Digital preservation policy of the KB, National Library of the Netherlands

Version 6

August 2024

KB national library of the netherlands

Colophon

Publisher:
KB, national library of the Netherlands,
2024

Editorial:
Sam Alloing
Daniël Steinmeier

Imagery:
Beeldstudio KB

Thanks to:
Simone Kortekaas
Elsbeth Kwant
Dennis Mark
Brigitte den Oudsten
Meta van der Waal-Gentenaar

More information:
jeffrey.vanderhoeven@kb.nl

© KB, national library of the Netherlands,
The Hague.

Executive summary

Our mission references long-term preservation by mentioning both current and future users. The vision of the KB extends this notion by explicitly mentioning long-term access and usability. This document will describe how we aim to achieve these goals at a strategic level for our digital collections.

Preservation is an active process, requiring constant monitoring and involvement of the whole organisation. It is first and foremost a way of looking at things from a long-term user-centred perspective. We cannot predict the future, but we can try to stay up to date with technical developments and evolving user requirements. We cannot be certain in advance that preservation solutions will be sufficient or will be productive even. Therefore, it is important that our policies are open to learning through evaluation.

The goal of this document is to define a clear direction that may help organisational decision making on topics that impact long-term accessibility. We realise that preservation is a sociotechnical system that not only involves IT, but also people. It needs both **stability and flexibility**. These opposing organisational values must be counterbalanced carefully if we want to profit from both. Our preservation policy contains principles, strategies, definitions and guidelines, but also our vision on preservation, background information and a rationale for the rules we try to establish. Our goal is to make this a document that can be understood by experts as well as non-experts. This is important to get the message about preservation across within the organisation.

Preservation is all about **adapting to change**. However, adapting to change might be an uncertain and ambiguous process depending on the magnitude of change required and on the level of familiarity with new circumstances. Uncertainty is not something to solve, but to gradually clear up by processes of organisational learning. Without this, uncertainty can become a reason for inaction.

We define preservation as the work required for **maintaining integrity, authenticity and long-term accessibility for our collections**. We should not only do the right things but also be able to provide evidence of this to our user groups: the **designated community**. Access is an important part of preservation because the access services are the channels via which our designated community can determine whether **our collections are accessible, discoverable and reusable**.

The designated community also needs information to check the **reliability** of our collections. For example, a researcher needs to understand the choices that were made during the life cycle of the content to understand the reliability of the information for their research.

The infrastructure that is required to run systems supporting preservation is also part of this policy, because **infrastructure is the fundament upon which preservation is built**. If something goes wrong in the infrastructure it can have major consequences for the preservation of our collections. Safeguarding the collection also means **proactively** identifying risks and dealing with them. This way we don't need to come up with ad hoc solutions.

Preservation is to a large extent also about **people**: users, producers and the larger organisation within a network of other organisations. The knowledge of professionals and their experience are as important as other aspects of preservation within the organisation.

In this document we also describe which adjacent fields within the organisation provide relevant preconditions for the sociotechnical preservation system to work.

1 Introduction

Strategy alignment

"The KB, National Library of the Netherlands, has been a source of inspiration and development for centuries. Since our foundation in 1798, we have developed into a broad, versatile organisation, that makes the National Library collection visible, usable and sustainable for all Dutch people, for any purpose, both now and in the future. We care for the written word, especially the Dutch publications, and enable everyone to read, learn and do research. This is how we contribute to a smarter, more skilled and more creative Netherlands."

Our mission references long-term preservation by mentioning both **current and future users**. The vision of the KB extends this notion by explicitly mentioning long term access and usability.

Furthermore, one of the KB core values, **trustworthiness**, is also linked to long-term accessibility of the collections. The above principles have taken more concrete form through the different documents that form our preservation policy, the strategic collection plan, the content lifecycle, the policy plan, as well as in procedures and technical solutions intended to sustain the processes of ingest, storage and access of data and metadata within the lifecycle of our digital objects.

