

The Vatican Apostolic Library's Digital Preservation Project

Paola Manoni

IT Service Management Coordination, Vatican Apostolic Library, Vatican City, Vatican State.

E-mail address: manoni@vatlib.it

Ángela Núñez Gaitán

Preservation and Conservation Workshop, Vatican Apostolic Library, Vatican City, Vatican State.

E-mail address: nunezgaitan@vatlib.it

Irmgard Schuler

Photographic Laboratory, Vatican Apostolic Library, Vatican City, Vatican State.

E-mail address: schuler@vatlib.it



Copyright © 2018 by Paola MANONI-Ángela NÚÑEZ GAITÁN-Irmgard SCHULER. This work is made available under the terms of the Creative Commons Attribution 4.0 International License: <http://creativecommons.org/licenses/by/4.0>

Abstract:

Preservation, conservation and free consultation are the main principles of the Vatican Apostolic Library from its inception, in the middle of the XV century. Precisely in compliance with this role, a project of long-term digital preservation began in 2010. The challenges encountered were twofold: on the one hand, the guarantee of the longevity of the technological product obtained and on the other, enabling for the dissemination of the Vatican Library's manuscripts in a networking digital library. The continuous cooperation between the Photographic, Preservation and Conservation and Information Technology Sections of the Vatican Library was a prerequisite and has been fundamental for maintaining the high quality of the whole process.

This paper will focus on three main areas:

1. The workflow begins in the Preservation and Conservation Workshop. During these years of experience, the conservators have carefully defined the discernment/standard needed for the first selection of the volumes to be digitized. If there is any risk of damage of a manuscript during the images acquisition process, the conservators declare the volume "not digitizable". In other instances, the conservators perform conservation treatments to enable safe photographic procedures. Together with the photographers, conservators have corrected and perfected the photo shooting process and they have defined the handling instructions to the digitizers and their preservation awareness.

2. *The planning, execution and evaluation of the large-scale digitizing project – that regards at least 82.000 volumes corresponding to 45 petabyte of data – is described.*

The project was preceded by a long period of preliminary study, in which guidelines based on international standards and best practice rules, but formulated according to the specific needs of the Institution, have been drafted, followed by a test bed where the proper design of infrastructure and the hardware and software resources needed were tested. Acquiring devices and shooting strategies, the digitizing workflow, color management of images, and quality control incorporated into the workflow process will be illustrated.

The peculiarity of the project lies in the fact that it has been conceived, designed and controlled thanks to the internal expertise of the Institution. As a result, a model for a high quality digitizing project has been elaborated, contemplating also strategies that ensure access to digital content over time.

3. *This session showcases the practice of using IT storage systems for digital preservation as well as the long-term preservation archive of FITS files at the Vatican Apostolic Library. The implementation in use is able to ensure continued access to digital information: it performs the ingestion of files, the conversion to derivative formats and the use of the PREMIS structure for supporting and recording any digital preservation action.*

All the information stored in PREMIS files is indexed and searchable. The database architecture in itself has been deployed according to the semantic units of the standard: the interrelated entities (Object, Event, Agent, and Rights) describing the preservation for each digital file ingested. The showcase will also describe how the information in PREMIS structure is derived.

Keywords: Vatican Apostolic Library, Conservation, Digital workflow, PREMIS, FITS

1st Session - Conservation towards large scale digitization

Since its inception, in the middle of the fifteenth century, the Vatican Apostolic Library has had as its primary task that of preserving and conserving its own collection, but also making it available to scholars in a free and open way¹. In the outset of the twenty-first century, these key concepts have easily led to the evaluation of a new technology that considerably widens the possibilities of diffusion of the Vatican's heritage: the digitization and consequent online publication of images.

The Vatican Library has begun to digitize its manuscript collections in order to preserve, but also to provide access to a wider range of users by disseminating the complete, high-fidelity digitized items online in a digital portal, building up digitization strategies by best practice methods to create the digital collection. The overall project is ambitious and intended to digitize approximately 40 million pages corresponding to 82.000 shelf marks, totaling 45 Petabytes of data in the highest possible quality.

