

Ville Kajanne

Disposal of Documents After Mass Digitisation

Abstract: The paper focuses on how a reliable implementation model, a meaningful schedule and a reasonable cost structure in terms of long-term digital preservation and usability of digital manifestations can be implemented in a large-scale digitisation production. Not all records are suitable for large-scale mass digitisation. The fragile condition of the material, its cultural-historical value or restricted personal data may be grounds for exclusion from the service. By exclusion, worst risk scenarios regarding disposal-oriented digitisation can be dealt with. Production can be effective and credible only if the material has been carefully reviewed, e. g. through pilot test production. This applies to both cost structure and qualitative target points. Competent staff can increase autonomy as turnkey solutions often do not exist and cut ongoing maintenance costs. Image quality, content recognition, file format and archival transfer package are key requirements but achieving them in a large-scale production requires sophisticated and both automated and human-based implementations throughout the digitisation process. The final disposal process requires quality control checkpoints at various stages of the digitisation process, complemented by customer testing of the access files and a secure disposal procedure.

This chapter focuses on the digitisation of the official documents created by ministries or other central government agencies. Documents with no cultural-historical value (typically less than hundred-year-old documents) are digitised in the mass digitisation process and disposed after that. Documents with cultural-historical value (typically more than hundred-year-old documents) are digitised in the retroactive digitisation process. In both cases the digital manifestations of these documents will be archived permanently.

1 Introduction: Which Documents Are Destroyed and on What Grounds?

Documents formed in Finland are now stored primarily in digital format. The main reason is that they are born digital and used mainly in digital services. Previously

Ville Kajanne, National Archives of Finland, Helsinki

the Archives Act required that documents stored permanently must be stored in analogue format (usually on paper). Because of this large numbers of born-digital documents were converted to paper format only so that they could be transferred to the National Archives after their original purpose of use had ended. The central government's decreasing archive storage facilities and the increasingly faster handling of matters has encouraged the authorities to transfer documents to the National Archives (NAF) more and more quickly as a result of which the National Archives waived the previous age of forty years as a condition for handing over documents. This accelerating circulation of documents, the general digitalisation of document management and the objectives of the Government Programme to significantly promote the digitalisation of paper materials led to amendments to the Archives Act in 2019.

Under current legislation paper documents can be converted and permanently preserved in digital format provided that their preservation, integrity, authenticity, cultural-historical value and legal weight as evidence are not compromised. Consequently, the National Archives has stated that permanently archived documents can only be transferred to the National Archives in digital format (i. e. born-digital or, in the case of paper documents, equipped with a contract regarding the digitisation responsibilities). Authorities can implement the digitisation themselves or purchase it as a service from private service providers or the National Archives.¹ Certain fixed-format documents are accepted into the mass digitisation service of the National Archives; in this case, the National Archives is practically responsible for the costs of digitisation. However, the current mass digitisation production capacity enables only a relatively limited annual amount of digitisation (approximately 1,500–2,000 metres of shelves) whereas over 125 kilometres of shelves of documents suitable for mass digitisation have been identified as still in the possession of the authorities.

Since then, the availability of documents in digital format has also been required by legislation other than the Archives Act. The relevant legislation is currently being updated² which is practically the new prevailing status quo; from the point of view

¹ As NAF approves official documents only digitally but there have been only a few operators so far, not to mention the governmental agencies themselves, with necessary expertise to provide digitisation services efficiently, it has been deemed appropriate that NAF provides these services to a certain extent as well. Requirements for the disposal-oriented digitisation are always the same regardless of the service provider.

² Act on Information Management in Public Administration (2019) changed somewhat the terminology of permanent preservation and/or archiving in Finland. According to this Act the 'data material' is first in the 'preserving stage' (still necessary for the records creator agency). Therefore, there can be permanent preservation for the original purpose of the document while it is still being

of Finnish legislation digital documents are as authentic as analogue ones. They are not copies as they are original and unaltered documents based on their data content and context. Actually, the focus of the debate in the governmental records management services has shifted from mere digitisation to digitalisation; providing citizen-centred and interoperable digital services is one of the key strategic goals of public administration in Finland. In addition to digitisation the current emphasis is thus on digitalisation which are prerequisites for each other. In addition to the actual services (systems) digitalisation refers to artificial intelligence supported systems or otherwise more versatile ways to search for content in born-digital and digitised material.

