Paolo Buonora*

Digitization, online utilization and preservation of cadastral very large format cartography

Keywords: Archivio di Stato di Roma; Large format carthography; Digitization; Digital preservation; WebGIS.

Summary
The aim of this paper is to give e brief idea of the effort that the Archivio di Stato di Roma, and then the Centro di Fotoriproduzione, Legatoria e Restauro degli Archivi di Stato, made in more than ten years in digitizing archival documents and implementing new services to exploit archival heritage, both for the general users and for the professionals or public administrators. The most advanced of these experiences deal with cartographic materials and our unique collection of cadastral maps of Pontifical State, XIXth century[1].

Imago Project in Italian State Archives

The State Archive of Rome (Archivio di Stato di Roma – AdSR) is placed near Navona square, in the old university "Sapienza" building (a Borromini's masterpiece), and holds the documentation produced in the past centuries from the Papal State, who included a large part of central Italy, and contemporary documents from the roman region. Document typology includes manuscripts, parchments, registers, volumes and maps[2]. The Gregoriano Cadastre is the first modern cadastre of the entire Pontifical State: it was promoted by Pio VII in 1816 and activated Gregorio XVI in 1835. It was made by the *Presidenza generale del Censo*, with the same technique that was adopted during the Regno d'Italia (1805-1814) in making the napoleonian cadastre in the Legazioni province (Bologna and Romagna) and in the Marche region. Two copies of it were made, one was kept in the *Presidenza generale del Censo* (now at the AdSR), one in the local offices of the Cancellerie del Censo (now in several State Archives of the ancient Pontifical State). It is composed by three related series: maps (1:2000, in one large roll in Rome, in single sheets in the local version), reduced maps (1:4000 o 1:8000), owners registers (*brogliardi* or *sommarioni*). The roman version of the collection is generally the most complete, original and well preserved comparing to the one in the local archives, where originals were often remade during more than 100 years. In other words, this cadastre offers the more complete vision of the land and urban shape relating to ancient Pontifical territories.
For preservation purposes a massive microfilming activity began since the '50, on the most precious series and on all most consulted and exposed documents. Microfilm allowed to let the originals safe, and access the surrogates on microfilm. But traditional photograph and microfilm never solved the problem of an efficient management of the surrogates, nor of making copies in colour of large format materials: this was provided only by digital imaging. Scanners and digital backs became more and more performing and allowed to reach a spatial resolution of 250-300 ppi in colour also on large format maps.

* State Archive of Rome [paolo.buonora@beniculturali.it]
[1] Part of this paper is a new version of the following: P. Buonora, *Digitization and online access strategies/ Numérisation et stratégies de l'accès en ligne*, International Congress of Archives, Vienna 21-29 August 2004, available at: http://www.wien2004.ica.org/fo/speakers.php?ctNv1=48&ctNv2=&IdSpk=388&AlphSpk=B&p=3&SpkV=2. About the same topics see also ID., *Digitalizzazione e accesso on-line per la cartografia storica*, in *Un accesso migliore è possibile ... verso l'integrazione delle risorse informative per l'architettura e l'urbanistica*, Atti delle ottave giornate di studio del CNBA, Venezia, 28 – 31 maggio 2003 (I quaderni del CNBA, n° 7), edited by Laura Casagrande, Serena Sangiorgi, Pierre Piccotti, Venezia CNBA - Casalini Libri, 2005; available at: http://www.iuav.it/CNBA/Giornate-d/2003-Le-Ot/Abstract--/Buonora.doc_cvt.htm , and the *Linee guida per la digitalizzazione del materiale cartografico*, edited by Gruppo di lavoro sulla digitalizzazione del materiale cartografico, Roma, ICCU 2006.
[2] http://www.archiviodistatoroma.beniculturali.it

