Umanistica Digitale – ISSN: 2532-8816 – n. 20, 2025: Special Issue
S. Conte – *Magic, a Service Center for Technologies Applied to Manuscripts and Printed Books*
DOI: http://doi.org/10.6092/issn.2532-8816/21038

Magic, a Service Center for Technologies Applied to Manuscripts and Printed Books. Project Overview and Progress Report

Stefania Conte

University of Naples "Federico II", Italy
stefania.conte@unina.it

Abstract

Magic, a three-year interdisciplinary project conducted by the Department of Humanities and the Department of Physics "Ettore Pancini" of the University of Naples "Federico II", has started and is developing a Service Center for the use of cutting-edge technologies in the sector of digital conservation and enhancement of manuscripts and ancient printed books. Through digitization, cataloguing, physical and biological diagnostics and with the aid of artificial intelligence, the project is making many digitized volumes, accompanied by cataloguing descriptions, available on a dedicated website: now there are almost twenty thousand digital reproductions, coming from the book collections of the libraries of the territory. Starting from the digital library, Magic aims to make digital resources accessible to the research community and beyond, also through modern technologies capable of creating value, knowledge, comparison, discussion, and sharing. The article describes the project in all its phases and its progress, highlighting its objectives, the target users, the procedures, and methods used, initial research results and prospects. The volumes of interest for the project are, presently, the illuminated codices of Dante Alighieri's Divine Comedy and the collection of incunabula and *cinquecentine* of the Accademia Pontaniana of Naples. Some issues related to the treatment of digital resources and the results of the research carried out by the project will be addressed: the use of formats for long-term preservation and the use of artificial intelligence to remove the phenomenon of ink penetration between the *recto* and *verso* of paper, known as the bleed-through effect.

Keywords: AIUCD2024, Magic, digitization, bibliographic resources, physical and biological diagnostics, digital conservation, F.I.T.S., artificial intelligence, bleed-through

Magic, un progetto interdisciplinare triennale condotto dal Dipartimento di studi umanistici e il Dipartimento di fisica "Ettore Pancini" dell'Università degli studi di Napoli "Federico II", ha fatto partire e sta sviluppando un Centro servizi per l'utilizzo di tecnologie all'avanguardia nel settore della conservazione e valorizzazione dei manoscritti e dei testi antichi a stampa. Attraverso la digitalizzazione, la catalogazione, la diagnostica fisica e biologica e con l'aiuto dell'intelligenza artificiale sta rendendo fruibili, attraverso un sito web dedicato, molti volumi digitalizzati, corredati di descrizioni catalografiche: al momento sono quasi ventimila le riproduzioni digitali, provenienti dalle collezioni librarie delle biblioteche del territorio. Partendo dalla biblioteca digitale, Magic mira

alla fruibilità delle risorse digitali da parte della comunità di ricerca e non solo, anche attraverso le moderne tecnologie in grado di creare valore, conoscenza, confronto, discussione e condivisione. L'articolo descrive il progetto in tutte le sue fasi e il suo stato di avanzamento, sottolineando i suoi obiettivi, l'utenza a cui si rivolge, i procedimenti e i metodi impiegati, i primi risultati di ricerca e le prospettive future. I volumi di interesse del progetto sono, al momento, i codici miniati della Divina Commedia di Dante Alighieri e la collezione di incunaboli e cinquecentine dell'Accademia Pontaniana di Napoli. Saranno affrontate anche alcune questioni relative al trattamento delle risorse digitali e gli esiti della ricerca portati avanti dal progetto: l'uso di formati per la conservazione a lungo termine e l'impiego dell'intelligenza artificiale per la rimozione del fenomeno della penetrazione degli inchiostri tra il recto e il verso delle carte, noto come effetto bleed-through.

Keywords: AIUCD2024, Magic, digitalizzazione, risorse bibliografiche, diagnostica fisica e biologica, conservazione digitale, F.I.T.S., intelligenza artificiale, bleed-through.

1. Introduction

Promoted by the Department of Physics "Ettore Pancini" of the University of Naples "Federico II", in collaboration with the Department of Humanities and three commercial companies with consolidated experience, SA Documents, EsseA Digit and NetCom Engineering, the Magic project embodies multidisciplinary expertise and the valorization of advanced transversal skills. It is a mix of industrial and scientific skills that lead to the development of innovative technologies in the field of protection and conservation, as well as the valorization and use of books and archive heritage.

We have designed the Service Center to generate culture through concrete activities in the fields of digitization, cataloguing and Digital Humanities. The activities of the Center include conservation, diagnostics, improvement of images for public use, through scanning and data processing activities, starting experimentally, and improving them with an evolutionary prototype approach.

The goal of the Magic laboratory is to make cultural heritage accessible to all, allowing the libraries involved in the project to create an open access digital library, applying technologies to provide easy access to digital collections, so that they are usable by different communities, such as scholars, researchers or a new audience of interested people.

Digitization is the only first step towards access and transmission of knowledge. Both in the prototyping phase and in the digital reproduction phase, we have focused on heterogeneous book objects. The selected books, due to their form and content, present different peculiarities and require a diversified and suitable approach. In fact, the subject of an initial prototyping phase was a printed musical score, dating back to the second half of the nineteenth century, Richard Wagner's Tristan and Isolde, and a printed monograph "Memories of the Royal Academy of Herculaneum of Archaeology" dating back to 1862.

The first phase of digital reproductions of the Magic project involved the follow-up of another project of our university, which studied the illuminated manuscripts of Dante Alighieri's Divine Comedy. These manuscripts, dating back to the XIV and XV centuries, are bibliographic objects with peculiar characteristics that require particular care and attention, given their value and their delicacy. It is a core of 283 illuminated codices, which come from various cultural institutions, Italian and foreign libraries, and archives (Fig.1). Among manuscripts, we highlight *Biblioteca nacional de España*, *Bibliothèque et Archives du château de Chantilly*, *Bibliothèque de l'Arsenal (Bibliothèque Nationale de France)*, *Archivio storico civico* and the *Biblioteca Trivulziana* in Milan, *Biblioteca nazionale*

centrale in Florence, *Biblioteca nazionale centrale* in Rome, *Biblioteca apostolica vaticana* in Vatican City, *Biblioteca nazionale Vittorio Emanuele III* in Naples.

Fig.1. Example of illuminated codices of the Divine Comedy. Madrid, *Biblioteca Nacional de España*, VITR/23/3 (Copyright of the image: Department of Humanities, University Federico II, www.dante.unina.it)

The second interest of the Magic project is to create a virtual collection of the 15th and 16th century editions preserved in the library of the *Accademia pontaniana*; a Neapolitan academy founded around the year 1443. The object of the digitization are six *incunabula* and 186 *cinquecentine*, coming from donation of the Academy's members Francesco and Luigi Torraca, respectively father and son. This book collection contains works of Greek, Latin, Hebrew, and Italian literature, as well as books of history, law, philosophy, religion, law, architecture, medicine, numismatics, astronomy, and geography (Fig. 2) ([2]:38-42).

