



KANSALLISARKISTO

Planning project for mass digitisation

Vipuvoimaa
EU:lta
2014–2020



Euroopan unioni
Euroopan aluekehitysrahasto

Planning project for mass digitisation

Final report



Table of contents

Summary of project results.....	4
1 Destruction after digitisation as an archival theoretical question	6
1.1 National starting points.....	6
1.2 Destruction after digitisation and the nature of records.....	7
1.3 Destruction after digitisation as screening	9
1.4 Assessing the intrinsic value of analogue records	10
2 Introduction	12
2.1 Background and goals of the project	12
2.2 Project limitations	13
2.3 Key concepts.....	14
3 International status of mass digitisation	15
4 Digitised material	18
4.1 Volume of analogue records s in the public sector	18
4.2 Digitised material	20
4.3 Sequence and schedule of digitisation.....	21
5 Mass digitisation process	23
5.1 Preconditions of the process.....	25
5.2 Preparations by government agencies.....	25
5.3 Logistics and transfer	26
5.4 Digitisation preparations.....	27
5.5 Digitisation.....	28
5.6 Transfer of digital objects.....	30
5.7 Processing analogue material after digitisation.....	31
6 Production of mass digitisation	31
7 Costs and advantages	36
7.1 Digitisation costs	36
7.2 Advantages of digitisation	37
7.3 Analysis of costs and advantages	38
8 Risks associated with mass digitisation	41



KANSALLISARKISTO

Planning project for mass digitisation

Vipuvoimaa
EU:lta
2014–2020



Euroopan unioni
Euroopan aluekehitysrahasto

9	Dependencies of mass digitisation and further activities.....	42
9.1	Personnel and skills	43
9.2	Data systems	45
9.3	Proposed further activities	46



Summary of project results

The aim of mass digitisation is to convert paper material in the field of administration into digital format and, therefore, improve its availability and usability. The digital storage of material and the improved usability of material serve to improve the efficiency of record management and archiving in the public sector and to produce cost savings.

The planning project for mass digitisation defines a new disposal-oriented digitisation process in which the analogue manifestation can be disposed of after its digitisation. The data content of the record will not be disposed; only the storage format will change. In the definition phase, a key requirement was the statutory task of the National Archives of Finland to ensure the preservation of official records belonging to the national cultural heritage and the information contained therein.

An extensive analogue archival material survey was conducted at the beginning of the planning project. As a result, the National Archives of Finland obtained a much more detailed overview of the volume of public sector material to be archived and the properties that affect digitisation. In the survey, the volume of material increased to nearly 400,000 shelf metres, being significantly higher than estimated previously. This volume is nearly twice the current volume of material held by the National Archives of Finland. In the project, the volume of material to be digitised was estimated to be 212,000 shelf metres. The majority of material left outside the scope of digitisation consists of material retained for a fixed period. The volume of the material to be digitised will be specified during the further preparation of mass digitisation.

The material to be digitised is located in different agencies and their archives. When planning the process, it was necessary to consider how the material can be controllably transferred from the archives of different agencies for centralised digitisation. In the digitisation process, government



agencies require much fewer actions to prepare and transfer their material than in the current transfer procedure of the National Archives of Finland. The starting point of the new operating method is to centrally prepare material. This improves the efficiency of the process.

Digitisation costs mainly consist of the manual labour required for the preparation and scanning of material. As a result, the properties of material have a significant impact on the workload and, therefore, on costs. A key challenge in digitisation is to preserve the integrity and authenticity of records, i.e. to carry out a high-quality digitisation process so that an analogue manifestation is converted into digital format while preserving its information content. This will be ensured by producing a high-quality digital image, the quality of which is defined on the basis of international quality requirements used in the archiving sector.

The analysis of the advantages of mass digitisation is based on the comparison of the current analogue storage scenario with the digital storage scenario. In the current state, government agencies slowly transfer analogue records dating back forty years or more to the National Archive of Finland, meaning that agencies will continue to maintain paper archives and related costs for a long time. In the digital scenario, mass digitisation significantly speeds up the transition of government agencies to fully digital data management. As a result, they can stop maintaining paper archives much earlier than originally expected. When the mass digitisation process ends in 2030, annual savings in the digital storage scenario compared with the analogue storage scenario are estimated to be fourteen million euros.

The mass digitisation process will take more than ten years, and it will require effective coordination and, above all, strong commitment. Government agencies must commit to mutually defined digitisation schedules and instructions set for the preparation of material. The National Archives of Finland must commit to leading the digitisation process, and the ministries that steer the process must commit to mutually set goals and operational continuity.



1 Destruction after digitisation as an archival theoretical question

Introduction to the English version

1.1 National starting points

The strategic goal of the National Archives of Finland is to promote the use of digital material and long-term preservation in digital format and medium. The goals of the National Archives support the digitalisation goals of public services in Finland to obtain better services for citizens and improve the efficiency and profitability of the public administration. A Letter of Intent from the Finnish Government defines that analogue records appraised for permanent retention¹ and in the custody of government agencies are to be converted into digital format and medium, and that the National Archives will only accept digital records for custody. The public administration is preparing a mass digitisation project to convert 212 shelf kilometres of analogue material into digital format and medium over the next few decades. The large-scale digitisation of public records has also raised the question of how much the two-fold preservation of large volumes of records will cost. The amendment to the Archives Act entered into force in 2017 authorises the National Archives to decide on the destruction of the original² copies of records appraised for permanent retention, and converted into digital format and medium, provided that the records as material objects are not considered to have a value in terms of cultural history and that their digitisation does not put the evidential value of the records at risk.

The National Archives has defined that destruction following digitisation only applies to records that have been produced after 1922 and meet the aforementioned conditions. It is estimated that, initially, 80 to 90 per cent of records younger than forty years in custody of government

¹ By permanent retention I refer to records appraised as having enduring value.

² The concept "original" is used in the Archives Act.



agencies are within the scope of destruction. Of the collections of the National Archives, some 70 per cent are younger than the limit set for destruction. However, it must be assumed that for these collections a higher percentage of records will be assessed as valuable in terms of cultural history. Finland is not the only country in which the legislation permits the destruction of analogue records after digitisation but a pioneer in the execution of destruction after digitisation. On a broader international scope, the professional dialogue has mainly addressed destruction after digitisation in a critical way.

1.2 Destruction after digitisation and the nature of records

In the international dialogue, destruction following digitisation is linked to the understanding of the nature of records and archival material. A key argument seems to be the material nature of analogue records, to which its cultural and research-related values are linked. An analogue record is an indivisible sum of its information content and preservation format and medium, and its evidential value can be bound expressly to its indivisibility. Being a material object, a record has also been part of the activities which it documents or in which its authors have been involved. In this material context, a record acquires its value as a source of the past. A digital manifestation, produced from an analogue record, does not have this same material quality, or its material integrity may decrease during digitisation. As a result, it is seen that a digital object cannot replace an analogue record. (Digitalisering for Kassarjon – en utredning – og en utfordring 2015). This traditional understanding of the material nature of records is evident in the execution of destruction after digitisation of analogue records, appraised as having enduring value, in that the national and international starting point is destruction under the condition that analogue records are not valuable in terms of cultural history and that digitisation does not put the value of records as evidence at risk.

However, US archivist Paul Conway (2014) has stated that the long-term preservation as archives of large digital collections produced by mass digitisation requires a new way of archival theoretical thinking. In general, technologies enabled by each era have always had a significant



role in the definition and development of archival theoretical principles. According to Canadian archivist Terry Cook (1997), the material nature of records and archives has had an almost sacred and unviolated status in the archival theory during the era of analogue technologies. As born-digital records have become voluminous, it was, however, understood that the relevance of traditional theoretical principles in ensuring the record-like quality of digital data demands that records and archives are regarded as conceptual entities (Cook 1997). The nature of records was no longer defined through material aspects, but conceptualised as characteristics of integrity, authenticity, reliability and availability of data and ensuring the evidential value of records (ISO 15489 Records management 2007).

