## The Vatican Library towards the year 2000 and beyond

0

0

۲

Paul Gabriele Weston Librarian of the Biblioteca Apostolica Vaticana, Italy

Libraries have traditionally been conceived as institutions established and organized to manage sometimes huge quantities of data. The procedures that regulate their activity are usually subdivided into four quite distinct categories: the acquisition of bibliographic and documentary resources; their organization; the identification of documents and access to their consultation; the analytical or synthetic description of their content with the subsequent insertion of the relevant data into the information circuit. Libraries have long since developed specific techniques for each of the aforesaid procedures, and subjected them to the necessary adaptations to enable them to cope with the changing requirements of new types of documents. The same process of adaptation has also occurred whenever the benefits deriving from the use of more advanced technologies have become clear.

The discovery of the potential of the World Wide Web as a means for the diffusion of information and for interaction between the users of the Web proved no exception. Indeed, its extraordinarily rapid growth has challenged librarians to develop in great haste solutions that might permit a rational organization to be given to the information resources that were gradually becoming available, before the excessive quantity of documents made any kind of search and retrieval system unmanageable. One fact was clear right from the start: the available techniques were no longer adequate; no longer would it have sufficed to make small-scale adjustments to the organization of work. The transition from written to digital culture was in no way comparable to the arrival of videocassettes or even the first CD-ROM's in libraries. These could still be treated in much the same way as bibliographic documents; to describe them, it was enough, for example, to supplement the existing cataloguing rules with one or two new chapters. Conceptually, such documents were (like books) physically present in the library; they were static, and available for consultation on request with the aid of specially provided apparatus. Digital culture, by contrast, is mainly based on documents contained in remote archives, and available for consultation on-line; documents, moreover, that are not at all static, but often subjected to a continuous process of transformation, augmentation or elimination in times that are hardly ever predictable. The need to give an organized form to this material through bibliographic cataloguing has led to the development of the conceptual structure of metadata, i.e. that series of elements that each document ought to include, duly marked, to permit the indexing of the document itself, its consequent retrieval during the search phase, and the links between different documents that we know by the name of "hyperlinks". More than one expert has expressed the view that libraries will only be able to survive and continue to play an active role in the information society if they are alert to the new requirements and able to absorb these concepts into the strategic vision of their own tasks, and that they do so as rapidly as possible, while the material is still in its initial stage and the applications are still relatively limited in diffusion.

The most complete application of all this so far is the so-called *digital library*, whose definition has undergone numerous adjustments over the last five years. Michael Hart, in presenting the Gutenberg Project, affirmed that "an electronic library consists of computer searchable collections which can be transmitted via disks, phone lines, or other media (at a fraction of the cost in money and time as with present day paper media), featuring electronic books that will not have to be reserved and restricted to use by one patron at one time. In other words, all materials will be available to all patrons from all locations, at all times." Such a definition can clearly be applied to a universal digital library, without walls, without shelves, without librarians, and hence a totally virtual library. However, it does not take into account the diverse practical aspects that make the individual library anything but a mere collection of documents. It is as if the enthusiasm caused by the immense prospects of global communication opened up by the Web have led it to being identified as the complex structure to which we give the name of library.

Far more conformable to the real nature of libraries is the alternative definition of the digital library given by the Digital Library Federation a short time ago:

Digital libraries are (according to this definition):

- organizations that provide the resources, including specialized staff,
- to select, structure, offer intellectual access to, distribute, preserve the integrity of, and ensure the persistence over time of
- collections of digital works
- so that they are readily and economically available
- for use by a defined community or set of communities.