Purpose

This document will describe how we aim to achieve these goals at a strategic level.[1] Preservation is an active process, requiring constant monitoring and involvement of the whole organisation. It is first and foremost a way of looking at things from a **long-term user-centred perspective**. We cannot predict the future, but we can try to stay **up to date** with developments and **evolving user requirements**. This is an important part of what it means to manage a collection. A long term perspective implies making assumptions about the future, but these should be grounded in empirical evidence. Long-term accessibility is not a goal to be achieved in the future, but something to keep in mind when making decisions in the present. It not only involves being up to date with technical evolution of file formats, tools and systems but also following documented procedures and being able to provide evidence for trustworthiness to our users. To this end we seek certification on a regular basis and provide information relevant to certification either via our corporate website or via one of our services. We have based our preservation goals on:

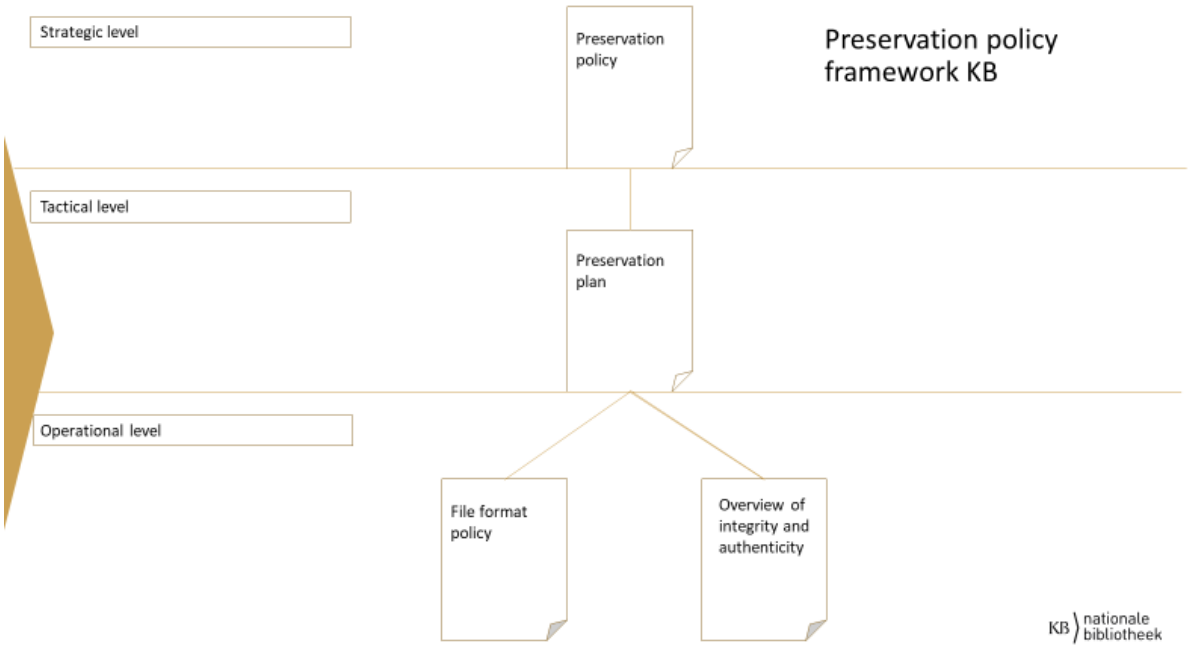
- The OAIS-reference model[2]
- The guidelines for trustworthy digital repositories (TDR)[3]
- FAIR[4]

[1] The structure of this document is partially based on the DPC template: Template for building a preservation policy - Digital Preservation Coalition (dpconline.org)
[2] https://public.ccsds.org/pubs/650x0m2.pdf
[3] https://public.ccsds.org/pubs/652x0m1.pdf
[4] https://www.go-fair.org/fair-principles/

The reference model provides a flexible structure that requires translation to our organisational context, before we can even think of implementing any of its functions. This document describes how we intend to **bridge the gap between model and practice**. The goal is to define a clear direction that may help organisational decision making on topics that impact long-term accessibility. We want to show our users not only that we follow documented procedures, but also that procedures may change due to innovation and new insights. This is how we intend to stay up to date.

Scope

The preservation policy concerns all digital collections of the KB, regardless of the system in which they are preserved. It defines high level policy guidelines that all digital material needs to adhere to in order for them to be considered long-term accessible. These guidelines will be worked out in more detail at the tactical level in the preservation plan. Finally, the guidelines in the preservation plan will be implemented through policies at the operational level.



2 Preservation policy

Preservation is all about adapting to **change**. However, adapting to change might be an uncertain and ambiguous process depending on the magnitude of change required and on the level of familiarity with new circumstances. Therefore, we will first establish four guiding principles that form the basis of our preservation strategies. These are the underlying values we believe can help us form **consistent strategies** even for situations we cannot yet foresee.