Currently, analogue records produced in earlier times exist in a digital world under digital rules, and technological development has enabled the conversion of these records into the digital format and medium. Paul Conway (2014) regards the conversion of analogue records into digital format and medium as a transformation process. As a result, the metamorphosis taking place during the digitisation process does not produce copies of analogue material, but surrogates, new digital records that replace analogue records under their own rights. According to Conway, converting analogue records into digital format and medium and documenting the process; ensuring the value of digital surrogates as evidence, as well as the other management activities throughout the existence of the surrogates; and, in particular, the use of digital surrogates in place of analogue records, provides these manifestations with a secondary provenience and a performative material quality. Conway's theory is based on Geoffrey Yeo's theory of records as permanent representations of events, the material manifestations of which may differ during their existence. Therefore, digital manifestations regarded as surrogates represent the material aspects and life history of analogue records. To extend these, they obtain cultural and research-based material value independent of analogue records through their management and use.



1.3 Destruction after digitisation as screening³

On a national level, destruction after digitisation is associated with screening in Finland. Destruction after digitisation has a clear analogy to the requirement for the destruction of records following the huge increase in analogue records in the early 20th century. The trigger at the time was the practical need to save costs, and the development of appraisal and disposal theories and the rise of these activities into a key position in the recordkeeping professions were merely consequences (Dollar 2011). The new way of archival theoretical thinking required by mass digitisation also includes a requirement for updating appraisal and disposal principles as demanded by technological development. With regard to this update, the professional understanding of the nature of records and archival material in the digital world plays a key part. The Finnish Archives Act authorises the National Archives to decide on records having enduring value and the preservation format and medium of these records. Appraisal is based on evidential and informational value and outlines records to be included in the national cultural heritage. In public sector, appraisal is a proactive process in relation to records creation. The objects of appraisal are not the existing material records, but the record system and its internal and external relationships as well as the unique nature and social significance of task-specific record types emerging from different activities of the records creator. (Screening policy of the National Archives of Finland 2012). Actually, this means that the appraisal of records in the Finnish public administration is focused on records as a conceptual entity, not on any specific material manifestations of these conceptual record types.

The decision of the preservation format and medium of records with enduring value comprises part of screening. The National Archives obtained its authorisation under the Archives Act of 1981 in order to ensure the long-term preservation of born-digital records. Initially, screening

³ I use term *screening* for the Finnish national proactive practice of professional activities named in English as appraisal and disposal.



decisions defined that born-digital records be transferred to the National Archives in analogue format and medium. The current starting point of screening at the National Archives is the preservation of public records in digital format and medium, enabling a more extensive appraisal for permanent retention. As a result of this, previous decisions on the analogue preservation format and medium have been substituted for decisions on preservation in digital format and medium. The destruction of the digital manifestation was implicitly accepted in conjunction with the issuance of decisions on converting born-digital records into analogue format and medium. Instead, the destruction of born-analogue records after digitisation is conditional and requires an explicit decision.

1.4 Assessing the intrinsic value of analogue records

The destruction of analogue records after digitisation requires that their value as cultural historical artefacts and their evidential value embedded in the preservation format and medium are assessed. In international professional literature, these values are referred to as the *intrinsic value*. Assessing intrinsic value of records has played a highly marginal part in archival theory, but its significance has increased with the growth of the volume of records to be preserved, and as a result of the development of preservation technologies (The Intrinsic Value of Archive and Library Material 1996). Intrinsic value has been defined differently in different sources, and in this text it is defined as the identification of records appraised for permanent retention that need to be preserved in their former format and medium after conversion into another preservation format and medium, because not all properties defining their value as cultural historical artefacts and evidential value as records can be recovered in the conversion process (The Intrinsic Value of Archive and Library Material 1996). Defined in this way, the intrinsic value covers all conversions, not exclusively the conversion of analogue records into digital format and medium.

The assessing of the intrinsic value has been considered to be a part of the appraisal activity. However, this approach has also been criticised. US archivist Shauna McRanor (1996) states that the assessing of the intrinsic value of records does not comprise part of appraisal as it does not



assess the enduring value of records as logical information entities. In addition, the report of the Archives School Marburg (The Intrinsic Value of Archive and Library Material 1996) states, that the assessing of the intrinsic value is retrospective in relation to the appraisal of the enduring value. In the appraisal process, it is decided *what* needs to be retained permanently, while the assessing of the intrinsic value determines *how* entities appraised for permanent retention are to be preserved. Accordingly, destruction after digitisation at the National Archives of Finland is associated with screening, i.e. deciding on the preservation format and medium of records with enduring value. The new preservation format and medium of records is digital, and the previous analogue manifestations are destroyed, unless they are assessed at having intrinsic value.

In the National Archives of Finland, the assessing of the intrinsic value is based on a formal criteria defined on the basis of international precedents. Furthermore, records to be also preserved in their current analogue format and medium are assessed on the basis of formal criteria, into two categories. The first category comprises analogue records assessed at having special cultural historical value. The material integrity of these records will be preserved meticulously. The second category comprises records that are to be preserved in analogue format for juridical reasons or because they may lose their evidential value in the digitisation process. However, the material integrity of these analogue records may be affected, for example, by means of unbinding in the digitisation processes.

Bibliography

Conway, Paul (2014). Digital transform and mediumions and the archival nature of surrogates. *Archival Science* vol. 15, no. 1.

Cook, Terry (1997). The Impact of David Bearman on Modern Archival Thinking: An Essay of Personal Reflection and Critique. *Archives and Museum Informat and mediumics* 11: 15–37, 1997. Kluwer Academic Publishers.

Digitalisering for Kassarjon – en utredning – og en utfordring (2015). (Digitalisation for disposal). Survey of Riksarkivet, Norway.



Dollar, Charles M. (2011). Solon. J. Buck: Historian, Archivist and Administrator. American Archivist Online Supplement to Volume 74

The Intrinsic Value of Archive and Library Material (1996). Final report of the Archives School Marburg.

Screening policy of the National Archives of Finland (2012).

McRanor, Shauna (1996). A Critical Analysis of Intrinsic Value. The American Archivist vol. 59/4.

ISO 15489 records management (2007).

2 Introduction

2.1 Background and goals of the project

The framework of mass digitisation has been defined in the Government's decision in principle of 21 June 2017, according to which the analogue archival material held by government agencies is to be digitised by 2030. According to this decision, the records will only be archived in digital format, meaning that the majority of analogue archival material will be disposed of after their digitisation.

Here, "mass digitisation" refers to the large-scale digitisation of material as an organised and industrial process. Mass digitisation differs from other digitisation methods, particularly in that the digitisation process is not customised according to the material type or organisation, which is typical in digitisation carried out in the form of a project.

The conditions set in the decision in principle have defined the planning project for mass digitisation. The goal of the project was to prepare an implementation plan for the digitisation process, starting from the transfer of material from government agencies for digitisation and ending at the conversion of material into digital format and the disposal of the analogue object.



The following were defined as project tasks:

- Producing a roadmap as a plan for digitising paper records in the public sector, presenting the sequence and schedule of the digitisation process
- Identifying the volumes and properties of material held by government organisations and transferred from government agencies and municipalities to regional governments
- Planning the mass digitisation process and defining the technologies to be used in mass digitisation
- Preparing calculations of the costs and benefits of mass digitisation
- Defining logistics, i.e. how material can be transferred from the archives of government agencies for centralised digitising
- Identifying the international status of mass digitisation and the best practices of digitisation

The project was carried out between 2 June and 31 December 2017, being partly funded by the European Regional Development Fund (ERDF). The main implementing party was the National Archives of Finland, and its other implementing party was South-Eastern Finland University of Applied Sciences (XAMK), which was in charge of identifying logistics. The City of Mikkeli also took part in project funding and the steering group.

The goal of this report is to provide an overview of the implementation of mass digitisation. Appendices to this report describe different studies and plans in more detail. The goal of the appendices is to provide a detailed description of specific parts of mass digitisation.