To summarise, the vast majority of the materials to be transferred to the National Archives are permanently stored only in digital format. Since its appraisal strategy in 2012 the National Archives itself has aimed to ensure that archived (permanently stored) documents are kept exclusively in electronic form whenever possible while digital preservation also allows for a larger share of the materials to be preserved. Previously, when the materials had to be preserved in paper form the aim was to dispose of up to 85–90% of them on a fixed-term basis. With regard to digital preservation a corresponding quantitative target for the share of material to be archived permanently has not been made, at least not so far.

When it comes to analogue documents their archival storage form is decided before they are transferred to the National Archives. The owner of the materials (an authority) prepares the proposal and the National Archives makes the actual decision. The assessment of the preservation form is based on the National Archives valuation and appraisal policy document and above all the guidelines for the categories and criteria for the valuation of the preservation form attached to it. In practice, therefore, authorities can propose to the National Archives that the document be archived in analogue (mostly paper) format on the basis of its cultural-historical value. This division of duties may change in the future so that the authority decides more and more independently both on the form of storage of its materials and on their possible cultural-historical value. However, the prerequisite is that the authority documents the decision and the reasons so that it can demonstrate, on request, how the valuation criteria of the National Archives were applied in the decision.

What is meant by cultural-historical value? Internationally, the term ‘intrinsic value’ has usually been used to refer to the material value of records. In Finland

held by the agency. After the original purpose has expired, NAF can stipulate ‘data material’ to be ‘permanently archived’. The situation is somewhat incomplete since the Archival Act is currently being reformed and consequently the concepts can change. Finally, it must be pointed out that even documents which are still in the ‘preserving stage’ can be transferred to NAF for digitisation.

the National Archives' definition of cultural-historical value includes both evidential and informative value. The values can therefore be tangible or intangible but still linked to the analogue manifestation of the document. Cultural-historical value, i. e. material value, is related to the features of a record – for example, a special analogue format such as a glass negative or parchment. The value can lie in the material itself or in the information it contains which cannot reliably be preserved in any other manifestation (format).³

Preservation in analogue format requires that the conditions of at least two valuation criteria are met. The first criterion that is required in all cases is the criterion according to which the condition of the document must be such that it can be permanently preserved in analogue format; for example, the record is not severely mouldy or damaged in any other way so that it cannot be used or preserved. Furthermore, the material of the record cannot be self-destructive. As far as the other criteria are concerned, perhaps the most commonly referred to is age: documents dated older than 1921 are kept in an analogue format. The criterion is not mandatory, which means that a record that dates from 1921 or later may be preserved in analogue format despite digitising if another criterion of the preservation format's cultural-historical value is fulfilled. There is a total of ten criteria including special material features, information (which the digital manifestation cannot convey), artistic or aesthetic qualities and rarity or uniqueness.⁴

However, it is worth noting that preservation in paper form does not necessarily mean that the original form is preserved completely. The decision on the storage format is therefore not strictly museological or related to the object value; instead, it can be related to paper quality, text with markings and standard forms (stamp, handwriting, form, ink, etc.).

In accordance with this principle the valuation criteria will possibly be refined so that only a certain part of archival units will be retained in their original form, and those whose digitisation would not be possible except by untying the bonds into loose sheets (by damaging the original structure of the archival unit) will be converted into loose sheets required for digitisation. In this case, culturally-historically valuable materials would be stored in paper form but not comprehensively as original objects.

³ For example, some archival volumes can be seen as exemplaries of a certain period and therefore will be preserved. In some cases, cutting an archival volume into loose sheet could lead to loss of information.

⁴ Cultural-historical value criteria (in Finnish): <<https://kansallisarkisto.fi/documents/141232930/149322637/Kult.+hist.+arvon+kriteerit+ja+soveltamisohje.pdf/5c4a7577-17d3-4ccc-5c0b-41514e3690d8/Kult.+hist.+arvon+kriteerit+ja+soveltamisohje.pdf?t=1705666066988>> (last accessed 23 April 2024).