¹ In the document *Lamdiu decrevimus* of April 30, 1451, the new library of the popes had as its purpose «pro communi doctorum virorum commodo» [«the common convenience of the learned»]: Pope Nicholas V made the papal collection also accessible to external scholars, which, until that moment, was used only by the papal court, increasing it also with his personal collection. In the bull *Ad decorem militantis Ecclesiae* of June 15, 1475, Sixtus IV reiterates the purpose of the library already expressed by the founder «eruditorum ... virorum commodum et honorem» [«the usefulness and honor of learned men»]. See Piazzoni, *Introduzione alla storia della Biblioteca Apostolica Vaticana*, *op. cit.*, pp. 15-32: 15-16.

The choice of the sequence of digitization of the manuscripts follows the criteria listed below:

- the relevance of the manuscript;
- the less known and studied collections;
- the most studied collections;
- the books with conservation's problems.

The peculiarity of the project lies in the fact that this large-scale digitization project has been conceived, designed and controlled thanks to the expertise of employees from within the Institution, from the beginning until the end of the process. The continuous cooperation between the Photographic, Preservation and Conservation and Information Technology Sections of the Vatican Library has been an essential element, almost a prerequisite to maintain the high quality of the results and to reach the goals previously set.

Although for any ancient volume, during the process of handling and digitization, some precautions that preserve its integrity must be taken, the needed attention and care increase when the object to be digitized is a manuscript. The particular physical characteristics of the materials used, as well as the intrinsic uniqueness of the manuscript are the main focus. The need for a substantial and indispensable involvement of conservators within the project is therefore logical and obvious.

The natural inclination of the conservators to protect the book and its integrity has sometimes clashed with the also genuine inclination of fellow photographers aiming to obtain an image of the highest quality, as faithful as possible to the original. This gap has been overcome with a strong, long-standing and very respectful mutual cooperation. To obtain the best possible image at the first and hopefully only shooting means also not having to repeat - at least for many years - the photographic campaign and the consequent extraordinary stress that this process would entail for the volume. The digitization contributes also to reduce the number of consultations of the original and, therefore, to its preservation.

The vastness and variety of the collection of manuscripts held in the Vatican Library have influenced the choice regarding the technology for large-scale digitization. Scanners were chosen as main acquisition devices, but also cameras with book cradles were added. The standard scanners also provide for a 180 ° opening of the volumes during the acquiring process, which might be a contraindication, from a conservative point of view. This issue has influenced the work of the conservators, for example in the selection criteria of the volumes and in the handling instructions. Therefore, as conservators, we felt the need to add personalized cradles that would allow opening the volumes up to 130 °. The volumes, whose opening for mechanical or conservative reasons is less than 130 °, have instead been excluded from the digitization project, at this time; this is a non-definitive exclusion, and we are confident that during the long duration of this project² we will have the opportunity to acquire innovative equipment that allow shooting at lower degrees of openness while maintaining the same high quality standard of the images.

The digitization workflow in the *Vaticana* starts right in the conservation studio. Each and every book passes through the studio, where the conditions are checked and any conservation treatments are carried out exclusively for the purpose of guaranteeing a safety

² With the current workflow, about 2.000 volumes are digitized each year.

digitization. The whole process is recorded in the database used for the conservation reports³. Once the shooting is over, each book goes through a second check up in the conservation studio.

The post-digitization check and the evaluation of any possible damage suffered by the volumes made it possible to understand what were the critical issues of the scanners from the conservative point of view, those that were not foreseen. During these five years of experience, the conservators have carefully defined the discernment standards needed for the first selection of the volumes to be digitized. Together with the photographers, conservators have corrected and perfected the photo shooting process and they have clearly defined the handling instructions for the digitizers and their preservation awareness.

Conservators determine the degree of opening angle of each volume, the scanning with or without the contact of the scanner glass, and if the acquiring process requires the assistance of a conservator. The conservators also have the power to declare “non-digitizable” a volume when its integrity would be clearly compromised during the capturing process.

All the digitization operators have been trained by photographers and also by conservators for everything related to the safe handling of the volume and to increase the sensitivity towards the manuscript.

The interface with the digitizers takes place through the fundamental web-based management software *Inside*, developed by the Library's IT colleagues for the management of the entire workflow and which has met the needs of photographers and conservators. In this information system, the conservators indicate the characteristics of each volume especially those conditioning their reproduction, they include the description of damage already done to the volume, and they give precise instructions for an *ad hoc* handling for each book.