2.2 Project limitations

The project was defined as a technical identification project that does not aim to answer any administrative or legislative questions related to mass digitisation. The project was carried out with the following limitations:



- The project does not address different administrative implementation methods of mass digitisation (outsourcing, official work, etc.).
- The project does not address any transfer of ownership as a result of the conversion into digital format.
- The project does not address the future storage solution for digital material and its development.
- The project does not produce descriptions related to material of municipalities that will not transfer to the ownership of regional governments.
- The project does not address the funding method and options of the future mass digitisation process.

Because of the tight project schedule, the project was limited to the process of converting material into digital format, in which case the storage and use of digital material were not covered in the project. An architectural project for an integrated? digital archiving service system (SAPA) was carried out alongside this project. Its goal was to define the solution architecture of the service. It was led by the Ministry of Education and Culture, and its members were the Ministry of Finance, the National Archives of Finland and CSC – IT Center for Science.

One of the goals of the planning project for mass digitisation was to identify what material related to tasks transferred from municipalities to regional governments will be digitised. During the project, the overall schedule of the regional government reform was postponed by a year. As a result, no statements were received from parties responsible for the regional government reform regarding the digitisation of material owned by regional governments. Therefore, the digitisation of material transferred to regional governments will be defined after this project.

2.3 Key concepts

- **Material:** A records or a group of records s. Not an unambiguous group in terms of an archival description.



- **Digitisation:** A process in which analogue information is converted into digital format.
- **Mass digitisation:** Large-scale digitisation of material as an organised and industrial process.
- **Scanning:** An activity with which analogue information is converted into digital format by using technology and software solutions intended for this purpose.
- **Preservation decision:** By means of a preservation decision, the National archives of Finland decides on what public sector data contained by records must be stored permanently, what their storage format is and what analogue archival material is disposed or kept after digitisation. This refers particularly to a decision to determine whether the information content of the record is also preserved in analogue format after its digitisation.
- **Roadmap:** A plan for digitising material, guiding the sequence and schedule for digitising analogue material held by government agencies, and a description of the material types and volumes to be digitised regarding each administrative field.

3 International status of mass digitisation

In October–December 2017, the National Archives of Finland conducted a survey to identify the status of mass digitisation in the national archives sector. Its key finding is that mass digitisation as referred to in this report has not been carried out anywhere in the world in the archival context.

As the analogue manifestations has not been disposed after digitisation, international references of good practices are not available.

The survey was conducted in the form of an online survey and mutual discussions with the national archives of Sweden, Norway, the Netherlands and the USA. In addition, digitisation quality specifications, digitisation instructions and contact details of digitisation contact people were collected from different countries. The survey was responded to by representatives of forty



archives, libraries and companies from all Nordic countries, the USA, Australia, the Netherlands, Belgium and Austria.

On the basis of the survey, the current annual digitisation volume of the National Archives of Finland is among the highest in the world, following the USA, at least in the light of figures in 2017. In the USA, roughly thirty million files are digitised every year, while the National Archives of Finland digitised eighteen million files in 2017. Other parties that digitise significant volumes of material (more than ten million images per year) are the national archives of Sweden, Denmark and Australia. Digitisation volumes are not definitive, and there can be notable variation from one year to the next, depending on projects.

What is important about scanning quality criteria and the storage format is the aim to retain the information content of the original record without any losses. As a result, more than 80 percent of respondents to the survey stored digitised files in uncompressed TIFF format. A significant majority scanned their files as 24-bit colour images with a resolution of 300 dpi. In some cases, 16- or 8-bit grey tone digitisation was used. The national archives of the Netherlands and the USA also saved JPEG files in addition to TIFF files. Neither has disposed their analogue archival material. Of all national archives, only the UK used a storage format other than uncompressed TIFF files. They have started to use the JPEG2000 format, as it enables lossless compression and, therefore, a smaller file size than the TIFF format. However, this format is problematic because it is not fully patent-free, it has no browser support and the software required is expensive.

Only four of all archives that responded to the survey have prepared instructions for disposing analogue storage objects after digitisation. These mainly concerned material hazardous to health or explosive material. The National Archives of Australia was the only respondent that prepared instructions for digitisation that enables the disposal of analogue material. New Zealand did not respond to the survey, but it published destruction instructions in May 2017.



In addition to Finland, three countries (Belgium, Norway and Portugal) are considering the preparation of disposal instructions, as their current national legislation permits the disposal of material after digitisation. Of these countries, Belgium and Norway, however, have a critical attitude towards any disposing. The national archives of the Netherlands, Iceland, Sweden and the Czech Republic, the EU Historical Archives and the Lüneburg City Archive emphasised in their responses that they are not ready to dispose any analogue archival? cultural heritage existing in the form of records s. The countries that responded to the survey have not yet started or planned the digitisation of very young material, and digitisation has focused on improving the usability of material already existing in their archives, the most frequently used material or the most valuable material in terms of cultural history.

The closest points of comparison to the mass digitisation project in the near future will be at least the Netherlands, Sweden, the USA, and the digitisation of medical records of *Helse arkivet* in Norway, in which case the plan is to dispose analogue originals after they have been digitised. The national archives of the Netherlands are planning to digitise some 130,000 shelf metres of its most frequently used material during the next decade. The national archives of Sweden aim to digitise 10 percent of its material in less than ten years. The USA is launching a project to digitise a large volume of material in a decentralised manner.

In conjunction with the digitisation survey, the plans prepared for the mass digitisation project were briefly described to the respondents. The survey was sent to persons responsible for digitisation at different national archives via email, the ICARUS network and social media channels. As a result, information about the planned digitisation project has spread widely throughout the archives sector. The survey was opened nearly nine hundred times. Its results will be sent in English to all respondents. Information about the planning project for mass digitisation, the volume of material to be digitised, the rough project schedule and the disposal of analogue material after digitisation has spread to all EU states.



The results of the international survey have been described in more detail in Appendix 2.

4 Digitised material

4.1 Volume of analogue records in the public sector

The volumes of analogue record material in the public sector and properties affecting the digitisation of this material were identified by means of a two-stage material identification survey. The surveys were extensive and provided a more detailed view of the state of archiving in the public sector and the volume of analogue records. The response rate to the first stage of the survey was 92 percent. At the second stage, responses were received from all recipients of the survey.

On the basis of the results of the first stage of the material identification survey, government organisations have approximately 387,000 shelf metres of analogue record material, of which roughly 171,500 shelf metres are stored permanently and 209,600 shelf metres are stored for a fixed period.⁴ This volume is much higher than any previous estimates. A majority of analogue material consists of paper records.

The second stage of the material survey was only carried out for some government organisations. The survey was mainly directed at organisations with the highest material volumes. Respondents to the survey conducted at the second stage represented roughly 80 percent of all respondents to the survey conducted at the first stage in terms of their material

⁴ A majority of analogue material held by government agencies is defined for permanent or fixed-term storage with previous decisions of the National Archives of Finland. In addition, a majority of permanently stored material has already been screened and separated from material stored for a fixed period. Since the 1980s, the goal of the National Archives of Finland has been to assign approximately 10–15 percent of an annual increase in paper records held by government agencies to permanent storage, also taking cost factors into account. However, costs cannot primarily determine what records are stored permanently. Even after screening, it must be possible to identify the functional, chronological, local and theme-related causality of records.



volumes. At the second stage, the aim was to identify properties that affect the digitisation of material: physical properties, quality and condition of material and the level of description and organisation. Responses were given at an organisational level⁵. As a result, more detailed information about material properties was obtained.

The total volume of material included in the results of the second stage of the material survey was 266,000 shelf metres, of which roughly 124,000 shelf metres of material is stored permanently and 140,000 shelf metres of material is stored for a fixed period. Of different material types, paper records (including index files) account for 86 percent, bound material makes up approximately 7 percent and special material (such as maps, drawings, microfilms and photos) and undefined responses represent 7 percent. On the basis of the results, some 9 percent of all material has not been organised and the value of 4.3 percent of all material has not been defined. Significant information considering digitisation planning was obtained about the volume of loose paper records (50,700 shelf metres), joined paper records (165,000 shelf metres), special records (12,600 shelf metres) and hand-written material (83,000 shelf metres).

