The role of project MA.G.I.C. in the context of the European strategies for the digitization of the library and archival heritage

S. Conte[1] P. M. Maddalena[1] A. Mazzucchi[2] L. Merola[1] G. Russo[1,3] and G. Trombetti[1]

[1] "Ettore Pancini" Physics Department of the Federico II University, Napoli, Italy
[2] Department of Humanities of the Federico II University, Napoli, Italy
[3] National Institute for Nuclear Physics, Napoli, Italy

Abstract
Considering the allocation of Next Generation Eu funds, part of which is intended for the digital transition through the national plans of the Member States of the European Union, this contribution focuses attention on the policies adopted by them on the digital transformation of heritage library and archive.
A European overview exposes the digitization initiatives, as well as the standards adopted, which lead to the creation of multiple and diversified digital platforms that display the results; the same ones that implement Europeana, the platform for European digitization projects.
A considerable part of the contribution hinges on IT projects affecting Italy, of which particular regard has the digitization activity, carried out by the Vatican Apostolic Library and the Federico II University of Naples: united by the choice of digital format, the F.I.T.S. format, they intended to use it, because it is suitable for the long-term conservation of files and, therefore, functional for the custody, enhancement and use of documentary complexes and bibliographic collections.

CCS Concepts
• *Computer Systems Organization* → *Digitization pipeline;* • *Data* → *Long time data preservation;* • *Information Systems* → *archival and bibliographic resource; cultural platform;*

1. Introduction: state of the art

The European Commission has designated this decade as the "European digital decade", so that on 26 January 2022 an interinstitutional solemn declaration on digital rights and principles was proposed. The focus is the application of new digital technologies to culture. Digitization of the archival and library heritage of humanity is now considered unavoidable, by means of the creation of real digital collections, for which innovative solutions are needed to achieve their conservation, promotion and accessibility: this is the challenge! Since 2014, the European Commission has reported Member States digital progress through the Digital Economy and Society Index (DESI). It is certain that, during the COVID-19 pandemic, they have made considerable efforts in digitization, but some gaps in digital skills still need to be filled. In this regard, the European Union has allocated considerable resources to deal with the economic and social crisis caused by the pandemic and to start ecological and digital transition processes. Italy has adopted the "P.N.R.R., National Recovery and Resilience Plan", to access the Next Generation EU funds, as approved by the European Council. The "Digitalization and innovation" section, which includes Mission 1 "Digitalization, innovation, competitiveness, culture and tourism", Component 3 "Tourism and culture 4.0" (M1C3), provides investments for the digitization of the heritage held by museums, archives, libraries and places of culture to promote their accessibility and increase their use on-line with the "Faro Convention" and the European action framework for cultural heritage. There are many digitalization initiatives, which invest in the digital humanities sector through the funds allocated at both European and national level. Launched in the last few years by research institutions, government institutions and private entities, some of them are particularly demanding. The multidisciplinary approach, the advanced interdisciplinary skills, are the main features of the MA.G.I.C. project, which was promoted by the "Ettore Pancini" Physics Department of the Federico II University of Naples, in collaboration with the Department of Humanities of the same University and the current Ministry of Business and Made in Italy. MAGIC aims at the creation of a "Service Center", starting from the digitization of the book and documentary heritage, owned by the "Accademia Pontaniana" in Naples as a prototype, as well as about 150 manuscripts concerning the singe topic of "Divina Commedia", and later on of other important Public Libraries.

2. European digitization projects and "Europeana" platform

The European Union recommendation of 10 November 2021, which helped define the objectives of the "European digital

decade", urges member countries to complete the digitization of cultural heritage by 2030. A group of professionals oversees monitoring progress in cultural innovation. The "Europeana" platform, launched in 2008, will be the common space for museums, archives and libraries to share high-quality scans of paintings, archival sources, collections, as well as 3D models of historic sites for the benefit of sectors such as education and sustainable tourism. The digital infrastructure Europeana, which houses digital cultural heritage data from 3700 institutions, currently provides access to 56 million resources, of which only 1.79% is audiovisual content and 0.01% from 3D. A network of collaborators verifies the collected data, associates them with others by common topics, people or places and reinforces them with geolocation. Europeana shares the actions of its cultural policy [Ist15].
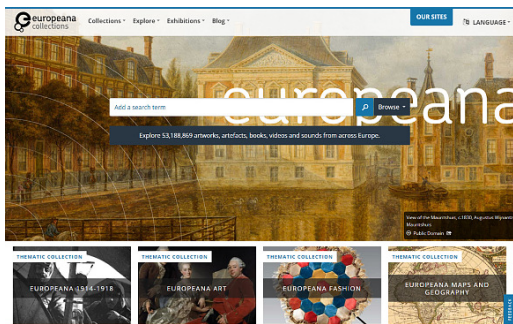

Figure 1: *"Europeana" platform*

2.1. "DARIAH": the pan-European infrastructure for arts & humanities scholars

Collaboration with other realities, including Europeana, is the basis of "DARIAH, Digital Research Infrastructure for the Arts and Humanities", a network of professionals, skills, tools, knowledge, sharing and technologies from all over Europe who, through collaboration, it supports and disseminates research on information and communication technologies, applied to artistic and humanistic activities, offering training and education opportunities. The transdisciplinary approach of DARIAH is visible in the four sectors involved, research, education, culture and economy, since humanistic research skills, in rapid technological evolution, play a decisive role in the digital age, such that they can be employed in a commercial context. The Italian office of DARIAH is located at the National Research Council-CNR OVI in Florence. "DARIAH.it, Developing nAtional and Regional Infrastructural nodes of dAriaH in Italy is also the name of the PON project, promoted by the CNR, for the updating of the research infrastructure involving 6 Italian geographical nodes, for the construction of as many Data Centers, of which 3 planned in the regions of southern Italy, for supercomputing, storage, virtualization, 3D rendering services at the service of digital humanities [RE19].