This definition, while it has the advantage of leaving almost unchanged the objectives of the conventional library, also poses a series of questions of fundamental importance: "How should libraries define what they wish to do with this technology? How will it serve their needs as institutions in fulfilling their mission in society?" To understand what approaches libraries have adopted in responding to these two questions, it is enough to consider in what way they have made use of the potential of the Web in the course of time. During an initial phase, appreciation was shown for the possibility of producing information of considerable communicational effectiveness by its simple presentation in graphic mode. This explains why the first data entered in the Web were those on library organization: information ranging from library opening hours to a listing of the services provided, to the latest accessions and the presentation of particular events. The ever more simple treatment of the electronic pages was matched by the availability on-line of a growing number of information services for the user: guides to bibliographic research, reference works such as encyclopaedias and dictionaries, specialized bibliographies. The next step was to permit access to the electronic documents present in daily increasing quantities in the library's holdings, through the institutional website. This resulted in the confines between the different types of information (bibliographic, textual, factual, etc.) becoming ever more fluid. With the result that the more developed sites today - I could cite as a model that of the Library of Congress - are far from simple on-line public access catalogues. The wealth of hyperlinks enables the user to pass effortlessly from the page of the diary of an American President, say, to the collection of images of his family and from these to books and articles devoted to his biography; he may even find himself placed in the midst of a virtual exhibition, from which he can obtain all the information he needs on how to consult or reproduce one of the manuscripts on display. In this case, the request may be made by compiling a special form directly on-line and entering onto it the details of one's own credit card for payment. And if the user's curiosity is still not satisfied, he can easily transfer to another site, anywhere in the world, and so continue his own investigations.

0

It has to be said, however, that despite the fact that the quantity of electronic resources has enormously grown over the last eighteen to twentyfour months, almost all libraries have not restructured themselves in any radical way to cope with them: all they have done is to organize their electronic documents more or less according to the previous structure, and simply transferred their own catalogue and other research tools to computerized form. This is only the first step in the realization of a true *digital library*. What has slowed down the widespread diffusion of digital library projects – however desirable – is not just the lack of resources at the disposal of libraries, but obstacles of a technical nature, the lack of precise standards of reference, and the inadequacy of many of the infrastructures to permit the distribution of digital information. In the light of these difficulties, it is easy enough to understand why the onus for spearheading the more demanding campaigns of document digitalization has been assumed by the great European and North American institutions, such as the national libraries, which are institutionally pledged *inter alia* to safeguard and conserve the cultural heritage of their own country.

As far as the Vatican Library is concerned, it has characteristics rather different from those of the national libraries. Its peculiar status is also recognised in its Statute, indeed in the very first line of its Statute: it is not the library of the Vatican, i.e. the library of Vatican City State, but the library of the Holy See. It is not therefore, or not yet at any rate, a national library of a separate nation, in this case Vatican City State. It does not yet play the role of a national deposit library: i.e. of preserving all the books that have been published in that particular state. Nor does it compile a national bibliography. The fact that it is the library of the Holy See, i.e. of an institution that far transcends the political divisions between states, gives it a particular status: it places it at the service of the common good of humanity, what the modern founders of the Library, Nicholas V and Sixtus IV, called the "communis commoditas doctorum virorum", the common convenience of learned men. It is therefore a research library for the advancement of learning.

It should also be borne in mind that the Vatican Library represents in some sense a library sui generis, given the sheer variety of its holdings: it possesses many tens of thousands of ancient engravings, tens of thousands of drawings and art objects, one of the largest numismatic collections in the world (comprising over 400,000 coins and medals), and many hundreds of thousands of books printed in the last two centuries. Nonetheless, its great richness, the principal reason why so many scholars from all over the world flock to it daily, undoubtedly consists of its holdings of ancient books: manuscripts, above all, but also incunables, and early printed books of the 16th century, as well as huge deposits of 17th and 18th centuries books. The manuscript codices amount to some 150,000. Half of these are archival in nature (they come, that is, from the private collections of some of the leading families of the past, such as the Chigi, the Barberini and others). The other half consists of what we call library manuscripts proper. There are some 75,000 of these, of which some 60,000 are Latin codices, some 5,000 Greek codices, 801 Hebrew ones, and 9,000 Oriental ones. This great variety in the nature of the collections represents, for those responsible for their transposition in electronic form, an unique occasion to apply, within the same project, sophisticated technological solutions in response to the specific needs that each type of

document poses. It goes without saying that the description for cataloguing purposes of a coin proceeds according to different criteria than those adopted for the description of a modern book. It also goes without saying that the needs of someone who consults the manuscripts are different from those interested in frescoes or archaeological artifacts. The digitalization of all material, whether it is a case of its description or the acquisition of the electronic image, or the realization of the procedures to permit the user to identity and retrieve it, must therefore be specifically geared to the individual characteristics of each class of objects. Each solution that fails to take this need into account from the start risks giving rise to a product in some sense hybrid: a solution that will probably dissatisfy in equal measure both those within the library who are responsible for the description of the documents, and those outside it who, for reasons of study or mere curiosity, are interested in consulting the electronic archive.

