National Library of the Netherlands (KB) digital preservation policy

Version 1.0, November 2019

KB \ National/Library

Colophon

Publisher:
KB, National Library of the Netherlands, 2019

Editorial team: Sam Alloing, Jeffrey van der Hoeven, Brigitte Brand

Image:
Beeldstudio KB

Acknowledgements: Astrid van Wesenbeeck, Maarten Steenhuis, Judith Rog, Daniël Steinmeier

More information: jeffrey.vanderhoeven@kb.nl

Index

1 Introduction

1.1 Purpose of this document

This document contains the policy principles governing the digital preservation activities that the National Library of the Netherlands (KB) performs with regard to the digital collections in the e-Depot. These principles are based on the situation that pertained in 2019.

The structure of this document is based on the catalogue[1] developed in the context of the Dutch Digital Heritage Network (NDE).

The preservation policy will be periodically evaluated and updated. Responsibility for this has been

[1] http://wiki.ncdd.nl/index.php?title=Hoofdpagina

assigned to the manager of the KB's Digital Preservation Department (DDT).

1.2 Scope of policy and certification

This document supports the certification process that, according to the KB's 2019-2022 strategic plan,[2] was scheduled for 2019. The certification relates to born-digital material stored on the existing technical infrastructure. In the remainder of this document, everything that the KB is currently using for the long-term preservation of these digital collections (including the associated infrastructure, resources and work processes) is referred to as 'the e-Depot'.

The e-Depot stores born-digital publications produced by publishers in the Netherlands and elsewhere, together with research publications from Dutch research institutions.

For the time being, the management of other KB digital collections is beyond the scope of the current certification process. This includes the web archive and any digitised publications stored outside the e-Depot.

The KB is expanding and renewing its e-Depot in the context of the New Digital Repository programme. In the upcoming years, the KB aspires to create a long-term preservation regime (together with the associated technical provisions) for all of its digital 'preservation collections'.

In March 2019, the KB's Board of Directors decided to phase out its role as the custodian of an international e-Depot. This means that the inclusion of collections from publishers outside the Netherlands will be brought into line with the KB's new content strategy. The implications of this change for publications that are currently held in the e-Depot will be specified in due course.

1.3 The KB's digital preservation activities

The KB endeavours to ensure that as much of the Dutch textual heritage collection as possible is held in the e-Depot for Dutch Publications.

The KB is not legally required to perform a repository function. However, the Higher Education and Scientific Research Act (WHW) states that the KB 'is responsible for the national library collection'. It does not state how the KB is required to do this, nor exactly what it includes. The KB has been performing this task independently since 1974, based on voluntary contributions from publishers. The relevant frameworks are set out in the KB's strategic plan and content strategy. Since 2010, policy has dictated that digital outweighs physical and that content outweighs form.

We use the term 'digital preservation activities' to broadly describe the KB's work in electronically

[2] https://www.kb.nl/sites/default/files/docs/kbnb beleidsplan-nl.pdf

collecting digital publications, managing them, and making them available. These activities are part of the KB's mission, which is responsibility for the written word.

1.4 The purpose of preservation activities

The textual heritage must remain accessible, both now and in the future, and be protected against any identified threats. To this end, the KB's preservation activities focus on authenticity, access, display and usability, fixity, comprehensibility, identity, and availability.

Section 2 describes the principles by which these goals are achieved.

1.5 For whom are we preserving the e-Depot's collections?

The KB holds its textual heritage in long-term digital storage for current and future users. The long-term preservation of the e-Depot's current collections is for the benefit of the following designated communities[3]:

1. the suppliers (publishers, government, and research institutions), and for
2. the Dutch public, including researchers & academics.

The KB wants the next version of the e-Depot to make preserved collections available by a variety of means. These would include open access as well as text and data mining (TDM), in full consultation with rights holders and suppliers, of course.

Given its public service remit, the KB aims to make these collections accessible in the intramural setting, at the very least. In effect – for today's users – this means on the KB's premises, in The Hague.

The KB is guided by the general principle that public access to collections that are subject to long-term preservation must not distort the market.

2 Preservation themes

2.1 Authenticity

Integrity

The integrity of any publications supplied to the KB will be safeguarded throughout the entire preservation process, to ensure that they cannot be inadvertently changed.