2nd Session - The digital work flow

In order to be able to establish a reproduction project of the copious and different material which should have long-term-preservation connotations, a preliminary feasibility study became necessary. This survey was undertaken by the Library's internal expertise during the first decade of the new millennium, exactly coinciding with the maturity and reliability that capturing and archiving technologies had finally reached at that moment.

³The computerized form was created by conservators and IT staff of the Vatican Library in 2009. It is subdivided in five parts: the description according to a codicological/bibliographic approach, the conservation status, the preservation projection, the conservation treatments (great importance is given to the description of the materials and products used) and the testing procedures. The forms are accompanied by digital photos. All the completed forms are organized in a database that will be available in the manuscript data archiving and consulting system, that is already working in the Vatican Library, called InForMA (*Informatics for Manuscripts and Archives*). The InForMA is conceived in an open source Java/XML based program that can be easily managed by a WEB interface. It is a well-constructed database for the description and the research of manuscripts and archival items, consistent with the international standards (TEI-MS for the manuscripts and EAD for the archive items). This search engine can find each and every information present in a single conservation form, facilitating the opportunities of historical and statistical research about conservation. Cf. Grimaccia, *The Vatican Library conservation workshop, op. cit.*

The main aspect to be analyzed was, first of all, acquiring technology: which acquisition devices should generate the master files that had to be virtual copies of the precious originals? During a 6-month-test-phase, two acquiring positions - one equipped with a *Hasselblad* single shot 50 MP digital back and flashes⁴, and the other with a *Metis* planetary scanner with an acquisition format of 50x70 cm at a resolution of 400 ppi - were installed.⁵ This test bed gave way to experiment the acquisition technology in response to the differentiated originals of the Library,⁶ to elaborate the proper design of infrastructure and the hardware and software resources needed, to draft an efficient but flexible work flow, and to make estimates of timing and productivity. The last aspect is strongly connected to the financial planning of the whole activity over time and required special attention.⁷ Logistics regarding the transport of the volumes, as the project takes place in an external building which is close to the historic library, were studied.

Guidelines based on international standards, but formulated according to the specific needs of the Institution, were drafted.⁸ The state of the art is capturing at highest possible quality, which means scanning and digital capture with high quality lenses and without compression in acquisition and in storing. Native resolution; high-bit-mode for elaborations; reproduction scale accuracy; absence of external elements, such as sticks to flatten the sheet, on the image; no stitching of master files, are some of the best practice rules adopted. Qualified and well coordinated operators guarantee the implementation of these quality objectives and are therefore an important aspect in planning.⁹

In the context of digital preservation of documentary heritage, color is an important source of information that must be reproduced and displayed as faithfully as possible to the original. The *Metis* scanners are provided with a color profiler software implemented into the

⁴ Flashes were chosen because they supply a little more accuracy than continuous lighting, are more powerful and can be better regulated. Furthermore, faster shutter speed is more suitable for the protection of the originals, avoiding overheating of objects and environments. See also the recommendations formulated by the European Photography Workshop at Girona, *op. cit.* Risks for cultural heritage objects arising from the use of flashes are not concrete and are rather a legacy of the past, when explosion of the bulbs was frequent: cf. Modolo, *Il dibattito sulla liberalizzazione della fotografia digitale, op. cit.*, p. 11.

⁵ The *Metis* scanners are advanced professional planetary reproduction devices for the cultural heritage for high-quality digital reproduction that perform very high resolution (from 400 up to 800 spatial ppi), generating images of accurate color fidelity and wide geometric accuracy. Information about acquiring devices, software and special features distributed by the *Metis Systems* company are available at <<http://www.https.metis-group.com/>>

⁶ No easy task: The manuscripts have medieval and humanistic, but also late-antique and modern origin. They are written in numerous alphabets and many languages, and their contents cover practically all the spheres of human knowledge. Regarding their materiality, bound books in various formats and partially illuminated are the majority, but the collections preserve also geographical maps and nautical charts, long parchment or leather scrolls, pre-Columbian books folded like an accordion, materials written on palm leaves, fragile papyri fragments, Japanese paper scrolls, and many more different materials. The manuscript collection is therefore a treasure and at the same time a constant challenge.

⁷ This aspect, at a certain point, led to the creation of the new asset of fundraising.