A fairly large proportion of the materials held by the National Archives are of cultural-historical significance but only a very small proportion of the materials held by the authorities. It is worth noting that the cultural-historically valuable materials that are currently being digitised are in practice stored both in analogue and digital permanent forms.

Some of the documents previously ordered to be preserved in analogue format can later be ordered to be archived in digital format only. In addition to the materials in the possession of the authorities this principle applies primarily to materials already in the possession of the National Archives, the total amount of which is about 240 kilometres of shelves at the moment.⁵ Digitisation has been done in the National Archives for more than twenty years but these materials have not been destroyed afterwards despite legislative changes in recent years. The main reason for this is probably that the quality of digitisation has not been monitored in the same way as in digitisation aimed at disposal because the aim of digitisation was to produce digital copies of the actual analogue documents to be stored for customers. Currently, the National Archives' own digitisation (retroactive digitisation) implements mainly the criteria for digitisation aimed at disposal in terms of image quality and file structure but content quality control (validation) has not yet been implemented as systematically as in the production of mass digitisation.

In practice only few archive series digitised in the National Archives have been destroyed after digitisation. However, no principle prevents the updating of the decisions on the preservation forms of materials held by the National Archives, that is, the decisions on the permanent preservation of materials in digital format. Before we can move on to this the quality control methods of retrospective digitisation and probably the whole implementation model and process of digitisation must be refined and combined with the corresponding methods and processes of mass digitisation. Increasingly decreasing storage space urges development in this direction but the lack of space could be alleviated more quickly by identifying materials that are not expedient to preserve at all.⁶ Regrettably, the National Archives has received interim archive material or documents whose permanent archiving could be challenged retrospectively. Of course, these measures would require additional resources for the identification of materials and practical appraisal work; nevertheless, it should be noted that a fairly significant part of the archive material held by the National Archives is dated before 1921 (cautiously estimated at 30–40%) and will therefore be preserved in analogue format in any case.

⁵ Estimated in 2022.

⁶ It is challenging to get funding in the state budget allocated for archive storage facilities. The NAF's latest archive facility was inaugurated in 2017 and free storage space is already becoming scarce.

To summarise, the analogue manifestations of materials to be transferred to the National Archives and converted into digital format are the primary objects of disposal in the current digitalisation planning.

2 Conditions and Terms for Disposal

Disposal-oriented digitisation (a digitisation process specifically designed for the disposal of digitised analogue documents) has been implemented in the mass digitisation production of the National Archives since 2019. It is not only about constructing a production and implementation model of digitisation but also about the administrative guidance required for it, preparing the storage form described above and designing a plan for the documents to be selected for mass digitisation (selection of archive collections/series and the order of digitisation). For the sake of clarity it is stated once again that the National Archives stores documents in digital format and thus disposes of only such analogue objects that are permanently stored. Permanently stored documents refer to documents in the archiving stage in accordance with the Finnish Archives Act, the Act on Information Management in Public Administration and the EU General Data Protection Regulation. However, it should be noted that such documents can be transferred to the National Archives and digitised before they enter the archiving stage. The documents must be digitised in accordance with the National Archives' regulation concerning the requirements for the disposal-oriented digitisation. These are, in a way, the starting points and basic prerequisites when starting to look at the disposal of analogue materials.

At present the National Archives is running a pilot of digitised analogue documents' disposal process. Why has it not been done before – indeed, why was disposal not started at the same time as other processing stages of digitisation aiming at disposal? Mass digitisation had to be piloted at the very beginning and it was necessary to ensure that the production would form compliant transfer packages that meet the quality requirements for disposal. We also wanted to ensure that digitisation proceeds at approximately the set speed so that it covers the entire dataset precisely.

