Co-financed by the Connecting Europe
Facility of the European Union

DELIVERABLE

EKA Digital Preservation Plan

Document Author(s)

Name	**Role**
Karin Oolu	EKA Records Manager
Andres Uueni	EKA Researcher
Anssi Jääskeläinen	XAMK Research Manager
Stephen Mackey	PIQL

Document Approver(s)

Name	**Role**
Radoslav Jakub	HaDEA Project Adviser

Digitalia Research Centre for Digital Information	EKA Estonian Academy of Arts
DISEC	piql

Contents

1. Introduction

The following plan describes the preservation of the digital resources of the EKA collections Electronic Records Management System (ERMS) Webdesktop and Content Management System (CMS) Digiteek. Due to the rapid growth and significance of digital resources, it is important that the authenticity, reliability, and long-term availability of these materials are ensured.

Digital preservation ensures that current and future academics, researchers, and students, as well as the management of the academy, may access these digital resources for the long term. The current plan provides strategies and action steps that comprise digital preservation by established best practices. The plan also describes the need for digital preservation, which requires more active management than traditional physical preservation. Digital objects are fragile as continually evolving hardware, software, standards, and file formats increase the risk that they become corrupted, inaccessible, obsolete, or lost. The plan also includes information about migration statistics from the deliverable M2.16 EKA Collections Migration Plan (FINAL).

The primary audience for this plan are those directly responsible for caring for digital materials in EKA, also including the management team to support investment decisions. The plan should be followed for any digital content that should be preserved in the long term. Having a plan, and regularly updating it, will increase the sustainability, accountability, and flexibility of digital projects.

1.1 **Preservation policy in EKA**

No single policy or strategy document can stand alone, so to achieve support for and successful implementation of a digital preservation plan, it is essential to embed it in the broader context. There is no specific digital preservation policy document in EKA that provides guidance and authorization on the preservation of digital materials to ensure their authenticity, reliability, and long-term accessibility. Instead, a range of policy documents already exist across EKA covering a variety of issues such as staffing, information management strategy, and financial plan. There are also a number of policies relating to more specific issues of records and EKA museum collection management plans that are relevant to digital preservation activities. It is essential to consider both the content and established style and structure of all relevant policies within the EKA as well as how digital preservation policy will fit within the wider landscape. Moreover, a policy should explain how digital preservation can serve the major needs of an institution and state principles and rules on specific aspects in accordance with the law. In the context of Estonia, EKA is guided by the following laws: Archives Act, Public Information Act, the Museums Act, and Principles for Managing Services and Governing Information. The main standpoints and principles that frame digital preservation in EKA are stated as follows:

- Scope: Provide access to all digital assets for an indefinite period unless there are restrictions from the legislation.
- Operating Principles:
  - Digital content will be **created with preservation in mind** (populating metadata for the content, where possible).
  - Preservation copies of material will be made on a **regular basis**.
  - An automated process will be used to create basic preservation metadata as preservation copies are made.
  - Preservation copies will be kept on multiple forms of physical media, **in multiple locations**.
  - Preservation copies will be periodically **verified for data integrity**.
  - Preservation copies will be **refreshed to new media on a regular basis**.
  - Preservation copies will be migrated to **new formats as required**.
  - Access to preservation copies will be limited to **specific staff/position**s.
- Selection & Acquisition
  - If a digital asset is unique and falls into any of the digital inventory lists then it should be preserved indefinitely.
- Challenges & Risks
  - The greatest challenge is related to limited resources.
  - There is a need for training.
  - In terms of risk, the greatest concern is regarding content that resides exclusively on read/write CDs and other access media. These need to be moved to a more permanent physical carrier.
  - There is also a risk of the unknown: currently, no standard exists for digital preservation in EKA, and thus any solution chosen, would involve a degree of uncertainty.
- Financial Sustainability
  - Although the proposed solutions require sustainable financing, finding additional qualified support staff is important.
- Technological & Procedural Stability
  - **This section will be completed in the future, once digital preservation technology has been selected.**
- System Security
  - **EKA will ensure a complete and accurate record of archived digital material by maintaining multiple copies on disparate physical media, in separate locations.**
- Procedural Accountability
  - An external audit may occur if requested (e.g. ISKE audit[1]). EKA is committed to periodic spot checks of archived digital content and will use **checksum software** to further verify the integrity of preserved content.

The digital preservation plan is based specifically on the aforementioned points and follows the schema of Becker et al, 2009[2].

[1] ISKE audit: https://www.ria.ee/sites/default/files/content-editors/ISKE/iske_audit_manual.pdf

[2] Becker, C. et. al. (2009). Systematic planning for digital preservation: Evaluating potential strategies and building preservation plans. Int. J. on Digital Library.