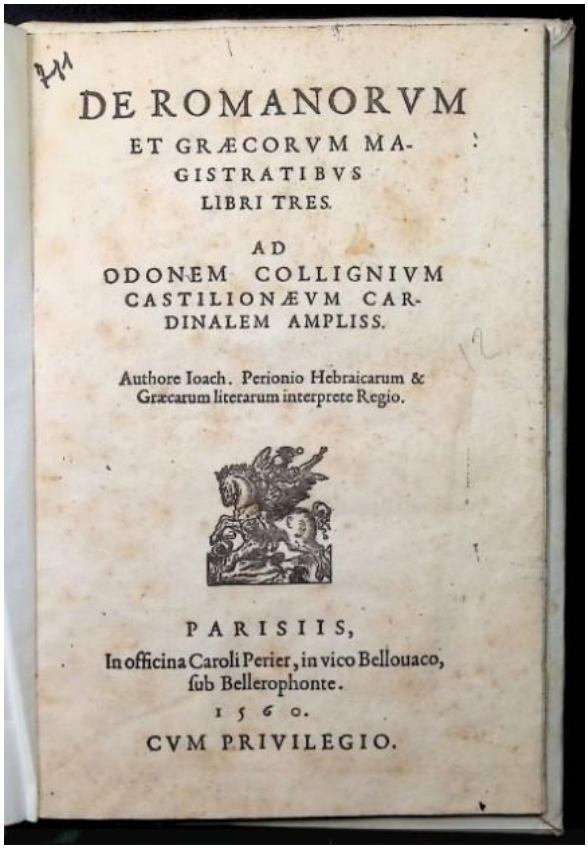
DE ROMANORVM
ET GRÆCORVM MA-
GISTRATIBVS
LIBRI TRES.
AD
ODONEM COLLIGNIVM
CASTILIONÆVM CAR-
DINALEM AMPLISS.
Authore Ioach. Perionio Hebraicarum &
Græcarum literarum interprete Regio.
PARISIIS,
In officina Caroli Perier, in vico Bellouaco,
sub Bellerophonte.
1560.
CVM PRIVILEGIO.

Fig.2. Example of *cinquecentina*. Naples, *Accademia pontaniana*, 5C40

In the conservation activities of the original books, the desire, and the need to preserve works that are often fragile and subject to the passage of time have led to the decision to limit direct consultation, favoring digitization activities. With a view to safeguarding the original material and its use, digital acquisition allows: i) the creation of digital copies, which, if performed correctly and for documentation purposes, take on an informative and cognitive value comparable to the originals; and ii) it promotes the virtual use of the originals themselves.

The Service Center is located at the University for experimentation and development, but the intention is to extend the services throughout the national and European territory, in fact the promoters are already taking action to ensure their continuity even after the deadline ([1]:1-8).

2. The subjects and institutions participating in the Magic project

The University is clearly present in the project as a Research body. The two local units are the Department of Humanities and the Department of Physics, both dedicated to research in their respective fields.

The Department of Humanities carries out research in the field of ancient volumes, making use of what is being done at the Advanced Training Program in "History and Philology of the Manuscript and the Ancient Book", promoted by the University of Naples under an agreement with the current Ministry of Culture.

The Department of Physics "Ettore Pancini" deals with the study of the conservation state and the related diagnostics of manuscripts and ancient books. In particular, the Materials Optics Group (MOG) of the Department has been conducting scientific activity for years, with dozens of 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 more advanced ones: we have the possibility of optical characterization at high spatial resolution thanks to the availability of various scanning optical microscopy devices (confocal microscopy, AFM, Micro/Nano-Raman, micro-FTIR). In the project, MOG is contributing to the analysis of the typology and state of conservation of papers and inks, diagnosing their state of degradation and contamination: it is essential to use techniques and procedures to monitor the progress of degradation and identify its causes.

The three commercial companies, SA Documents, EsseA Digit and NetCom Engineering, involved in the project, are all equipped with software research and development centers ([2]:87-93). They contribute to the project in six of the eight goals, as described in section 7.

3. The spending amounts

The three-year Magic project (2023-2026) was funded by the Ministry of Business and Made in Italy (MIMIT), thanks to the Sustainable Growth fund, intended to finance programs, research, development and innovation projects to relaunch the competitiveness of the production system, also through the consolidation of corporate research centers and structures. The overall value of the project for the entire three-year period is € 15,675,500, of which € 5,681,720 from by MIMIT as work progresses, based on periodic reporting.

The creation of the Magic Service Center, capable of combining scientific research in the sector and technological innovations, can offer an essential contribution to local development processes, in terms of economic progress of the cultural industry, direct generation of qualified employment, raising territorial quality with effective planning of the offer of cultural services that guarantees easy accessibility to the bibliographic and archival heritage, promoting its use and dissemination. The strategic objective of the Magic project partnership aims to strengthen the trust of companies towards universities and research centers in general: the research system based on the use of technologies can be considered a fundamental industrial asset for the development of demonstrators and their experimentation which, in turn, transforms into competitive advantages compared to other companies not involved in the project.

4. Management of the project

The project leadership of Magic is with the company SA Documents, but two other figures are foreseen: the scientific manager and an administrative assistant, who support the manager in conducting the activities delegated to him.

The key elements of the entire governance are the Management Committee, composed of one member for each subject, who has the task of verifying the good progress of the project, and an internal Scientific Committee for direction and control for the research institution "Federico II", composed of representatives of the two Departments involved. Specifically:

- Prof. Andrea Mazzucchi of the Department of Humanities (President),
- Prof. Guido Russo of the Department of Physics (General scientific director),
- Prof. Guido Trombetti, Rector Emeritus (member),
- Prof. Pasqualino Maddalena of the Department of Physics (member).

5. State of the art

The European Commission has designated this decade as the "European Digital Decade" and on 26 January 2022 a solemn interinstitutional declaration proposed on digital rights and principles. The focus is on the application of new digital technologies to culture.

The complex changes of contemporary times urge libraries and archives to rethink paths, approaches, reference models, to continue to serve the community, ensuring not only in-person consultation, but offering innovative and heterogeneous services, to promote easier, more advantageous and functional reading, study, and research activities.

A territory, such as Italy, characterized by the presence of hundreds of libraries and archives, requires adequate planning of the offer of cultural services that can promote their use, information, and communication.

The digital revolution has not taken libraries and archives by surprise, already attentive to the needs of their users. The digitization initiatives, conducted by archives and libraries belonging to research institutions, state institutions and private entities, are diverse and particularly challenging. Certainly, the restrictions imposed by the COVID-19 emergency have had a profound impact on the demand for access to cultural heritage, to which the initiatives of the Ministry of Culture for the virtualization and sharing of cultural content via the web have been able to provide answers.

It is no coincidence that the National Recovery and Resilience Plan (P.N.R.R. - section "Digitization and innovation", mission 1 "Digitalization, innovation, competitiveness, culture and tourism", component 3 "Tourism and culture 4.0") adopted in Italy, provides for investments in the digitization of the heritage held by archives and libraries. In 2020 the "Central Institute for the digitization of cultural heritage - Digital Library" was born, as a coordination structure for digitization programs, developing the National Plan for the digitization of cultural heritage.

In line with the digital transition referred to in the P.N.R.R., a further objective of the Magic project is to invest in the multidisciplinary relationships of Digital humanities, which concern the sciences of the book, the document and cultural heritage, in the light of recent developments in artificial intelligence. According to The Digital humanities manifesto 2.0, researchers called to build native digital methodologies, models, tools, and perspectives to shape innovation in information sciences and to promote the formation of networks of culture and diffusion of knowledge with methodologies, languages, and IT tools.

It is true that during the COVID-19 pandemic, significant efforts have made in digitization, but dozens of gaps in the digital transition still need to fill. The spread of digital technologies has encouraged and prompted new ways of enjoying cultural heritage. Archives and libraries are entering the fast and dynamic world of digital, rethinking their functions. For the purposes of physical conservation, the limitation of direct consultation of originals, but also of a more widespread promotion and accessibility of the same, we now consider unavoidable the creation of digital collections and innovative solutions for cultural heritage.