Disposal was not prevented by a lack of a long-term storage system required for permanent storage. The KP-PAS service (long-term storage system for the cultural-historical objects), which is a ministry-managed system and therefore external to the National Archives, has been available before and since the mass digitisation was launched the National Archive's own electronic archiving system SAPA was developed between KP-PAS and digitisation production (the digitisation system was directly integrated to SAPA at the end of 2020). SAPA has been found to be highly functional and reliable; in practice it receives the digital materials produced by

digitisation and creates the necessary user copies to digital services intended for customers through interfaces in accordance with the information system model. From the point of view of digitisation SAPA is one of the guarantors and metrics of quality as it only accepts transfer packages that comply with the requirement specifications. In other words, the introduction of SAPA has in its way created a better starting point for launching disposal.

Waiting for customer feedback as well as seeing and evaluating the end product in the customer interface have been very important reasons for the delay in launching the disposal. Mass digitisation began before the required changes that would allow the user to view the materials or images had been properly made to the customer system. ASTIA (customer service developed to search and view all the documents transferred in the NAF) allows you to not only view the material but also search for content using the OCR content identification search carried out in mass digitisation. The challenge has been that the database information (e. g. latest archive material available in digital form) is not updated to the ASTIA system in real time but with certain intervals. The difference between completing the digitisation and updating the ASTIA database may have been up to several months. Content recognition is updated as a separate entity and there have also been delays. In addition, the OCR search had to be limited to only concern the archive unit (not, e. g., the whole archive). The quality of the recognition has not been scrutinised efficiently so far either (there is an AI development project going on at present to improve the reliability of the identification). For these reasons it has been difficult to assess the quality of the content identification of digitisation and the reliability of digitisation itself through wider datasets. After all, digitisation production software has no system available that can easily and comprehensively cover the resulting image sets. Images are reviewed in scanning and validation applications but only in limited batches and momentarily. It is somewhat paradoxical that even the employees working with mass digitisation have not been able to comprehensively see the digital materials they have produced.

For the same reason we have only very recently been able to receive feedback on the 'end product' – the images – from the customers themselves. However, there are also other reasons for this: in mass digitisation the material to be digitised is mainly still in the preserving stage and as a rule contains information that is restricted for use on various grounds.⁷ Therefore, it is not freely available but principally only accessible to identified users of the official organisation that handed it over. Even after the materials have become more available in the ASTIA system, their use and the evaluation of quality has been dependent on how the relevant

⁷ See footnote 2.

authority needs to use these materials. The situation is therefore different from that of completely public materials where customers typically provide the National Archives with immediate and abundant feedback on deficiencies or quality deviations.

This does not mean that the rate of use was low. On the contrary, the rate of use of the mass digitisation materials has been steadily increasing, which is logical in the sense that mass digitisation seeks to identify and include materials that are frequently consulted. Nevertheless, it has been necessary to wait several years so that customers could have the opportunity to give feedback on the quality of digitisation before disposing of the original analogue materials.

Finally, the most important reason for the delay in launching the disposal has been that the disposal itself requires its own process – both logistically in terms of the implementation of the actual physical disposal of the analogue material and in terms of the related information system functionalities. The first years of mass digitisation have focused on building steady, stable and reliable digitisation and there have been no resources for disposal process issues; disposal has not been a main priority. It has also been considered that disposal – as an extreme, irreversible measure – requires a certain maturity of the organisation and a deeper understanding of the fundamentals of its activities, although disposal like any process or part of a process inevitably requires a certain level of experimentation and piloting at the initial stage. It is not an area of digitisation that one would want to implement more generally on digitisation production that is in the start-up or pilot stage. Ultimately it is a matter of assessing and identifying when the time is truly ripe for activities such as disposal.

The development of recent years has forced us to reflect on this question and brought new perspectives to its evaluation. The National Archives' storage facilities are filling up and at the same premises could be vacated. In the future there will therefore probably be less storage space available although the authorities still have at least a few hundred kilometres of documents in their possession that will be transferred to the National Archives (mass digitisation-oriented and others). By the end of 2023 more than six kilometres of materials will have already been digitised in mass digitisation. As the storage space decreases the pressure to dispose of materials that have already been digitised is continuously increasing. There is no actual justification for preservation as the use of the mass-digitised documents has genuinely moved to digital forms.

Of course, the launch of the disposal is also related to the public image of the National Archives and in particular the mass digitisation function. The key difference between mass digitisation and traditional digitisation is precisely the disposal of the original format. As long as it is not done the mass digitisation is not materially different from other forms of digitisation.