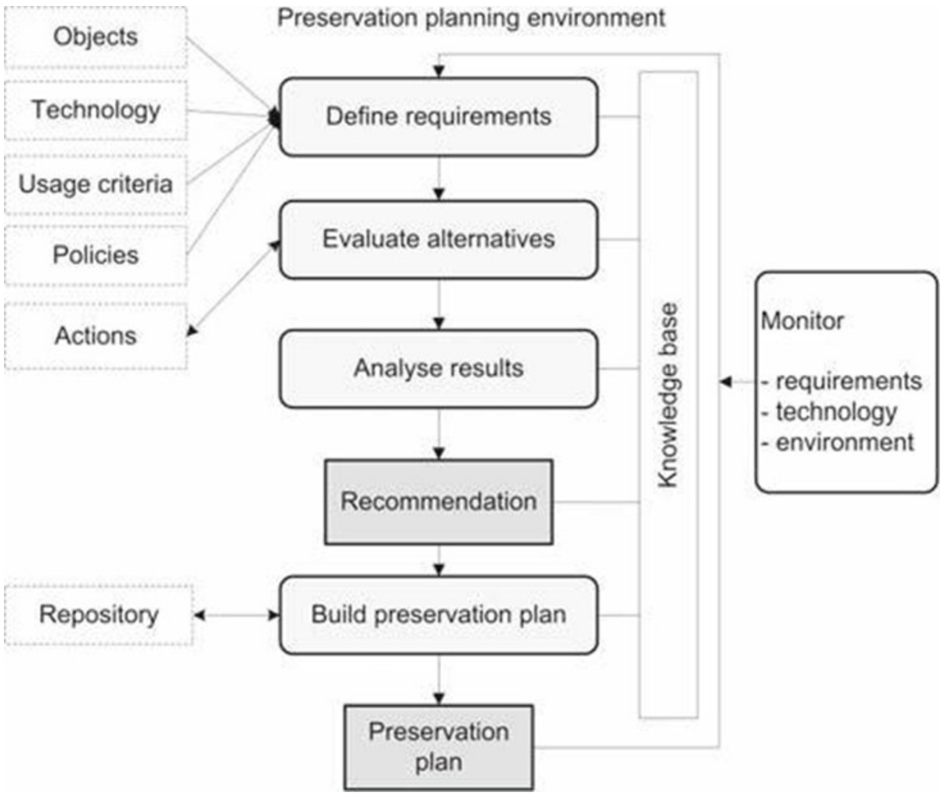

Figure 1. Systematic planning of digital preservation

Regular updating of the preservation plan increases the sustainability, accountability, flexibility, and trustworthiness of digital content. The schedule for review/update of the Digital Preservation Plan in EKA is agreed to perform annually.

1.2 **The main challenges of long–term preservation**

Due to the fragile nature of digital materials along with continually evolving hardware, software, standards, and file formats, there are recognized challenges in implementing an effective digital preservation plan.

A long-term preservation plan combines several strategies and actions to ensure long-term access and authenticity of digital content despite the challenges of technological changes and media failure. For long-term digital preservation, it is vital to identify the bottlenecks at the organizational and technical levels.

File storage - including information about primary and secondary storage solutions, identifying the responsible departments, sustainable funding, also backup schedule, disaster plan, technical support, and identifying the versions of files.

Long-term systematic and sustainable preservation are different components of the preservation plan and are highly dependent on each other: file integrity and file access.

File integrity - including workflow mapping and identifying responsible departments/positions, how to ensure authenticity, security, virus checking, and managing technical 'triggers' for data migration and verifying.

File access - including identifying responsible departments for access, information about metadata structure, metadata standards followed and levels (administrative, descriptive, technical), file types and versions, data migration and conversion procedure description, and storage and file type migration interval.

> *A key consideration for tools is to locate where they sit on an overall workflow which is presented in Figure 2, so considering and mapping out the entire workflow helps in selecting tools. In addition, EKA as an organization has multiple workflows that may have different requirements which might conflict in some way. Describing a workflow provides a basis for anticipating difficulties and can provide a roadmap for ongoing development. It is important to follow the workflow described in the following chapters to solve possible bottlenecks, possible tools to support the current preservation plan can be found at the Community Owned digital Preservation Tool Registry COPTR.*

1.3 **Institutional commitment**

The following resources are needed to successfully implement a digital preservation plan.

1. **Dedicated staff time:** Dedicating time to specific staff to the practical work of digital preservation is necessary for the digital preservation plan to succeed in the long term. This may include hiring additional staff or reallocating staff time mostly in regard to the EKA Digiteek collection. This also includes time for training and adapting current workflows to incorporate digital preservation standards among staff in all academic units.

2. **Dedicated funds and other resources**: A financial commitment to support long-term digital preservation actions is necessary. This includes the maintenance, acquisition, and management of technology to support digital preservation activities. The implementation of the new software for EKA Digiteek collection will initially require more investments.

3. **Digital storage capacity**: Digital resources must be stored in a manner that is consistent with accepted best practices. This includes technical infrastructure such as hardware, software, network access, data backup, and facilities. Best practice in digital

preservation requires duplication of digital objects in local systems and means continuous assessment and monitoring of EKA storage capacity.

4. **Support/cooperation/communication**: The importance of collaboration between all stakeholders to successfully accomplish a digital preservation plan.

2. Digital preservation workflow steps to consider in EKA

The digital preservation workflow steps to consider in EKA will be described according to the following schema, starting from the bottom up.

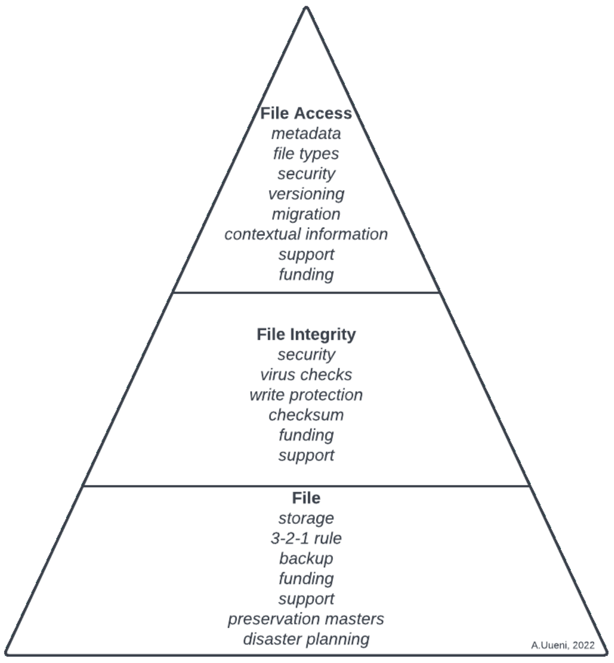

Figure 2. Digital preservation workflow in EKA (Uueni, A.)