The survey conducted at the first stage also included municipal organisations that are within the scope of the regional government reform. With regard to these organisations, the survey also identified the volume of material to be transferred to new regions established as a result of the reform in order to secure operational continuity and carry out their tasks. On the basis of the results, approximately 22,000 shelf metres of material will transfer to new regions. Because it has not been clearly defined what material will transfer to new regions, it was difficult for municipal organisations to respond to the survey. This has an impact on the reliability of the results, which should be taken into account when processing the results of the first stage of the survey.

⁵ Here, "archive" refers to a group of records resulting from the activities of the party generating the archive (not an archive room or building).



The results of the material identification surveys have been described in more detail in Appendix 1.1.

4.2 Digitised material

The volume of material to be digitised, material types and the digitisation schedule are presented in the roadmap presented on the basis of the material identification surveys. The roadmap is a planning tool, and it will be updated and maintained throughout the planning and implementation of mass digitisation. It is estimated that the volume of material held by government agencies and included in mass digitisation and the roadmap totals 212,000 shelf metres. The material suitable for mass digitisation has been divided into three main categories on the basis of material types: Paper material in A4/folio size (72 percent), index file material (14 percent) and bound material (14 percent).

As a result of the material identification surveys, it was discovered that some of the analogue material held by government agencies is not suitable for the mass digitisation process that requires that material can be digitised quickly and easily. The following material has been excluded from the scope of mass digitisation (212,000 shelf metres):

- Special material (such as maps, drawings, microfilms and microcards, as well as AV material) as separate collections. Special material, such as maps and drawings, can be included in the mass digitisation process when it is among other material (e.g. in a single storage unit or record).
- Material not suitable for mass digitisation due to their physical properties (e.g. special papers, particularly fragile material and mouldy material).
- Material in analogue and digital formats corresponding with each other in terms of their data content.
- A significant volume of records to be stored for a fixed period.
- A majority of security-classified material and material of protection levels ST I and ST II.



- Material of specific organisations: An organisation is not obligated, under the Archives Act, to transfer material to the National Archives of Finland, an organisation has a special exemption from transferring material to the National Archives of Finland, an organisation only has a very low volume of permanently stored material or an organisation's material comprises special material that is slow to digitise or is otherwise unsuitable for digitisation in terms of its digital properties.

The material volume of 212,000 shelf meters planned for mass digitisation is a highly significant volume considering the international archives sector. This volume is many times higher than the volume of material digitised by the national archives of the USA, which has so far digitised the most material. (As a result of mass digitisation, the digitisation of 212,000 shelf metres would result in roughly 3.4 billion images; the national archives of the USA has digitised three hundred and ninety million images in total, with an annual production of thirty million images.)

During full production, the aim is to digitise 22,000 shelf metres a year through mass digitisation. This means that roughly three hundred and fifty million images will be created every year, being more than the volume of material digitised so far by all national archives of the Nordic countries combined. In summary, the volume of material planned for mass digitisation is many times higher than current digitisation volumes.

4.3 Sequence and schedule of digitisation

In the roadmap that guides the sequence and schedule for digitising material of government agencies, material has been divided into four groups on the basis of the schedule of mass digitisation and criteria affecting the digitisation of material. These criteria are based on information collected by means of material surveys, such as the number of loose papers and the level of description and organisation. The purpose of these criteria is to offer guidelines for dividing material of government agencies into these four groups. These are not absolute criteria,

and the final sequence will be based on an overall estimate. The schedule of mass digitisation and the division of material into groups included in the roadmap, including material volumes and criteria, are presented in Table 1.

Table 1. Schedule for digitising material through mass digitisation and division of material into groups in the roadmap (groups 1–4)

Schedule	7/2019–7/2021	7/2021–2023	2024–2026	2027–7/2030
Groups	Group 1	Group 2	Group 3	Group 4
Volumes (shelf kilometres = km) Total 212 km	12 km	~52 km	~67 km	~81 km
Criteria	1. Quick and easy digitisation of material 2. Ability of government agencies to prepare material for transfer to the digitisation process 3. High need to use material	1. High need to use material 2. Changes in organisation and archive room 3. Organisations with largely digital operations → digitisation of remaining papers	1. Slowly digitised material 2. Low need to use material	1. Slowly digitised material 2. Low volume of material / not possible to stop using archive rooms 3. Organisation not yet digital

The preparation of the roadmap will continue during spring 2018 by defining an organisation- and archive-specific digitisation sequence in cooperation with government agencies. As a result, a concrete schedule for digitising the material of each organisation will be defined. Currently, the roadmap has only been prepared by government organisations, as no guidelines have yet been set for the digitisation of material transferred from municipalities to regions.



The digitised material and their digitisation sequence have been presented in more detail in Appendix 1.2.

5 Mass digitisation process

During the planning project for mass digitisation, one process was defined, according to which all material within the scope of mass digitisation will be digitised. The process consists of the following roles and their tasks:

- *Government agencies* own the material to be digitised. Government agencies are responsible for preparing their material.
- *The National Archives of Finland* is responsible for steering the preparations made by government agencies. The National Archives of Finland coordinates and supervises the fulfilment of the schedule set out in the roadmap plan with regard to preparations made by government agencies and the progress of digitisation. Furthermore, the National Archives of Finland are responsible for preparing preservation decisions.
- *Digitising party* refers to a service provider that is responsible for the practical implementation of mass digitisation, ranging from the planning of the logistics and transfer of material to the disposal of the analogue material digitised or its storage at the National Archives of Finland.

The overall digitisation process (Figure 1) is divided into six separate functional processes. The end of each process launches the next process.

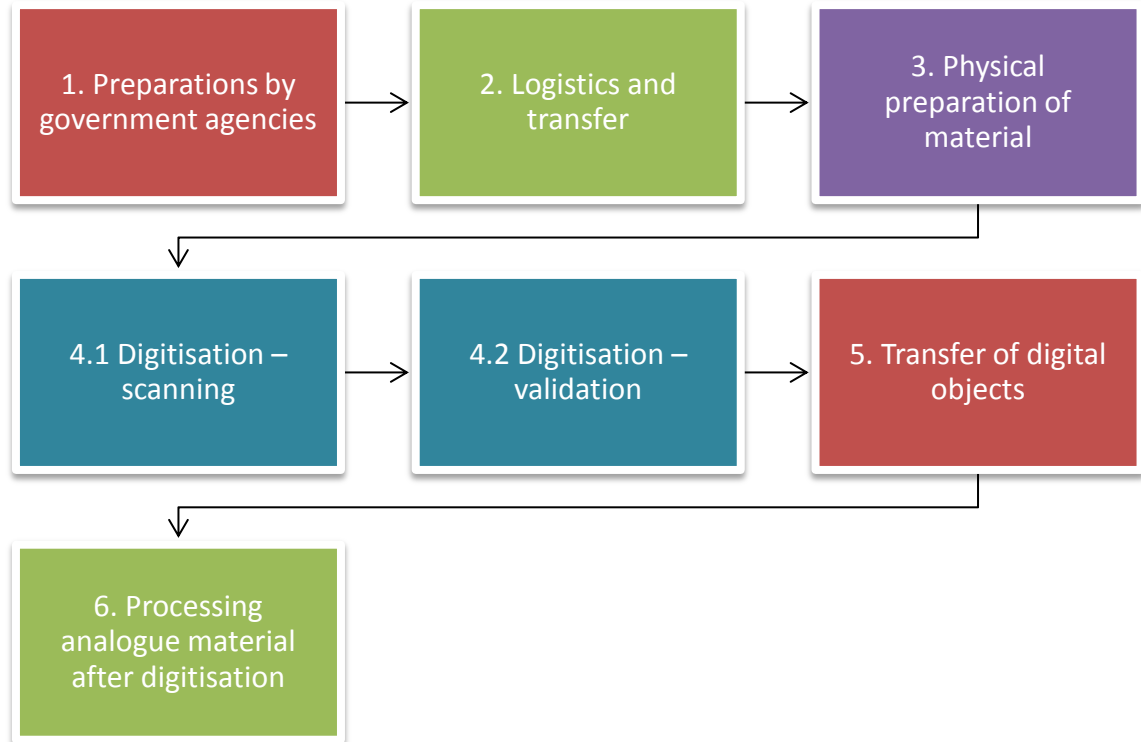


Figure 1. Mass digitisation process

A majority of the material to be digitised through mass digitisation comprises cultural heritage in the form of records defined for permanent storage. Its analogue manifestation will be disposed after it has been converted into digital format. As a result, the digitisation process has many different requirements, ranging from the methods of processing analogue material to the production of a probative and authentic digital object with a full data content. An archival master file of as high a quality as possible must be produced for long-term storage, and it must be possible to reproduce different access files for different needs.