The Imago II project developed in several years. The first step (1997-1999) was producing a significant digital asset, digitizing complete series of the most precious and consulted documents (130.000 images); many of them were maps (15.000), and the required resolution was reached with roll scanners for contemporary maps.
When Imago Project began, digitization of documents was mostly performed as a further development of microfilm, i.e. digitizing security microfilm. Obviously, the choice of an indirect capture has a limit: if we choose to start with a 35mm. film, we will never produce a large amount of graphic information. It does not matter how many pixels we generate, the problem is what they represent, and what we can see in the digital image. Also if we digitise large format films, in the case of detailed maps we will be able to read details up to a A0 format (cm. 70*100), not more. The next step was necessarily the adoption of more performing devices of direct digital capture of the original, scanners and digital backs. To improve the performance of the digital back required increasing the number of CCD of the sensor, and consequently the spatial resolution (number of points represented in the space unit, currently ppi or dpi). On the other side, scanner results are improved by augmenting correspondingly the dimension of the original that could be captured in one shot. Starting with commonly used A4 and A3 format scanners, technology arrived at flatbed scanners of large format, such as the Tangent scanner that was used by the Library of Congress for its map collections[3]: this device can scan originals up to 24" * 36" inches, a little less than the A0 format. A resolution of 300 ppi (points per inch) was adopted for medium size maps, and it represents until now a "de facto" standard for high-resolution digital imaging in graphic documents. To keep this spatial resolution for larger documents, assuming the dimension of available sensors (5.000 CCD at the end of the 90th, up to 14.400 CCD today), the technique of coupling more images to produce an unique final image was employed: both with different sensor and different optics working one beside the other, both with a unique sensor moving precisely behind a unique optic, and taking different shots. In the first case we have roll scanners, very useful if we have to scan long originals such as maps[4]. The second approach was used by important firms in professional photography, as Sinar, or in Italian archives for very large maps, up to 3 m * 4: on request of some Imago projects in the State Archives, a small company in Rome created a special device[5]. Both these kinds of devices can generate an impressive output, from 500 Mbytes for the roll scanners, up to 3,5 Gbytes for the Metis DMC (Digital Macro Camera). Such large digital objects pose a management problem which will be discussed later.
The second step (2001-2002) was providing full resolution images on the web. While most of the digital libraries choosed to provide only a JPEG copy of the TIFF master file, with limited resolution, the AdSR followed the experience of the *American Memory project* on maps at the Library of Congress, which used at that time a MrSID format from Lizardtech. All our imagery was converted to MrSID, and a web interactive application was implemented integrating a Lizardtech image server. The first important experience of virtual digital access to historical archive documents was conceived many years ago, at the *Archivo General de Indias* in Seville[6]. There it was carried out, for the first time, the ambitious project to substitute the direct consultation of the originals with the access to digital surrogates; entire archival series were digitised employing flatbed scanners, and several hundred thousands images were done. At that time virtual access was thought and carried out as a "virtual reading room" to be used on-site.
After that, the Internet crossed many borders, including virtual access to historical documents with no distance limit. The first in Italy to adopt an "image server" technology, and to open a virtual access to its digital docu-

[3] See David Yehling Allen, *Creating and Distributing High Resolution Cartographic Images*, «RLG Diginews», 2/4 (15 agosto 1998):
http://worldcat.org/arcviewer/1/OCC/2007/09/28/0000073852/viewer/file1145.html#feature
[4] The roll scanner we used in the late '90 the AdSR was a Rigoli/Colortrac http://www.colortrac.com . This kind of scanners can also be adapted to a glass table to scan registers, as for the SMA scanner: http://www.sma-edoc.com
[5] Metis DMC was first successfully employed on very large format maps of the State Archive of Turin: www.metis-group.com
[6] P. Gonzalez Garcia, Long term projects and changing methods: the case of the Archivo General de Indias, in Digitisation of European Cultural Heritage. Products-Principles-Techniques, Utrecht University, Netherlands, 21-23 October 1999, edited by Hans Mulder and Frans Wiering: http://www.cs.uu.nl/events/dech1999/dech.htm

ments was the State Archive of Florence, who in march 2000 made a web access to the important fond *Mediceo avanti il Principato*, that had been digitised with the same techniques than were used in Seville. In 2003 State Archives of Spain experimented a new web access for their digital assets, using the same approach as the State Archive of Florence, but offering to the users about 14 million scanned pages of documents from several Spanish archives. AER (Archivos Espanoles en Red)[7] is still one of the most relevant archival digital asset available online. Both Florence and Spain allow a free access after the registration of the user.