2.1 **File storage**

It is crucial to identify all file storage in EKA, also responsible departments and options, focusing on WD and Digiteek. The next standpoints must be considered to ensure sustainable storage capacity in EKA in the future:

a) **Responsible** positions and schedules for storage and backup of digital files must be determined.
b) Due to the expansive nature of EKA collections, we face challenges securing sufficient long-term digital storage space for digitized and born-digital materials. As the amount of information is growing, the storage space requirements will increase over future years. It should be considered to follow the 3-2-1 rule (3 copies of data, stored on 2 different types of media, 1 in a different geographic location).
c) **Backup schedule**: It is critical to ensure that files are to be backed up to the three storage steps (daily, weekly, monthly, etc.). The responsible department for backup is IT.
d) **Funding for storage**: The funding for purchasing storage media must be decided as the cost for storage is regular. The cost of new equipment depends on the infrastructure and technical solution. E.g. One disk system with 50Tb of storage can cost 50k€ while another with the same space but more feature/more capable can cost 150k€ if done on-premise.
e) **Technical support for storage**: It is critical to pay attention to positions responsible for the setup and maintenance of storage media.
f) **Versions of the files** (preservation masters, access copies, derivatives): It must be specified what versions of files will exist, and then which of those versions will be backed up according to the preservation plan. It is critical to pay attention to this in the preservation plan now, as only preservation masters are currently preserved in EKA (Appendix B).
g) **Disaster planning**: A short outline of what will happen in the case of disaster must be detailed: types of disaster that can occur can include: natural, human error, or media failure. Then dealing with a disaster will be described: who will be responsible for damage assessment and recovery, how files will be recovered and within what time, what are the procedures if the files cannot be recovered, etc. This type of information should be thought out, tested and then included in EKA’s overall disaster plan. Also, EKA needs to create a document or tool which allows following the performance on a regular basis.
h) **Digital trash/ green/sustainable:** EKA as one of the Estonian universities is concerned about its impact on the environment. Keeping digital content viable requires not only energy use, but also refreshing the digital storage media and technologies, it includes topics like Data Storage Materiality, E-waste, and Mitigating the Environmental Impact of Audiovisual Digital Content. It is important to follow the market trends and available solutions, e.g. energy efficient close-by data centre solutions. In addition, spinning disk compared with tape solution is several times more costly and

energy consuming. However, there is still no conclusive research about the environmental impact of long-term digital preservation, however, there are several guidelines[3] and surveys/reports[4].

2.2 **Long-term file integrity**

The content of digital files must also be protected for long-term preservation. Digital data can easily become corrupted, erased, contain errors, or become infected with malware which would compromise its accessibility. In this plan, the methods for regular file integrity monitoring should be outlined, also the replacement or correction of any detected file errors or degradation over time. Commonly, this integrity monitoring with appropriate actions is automatically handled by the digital archive. Otherwise, a common approach is to calculate a simple hash value from the file and compare it to the originally calculated hash value. If the values mismatch, the file is either changed or corrupted.

a) **Responsible** departments and positions in EKA must be determined.
b) **Security**: Responsibility for the security of digital files with access rights to move, delete, or change digital files is currently distributed. Detailed regulations through digital policies, procedures, and action plans should be settled. In EKA this might include everything from access points, and passwords used on office computers to digital storage media being kept locked in a secure area.
c) **Virus checks**: Any kind of vulnerability must be managed and controlled. Regular virus checks run on all office computers/server side should be enforced. The position responsible for fixing a virus that is found should be settled.
d) A non-networked computer is necessary for receipt of file transfers of donated digital materials. For example, flash drives are vulnerable to viruses and must be processed carefully.
e) **Checksums** demonstrate authenticity and that no files were changed when copying from storage from one media to another. Oneclick SIP creator adds, according to the SIP specification, an SHA-256 checksum of each included payload file to the package. This way it can be ensured that files were not altered or damaged during the transfer. Also, the metadata file containing the checksums is checksummed so we can be sure that the package is intact.
f) **Funding**: The source of funding should be identified for file integrity steps for any software or hardware needed.
g) **Technical support**: Responsible positions for setting up, managing, and implementing file integrity hardware, tools, or processes must be determined.

2.3 **Making digital content publicly accessible**

Delivering digital content to users is of prime importance. Public audience access may be limited; some materials may also be under restriction until a later date determined by donor directives, copyright, or legislative mandates. These kinds of materials require different curatorial considerations than publicly accessible content.

a) **Responsible** departments and positions in EKA must be determined.
b) **Metadata**: It is important to identify what preservation metadata is collected for digital files, what standards are followed, and where/how is the metadata stored. Who is responsible for creating and checking metadata? There is a need for multiple levels of metadata (for example basic inventory, descriptive, technical, preservation, and administrative). Oneclick solutions harvests high-level metadata to the package-level DC.xml file including the basic file-level metadata to the representation METS.xml and stores the original metadata files as payload files. This way the most important things can be accessed directly with package exploration tools and the complete original metadata can be seen as a payload.
c) **File types**: A wide selection of various file formats is used in EKA: there is a need to unify preservation copy file types, especially for audio, video, and 3D model formats. File types for the different formats and versions of digital files in both EKA collections should be detected. Moreover, it should be decided how to instruct donors about file types for digital donations and if there is a need to convert files to standardized formats. A wide variety of materials, standards and methods exist depending on the item format. Each format requires its own set of guidelines and resources and international registers[5] should be followed, including the requirements of NAE. In general, this issue can be partially tackled by defining acceptable formats. However, in the case of art content, it may be unacceptable to restrict the artist's freedom via format definitions.
d) **File format migration**: As digital formats continue to change, a preservation plan must include provisions to convert existing file formats into formats that avoid obsolescence and promote continued accessibility. It also needs to be kept in mind that format migration will always lose or add some information or metainformation to the file. Therefore, in the case of art content, it's a suggested approach to keep the original file type along with the migrated format.
As the active file migration plan is currently missing, such an analysis is necessary to be created (e.g. 2023 GANTT).
e) **Contextual information**: There is often minimal metadata for digital files, which results in the loss of information about their content. The lack of metadata also results in the loss of background information about the files and decreases the ability to understand the significance of the material. A digital preservation plan should include methods by which digitized materials can maintain their associated metadata.
Example: A digitized physical photograph with no writing on the back features three people standing in front of a building. The metadata package that is digitally attached to that picture includes the names of the three people, the photographer's name, the building location, and the date the photograph was taken.
f) **Technical support for file access**: Responsible positions/partners for setting up and managing file access tools or processes must be fixed.

[3] Tape Storage vs. Disk Storage: Which is best for backup? https://www.altaro.com/backup-dr/tape-storage-vs-disk-storage/
[4] Greenpeace, ClickClean, http://www.clickclean.org/
[5] Example of indexes: https://www.nationalarchives.gov.uk/PRONOM/Default.aspx

g) **Funding for file access**: support digital access continuity and identify the source of funding for any software or other tools needed for file integrity steps.

