THE DIGITISATION POLICY OF THE NATIONAL LIBRARY OF FINLAND

Contents

The vision of the National Library:

a national gateway to information

Comprehensive digital collections

Authentic digital surrogates as source material

Sustainable digitisation and lifecycle management of digitised content

Maximum use and reuse of digitised content

Freeing digitised content for online distribution

1   PRINCIPLES OF THE DIGITISATION POLICY

1.1   Objectives

The strategic objective of the National Library is to enable cultural heritage to be widely accessible online.

Through digitisation, the National Library aims to enhance the visibility, accessibility and usability of its unique collections. Digital collections facilitate the use of cultural heritage materials in virtual environments, particularly for research, education, new kinds of public use and creative endeavours. The National Library therefore strives in its digitisation activities

- to provide comprehensive digital collections for research, learning and public use that are as freely available as possible,
- to ensure the authenticity of the digitised collections,
- to ensure permanent and sustained access to and usability of its digitised collections,
- to improve the preservation of the original items in its collections by offering the public digital surrogate copies, and
- to generate new data and content through digitisation, enabling new uses and use for new purposes.

1.2   Regulations and recommendations

The digitising activities of the National Library are regulated by government legislation, the National Library's own guidelines and various international and national standards.

1.2.1 Legislation

- The Act on Collecting and Preserving Cultural Material (1433/2007)
- The Copyright Act (404/1961), the Copyright Decree (574/1995) and the Universities Act (645/1997; amended Sections 25 and 25a 586/2006)
- The Personal Data Act (523/1999)
- The Act on the Openness of Government Activities (621/1999)

1.2.2 National Library guidelines

- Finnish National Library Strategy 2006–2015
- Preservation Policy of the National Library 2009
- Collection Development Policy of the National Library 2009
- Metadata Policy of the National Library 2010/2011
- Digitising Plan 2009
- Digitisation principles adopted in the OPM KD project 2007–2009

1.2.3 International and national standards and guidelines

The National Library complies with international and national standards and good practices in its digitisation production and processes, and follows the validated EU guidelines.

The National Digital Library project (NDL) has defined a portfolio of metadata standards and formats to be used in digitisation. The National Library complies with the NDL specifications and contributes through its expertise to the development of this portfolio by actively participating in the relevant development work.

The National Library participates in the development of standards and guidelines as they pertain to digitisation.

1.3 Interest groups

In implementing its Digitisation Policy, the National Library co-operates with the interest groups specified in the National Library Strategy, these being information seekers, libraries, the publishing industry, and the consumers of cultural products. These interest groups may be categorised further based on the interaction roles they adopt with digital content in virtual environments into producers of original content, users of digitised content, and consumers of digital content, reusers or users for new purposes.

- **Producers of original content** include publishers, authors and copyright holders (persons or organisations).
- **Users of digitised content** include persons or organisations acting as clients who retrieve, access or use digital content, together with organisations which aggregate metadata or render digitised content accessible (such as the National Digital Library, Europeana or the World Digital Library).
- **Reusers or users of digitised content for new purposes** include researchers, students and citizens as individuals or as parts of online communities who reuse content, exploit it or use it for new purposes.

1.4 Funding and partnerships related to digitisation

The public sector is the primary source of funding for sustained digitisation, but public funding can be supplemented by means of cooperation agreements with the private sector. In such cases, the terms of the contracts must be in compliance with the digitisation policy of the National Library. In addition, national and EU guidelines related to public administration and digitisation must be considered:

- The National Library also strives to ensure maximum wide, open use, distribution and utilisation of its digitised collections and content in the case of items produced or processed with partners.
- The National Library defines the quality criteria for digitised content in terms of authenticity, accessibility, searchability and usability. The National Library's specifications and criteria for the long-term preservation of material and its principle for the sustainability of digitisation are significant concerns when considering cooperation contracts relating to the digitisation of its collections.
- The National Library complies with the Copyright Act.
- The National Library strives to follow the guideline issued by the European Commission stating that content which is freely available in analogue formats must remain freely available in digital environments. This means that the National Library will not grant private organisations exclusive rights to material digitised on the strength of public funding. Content that becomes copyright-free and is or has been digitised on the strength of private funding is to be freed and provided for open use comprehensively and with the minimum of delay whenever possible.