Preservation principles

1. *We believe in an agile approach*, so we try to use or create preservation strategies that enable us to progress towards our goals **step-by-step** and based on **empirical decision making**. Preservation is not an all-or-nothing situation, so the models we use should reflect this reality. Rather than nominally adhering to an ideal while in practice making exceptions, we prefer to reach our goal by taking small but consistent steps towards our goals while communicating and evaluating our methods and even our underlying values. This will enable us to be flexible and adaptive while at the same time creating transparency about the state of the implementation process.
2. *We have a holistic view on preservation.* Preservation cannot be an isolated function within a larger process but needs to be a **part of every process** that might impact long-term accessibility of the collections. This means preservation specialists also need to be able to get involved and create change in areas that impact the digital collection indirectly. They do not need to take over responsibility but should be able to address problems and effect change, if required by preservation standards. This means preservation strategies can be relevant to any area of the organisation as long as the relevance of change within these areas can be related to aspects of long-term accessibility. It also works the other way around: people from other parts of the organisation may be involved in thinking about preservation solutions or implementing them.
3. *We strive to keep up with change.* Over the course of years, we should be able to demonstrate **user-oriented improvement** in all facets of preservation: policy, expertise, services, systems, documentation. We do not want to freeze things or predict the future. Instead, we intend to create strategies that enable us to report on improvements made. What direction to take is informed by risk analysis and monitoring of the designated community.
4. *We focus on what is needed urgently.* We do not attempt to pre-emptively take action based on uncertain forecasts or rules-of-thumb because these might well turn out to be unnecessary. Instead, we define strategies through **risk assessment and risk calculation** based on reliable information about current developments and urgent user requirements without losing the longer-term developments out of sight.

Preservation strategies

In accordance with these principles, we have defined a number of strategies that describe how we plan to move preservation forward in the coming years.

In line with our principle about risk assessment, we use a **just-in-time-strategy** for implementing preservation strategies, like file format migration. We do not normalise files during ingest nor will we consider migration based on rules-of-thumb, like open formats being deemed easier to preserve. We will consider this as a viable strategy only when there is an absolute necessity to migrate to a new file format.

We have defined a **minimal ingest strategy** to make sure files are safely preserved before they are analysed. This means we preserve files as we receive them. The minimal ingest strategy will ensure a faster and more reliable ingest process with minimized risk of files ending up unmanaged. This also means the files can be used sooner in other processes, like providing access, without having to wait on validation processes or error handling. There are however important checks that do need to be carried out before ingest. At a minimum, these are verifications whether the delivery is according to agreements with the producers and whether files are not corrupt. This way the content can be preserved and used within the KB services. A disadvantage of minimal ingest is that file format identification is postponed and needs to be executed afterwards. Therefore, file characterization requires special attention as part of another preservation strategy described below.

To improve our control on risks associated with file format obsolescence and other risks for example encryption, we will implement our strategy of **knowledge levels**. After ingest, we will characterize the files and thus over time gradually expand our knowledge about formats in the collection. Technical information will be extracted and used for risk calculation, so obsolescence threats and other threats can be detected and mitigated. We don't expect to know all risks and threats right away it will be expanded step-by-step. This strategy will be one of the stepping stones towards being able to do functional preservation[5] in the future, so we can preserve the content without being dependent on the original format. The original files are migrated to a new file format. It also enables us to be transparent about the current state of functional preservation because we can report to our users exactly which file format is known at which level.

Other areas of preservation will also use growth models for gradual improvement. For instance, **representation information** is also an area where this approach seems suitable. This term from the OAIS-model is about the context information needed to ensure both contents and technical aspects of our collections are understandable to users. Based on feedback from the designated community and internal expertise, we can determine whether current representation information is deemed sufficient for understanding the data or whether information needs to be added. So, these strategies are important for preservation. But how do we define preservation?

Core values

Preservation is best summarized by introducing our three core values: integrity, authenticity and long-term accessibility.

[5] Functional preservation is the next level in CoreTrustSeal certification (previous is bit level preservation).

Integrity

Integrity is related to the completeness of the digital objects and the collection. We have defined five aspects of integrity that we use to operationalise this concept.

- **Bit integrity** defines completeness of individual files. It is important to ensure files do not become corrupted over time.
- **IP-integrity** is used for determining completeness of the Information Package. This means that all the information provided by the Producer and created during ingest, is kept together in an Information Package so it can be used in other processes. By grouping the information in this way, we can ensure no information is missing. After all, when information is missing the publication can't be understood properly. Information Package is a reference to the concepts of SIP, AIP and DIP in OAIS.
- **Version integrity** as a concept is used for ensuring connections between versions of the content are available. These connections can be used by the designated community and by administrators to inspect the chain of custody.
- **Information integrity** is defined as a concept to safeguard understandability of the content. Users should be able to understand the content without dependence on expert help. This means that representation information can be added if necessary to improve understanding of the content. Only information that is not already part of the knowledge base of the designated community is provided as representation information.
- **Collection integrity** is used for determining completeness of the collection as a whole.

