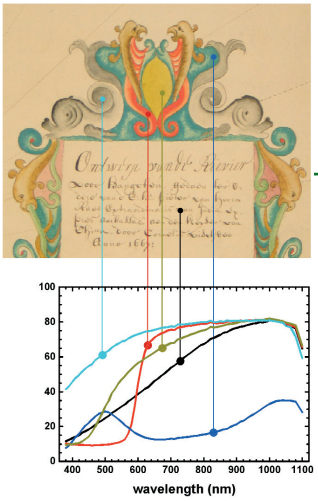
International Preservation News

A Newsletter of the IFLA Core Activity
on Preservation and Conservation

IFLA PAC

No. 40
December 2006

Contents



4 **The Quantitative Hyperspectral Imager A Novel Non-destructive Optical Instrument for Monitoring Historic Documents**
M.E. Klein, J.H. Scholten, G. Sciutto, Th. A.G. Steemers, G. de Bruin

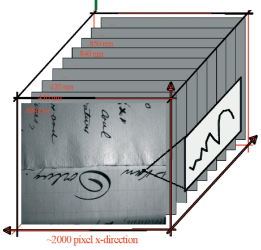
16 **Des identifiants pérennes pour les ressources numériques : l'expérience de la BnF**
Emmanuelle Bermès

27 **Web Archiving at the BnF**

35 **News**

36 **Publications**

37 **Events and Training**
- *Announcement*
- *Reports*



INTERNATIONAL PRESERVATION NEWS
N° 40
December 2006

ISSN 0890 - 4960
International Preservation News is a publication of the International Federation of Library Associations and Institutions (IFLA) Core Activity on Preservation and Conservation (PAC) that reports on the preservation activities and events that support efforts to preserve materials in the world's libraries and archives.

IFLA-PAC
Bibliothèque nationale de France
Quai François-Mauriac
75706 Paris cedex 13
France

Director:
Christiane Baryla
Tel: ++ 33 (0) 1 53 79 59 70
Fax: ++ 33 (0) 1 53 79 59 80
E-mail: christiane.baryla@bnf.fr
Editor / Translator
Corine Koch
Tel: ++ 33 (0) 1 53 79 59 71
E-mail: corine.koch@bnf.fr
Spanish Translator: Solange Hernandez
Typewriting: Isabelle Fornoni
Layout and printing: AXPRO, Paris

PAC Newsletter is published free of charge three times a year. Orders, address changes and all other inquiries should be sent to the Regional Centre that covers your area. See map on last page.

ISO 9706

editorial

The year 2006 ends with the publication of "International Preservation News" n° 40. For administrative reasons, we've had to complete very urgently this last issue of the year that is available today on the IFLA website before being printed and delivered in January. This editorial allows me to propose to our readers to develop in the future a different presentation of our information. 'Online' means immediacy whereas the printed copy is available with a delay that is more or less important according to the region delivered. The IFLA website should be restructured in Spring 2007; then, it will be easier for us to update it. The paper copy will be devoted to basic articles, book reviews and conference reports. On the worldwide Web, we will have the opportunity to add, regularly and more easily, short-lived announcements and news and first of all, to propose the links to other online publications.

The content of this last issue of the year might seem rather heterogeneous: however, it is mainly dedicated to topics relating to the digital field. First, several researchers working with the National Archives of The Nertherlands present a new tool used to measure the level of paper deterioration; this tool will certainly be used for other applications in the future. Then, the article by Emmanuelle Bermès entitled « Persistent Identifiers for Digital Resources: The Experience of the National Library of France » analyses the problems linked to Web archiving. Eventually, the last paper summarizes the major points of Web archiving at the Bibliothèque nationale de France. This article is the first of a series that will propose a state of the art of the situation in other countries.

On the other hand, I wish 2007 to be a year to rethink the editorial choices we take for "International Preservation News". Our newsletter, that will keep on being published three times a year, should be more clearly based on specific themes and follow the orientations defined by PAC and IFLA Strategic Plans. This is why the next issue will focus on preventive measures. Several papers by experts will allow us to wonder about the risks of contamination and damages that still exist in new facilities that are *a priori* protected against the dangers usually threatening our traditional libraries. The next issue should focus on training.
2007 will also be the year for regional meetings. In January, I will attend in China a meeting gathering the Chinese, Japanese and Australian PAC Centers. In April, I hope sincerely that the conference organized in Santiago by the IFLA Section on Newspapers and the PAC center in Chile will allow me to gather PAC directors from Latin... and North America. In August, we will meet our colleagues from Africa, in Durban. Parallel to these meetings, I think it is important to publish regularly more numerous articles about the regions covered by PAC regional centers.