Nevertheless, we are still dealing with small size originals, scanned at a medium resolution: more efforts had to be made to give a remote access to more large format documents and higher resolution images. Maps and plans in our archives seemed to challenge any chance of digital technology, in consequence of their size and and of the total number of items: it is really hard to reproduce them in detail, to manage such large image files, or to deliver the image on the web, assuming a usual bandwidth limit. The usual approach is since a long time to use three digital images: a thumbnail for the catalogue record, a medium resolution image to be delivered on Internet, and finally a full resolution high quality image to keep and deliver only on site. This solution is common in most commercial systems of image delivery and selling online, but it has no use in historical documents. Archive users do not need simply to recognise a catalogue item, as it happens in a picture or a sculpture: they need to read the archival document, no matter how large and detailed originals are.

The solution to the technical problem of managing extra large format images came from the space satellite world: in this context, assembling and managing this kind of image, for military or geographical purposes, became common practice. The development of sensors able to record substantial levels of detail required newer software technologies capable of providing for building a large image piece-by-piece and the ability to zoom into a part of a single image. This was the origin of the "multi-resolution" formats, which store in a single file – or in a hierarchical structure of files – several different resolutions of the subject.

The American Memory mentioned was the first in the cultural heritage community to adopt this solution to give a web access to cartographic materials, just because of the remarkably high quality digital asset produced in the digitisation of the map collections. This work had a considerable impact on the cartography stakeholders all around the world[8], as it had an extremely liberal approach, offering at the same time a free and high quality access, and the ability to freely download a full image. I would stress this point, as the multiresolution technology gives also a good solution to the management of property rights on the image. The user can display only a small portion of the image, and cannot download the original full resolution image unless the system administrator allows him to do it. This opportunity drives a correct distinction between the user who accesses the document free, and the professional who would pay for professional and commercial uses.

The Library of Congress adopted the Lizartech MrSID (Multi-resolution Seamless Image Database) file format[9]. Other firms adopted proprietary formats with a multiresolution approach: ECW format by ER Mapper[10], the main competitor of Lizardtech in cartographic services, or XL Image format by the Italian Centrica, mostly oriented to high colour quality reproduction[11]. We also followed this choice in the AdSR, and in 2002 we moved our local digital assets management systems to a new system accessible on the web, using MrSID

[7] The AER project is officially illustrated in the International Congress of Archives in Vienna, 2004.
[8] Just to mention some others National Libraries, see the rare maps project in Australia (http://www.nla.gov.au/rmaps/) and the maps in the National Library of Scotland (http://www.nls.uk/collections/maps/index.html).
[9] http://www.lizardtech.com .
[10] http://www.ermapper.com .
[11] The format was used in the Uffizi Gallery and Alinari Photographic archive in Florence: http://www.xlimage.it.

image format and Lizardtech software[12]. These proprietary formats remained the only solution until the new JPEG 2000 really became an ISO standard in late 2002[13].