۲

When the decision to automate the Vatican Library was taken in 1985, a great deal of long hard thought was given to the question whether it was appropriate to install several distinct electronic systems, in line with what had been the standard practice in the library for many years and that had led to the realization of a series of heterogeneous printed, typewritten and manuscript catalogues and inventories: each very different in nature due to the descriptive criteria each adopted and the information each made available; or, alternatively, whether it was not better to create a single information system, searchable according to the same procedures, irrespective of the nature of the document being sought. Among the considerations that led, finally, to this second alternative being preferred, one of the most important was that a complex system, conceived in this way, integrated with images and possibly with other multimedia information, would give rise to a very powerful research tool, whose utility would be greater than, and whose potential far outstrip, the mere sum of a series of separate catalogues. The project of the digital library conceived along these lines, and based on the Vatican Library's holdings, can indeed be described as the realization of a virtual encyclopaedia of humanistic scope, in which the Renaissance spirit that inspired the first years of the library's existence is given a new lease of life.

No matter how fascinating in prospect, the enterprise was not, at the outset, without unknown quantities and unresolved problems of various nature. So there was no other choice but to carry out a pilot project first: only in this way could the succession of procedures be formulated and the most appropriate solutions identified to the problems encountered. My previous remarks on the nature of the library and on the characteristics of its most valuable collections will explain why it was decided to regard the manuscripts as the most challenging test bed, on which the efforts and initial objectives of the pilot project should be concentrated. The manuscripts in fact comprise an incredible variety of supports (parchment, paper, papyrus and other materials), inks and other pigments, forms and dimensions. Many manuscripts are bound, others are particularly fragile. Their electronic scanning and the subsequent storage of the images thus obtained represent elements of evaluation of the highest importance, with a view to the definition of the criteria to be adopted in the final project.

An advisory committee, formed of a select group of university professors, chosen on the basis of their academic interests and the different geographical provenance of each, was set up. To it was given the task of identifying a first significant sample of manuscripts to subject to the initial experimentation, chosen from among the better known and more valuable ones in the Vatican Library, and of formulating a series of recommendations of objectives to be pursued, or procedures to be adopted, in the realization of the electronic archive. The list of recommendations thus drawn up included the following:

- each document reproduced should be accompanied by a sufficiently detailed catalogue description to represent in an unambiguous way the document itself and permit an adequate number of searches
- the digital images should be of the highest possible quality

0

0

- access to the electronic descriptions and reproductions should primarily be possible via Internet
- the system should provide very reduced response times
- the adoption of the more widely used data and image storage formats should permit utilization of the more common hardware and software products
- in planning the systems for archive interrogation and for the presentation of archive search results, due account should be taken of those users with little or no computer literacy.

To this list the Vatican Library added three further recommendations:

- the reproduction of documents should never jeopardise the integrity of the originals
- access to the electronic archive thus constituted should also be possible in the Library's reading rooms
- the project should take appropriate steps to safeguard the intellectual property rights of the electronic reproductions.

With the intervention of a team of IBM computer experts, who developed the appropriate technologies and prepared the necessary programmes, the go-ahead was given to an initial campaign of reproductions. In its first phase, between 1994 and 1995, this permitted the digitalization of 60 manuscripts in their entirety and a further 60 in part, comprising a total of approximately 20,000 pages. The digital scanning continued in the following years, and over a hundred further manuscripts were reproduced in electronic form.

