Strategy for long term preservation of digital collection materials at the Royal Library

This document is the second revision of the Strategy for long time preservation of digital collection material at the Royal Library.

1. Introduction
This document, "Strategy for long term preservation of digital collection materials at the Royal Library", describes the principles behind the implementation of the library’s policy in this area and how priorities are assigned in digital preservation. This strategy also forms the basis for the specific preservation plans for the individual digital collections.

Digital preservation involves a series of activities, which must ensure that the library’s digital collections can be preserved over time, across changing technologies and user needs. The activities include preservation of the physical bits (bit preservation) as well as ensuring that the preservation system continues to be able to interpret these data and convert them into understandable information for users (logical preservation). In addition, the activities must ensure preservation of the authenticity of the collections and their integrity over time, so that users can have full trust in the information that the Royal Library makes available.

The strategy is expected to be implemented gradually in the coming years.

1.1. Purpose
The overall purpose of the strategy is to ensure that the Royal Library at all times employs the most appropriate modes to achieve the goals set in the policy for digital preservation, so that the digital collections are preserved as well as possible within the given framework of legislation, institutional policy and economy. The strategy must ensure that there are clearly stated priorities that make it possible to carry out an optimal policy in relation to the budgets of the Royal Library. This includes
ensuring that the library must continuously carry out the investments in infrastructure and knowledge, which are necessary to accomplish these preservation activities.

The strategy must also ensure that the development of competences and choice of system solutions are such that the Royal Library remains an attractive cooperative partner within digital preservation both nationally and internationally, and that the institution thereby can both contribute to, and gain by, common developments.

Finally, the strategy is an element in the Royal Library’s efforts to document and create transparency with regard to preservation, so that the institution’s preservation systems through audit and certification can achieve the status of a Trustworthy Digital Repository.

1.2. Activities covered by the strategy
The strategy covers all the activities needed to ensure the Royal Library’s digital collections are preserved over the long term in such a way that gives the users the best possible access to use of the collections. These activities range across organisational activities, which support fulfilment of the Royal Library’s mission and current legislation in this area, as well as ensuring that the Royal Library has the necessary economic and personnel resources to carry out these responsibilities. In addition, it covers management of the digital objects from ingest, qualification and preservation, and until they are made available for the users of the library. And finally the activities include the supporting technical and security infrastructure.

The activities within digital collection preservation are thus an integrated part of the functions of the library, with interfaces in common with many of the other departments and tasks, which requires a clear division of roles and areas of responsibility.

1.3. Collections covered by the strategy
The strategy covers all the Royal Library’s digital collections deemed worthy of preservation, both those of born digital materials and digital copies of physical (see also “Appraisal policy for digital materials at the Royal Library” [document in preparation]).

Preservation of physical material, even physical media with digital information such as tapes, cds and dvds, is not covered by this strategy. The document will eventually contain a section, which describes the Royal Library’s strategy for media conversion.

Nor is the preservation of legal deposit Internet material included, since it requires its own policy and strategy, which must be established in close cooperation with the State and University Library.

1.4. Relationship to the Royal Library’s other planning documents
This document is a part of a series of top-level planning documents, which in addition includes:

- "Policy for long term preservation of digital collection materials at the Royal Library"
1.5. Updating the Document
This document is updated and approved by the Board of Directors of the library bi-annually. Responsibility for the revision rests with the person who has the operative responsibility for the activity of digital preservation, at present the department head of Digital Preservation in the National Library area.

This document is published on the library’s website in both Danish and English versions, in order to facilitate general international evaluation and development of policies and strategies within digital preservation.

1.6. Guide to reading this document
Section 2 describes general principles for how the Royal Library implements the policy for digital preservation. Section 3 describes the strategies in the organisational framework, section 4 the strategies for how the Royal Library manages the digital material and section 5 describes the strategies regarding the technical infrastructure on which the preservation systems are based. Finally you will find vocabulary and literature in section 6 and 7.

2. General principles

2.1. Risk Management
The Royal Library employs risk management to protect its investments and assess the factors, which can affect the ability of the institution to deliver credible long term preservation. Through the identification of risks, evaluation of their probability and consequences, as well as the possibilities for action, the Royal Library analyses the potential problems and describes the challenges. Such risk evaluations are performed bi-annually for the directors and form the basis for decisions on adjustments of the preservation policy and strategy. In addition they deliver input to the Royal Library’s Security Policy.