[3] Glossary in the annex on page 19

For example, the KB carries out checksum checks, and reaches agreements with the suppliers concerning the integrity of their data. From the moment they are delivered, digital objects cannot be inadvertently changed. Any changes that do occur are planned. The KB carries out periodic integrity checks (bit preservation).

Authenticity

The KB guarantees that, at all stages of the preservation and access process, the publications received are a complete and accurate representation of what they prove to be, or intend to be. To this end, the KB employs transparent procedures and guidelines (such as preservation strategies) and uses metadata to describe the object's content, context, provenance information, and preservation history.

The KB aims to store all versions of each digital collection object (whether it was born digital or digitised). The KB also records details of every action to which the data have been subjected, to show that these have taken place correctly (chain of custody). The details of all relevant actions are recorded in event metadata, for example.

The KB endeavours to preserve the look and feel of the publications it receives. This means that, during any future use, it should (ideally) be possible to reproduce the original form in which the work was made available to the consumer, based on the objects stored by the KB. If priorities are to be set, however, the preservation of the textual content and structure of publications takes precedence over the preservation of the wide variety of publication formats (including online publications).

Reliability

The KB aspires to be a reliable source for users of the collections, both now and in the future. It must be possible for the designated community to have confidence in the KB's ability to store all data correctly, and to handle the data provided by suppliers in a reliable manner. To that end, the KB will diligently document each of its preservation processes.

Security

The KB endeavours to guarantee the continuity, integrity, and reliability of its digital collections, and to protect them against any internal and external threats. The collections are stored in data centres that are diligently managed and monitored. The infrastructure is subject to a strict information security policy. There must be no data loss. To this end, the KB classifies its data and digital collections into various security classes. Based on these security classes, the KB takes appropriate measures to secure the digital collection. The management of the systems and of the data are separate responsibilities, accordingly the associated roles and authorities are authorised separately.

Any information held on data carriers that are to be destroyed is diligently deleted, or made unusable, during the process. This prevents data from accidentally falling into the wrong hands.

The KB has an Information Disaster Recovery Plan and a Disaster Recovery Plan for emergencies or major incidents in the area of information security.

Provenance information

Users must always be made aware of the sources of information. Accordingly, for each object, the KB records information (in the preservation metadata) that can be used to trace the object's provenance. This enables users to assess an object's authenticity.

2.2 Bit preservation

Bit preservation is the basic level of preservation. It ensures that the stored digital objects remain unchanged during the preservation process, and enables preservation to be monitored.

The KB preserves all objects in the form in which they are delivered, at bit level, which enables it to guarantee the integrity of every object it receives. This process involves triple replication, the use of Write Once Read Many (WORM) technology with quadruple redundancy per node, and periodic checksum checks.

The KB preserves the collections in at least two geographically separated physical locations. It also wants at least two different technologies (forms of storage) to be involved, to reduce any vulnerability to software failures and operator errors. The KB draws a distinction between system management and data management.

The KB's aspiration is to put disaster recovery and restore procedures in place, such that, in the event of a major emergency, the e-Depot can be fully restored within a period of three months.

The KB uses one level of bit preservation, and does not distinguish between different collections and objects in this regard.

The KB checks all material upon arrival for completeness and bit integrity (a zero-byte check and a completeness check).

Persistent identifiers

The KB assigns a persistent identifier (PI) to each object (Archival Information Package, AIP) in the e-Depot. These PIs are actively managed, to ensure that objects remain accessible via the PI. In the future, the KB wants to be able to assign a PI to each logical level in a publication. The PI currently

assigned by the KB is the National Bibliography Number (NBN)[4]. Any PIs provided by the supplier (such as a Digital Object Identifier, DOI) are held in long-term storage (in the AIP).

2.3 Functional preservation

Planning functional preservation

The KB currently has no functional preservation regime. The KB's guiding principle is to ensure that it has relevant knowledge concerning any material that is brought in and deposited (preservation watch). If a preservation action is required, then a preservation plan is drawn up. If necessary, the significant properties of the various information objects and collections are also identified.

Preservation strategies

The KB endeavours to pursue strategies that are appropriate to the preserved material, if the situation so requires. This is based on best practices.