⁸ Almost all procedures of the established photographic process correspond to the *four star quality level of imaging* of the FADGI-Guidelines. Also the *DFG-Praxisregeln "Digitalisierung"* of Deutsche Forschungsgemeinschaft, and the IFLA Guidelines *Linee guida per pianificare la digitalizzazione di collezioni di libri rari e manoscritti*, have been used.

⁹ The technical staff works in shifts of five hours and is supplied by service companies; tutors and supervisors are members of the internal Library staff. For outsourcing, DFG-Guidelines foresee that the required performance parameters of production must be exactly specified (*DFG-Praxisregeln "Digitalisierung"*, *op. cit.*, p. 41). We realized that dealing with manuscripts in general, and especially with such different materials as those of the Vatican Library, it was not realistic to pretend to follow this rule exactly.

scanning software.¹⁰ It is able to create an ICC scanner profile by using different types of color checker targets. The optical ICC profile can then be converted to different working color spaces. We use the standard wide-gamut AdobeRGB color space¹¹ for visualization, while sRGB¹² is the color space provided to the JPEG2000 images distributed through the Library's digital portal. In front of this complex and not completely standardized background,¹³ being aware that the color space triangle is, in any case, always smaller than the spectrum which is visible to the human eye, and that color management cannot compensate poor imaging, we follow best practice by capturing, for every volume, reference images with color charts (*x-rite* ColorChecker) that permit the reconstruction of the original color ratios.

After the preliminary study, the project started in 2010 with an initial collaboration with the University Library of Heidelberg.¹⁴ Turning plans into actions, it took about one year until the desired output (about one codex per day per acquiring device) was reached. The project continued and grew, and currently it is carried out through different sub-projects that are for the most part conducted in an international context of collaboration with academic institutions, foundations or technical partners.¹⁵ Within these sub-projects, groups of manuscripts, parts of collections or entire funds which are chosen together with the partners, are reproduced.¹⁶ Presently the capturing system consists in six different models of *Metis* scanners, three *Metis* scanners equipped with book cradles, and two digital cameras mounted over book cradles.¹⁷

¹⁰ The scanners comply with the requirements of *Metamorfoze* and *FADGI* quality digitization guidelines; cf. <https://www.metis-group.com/color-accuracy>. For further information regarding the technical specification to analyze image quality according to these two guidelines, see *Standardization of Image Quality Analysis – ISO 19264*, *op. cit.*

¹¹ Recommended because it is a widespread standard working color space with wide color range, particularly suitable for visualization on professional screens.

¹² The range of color included in this smaller color space is viewable on most current color monitors and for this reason it is appropriate for large distribution contexts, as well as in printing environments.

¹³ Cf. *FADGI-Guidelines*, *op. cit.*, p. 65: «It may be necessary to deviate from the normal, simple color managed workflow to achieve the best results. There are many options possible to achieve the desired results [...]». See as well Falcinelli, *Cromorama*, *op. cit.*, pp. 437-438.

¹⁴ The entire collection of the *Palatini latini* codizes has been digitized and provided with structural metadata in this common project.

¹⁵ Further cooperations involve, for example, the Polonsky Foundation (thanks to which it was possible, in a joint venture with the Bodleian Libraries, to reproduce 1.500.000 images of incunabula, and Greek, Hebrew and Latin manuscripts); with *NTT Data* (3.000 volumes of various collections); with the Republic of Azerbaijan (conservation and reproduction of codizes regarding the history of Islamic civilization in Central Asia); with the Brigham Young University (digitization and implementation of metadata for Syriac manuscripts); with the Alamire Foundation of the Catholic University of Leuven (about 50 Renaissance musical manuscripts); with the Chinese National Committee for the compilation of the history of the Qing Dynasty (ca. 1.200 book units); with institutions related to the Japanese National Institute for Humanities (papers concerning the persecution of Christians in Japan from 16th to 19th century); and numerous other projects.

¹⁶ When selecting the volumes or collections to be digitized, priority is reserved, as already said, to manuscripts that are fragile or run into danger of deterioration, and to the particularly precious ones. Productivity varies, depending on the volume's materiality and conservation state. An added value to the project is constituted by special imaging: texts which are damaged, difficult to read or inaccessible, are captured with particular techniques and technologies (ultraviolet fluorescence, infrared reflectography, or with optical prisms to reach the writings in the gutter); palimpsests are reproduced in a specific project.