2.2. Digitization of spanish archives and libraries

PARES is the main platform for the dissemination of the Spanish historical documentary heritage, created and managed by the
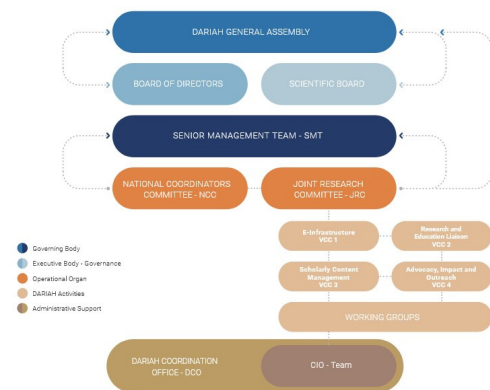

Figure 2: *DARIAH Organization*

General Sub-Directorate of the State Archives of the Ministry of Culture and Sports in 2007 and continuously implemented, thanks to the identification, description and digitization work of the documents conserved at the state archives, including the National Historical Archive and many others [L.B20]. From the website of the *Biblioteca Nacional de España* it is possible to reach the Hispanic digital library, created in 2008 to disseminate cultural heritage, made up of printed books between the 15th and 20th centuries, manuscripts and many others. The file "process of digitization in the national library of Spain" shows in detail the stages, formats, technical characteristics, metadata standards METS (Metadata Encoding and Transmission Standard)-PREMIS used. There are many other digital resource containers in Spain: the Miguel de Cervantes virtual library, the digital library of the Royal Academy of History, the digital portal Hispana, all working to power the Europeana platform. According to the business newspaper Expansion, Spain has recovered two places in the DESI for 2022, considering that state investments in digitization in the period 2021-2022 have increased nine-fold compared to the previous two years. On the other hand, Spain, together with France, is among the European countries ahead in the implementation of its PNRR.

2.3. "Gallica" and "FranceArchives": portals of the french archival and book heritage

France, unlike Italy, has arranged to integrate the Next Generation EU grants with national funds: the *Plan national de relance et de résilience* presented by the French Government is part of the overall 100 billion euro plan "France Relance". The Ministry of Culture has been encouraging the digitization of book collections since 1996 and since 2007, in collaboration with the National Book Center, has promoted the development of "Gallica", the digital library of the *Bibliotèque nationale de France* and its public partners. Gallica currently allows the consultation of 9,427,592 digital contents, including manuscripts, printed publications, audio, video, maps, photographs. International cooperation prompted the *Bibliotèque nationale de France* to collaborate in the implementation of Europeana and to launch, in 2011, "Data" which aggregates multiple national and international databases. On the con-

trary, "FranceArchives", born in 2017, is the national portal of the central and peripheral, public and private archives of France. While not allowing direct access to the digitized images, but the connection to the portals of the individual archives, the richness of the digitized documentary collections allows us to range between parish and civil status registers, personnel registers, municipal or departmental resolutions, photographs and audiovisual documents, maps and plans [Min22].

2.4. German cultural heritage portal: "Deutsche digitale bibliothek"

Among the major beneficiaries of European funds, there is also Germany, whose *DARP, Aufbau und resilience plan* allocates 52% of the total to digitalization, affecting many sectors, from industry to education, from social and health care to the public administration. From the site of the Ministry of Culture and the Media, there is a reference to the "Deutsche Digitale Bibliothek", the platform intended for all the cultural and scientific institutions of the federal states and municipal authorities; museums, archives and libraries put their heritage into it in digitized form. Born in 2014, today it has the participation of more than 4,400 institutions and the presence of more than 45 million digital objects. The platform is enriched with specific portals, "Archivportal-D" and "Deutsches Zeitungsportal", which respectively reveal archival data and periodical material. The Deutsche Digitale Bibliothek contributes to Europeana [BD20].

3. The italian projects

The Ministry of Culture, aware of the need to enhance digital collections, to promote the long-term use of digital archives and to adopt open access policies, has considered the potential of the P.N.R.R. for the relaunch of cultural heritage digitization policies cite [Pir22].

3.1. The P.N.R.R.: National Recovery and Resilience Plan. The role of the "Central Institute for the Digitization of Cultural Heritage – Digital library"

Mission 1, Component 3 (M1C3) of the Ministry of Culture is divided into three measures, including one reserved for "Cultural heritage for the next generation" and aimed at creating a digital heritage of culture: the use of 500 million euro will serve the coordinating structure of digitization programs, the "Central Institute for the digitization of cultural heritage - Digital library". Founded in 2020, it expresses mandatory and binding opinions on the initiatives of the Ministry of Culture, prepares censuses on the state of digitization, supports and defines the reference space within which each institution will be able to implement its own digital transformation, elaborating the National plan for the digitization of cultural heritage.

3.2. The digitization activities of the "Direction general for archives" and the "Direction general for libraries and copyright"

On the other hand, the Directorate General Archives of the Ministry of Culture, since 2000, has undertaken the digitization of its

printed publications, making them available online. Furthermore, in the "Catalogue of publications of the Directorate General" section, it is possible to consult numerous series and periodicals of an archival nature, as well as forthcoming publications, those of archival institutions and those of other publishers, including those promoted by the Italian National Archival Association. Since 2005, the Directorate General for Libraries and Copyright of the Ministry of Culture has been sharing the results of the digitization projects of Italian libraries and prestigious Italian cultural institutions on the "Internet Culturale. Catalogs and digital collections of Italian libraries", edited and directed by the Central Institute for the single catalog of Italian public libraries, ICCU. In continuous expansion, it presents itself as an aggregator of heterogeneous digital bibliographic information: manuscripts, books, scores, geographical maps, images and sound recordings are indexed with metadata in MAG (*Metadati Amministrativi Gestionali*) format [Cer19]. Quite recently has been set up *Alphabetica*, the new digital ecosystem that collects millions of contents and bibliographic information from over 6,500 Italian libraries and institutions, members of the National Library Service. The project sees the use of Application Programming Interfaces, APIs, and the IIIF standard, which make the bibliographic ecosystem interoperable with other external resources.