The KB could potentially pursue strategies such as the characterisation and validation of file formats, the migration of file formats, preservation watch, technical metadata extraction for risk analyses, and preservation actions based on identified risks.

Ingest and preservation actions

The KB records the provenance information (if supplied by the publisher) and checks the authenticity of the content that has been received. Full details of any processing to which the information objects and metadata have been subjected from the moment of receipt are recorded in an event log.

The items recorded by the KB during the ingest process include intellectual property rights, descriptive metadata, technical metadata, and structural metadata.

Version management during storage migration

The KB employs full version management. This means that the KB stores each modification of digital objects, metadata, and other content as a new version of the original object.

2.4 The digital object

Original object

The KB always preserves the original object. The form in which the objects were originally supplied

[4] https://www.kb.nl/organisatie/onderzoek-expertise/informatie-infrastructuur-diensten-voor-bibliotheken/registration- agency-nbn/principes

remains accessible, wherever possible. This means that, in theory, it is possible to inspect the objects and metadata relating to the original structure and content. Any changes made to the object are recorded.
As stated in the data storage contract, the KB does not accept any objects that involve Digital Rights Management (DRM).

Removal of objects

In general, the KB does not remove any digital objects once they have been stored. In exceptional cases, an object can be rendered inaccessible to public inspection. If a supplier expressly requests that an object be removed from the e-Depot, the KB will comply with this, in accordance with the relevant agreements in the data storage contract.

Monitoring file-format developments

The KB conducts a limited preservation watch, to ensure that objects remain available to future users. Based on this, guidelines and procedures are drawn up to accommodate any changes or adjustments that may be required.

Rendering objects inaccessible

The KB makes material available, within legal constraints and subject to any agreements reached with rights holders. It will only render material inaccessible in exceptional circumstances, such as those described in the data storage contract, or in the event of a court request. In some cases, the KB itself may determine that legislation and regulations have been infringed (in the case of major violations of copyright, privacy, or decency legislation, for example).

2.5 Metadata

Metadata management

The KB collects and creates metadata on all digital objects and collections held in the e-Depot. It also monitors the quality of this metadata. In every case, metadata is stored in an AIP, together with the preserved information object. This serves to guarantee the object's future usability. The recognisability and comprehensibility of the collections and information objects for the designated communities is safeguarded by means of descriptive metadata (see subsection 1.5). Metadata relating to rights, and to rights holders, is used to safeguard the appropriate access, use, and processing rights. Full details of any preservation actions are recorded as metadata, which is stored with the object.

The KB uses an in-house data model. It is developing an information model which links the digitally preserved collection to other KB services. Its in-house data model requires the KB to adhere to the following standards: PREMIS (for preservation metadata), METS (for structural metadata), MIX (for technical metadata) and MODS (for descriptive metadata).

Original metadata

The KB stores the original metadata as a file in the AIP, as provided by the supplier. For reasons associated with management and use, the original metadata is embedded in the AIP metadata. The KB works with the standards provided by suppliers, such as ONIX, NL EDU Standard for repository harvesting, NLM, and the suppliers' own standards.

Descriptive metadata

In the context of its e-Depot, the KB uses the MODS standard to standardise or normalise any descriptive metadata it receives. The resultant general format is used throughout the entire e-Depot, for all formats. MODS is one of the standards that have been officially recommended for use in METS' descriptive metadata section. The MODS standard is also widely used in libraries.

Preservation metadata

The KB adheres to the PREMIS model, and works with its own PREMIS application profile. This contains details of the full interpretation of the PREMIS model, tailored to the preservation of the collection.

Accordingly, this is in line with other KB work processes (such as harmonisation with the RDA model). We record all events, for purposes of authenticity (see also page 5). The KB can make this data available to its designated communities. This policy assures us of a better exit strategy as it means that we have an archive log that we can continue in any subsequent version of the preservation system.

Structural metadata

The focus of the KB's policy is to ensure that any digital objects it receives remain intact. The KB creates an AIP manifest, based on what it received. We use the METS standard for this purpose, together with any associated files that we have received from the supplier.