Sending all of our readers my best wishes for Christmas and the New Year, I am also glad to announce the publication online of the proceedings of the Symposium entitled "The 3D's of Preservation" that took place in Paris in March 2006.
Please see at: <www.ifla.org/VI/4/ipi.html>

Christiane Baryla
IFLA-PAC Director

éditorial

© DR.

Avec 2006 qui s'achève, paraît le numéro 40 de nos « International Preservation News ». Cette dernière livraison de l'année, que nous avons dû boucler rapidement pour des raisons administratives, paraît aujourd'hui en ligne sur le site Web de l'IFLA. Elle sera imprimée et diffusée en janvier prochain. A l'occasion de cet éditorial, je voudrais proposer à tous nos lecteurs d'évoluer, pour l'avenir, vers une présentation différente de nos informations. Qui dit « online », dit immédiateté ; l'imprimé, en revanche, vous est adressé avec un retard plus ou moins important selon la région du monde desservie. Le site de l'IFLA devrait être réaménagé au printemps 2007 : il nous sera dès lors plus facile de le mettre à jour. Nous réserverons le papier aux articles de fond, aux présentations d'ouvrages et aux comptes rendus de colloques. Sur la toile, nous pourrons ajouter régulièrement et plus souplement des annonces et des chroniques éphémères et surtout produire les liens avec d'autres publications en ligne.

Le contenu de cette dernière livraison pourra sembler un peu hétérogène : le numéro est dédié principalement, mais très généralement, à des sujets qui ont à voir avec le numérique. Ainsi un groupe de chercheurs liés aux Archives nationales des Pays-Bas nous présente-t-il un nouvel instrument d'analyse de l'état de dégradation du papier, outil qui verra sans doute bien d'autres applications dans le futur. Puis, ce sont des problèmes associés à l'archivage électronique du Web qui sont explorés par Emmanuelle Bermès dans un article sur « Les identifiants pérennes des ressources numériques ». Nous terminons par une synthèse de l'état d'archivage du Web à la Bibliothèque nationale de France, premier épisode d'un feuilleton qui proposera l'état de l'art dans d'autres pays.

Je souhaiterais par ailleurs que l'année 2007 marque une inflexion dans les choix éditoriaux de notre revue. « International Preservation News », que nous allons publier au même rythme (trois numéros par an), devra être plus clairement thématique et suivre les directions données par les plans stratégiques du PAC et de l'IFLA. Aussi le prochain numéro traitera-t-il de conservation préventive. Nous nous interrogerons, en compagnie de spécialistes, sur les risques d'infestation et de dégâts qui demeurent dans un bâtiment neuf *a priori* protégé et hors des dangers habituellement courus par nos bibliothèques traditionnelles. Le numéro suivant devrait traiter de formation.
2007 sera aussi l'année des réunions régionales. Je me rendrai en Chine en janvier pour assister à une réunion PAC Asie avec les centres PAC chinois, japonais et australien. En avril, à l'occasion de la Conférence organisée à Santiago par la Section des journaux de l'IFLA et le PAC Chili, j'espère bien pouvoir réunir les directeurs PAC d'Amérique du Sud et… du Nord. En Août, à Durban, nous retrouverons nos collègues africains. Parallèlement à ces réunions il me paraît souhaitable de publier régulièrement et en bien plus grand nombre des articles concernant les régions irriguées par les centres régionaux PAC.

En souhaitant à tous nos lecteurs une excellente fin d'année 2006, j'ai le plaisir d'annoncer aussi la publication en ligne des Actes du Symposium de Paris sur « La Conservation en trois dimensions ».
Rendez-vous sur <www. ifla.org/VI/4/ipi.html>.

Christiane Baryla
Directeur d'IFLA-PAC

The Quantitative Hyperspectral Imager A Novel Non-destructive Optical Instrument for monitoring Historic Documents


© DR

by M.E. Klein Doctor in Physics


© DR

J.H. Scholten Mechanical Engineer


© DR

G. Sciutto Conservation scientist


© DR

Th. A.G. Steemers Head of Conservation and Restoration (National Archives of The Netherlands)


© DR

G. de Bruin Senior Conservation Consultant (National Archives of The Netherlands)

The National Archives of The Netherlands have commissioned the development of an advanced optical device for a quantitative analysis of historic documents. The quantitative hyperspectral imager measures optical reflectance spectra at millions of different points simultaneously. This can be exploited to identify pigments and reproducibly measure location-dependent differences in paper yellowing.