¹⁷ The need to obtain images of excellent quality while respecting the integrity of the originals induced to provide book cradles (opening angle 130° – 180°) to the scanners, which were designed in collaboration with the Library staff, or stand-alone cradles like the *Grazer Kameratisch*, such that it is technology that adapts, as far as possible, to books. *PhaseOne*, *Hasselblad*, *Nikon* and *Canon* cameras are the preferably used acquiring devices, being aware that high resolution of the sensor does not automatically mean high image quality.

The execution of the project involves the Library's scientific, conservation, photographic, cataloging, and IT departments. It starts, as already shown, with the analysis of the conservators on the books to be reproduced, proceeds with their transport and digitization,¹⁸ the control of the images,¹⁹ the addition of metadata (descriptive metadata is retrieved from catalogs), and the dissemination of these virtual copies in a digital library platform, permitting the provision of free, worldwide access to these research resources, together with the possibility for users to download the digitized objects.²⁰

High level workflow is a prerogative for the functionality of all processes. Having been worked out from the test bed experience, it is continuously refined and adapted to new needs, as digitization is a permanent work in progress. At a certain point, it became necessary to manage it through a database. The already mentioned electronic management tool *Inside*²¹ connects the single departments and staff members involved, permitting the monitoring of the preparation and the single movements of the physical volumes,²² as well as the management of the various certification and dissemination steps which their reproductions go through, until the permanent storage of the data in the Library's digital object repository, and the return of the tomes to the Library's deposit.

Among the questions to be asked with regard to the management of the digital workflow process, there is the one of how to measure "success".²³ The answer was found to lie in the expansion of *Inside*, that became the instrument which is also used for the evaluation of the project's advancements and results. It permits us to perform statistical surveys of the production (i. e. how many images have been acquired when and by which scanner operator, of which manuscript or collection; who replaced which images; which images have been cancelled or re-named how many times; how many images of which project controlled by which tutor show not numbered folios or back pastedowns; how many digital objects are disseminated, stored or converted in which period), enabling us to acquire

¹⁸ Digitization is facilitated by a codified file nomenclature written especially for the Library's needs (including file names for the exceptions in foliation of the original). It is directly implemented in the acquiring software of the scanners. High quality digitization strategy has been chosen also when it was established to interleave every single manuscript page by white paper. Every reproduction of an entire volume, furthermore, is provided by txt and log files automatically generated by the acquiring software which permit interoperability of the reproduction with the databases and softwares involved in the whole workflow process.

¹⁹ Aforementioned high quality digitization strategy is, again, adopted especially during the quality review of the photographs whose validation – done by checking of photographic quality, aesthetic standards, and correct file naming - is made twice by two different tutors and performed with different programs. Post processing of the scanned TIF images is not done, while it is performed on the RAW data of the camera shots which are elaborated and transformed in non-compressed TIF.

²⁰ The Digital Library *DigiVatLib* is accessible at <https://digi.vatlib.it/?ling=en>, giving access at the moment to approximately 16.000 digitized manuscripts linked to online catalogues and bibliographic references. There are also images in different quality levels of printed books, archives, special materials, coins and medals, visual materials. Images are provided with an electronic watermark and disseminated in a leaf-through presentation version.

²¹ *Inside* is a web application that has been designed with open source languages. It is a system that is interoperable with third-party softwares, each of which undertakes specific but limited activities in the digitization process.

²² After the conservator's check, the volumes are provided with an RFID device chip, permitting to trace them at any time. After the transport to the digitization building, they are stored in a temperature and humidity controlled safety vault.

²³ Assess success doesn't exclusively mean to produce Key Performance Indicators, or to verify if the goals that were set, have been achieved. New considerations may emerge (for example, which kind of volume or collection is the most likely to incur in a specific kind of damage during the shooting phase), constituting new parameters of evaluation. In our case, it turned out that manuscripts damaged by iron gall ink need a different/additional acquisition technology to the standard one.

hundreds of queries by crossing different data. It is also the communication tool among the professionals involved. Quality control, in this way, is incorporated into the workflow process allowing weekly detection of production, of machine and human errors during the work phases, and of many other parameters.²⁴

As the project is developed and managed internally (collaborations with external expertise occur for specific project matters, usually technological ones), its constant adaptations to new requirements are easier to realize. The result can be considered a model for a high quality digitizing project.