In order to start the implementation of the disposal it is necessary to first define its conditions. These are the so-called criteria for the disposal of mass digitised analogue documents. Each criterion must be fulfilled before the dataset can be considered to have been properly disposed of. The following briefly describes the conditions for disposal in the order of the digitisation process.

2.1 Ensuring Storage and Preservation Format

Mass digitisation contracts have been signed. National Archives and the transferring authority have drawn up a mass digitisation agreement for the materials. The agreement is standardised and will only be drawn up for datasets whose permanent preservation value and permanent preservation in digital format has been confirmed. On the basis of the agreement the National Archives agrees to be responsible for the digitisation of the materials in accordance with the criteria for digitisation aimed at disposal and on the other hand receives permission from the authority to dispose of the analogue materials after ensuring proper digitisation without a separate notice or permission. As part of the agreement the materials are also specified and a joint risk assessment is usually performed, which must be taken into account when digitising and disposing of the material.

2.2 Metadata

The metadata must have been fed into the AHAA (the NAF's cataloguing metadata system) and been approved by the NAF and the 'possession status' must have been changed from the governmental agency to the NAF (for further measures). In practice the only action required from the authority is to add metadata on the content of the materials into the AHAA; this is the authority's task as it knows the materials. However, it is also important because the authority decides and is responsible for the order of the materials and the metadata related to them. The National Archives does not change the order or metadata in any way at any stage of digitisation. This principle is important in terms of the integrity, originality and authenticity of the materials. The authority also prints out the barcodes it generates from the AHAA database and attaches them to the archival units. The materials are monitored during digitisation with the help of barcodes; they also help to ensure that the order is maintained and that the materials are not lost during the process.

2.3 Reception Inspection

After the metadata is approved the materials are transferred to the National Archives. NAF checks and compares the analogue materials quantitatively and partly in terms of their content with the information presented about it in AHAA. When the inspection is complete the approval certificate is delivered to the authority that transferred the records. If the shelf and database information differ the necessary modifications are made to the database or the materials forgotten from the transfer are submitted to the National Archives.

2.4 The Contents of the Digital Materials Have Been Verified

Mass digitisation uses a validation programme built by the National Archives' application development unit. The validation programme reviews 100% of digitised material. In practice each image is checked by the operator (in the near future, at least some of the materials will be pre-screened using AI-supported API components). In order to ensure integrity, originality and authenticity it is not enough that the actual data content of the image (document) is readable – the digital object must perfectly correspond to the visual recognition of the original analogue document. The image file must not include any elements that are not included in the analogue counterpart. It should be noted that the image quality is also not improved with AI or photo editing solutions even if the original document is in poor condition (for example, if it is a photocopy or a sheet where the text is hard to read).

The aim is to ensure that the front and back sides of the sheets are scanned by using optical readers to identify possible double and error inputs and by carefully monitoring the image queue during scanning.⁸ We mean to identify and mark any small-sized labels and loose papers already at the material's preparation stage before the scanning stage. Overall, it is a question of staff training and experience, that is, professional skills. A professional operator can identify any missing sheets even in the image queues during validation.

⁸ When purchasing equipment, preference is given to equipment that has a higher performance in separating documents from each other.

2.5 File Structure Approval and Transfer

The file structure and usually the technical process of image creation are based on the National Archives' regulation on digitisation aimed at disposal, which has been drawn up taking into account the standards commonly used in the archives sector and the quality requirements for digitisation of other national archives. Additionally, the acceptable formats and the package to be formed in the digitisation process have been defined with the national requirements set by the long-term storage services (KP-PAS) for stored materials in mind.