Introduction

In recent years, major public archives and libraries worldwide have been redefining their societal roles: traditionally being institutions dedicated to the conservation of documents of cultural value, their role as information providers has gained considerable importance. The scope of making our cultural heritage accessible has widened and acquired a more educational character, as it addresses not only specialists such as historic researchers, but also the general public.
As a result of this transition, the National Archives in The Netherlands are being confronted with the challenges of balancing two fundamentally contradicting tasks: the first task is to manage the collections for optimum conservation of the objects, which typically means that they are locked away and stored in a well-conditioned environment. The second task, however, is to make the objects more easily accessible to the public, and in particular to satisfy the increasing demand for object loans to public exhibitions in The Netherlands and abroad. In addition, the National Archives in cooperation with the Dutch Royal Library in The Hague have established their own permanent exhibition space, called 'De Verdieping', where continuously a selection of top pieces of both institutions is on display.

With each loan and exhibition the constant struggle between conservation and presentation becomes more apparent. Although there are general guidelines for the exhibition and transportation protocols concerning light levels, total light exposure, humidity, temperature, vibration levels, etc., the applicability of these guidelines and the validity of certain limit values for the particular cases at hand is discussed frequently and often very controversially. The fundamental reason for these ever-repeating discussions is that there is no generally accepted set of quantitative measures that reliably describes the overall condition of a historic document. Consequentially, the influence of any particular combination of intrinsic (object material, paper acidity, etc.) and environmental parameters (humidity, illumination level, etc.) on the object condition cannot be quantified, either. Therefore, most arguments brought forth when negotiating the needs of conservation and presentation have actually a fairly weak basis, which, however, does not necessarily mean that a consensus is found easily.
A major problem is the verification of the regulations on which the conservation and the presentation parties could finally agree, and the actual effect that the fulfillment (or the non-fulfillment) of these regulations has on the object condition. The effort to safeguard the objects on loan and display places an enormous workload on the conservation staff, as they have to prepare extensive condition reports before and after exhibition or transport. It is hoped that by comparing these condition reports one can detect any object damage resulting e.g. from an accidental maltreatment of the object that may involve an insurance case.

Conservators take great care in compiling such condition reports, and make use of a number of modern instruments such as digital cameras for documenting the condition of an object at a given time. However, conventional methods of evaluating the condition of an object are not at all satisfactory, because they provide only qualitative results, the interpretation of which is highly subjective and thus very difficult to reproduce and to compare with earlier results. The National Archives have recognized the need for developing techniques and instruments with which aspects of the object condition can be measured and recorded in objective ways.
Any such instrument has to fulfill a catalogue of requirements in order to be suitable as a practical technique for monitoring the condition of historic documents and similar objects. First of all, it has to be completely non-destructive so that measurements can be repeated frequently without any risk of inducing or accelerating degradation processes. Secondly, the results should be quantitative, objectively reproducible, and therefore comparable with measurements on the same object and on different objects. Since the objects in general have a quite inhomogeneous surface, the measurement has to be done with sufficient resolution, in order to make sure that in a later measurement, the relevant locations on the object can be re-found reliably. Thirdly, the instrument has to be time-efficient, so that it can be used as a standard method for measuring a reasonable number of objects that e.g. return from an exhibition.

We have developed a so-called quantitative hyperspectral imaging instrument for truly non-destructive paper and writing durability research, which combines the high spatial resolution of a digital camera with the large number of wavelength bands required for high-resolution spectral reflectance and accurate colour measurements. The instrument is to be used for application studies aiming at reliable identification of different types of ink, early-detection of ink corrosion and local substrate discoloration, and an objective quantification of the resulting degree of document degradation.
In this contribution, we discuss the operating principle of this so-called hyperspectral imaging system and we present initial experimental results.

The Operation Principle of The Quantitative Hyperspectral Imager

The term 'hyperspectral imaging' (HSI) refers to the acquisition of a series of digital images at a large number of different, well-defined optical wavelengths in the ultra-violet, visible and near-infrared. By applying proper calibration procedures, the value of any pixel of a digital image of the series represents a precise measurement of the portion of light that is reflected from the corresponding location on the target at a particular wavelength. The result of a hyperspectral imaging measurement is a stack of such spectral reflectance images, which contains one image for each wavelength band and which is often called the hyperspectral data cube (see Fig. 1). For a given pixel coor-