2.2. Audit and Certification
As a part of risk management the Royal Library works to implement internationally recognized standards and methods for audit and certification, such as TRAC or DRAMBORA, in order that the Royal Library’s preservation systems can be certified as a Trustworthy Digital Repository.

In the first place the Royal Library intends to carry out a self-certification, in order to ready the preservation systems for an actual external certification.

2.3. Quality Control and Documentation
The Royal Library works continuously to control the quality of all preservation activities, including work processes and tools. An important element here is to ensure complete documentation for all activities and that the right persons have access to it.
2.4. Standards and Open Source Programs
The Royal Library works to live up to authoritative national and international standards within digital preservation, and therefore follows developments of these in relevant forums.

Moreover the Royal Library supports the development and use of open source programs, which is a clearly necessity in this area, where as yet no complete packaged solutions are available.

3. Organisational infrastructure

3.1. Legislation
The Royal Library determines for the individual collections and objects, whether there are intellectual property, personal data or contractual issues which require registration and control of access. In order to meet these legal requirements the Royal Library uses exclusively solutions where access can be registered and regulated on the object level as well as at the level of the collection.

3.2. Economy
The Royal Library operates with a four year planning horizon with associated gross budgets, as well as with a one year horizon with a detailed budget for the area of digital preservation. The specific costs of preservation are assessed on the basis of data volume and expected growth, data formats, preservation level, expected frequency of platform migrations, and from the requirements for ingest and access to the collections. These data are collected annually for statistical purposes from a specified date.

The Royal Library makes every effort to achieve balance between needs and resources, and clarifies for the directors when this cannot be achieved.

The Royal Library is working to implement a cost model which calculates all direct and indirect life cycle costs of digital administration and preservation, and which, among other things, can be used to evaluate alternative preservation scenarios.

3.3. Personnel and Competence Development
The Royal Library is working to ensure that the library always has sufficient, up to date and professional personnel so that the preservation policy can be maintained and carried out responsibly whether it is done in-house or by outsourcing.

The subject field of digital preservation is not thoroughly supported by institutions of higher learning and it is only dealt with at a few institutions. Therefore the number of graduates with prior knowledge of the field is very limited, and long introductory periods of on-the-job training are most common. This means that the Royal Library attempts to prevent a high flow-through of staff, and, as well, the ordinary practice of documentation and doubling up of knowledge is intensified.

11.september 2014
Participation in national and international research projects comprises one mode of attracting and retaining the necessary qualified personnel at the Ph.D. level. Active membership in national and international specialist networks similarly develops competence.

The Royal Library maintains an overview of which competencies are required for problem solving in relation to digital preservation, including job descriptions and competence development plans. In addition there is a suitable budget for training and further education of personnel in digital preservation.

3.4. **Roles and Responsibility**

Responsibility and roles in the tasks of digital preservation are divided according to the Royal Library’s Security Policy.

The vice-directors for the National Library area and Copenhagen University Library respectively have the overall responsibility for the digital preservation of collections, which belong to their areas. The National Collection Department maintains an overview of all the digital collections and the decisions, which have been taken about them and have responsibility for coordination of the preservation activities between the vice-directors of the National Library area and Copenhagen University Library and the functions ”Digital Preservation” (DB) and ”Digital Infrastructure and Services” (DIS).

The National Collection Department is responsible for the digital collections and decide in consultation with the Department of Digital Preservation issues regarding preservation formats, normalisation of data at ingest, access controls etc.

The National Collection Department has responsibility for overview and control of the holdings and the access arrangements, as well as carrying out regular revisions.

The National Collection Department is responsible for contact to producers with regard to delivery and acquisition of digital collections, including identification of all legal relations associated with the collections and for making the necessary agreements.

The National Collection Department is also responsible for identifying and contacting current and future user groups, which wish to have access to the digital collections.

DIS and DB share the responsibility for the technical implementation and operation of the administration and preservation systems.

