made in the late 1960s but by the preservation attempt both the software and the hardware needed for running the tapes were no longer available.

[Arora(2006)]

---

<sup>4</sup>National Aeronautics and Space Administration, <http://www.nasa.gov/>

<sup>5</sup>National Science Foundation, <http://www.nsf.gov/>

<sup>6</sup>National Oceanic and Atmospheric Administration, <http://www.noaa.gov/>

<sup>7</sup>TIROS Operational Vertical Sounder, [http://www.class.ngdc.noaa.gov/data\\_available/tovs\\_atovs/index.htm](http://www.class.ngdc.noaa.gov/data_available/tovs_atovs/index.htm)

<sup>8</sup>Advanced Very High Resolution Radiometer, [http://www.class.ngdc.noaa.gov/data\\_available/avhrr/index.htm](http://www.class.ngdc.noaa.gov/data_available/avhrr/index.htm)

A more currently example is provided by Chen, where the U.S. government spent over \$15 million to save email archives from the Regan and Bush administrations. The three examples presented by Arora are also mentioned. [Chen(2001)]

For a big company or institution it may be possible and cost-efficient to use large resources on finding the right technology to access content that is currently lost, but to a private person this may not be feasible due to the costs, work and time involved. As such it can be argued that the initial archiving and subsequent follow-up is more crucial for personal archives than for big institutions in order to ensure the longevity of the material.

As previously mentioned it is argued that the biggest threat to the sustainability of digital content is technological obsolescence. It can affect all parts of digital technology, including file format, hardware and software. Physically the digital content may survive for a long time but without the right technology it will be inaccessible and the information irretrievable. Backward compatibility refers to the possibility of new software reading formats from earlier versions. Software developers have not always taken this into account and over the years a lot of information has been left obsolete. [Arora(2006)]

In most cases institutions, organizations as well as individuals want to secure their digital assets. Digital piracy, unauthorized use and unwanted access are some of the issues being protected against by use of passwords, encryption and DRM<sup>9</sup>. But there is a conflict of interest between protection and long-term preservation. Password protection and encryption can keep documents safe for the time being but may render information inaccessible in the future when the passwords and encryption keys are lost. [Marshall(2007)] The dilemma between accessibility and privacy can also occur upon death or incapacity and the person leaves behind a significant volume of digital records that are now inaccessible due to password protection or other measures.

One of the best-known example of information becoming unavailable because of lost passwords is from the Ivar Aasen Centre<sup>10</sup> in Norway. Reidar Djupedal<sup>11</sup> had compiled over 11000 titles in a database, but took the password to his grave. The Centre's director appealed hackers to help obtain the password and they successfully unlocked access to the database in five hours. Recreation of the database would have taken four years and the cost this incurs. [Beagrie(2005)]

As Beagrie predicted in 2005 we are now seeing digital *safe-deposit boxes* for storing of passwords

---

<sup>9</sup>Digital rights management, [http://en.wikipedia.org/wiki/Digital\\_rights\\_management](http://en.wikipedia.org/wiki/Digital_rights_management)

<sup>10</sup><http://www.aasentunet.no/default.asp?menu=94&id=453>

<sup>11</sup>[http://no.wikipedia.org/wiki/Reidar\\_Djupedal](http://no.wikipedia.org/wiki/Reidar_Djupedal)



and access rights offered. [Beagrie(2005)] Current examples of such service providers is Clipperz<sup>12</sup> and Passpack<sup>13</sup>. They both offer a service where one can collect several passwords as well the opportunity to share access with family members or colleagues in case of death or end of employment. However the risks associated with transience of commercial services are still applicable in this case and could jeopardize long-term accessibility.

Digital rights management (DRM) and other copyright protection may secure creativity and privacy, but can prevent sustainability in the long run. It can make preservation difficult by not allowing copying and migrating to new media or formats. [Marshall(2007)]

### 2.2.1 Management strategies

This section will examine some of the prominent trends in personal archiving. In other words what are the common ways in which individuals are trying to catalog their everyday life, and more importantly is it working?

Catherine Marshall (2008) have done an in depth study on this topic and her findings are discussed below to give an example of contemporary archiving strategies amongst individuals. People do take action to preserve their digital documents. Encouraging signs are encountered for some level of data management and backup procedures both online and offline. Most people have some level of knowledge of the risk digital documents endures and that doing nothing is a dangerous approach. [Marshall(2008a)]

Marshall goes on to identify six prevailing strategies for personal archiving;

- Relying on backup as long term archiving.
- Duplicating document structures.
- External backup on CDs and DVDs
- Emails with file attachments are used for archiving.
- Combining social media sites.
- Platform backup for rebooting and accessing files when needed.

---

<sup>12</sup><http://www.clipperz.com/>

<sup>13</sup><http://www.passpack.com/>

[Marshall(2008a)]

System backups are not designed to be stored away and retrieved later. CDs and DVDs are fragile and there is no easy way to find out what was on them. As will be examined further in the next section, social networks and other commercial services are not always reliable and as already discussed hardware and software obsolescence is always a risk. [Marshall(2008a)]

Using email as an archiving strategy may seem to be the best of these solutions. In this sense email is used as a tool for communicate with oneself in the future and some email providers have presented services specifically for this. But obsolescence and the risk of the email provider going under is still an issue.

Marshall also identifies four assumptions individuals make about their digital assets.

- Leaving the material be is adequate for retrieval when needed.
- Will be able to keep track of material scattered online and offline.
- Will be able to distinguish between different material.
- Technology will be available in the future to access the material.

[Marshall(2008a)]

There are many other similar assumptions leading to difficulties. Encryption and copyright difficulties are some issues that may occur. Neglect is not a good strategy when it comes to digital archiving and the rapid accumulation combined with the strategy of storing material in many different places makes it difficult for people to keep track of their digital documents, specially if it is only being sorted out every decade or so. [Marshall(2008a)]

As can be seen several different personal storage strategies are often employed by the average person. These all come with their own drawbacks and risks. While personal archiving is often done through benign neglect one can argue that as information accumulates more people would likely be looking to improve and simplify their archiving efforts. As such the time could be ripe for the introduction of a centralized service. This would not only cater for the ones who are looking for said service but also educate those who do not actively archive at the present and raise awareness of the need to document the present and the best way to do it.

## 2.3 The transience of commercial services

The research for this project have identified that an increasing number of both individuals and companies turn to commercial service providers for their storage needs. This section will examine what these commercial services can provide, what additional risks they represent and to what degree they represent a long term storage solution.

To store, create and share files people are increasingly relying on online services and social networks. Google<sup>14</sup>, Flickr, Facebook<sup>15</sup> and Carbonite<sup>16</sup> are examples of some of these services. The content on several of these sites are unavailable to the general public and is therefore not undertaken by web harvesting. Jeremy John asks some interesting questions in his Nature article from 2009: *'Is this information being archived at all or simply lost? If kept, who maintains legal ownership of it or could make it available for research?'* [John(2009)]

While storage in the so called "cloud" through commercial services removes some of the risks tied to physical damage, hardware malfunction or loss of information due to fire, theft etc, it also presents an additional set of risks that have to be evaluated. Internet service providers shutting down, problems associated with affiliations change, casual system and data management. [Marshall(2008a)]

A survey on data loss conducted by Marshall, McCown and Nelson in 2007 shows that discontinued services, internet service providers policies and practices as well as accounts being deactivated are among the primary causes of data loss. [Marshall et al.(2007)Marshall, McCown, and Nelson]

Most online services are free of charge but provide extra services for a 'small' charge such as additional or 'unlimited' storage space. JISC suggests that this money may be better spent on systems giving the user control over their own collection. The user having access to the material and sustainability of the data rely on the credibility and longevity of the service. One may experience not having access to ones own material when needing it or losing it all together. Access control related to other users and copyright may also be an issue. The service may not relate to the same copyright law as the user is accustomed to, and complications may occur. [JISC(2009c)]

To illustrate the aforementioned points a few case examples will be highlighted to show the real

---

<sup>14</sup><https://plus.google.com/>

<sup>15</sup><http://www.facebook.com/>

<sup>16</sup><http://www.carbonite.com/>

risks experience by those who had the misfortune of choosing the wrong service provider for their storage. It should be noted that these are the worst case scenarios and that several services have operated without any such incidences, but the examples below are presented to illustrate the risks involved with relying on commercial services for long term storage.

In 2008 the Washington-based company Big Canvas Inc <sup>17</sup> lost photos belonging to 50 of their 50 000 costumers. The reason for the loss was failure at their third party service. The company used Amazon.com Inc's Simple Storage Service (S3)<sup>18</sup>, to store their costumers belongings. When this service got disrupted for seven hours the material was lost. [Waxer(2009)]

In an even more serious event a system administration error deleted up to 45% of the data stored at MediaMax / The Linkup<sup>19</sup>. The scandal resulted in the company shutting down, raising serious questions about the reliability of cloud storage. [Krigsman(2008)]

When Amazon's EC2<sup>20</sup> cloud server service got disrupted for more than a day, its customers lost access and in some cases data. This was further aggravated by the fact that other companies such as Foursquare<sup>21</sup>, Reddit<sup>22</sup>, Quora<sup>23</sup>, Hootsuite<sup>24</sup> and Moby<sup>25</sup> relied on the server and were also disrupted. [Bellan(2011)]

As can be seen the potential of information loss is magnified by the fact that several commercial services rely on larger companies to provide the storage infrastructure before marketing and selling it to private consumers.

When considering the transience of commercial services the physical risks are one aspect of the overall assessment one also have to consider the lifespan of the company itself. Longevity is a key factor in archiving and from research conducted on the lifespan of companies in general, it is apparent that historical evidence suggests that even the best commercial services only lasts for a generation. A study researching companies of all sizes, indicates a life expectancy of 12,5 years. Fortune 500<sup>26</sup> corporations and other big companies are shown a life expectancy between 40 and

---

<sup>17</sup><http://bcphotoshare.com/>

<sup>18</sup><http://aws.amazon.com/s3/>

<sup>19</sup>[http://en.wikipedia.org/wiki/The\\_Linkup](http://en.wikipedia.org/wiki/The_Linkup)

<sup>20</sup><http://aws.amazon.com/ec2/>

<sup>21</sup><https://foursquare.com/>

<sup>22</sup><http://www.reddit.com/>

<sup>23</sup><http://www.quora.com/>

<sup>24</sup><http://hootsuite.com/>

<sup>25</sup><http://www.moby.com/>

<sup>26</sup>[http://en.wikipedia.org/wiki/Fortune\\_500](http://en.wikipedia.org/wiki/Fortune_500)

50 years. [Geus(2002)]

There are however some countermeasures against the transience of commercial services. Some of these are in the form of emergency plans by the storage providers, and include offloading data when there is a failure and caching the data upon initial transfer. [Waxer(2009)]

To conclude it seems apparent that using commercial services for long-term storage is a risky proposition. Even a best case scenario is only likely to last for 40-50 years, considering that is the high end of a company's assumed lifespan. One potential solution could be to rely on a public-private partnership<sup>27</sup> between a public institution and a private one, where the public institution is responsible for carrying on with the project if the private institution fails.

## 2.4 Measures against ephemerality

The previous sections have outlined various manifestations of ephemerality. However seemingly from the beginning of archival endeavors man has tried to circumvent said ephemerality. This section will briefly outline measures against ephemerality throughout history, before examining what is currently being done in some of the major archival institutions both in Norway and abroad.

### 2.4.1 Historical countermeasures

Public organizations with preservation as one of their primary objectives have been carrying out digital preservation initiatives since the late 1960s. These programs have gone from focusing on the technology of the times such as paper tapes or punch cards, to evolve into the initiatives we see today from institutions such as the Library of Congress or the British Library. [Arora(2006)]

An example from another part of the world comes from India where for over a thousand years Indians scholars, poets and philosophers have recorded their contemporary times through scriptures, engravings and songs. In the times before prints were available information was handwritten and stored for future use. [Ramana(2004)]

It should be noted that the knowledge of historical archiving practices is generally considered to be incomplete, and it is likely that only a small portion of ancient storage sites have been uncovered by researchers. [Pedersén(2007)]

---

<sup>27</sup>[http://en.wikipedia.org/wiki/Public-private\\_partnership](http://en.wikipedia.org/wiki/Public-private_partnership)

Even in ancient times those looking to preserve information had to choose between several different means of preservation, not unlike the choices one has to make today. One of the first methods of recording writings were through using a clay tablet. The Egyptians preferred to record using ink on papyrus or leather, though papyrus proved to be a very fragile material which could only survive in very specific climate conditions. Clay tablets on the other hand were much more durable and as a result hundreds of thousands of such recordings have been excavated from the Middle East. [Pedersén(2007)]

Another archiving issue which is still relevant today is that of authentication. In ancient times documents were not authenticated by a signature, rather a personal seal was used. On recovered clay tablets it can be seen that these authentication imprints were often input onto the tablet itself or through marking a clay envelope. [Pedersén(2007)]

Research and archeological efforts have shown that nearly all archives or libraries have been located in cities, often with numerous sites in a single dwelling. Examples of ancient archives have been found both in public buildings such as temples and palaces, but also in private houses. The intended lifespan stretches from a couple of years for administrative archives to several generations for family archives. [Pedersén(2007)]

Categorization was attempted in larger libraries and archives through the use of labels and tags, and in some cases entire catalogues. In later Egypt and Mesopotamian works one could even find references to the author of the journals along with other information. [Pedersén(2007)]

It is reported that ancient Greek archives, in contrast to many of their counterparts at the time, took great care in preserving personal information. These were often recordings of gifts, loans or emancipation of slaves. It is suggested that this information was recorded for the person's record or that the state kept it for tax purposes. [Brosius(2003)]

The first large scale discovery of private archives occurred in 2000-1500 BC. An excavation of a set of old Babylonian houses revealed a large family archive placed in an inner room of a house. The contents of ancient private archives cover such information as trading activities (Assyria) and even personal letters, such as the case of a refugee writing to relatives. [Pedersén(2007)]

From this brief overview of historical countermeasures against permanence it would appear that ancient archivists faced similar challenges as we do today in areas such as technology, longevity, organization and authentication. It is also interesting to note that certain cultures have been more thoroughly persevered and thus documented due to their choice of recording material, ie clay tablets

over papyrus and ink. One can only hypothesize if similar parallels will be drawn to the choice of digital storage formats and practices of our times in a thousand years or so. [Pedersén(2007)]

### 2.4.2 The National Library

Like national libraries all over the world the National Library (NB) is preserving the contemporary. They strive to preserve all published material and give the public access to this content. The Legal Deposit Act<sup>28</sup> is the main pillar in this work and leads to NB being '*the premier source of information about Norway, Norwegians and Norwegian culture*' [Nasjonalbiblioteket(2011)]. In addition they collect and archive Norwegian media. [Nasjonalbiblioteket(2011)]

Their visions and goals are based around being a multimedia centre of high quality knowledge and experience, and among one of Europe's most modern national libraries with room for change and development. They want to be the core in Norwegian Digital Library assisting with the understanding of culture and technology. [Nasjonalbiblioteket(2011)]

NB want the public to have easy access to the material. The division in Oslo is recently renovated and welcomes users to their reading room, exhibits and events. [Nasjonalbiblioteket(2011)] NB stores the material in a mountain in Mo I Rana, see figures 2.4 and 2.5. Other copies are made available to the public in five libraries throughout the country. One such location is the NB's division in the Oslo, and the rest being the four university libraries in Norway. NB is also in the process of digitalizing the material. Everything will be available online once in a library and the material that is freed from copyright is available on nb.no<sup>29</sup>. [Nilsen(2011b)] An example of historical material being digitalized and made available online is Norway's first newspaper Norske Intelligenz-Sedler<sup>30</sup> from 1763.

---

<sup>28</sup>Pliktavleveringsloven, <http://www.lovdatab.no/all/hl-19890609-032.html>

<sup>29</sup><http://www.nb.no/bokhylla>

<sup>30</sup><http://www.nb.no/avis/intelligenssedler/>



Figure 2.4: Entrance to the archive inside the mountain. Photo: TRN



Figure 2.5: Example of the archives shelf system. Photo: TRN

The pictures utilised in this section were taken during a field study to Mo i Rana in the summer of 2011.



The digital material is also kept in NB's storage facilities in Mo I Rana. The material is stored on several hard drives, see figures 2.7 and 2.6. The material is migrated every 3 years. It is kept in blocks so if one hard drive shuts down nothing is lost. One of the biggest challenges with the digital storage is that as the amount of material grows it will soon take longer to migrate the material than the frequency between migrations. [Nilsen(2011b)]



Figure 2.6: NB's digitized archival library. Photo: TRN



Figure 2.7: NB's 2Pb digitized library backup. Photo: TRN

NB also archives material from Norwegian online publications. They archive material affiliated with events related to sport and politics, ten Norwegian online newspapers and three blogs. The last time they collected all Norwegian domains was in 2008. The reason being disagreements with The Data Inspectorate. [Nilsen(2011b)] Initially NB wants to harvest all Norwegian domains on a regular bases. But The Data Inspectorate have made requirements of approval and disclosure from the contributors, which NB finds difficult to adhere to, leading to them only harvesting the aforementioned material. [Rønnevig and Bendiksen(2010)]

### 2.4.3 Practices in other nations

Both the U.S and U.K seems to have taken the challenges of digital preservation seriously. Their respective archival institutions have been tasked with addressing the issue at hand. This section will briefly review what the Library of Congress and the British Library have initiated to establish a working strategy for the preservation of digital material.