3.3. The digital library of the "Central State Archive"

Remaining within the framework of the offices of the Ministry of Culture, endowed with scientific, financial, organizational and accounting autonomy, on 1 March 2023, the Central State Archives presented the digital library to the community of users and scholars, a platform, which facilitates access to digital collections, with a simple click from the "online consultation" section. Massive digitization of 1,500 inventories that can be searched by keywords, hundreds of photographs, files, documents, thousands of images including technical drawings, posters and registers was launched in 2022, it is estimated that between 2023 and 2024 the digital library will be further enriched, in view of specific thematic itineraries. A link also allows consulting the entire archival heritage of the Central State Archive. The digital contents, freely downloadable for study and research purposes, are published in the IIIF framework, which allows their interoperability and wider sharing. Among other things, the Central State Archives will be the implementing body of the sub-investment 1.1.8 of the M1C3 for the creation of a digital conservation hub, which will bring together two conservation systems: 1. the digital preservation pole of the State Archives, which concentrates the historical digital archives of the central and peripheral administrations of the State, of public bodies of national importance and of private archives declared to be of particularly important historical interest; 2. the digital archive of the Ministry of Culture.

3.4. "I.C.AR., Central Institute for ARchives": the Digital Archive project

The other institute with special autonomy, the "Central Institute for Archives" (I.C.AR.), gave birth in 2018 to the Digital Archive project, which allows the consultation of the digitized documentary heritage of the State Archives and archival and bibliographic

Superintendencies. Now there are 19 affiliated institutes. MetaFAD developed the open-source platform. A user-friendly graphical interface leads to the exploration of individual digitization projects, accompanied by descriptive, structural, administrative and management metadata.

4. The digitization project of the Vatican Apostolic Library and the "DigiVatLib" web portal

In 2012, the Vatican Apostolic Library started an intense digitization and long-term conservation project, after a period of 2 years, during which preliminary analyzes and acceptance tests of the hardware infrastructures and conservation formats were required [Ent22]. The European Space Agency and the National Institute for Astrophysics, with a view to integrating resources, have supported the ambitious project to digitize over 80,000 manuscripts using, instead of the usual image formats for digital archiving (TIFF and JPEG), the F.I.T.S. (Flexible Image Transport System), which was adopted by N.A.S.A. since the 1970s for long-term data preservation. Fundamental is the design of a digital archiving service called Advanced Museum Library Archives Deposit, AMLAD, a cloud computing technology, compliant with various international standards for both archiving and long-term preservation of digital documents, such as Open Archival Information system framework, OAIS, regarding metadata, such as Dublin-core, METS, MODS (Metadata Object Description Schema). The interoperability of the resources is guaranteed by the OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) and SRU/SRW SRU (Search Retrieve via URL/Search Retrieve Web Service) protocols. The digitized collections, divided between manuscripts, incunabula, graphic materials and art objects, coins and medals, archival fonds, printed-special projects, can be used through the "DigiVatLib" portal which satisfies the IIIF framework, designed to facilitate access by researchers from all over the world. Please refer to the appendix, by way of example, for the description of the configuration of the Data Center of the Vatican Apostolic Library (see section 6).

Figure 3: *"DigiVatLib"*

5. MA.GI.C.: contribution to technology research applied to cultural heritage

The Federico II University of Naples, through two of its Departments, the "Ettore Pancini" Physics Department and the Humanities Department, and in collaboration with private companies, pre-

sented an important digitization project in 2019, addressing the prototyping of a selection of the manuscripts conserved in Public Libraries. As anticipated, the goal is to create a Service Center prototype, capable of developing technologies and processes for the treatment of ancient texts and manuscripts, aimed at their conservation and use. The service center will be located in the Federico II University of Naples, with the clear intention of contributing to "information and communication technologies" and extending its services to the national and European territory [RAA*20]. To achieve this goal, the project combines the industrial skills with the purely scientific skills of the largest university in the south, the Federico II University of Naples, in order to develop cutting-edge technologies in a sector that for the first time intends to develop a digitization process of ancient and non-ancient texts that is accompanied by technologies:

- of IoT for volume monitoring and protection;
- of IT for the use of the text based on his story;
- Artificial Intelligence algorithms for recognizing different types of writing and printed and handwritten characters; for image extraction; for the pre-cataloguing of texts and images;
- big data applied to information retrieval for an accurate access system to digitized materials.

The Center is designed so that it can generate culture and create value, through concrete activities in the fields of restoration, digitization, cataloguing, also training young people capable of exploiting diversified methodologies and skills, all of which can be found in the subjects proposing MAGIC.

5.1. The F.I.T.S. format: innovation and experimentation

One of the goals of the project lies in the choice of format, F.I.T.S. (Flexible Image Transport System), which ensures the long-term preservation of digital content, in this sense of the digitized two-dimensional images of manuscripts (X, Y). Nonetheless, on 20 January 2022 the UNI 11845:2022 standard "Processes for managing the long-term conservation of digital images with the use of the F.I.T.S. format" came into force. which outlines the requirements of integrity and durability over time that an archive must satisfy, in relation to the properties of the format, i.e. freedom of use, self-referentiality, transparency, and autonomy. The F.I.T.S format it is open-source, therefore it is free to use and does not require any user licence [Bar07]. A set of rules of the F.I.T.S. requires you to write data in a standard format, which makes it compatible with any type of operating system. A F.I.T.S. consists of one or more Header + Data Units (HDUs): the first HDU is called the primary HDU or primary array and is followed by other HDUs called F.I.T.S. extensions, in fact each HDU consists of a Header unit, readable in ASCII characters, with the description and information on the data, in binary form, and an optional Data Unit. The primary HDU can be empty or contain a mix of pixels, 1-D spectra, 2-D or 3-D images, and support up to 64-bit data types. The headers include fields that start with mandatory keywords of 80 characters that specify the size and format, as well as optional keywords that can specify Comment and History, useful for better describing the data content of the files and, above all, providing metadata. You can provide up to hundreds of lines of comments and observations that accompany the image and are stored in the next Data Unit; for this reason

it is said that the F.I.T.S. it enjoys the important characteristic of being self-documented. Moreover, the search for keywords is facilitated by the reading, by the search program, exclusively of the text part of the file, omitting the image. The FITS format does not consider data compression: any compression would only affect the image, which in itself remains a FITS file, leaving aside the header with the keywords; also in this case the speed in the search is guaranteed. Also, the F.I.T.S. format can store multiple images in the same file. In addition, three types of standard extensions are included: images, ASCII character format and binary representation. The International Astronomical Union F.I.T.S. Working Group is responsible for managing, enriching and improving this format and always keeping it in step with the needs of modern science, in fact in July 2016 version 4.0 was adopted, and officially released in August 2018, with new features, still while maintaining full backwards compatibility. The watchword "once F.I.T.S, always F.I.T.S" guides The International Astronomical Union.
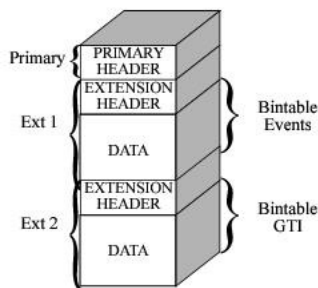