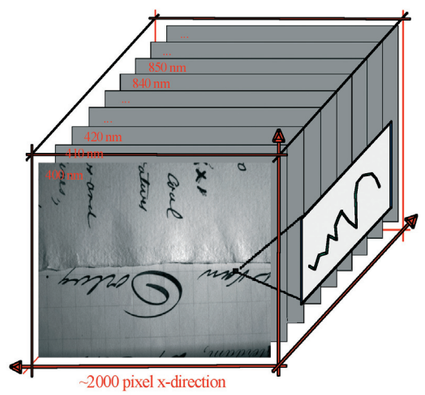


Fig. 1: The hyperspectral data cube. Each image of the data cube represents a precise measurement of the reflectance of the target object at a particular wavelength. For each pixel, the series of images corresponds to an entire spectral reflectance curve of the corresponding target point.

dinate, the sequence of pixel values in the images of the stack thus corresponds to an entire reflectance spectrum of the document at this location.
The quantitative hyperspectral imager is based on two identical wavelength tunable light sources (TULIPS) and a monochromatic CCD camera, as shown schematically in Fig. 2. The historic document or other object under investigation is illuminated by these two TULIPS under an angle of 45°, which ensures a homogenous light intensity and non-specular imaging by the camera under an angle of 0°. The entire setup is enclosed in a light-proof cabinet in order to avoid any stray light from external sources that would disturb the measurement.

Fig. 2: Photograph and setup scheme of the quantitative hyperspectral imager. The instrument is based on two identical wavelength tunable light sources (TULIPS), covering the spectral range from 365 to 1100 nm, and a monochromatic 4 megapixel CCD camera. The entire setup is placed in a light-proof cabinet that ensures a maximum quality of the measurement data by excluding any uncontrolled illumination of the investigated document by external light sources.

In this approach the spectral filtering required for a hyperspectral measurement is applied to the light illuminating the object, as opposed to the conventional approach where spectral filtering is applied to the light reflected from the object towards the CCD sensor. The great advantage of the new approach is that at any time during the measurement the object is illuminated only by the small portion of the light spectrum actually required at this time. This results in a drastic reduction of the total amount of light hitting the object, so that there is virtually no risk of any light-induced damage even of very fragile objects.

For a hyperspectral measurement, the TULIPS are tuned synchroneously through up to 70 different wavelength bands, ranging from the near ultra-violet (365 nm), via the entire visible into the near-infrared (1100 nm) spectral range. Over most of this wide spectral range, the center wavelength of the light emitted by the TULIPS can be tuned in small wavelength steps of 10 nm, while the spectral bandwidth ranges from 10 to 16 nm.
At each spectral band, the document is imaged by a high-quality monochrome digital camera, which is sensitive over the entire spectral range covered by the TULIPS. The camera has a resolution of 2048 x 2048 pixels (4 megapixels) and a field-of-view of 125 mm x 125 mm, so that each camera pixel represents an area of 60 µm x 60 µm on the document (resolution > 400 dpi). Due to this very high resolution, the instrument is capable of providing reliable measurement data even from within the areas of small features such as very thin lines of handwritings or prints.
After taking the series of up to 70 digital images from the document, data obtained from reference measurements is used to correct each of the images for imperfections of the light intensity distribution, transmission of the camera lens, and sensitivity of the CCD sensor. The result of this calibration procedure is the hyperspectral data cube mentioned before, which contains for each of the four million object points a spectral reflectance curve consisting of values from up to 70 different spectral bands. In total, when using all spectral bands, a hyperspectral imaging of a document provides about 550 megabytes of measurement data in digital format, which is then available for quantitative analysis, such as calculating CIE colour coordinates of substrate and pigments, etc.
In the following, we are going to describe two of the initial measurements that we performed with the quantitative hyperspectral imager, addressing the investigation of spectral curves of pigments and yellowing of paper, respectively.

Initial Experimental Investigations

1. Hyperspectral image of a historic map

As a first application, we made a hyperspectral imaging measurement of the legend of a Dutch map depicting the riverside of the city of Hangchiu (modern day Hangzhou). The map was drawn in 1667 by a cartographer of the Dutch United East India Company as a present for the Chinese emperor and it belongs now to the top pieces of the collection of the National Archives in The Hague (The Netherlands).