#### 2.4.3.1 Library of Congress

The Library of Congress was among the institutions created by the founding fathers to allow later generations access to the nations accumulated information. In 2002 the Library was tasked by the U.S. Congress to address what was seen as a growing challenge of digital preservation. Armed with \$100 million they established the National Digital Information Infrastructure and Preservation Program (NDIIPP)<sup>31</sup> whose task it was to establish a strategy for digital preservation, including standards, collecting strategies and the creation of a national repository for digital information. [Library of Congress(2011)]

The concluding remarks regarding the substantial efforts of the Library of Congress and the NDIIPP is the final results of its ongoing review of the state of digital preservation in the U.S. It notes that “there are few incentives, and too many disincentives, to preserve digital content in the public interest”. [Library of Congress(2011)] It continues to state that while satisfactory policies and incentives are present for analog preservation, the policies that apply to digital preservation are often serve as a hindrance and create major obstacles for the preservation efforts of individuals and institutions. [Library of Congress(2011)]

---

<sup>31</sup>[http://en.wikipedia.org/wiki/National\\_Digital\\_Information\\_Infrastructure\\_and\\_Preservation\\_Program](http://en.wikipedia.org/wiki/National_Digital_Information_Infrastructure_and_Preservation_Program)

### 2.4.3.2 The British Library

The British Library located in London, has a similar role in the UK as NB has in Norway. Because of the British legal deposit<sup>32</sup> they receive one copy of everything published in the UK and Ireland, resulting in a collection of over 150 million items. In all they have 16000 users online and on site everyday. [The British Library(2012)]

The British Library works towards securing perpetual access to the digital contents within its collection. Their goal is to create a digital repository by 2016 which will hold all its digital collections. The institution aims to be a leader in digital preservation. [The British Library(2006)]

To achieve the aforementioned the British Library is using a digital object management system to conserve its digital resources. It is working to establish best practices within the field and is collaborating with several key players both internationally and at home. [The British Library(2006)]

The British Library has chosen to pursue a diverse set of preservation methods. This is done in order to achieve flexibility and to mitigate the risk of one method failing in the long-term. They also put an emphasis on applying life cycle modeling to their efforts and are active in shaping public policy and sharing their experiences with others (as well as taking inspiration themselves). [The British Library(2006)]

---

<sup>32</sup><http://www.legaldeposit.org.uk/>



## Chapter 3

# Preservation

*'Digital preservation refers to a series of managed activities designed to ensure continuing access to all kinds of records in digital formats for as long as necessary and to protect them from media failure, physical loss and obsolescence'* [Cornell University Library, 2005, Cited in [Arora(2006)]]

This section will outline the main components of digital preservation and what elements need to be present to foster said initiatives.

### 3.1 Prerequisites

In digital preservation, issues associated with storage media, hardware and software, as well as formats, application functionality and display capabilities has to be addressed. But there are also many other issues that should be looked at when thinking of long-term archiving. [Marshall(2007)]

One such issue is the fact that digital information accumulates very fast. Selection can therefore be a challenging aspect of digital preservation. One common approach to this challenge is to keep everything, while others eliminates as much as possible and as often as they see fit. With no clear strategy it can be difficult keeping track of all digital assets effectively. [Marshall(2008a)]

Figur 3.1 illustrates various concerns relating to digital archiving. It is divided among different forms of archiving and their related concerns.

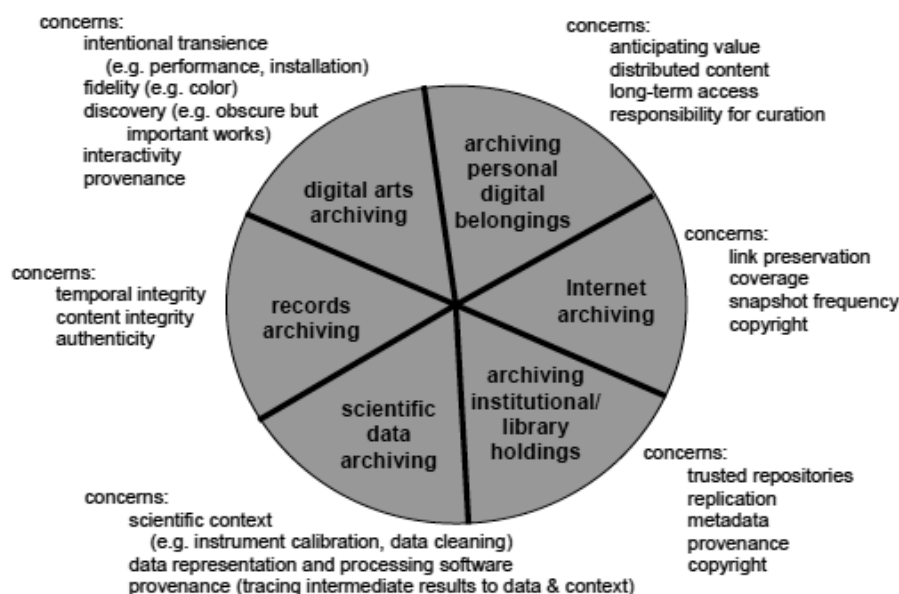


Figure 3.1: Digital archiving concerns arising from different disciplinary practices and institutions. [Marshall(2007)]

Technology can facilitate digital preservation and longevity can be enhanced, but valid organization of objectives and procedures must be combined within a solid preservation environment. In addition, financial, legal and management are important issues to address. [Beagrie et al.(2008)Beagrie, Semple, Williams, and Wright]

As mentioned active management is a necessity in preservation of digital content and the management choices will determine the sustainability and availability of the content over time. To enable long-term access, open and widely used formats and well organized management is essential. This can also help avoid expensive recovery projects in the future. [Library of Congress(2011)]

It is also important that the customer and creator clearly define what is expected of the project, as user requirements take a leading role in the development of digital systems. [Anderson et al.(2006)Anderson, Pringle, Eadie, Austin, Wilson, and Polfreman]

Anderson, Pringle, Eadie, Austin, Wilson and Polfreman (2006) states that there is no decisive approach to digital preservation and that no one single person has the opportunity to gain all experience needed. Information on this subject is far from complete and with the constant change and development, digital preservation has to adapt over time. This is reflected in figure 3.2 [Anderson

et al.(2006)Anderson, Pringle, Eadie, Austin, Wilson, and Polfreman]

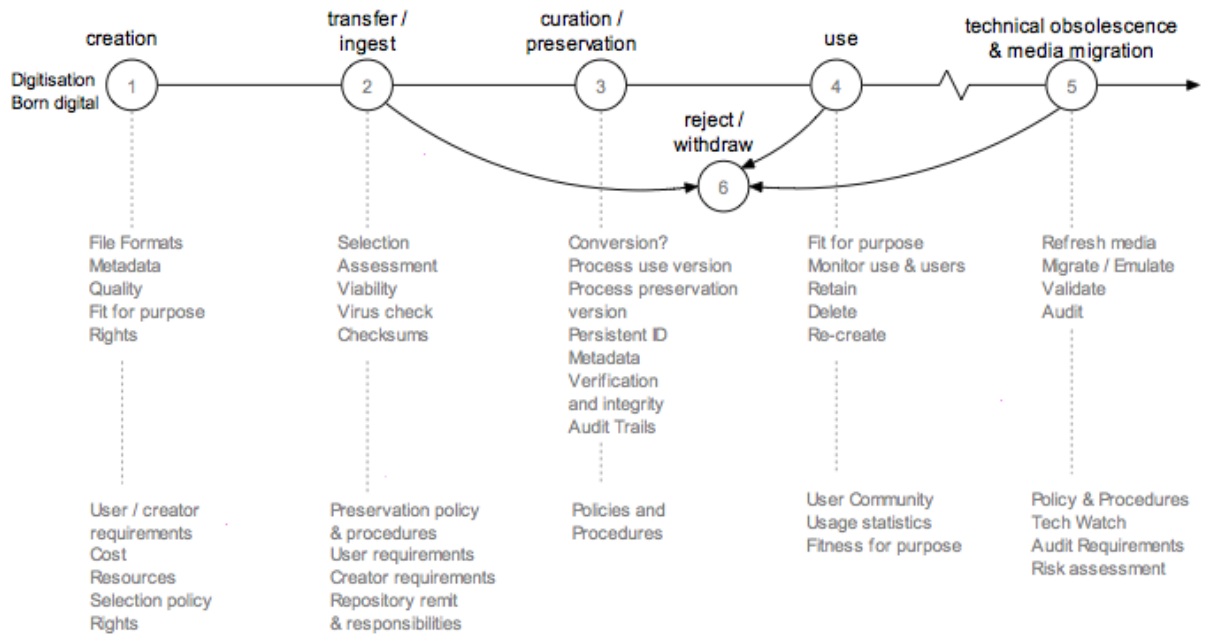


Figure 3.2: Life-cycle identifies six key events [Anderson et al.(2006)Anderson, Pringle, Eadie, Austin, Wilson, and Polfreman]

### 3.1.1 Strategies

Personal archiving strategies have already been discussed in this thesis. The next step is to move on to preservation strategies on a grander scale. The section below will highlight some of the established strategies for digital preservation and discuss the merits of each.

One example of a general strategy is the Four point strategy from Online Computer Library Center (2006) for long-term preservation:

- Open and widely used formats and applications.
- Evaluating content before determining possible preservation actions.
- Standards for metadata.
- Providing content access.

[Ram and Mishra(2008)]

However, there are many different strategies when it comes to digital preservation and there is no single right procedure for archiving digital content. [Arora(2006)] The various strategies are combined to try to guarantee accessibility and usability in the future. [JISC(2009a)]

Some strategies are necessary to a digital preservation process but may not be adequate by themselves. In most cases a combination of different strategies are required for the best results.

**Bit Stream Copying / Backup.** Bit-stream copying or backup is a common strategy among institutions and individuals alike. It is a necessity in the archiving process, but will not be adequate alone because of technological obsolescence, hardware and media failure as well as environmental influence. [Arora(2006)]

The next strategies may to some degree help preserve information, but in the long run they may be more harmful than helpful.

**Technology Preservation.** Preserving technical environments such as operating systems, applications and media drivers may extend accessibility. However, it is not a good long-term strategy because indefinite function is not possible for obsolete technology. [Arora(2006)]

**Analogue Backup.** Having printed copies of digital material is a good way of ensuring sustainability of the content. [Arora(2006)] However, it is a tedious job and may desolate the benefits of digital technology and in the long run it may not be a possibility because of accumulation.

The following are the most common strategies in digital preservation and as mentioned a combination may lead to the best results.

**Refreshing.** Refreshing entails copying data into a newer version of the same format without any changes being made to the data. [JISC(2009a)] Copying data from one CD-ROM to a new CD-ROM of the same standard is an example of refreshing. This is effective as long as the hardware and software involved is still accessible or still in use. [Ramana(2004)]

**Migration.** Migration is more extensive than refreshing, and refers to copying content into a newer format. The organized tasks are done periodically, transferring the data from one configuration or generation of technology to another. [Arora(2006)] An



example may be copying material from a CD-ROM to a flash drive. It leaves the data in a fresh state and instantly accessible. However, copies may degrade over time. [Tristram(2002)]

**Emulation.** An even more involved strategy is emulation. Data and old programs are made accessible for other platforms, mostly used when the original system is no longer available. There are many examples of this in the gaming industry, when old games are made available on contemporary consoles. [JISC(2009a)] The data don't have to be altered, but the mimicking is rarely flawless and in the long run the chains of emulators will break down. [Tristram(2002)]

**Encapsulation.** The technique for packaging digital content with the metadata necessary to access it is called encapsulation. [Arora(2006)] This strategy is mainly for textual data, and has to be redone when new format and software is released. [Tristram(2002)]

**Universal Virtual Computer.** Universal Virtual Computer is affiliated with emulation. It is a device independent program, with the ability to mimic every generation of computer architecture to simulate memory, sequence of registers and information movement. [Arora(2006)] It has also been described as the act of archiving machine specifications on paper and saving data in a format readable by that machine. Paper is of course a very sustainable media, but it can be quite difficult fitting such a specification onto a fairly brief paper document, which may lead to complications. [Tristram(2002)]

**Replication.** Many digital preservation strategies can be called replication. Backup is one of them. It refers to keeping several copies of a document in multiple facilities and locations. Lots of Copies Keeps Stuff Safe (LOCKSS)<sup>1</sup> is a project at Stanford University and can be defined as a replication consortium. Peer-to-peer<sup>2</sup> data trading can be seen as an open and free form of replication. The goal is to enhance longevity and maintain authenticity and integrity of digital material. [Arora(2006)]

Some other approaches are:

**Digital Archaeology.** Digital archaeology refers to methods to rescue digital content

---

<sup>1</sup><http://lockss.stanford.edu/lockss/Home>

<sup>2</sup><http://en.wikipedia.org/wiki/Peer-to-peer>

subjected to damaged or obsolete technology. [Arora(2006)]

**Reliance on standards.** Digital objects follow set standards and avoid using proprietary ones. [Arora(2006)]

**Normalization.** All digital objects of a particular type, ie photos, are converted to one format which is judged to be the optimal one for this type of object. [Arora(2006)]

**Canonicalization.** This technique evaluates whether the key elements of a document are still in place after a conversion have been conducted. [Arora(2006)]

### 3.1.2 Sustainability

Several sustainability factors are important to make sure our digital information is available in the future. In this case the focus lies on sustainability for digital content and in relation to a personal digital archive. The project has to focus on the sustainability of digital information but it is also important to focus on the sustainability of the project. If the archive is not sustainable, neither is the collection. In both cases it is important so plan for sustainability from the start. [JISC(2008)]

#### 3.1.2.1 Ongoing maintenance

Ongoing maintenance in some form is required for digital objects to sustain over time. The level of maintenance will vary depending on sustainability and general project design. Required maintenance may be;

- Ensuring availability to users during all hours of the day.
- Security procedures such as virus protection, backup systems, user authentication and management of access.
- Digital rights management in form of systems ensuring correct use of digital assets.
- Maintaining the content by keeping it dynamic, relevant and up-to-date.
- Maintaining technology, ensuring up-to-date systems and platforms.
- Administration of maintenance, with a straightforward transition between project and service.

[JISC(2008)]

### 3.1.2.2 Authentic preservation

With the purpose of authentic preservation of digital information sustainability of individual formats must be planned out. The Library of Congress has listed seven important sustainability factors that affect choices and can influence feasibility and cost. They can be applied to digital formats of all information categories. [Library of Congress(2007)]

- **Disclosure.** The degree of specification availability and access to technical integrity validating tools is referred to as disclosure. It is necessary to understand how the digital information is encoded to preserve it long term.
- **Adoption.** To which degree a format is in use is referred to as adoption. A widely used format is less likely to become obsolete and preservation tools fitting that format are more likely to become available.
- **Transparency.** Transparency suggests the degree of openness to which the digital format can be analyzed with a common toolset. Formats with simple and direct underlying information will be more susceptible to migration and other preservation efforts.
- **Self-documentation.** Documents stored with all necessary metadata are referred to as self-documented. Self-documentation will make a digital object more manageable and will lead to higher sustainability. In the future explanatory metadata can also give higher understanding to how an object originated and functioned in its original technical environment.
- **External dependencies.** Hardware, software and OS dependency is referred to as external dependencies. The more dependent an object is on external entities the more likely it is that something becomes obsolete leaving the object inaccessible.
- **Impact of patents.** Digital formats with patents related to them may be difficult to sustain. The development of open source encoders and decoders can be slowed down because of terms and fees given by the patent-holders.

[Library of Congress(2007)]

### 3.1.2.3 Digital infrastructures

Misund and Høiberg (2001) have outlined the criteria they consider to be applicable for a digital infrastructure to sustain over time and for a project to succeed, both in the present and in the long run. [Misund and Høiberg(2001)]

- **Longevity.** A system has to be able to survive over time.
- **Demand.** A system has to offer a service or services in demand to build a significant and stable user group. High demand will help the sustainability of a system.
- **Simplicity.** A well planned system with a design and functionality that is easy to use and understand will be more sustainable.
- **Quality.** A low-quality system will fail sooner or later.
- **Accessibility.** Easy and ubiquitous access is important for sustainability.
- **Responsiveness.** User feedback, participation and contribution will strengthened the system.
- **Adaptivity.** With the rapid change and development in technology there must be a underlying structure for innovation and improvement.
- **Scalability.** A system has to be able to handle increased interest, requests and work load, as well as a rapid growth of content.
- **Robustness.** A system should have tolerance for part system failure without total system shut down and there should also be possibility of improving modules of the system without having to redesign everything.
- **Stability.** To ensure sustainability and longevity of a system stabile management is necessary. Some one has to make sure the system sustain, even if it becomes financially uninteresting.

[Misund and Høiberg(2001)]

## 3.2 Current approaches

The importance of personal digital archiving is increasing exponentially as the number of documents created and stored in digital form grows. At the same time the fragile environment in which these

files are created jeopardizes its long-term survival. This is more so the case for personal documents which are in the care of individuals of varying means and skills, than those present in an institutional archive where it is likely that professionals are responsible for the materials. [Kim(2010)] As such it can be argued that digital personal information is under imminent threat and clear initiatives are needed to put them back on a track where its long-term survival is ensured.

These issues have not gone unnoticed and the work towards a sustainable strategy for personal digital preservation is getting attention from the archival profession. [Kim(2010)] This section will look at the overall concept of general digital preservation projects to provide insight into techniques or process that could also be relevant when attempting to establish a digital personal archive.

### 3.2.1 Digitalization projects

Digital preservation is described as process far encompassing the act of digitalization itself. It should be considered an ongoing process not limited by the timeframe of a single digitization project where the main goal might be to move analog material into a digital state. [JISC(2009b)] Rather, a digital preservation strategy should be in place well before the digitization occurs and it should cover the long-term preservation of the digital content as well as its metadata. It is also noted that documenting the digitalization projects themselves are extremely important as this will give the practitioners insight into the overall goals of the projects, and thus facilitate their work. [JISC(2009a)]

Due to the complex nature of digital preservation it is recommended that digitalization projects are supported by an digital preservation policy that covers the following issues;

- Ownership and preservation responsibility.
- A plan for refreshment/migration including timeline and responsibility.
- A budget and funding plan.