Digitization and online delivery of very large cadastral maps

In 2005 the AdSR added a new relevant digitized asset to his online access service: the original big roll maps of the Roman region, from the Gregoriano Cadastre, were scanned in one single file. As this maps can measure up to 3*4 meters and the design is very detailed, the special camera we mentioned for very large formats (Metis DMC) was used: in this device, the sensor moves inside the camera putting three scans in one single file that can be larger than 3 Gb, providing a 256 ppi resolution. All these 500 large images were provided directly in a Lizardtech MrSID format. Considering the amount of storage needed, but also the high quality of the original scanned images, this experience confirmed our choice of adopting a 1:20 compression also as a master copy of the image: viewing, managing and preserving these giant images would be very difficult on the original TIFF. The final step (2007) was migrating all the images in the new JPEG 2000 ISO standard, always using Lizardtech software: a price to be paid to get our image asset free from a proprietary format, as Lizardtech MrSID was. The Imago service was charged to the *Centro di Fotoriproduzione, Legatoria e Restauro degli Archivi di Stato*, where new powerful servers with 4 Gb RAM, and open source applications (PostGres database) were implemented (www.cflr.beniculturali.it / Digital Library), using Lizartech Express Server for delivering the images: in this way acces online was provided also for the very large maps of the "Campagna romana".

GIS implementations

A significant cartographic asset of the AdSR is the Urban cadastre, which is a part of the Gregoriano cadastre, and was used for implementing a GIS application on historical maps. Urban cadastre of Rome was born both for maintaining city roads, both in the context of the State cadastral operation which focused on new taxes of XIXth century on real estate properties. The architects that were charged for the cartographic survey, in 1818, used the Nolli map (1748), dividing it in house blocks and composing blocks in map sheets for each neighborhood (*rione*). The design of the maps in scale 1:1000 was the starting point to compile owners' registers (*brogliardi*); as any litigation was satisfied, cadastre was activated in 1824. The neighborhood maps must be compared with two series of original subdivision maps and of later upgraded maps of 1871, as well as original registers are to be compared to a second series of upgraded registers and a third series of new registers made in 1871. The whole documentation provides – as for others towns of the Pontifical State – a faithful and extremely detailed image of the urban shape, of functional and economic aspects of modern Rome before it became the capital of the Italian State. The representation of monuments, horticulture and gardens, of fountains and of the Tiber river is very relevant by the artistic point of view.

In 2006 a new project started, financed by the CARIPLO bank, focused on the GIS techniques implementation on our digitized cartography. Partners in this project were the Urban Studies Department of the University of Roma 3 (DIPSU), the Soprintendenza Archeologica of the city of Rome, and our *AdSR*[14]. These institutions assumed that, although specific missions were different, a common perspective was possible to reach some common goal. DIPSU has to study architectural and urban evolution of Rome, the Soprintendenza must ingest

[12] http://www.cflr.beniculturali.it/Imago/index.html .
[13] See *Editor's Interview: JPEG 2000. Dr. Daniel Lee, ISO SC29/WG1 (JPEG)*, in RLG DigiNews, vol. 6/6, December 2002: http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070513/viewer/file882.html#interview .
[14] Project director is prof. Paolo Micalizzi, teaching in DIPSU Roma 3; DIPSU laboratory director is Stefano Magaudda; responsibles in the Soprintendenza are Susanna Le Pera and Luca Sasso d'Elia, who provided many of the technical information below.

data, to manage and preserve archeological heritage, and the *AdSR* has to preserve and make accessible the original documents. All of them have the same interest to make historical cartography easy to access, specific information easy to retrieve, provide integration in the complex information that can be inferred by different series of documents. The implemented model considered 3 spatial dimensions (each property on the map may have several floors) and 1 temporal dimension; in consequence of that we decided to make a GIS, a geo-database that could be used for common or specific goals of three institutions. The implementation of a WebGIS was considered a primary goal, both to share results among partner institutions, and to deliver a new useful service to several range of users. The challenge of making a GIS with this material had to face with many problems: ancient maps do not match easily with the contemporary cartographic instruments. So to the original digital object 3 morphing operations (on the single map sheet, within the neighborhood, among neighborhood) were applied in progress to make the matching, which makes also the final digital object something very different from the archival original. Using open source software, a database was created with more than 16.000 records, copying data from the owners' registers, and indexes were made of the owners, or of property addresses. An open source software application was also made implementing the WebGIS, which provides a free access online, as a general service for all the different goals of our institutions[15]. In this WebGIS other materials are available too, as a digital version of the Nolli map of Rome and, soon to come, archival documents on specific properties, taken from notaries and municipal funds. An extension in time is also planned, including in the GIS more contemporary version of cadastral cartography. These extensions are included in the new CARIPLO bank financial plan that deals not only with the city of Rome but also with Milano and Bologna. We expect a very interesting development of our work by exchanging experiences and knowhow with our colleagues of the archives and universities involved in this new project. Since several years, other GIS projects based on Gregoriano cadastre are undergoing in many Italian provinces: Pesaro, Ancona, Ascoli Piceno, Macerata, Bologna. Very recently a coordination network has been proposed in a meeting in Pesaro[16]. The *AdSR* is also planning with Regione Lazio a complete digitization of our maps and registers related to the region, to build a GIS on the whole territory. We hope that similar initiatives will start with other Regions (Marche, Umbria, Emila-Romagna) until all our Gregoriano cadastre will be digitized and available online.