In the controlled completion of the process, the production management system plays a key part. It is used to control material sent for digitisation, the transfer of material and all activities carried out for the material in the digitisation location. The production management system ensures that a specific function does not receive an excess volume of material at any single time



and that all functions receive a sufficient volume of material. Therefore, the aim of the system is to avoid any bottlenecks and idle time.

The mass digitisation process has been described in more detail in Appendices 3.1 and 3.2.

5.1 Preconditions of the process

The material to be digitised must have the preservation decision of the National Archives of Finland, in which the storage of the analogue records to be preserved permanently or their disposal after digitisation is defined. The preservation decision defines the storage format of digitised records on the basis of their weight as evidence and cultural historical value. It is estimated that approximately 10 percent of the digitised material will also be preserved in analogue format after the digitisation process. The ground rule is that any previous value definition (storage period) set by the National Archives of Finland for the record will not be changed in the preservation decision.

The sequence of material to be digitised and the transfer schedule are defined in accordance with the roadmap plan. Government agencies must commit to the schedules set out in the roadmap plan and time their preparations so that material can be transferred for digitisation in accordance with the schedules set out in the roadmap. Commitment to roadmap schedules is a critical factor for the success of mass digitisation.

5.2 Preparations by government agencies

Government agencies must prepare their material for logistics and transfer. Government agencies identify the material to be digitised on the basis of the roadmap. The roadmap defines the digitisation schedule.



The aim of the preparatory stage is to enable the reliable transfer of material to the digitisation location. Its aim is not to prepare material ready for any direct digitisation. Preparations by government agencies require fewer actions from the National Archives of Finland compared with the current transfer procedure, and the physical preparation process can be lighter. Government agencies must still catalogue their material in the catalogue system of the National Archives of Finland and protect their material. However, storage units do not need to be suitable for archiving; it is sufficient that they protect the material during transportation and prevent the material from being mixed with other material. In addition, government agencies must also identify their storage units by using barcodes available from the catalogue system. The material will be prepared for scanning in the digitisation location.

The National Archives of Finland will issue more detailed instructions for new preparatory procedures. It should be noted that the aim of preparations by government agencies will change fundamentally. Previously, the aim of preparations has been to preserve the analogue format of material. The starting point of the new procedure, however, is that analogue objects are disposed after digitisation.

At the final stage of preparations by government agencies, an expert from the National Archives of Finland, together with each government agency, will prepare a material processing plan, in which any special features related to preparation and digitisation processes are described. The aim of the processing plan is to facilitate the planning of different process stages in the digitisation location and, therefore, increase proactivity. This process results in *material ready for transfer*.

5.3 Logistics and transfer

The material to be digitised is scattered in the archives of different government agencies. The material must be transferred reliably and under proper coordination from different archives to a



single central digitisation location. A large interim storage will not be set up in the digitisation location. The production management system plays a key part in terms of logistics, and it maintains information about the volume of material *ready for transfer* in the facilities of different government agencies.

The digitising party will be wholly responsible for logistics and transportation. It must plan and order transportation for material located in the facilities of government agencies. Government agencies will not take part in the organisation of transportation. During logistics and transfers, it must be possible to track the material at all times, from the facilities of government agencies to the mode of transport and interim storage in the digitisation location, on the basis of status data controlled in the production management system and by using GPS for transportation tracking.

In the digitisation location, the material will be transferred for interim storage in trolleys to wait for digitisation. This process results in *transferred material*.

Logistics for the mass digitisation process has been described in more detail in Appendices 3.4 and 3.5.

5.4 Digitisation preparations

Digitisation preparations ensure that digitisation is an efficient and safe process. Preparations should be carried out in the digitisation location, as they require, for example, special expertise in the handling of different materials and competence in the use of scanning systems. Any improperly prepared material will stop the digitisation process and may damage equipment or, in the worst case, cause information to be lost.

Preservation decisions issued by the National Archives of Finland and the storage platform of the record (i.e. material type) must be taken into account during preparations. If the record is also



preserved in analogue format after digitisation, preparatory activities and any material added must be suitable for archiving. Preparations should be as minor and clear as possible. If possible, the efficiency of preparations will be improved; however, without damaging the data content of material.

Preparations take a lot of time in the digitisation process, as they mainly need to be carried out manually. Automation cannot often be used, as it increases the risk of losing information before digitisation. The processing plan prepared during preparations by government agencies serves to estimate the resources and activities required.

At the preparation stage, it must be decided which system is used to scan the material. Mainly, material must be prepared as separate sheets for a high-volume scanner. Bound material can be digitised after unbinding. If any material cannot be scanned using a high-volume scanner, it must be prepared for a special scanner. This process results in *material ready for digitisation*.

5.5 Digitisation

Digitisation consists of the conversion of analogue material ready for digitisation into digital format (scanning into images), the validation of the image quality, the recognition of the image content using optical character recognition (OCR) methods and the generation of a transfer package.

The aim of the scanning process is to produce an archival master file and an access file by means of a single scan. During scanning, every archival unit has been identified, and the content of each archival unit will be scanned in the original sequence.



Validation means that the quality of each image is verified after scanning. During validation, it is ensured that digital objects generated as a result of scanning are in accordance with specifications and the digital manifestation is intact.

The content is recognised using OCR methods that, in initial phases, are only targeted at those digital objects, the data content of which is mainly machine-encoded. OCR methods will also be used as part of the validation of the image quality.

It should be emphasised that quality assurance is a more extensive process than validation alone. The overall process does not have any specific single point, at which quality is assured. Quality assurance also includes the verification of the functionality of equipment and software, already before any scanning. The preservation, integrity, and authenticity of records and the data they contain will be ensured by keeping the entire process, ranging from preparations by government agencies to the final disposal of material, under control and saving information about activities carried out at different stages of the process.

At the end of the process, a transfer package will be generated of digital objects, the OCR result and metadata. The process results in *digitised material ready for sending*.

Digital objects

Table 2. Digital objects produced from analogue manifestations

File role	File format	Compression	Resolution	Bit depth	Suitable ICC profile
Archival master file	TIFF 6.0	LZW (lossless)	300 ppi	24	sRGB (or higher)
Access file	JPEG	JPEG quality 60 percent	300 ppi	24	sRGB

Recognised text	ALTO (3.0 or 3.1) XML ⁶	-	-	-	-
Metadata describing the digitisation process	XML (MIX schema ⁷)	-	-	-	-

Of each document page with a data content (front and back sides), archival master files and access files with the same data content will be produced. The image resolution, bits and colours are based on securing the weight of the digital object as evidence and the lossless preservation of the information contained by analogue objects. Storage objects will be lossless files, so that the quality of the digital image and, therefore, the information it contains and its weight as evidence do not suffer from any future file migrations. In addition, file formats with loss, such as JPEG, reduce the OCR accuracy⁷. An XML file in accordance with the MIX schema will be produced for each storage object to document the digitisation process. Of each document page containing machine-encoded data content, a separate XML file will be saved in Analyzed Layout and Text Object (ALTO) format. This enables text searches for digital objects.