In addition, if the access copy is the only copy of a digital resource, then the danger of loss from theft or damage is clearly very high. If this approach is taken a risk assessment needs to be undertaken consisting of some of the following questions

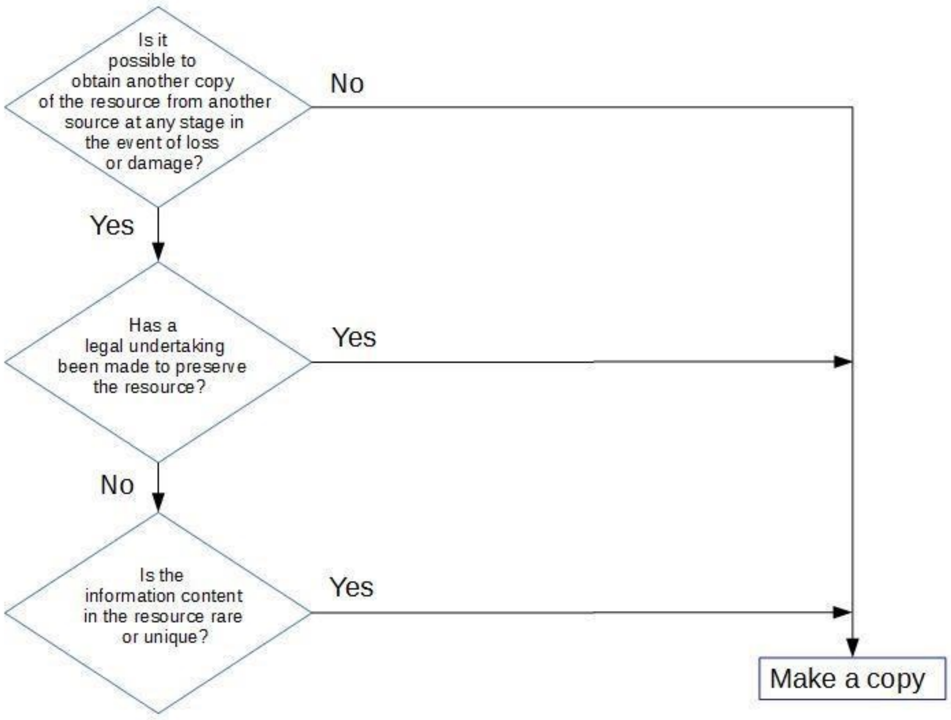
Is it possible to obtain another copy of the resource from another source at any stage in the event of loss or damage?
No
Yes
Has a legal undertaking been made to preserve the resource?
Yes
No
Is the information content in the resource rare or unique?
Yes
Make a copy

Figure 3. Decision schema (Digital Preservation Coalition)

3. EKA digital assets

The following chapter gives an overview of the main digital assets of EKA, what data is included in the main collections and the reasons, as well as challenges of preservation.

Digital resources described in the preservation plan from the EKA collections Webdesktop and Digiteek comprise a broad range of content, including digitized materials and born-digital resources. These are the EKA's two biggest collections of different varieties of materials and the need for long-term preservation. Types of digital materials include textual documents (in both systems) and multimedia (in Digiteek). Both collections, but especially Digiteek, will likely acquire materials in additional formats in the future. Therefore, preservation strategies will be designed to be format-, and type-independent to be able to accommodate new formats with the minimum possible effort.

In addition to clearly defined collections ERMS Webdesktop and CMS Digiteek, there is a lot of unclassified information around EKA. The initial research was conducted in the period of March-May 2022 to locate and classify all the digital material in four academic and supportive units[6]. The research was based on oral interviews with key academics, also with key administrative staff members. The results have revealed a large amount of uncategorized digital material which can broadly be divided into three main groups as follows:

1. Graduation thesis/ both BA and MA – partly preserved in Digiteek, partly in the cloud, partly in the Museum Information System MUIS
2. Historical digital assets of academic units (photo, video) – uncategorized in the cloud
3. Digital material evolving during the study process –uncategorised in the cloud and e-learning platform Moodle.

Analysis of research results showed the need for the implementation of the new digital asset management (DAM) software in EKA. In addition to the problem of Digiteek software becoming obsolete, the new software implementation also aims to solve the problem of storage space (there are two collections with a capacity of 3 TB already), and also organize digital assets from group 2, historically valued digital materials from all academic units. Considering the future, the new software should ensure the preservation of all records of cultural-historical value.

The overall description of the collections Webdesktop and Digiteek will follow, the more detailed statistics of both systems can be found in Annex B.

3.1 EKA ERMS collection preservation

Electronic records are consulted as proof of activity by anyone inquiring about the decisions, processes, or performances of an organization. In Estonia, the archival value of records is determined by the National Archives of Estonia (NAE) during the macro appraisal process. When evaluating documents, the NAE captures all the information generated in the process of the activities of the institutions, regardless of the medium on which it is recorded or in which databases. It also seeks to cover as long a period as possible for both existing and future documents. As a result, the list of classified information with assigned archival value will be provided for

[6] EKA structure: https://www.artun.ee/app/uploads/2022/09/EKA-struktuur-190922-ENG.png

the organization (Archives Act[7], 2011). The institutions must ensure that the digital content transferred to NAE is technically compliant, authentic, and can be archived with context. Upon another, evidential value (business value), the institution can decide for itself, ensuring that all procedural and legal obligations are fulfilled.
In Estonia, documents with archival value are usually located in ERMS together with metadata. Furthermore, until the documents are transferred to NAE, the management of metadata and file formats takes place in ERMS. In EKA ERMS Webdesktop (WD) is the system that holds the corporate management records and processes of EKA.

The records in WD can broadly be divided into two categories:

1. **Records with archival value (AV)** with a permanent retention period (according to the macro-appraisal done by NAE) will be transferred to NAE after 10 years.
2. **Records with evidential/business value** with different retention periods (the institution decides itself, ensuring that all procedural and legal obligations are fulfilled).