Figure 4: *F.I.T.S. header*

5.2. Workflow. Processes and prototype results

The workflow will unfold through various phases, as follows. In the prototyping phase, the parts that can be automated and/or robotized will be highlighted.

- **Handling of the material**. For safety reasons and in order to minimize the physical stress to which ancient books, with a particularly fragile support, are subjected, a preliminary evaluation is of their location, i.e. their careful transport from the warehouses, where they are usually stored, up to the place where the digitization operations will take place.

- **Prototype creation**. Once the corpus of volumes to be digitized has been identified, a prototype must be created for each different digitization activity; periodic checks will allow to eliminate any manufacturing errors. Using a planetary scanner, we made the digital acquisition of the pages of the printed musical score "Tristan und Isolde" by Richard Wagner, datable around the second half of the 19th century. Good quality jpeg files have been prepared with 24-bit color depth and 330 dpi optical resolution. Furthermore, the digital acquisition of the individual pages of the printed monograph "Memorie della regale Accademia ercolanese di archeologia", dating back to 1862, was followed, preparing additional jpeg files with 24-bit color depth and 300 dpi optical resolution.
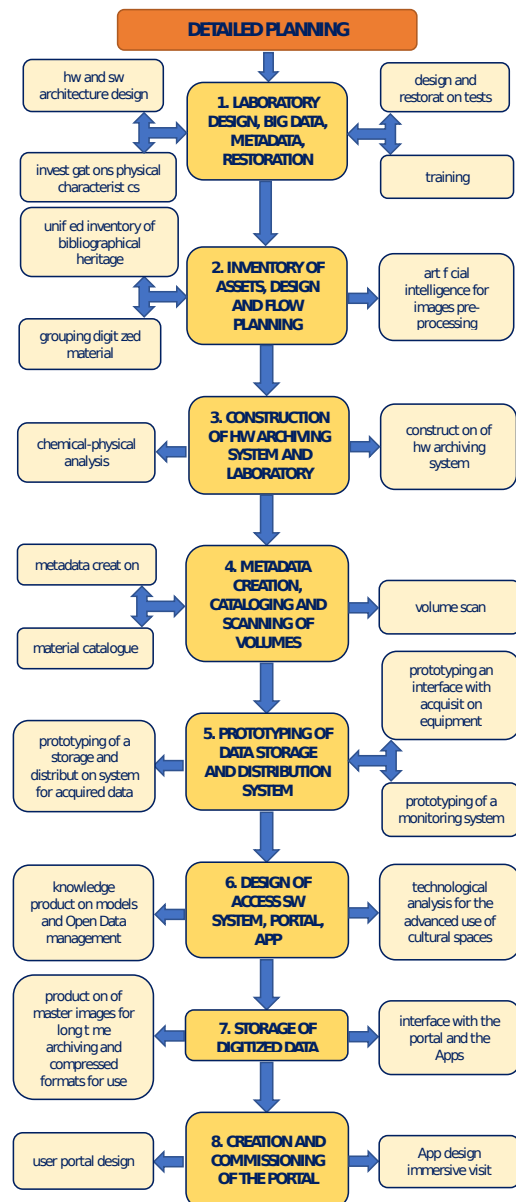

Figure 5: *Ma.G.I.C flowchart*

- **Digital capture**. The digital acquisition will be carried out by personnel capable of making the most of the available instruments, i.e. planetary scanners, with lighting contained within safe levels, able to scan very large documents or documents whose binding is fragile. Such equipment captures each paper of the manuscript in JPEG2000 format and stores it on a staging area. The high-definition digitization of manuscripts, incunabula and sixteenth-century books will be accompanied by a campaign to raise awareness of library objects, for each of which a detailed file will be prepared which will describe its content, material consistency, cultural history, state of conservation, also making use of what is already being achieved at the advanced

training course in "History and philology of the manuscript and of the ancient book" promoted by the University of Naples Federico II. In order to plan the processing flows and verify the optimal data production, an artificial intelligence system will structure the processing batches in real time, grouping them according to common characteristics. Eight months will be needed to divide the bibliographic material according to size, state of conservation, amount of digital data produced, state of inks and papers, able to agglomerate it in reference to the coefficient of difficulty.The main agglomeration criteria will be:

–   the state of conservation of the paper support;
–   the readability of the ink, with particular attention to the oozing of the latter;
–   the physical characteristics of the volumes, including dimensions and color shades of both the paper and the ink;
–   the regularity and orientation of the writing style, trying where possible to group together any collections of works written by the same writer;
–   the projection of the amount of digital data produced, to ensure that each batch can be managed in the best possible way by the subsequent IT systems and completed in a short time;
–   the degree of complexity of the work, mainly determined by the state of conservation of the volumes.