Requirements for digital objects produced in the digitisation process have been described in more detail in Appendix 3.3.

5.6 Transfer of digital objects

Digital objects will be transferred automatically so that the production management system is responsible for transferring transfer packages to the receiving system that checks the transfer packages and acknowledges that the transfer succeeded or failed to the production management

⁶ The Library of Congress » Standards » ALTO. Website of the Library of Congress. Accessed 19 December 2017. <https://www.loc.gov/standards/alto/>

⁷ The Library of Congress » Standards » MIX. Website of the Library of Congress. Accessed 20 December 2017. <http://www.loc.gov/standards/mix/>



system. It has been planned that the digital archiving service system (SAPA) will act as the receiving system.

After an approved receipt, the process is completed and results in *a completed transfer of digitised material*.

5.7 Processing analogue material after digitisation

Processing analogue material depends on the preservation decision issued for each material, i.e. whether analogue material is disposed after digitisation or whether it is preserved. Analogue material received from scanning is stored for a pre-defined period, after which it is destroyed in a secure manner or encased and stored at the National Archives of Finland. If analogue material is preserved after digitisation, the digitising party will be responsible for encasing the material after it has been digitised. The National Archives of Finland will be responsible for storing the material.

6 Production of mass digitisation

Production planning for mass digitisation has been carried out on the basis of the material volume identified in the roadmap (212,000 shelf metres). The goal is that all material has been digitised by 2030. The first two years will be spent on production preparations, and production will start up to its full capacity starting from the third year. In terms of material volumes, this means that 12,000 shelf metres will be digitised during the first two years (*production start-up stage*) and 200,000 shelf metres will be digitised during the next nine years (*production stage*). In total, some **1.7 billion sheets** will be digitised. Considering that some sheets are two-sided, there will be nearly **2.55 billion images**, i.e. **40 petabytes of data**. These figures are based on an estimate that there are 8000 records per shelf metre. However, there can be great variation in the number of records per shelf metre, depending on the form of each analogue object. Regardless of any inaccuracies in estimates and on the basis of an international survey, this is a much broader scale than in any previous digitisation projects.

Resulting requirements for production capacity are presented in Tables 3–4. The figures are based on surveys conducted for the roadmap and estimates of the production capacity of different equipment and personnel.

Table 3. Production volumes at the production start-up stage

Digitised material per year	6000 shelf metres, approx. 48 million sheets
Digitised material per working day (250 working days)	24 shelf metres, approx. 192,000 sheets
- of which high-volume material	22 shelf metres, approx. 176,000 sheets
- of which special material not suitable for high-volume production	2 shelf metres, approx. 16,000 sheets

Table 4. Production volumes at the production stage

Digitised material per year	22,250 shelf metres, approx. 178 million sheets
Digitised material per working day (250 working days)	89 shelf metres, approx. 712,000 sheets
- of which high-volume material	80 shelf metres, approx. 640,000 sheets
- of which special material not suitable for high-volume production	9 shelf metres, approx. 72,000 sheets

Scanning equipment used in mass digitisation

The scanning infrastructure designed for mass digitisation enables paper material to be scanned. Table 5 present the equipment required by the planned mass digitisation process.

Table 5. Scanning equipment used in mass digitisation

Device	Technology	Purpose of use
National Archives of Finland	Rauhankatu 17	PO Box 258, 00171 Helsinki
Riksarkivet	Fredsgatan 17	PB 258, 00171 Helsingfors
		Tel. Tel. +358 29 533 7000
		Fax +358 9 176 302
		kirjaamo@arkisto.fi
		http://www.arkisto.fi



High-volume scanner	An open-track scanner where analogue material does not pass through two rolls. Enabling the simultaneous scanning of both sides of a document.	Loose sheets prepared for a high-volume scanner. The primary scanning method.
Large scanner	A feed-through scanner for large-sized material, enabling the scanning of at most 140 cm analogue material. Enabling the simultaneous scanning of both sides of a document.	Special material. Paper sheets larger than A3.
Book scanner	A scanner for enabling bound material without any unbinding. Also suitable for scanning other analogue material that is not suitable for other scanner types.	Special material. Analogue material that cannot be scanned with other scanning technologies.
Document scanner	A desktop feed-through scanner. Enabling the simultaneous scanning of both sides of a document.	Special material. An alternative for high-volume scanners to scan loose sheets of under A3. To be used at discretion.

Well-planned and properly controlled logistics and material preparations enable the most effective and productive scanning stage. When preparing material for digitisation, the material is sorted and assigned to the correct scanner for scanning. High-volume scanners are independent scanning lines, in which a single scanner forms a single production line. The plan is to use high-volume scanners to scan suitable material continuously and as efficiently as possible. Any material, the properties of which would require a high-volume scanner to be slowed down, will be assigned to another scanner type. Any material unsuitable for a high-volume scanner will be scanned using a special material line, consisting of document scanners, large scanners and book scanners.

Other equipment

In addition to scanning equipment, the process requires other equipment and tools to be acquired. Different equipment will be needed to prepare material, such as cutters for unbinding, heated spatulas or similar for straightening any folds, vacuums and metal detectors to identify



KANSALLISARKISTO

Planning project for mass digitisation

Vipuvoimaa
EU:lta
2014–2020



Euroopan unioni
Euroopan aluekehitysrahasto

staples and other metal binders. The digitisation locations also requires, for example, a fume hood if mouldy material is prepared for digitisation. In addition, proper personal protective equipment must be available. For logistics, several hundreds of trolleys will be needed to store and move material during the production process. The disposal of analogue material may also require the acquisition of shredders and a baling press, unless this process is outsourced. In addition, the IT infrastructure will be significant, ranging from server solutions to individual terminal devices.

Separate systems and software will also be needed in production. The entire production process will be controlled by means of a production management system. Other significant software needed includes scanning and OCR software.



Personnel needed in the process

To complete the process, the roles presented in Table 6 have been defined.

Table 6. Roles defined for the production of mass digitisation

Role	Task
Storage employee	Receipt and relocation of material at different process stages. For example, removing scanned material from high-volume scanners and storing it in interim storage.
Preparation employee	Preparation of material for different scanning processes.
High-volume scanner operator	Scanning material using a high-volume scanner. Loading the scanner and scanning bundles.
Special line operator	Scanning using a special line. Loading the scanner and scanning material.
Validator	Validation, i.e. inspecting the quality of scanning work visually.
Supervisor	Regular supervisory tasks and day-to-day decisions.
Information service employees	Responsibility for responding to material requests for material in the digitisation location. Typically, material requests concern material that is in the interim storage waiting for preparations.

In addition to the aforementioned roles, employees will be needed in production planning, management and logistics. Furthermore, technical employees will be needed to see to the continuous maintenance of scanning equipment, the IT infrastructure and software.

Facilities in the digitisation location

The start of mass digitisation requires sufficiently large facilities suitable for the purpose. The need for facilities will increase in stages as digitisation volumes increase. Special attention must be paid to the space needed for the interim storage of incoming material and material waiting for destruction, the easy transfer of material inside the facilities and the social facilities required for a large number of employees. There are also other special needs, for example, relating to power supply and air conditioning.

The planning project for mass digitisation has identified the start of the digitisation process in the facilities of the Mikkeli central archives of the National Archives of Finland. The start of the



digitisation process in the Mikkeli central archives is a good option as all facilities in the central archives can, in practice, be assigned to the digitisation process. As a result, no additional costs will arise from facilities. In addition, the central archives has a vacant warehouse for interim storage. Situating the digitisation process in Mikkeli also has other advantages, such as the location the digitisation function of the National Archives of Finland in Mikkeli and the strong competence centre specialising in digital data management in the region.

During the project, digitisation facilities were designed in the facilities of the central archives. On the basis of this design, the central archives only need minor modifications in order to be used as mass digitisation facilities at the production start-up stage. After the start-up stage, the full digitisation production capacity will, however, require larger facilities than exist in the central archives for this purpose.

