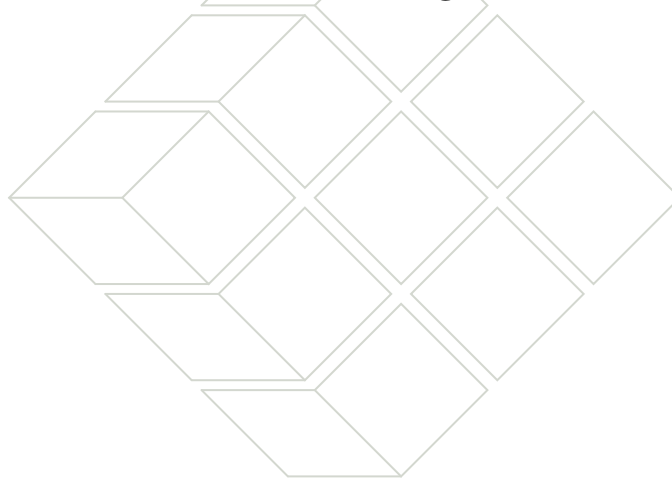


Digital Library Sector
Working group
PHAIDRA

Guidelines on Digitisation

First version 2011
Lorisa Andreoli
Marina Cimino

Revision 2017
Lorisa Andreoli
Gianluca Drago



October 2017



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

SBA SISTEMA BIBLIOTECARIO
DI ATENEIO

Index

Premise	2
1 Objectives	3
2 Selecting documents	3
3 Legal Aspects	4
4 Preservation	4
5 Digitisation	4
5.1 In-house or outsourced digitisation	4
5.2 Choice of equipment	6
5.3 Digital acquisition	7
Image files	7
Master file	7
Derivative file	10
Texts to be subjected to OCR (<i>Optical Character Recognition</i>)	12
5.4 File Names	12
Books	13
Journals	13
Photos, posters, maps (not bound in an atlas), parchments and other materials in loose sheets	14
Archive material	15
5.5 Data storage and conservation	15
5.6 Quality control	16
6 Archiving in Phaidra	17
7 Further details	18
7.1 Planning	18
7.2 Preservation	19
7.3 Digitisation	19
7.4 Legal aspects	20
7.5 Metadata	20
8 Contact	21
Attachment 1. Specifications for XML files of texts to be subjected to OCR	22
Attachment 2. Digitisation project information sheet	24
Attachment 3. Digitisation workflow and the professionals involved	28

Premise

Through various strategies and instruments and in compliance with current legislation on copyright, the University of Padova Library System aims to preserve and make accessible on Internet individual documents as well as important digitised collections related to a broad spectrum of disciplines, in the wake of scientific and experimental tradition that has always characterised our University.

To promote its own ancient, prestigious documentary heritage and to meet customer needs for quick and easy access to digital information content, the University of Padova Library System has established the Phaidra platform: a digital object management system with long-term archiving functions to meet specific aims for the conservation and access to digital collections.

In principle, all digitisation projects must be consistent with one or more of the following general purposes:

- valorise the documentary heritage preserved at the University of Padova and, ultimately, in major city libraries (public, ecclesiastical, etc.)
- expand public access to segments of Paduan documentary heritage relevant for scientific and cultural purposes
- promote interdisciplinary studies and promote collaboration between different local institutions
- promote knowledge of local or unique collections, through their widespread dissemination
- create virtual collections through the integration of various formats or materials located in different places
- limit the direct consultation of original documents in particularly critical conditions
- facilitate access to material which is difficult to attain
- ensure that the documentary material is available to future generations of students and scholars

This document is intended to be a reference tool for use of Phaidra and therefore refers to a set of procedures for digitisation of two-dimensional formatted documents, consistent with best practices and national and international standards for the quality reproduction of documents.

1 Objectives

Defining the objectives of a digitisation project makes it possible to establish the operational framework of the project from the beginning. The reasons and purposes can be various:

- to expand access to documents and their content
- to improve services to users, with the possibility of consulting resources collected and sorted into virtual collections, physically distant, inaccessible, little known
- to reduce the consultation of original documents in particular conditions (ancient and valuable documents, fragile, in poor condition, in high demand, difficult to handle)
- to develop collaborative activities with other institutions by creating virtual collections with greater access

2 Selecting documents

The documents are selected on the basis of the selection criteria¹ defined by the project, paying particular attention to legal issues (laws on copyright, privacy...). From this point of view, any concerns must be submitted to the opinion of legal counsel.

The selection criteria generally measure:

- historical and cultural value
- uniqueness and rarity
- high demand
- material without legal constraints or digitisation permits obtained
- restricted access due to the condition, value and location
- value added through online access, the creation of virtual collections, increased interest in little known or unknown material

In some cases, it may be useful to carry out an inventory of documents for identifying the quantity, type, size, state of preservation, and the net asset value of documents and any other information. The project sheet (see *Attachment 2*) may be used as a reference.

This information may be used for subsequent activities of conservation, cataloguing and digitisation.

¹ Selection for digitizing: a decision-making matrix <http://www.clir.org/pubs/reports/hazen/matrix.html>

3. Legal Aspects

In the digitisation of documents, it is essential to pay attention to copyright issues for both the original materials and for digital assets.

It is necessary to consider: the characteristics of the work to be treated, the ownership rights (who the owner of the rights is, the type of protection, if any), the actions to be performed with the work (which, what rights are involved, permissions to proceed), the potential problems and possible solutions.

Works that fall under the protection of copyright and works already digitised in other collections and accessible to the public through the network in order to avoid duplication and reduce costs.

4. Preservation

Digitisation does not replace the commitment to the care and preservation of the originals. It is important that an evaluation of the original state of preservation be undertaken before the digitisation and that any treatment on documents be performed after a survey by expert staff.

The restoration of the documents must be authorised by the relevant Superintendent and communicated to the Rector. The return of documents should be reported to the Rector and the Superintendent.

5. Digitisation



Digitisation is the process of transformation/conversion of an analogue object (text, image, audio, video) into a digital format, interpretable by a computer.

The nature and size of the originals determine the choice of the recording system, the lighting system and methods of treatment (transport, opening of the volumes, handling).

The quality of images defined in the project determines the hardware and the recording software requirements, the acquisition times and image processing, and the memory usage in the storage media to manage and maintain.

5.1 In-house or outsourced digitisation

The choice of digitisation within the institution (in-house) or the use of outside services (outsourcing) has to consider the advantages and disadvantages of the two methods.