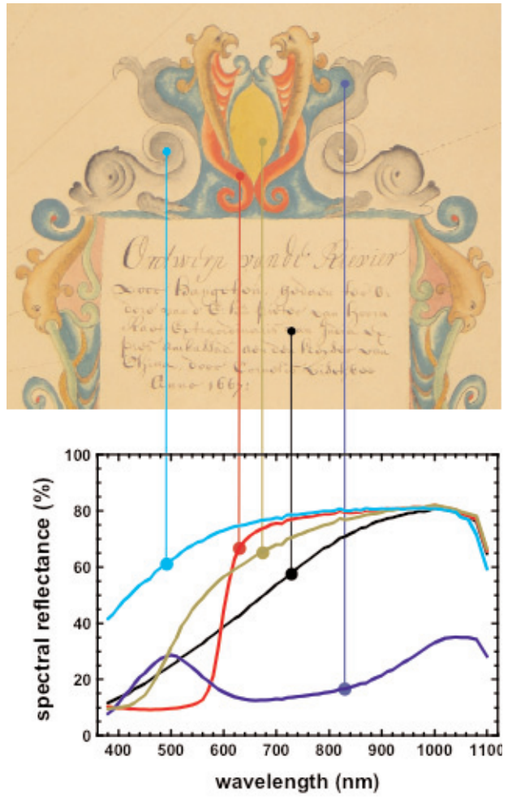


Fig. 3: Upper part: Photograph of the legend of a historic map, drawn in 1667 by a Dutch cartographer as a gift to the Emperor of China. The map belongs to the collection of the Nationaal Archief (The Hague, The Netherlands). Five regions of interest were defined in the marked areas of writing and illustration. Lower part: Average spectral reflectance curves of all pixels of the regions of interest, extracted from the hyperspectral data cube.

For illustration, the top of Fig. 3 shows a conventional colour image of the area of which we made a hyperspectral measurement that included all 70 spectral bands of the quantitative instrument. In this image, five regions of interest (ROIs) are marked lying within the ink writing and within various areas of the colourful illustrations. For each of these ROIs, the average spectral reflectance curves of all their pixels were calculated from the hyperspectral data cube of the measurement. The resulting spectral curves, which are shown in the diagram below the map in Fig. 3, show not only significant differences in the visible spectral range from 380 to 780 nm, which is responsible for the colour impression to the human eye. The curves show also different behaviour in the near-infrared range beyond 780 nm, which is not accessible by direct viewing or conventional colour photography. In particular, the curve belonging to the blue pigment area shows a much lower reflectance in the infrared than the other curves, which all increase, albeit with different slopes, towards the infrared. This means that as opposed to the other areas the blue pigment area does not become transparent in the infrared. For comparison, we measured the reflectance spectra of 9 different blue pigments (azurite, cobalt blue, manganese, indigo, Prussian blue, smalt, light and deep ultramarine) that were available as a preparation with the medium Paraloid[1] on canvas. The spectrum of the blue pigment of the map is most similar to the characteristic reference spectrum of the mineral pigment azurite, where residual differences of both spectra may originate from the differences in medium and substrate materials. Taking into account the period of time when the map was drawn, azurite is indeed a very likely candidate for the blue pigment used in the map.

2. Yellowing of paper

Yellowing of paper is one of the most important degradation processes, and therefore measuring the present levels and rates of yellowing is a very important task for investigating environmental influences when storing and exhibiting documents.
For fundamental research on this subject, the National Archives store reference paper sheets under different conditions for several years, and measure periodically the progress of yellowing. For the first time, the quantitative hyperspectral imager was used to measure and compare the local variation of the yellowing of papers that had been stored for several years under different conditions. In the following, we discuss the results obtained with two sheets of acid mechanical pulp paper which have been stored in the same depot since 1997 at the National Archives: one sheet as part of a bound volume and the other sheet stored in a special cardboard box.

Yellowing, or in general a discoloration of paper seen by the human observer, corresponds to a change of the visible part of the spectral reflectance curve of the corresponding area on the paper. Fig. 4 shows for two areas on the bound paper sheet the entire spectral reflectance curves, which were extracted from a single hyperspectral measurement with the quantitative instrument and include not only the visible but also the near-infrared spectral region. As compared to the first paper area in the sheet centre, the area at the border of the sheet clearly exhibits yellowing indicating the first stages of degradation. The reflectance curves

1. Provider Paraloid: Kremer Pigmente GmbH & Co. KG, Hauptstraße 41-47, D-88317 Aichstetten/Allgäu.

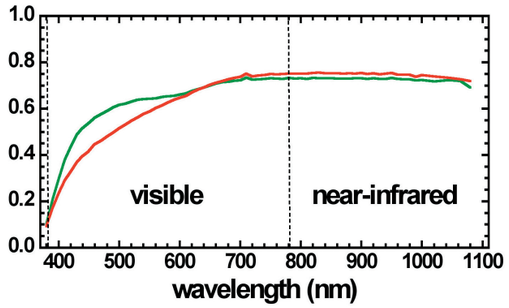


Figure 4: Spectral reflectance curves of two areas of a paper sheet stored in bound form in a special depot of the National Archives as a reference sheet for researching paper degradation. Green curve: spectral reflectance of an area in the center of the sheet. Red curve: spectral reflectance at the border of the sheet. In the blue-green visible spectral range (<540 nm wavelength), the red curve is significantly lower than the green curve, which is seen by the human eye as a darkening and yellowing of the corresponding border area as compared to the area in the sheet center.

of both areas are fairly flat and have similar values in the near-infrared ranging from 780 to 1100 nm. However, over most of the visible spectrum (380-780 nm), the more yellow paper area has significantly lower reflectance values, in particular in the green and blue spectral range at wavelengths < 540 nm. By the human eye this reduction of the blue-green reflectance values is seen as the typical darkening and yellowing of the paper colour indicating degradation.