Since 2005, the General Direction 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 "*Cultural Internet. Catalogs and digital collections of Italian libraries*", edited and directed by the Central Institute for the single catalogue of Italian public libraries, ICCU. Continuously expanding, it presents itself as an aggregator of heterogeneous digital bibliographic information: manuscripts, books, scores, geographical maps, images, and sound recordings indexed with metadata in MAG (Administrative Management Metadata) format. 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.

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 forty-three affiliated institutes. A user-friendly graphical interface leads to the exploration of individual digitization projects, accompanied by descriptive, structural, administrative and management metadata. ([3]:119-128)

In this regard, we believe it is useful to mention Cristina Marras' contribution to the XIII Annual Conference AIUCD2024, in which she presents a "conceptual map" for the creation of interdisciplinary spaces and to support the theory and practice of modeling in Digital Humanities ([4]:43-47).

6. The goals of the Magic project

It is essential that libraries and archives ensure not only the traditional services of access, reference, and lending, but further expand the ability to arouse interest especially in young people, who are increasingly showing an appreciation for information and technology services, with the understanding that the printed word is not the only means of producing and distributing knowledge, but it is necessary to keep up with the times, also feeding into the processes of modern artificial intelligence.

Keeping this in mind, the multidisciplinary approach and the advanced interdisciplinary skills of the Magic project have considered the potential of Digital Humanities policies for cultural heritage. The Magic Service Center proposes and envisages a Smart Library model, capable of supporting the strong demand for culture and functioning well in a forward-looking way: an information hub, which provides access to information and improves information literacy.

A smart library is an information hub with innovative services that become intelligent only to the extent that they are intuitive and user-centric: for the user, the smart library is more user-friendly than intelligent ([6]:282-291).

Magic shares Rudolf Giffinger's definition ([7]:10-12), when it refers to the search and identification of intelligent solutions that allow modern libraries to improve the quality of their services. The project does not limit the sole acquisition of bibliographic and archival materials, but includes the use of artificial intelligence, automatic recognition of handwritten characters, interoperability, and interconnection with other information services.

On the other hand, another objective of the Magic project is to increase the tools that can guarantee the development of the cultural industry: the involvement of the three commercial companies and their internal staff, who deal with research and development, is giving a significant boost to the adoption of increasingly innovative solutions in line with the new needs expressed by the market of digital transformation of cultural heritage, placing culture in the fundamentals of the economy: the aim is to contribute to the creation of new professions, without neglecting the relaunch of the local territory and the development of cultural tourism, exploiting the benefits of public-private collaborations to support innovation, collaborations required and also stimulated by community directives.

Therefore, this scenario highlights the role of university research and serves to strengthen the trust of companies towards universities and research centers in general, with an evaluation of research as a fundamental industrial asset.

The activities conducted to achieve the final objective of the project divided into Realization Objectives (OR), divided between industrial research and experimental development. Eight intermediate Realization Objectives are foreseen, based on the technical feasibility and the scientific objective achieved and assigned in the amount of two for each of the four participating subjects.

- OR1: executive design of hardware and software architecture for the digitization of books and subsequent processing. This followed by the definition of analysis models for the processing of images and texts for the extraction of metadata and physical and biological analysis of the media (responsibility: University).
- OR2: activation of an Artificial Intelligence system for the removal of the bleed-through effect and subsequent character recognition (responsibility: EsseA Digit).
- OR3: experimental development of the image acquisition system and the data storage system described in OR1 (responsibility: University).
- OR4: identification and creation of metadata and cataloguing parameters for library material (responsibility: EsseA Digit).
- OR5: creation of a prototype of a conservation system, public use, interface and monitoring adequate for data to manage (responsibility: SA Documents).
- OR6: creation of knowledge production models and Open Data management and technological analysis for use, through augmented reality (responsibility: NetCom Engineering).
- OR7: production of image master for long-term conservation, using the F.I.T.S. format, and compressed formats for use and consultation (responsibility: SA Documents).
- OR8: design and creation of the web portal for general users (responsibility: NetCom Engineering).

7. Target audience of the Magic project: usability of web services

Web communication has led to an increase in the visibility of libraries and archives, effectively expanding the number of users who access them: the intent of digital collections and related digital services is precisely to favor the expansion of the user base, moving away from the problem of territorial limitations and adapting new technologies to the needs of users. In fact, the new challenge is to adapt libraries and archives to modern technologies, new communities, new user needs and added information behaviors. Smart libraries made for and with smart people. Not only are smart library services intuitive and user-centered, but they are also based on the vision or assumption that smart library users are an active producer of knowledge. To this end, the Magic lab is making information available through the provision of digital content, both for researchers, teachers, students, scholars in the field, and for a new audience of interested readers. This vision of a digital library/archive focuses on the user profile and the cultural context in which it operates, with a systemic and multidisciplinary approach, which allows intelligent interactions between the user and the cultural asset. Each library user is a producer of knowledge together with the other users ([8]:171-183).

Among the first objectives of the Magic project is the creation and implementation of the fruition website (https: www.magic.unina.it), whose main functions, as of today, are the display of informative content, the registration of users that brings additional possibilities, the catalog of digitized works.

These new tools have an incredible inclusive potential. The principles of universality and democratization on which this new environment is based lead to considering it as the ideal space in which people, from different cultures, can now enjoy greater representation in the spectrum of the library and archive audience ([15]:73-83). Furthermore, the addition of a second language to the Magic website allows us to expand the reach to a new national and international audience. In the future, it plans to develop web applications to support user interaction, taking advantage of augmented reality and virtual reconstruction.

Making bibliographic heritage accessible requires that the information be also adapted to an easy-to-read format. The identifying bibliographic information of each book arranged in such a way as to ensure that it can transmitted in a complementary way and with a clear information hierarchy. Each book displays an initial label with basic information adapted for easy-to-read purposes: title, author, collocation. The same information enriched by more detailed and specialized cataloguing when you click on the "Description" button. The narration designed from the very beginning with this inclusive approach guarantees effective communication with the target audience.

High usability will therefore be the winning element with the availability of multiple resolutions for all needs, obviously all obtained from high-resolution digitization. Not a mass digitization of books, therefore, but an operation for the masses, starting with students and ending with scholars who, in most cases, are not able to access the manuscript *de visu*, for obvious reasons of safeguarding the integrity of the volume, since digitization is part of both the conservation and fruition processes. For these reasons, the application of multimedia and new forms of participatory communication to the cultural heritage sector is considered an essential condition to ensure the definitive transformation of cultural institutions into "socio-cultural platforms for integrated development", capable of allowing active communication with their audience and fruition of their cultural heritage, without geographical borders and projected towards a future in which sharing and the open access model will be increasingly greater.

After all, the aim of the Conference AIUCD2024 has been to create connections between texts and people, establish communication between distinct cultures and create virtual contexts for sharing texts.

8. Approaches and methodologies used for the Magic project

The Magic project is using different procedures, technologies, and software tools, which over time integrated into a single structure, thus characterizing all the results as deriving from the new Magic approach. Existing tools used as much as possible, but there are cases, such as correction of bleed-through effect and use of long-term preservation format, where new ad hoc tools introduced by the Magic group and discussed in the following paragraphs.

8.1. Digitization processes

First, it is necessary to distinguish the work on manuscript books from that of the printed books, which require two different methodologies.

The first phase of the Magic project involved the follow-up of another project of our university, which studied the illuminated codices of Dante Alighieri's Divine Comedy, bibliographic objects with peculiar characteristics and, in dozens of cases, very fragile. With a view to safeguarding the original material, which limits direct consultation, the first step to take is the digitization activity with the creation of digital copies, which facilitate fruition. For a conscious, coordinated, and valid digitization, a precise procedure followed, described below.