2.6 Rights

Legislation and regulations

The KB has drafted its own copyright policy, based on current legislation and regulations. The existing copyright policy was adopted in 2011. The KB plans to update that policy in 2020. The details of any departures from that policy are submitted to the Board of Directors, for decision-making purposes.

The KB's digital preservation activities are subject to the following legislation and regulations:

The Higher Education and Scientific Research Act (WHW)[5]. The WHW provides no guidance concerning the precise interpretation of the KB's remit (and digital remit).

> Copyright Act[6] – the KB's entire born-digital collection is subject to copyright. Articles 15h + 16n, in particular, are relevant to its digital preservation activities:
– Article 15h enables the KB to make content available in the intramural setting. – Article 16n enables the KB to create digital copies for long-term preservation.

> Directive on Copyright in the Digital Single Market (DSM Directive)[7] – It is anticipated that this directive will be implemented in the Copyright Act by mid-2021. The following articles in the Directive are particularly relevant: Article 3, Text and data mining, Article 5, Preservation of cultural heritage (= preservation copy), and Article 7, Cultural heritage institutions' use of works that are no longer on the market (= extended collective licensing, ECL for out- of-commerce works).

The Diligent Search Requirement for Orphan Works Decree (*Besluit zorgvuldig onderzoek verweesde werken*)[8] is based on the Copyright Act, Articles 16o to 16r.). Orphan works are always protected by copyright. The KB also incorporates orphan works into its e-Depot. However, this Act is not particularly relevant to the born-digital collection in the current e-Depot, as suppliers/publishers are unlikely to supply any orphan works.

> The General Data Protection Regulation (GDPR)[9] and the General Data Protection Regulation

[5] Higher Education and Scientific Research Act)
[6] Copyright Act
[7] Directive on Copyright in the Digital Single Market (DSM Directive)
[8] Diligent Search Requirement for Orphan Works Decree (*Besluit zorgvuldig onderzoek verweesde werken*)
[9] The General Data Protection Regulation (GDPR)

Implementing Law (GDPRIL)[10]

> Databases (Legal Protection) Act[11]: applicable to databases supplied to the KB (a database is released 15 years after production, unless notification of changes is given).

> Temporary Decree on the Digital Accessibility of the Government (1-7-2018)[12]: applicable to the website through which the KB makes publications from the e-Depot accessible. This website must be digitally accessible to everyone before 23 September 2020.

Documentation of object creators and rights holders

For each object, the KB records the identity of the supplier and details of any rights holders. It uses the rights metadata for this purpose, as part of PREMIS.

Deposit contracts and data storage contracts

An umbrella agreement has been reached with publishers affiliated with the Media Federation, who have each committed themselves to deliver to the depot. The KB draws up a data storage contract with all supplier parties, setting out agreements concerning the delivery and processing of content, the long-term preservation of content, and the availability of content.

A standard data storage contract applies for those suppliers with whom no separate agreements have been reached concerning bulk delivery, and who deliver individual publications via the *Webloket* (Web Service Point). This standard data storage contract is based on the agreement reached with the Media Federation (see above).

Legal context of preservation activities

The Netherlands does not legally require the National Library of the Netherlands to perform a repository function. Accordingly, publications are held in long-term storage wherever possible, with the express approval of the party that supplied them. As part of the agreement, KB may make any changes to the objects that are required for the purposes of long-term accessibility. An exception is the web archive, which is subject to an opt-out procedure. The data storage contracts set out conditions for processing, preservation, management, and availability.

Any works that are free of rights (public domain) are, where possible, made freely accessible (in consultation with the supplier or the former rights holder).

Any works that are published in open access are subject to long-term preservation. The KB makes them available in the e-Depot. If a rights holder should make themselves known to the KB, they can request

[10] The General Data Protection Regulation Implementing Law (GDPRIL)
[11] Databases (Legal Protection) Act
[12] Temporary Decree on the Digital Accessibility of the Government (1-7-2018)

that access to the work in the e-Depot be made compliant with the applicable rights and agreements.

2.7 Standards

Principle concerning the use of standards[13]

The KB acknowledges that:

> with regard to the development of the digital archive, its digital preservation activities comply with the standard reference model – the Open Archival Information System (OAIS)[14];
> when developing and maintaining its organisational and technological context, it will adhere to the high standards of the community's standards and best practices;
> it will participate in the development of digital preservation standards and their dissemination.