	 In-house digitisation	 Outsourced digitisation
Výhody	<ul style="list-style-type: none"> - organizácia má priamu kontrolu nad celým procesom - zamestnanci sa učia prácou - požiadavky na digitalizáciu sa nemusia stanoviť vopred a priebežne sa práca zlepšuje - zaručené je bezpečné a správne zaobchádzanie - materiál a pomôcky sú dostupné 	<ul style="list-style-type: none"> - the institution pays for the product, usually at an established price per image - containment of costs and limited risks - the supplier can handle large amounts of material - the supplier absorbs the costs of expertise, training and technological obsolescence - availability of a wide range of options and services
Nevýhody	<ul style="list-style-type: none"> - the institution pays for expenses instead of for products, which include training costs, technological obsolescence and downtime - investment in purchasing and maintaining equipment - need for specialised staff - cost per image not defined 	<ul style="list-style-type: none"> - the institution eliminates one phase of the process; it does not develop in-depth knowledge on digitisation - issues of security, transport and handling of specimens
Odporúčania s	<p>Digitalizácia vlastnými silami sa odporúča, ak:</p> <ul style="list-style-type: none"> - zbierky nie je možné dať mimo inštitúcie - ak je digitalizácia nenáročná - ak je v inštitúcii špecializovaný personál a infraštruktúra 	<p>Outsourcing is recommended if:</p> <ul style="list-style-type: none"> - it is not possible for the originals to be digitised within the institution - the planning involves a large quantity in a short timeframe - there are constraints of space, infrastructure and personnel

Outsourced digitisation can be performed in the premises of the library or at the selected company's location. If the digitisation is performed at a company, the moving of documents must be authorised by the Rector and the relevant Superintendent. The return of documents must be communicated to the Rector and the Superintendent.

The flow of outsourced digitisation activities includes:

- definition of the scanning parameters
- preparation of a market study or a tender
- examination of the technical and logistical aspects
- arrangement of the digitalisation set
- preparation of documents
- training of staff and operators involved for quality control
- creation of a prototype

- digitisation
- quality control
- correction of defects and errors
- relocation of documents
- product delivery
- final quality control

The flow of in-house digitisation activities includes:

- definition of the scanning parameters
- purchase of equipment
- training of staff and operators involved
- examination of the technical and logistical aspects
- arrangement of the digitisation set
- preparation of documents
- creation of a prototype
- digitisation
- quality control
- correction of defects and errors
- relocation of documents

5.2 Choice of equipment

The data acquisition system (light source, optics, sensor, capture and calibration software) should ensure the image quality required by the project and not damage the original documents. In particular, the lighting system must be cold-light without emission of UV and IR. For ancient or valuable documents the use of suitable supports is required in order to not damage the document (facing the surface to be scanned upwards and using a tilting platform or V support).

These are some general indications on scanning systems:

Flatbed scanner: for single-sheet documents, or bound documents that can be opened easily, smaller or equal to A3 size.

These documents include: printed materials (e.g. leaflets, posters, brochures), manuscripts (e.g. letters), maps in good condition, printed music, prints (e.g. engravings, etchings, lithographs), pen and ink drawings without added watercolour or gouache (e.g. cartoons), photographic material (e.g. gelatin prints in black and white and in colour, albumen prints).

Film Scanner, negatives and slides

Planetary scanner or digital camera: for bound documents, documents of a particular nature, documents larger than A3 size.

These documents include: bound volumes (e.g. books, albums, printed music, atlases), fragile documents, oil paintings, most works of art on paper (e.g. watercolours, drawings), graphic material and artworks made with flaked and friable substances (e.g. crayons, charcoal, soft pencil), wa-

tercolours with thick drafting, tempera or with paints, large or fragile maps, manuscripts (e.g. bound diaries, folded documents), parchments, photographic material (e.g. large prints, historical photographic processes, such as daguerreotypes and ambrotypes), three-dimensional material (e.g. textiles, sculptures, objects).

5.3 Digital acquisition

The result of digitisation is the creation of files intended for long-term storage, “master” files, and files resulting from further processing, “derived” files, intended for use by users, typically via the Web.

The master file (“preservation master file” or “archival master file”) is the file that represents the best-copy output from digitisation, where “best” means that it meets the objectives of a particular project. These objectives may vary depending on the type of document. The criteria to be used in creating the master file must ensure faithful reproduction of the document in view of its long-term digital preservation or the need for high-quality printing, ensuring that there be no need to repeat the digitisation in the future.

Derivative files are produced from the master file and optimised for different fruition by the user, for example for display in a browser, to be converted to text via OCR, or for viewing on a dedicated workstation. They are normally resized and compressed, even with loss of information (i.e. JPEG images, MP3 audio format), for more convenient use achieved without excessive loss of quality. Below are guidelines for the digitisation of image files, i.e. the product of the digitalisation of text, graphic or three-dimensional documents.

Image files

The following specifications are to be taken as general guidelines, to be tailored in each case to achieve the best compromise between quality and cost.

High quality images, both in terms of resolution and in terms of colour depth, also imply higher costs of acquisition (equipment and qualified personnel) and of management (file size to be kept). On the other hand, the choice of the digital parameters must be sufficient enough to faithfully reconstruct the level of detail of the document.

The sampling density, or the number of pixels that represents the unit of length, must therefore be assessed not only based on the size of the document, but also based on the importance of the original document and the available resources.

“It is important to keep in mind that there are multiple factors that influence image quality: among these, in addition to the sampling density, we maintain colour accuracy, dynamic sensor and its noise.

Establishing a certain sampling density is therefore conceptually wrong because, depending on the shooting system that is used, equal to pixels-per-inch, the final quality of the scan can be very different.”²

Master file

² F. Lotti, M. Lunghi, G. Trumpy, *Procedure per un laboratorio fotografico digitale*, 2009

- The image is archived as it has been captured by the scanning instrument.
- The document must be taken in its entirety. Around the document, it is necessary to leave a border of a few millimetres in order to make it possible to read the contours of the document.
- For books, an image file is produced for each page: each side, recto and verso, of each page, including flyleaves, even if there is no information, and blank pages; all parts of the binding: endpapers, spine, textblocks, (in order to show headbands, clasps, hinges, borders). For maps, photographs and archive material, the verso is scanned only if there is information present.
- If the original is mounted on a support which contains information (e.g. a photograph mounted on cardboard with the photographer's trademark), digitisation must also include the support.
- Each document must be scanned alongside a chromatic scale, a greyscale and a metric scale, placed outside of the reproduced image and within the overall frame. In the case of volume, it is sufficient to place the scale once on a paper or page (which will be scanned two times, one with the scale and one without).
- In the presence of scratches, wormholes or oxidation of the inks, the papers must be masked with white paper in order to avoid capturing the underlying content.

Depending on the data capture tool, the master files can be of two different types:

- TIFF images
- RAW images (so-called digital negatives), in one of several proprietary camera formats such as NEF for Nikon or CR2 for Canon

If the master was RAW format, a copy should be made in an uncompressed TIFF 6.0 format to ensure long-term readability in commonly used software. These TIFF images must be faithful to the original RAW images and therefore should not be processed, except for colour correction, an operation that is performed with greater effectiveness and security with RAW files.