Authenticity

Authenticity concerns the question of how we can prove the content is what it purports to be. The provenance needs to be recorded, so our users can get to know the origin and the history of the content. To this end we make sure:

- The producer as **source** of the content is recorded.
- Relevant actions within the lifecycle of the content are recorded as **event history** within the preservation metadata. Relevant actions are for example receiving publications, validation of the received package or changes to the information we originally received.
- The original **intention** of the content is maintained.

Source, event history and intention together make up our concept of authenticity. Authenticity is based on trust and should be available as verifiable evidence for the designated community to inspect. There can be various levels of granularity when it comes to the evidence presented for authenticity and more can be added when required by the designated community.

Sustainable accessibility

Sustainable accessibility is determined according to the DUTO-model[6] by the following aspects: findable, readable, interpretable, reliable, and available.

This means users must be able to:

- find material using metadata and persistent identifiers
- display any object found
- comprehend an objects' contents
- determine an objects' integrity and authenticity
- issue an object as a DIP

Designated community

Strongly related to sustainable accessibility is the concept of the designated community. In the OAIS-model this is defined as the target audience of the collection. This community is important because their feedback is used for **verifying** that preservation solutions are sufficiently implemented. This means providing sustainable access or more precisely findable, readable, interpretable and reliable collections. The designated community does not have to be a single group but is often divided into several subgroups.

Subgroups can be distinguished based on special services or specific requirements that are appropriate for a certain group. This concept is not meant to exclude people, but rather a means to **explicate the needs** of certain groups. We need to enter into dialogue with these groups so that a relationship may be built that is based on mutuality. Without this **relationship**, reliability and trustworthiness can not be guaranteed because bias may be implied in collections or access services in ways that we are not aware of. The concept of Designated Community also includes people who currently are not users of the library. We have an explicit intention to make inclusion part of the concept of Designated Community. Helping subgroups with limited access to the library to get access is also part of this.

Offering a variety of services also implies having various subgroups of the designated community. All these subgroups may have different requirements, a different knowledge base and different ways of considering understandability. For instance, linked data is targeted to machine-readable forms of metadata that can be linked to online resources. The people belonging to this subgroup have vastly different needs compared to, for instance, a subgroup of researchers who come to the reading room to inspect the web archive. As subgroups we distinguish:

- Researchers

[6] https://www.nationaalarchief.nl/archiveren/kennisbank/duurzaam-toegankelijk

- Network partners
- The general public
- Linked data users
- Producers
- Internal users

Content and metadata

We use the term content in this document as a blanket term for the material that we want to preserve for the long term. In digital preservation it is common to distinguish between different composition levels of information. For instance, the PREMIS-model that distinguishes between bit streams, files, representations (bundles of files for a certain purpose, for instance access copy) and intellectual entities. The OAIS model distinguishes between data, packages and objects. However, as can be seen from the case of web archiving, what constitutes an object cannot always be clearly defined. An archived website is very much a creation of the archiving institute itself. It is the capture of the state of an object at a point in time, rather than the object itself.[7] Therefore, the notion of object might also be fluid, depending on content type. To cover all these different types and granularities we therefore use the term "content". The data describing the content is the metadata, which may consist of different types of metadata, for example descriptive information, provenance information, technical information, rights information and context information. In practice the boundary between data and metadata might also be fluid. A case in point would be an ALTO-file that stores both the full text of the content as well as data about the content, like coordinates.

According to the FAIR principles, **metadata should be as rich as possible**. It should describe the resource adequately for current and future use. This helps in understanding and finding the resource.[8]

We try to adhere to this principle by capturing as much metadata as relevant for long-term accessibility from the source and by generating technical metadata and provenance information. When it comes to **manual creation of metadata**, what is possible is determined by cost-benefit analysis and **may differ**. Analysing the quality of the metadata is a way to monitor the richness of the metadata and follow the evolution of the metadata. By quality analysis of the metadata, we also mean the contents of the metadata fields and not just whether the metadata validates technically according to, for example, a schema. This is of course also important and can be considered the basic level of analysis to build further quality analysis upon. The following points apply:

- The designated community needs to be able to inspect the trustworthiness of the publication. For example, they should be able to use the metadata and the publication description for this