4. **Object Management**

The Royal Library works to establish well-defined and quality assured workflows for ingest, preservation and access to the collections, which ensures that both the objects and the preservation processes are supervised and documented.

Long term digital preservation can only be done through the Royal Library’s administrative systems since these are used to register the library’s own digital collections.
The Royal Library must have an object model (data model), which supports the object types for which the Royal Library has preservation duties.

4.1. Ingest

At ingest into the Royal Library’s preservation systems, a concrete preservation plan must be prepared for each collection, which, i.e. fixes workflow, preservation method and bit preservation level.

As a starting point, born digital material produced outside the Royal Library is always preserved in the long term in its original format to ensure authenticity.

At ingest of a digital collection into the Royal Library’s preservation systems an evaluation is made of whether there is a further need for normalization to another data format than that in which it was delivered, in order to adapt it to the Royal Library’s current preservation policy. Resources limit the number of preservation formats to as few as possible, unless it is a case of collections which have been chosen for preservation through an emulation strategy.

The Royal Library is working to establish an automatic quality control and validation of all objects in connection with ingest into preservation systems, e.g. with tools such as DROID and JHOVE2. As a part of this, checksums are generated on all objects and these are documented in system logs. In this way the Royal Library will in future be able to control the integrity of the objects.

For some more dynamic types of objects, e.g. teaching games and online games on the net, it can support the preservation activities to establish an extra representation of the object in the form of a filming. The interaction in the dynamic element itself disappears, but the new representation documents parts of the object and can be more easily preserved than the original object. The result of the filming, the new digital representation, is itself subject to digital preservation.

In connection with ingest each object is also given a unique identifier (UUID).

Insofar as it is needed, objects are also given a persistent identifier (PID).

A listing of the full contents is added to all new collections.

Digital material designated for long term preservation is submitted to the following four preservation activities:

- Preservation of the physical bit sequence
- Preservation of the ability to retrieve the materials through the use of unique identifiers
- Preservation of information or structures which tie the elements together
- Preservation of the ability to interpret the physical bit sequence
4.2. Bit Preservation

Long term preservation includes as a minimum secure storage of the physical bits, here understood as storage in systems with well known risks, as well as control of integrity and restoration of possible errors. The Royal Library’s strategy is to offer several levels of preservation in order to be able to respect the collections and the users’ various requirements for security, privacy and speed of access. This differentiated storage moves from a secure copy with backup to multiple copies with active bit preservation. By bit preservation is understood continual integrity check and error correction.

For the solutions which are part of the Royal Library’s bit preservation the specifications will be updated continuously to ensure organisational, technological and geographic diversity. There are fixed established routines for re-establishing and logging in line with the Royal Library’s security policy.

4.2.1. Bit preservation levels

The Royal Library’s different bit preservation levels are listed below with a rising degree of security and costs:

Level 1: The digital material is placed in a secure storage with backup and with a quarterly backup located over 150 km from the Royal Library.

Level 2: The digital material is stored with active bit preservation in a minimum of two complete copies supplemented by an extra copy of the checksum, so that the three match in the integrity check. Moreover, there must geographic separation between the complete copies (separation of more than 150 km).

Level 3: Level 2 supplemented with one or more copies of the checksums, where these are geographically separated from the data. Further security can be obtained by the use of different checksum algorithms, conceivably in combination.

Level 4: Level 2 storage of the digital material, supplemented by active bit preservation in a minimum of two copies where there is organisational, technological and geographical independence between the copies.

Level 5: Level 4 with one or more copies of the checksums geographically separated from the data.

For level 4 and 5, the preservation system must ensure that bit preservation cannot be corrupted by one organization.

If compression is to be used, this requires a new level of bit preservation.
4.2.2. Integrity check

The number of integrity checks must be set in relation to the number of errors which occur. This means that statistics must be continuously kept and all errors logged, so that there is a basis for adjusting the frequency of integrity checks. If the preservation format is also used for dissemination, then the frequency of the integrity checks can be reduced, if all else is equal, because the preservation formats are controlled in connection with continual access. However, the Royal Library uses derived formats for dissemination in the case of many types of materials, which increases the requirements for frequency of integrity checks.

