



International Federation of
Library Associations and Institutions

Guidelines for Planning the Digitization of Rare Book and Manuscript Collections

Written by the IFLA Rare Book and Special Collections Section

September 2014

Endorsed by the IFLA Professional Committee



International Federation of Library Associations and Institutions, 2014

International Federation of Library Associations and Institutions
Guidelines for Planning the Digitization of Rare Book and Manuscript Collections

Revised in January 2015 to reflect a change in the name of the author from Rare Books and Manuscripts Section to Rare Books and Special Collections Section.

©2014 by International Federation of Library Associations and Institutions. This work is licensed under the Creative Commons Attribution 3.0 (Unported) license. To view a copy of this license, visit: <http://creativecommons.org/licenses/by/3.0>

IFLA
P.O. Box 95312
2509 CH Den Haag
Netherlands

<http://www.ifla.org/>

Table of Contents

Introduction and scope	4
Acknowledgments	6
1. Designing the project	7
2. Selection of originals	8
3. Workflow for creating the collection	9
4. Metadata	13
5. Display	14
6. Dissemination, promotion, and reuse	15
7. Evaluation	15
8. Long-term preservation of the digital collection	16
9. Summary of recommendations	16
References	17

Introduction and scope

Introduction

The digitization of library collections is transforming the ways that people discover information and conduct research. Libraries have a responsibility to provide global access to their digital collections: the public demands it and scholars expect it. The Internet has broken down traditional barriers to access brought about by geographic distance, economic circumstances, political boundaries, and cultural sensitivities. Researchers are developing new fields of inquiry, often stimulated by the capacity of new technologies to juxtapose research resources from diverse sources and disciplines, and to manipulate digital texts and images in new ways.

Digitization transforms the discoverability and use of rare and special collections to a greater extent than it does for the general library collection. Once these collections are accessible, they become a core resource. Without digitization, rare and special collections remain obscure and hidden.

While many libraries have procedures in place for participating in “mass digitization” projects, the needs of unique, rare, and non-print format materials require special consideration and different procedures. One of the purposes of these guidelines is to address this gap.

These guidelines are intended for anyone who is involved in planning a digitization project for rare and unique materials including library leaders who are proposing projects, librarians and researchers who are planning and executing projects, and funding organizations that are considering supporting the digitization of special collections.

Scope

Digitization has profoundly changed the way we view library collections, services, and strategic planning. As a result, digitization guidelines have proliferated in many forms over the last fifteen years at the international, national, local, and institutional levels, and all of them reflect a set of best practices that continues to evolve. Some of these guidelines attempt to be comprehensive,¹ while others are more minimal and refer to other resources.² Still others concentrate on digital capture standards for specific media,³ while some call attention to the need for more robust digital preservation programs.⁴ Each successive document tends to enhance the previous ones rather than completely replace them. The Working Group that prepared these guidelines studied these previous efforts and attempted to write a complementary document that addresses the specific needs related to planning digitization projects for rare and special collections.

Rather than concentrating on technology issues, specific methods of digital capture, or digital preservation, these guidelines focus on conceptual planning and collaboration with potential users in order to achieve desired outcomes and sustainable results. They are written from the point-of-view of special collection managers, curators, and researchers who study the physical object as an artefact bearing intrinsic historical evidence as much as for the intellectual content that it contains. They also attempt to foreshadow the needs of users who are now conducting digital-based research that requires full-text analysis or large aggregations of “big data.” They will also help professionals and non-professionals create sustainable and adaptable digital collections that will be re-useable and continue to have research value into the future.

Emphasis is given to the creation and discovery of digitized collections rather than individual objects. The digitization of rare and unique collections tends to be complex, because it is important to maintain physical, intellectual, and contextual relationships both within the various parts of a single physical object, but also among the parts of the collection as a whole. Maintaining these relationships for the user often involves special metadata as well as different discovery and use platforms. In addition, the user must have enough contextual information and metadata to be able to evaluate the authenticity and reliability of the digital surrogate. The ability to present an entire collection in context, while maintaining authenticity and communicating evidence of the physical artefact, requires careful conceptualization and planning, advanced preparation, the execution of a great many related activities, and collaboration across many library units.