[7] Masanès, Julien. "Web archiving methods and approaches: A comparative study." *Library trends* 54.1 (2005): 72-90.
[8] https://www.go-fair.org/fair-principles/f2-data-described-rich-metadata/

purpose. Because of the event information and the richness of the metadata, this type of inspection is possible.
- The quality of the metadata is ensured by checks and during transformations in the transfer phase.
- Submission agreements will be drawn up for depositors. When setting up a new ingest, these agreements will form the basis for an inventory of what metadata can be obtained.
- These documents are also used for knowledge sharing so we can inform depositors about requirements for proper processing.
- If the content or metadata is not according to the agreements, the depositor is asked to redeliver. For example, corrupt content will not be preserved in that case. Instead, the business information manager will request the correct version.
- The KB currently preserves all versions of each digital collection object. During implementation it is important to take into account the efficient storage of multiple versions. Further policy making on this topic will be expected in the future.
- The acceptance of content and metadata according to the agreement is recorded as part of a preservation event. This is an important event within the lifecycle, because this is the point in time where responsibility is transferred from the producer to the digital archive.

Discoverability and persistent identifiers

Discoverability[9] is an important part of the digital collection. The KB doesn't intend to create a Dark Archive. Therefore, **all digital collections should be discoverable**.

- The designated community should always be able to inform itself on the contents of any collection the KB holds.
- Access rights are specified according to agreements with the producer.
- There is monitoring of unauthorised access.
- Users should be able to discover what is not available anymore. The metadata needs to be accessible even if the item is not in the collection anymore. This is in accordance with one of the FAIR principles[10].
- Retaining the metadata of removed collection items is also useful in case a persistent identifier links to that item. The persistent identifier can be made to resolve to a new location in case an item is transferred to another institution.
- To make the collection discoverable or findable the persistent identifier must also be part of the metadata and should be presented to the user. It functions as an important link between systems and when executing workflows.
- The persistent identifier functions as a guarantee to our users that this identifier will not change and that it is being maintained for the long term.

[9] https://www.go-fair.org/fair-principles/f4-metadata-registered-indexed-searchable-resource/
[10] https://www.go-fair.org/fair-principles/a2-metadata-accessible-even-data-no-longer-available/

- The persistent identifier needs to be globally unique, so it can be cited and consulted by any user, regardless of the technical context of the underlying resource. This is a FAIR principle[11] as well.

Reuse

Reuse of the digital collection is an important goal of the KB. Reuse means that the designated community can use the digital collection as they see fit, in accordance with the rights determined by the producers.

- The reuse conditions need to be clearly visible for all users and need to be preserved with the items of a collection.
- To support reuse, appropriate derivatives need to be created.
- Reuse is also the R in FAIR. Richly described metadata is needed in accordance with the FAIR principles regarding reuse.
- All available metadata should be preserved and provided to the user when requested and if relevant for long-term accessibility.
- There should be a clear indication of the version of the preserved content.
- The user should be informed of what is available and what is not, as well as any other limitations known in regard to the collection.
- A description of the condition of the collection should be made available in collection profiles.
- Appropriate licensing information needs to be present in the metadata. This information should also be machine readable so machines can automatically enable reuse of the collection in accordance with the specified license.
- In accordance with FAIR[12], the protocols used in granting access to the collection should be free and open, so anyone may implement a retrieval process as they see fit. This doesn't mean the collection needs to be freely available. Verifying access rights or account creation may be part of the process.

Standards and models

The use of standards is important for preservation. The definition of standards can differ, depending on context. In this document the term standard refers to **de jure standards**.[13] **Open specifications** can also be considered standards. It is important that the specification that is used is openly available. This level of availability will enable us to look up any information we may require from the specification at any time, so we can take action or expand our knowledge if needed. What we cannot consider standards, are de facto standards. These are standards that are used often but are not de jure standards or don't have any open specification available online.

Standards can also be important as a source of **inspiration on good practices** of other institutions. This helps the knowledge exchange in digital preservation and related fields. It can function as a

[11] https://www.go-fair.org/fair-principles/f1-meta-data-assigned-globally-unique-persistent-identifiers/
[12] https://www.go-fair.org/fair-principles/a1-2-protocol-allows-authentication-authorisation-required/
[13] https://dictionary.archivists.org/entry/de-jure-standard.html

common language that is better understood by others. It also helps in becoming less dependent on specific solutions and is therefore more future proof. Standards need to be implemented in the organisation and any implementation of standards in the KB needs to be documented. Standards often provide room for interpretation. Therefore, specific implementation profiles should be documented.

Currently used standards are:

- OAIS[14]
- PREMIS[15]
- METS[16]
- ALTO[17]
- PAIMAS[18]
- ISO-16363[19]

Protecting the collection

The KB protects the digital collection following the concept of Three lines of Defence. The highest level of defence is **the strategic level** as this specifies the direction of the collection management. This line ensures commitment, multi-year budget and the development of policy. Monitoring and evaluation are important aspects of governance because they help to determine the right direction or correct the chosen course.