Although the system does not support maximum digital information management, it is sufficiently widespread both in Estonia and abroad for corporate information management. Digital content and associated metadata are developed according to the current standards and best practices, but the system is not developed with the functionality of a long-term repository.
What will be captured in the WD is declared in the information management regulations and strategies in the organization in accordance with the local legislation of Estonia, also following the macro-appraisal done by NAE. At the institutional level, these documents are critical to the life cycle of EKA's information, regulating how information is created, stored, disseminated, managed, and protected.
EKA implemented WD in 2015 and had a total of 257 active users in Mar 2022. The amount of items currently in use is 96055 (incl records/items, users, groups, and routings). The total amount of files is 169097 (53173 GB) (incl old versions, DHX registered files, and file creation templates). 101092 files (41053GB) are related to records/items. Digital signatures given during Feb 2022 were 956 (smart-ID not included).
Records are classified according to a functionality-based classification schema and there are a total of 220 different data types, of which 183 are active for new documents captured to the system. Data-type settings play a crucial role in locating documents in classification schema assuring the capture and storage of the data. Detailed characteristics of WD can be found in Annex C.

Roles and responsibilities

Overall responsibility of the preserved content of WD relies on the EKA records manager. At the operational level, the administrator supports an environment in which digital preservation is recognized as a critical step in caring for materials. This includes providing adequate managerial, technological, and financial resources to establish and maintain the digital preservation plan. In EKA the records manager has accountability at the operational level.

In general, all staff including those who care for content, create it, use it, or administer it, have the responsibility of preservation by following the regulation and legislation of their everyday work.

The Department of IT manages the technical infrastructure needed to care for digital resources in EKA. They create, install and maintain software as needed, providing support for staff using the tools. Software vendor Webware takes care of all developments of WD according to the requests from the WD administrator.

Workflow steps to consider for preserving the content of ERMS Webdesktop

a) **File formats:** Even though the software supports different formats the most used are PDF, .doc, .docx, .DDOC, .BDOC and ASICE, the most challenging to preserve, are digital signature formats. If one signed capsule in Digidoc[8] software is inside another capsule, which is also signed, causes more clicks to open signatures meaning difficulties in interpreting the file. This kind of file is more fragile and complicated, which increases the risks of digital preservation.
If a zipped file is digitally signed, the BDOC and ASICE format files may appear too large. Zipping the files would only make sense if there is a need to preserve the structure of the catalogue. Otherwise, files can become too large due to ASICE and BDOC formats being zipped themselves and therefore preserving one zip inside the signed capsule makes the overall file larger.
In the EKAs case, where all system users are creators of records and able to create digital signature files, extra training must be provided to keep BDOC and ASICE file creation under control.
b) **File format migration:** DDOC signature files need to be converted to a newer BDOC or ASICE format to be accessed and valid by today's software.
c) **Long-term file integrity**: File integrity issues should be managed by the service provider, however, agreements do not cover this yet.
d) **Making digital content publicly accessible**: The content of ERMS is published according to the Estonian legislation and Public Information Act[9]. In the active phase of documents, the responsibility for correct settings relies on the records manager. As access restrictions may have a long validity period the access metadata must be correct.
The public part of EKA ERMS is available through WD Public Document Registry (PDR) (https://adr.artun.ee/?desktop=57835&tid=1) and public access is ensured through an automated solution built into the system.
e) **Storage space**: Data hosting service is provided by a private company which has agreed to 5 GB of storage capacity. The volumes will be reviewed regularly and, if necessary, the contract will be amended.

[7] Archives Act of Estonia: https://www.riigiteataja.ee/en/eli/521032019019/consolide
[8] https://www.id.ee/en/rubriik/ria-digidoc-application/
[9] https://www.riigiteataja.ee/akt/112072014033

f) **Extraction of data from the system**: Oneclick eArchiving project activities have shown that the biggest bottleneck is extracting the data from Webdesktop for the later transfer to NAE. There is no solution yet for the user to choose records to be extracted, however, the work is started and considered in the Action Plan Summary in Annex A.

3.2 EKA Digiteek collection preservation

EKA Digiteek Metfond (https://metfond.artun.ee/metfond/est) is one publicly available database of the whole Digiteek system (https://digiteek.artun.ee/est). The Digiteek system is designed and optimized for the needs of the Cultural Heritage and Conservation Department, especially for condition reports, asset management, and Graphical Documentation (GraDoc) of heritage artefacts. Currently, in the image bank, there are more than 8700 records, 71 conservation condition reports, and 85 GraDoc reports (August 2022).
The Metfond was developed inside of the system in 2016, and it is a part of the system which contains mainly image material, more than 4400 image records.
Since 2014 Digiteek has had some updates and technological changes (e.g. in 2018-2019 transferring technology from a Flash-based solution to HTML5), and the system has proven to be firm for cyberattacks and server updates.
EKA's digital strategy roadmap is a plan to implement a new Digital Asset Management (DAM) system during the next few years. Nevertheless, aside from the new manifest-based DAM is still a place for Digiteek GraDoc, which meets the needs and expectations of the conservators. Technical characteristics can be found in Annex C.

Roles and responsibilities

A number of Digiteek stakeholders have critical digital preservation responsibilities, including those who care for content, create it, use it, or administer it.

1. **Digiteek Management Team:** The team provides oversight of the digital preservation plan implementation. The team evaluates the plan on a regular basis, revising it as processes, standards, and formats evolve. The team also oversees outreach and promotion efforts for digital preservation.
   The management team includes a records manager, data curators and a representative from the IT department.

2. **Data curators:** Data curators determine digital preservation priorities and are responsible for developing processes and workflows implementing digital preservation activities. Data curators are responsible for assisting producers with understanding and complying with established deposit requirements.
   In terms of Digiteek and all cultural-historical digital assets, the data curator's team includes the members of EKA Museum and Gallery, also, the data curator(s) from faculties must be fixed to ensure a cross-organisational view and higher data quality.

3. **Users:** Any individual or group who uses the services to discover and access digital materials. Sometimes referred to as consumers. Users include students, researchers, visitors, and online patrons.