[JISC(2009a)]

The main goal for digital preservation should be to provide storage, recovery and access to documents saved in digital format. This will allow researchers to consult material without jeopardizing the physical properties of said objects while at the same time preserving the material beyond what the original format could have done. In general it is the responsibility of National Archives around the world to facilitate this process. [Ramana(2004)]

It is of key importance that a digitization project is compatible with the overall mission of an institution. This will likely ensure that the project is well funded and supported and thus have an increased chance to complete its objectives. In addition it is also recommended that the individuals involved have knowledge of both collections and technical expertise. [JISC(2010)]

### 3.2.1.1 The First World War Poetry Digital Archive

The First World War Poetry Digital Archive is among the projects founded by the JISC. It set out to digitize and provide access to over 4500 documents from British poets of World War One as well as supporting material. In order to achieve this, the project actioned the following list of items;

- To migrate digital content from an existing archive into a new system for managing the content.
- To digitize 2000 analog source material.
- To create metadata for each item in the archive.
- To make the poems readable by machines.
- To develop an interface where users could search, acquire and display items.
- To allow the public to donate material to the archive through a web based solution.
- To allow users to personalize the access through association prompts and personal pathways through the archive.

[Lindsay(2009)]

One of the takeaways from this project is how involving the public in digitization and curation of these poems, resulted in not only a finished product that exceeded its original goals but also at lower cost than anticipated. For example the project outsourced the digitalization to the institutions which held the original works. Once the digitized copies were received they were subsequently archived using Portfolio software. It also appealed users to donate material through a web based solution, this resulted in a vast amount of material being included in the archive. [Lindsay(2009)]

To support the creation of this archive the project involved stakeholders such as experts in the field, students, teachers and World War One interest groups. By holding events such as conferences and

symposiums they were able to generate interest in the project as well as obtaining material. It is also interesting to note that the systems is built using open source tools and has been made available for others to use. [Lindsay(2009)]

Figure 3.3 illustrates the digitization workflow in the First World War Poetry Digital Archive.

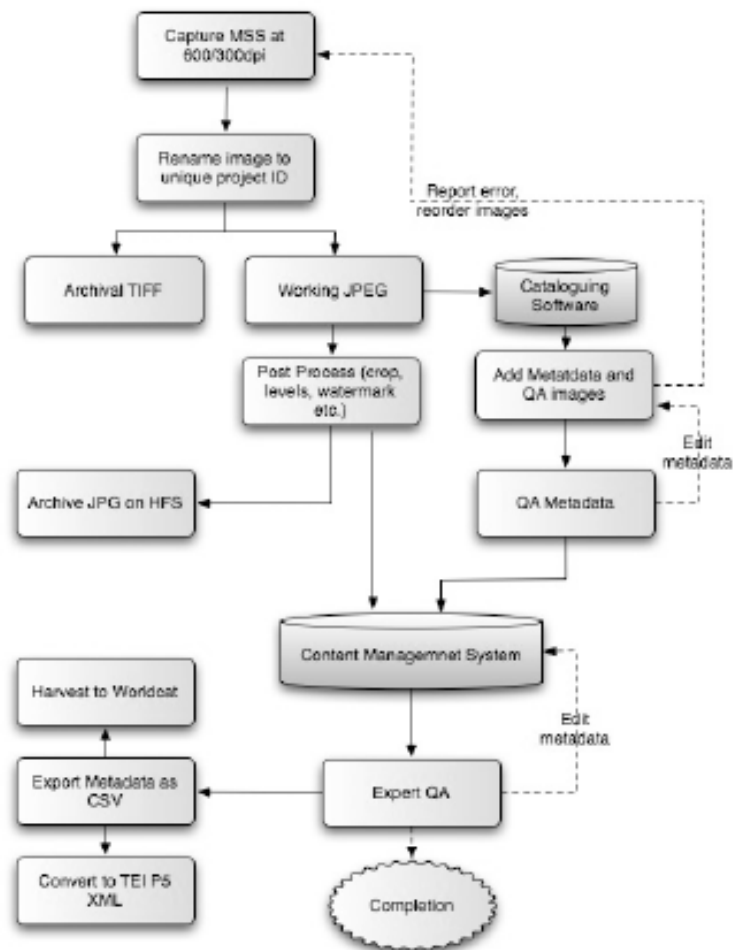


Figure 3.3: Digitization workflow of the First World War Poetry Digital Archive [Lindsay(2009)]

### 3.2.1.2 ABM

ABM<sup>3</sup> was a Norwegian organization tasked with archiving, library and museum efforts. It served in an advisory capacity to the Norwegian state in matters pertaining to the three aforementioned elements. The organization was reorganized in 2010 and is not present in its original form. [Norsk Kulturråd(2012)] The Arts Council Norway<sup>4</sup> now have their responsibilities.

One of the institutions efforts were to recommend a format for archiving digital photographs. The goal was to ensure a consistent and prudent cataloging of photographs. Its purpose was to eventually serve as a foundation for future exchange of data across institutions, as well as, being valid for both individual photographs and collections/archives. In addition it was also intended as a guideline in the evaluation of archiving solutions for photographs. [Torgnesskar(2007)]

It is recommended that the archive is based on the ISAD(G)<sup>5</sup> standard for cataloging archival material. This standard consists of four basic principles;

- To describe from general terms to specifics.
- To only describe what is relevant in the current level of archive.
- To link each level of description with each other.
- To avoid repetition of information.

[Torgnesskar(2007)]

The standard itself contains of 21 fields which covers such elements as, title, creation date, condition and photographer. It is argued that this standard will be able to cover both administrative as well as preservation functions for most archiving institutions. It should also ensure that the cultural and historical documentation is present and facilitate searches between collections. However, to ensure this the 21 fields of which the standards consists must be filled out to the full extent possible. [Torgnesskar(2007)]

In regards to copyright and ownership it is highly recommended that instances where these are present are recorded in a special field. This to ensure explanations for the conditions present as well as ensuring that only the elements that are covered by such restrictions are filtered out when

---

<sup>3</sup><http://no.wikipedia.org/wiki/ABM-utvikling>

<sup>4</sup>Norsk kulturråd, <http://www.norskkulturrad.no/>

<sup>5</sup>General International Standard Archival Description, [http://en.wikipedia.org/wiki/ISAD\(G\)](http://en.wikipedia.org/wiki/ISAD(G))



needed. [Torgnesskar(2007)]

### 3.2.1.3 NDIIPP

The NDIIPP was created by Library of Congress in 2000 and were structured to run for a ten year period. Its main tasks were;

- To identify and preserve digital content considered at risk.
- To develop a strategy for collection and preserving digital content at a national level.
- To actively support the development of models, tools and methods for digital preservation.
- To engage libraries, industry, research institutions and federal agency in digital preservation initiatives.

[Fleischhauer(2008)]

One of its primary concerns has been to work together with the Copyright Office to improve the current legislation to better facilitate digital preservation of material which falls under the current statutes. [Fleischhauer(2008)] The NDIIP identified their four key goals to be;

- Developing a national preservation network.
- Establish a national digital collection.
- Develop a sound technical infrastructure to facilitate the collection and preservation of the digital material amassed.
- Improve public policy in order to better facilitate the strategy due to be implemented.

[Library of Congress(2011)]

With these overall goals in mind the NDIIPP set in motion some concrete actions to address the concerns raised by congress regarding the state of digital preservation in the U.S. [Library of Congress(2011)]

The next steps for the NDIIPP is to continue to promote the National Digital Stewardship Alliance<sup>6</sup> and establish it in all 50 states. It considers preservation as an action to be undertaken by committed

---

<sup>6</sup><http://www.digitalpreservation.gov/nds/>

organizations to create a societal good. To achieve this it is dependent on its partnership of companies and institutions which are also looking to preserve the digital content of the nation for future generations. [Library of Congress(2011)]

### **3.2.2 Social networking platforms**

An increasing amount of storing and publication of personal digital information takes place on social networking platforms. As already mentioned many social networks sites offers users the opportunity to upload and store photos and other material. To share with their friends and as a security for their memories. The following will give an overview of a small section of services.

#### **3.2.2.1 Flickr**

Flickr is a photo and video sharing community run by Yahoo<sup>7</sup>. The users are allowed 300 Mb every month and can get unlimited capacity for an extra fee. [Yahoo(2012)] The user can specify accessibility to their material by choosing whether or not a photo is public, limited to friends or private. The infrastructure is based on meaningful metadata by tags, sets and galleries. Anyone can browse public material, but only members can upload to the site. [Flickr(2011)]

#### **3.2.2.2 Photoblog**

*'Photoblog helps you preserve your memories and share your life through photographs'* [Photoblog(2012a)] This self proclaimed statement gives a good indication to what Photoblog is. The photo sharing network was founded in 2006 by Michael Zhang. [Photoblog(2012b)] The functions of the site is similar to other sites although a bit more straightforward.

#### **3.2.2.3 Photobucket**

Photobucket was established in 2003, and after takeovers the company is now know as Photobucket Corporation. [Photobucket Corporation(2012a)] Photobucket is one of few sites that lets users upload unlimited (but not excessive) material. [Photobucket Corporation(2012b)]

---

<sup>7</sup><http://yahoo.com/>

### 3.2.3 Web harvesting

One trend in digital archiving is web harvesting<sup>8</sup>. Web harvesting is in effect the download, storage and indexation of complete websites. An example of a program made to facilitate this is The Internet Archive and Archive-it.

Whole domain harvesting refers to harvesting complete internet domains at regular intervals, to ensure its availability should it later be taken offline. This can be done in a fairly automated process thus reducing costs and staff requirements. However, disadvantages include the need for considerable computer and storage requirements limiting the frequency of harvesting. Indiscriminate harvesting also runs the risk of being left without manual quality controls, and as such the value of the information retrieved may be compromised. [Ram and Mishra(2008)]

Among those institutions that have pursued whole domain harvesting are the National Libraries of Finland, Iceland, Norway, and Austria. [Ram and Mishra(2008)]

#### 3.2.3.1 The Internet Archive

The Internet Archive was founded in 1996 and has with time built an Internet library, offering permanent access to digital historical collections. This includes text, moving images, audio, software and web pages. Collaborating with Library of Congress and the Smithsonian, among others, to prevent digital material from disappearing and to preserve a record of the Internet for the future. [The Internet Archive(2012)]

At the moment using the collection requires programming skills, though tools and methods are under development giving the general public easier access. [The Internet Archive(2012)]

#### 3.2.3.2 Archive-it

Archive-it is a subscription service which facilitates the creation and preservation of collections of born digital content. It was initiated by the Internet Archive in 2006, allowing for harvesting, cataloging, managing and browsing of archived content with full text search. [Archive-It(2011)]

---

<sup>8</sup>[http://en.wikipedia.org/wiki/Web\\_harvesting](http://en.wikipedia.org/wiki/Web_harvesting)

At the present time they work with 190 organizations worldwide, together working to archive websites and online publications. This includes such items as standard online publications as well as data from social networking sites, blogs, tweets and commentaries. These can then be assembled to create an archive of a unique event or topic. In addition Archive-it is used to preserve important government and scholarly records. [Archive-It(2011)]

### **3.2.4 Other projects**

In addition to the aforementioned type of projects, other notable digital preservation initiatives include the following.

#### **3.2.4.1 The Digital Manuscripts Project**

The British Library have coined the term eManuscripts (eMSS) for the digital versions of letters, diaries, family photos and the like. It has also started to add a variety of computer media to its manuscripts archives. [Summers and John 2001, John 2005, Cited in [John(2008)]]

The goal of the Digital Manuscripts Project is to create and install systems which can secure personal archives and ensure long-term access to these materials. In other words, to efficiently capture peoples digital information alongside their analog output. [John(2008)]

It is noted that the requirements to check for confidentiality and data protection in the personal material is a considerable challenge for curators. John (2008) calls for software that could automatically perform these tasks, and argues that this will go a long way to ease the burden currently placed on curators shoulders. [John(2008)]

The Digital Manuscripts Project has not relied on one solitary technology for its capture of digital content. The project has drawn on both software and hardware from ancestral computer, forensic and bioinformatic communities to facilitate digital capture. [John(2008)]

#### **3.2.4.2 Digital Lives**

The Digital Lives research project consisted of archivists and researchers from the British Library as well as scholars and information technology and law institutions. Its four part mission consisted of the following:

- Facilitate personal digital archives to reach their potential.
- Comprehend how academics and the general public use personal computers.
- To raise awareness of how one can record, store and utilize personal archives over the course of ones life.
- To educate individuals how to deposit a personal archive to an institution or a family member to ensure its continued life.

[John et al.(2010)John, Rowlands, Williams, and Dean]

The theme of advocacy is a red thread throughout the project. The need to raise awareness of the urgency to preserve personal digital information is highlighted in the project report. The brunt of this responsibility is likely to fall on the archival community, libraries and government institutions. [John et al.(2010)John, Rowlands, Williams, and Dean]

### 3.2.4.3 Lifelogging

One of the recent interesting projects within digital preservation is Microsofts ambitious MyLifeBits project, which will be detailed in the next section. In the spirit of Vannevar Bush's Memex approach, the project sets out to record and store close to every aspect of daily life. [Kim(2010)] This form of initiative is commonly referred to as lifelogging. Potential benefits of the approach are for instance, serving as a memory device, improve personal time management, security and medical health. [Czerwinski et al., 2006; Bell & Gemmell, 2007; Sellen and Whittaker, 2010, Cited by [Kim(2010)]] The data recorded can also be used to facilitate research and even act as an advertising tool. Among the challenges to this approach is how to efficiently store such vast amounts of data, as well as how to mine and extract key pieces of information from the sizable data that could build up over time. [van Dijck(2007)]

John (2008) provides an opinion on where lifelogging might lead us in the future. He states that *'It is possible to anticipate in the not too distant future an ability to identify patterns that enable the eMSS to be provisionally classified according to key phases of a person's life: associated with childhood stages (eg starting school), coming of age, initiation rites, process of a job application, a resignation, a promotion, communications leading to weddings or partnership, professional collaborations, retirement, reminiscence and reflection, births and deaths, memories and remembrance, and so on'*. [John(2008)]

To give some insight into the inspiration of lifelogging, lets take a quick look at Vannevar Bush's 1945 work *As we may think*. In this work Bush raises concern regarding the amount of information we are able to possess and the subsequent difficulties in making sense of them. He rightly predicts that an entire encyclopedia could be put onto something the size of a matchbook. However this is all inconsequential if it is not recorded or indexed properly so it can be efficiently accessed. [Bush(1945)]

To address these issues Bush raises the concept of a Memex. This is a device which would store an individuals records, books, and communications. It would act as an extended memory. The information should be selectable through associated rather than by indexing. [Bush(1945)]

The end goal is to relieve man of remembering unnecessary information and provide a complete record of past mistakes or fortunes. This extended memory should help man retrace his steps, and allow for the limited capacity of memory to be put to better use. [Bush(1945)]

#### 3.2.4.4 MyLifeBits

MyLifeBits is intended as a digital personal lifetime storage. It encompasses the recording and storage of digital media of all variety such as email, documents, video and audio. Its metadata is based around dates, location, and links. This is recorded using an SQL server database. [Gemmell et al.(2005)Gemmell, Aris, and Lueder]

The four basic principles of the MyLifeBits project is

- Collections and search must be the primary method of organization, not static hierarchy.
- Visualizations in many forms must be supported.
- Annotations will serve as critical to non-text media and must be easy to enter/use.
- Authoring should occur via transclusion<sup>9</sup>

[Gemmell et al.(2002)Gemmell, Bell, Lueder, Drucker, and Wong]

It is argued that for the archive to have value over several generations annotations must be present to frame the meaning and setting of the material for future generations, who cannot rely on their own memory to give the data meaning. [Gemmell et al.(2003)Gemmell, Lueder, and Bell]

---

<sup>9</sup>the inclusion of a document or part of a document into another document by reference.<http://en.wikipedia.org/wiki/Transclusion>

MyLifeBits evolved from seeking to store archival material into storing absolutely everything about a person's life right down to keystrokes and mouse clicks. Storing too much is not an issue for the project, the focus is rather put on how cost and copyright issues prevent them from storing absolutely everything. The argument is that one cannot predict the future value of an item at the present time so storing everything ensures that nothing that could prove useful in the future is being lost. Another positive aspect of keeping everything is that as more information is kept (and tagged properly) more links emerge, potentially making it easier to find material through association. In order to fulfill its full potential, users of the systems have to act as filing clerks, duly tagging the masses of information that is being captured. In the future it is expected that an increasing amount of tagging and metadata recording will be done automatically. [Gemmell et al.(2006)Gemmell, Bell, and Lueder]

The MyLifeBits software utilizes a set of queries, hyperlinks, annotations and fast search to make sense of the masses of information accumulated. It can record anything from web pages, chat transcripts, radio and television. [Gaudin(2008)]

In 2007 Gordon Bell's (the alpha user of MyLifeBits) archive held 122,000 emails, 58,000 photographs, thousands of recorded phone calls, every webpage he had visited and every instant message chat from 2003, a record of desktop activity, 800 pages of medical records, various other items such as labels from favorite wines. Bell's actions have been called an extreme version of the same practice Samuel Pepys pursued in the 1600s. Bell has highlighted software and technological obsolescence as a key threat to his lifelogging efforts. When the goal is to store every aspect of daily life over a lifetime, the threat of obsolescence is Bell's number one worry. [Wilkinson(2007)], [Gemmell et al.(2003)Gemmell, Lueder, and Bell]

Another problem that Bell has faced is being able to extract exact data from his vast archive. He states that he can recall an event but not necessarily obtain the exact records of it from his digital archive. The issue of privacy is also present as one cannot know if the information recorded today will be used against a person ten years from now. [Anderson(2007)]

#### **3.2.4.5 Paradigm**

The Personal Archives Accessible in Digital Media (Paradigm) was a project run by research libraries at the University of Oxford and the University of Manchester. The project explored different issues related to digital preservation of private information. Drawing on practical experience in

accessioning by taking in and processing private digital material in line with digital archiving requirements. [The University of Manchester(2008c)]

From 2005 to 2007 Paradigm dealt exclusively with personal archives engaging in curation and preservation of mixed media. The project involved itself directly with contributors as well as future researchers. To keep the project manageable Paradigm focused on the personal archives of contemporary politicians. Testing various tools, standards and techniques they have targeted an area of importance and raised awareness of digital preservation in the political sphere. [Thomas(2007)]

### 3.3 Open formats and standards

This section will cover the general arguments concerning open formats and standards to illustrate which elements to focus on once the standards of today ceases to be relevant.