TIFF master

Type of document	File format	Colour	Optical resolution
Graphic material (Photography, Prints, Drawings, Paintings, Posters, Maps, Geographic Maps...)	TIFF 6.0, uncompressed	Colour profile "Adobe RGB" to 24 bit (8 bits per channel). For documents requiring the highest quality: Colour profile "ProPhoto RGB" to 48 bit (16 per channel)	Format up to A4: 600 dpi. Larger than A4: 400 dpi. For large and small formats, adjust the resolution in order to get the best results
Books, journals and manuscripts, rare or valuable (e.g. illustrated or painted) or with poor readability (faded characters, low contrast, margin notes in pencil, stained)	TIFF 6.0, uncompressed	Colour profile "Adobe RGB" to 24 bit (8 bits per channel). For documents requiring the highest quality: Colour profile "ProPhoto RGB" to 48 bit (16 per channel)	Format up to A4: 600 dpi. Larger than A4: 400 dpi. For large and small formats, adjust the resolution in order to get the best results
Books, journals, manuscripts, typed and mimeographed, not rare nor valuable, easily readable	TIFF 6.0, uncompressed	Colour Profile "Adobe RGB" to 24-bit (8 bits per channel) or to 16-bit greyscale	Format up to A4: 400 dpi. Larger than A4: 300 dpi. For large and small sizes, adjust the resolution in order to get the best results
Negatives, Black and White Slides	TIFF 6.0, uncompressed	16-bit greyscale	From 35 mm to 10x12 cm: 800-2800 with a resolution based on 4000 pixels on the longest side. From 10x12 to 20x25 cm: 800-1200 with a resolution based on 6000 pixels on the longest side. > 20x25 cm: 800 with a resolution based on 8000 pixels on the longest side.
Negatives, Colour Slides	TIFF 6.0, uncompressed	Colour profile "Adobe RGB" to 24 bit (8 bits per channel). For document requiring the highest quality: Colour profile "ProPhoto RGB" to 48 bit (16 per channel)	From 35 mm to 10x12 cm: 800-2800 with a resolution based on 4000 pixels along the long side. From 10x12 to 20x25 cm: 800-1200 with a resolution based on 6000 pixels along the long side. > 20x25 cm: 800 with a resolution based on 8000 pixels on the longest side.

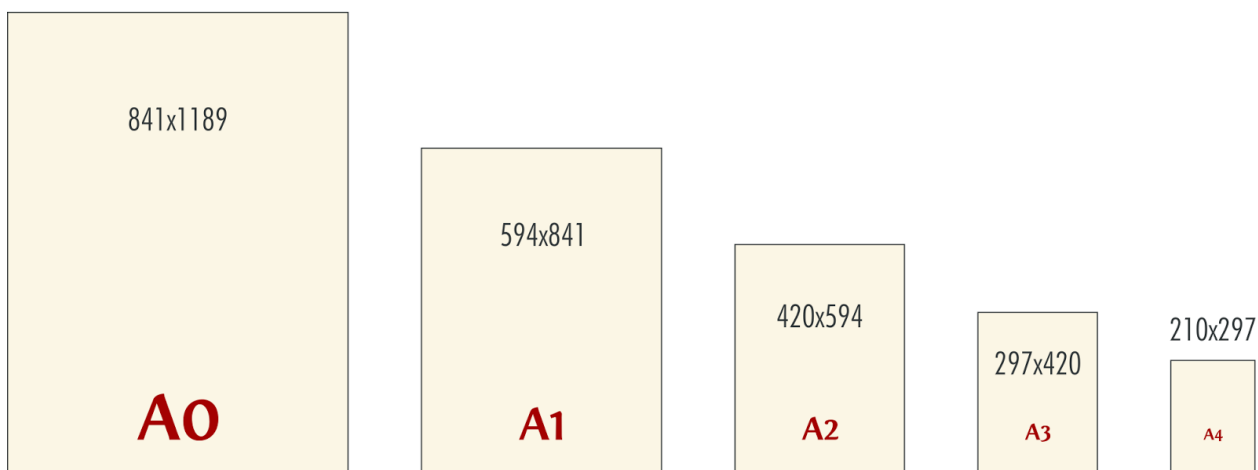


Figure 1: Paper size (mm). A0–A4 of Series A, International Standard ISO 216

Derivative file

Chromatic scales, greyscales and metric scales should be removed from derivative files.

Derivative files must be:

- Balanced for brightness, contrast and saturation in order to correct any chromatic aberrations due to the conditions of capture, on the basis of samples resulting from the colour scales and greyscales. This balancing should aim to achieve faithful reproduction of the original colour characteristics, not to an arbitrary aesthetic improvement.
- Straightened and cropped for the best visualisation

The choice of the type of derivative file to be created depends on the needs of the digitisation project, taking into account the availability of “in-house” tools and skills able to process the files as needed, of the different intended uses, as well as the quality of the images that you wish to upload in Phaidra.

The characteristics of the derivative files for different uses are described in the following tables.

Derivative TIFF

Type of document	File format	Size	Colour	Optical resolution	Use
All documents in the Master File Table	TIFF 6.0, uncompressed	Approximately 2400 pixels on the longest side	Colour profile Adobe RGB (1998) and depth of 24 bits (8 bits per channel)	The same of master	Print

High-quality JPEG

Type of document	File format	Size	Colour	Optical resolution	Use
All documents in the Master File Table	JPEG compressed at the best quality (100%)	The same of master	sRGB colour profile	300 dpi	For high-definition viewing of images in Phaidra. It can be adopted for maps and other objects requiring viewing of small details.

Medium quality JPEG

Type of document	File format	Size	Colour	Optical resolution	Use
All documents in the Master File Table Compressed JPEG at the best quality (100%)	JPEG compressed at the best quality (100%)	Approximately 2400 pixels on the longest side Colour profile Adobe RGB (1998) and depth of 24 bits (8 bits per channel)	Colour profile sRGB IEC-61966-2.1 and depth of 24 bits (8 bits per channel)	300 dpi	For average quality printing or uploading to Phaidra

Low quality JPEG

Type of document	File format	Size	Colour	Optical resolution	Use
All documents in the Master File Table	JPEG compressed at a quality between 90% and 100%	Between 1200 and 1500 pixels on the longest side	sRGB colour profile	150 dpi	For uploading to Phaidra

Texts to be subjected to OCR (*Optical Character Recognition*)

If you want to make text-searchable files available, the digitised images must be subjected to OCR

In this case, you can create a searchable PDF³, as well as various other formats depending on your needs (TXT, ODT, DOC, EPUB, MOBY...)

If you want to upload a “Book” in Phaidra as searchable text:

- the OCR must be performed at the same image size as those that will be uploaded to Phaidra
- an XML file must be created for each image with the same image file name, following formatting described in *Attachment 1*
- a searchable PDF must be created

5.4 File Names

In general, the name of each file will be a character string composed of several parts, having therein the information necessary to uniquely identify the project document to which the image refers. File names will be completed with the appropriate extension (tif, jpg, pdf, xml).

In mass storage, image files will be organised in multiple folders, in order to preserve the overall ordering of materials.

The nomenclature of the folders and files is a string of fields (library code, shelf mark...) separated by a hyphen (-). Where the shelf mark contains a hyphen (-), spaces or special characters, they are replaced by a dot (.).

