2019-2022 Preservation plan

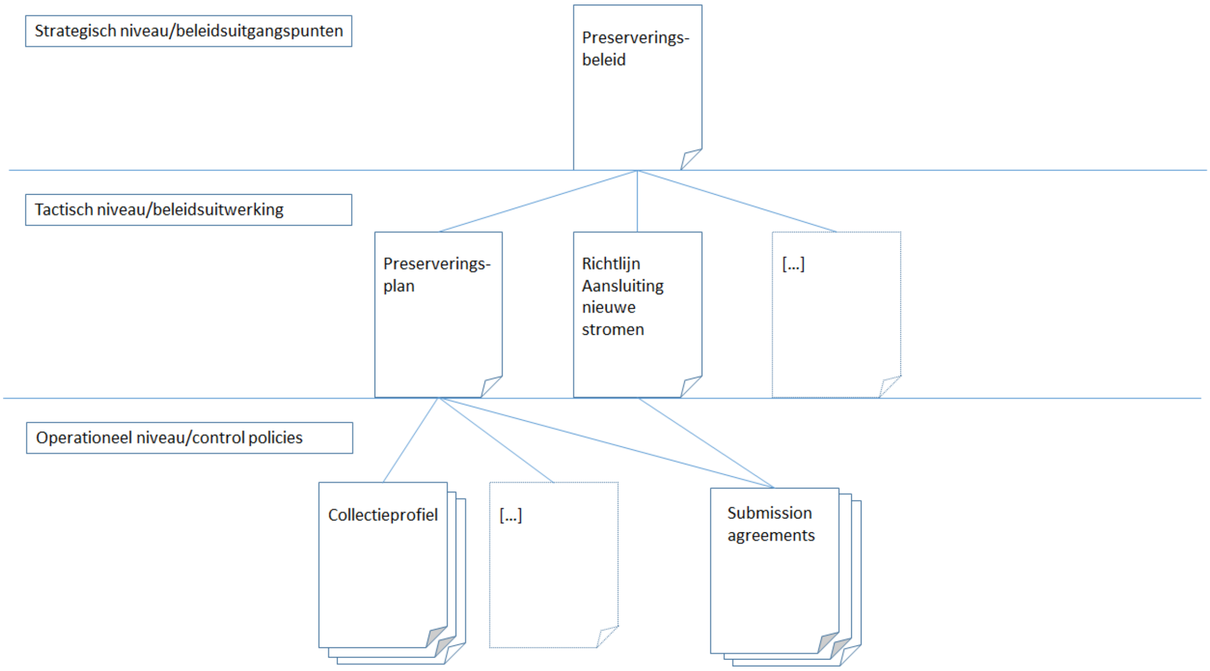
Introduction

The purpose of this document is to give further substance to the National Library of the Netherlands' (KB) adopted preservation policy for the period from 2019 to 2022[1]. Whereas the preservation policy sets out principles at strategic level, this document provides guidance on how to formulate the aforementioned principles in detail at tactical level. The level of the present document also corresponds to what the Netherlands Coalition for Digital Preservation's catalogue refers to as the level of policy development. The preservation policy then formulates the policy principles, using that terminology. The present document's structure deviates from that of the preservation policy, as we expect that it will appeal to a different target group. We also feel that it should be possible to read it as an independent document.

The topics formulated in detail in this document involve preconditions (cited in the preservation policy) for the collection's long-term preservation. To facilitate this, reference is also made to several broader preconditions (concerning the legal and organisational frameworks, for example) without which there could be no guarantee of preservation[2]. While these preconditions are derived from the preservation policy, their scope is not limited to preservation alone. For this reason, they will not be formulated in detail in the preservation plan.

The current plan is not intended to pool all items of information about preservation at a tactical level. However, it does include references to other documents at the same level, in which mention is made of matters that have not been formulated in detail, if at all. This document also contains references to documents at operational level (or suggestions for drafting documents at this level), which give practical expression to points relating to policy development. For clarification, see the proposed document hierarchy depicted in the illustration below. Preservation policy underpins all tactical-level documents. These, in turn, underpin operational-level documents. Operational documents, such as the submission agreements, can give substance to the principles enshrined in more than one tactical-level document.