Information is encoded in different forms for processing and rendering by combinations of software and hardware. Accessibility of digital information is therefore dependent on access to the right hardware, software and sometimes operating system. The file formats has different degrees of openness depending on how much it is dependent on external programs to be accessible. This is called external dependencies and as previously mentioned the higher the extent that a file format is dependent on other programs the higher the likelihood that something becomes obsolete leaving the information inaccessible. [Library of Congress(2007)] This raises major concern for short term information management as well as sustainability and long term digital archiving. [Brown(2008)]

Classified depending on format specification, file formats can be divided in three categories;

- Proprietary file formats are obtained by software producers and the specifications are not made public.
- Open / public file formats have public specifications. In some cases a consortium has defined a standard for several companies to develop compatible formats.
- Standards are file formats that are developed and used as international standards with public specifications.

[Anderson et al.(2006)Anderson, Pringle, Eadie, Austin, Wilson, and Polfreman]

Selection of file formats can be crucial when it comes to long-term sustainability. To simplify the



archiving process minimizing the number of file formats involved can be a good strategy. It can be useful to identify file formats that meet certain criteria for long-term archiving and sustainability, and choose a small set of formats to recommend or require for archiving. Criteria to consider should be; Ubiquity, Support, Disclosure, Documentation quality, Stability, Ease of identification and validation, Intellectual Property Rights, Metadata Support, Complexity, Interoperability, Viability, Re-usability. [Brown(2008)]

It is argued that open standards can give more people access to public information. It is a good counteract against monopolization and unfair competition. Information should be portable across platforms and commonly available software. Thus making it accessible today and easier to archive for the future. [Vaaler and Ludvigsen(2008)]

In Norway determining standards for the library field falls on the National Library. Digitalization and digital preservation are areas in constant development. Still it is necessary to formulate standards for quality levels of digitalization, metadata and formats in use. [Kulturdepartementet(2009)]

### 3.4 Legal aspects

This section will detail the current prevailing laws which are pertaining to a digital personal archive. The main area of focus will be Norwegian law but aspects from other legal environments will also be discussed to highlight initiatives made in other nations. The five main legal works pertaining to digital archiving in Norway are considered to be;

- Lov om behandling av personopplysninger (personopplysningsloven)
  - Data Protection Act [Justis-og beredskapsdepartementet(2000)]
- Lov om opphavsrett til åndsverk m.v. (åndsverkloven)
  - Copyright Act [Kulturdepartementet(1961)]
- Lov om kulturminner (kulturminneloven)
  - Cultural Heritage Act [Miljøverndepartementet(1978)]
- Lov om arkiv (arkivlova)
  - Archives Act [Kulturdepartementet(1992)]

- Lov om avleveringsplikt for allment tilgjengelege dokument (pliktavleveringslova)  
- Legal Deposit Act [Kulturdepartementet(1990)]

### 3.4.1 Review of relevant laws

The following section highlights the parts of the aforementioned laws viewed to be relevant for a personal national archive. It has been translated from Norwegian and will serve as a reference point for later discussions within the thesis.

#### 3.4.1.1 Data Protection Act

§ 1. *The purpose of the law.*

*This laws purpose is to protect the individual against infringement on person rights during the processing of personal information. [Justis-og beredskapsdepartementet(2000)]*

*The law should ensure that personal information is treated in accordance with basic personal right concerns, here within the need for personal integrity, privacy, and sufficient quality of the personal information. [Justis-og beredskapsdepartementet(2000)]*

§8. *Conditions for processing personal information*

*Personalinformation (jf. §2 nr. 1) can only be processes if the person has consented or it is established in law that processing of the material is valid, or if the processing of material is needed:*

- a** *to honor an agreement with the person, or to execute the persons request prior to said agreement begin made,*
- b** *for the case handler to be able to honor a legal commitment,*
- c** *to protect the vital interest of the registered person,*
- d** *to perform a task of public interest,*
- e** *to execute public authority, or*
- f** *that the case handler or third parties which receive the information has a vested interest and that the responsibilities regarding the registered persons third party rights does no overrule this interest.*

[Justis-og beredskapsdepartementet(2000)]

**§9. Processing sensitive information**

*Sensitive personal information (jf. §2 nr. 8) can only be processed if the conditions under §8 is fulfilled and*

- a** *the registered person consents to the processing,*
- b** *it is established in law that processing of such material is lawful,*
- c** *the processing is needed to protect a persons vital interests, and the registered person is in no condition to consent,*
- d** *it is solely information that the registered person himself have made public that is processed,*
- e** *the processing is needed to establish, enforce or refute a legal claim,*
- f** *the processing is needed for the case handler in order for the case handler to preform his duties or rights,*
- g** *the processing is needed for preventive treatment of illness, a medical diagnosis, care, treatment or management of health services, and the information is handled by health care professionals bound by confidentiality, or*
- h** *the processing is needed for historical, statistical or scientific endeavours, and societies interest that the processing takes place overrules the possible inconcenance for the individual.*

[Justis-og beredskapsdepartementet(2000)]

*The Data Inspectorate can rule that sensitive personal information can be processed also in other cases than the aforementioned if significant societal interests so dictates and efforts are made to protect the interests of the* [Justis-og beredskapsdepartementet(2000)]

**§11. Basic tenements for processing personal information**

*Subsequent processing of personal information for historic, statistic or scientific purposes are not at odds with the original purpose of collecting information, ref first paragraph letter c, if societies interest in the processing occurring trumps the inconvenience the individual might be exposed to.*

[Justis-og beredskapsdepartementet(2000)]

**§13. Security of Information**

*In order to obtain satisfactory information security the case handler and data processor must document the information system and security measures. [Justis-og beredskapsdepartementet(2000)]*

#### **§14. Internal Controls**

*The person in charge of processing should establish and maintain systematic and well planned measures needed to fulfill the criteria of or in this law, and through this ensure the quality of the personal information. [Justis-og beredskapsdepartementet(2000)]*

#### **§28. Ban against storing superfluous personal information**

*The person in charge of processing should not store personal information for longer than what is necessary to complete the purpose of the processing. If the personal information should not be stored in accordance with the Archiving Law or other relevant legislation, then it should be deleted. [Justis-og beredskapsdepartementet(2000)]*

*The processing responsible can without obstruction from the first paragraph store personal information for historical, statistical or scientific purposes, if societies interest of storing the information outweighs the potential inconvenience of the individual. The processing responsible should in this case ensure that the information is not stored in a manner that makes it possible to identify the individual for longer than necessary. [Justis-og beredskapsdepartementet(2000)]*

*The Data Inspectorate can - after the National Archives have been consulted - enact rulings that the right to deletion after the third paragraph trumps the rules of the archiving law 4 December 1992 nr. 126 §9 og §18. [Justis-og beredskapsdepartementet(2000)]*

### **3.4.1.2 Copyright Act**

**§1.** *He who creates a work has the copyright to that work. [Kulturdepartementet(1961)]*

**§2.** *The creation of a work also covers transition to a device which can recreate the work. [Kulturdepartementet(1961)]*

**§40.** *The copyright is valid in the authors lifespan and 70 years after the year of his death. For the works covered in § 6, the 70 years is valid from the year of the longest living authors deathyear. For cinematic work the protection time spans from the end of the death year of the longest living principal director, screenwriter, dialog author or composer of the movies score. [Kulturdepartementet(1961)]*

**§43a.** *He who creates a photograph has the sole right to produce a copy either through photography, print, drawing or through any other means making it available to the general public. [Kulturdepartementet(1961)]*

*The copyright to a photograph lasts in the photographer's life and 15 years after the end of his death year, though at least 50 years from the year the picture was created. If two or more people share the copyright, then the protection period runs from the end of the death year of the longest living. [Kulturdepartementet(1961)]*

*The conditions in §§2 second to fourth paragraph, 3, 6 to 9, 11 to 21, 23 to 28, 30 to 39f and 39j to 39l, is valid for photographic images in the same regard as for photographic works. [Kulturdepartementet(1961)]*

*If a photograph is covered by copyright, it can also be [Kulturdepartementet(1961)]*

**§45c.** *A photograph which depicts a person cannot be reproduced or displayed in public with the consent of the depicted individual, except when*

- a** *the picture has current and public interest,*
- b** *the depicting of the person is less important than the main theme of the picture,*
- c** *the picture depicts congregations, open air rallies or conditions or events of public interest,*
- d** *copies of the picture is used to promote the photographer's business and the depicted does not veto it, or*
- e** *the picture is used as covered by §23 third paragraph or §27 second paragraph.*

[Kulturdepartementet(1961)]

*The protection is valid in the depicted lifespan and 15 years after the year he died. [Kulturdepartementet(1961)]*

### **3.4.1.3 Cultural Heritage Act**

**§1.** *The purpose of the law.*

*Cultural heritage and culture environments with their uniqueness and variety shall be protected as part of our cultural heritage and identity and as part of a general environment and resource management. [Miljøverndepartementet(1978)]*

*It is a national responsibility to protect these resources as scientific source material and as a permanent foundation of living and future generations experience, sense of self, enjoyment and enterprise. [Miljøverndepartementet(1978)]*

#### **3.4.1.4 Archives Act**

##### **§1. The purpose of the law.**

*The purpose of this law is to secure archives which have mona cultural or scientific worth or which contain legal or important historic documentation, so these are preserved and made available for the coming generations. [Kulturdepartementet(1992)]*

##### **§13. Management and registration.**

*The National Archiver shall;*

- a** *keep overview of private archives worthy of preservation.*
- b** *keep a register of private archives which have been acquired by public and private archiving institutions.*

[Kulturdepartementet(1992)]

*The National Archiver can register select private archives as especially valuable. The notice of such registration should be sent to the owner of the archive. [Kulturdepartementet(1992)]*

##### **§14. Guidelines.**

*The National Archiver provides further guidelines for working within personal archives in the "Archiving Work". The National Archiver can set these guidelines wholly or partly and should cover*

- a** *other public institutions which works towards the preservation of personal archives.*
- b** *private institutions which work towards the preservation of personal archives and have received public funding for this work.*

[Kulturdepartementet(1992)]

##### **§15. Duty of disclosure.**

*The owner of the archive which have received notice from the National Archiver of the special value of the archive, jf. §13 second paragraph, has the duty to notify the National Archiver once the*

*archive changes owner; wholly or in part is exported from the country or is in danger of being lost. [Kulturdepartementet(1992)]*

**§16. Transfer and deposit.**

*If the valuable private archive is transferred or deposited to the Archive Work or another archiving institution, then the owner of the archive can set forth limitations to the accessibility to the archive. Such limitations cannot be valid for over 100 years after the transfer or deposit. [Kulturdepartementet(1992)]*

*When transferring the rights to the archive transfers to the receiving institutions. [Kulturdepartementet(1992)]*

*When depositing the depositor and his heirs have the legal rights to the archive. The legal rights transfers to the receiving institution when the line of heritage is broken or when 100 years have passed since the deposit date. [Kulturdepartementet(1992)]*

**§17. Copy.**

*When a valuable archive is being exported out of the country or is in the danger of being lost, the National Archiver can demand copies of the original documents. The National Archiver have the same right in case a valuable archive which have been deposited is in the process of being handed back. [Kulturdepartementet(1992)]*

*The copies should be kept in the Archive Work or another archiving institution, which have to cover the cost of copying. The owner of the archive can set terms of the copies in terms of §16 first paragraph, but cannot demand compensation for the copies being created. [Kulturdepartementet(1992)]*

**§18. Limited rights.**

*Archives which have been registered as valuable, jf. §13 second paragraph, cannot be partitioned, exported, damaged, or destroyed without the consent of the National Archiver. The rules in the personal information law and the health registration law about correcting and deleting information will still be completely valid. [Kulturdepartementet(1992)]*

**§19. Personal archives with a public link.**

*If a private individual receives public funds which is important to the business or have the authority to make or execute "forskrifter", then the National Archiver can establish the archive to part or wholly follow the terms of a public archive given with this law. [Kulturdepartementet(1992)]*

### 3.4.1.5 Legal Deposit Act

#### §1. Purpose.

*The purpose of this law is to secure the deposit of documents with public information to national collections, so these testimonies of Norwegian culture and society can be made available as source material for research and documentation.* [Kulturdepartementet(1990)]

### 3.4.2 Legal impact

*Samkatalogen for privatarkiv*<sup>10</sup> is a national *union directory for private archives*, and has been available online since 1998, before the law was put into motion. The registering of data in the directory is done on the basis of information reported from the institutions holding the material to the National Archive. By the end of 2008, 54 institutions had incorporated their archives; with 14 821 record creators and 15 555 archives in the database. [Kulturdepartementet(2009)]

Although quality assured by the National Archive, the individual institutions are responsible for the authenticity of the reported information. The information in the directory is only at a descriptive level, if the users want more information they must contact the individual institutions. [Kulturdepartementet(2009)]

It is not uncommon that some of or all of the material in an archive is restricted by the owner to protect personal information, third parties or trade secrets. [Kulturdepartementet(2009)]

Because increasingly more digital material will be made available 100 years after being deposited (if it survives), there is an urgent need for higher levels of competence within privacy policy, including regulation standards for ensuring legal and ethical requirements. [Kulturdepartementet(2009)]

Web-based communication of digital material have increased the extent and the complexity of the legal challenges the libraries, archives and museums endure. In 2001 the National Library in Norway engaged their own lawyer, expanding till two in 2005, to encounter these challenges. In collaboration with ABM these representatives have compiled updated guides and held conferences on the topic of copyright. They also spend time assisting other institutions throughout the country. [Kulturdepartementet(2009)]

---

<sup>10</sup><http://www.arkivverket.no/arkivverket/Bruk-av-arkiv/Arkiv-og-samlinger/Arkivkataloger/Samkatalogen>



The National Archiver is required by the Archives Act to strive to preserve, maintain and make available archived material created by private individuals. But constraints set by the individual owners of the material at the time of deposit can obstruct digitalization and availability. [Kulturdepartementet(2009)]

In the U.S both copyright and regulatory environments can discourage the best practices for digital preservation. In some cases even making them illegal. Library of Congress calls for updates consistent with the digital era. [Library of Congress(2011)] In Norway revisions of the Copyright Act started in 2009, however the main focus have been measures against illegal file sharing<sup>11</sup>.

Copyright and IPR<sup>12</sup> in general are causing difficulties in many areas, including education and preservation. Copyright and other related issues are similar for long-term preservation and for use in general, leading to difficulties. [Anderson et al.(2006)Anderson, Pringle, Eadie, Austin, Wilson, and Polfreman] The right to create and keep preservation copies are one of this issues. This can lead to privacy claims which in turn cause anguish of collection and documentation, especially of digital material. [Library of Congress(2011)]

The Ministry of Culture<sup>13</sup> stated in St.meld. 24 in 2009 that there will be an evaluation and review of the Legal Deposit Act. The reason being the law have been in effect for 20 years without review in a time of rapid technological development. Leaving it outdated in several areas, legal deposit of digital and online material being one of them [Kulturdepartementet(2009)]

### 3.4.2.1 Intellectual Property

Intellectual property rights is a crucial legal issue in the area of digital archiving. Beagrie (2008) have detailed how one can address the issue of property rights in a satisfactory manner.

- *'Agreements with authors and data owners made clear and recorded. A commitment to keeping the data secure should be stated. Any changes to digital object tracked.'*
- *'A registry of object creators and owners should be created, and their details tracked.'*
- *'Legal context: Can the digital object be reproduced? Make clear explicit agreements with authors on rights for preservation and reproduction of the object.'*

---

<sup>11</sup>Høringsnotat: Endring i åndsverkloven (tiltak mot ulovlig fildeling og andre krenkelser av opphavsrett m.m. på Internett), [http://www.regjeringen.no/pages/16468635/Horingsnotat\\_2011.pdf](http://www.regjeringen.no/pages/16468635/Horingsnotat_2011.pdf)

<sup>12</sup>Intellectual Property Rights, [http://en.wikipedia.org/wiki/Intellectual\\_Property\\_Rights](http://en.wikipedia.org/wiki/Intellectual_Property_Rights)

<sup>13</sup>Kulturdepartementet, <http://www.regjeringen.no/nb/dep/kud.html>

- *'Access issues: Routine access levels should be explained and different levels assigned to different collections or a similar procedure outlined.'*
- *'Deposit agreements and methods of depositing, e.g. self archiving, mediated by staff, tightly controlled.'*

[Beagrie et al.(2008)Beagrie, Semple, Williams, and Wright]

### **3.4.2.2 Public Policy**

Advocacy and changing public policy is by many viewed as a key aspect of improving conditions for archiving. It seems that in many cases the environment for archiving analog material is well established, while the digital realm faces significant hurdles.

In addition to shaping public policy to facilitate digital preservation, the Library of Congress has also set out to adapt the copyright law to better suit digital perseverance. Another important legal issue is that of creating official incentives for digital perseverance. A key area for the Library of Congress has been to examine how tax related incentive could encourage digital preservation and make the process easier. [Library of Congress(2011)]

## Chapter 4

# Trust

This chapter will encompass various themes relating to trust. In other words how an archive should be set up in order to establish trust among users and stakeholders as well as debating various ethical issues related to the establishment of a personal digital archive on a grand scale. The chapter will cover several topics including the three topics of relevance to digital archives highlighted in the Society of American Archivists (SAA)<sup>1</sup> ethical code of 2005 namely; selecting material, restrictions regarding access and publication, and privacy rights of users and subjects. [Baker(2011)]

### 4.1 Ownership

Why is a public institution better suited for the task of establishing and running a personal digital archive on a national scale than a private enterprise? Would not a private company who enjoy more freedom and have a greater incentive when it comes to economical gain, make the archive work in order to obtain profits? This goes back to many of the issues discussed in *Ephemerality*, especially issues relating to the transience of commercial services. This section will examine the key arguments made by scholars on the topic and in turn present an argument that a public entity is vital for the establishment and longevity of a national archive of personal digital collections.