The first operational step of the project will be the digitization of selected manuscripts and books which will be carried out at high resolution (600 dpi) using planetary scanners and according to the guidelines of the International Federation of Library Associations and Institutions. Taking into account the value and extreme delicacy of the originals, the shooting methods (opening of the volumes, lighting, handling) will be such as not to cause them damage, and to this end the operator will verify in particular for each volume the positioning, opening and the possibility of single-page or double-page shooting. To carry out the digital shots, highly professional scanners are used for the reproduction of ancient volumes, i.e. equipped with tilting shelves to accommodate the books and, if necessary, usable with the addition of V-shaped book lecterns that allow suitable placement of books with a limited opening angle. With specific reference to the material being processed, a process may be appropriatewith a tilting system opening at 90° which allows you to preserve the material itself as much as possible any form of stress or treatment deriving from the classic 180° process. The tilting panel (in the figure) allows it to be adapted according to the original to the specific needs, in particular it can reach an opening of 90° but also work at 120° or at other levels. Through a manual movement mechanism, the volume is positioned in the recovery area and the subsequent digitizations are carried out by progressively leafing through the pages. All operations are carried out in manual mode and therefore as the volume scrolls during the reproduction phases, adjustments can be made to guarantee its correct housing. The system just illustrated represents the maximum from the point of view of safeguarding and treating rare and valuable books. For each volume, a preview will be performed to delimit the scan area, leaving a small margin to compensate for any micro-rotations. During all operations related to digitization, the personnel in charge must wear special cotton gloves. The reproduction will take place with the use of cold light lamps (5400

Kelvin), with no ultraviolet component. Resolution, color depth and illumination are parameters which should be decided taking into account the specific widely accepted standards for the type of material in question, based on the requirements of the master file retention and the requirements of presentation and use provided for digital objects. Other elements to consider regarding image quality are color saturation, brightness, image integrity, and the absence of halos and other optical defects.

●   **Digitization operations**. The double image will be taken with a single scan but an image file (digital object) will be produced for each of the two parts relating to:

–   each verso and recto of each paper for manuscripts; each page for printed texts, including endpapers, even if without information, and both interpolated and consecutive blank pages; periodicals are treated on the page and as files, not as bound volumes;
–   all the component parts of the binding: plates, spine, cuts if significant;
–   the chromatic scale and millimeter scale will be placed only once on a non-significant original paper of each document; the paper to be scanned will be agreed with the project manager and placed as the last image file in the digital package; the millimeter scale must be positioned along the lower edge with the "zero" aligned with the vertical edge of the paper;
–   the leaves will be reproduced in a "natural" way, including the margins of the leaves and parts of the underlying volume, to show the materiality of the document in its entirety, without interfering with the quality of the colors on the depth of field. The shots will be taken on a black background so as to include a minimum dark area surrounding the scanned paper which allows it to be viewed in the context, in any case in such a way that no part of the document is ever excluded from the shot. Papers/pages in the binding area should be trimmed with a margin to show the binding and to a small extent the facing page;
–   in case of lacerations, woodworm holes and oxidation of the inks, the papers must be masked with white paper in order to avoid copying the underlying content;
–   In the presence of papers to be resumed, smaller than the underlying sheets, place under the resumed sheet, a Japanese paper (not a common white sheet), thick enough to allow the visibility of the underlying pages so that they do not interfere with the reading of the scanned sheet, of a dimension equal to the measures of the document; The scan folder will be organized by placing in order: front plate, spine, back plate, and the internal papers in sequence, at the bottom of the entire package the chromatic and millimeter scale.

●   **Image capture formats**.The acquisition must be performed respecting the following parameters: High resolution uncompressed TIFF 6.0, 600 dpi optical, 24-bit RGB color depth for size less than or equal to A4; or 400 ppi, 24-bit RGB color depth for significantly larger than A4 size. The acquisitions will essentially have to guarantee the 1:1 ratio between the dimensions of the analog object and those, "print size" (DPI), of the TIFF version of the image This digitization is intended for offline storage and as a backup copy, it is called (master). Once the quality and completeness has been verified by the master, the follow-

ing will subsequently be produced: JPEG in medium resolution compressed format, 300 dpi optical and 24 bit RGB color depth, intended for intranet consultation; JPEG in low-resolution compressed format, 150 dpi optical, 24-bit RGB color depth, 85 percent compression factor, with image resampling, a method by which the "print size" (DPI) of the JPG file will be identical to that of the TIFF file. The value of the dpi resolution resizing can be increased to 200 dpi, or decreased to 100 dpi, its value must be defined in the prototyping phase, according to easy consultation, depending on the quality of the material (inks, quality of writing , of the print, body of characters), also in relation to the size of the original, always taking into account not to underestimate the too much weight of the final digital file for the purpose of uploading it to the repository and being consultable by users. Particular attention must be paid to the choice of dpi for periodicals, determined by the font size when very small and by the size of the sheet; in fact, good legibility on video with the zoom is essential. However, an evaluation on the use of this modality must be analyzed by all interested parties in the pre-prototyping phase.

- **Post-production treatment**. Any corrections to the files, minimal and only if necessary, must be established at the beginning of the project. Images will undergo post-capture quality control to ensure accuracy and integrity of the final product, with the aid of calibrated monitors. Images that do not meet project quality standards will be re-scanned and replaced; likewise the missing images will be retrieved and inserted at the correct point in the sequence. Structural metadata will be reviewed and modified if necessary. Also in the prototyping phase, the external margins of the cards will be trimmed for a correct framing and exposure of the same that did not jeopardize the conformity of the reproduction.

- **Metadata**. At the same time, the metadata will be added, the structure of which was previously defined in agreement with the Library managers, and F.I.T.S. files will be created. which will be stored in storage. As specified, the header of the F.I.T.S. they can include multiple keywords, which can provide metadata. The creation of metadata will be completed by the creation of RFID (Radio-Frequency IDentification) tags, capable of remotely detecting the information contained therein and allowing for the management of the library heritage very effectively: the affixing of a small two-dimensional microchip which, thanks to an RFID reader, will activate the data exchange. The enormous amount of data, which can vary from tens of terabytes of data to hundreds of petabytes, requires accurate technology with a Big Data approach: the Data Center of the Federico II University of Naples will archive the project storage. Please refer to the description of the configuration of the Data Center located in the Monte Sant'Angelo Campus university complex, owned by the University of Naples Federico II (see section 7). However, the digitization operation will not be the supporting element of the project, but rather the creation of an ample metadata base relating both to the description of the object-manuscript and to all the existing studies on the subject, a sort of cross-reference of excellence that will elevate the digitized document to a real entry-point in history, sociology, philology, for a complete, multilingual and multicultural path that has no equal in Italy. And the high usability will be the winning element: from print reproduction, to