[1] https://www.kb.nl/sites/default/files/docs/digitaal_preserveringsbeleid_kb_def.pdf
[2] Ibid, see subsections 2.6 Rights and 2.9 Organisation, for example


Strategisch niveau/beleidsuitgangspunten
Preserverings-beleid
Tactisch niveau/beleidsuitwerking
Preserverings-plan
Richtlijn Aansluiting nieuwe stromen
[...]
Operationeel niveau/control policies
Collectieprofiel
[...]
Submission agreements

Finally, it is useful to remember that this document does not contain a specific preservation action plan. The principles of preservation planning as a process are, however, defined in the tactical policy. Nevertheless, specific action plans - based on risk analysis and preservation watch - need to be drawn up at operational level. Based on the Open Archival Information System (OAIS) model, this document can be seen as administration policy. The functional entity of preservation planning can then use this as a basis for developing proposals for action plans.

How this document is related to other policy documents and to ISO 16363

This document is related to a number of policy documents adopted by the KB.

Firstly, the preservation policy. This document sets out the strategic principles that are given tactical expression in the present document. The strategic policy can thus be regarded as what ISO 16363 refers to as a Preservation Strategic Plan. Accordingly, this meets requirement 3.1.2.[3] Therefore, at the tactical level, the present document is an initial response to requirement 3.3.2, which states that a repository must possess policy documents that were formulated on the basis of a strategic plan.[4] ISO 16363 refers to such documents as preservation policies. In this document, however, we will refer to such documents as preservation plans, to avoid any confusion with strategic policy.

Secondly, there is the KB’s Content strategy. Broadly speaking, this document sets out the main objectives and principles for collection development and access. Accordingly, it sets preconditions that govern the selection of items for long-term preservation. This document, therefore, complies with requirement 3.1.3, which states that collection policy documents

[3] “September 2011 Magenta Draft of ISO 16363”, https://public.ccsds.org/pubs/652x0m1.pdf: pp. 3-1
[4] Ibid. pages 3-6

must be available.

Finally, there are various documents in the areas of security, privacy and risk management, such as the Company Emergency Response Plan[5], the Company Emergency Plan[6], the Information Disaster Recovery Plan[7] and the Collection Disaster Recovery Plan[8]. Broadly speaking, these documents are a response to requirement 5.1.2, which states that provisions must be made for mitigating security risks in the areas of systems, staff, and physical conditions.[9] For more details, see the Information Security section below.

The processes used to ensure that data remains authentic and accessible need optimum environmental conditions if they are to be really effective. This means that any staffing, financial, and infrastructural factors that have an indirect impact on the collection's preservation must be taken into account.
These considerations have been broadly incorporated into the abovementioned documents. However, with a view to the future, it is important that they should also include the issue of digital preservation. For instance, the preservation policy is duly compliant with requirement 3.1.2 (as stated above) but additional documents must be supplied if this policy is also to be compliant with the underlying requirements. The latter state that there must be a succession plan, for example, and that disaster recovery measures must have been implemented.[10]

Finally, there must also be a clearly documented process by which changes to the infrastructure can be initiated, based on projected risks to the collection's preservation. There should also be a financial buffer, to fund any innovations that may be required.[11] These points represent the preconditions for effective digital preservation.

[5] https://plein.kb.nl/documents/51711
[6] https://plein.kb.nl/documents/51710
[7] [Details of the location to follow]
[8] https://plein.kb.nl/documents/51712
[9] "September 2011 Magenta Draft of ISO 16363", https://public.ccsds.org/pubs/652x0m1.pdf: pp. 5-12
[10] Ibid. 3.1.2.1, pp. 3-2
[11] Ibid. See requirements under 5.1.1, pp. 5-1 ff

Policy development in terms of preservation themes

A tactical interpretation of the preservation policy's preservation themes is set out below. The interpretation of these concepts must be subject to regular review. Furthermore, the chosen solutions must be re-evaluated in line with changes in legislation, software development, software contracts, IT infrastructure, the expectations of users, and the requirements of depositors.

Integrity

The term 'integrity' refers to the guarantee that objects and collections are complete, and that any changes are verified and documented.

There are several levels of integrity, and we use a range of measures to verify them.

- Bit integrity[12]
- Version integrity[13]
- IP integrity[14]
- Information integrity[15]
- Collection integrity[16]

Bit integrity

We verify the bit integrity of files to demonstrate that a given digital file – as stored within the IT infrastructure – is identical to the publication that was received (or to the latest version thereof). A file's bit integrity is verified by means of a checksum. Depositors provide a checksum upon delivery. If a checksum cannot be provided then, as a minimum requirement, a comparable mechanism must have been agreed with the depositor in advance, to guarantee reliable transfer. In any event, a checksum is then calculated and stored after receipt. Whenever changes are made to a file, a new checksum must always be calculated and stored.