From 1994 to 1996 Donald Waters participated in the Task Force on Archiving of Digital Information's creation of the report '*Preserving Digital Information: Report of the Task Force on Archiving*

---

<sup>1</sup><http://www2.archivists.org/>

*of Digital Information*<sup>2</sup> This report set out to answer the question 'Why should we preserve digital information, and who should do it?' [Waters(2002)]

For the purpose of this segment of the report the second part of the question is of high importance, *who should do it?* For one it has to be an entity that can provide trust. An infrastructure capable of establishing trustworthy solutions for long-term storage. [Waters(2002)] The users have to be sure that their belongings are handled in a secure way. That the material is not being abused in any way, that their conditions are respected, that the material is secure from outside treats and that the infrastructure and the material is there in 100 years when today's users great grandchildren are looking for it.

Walters continues to highlight that up until now much of the archiving responsibilities have fallen on government organizations. One justification for this is given as the importance of *cultural memory*<sup>3,4</sup>. Waters defines cultural memory as a public good, and because digital information, like all information, contributes to societies cultural memory archiving it should also be a public good. [Walters(2002)] Which in turn can be seen as a service that should be provided and distributed by the state of a democracy.

Another argument for public involvement in a personal digital archive is the free rider problem<sup>5</sup>. Excluding beneficiaries will be difficult and can lead to large expenses that is unlikely to improve the functionality of the archive. [Waters(2002)] Investors interested in profit may see this as a reason not to participate, or insist on systems that exclude free riders. But an archive built on a foundation for financial gain may not be as sustainable as an archive with governmental funding, built on the foundation of sustainability.

Another report with a focus on the responsibility of digital archiving is the '*JISC/NPO Studies on the Preservation of Electronic Materials*'. This study was part of a program funded by JISC as a result of a workshop on the Long Term Preservation of Electronic Materials held at Warwick in November 1995. While its focus was not specifically on personal archives, rather digital archives of copyrighted material, its findings are still presented as they can hopefully add to discussion on whether or not a public entity is best suited for the tasks at hand. [Haynes et al.(1997)Haynes, Streatfield, Jowett, and Blake]

---

<sup>2</sup> <http://www.clir.org/pubs/reports/pub63watersgarrett.pdf>

<sup>3</sup> [http://en.wikipedia.org/wiki/Cultural\\_memory](http://en.wikipedia.org/wiki/Cultural_memory)

<sup>4</sup> <https://tspace.library.utoronto.ca/citd/holtorf/2.0.html>

<sup>5</sup> [http://en.wikipedia.org/wiki/Free\\_rider\\_problem](http://en.wikipedia.org/wiki/Free_rider_problem)

Some of the main arguments presented for leaving the task of digital archiving on a national scale up to the government included:

*'Need collaboration of different bodies to ensure preservation of the widest variety of archives. To rely on marketplace actors is probably a mistake. Need a public institution with ultimate responsibility.'* [Haynes et al.(1997)Haynes, Streatfield, Jowett, and Blake]

*'In the end it has to be a government responsibility because of resources and long-terms interests. Government includes the British Library.'* [Haynes et al.(1997)Haynes, Streatfield, Jowett, and Blake]

*'Cynical answer is not the corporates. They would be always doing it for own gain, could be hidden agendas. Wouldn't want that responsibility themselves, nor even less think that their competitors had responsibility for archiving. Has to be in public domain. Corporates would support it financially.'* [Haynes et al.(1997)Haynes, Streatfield, Jowett, and Blake]

One can see that the need for variety, resources, long-term interests and non profit motives are mentioned as some of the primary reasons why the task of digital preservation should be left to a governmental institution or collaboration.

Another interesting excerpt is one concerning financial incentives;

*'Make it economically rewarding to deposit data. As people realise the value locked up in the data they will do it more.'* [Haynes et al.(1997)Haynes, Streatfield, Jowett, and Blake]

It can be assumed that the government would have the ability to do so through tax breaks for participation, grants or in other ways support participation even if it does not make fiscal sense (assuming that a private actor would not).

The most common answer regarding funding consisted of;

*'A digital archive should be funded by government. It should be a national responsibility.'* [Haynes et al.(1997)Haynes, Streatfield, Jowett, and Blake]

It is also interesting to note that many respondents suggested that a digital archive should follow the model of the British Libarary and that the current system for print archiving could be carried over

to the digital realm.

The report concludes with recommending that a digital archive for the public should be publicly funded to ensure continuity and users should not have to pay to access the material. [Haynes et al.(1997)Haynes, Streatfield, Jowett, and Blake]

As can be seen from the arguments presented above some key arguments for government involvement in a national archive of personal digital information include; *The ability to provide a deep infrastructure which can ensure the long-term storage and retrieval of vast amounts of data.* The government might already have experience in this from its analogue archives or should be assumed to have the resources to establish such infrastructure should it so desire. Another argument is that the information stored should be considered a public good and as such it is in the public's interest to preserve it. As such it can be argued that this kind of responsibility would logically fall in the shoulders of a public entity. A final point is that of the free rider problem as long as one entity stores the information others are less incentivized to do so, and this can dissuade commercial actors if they assume that someone else will take the initiative. Or it may harm the sustainability and longevity of a set project. And if one cannot trust the sustainability of this type of project the project itself in many ways is doomed to fail. Therefore it is reported that responsibilities which are affected by the free rider problem often falls into the hands of the government as there are few other stakeholders willing to make the investment needed.

#### **4.1.1 Governmental trust**

If one follows the argument made in the preceding section that a public institution is best suited for the task at hand another question emerges, *why should we trust the government?* Can the government be trusted to store your personal digital collections refrain from manipulation, ensure access is given in an appropriate manner, and can one trust that the material is not being surveilled or misused in any way? The arguments discussed in this section will have a basis in digital archiving so as not to venture of on a discussion as whether or not one can trust the government in general, as that is outside the scope of this thesis.

The Merriam Webster Dictionary defines trust as '*assured reliance on the character, ability, strength, or truth of someone or something - one in which confidence is placed - a charge or duty imposed in faith or confidence or as a condition of some relationship - something committed or entrusted to*

*one to be used or cared for in the interest of another.*<sup>6</sup>

As such it can be argued that many of the cultural archival institutions already command a certain level of trust among the public. Museums, and national libraries and archives are currently trusted with the responsibility of storing and maintaining historical artifacts, analog documents and digital material, preserving these items in order to promote growth of knowledge and make history available for later generations. [RLG/OCLC(2002)]

The work by Flecker (2002) also touches upon the topic of trust. Because of the complex nature and the cost associated with it, digital archiving needs a long-term commitment preferably by a recognized institution. In the past such institutions have typically been national and research libraries, museums or other establishments looking to document the past. Flecker argues that archiving is a serious task and as such it is often the duty of establishment who inherently seek to preserve the past and is expected or trusted to do so by the general public. [Flecker(2002)]

In addition, certain material is protected by copyright or other forms of licensing, and in order to archive such material one needs the permission to store or expertise to negotiate said permission. A government institution such as a national archive is also more likely to receive gifts and donations in the form of archival material from commercial actors or other third parties. [Flecker(2002)] Governmental institutions can be assumed to be experienced in licensing (or the clout to command permission), and could therefore be assumed to ensure that the material supplied is protected by current legislation. Consequently the dual role of government as an archiving entity and as a legal entity can create trust that it will attempt to align laws with the needs of the archive, or at least the possibility is there unlike with private facilitators.

Large scale archives can be both cost and resource intensive. In the past governments have taken the role of custodian for analogue archives and have often been trusted to support the digital equivalents with the continual funding needed to preserve them. [Flecker(2002)] Essentially the governmental experience in archiving, both long-term of running analogue archives for decades (some even centuries) and with running digital archives in other fields should lead to increased trust in the institution (if the track record has been good in past ventures).

The fact that some of the archival infrastructure might already be in place as part of a current governmental setup, might help reduce the cost of establishing a digital archive. [Flecker(2002)] It can therefore be argued that a governmental entity often will be suited to embark on digital archiving

---

<sup>6</sup><http://www.merriam-webster.com/dictionary/trust>

tasks as it has experience and infrastructure from analogue archive and will likely not have to build such facilities from scratch.

Archiving is often aimed at filling an exact need for a specific segment of the population. Trust is a paramount concern when it comes to deciding who to grant the responsibility of preserving the material in question. The desire is to find an institution which can provide expertise, values and resources need to support the archive. [Flecker(2002)] This can be interpreted so that a set institution have to provide a clear and suitable mission statement for the intended archive to establish additional initial trust. Furthermore trust will be built over time if the archive fulfills the promises given and meets the needs of the community.

As government agencies often possess a larger resource pool than nonprofit entities in the private sector, they are more likely to support large scale efforts to document history, technology and culture, for example in a digital archive. [Flecker(2002)] Based on this one can consider that the community will likely trust the government to keep funding the archive if the initial initiative of establishing it have been met.

It should be mentioned that one caveat of governmental archives is the potential for political or ideological bias. If a doctrine exists to promote a certain historical view, it can influence what is collected and stored in the archive. Political concerns may also influence censorship of materials such as hate propaganda, pornography or other material which may be deemed controversial at the time or in hindsight. [Flecker(2002)]

Another approach to quantify trust is the RLC/OLC paper *Trusted Digital Repositories: Attributes and Responsibilities*. The focus lies on defining which qualities make up a *trusted digital repository*. Therefore by detailing their findings and analyzing whether or not this is a fit with government in general and subsequently the proposed archive it should offer another indication on the level of trust one can expect to encounter should the project be initiated. [RLG/OCLC(2002)]

It is stated that a trusted digital repository delivers a trusted, perpetual access point to digital material under active management to the population both at the present and in the future. However to fulfill the criteria of a trusted digital repository the following must be a part of the institution;

- Be responsible for long term management of digital material for both depositors and users.
- Incorporate a system supporting long-term viability of the archive and the digital material within.



- Exhibit financial responsibility and longevity.
- Have systems created within accepted standards and conventions in order to ensure smooth management, publication and security.
- Evaluate the system in a trustworthy manner.
- Be open and explicit in carrying out its responsibilities.
- Supply clear practices, policies and performance that can be checked and audited.

[RLG/OCLC(2002)]

Furthermore, it is stated that the management of integrity, identity and the overall quality of the archive is instrumental in obtaining the trust of the users. They have to be confident that the archive is fulfilling its promise and that whatever they deposit is safely stored for the long-term. [RLG/OCLC(2002)]

Another aspect of trust is the authentication of documents. An individual accessing stored material must be able to trust the authenticity of the digital document. As such the institution needs to have a solid system in place to ensure authenticity upon deposit and during maintenance and storage. [Peter Graham(2000), cited in [RLG/OCLC(2002)]]

The list below denotes what a panel of experts considered to be elements of a trusted digital repository;

- '*Compliance with the Reference Model for an Open Archival Information System (OAIS)*'<sup>7</sup>
- '*Administrative responsibility*'
- '*Organizational viability*'
- '*Financial sustainability*'
- '*Technological and procedural suitability*'
- '*System security*'
- '*Procedural accountability*'

[RLG/OCLC(2002)]

---

<sup>7</sup>[http://en.wikipedia.org/wiki/Open\\_Archival\\_Information\\_System](http://en.wikipedia.org/wiki/Open_Archival_Information_System)

It is argued that trust is more than the responsibility of managing and storing data files. The institution have to build trust from the ground up through its mission statement, systems, track record, and experience in adjacent areas of archiving. Due to this it is proclaimed that a certification process is one possible solution to fast track the establishment of trust to an institution. [RLG/OCLC(2002)]

As can be seen attributes of government itself such as acting in public interest, significant resources, experience in the archiving field (both analogue and digital), and existing infrastructures, should lend a certain element of trust in its sustainability for the task at hand. Furthermore, an individual will have past experiences with government, good or bad, to base its initial trust on. The potential for further building on this trust and eventually spreading it by the word of mouth, lies in the work around the archive itself.

#### **4.1.2 Archiving selection**

Personal archiving strategies have been covered in an earlier chapter of this thesis, this section will focus on what to store in terms of content and associated risks.

An important aspect to consider when discussing what to include in a personal digital archive is the risk of potentially storing libel or slanderous information. While this information might be a valid and illustrative part of a persons communication it also presents the person and potentially the archive to legal claims. The Paradigm project defined cyber libel as defamation which occurs online regardless of its format. (ie blog, email, website, message board etc). The court generally considers the offense to have been committed in the jurisdiction where the material was accessed. As such the legalities can be quite complex. [The University of Manchester(2008a)]

One recommended processes a personal archive must incorporate if it is to catch such material before it unknowingly shares it with the public, is to provide access to private archives of living individuals in supervised research rooms and to accredited readers only. [The University of Manchester(2008a)]

Unfortunately personal archives are especially exposed to libelous content due to the nature of personal communication and its expressive and often informal nature. This puts extra demands on the archivists to review material before making it available to the general public. [The University of Manchester(2008a)]

The recent case of Twitter and the Library of Congress can serve as an illustration of what might be published, and as such provide the individual with an indication of what he or she might be comfortable sharing on the basis of past activity on social media sites. The decision to create an archive of the donations made by Twitter was not made by consulting the users who created the content. Said content was donated by a commercial entity, and it is likely to get published with a six month delay from the date of creation. [Baker(2011)]

As one can see the decision as to what to store puts demand on both the individual and the archiving institution. Not only will one have to make choices as to what is stored and how but one also have to consider potential publication of the material in the future. [Baker(2011)]

## 4.2 Publication

Another aspect of trust is how the archiving institution goes about publishing the content it receives. The rights of the individual and associated third parties will have to be factored into the ambition to share the material with the world. This section will examine some common ways of publishing, and what to take into consideration when deciding what material to publish and when.

Publication seems to be a fundamental part of a personal digital archive. In fact, Waters (2002) states that *access is the key*. By this he means that access is one aspect where investment can be drawn in, as the control of material will likely have a value for commercial players as well as governmental institutions. [Waters(2002)] Therefore it can be argued that the decision to publish and how can be crucial for the success of the archive.

Having identified some of the prevailing arguments on publishing in digital archiving the *JISC/NPO Studies on the Preservation of Electronic Materials*'s findings also emphasizes the importance of access. The majority of the responses indicated that access should be granted but with a set of restrictions. One suggestion is to use the setup currently in place for paper archives, where users generally have to demonstrate a clear need in order to gain access to the material. [Haynes et al.(1997)Haynes, Streatfield, Jowett, and Blake]

The Paradigm project recommended that a balance must be struck between preservation and access. The need of the material contributors who may value privacy and property rights must be weighted against the need of the users who wants access. Another aspect of publishing is the possibility of opening the institution up to libel and slander charges. Digital material can be spread quickly

through the Internet, any libelous or slandering material have the potential to reach a vast audience very quickly. This means that certain parts of the archive can be released while others remain private. The danger of potential lawsuit might put an extra burden on curators and restrict what they may publish. [The University of Manchester(2008b)]

Baker (2011) provides an in depth discussion on ethics within her research journal *Ethical Considerations in Web 2.0 Archives*. The journal touches upon several aspects of access and publication. An *Web 2.0*<sup>8</sup> archive will normally have some identifying properties connected to its material. While general inferred statistics might be useful, the real value may lie in the individual stories being told. At the same time it is important to make sure that sensitive information such as passwords and credit card details do not become part of a public archive. [Baker(2011)] Limiting access and anonymizing certain data is suggested by MacNeil (1992) as a means to publish while protecting the individuals rights, also suggested is restricting access only to scholars or a similar set of individuals. [MacNeil(1992), cited by [Baker(2011)]] However it is argued that such a policy often goes against the mission of many archives as well as governing ethics documents. A scholar only access policy increases the burden on curators to act as gatekeepers and credential checkers. [Baker(2011)]

The timing of when to make specific content in an archive available is a subject of much debate. Hodson (2005) argues for the established method of sealing archives for a period of time, even past the lifetime of the author. This also is suggested by Gaudette (2003) when discussing blind donor material. Her suggestion was to go by the example of the United States Copyright Law, meaning the life of the author plus 70 years, unless consent was provided. It can be argued that Web 2.0 users have a lot in common with blind donors. One can never know who stores your posts or who decides to donate the content of your third party stored archive. Postings that were meant for friends and family may be shared with the general public without the authors knowledge or expressed consent. [Hodson(2005), Gaudette(2003), cited in [Baker(2011)]] As such Baker argues that in a Web 2.0 environment the arguments of Hodson and Gaudette may not be valid. The material created consist of a vast amount of blind donors and it is not possible to track the identities of contributors and subsequently nor their time of death. If one was to impose a time delay in the manner of an analogue archive or copyright provisions the value of the digital archive will likely be lessened. It is argued that the time restriction of six months may be suitable for publishing Web 2.0 material. This would allow users to delete their profile before publication could occur (an implied six month timeframe) while still keeping the material relevant. [Baker(2011)]

---

<sup>8</sup>[http://en.wikipedia.org/wiki/Web\\_2.0](http://en.wikipedia.org/wiki/Web_2.0)

Another proposed solution is to leave the issue of access and potential restriction in the hands of the donor. Greene (1993) advocated the use of donor-imposed restrictions. The argument made was that the donor would understand the nature of the material and the associated risks. [Greene(1993), cited in [Baker(2011)]] However one cannot be sure of the motivations of the donor, or in the case of corporations it may not be in their best interest to create extra protection for third party users, as they may well prioritize their own interest over that of an unknown third party. [Baker(2011)]

A recent case with relevance for a personal digital archive is the case of Twitter and the Library of Congress. This case can be used to highlight how access was handled in one of the most significant donations of personal correspondence in 2011. The published material would not contain any account information or Tweets that had been deleted by the user. In addition a six month delay would be enforced from creation to publication of a Tweet. It is also the intention of the Library of Congress to make the archive available to researchers, but it is not yet known if this is a policy of the institution or a condition of the donation. [Baker(2011)]

### 4.3 Rights of third parties

Protection of third party rights is a vital aspect of the proposed archive. Due to the personal nature of the material it should be assumed that friends and acquaintances of the contributor could be depicted in pictures or discussed in personal correspondence. The section will therefore detail some of the prevailing themes in the field of digital archiving when it comes to protecting third party rights.