1. Stipulation of concessions and scientific partnership agreements with the conservation institutes involved having permission to reproduce online, with free access, in high definition and in compliance with the protocols of the International Image Interoperability Framework (IIIF) all the codices owned.
2. Setting up, in each conservation institute, of an acquisition laboratory, equipped with a planetary scanner and suitable equipment, followed by the calibration of the scanner itself: The Department of Humanities conducts the optical acquisition of Dante's manuscripts.
3. Compliance with international technical and quality standards in the digitization activity, for which there was an in-depth analysis of all the volumes to identify common characteristics that can be used to divide the works into processing batches according to the physical characteristics of the volumes and the degree of complexity of the processing. The creation of the processing batches minimizes the physical movements of the volumes, simplifies the monitoring processes and favors artificial intelligence systems, allowing the training of neural networks which will take place in the next steps. Subsequently, the formats and the relative resolution, the file nomenclature and the recovery methods that would preserve the volumes defined. As regards the shooting methods, the positioning, the opening, and the possibility of shooting on a single or double page evaluated for each volume, also thanks to the aid of V-shaped lecterns, which have a book opening angle limited to 90°. For each volume, a preview performed to delimit the scanning area, leaving a small margin to compensate for any micro-rotations. All operations conducted manually, wearing special cotton gloves: turning the pages of the volume required adjustments to ensure their correct positioning on

the lectern. The reproduction conducted using cold light lamps (5400 Kelvin), without ultraviolet components. Periodic checks allowed us to eliminate any processing errors.

What stated for the optical acquisition methods of Dante's manuscripts observed for the reproduction of the Torraca book collection, preserved at the *Accademia pontaniana*. In this case, an agreement stipulated between the Department of Humanities, the Department of Physics, and the *Accademia pontaniana*. The Department of Physics conducts the digitization activity. Thanks to MIMIT funding, it was possible to purchase a Metis EDS-Alpha scanner, equipped with a 24 megapixels Canon EOS R10 DSLR digital camera with a resolution of up to 400 ppi, an adjustable V-shaped lectern, automatic shape recognition functions, tilt and curvature correction tools (Fig. 3).

We want to point out that all digitization processes in the world use planetary scanners as ours, with a resolution which varies according to the price of the scanner itself. The original images are (almost) always in tiff format, but these images need reprocessing in order to change the format (jpg for the web access, etc.) and, more important, to calibrate the images on the basis of a color checker image and a metric scale images, usually on the last image of the book. The calibration will correct both chromatic aberrations and geometric aberrations, the most typical problems in image acquisition [9],[10]. This post-processing job is often under-evaluated in its importance.

Fig.3. Digitization activities at the *Accademia pontaniana*

The same funds have allowed the purchase of a second Metis EDS-Alpha scanner and a third OS 15000 Advanced Plus-Zeutschel scanner, equipped with automatic glass plate opening, automatic book positioning, CCD line sensor and 43200 pixel scanning mode.

8.2. Physical and biological diagnostic criteria

As anticipated, the Magic project envisaged conducting investigations on the physical and biological characteristics of book materials, to evaluate the degradation agents within the paper, oxidation, and the effects of humidity, temperature, and radiation. We used a portable spectrophotometer, shown on Fig. 4.

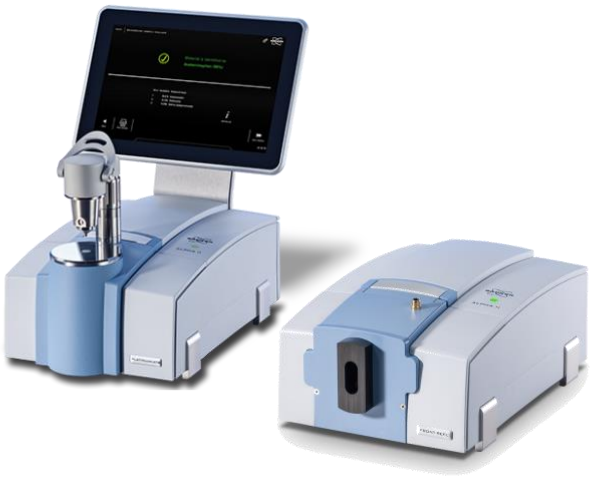

Fig.4. The Alpha-II spectrophotometer by Bruker used in the Magic project

The ATR-FTIR spectroscopic analysis conducted on four fragments of an incunabulum preserved at the *Accademia pontaniana* in Naples. The physical analysis consisted of selecting unprinted portions of the paper, preferably from the margins or non-visible areas, to avoid damaging the texts or illustrations.

Before starting the study on the incunabulum, a preliminary study conducted to simulate the natural aging of paper materials. Various artificial and accelerated aging techniques adopted to replicate the degradation processes of cellulose. The samples subjected to four different procedures:

- Humid aging,
- Elevated temperature aging,
- Chemical oxidation,
- Stain formation followed by thermal aging.

For each of these procedures, we used standard laboratory machines.

FT-IR spectroscopy was performed using the attenuated total reflectance (ATR) technique, which allows the direct analysis of the surface and to identify the main molecular vibrations of cellulose, the primary constituent of paper, and of any additives such as gelatin, starch or glues used during the manufacturing of the volumes. The graph produced by the spectrometer allows

us to get direct information on contaminants, through identification of the position of the peaks on the graph.

On the contrary, the characterization of microbial contaminants conducted with the contribution of professional skills in the biological and biotechnological fields of the Department of Molecular Medicine and Medical Biotechnology of the University of Naples. The advanced high-throughput third-generation Nanopore procedures have allowed the sequencing of long DNA molecules with high speed and accuracy.

The first sampling conducted using sterile nylon swabs, which have a rough surface capable of capturing the smallest particles just by touching them and, therefore, without damaging the touched surface. We thought they could be ideal for this type of sampling. The incunabulum had pages damaged by environmental and biological agents, it had traces of mold due to humidity, blackened areas, and small holes. After having viewed the book, five samples taken, different for the pages and the areas sampled, with a brushing movement on the pages.

Subsequently, the swabs placed in sterile bags and stored in the laboratory in the refrigerator at +4°C. DNA extraction performed using standard procedures.

To optimize the quantitative yield to obtain enough for an NGS sequencing analysis, another sampling method evaluated. For this reason, the second sampling performed with two fundamental modifications:

[1] both nylon swabs and nitrocellulose sheets used,

[2] the collected samples were immediately immersed in the lysis buffer.

All measurements performed with standard instruments of the microbiology laboratory.

8.3. Artificial intelligence algorithms for the removal of ink penetration in the paper

The Magic research project aims to evaluate new working techniques and highly innovative methodologies, sharing them with the world of research and industrial partners. Artificial intelligence is part of the project, as a tool to support the analysis of data obtained from the digitization work. We used GPU boards from nVidia, namely a L40S model, shown in Fig. 5. The board mounted on a top-level PC (24 cores, 128 Gbyte memory, big SSD disks) to avoid any bottleneck.

Fig.5. The GPU board used in the Magic project for bleed through removal

The Magic project is applying an artificial intelligence system for image processing to remove the effect known as bleed-through, without affecting the integrity and readability of the content. Bleed-through is a phenomenon whereby the ink of a page passes through the paper or parchment, penetrating the other side of the paper or parchment itself. This can happen due to the quality of the support, the ink used, or the humidity accumulated over time. Each page of the manuscripts comprises three levels of information: foreground text, background, and unwanted degradation.