Each file has its own checksum. The IT infrastructure's design ensures that files are checked regularly and that they are verified by means of a checksum, to detect any bit rot. Regular reports on this process are provided for review purposes. If any errors are detected within the system, self-healing and replication processes are used to maintain the file's integrity. A report is also drawn up. When transitioning to a new storage medium, it is essential that a comparable verification and reporting process be put in place.

Bit integrity also means that bits are stored exactly as they were received, and that they have not been changed without due verification. Accordingly, such data must not be encrypted, compressed, or deduplicated at storage level, to ensure the preservation of bit integrity. The goal is end-to-end verification, to be sure that the file remains unchanged from the moment of delivery until it is made available. A standard verification process is then used

[12] This is in compliance with ISO 16363, requirements 4.4.1.2 and 5.1.1.3

[13] This is in compliance with ISO 16363, requirement 4.4.1.1, and is comparable to the model implementation method presented there. See also requirement 5.1.2.

[14] This is in compliance with ISO 16363, requirements 4.1.5 (Submission Information Package – SIP), 4.2.1 and 4.2.9 (Archival Information Package – AIP)

[15] This is in compliance with ISO 16363, requirements 4.1.2 (SIP) and 4.2.5 (AIP)

[16] This is in compliance with ISO 16363, requirement 4.2.9

to ensure that the file remains unchanged during storage.

Version integrity

The policy principle that the original (and every version of a publication) must always be preserved, means that it is important to safeguard *version integrity*. This is because some files are the originals (as originally received from the depositor) while others are new versions supplied by the depositor to correct an error. In such cases, the integrity of the material requires that any relationships between different versions of an intellectual entity must be preserved. Preserving relationships between versions, and rendering them transparent, makes it possible to manage an object's chain of custody. This can be used to demonstrate the integrity of an object's versions to depositors and to the designated community, and to verify that these versions can be traced back to the original. As things stand, versions are stored but the relationship between them is not specifically recorded in any systems.

IP integrity

Another level of integrity is *IP integrity*. This means that the Information Packages' completeness is verified as they pass through the various stages. During the ingest process, checks are carried out to verify that every file the SIP was expected to contain has, in fact, been delivered, based on agreements with depositors. With regard to the AIP, it must be possible to trace every file (as originally supplied) back to one or more AIPs. The same applies to the Dissemination Information Packages (DIPs). Here, it is important to point out that, from the viewpoint of preservation, no distinction is drawn between the main files and any supplementary files. Details of the information package's (IP) stipulated contents are recorded in a manifest file. This file also serves as a reference in completeness checks, during the ingest process, or during migrations, for example. In the future, IP summary structures should also be used at collection level to record details of an IP's stipulated contents at each of the various stages. This would provide a baseline that could be used to verify the completeness of individual IPs.

Information integrity

Information integrity is another level that needs to be monitored. This concept is about ensuring the long-term preservation of the representation information needed to interpret an object, and that it is always persistently related to that object. That involves storing extra information as metadata or as a document. In practice, any documentation needed to understand the information contained within an AIP is stored with the AIP itself, or as a separate AIP with an appropriate reference.

In practice, issues such as the composition of an AIP or the features of different versions of an object differ from one collection to another. These differences are clarified and recorded by means of specially prepared collection profiles and IP summaries. This is to ensure that there is just a single reality concerning an AIP's stipulated contents and the status of objects within a collection. That documentation will then contribute to the information integrity of every object within the collection.

Collection integrity

The latter point touches on the overarching level of *collection integrity*, which involves verifying the integrity of the collection as a whole. This involves verifying that every object that should be present within a collection has actually received a definitive status. Definitive status is allocated if, at the end of the ingest process, it can be confirmed that a complete and verified AIP has been stored at the final location. Alternatively, based on documented procedures, it may be decided that an object will not be definitively included. In certain cases, a tombstone record must be kept, to ensure that the reasons behind this decision remain transparent over the long term. Many flows currently lack a specified, independent mechanism for objectively confirming a collection’s completeness. Accordingly, mechanisms of this kind should be specified in detail in the future.