One treat to the rights of third parties in archived material comes in the form of the *U.S. Patriot Act*. If archives are stored *in the cloud* and some of these servers are on U.S. soil it could have an effect on who can gain access to the material. The location where one stores, process' and collects information can open up to U.S. jurisdiction and the potential risks that involves. [Gallagher(2011)]

Baker (2011) highlights the protection of third party rights under *The International Council on Archives (ICA)*<sup>9</sup> code of ethics of 1996. In this document it is stated that '*they must respect the privacy of individuals who created or are the subjects of records, especially those who had no voice in the use or disposition of the materials*'. Suggesting that special attention must be taken to preserve rights of third parties especially those of blind donors or whom have been depicted or

---

<sup>9</sup><http://www.ica.org/3/homepage/home.html>

described as part of an acquaintances personal archive. In other words it is likely that when the ethic code highlights the need to protect third parties, it can in turn create trust of the archive if it follows said code. Such official sources of guidance as the ethics code can be a valuable signpost for both archivist and potential donors. [Baker(2011)]

Baker concludes that digital archives containing Web 2.0 information are important and worth preserving. However to protect the rights of third parties she recommends the following measures to be put in place.

- *'Anonymize sensitive personal information, such as account registration information, that will not be useful to users, keeping in mind the mission of the archive;'*
- *'Seal archives for an appropriate period of time;'*
- *'Implement technical solutions to permit access to the archive while protecting privacy;'*
- *'Establish policies for addressing complaints and monitoring potential privacy violations;'*
- *'Revise and update professional ethics codes to address Web 2.0 archive challenges; and'*
- *'Educate other archivists about Web 2.0 archive problems and solutions. Archivists should use their expertise to educate and influence donors of archive materials. In the case of a corporation donating material created by third parties, archivists should work with corporate donors to create policies that allow third parties to make informed decisions, such as:*
- *'The archival institution, the existence of the archive, and the restrictions on its use should be clearly disclosed on the website's terms of service or privacy policy;'*
- *'Users should be given an opportunity to opt-out (or opt-in) to archives via a checkbox;'*
- *'Changes to the archive policy should be disclosed to users in the same way changes to the website's terms of service or privacy policy are disclosed; and'*
- *'In the event that archivists cannot persuade corporate donors to adopt measures reasonably protective of third-party privacy, and cannot protect third-party privacy completely in the archive in practice, the archive gift should not be accepted.'*

[Baker(2011)]

OpenFran<sup>10</sup> provides a set of guidelines formed on the principals of the major legal documents on

---

<sup>10</sup>A resource archive for franchises

third party rights within archives in the US.<sup>11, 12, 13</sup> as well as applicable law. These guidelines deal with instances where the archive is approached by a person requesting removal of material due to privacy concerns. The list of common requests and suggested responses can be found in table 4.1. [The Franchise Openness Project(2002)] A similar set of transparent guidelines could be a valuable addition to a personal national archive as it provides a clear way to handle third party grievances.

Type of removal request	Response
Request by a private (non-governmental) group, typically for reasons of privacy, defamation, or embarrassment.	<ol style="list-style-type: none"> <li>1. Archivists should provide a "self-service" approach private groups use to request removal of their materials.</li> <li>2. Requesters may be asked to substantiate their claim of ownership.</li> <li>3. This allows archivists to ensure that material will no longer be gathered or made available.</li> <li>4. These requests will not be made public; however, archivists should retain copies of all removal requests.</li> </ol>
Third party removal requests based on the Digital Millennium Copyright Act of 1998 (DMCA).	<ol style="list-style-type: none"> <li>1. Archivists should attempt to verify the validity of the claim by checking whether the original document has been taken down, and if appropriate, requesting the ruling(s) regarding the original electronic document.</li> <li>2. If the claim appears valid, archivists should comply.</li> <li>3. Archivists will strive to make DMCA requests public via Chilling Effects, and notify searchers when requested documents have been removed.</li> <li>4. Archivists will notify the submitter of the affected document, generally via email.</li> </ol>
Third party removal requests based on non-DMCA intellectual property claims (including trademark, trade secret).	<ol style="list-style-type: none"> <li>1. Archivists will attempt to verify the validity of the claim by checking whether the original documents have been removed, and if appropriate, requesting the ruling(s).</li> </ol>

<sup>11</sup>American Library Association's Library Bill of Rights, <http://www.ala.org/work/freedom/lbr.html>

<sup>12</sup>The Society of American Archivists Code of Ethics, [http://www.archivists.org/governance/handbook/app\\_ethics.asp](http://www.archivists.org/governance/handbook/app_ethics.asp)

<sup>13</sup>The International Federation of Library Association's Internet Manifesto, <http://www.ifla.org/III/misc/im-e.htm>

	<ol style="list-style-type: none"> <li>2. If the original documents have been removed and the archivist has determined that removal from public servers is appropriate, then the archivists will remove the documents from their public servers.</li> <li>3. Archivists will strive to make these requests public via Chilling Effects, and notify searchers when requested documents have been removed.</li> <li>4. Archivists will notify the submitter of the affected document, generally via email.</li> </ol>
<p>Third party removal requests based on objection to controversial content (e.g. political, religious, and other beliefs).</p>	<p>As noted in the Library Bill of Rights,</p> <p>”Libraries should provide materials and information presenting all points of view on current and historical issues. Materials should not be proscribed or removed because of partisan or doctrinal disapproval.”</p> <p>Therefore, archivists should not generally act on these requests.</p> <p>Third party removal requests based on objection to disclosure of personal data provided in confidence.</p> <p>Occasionally, data disclosed in confidence by one party to another may eventually be made public by a third party. For example, medical information provided in confidence is occasionally made public when insurance companies or medical practices shut down.</p> <p>These requests are generally treated as requests by authors or publishers of original data.</p>
<p>Requests by governments.</p>	<p>Archivists will exercise best-efforts compliance with applicable court orders.</p> <p>Beyond that, as noted in the Library Bill of Rights,</p> <p>”Libraries should challenge censorship in the fulfillment of their responsibility to provide information and enlightenment.”</p>
<p>Other requests and grievances, including underlying rights issues, error correction and version control.</p>	<p>These are handled on a case by case basis by the archive and its advisors.</p>

Table 4.1: List of responses [The Franchise Openness Project(2002)]



## **Chapter 5**

# **A personal national archive**

Figure 5.1 illustrates the concluding remarks on a proposed national archive. It is divided into internal factors which the archive itself remains in control of. These include technology choices, security, accumulation and ethics. Outside influences can however aid or impede the efforts of the archive these include, technology development, media attention, legislation, Norwegian government, users and society in general. Both the internal and external factors will likely contribute to the overall success of the archive. The major elements of the model will be detailed in the segments below.

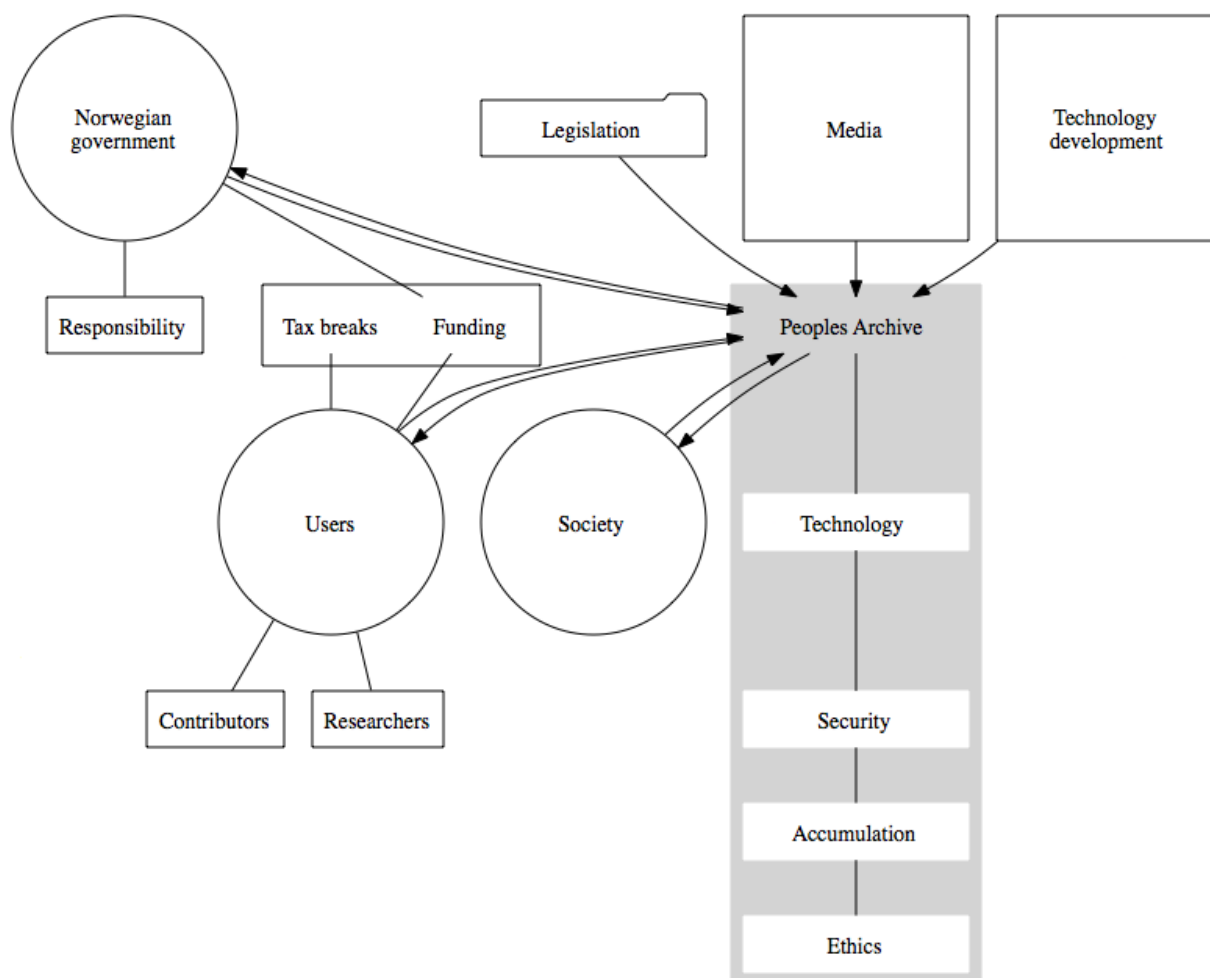


Figure 5.1: Model of a personal national archive.

## 5.1 Overview

The archival sector in Norway is structured by 44 archival institutions. Most of the archived material is stored at relatively few public and private institutions many of whom do not have archiving as a primary task. Unlike state entities, private companies, organizations and individuals can themselves decide how to dispose of their records. The private records that are preserved are distributed at various museums, libraries and archives throughout the country, and many of them in an unorganized

manner. [Kulturdepartementet(2009)] The current state is that the life of private individuals and organizations are poorly documented and the effort related to private archiving has a long way to go. Adequate funding and more coordinated and systematic infrastructures for the whole country could help better the conditions. [Østby and Ertesvåg(2007)]

One potential helpful trend is peoples increased openness to organize and share personal information. The increased adaptation of such activities could suggest that the general public would be open to continue sharing personal information also for the long-term. The familiarity with categorizing images, grouping and publishing material regularly may go some way in making people aware of the basic concepts of digital archiving. A personal digital archive could aim to capitalize on this trend and make users take the next step in personal archiving.

In the recent past the archival sector in Norway has been criticized for a lack of responsibility and authority when it comes to personal archives. While both the state and county have dedicated entities responsible for their archives, personal archives are left in limbo. [Brekke(2010)] As such there is a clear need for an initiative like the project *Norske Mediearkiv* are aiming for, and the timing seems to be good both in connection with current trends amongst users as well as increasing government attention into the issue.

In regards to the adjacent work on the project there are currently two other master projects on the subject at Østfold University College, such as the creation of a prototype.

## 5.2 Legislation

One can assume that for a personal digital archive to have any affect on the legislative process it should be a part of the overall archival initiative. The best indication of this is the governments current plans to alter the laws to better suit archival needs. [Kulturdepartementet(2009)] The governmental strides to adapt the current legislation is in line with one of Library of Congress main goals, namely to change public policy. [Library of Congress(2011)] The institution has considerable leverage and support and their initiative provide an interesting example of how a Norwegian archive could seek to influence key stakeholders.

Some of the most important aspects of Norwegian law which could impede the archival efforts and should be considered include;

- Data protection act:
  - Consent needed or greater societal need
- Copyright act:
  - Authors lifespan and 70 years after his death, 15 for photo, 50 years after it was taken
  - Consent of the depicted person
- Archives act:
  - May have to follow guide lines from the national archiver if funding is received

### 5.3 Technology and security

An online portal for uploading and accessing material could be a good solution for a private digital archive. The example of the First World War Poetry Digital Archive could serve as a guide in how to setup such a service.

As a part of this project a visit was paid to UNINETT's offices in Trondheim for a meeting with Olaf Schjelderup. He listed a range of issues to consider when forming a national archive. A full list of highlights can be seen in appendix B

In terms of structuring the material he focuses on ensuring its authenticity through digital stamps, metadata control, controlling legitimacy and providing user friendly upload environments. He lists diversity, distribution and peer-to-peer technology as measures to ensure security. He also mentions three aspects that should be in place to ensure longevity. They can be seen in figure 5.2 as a model for distributed storage. Geographical distribution refers to the importance of retaining copies of the material in different locations to ensure protection against physical threats. It can also be a good idea to keep a frozen copy of the material which is not tampered with in any way. The last issue is to ensure technological diversity. This refers to keeping copies of the material in different technological states to best protect against obsolescence. [Nilsen(2011a)]

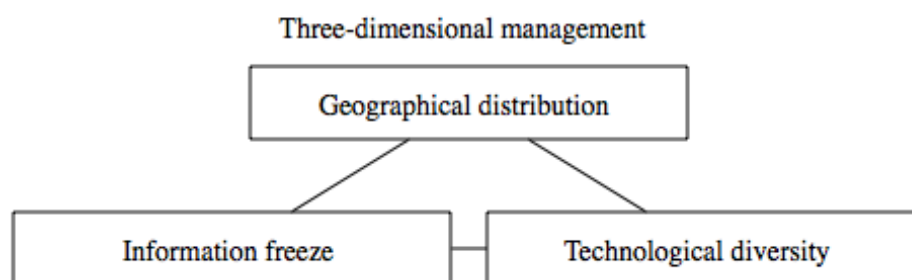


Figure 5.2: Model for sustainable preservation.

In selecting which file formats to include in a personal archive it is important to minimize the number of file formats. The chosen formats should fulfill criteria such as Ubiquity, Support, Disclosure, Documentation quality, Stability, Ease of identification and validation, Intellectual Property Rights, Metadata Support, Complexity, Interoperability, Viability, Re-usability. [Brown(2008)]

In order to establish trust the treat of hacking should be taken seriously and appropriate precautions put in place. As several examples, such as the Playstation Network hacking<sup>1</sup> shows, the trust in a service can be greatly diminished if hackers can gain access to personal information. This would be even more severe in the case of a personal archive as not only will personal information be compromised but also entire collections of personal memories. One potential solution to this is to let benign hackers test the system prior to launch.

In addition, simple and user friendly design is paramount in order not to exclude any willing participants. A function that would allow users to upload all email or all photos on a home computer to the archive for selection and tagging could be a good idea. If the system is too complex it could alienate or exclude older members of society and as such lose an important input into the archive.

The archive should offer the same level of service as commercial services while the intention at the present time is to keep it free, or even subsidized. Several sources mentioned tax breaks as a means to not only promote the creation of an archival institution but potentially also to entice contributions. Could a scheme where users are incentivized to participate against tax credits be a viable solution? Or is the fiscal reality such that users would actually have to pay to participate? This is outside the

<sup>1</sup>[http://en.wikipedia.org/wiki/PlayStation\\_Network\\_outage](http://en.wikipedia.org/wiki/PlayStation_Network_outage)

scope of this thesis, but it is an interest question for further research.

## 5.4 Association and access

In Norway all citizens are given their own identification number at time of registration after birth or legal immigration. This identification number consist of data of birth and a personal number consisting of five figures. [Skatteetaten(2012)] The identification number is used to distinguish one person from the next and are used as identification in many public registration forms, as well as by banks, schools and others institutions. [Datatilsynet(2011)] MinID is a public login service based on the identification number and can be used to administer 50 public services. [MinID(2012)] Within a personal archive where government is involved the use of identification number and MinID could be a possible solution for the identification of users. Not only could this provide a level of trust in an already established system, trusted by 50 public institutions. In addition many people already use it for their management of public services and this familiarity and notion of security could help establish initial trust in the archive.

The system should facilitate the possibility of a user providing a list of heirs which would get access to the material upon the users death. This would ensure that the material is made accessible to those that would value it, and they could provide further content. Publication has been identified as a key element of an archive of this scale. Publication issues will likely be handled through donor agreements or as per the current law which at the time of writing would indicate 70 years after owner death for most material and 15 years after death as well as 50 years after a image was taken for photographs, unless other provisions have been requested.