Traditional restoration methods often adopt a recto-verso approach, which requires precise alignment and subsequent removal of ink from paired pages. However, this method is ineffective in real-world applications due to the common presence of severe document degradation and spatial distortions, complicating the alignment process and reducing the practical applicability of such techniques. Magic proposes a "blind" approach, which operates on single pages without requiring their recto-verso counterparts, thus providing a more versatile solution suitable for documents subject to significant wear or deformation. This method employs a complete pipeline, using various color spaces and techniques such as contrast enhancement and white edge inpainting to improve the segmentation accuracy of different regions of the document.

Mitigation occurs in two sequential phases, an identification phase and a removal phase, also called inpainting. The computational pipeline is composed of the following blocks (Fig. 6).

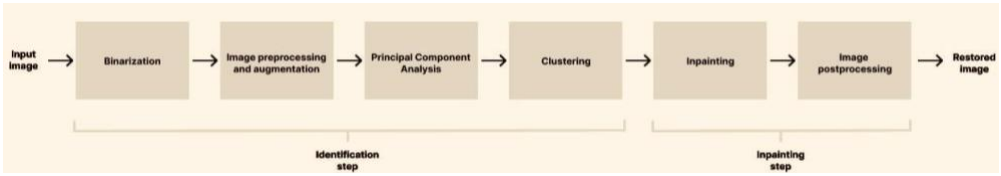

Fig.6. Bleed through removal pipeline

The purpose of the identification phase is to find the bleed-through component without having a ground truth: this requires the use of unsupervised machine learning algorithms. This phase consists of the following steps:

1. Binarization, where the input image is binarized through a neural network, to improve the identification of bleed-through pixels;
2. Image pre-processing and augmentation, where the original image normalized, pre-processed, and subsequently enriched by adding information from distinct color spaces, optimizing its information content for subsequent processing;
3. Principal component analysis, which aims to contract the size of the features extracted from the image without reducing their information content, to improve and speed up the subsequent clustering phase;
4. Clustering, combined with Gaussian mixture models, where the information content extracted through the previous steps processed through an algorithm to classify pixels as text, background and bleed-through.

In the inpainting phase, pixels labeled as bleed -through replaced with pixels that mimic the background tone. It consists of further steps:

1. Inpainting, in which pixels identified as bleed-through assigned a color that mimics the peculiarities of the text background, so parchment or paper;
2. Post-processing of the image, to further attenuate any graphic artifacts, making the colors uniform.

We have to put in evidence that there are several methods in the literature that cope with the removal of effects like the bleed-through; these methods are classified as "blind" and "non-blind", meaning that they can run unattended (automated) ([11]; [12]:5702-5712) or they need human intervention to align the *recto* and *verso* of each page ([13]:281-286). Our method lies in the first category and has the advantage of a solid mathematical base [14]. We choose the blind approach because we plan to apply it to about 150,000 images, already available to our project, and a non-blind approach is not affordable.

9. First search results

9.1. Digitization and cataloguing operation

The first phase of digitization (conducted by the Department of Humanities of the University Federico II) makes available all 283 illuminated manuscripts of Dante's Divine Comedy: the images acquired at a resolution of 400 optical dpi in tiff format. After scanning the books, a directory produced for each individual volume containing images in high-resolution tiff format. The tiff format is the most suitable solution for its extreme flexibility and the possibility of using lossless compression techniques to reduce files and make more effective use of data. The tiff images then converted into jpg files, more suitable for web consultation.

The digitization activity is not the focus, but the production of a large metadata base, both relating to the description of the manuscript object and to all existing studies on the subject, which make the digitized document the starting point for a complete, multilingual and multicultural journey. The digitized manuscripts accompanied by MAG management metadata in Administrative Management Metadata (MAG) format and METS standards in .xml format.

The descriptive cards consider the bibliographic object from a codicological and iconographic point of view, reaching a very detailed expository analysis that reaches the examination of the individual miniatures, which accompany and embellish the codices, going beyond the boundaries of the codicological, paleographic and historical-artistic descriptions provided by Manus OnLine (MOL).

The digitized images and the relative descriptions flow into the dedicated web portal, IDP-Illuminated Dante Project, managed by the Department of Humanities. The considerable digital archive and codicological and iconographic database aimed at specialized and non-specialized users: it currently allows the consultation of high-definition digital images in jpeg 2000 multi-resolution format of the first 101 manuscripts, through the Mirador viewer (https://www.dante.unina.it/) (Fig. 7).

The advantage for researchers, academic specialists, but also for a wider audience of enthusiastic readers, is to have the largest archive of Dante codices in free access. Furthermore, the images of all the manuscripts, for which copyright granted (thanks to an agreement between the University of Naples Federico II, the General Direction of the State Libraries of Italy and important international libraries - https://www.dante.unina.it/idp/public/pagine/progetto), are interoperable in the IIIF web community with a specific web app manifest.

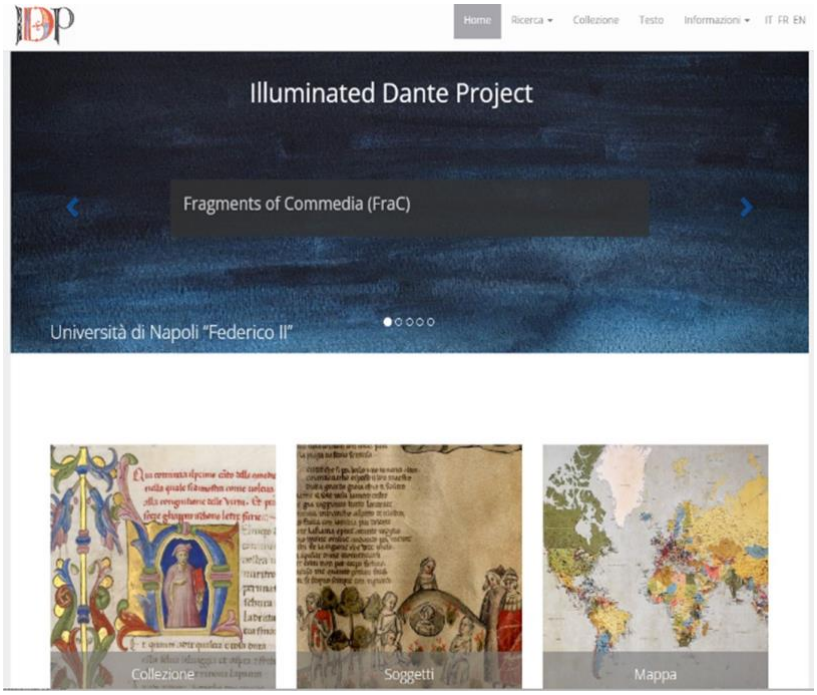

Fig.7. Home page of web portal IDP-Illuminated Dante Project (Copyright of the image: Department of Humanities, University Federico II, www.dante.unina.it)

For the second phase of digitization, which takes into consideration the Torraca collection of the *Accademia pontaniana*, the digitization of the six incunabula completed, while that of the 186 *cinquecentine* has just undertaken. For the digital acquisition of the 192 total volumes, the tiff format used for conservation with a high resolution of 400 dpi and the jpg format for consultation with a resolution of 300 dpi.

This initiative of valorization and diffusion makes usability possible and guarantees access to book heritage by a specialized and public, thus guaranteeing benefits for researchers, scholars, and general users.