The National Library can also offer digitisation services for a fee, subject to separate negotiations in each case. The National Library's position is that responsibility for the long-term preservation of the material and the maintaining of sustainability should lie with the customer in such cases.

1.5 Responsibilities

The National Library is responsible for the digitisation of the cultural heritage material in its collections and for the accessibility, preservation and ongoing use of its digitised collections. The

National Library is responsible for enforcing the rights and access restrictions applying to its digitised content.

The Centre for Preservation and Digitisation is in charge of drafting and updating the National Library's Digitisation Policy and Digitisation Programme together with the other functional areas. The Centre is also responsible for the National Library's digitisation operations, processes and instructions and cooperates with Library Network Services in making the collections available and with Administration and Development in managing potential partnerships.

The Research Library is in charge of selecting material for the digital library together with the Centre for Preservation and Digitisation, in accordance with the National Library's collections management, preservation, digitisation and metadata policies. The Research Library is responsible together with Administration and Development for ensuring that adequate measures are put into place for obtaining copyright clearance for digitised items.

The Centre for Preservation and Digitisation is in charge of attaching the copyright and use restriction metadata to digital objects.

Library Network Services is responsible for national infrastructure development, the long-term preservation of the National Library's digital material and development of the integrated production of digitised content together with the Centre for Preservation and Digitisation.

2   CONTENT TO BE DIGITISED

2.1   Quality criteria for digital surrogate copies

As a trustworthy custodian of the cultural heritage, collecting, preserving and providing access to its physical collections, the National Library creates digital surrogates for physical originals for the purposes of online research and public use. The National Library's reputation as a trustworthy, reliable reference source is to be extended into its production of digital collections and content and their provision in virtual environments.

All digital surrogates created are to be authentic. This will ensure that researchers can confidently quote and cite digital sources. The collections and the items in them can be uniquely identified and recognised as having been produced by the National Library.

2.2   Digital collections

The creation of a comprehensive public and research resource requires a critical mass of content in digital format. The National Library seeks to promote the discovery and accessibility of its unique collections through digitisation. The main aspects criteria governing selection for digitisation are scope, preservation, demand and content.

- **Scope:** The National Library aims through strategic, co-ordinated digitisation to create a critical mass of content for research and public use in the form of cohesive digital collections grouped, for example, by type of material (newspapers), subject matter

(geography), historical period (wartime), or content (dissertations from the Royal Academy of Turku, manuscripts from an individual author). The collections of the National Library to be digitised can be supplemented from other collections external to the National Library.

- **Preservation:** The preservation criterion is based on the condition and security aspects of the material. Digitisation can reduce wear and tear on original material by providing reliable surrogate copies.

- **Demand:** User-centred digitisation must take the needs of interest groups into consideration. Material can be selected on the basis of demand, especially since heavily used material is often in poor condition. Digitisation can also improve the availability of works which are out of print, and can promote regional equality when material of interest to both researchers and the general public is provided for online use.

- **Content:** Selection based on content can emphasise the educational, cultural, historical, aesthetic or entertainment value of items. The criteria in such cases can be defined in the framework of national or international digitisation initiatives or agreed upon with interest groups. Content selection should also take into account the fact that digitisation provides global access to unique and valuable cultural items.

2.3   The Digitisation Programme

The National Library's Digitisation Programme covers three years. The annual digitisation plans are included in the library's general operational planning. Separate project plans will be drafted for digitisation projects that are funded or initiated from outside the digitisation programme.

3 LIFECYCLE MANAGEMENT OF DIGITISED COLLECTIONS

The National Library carries out comprehensive digitisation in the course of managing the whole lifecycle of its material from the initial production processes to access and long-term preservation.

3.1 Digitisation

The National Library develops and maintains the processes and production environments necessary for mass and specialised digitisation.

The National Library ensures the reliability and authenticity of the digitised items by managing the whole production process, including digitisation and post-processing. The processes include management and control of the chains involved in the production of the physical item, the digital object and the ingesting of metadata and its attachment to the newly created digital surrogates. The digitisation process ensures the attachment of comprehensive technical and administrative metadata and authentic and consistent provenance information with regards to the physical and digital objects managed and created.

The National Library creates METS (Metadata Encoding and Transmission Standard) packages for its digitised content which make use of interoperable standards and incorporate long-term preservation metadata, so that they can be easily transferred from back-end operational systems to presentation and long-term preservation systems. Separate METS profiles are created for each type of material (such as monographs and newspapers). The metadata are created and attached throughout the digitisation process according to these profile specifications.