To facilitate quality control, it is recommended not to include more than 200 pictures in folders for TIF files, or more than 100 images if they are large format documents. In these cases, subdivide the folder into more, consecutively-numbered folders.

For graphic material and archive material that are scanned on both sides, follow the progressive numbering of “-r” files for the recto, and “-v” files for the verso.

For books, front and back covers are named so that they occur in the same order they have in the physical document. The spine or other parts of the original document (textblocks, binding details ...) must be included at the end.

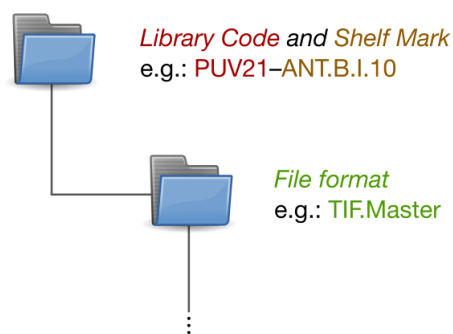
³ The PDF may be one of three types: a “normal” or digitally created PDF, for example by exporting the text from Microsoft Word; an “image-only” or scanned PDF; or a searchable PDF created by performing OCR on the images it contains (see: <https://www.abbyy.com/it-it/finereader/types-of-pdfs/>).

The image that includes the colour scale, the greyscale and the metric scale, must be named so that it is the last file in the folder and a “-c” is added to the progressive numbering of the file.

Books

The main folder, named “*Library Code – Shelf Mark*”, will contain the following subfolders: *TIF.Master* (or *RAW.Master* depending on the native format produced by the capture tool), *TIF.Derived*, *JPG300*, *JPG150* and, if required, *OCR*, even the *PDF* e *XML* subfolders, as well as a folder for each type of text file that may be present (*TXT*, *EPUB...*)⁴

The file name will follow the following schema: “Library Code – Shelf mark – Progressive Number.extension”



Example of folders structure and file name:
PUV21-ANT.B.I.10\TIF.Master\PUV21-ANT.B.I.10-0001.tif

In the following case the folder containing the master file has been subdivided into consecutively numbered folders:

PUV21-ANT.B.I.10\TIF.Master-1\PUV21-ANT.B.I.10-0001.tif
PUV21-ANT.B.I.10\TIF.Master-2\PUV21-ANT.B.I.10-0101.tif
PUV21-ANT.B.I.10\TIF.Master-3\PUV21-ANT.B.I.10-0201.tif

Journals

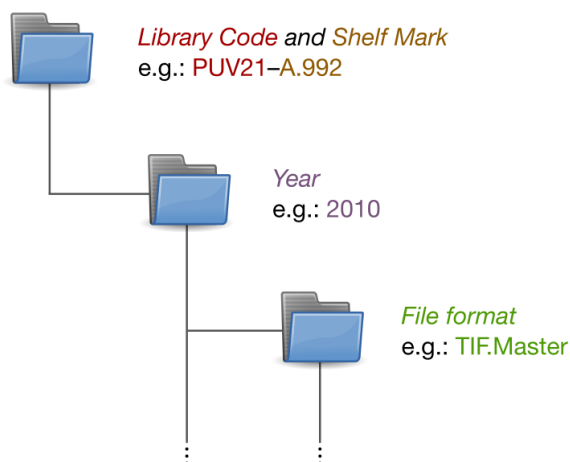
The main folder, named “Library Code - Shelf mark”, will contain a subfolder for each year of the journal.

Within individual years, there will be different folders for different types of files, named *TIF.Master* (or *RAW.Master* depending on the native format produced by the capture tool), *TIF.Derived*, *JPG300*, *JPG150* and *OCR*, if required, even the *PDF* and *XML* subfolders as well as a folder for each type of text file that may be present (*TXT*, *EPUB...*)⁵

⁴ If the master is a RAW file, a *TIF.High.Quality* folder will also be created to contain the exact copies of the RAW files for which colour correction has been applied.

⁵ See previous note.

The files will be named as follows: “Library Code – Shelf mark – Year – Month – Issue – Progressive Number.extension”



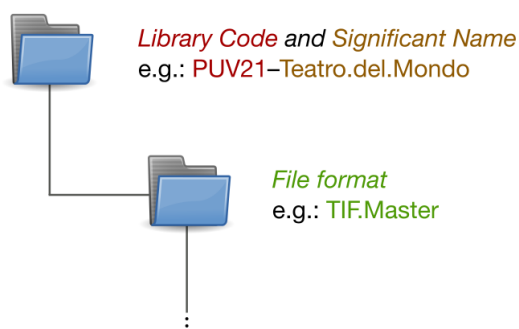
Example of folders structure and file name:

PUV21-A.992\2010\TIF.Master\PUV21-A.992-2010-12-24-0001.jpg

Photos, posters, maps (not bound in an atlas), parchments and other materials in loose sheets

The main folder will be called “*Library Code – Significant Name*”. The significant name will be created case by case at the time of digitisation. This folder will contain the following subfolders: *TIF.Master* (or *RAW.Master* depending on the native format produced by the capture tool), *TIF.Derived*, *JPG300*, *JPG150* and *OCR* if requested, and *PDF* e *XML* subfolders, as well as a folder for each type of text file that may be present (*TXT*, *EPUB*...)⁶

The file name will follow the following schema: “Library Code – Significant Name – Progressive Number.extension”



Example of folders structure and file name:

PUV21-Teatro.del.Mondo\TIF.Master\PUV21-Teatro.del.Mondo-0001.tif

⁶ If the master files master are RAW, a *TIF.High.Quality* folder will also be created to contain exact copies of the RAW files for which colour corrections have been applied.

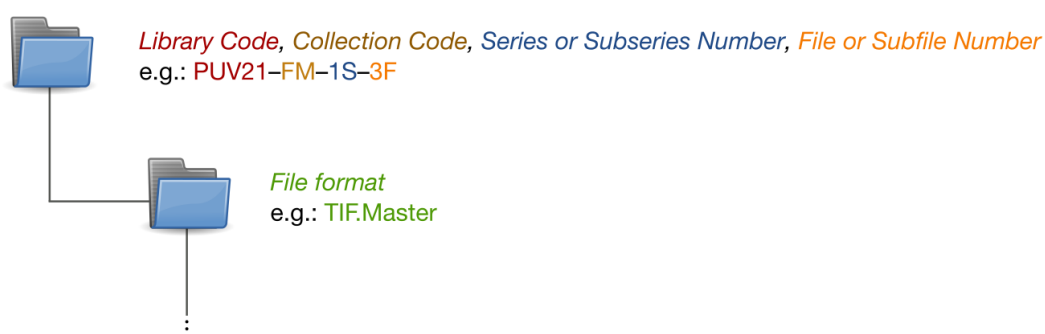
If necessary, distinguish recto from verso (e.g.: photography with information on the back):
PUV21-IB.Y.1\TIF.Master-3\PUV21-IB.Y.1-0001-r.tif
PUV21-IB.Y.1\TIF.Master-3\PUV21-IB.Y.1-0001-v.tif

Archive material

The main folder, named “*Library Code – Collection Code – Series or Subseries Number – File or Subfile Number*”, will contain the following subfolders: *TIF.Master*, *TIF.Derived*, *JPG300*, *JPG150* and *OCR* if requested, *also PDF and XML subfolders*, as well as a folder for each type of text file eventually present (*TXT, EPUB...*)⁷

The file name will follow the following schema:

“*Library Code – Collection Code – Series or Subseries Number – File or Subfile Number – Progressive Number.extension*”



Example of folders structure and file name:

PUV21-FM-1S-3F\TIF.Master\PUV21-FM-1S-3F-0001.tif

5.5 Data storage and conservation

The image collection consisting of folders and files will be stored on optical or magnetic storage media, such as CDs, DVDs, and external hard drives. It is recommended to store data on two different supports – of different brands or different series – and to keep the media in two locations, to verify the data periodically, and to transfer data periodically to new media.