4. **Technology Support:** Technology Support manages the technical infrastructure needed to care for digital resources. They create, install and maintain software as needed, providing support for staff using the tools. Examples include the EKA department of IT, and software vendors (Webware LLC, TrinidadWiseman LLC).

5. **Administrators:** Administrators support an environment in which digital preservation is recognized as a critical step in caring for materials. This includes providing adequate managerial, technological, and financial resources to establish and maintain the digital preservation plan.
   Members should include the super user of Digiteek and the new DAM technologies and content manager.

Workflow steps to consider for preserving the EKA Digiteek

a) **File formats**: A very diverse spectrum of different formats can be included as the artwork in EKA varies from audio-visual and multimedia to the restoration of centuries-old artefacts. In the Metfond section, there are JPEG, TIFF, PDF formats. In the whole Digiteek, there are JPEG, TIFF, PDF, PDF/A, MP4, MP3, M4A, WMA.

b) **File format migration**: File format migration of the image files of the Metfond section can be arranged when new technological solutions have been tested and proven with no quality loss migration available.
File format migration of the Digiteek, especially video and audio file formats, is important to test continuously, as some of the formats are contained inside additional containers and during the automatic migration can be either declining quality level or even information loss.
PDF and other possible textual formats are important to convert and migrate into PDF/A format which is the universal archival solution for all e-file storage needs. PDF/A eliminates the chaotic "format zoo," creating a streamlined and uniform archive.
PostgreSQL database provides several solutions for migration without data loss or corruption.

c) **Long-term file integrity**: The fewer different formats the easier it is to arrange to monitor and provide continuous public service. It is necessary to implement checksum solutions - demonstrating authenticity, and that no files were changed when copying storage from one media to another.
About PostgreSQL database: it is very difficult to enforce business rules regarding data integrity using Read Committed transactions because the view of the data is shifting with each statement, and even a single statement may not restrict itself to the statement's snapshot if a write conflict occurs.

d) **Making digital content publicly accessible**: The content of Metfond is published according to Estonian legislation and Public Information Act[10]. There can be materials that are accessible via different levels of restrictions. The access metadata must be

[10] https://www.riigiteataja.ee/akt/112072014033

correct because access restrictions may have a long validity period. The public part of EKA Metfond is available here https://metfond.artun.ee/metfond/.

e) **Storage space**: For the first year, we intend to capture only the thesis and illustrative materials of the final works for permanent preservation and retention. In the following years, the bachelor's theses will be added, as well as other art collections. The staff, system architecture design, and designated server space need continuous funding to accommodate the growth of storage space each year.

f) **Criteria for prioritization**: Given the breadth and complexity of materials within a repository, it is essential to prioritize digital preservation activities of collection materials. The first priority is to capture and preserve theses/final works ensuring the integrity of these collections of each faculty. In the next few years, each faculty will identify collections and projects as leading digital priorities. This list will be updated as progress is evaluated and priorities are reviewed over time.

g) **Copyright issues**: EKA needs some fundamental analysis, understanding, agreement and documentation. "Those engaged in digital preservation must work within the law as it stands. This requires both a good general knowledge of what the law is and a degree of pragmatism in its application to preservation work. Such knowledge enables the archivist to avoid the pitfalls of over-cautiousness and undue risk aversion, and to more accurately assess the risks and benefits of taking on the preservation of new iterations of digital work" (Charlesworth, 2012., p. 3)[11].

In conclusion, there are a lot of challenges and further actions needed in terms of the long-term preservation of EKA's digital assets. In regards to the first collection, ERMS WD's most critical issues will be solved after the extraction problem and development of the direct interface between ERMS and NAE are completed.

Secondly, records of cultural-historical value from academic units need to be categorized and functional requirements for the procurement of the new DAM solution need to be finalized. The migration of CMS Digiteek Metfond to the new DAM software is dependent on what is planned and on DAM implementation. Moreover, the organizational structure for managing storage, maintenance and long-term preservation should be fixed.

4. Suggestions and consideration of possible Action Plans

The following options are to be considered to ensure a secure and safe digital future:

4.1 Option 1 – Multiple Backups

- Finish digitizing from analogue media (film, etc.) as time and budget permit. Make copies to the office machine hard drive, and two external hard drives.
- For digitized images, audio, and film: Create a naming system that ensures each resource has a unique filename. Record that unique filename in the archival database under the record for that resource. This allows digitized resources to be searched using the database.
- Keep all working copies of all digital assets (all groups) on the internal hard drives of the office computers. These computers are password protected at the operating system level (i.e. one must provide a password to use the machine).
- Make regular backups of the relevant directories for all media groups to two external hard drives.
- Refresh media by replacing the external hard drive every five years.
- Migrate to new file and software formats as required.
- At any time that migration takes place, ensure the migrated files are backed up to the external drives.
- Keep at least one external hard drive offsite.

Option 1 Pros

- Addresses immediate need to move digital assets from non-archival CDs.
- The simplest solution – can be implemented without significant training.
- Affordable – requires only the purchase of a second external hard drive.
- Provides better archival protection than the status quo.

Option 1 Cons

- No Fixity checking (no process to determine if a file has changed from its original state).
- Keeping all content on hard drives does not observe the 3-2-1 rule
- No audit trail if archival content changes.
- The archive is susceptible to viruses being transferred to it.
- Handling/referencing between old and new ones

4.2 Option 2 – Multiple Backups including backup checksums

Checksum is one of the important factors in the fixity of the files which are currently missing in the EKA preservation system. Checksum generators use a function to produce a unique value based on the contents of a file. If the file's contents change even by a single bit, the

[11] Charlesworth, A. (2012.). Intellectual Property Rights for Digital Preservation. *DPC Technology Watch Report 12-02.*, 3. Retrieved from Available: http://dx.doi.org/10.7207/twr12-02

same checksum will not be produced the next time the generator is run, indicating that the file is not the same; i.e. its “fixity” can be assessed. EKA workflow needs to be designed in compliance with modern long-term preservation tools, e.g. E-ARK tools[12].