[13] http://wiki.ncdd.nl/index.php?title=Duurzaamheidsbeleid/Beleidsuitwerking-Legal context for preservation activities
[14] https://www.kb.nl/sites/default/files/docs/sierman oiasmodelned.pdf

Reference model

Where appropriate, the KB endeavours to comply with the OAIS reference model. We also operate, wherever possible, in keeping with DUTO (Standards Framework for Sustainable Accessibility to Government Information) and DERA (Digital Heritage Reference Architecture).

Use of specific standards

With regard to its digital preservation activities, the KB applies current community-based standards to the development and management of its organisational and technological context. Where possible, open, international and documented standards are used. Open standards guarantee the interchangeability of data, information, and collections between different systems, both now and in the future, while also contributing to digital preservation. A key side effect of open standards is greater freedom of choice in terms of suppliers (and, as a result, reduced dependency on them).

2.8 Access

Usability

The KB makes every effort to ensure that future users will be able to inspect the contents of preserved objects, in the context in which those objects were published. The KB endeavours to preserve the look and feel of publications. This means that, ideally, it should be possible for the original form in which the work was made available to the consumer to be reproduced during any future use, based on the objects we have stored.

Digital Rights Management (DRM)

When granting access, the KB respects the rights of the supplier, but it does not store any DRM-protected objects for the purposes of its own preservation activities. Where appropriate, we remove the DRM if preservation activities so require. Agreements are reached with the publisher about this.
The KB uses metadata to record the user's rights, in terms of what they can do with the content. It also monitors compliance with these rights in its processes and systems.
The KB makes its preserved collections available in accordance with agreements reached with the publishers/suppliers in question. The basic level of access is the intramural setting.

The Dissemination Information Package (DIP) always contains information concerning the origin of the object being inspected. The user can see that it is a KB version and that the original publisher accepts no liability for any deviations from the original.

Comprehensibility

The KB endeavours to ensure that the preserved objects remain readable and understandable for the

designated communities.

Search facilities/access method

The KB makes the preserved objects available via a simple user interface.
The KB aims to make the preserved objects/collections available in a way that meets the needs of the designated communities (see: For whom are we preserving the e-Depot collections?).

2.9 Organisation

Staffing policy, roles & responsibilities

We operate a KB-wide job classification system, in which various roles are defined. Each role consists of multiple levels, in which details of the requisite knowledge and skills are recorded. This also applies to digital preservation.

The manager of the Digital Preservation Department (DDT) is responsible for the comprehensive vision of digital preservation at the KB, and for performing the tasks associated with that vision. This includes an adequate staffing policy, to ensure that all of the processes necessary for the sustainable archiving of digital collections in the e-Depot can be actually carried out.

Since 2016, the KB has been using product portfolios, product management, and product life-cycle management. Each product involves a number of fixed roles and responsibilities. Product managers are responsible for the product's continuity and further development at tactical level (in this case the e-Depot). They are supported at operational level by business information management, IT management, and service management. All business information managers are BiSL certified. Change management involves several standard processes for renewing (changing) the e-Depot. These are safeguarded in a KB-wide change process, which is supervised by the Service Management department.

The KB (more specifically, the DDT department) employs metadata specialists and digital preservation officers. These members of staff interpret the KB's metadata policy and preservation policy at strategic and tactical levels. They also advise the entire organisation about the preservation of digital collections, the preparation of preservation plans, and the implementation of preservation actions. In addition, the Research department's staff includes specialists in digital preservation who focus on the areas of audit & certification, preservation tooling, standardisation, and formats.

The preservation of digital collections touches on many of the organisation's business processes. As a result, the Digital Preservation Department is not the only department involved in this area. The KB aims to further define the associated roles and responsibilities in the upcoming years.

The KB has an active Result & Development policy, to further the development of knowledge and skills.

The policy encourages every employee to set clear results and development goals, in the context of a yearly cycle. The KB has a budget for organisation-wide training and education. Staff who are involved in digital preservation are expected to follow the 'Learn how to Preserve' training course. They regularly have opportunities to attend conferences and workshops in their own field.