Changes in colour and brightness of papers (and other objects), as seen by the average human observer under well-defined conditions of object illumination and viewing, are expressed precisely by the system of XYZ colour coordinates defined by the Commission internationale de l'éclairage (CIE).[2] The three XYZ colour coordinates can be calculated from the spectral reflectance curves resulting from the hyperspectral measurement.

Based on the XYZ colour coordinates the Deutsche Industrienorm (DIN) 6167 defines the so-called yellowness index,[3] in order to facilitate the comparison of papers. The yellowness index is a single value that measures in particular the yellowness of paper independent from its apparent brightness. From the hyperspectral data cube of each of the two (bound and loose) paper sheets, the yellowness indices on all locations of the papers were calculated, and the results were rendered into the two false-colour images shown in Fig. 5. In these images, paper areas with a low yellowness index have a green colour, whereas paper areas with a high yellowness index have a red colour. Both paper sheets have basically the same low yellowness index in the sheet centre, however, the about 20 mm wide border area of the bound paper (Fig. 5B) shows a much stronger increase of the yellowness index than the border area of the loose paper (Fig. 5A).

For a more quantitative comparison, for each paper the variation of the yellowness index along a cross section (as indicated in Fig. 5) was determined. In Fig. 6 the green and the red curves, corresponding to the loose and bound paper sheet, respectively, fluctuate around the same yellowness index of about 19 inside the paper sheets. For both papers, an approximately 20 mm wide border area can be identified, where the yellowness increases drastically. For the bound paper, however, this increase is much stronger than for the loose paper, reaching a value of about 38 at the sheet border, as compared to the value of about 30 for the loose paper.

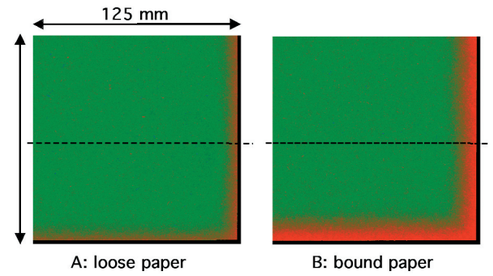


Fig. 5: False-colour images visualizing variation of the yellowness index over the paper sheet for A: loose and B: bound paper. Green colour corresponds to a low, red colour to a high yellowness index. For the paper stored in a bound volume the increase of the yellowness index towards the border of the paper is much stronger than for the papers stored as loose sheet. The dashed lines indicate where cross sections were extracted.

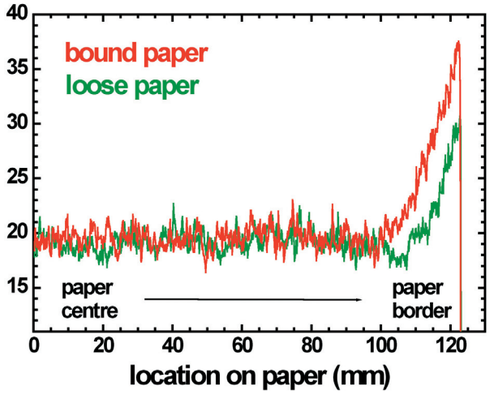


Figure 6: Cross sections of the yellowness index from the centres towards the border of the paper sheets. Red curve: yellowness index of bound paper. Green Curve: yellowness index of loose paper. Yellowing occurs for both types of paper about 20 mm from the border, however, the increase in the yellowness index is much stronger for the bound paper.

2. R.W.G. Hunt, *Measuring Colour*, Fountain Press Kingston-upon-Thames England, Third Ed., 1998, 344 pages.
3. DIN 6167, "Beschreibung der Vergilbung von nahezu weißen oder nahezu farblosen Materialien", January 1980.