4.2.3. Packing of data

Packing of data must be done simply, and as independently of the system as possible. To reduce maintenance costs and endure robustness, data should be packed in a simple format with good chances for a long term existence, e.g. due to frequent use and ISO standardization. Files should be uniquely identifiable, independent of the formats they contain, and references (e.g. between files and metadata) should be easy to retrieve independent of the systems. In accord with the Royal Library’s policy for digital preservation, not only must the data be preserved in the long term, but also the relevant metadata. The packing format must thus support that updated metadata from the administration system can be transferred to long term preservation after data has been long term preserved.

The Royal Library uses the WARC format for packing digital material.

4.3. Data Management

The Royal Library’s data model must strive to be as simple and system independent as possible with regard to future maintenance. At the same time, it must be sufficiently flexible to be able to manage new types of material and new technologies.

All digital objects must have an identifier, which ensures that the Royal Library retains the ability to retrieve objects and refer to them.

Preservation activities on the individual objects are documented by audit trails.

The Royal Library maintains a complete listing of the holdings of digital collections, where these are described on a general level, which includes the owner of the collection, its size and the chosen level of preservation. This list will eventually also cover access issues and legal conditions.

The Royal Library employs, insofar as it is possible, international standards, such as METS, MODS and PREMIS, which are suitable in relation to preservation and dissemination systems, and which can be integrated with the Royal Library’s object management system (SIFD) and the underlying data model.
4.4. Preservation Planning

4.4.1. Logical preservation
Beyond bit preservation, the Royal Library’s preservation activities must preserve the digital materials from technological obsolescence, so that both now and in the future it will be possible to read, understand and display/play the materials with standard programs and equipment.

4.4.2. Logical Preservation strategies
The Royal Library operates with two logical preservation strategies: migration and emulation. The strategy chosen depends on the specific type of material, and the functionalities and properties of the material that the Royal Library wishes to preserve (significant properties). As long as there are no packaged emulation solutions, migration is preferable. For example, the Royal Library employs a migration strategy for digital images. For the collections where migration is not a realistic possibility because significant properties cannot be preserved, management of the collection is organized with regard to future emulation.

For each collection a preservation strategy to use must be determine. Insofar as the plan is for future emulation, secondary material (programs, licenses, documentation etc.) must be identified and collected.

4.4.3. Media Conversion
A strategy should be developed for the media conversion of all relevant collection materials. (Sound, video formats; physical negatives and paper materials for substitution digitalisation; computer games; artists’ and researchers’ archives in electronic media)

4.4.4. Preservation Plans
It is the Royal Library’s policy that there must be preservation plans for all digital collections, which describe specific conditions with regard to how the collections are to be preserved and used.

The preservation plans must at the latest be prepared in connection with the ingestion of the collections into the preservation system. Thereafter the preservation plans must be brought up-to-date according to need, at a minimum in connection with preservation activities.

To support this activity the Royal Library uses an up-to-date preservation planning tool supported by OPF.

4.4.5. Data Formats
In choosing data formats the Royal Library as far as possible chooses open, widespread (global and used by memory institutions), standardized, error tolerant formats with a reasonable assumption that they will be able to be used within one of the preservation strategies which the library chooses to employ.
The Royal Library is working on preparing a list of formats which are suitable for preservation of the library’s various types of materials. Until this list is completed, the Royal Library follows the guidelines made by the Library of Congress (LOC).

The Royal Library views formats to be threatened if the format’s global distribution and use is declining too strongly or if the number of tools that support reading from or writing to the format is falling dramatically. If an earlier suitable format no longer is suitable, the Royal Library must ensure that the preservation plans are revised.

The goal is to keep annual statistics over which formats the collections include (the number of files and the number of TB).

4.4.6. Compression and encrypting

If possible, data is not compressed. If compression is viewed as necessary, in choosing the algorithm (and implementing the algorithms) the choice should be lossless, open, efficient (minimal time for packing and unpacking of both large and small files) and widely used algorithms.

Insofar as a digital object must be compressed for storage the process should be carried out and documented by the Royal Library itself.

If the data is to be encrypted, the Royal Library should use standardized encrypting algorithms.