Authenticity

The question addressed by the abovementioned mechanisms associated with the concept of integrity is: “Is the material complete and has it not been unintentionally changed?” However, the concept of authenticity concerns verification that an object is, in fact, what it seems to be[17] or what it should be, according to Submission Agreements, and that steps within the IP lifecycle have been implemented in accordance with standard procedures. In this connection, certain attributes of an object (i.e. intention, provenance information, history) must be highlighted. These aspects are also highly dependent on the ingest process, as all items of information required to guarantee every aspect of authenticity need to be obtained at this stage.

Intention

It is important to be able to demonstrate that the object corresponds to what the producer intended to deliver.[18] This also involves the above-mentioned concept of integrity because, if an object is incomplete, this may lead to a situation in which it can no longer be displayed as intended. At the same time, this is also a matter of the files’ authenticity, in terms of their contents. This does not necessarily mean that a file’s name or contents must be entirely error-free. However, the minimum requirement is that the data has actually been supplied in the form in which it has to be stored. For example, if a file has a ‘.doc’ extension but is actually a PDF, this does not necessarily mean that it is not authentic. Renaming the file could then be considered a preservation action. However, if a file contains an error message instead of descriptive metadata, then the object will not correspond to the producer’s intention. The latter category is relevant for verification in the context of authenticity. To this end, current ingest processes use special file-identification software. It is important to learn more about formats, to signal any future risks in terms of the format’s accessibility (as described below). This aspect of authenticity is also important in terms of format migration. Here, it must be confirmed (based on significant properties) that a new form of the object still corresponds, as far as possible, to the form that was intended by the maker.

Provenance information

Authenticity can also be demonstrated by storing the object’s *provenance information* as metadata within the AIP. The contents of an object’s *provenance information* are determined in consultation with the producer, who is also part of its provenance. Producer verification

[17] For instance, the contents of a file may not correspond to the format indicated by its file extension, perhaps because it was renamed incorrectly.

[18] This is in compliance with ISO 16363, requirements 4.1.3, 4.1.5

currently takes place during the ingest processes. In this way, the identity of the party who originally supplied the object is always clear.[19] This aspect of authenticity is further safeguarded by concluding contracts with depositors. We also store these details, and any references to the supplying party, within the object itself.

History

Finally, the *history* of the object is established, as further confirmation of its authenticity. This is established by adding metadata associated with specific events that have taken place throughout the object's entire life cycle at the KB. These events are currently recorded in the manifest file contained in each AIP.[20]

Long-term accessibility

If access is to be regarded as 'long term', certain aspects must be guaranteed. Users must be able to find material using metadata[21], it must be possible to display any object found[22], an object's contents must be comprehensible[23], the user must be able to determine an object's integrity and authenticity, and it must be possible to issue an object as a DIP[24]. This corresponds to the following concepts – retrievable, readable, interpretable, reliable, and available.[25] Thus, the definition of 'reliable' is directly linked to the way in which the above-mentioned concepts of integrity and authenticity are interpreted. Accordingly, integrity and authenticity are preconditions for long-term accessibility.

Each of these points also involves a legal dependence, as all of these processes require guarantees that there are sufficient rights to implement actions, that these rights are transparent, and that they are also enforced within the processes.

This currently involves documenting agreements on this matter in the form of contracts with depositors, recording details of rights (as metadata) and, where relevant, translating this into access rights within the systems. The contract management process is now a separate process. This process should be better integrated in the future, to indicate more clearly which contract is linked to which publication. In this way, any rights that can be derived from the contract with regard to preservation are also clearly laid down, thus ensuring the long-term preservation of the underlying agreements.[26] The processes related to preservation must be further formalised and documented. A strategy for modifying these processes in response to changes also needs to be developed. The suitability of using ISO 9001 as a guideline for the implementation of these measures will be assessed.

Input from the following three information sources is used to determine whether objects can be deemed accessible:

[19] This is in compliance with ISO 16363, requirement 4.1.4

[20] This is in compliance with ISO 16363, requirements 4.1.8 (SIP) and 4.2.10 (AIP)

[21] This is in compliance with ISO 16363, requirement 4.5.1

[22] This is in compliance with ISO 16363, requirements 4.3.2/4.2.7

[23] This is in compliance with ISO 16363, requirements 4.3.2/4.2.7

[24] This is in compliance with ISO 16363, requirement 4.6.2

[25] As formulated within the definition of long-term access described in the Sustainable Access to Government Information (DUTO): https://www.nationaalarchief.nl/archiveren/kennisbank/duurzaam-toegankelijk

[26] This is in compliance with ISO 16363, requirement 3.5.1

1. Agreements with depositors (submission agreements)[27]
2. User survey (monitoring designated communities)[28]
3. Preservation watch (risk analysis based on aspects such as technology watch and monitoring designated communities)

The input from user surveys and preservation watch will be used to determine whether compliance with the above-mentioned aspects of long-term accessibility is possible, over time. Agreements with depositors and preservation watch information are used as input for the legal preconditions that will need to be taken into account in any future solutions for safeguarding accessibility.