7 Costs and advantages

7.1 Digitisation costs

The starting point of the cost estimate for mass digitisation is the material volume of 212,000 shelf metres as defined in the roadmap and the decision in principle on the completion of the digitisation process by 2030. Costs are based on the production described in section 5.

Nearly 80 percent of mass digitisation costs come from manual production work. With regard to remaining costs, the determining factor is the amount of work required for the preparation of material and scanning. The workload is affected by the methods and technologies used, i.e. how the efficiency of preparations and scanning can be improved. Above all, the material to be digitised has a significant impact, as the workload required for preparing and scanning material that is difficult to process may be more than ten times higher than that required for material that



is easy to prepare. At the start-up stage, costs also come from necessary investments in equipment, data processing capacity and software.

The starting point of cost estimates is a digitisation process by using current commercially available technologies (e.g. scanning equipment and OCR methods). Costs have not been estimated on the basis of any solutions introduced in the future. However, the cost estimate involves uncertainties, the most significant of which is associated with the digitised material: The volume and properties of material may differ considerably from estimates, which may have a significant impact on the workload required for preparing and scanning the material. Correspondingly, if a larger volume of material is easier to digitise than expected, the workload and costs will be lower.

7.2 Advantages of digitisation

The key advantage of mass digitisation is the significantly improved availability and usability of material as a result of digitisation. Digital material is available independent of time and place, and information can be searched for using modern search tools, such as full-text searches and various ontologies. Public material is also better available directly online, which allows users to search for information and produce services as a self-service.

Financial benefits start from the benefits that can be directly assigned to budgets of government agencies. Therefore, advantages related to the improved availability of information or the availability and usability of information used for research purposes have not been analysed as part of financial benefits. However, these are significant advantages and their significance increases as more material is made available in digital format. A good example is that different types of material can be combined for research purposes. In addition, longer time series can be used in digital format.



7.3 Analysis of costs and advantages

The starting point of the analysis of costs and advantages is the comparison of two storage scenarios. *In the scenario of analogue storage*, current archiving practices are continued, so that government agencies slowly transfer analogue records dating back forty years or more to the National Archives of Finland. In this scenario, the National Archives of Finland must start designing new central archives in the near future as the current facilities will be full. Furthermore, government agencies will maintain their paper archives for dozens of years, forcing them to tie facility and employee costs to their management.

In the scenario of digital storage, the transition of government agencies to a fully digital data management system will be sped up by mass-digitising analogue material by 2030. As a result, they can stop maintaining paper archives much earlier than originally expected. In addition, the analogue records required in operations are within the scope of the same data management processes and procedures as digital material. In digital administration, it is not purposeful to process and control extensive paper archives. Carrying out two overlapping processes is inefficient and generates unnecessary costs.

The analysis of costs and advantages is based on an estimate of current costs arising from paper archives of government agencies and an estimate of the development of costs until 2040. Furthermore, the analysis of costs and advantages compares the cost savings resulting from the scenario of digital storage and those resulting from the scenario of analogue storage.

- **Costs saved by government agencies from archive facilities** As a result of digitisation, the archives of government agencies will become empty in stages. Costs arising from archive facilities of government agencies, not including the facilities of the National Archives of Finland, the National Board of Antiquities and certain security agencies, total approximately seven million euros a year. According to the estimate of Senate Properties, it is possible to save roughly 70 percent from facility costs after the digitisation process



and the facilities have become vacant. The realisation of these savings requires that the digitisation process is coordinated so that archive facilities can be emptied as uniform facilities.

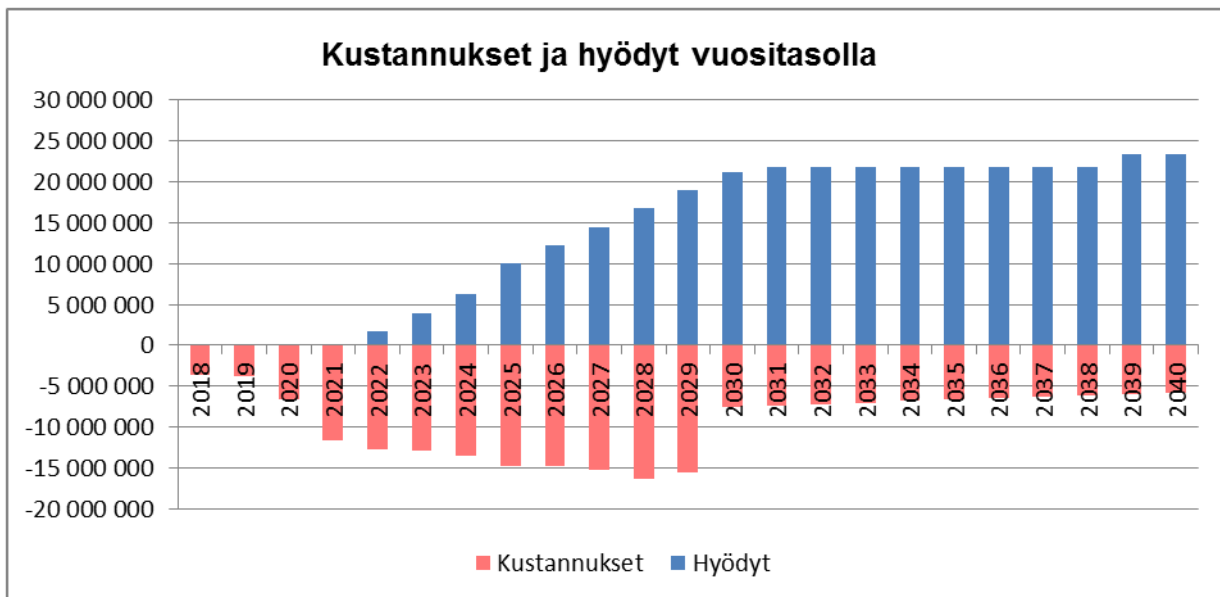
- **Decrease in person-years assigned to record management and archiving in government agencies** As a result of digitisation, tasks related to the use and management of material will change in government agencies. All tasks related to the manual management of archives will no longer be necessary (archiving, pick-up from archive and arrangement of archives). Significant savings can be obtained from tasks related to the use of material (information service) and time spent on searching for and finding information. In the public sector, personnel costs related to record management and archiving tasks total fifty million euros a year. By discontinuing paper archives and related processes, it is estimated that 30 percent of person-years related to record management and archiving can be saved. The realisation of these savings requires that government agencies change their operating methods and processes suitable for the digital operating environment. Organisations that modernise their processes and are able to develop completely new operating methods will achieve the highest savings.
- **No need to build new archive facilities for the National Archives of Finland** If paper archiving is extended, the National Archives of Finland will have to restart the receipt of paper records . In the scenario of analogue storage, the National Archives of Finland should open new central archives in 2025 and the third central archives in 2039. The total costs of these new buildings at current prices would be 32.8 million euros.

In the scenario of digital storage, costs are made up of the mass digitisation process over the next eleven years and digital storage costs that are continuous and will increase as the volume of

digital material increases. Digital storage costs are based on the storage of archival master files in the long-term storage service of the National Digital Library (KDK-PAS).⁸

When the mass digitisation process ends in 2030, annual savings in the analogue storage scenario compared with the digital storage scenario are estimated to be fourteen million euros a year. On one hand, cost savings are affected by the realisation of facility and personnel savings in government agencies and, on the other, by digital storage costs that depend on the data volume stored and the unit price of storage space.

Table 7. Costs and benefits of mass digitisation at an annual level. Costs on red and benefits on blue.



⁸ The estimate of storage costs until 2030 is based on CSC's estimated budget, with which total operating costs can be calculated. Any decreases in the prices of storage technologies have been taken into account so that acquisition costs will decrease by 10 percent in 2019–2029. It is much more difficult to estimate costs from 2021–2022 onwards, as technological changes are currently in progress regarding storage technologies.

8 Risks associated with mass digitisation

Identified risks associated with the mass digitisation process presented herein are listed below. Risks associated with different implementation options have been analysed at a general level, not in greater detail.