The internal staff of the team has prepared a cataloging protocol for the incunabula and the *cinquecentine*, such as to consider the bibliographic object from a codicological, historical and content perspective: the description, present on the Magic site, in fact considers author, title, publication, dimensions, signature, imprint, text layout, lines, references, binding, language, state of conservation, decoration, location. Since researchers active in the literary or historical field very often dedicate themselves to the in-depth study of the volumes, the descriptive analysis moves towards the analytical, also considering the individual internal parts of the content of the volume, of which the title, the names of the significant characters, the incipit and the explicit are reported. From the detail tab it is possible to connect to resources like the one selected, thanks to the implementation of hyperlinks that activated by clicking on the keyword: the utility for the user is to encourage the discovery of other external contents. As the reproduction activity proceeds, the digitized images and the relative cataloging find a place on the Magic website.

Another advantage for users offered by the Magic website is the registration of users, who will also have the possibility, after proper evaluation, of downloading the complete volume in pdf format (Fig. 8) ([2]:66-69).

The display of the digitized books takes place through special software within the site, the Mirador viewer: it offers users the advantage of viewing the object in single or gallery view, the thumbnails of the volume; there are also enlargement and full screen functions.

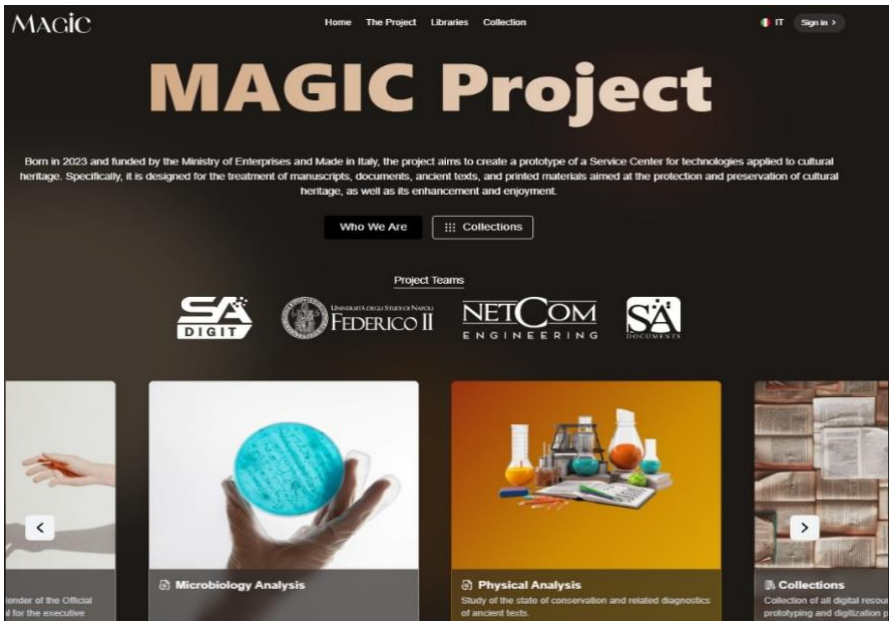

Fig.8. Home page of Magic website

9.2. Chemical and physical investigation

The objective of the investigation was to identify the chemical composition of the paper, highlight any signs of degradation, and verify the presence of foreign materials resulting from paper deterioration or environmental contamination. The graph produced by the spectrometer allows us to get direct information on contaminants, through identification of the position of the peaks on the graph. The main identified signals (Fig. 9) confirm the presence of cellulose as the primary component of the paper support (e.g., peaks at 3300, 1150 and 1640 cm^{-1}). The expert eye sees indications of paper degradation attributable to oxidation and hydrolysis processes: interaction with ambient moisture; formation of carbonyl compounds, typical of cellulose degradation; traces of isocyanates, nitriles, or carbonyl compounds resulting from material aging. Some signals may suggest the presence of other substances used in bookbinding or restoration interventions: proteins, suggesting the use of animal glue for paper treatment starches or adhesive substances used in bookbinding or paper consolidation.

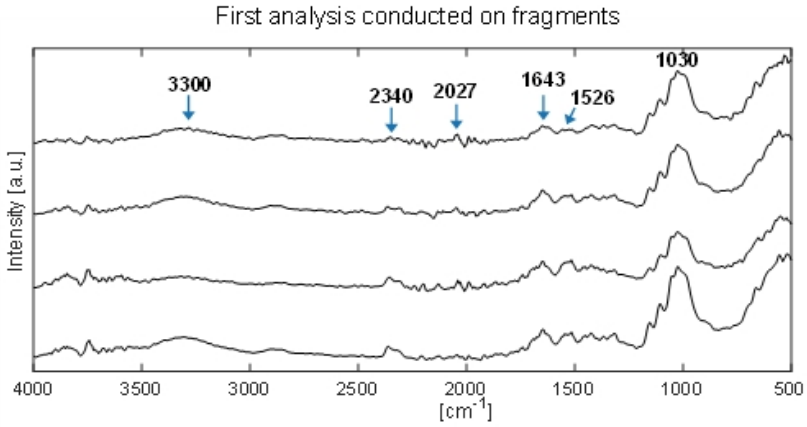

Fig.9. Result of the FT-IR analysis on an incunabulum dated 1478. In abscissa there are the "wave numbers" (a wavelength of twenty μm corresponds to a wave number of 500 cm^{-1}).

Each incunabulum subjected to FT-IR documented in a digital database containing the chemical-physical information detected, to monitor the conditions of the volumes over time, but also to adopt preventive measures.

Furthermore, the results obtained with the second sampling for the characterization of microbial contaminants, the one taken with nylon swabs, showed a significantly higher yield both compared to the samples obtained from nitrocellulose sheets and compared to the values of the first sampling.

One of Magic's goals is conservation: it is the set of direct and indirect actions aimed at slowing down the effects of degradation caused by time and use on the material components of cultural heritage. It is our intention to conduct a real-time monitoring activity of environmental conditions and the main atmospheric pollutants, inside the warehouses where the books stored. The measurements of environmental parameters that influence microclimatic conditions lead to the advantage of having a report on the state of degradation/conservation of the books, avoiding having to resort to restoration. Resorting to restoration is an admission of defeat or at least an inability to implement adequate protection activities.

9.3. Bleed-through effect removal results

The proposed method efficiently removes bleed-through degradation, leaving the foreground text intact and preserving the background texture of the document, without unpleasant visual effects. The performance of the method compared to the figure, where there are input image and restored image.

This developed technology produces dozens of benefits to general and non-expert users because, by removing or significantly reducing this widespread degradation, we allow the readability of handwritten texts and the understanding of their content.

The multiplicity of supports and inks used for manuscripts and books does not allow us to identify a universal method to remove bleed-through. For this reason, the Magic group is developing three other different algorithms, all effective in bleed-through, but with different results. Always with a view to interacting with the user, the latter will have the possibility to view

the results of the 3 methods, all of which use artificial intelligence systems, on the Magic website, evaluating the efficiency of the algorithm not only in theory but also and above all in practice: the results differ from each other with regard to the visualization of the writing support and the textual content, leaving the users themselves the task of identifying the best method to remove this effect. (Fig. 10).

