MAGIC: Manuscripts of Girolamini in Cloud

Guido Russo (University of Naples) for the MAGIC Collaboration

14-16 October, Florence

The MAGIC project

In 2019, a collaboration between private companies and a research institutions, has submitted a project in the filed of digital libraries.

The project proposal has as the ultimate goal the creation of a Service Center for ancient books and manuscripts, considering as a first step an entire section in the Library of Girolamini. Main topics are:

➢ long-time preservation of the contents of the ancient books through the digitization

➢ An adoption of an open file format

➢ Provide smart access to the digitalized contents

➢ Strong categorization of the contents

➢ Provide a fast way to physically locate the ancient document using the Internet of Things paradigm

The MAGIC collaboration

UNIVERSITÀ DEGLI STUDI DI
NAPOLI FEDERICO II

SA DOCUMENT S.r.l.

SA Lombardia S.r.l.

NETCOM GROUP S.p.A.

The Library of Girolamini

The Library of Girolamini is a rare episode in the history of books and culture in Italy and in the world. Carried out over the centuries through the subsequent composition of private funds

➢ The Library has a collection of books of about 159,700 units

➢ 137 musical prints and 120 incunabula

➢ 5,000 editions of the sixteenth century

➢ 10,000 rare and valuable editions

➢ large quantity of microfilms and portraits

The process of digitization and archiving

The project team has identified a number of tasks and subtasks, which are schematically represented below, and which will be followed for any single manuscript of the Girolamini's collection.
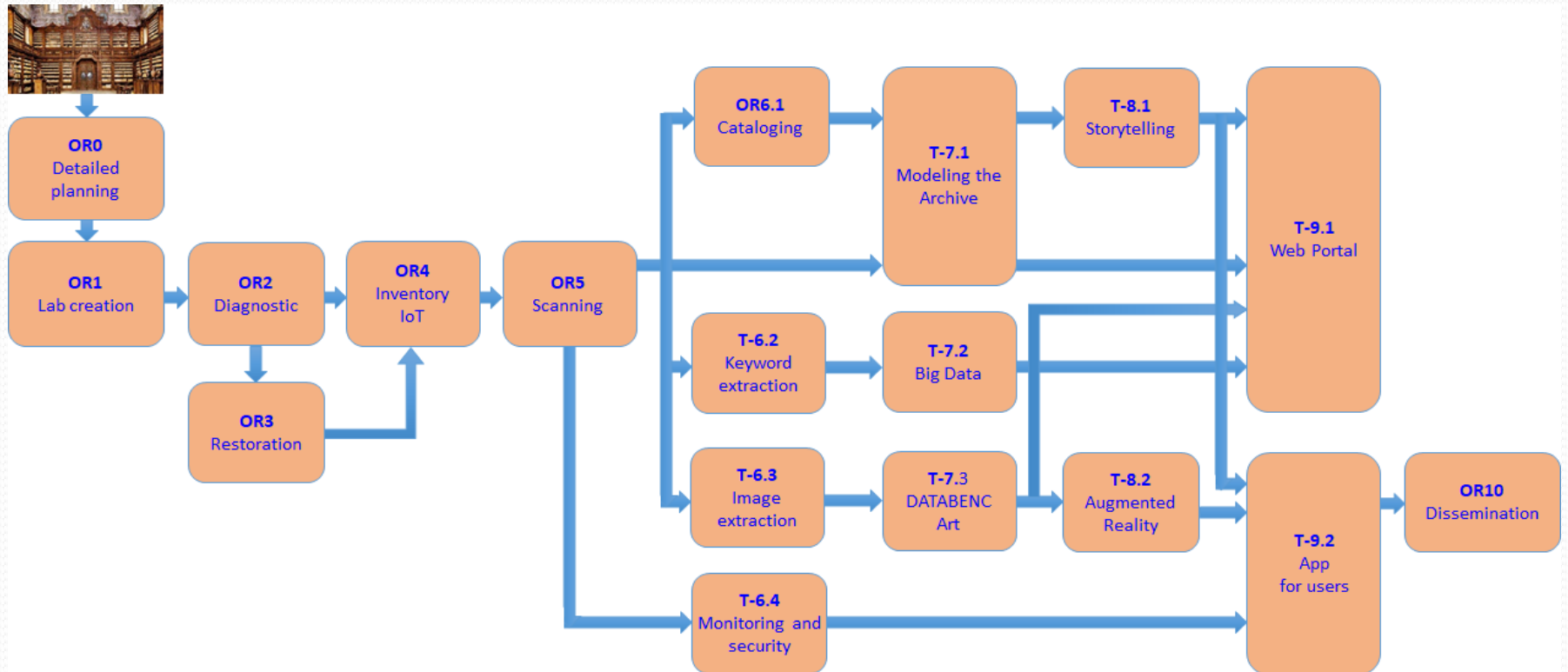
The process of digitization and archiving

➢ *Detailed planning:* arrange how to handle of the documents and obtain all the authorizations to move the documents temporarily.