access on smartphones, tablets, PCs with the availability of multiple resolutions for all needs, obviously all obtained from a single high resolution scan. The set of information that represents the document forms the basis of the organization of the document repository as well as the main criteria for searching and consulting the digitized works. Thanks to software specifically developed to fully meet the needs of the project, the operators involved in the indexing process will be able to view the scanned pages and, by filling in a simple guided form, be able to generate a metadata dataset which can subsequently be exported to the main standards in use. For each volume of the library, both descriptive metadata, therefore suitable for describing and facilitating the search for the resource, and any structural metadata, i.e. aimed at highlighting the relationships between the volumes, will be acquired. The choice of the fields to be surveyed in this phase will be the union between the classic fields common to book works (for example title, publisher, year of publication, etc.) and the fields necessary to allow the completion of the most frequent searches, as indicated by the operators of the library itself. The main metadata standards for the cataloging of books and texts will be considered, among which the METS standard and MAG 2.01 will be analyzed with particular attention. Cultural Internet metadata is written in xml format and according to the schema MAG standard, version 2.01.

- **Investigations on the chemical-physical characteristics**. On selected volumes investigations will be carried out on the physical, chemical and biological characteristics of the materials that make up the ancient books, on the agents of degradation inside the paper, on oxidation, on physical degradation, on the effects of humidity, temperature of radiation, as well as on the physycal characteristics of the location and of its wood furniture. On these specific manuscripts, a microscopic analysis of the paper fibers will be carried out, on the chemical-physical characteristics of the cellulose, on the cellulose molecules, on the structure of the fibers, on colloidal adsorption, imbibition, capillary absorption, on the degree of sizing and on the angle of contact. We will also proceed with traditional but highly technological techniques. The study of the state of conservation and the relative diagnostics of the manuscripts can only pass through a non-destructive analysis of the work: in this respect the optical techniques of spectroscopic investigation, coupled with traditional or confocal microscopy techniques, are irreplaceable. Portable versions of the spectroscopic characterization instruments will be purchased, to allow for in situ analysis of the works, where the dimensions or conditions prevent transport or observation with traditional instruments. The Material Optics Group (OdM) has been carrying out a scientific activity for years, attested by numerous publications in international journals, aimed at the optical characterization of materials and surfaces. The techniques used range from the more traditional ones (visible and infrared spectrophotometry, photoluminescence, Raman spectroscopy, FT-IR) to the more advanced ones with the possibility of optical characterization at high spatial resolution thanks to the availability of various optical microscopy apparatus at scanning (confocal microscopy, AFM, Micro/Nano-Raman, micro-FTIR). In the project, the OdM Group will contribute to the analysis of the typology and state of conservation of inks and binders, diagnosing their possible state of chemical, biological or light-induced

degradation. Portable versions of the spectroscopic characterization tools will be purchased, to allow the in situ analysis of the works, where the dimensions or conditions prevent their transport or observation with traditional instruments. Manuscripts are subject, during public display or conservation, to threats to their integrity from environmental factors.Among these, the atmospheric conditions and the presence of pollutants in the gaseous phase are the main factors of degradation. Accurate monitoring of these conditions, both for the purpose of punctual detection and continuous environmental control, could be decisive in ensuring correct conservation of the asset. Moreover, the conditions, both externally and internally, of atmospheric pollution are affected by localities due to the effects of turbulent transport or even diffusion in the presence of significant temperature gradients and therefore require pervasive or in any case punctual monitoring in the vicinity of the work to be preserved. It is our intention to carry out a monitoring activity of the environmental conditions in indoor environments in which works of significant cultural interest are hosted, using transducers capable of monitoring in real time the environmental conditions (RH, T) and the main atmospheric pollutants relevant to the works in question (NO2, O3, NO, SO2). The activity will be accompanied by a campaign, extended to the areas of interest and carried out with commercial instruments, of measurements of the environmental parameters that influence the microclimatic conditions.

- **Storage**. The archiving and analysis process requires the use of a Data Center that guarantees data security and adequate quality of service. In the first phase, throughout the project, the scientific Data Center of the University Complex of Monte S. Angelo will be used, currently being upgraded thanks to the PON IBISCO project (Infrastructures). This Data Center is monitored 24/7 and is equipped with a very high-speed Internet connection. The system will use a fixed filename system, so that filenames and all corresponding metadata can then be entered into a database for ease of searching. This Data Center is able to provide adequate computing power for the automatic analysis of texts (through the use of particular calculators equipped with a graphics processor) and has specialized systems for archiving and managing large amounts of data. The archiving will concern both the original high resolution image (raw data), and the images created starting from this with special software, for example to guarantee a resolution suitable for the web, or to guarantee good readability on portable devices such as smartphones. The mass storage systems that will be present in the aforementioned Data Center can also be used for long-term archiving, coupled with a Cloud access paradigm (on the OpenStack platform) which will allow them to be used by scholars.

- **Artificial Intelligence**. A further activity, the easy recognition of the library heritage, will involve artificial intelligence: the cataloging of manuscripts. From the analysis of the images in progress, a system of algorithms will be able to carry out a self-learning process, which will lead to the pre-cataloguing of texts and images and to the codicological and paleographic recognition of ancient volumes. Starting from an on-site comparison, it is estimated that the work will take 18 months, within which to proceed, first, with the digitization of the handwritten and printed historical catalogs that document the subsequent acquisitions, and then with the establishment of a unified digital inventory. A

summary file for each manuscript, incunabula and sixteenth century will illustrate the extrinsic and intrinsic characteristics: writing support, dimensions, consistency, presence of decorations, such as miniatures and engravings, author, title and date. The digitization phase will be followed by the research and development of the recognition application and the semi-automated transcription tool of ancient scripts and characters. An initial division of the volumes, characterized by ink in a good state of conservation and clearly legible writings, will produce immediate optimal results, capable of providing solid learning bases for the subsequent and more complex processing phases. The artificial intelligence will perform an additional function: the algorithmic removal of the degradation of the support, due to the state of the inks which, penetrating the papers (bleedthrough effect), compromise the readability between the recto and verso of the paper and, therefore, the accuracy in automatic character recognition.