Not covered in these guidelines are practices for digitizing parts of a complete object for special uses or the use of advanced spectral imaging required for conservation analysis or technology-driven research. They are also not intended to guide imaging for long-term preservation, which requires much more detailed information on imaging techniques, file formats, image resolution, and special equipment.

Values, general tenets, and desired outcomes

As the members of the Rare Books and Special Collections Section deliberated these guidelines, a number of shared professional values and desired outcomes emerged that informed the final text. They are expressed here as general tenets to guide decision-making:

- Diligently continue and defend the library's obligation to preserve and provide access to the original materials.
- Provide free, worldwide access to research resources, and, when possible, provide users the ability to download digitized objects and collections.
- Be sensitive to the needs of scholars; encourage a dialogue with scholars and users; invite participation in the planning process.
- Strive for achievable outcomes, open access, improved preservation, and added value to collections when possible.
- Build on evolving best practices and successful projects; build quality control into all phases of the project, from initial planning to final evaluation.
- Preserve both the structure and context of the original, document which copy has been digitized, and maintain a link back to the catalogue record.
- Administer "on demand" requests in a way that adds to the growing body of accessible, digitized collections.
- Leverage digitization projects to provide access to collections that have been hidden in the past.

Acknowledgments

The IFLA Rare Books and Special Collections Section gratefully acknowledges the contributions of the Working Group who dedicated themselves to developing the guidelines over a 3-year period, and to many others who helped to refine and improve several draft versions.

These guidelines were identified as a priority strategic task at the Gothenburg meeting in August 2010. The Working Group chair, Isabel Garcia-Monge (Spain), hosted a 2-day midterm meeting in Madrid in April 2011, during which the group produced a first draft. The group reconvened for a full working day at the Puerto Rico meeting in August 2011 and produced a second draft in January 2012. A third draft was issued following a midterm working group meeting in Antwerp in February 2012. The Working Group then presented a 2-hour open conference session in Singapore in August 2013 to introduce the draft guidelines and solicit comments from the audience. More than 200 people attended the session and participated in an animated 45-minute question-and-answer discussion which resulted in much valuable feedback. This work would have benefited from receiving comments from members in other geographical regions who were not members of the Working Group and who did not respond to our call for review. The Working Group acknowledges that the views of all regions are essential in providing a balanced perspective on the planning of digitization projects. We positively encourage contributions from other parts of the world that we will include in future editions.

The Working Group members are listed below. (An asterisk (*) denotes a member of the Rare Books and Special Collections Standing Committee.)

- Isabel Garcia-Monge, Chair (Spain)*
- Ivan Boserup (Denmark)
- Anne Eidsfeldt (Norway)*
- Pilar Egoscozabal Carrasco (Spain)*
- Claudia Fabian (Germany)*
- David Farneth (United States)*
- Sirkka Havu (Finland)*
- Wolfgang-Valentin Ikas (Germany)
- Raphaële Mouren (France)*
- Angela Nuovo (Italy)*
- Krister Östlund (Sweden)*
- Edwin C. Schroeder (United States)*
- Garrelt Verhoeven (Netherlands)*
- Marina Venier (Italy)*

Other people who participated, read, or commented on various drafts include:

- Jan Bos (The Netherlands)*
- Mark Dimunation (United States)*
- Joanna Escobedo (Spain)
- Cristina Guillén Bermejo, (Spain)
- Pilar Moreno García (Spain)*
- Alexander Samarin (Russian Federation)*
- Carolin Schreiber (Germany)
- Winston Tabb (United States)*
- Marta Torres Santo Domingo (Spain)
- Olga Vega (Cuba)

1 Designing the project

Careful consideration of a number of overarching questions will facilitate the design and implementation of successful projects. A digitization project can result in different products and services. Available resources need to be considered. Clearly defined goals and outcomes at the beginning of the project will assist in advocacy and fundraising, and at the end will provide meaningful benchmarks for evaluation. It is important to begin the selection of a project by answering some basic questions.