Digital preservation and JPEG 2000[17]

Digital preservation requires a strategy for the storage of large quantities of data, which increases dramatically when dealing with high resolution images. Typically, decision-makers must choose whether to keep terabytes of images in their original Tiff format or compress them. This can be a very difficult decision: to lose visual information though compression could be a waste of the money expended in the creation of the digital assets; by choosing to compress, we will reduce however the costs of storage. Wavelet compression of Jpeg 2000 produce an high quality image: it is an acceptable alternative to Tiff, and a good strategy for the storage of large image assets. Moreover, Jpeg 2000 may be considered a format which is able to guarantee an efficient robustness to bit-errors and offers a valid quality with transmission or physical errors: this point of view is confirmed by the case study we report further, concerning image quality after occurrence of random errors by a comparison among different file formats. Easy tools and freeware software can be used to improve format robustness by duplicating file headers, inside or outside the image file format, enhancing the role of Jpeg 2000 as a new archival format for high quality images. Among different features, error resilience, offered by the JPEG2000 coding, is considered fundamental for digital preservation. Typically, two types of errors are considered: bit errors and packet losses (during transmis-

[15] http://www.dipsuwebgis.uniroma3.it/gregoriano/

[16] http://www.provincia.pu.it/index.php?id=3518&cHash=d548324085&MP=3491-4348&tx_ttnews[tt_news]=12623&tx_ttnews[backPid]=5492

[17] See P. Buonora, F. Liberati, A Format for Digital Preservation of Images. A Study on JPEG 2000 File Robustness, on D-Lib Magazine, July/August 2008.

sion). The error control offered by the JPEG2000 operated in these cases. The impact of bit errors heavily depends on the error location while packet losses have a catastrophic result on the coded image because they lead to a synchronization loss between coder and decoder. Code-blocks are coded independently in EBCOT; bit errors hence affect only the concerned code-block. For packet lossless, JPEG2000 offers the opportunity to use two kinds of strategies: enabling the detection of bit errors and the deletion of corrupted data. Using smaller precincts increases the robustness against bit errors although the coding efficiency is decreased. The adoption of markers for error resilience in Jpeg 2000 implies a robustness which makes this format better than any other for the digital preservation. In the Digital Laboratory of the Centro di Fotoriproduzione Legatoria e Restauro (CFLR), we compared the effects of introducing random errors (excluding the headers) in a set of images files, proposed by Kodak, presenting all possible chromatic cases. The error numbers was in according to image dimension: 0,01% (around 10 byte); 0,1% (around 100 byte) and 1% (around 1000 byte). In our tests, a Jpeg 2000 image behaviour as follows depending on the errors introduced:

- 0,01 % of errors does not produce a great effect on the image;
- 0,1 % implies a noise distributed on a huge part of image, but the subject of image and principal characteristic are preserved;
- 1% spread the noise on the entire image and the subject is not distinguished.