Table 8. Risks associated with mass digitisation

#	Description of risk	Impact
1	Management and cooperation: Cooperation with different parties (the National Archives of Finland, the Ministry of Education and Culture, the Ministry of Finance, government agencies) is not functional. The steering roles of ministries are not clear. Parties are not sufficiently committed to the operating model.	The goals of mass digitisation change during the process, the continuity of mass digitisation is uncertain and any funding is discontinued.
2	Administration of the operating model fails: The management and administration of the new function and operating model are not functional. The National Archives of Finland faces problems steering the operating model and implementing it with government agencies.	Problems in the operating model; the goals set cannot be reached. Government agencies are unclear over the operating model.
3	Abilities of related systems: Several different data systems (see section 8) are related to mass digitisation. The risk is that the implementation of these data systems is delayed or their ability is not sufficient considering mass digitisation.	Significant deviations may have an impact on the schedule for starting the digitisation process. As a result of minor deviations, replacement solutions are needed, and these are reflected in the smoothness of the process.
4	Accuracy of digitised material: The volume of digitised material is larger than expected or its quality is more challenging than expected. The material may also be easier to process than expected.	The volume of digitised material and digitisation costs increase. Impact on the digitisation schedule.



5	Challenging in starting production: No practices that directly apply to the start of production exist. There are no service providers that have carried out a similar project in terms of scale and quality requirements. The risk is that product starts slower than expected.	Impact on the digitisation schedule and production costs.
6	Challenges in the expertise required: The digitising party does not possess sufficient expertise in digitisation. The National Archives of Finland does not have sufficient resources for steering government agencies and preparing preservation decisions.	Impact on the digitisation production rate and, in the worst case, the quality of digitisation. Impact is reflected in defective preparations by government agencies or slow preservation decision processes.
7	Problems in preparations by government agencies: Government agencies do not have the resources to prepare their material within the schedule set out in the roadmap.	Impact on the availability of digitised material. Material is not sent for digitisation at the planned extent. Significant impact on the whole digitisation process.
8	Problems in the process: The process does not work reliably and without problems.	Problems in the process lead to changes in the process, possibly to new additional stages and, for example, to more quality assurance tasks.

9 Dependencies of mass digitisation and further activities

The start of mass digitisation production requires many different abilities. Some of these are directly related to the operating model presented in this report, while some are not directly within the scope of mass digitisation, but they are external dependencies from other data systems.

Here, abilities and external dependencies related to the operating model are examined from the point of view of *architectural capabilities*. "Capabilities" refers to abilities to act purposefully in a specific field and to use individual skills and resources in order to reach goals.



To fulfil the capabilities of organisations, usually combinations of the following three areas are needed: *operating models and processes, personnel and skills, and information and systems*. The capabilities required in these areas in order to implement the operating model and process presented herein are described below from the points of view of *personnel and skills* and *information and systems*. The progress of external dependencies requires systematic monitoring and a clear risk matrix.

9.1 Personnel and skills

National Archives of Finland:

- The capability to steer government agencies to prepare their material for digitisation and draw up material processing plans.
- The capability to maintain the digitisation roadmap together with government agencies. The roadmap defines the sequence and schedule for digitising the material of government agencies.
- The capability to make preservation decisions on digitised material within a tight schedule and in large quantities.
- The National Archives of Finland has experience in steering the activities of government agencies, transfer planning and preservation decisions. Key challenges in the development of capabilities come from the requirement to modify processes to correspond with new needs and from significantly increasing material volumes, the steering of which will require new methods.

Digitising party:

- The capability to implement the digitisation process in accordance with the specifications of the National Archives of Finland. In particular, the digitising party must be able to



product digital objects in accordance with the specifications of the National Archives of Finland and to ensure a high level of quality by using the specified methods.

- The capability to implement mass digitisation with the material volumes and to the extent defined in the roadmap, i.e. the capability to carry out a large-scale digitisation process.
- The capability to steer the production of mass digitisation and the transfer of material from government agencies to digitisation facilities.
- The requirements set for the digitising party regarding the quality of digitisation and material volumes are unique, also on an international scale, and no service provider directly has the abilities required for mass digitisation.

Government agencies:

- The capability to prepare material in accordance with new transfer instructions to wait for transfer for digitisation. Government agencies must carry out their preparations within the schedule set out in the roadmap plan. Material must be prepared in accordance with new starting points, as the goal is no longer the preservation of the original analogue material, but its transfer and digitisation.
- The capability to use digital material as part of processes and adapt individual operating methods to digital data management.
- The mass digitisation process is a significant change for government agencies compared with the current situation. The rapid transition from paper archiving to a fully digital data management can be a massive change in some organisations, requiring advance preparations. For many government agencies, the replacement of conventional paper archiving tasks by digital data management tasks may require new skills.



9.2 Data systems

Master data systems considering the mass digitisation process are presented below. Services carried out by data systems are governed by various capability requirements that have not been explained in detail herein. The starting point is to show dependencies to master data systems.

- *The production management system* is a mass digitisation management system, the interfaces, and information of which are used to steer the overall mass digitisation process and to document actions carried out for the material. The system must be closely integrated into systems of the National Archives of Finland. The implementation of the system has not yet been started.
- *The metadata system of the National Archives of Finland (AHAA)* is a data reserve which contains key information about descriptions and catalogues of archival material. In the mass digitisation process, the system will be used to catalogue material of government agencies and as a core data reserve for material data, from which data is transferred to the production management system. The system will be deployed by the National Archives of Finland during 2018. The capability for mass digitisation may require the development of system interfaces.
- *The material management system of the National Archives of Finland (AHJ)* is a data reserve which contains key information about the location of analogue material. The system will be used in the background of the production management system and, in particular, for the positioning of material to be preserved in analogue format. The system will be deployed by the National Archives of Finland during 2018.
- *The receiving system* receives transfer packages generated during the digitisation process, enriches them and distributes them for use and storage. The SAPA service system has been planned to carry out the tasks of the receiving system. The implementation of the system has not yet been started.



- *The long-term storage system* is responsible for the long-term storage of archival master files generated during the digitisation process. The system exists as the KDK-PAS service. The capabilities required for mass digitisation may require that the system is developed.
- *Content analysis* uses text recognised by means of OCR methods and automatically produces metadata from text by means of analysis to make the use of material easier. The implementation of the system has not yet been started.

9.3 Proposed further activities

To improve mass digitisation abilities and requirements, at least the following further activities are considered to be necessary. Any delays in these activities have an impact on the schedule for starting the production stage of mass digitisation.

- All parties must be committed to the goals of mass digitisation. Government agencies cannot start their preparations, unless clear and binding decisions on the mass digitisation process are made. This requires that the ministries that steer the process issue clear statements, are committed and make decisions on funding.
- Government agencies must commit to the schedules defined in the roadmap regarding material transfers. The National Archives of Finland must specify the roadmap plan together with government agencies. However, government agencies cannot commit to the roadmap, unless steering ministries issue a sufficiently strong message regarding the start of mass digitisation.
- The transfer abilities of government agencies must be developed. The critical point in the mass digitisation process is the transfer of material from government agencies to the digitisation location. Government agencies require straightforward instructions, defining what actions must be carried out for their material and how the material can be prepared for digitisation. The National Archives of Finland must prepare instructions for the preparation stage.



- The SAPA service system for the digital archiving of material must be functional so that material converted into digital format can be used. The SAPA service system is needed so that records converted into digital format are available to government agencies as part of their tasks and so that material can be offered to other parties via easy-to-use interfaces. Analogue material cannot be disposed after digitisation if it is not widely and reliably usable.
- *A proof of concept* regarding high-volume scanners and OCR methods. The National Archives of Finland must test the selected technologies in order to obtain a better understanding of preparation, digitisation and scanning activities.
- The production management system must be defined, implementation must be started and interfaces must be built in the AHAA and AHJ systems of the National Archives of Finland. Mass digitisation cannot be carried out reliably without a production management system designed for the specific purpose.