What is the vision for the project? What are the goals and objectives? Who will use it? How will they use it?

Is the goal to provide general research access or to meet the needs of a specific research agenda? Are you digitizing for access, or for preservation, or for both? How does the project fit into the overall goals of your institution?

Librarians value building long-term, comprehensive collections; scholars and users desire delivery tools tailored to their immediate needs; administrators look for projects that increase the institution's visibility and stature. Initial planning should address the needs of all of the interested parties.

Who should be involved in the planning?

Teamwork is essential for an effective project. Most successful projects include scholars, users, administrators, library staff (cataloguers, curators, conservators) and technologists in the planning process. Workflows affect many departments within the library, and it is important to include everyone in making decisions that will affect their on-going work.

Are there external funding opportunities?

Many digitization projects are funded by grants or are achieved through collaboration with other institutions. External funding sources may have their own guidelines for selecting the collections to be digitized. Ensure that the project aligns with the priorities of all of the partners.

What level of complexity is desired? What level of complexity can be achieved?

Often an ambitious project will require a phased development approach. If so, concentrate on building good, robust, and—most importantly—usable resources. Enhancements can be made over time, often involving the user community.

What do you want to digitize, and why?

Selecting materials to digitize can often be the most difficult part of the process, especially if there is high demand and limited resources (see chapter 2).

Are there any copyright issues regarding the materials?

If there are, you will need to consider them (see chapter 2).

Should the digitization be accomplished in-house or by external service providers? Do you have the space, money, and equipment, and expertise? What can an external vendor provide?

The answers to these questions will depend on the capacities of the institution and the nature of the project. Advantages of doing the digitization in-house may include building expertise and infrastructure to accomplish future projects, less risk of damage to original materials (value and insurance issues are to be considered), and more control over the quality of the images and transfer of the data. On the other hand, using external service providers may provide expertise and better equipment, lower cost, and/or a faster completion time with guaranteed results, thus allowing in-house staff to accomplish other tasks. Remember that outsourcing always requires a significant amount of in-house staff time to manage the project.

What is the final format of the project? Do you have the means to achieve it?

It is important to consider if the final format of the project will achieve the overall goals. Determine if you have the equipment, software, facilities, expertise, and funding required for completing the project successfully.

Is a social-networking component envisioned, such as crowd-sourced transcription or metadata enhancement?

Digitization projects have the potential to reach a broad public, and this audience should be considered when designing a project. Think about how your project might be enhanced by future contributions from scholars and other users.

How will you incorporate quality management into all stages of the project?

Quality management is not limited to the creation of digital images; rather it should be considered and incorporated into every phase of the project. The overall quality parameters include: high quality optical images, completeness and reliability of the metadata (including links), faithfulness to the original, and ease of use.

2 Selection of originals

The selection of originals is an essential task in the development of a digital collection and one that relies heavily on specialist librarians. Collections, works, editions, and copies are to be studied and checked against the scope of the new digital collection. Chronology, geography, author, subject, format, owner, etc. may also need to be considered.

The following questions might help to determine the scope of the project and assist in the selection process:

- Do you want to digitize single, miscellaneous items, such as “treasures”?
- Do you want to digitize an existing collection?
- Do you want to create a new “virtual” collection? For example, a collection of items from different institutions that have something in common, such as provenance.

The design of the new digital collection at this level will be determined by the goals of the institution, its functions, and intended users. As digital collections and projects grow over time, it is useful to contemplate the future development and interaction with other collections from the same or other institutions.

Manuscripts and the majority of hand press books are unique items. Printed books may have significant variation within the same edition. Some copies will have specific interest because of a seal, manuscript annotations, or a bookplate. The number of editions and copies of the same work to be digitized must be decided according to the objectives of the institution and of the project, its public, and the available resources. At minimum, the best physical copy in hand should be selected for digitization.

Selection should include a physical review of the material, as its condition will influence the digitization process. Communication between staff doing the selection and staff doing all other activities is essential.