In the future, agreements with depositors will be explored and recorded in accordance with the steps set out in the PAIMAS model. Risk management is an important aspect of preservation watch, in terms of safeguarding the digital collection's preservation. Risk management is important for the preservation of collections throughout the KB. In addition, technology watch will also need to be set up systematically. This will include monitoring that focuses on giving timely notification of any relevant developments in the areas of technology and software. Links will be established with existing KB user survey initiatives in this area, which can then be expanded in a way that is in line with preservation. The definition of a designated community will also be expanded. Finally, the internal processes used to analyse file formats will also become an important source for preservation watch. See the Knowledge Levels section for further details.

Preservation strategies

The preservation policy defines two preservation levels – bit preservation and functional preservation. The ultimate goal is the functional preservation of every single file. Thus, the preservation levels that can be assigned to individual objects are temporary qualifications.

The point of *bit preservation* is to ensure that, following the initial determination, the bits contained within a file remain verifiably unchanged. The above-mentioned measures for guaranteeing the integrity, authenticity, and accessibility of objects are of key importance in this respect. Storing something at the bit-preservation level signifies that very high preservation requirements are still being set. However, the very nature of bit preservation precludes full compliance with every aspect of integrity, authenticity, and accessibility. To be more specific, information integrity cannot be fully guaranteed. Nor, with regard to accessibility, can the concepts of 'readable' and 'interpretable' be guaranteed. Accordingly, this level is sustainable over the short term, as the lifespans of form and content are directly linked to the lifespan of the format and to that of the software environment in which it functions.

The point of *functional preservation* is to ensure that a representation of the original file can be made accessible. Here too, the definitions of integrity and authenticity play a key role. Extra emphasis is also given to preserving form and content, independently of the format. Bit

[27] This is particularly important with regard to the legal aspect of accessibility

[28] For example, ISO 16363 requirement 4.2.7 makes reference to comprehensibility testing

preservation is a standard feature of functional preservation, which also involves measures to maintain accessibility over time. Accordingly, this level is sustainable over the long term. One of the policy goals is to shift the focus from the exclusive use of bit preservation to functional preservation. The first step in this process is to define *preservation strategies*[29], *preservation levels*, to design a process for *preservation planning,* and to establish the necessary monitoring mechanisms (as mentioned above, under accessibility).

The two most important strategies for functional preservation are emulation and format migration. Given the diverse nature of the KB's collections, it will be appropriate to explore both strategies in the context of long-term accessibility.

Emulation can be used to guarantee accessibility, by maintaining a software environment in which an original file remains accessible. In this context, it may also be important to keep the original hardware and/or software accessible for a fixed period of time.

Format migration involves transferring an object's contents to a new format, to ensure that they remain accessible. For a proper understanding of this concept, it is important to draw a distinction between the four different types of migration, in line with the OAIS-model. These types can be categorised on the basis of their intended purpose and their impact on the data being migrated.

1. Refreshment involves migrating objects to the same medium, without modifying them.
2. Replication involves migrating objects to another medium, without modifying them.
3. Repackaging involves migrating objects to another information package, without modifying their contents.
4. Transformation involves migrating objects to another file format, and modifying the Content Information within the file format.

The first three types of migration are important for both bit preservation and functional preservation. The latter type is only used in connection with functional preservation. In this case, there are changes to the files, so the checksum changes too.

The latter type of migration is referred to here as format migration. This type can involve reversible and irreversible transformation.[30] In the first case, the new format can be converted back to the old format without any loss of data. In the second case, the original data can no longer be accurately reconstructed. It is then necessary to verify that a representation of the original data has been preserved (which involves the use of *significant properties*).

The term 'significant properties' refers to the properties or components of an object without which that object can no longer be regarded as entirely authentic, if at all. These are initially defined at the level of the intellectual entity, in abstract and substantive terms. A migration plan must include a substantiation of the method used to translate these significant properties into technical properties within the new format. This information is also recorded in collection profiles, for long-term preservation.