The METS profiles contain the descriptive, structural and administrative metadata that are included in the digital object.

- **Descriptive metadata**

  In its creation of descriptive metadata for analogue material in conjunction with the cataloguing process the National Library complies with international and national standards in order to ensure compatibility between systems. (The intellectual content of the item or document is described (e.g. author, publisher, year), enhancing discovery and aiding relevance judgments of content.)

  The National Library will anticipate the future use of descriptive data in online environments by adopting a two-record model for digitised items in which the physical and digital copies will each have their own bibliographical metadata records, with links connecting them..

  The library can allow a minimum level of cataloguing in connection with digitisation of previously uncatalogued items in order to facilitate digitisation production. Metadata not previously included in the library catalogue can be reliably created in the course of digitisation and can be used to enrich and supplement existing descriptive data maintained by the library.

- **Structural metadata**

  The National Library creates structural metadata during the post-processing stage of digitisation by segmenting digitised items into their component parts. Such metadata can be used to describe the physical and logical structure of the digitised object or document. With structuring the original item or content can be described as a separate entity down to article and image level, for example. The metadata created through structuring is captured and attached to the digital object and can result in enhanced search functionality online and improved usability and searchability in terms of content.

  The National Library will define quality criteria for the levels of structuring to be used with different materials. For some materials these can be defined at the level of constituent units – e.g. at the article level for magazines. Metadata related to access rights will be attached to digitised items and their component parts wherever possible to clarify the terms of use in virtual environments.

- **Administrative metadata**
  The administrative and technical metadata created during digitisation will be expressed at an equivalent quality level and attached to all items. Such metadata will include information about the origin, processing history and technical environment of the item that has been digitised.

  The processing history includes information on the equipment used, people involved, settings applied and any measures taken after conversion to the digital format, such as image or sound processing. The attaching of processing history metadata to the digital object is a central quality criterion for long-term preservation.

  Reliability and authenticity are ensured through the use of unique identifiers for the physical and digital items and the use of technical integrity data for digital objects.

  Copyright data and other restrictions on the use of the material are described in the rights metadata.

3.2 Access to digitised content

The digital collections administered by the National Library will be widely distributed online, in the optimal case forming cohesive entities and collections. The same principle shall apply to digital collections or items digitised with private sector funding.

The provision of access to content requires constant maintenance and development since the services provided must correspond to the rapidly advancing technology and to changes in online media. Accessibility guidelines are followed to ensure that the requirements of special groups in different situations are met.

The National Library can use other distribution channels in addition to its own, such as the National Digital Library, international digital libraries (Europeana, the World Digital Library), online communities, social media tools and services and commercial organisations.

3.3 Long-term preservation of digitised content

The National Library complies with the long-term preservation strategy for digitised material as defined in national resolution. The long-term preservation metadata portfolio of the National Digital Library is to be taken into account and implemented in digitisation production. The 2009 Preservation Policy of the National Library provides methodological guidelines for digital preservation.

During digitisation production, interoperable METS packages complying with accepted standards are created for long-term preservation according to the principle of sustainable digitisation. These packages include archival-quality files and metadata that will ensure their reuse in the future.

4 USE AND REUSE OF DIGITAL INFORMATION RESOURCES

4.1 New information resources

The National Library provides access to and preserves traditional information resources (printed books, newspapers, periodicals and ephemera) and improves their searchability and discovery through measures such as cataloguing.

The searchability, retrieval, accessibility, use and reuse of content can be considerably improved by creating digital collections that are unrestricted by place and time.

Through digitisation the National Library can produce not only surrogate copies but also entirely new information resources that can be used innovatively by virtual communities and with greater efficiency than is allowed by physical items. This will enable its digitised collections to function as new resources in the information society, contributing to the common national goal of promoting the reuse of content in the possession of the public sector.

Resources produced by the National Library through digitisation:

- Digitised image, text and audio content for online use
- Structured content for online use
- Comprehensive, cohesive digital collections for research use
- Descriptive, structural and administrative metadata that will enable the long-term preservation of items
- Submission Information Packages (SIPs) complying with the requirements of the National Digital Library project, for easy transfer to a joint long-term preservation system to be maintained by Finnish memory institutions, in order to promote the sustained use of digital collections
- Opportunities for the collective enrichment of digitised material

4.2 Copyright and accessibility

The National Library aims to maximise access to digitised content and to make metadata as freely available as possible. To achieve these aims, it will strive to influence the drafting of copyright regulations at a national level and within EU by issuing statements and cooperating with relevant interest groups.