The lifespan of the storage media is affected by various factors (the ISO standards 18923:2000 and 18925:2013 indicate the parameters for the proper maintenance of the storage media).

It is essential to maintain digital assets created over time in order to avoid repeating the costly work of scanning, so procedures must be put in place to ensure that digital objects remain usable and accessible regardless of future changes in technology.

The usability and accessibility of digital objects over time is guaranteed by file format (format standard, file size, network transmission time, how the images are displayed...), by media storage and

⁷ If the master files are RAW, a *TIF.High.Quality* folder will also be created to contain exact copies of the RAW files for which colour corrections have been applied.

by the digital repository. It is essential to use open standards to facilitate interoperability with other systems and thus access to metadata through other service providers (e.g. Europeana).

The files of the digitisation project must be delivered to UCT (University Library Centre) in accordance with the established archival procedures⁸.

CAB preserves digital data mainly in its “Storage and Backup” infrastructure and uses the services of the University of Padova for replication of its digital *assets*. The latter are validated in order to preserve their integrity.

The hardware infrastructure is equipped with modern deterioration detection systems, capable of quick change and recovery.

5.6 Quality control

Quality control is aimed at ensuring good screen readability of the entire information content present in the original, this should be documented and maintained during the entire digitisation process. Besides the on-screen control, it can be useful to do print tests to verify the quality of the image on paper.

Quality control planning includes:

- proper preparation of the environment (hardware configuration, visualisation software, viewing conditions, etc.)
- a priori definition of “acceptable” and “unacceptable” characteristics
- verification mode (any product or a sample, all files or only the master, visual screen quality and printing quality, etc.).

The visual inspection of an image usually involves:

- correctness of framing and exposure, the absence of any deformation and/or optical aberrations
- control of the chromatic tolerance
- depth and colour profile
- digital size and format
- the presence of any elements which compromise the fidelity of the reproduction (light reflections, etc.);
- file name

⁸ https://bibliotecadigitale.cab.unipd.it/collezioni_navigazione/Members/bibliotecari/materiali_settore_bd/gl-biblioteca-digitale/gruppo-phaidra/factory/digitalizzazione/READMEArchiviazioneprogettigitalizzazioneSBA3.txt (Accesso riservato)

6 Archiving in Phaidra

Archiving in Phaidra consists of uploading digitised files and entering the necessary data for the identification and description of the digital item.

It is possible that the object being archived is catalogued in other systems, such as an online catalogue or other platforms, so it is recommended to contact the Phaidra Project Team⁹ to determine the procedure for the possible migration of data.

For compilation of metadata, please refer to the *Guidelines for the compilation of metadata*¹⁰, for the storage of objects, please refer to the *Guidelines for creating an object*¹¹.

⁹ <https://bibliotecadigitale.cab.unipd.it/aiuto>

¹⁰ <http://phaidra.cab.unipd.it/static/linee-guida-compilazione-metadati.pdf>

¹¹ <http://phaidra.cab.unipd.it/static/guida-completa-oggetto.pdf>

7 Further details

Selection of resources divided by topic.

7.1 Planning

ATHENAWP3 (edited by), *Digitisation Standard Landscape*

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

Cohen, Daniel J. – Rosenzweig, R., *Digital history : a guide to gathering, preserving, and presenting the past on the web*

<http://chnm.gmu.edu/digitalhistory/index.php>

Europeana Pro

<https://pro.europeana.eu/:/pro.europeana.eu/web/guest/home>

International Federation of Library Associations and Institutions (IFLA), *Guidelines for digitisation projects*

<https://www.ifla.org/publications/guidelines-for-digitization-projects-for-collections-and-holdings-in-the-public-domain>

Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU), *Linee guida e standard*

<http://www.iccu.sbn.it/opencms/opencms/it/main/standard/>

Lunati, Gabriele – Bergamin, Giovanni (edited by), *Manuale virtuale per la progettazione digitale*

<http://www.regione.toscana.it/-/manuale-per-la-progettazione-digitale>

Ministerial network for valorising activities in digitization (MINERVA), *Guida alle buone pratiche*

<http://www.minervaeurope.org/publications/buonepratiche.htm>

Ministerial network for valorising activities in digitization (MINERVA), *Linee guida tecniche per i programmi di creazione di contenuti culturali digitali*

http://www.minervaeurope.org/publications/technicalguidelines_it.htm

National Information Standards Organization (NISO), *A framework guidance for building good digital collections*

www.niso.org/publications/rp/framework3.pdf

The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials

<https://chnm.gmu.edu/digitalhistory/links/pdf/chapter1/1.17.pdf>

Northeast Document Conservation Center (NDCC), *Handbook for digital projects*

<https://www.nedcc.org/assets/media/documents/dman.pdf>

7.2 Preservation

International Federation of Library Associations and Institutions (IFLA) Core Programme, Preservation and Conservation, *Principi dell'IFLA per la cura e il trattamento dei materiali di biblioteca*
<https://www.ifla.org/files/assets/pac/ipi/ipi1-it.pdf>

The Library of Congress, *Preservation, Collections Care*
<http://www.loc.gov/preservation/care/>

7.3 Digitisation

Association for Library Collections & Technical Services (ALCTS), *Minimum Digitization Capture Recommendations*
<http://www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations>

Besser, Howard (revised by S. Hubbard, D. Lenert), *Introduction to Imaging*
http://www.getty.edu/research/publications/electronic_publications/introimages/index.html

Cornell University Library, *Digital preservation management resource*
<http://www.icpsr.umich.edu/dpm/>

Cornell University Library, *Moving theory into practice: digital imaging tutorial*
<http://www.library.cornell.edu/preservation/tutorial/contents.html>

Digital Library Federation (DLF), *Draft benchmark for digital reproductions of printed books and serial publications*
<http://old.diglib.org/standards/draftbmark.htm>

Digital Library Federation (DLF), *Guides to quality in visual resource imaging*
<http://www.diglib.org/pubs/dlf091/dlf091.htm>

Federal Agencies Digitization Initiative (FADGI) - Still Image Working Group, *Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files*
http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf

JISC Digital Media
<http://www.jiscdigitalmedia.ac.uk/>

National Library of the Netherlands, *Alternative File Formats for Storing Master Images of Digitisation Projects*
http://www.kb.nl/sites/default/files/docs/alternative_file_formats_for_storing_masters_2_1.pdf

Osservatorio Tecnologico per i Beni e le Attività Culturali (OTEBAC), *Schema di capitolato per attività di digitalizzazione*
<http://www.otebac.it/index.php?it/127/capitolato-tecnico-digitalizzazione>

RLG Guidelines for creating a request for proposal for digital images services
<https://www.oclc.org/content/dam/research/activities/digimgtools/rlgmodelrfp.pdf>

University of North Texas Libraries (UNT), *Digital projects unit*
<http://www.library.unt.edu/digitalprojects>

7.4 Legal aspects

Legge 22 aprile 1941 n. 633, *Protezione del diritto d'autore e di altri diritti connessi al suo esercizio*
http://www.interlex.it/testi/l41_633.htm

Portale sul Diritto d'Autore per l'Università
<http://dirittoautore.cab.unipd.it/documentazione/dd>

7.5 Metadata

Baca, M. (edited by), *Introduction to metadata*
http://www.getty.edu/research/publications/electronic_publications/intrometadata/index.html

Dublin Core Metadata Initiative Wiki, *User guide*
http://wiki.dublincore.org/index.php/User_Guide

IEEE, *Standard for Learning Object Metadata*
<http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=8032>

Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU), *Linee guida e standard*
<http://www.iccu.sbn.it/opencms/opencms/it/main/standard/>

The Library of Congress, *Standards at the Library of Congress*
<http://www.loc.gov/standards/>

National Information Standards Organization (NISO), *Understanding metadata*
<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>

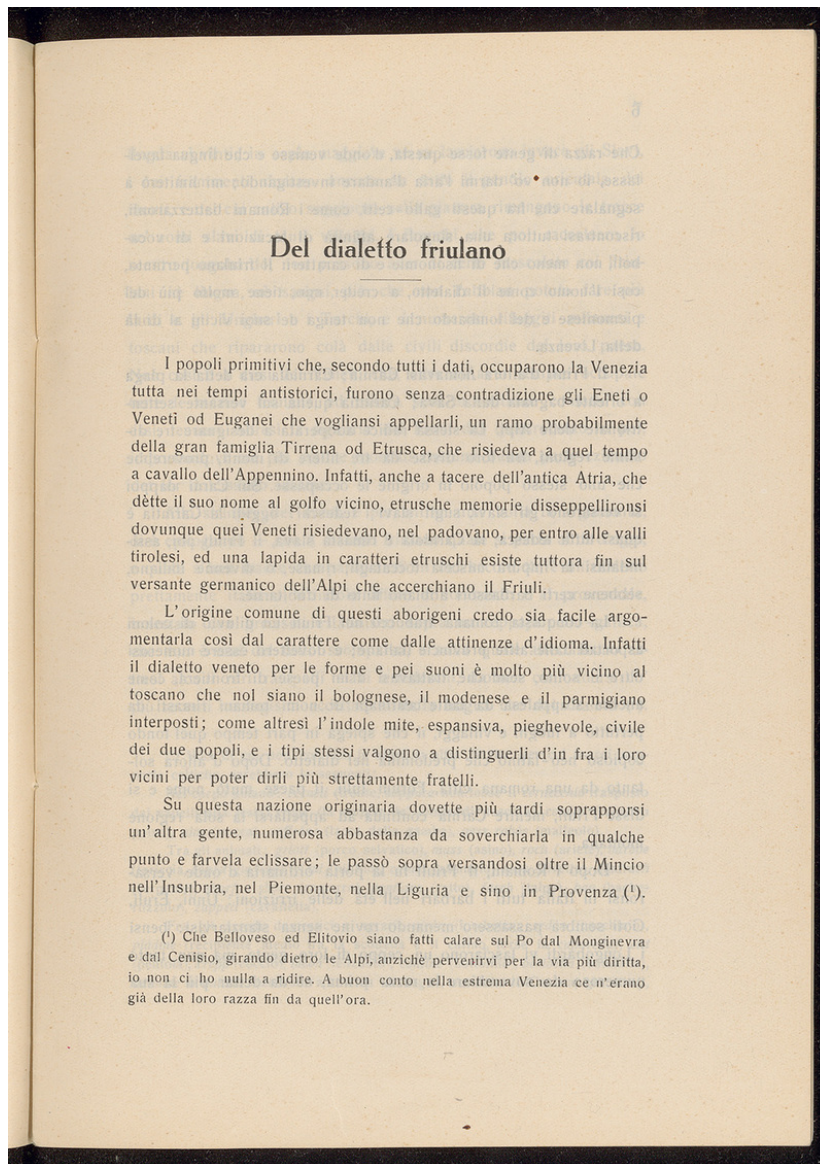
8 Contact

For more information about digitisation, contact Lorisa Andreoli lorisa.andreoli@unipd.it or Gianluca Drago gianluca.drago@unipd.it

For information about Phaidra, contact the [Help](#) provided by the University of Padova Library System.

Attachment 1. Specifications for XML files of texts to be subjected to OCR

The XML file must have the same name as the image file to which it refers (e.g. the image *page1.jpg* must correspond to an XML file named *page1.xml*).



From an image like the one above (https://phaidra.cab.unipd.it/detail_object/o:83943) an XML file formatted like this should be obtained:

```
<?xml version='1.0' encoding="UTF-8"?>
<ns0:ocrtext xmlns:ns0="http://phaidra.univie.ac.at/XML/book/ocrtext/V1.0">
  <ns0:page pid="o:4" abspagenum="9">
    <ns0:ocrword word="Del" x1="542" x2="620" y1="471" y2="515">
      <ns0:ocrchar char="D" x1="542" x2="581" y1="471" y2="514"/>
      <ns0:ocrchar char="e" x1="585" x2="606" y1="492" y2="515"/>
      <ns0:ocrchar char="l" x1="609" x2="620" y1="475" y2="515"/>
    </ns0:ocrword>
    <ns0:ocrword word="dialetto" x1="652" x2="821" y1="475" y2="517">
      <ns0:ocrchar char="d" x1="652" x2="676" y1="475" y2="515"/>
      <ns0:ocrchar char="i" x1="681" x2="690" y1="478" y2="514"/>
      <ns0:ocrchar char="a" x1="694" x2="718" y1="492" y2="515"/>
      <ns0:ocrchar char="l" x1="722" x2="731" y1="475" y2="514"/>
      <ns0:ocrchar char="e" x1="735" x2="757" y1="492" y2="516"/>
      <ns0:ocrchar char="t" x1="761" x2="774" y1="478" y2="516"/>
      <ns0:ocrchar char="t" x1="778" x2="792" y1="477" y2="516"/>
      <ns0:ocrchar char="o" x1="794" x2="821" y1="493" y2="517"/>
    </ns0:ocrword>
```