- **Augmented reality**.Augmented reality is a blend of physical and digital worlds, which allows a user to intuitively perform 3D human, computer and environmental interactions. In recent years the use of mixed reality to augment and improve a learning experience has seen a steep rise, and many applications have been developed by museums and other cultural institutions (some examples: https://www.museumnext.com/article/how-museums-are-using-augmented-reality/). The MA.G.I.C. project has chosen to employ augmented reality to increase the amount of information which is given to the user. Instead of accessing a database of digital documents which only brings the content of the manuscript, incunabula and sixteenth century a user, equipped with a augmented reality headset such as the HoloLens 2 or a standard smartphone, may choose to project a digital twin of the manuscript directly to its table, enjoy the beauty of the manuscript itself, and appreciate the work that medieval monks put in crafting those ancient books. Through hand and gesture tracking the user may interact with the manuscript, turn pages, and access additional information by opening paper notes and bookmarks left in the books by the monks. These may include personal notes, translations, and links to in-depth analyses. All development will be carried out with Unity one of the leading real-time development platforms on the market which allows for a high level of compatibility and portability between devices. The final goal of the augmented reality MA.G.I.C. project is to generate an interactive bookshelf which will allow, only using your hands, to experience a book as it was standing in front of you. The goal is to bridge the gap between the user and the cultural asset, be it a book and/or an ancient manuscript, through a stimulating and innovative use that can bring the young public closer to this type of asset. The idea is to characterize the use by guaranteeing everyone the possibility of activating a mobile application on their smartphone and/or renting it at the entrance to the place of interest, and starting a visit inside an "animated" place. in a library that comes to life. The enhancement and use of a library pass through the use of innovative technologies that allow an immersive and experiential visit to the site itself. The smartphone/tablet becomes the device par excellence for learning about ancient books and manuscripts, with the help of augmented reality and virtual reality. The goal is to create a virtual assistant that helps the user while visiting the entire complex,

providing information on routes and objects, based on the time available, profile, preferences, context. The possible use of Bluetooth Low Energy (BLE) proximity sensors, combined with augmented reality, will make it possible to identify the position and points of interest. The activity also has the purpose of analyzing and identifying the best 3D reconstruction techniques applied to cultural heritage, with particular attention to the analysis of photogrammetry algorithms. The latter technique which allows to obtain, from a sequence of appropriately taken photos, a convincing digital model of the object which, inserted into a graphic engine, will allow the user to manipulate the object freely, via mouse or touch screen . With typical extremely intuitive gestures, the object can be rotated, enlarged or moved as indicated in the Europeana Media Policy. The extreme realism of the objects obtained with this technique, not being given by the polygonal density of the model typical of 3D scanners but by the extraction of textures directly from the scene photographed using high-performance cameras, allows photogrammetry to be applied in most cultural contexts in which will be operated. Finally, the activity will see the analysis of innovative fruition models capable of also using holographic image projection systems. The replication and contextualization of 3D holographic images of cultural heritage, such as ancient books and/or manuscripts, lends itself to various uses: reuniting the fragments/pages of a book that have been physically separated; use the reserve of books/manuscripts housed in the deposits; allow virtual exhibitions, or of individual works not available because they are being restored or on loan to other institutions.

- **Access**.The resulting archive, which will gradually be expanded with other digitizations, will conform to the most widely used international standards, and will also be accessible "via software", i.e. with the appropriate standard Application Program Interfaces, which will allow the development of applications that comparing the archives of various digitized libraries, a particularly effective tool for scholars in the sector. The industrial research activity starts from the analysis of the specific application of the MAGIC project, which today, given its complexity, sees the coexistence of different disciplines that deal with the description of cultural heritage.This coexistence, not always simple, is however necessary if the common objective of maximizing the value of the asset itself is to be achieved, through:

  – a rational use of the information collected;
  – more and better exchange and sharing;
  – the integration of different archives;
  – data retention.

From a technological point of view, the tools available today, such as the semantic web, represent an important starting point for achieving these objectives. It is necessary to create significant connections between data and to do this it is necessary to follow open and shared standards and to work on a model of sharing data and building mutual relationships between cultural objects so that the perspective of Linked Open Data can become truly effective and contribute to the real sharing of knowledge. The implementation objective is, therefore, aimed at defining the entire knowledge representation and management process, which assumes a pivotal role for the remaining implementation objectives. For this purpose, a methodology will be developed that encom-

passes structured knowledge in its entirety and stratification, capable of feeding an open knowledge network that rests its foundations on the Open Data and Linked Open Data paradigm. The production and management of knowledge will be entrusted to an Authoring platform, capable of collecting information from heterogeneous sources, structuring data entered voluntarily by domain experts, capturing what is now defined as social knowledge, in order to bring knowledge together appropriately. A knowledge that also aims to open up to the community according to the Open Data model, but which must also be interoperable with national and international standards. The activity will see the analysis and modeling of the links between different datasets which are graphically represented in the form of a large 'cloud' called "LOD cloud diagram", in which there is an interactive visualization of the groups of interoperable datasets. The goal is ambitious, as it aims at the creation of a single large data space connected to each other and accessible for different users and software applications that can discover new information, create new knowledge and, in the case of Open Data, collect and republish it freely. Linked Data marries with the vision of the Open Data movement which aims to break down the social, cultural, legal and economic barriers that hinder the free sharing of data between people and software agents. The possible scenarios in which LODs can promote interoperability between datasets are endless.

[ACD22].

Figure 6: *Augmented reality platform at Federico II University of Naples*

6. Example of applications

6.1. The Data Center of the Vatican Apostolic Library

During the planning stage, such a digitization project required the development of an appropriate Data Center for the conservation and management of digital data, considering that the size of each digitized page is around 200 megabytes and each volume includes about 500 pages. As far as storage is concerned, a range of around 40 petabytes was assumed.