Donor agreements have been mentioned by the Data Inspectorate as one way to design a system that would not be restricted by current laws. There are drawbacks to this approach but for a personal archive based on donations by individuals it should be evaluated as a possible solution if the current legislative environment is found to be to restrictive. Greene made the argument that the donor would understand the nature of the material and the associated risks. [Greene(1993), cited in [Baker(2011)]] However one cannot be sure of the motivations of the donor, or in the case of corporations it may not be in their best interest to create extra protection for third party users, as they may well prioritize their own interest over that of an unknown third party. [Baker(2011)] As such the benefits of donor agreements have to be judged against the risks in relation to a personal archive. If the intention is for individuals to donate person material donor agreements could work

but the rights and consequences would have to be spelt out clearly to inform the users of his or her options.

It would have to be decided if an archive should take on a private sponsor or if it would accept corporate donations of material such as the case of the Library of Congress and Twitter. This would then require appropriate resources to protect third party rights and evaluate collections. Another choice to make is whether to make aggregated statistical material from the archive available. This could be of value to researchers and it does not necessarily have to infringe on the rights of those who have participated with material.

## 5.5 Accumulation

Accumulation is an issue in Norwegian archives today. One important aspect for a personal digital archive is to create a dedicated entity so personal archives does not have to compete for attention and funding with the significant digitalization backlog present in state and county archives today. For instance, the analog record stock in the Norwegian archival system today, include maps, drawing and photographs and was at the end of 2010 approximately 230 000 self meters; of this was 195 000 self meters public material and 34 000 self meters from private archives. Currently the private archives constitute 14,9 percent of the overall official archival holdings in Norway and in 2010 just 258 private archives were deposited. [Kulturdepartementet(2011)]

Analog archives can take up much physical space, and will likely create lasting expenditures for management, rent and energy. [Kulturdepartementet(2011)] As a digital archive accumulates the same issues will occur but on a different scale. Because submission of private archives are voluntary there are no demands to organizations or individuals. [Kulturdepartementet(2009)] Consequently, 70 % of all private records deposited to archival institutions are not organized in a systematic manner. [Østby and Ertesvåg(2007)]

It can be argued that to prevent a personal digital archive from growing out of proportions it will be important not to get the image of being a national back-up system. It is important to make the users understand that this is an archive and for them to treat it as such. Although it will be crucial with a user friendly uploading system that helps capture, upload, share and publish in a simple way, it is equally important to encourage the users to categorize, group, tag and so on to get as much metadata per item as possible for future research and retrieval.

## 5.6 Integrity and ethics

Archiving is often aimed at filling an exact need for a specific segment of the population. Trust is a paramount concern when it comes to deciding who to grant the responsibility of preserving the material in question. The desire is to find an institution which can provide expertise, values and resources need to support the archive. [Flecker(2002)] This can be interpreted so that a set institution has to provide a clear and suitable mission statement for the intended archive to establish additional initial trust. Furthermore trust will be built over time if the archive fulfills the promises given and meets the needs of the community.

In order to establish trust and fulfill the mission of a personal archive it is important to adhere to the recommendation of focusing on long-term preservation, ensuring strong financing, use accepted standards, be open and allow policies and performance to be audited. [RLG/OCLC(2002)]

As previously mentioned delayed publication is dictated by today's laws in cases where other indications have not been given. This might, in some cases, also be the best way to protect a third party, and may therefore be necessary even if the owner is willing to publish sooner. Other measures to protect the rights of third parties include; anonymizing sensitive personal information, monitoring privacy violations and establishing policies for addressing complaints. [Baker(2011)]

## 5.7 Roles

A personal national archive is intended as an easy to use archive which eventually will be available for all Norwegians. While it might be rolled out in a smaller scale initially (per county, town or even academic institution) the ultimate goal must be for all Norwegians to be able to participate if they want to. Other stakeholders is the government who is looking to preserve the contemporary for the future, and improve on what seems to be a neglected area of public services. Researchers and potentially private companies might also have an interest in the archive.

The media might also serve an important role in educating people about the archive as well as building trust and explaining its services.

**Contributors** This may slowly be changing and John(2008) and Cox (2008) highlights the potential of *citizen archivists* where people grow accustomed to administrate and manage personal digital



collections through the adaptation of software specially designed for these tasks. The importance of having regular people acting as archivists of their own personal digital collections is considered crucial if the information is to be preserved longterm. [John(2009)], [Cox(2009)]

In addition companies might want to donate material to the archive. This could be dedicated archival type services (facebook, nettby etc) or private companies wanting to ensure their place in the common history.

**Researchers** The archive has to make the decision on whether or not to restrict access to those with a demonstrated need, ie researchers, or open it up to the general public. One solution could be to let the users tag material as private, network only, research only or public, with potentially a deadline of 70 years after that.

**Facilitators** Facilitators or partners identified through this research include;

- The National Library.
- The National Archives.
- Universities.
- The Data Inspectorate.
- UNINETT.
- Arts Council Norway.
- Ministry of Culture.
- Agency for Public Management and eGovernment (Difi)<sup>2</sup>.
- Ministry of government administration, reform and church affairs<sup>3</sup>
- The Research Council of Norway<sup>4</sup>.
- Preus museum.
- Landslaget for lokal- og privatarkiv (LLP)<sup>5</sup>.

---

<sup>2</sup>Direktoratet for forvaltning og IKT, <http://www.difi.no/>

<sup>3</sup>Fornyings-, administrasjons- og kirke departementet, <http://www.regjeringen.no/nb/dep/fad.html?id=339>

<sup>4</sup>Forskningsrådet, <http://www.forskningsradet.no/no/Forsiden/1173185591033>

<sup>5</sup><http://llp.no/>

**Norwegian government** The Norwegian government through the culture minister Anniken Huitfeldts indicated at a conference in november 2011 that the Ministry of culture is planing to publish a *stortingsmelding* on archival issues some time this year. She also stated that archiving is one of the most intangible parts of cultural life, but that in the future neglect will be more visible here than in many other parts of society. Because of this there will be an increase in the budget for archival needs, as well as increased assistance to private and county archives. However, establishment of new institutions is not an option, instead she is promoting new activity to take place within existing structures. [Huitfeldts(2011)] A personal national archive would therefore most likely have to find its place in the existing archival structure.

One of the main initiatives in the budget proposal for 2012 is an increase of 24,4 million kr. to the archival sector. The Art Council Norway who are responsible for the archival field are also responsible for subsidies given to institutions working to preserve and maintain records from private companies, organizations and individuals. This is done to promote a more cohesive documentation of contemporary society. [Kulturdepartementet(2011)]

## 5.8 Contribution to society

There has been an identified need for a dedicated personal archive in Norway. This has been show in the current archival structure where personal archives are lagging behind state and county archives. It has also been supported by current research into the field stating that there is an immediate need for attention to digital personal archiving. This project aims to address these issues by providing the foundation of a personal archive for the people of Norway.

Furthermore a dedicated effort to archive personal information and entice people to participate through intuitive interfaces and public awareness campaigns a broader fabric of society will get its past recorded. This will ensure that the present will be recorded in a more complete manner and thus have significant value in the years to come. . [John(2009)] In addition the ease of which digital information can be created and stored allows insight into new aspects of human life. [Kim(2010)]

The importance of personal digital archiving is more important than ever as the number of documents created and stored in digital form grows. However, the fragile environment jeopardizes the long-term survival of digital artifacts. This is more so the case for personal documents which are created by individuals of varying means and skills, and if not deposited might not get the care and

attention needed to preserve it. [Kim(2010)] Digital personal information is threatened and clear initiatives such as a dedicated national archive are needed to put them back on a track where its long-term survival is ensured.

## 5.9 Conclusion

It would seem that the timing is right for the initiative of a personal national archive in Norway. Not only is the area getting more attention from the responsible governing bodies, there is also no clear alternative which can fill this need at the present time. Regardless of what is proposed by the government next year, it is not inconceivable that input into the process of establishing a national archive for personal information would be taken on board. If one also considers the growing trend among the general public to start archiving their lives through Facebook, Twitter, Flickr and the like it seems that both sides of the equation is about to match.

However it would also seem that any initiative would have to be within the existing archival structures. The government has made it clear that it does not see the need for establishing new archival institutions at the present time. Given its recent disbanding of ABM it is a challenge to see how a personal national archive would be setup within the current structure. As such this area is recommended to be explored in future research.

This thesis should be viewed as one building block towards the goal of establishing a personal national archive in Norway which will stand the test of time. By reviewing the current state of affairs and the established literature in the field through research and field studies the goal has been to build a foundation for continued research. As such it will be exciting to see how the current endeavours at Østfold University College will benefit from the work presented here. It is encouraged that this document can be used freely as reference and is continued to be updated as the theme progresses. Further steps could include the prototype already under development, more insight into the financial aspects of the archive, and potentially launching a test archive in a closed environment such as Østfold University College.

To conclude this paper has established the need for a personal digital national archive in Norway. It should have the involvement of government and it should aim to include all facets of Norwegian society from the very core of its mission statement all the way to an initiative interface for depositing material. The prize for succeeding would be a representative tapestry of the present stored for

posterity while the cost of failure is potentially a lost generation. Nothing would please this author more than to be among the first to deposit into a personal national archive in the future.

# Bibliography

- [Anderson(2007)] Nate Anderson. Lifeblogging: Is a virtual brain good for the real one? *Ars Technica*, February 2007. URL <http://arstechnica.com/old/content/2007/02/8790.ars>.
- [Anderson et al.(2006)Anderson, Pringle, Eadie, Austin, Wilson, and Polfreman] Sheila Anderson, Mike Pringle, Mick Eadie, Tony Austin, Andrew Wilson, and Malcolm Polfreman. Digital images archiving study. *Arts and Humanities Data Service*, March 2006. URL <http://www.jisc.ac.uk/media/documents/programmes/preservation/finaldraftimagesarchivingstudy.pdf>.
- [Andressen(1986)] August Andressen. *Krigen sett fra Lista*. Forlaget Lister, Farsund, 1986.
- [Archive-It(2011)] Archive-It. Archive-it, learn more, 2011. URL <http://archive-it.org/learn-more>.
- [Arora(2006)] Jagdish Arora. Digital preservation and management: An overview. November 2006. URL <http://ir.inflibnet.ac.in/dxml/bitstream/handle/1944/1187/1-20.pdf?sequence=1>.
- [Arvidson(1999)] Tommy Arvidson. Skildra samtiden nu - i morgon är den borta. In Eva Dahlman, editor, *Verklighets bilder? Om fotografer och fotografier i museernas samtidsdokumentation*, pages 101–104. Nordiska museets forlag, 1999. ISBN 9171084487.
- [Baker(2011)] Antoinette E. Baker. Ethical considerations in web 2.0 archives. *SLIS Student Research Journal*, 1, July 2011. URL <http://scholarworks.sjsu.edu/cgi/viewcontent.cgi?article=1010&context=slissrj>.

- [Barksdale and Berman(2007)] Jim Barksdale and Francine Berman. Saving our digital heritage. *washingtonpost.com*, May 2007. URL [http://msl1.mit.edu/furdlog/docs/washpost/2007-05-16\\_washpost\\_digital\\_libraries.pdf](http://msl1.mit.edu/furdlog/docs/washpost/2007-05-16_washpost_digital_libraries.pdf).
- [Beagrie(2005)] Neil Beagrie. Plenty of room at the bottom? personal digital libraries and collections. *D-Lib Magazine*, 11(6), June 2005. URL <http://www.dlib.org/dlib/june05/beagrie/06beagrie.html>.
- [Beagrie et al.(2008)Beagrie, Semple, Williams, and Wright] Neil Beagrie, Najla Semple, Peter Williams, and Richard Wright. Digital preservation policies study, part 1: Final report. October 2008. URL [http://www.jisc.ac.uk/media/documents/programmes/preservation/jiscpolicy\\_plfinalreport.pdf](http://www.jisc.ac.uk/media/documents/programmes/preservation/jiscpolicy_plfinalreport.pdf). A study funded by JISC.
- [Bellan(2011)] Greg Bellan. Can you rely on the cloud for your business? *Acroment*, May 2011. URL <http://www.acroment.com/2011/05/can-you-rely-on-the-cloud-for-your-business/>.
- [Brekke(2010)] Åshild Andrea Brekke. Jo betre søl, medverknad i arkiv, bibliotek og museum. *ABM-skrift*, 62, 2010. URL <http://www.abm-utvikling.no/publisert/abm-skrift/abm-skrift-62>.
- [Brosius(2003)] Maria Brosius. *Ancient archives and archival traditions: Concepts of record-keeping in the ancient world*. Oxford University Press Inc, 2003. ISBN 0-19-925245-9.
- [Brown(2008)] Adrian Brown. Digital preservation guidance note, selecting file formats for long-term preservation. August 2008. URL <http://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf>.
- [Bush(1945)] Vannevar Bush. As we may think. *The Atlantic Monthly*, July 1945.
- [Chen(2001)] Su-Shing Chen. The paradox of digital preservation. *Computer*, March 2001.
- [Cohn and Hibbitts(2004)] Ellen R. Cohn and Bernard J. Hibbitts. Beyond the electronic portfolio: A lifetime personal web space. *Educause Quarterly*, 27(4), 2004. URL <http://www.educause.edu/EDUCAUSE+Quarterly/EDUCAUSEQuarterlyMagazineVolum/BeyondtheElectronicPortfolioAL/157310>.

- [Cox(2009)] Richard J. Cox. *Personal Archives and a New Archival Calling: Readings, Reflections and Ruminations*. Litwin Books, LLC, January 2009. ISBN 978-0-9802004-7-8. URL <http://litwinbooks.com/personalarchives.php>.
- [Datatilsynet(2011)] Datatilsynet. Fødselsnummer, 2011. URL <http://www.datatilsynet.no/Sikkerhet-internkontroll/Fodselsnummer/>.
- [Finney(1986)] Andy Finney. The bbc domesday project, November 1986. URL <http://www.atsf.co.uk/dottext/domesday.html>.
- [Flecker(2002)] Dale Flecker. Organizational models for digital archiving, March 2002. URL <http://www.clir.org/pubs/reports/pub114/appendix1.html>.
- [Fleischhauer(2008)] Carl Fleischhauer. Digitizing history (and preserving history in the making), Desember 2008. URL [http://www.digitizationguidelines.gov/audio-visual/documents/GovVideo\\_Fleischhauer\\_2008.pdf](http://www.digitizationguidelines.gov/audio-visual/documents/GovVideo_Fleischhauer_2008.pdf).
- [Flickr(2011)] Flickr. Flickr privacy, Desember 2011. URL <http://info.yahoo.com/privacy/us/yahoo/flickr/details.html>.
- [Fraser and Schewe(2009)] Bruce Fraser and Jeff Schewe. *Real World Camera Raw with Adobe Photoshop CS4*. Peachpit Press, 2009.
- [Gallagher(2011)] Sean Gallagher. Patriot act and privacy laws take a bite out of us cloud business. *Ars Technica*, Desember 2011. URL <http://arstechnica.com/tech-policy/news/2011/12/patriot-act-and-privacy-laws-take-a-bite-out-of-us-cloud-business.ars>.
- [Gaudin(2008)] Sharon Gaudin. Total recall: Storing every life memory in a surrogate brain. *Computerworld*, April 2008. URL [http://www.computerworld.com/s/article/9074439/Total\\_Recall\\_Storing\\_every\\_life\\_memory\\_in\\_a\\_surrogate\\_brain](http://www.computerworld.com/s/article/9074439/Total_Recall_Storing_every_life_memory_in_a_surrogate_brain).
- [Gemmell et al.(2002)Gemmell, Bell, Lueder, Drucker, and Wong] Jim Gemmell, Gordon Bell, Roger Lueder, Steven Drucker, and Curtis Wong. Mylifebits: Fulfilling the memex vision. 2002.

- [Gemmell et al.(2003)Gemmell, Lueder, and Bell] Jim Gemmell, Roger Lueder, and Gordon Bell. Living with a lifetime store. September 2003.
- [Gemmell et al.(2005)Gemmell, Aris, and Lueder] Jim Gemmell, Aleks Aris, and Roger Lueder. Telling stories with mylifebits. October 2005.
- [Gemmell et al.(2006)Gemmell, Bell, and Lueder] Jim Gemmell, Gordon Bell, and Roger Lueder. Mylifebits: A personal database for everything. February 2006.
- [Geus(2002)] Arie P De Geus. *Living Company, The: Habits for Survival in a Turbulent Business Environment*. Harvard Business School Press, May 2002.
- [Haynes et al.(1997)Haynes, Streatfield, Jowett, and Blake] David Haynes, David Streatfield, Tanya Jowett, and Monica Blake. Jisc/npo studies on the preservation of electronic materials : Responsibility for digital archiving and long term access to digital data, 1997. URL <http://www.ukoln.ac.uk/services/papers/bl/jisc-npo67/digital-preservation.html>.
- [Huitfeldts(2011)] Anniken Huitfeldts. Innlegg under kulturrådets årskonferanse, 16. november 2011. *Norsk kulturråds årskonferanse*, November 2011.
- [JISC(2010)] JISC. Lessons learned and advice for future digitisation projects. *JISC*, July 2010. URL <http://www.jisc.ac.uk/whatwedo/programmes/digitisation/evaluation/lessons.aspx>.
- [JISC(2008)] Digital Media JISC. Sustainability of digital collections, November 2008. URL <http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/sustainability-of-digital-collections>.
- [JISC(2009a)] Digital Media JISC. An introduction to digital preservation, January 2009a. URL <http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/an-introduction-to-digital-preservation>.
- [JISC(2009b)] Digital Media JISC. Establishing a digital preservation policy, January 2009b. URL <http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/establishing-a-digital-preservation-policy>.