➢ *Lab creation*: the construction of the lab, including the clean room.

➢ *Diagnostic:* a diagnosis of each document in order to validate the feasibility of the scanning process.

➢ *Restoration*: if the damage is not too critical will be performed a restoration.

➢ *Inventory-IoT*: the cataloguing will include a refined metadata creation and the creation of RFID tag.

➢ *Scanning*: the most time-consuming task, which will require special scanners.

OR0 Detailed planning

OR1 Lab creation

OR2 Diagnostic

OR3 Restoration

OR4 Inventory IoT
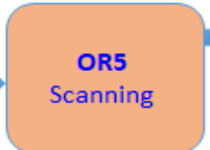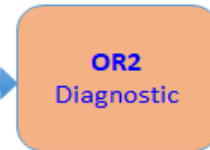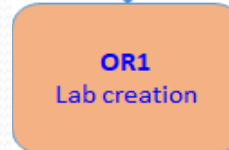
OR5 Scanning

The process of digitization and archiving

➢ *Cataloging, keyword and image extraction, monitoring*: in parallel with scanning, people of the University will define a complete keyword set for each document to enhance the metadata database.

➢ *Modeling and Big Data*: the huge dataset which will come out from the digitization process will require an accurate modeling but also a Big Data approach

➢ *Storytelling, augmented reality*: before opening the archive to the public, a storytelling process will define the information to put online

➢ *Web Portal, App*: under this task we will create a modern web portal and smartphone Apps

➢ *Dissemination:* the whole project, and the availability of data, will be supported by a large information campaign
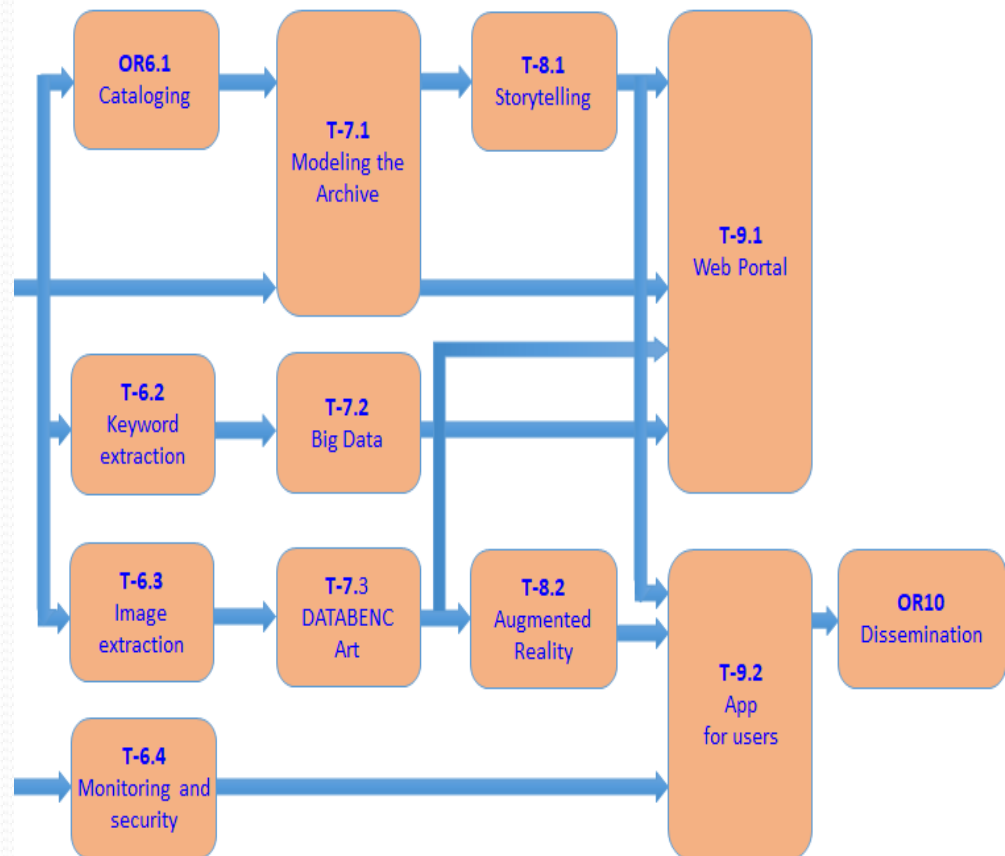
Technical aspects
## The choice of the FITS file format:

Flexible Image Transport System (FITS) was initially developed by astronomers in the USA and Europe in the late 1970s to serve the interchange of data between observatories.