At the **tactical level**, the next line of defence is designing and implementing process management, risk management, budget follow-up and translating policies in guidelines. Designing process management includes assigning responsibilities, dealing with financing and structuring the decision-making process. The core business process features process managers who will manage the process and resolve any problems that may arise in the core process. Digital preservation is part of this business process and sets the agenda within the frameworks of the business process. It is important to keep improving these processes and making sure they are well documented. Having well documented processes and continuous improvement is an important precondition to digital preservation. In this way everybody can understand the process and evaluate, improve and verify processes related to digital preservation. This is important from the perspective of accountability and trustworthiness.

[14] OAIS: https://en.wikipedia.org/wiki/Open_Archival_Information_System
[15] PREMIS: https://www.loc.gov/standards/premis/
[16] METS: https://www.loc.gov/standards/mets/
[17] ALTO: https://www.loc.gov/standards/alto/
[18] PAIMAS: https://www.dcc.ac.uk/guidance/standards/diffuse/show?standard_id=154
[19] ISO-16363: http://www.iso16363.org/standards/iso-16363/. Previously in this document also called guidelines for trustworthy digital repositories (TDR)

At the **operational level**, process management, risk management, budget and guidelines need to be implemented to support digital preservation. For all these topics evaluation and monitoring need to be implemented as well.

Documentation

All the processes related to digital preservation should be documented, so they can be used to **evaluate and improve processes**. This is a starting point to continuously improve digital preservation and make processes more reliable. The documentation can also be used to inform third parties about our digital preservation processes. It will improve knowledge and understanding of what is needed to create processes that support the digital preservation of the digital collection.

It also **gives producers confidence** that the information is preserved in a reliable and trustworthy way. As stated before, an important part of documentation is representation information. By having representation information preserved and linked with our digital objects, the designated community can get to know the **context and history of our collections**. Collection profiles and collection descriptions are some of the forms that representation information may take. Other solutions might become necessary in the future. This will be determined based on user feedback and internal expertise.

3 Adjacent policy areas as preconditions for implementation

Preservation is often seen as a technical field: file formats, storage, technical metadata. However, it requires **cooperation** with important stakeholders in other areas of the organisation. Some of the requirements in these areas are also explicitly mentioned in the **guidelines for Trustworthy Digital Repositories**, **CoreTrustSeal certification** and the **DDHN preservation policy guidelines**[20]. This is why it is important to mention these topics as part of the preservation policy. The following paragraphs will describe some of the policy areas adjacent to preservation. These areas of responsibility are important preconditions for the preservation policy to function as intended.

Expertise

It is important to guarantee sufficient in-house expertise in the field of digital preservation. All stakeholders involved in the long-term acquisition, processing, preservation, and accessibility of digital collections in the KB must be represented in organisational decision making on topics that affect long-term accessibility. Preservation requires an attitude of life-long learning. Expertise can be improved by various means, for example training, workshops and conferences on the subject of digital preservation. Another way of improving expertise is by experimenting. Hands-on experience with new topics can improve understanding and expand our expertise. It is particularly useful for things that are innovative and topics we don't have any experience with. All stakeholders should get digital preservation training to improve understanding and get a feeling for the subject. Preservation specialists should get specialised training to improve knowledge in specific fields and/or actively participate in working groups in the digital preservation community. We strive towards equal collaboration and participative decision making within the organisation. To this end, we also want to make sure stakeholders can make informed decisions based on shared knowledge. We try to further this idea of a common knowledge base by documenting background information and implementation suggestions on important preservation topics, such as authenticity.

By improving expertise, people in the organisation are encouraged to become involved and take responsibility in reducing identified risks or proposing ways of reducing risks and creating opportunities to improve digital preservation. Everyone involved in handling the digital collection needs ton get training on preservation. This way employees who work with the digital collection are made aware of the importance of keeping the information safe for the long term. Having a code of conduct as part of preservation work might be a way to further guarantee responsible handling of the data.

20 https://kennis.cultureelerfgoed.nl/index.php?title=Speciaal:Vragen&limit=500&offset=20&q=%5B%5BCategorie%3AArtikelen%5D%5D+%5B%5BLid+van%3A%3AThema%2FWegwijzer+duurzaamheidsbeleid%5D%5D&p=mainlabel%3D%2Fformat%3Dul&sort=Voorkeurslabel&order=asc&eq=no#search

Networks

Contributing to (international) partnerships is also important in order to support the digital preservation community and build a professional network. For digital preservation important partnerships are, for example:

- Open Preservation Foundation (OPF)
- International Internet Preservation Consortium (IIPC)
- Dutch Digital Heritage Network (NDE)
- METS Editorial Board
- ALTO Editorial Board.