The use made of content is ultimately dependent on copyright. Wherever possible, public domain content should be made available without restriction, while content that is protected by copyright should be made accessible to the public in compliance with existing agreements governing its use.

- **Content in the public domain**
  Digitised material in the public domain which is not subject to any other statutory restrictions on its display or use can be made freely accessible regardless of the source of funding for its digitisation. Such material can be made available through a variety of dissemination channels.

  The reuse of content in the public domain is regulated by licences such as the Creative Commons type of licence. The National Library will prepare separate instructions on terms of use and the licence types.

- **Content protected by copyright**
  Content submitted to the National Library via the legal deposit system and content that is protected by copyright can be made available to the public in a restricted area, on the premises of the National Library, or within other organisations as defined in the Act on Collecting and Preserving Cultural Material.

  The National Library can agree on the conditions of access to and use of content protected by copyright either through separate agreements with individual copyright holders or with an organisation with extended collective licensing rights that represents copyright holders. Should the conditions of Section 26 of the Copyright Act be met, an extended collective

licence agreement with such an organisation may take effect, in which case the agreement will also extend to copyright holders who cannot be reached.

The National Library supports agreement-based uses of digitised collections for research purposes.

The National Library supports measures towards cross-border licensing agreements allowing citizens and researchers access to cohesive digitised collections across national boundaries.

During digitisation production, the National Library attaches the relevant metadata to copyright-free digital objects to level of granularity required, defining the conditions of use and reuse of the content in virtual, learning and personal environments.

4.3   Reuse of digitised material

Thoroughly executed digitisation allows innovative use and reuse of material. Users benefit from the digitisation of comprehensive collections in terms of critical mass and in terms of the high quality of the structural analysis: so that instead of searching for information in a single issue of a single periodical, they can simultaneously search through in hundreds of thousands of pages, for instance, or alternatively, they can search for an article by means of its author or title alone, without any data on when or where the article was published.

The objective of the National Library is to maximise the exploitation, or reuse, of content and its constituent units. Digitised items and collections, metadata and other new resources created in the digitisation process are to be enhanced and made available for innovative new uses and purposes.

- **Structured constituent units**
  Structuring can be used in digitisation to mark up articles in a single issue or the images in an article, for example, thus enabling displays and searches to be developed to the level of granularity of the mark-up. Structuring can also be used to differentiate constituent units

from the whole so that they can then be used as independent entities. The level of structuring will be determined by the library for each type of material or content separately, in line with users' needs and use situations.

- **Community enrichment of content**
  The comprehensive structuring of digitised content requires considerable human resources, and in this sense the community enrichment of digitised collections can be regarded as a means by which communities can retrospectively improve the quality of digitised items through distributed, collective efforts. Crowdsourcing has been used, for example, to correct errors in text recognition, and as tools develop it should be possible to do much more than merely the structuring of documents. The National Library supports community-based enrichment and crowdsourcing for digitised items as widely and freely as possible. For these purposes, individuals and special communities need to be granted specified rights to enrich digitised collections.

- **Digitised collections for research use**
  The National Library is creating extensive digital collections of physical sources which will substantially improve researchers' access to and retrieval of individual items. In addition, the National Library can offer copies of the master files of its digitised collection for research use, thus enabling new kinds of research needs to be met. Automated methods can be used for digitised collections, e.g. in order to develop and enlarge multilingual glossaries, ontologies and authority databases. The National Library supports agreements by which researchers can be granted access to digitised corpuses in its collections regardless of copyright, on the proviso that a written agreement is made protecting the digitised items from unauthorised use.

- **Metadata for digitised material**
  The digitisation process yields reliable metadata which can be used to enrich and supplement the descriptive metadata contained in library catalogues and to improve searchability. Descriptive metadata (e.g. the titles of short stories) can be created during digitisation and attached to the resulting digital objects. Such metadata can be used and reused in various environments (National Digital Library, Europeana, Google, WorldCat), thus maximising searchability.