6.2. The S.Co.P.E.- Re.Ca.S.- I.Bi.S.Co. Data Center in the Federico II University of Naples

The University of Naples Federico II has great experience in the field of Data Centers, as it manages petabytes of data from multidis-

ciplinary experiments. With an initial financial contribution from the MIUR (Ministry of Education, University and Research), the project "S.Co.P.E., Cooperative system for multidisciplinary scientific elaborations" began in 2006, which envisaged the construction and activation of a Data Center that housed approximately 300 eight-core blade servers, but capable of hosting another 500, and 220 terabytes of storage. It consists of 33 racks. Each rack is equipped with 20 servers Rittal s.p.a. handled the liquid cooling systems of the racks, with two redundant chillers and a distribution system that supplies cold water. Each rack consists of liquid cooling package (LCP) units [BMF09]. A subsequent project "Re.Ca.S.-Calculation network for Super B and other applications" was further funded by the MIUR (Ministry of Education, University and Research) in 2011 under the National Operational Program PON 2007-2013 "'Research and Competitiveness". The project implementers were the University of Naples Federico II, the University of Bari Aldo Moro and the Italian National Institute of Nuclear Physics (INFN) with its offices in Naples, Bari, Catania and Cosenza; the operational synergies with the University of Catania and the University of Calabria were also fundamental. Completed in two years, the Data Center has reached 4960 terabytes of storage, which conforms as an "archiving service" with two system nodes: a front-end that allows access to data, recorded in an FC box and a second which implements the software, maintaining the file name and location for all datasets [ABN*17]. A further contribution from the National Operational Program PON 2014-2020 "Research Infrastructure" and the MIUR Ordinary Operational Fund (DHTCS project) was addressed to the "I.Bi.S.Co.-Infrastructure for Big data and Scientific COmputing project". The project, coordinated by the National Institute of Nuclear Physics (INFN) sees other partners: the University of Naples Federico II, the University of Bari Aldo Moro, the Institute of Applied Sciences and Intelligent Systems "Eduardo Caianiello" (ISASI) of the National Research Council (CNR), the Institute for Electromagnetic Sensing of the Environment (IREA) of the National Research Council, the National Institute of Astrophysics (INAF), the National Institute of Geophysics and Volcanology. The goal was to consistently upgrade and upgrade the available scientific computing and storage infrastructure to keep the infrastructure competitive, funded with previous S.Co.P.E. and Re.Ca.S., as well as to create a digital platform aimed at collaborating with other scientific fields: the MA.G.I.C. Project, precisely, focused on transdisciplinarity, will initially use, for data archiving, 10 servers and 1 storage system with 1 petabyte.

Figure 7: *Data Center in the Federico II University of Naples*

7. Acknowledgments

References

[ABN*17] ANDRONICO G., BELLOTTI R., NARDO G. D., LACCETTI G., MAGGI G., MEROLA L., RUSSO G., SILVESTRIS L., TANGARO S., TASSI E.: *High performance scientific computing using distributed infrastructures. Results and scientific applications derived from the italian PON ReCaS project*. World Scientific Publishing, 2017. 10

[ACD22] ANDREOLI L., CIMINO M., DRAGO G.: *Linee Guida sulla digitalizzazione di documenti bidimensionali*. Università degli Studi di Padova - Sistema Bibliotecario di Ateneo, 2022. 9

[Bar07] BARVE S.: File formats in digital preservation. In *International Conference on Semantic Web and Digital Libraries '07* (2007), pp. 239–248. 4

[BD20] BERTA C., DUNKHASE G.: 33 million objects and counting: a conversation about the Deutsche Digitale Bibliothek. *DigItalia 15*, 2 (2020), 174–177. 3

[BMF09] BARBERA R., MANNO M. I., FARGETTA M.: The S.Co.P.E. Project. In *Proc. Final Workshop of the Grid Projects of the Italian National Operational Programme 2000-2006 Call 1575 '09* (2009), pp. 18–35. 10

[Cer19] CERIANI D.: Internet culturale: Catalogues and Digital Collections of Italian Libraries. *Nineteenth-Century Music Review 16*, 1 (2019), 161–171. doi:10.1017/S1479409817000805. 3

[Ent22] ENTE NAZIONALE ITALIANO DI UNIFICAZIONE: *UNI 11845: Processi di gestione della conservazione a lungo termine di immagini digitali con l'uso del formato FITS. Processes for managing long term preservation of digital images using the FITS file format*. UNI, 2022. 4

[Ist15] ISTITUTO CENTRALE PER IL CATALOGO UNICO DELLE BIBLIOTECHE ITALIANE E PER LE INFORMAZIONI BIBLIGRAFICHE: *Athena Plus. Access to cultural heritage networks for Europeana. Innovative tools for the creation, reuse and retrieval of digital cultural content*. ICCU, 2015. 2

[L.B20] L.BERGMAN T. L.: Portal de Archivos Españoles (PARES; Portal to the Spanish Archives). *Early Modern Digital Review 3*, 4 (2020), 230–233. doi:10.33137/rr.v43i4.36392. 2

[Min22] MINISTÈRE DE LA CULTURE: *Chiffres clés, statistiques de la culture et de la communication*. DEPS, 2022. 3

[Pir22] PIRRELLI M.: Corte dei conti: la digitalizzazione della cultura ancora in ritardo. *Il Sole24 ore* (ottobre 2022). 3

[RAA*20] RUSSO G., AIOSA L., ALFANO G., CHIANESE A., CORNEVILLI F., DOMENICO G. D., MADDALENA P., MAZZUCCHI A., MURAGLIA C., RUSSILLO F., SALVI A., SPISSO B., TROMBETTI G., ZOLLO G.: MA.G.I.C.: Manuscripts of Girolamini In Cloud. *IOP Conference Series: Materials Science and Engineering 949*, 012081 (2020), 1–8. doi:10.1088/1757-899X/949/1/012081. 4

[RE19] ROMARY L., EDMOND J.: A tangential view on impact for the arts and humanities through the lens of the DARIAH-Eric. In *Stay tuned to the future: impact of the research infrastructures for social sciences and humanities* (2019), pp. 149–158. 2