- [JISC(2009c)] Digital Media JISC. Systems for managing digital media collections, September 2009c. URL <http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/systems-for-managing-digital-media-collections>.
- [John(2008)] Jeremy Leighton John. Adapting existing technologies for digitally archiving personal lives, digital forensics, ancestral computing, and evolutionary perspectives and tools. 2008.
- [John(2009)] Jeremy Leighton John. The future of saving our past. *Nature*, 459, June 2009.
- [John et al.(2010)John, Rowlands, Williams, and Dean] Jeremy Leighton John, Ian Rowlands, Peter Williams, and Katrina Dean. Digital lives, personal digital archives for the 21st century. March 2010.
- [Jones(1992)] John Chris Jones. *Design methods*. David Fulton Publisher, second edition, 1992. ISBN 0-471-28496-3.
- [Justis-og beredskapsdepartementet(2000)] JD Justis-og beredskapsdepartementet. Lov om behandling av personopplysninger (personopplysningsloven). Lovdata, April 2000. URL <http://www.lovdata.no/cgi-wift/wiftldles?doc=/usr/www/lovdata/all/nl-20000414-031.html&dep=alle&kort+,+titt=personopplysninger&>.
- [Kim(2010)] Sarah Kim. *Personal Digital Archives: Preservation of Documents, Preservation of Self*. Dissertation proposal, University of Texas at Austin, November 2010. URL <http://srhkim.com/files/PhD-proposal-final-sarahkim-dissemination.pdf>.
- [Krigsman(2008)] Michael Krigsman. Mediamax / the linkup: When the cloud fails. *ZD-Net*, August 2008. URL <http://www.zdnet.com/blog/projectfailures/mediamax-the-linkup-when-the-cloud-fails/999>.
- [Kulturdepartementet(1961)] Kulturdepartementet. Lov om opphavsrett til åndsverk m.v. (åndsverkloven). Lovdata, May 1961. URL <http://www.lovdata.no/all/nl-19610512-002.html>.
- [Kulturdepartementet(1990)] Kulturdepartementet. Lov om avleveringsplikt for allment tilgjengelege dokument (pliktavleveringslova). Lovdata, July 1990. URL <http://www.lovdata.no/all/hl-19890609-032.html>.

- [Kulturdepartementet(1992)] Kulturdepartementet. Lov om arkiv (arkivlova). Lovdata, Desember 1992. URL [http://lovdata.no/cgi-wift/wiftldles?doc=/app/gratis/www/docroot/all/nl-19921204-126.html&emne=arkivlov\\*&](http://lovdata.no/cgi-wift/wiftldles?doc=/app/gratis/www/docroot/all/nl-19921204-126.html&emne=arkivlov*&).
- [Kulturdepartementet(2009)] Kulturdepartementet. St.meld. nr. 24, (2008-2009), nasjonal strategi for digital bevaring og formidling av kulturarv. April 2009. URL <http://www.regjeringen.no/pages/2176228/PDFS/STM200820090024000DDDPDFS.pdf>.
- [Kulturdepartementet(2011)] Kulturdepartementet. Prop. 1 s, (2011–2012), proposisjon til stortinget (forslag til stortingsvedtak). October 2011. URL <http://www.regjeringen.no/pages/35168579/PDFS/PRP201120120001KUDDDDPDFS.pdf>.
- [Larsen and Lien(2007)] Peter Larsen and Sigrid Lien. *Norsk fotohistorie - Frå daguerreotypi til digitalisering*. Det Norske Samlaget, 2007. ISBN 078-82-521-6291-2.
- [Library of Congress(2007)] LOC Library of Congress. Sustainability of digital formats, planning for library of congress collections, sustainability factors. March 2007. URL <http://www.digitalpreservation.gov/formats/sustain/sustain.shtml>.
- [Library of Congress(2011)] LOC Library of Congress. Preserving our digital heritage: The national digital information infrastructure and preservation program 2010 report, a collaborative initiative of the library of congress. *Digital Preservation, National Digital Information Infrastructure and Preservation Program*, January 2011. URL [http://www.digitalpreservation.gov/library/resources/pubs/docs/NDIIPP2010Report\\_Post.pdf](http://www.digitalpreservation.gov/library/resources/pubs/docs/NDIIPP2010Report_Post.pdf).
- [Lindsay(2009)] Kate Lindsay. Jisc final report, the first world war poetry digital archive. May 2009. URL <http://www.oucs.ox.ac.uk/w1lit/documents/w1finalreport.pdf>.
- [Ludvigsen(2011)] Børre Ludvigsen. Mediarkiv. April 2011. URL <http://abdallah.hiof.no/mediarkiv/prosjektbeskrivelse/mediarkiv.pdf>.
- [Marshall(2007)] Catherine C. Marshall. How people manage information over a lifetime. *University of Washington Press*, 2007. URL <http://www.csd1.tamu.edu/%7Eemarshall/PIM%20Chapter-Marshall.pdf>.

- [Marshall(2008a)] Catherine C. Marshall. Rethinking personal digital archiving, part 1, four challenges from the field. *D-Lib Magazine*, 14(3/4), March/April 2008a. URL <http://www.dlib.org/dlib/march08/marshall/03marshall-pt1.html>.
- [Marshall(2008b)] Catherine C. Marshall. Rethinking personal digital archiving, part 2, implications for services, applications, and institutions. *D-Lib Magazine*, 14(3/4), March/April 2008b. URL <http://www.dlib.org/dlib/march08/marshall/03marshall-pt2.html>.
- [Marshall et al.(2007)Marshall, McCown, and Nelson] Catherine C. Marshall, Frank McCown, and Michael L. Nelson. Evaluating personal archiving strategies for internet-based information. *Proceedings of IS and T Archiving 2007*, May 2007. URL <http://arxiv.org/pdf/0704.3647v1>.
- [Miljøverndepartementet(1978)] Miljøverndepartementet. Lov om kulturminner (kulturminneloven). Lovdata, June 1978. URL <http://www.lovdata.no/all/nl-19780609-050.html>.
- [MinID(2012)] MinID. Minid - felles innlogging til offentlige tjenester, 2012. URL <http://minid.difi.no/minid/minid.php?lang=nb>.
- [Misund and Høiberg(2001)] Gunnar Misund and Jan Høiberg. Sustainable information technology for global sustainability. 2001. URL [http://www.ia.hiof.no/~gunnarmi/om/dig\\_earth\\_03/](http://www.ia.hiof.no/~gunnarmi/om/dig_earth_03/).
- [Nasjonalbiblioteket(2011)] NB Nasjonalbiblioteket. About the national library, 2011. URL <http://www.nb.no/english/facts>.
- [Nilsen(2011a)] Therese Roksvåg Nilsen. Interview with olaf skjelderup, March 2011a.
- [Nilsen(2011b)] Therese Roksvåg Nilsen. Minutes from meeting with nb, June 2011b.
- [Norsk Kulturråd(2012)] NK Norsk Kulturråd. Om oss, 2012. URL <http://www.norskkulturrad.no/om-kulturradet/>.
- [Østby and Ertesvåg(2007)] Jon Birger Østby and Randi Ertesvåg. Til kildene! *ABM-skrift*, 40, 2007. URL <http://www.abm-utvikling.no/publisert/abm-skrift/abm-skrift-fulltekst/abm-skrift-40-til-kildene.html>.

- [Pedersén(2007)] Olof Pedersén. Archives and libraries: the ancient near east, the classical world, and the old testament. *ERB*, March 2007. URL <http://www.anst.uu.se/olofpede/ArchivesLibraries070308.pdf>.
- [Pepys(2003)] Samuel Pepys. *The diary of Samuel Pepys*. Modern Library, 2003. ISBN 978-0-8129-7071-5.
- [Photoblog(2012a)] Photoblog. Photoblog - your life in photos, 2012a. URL <http://www.photoblog.com/>.
- [Photoblog(2012b)] Photoblog. About photoblog, 2012b. URL <http://www.photoblog.com/about/>.
- [Photobucket Corporation(2012a)] Photobucket Photobucket Corporation. About photobucket, 2012a. URL <http://photobucket.com/about>.
- [Photobucket Corporation(2012b)] Photobucket Photobucket Corporation. About photobucket, 2012b. URL <http://photobucket.com/learnmore>.
- [Ram and Mishra(2008)] Mange Ram and J K Mishra. Digital collections: Preservation and problems. 2008. URL <http://ir.inflibnet.ac.in/dxml/bitstream/handle/1944/1274/34.pdf?sequence=1>.
- [Ramana(2004)] Y V Ramana. Digital archiving : Indian scenario. November 2004. URL [http://ir.inflibnet.ac.in/dxml/bitstream/handle/1944/417/04Planner\\_3.pdf?sequence=1](http://ir.inflibnet.ac.in/dxml/bitstream/handle/1944/417/04Planner_3.pdf?sequence=1).
- [RLG/OCLC(2002)] RLG/OCLC Working Group on Digital Archive Attributes RLG/OCLC. Trusted digital repositories: Attributes and responsibilities. *An RLG-OCLC Report*, May 2002. URL <http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf>.
- [Rønnevig and Bendiksen(2010)] Cecilie Rønnevig and Aslaug Bendiksen. Revidert konsesjon - innsamling av norske sider på internett. August 2010.
- [Skatteetaten(2012)] Skatteetaten. Fødselsnummer, 2012. URL <http://www.skatteetaten.no/no/Alt-om/Folkeregistrering/Fodselsnummer/>.
- [Stokke(2010)] Regine Stokke. *Regines bok, en ung jentes siste ord*. Gyldendal, 2010. ISBN 978-82-05-40187-7.

- [Stowe(2002)] Steven Stowe. Making sense of letters and diaries. *Making Sense of Evidence, History Matters: The U.S. Survey on the Web*, July 2002. URL <http://historymatters.gmu.edu/mse/letters/letters.pdf>.
- [The British Library(2006)] BL The British Library. British library digital preservation strategy. 2006. URL <http://www.bl.uk/aboutus/stratpolprog/ccare/introduction/digital/digpresstrat.pdf>.
- [The British Library(2012)] BL The British Library. Facts and figures, 2012. URL <http://www.bl.uk/aboutus/quickinfo/facts/index.html>.
- [The Domesday Book Online(2011)] TDBO The Domesday Book Online. The domesday book online, 2011. URL <http://www.domesdaybook.co.uk/index.html>.
- [The Franchise Openness Project(2002)] OpenFran The Franchise Openness Project. Archival integrity, Desember 2002. URL <http://www.openfran.org/about/integrity>.
- [The Internet Archive(2012)] archive.org The Internet Archive. About the internet archive, 2012. URL <http://www.archive.org/about/about.php>.
- [The University of Manchester(2008a)] Manchester The University of Manchester. Defamation. *Paradigm*, January 2008a. URL <http://www.paradigm.ac.uk/workbook/legal-issues/defamation-digitalarchives.html>.
- [The University of Manchester(2008b)] Manchester The University of Manchester. Policy and legislation aimed at enhancing access to information. *Paradigm*, January 2008b. URL <http://www.paradigm.ac.uk/workbook/legal-issues/access.html>.
- [The University of Manchester(2008c)] Manchester The University of Manchester. Paradigm, welcome to paradigm, October 2008c. URL <http://www.paradigm.ac.uk/>.
- [Thomas(2007)] Susan Thomas. Paradigm, a practical approach to the preservation of personal digital archives. *Bodleian Library*, March 2007. URL <http://www.paradigm.ac.uk/projectdocs/jiscreports/ParadigmFinalReportv1.pdf>.
- [Tomalin(2003)] Claire Tomalin. *Samuel Pepys, The Unequalled Self*. Penguin Books, 2003. ISBN 978-0-14-028234-4.

- [Torgnesskar(2007)] Per Olav Torgnesskar. Standard for fotokatalogisering. July 2007. URL <http://www.abm-utvikling.no/horinger/standard-for-fotokatalogisering>.
- [Tristram(2002)] Claire Tristram. Data extinction. *Technology review, MIT*, October 2002. URL <http://www.technologyreview.com/communications/12975/page1/>.
- [UNINETT(2011)] UNINETT. This is the uninett group, 2011. URL <http://www.uninett.no/english>.
- [Vaaler and Ludvigsen(2008)] Audun Vaaler and Børre Ludvigsen. Åpne standarder for grafikk, bilde, lyd og video på offentlige nettsider: Dokumentering og evaluering. September 2008. URL [http://media.hiof.no/diverse/fad/rapport\\_2.pdf](http://media.hiof.no/diverse/fad/rapport_2.pdf).
- [van Dijck(2007)] José van Dijck. *Mediated memories in the digital age*. Stanford University Press, 2007. ISBN 978-0-8047-5623-5.
- [Waters(2002)] Donald Waters. Good archives make good scholars: Reflections on recent steps toward the archiving of digital information. *The State of Digital Preservation: An International Perspective*, July 2002. URL <http://www.clir.org/pubs/reports/pub107/waters.html>.
- [Waxer(2009)] Cindy Waxer. Is the cloud reliable enough for your business? *Computerworld*, June 2009. URL [http://www.computerworld.com/s/article/339152/Can\\_You\\_Trust\\_the\\_Cloud\\_](http://www.computerworld.com/s/article/339152/Can_You_Trust_the_Cloud_).
- [Wilkinson(2007)] Alec Wilkinson. Dept. of technology: Remember this? *The New Yorker*, May 2007. URL [http://www.newyorker.com/reporting/2007/05/28/070528fa\\_fact\\_wilkinson](http://www.newyorker.com/reporting/2007/05/28/070528fa_fact_wilkinson).
- [Yahoo(2012)] Yahoo. Flickr, 2012. URL <http://www.flickr.com/>.

# Appendix A

## Design methods

### A.1 Systematic search

**Aim** To solve design problems with logical certainty.

**Outline**

1. Identify the constituents of the problem:
  - a The variables that can be controlled by the designer.
  - b The variables that cannot be controlled by the designer.
  - c The variables that are to be controlled by the designer.
2. Identify relationships between the variables.
3. Predict values that the context variables are likely to assume.
4. Identify constraints, or boundary conditions, i.e. limiting values to any of the variables.
5. Adjust the values of each decision variable and calculate the values of the dependent variables.
6. Select values for the decision variables to achieve the greatest combined value of the weighted objectives, or at least an acceptable value for each.

[Jones(1992)]

## A.2 Literature searching

**Aim** To find published information that can favourably influence the designers' output and that can be obtained without unacceptable cost and delay.

### Outline

1. Identify the purpose for which published information is being sought.
2. Identify the kinds of publication that are likely to contain information that can be relied upon for such purposes.
3. Select the most relevant of the standard methods of beginning a literature search.
4. Minimise the search cost by allowing for retrieval delays and by continuously evaluating both the choice of sources and the applicability of data collected.
5. Keep accurate and complete references to documents that are found to be usable.
6. Keep local collections of publications sufficiently small and temporary to permit rapid retrieval.

[Jones(1992)]

## A.3 Classification of design information

**Aim** To split a design problem into manageable parts.

### Outline

1. Record, on a separate card, each item of information collected while exploring the design.
2. Sort the cards into alternative sets of categories until a set is found that appears to fit both the recorded data and one's own view of the problem.
3. Use the selected set of categories as a basis for indexing information collected later, for splitting up the problem for serial or parallel working, and perhaps as a tentative identification of variables and of relationships between them.



4. Revise the classification at a later stage if contradictory evidence accumulated, if the objectives change or if one's view of the problem changes.

[Jones(1992)]

## **A.4 Specification writing**

**Aim** To describe an acceptable outcome for design that has yet to be done.

### **Outline**

1. Provisionally identify a range of possible outcomes, of varying levels of generality.
2. Select the lowest level of generality that leaves the designers with sufficient design freedom.
3. Define the expected design outcome without reference to the design features that the designers are free to change and with reference to measures of performance that the designers will be able to predict.

[Jones(1992)]

## **Appendix B**

# **Interview, Olaf Schjelderup at UNINETT**

UNINETT is a group consisting of UNINETT FAS, UNINETT Norid and UNINETT Sigma. Their function is to operate network solutions for university collages, universities and research institutions in Norway. They also handle other national ICT tasks. [UNINETT(2011)]

Here are the highlights from meeting with Olaf Schjelderup, the listing is of a range of issues to consider when forming a national archive;

### **Participants**

- NB
  - Cultural information
- Riksarkivet
- Universities
  - Historical institutes
  - Humanistic
  - Philosophy of history
  - Stakeholders

- Government
  - Financing
  - Organization
- Datatilsynet
  - Privacy
  - Rights
- Law
  - Mechanisms for accessibility
- UNINETT
  - Academic interests
- Historians and social anthropologists
  - Digital archaeologists

### **Financing**

- Financing for:
  - Start-up
  - Power
  - Disks
  - Maintenance
  - Management
  - Labor
- Governmental
- Annual fees
- Profit motive
- Employment

- Third party stakeholders

### **Structure**

- Usability
- Identity
- Authorization
- Authenticity
  - Digital stamps
  - Metadata revision control
    - \* Correction possibility
    - \* Revision history
- High level variability of metadata
- Control of legitimacy
- Dynamic application technology in the background
  - User friendly uploading environments.

### **Distribution and management**

- Hardware
  - Mechanical wear, disk replacement after approximately 8 years
  - Importance of proper temperature
  - Importance of regular use
  - SSD<sup>1</sup>
  - Bit rate
- Content management
- Distribution of geographic duplication

---

<sup>1</sup>Solid state drive, [http://en.wikipedia.org/wiki/Solid-state\\_drive](http://en.wikipedia.org/wiki/Solid-state_drive)

- Ensure availability
  - Synchronization of geographically management
- Format updates
  - Copies must be frozen before anything else
  - How long should frozen copies and technology to open it be kept?
- Multi-dimensional management
  - Distribution
  - Frozen data
  - Technological diversity

### **Security**

- Securing the data from:
  - Natural disasters
  - Theft
  - Vandalism
- Diversity and distribution
- Peer-to-peer

### **Ownership**

- After death
  - Descendants
  - Foundations
  - 70 years

[Nilsen(2011a)]