When scanning materials only equipment whose performance meets the requirements of the aforementioned regulation – both in terms of manufacturer specifications and actual tests – is used. The performance of the equipment is optimised with the quality of the image files verified on each production day using measurement tables designed for it (UTT target). The test is repeated as many times as necessary until the requirements are passed (if not, the equipment is subjected to the necessary cleaning or repair procedures).⁹

It is perhaps more important than the technical specifications of image files that the image files are in lossy compression format (JPEG) and that they are not processed and saved again after the first processing. Content identification is performed in ALTO format and metadata concerning the image file's creation history is presented in accordance with the MIX meta scheme. It is very important that digitisation does not only produce high-quality images but also stores key information about the digitisation process itself. A transfer package compliant with the requirements includes the master file, the OCR file (XML-ALTO) and the MIX metadata files. The transfer package is in TAR format (no compression used). Each batch has a separate JSON file containing batch-related information such as transfer type and TAR package checksum which is made by MD5 hashing. After a successful SAPA transfer the original files are automatically deleted from the Mass digitisation servers (but only after a short retention period).

The verification of the technical characteristics of image packages is fully automated in the transfer application devised by the National Archives' product development. Images will not proceed to later processing stages such as content identification if the figures in the MD5 inspection do not match. The TAR package is checked when transferring batches to the SAPA system. In a way, there are three verification steps because the information is also checked when transferred to the KP-PAS long-term storage. It is worth comparing the number of inspections of the

⁹ The quality is measured with iQ-Analyzer software: Metamorphose extra-light standard with few adjustments.

digital material with the corresponding number of inspections of analogue material, for example when transferring materials traditionally from an authority to an archive – it is clear that digital materials are inspected much more comprehensively than analogue materials. Although there are always risks with the transfer and processing of all materials the risks can be considered smaller thanks to the comprehensive inspection of digital materials.

2.6 Metadata Life-Cycle Information

As part of the technical package's transfer, the information concerning it must also be stored in the metadata system. SAPA transmits information to the AHAA system about the successful transfer of the materials to the KP-PAS long-term storage.¹⁰ What is most important is that information regarding the digital manifestation must be connected in the archival unit in the AHAA metadata system.

2.7 Image Content Inspection (Part II)

In the event that the materials have not been used significantly by the customer random sampling will be used to check that all the materials are in the ASTIA system after the end of digitisation in the mass digitisation process. This constitutes the third content inspection stage after scanning and validation. This sampling inspection has been used for a long time in the process of mass digitisation and it is possible that when the disposal starts, the scope of the sampling will be further increased.

2.8 Retention Period

The materials will not be disposed of directly after the completion of the digitisation, but after a retention period. Its final duration is still to be determined but for the time being the minimum retention period is set to six months. At the system level the retention period will be connected to the content inspection of the materials in the ASTIA system after the digitisation is completed.

¹⁰ PAS-AIP-ID (storage package identifier in the KP-PAS system).

3 Piloting the Disposal

Before the actual regular disposal operations begin the operations are tested by piloting the planned disposal implementation model and ensuring that the changes it requires for the information system function as intended. The pilot focuses on assessing whether the metadata of the life cycle of the materials and information on the measures of disposal itself are saved. It is important to gain experience on how easily manageable and controlled the launch and control of disposal is from the point of view of using the application controlling it on one hand and on the other hand as a logistic entity (collection of materials and physical disposal). The effectiveness of the operations and the appropriateness of the implementation method are also evaluated. As a whole the disposal pilot evaluates whether the outlined implementation method is such that it works in regular and significantly larger-scale disposal than the pilot's scope. The aim is to dispose of an estimated 1–3 kilometres of shelves of material per year over the next five to ten years; the implementation model must therefore not only be very simple but also accurate and reliable. Some areas have already been identified that can only be developed and implemented after piloting. It is clear that a fully completed model cannot be created all at once; the process will be developed slightly using the Deming circle model (Plan – Do – Check – Act). It is essential to create a reliable system on top of which more functionalities and layers can be built in the future.

Disposal requires changes in key metadata systems and information systems of digitisation production. The changes that are to be implemented as part of the disposal pilot are briefly described below.