[29] This is in compliance with ISO 16363, requirement 4.3.1

[30] "Reference Model For An Open Archival Information System (OAIS)", https://public.ccsds.org/pubs/650x0m2.pdf, page 5-4ff

The collection regularly undergoes the first two types of migration. As these actions do not result in any changes to the objects, it has been decided that they should not be recorded as event history in the object's metadata. Of course, it is important to demonstrate that the data's integrity has been preserved. This is achieved by means of integrity checks. Integrity, therefore, refers to all integrity checks and not just checksum checks. In addition, details of migration projects – including planning and procedure – are recorded in documentation for long-term preservation.

The third form of migration, which is also in current use, is recorded as event metadata. In the case of newly created objects, that would be a 'creation' event. This can also be a feature of migrations if any changes have been made to the AIP's structure or to the way in which that structure is defined within a system. In cases like this, a preservation action plan defines how the structure is to be renewed, whether there might be any different scenarios, and how the new structure takes account of the necessary preconditions for integrity, authenticity, and long-term accessibility.

The fourth type of migration comes into play if preservation considerations result in a decision to store certain formats differently. One example would be when files that were originally packaged are stored as separate files. This is a reversible transformation. Event metadata concerning this will also be stored, and a new checksum will be generated, to enable the new files' bit integrity to be monitored in the future.

Knowledge levels

As mentioned above, the aspiration is to find out more about the range of formats that are available in the collection. This will involve the use of functional preservation to obviate any future risks with regard to accessibility. The principle here is that, ultimately, technical information will be available on all of the formats in the collection. In addition, there will be sufficient in-house knowledge about these formats to ensure their long-term accessibility, by means of various preservation strategies. Gaining knowledge and analysing objects is a growth model, in which each format is assigned a specific level. In the near future, this growth model will be initiated by taking stock of the file formats that we are currently storing, and of the corresponding level of storage. In addition, examples of good and bad file formats are recorded.

Ultimately, the level of knowledge involved will progress from stored (level 1) to known (level 3). Accordingly, the knowledge levels that we use here are a snapshot. The aim is to raise objects' levels, yet some file formats may never reach the highest level.

The following classification system is assigned per format:

- Stored: stored without any file format identification
- Identified: stored with basic file format identification (e.g. MIME type)
- Known: stored with extensive technical metadata and preservation watch

For a more detailed description, see the document entitled '*Richtlijnen bestandsformaten*' (File format guidelines)[31].

Functional preservation is only possible if the file format is known. Thus, a collection's

[31] [Details of the location to follow]

preservation level is linked to the level of knowledge regarding the formats that it contains.

Minimum ingest

Many checks are currently carried out during the ingest process. If any errors are detected, this can sometimes lead to material not being stored. The aim is to schedule and expand these processes after ingest, wherever possible. This will mean, for example, that more technical metadata will be available on each format, and that this could also be linked to an external register.[32] It also has the advantage that material is securely stored right away, rather than being stuck in a processing server's error directories.

This is referred to here as minimum ingest. We want to enforce checks, and we will continue to carry them out. However, most of the checks will shift from the SIP of the IP life cycle to the AIP stage, to ensure that making the material secure is given top priority.[33] As a result, the checks will take place once the material has been securely stored. Any error handling can then be performed as a documented change, in which the original file is preserved. This will enable the associated events to be recorded as well, to guarantee authenticity. That is one reason why objects need to be ingested as quickly as possible. It must also be possible to inspect these events, so that the chain of custody can be evaluated. There must be clarity about what has happened to any given object, which is why each and every file needs to be preserved. An audit log is also kept, to document all actions. Even if objects are deleted, the metadata must be retained, to provide clarity about what was once there.

The checks performed during ingest must be designed with a view to taking over the management of an object (i.e. they must be sufficient to guarantee bit preservation). Accordingly, this means that the object will initially be stored at knowledge level 1. The identification and validation of file formats within the SIP (as required for functional preservation) are carried out on the AIP after ingest. Once this has been done, the knowledge level and preservation level can be modified. The aim is to optimise the future ingest process, to further clarify the division between bit-preservation actions and functional preservation actions. In the future it will also be important to document (by means of a standard procedure) the process used to convert SIPs to AIPs, and to record these details for inspection purposes.[34] The ultimate goal of this process is to ensure that certain ingest-process requirements (e.g. material may not be encrypted, compressed, or duplicated) are met, by means of automated checks.