[omissis]

```
<ns0:ocrword word="fin" x1="528" x2="557" y1="2038" y2="2058">
  <ns0:ocrchar char="f" x1="528" x2="536" y1="2038" y2="2057"/>
  <ns0:ocrchar char="i" x1="536" x2="543" y1="2038" y2="2058"/>
  <ns0:ocrchar char="n" x1="545" x2="557" y1="2044" y2="2057"/>
</ns0:ocrword>
<ns0:ocrword word="da" x1="575" x2="604" y1="2038" y2="2059">
  <ns0:ocrchar char="d" x1="575" x2="589" y1="2038" y2="2058"/>
  <ns0:ocrchar char="a" x1="592" x2="604" y1="2045" y2="2059"/>
</ns0:ocrword>
<ns0:ocrword word="quell'ora." x1="621" x2="746" y1="2038" y2="2064">
  <ns0:ocrchar char="q" x1="621" x2="635" y1="2045" y2="2064"/>
  <ns0:ocrchar char="u" x1="639" x2="651" y1="2045" y2="2058"/>
  <ns0:ocrchar char="e" x1="654" x2="666" y1="2043" y2="2058"/>
  <ns0:ocrchar char="l" x1="669" x2="674" y1="2038" y2="2057"/>
  <ns0:ocrchar char="l" x1="678" x2="683" y1="2038" y2="2057"/>
  <ns0:ocrchar char="'" x1="686" x2="691" y1="2038" y2="2046"/>
  <ns0:ocrchar char="o" x1="697" x2="711" y1="2044" y2="2058"/>
  <ns0:ocrchar char="r" x1="714" x2="723" y1="2045" y2="2058"/>
  <ns0:ocrchar char="a" x1="725" x2="738" y1="2045" y2="2060"/>
  <ns0:ocrchar char="." x1="741" x2="746" y1="2056" y2="2060"/>
</ns0:ocrword>
</ns0:page>
</ns0:ocrtext>
```


Projekt digitalizácie Evidenčný list	
Pracovisko (Oddelenie, Centrum, Knižnica...) Názov inštitúcie a organizačného útvaru	
Vedecký manažér	The scientific manager (an expert or scientific committee) is the person who assumes responsibility for the selection of the materials and defines the quality of the metadata. In the selection phase, he/she is supported by the project manager, particularly for the examination of the materials and for legal aspects.
Priezvisko, Meno	
Telefón	
fax	
e-mail	
Manažér projektu	The project manager cooperates with the scientific manager, supports the scientific manager in the analysis of the legal issues, coordinates the activities related to the digitisation and guarantees the quality of the metadata.
Priezvisko, Meno	
Telefón	
fax	
e-mail	
Technical coordinator	As coordinator for technical-operational activities, he/she collaborates with the Phaidra Team, which in turn provides technical assistance.
namPriezvisko, Menoe	
Telefón	
fax	
e-mail	
Názov projektu	
Stručný popis zbierky určenej na digitalizáciu	

Stručný popis etáp projektu a osoby zodpovedné za etapy.	
Trvanie projektu: Začiatok - Koniec RRRRMMDD - RRRRMMDD	
Informácia o dokumentoch určených na digitalizáciu (knihy, noviny, časopisy, atlasy, mapy, fotografie atd.).	
Dating	od _____ do _____
Typ:	Približné množstvo:
printed text	
handwritten text	
printed and handwritten music	
map	
poster	
postcard	
drawing	
painting	
print (engraving, etching, etc.)	
parchment	
negative b/w	
negative col.	
photograph b/w	
photograph col.	
slide b/w	
slide col.	
other (specify) _____	
Forma dokumentov určených na digitalizáciu	voľné listy zrolované listy, zvitky zviazané album na kartóne alebo iných materiáloch v ráme obálky zložky krabice iné ? _____
Rozmery dokumentov	< A4 _____ A4 _____

	A3 _____ A2 _____ A1 _____ A0 _____ > A0 _____ iné (špecifikuj) _____
Celkový počet dokumentov	
Informácia o digitálnych objektoch	
Odhad počtu digitálnych objektov	
Použitie digitálnych objektov	otvorený, voľný prístup na webe obmedzený prístup na webe prístup v lokálnej sieti CD-ROM alebo DVD tlač iné (špecifikuj) _____ * V sprístupnení sa musí dbať na práva duševného vlastníctva. Otvorený prístup je k deskriptívnym metadátam a môže byť aj k náhľadom objektov, voľným alebo osirelým dielam ap. Práva sprístupnenia sa môžu viazať na každý jednotlivý dokument.
Predbežná kontrola	
Zdroj dokumentov (odkiaľ pochádzajú)	akvizícia dar neviem iné (špecifikuj) _____
Urobil sa predbežný výber?	áno čiastočne nie Ak áno, aké boli výberové kritériá? historická a kultúrna hodnota unikátnosť a rarita často žiadaný materiál bez právnych obmedzení prístupu alebo so získaným povolením na digitalizáciu a sprístupnenie

	prístup obmedzený z dôvodu stavu ochrany, hodnoty alebo miesta pridaná hodnota prostredníctvom online prístupu, vytvárania virtuálnych zbierok, zvýšeného záujmu o výskum pre málo známy alebo neznámy materiál atď. iné (špecifikuj)_____
Bola vykonaná kontrola	áno čiastočne nie
Existuje digitalizovaná verzia?	áno nie Ak nie, tak vypíšte, v ktorých organizáciách ste vykonali kontrolu, či je dokument alebo zbierka digitalizovanáe ? _____ _____ _____
Sú nejaké právne obmedzenia (copyright, ochrana súkromia, donorské práva atd)?	<input type="checkbox"/> áno čiastočne nie Podrobnejšie informácie: _____
Sú už dokumenty katalogizované?	<input type="checkbox"/> áno všetky; <input type="checkbox"/> áno čiastočne; <input type="checkbox"/> nie; <input type="checkbox"/> neviem; Ak áno, tak ako? <input type="checkbox"/> tlačný zoznam <input type="checkbox"/> elektronický zoznam <input type="checkbox"/> tlačný katalóg <input type="checkbox"/> elektronický katalóg <input type="checkbox"/> tlačný archívny zoznam <input type="checkbox"/> elektronický archívny zoznam <input type="checkbox"/> iné (špecifikuj)_____
V prípade, že ide o tlačný text, máte v úmysle robiť OCR ? (Optical Character Recognition)?	<input type="checkbox"/> áno <input type="checkbox"/> čiastočne <input type="checkbox"/> nie
V prípade že ide o rukopisný text, máte v úmysle transkribovať, prepísať text dokumentov?	<input type="checkbox"/> áno všetky; <input type="checkbox"/> áno čiastočne; <input type="checkbox"/> nie; <input type="checkbox"/> neviem
Približné náklady na digitalizáciu	Ak robíte digitalizáciu doma vlastnými silami, napíšte: - náklady na technickú infraštruktúru _____ _____ - prevádzkové náklady _____