Digitization on demand is another way of developing digital content but it cannot be considered as the basis of a collection.

It is always preferable to digitize a complete intellectual entity, rather than part of it. Therefore, digitize an entire book or document rather than a chapter or a page. Digitizing parts of objects for one-time use are not the objective of these guidelines.

Copyright

The digitization, dissemination, and reuse of collections depend on intellectual property and privacy issues related to the content; laws that regulate these issues vary significantly from country to country. Identify which items are in the public domain and which are not; this activity should be accomplished when new physical collections are acquired for the institution. Copyrighted materials can also be digitized under adjudicated regulations and agreements. Rights ownership and “terms of use” are also very important in projects developed through agreements with collaborators and other third parties; they must be discussed and agreed upon in advance and in writing before the project starts. In addition, a digitized object can be considered a new edition in itself. As a result, the availability and terms of use for every digitized object and collection should be clearly stated to users.

3 Workflow for creating the collection

The process of creating a digital collection can be broken down into various phases or steps. Each institution may organize these activities slightly differently, but the basic processes remain the same. All of the processes should be adequately documented. If the general design of the project and the tasks of the participants are clearly defined and known, mistakes and misunderstandings can be prevented.

Step 1: Inspecting and preparing the materials for digitization: physical condition and existing metadata

Materials that are to be digitized will usually be removed from their permanent storage location and their subsequent movements should be tracked as closely as possible. Fragile materials should be evaluated by a conservator to minimize the potential for physical damage during the

digitization process. Scanning technicians should receive proper training to ensure the safe handling of the materials, and they should be instructed to notify a supervisor if an item could be damaged by the digitization process. All items should be inspected at the beginning and at the end of the process.

Librarians should evaluate if the current bibliographic metadata is adequate to support user discovery. The bibliographic metadata should be in a system that permits user searching and accommodates links to and from the digital surrogate. Cataloguing is an essential part of the creation of the collection, and it should include decisions about the protocol, level and detail of the description, and the language or languages in which it is going to be codified. As a minimum, there should be one description in the language of the main catalogue; translations to any other language could be added in order to participate in national and international projects and to provide wider access to the materials. Other multilingual tools or protocols can also enhance access to records. The length and depth of the descriptions should be balanced: their anticipated public and dissemination will influence these decisions. If adequate metadata do not exist, it should be created before digitization occurs. (See Chapter 4 for a more detailed discussion of metadata.)

In addition, adequate structural metadata should document the various physical parts of the object, and the object should be checked against the metadata to make sure that all of the parts are in order. Items should be foliated if necessary. Manuscripts should be checked against the foliation. In the case of archives, the order of the folders in the boxes, and items in the folders, should be checked against the inventory or finding aid. Items to be omitted from digitization should be flagged.

Step 2: Digitization process

Choice of equipment

Choose digital capture equipment suitable for the items to be digitized and appropriate to the goals of the project. For instance, high-resolution digital cameras are recommended for medieval manuscripts and other materials for which researchers will want to study minute details. A flatbed scanner may be well suited to modern photograph collections. Special book scanners can be used for a wide range of printed books.

Whenever possible, the same process should be used for an entire object to provide uniformity. Some of the automated equipment that has been developed for mass digitization projects may be inappropriate for digitizing rare and fragile materials due to the risk of damage.

Image Quality

Resolution, colour depth, and lighting should be decided by taking into account any specific standards and recommendations generally accepted for the particular material, requirements for the archival master, and the requirements of display and use, according to the general design of the project. Other elements of image quality to consider include colour saturation, image brightness, image integrity, and the absence of halos and other optical flaws.

If foldouts are photographed with different equipment, the images should be inserted into the proper sequence.

When planning capture resolutions, be sure to calculate how much storage space you will need, and consider the download time that researchers will experience. Institutions that wish to digitize collections only once to accommodate current and future needs should consider capturing at 1.5 times above the currently desired final format. The higher the image quality, the more uses the file will have in the future.