Disposal is built around the logic of archive series defined in a certain way. In the AHAA metadata system the archive structure is usually based on the traditional archive hierarchy levels (collection – series – archive units). However, documents related to the same archive in AHAA may have been transferred to the National Archives at different stages and at very different times. Before the digitisation can commence the metadata from AHAA must be transferred further to the Logistics application of mass digitisation which is a NAF in-house solution and serves as the central system for material management in digitisation. Disposal constitutes one of the logistical processing stages of logical data so the disposal will be guided mainly from the Logistics application. As the archive series and units in the AHAA system and the logistics application are therefore not completely identical there are risks associated with the exchange of information between them.¹¹ As part of the

¹¹ For example, an archive series may include several thousand archive units. Some of them may have been digitised earlier. Therefore, not the whole series but only the metadata of those units which have not yet been digitised will be transferred into the Logistics application.

development of the processing steps for disposal there will be a new metadata level created, the Release Batch that allows the documents to be processed in the form in which they were transferred to the National Archives, from the digitisation and all the way to disposal in future. In practice, after the National Archives has received and inspected the documents, information of the Release Batch (a Release Batch tag number) will be added to the documents using their unique ID information in the AHAA system. After this the data can be collected and transferred to digitisation systems quite easily and automatically.¹²

In the Logistics application disposal is controlled by the Disposal Batch entity and the desired Release Batches (one or more) can be tagged to the Disposal Batch. When the analogue documents are completely disposed of this stage will be recorded in the Logistics application. From there the information on the Release Batches included in the Disposal Batch will be sent to the AHAA system where information on the life-cycle event of disposal will be automatically recreated in the documents under the respective Release Batch ID and the analogue storage units linked to the documents will be deleted.

The permanent and up-to-date information on the documents – such as the fact that the unit only exists as a digital manifestation and its analogue manifestation has been destroyed – are thus in the AHAA system but the more detailed stages of the disposal itself and the actual control are included in the Logistics application. There are several processing stages for disposal where the documents to be destroyed are defined and assembled and where the documents are determined for different processing steps of disposal.

In general materials have several stages or ‘statuses’ in the Logistics application as soon as they are retrieved from the AHAA system. In the first stage, for example, the documents are marked with the status ‘Waiting in the repository’. Information is mainly updated on the basis of the barcode reading and partly by using automation. Once the digitisation is completed the documents have been validated and transferred to the SAPA system through automated processing and further to KP-PAS long-term preservation they go through the stages ‘Digitisation completed’ and ‘Returned to repository’ in the Logistics application. Once the contents of the documents have been checked again in the ASTIA system they will be transferred (as Release Batches) in the Logistics application to the ‘Retention period in progress’ tab. Here the documents can be viewed as a whole, showing which materials will

¹² For example, a transfer to NAF could include fifteen archive units to ten series in the collection which would have been created earlier in the AHAA and include other units dating to previous transfers. These fifteen units in ten series constitute the Release Batch which will be delivered to the Logistics application to be processed in digitisation and finally in the disposal stage.

be transferred to disposal. When the retention period expires the application automatically transfers the materials to the 'Ready for disposal' tab. Here, the materials can be transferred to disposal at the desired time.

Therefore, when the disposal is initiated a Disposal Batch will first be created in the Logistics application. Then all the needed Release Batches as well as the information of the premise from which the disposal takes place are tagged to the Disposal Batch. When these selections have been confirmed the application automatically transfers the materials to the 'Delivered for disposal' tab.

The steps described above take place solely at the information system level. When the documents have been checked to 'Delivered for disposal' their bar codes are also physically read in the storage facilities. This is to check that all units the system has collected for the Disposal Batch are actually found in the warehouse.

When the result of the barcode reading is 100%, meaning that all the units listed in the application and the barcodes read in the warehouse correspond to each other, the documents are automatically transferred to the tab 'Transport'. After this the archive units (or physical unit boxes) are stamped with the 'To be destroyed' stamp.

You can now choose the desired units for transport from the documents in the Transport tab. Of course, the number of transports depends on the material volume; typically about 100–200 metres of documents (or appr. 1250–2500 archival units) are transported for disposal at one time and transport continues until the selected release batches have been completely destroyed. Once all the materials have been checked to the 'Transport' tab, the respective disposal batch is automatically transferred to the 'Destroyed' tab.¹³

The 'Destroyed' tab lists all the Disposal Batches and the disposal entities within that have been delivered for disposal. The tab contains information on the Disposal Batches, Release Batches, metres of shelves in the Disposal Batch, the date and time when the last transport has been recorded as complete as well as the date and time when the Disposal Batch has been recorded as destroyed.