Memory institutions have to guaranty strategies that ensure access to their digital assets over time. The Vatican Library's ones are illustrated in the following section.

3rd Session - FITS conversions and the use of PREMIS structure for recording digital preservation actions

Manuscripts and antique books of the Vatican Library's collections are being digitized to preserve them for future generations adopting a file format developed in the context of space missions and storing satellite images of the sky during the end of 70s of the last century: the *Flexible Image Transport System* (FITS), an open format, fully documented, without royalties or copyright, based on a series of specification publicly available and managed by a non-profit scientific authority.

The challenge of the implementation of Vatican Library preservation project was the conversion of the digitized images to the *Flexible Image Transport System* (FITS) with the aim to preserve images by using the long-lived technology evolved out over decades in the astronomical field and with an experimental approach to adjusting this format for libraries' purposes with the belief that:

*“An archival format must be utterly portable and self-describing, on the assumption that, apart from the transcription device, neither the software nor the hardware that wrote the data will be available when the data are read.”*²⁵

A FITS file is made of 2880-byte records called FITS blocks divided between a header and a data area. The major feature of the FITS format is that image metadata is stored in a human-readable ASCII header, so that an interested user can examine the headers to investigate a file of unknown provenance. Each FITS file consists of one or more headers containing ASCII card images (80 character fixed-length strings) that carry keyword/value pairs, interleaved between data blocks. The keyword/value pairs provide information such as size, origin, coordinates, binary data format, free-form comments, history of the data, and anything else the creator desires: while many keywords are reserved for FITS use, the standard allows arbitrary use of the rest of the name-space.

²⁴ Productivity (an important criterion, as closely connected with budget) has increased since the beginning of the project, while the error rate that determines the rejections of images remained, in proportion, nearly the same (on average, 0,09 % to 0,11 % of the images - corresponding to 10 % of the volumes – are replaced). Audit and certification of the integrity of the digital content through Jhove (JSTOR/Harvard Object Validation Environment) is about to be implemented.

²⁵ Steering Committee for the Study on the Long-Term Retention of Selected Scientific and Technical Records of the Federal Government, [US] National Research Council, *op. cit.*, p. 60.

The choice of FITS file as long-term preservation format implied the implementation of a conversion tool because the FITS file, in the Vatican digitization project, is not set as a publishing format but only as storage file format.

The Vatican Library, in collaboration with experts of the INAF (the Italian National Institute for Astrophysics) has implemented a tool, able to perform lossless conversion of data with particular regards to the TIFF / FITS and vice versa conversion (from FITS to derivative formats such as TIFF, JPEG, PDF). In this way the Vatican Library is able to store manuscripts digital images in FITS file in its WORM data storage device, ensuring the availability of the other formats for the needs of scholars and end-users.

This conversion is able to take into account all the characteristics of technical metadata (for example in TIFF format) in order to identify matches in the keywords of the header portion of the FITS file.

Conversions are performed after completion of the quality control review of images, described in the previous section.

The converter is able to gather all the information related to the creation of a master file in order to collect them into keywords of FITS files.

The FITS community has developed many conventions for using certain keywords or FITS file structures. The Registry of FITS Conventions provides a central and authoritative repository for documenting conventions that have been developed by the FITS user community for storing and transmitting various types of information in FITS format data files. The International Astronomical Union FITS Working Group is responsible for this Registry and for the rules and procedures. The Group has recognized the keywords of the Vatican Library for the use of the FITS format for library holdings.

After conversion from TIFF master files, FITS files, beyond the availability of technical metadata (coming from the native TIFF headers) as well as the use of FITS standard keywords (such as the CHECKSUM and DATASUM used to verify the integrity of files), are able to embed the following elements which represent the new keywords currently adopted in the Vatican Library's project.

KEYWORD: COLORMAP

DEFINITION: It indicates how colors are represented in the matrix Image. Only the value RGB it is currently used. Other possible values could be, for example, CYMK (four-colour printing) or BW (for black and white images).

KEYWORD: IMGURESL

DESCRIPTION: Indicates the resolution unit referred to IMGXRESL and IMGYRESL values.