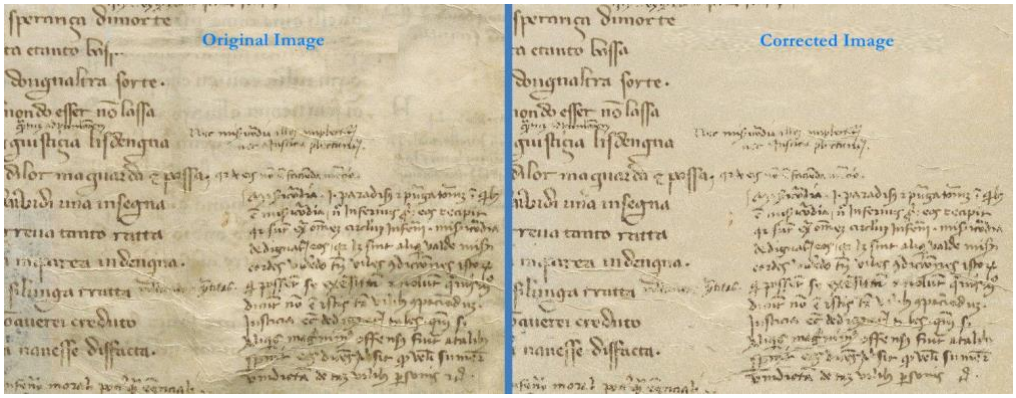

Fig. 10. Example of removing the bleed-through effect

10. Long-term storage of formats

Protection and enhancement, pursuant to art. 9 of the Italian Constitution, are terms that identify specific practices of conservation and relative enhancement of cultural heritage, as defined by the Code of cultural heritage and landscape. Magic has addressed these issues, engaging in the sector of long-term preservation of formats, overcoming the obsolescence of data and systems. The technology used by the Magic project involves the use of the F.I.T.S. (Flexible Image Transport System) format, a format created by NASA in the 1970s for the storage and exchange of images and data between scientists in the field of astronomy and space astrophysics. F.I.T.S. is an open standard that aims at the long-term archiving and relative transmission of images, in the name of the principle "once F.I.T.S., always F.I.T.S." This means that the data saved in this format will always be accessible and compatible with any evolutions of the standards. The structure of an F.I.T.S. file is very simple, but effective. A file is composed of two distinct parts, which can repeat hundreds of times in sequence:

[3] "Header", in ASCII characters, contains the description and information about the data, in binary form, i.e., the keywords (author, format, dimensions, history of the book object, observations, etc.);

[4] "Data Unit".

The file is self-documented because the metadata included within it (Fig. 11).

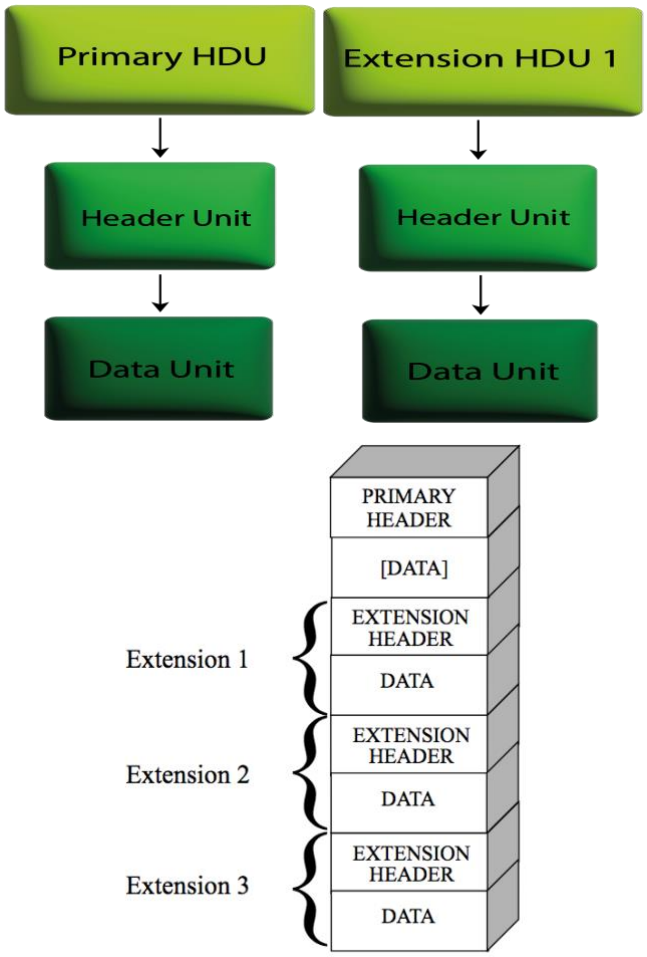

Fig.11. F.I.T.S. format structure

In 2022, the Italian National Unification Body recognized F.I.T.S. as the standard format for the digital preservation of ancient books and manuscripts (UNI 11845, Processes for the management of long-term preservation of digital images using the F.I.T.S. file format).

By creating a back-end interface, it is possible to transform tiff images into F.I.T.S. format, through a private server and with a method of minimal interaction for the operator who immediately:

[5] inserts the scanned images into tiff,

[6] transforms the tiff files into the F.I.T.S. format,

[7] inserts the keywords into the F.I.T.S. header,

[8] uploads the images into the storage system.

The storage system uses the IBiSCo Data Center, located in the Monte Sant'Angelo Complex of the University of Federico II, which can guarantee data security, image input speed even by non-IT personnel, adequate storage capacity, thanks to a dedicated server, with double CPU, 24 terabyte disk space and Alma Linux operating system, on which the web server is also installed. Other servers and disks are available for an increase in data, even if the current system already guarantees storage of over 1,000 complete volumes. As for the network connection, Magic uses the connection to Garr, the national research network, widespread throughout Italy and, in turn, connected to the general Internet network. The available bandwidth is twenty gigabits per second bidirectional, sufficient to guarantee access even by one hundred simultaneous users, without slowing down operations ([2]:197-205) ([18]:43-72).

11. Prospects

The Magic project is now halfway through its journey: it has brought together different experiences from the large multidisciplinary University and the regional production world. What has achieved so far is the embryo of a Service Center that, with the support of the institutions, will be able to develop and become a point of reference for multidisciplinary activities on manuscripts and ancient books.

During the design phase, dozens of standards, guidelines, encodings, protocols, and formats identified and applied to ensure description, metadata, long-term archiving, interoperability, which will evaluated in the future. Among them, the open source protocol of the International Image Interoperability Framework (IIIF), different levels of Application Program Interface (API), the Text Encoding Initiative (TEI) model for the XML file format of metadata, metadata encoding standards, with the aim of making access, consultation and interoperability effective, providing for the possibility of representing the semantic relations between the metadata created, through the Linked Open Data (LOD) technology and, in particular, through ontological languages (RDFS) and graph technologies (RDF).

The Magic project will soon follow the FAIR Data principles, aimed at ensuring that research data are Findable, Accessible, Interoperable and Re-usable. The benefits for users are that digital resources on the Magic site equipped with a specific web app manifest, the manifest Json, in compliance with the protocols of the International Image Interoperability Framework (IIIF). The goal is to make access, consultation, and interoperability of data as simple and effective as possible, through description formats interoperable with other national and international library resource aggregators.

There will also be some innovative services, driven by artificial intelligence, automated rules, natural language processing (NLP) and machine learning (ML): a chatbot software that simulates and processes human conversations, allowing users to interact with digital content. This is a very useful software for users, because it can answer their questions effectively and accurately, as well as provide information on the digital services offered, bridging the gap between user and cultural heritage, through a stimulating and innovative use, which can bring the young public closer to this type of heritage.

The application, within MAGIC, of new artificial intelligence techniques, for example for the removal of bleed-through, will open new scenarios also for the transcription of handwritten texts, also providing a glimpse of the future of reprocessing of the hundreds of existing digital collections.

For the Handwritten Text Recognition (HTR) of manuscripts, we will compare ourselves with the research underway at the Transkribus project, which presents about eighty-six public artificial intelligence models, available for training the software developed by the project. Given the progress of the Transkribus development, Magic will benefit from its results, verifying those that already fit the models for formal, typological, and historical characteristics and how many others instead require a specific artificial intelligence model. The automatic recognition of handwritten characters is an open research topic and subject to continuous progress in extraction methods and classification algorithms: the algorithms can supervise or unsupervised and the applications include pattern analysis (unsupervised learning) and classification (supervised learning). Tests will also perform with the new system eScriptorium.