Expertise

Our move from bit preservation to functional preservation will require additional knowledge about digital preservation. Some of this knowledge already exists at the theoretical level, but it needs to be implemented in a workable and usable way.[35] This will involve a number of actions, such as:

- active participation in the preservation community or in communities that focus on specific solutions, such as the tools we use for quality control or file identification. We actively contribute to these communities through knowledge transfer. Not only does this provide valuable insights, it also gives us a platform that can be used to raise

[32] This would help to demonstrate compliance with requirements such as ISO-16363 requirement 4.2.5

[33] Our minimum ingest concept is derived from the Royal Danish Library's interpretation: https://en.statsbiblioteket.dk/about-the-library/projects-1/MinEffortIngest_iPRES2015.pdf. This document also shows that processes of this kind are compliant with OAIS principles.

[34] This is in compliance with ISO 16363, requirement 4.2.2

[35] This is in compliance with ISO 16363, requirement 3.2.1.3

awareness of digital preservation.

- set up a better test infrastructure featuring problem files, for example, which can be used to test tools and to perform regression tests during tool updates. This enables any potential problems in the areas of processing and long-term preservation to be quickly detected. In addition, the experience gained with new tools can be used to enhance the level of knowledge about the collections.

Collection profiles

Much of the above information about preservation levels, strategies, knowledge levels, and significant properties, as well as special context information, is constantly changing, and often varies from one collection to another. In many cases, the same information applies to large groups of objects, so there is no point in recording these details for every single object. For this reason, we deliberately avoid recording such information as metadata within the AIP. Instead, we include it in overarching collection profiles. As stated above, these collection profiles will also be important in terms of the object's information integrity. This is because they can be used to record the information needed to understand these objects in the context of a collection. This should also be in keeping with the 'interpretable' aspect of long-term accessibility. The collection profiles can also be used to indicate which objects the AIP should contain, thus clarifying this matter for internal and external users.

Preservation planning

Finally, 'preservation planning' combines all of the above-mentioned concepts. The aim is to implement functional preservation and to guarantee the long-term accessibility of objects by means of a preservation strategy, while ensuring that integrity and authenticity are conserved. In this connection, evidence must be provided to show that implementation of the preservation strategy does indeed preserve the significant properties. As a precondition, input from the various monitoring mechanisms is required, and the highest knowledge level must be assigned to the objects in question.

A preservation action plan must then be drawn up, listing the above-mentioned points, for substantiation purposes, together with details of the specific preservation action to be carried out. This document serves as supporting evidence for the object's authenticity. It can also be used as a report to show that preservation actions have been carried out in accordance with standard procedures.[36] Technical metadata will be used as input for this preservation action plan. This information can be used for various purposes, such as assessing preservation risks and formulating potential measures as part of the preservation action plan.

Information security

We use a potential threat taxonomy[37] to show how the preservation policy highlights various aspects of information security. In this context, we either refer to documents in which these aspects are discussed or we put forward proposals for any additional measures that might be needed to avoid as-yet unidentified risks.

[36] This is in compliance with ISO 16363, requirement 4.3.4

[37] Derived from http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html

The basic principles pertaining to the security of the IT infrastructure (as set out in the Security policy) are seen as preconditions for safeguards against threats such as media failure, software failure, hardware failure, operator error, internal attack, and external attack.[38] The server rooms must be secured, to ensure that unauthorised individuals cannot gain physical access to the data.

Incidents such as natural disasters are described in the Company Emergency Plan and in the Disaster Recovery Plan. In addition, the infrastructure must be designed to spread any potential risks. The principle here is to eliminate any single points of failure from the infrastructure or to draft plans that could provide solutions in the event of service interruptions. In no event must a single point of failure result in data loss. A list of single points of failure must be drawn up, to clarify the situation. Once these factors have been identified, action can be taken to resolve any problems or preventive measures can be implemented to contain them. Furthermore, no encryption, compression, or data deduplication must be used. This will facilitate simpler recovery and enable emergency response processes to be streamlined. Setting up the preservation planning process will facilitate the management of risks related to hardware and software obsolescence. Finally, a long-term policy must be developed to address any economic and organisational risks within the organisation. It should also feature a succession plan. This is a plan for transferring the collection if there are insufficient resources available to ensure the continued long-term preservation of all or part of that collection.

This list shows that information security is much more than just an IT issue, it also needs to be considered in terms of the digital collection’s long-term preservation.