We also want to encourage contributions to partnerships outside of digital preservation in fields that are relevant to the domain of digital preservation. For example, communities for tools we use in the digital preservation workflow. The term contributing needs to be interpreted broadly and can also mean financial support.

Another important reason to contribute to (inter)national communities is that it provides an opportunity of challenging current practices and learning new practices in collaboration with peers. This stimulates discussion and implementation of new ideas. To evolve these new ideas and validate them against current practices, we intend to use different techniques. For instance, setting up experiments.

Infrastructure

The infrastructure is an important underlying component of the technical side of digital preservation. Problems of data loss may occur in a way that is not immediately apparent. The term data loss should be interpreted in broad terms. All loss needs to be considered, not only bit rot. Bit rot may happen on storage systems: the bits on a system are inadvertently changed by actions on the data. It is the most commonly known form of data loss. Current IT storage systems have multiple protections against bit rot, like using some form of erasure coding or RAID[21] to name one strategy. But there are other forms of data loss as well. For example, data corruption.

- To prevent data loss or corruption, it is important to verify **checksums** regularly and to ensure that no changes occurred. There needs to be a process to restore the uncorrupted file. Also, during moving or copying of files it is important to ensure there is no data corruption.
- The status of content for long-term preservation must be **documented**. If the collection is spread over different systems with no common view on the status of the data, the content can be at risk to data loss. If no one knows that content is stored in a certain location, the content may get thrown away along with the server for example during media refreshment or when storing on external media.

[21] https://www.techtarget.com/searchstorage/definition/erasure-coding

- When data loss or corruption has occurred, or when it was imminent, a root cause analysis should be executed to **learn from the experience**. An important role of the infrastructure in countering data loss or corruption is storage.
- To ensure reliability, the infrastructure needs to be supported by the vendor or similar party with **appropriate support** contract and to be implemented according to best practices of the vendor providing the solution. The vendor can also be the KB itself. The same principles apply.
- The infrastructure also needs to be reliable in the long term, therefore **lifecycle management** on the components of the system is an important issue. When the technical lifespan of a system is in the end stages, the system needs to be replaced. The procedure to replace the system needs to be defined when starting a project, so that systems are replaced or upgraded before support expires. Systems that are too old are less reliable and are more expensive to maintain. Also, the systems the KB creates itself need to be included in the lifecycle management
- The reliability of the infrastructure is also something that needs to be **audited and tested**. Disaster recovery procedures need to be audited and practiced in real life situations. When a calamity occurs, everyone should be prepared to take well considered action. The disaster recovery plan is part of a wider Business Continuity plan.
- To make the storage reliable for long term preservation the storage component needs to provide for **multiple replicas** or similar techniques stored in different locations with enough distance between the locations. This way a problem in one location doesn't create data loss or corruption. The copies as described this section are defined as redundant data for protection purposes, regardless of whether multiple versions of the content are available in the collection.
- There should be only one **preservation copy** of an item in the digital collection, there are possibilities of derivatives for specific purposes, for example for access. They are only used for this purpose and are no substitute for the collection.
- All files are stored in accordance with the same storage policy.
- The storage environment is **monitored**, so timely notification is given of any storage medium failures.
- Replication functionality, together with storage redundancy and monitoring, ensures that no data is lost.

Security

Security is a broad topic and includes **IT**, **buildings**, **people**. Security is handled by a wide range of specialized personnel. Recent legal developments, such as the Network and Information Security (NIS2) Directive expected to be applicable to our services at the end of 2024, necessitate taking action on this topic.

- An important aspect of security is countering **cybersecurity attacks** like ransomware attacks. It is important to protect the digital collection against these kinds of attacks, but also to have tested measures in place when such an attack occurs.
- It is important to **revise** the security measures regularly and do an **audit** of the procedures. This is important for evaluating the procedures, but also for practicing the procedures. This includes training on cyber security breaches.