FITS format has a strong the metadata representation

FITS files can be used to ensures long term preservation including backwards compatibility "once FITS, forever FITS"

FITS files can contain many other information about the manuscript (dimensions, material ...); it is also free from legal restrictions.

The whole Vatican Library is being archived in this FITS format.
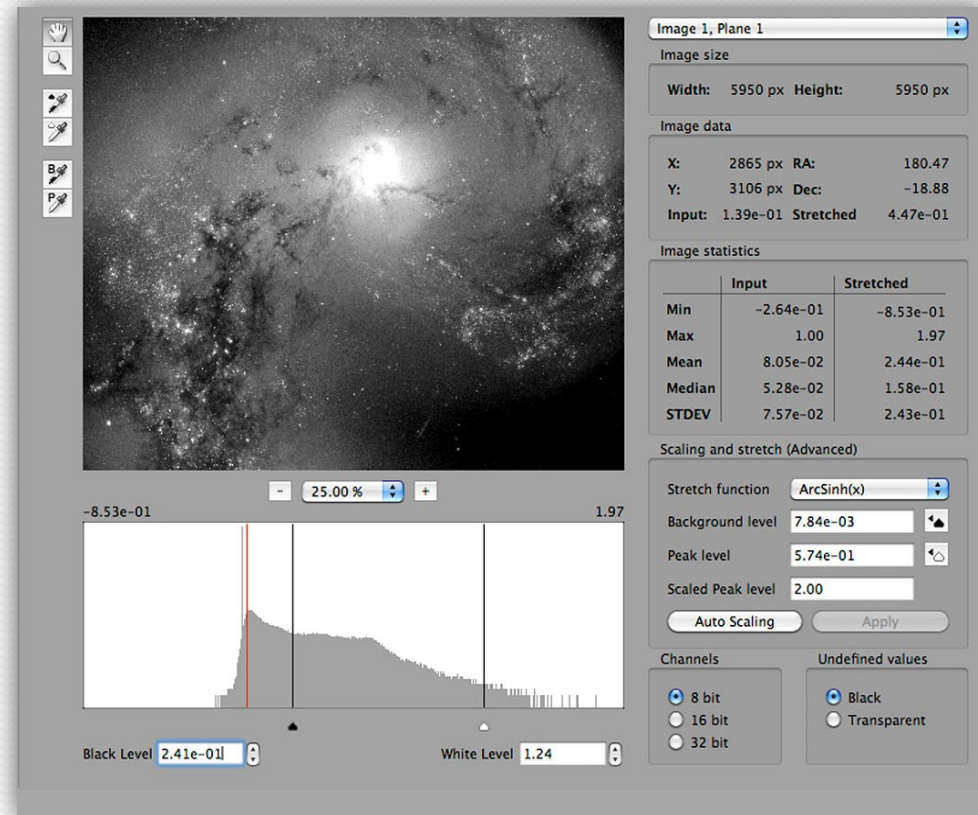
Technical aspects
The physical infrastructure

The physical infrastructure for the MAGIC project is based on two different sites:

The *first site* will be within the Library of Girolamini.

The *second site* is located in the university complex of Monte Sant'Angelo, owned by University of Naples Federico II.

Technical aspects
Library of Girolamini site

The reason for this choice lies in the need to minimize the ancient books physical stress and for safety reasons. The manuscripts have to be scanned in a room with controlled humidity and temperature.

The technical choice is to install a clean room with humidity and temperature controlled, on the ground floor of the library room.

The books will be entered into the room a few at a time until the complete digitization, which will use the so-called planetary scanners able to scan also the fragile bound documents like the ancient books.

Technical aspects
Monte Sant'Angelo site

In this large academic site the University has built, with financial support form the MIUR, a Data Centre for scientific applications.

The Data Centre has 33 racks, water refrigerated, each of them capable of up to 20 servers fully equipped. The MAGIC ptoject will initially use 10 servers and 1 storage system with 1 Pbyte capacity.

The whole Data Centre is connected to the Internet through a 2 Gbit/sec optical link, fully symmetric, and the bandwidth will soon be upgraded to 10 Gbit/sec; this fast connect will allow access to the private cloud of the Data Centre to any user around the world.