Various checksum generators can be considered:

a) The **Duke Data Accessioner**, a free application that ingests content by making a copy of a file (or several files) into an archive, generating basic preservation metadata (such as: who preserved the content, when, a short description of the content), and then generating checksum information. Versions of this software can be run on several common operating systems and it is accompanied by an operator’s manual.
b) The **MD5 Checker**, is also free. It creates checksum information only – no other preservation metadata. However, it is easier to use than the Duke Data Accessioner, and sufficiently intuitive that no consultation of the manual (which is available online) was required.
c) If the OS is a Linux-based system a scheduled cron job with a native checksum tool could also be utilized.
d) All programming languages have the possibility to calculate checksums.

Option 2 Pros:

- Fixity of files can be easily assessed
- Using a checksum is in keeping with archival standards

Option 2 Cons:

- Some amount of training may be required
- The Checksum generation represents an extra step in the archival process

4.3 Option 3 - Backups stored as SIP packages

Multiple backups stored as SIP packages allow a simple check of consistency among copies with E-ARK tools and there would be a possibility to transfer those directly into a compliant archive in the future.

Option 3 Pros:

- Compliance with standards
- Simple to take into use
- Workflow design supports the organisational structure

Option 3 Cons:

- Preparation works for homogeneous quality level might be the threshold
- Implementation cost if connected directly to vendor systems
- Some amount of training may be required

4.4 Option 4 – Create an internal archive using the OAIS[13] model

In this option, the simplest form of software necessary to create an OAIS-compliant model is considered. Archivematica[14] and ResourceSpace[15] were reviewed for this purpose (both in installation costs, and operational complexity) to be practical for EKA. Other OAIS-compliant applications such as Roda[16] and ESSArch[17] and Disec OAIS-compliant archive solutions as a result of the Oneclick project should be reviewed, compared and considered to select the most suitable and affordable to implement and manage in EKA.

Option 4 Pros:

- The whole OAIS workflow will be covered
- compliance with standards is guaranteed

Option 4 Cons:

- possible implementation is difficult and needs strong organizational effort most critically by the IT department
- substantially too onerous

4.5 Option 5 - Use an external archive

[12] E-ARK Tools: https://www.eark-project.com/resources/eark-tools.html
[13] OAIS Model: https://public.ccsds.org/pubs/650x0m2.pdf
[14] Archivematica: https://www.archivematica.org/en/
[15] Resource Space: https://www.resourcespace.com/#
[16] Roda: https://demo.roda-community.org/#welcome
[17] ESSArch: https://www.essolutions.se/essarch/

If an external organization could manage an OAIS-compliant archive for use by the EKA and similar organizations, the responsibility of refreshing and migrating could be completely offloaded. Moreover, the work involved for the EKA in accessioning and accessing content might be reduced. No known service exists at present in Estonia, but services offered by Keep and ESSArch should be considered and a cost/benefit analysis would have to be performed.

Option 5 Pros:

- Does not require specific infrastructure
- Short-term maintenance costs are under control
- The ease with which content can be searched and retrieved

Option 5 Cons:

- The archive's ability to manage disparate forms of data
- Bandwidth/time considerations for uploading and downloading content
- Sustainable finances for the EKA (where fees for using the archive exist), sustainable finances for the archive, and assurances (legal obligations) of the archive to maintain the EKA's content
- highly dependent on third-party service providers
- National autonomy hard to control

Annex A – EKA Action Plan summary and General Procedures

The following is an itemisation of actions and procedures to be carried out under the recommended plan which is based on the traditional backup option model. The preservation needs of the collections Webdesktop, Digiteek Metfond and all EKA's records of historical-cultural value have been considered. The action plan is coherent with Oneclick eArchiving project deliverable M2.16: EKA Collections Migration Plan and M2.2: CMS/CRM/ERMS Exports Report, where more detailed descriptions can be found about each collection's migration actions.

To do short-term:

- Purchase, or ensure the EKA has in its possession three new external Hard Drives, each with at least 2 Terabytes (TB) of disk space (PRICE total). Label the drives as
  - Preservation Copies 1 Working Machine – Replace on YYYY/MM
  - Preservation Copies 2 Working Machine – Replace on YYYY/MM
  - Preservation Copies Archives – Replace on YYYY/MM

  …where YYYY/MM is the date five years from the purchase date of these drives.
- Establish a protocol to ensure the content on the archives machine and the database remain synchronized on a regular basis.
- Snapshot/backup Digiteek and Webdesktop as follows:

  A. Extract the content from the systems as three collections:
  1. **ERMS WD PDR** (Public Document Registry), which includes information provided for the public and does not include sensitive data
  2. **ERMS WD AV** (archival value), which consists of records appraised by NAE to be transferred to NAE and preserved for long-term
  3. **CMS Digiteek Metfond** which is EKA's web-based digital collection of graduation works/theses created since 1914

  B. Create SIPs by using Oneclick SIP Creator[18]

  C. Create AIPs and DIPs by using EARK Archivematica AIP and DIP Creator[19]
- Identify source directories, all those containing content.
- Ensure deleted files on the source drive are archived on the target drive.
- Name the destination folder on the target drive as "BackUp YYYY/MM/DD"
- Download and install a copy of MD5 Checker, on the office machine.
- Copy all contents of existing CDs (and in particular, contents of all non-archival CDs) to the hard drives on the office computers (archival content being copied to the archival computer, and all other content being copied to the office working computer).
- After the first scheduled backup (the first backup can be run early if desired) rename the "Current Backup" folder to "Preservation copy backed up YYYY/MM/DD" whereYYYY/MM/DD is the date on which the backup was run.
- Open MD5 Checker, and run it on all files under the new "Preservation copy" folder. Store the MD5 checksum information in that folder.
- Plug in the second external hard drive and copy to it the contents of the first external hard drive.

To be performed routinely:

Extract the content from the ERMS WD AV and create SIP. Transfer SIP to NAE for long-term preservation according to the requirements of NAE.