4.4.7. Technology Watch

Preservation strategies based both on emulation and migration require constant follow-up on the developments in standards, formats, tools, system solutions, as well as the continuous requirements for support of dissemination which DIS demands.

The formats included here are data formats, metadata standards (descriptive, technical and administrative), packing formats and exchange formats. The tools cover both tools for quality control (characterization and validation) and integrity checks, as well as migration and emulation. Systems solutions focus on object management and storage.

The Royal Library carries out technology watch by studying international watch reports and participating in international conferences and OPF activities.

4.5. Access

The Royal Library endeavours to make the collections continually available in up-to-date data formats.

When the decision is made as to how a collection is to be made available, a decision is also made as to whether the preservation format is suitable for dissemination or whether a derived format must be generated. If a special dissemination format must be generated, a decision must be taken as to whether it is to be generated ahead of time, or on-demand.
Access to data must ensure compliance with the laws of copyright, personal data, public administration and other legal contracts and agreements.

As a part of management of the collections, user access to create, update, delete and view the digital objects must be regulated on an individual level with the use of user ID and password.

Moreover, quality controlled interfaces (APIs) must be developed for machine access to data, e.g. to use for results for search interfaces, document delivery, generation of derived copies and performance of preservation processes such as integrity check and format validation. Finally there may be access requirements for mass processing of large quantities of data (data mining).

The aim is that both access and changes in data are to be registered in a system log, regardless of mode of access.

5. Technical Infrastructure
The Royal Library establishes the necessary technical infrastructure to preserve its digital collections. The infrastructure supports the necessary surveillance of systems, media and formats to ensure a proactive effort. The infrastructure includes tools both to support decision-making processes as well to supervise performance of specific preservation activities. The technical infrastructure must live up to the Royal Library’s security policy.

Tools to both support the decision-making processes and supervise performance of specific preservation activities should be established through planning and development of new systems and components for managing of digital collections.

6. Vocabulary
Authenticity: Quality that assures that a resource or a person is what it claims to be (DS 484:2005).

Bit preservation: Bit preservation ensures that the individual bits in a digital material are preserved. In order to achieve bit preservation, at least two copies of the material are required, so that continual checking shows that the number of bits has not changed (and that bits are thus lost) over time.

Certification: International efforts are being made to establish standards for digital archives, so that in the future external auditing and certification of the digital archive will be possible. Such certification can provide visibility, both to others and the organization itself, that the digital archive lives up to the requirements with which a digital archive should comply.

Checksum: A checksum is the fingerprint of a file, and it can be computed in several ways. The fingerprint can be used to validate whether or not the file is intact.

Data format: A data format is the way that data is structured or arranged in a given context.

Data model: A model that describes the organization of data and the relations between data.
DOMS: "Digital Object Management System"

DRAMBORA: "Digital Repository Audit Method Based On Risk Assessment" is a tool for risk assessment in preservation institutions, which sets measurable criteria in order to evaluate the abilities of the institution, identify weaknesses and recognize strengths. DRAMBORA sets the stage for the use of the institution’s risk profile to be used as a management tool, in order to ensure that the institution navigates safely around the potential risks that may threaten its collections. http://www.repositoryaudit.eu/

DROID: "Digital Record Object Identification" is an automatic file identification tool developed by The National Archives. DROID uses internal and external signatures from the The National Archives PRONOM-project. http://sourceforge.net/projects/droid/

Emulation: Through emulation an imitation of the original context is secured in software and/or hardware. Emulation is achieved by imitating the original presentation of the bits in the digital material. I.e. new contexts are created that provide the possibility of accessing the digital materials in their original form and with the appearance of the original experience of access to them.

Filming: With filming a new digital object is created which preserves possible interaction in a form that documents significant aspects of the interaction, but does not allow reactivation. An example of a filming is a video capture of an online computer game. The new digital object established by the filming is subject to digital preservation.

Functional preservation: See logical preservation.

IIPC: "International Internet Preservation Consortium". IIPC’s mission is to collect, preserve and provide access knowledge and information from the Internet, as well as to promote global exchange and international relations. http://netpreserve.org/

Ingest: In OAIS the function which provides the services that make it possible for an archive to receive information packages from producers and prepare them for further preservation.

Integrity: The property which ensures the precision and completeness of assets. (ISO/IEC 13335-1:2004).