The Disposal Batch is recorded as destroyed manually when the National Archives has received a certificate of disposal from the disposal facility. After that the documents of the respective Disposal Batch can be recorded as destroyed from the drop-down menu intended for that purpose. The Logistics application sends the AHAA system a notification to delete the storage unit information (physical storage media) of the disposed materials as well as the information regarding the analogue manifestation. This will also update the data in the ASTIA system where the docu-

¹³ The application also has a hidden intermediate stage between Transport and Destroyed for monitoring the material's progress during transport.

ments can be consulted – available only in digital format from then on. The AHAA system is also connected to the National Archive's storage management system (AHJ) from which the facilities so far reserved for Disposal Batches are released for other uses and are added to the amount of empty shelf space after the disposal of the documents.

To summarise, the stages of disposal have been modelled very accurately but the stages progress clearly and logically and can be used easily in the Logistics application guiding the disposal. Importantly, all processing steps are verifiable and always apply to the entire Disposal Batch, that is, one or more Disposal Batches, and never to individual archive units. The commands to move from one stage to the next are manual but the information is updated collectively and automatically (individual series or other hierarchical levels do not need to be clicked, selected or opened).

The physical method of disposal of the materials will most likely be incineration and further use as an energy source. The same monitoring model based on barcodes as is used in the information system will be applied as much as possible also in logistics; trolleys or other means of transport by which the materials are transported to disposal will also be bar-coded and the materials will be included in them. Likewise at the incineration plant, the materials can still be registered for actual disposal, which can also verify that the transport has delivered them as planned.

The piloting of the disposal itself will be carried out in three to four stages. First, we ensure the functionality of the changes made to the systems. The disposal is therefore simulated with a database duplicated (cloned) for this purpose, which is used to practise the disposal without transferring or reading the barcodes of the actual materials. Thanks to cloning disposal can be tested in a real production environment without disrupting the actual simultaneous digitisation production. If everything goes according to plan and no disturbances occur the first Disposal Batch is disposed of with a single Release Batch. The first stage also models different ways to calculate and place archiving units in the means of transport in the best possible way.¹⁴ In the next stage the Disposal Batch is assembled from different Release Batches where the system's ability to manage the compilation of data from different materials into one Disposal Batch is tested in the real environment, with several transports also carried out. In the third stage disposal is carried out with a different transport model and the size of the Disposal Batch is increased so that it includes several Release Batches and is carried out in several different transports. In a possible fourth phase material monitoring will be introduced during transport and at the disposal facility.

¹⁴ The storage warehouse from which the materials to be disposed of are collected only stores materials that have already been digitised and which are to be destroyed after the digitisation.

After the pilot we will assess the need for further development, specify the risk assessment and immediately start to model the additional measures required for regular disposal. The significant increase in volumes should mean at least more specific solutions for the implementation of transport and processing of physical materials (for example, intermediate storage of materials waiting for disposal transport).

4 Conclusion Remarks

In conclusion, the National Archives is enthusiastic but also aware of its responsibilities and risks regarding the launching of the disposal process. As with all handling of documents (in this case containing a large amount of personal data) there are risks, which have been estimated, and we will continue to analyse and specify them after the pilot as well. However, the starting point and foundation of everything is high-quality digitisation – that the documents truly exist in a reliable digital format of good quality. Another important part is to design the information system that directs the disposal so that the number of human errors is as minimal as possible – nothing can happen ‘accidentally’ and the number of stages, measures and information to be monitored is moderate. Of course, the use of the system requires good and sufficiently long training – disposal will probably be one of the few stages of digitisation that cannot be carried out by all operators but rather only by persons trained and authorised to do so. The third important element is to ensure the data protection of disposal. It is also the only stage of disposal for which the National Archives is not solely responsible but due to more operators its risks are also greater.

NAF will move on from the disposal pilot to actual regular disposal only after it has been ensured that everything works. We in the NAF believe that the launch of disposal as the last step of the digitisation process aimed at disposal will further increase the interest of the authorities and other organisations in obtaining the digitisation services of the National Archives.