To be performed on an annual basis:

Extract the content from the ERMS WD PDR and create SIP, AIP and DIP + Recovery test
Extract the content from the new DAM solution and create SIP, AIP and DIP + Recovery test

[18] SIP Creator codes: https://gitlab.com/jaaskela79/oneclick-full, Tutorial: https://www.youtube.com/watch?v=Vns-GwuyfWI

[19] PIQL codes: https://github.com/penwern, Tutorial: https://www.youtube.com/watch?v=aXgCv_Zlrps

Annex B - Summary of the Inventory

It is important to mark down all the digital assets and demonstrate the readiness to adapt the options based on the certain asset. One possible and easy-to-follow example is the following inventory sheet:

Name of Digital Collection	Preservation master filetypes	Approximate Number of Digital Assets	Approximate Amount of File Space	Minimum Number of Copies of Assets
EKA ERMS Webdesktop	*asice, bdoc, ddoc, odc, docx, gif, html, jpeg, jpg, odt, pdf, png, ppt, rtf, rtx, tiff, txt, xml, xls, xlsx, zip*	*Items in total: 96055* *Files: 169097 (53173 GB)* *Digital signature files: approx 34000 (smart-ID not included)*	*5 GB*	*NA*
EKA CMS Digiteek Metfond	*tiff, jpeg, jpg, pdf*	*objects 11435*	*87 GB*	*2 - hard drive on server & external - all onsite*
Some examples of unclassified collections to be ingested to the new DAM:				
Artproject 3 video works (since 2005)	*videoformats, converted H264 MP4, etc*		*1 TB*	
Estonian history of photography (since 1990-ies). In cooperation with KKEK[20]).	*tiff, jpg, dng*	*incremental*	*1 TB*	
Architecture Graduation Theses (since 2012)	*mostly PDF, but also mov, mp4, photos (large files)*	*incremental*		
EKA building film-chronicles (since 2007,2008)	*film, video*		*2 TB*	
…				

[20] KKEK: https://cca.ee/

Annex C - Technical characteristics

Corporate ERMS Webdesktop	
	Linux or FreeBSD based operating system
	Apache WWW server
	Mod_SSL security system
	PostgreSQL database
	Clam-AV antivirus software
	OpenOffice.org office software
	Programming languages PHP (main part) and Python (auxiliary scripts)
	The main functionalities: ● the creation of digital documents based on templates or digitized, ● internal workflow processing, ● digital signing, ● retention, and disposal within the EDRMS, ● standard interfaces with systems used to perform support functions (such as financial and personnel systems).
CMS Digiteek	
	Linux (Fedora 10, CentOS 7) either server / virtual machine or (preferably) Docker container
	Apache2 web server
	Perl, Embperl, Modperl
	PostgreSQL database

Glossary

AIP (Archival Information Package)	The package to transfer and store digital objects and associated metadata to enable access and preservation for the long term.
Analog	Data or information is created and maintained in such a way as to require transference to a digital format to make it digitally available to users.
Authenticity	Assurance that the digital object is complete and unaltered since its creation. Authenticity is established through metadata.
Bit	The fundamental unit of digital information is 1 or 0.
Bitstream	A sequence of bytes, which has meaningful common properties for the purposes of preservation. A bitstream may be comprised of a file or a component of a file.
Byte	A unit of digital information (eight bits).
Born Digital	Data and information are created and maintained in a digital format and are not intended to have an analogue equivalent either before or after creation.
Chain of Custody	Documentation of the acquisition, transfer, ingest, and ongoing preservation of archival material.
Checksum	Typically expressed as a text string or hash value, checksums are outputs generated by an algorithm and compactly express the data in a file or other data Checksums can be used to detect errors or changes to digital files.
Digital Object	An entity in which one or more files and their corresponding metadata are combined, physically and/or logically by means of a digital wrapper
Digital Wrapper	A system of encapsulation that combines administrative, preservation, technical, structural, or descriptive frames of data to be combined together into a single entity
Digitized Materials	Analog materials have been transformed into digital form, especially for storage, access, and use in a computer environment.
DIP (Dissemination Information Package):	The package of digital objects and metadata is provided to the user for access.
Emulation	The provision of functionality in one computing system that is equivalent to one found in another (often obsolete) system. This could be done at the hardware level or the software level.
Encapsulation	The inclusion of additional information around data, allows a system to access that data without any prior knowledge of the data's format. This typically involves a file of a known format (known as a "wrapper") that contains data in a format that may not be known to the accessing application in advance, as well as instructions for accessing this data.
File	A bitstream is managed by a file system as a single, named entity.
File Format	An attribute of a file that describes its encoding and is typically identified by the extension at the end of the file name.
Fixity	Refers to a digital file that has been unchanged. Archives want to ensure that files are not altered or corrupted, and will run fixity checks, usually through a checksum.
Ingestion	The process of transferring data into an archive or repository for long-term preservation. E.g the OAIS entity that contains the services and functions that accept SIP from producers, prepare AIP for storage and ensures that AIP and their supporting descriptive information become established within the OAIS-compliant repository.
Long-term preservation	The act of maintaining correct and independently understandable information over the long term.
Life Cycle	A series of stages through which digital information passes. The lifecycle for digital information includes creation, use and reuse, migration or emulation, and storage.
Hardware Dependency	The degree to which a specific piece of hardware is required in order for another object (hardware, software, file, etc.) to be accessed and used successfully.
Metadata	Structured information about content that enables long-term use. Metadata comes in multiple categories including administrative metadata, technical metadata, descriptive metadata, and structural metadata.

Migration	The process of converting data from an obsolete structure to a new structure to counter software obsolescence. OR: The transferring of data to newer system environments. This could include new file formats, new operating systems, or new physical carriers (media).
OAIS (Open Archival Information System)	Reference Model developed by the Consultative Committee on Space Data, a conceptual framework and reference tool for defining a digital repository. It provides a model of the environment, functions, and data types for implementing a digital repository.
Provenance	The chronology of the ownership, custody, and location of archival materials.
SIP (Submission Information Package):	The package of materials that are sent to digital storage for preservation, and then converted into an AIP through archival processes.
Software Dependency	The degree to which a specific piece of software is required in order for another object (software, hardware, file, etc.) to be accessed and used successfully
Refreshing	The transfer of data onto a newer version of the same storage media (that is, the same type of physical carrier) so that degradation of the older media does not pose a risk of data loss
Replication	The creation of copies of data on one or more systems