Fidelity to the Original

When digitizing rare and unique material, it is important to preserve and recreate as much as possible the look and feel of the original object. The entire physical object should be captured and not just the intellectual content. It is necessary to photograph entire pages front and back (including beyond the edges) and not to crop images within the page edges. Bound volumes should be photographed cover-to-cover, including flyleaves, empty pages, pastedowns and bindings (front and back covers and their interiors, the spine and the edges). No matter the format, the entire work or artefact should be reproduced. Special features (such as watermarks), on the other hand, are usually digitized with a different process and are often inserted at the end of the sequence or as a separate file.

To convey to the researcher the size of the original object, a linear scale should be included in the image. The orientation of pages within one binding unit should not be altered by image processing or turning the object.

During the image capture, scanning individual pages of books is usually preferred because it accommodates items with tight bindings and facilitates the display of pages with page-turning software. Scanning two facing pages sometimes may be necessary to display its content or preferred in order to maintain orientation and coherence; however this method can lead to indexing problems later on.

Paper or cardboard should be inserted behind damaged pages. Translucent pages should be backed with beige or white backgrounds to minimize “bleed-through”; some institutions use black for certain types of originals, but this method usually reduces the contrast of the image. Most institutions attempt to use the same background for the entire project.

Colour is one of the most important and complex issues when attempting to recreate the look and feel of the original object. At least one selected page or image from an object should contain a colour target to facilitate colour calibration. Each piece of equipment should be calibrated to the same colour value standard (for example, the CIR-Lab system) and re-calibrated periodically. Display hardware should also be calibrated.

Issues related to conservation

Conflicts between requirements of image quality and preservation concerns are inevitable and should be worked out in advance. Many special materials such as manuscripts are unique, often priceless, artefacts. Once digitized, however, a digital copy may help to ensure preservation of the original by serving as a surrogate. Digital files can also be used to create preservation microfilm of fragile originals.

Conservation issues to be considered for all projects—and especially in the case of out-sourced projects—include:

- The availability of trained conservators for consultation before and during digitization
- Control of environmental and security conditions during digitization and transport
- Use of special equipment to minimize damage (such as humidifiers, book cradles, etc.)
- Specific instructions and training for the scan operators, such as the appropriate opening angle for books and the appropriate handling of fragile objects

Conservators often disagree on the use of glass plates to keep the original flat during digital capture. Some consider the risk of damaged pages or spines to be too great. Others accept the practice. Newer manual book scanning devices are being developed that minimize the pressure exerted by the glass plates. Unbinding volumes for any kind of reproduction is no longer considered a best practice.

All original materials should be returned to their permanent storage locations as soon as possible, after being carefully checked for possible losses or damages. All the movements of these items must be tracked and documented.

Step 3: Post-capture image processing and system ingest

Images should go through a post-capture quality control process to ensure the accuracy and integrity of the end product (with the help of calibrated monitors). Images that do not meet the quality standards of the project should be photographed again and replaced. Missing images should be supplied and inserted into the proper sequence. The structural metadata should be reviewed again and revised as necessary.

No image processing should be done at this stage other than colour correction. Institutions should have a policy about if, or when, colour correction is permitted. The policy should also indicate how information about post-capture image colour manipulation is communicated to the user.

File naming should be standardized according to institutional policy. Some institutions have developed file-naming policies designed to associate the digital file with the institution and/or with the physical object. Each digitized object should have a persistent identifier.

Some projects may also employ additional processes to enhance access and usability, such as Optical Character Recognition (OCR) processing, text mark-up, and/or the inclusion of geospatial coordinates.

Institutional watermarks hamper and prevent the use of images. If added, they should not interfere with the main part of the image.

Images should be transferred to staging servers for system ingest. At the appropriate time, the archival “master” images should be transferred to a permanent digital repository and deleted from workstations and other temporary storage devices.

These guidelines will not cover system ingest (the process for transferring digital images and metadata into management and discovery systems) because the process is dependent on local technology infrastructure and/or the requirements of specific digital repository technologies.