JHOVE2: Open source tool developed for the characterization of digital objects. This tool can be used to ensure that files comply with the applicable format specifications, in order to ensure that programs can read the format correctly. https://bitbucket.org/jhove2/main/wiki/Home

Level of preservation: Digital objects can preserves at different levels of quality that reflect the value of the materials and requirements for confidentiality, bit security, processing and access, as well as economic necessity.
Logical preservation: Securing the integrity of digital materials so that they can be read, understood and displayed/played by a future program on a future computer and extra equipment. The methods for logical preservation can be emulation, migration or museological strategy. Within the library world, the term functional preservation is also used for this process.

Management system: A system for the administration of digital objects with all of types of related metadata which are intended partly to ensure a correct handling of both the object and the metadata with regard to administration, dissemination and preservation, and partly to function as an inventory registrant.


Open source programs (software): Covers a series of software licenses, which have in common that the user of the software is allowed access to the source code, is allowed to change the source code and make a new version of the program, as well as pass on the program and/or new versions of the program. (Wikipedia.dk).

OPF: "Open Planets Foundation" is an interest organization, which works to make practical solutions and expertise in digital preservation available. http://www.openplanetsfoundation.org/

Persistent identifier (PID): Unique persistent identifier (key), which ensures permanent access to a digital object independent of the physical location and ownership of the digital object.

Plato (Planets Preservation Planning Tool): A planning and decision management tool developed within the framework of the joint European Planets project, which supports the work of designing preservation plans in preservation institutions. http://www.iifs.tuwien.ac.at/dp/plato/intro.html (digitalbevaring.dk)


Preservation plan: A preservation plan is a central part of the task of preserving digital materials. The plan is the starting point for the daily work with the digital collection, and its purpose is to ensure, among other things, that the preserved data can also be accessed and understood in the future. This is achieved by ensuring the implementation of a series of goals that start in preservation policies and strategies.

Preservation policy: A preservation policy describes the general framework for an institution’s work in digital preservation.

11. september 2014
Preservation strategy: A preservation strategy describes how the institution’s preservation policy is to be implemented, as well as which principles and priorities are the basis of the digital preservation processes.

Preservation system: An organization made up of people and technology, which together perform preservation of the digital collections.

PRONOM: Online register which contains technical information on file formats and the products (software and other technical components), used to support long term preservation of digital objects. http://www.nationalarchives.gov.uk/PRONOM/Default.aspx

Significant properties: The aspects of a digital object that must be preserved over time so that the object can continue to be accessible and meaningful.

TRAC (Trustworthy Repository Audit and Certification): TRAC is a project developed in cooperation between several major layers within digital preservation, now under the leadership of The US Center for Research Libraries. TRAC provides a series of guidelines for how to assess and evaluate a digital archive, and it can also assist in actual certification. http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying-0 (digitalbevaring.dk).

Trustworthy archive: The concept of Trustworthy Digital Repository is used about an organization made up of people and systems which have taken responsibility for the long term preservation of digital information and for making this information accessible for the users of the archive. In order to be “trustworthy”, it is further necessary that the institution ensures and documents that the digital collections and objects are preserved in such a way that they can be accessed in as authentic a form as possible. The concept corresponds to the OAIS-model’s definition of an OAIS archive and it is used in connection with audit and certification of archives (digitalbevaring.dk).

7. Literature


The Royal Library: Policy for long term preservation of digital collection material at The Royal Library (2014)
A Framework for Applying OAIS to Distributed Digital Preservation
http://www.metaarchive.org/ddp/index.php/Main_Page

Eld Zierau, Det Kongelige Bibliotek: Package Formats for Preserved Digital Material. (Ipres 2012)

JISC report: Digital Preservation Study (2008)
http://www.jisc.ac.uk/media/documents/programmes/preservation/jiscpolicy_p1_finalreport.pdf

Library of Congress: Recommended Format Specifications 2014-15

Statens Arkiver: Strategi for bevaring af digitale arkivalier (2013)
http://www.sa.dk/media(4658,1030)/Statens_Arkivers_strategi_for_arkivering_af_digitalt_skabte_arkivalier.pdf