Risk management

The KB's Governance, Risk and Compliance (GR&C) department is responsible for structuring risk management for the entire organisation. With regard to preservation, it is especially important to spotlight any risks that might impact the digital collections' long-term accessibility. In theory, this is part and parcel of the DDT department's remit. Details of the organisational and financial aspects of preservation are presented under 'staffing policy' and 'budgets'.

Any risks in the areas of unlawful use of – and unlawful access to – digital collections are managed by recording rights metadata with the collection, and by making retrieval subject to access rights.

The KB has an up-to-date information security policy, which will enable the organisation to take effective action in the event of an incident, emergency, or crisis. This is intended to counter any risks and threats aimed at the digital collection and, more broadly, the IT infrastructure (such as system failure, a data breach, or cyber crime). To this end, the KB has a crisis management team, relevant scenarios, a Company Emergency Plan, a Disaster Recovery Plan, and an Information Disaster Recovery Plan. However, these plans do not yet take sufficient account of digital preservation. The aspiration is to bring these plans into line with the KB's expanded digital function.

The above measures will be further formalised in the future, and details of any associated decision-making will be recorded. The KB will also continue to evaluate measures and will set up notification processes in various key focal areas.

Budgets & cost estimates for preservation

Any costs related to acquisition, processing, preservation, and keeping digital collections accessible are covered by the KB's general budget. Accordingly, these activities are not dependent on temporary resources.

Since 2017, the KB has used the NDE cost model for digital preservation, which enables it to make accurate forecasts and to measure implementation more effectively. The costs calculated using this model are updated annually.

Since 2012, the KB has also used a fixed cost price calculation (Total Cost of Ownership, or TCO) for digital storage, which is assessed by an accountant. This TCO covers a total of 15 different cost components for the storage of digital data. These include costs relating to hardware, software, power

and cooling. They are broken down into various forms of storage – temporary archiving, long-term archiving, and rapidly accessible storage. Based on this cost price (which is expressed in terabytes per year), it is possible to calculate the anticipated costs of digital storage in the upcoming years.

In the upcoming years, the KB will substantially revise the cost price calculation for preserving the digital collection. On the one hand, efficiency gains in the IT infrastructure and economies of scale are expected to cut costs. On the other hand, more effective preservation of the digital collection and the growth of the collection itself will push costs up. The KB is preparing these forecasts in a bid to prevent or absorb potential cost price fluctuations in future. Forecasts are drawn up with a five-year horizon.

Preservation goals

The KB has set itself a number of preservation goals for the upcoming years:

Switch from bit preservation to functional preservation.

> Further our knowledge of the collections (including knowledge of the formats in use) by means of an inventory, as well as research, tools, technology watch, and by recording this information in collection profiles.
> Perform more checks, based on reports.
> Implement processes that impose preconditions on functional preservation,
> such as technology watch, monitoring designated communities, risk management, and preservation planning.
> Formalise and embed OAIS functions in the organisation's business process.

3 Certification

Standard for Audit and Certification

The KB is endeavouring to obtain certification for the born-digital material that is stored in its e-Depot. As a first step, we are aiming to achieve certification via the Core Trust Seal (CTS)[15]. This aspiration is set out in the 2019-2022 strategic plan.

4 Annex: glossary

Archival Information Package (AIP):
the information package that has been selected for inclusion in the archive.

Business information Service Library (BiSL):
a framework for business information management.

[15] https://www.coretrustseal.org/why-certification/

Bit preservation:
a term denoting a very basic level of digital material preservation – preservation of the original order of ones and zeros.

Born digital:
material with no analogue equivalent.

Chain of custody:
the chronological documentation or paper trail that records the order of custody, monitoring, transfer, analysis and possession of physical or electronic evidence.

Checksum check:
a unique numeric signature derived from a digital file.

Designated community:
an identified group of potential consumers who are considered to be capable of understanding a given set of information.

Dissemination Information Package (DIP):
the information package that is retrieved and viewed by the user.

Event metadata:
Metadata recording the chain of custody.

Fixity:
proof that an object is unchanged.

Ingest:
the process of incorporating a digital object into an archive, which focuses on the long-term storage and management of the selected material, and on keeping it accessible.

Node:
a device or structure that can be considered to be an independent unit.

Persistent Identifier (PI):
a long-term reference to a digital source.