	Ak využívate dodávateľský spôsob, napíšte - náklady na jednotky _____ - celkové náklady _____
Poznámky	
Popis projektu vytvoril	
Dátum	

The undersigned are aware that they must operate in accordance with local regulations on copyright. They declare that the documents of this project are (tick one of the options):

- owned by the University of Padova and protected by current legislation on copyright and industrial property
- owned by third parties who have, however, granted the University of Padova, the right to use them
- in the public domain

Signature of Scientific Manager _____

Signature of Project Manager _____



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

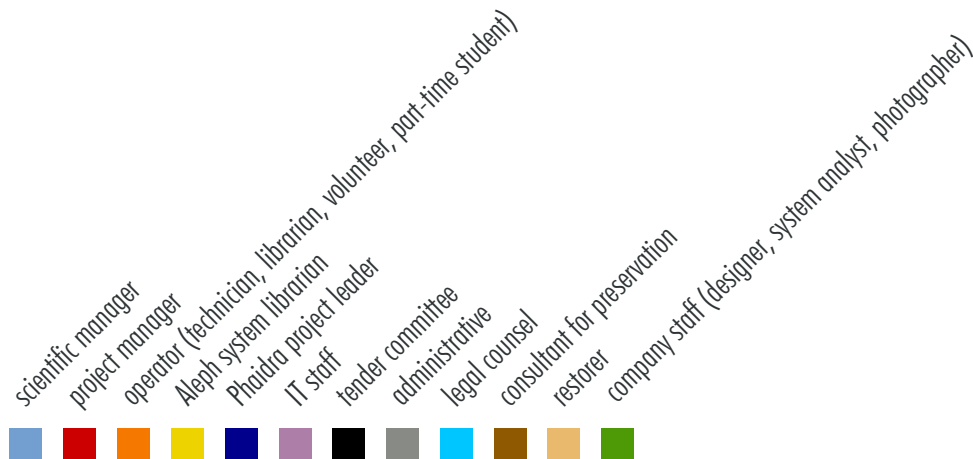
SBA SISTEMA BIBLIOTECARIO
DI ATENEIO

Digitisation workflow and the professionals involved

Lorisa Andreoli, Gianluca Drago
University of Padua – University Library System

Padua, 2014

KEY



1. Planning and preparation

Material Selection



The scientific manager and the project manager choose the documents on the basis of the selection criteria defined by the project, paying particular attention to legal issues (copyright, privacy laws ...). From this point of view, any problem must be submitted to the opinion of legal counsel.



Survey



Overview of quantity, size, type, conservation state of the documents, asset value and any other useful information. The *Scheda di progetto* [1], Project Information Sheet, may be used as a reference.



Preservation



Evaluation of the condition of the originals and preparation of a conservation plan (see *Raccomandazioni su come maneggiare i materiali per la digitalizzazione* [2], Recommendations on how to handle the materials for digitisation). The restoration of the documents must be authorised by the Superintendent for Library Heritage and communicated to the Rector (see *Richiesta di autorizzazione per restauro* [3], Authorization request for restoration, *Comunicazione restauro al Rettore* [4], Restoration communication to the Rector). The return of documents must be reported to the Rector and the Superintendent (see *Comunicazione rientro documenti da restauro/digitalizzazione al Rettore* [5], Communication on return of documents following restoration/digitisation to the Rector, *Comunicazione rientro documenti da restauro/digitalizzazione alla Sovrintendenza* [6], Communication on return of documents following restoration/digitisation to the Superintendent).



Cataloguing



Check for the presence of the cataloguing records in Aleph (or in other local databases in view of a possible import). Definition of standards and level of cataloguing and of metadata required depending on the type of document. Cataloguing of documents not found in Aleph.



Digitisation



Definition of the digitisation parameters (equipment, resolution, size, colour depth, file formats, file naming) (see *Linee Guida sulla digitalizzazione* [7], Guidelines on Digitisation).

Evaluation of advantages and disadvantages of outsourced (2a) or in-house (2b) digitisation.



2a. Outsourced digitisation



This can be performed in the premises of the library or at those of the selected company.
Preparation of the market research or definition of the tender (see *Documenti su indagini di mercato e gare di appalti* [8], Documents on market investigations and tenders) and the flow of activities.

If the digitisation is undertaken at a company's premises, the transfer of the documents must be authorised by the Rector and the Superintendent of Library Heritage (see *Comunicazione spostamento temporaneo al Rettore* [9], Communication to the Rector on temporary displacement, and *Comunicazione spostamento temporaneo alla Sovrintendenza* [10], Communication to the Superintendent on temporary displacement). The return of documents must be reported to the Rector and the Superintendent (see *Comunicazione rientro documenti da restauro/digitalizzazione al Rettore* [5] and *Comunicazione rientro documenti da restauro/digitalizzazione alla Sovrintendenza* [6], Communication to the Superintendent on return of documents following restoration/digitisation).

- review of the technical and logistical aspects
- possible preparation of the digitisation set
- preparation of the documents
- training of staff and the operators involved in quality control
- creation of a prototype



- digitisation batch
- quality control
- correction of defects and errors

- relocation of documents
- delivery of the finished product
- final quality control



2b. In-house digitisation



Defining the flow of activities. The flow includes:

- purchase of equipment
- training of staff and the operators involved
- review of the technical and logistical aspects
- preparation of digitisation set
- document preparation
- creation of a prototype



- digitisation batch
- quality control
- correction of defects and errors

- relocation of documents



3. Archiving in Phaidra



Import records from the Catalogue or import from other databases or direct archiving in Phaidra (see [Printable guides in Phaidra \[11\]](#)).

Storage of files produced.



LINKS

Most links require authentication at: <http://bibliotecadigitale.cab.unipd.it/>

- [1] *Scheda di progetto* <http://tinyurl.com/ljwux7d>
- [2] *Raccomandazioni su come maneggiare i materiali per la digitalizzazione* <http://tinyurl.com/ow985y3>
- [3] *Richiesta di autorizzazione per restauro* <http://www.regione.veneto.it/web/cultura/modulistica2>
- [4] *Comunicazione restauro al Rettore* <http://tinyurl.com/mu89pf5>
- [5] *Comunicazione rientro documenti da restauro/digitalizzazione al Rettore* <http://tinyurl.com/mj7xcm>
- [6] *Comunicazione rientro documenti da restauro/digitalizzazione alla Sovrintendenza* <http://tinyurl.com/kcnd7hm>
- [7] *Guidelines on Digitisation* <http://phaidra.cab.unipd.it/static/linee-guida-digitalizzazione-EN.pdf>
- [8] *Documenti su indagini di mercato e gare di appalto* <http://tinyurl.com/lqykkmc>
- [9] *Comunicazione spostamento temporaneo al Rettore* <http://tinyurl.com/mqyucpu>
- [10] *Comunicazione spostamento temporaneo alla Sovrintendenza* <http://tinyurl.com/lncxfm>
- [11] *Printable guides in Phaidra* https://phaidra.cab.unipd.it/help_long#printed-guides