KEYWORD: IMGXRESL

DEFINITION: The number of pixels per IMGURESL unit in the image width direction (horizontal resolution of the image)

KEYWORD: IMGYRESL

DEFINITION: The number of pixels per IMGURESL unit in the image height direction (vertical resolution of the image)

KEYWORD: ICCPDATA

It defines the ICC profile of the device that created the image (scanner or digital camera) and contains information that allows the faithful reproduction of the original color in the devices for display and printing.

DEFINITION: The sequence of bytes that defines the ICC profile is represented in hexadecimal. To rebuild the ICC profile you need to convert the hex values in bytes and use a software able to interpret it correctly.
Because of quantity of information saved in this keyword generally has to be used CONTINUE. (1)

KEYWORD: OPERATOR

DEFINITION: Identify by full name (or an abbreviation/code) the operator who carried out the acquisition.

KEYWORD: TUTORNUM

DEFINITION: After each scan the captured image will be reviewed by one or more trained operators, called "Tutor", who make a detailed check and quality control on the image. N indicates the levels of quality control(TUTORNUM = 0 means no check).

KEYWORD: TUTORn

DEFINITION: Identify by full name (or an abbreviation/code) the tutor who made the quality control in the level [n]

KEYWORD: TUTORnAP

DEFINITION: Date and time when the quality inspection was passed by the tutor in level [n]

KEYWORD: HOSTPC

DEFINITION: Identify the host pc used for the image acquisition

KEYWORD: SCANNER

DEFINITION: Identify the device used for the image acquisition. It may contain the device serial number or a value that uniquely identifies the device anyway.

The Vatican Library's IT storage system for long-term preservation, in which conversion are currently performed, includes:

- a Network Attached Storage system (NAS) for the storage of digital images, able to provide highly available access to data while maintaining high performance;
- a Storage Area Network (SAN) for the virtual infrastructure.

All this equipment is provided by DELL EMC Corporation.

The Vatican Library's EMC ISILON scale-out NAS storage is a clustered storage system composed of 35 + 5 nodes. Each node is a self-contained, rack-mountable device that contains industry standard hardware, including disk drives, Central Processing Unit (CPU), memory and network interfaces, and is integrated with operating system software called

OneFS. The current size of the ISILON storage is 5.6 PetaByte (2 Elastic Cloud Storage of 2,8 each).

SAN is composed of 2 unified hybrid storage array EMC VNX (30 TeraByte each).

The entire storage infrastructure is supervised and managed by a virtualizer, EMC ViPR, able to dynamically modulate the space available in the Data Center of the BAV.

ViPR software is integrated with VMWare and it currently manages more than 100 virtual machines with UNIX, LINUX and Microsoft OS.

The computational optimization of the infrastructure is governed by 7 main + 3 backup nodes. The 7 main nodes have a total capacity of: 212 GigaHertz for the CPU and about 1 TeraByte of Random Access Memory (RAM).

The conversions are performed in the above mentioned ISILON storage by the way of a software controller called Omnes (powered by Seret s.p.a. who has implemented the requisites established by the IT Service Management Coordination of the Vatican Library).

After ingestion of images, the software automatically tracks the content of FITS keywords in order to provide the redundancy of the information, embedded in FITS headers, into XML files, according to the PREMIS schema. It means that in the ISILON system, for each FITS file stored in it there is an XML file associated to it.

The long-term preservation database, deployed according to the semantic units of PREMIS, is able to manage, for each digital file ingested, the information, derived from the data contained in headers of FITS files and aggregated in the following sections of information:

Object:

The discrete unit of information subject to digital preservation is recorded here including elements describing physical attributes and properties of the digital objects:

In addition, it has to be noted that the search front-end of *Object* includes also codes of exceptions (that are assigned when there is an exception in the foliation of a manuscript), automatically derived from file names as well as all the other elements describing the *Object*.

Event:

Each action that involves or affects at least one Object or Agent associated with or known by the preservation repository are here registered, including the conversion event from TIFF files.

Agent:

Person, organization, or software program/system associated with Events in the life of an Object.

Rights

Statements of legal rights and permissions.