A close inspection of the false-colour images in Fig. 5 reveals that locally strong variations of the yellowness index occur also in the inside areas of both paper sheets, seen as small red spots on a green background. These variations of the measured yellowness are caused by the inhomogeneity of the mechanical pulp paper, which contains clearly visible bits of wood of up to several millimeters in size. In such cases of an inhomogeneous substrate, measurements of the yellowness or colour with a conventional photospectrometer typically suffer from an unsatisfactory reproducibility of representative results. This is because with such instruments, measurements can be performed at only a few locations, and each measurement yields values averaged over an often not very well-defined small area. As opposed to the conventional single-point instruments, the hyperspectral imager measures with high resolution at millions of points simultaneously. This is a great advantage, because the local variations of the paper colour can be accurately measured and taken into account during the data analysis, so that the measurements should become much more reproducible and comparable.

Conclusion

In summary, we report on the development of the quantitative hyperspectral imaging instrument, which is dedicated to research in paper and writing durability. The wavelength-dependent reflectance is measured with high spatial resolution by illuminating documents with a wavelength tunable, monochromatic light source and imaging them with a high-resolution digital camera. The advantage over alternative approaches is that the total light intensity on the document is minimal at any time, and that there is no risk of heat or UV radiation induced damage.

As a first demonstration, we performed a hyperspectral measurement of the legend of a valuable 17th century map. From the hyperspectral data, the average spectral reflectance curves of a number of freely-defined regions of interest were extracted. These curves show significant variations not only in the visible spectral range, but also in the near-infrared.

In a second experiment, we demonstrated the use of the instrument for measuring and comparing the local variation of the yellowness index of reference papers that had been stored in a bound volume and as loose sheets, respectively. The cross section curves of the yellowness index and its visualization by false-colour images show that, for both papers, yellowing occurs in an about 20 mm wide area at the sheet border. The increase of the yellowness index in this area towards the sheet border was found to be much higher for the bound paper than for the loose paper.

In conclusion, the initial investigations reported here demonstrate the advantages of measuring the optical reflectance spectra, and thus the local colour coordinates and yellowness, of historic documents with high spatial resolution. We believe that the quantitative hyperspectral imaging instrument has a large potential to become a standard research tool for a fast, non-destructive analysis of historic documents.

Authors

M.E. Klein, J.H. Scholten
Art Innovation BV
Zutphenstraat 25
7575 EJ Oldenzaal
The Netherlands
Phone: +31 541570720
Fax: +31 541570721
E-mail: info@art-innovation.nl
Website: <www.art-innovation.nl>

G. Sciutto
Microchemistry and Microscopy Art Diagnostic Laboratory
University of Bologna
Via tombesi dall'ova 55
48100 Ravenna
Italy
E-mail: giorgia.sciutto@studio.unibo.it

Th. A.G. Steemers, G. de Bruin
Nationaal Archief, Den Haag
Postbus 90520
2509 LM Den Haag
The Netherlands
Phone: 070-3315400
Fax: 070-3315540
E-mail: gerrit.de.bruin@nationaalarchief.nl

Sensor cuantitativo de imágenes hiperespectrales – Nuevo instrumento óptico no destructivo para monitorear documentos históricos

Este artículo presenta un informe sobre el desarrollo de un nuevo instrumento óptico, el sensor cuantitativo de imágenes hiperespectrales y los experimentos iniciales que demuestran el potencial del equipo para el análisis cuantitativo no destructivo de los documentos históricos. El instrumento ofrece una serie de imágenes digitales calibradas de un documento a 70 longitudes de onda óptica bien definidas distintas, en el espectro ultravioleta, visible y cercano a infrarrojo. Para cada uno de los cuatro millones de pixeles de una imagen, la medición hiperespectral contiene una curva completa de reflectancia espectral de un pequeño punto (de 60 micrones por 60 micrones) en el documento investigado.

Como primera aplicación del instrumento, se tomó la medida de la leyenda de un mapa histórico, trazado en 1667 por un cartógrafo holandés como regalo al Emperador de China. Para la medición, se extrajeron las curvas de reflectancia espectral calibrada en un número de regiones de interés definidas libremente.

En un segundo experimento, se midió hojas de papel blanco que habían estado almacenadas en lotes de hojas sueltas o encuadernadas bajo condiciones bien conocidas a fin de que sirvieran como referencia para los estudios sobre la descoloración del papel. Mediante el uso de, solamente, la parte visible de las curvas de reflectancia espectral local, se calcularon los valores estándares de color CIE. A partir de los valores de color CIE, se determinó el índice de amarillo local en cada papel, y se transformaron en una imagen falsa de color que muestra la distribución del amarillamiento.