The decision to reproduce the manuscripts in their entirety, and not to limit reproduction to individual details, such as title pages, miniatures and the like, was based on the need to realise a genuine electronic replica of the Library's documentary holdings. The aim of the operation is to permit a scholar hooked up by terminal from a remote site to carry out the same research that he or she would have been able to conduct if physically present in the manuscript consultation room. Another more general aim is to produce a digital system addressed not to a specific category of users, for example art historians, but to all those who have occasion to frequent the Vatican for reasons of study. Provision must also be made for the establishment of an image archive that (by its very nature) is destined to grow: each time a new manuscript is added to those already reproduced in electronic format a proportionate number of images will be added to it. A plausible estimate of 400 images per manuscript would produce a total of 20 million images, first to be acquired in electronic form and then to be stored, assuming it is decided to enter into the electronic archive at least two thirds of the 75,000 library manuscripts proper. So the electronic storage capacity must be commensurate.

0

For what reason, we may ask, should valuable resources in terms of personnel, money and time be allocated to an operation which might superficially appear as no more than a glorified form of photographic duplication? And what possible uses might arise from a database in which tens of thousands – perhaps one day even millions – of pages of manuscript are collected? Apart from such arguments as the facility of accessing the archive at a distance and the safeguard of unique documents for which the process of cataloguing is still far from having been completed, other reasons can be adduced to justify the utility of digitalization. In particular, there exist developments of a certain importance in the scientific study of manuscripts, in particular in palaeography, as well as in codicology and philology, on which it is worth adding a few remarks.

As regards palaeography, there are two main reasons why the digitalization of images seems to open up interesting prospects: first, the electronic visualization and analysis of documents and, second, the collecting of large quantities of images in electronic format. Palaeography is in fact based on the principle of comparing scripts whose provenance and/or dating are known with other scripts for which this information is lacking. It is therefore obvious that the more examples of script are available, and the simpler it is to produce their simultaneous presentation, the greater will be the facilities at our disposal for reconstructing the transmission of the manuscript we are studying. Ever since palaeography began to be developed as a science with the investigations of Jean Mabillon in the later seventeenth century, the objective of scholars has been to have at their disposal as large a number as possible of plates of examples of scripts which would enable them to date, locate and study manuscripts. The first such plates were in the form of engravings produced by hand on the basis of copies drawn as accurately as possible from the originals. The advent of photography meant that the study of palaeography passed to a type of reproduction which was immediately considered more faithful and reliable, in spite of the really miraculous perfection achieved by manual reproductions. Photography was considered, not unjustly, as more objective, in that it did not depend in any way on the skill or idiosyncrasy of the draughtsman and ensured considerable homogeneity in the treatment of reproductions produced from different manuscripts in different libraries. Photography has, moreover, made great progress in recent years, through the application of highly sophisticated photographic techniques. It is enough to think of the use of special lights or coloured filters, x-rays and other electromagnetic techniques, such as betagraphy which permit penetration into the very make-up of the paper (to reproduce the watermarks), or of the parchment (to read any cancelled texts or texts rendered illegible by pathogenic agents). The application of digital scanning to manuscripts represents another major step forward: it permits manuscript analysis to be pushed to hitherto unattainable levels, without the quality of the image being thereby compromised. By analysing the colour of the individual point, for example, in combination with its wave length, we are able to filter the image in virtual mode, i.e. remove a certain wavelength from it; this may permit the removal of an ink stain or traces of

humidity from the reproduction of the folio. In our experimentation of this technology we have been able to "restore" numerous documents in virtual mode, and so give them a degree of legibility a great deal higher than that of the originals. And it goes without saying that this whole process does not in the least way alter the original document: hence the term 'virtual restoration'. A very similar technique may be applied to the reading of palimpsests, i.e. to those manuscripts which were used more than once: that is to say, by first rubbing out the original script and then superimposing over it a later script. In the case of Vat. grec. 2061 the palimpsest turned out to be double. The original text of the 5th century (Strabo) had been cancelled in the 7th/8th century, to make way for a juridical canon, which in turn was superimposed two centuries later by a work of Gregory Nazianzen. Thanks to the scanning techniques used (infrared and ultraviolet), the three scripts were separated and all three made available in legible format, the intermediate one for the very first time.