Code of conduct

The security policy will have to be supplemented by other measures, to guarantee information security that is tailored to the digital collection in question. This can be achieved partly by adding preservation requirements to existing documents and partly by drafting new documents. For example, it is essential to show that the staff know which of their daily work-related actions could impact the preservation of the digital collection. At the operational level, this will involve drawing up a code of conduct and incorporating preservation-related aspects into emergency plans. In the case of migrations, for example, steps must be taken to ensure that storage media are completely erased, to ensure that objects are not made available in unforeseen ways[39]. This point will also be included in migration plans.

Disaster recovery

In addition, a disaster recovery document must be drawn up. This must clearly specify the actions to be taken when objects or information are at risk, together with details of the procedures to be followed to restore the objects in question. It is also important for these procedures to be regularly tested. The aim is to ensure that objects can indeed be restored, based on the measures that have been described. Furthermore, in all relevant projects, disaster recovery must become a standard feature of the test procedure. Here too, while it is important that provision has been made for the backup itself, the effectiveness of the backup procedure must also be verified. This could involve a restore to an acceptance environment, for example.

[38] Ibid.:Media failure, Software failure, Hardware failure, Operator error, Internal attack and External attack

[39] It is also a requirement of the Privacy and Security Directive (https://plein.kb.nl/thoughts/9360) which refers to DIN 66399, HMG Infosec Standard 5 or similar, in connection with erasing media.

Risk management as a process

Finally, in some cases, information security must also be heightened, by transforming certain initiatives into a process. For example, detailed plans for drawing up a data classification system have already been prepared. This will spotlight the requisite security level for data and for the associated systems. In addition, sessions have been held to identify the risks associated with the Digital Repository, and to assess them in terms of their probability and impact. With regard to information security, it is important that the risks identified by this means are actually addressed, based on the results of the data classification. They must also be regularly re-assessed, to verify that any risks that have been identified are indeed followed up. The same applies to the above-mentioned code of conduct for the reliable handling of a collection's data. It is not just a question of drawing this up, it must also be made verifiable and must be implemented in the context of current processes, to verify that it is actually being enforced.

Roles and responsibilities

Included in the preservation policy is the aspiration to establish roles and responsibilities for the entire organisation, in terms of preserving the collections. Care must be taken to ensure that job profiles feature the individual's role and responsibility in preserving the digital collection for jobs that involve dealing with data from the collection.
There is currently a results and development policy, as well as a personal development budget that gives staff the opportunity to pursue professional development. With regard to preservation, there is a need for clarification about how the organisation ensures that there is sufficient funding to retain substantive expertise, as well as training opportunities to boost and renew knowledge specific to that area.[40] This is also mentioned in the introduction, as one of the preconditions for preservation. The SHAMAN model was used to demonstrate that there are sufficient roles within the organisation to ensure effective preservation.[41]
The ultimate aim is to document the individual roles and responsibilities of each of the parties involved. That will provide a basis for granting access rights within the architecture's various systems.[42]

Action points

Based on the above-mentioned details, the following is a list of specific action points, relating to the above-mentioned actions, that will need to be implemented in the near future to improve the preservation of the collection and the sustainability of the related processes.

1. Drawing up collection profiles featuring details of integrity, authenticity, long-term accessibility, preservation strategy, and significant properties
2. Drawing up an IP summary structure
3. Implementation of every aspect of integrity, such as end-to-end verification for bit

[40] Ibid. 3.2.1 and underlying requirements, pp. 3-3 – 3-5
[41] Operator error and Internal attack, in particular, are defined in http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html.
[42] This is in compliance with ISO 16363, requirement 5.2.3

integrity and defining relationships for version integrity

4. Implementing the principles of minimum ingest
5. Working out the details of the interface process according to PAIMAS
6. Setting up processes for monitoring designated communities
7. Formalising/expanding preservation watch
8. Integrating contract management
9. Expanding knowledge of file formats by:
    a. participating in communities
    b. analysing the collection
    c. collecting good and bad examples
10. Defining knowledge levels regarding file formats, and drawing up an action plan to raise these levels.
11. Formalising preservation planning by drawing up preservation action plans
12. Adding preservation requirements to the Company Emergency Plan and the Disaster Recovery Plan
13. Drawing up a succession plan
14. Designing verifiable processes (that must be kept up to date) concerning:
    a. Digital preservation
    b. Data classification
    c. Risk management
    d. Disaster recovery
    e. Code of conduct
15. Defining roles and responsibilities, in connection with access rights