If we compare with the effects on a sequential compression format (i.e.: JPEG file), the image can be irremediably corrupted at only 0,01 % of errors; a simple error can go ahead on the entire bit-stream and can prevent correct sequential decoding. Increasing errors, file image cannot be open. In other not compressed formats as TIFF the structure and organization of the file can be ruined by some sequential byte errors or error bytes in particular areas (i.e.: end-line, end-strips). At the same time, in all formats, the headers in the image file are a crucial element. The header has the scope to supply the key information to visualize the image. It is a sequence of binary values, usually at the beginning of the file, in which are stored width, height, colour information and many others elements. It is possible to have corrupted or not-visible image losing or corrupting only some bits in the header. As the headers are a very small percentage of the whole file, it is very easy to improve the robustness of the file making a copy of this information inside or outside the file: information technology uses this principle of redundancy to make data safer. The Digital Laboratory of the CFLR implemented a freeware tool, called FixIt!, downloadable at its website[18]. With this tool the user can save, check, and eventually restore the main header of a Jpeg 2000 file: it is very easy to use, and it really improves the robustness of your image files. It is particularly useful if user cannot afford to make many copies of its digital assets. In the next future we could go further, and save with our FixIt! tool also tile headers, and have a more advanced security.

References

David Yehling Allen, *Creating and Distributing High Resolution Cartographic Images*, «RLG Diginews», 2/4 (15 agosto 1998): http://worldcat.org/arcviewer/1/OCC/2007/09/28/0000073852/viewer/file1145.html#feature.

Paolo Buonora, *Digitization and online access strategies/ Numérisation et stratégies de l'accès en ligne*, International Congress of Archives, Vienna 21-29 August 2004, available at: http://www.wien2004.ica.org/fo/speakers.php?ctNv1=48&ctNv2=&IdSpk=388&AlphSpk=B&p=3&SpkV=2.

Paolo Buonora, *Digitalizzazione e accesso on-line per la cartografia storica*, in *Un accesso migliore è possibile ... verso l'integrazione delle risorse informative per l'architettura e l'urbanistica*, Atti delle ottave giornate di studio del CNBA, Venezia, 28 – 31 maggio 2003 (I quaderni del CNBA, n° 7), edited by Laura Casagrande, Serena Sangiorgi, Pierre Piccotti, Venezia CNBA - Casalini Libri, 2005; available at: http://www.iuav.it/CNBA/Giornate-d/2003-Le-Ot/Abstract--/Buonora.doc_cvt.htm

[18] http://www.cflr.beniculturali.it/Progetti/JPEG2000.php

Paolo Buonora, F. Liberati, A Format for Digital Preservation of Images. A Study on JPEG 2000 File Robustness, on D-Lib Magazine, July/August 2008.

Editor's Interview: JPEG 2000. Dr. Daniel Lee, ISO SC29/WG1 (JPEG), in RLG DigiNews, vol. 6/6, December 2002: http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070513/viewer/file882.html#interview .

Pedro Gonzalez Garcia, Long term projects and changing methods: the case of the Archivo General de Indias, in Digitisation of European Cultural Heritage. Products-Principles-Techniques, Utrecht University, Netherlands, 21-23 October 1999, edited by Hans Mulder and Frans Wiering: http://www.cs.uu.nl/events/dech1999/dech.htm .

Linee guida per la digitalizzazione del materiale cartografico, Gruppo di lavoro sulla digitalizzazione del materiale cartografico, Roma, ICCU 2006.

Institutions URL	*Companies URL*
http://www.cflr.beniculturali.it/Imago/index.html	www.colortrac.com
http://www.cflr.beniculturali.it/Progetti/JPEG2000.php	www.ermapper.com
http://www.nla.gov.au/rmaps/	www.lizardtech.com
http://www.nls.uk/collections/maps/index.html	www.metis-group.com
http://www.provincia.pu.it/index.php?id=3518&cHash=d548324085&MP=3491-4348&tx_ttnews[tt_news]=12623&tx_ttnews[backPid]=5492	www.sma-edoc.com
	www.xlimage.it