4 Metadata

Librarians have been using bibliographic metadata (cataloguing) and some types of structural and administrative metadata to manage and provide access to physical collections. Now librarians are creating updated metadata models that not only provide access to digital objects but that also provide the information needed for long-term preservation and to facilitate access in networked discovery systems. The following section describes four types of metadata appropriate to digital collections.

Bibliographic (or descriptive) metadata

Bibliographic metadata describes the physical object being digitized, including information about its intellectual content. All materials selected for digitization should have some bibliographic metadata before they are digitized. After the items are digitized, links should be added to the catalogue records pointing to the digital version, and from the digital version pointing back to the catalogue record.

These records should be created using accepted international standards and can be minimal level records or complete bibliographical records. As mentioned, physical description (more or less detailed) of the item or holding should also be created.

Each object should be addressed by a catalogue record in the online catalogue (it can be a collection-level record), which then can be mapped and/or harvested into other systems. The resulting digital object can also have its own description in the catalogue if it is considered convenient or necessary.

Structural metadata

For studying complex objects such as medieval manuscripts, archives, correspondence, or photo albums, the researcher must be able to recreate (or reassemble) the physical item from the individual digital page images. Libraries that are digitizing medieval manuscripts and other similar complex objects should take care to provide excellent collation and other forms of structural metadata. At minimum, the researcher should be able to determine the original sequence of the pages or images. Older and newer foliations should be provided (if none existed, the book should be foliated prior to digitization). The total number of pages should be given as well as identifications of recto and verso. Other important elements include numbering schemes, textual divisions, important quotations, and illustrations.

Image (or technical) metadata

The image metadata (sometimes called technical metadata) is usually captured automatically by the camera or scanner and appears in the header of the file. It should include:

- The length and width in pixels
- Sampling
- Compression
- Resolution
- Size in bytes
- Production information (such as the brand and model of the camera or scanner)

- Date of creation

If the image has been manipulated, this information must be included and accessible to the users.

Administrative metadata

Many libraries require additional administrative metadata to assist in managing access to the digital files. Administrative metadata might include the name of the photographer, the owner of the image or of the original item, copyright information, and credit lines. Administrative metadata might also include information about collection items that were not digitized and why, decisions made during the digitization process, and descriptions of routine image manipulation applied uniformly to the collection.

5 Display

Users are looking for open and free access, easy discovery through common search engines, unproblematic interaction and display using standard web browsers and plug-ins, viewing options (including two-page display and zooming capabilities), tagging functionality (especially for later retrieval), individualized annotation capabilities, printing capabilities, and the ability to download, reuse, and combine. Presentation formats should be easily browsed, downloadable, and easily manipulated.

Presentation files are usually created from the “master” files. The display and format of the files depends on the objectives of the project; for example, TIFF or JPEG 2000 files are appropriate when a high degree of zoom is needed by the user. The master high-resolution TIFF files are usually saved for archival purposes and not used for presentation because they are too large for rapid display and easy manipulation while JPEG and PDF are popular presentation formats. (See Chapter 8 for long-term preservation of the digital collections.)

The user should be informed if the presentation formats have been manipulated in any way (e.g. cropped). Secondary products of the digitization, such as fictitious or reconstructed editions for incomplete holdings should be clearly stated. Users should be able to easily determine where they are located in the document or collection, and should be able to return to logical places easily. At a minimum, strive to recreate the reading room experience, and, if possible, go beyond with additional features.

Clearly display to the user any copyright restrictions or other “terms of use.” Consider using an established standard for conveying this information, such as a Creative Commons license.⁵

In order to insure persistent access to a resource, the use of a Persistent Identifier (PI) is strongly recommended, such as PURL (Persistent Uniform Resource Locator), URN (Uniform Resource Name), DOI (Digital Object Identifier), or Handle. A Persistent Identifier not only facilitates access when local URLs to the unique digital object are changed, but it also provides a convenient method for citation and future verification. It will also be critical in any future linked data environment.

6 Dissemination, promotion, and re-use

Bibliographical descriptions of the digitized items and bibliographical records of the digitized files (if they have been created) should be included in the main catalogue and linked to the files. These descriptions will also be stored in the digital library system, and they should be linked back to the main catalogue.