- The measures taken should be **tested** and **evaluated** for effectiveness.
- A **Business Continuity Plan** should describe the measures that need to be taken to ensure safeguarding of the collection. It is important to be prepared before a crisis occurs.
- **Risk management** is important as a method to identify predictable risks and opportunities and for evaluating which risks to mitigate, which risks to accept and which opportunities to seize. The accepted risks need to be clear to everybody involved and their acceptability needs to be evaluated. It is also important to improve knowledge of risk management and risk leadership. This way everyone can seize opportunities and take responsibility for risks in order to keep the digital collection safe.
- For **physical security** the digital collection should be part of the procedures that exist. This includes access to the data centres, access to computers in the reading room and access to network access points throughout KB locations.
- The content can be stored on physical digital media. This can be hard disks, tapes or optical media. Digital media that is decommissioned needs to be **purged** according to the existing standards. If external parties are involved, proof of destruction needs to be provided. The evidence of this action needs to be preserved.

Legal

Long-term access requires that licences to the content are clear. These determine what actions we can take to preserve the content and also under which conditions we can provide access. For this reason, we preserve rights information in the metadata, that agreements with producers are preserved along with the content and that users can access the rights information. These considerations should inform the process of contract negotiation. The KB has no legal deposit. Accordingly, publications are held in long-term storage with the express approval of the party that supplied them. As part of the agreement, we are allowed to make any changes to the objects that are required for the purposes of long-term accessibility.

- For authenticity purposes the changes need to be recorded as an event in the metadata of the publication. The terms of the agreement are checked when publications are provided, and the outcome of these checks are recorded as an event as well. The data storage contracts set out conditions for processing, preservation, management, and availability.
- The KB's guiding principle is to offer all content as much open access as possible, for current and future use. The strategy is based on the assumption that the KB will at least make publications available on-site (i.e., on its premises). The extent to which the KB can make publications openly available depends on the terms of any agreements reached with the publishers and on legislation and regulations (Copyright Act, General Data Protection Regulation). Statutory provisions and the agreements reached both determine the range of options that the KB can offer its users.
- The metadata is always accessible, also for publications that are not accessible. It is indicated in the metadata that the publication is not accessible, so that it is clear for the designated community what the KB holdings are.

Financial

Having planned and structured budgeting processes for preservation is important to ensure continuity of services and keeping in-house expertise. A long-term view on budgeting is required so that the quality of services is not overly dependent on the current financial situation of the organisation.

- Creating a financial buffer will help in times of financial shortage to ensure collections can still be preserved. If a buffer is not available, removing or pruning collections to save storage space will seem a more feasible option. However, this may turn out to be a regrettable decision in the long term. This is an often-overlooked risk in digital preservation because it is tied up with larger organisational processes and the less tangible, subjective question of value.
- A cost analysis and forecasting can help understand the wide range of costs and the expected evolution of these costs. The goal within preservation is not to put quality above all other considerations so that it becomes unaffordable, but to make decisions based on reliable information that can also be inspected by the designated community. Therefore, a cost-benefit analysis might be part of preservation decisions, but it must be clear that long-term benefits of any cost-cutting strategies are certain, necessary and in line with defined policies.

Collection policy

This brings us to another important area: the collection policy. This is also an area that has a large impact on digital preservation since it defines the scope of what needs to be preserved. The TDR guidelines do not necessarily define how to write a collection policy - or what should be in it - but it does have requirements about transparency and availability of collection policies, so the designated community may inform itself on what is and isn't in the collection. If the execution of our collection policy entails changes to our current collection, the following implications should be considered:

- In case a decision has been made to remove a collection, this should be clearly documented and communicated, so the designated community is informed of this decision.
- Before removal, there should be evidence that the content will be preserved in another trustworthy digital repository and that a 'tombstone' record will be available to inform users of the new location.
- The persistent identifier of a publication must not break. This can be done, for example, by redirecting the persistent identifier to the new location. There needs to be a succession plan in place where parties agree on the transfer of responsibility for preserving the information and keeping resolvable persistent identifiers.

4 Glossary

DUTO: is an abbreviation that means 'Duurzaam toegankelijke overheidsinformatie' in Dutch or in English: ' Sustainable accessible government information'. This model is defined by the National Archives in the Netherlands. The KB doesn't have government information and is not a government agency, but the practical framework is also applicable to other types of material. We use the following concepts according to this practical framework: findable, readable, interpretable, reliable, and available.

FAIR: guidelines for improving Findability, Accessibility, Interoperability and Reuse

OAIS: Open Archival Information System, the model that defines which functions a trustworthy digital repository should implement

PREMIS: Preservation Metadata dictionary

SIP, AIP, DIP: Submission Information Package, Archival Information Package, Dissemination Information Package. These terms are taken from the OAIS model and reflect the different forms of an information package during the stages of ingest, storage and access in the object lifecycle

TDR: the guidelines for Trustworthy digital repositories (ISO-16363), the standard that defines which requirements should be followed and what evidence should be provided to be considered a trustworthy digital repository