The guiding thread of the Magic project follows the title and subtitle of the XIII Annual Conference AIUCD2024, because it is necessary to continue towards a reticular paradigm with computer science, which allows the design and development of a methodology that affects the way of doing, producing, disseminating humanistic research. The network, understood as a large laboratory of experiences, skills, competences, experiments, can become an inclusive, multidisciplinary digital destination, projected towards greater accessibility, so that users can be active protagonists in the design of resources, tools and paths.

12. Acknowledgements

The project funded by Ministry of Business and Made in Italy, (code n. F/130093/03/X38 and CUP: B69J23000560005) and by Ministry of University and Research (code n. PIR01_00011, CUP: I66C18000100006).

References

[1] Russo, Guido, Aiosa Luciano, Alfano Giancarlo, et al. 2020. "MA.G.I.C.: Manuscripts of Girolamini In Cloud". *IOP Conference Series, Materials Science and Engineering* 949 (012081): 1-8. https://doi:10.1088/1757-899X/949/1/012081.

[2] Conte, Stefania, Di Domenico Gian Marco, Mazzei Andrea et al. 2024. "The MAGIC project: first research results". In Proceedings of the 20th Conference on Information and Research science Connecting to Digital and Library science, Bressanone, Brixen, 22-23 February 2024, edited by Eleonora Bernasconi, Andrea Mannocci, Antonella Poggi, Angelo Salatino, and Gianmaria Silvello. CEUR Workshop Proceedings, vol.3643. https://doi:10.1088/1757-899X/949/1/012081.

[3] Conte, Stefania, Maddalena Pasqualino Maria, Mazzucchi Andrea et al. 2023. "The role of project MA.G.I.C. in the context of the European strategies for the digitization of the library and archival heritage". In *Eurographics Workshop on Graphics and Cultural Heritage*, edited by Alberto Bucciero, Bruno Fanini, Holger Graf, Sofia Pescarin and Selma Rizvic. The Eurographics Association. https://doi: 10.2312/gch.20231167.

[4] Marras, Cristina, Ciula Arianna, Eide Øyvind et al. 2024. "Towards a resemantisation of the concept of modelling in Digital Humanities". In Me.Te. Digitali. Mediterraneo in rete tra testi e contesti. Proceedings del XIII Convegno annuale AIUCD, Catania 28-30 maggio 2024, Università di Catania, edited by Antonio Di Silvestro and Daria Spampinato. AIUCD Associazione per l'Informatica Umanistica e la Cultura Digitale.

[5] Conte, Stefania, Mazzucchi Andrea, Russo Guido et al. 2024. "The organization and management of the MAGIC project for ancient manuscripts digitization: connections between Mediterranean cultures". In Me.Te. Digitali. Mediterraneo in rete tra testi e contesti. Proceedings del XIII Convegno annuale AIUCD, Catania 28-30 maggio 2024, Università di Catania, edited by Antonio Di Silvestro and Daria Spampinato. AIUCD Associazione per l'Informatica Umanistica e la Cultura Digitale.

[6] Nam, Taewoo and Pardo Theresa A 2011 "Conceptualizing smart city with dimensions of technology, people, and institutions". In Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times, College Park, MD, USA, 12–15 June 2011, edited by John Bertot and Karine Nahon, Association for Computing Machinery New York, NY, United States. https://dl.acm.org/doi/10.1145/2037556.2037602.

[7] Giffinger, Rudolf, Fertner Christian, Kramar Hans et al. 2007. Smart Cities. Ranking of European medium-sized cities. Centre of Regional Science (SRF), Vienna University of Technology, Austria. https://www.smart-cities.eu/download/smart_cities_final_report.pdf.

[8] Desfarges, Pascal. 2017. "La bibliothèque distribuée: Fabriquer ensemble un territoire des communs de la connaissance". In *Communs du Savoir et Bibliothèques*, edited by Lionel Dujol. Electre Editions du Cercle de la Librairie. Paris, France. https://www.retiss.com/la-bibliotheque-distribuee-fabriquer-ensemble-un-territoire-des-communs-de-la-connaissance/.

[9] Kugler, Elisabeth and Reynaud Emmanuel G. 2020. "LSFM series – Part III: Image acquisition: Calibration and acquisition". https://focalplane.biologists.com/2020/12/17/lsfm-series-part-iii-image-acquisition-calibration-and-acquisition/

[10] Baltsavias, Emmanuel P. 1994. "Test and calibration procedures for image scanners". In *Proceedings of the ISPRS Commission I Symposium, Como, Italy, September 12-16, 1994*. Institute of Geodesy and Photogrammetry, Swiss Federal Institute of Technology. https://doi.org/10.3929/ethz-a-004334530

[11] Hu, Xiangyu, Lin Hui, Li Shutao and Sun Bin. 2016. "Global and local features based classification for bleed-through removal." *Sens Imaging* 17 (9). https://doi.org/10.1007/s11220-016-0134-7.

[12] Sun, Bin, Li Shutao, Zhang Xiao-Ping Sun and Sun Jun. 2016. "Blind bleed-through removal for scanned historical document image with conditional random

fields." *IEEE Trans. Image Process* 25 (12): 5702–5712. https://doi: 10.1109/TIP.2016.2614133.

[13] Hanif, Muhammad, Tonazzini Anna, Savino Pasquale, Salerno Emanuele and Tsagkatakis Gregory. 2018. "Document Bleed-Through Removal Using Sparse Image Inpainting". In Proceedings of the 2018 13th IAPR International Workshop on Document Analysis Systems (DAS), Vienna, Austria, 24–27 April 2018. IEEE: Piscataway, NJ, USA. https://doi: 10.1109/DAS.2018.21.

[14] Ettari, Adriano, Massimo Brescia, Stefania Conte, Yahya Momtaz, and Guido Russo. 2025. "Minimizing Bleed-Through Effect in Medieval Manuscripts with Machine Learning and Robust Statistics" Journal of Imaging 11, n. 5: 136: 1-28. https://doi.org/10.3390/jimaging11050136.

[15] Olmedo-Pagés, Elena Loreto e Arquero-Avilés Rosario. 2024. "Costruire ponti verso una cultura accessibile a tutti: l'integrazione di Easy-to-Read nelle mostre virtuali" AIB Studi 64 (1):73-83. https://doi.org/10.2426/aibstudi-14069.

[16] Conte, Stefania, Ferrante Gennaro, Laccetti Lorenza et al. 2024. "Content representation and analysis: the Magic Project and the Illuminated Dante Project integrated systems for multimedia information retrieval". In *Proceedings of the 14th Italian Information Retrieval Workshop Udine, Italy, September 5-6, 2024*, edited by Kevin Roitero, Marco Viviani, Eddy Maddalena and Stefano Mizzaro. CEUR Workshop Proceedings.

[17] Conte, Stefania, Russo Guido, Salvatore Marcella et al. 2024. "Application of the IBiSCo Data Center for cultural heritage projects". In *Proceedings of the Final Workshop for the Italian PON IBiSCo Project, Napoli, 18-19 aprile 2024*, edited by Giovanni Cantele, Gianpaolo Carlino, Luisa Carracciuolo, Alessandra Doria, Giorgio Pietro Maggi and Guido Russo. Zenodo. https://zenodo.org/records/13120590.

[18] Allegrezza, Stefano. 2011. "Analisi del formato FITS per la conservazione a lungo termine dei manoscritti. Il caso significativo del progetto della Biblioteca Apostolica Vaticana" DigItalia 6 (2): 43-72. https://digitalia.cultura.gov.it/article/view/476.