Most libraries and grant-funded projects have a mandate to provide global access to digitized collections, which often includes the distribution of metadata, links, and sometimes the images as well.

One way to increase access and visibility is through the use of portals and other aggregators, which might be international, national, or subject-specific.⁶

Many libraries also promote and advertise digital collections through the use of blogs and social networking sites as a way of reaching out to interested organizations and associations. Before sharing content with external services and aggregators, libraries should carefully review the “terms of participation” and legal rights of images to make sure that they do not conflict with library policies or other collaboration agreements already in place.

Many digitized collections form part of larger research projects. Libraries may provide annotations, commentary, and other important contextual information to add value to digitized collections and make them more useable. Newer distribution systems are offering technology to facilitate transcription, annotation, and image analysis.

7 Evaluation

Most libraries wish to evaluate the success of digitization methodologies as well as evaluate the diffusion and impact of their digitization projects and programs. Several organizations have developed forms to facilitate the evaluation process. Statistics related to production and use can provide a good starting point for quantitative evaluation, such as: the number of books/objects digitized, the number of visits to portal pages, the number of times a digital object is viewed and/or downloaded, the number of times it is cited or linked to, etc.

More important and difficult, however, is qualitative analysis, which often requires feedback from users. Consider the following questions:

- How faithful is the surrogate to the original? Is the image an effective substitute for the original, or does the researcher need to see the original at least once to accomplish his objectives?
- Is the product readable and usable?
- How well does the technology fulfil the research needs?
- How is the resource being used, and by whom?
- How is the resource being reused or repurposed?

- What is the impact on the use of the physical collection?

Statistics about the digitization program should be incorporated into all regular reports of library activities and use.

8 Longterm preservation of the digital collection

Institutions should develop strategies for the long-term preservation of their digital collections, especially in light of the cost of digitization, the investment of staff, and the stress that is put on rare and unique materials. Preservation can be accomplished in-house, out-sourced to vendors or service organizations, or accomplished using a distributed, consortium model. Standard requirements are in place for Trusted Digital Repositories (TDRs), but they are difficult for most libraries to achieve on their own.⁷

At minimum, a library should maintain their digital collections in high resolution on regularly backed-up network servers and have processes and systems in place to monitor the integrity of the digital files over time. Storing multiple copies in geographically dispersed locations is also an accepted preservation strategy. A process should be in place for regularly evaluating the need to migrate the collection or emulate software functionality.

9 Summary of recommendations

- Plan projects carefully, and consider goals and objectives, copyright restrictions, funding, and institutional capabilities. Involve a mix of librarians, technologists, conservators, researchers, and administrators.
- Anticipate the needs of scholars and digital research methodologies, which might include big data, resource aggregation, detailed image analysis, data visualization, geospatial mapping, social media, etc.
- Leverage digitization projects to provide access to collections that have been hidden in the past and strive to unite them virtually with related materials at other institutions.
- Always digitize a complete intellectual entity, rather than part of it. Preserve the look and feel of the complete original object, and enhance its study as an historical artefact. Preserve both the structure and context of the original, document which copy has been digitized, and maintain a link back to the catalogue record.
- Never discard the original collections after digitization. Interaction with a digital surrogate can never provide the full range of knowledge that is gained from interacting with physical collections.

- Provide high-resolution colour images that include technical information about the images, information about copyright, a scale and colour chart, zooming capability, and, when possible and appropriate, full-text searching, full or partial transcriptions of the texts, and geospatial coordinates. Use persistent identifiers for the digital resources.
- Provide as much descriptive, structural, technical, and administrative metadata as you can. Provide a link back to a description of the physical object at its permanent repository.
- Document the creation of a digital collection the same way you would a new physical collection and make this information available to researchers.
- Provide free, worldwide access to research resources and provide users the ability to download digitized objects, collections, and metadata in a way that is easily aggregated, used, and imported into digital research and publication platforms.
- Expose digital collections through external portals and aggregation services.
- Evaluate and document the results of the project.
- Ensure the longterm preservation of digital collections.