In accordance with the specifications of the National Digital Library project, the digitisation process will create administrative metadata by adding access restriction and use metadata to digitised items and their constituent parts, for example, thus improving their usability. The assignment of persistent identifiers and url addresses in the course of the digitisation process enhances the reliability and authenticity of items and is beneficial for the long-term preservation, use and reuse of their content. The National Library generates and maintains such persistent identifiers and url addresses, and it also supports the freest possible reuse of metadata.

- **Long-term preservation of digitised material**
  The National Library ensures that long-term preservation packages will be available for reuse as freely as possible.

5   SECURITY

The digitisation process includes both production and preservation. Security in terms of production is managed through a specific digitisation control and tracking system, while in terms of preservation it is managed by a temporary archiving system that is employed prior to transfer to the joint long-term preservation system for the Finnish memory institutions that is at present under development.

It is essential for comprehensive lifecycle management that the security of all digitised material should be ensured. The National Library attempts to identify and minimise both the risks associated with the creation of digital copies and those associated with the temporary archiving of digitised material prior to transfer to the national long-term preservation system.

The National Library's security system includes systematic, regular back-up for files and the database and a distributed preservation strategy that includes checking of the integrity of the digital packages. The National Library ensures that material (including its processing history) can be transferred in its entirety from one long-term preservation system to another, such as the LTP, by means of open interfaces and standards.

6   TRAINING AND INFORMATION

The professional skills and networking of the staff are promoted by offering them training in the latest developments in the field, organised either in Finland or abroad.

7   RESEARCH AND DEVELOPMENT

The National Library cooperates with others in planning and implementing community-based uses of material and its further enrichment by means of the latest technologies, services and dissemination channels.

The National Library and other actors and interest groups are engaged in developing new network technologies such as ontologies and methods for a semantic web.

The National Library promotes the sharing of knowledge and skills in digitisation and metadata with Finnish universities and the library sector as well as with other memory institutions.

The National Library supports the development of the infrastructure for a national long-term preservation system and the development and maintenance of easily transferrable, profiles for types of digitised material based on accepted standards.

International contacts and cooperation are actively being sought in matters of refining systems and skills, providing a basis for continuous operational development.

GLOSSARY

Administrative metadata	Consist of technical metadata and metadata concerning the lifecycle of an object and its terms of use.
Constituent part	A separate component (article, recording track, map, image, etc.) included in a larger work.
Creative Commons	Copyright holders can use Creative Commons licences to waive certain rights in a controlled fashion.
Descriptive metadata	Metadata describing the content, origin and context of an object that can be used to identify and locate it.
Exploitation	Digitised material, their constituent parts and/or related metadata can be published in new formats and contexts and be used for various purposes.
Extended collective licence	A licence for the use of material that is granted by a copyright organisation on behalf not only of the copyright holders that it represents, but also of other copyright holders.
Licence	A licence grants the right to use something which would otherwise be forbidden.
Long-term preservation	The reliable preservation of digital information (a digital object); an information management process that ensures the future usability of material.
LTP	Long-term preservation
Master file	A preservation copy of the highest quality, including a digital object of top-level quality.

Metadata	Data on data, i.e. data that describe or define a data resource or content unit.
METS	Metadata Encoding and Transmission Standard; the XML standard for objects in the National Digital Library which can be used to package a digital object and its related metadata (e.g. descriptive, administrative and structural metadata).
METS profile	An XML document describing the structural and content requirements of a METS file used for packaging; the Library of Congress maintains an international list of METS profiles.
NDL	National Digital Library
Ontology	Definitions for a key set of concepts and their relations in a particular field.
Provenance	Describes the origin and history of both physical and digital material; 1) physical provenance: origin, source, history of ownership or location, 2) digital provenance: processing and history of the creation of a digital copy.
Reuse	See Exploitation
Semantic web	Consists of objects (e.g. concepts or images in a document) with defined inter-relationships.
Structural metadata	Metadata that describe the physical and logical structures of an object and their inter-relationships.
Structuring	Differentiation of the logical structure and constituent parts of a digital object into its independent components.
Surrogate	A digital or micro-format copy with the same information content as the original publication.
Technical metadata	Metadata that describe the formats and technical features of the files and information flows in a digital object, including the colour mode of an image file or the sampling frequency of a recording file.

Two-record model	Separate bibliographical records for printed and digital material, with links between them.
Unique identifier	A barcode in a physical copy, for instance, or one of the identifiers maintained by the National Library in a digital copy.