Thus, the elements indexed and searchable are

<*objectIdentifierValue*> (file name)

<*dateCreatedByApplication*> (creation date)

<*relatedObjectIdentifierValue*> (shelfmark)

<*eventIdentifierValue*> (code of the event)

<*eventType*> (type of the event, e.g. “scan”, “approval”)

<*eventDateTime*> (date and time of the event)

<linkingAgentIdentifierType> (type of values, distinguishing between, person/organization, software/hardware)

<linkingAgentIdentifierValue> (code of the agent)

while the other elements composing the PREMIS file as well as the extension container element (<objectCharacteristicsExtension> describing the technical metadata, previously defined in FITS keywords) are displayed but not provided as search indexes in the database.

From a list of results users can download PREMIS XML files by selecting items in list or performing conversions from FITS to derivative file formats.

The Vatican Library, in making the choice of using FITS format, has encountered willing collaboration on the part of researchers in the field of astrophysics, and at the same time it has enjoyed the valuable partnership of the international institutions which handle the FITS standard, with the goal of making it fully adequate for the preservation of library holdings. The European Space Agency (ESA) is continuing its partnership with the Vatican Library to preserve, manage and share the library's extensive collection of documents and texts. The partnership between the two organizations follows a five-year effort to digitize the library's collection using FITS format, in the light of an agreement signed in 2016.

References

Biblioteca Apostolica Vaticana, *BAV FITS Keyword Dictionary (Version 1.0, 2015)*, available at: <https://www.vaticanlibrary.va/home.php?pag=KeywordsFITS>

Deutsche Forschungsgemeinschaft, *DFG-Praxisregeln "Digitalisierung"*, 2016, available at: http://wissenschaftliche-sammlungen.de/files/5814/8759/5924/12_151_de.pdf

European Space Agency, *ESA and Vatican work to preserve heritage data*, available at: http://www.esa.int/Our_Activities/Observing_the_Earth/ESA_and_Vatican_work_to_preserve_heritage_data

Europeana Photography Workshop 2012 at Girona, available at: http://www.photoconsortium.net/wp-content/uploads/2017/06/Girona_Workshop_Recommendations.pdf

Falcinelli, Riccardo, *Cromorama* (Turin, 2017).

Federal Agencies Digital Guidelines Initiative (FADGI), *Technical Guidelines for Digitizing Cultural Heritage Materials, Creation of Raster Image Files*, 2016, available at: http://www.digitizationguidelines.gov/guidelines/FADGI%20Federal%20%20Agencies%20Digital%20Guidelines%20Initiative-2016%20Final_rev1.pdf

Grimaccia, Marta, *The Vatican Library conservation workshop*, communication presented in the satellite conference in Rome at the IFLA-PAC, August 31, 2009, available at: http://www.ifla.org/files/assets/pac/Satellite_Meeting_Rome_2009/Grimaccia_eng.pdf

IFLA, *Linee guida per pianificare la digitalizzazione di collezioni di libri rari e manoscritti*, 2015, available at: <https://www.ifla.org/files/assets/rare-books-and-manuscripts/rbms-guidelines/guidelines-for-planning-digitization-it.pdf>

The Library of Congress, *PREMIS Data Dictionary for Preservation Metadata*, available at: < <https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>>

Metamorfoze Preservation Imaging Guidelines, available at: https://www.metamorfoze.nl/sites/metamorfoze.nl/files/publicatie_documenten/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf

Metis Systems S.r.l., available at: <<http://www.https.metis-group.com/>>

Modolo, Mirco, *Il dibattito sulla liberalizzazione della fotografia digitale in archivi e biblioteche quattro anni dopo l'appello di Reti Medievali*, in *Reti Medievali Rivista*, 18, 1 (Florence, 2017), available as well at: <http://rivista.retimedievali.it>

NASA, *FITS resources*, available at: < https://fits.gsfc.nasa.gov/fits_resources.html>

Piazzoni, Ambrogio Maria, *Introduzione alla storia della Biblioteca Apostolica Vaticana*, in *Conoscere la Biblioteca Vaticana. Una storia aperta al futuro* (Vatican City, 2010).

Steering Committee for the Study on the Long-Term Retention of Selected Scientific and Technical Records of the Federal Government, [US] National Research Council, *Preserving Scientific Data on our Physical Universe* (National Academy Press, 1995).

Wueller, Dietmar, Bøgvad Kejser, Ulla, *Standardization of Image Quality Analysis – ISO 19264*, available at: <https://www.image-engineering.de/library/conference-papers/883-standardization-of-image-quality-analysis-iso-19264>