References

Online resources were accessed on 22 June 2014 unless otherwise noted.

1. Some standards that take a more comprehensive approach include:

Detusche Forschungspemeinschaft. (2013, February). *DFG Practical Guidelines on Digitisation*. Retrieved from http://www.dfg.de/formulare/12_151/ (English)

Bibliothèque nationale de France, Bayerische Staatsbibliothek, Universitat de València, Herzog August Bibliothek Wolfenbüttel, and Bibliothèque Royale de Belgique. *Project Europeana Regia*. The following three documents are available from:

<http://www.europeana-regia.eu/en/project-europeana-regia/project-documentation>

- *Attractive Guidelines for Users*. (2009)
- *State of the Art in Image Processing* (2011)
- *Quality Management [in the field of medieval manuscripts and other prestigious objects]* (2011)

IFLA and UNESCO. (2002, March). *Guidelines for Digitization Projects: For Collections and Holdings in the Public Domain, Particularly Those Held by Libraries and Archives*. Retrieved from http://portal.unesco.org/en/ev.php-URL_ID=7315&URL_DO=DO_TOPIC&URL_SECTION=201.html

2. More minimal approaches with links to other resources include:

Kavčič-Čolič, A. (2001, August). *Selected Literature in the Field of Digitization*. Retrieved from
http://www.ictparliament.org/sites/default/files/2011_training_puerto_rico-alenka_kavcic_selected_literature_in_the_field_of_digitization.pdf

UNESCO. (In progress). *Fundamental Principles of Digitization of Documentary Heritage*. [“The purpose of this text is to collate in one document, the basic information needed to understand the requirements of digitization.”] Retrieved from
http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/digitization_guidelines_for_web.pdf

3. Guidelines that focus on digital capture and specific media include:

American Library Association, Association for Library Collections and Technical Services, Preservation and Reformatting Section. (2013, June). *Minimum Digitization Capture Recommendations*. Retrieved from
<http://www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations>

Federal Agencies Digitization Initiative Still Image Working Group. (2010, August.) *Technical Guidelines of Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files for the Following Originals: Manuscripts, Books, Graphic Illustrations, Artwork, Maps, Plans, Photographs, Aerial Photographs, and Objects and Artifacts*. Retrieved from
http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf

4. Digital preservation remains a pressing concern, as evidenced by this UNESCO declaration:

UNESCO and University of British Columbia. (2012, September 26-28). *UNESCO/UBC Vancouver Declaration: The Memory of the World in the Digital Age: Digitization and Preservation*, 26-28. Retrieved from
http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/unesco_abc_vancouver_declaration_en.pdf

5. IFLA is now using Creative Commons licenses for its publications.

Creative Commons. (n.d.). Retrieved June 27, 2014 from Creative Commons website,
<http://creativecommons.org/>

6. Aggregators and portals are proliferating. Two common international efforts include Europeana (<http://www.europeana.eu/>) and the Internet Archive (<https://archive.org/>). Examples of national efforts include Gallica in France (<http://gallica.bnf.fr/>), Trove in Australia (<http://trove.nla.gov.au/>), and the Digital Public Library of America in the United States (<http://dp.la>).

7. The Primary Trustworthy Digital Repository Authorisation Body (ISO-PTAB) plays a major role in training auditors and repository managers. There are three important ISO standards related to digital preservation:

International Organization for Standardization. (ISO 14721:2012). *Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model*. [A reference model for what is required for an archive to provide long-term preservation of digital information.] Retrieved from http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=57284

International Organization for Standardization. (ISO 16363:2012). *Space data and information transfer systems -- Audit and certification of trustworthy digital repositories* [Sets out comprehensive metrics for what an archive must do, based on OAIS.] Retrieved from http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510

International Organization for Standardization. (ISO 16919, soon to be published). *Space data and information transfer systems - Requirements for bodies providing audit and certification of candidate trustworthy digital repositories*. Retrieved from http://www.iso.org/iso/catalogue_detail.htm?csnumber=57950