Technology developments

Beside the primary activity MAGIC project proposal is characterized, from the point of view of technological developments, on three pillars that are increasingly characterizing the world of work:

Big Data

Internet of Things

Artificial Intelligence

Technology developments
Big Data.

Big Data can be defined as a combination of structured, semi structured and unstructured data collected in order to be mined for information. Big Data is often characterized by the so called 3Vs: the large *Volume* of data, the broad *Variety* of data and the *Velocity* at which the data is collected and processed.


Within MAGIC, the Big Data approach will allow a data mining process through all the metadata created, and afterwards through all the images themselves, in order to extract new information. MAGIC will deep drive into the data to extract the key knowledge/Pattern/Information and to create new information from this process, and eventually to make these new information an open access database.

Technology developments
Internet of Things.

Internet of Things (IoT) is defined as an ecosystem of connected physical objects that are accessible through the internet. The "thing" in IoT could be any device that has the ability to record and transfer data over a network.
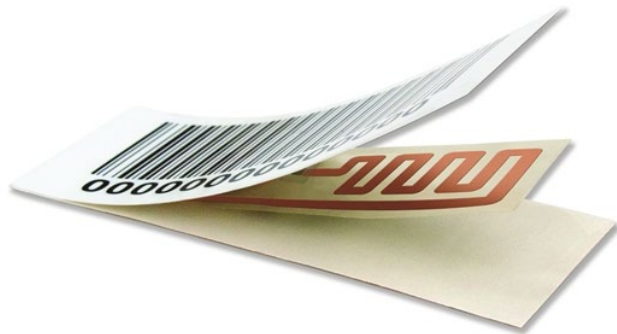
In MAGIC project will be experiment a very efficient microchips affixed to the original document. This small microchips are named RFID (Radio-Frequency IDentification) are one of the more widespread information repository of the IoT. This "Tags" allow to activate a much more complete and advanced path for information traceability.

Technology developments
Artificial Intelligence

At present time, the most successful realization of artificial intelligence are the Machine Learning and Deep Learning techniques.

Few examples are: Google search, Image recognition software and Self-driving cars.

input layer   hidden layer 1   hidden layer 2   output layer
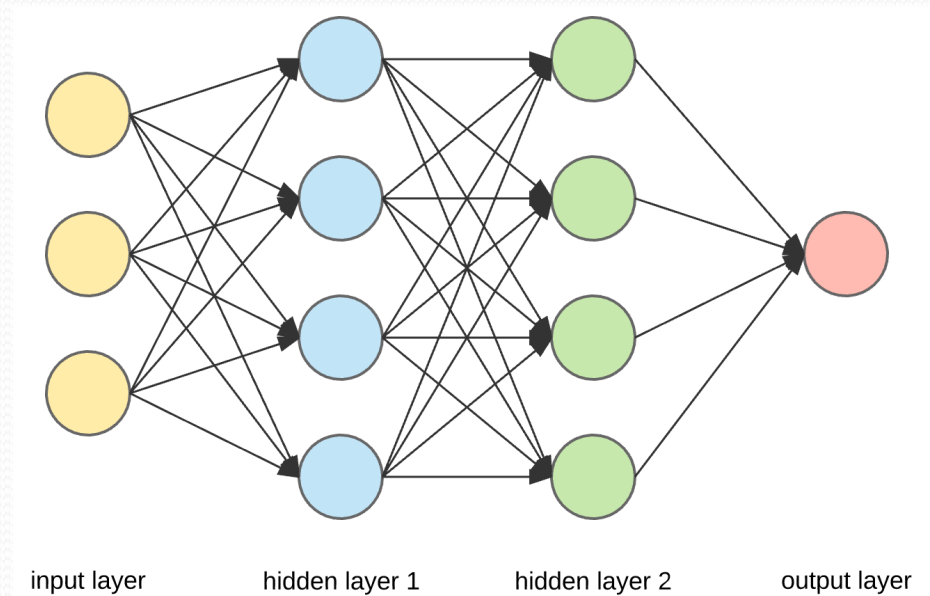
Within MAGIC project, AI will be used for the image recognition task, which means that within a large set of images, e.g. all the pages of a manuscript or book, AI can reveal if different hands have written it, or can recognize objects within the illustrations of the manuscript. If the analysis is extended to a set of books or manuscripts, AI can help in finding the similarities, if any, in the technique used to write the documents or can help in dating the document itself.

Conclusions

The MAGIC project is now just at the beginning, it will allow in the near future the digitization and fully availability to a large community of the many manuscripts and ancient books.

The adoption of many advanced technology, such as IoT, cloud paradigm and machine learning classification, will produce an integrated platform which will ease the work of scientists who want to make a comparative study of ancient documents.

Thank you for your attention