0

To make all this possible, the simple scanning of the documents is not enough. A large and sophisticated array of instruments and programmes is needed to facilitate not only the reproduction of manuscripts, but their accessibility in easily searchable formats, using a simple and uniform language. To this end various projects were formulated and are still in progress. One I would particularly like to mention is the "Digital Scriptorium", a joint project of the University of California at Berkeley and Columbia University in New York, which is aimed at reproducing in digital form over ten thousand manuscripts and subsequently making them available to the public via Internet. Another project has been aimed at defining the international standards for the specific requirements of the description of manuscripts: this is the project known as EAMMS (Electronic Access to Medieval Manuscripts). After first identifying the necessary criteria, it is now preparing the applications in MARC and SGML formats, also with the help of a special work group in the framework of TEI (Text Encoding Initiative). At a later stage, availing itself of the descriptive criteria established by EAMMS, the MASTER (Manuscripts Access through Standards for Electronic Records) European project, a partnership between a number of leading European libraries, will make available on the Web a database containing thousands of manuscripts in electronic format. Due to the great interest that both projects have aroused, the Vatican Library has lent its support by ensuring the active participation of its own representatives in the various work groups, and by making available a group of manuscripts on which to conduct the experiment.

While the application of technologies to the reproduction and diffusion of documents enables some problems to be solved, it also poses a new series of questions, not all of them satisfactorily resolved. I have already alluded to one such problem: the huge memory capacity needed for the storage of the images that a digital library project entails. Unknown quantities are the duration in time of the electronic supports and the consequences of changes, even minimal, in the standards on electronic archives, which have in the meantime become of mind-boggling extent. I would like lastly to recall here the aspect of the protection of copyright. The importance of this issue stems from the great facility of copying digital images, once these freely circulate on the Web. There are those who seek to counter this phenomenon by putting into circulation only very compressed images, with a resolution so low as to preclude reproduction and to permit only on-screen visualization. Various encrypting techniques are also being studied. One such technique being studied in the Vatican is the possibility of protecting each image with an electronic watermark, such as not to impede the reading of the document, but to render it unusable for purposes of publication, unless the electronic watermark is subsequently cancelled by the application of

complex algorithms. Clearly the protection of copyright is not a question limited to manuscripts; any kind of digital image could be subjected to unauthorized reproduction. In the case of manuscripts, however, it is more difficult to reach an acceptable compromise between the visualization of a poor-quality image and the need to guarantee in all cases the legibility of the script and of any other element worthy of consideration and study.

۲

Round this dilemma revolves a question that each institution that is a depository of a part of the world's cultural heritage is bound to pose: in what way should knowledge of this heritage be fostered and disseminated, without this putting at risk the very survival of the often invaluable, and always irreplaceable, testimonies on which it is based? If today's present is yesterday's future, on us is placed the obligation to transmit intact to tomorrow everything our predecessors have realized with such great skill and immeasurable love. The solution of placing all documents under lock and key, and rendering their study practically impossible, can in no case be valid. All too numerous are the books and manuscripts, of which all trace have been lost as a consequence of the fact that the only exemplars of them were jealously guarded, and hence inaccessible to study, only to be destroyed as a result of the periodic accidents that bedevil human history: fire, plunder, war, earthquake, etc. However paradoxical it may seem, the best way of ensuring a document's survival is to guarantee its accessibility to the viewing and study of a large number of people, to ensure that they, in turn, may, by their own work, transmit their testimony of it to posterity. The judicious application of information technology opens the way to "sustainable development" in the use of documents: not only does it reduce the need for the direct use of the originals, but enables multiple copies to be made and disseminated from these originals without causing the least damage to them. Where the reproductions are sufficiently accurate, it also promotes the objective of the advancement of learning. If - thank goodness - the day is still far distant when scholars will renounce coming to the Vatican to consult ancient manuscripts in person and so experience the indescribable thrill of having below their very eyes the writings of countless humanists and illustrious men who have shaped our past, we are convinced that by diffusing the use of electronic techniques in the management of our cultural heritage we are remaining faithful, even on the threshold of the third millennium, to the task that Pope Nicholas V enjoined on all those who, ever since the foundation of the Vatican Library, have succeeded each other in the curatorship of its holdings: namely, that of ensuring that the library serve "the common convenience of learned men".