En conclusión, los experimentos iniciales realizados con el sensor cuantitativo de imágenes hiperespectrales ya dan una indicación de sus potencialidades analíticas. Las futuras investigaciones para su aplicación que conducirán los Archivos nacionales de los Países Bajos estarán dirigidas a usar el instrumento para la cuantificación local exacta de los procesos de degradación del papel tales como el amarillamiento y la corrosión de tinta.

L'imageur quantitatif hyperspectral : un procédé optique novateur et sûr pour contrôler les documents anciens

par M.E. Klein, Docteur en physique,
J.H. Scholten, Spécialiste en ingénierie mécanique,
G. Sciutto, Expert en conservation,
Th. A.G. Steemers, Chef du service de conservation et de restauration des Archives nationales des Pays-Bas,
G. de Bruin, Expert en conservation, Consultant pour les Archives nationales des Pays-Bas.

Les Archives nationales des Pays-Bas ont choisi de développer un procédé optique sophistiqué pour effectuer une analyse quantitative des documents anciens. L'imageur quantitatif hyperspectral mesure la courbe de réflectance en des millions de points différents simultanément. Cette technique peut être utilisée pour identifier les pigments et mesurer de façon reproductible le jaunissement du papier qui affecte certaines zones du document plus que d'autres.

Introduction

Depuis quelques années, les principaux services d'archives et bibliothèques du monde entier redéfinissent leur rôle sociétal : ces institutions, traditionnellement chargées de conserver des documents de valeur culturelle, ont gagné une importance considérable en tant que pourvoyeurs d'information. Notre patrimoine culturel se veut aujourd'hui accessible et ceci de façon plus pédagogique dans la mesure où l'on s'adresse non seulement à des spécialistes, historiens par exemple, mais aussi au grand public.
Face à cette évolution, les Archives nationales des Pays-Bas sont confrontées à la nécessité de faire cohabiter deux missions fondamentalement opposées : la première implique de gérer des collections dans des conditions optimales de conservation, ce qui signifie normalement de mettre les objets sous clef et de les stocker dans un environnement bien conditionné. Cependant, la seconde mission consiste à rendre les objets plus facilement accessibles au public et en particulier, à satisfaire la demande croissante de prêts d'objets destinés à des expositions publiques aux Pays-Bas et à l'étranger. De plus, les Archives nationales en coopération avec la Bibliothèque royale, à La Haye, ont mis en place leur propre espace d'expositions permanentes, « De Verdieping », où un choix d'œuvres majeures des deux institutions est exposé de façon continue.

Chaque prêt et chaque exposition rendent plus évident le conflit éternel qui oppose conservation et présentation des œuvres. Certes, des directives existent qui préconisent des protocoles à suivre en matière d'exposition et de transport des œuvres (niveaux d'intensité lumineuse, durée totale d'exposition, hygrométrie, température, niveaux de vibration...) ; néanmoins, la possibilité d'appliquer ces directives et la validité de certaines valeurs limites dans les cas particuliers à traiter sont fréquemment remises en cause, souvent de façon très controversée. Ces discussions se répètent sans fin, essentiellement parce qu'on ne dispose pas d'un ensemble de mesures quantitatives standard pour décrire de façon fiable l'état général d'un document ancien. En conséquence, les effets de la conjonction spécifique de paramètres intrinsèques (matériau, acidité du papier, etc.) et environnementaux (humidité, niveau d'éclairage, etc.) sur la condition de l'objet ne peuvent même pas être mesurés. La plupart des arguments avancés lorsqu'on met en balance les mesures nécessaires de conservation et la présentation des œuvres sont donc en fait assez faibles, ce qui ne signifie pas pour autant nécessairement qu'un consensus soit aisément trouvé.
Un problème majeur consiste à vérifier les principes qui verraient finalement s'entendre les partisans de la conservation et ceux de l'exposition, et le résultat du respect (ou du non-respect) de ces principes sur la condition de l'objet. L'effort de sauvegarde des objets prêtés et exposés représente une charge de travail conséquente pour le personnel de conservation, parce qu'il faut préparer des rapports détaillés avant et après l'exposition ou le transport. La comparaison de

ces rapports d'état doit permettre de détecter tout dommage provoqué par exemple par une manipulation maladroite de l'objet, qui impliquerait une prise en charge par la compagnie d'assurance.

Les conservateurs rédigent ces rapports d'état avec grand soin et utilisent un certain nombre d'outils modernes tels que les appareils photo numériques pour illustrer la condition de l'objet à un moment donné. Néanmoins, les méthodes traditionnelles qui permettent d'évaluer la condition d'un objet ne sont pas toutes satisfaisantes parce qu'elles fournissent seulement des résultats qualitatifs dont l'interprétation est particulièrement subjective ; il est donc très difficile de les réutiliser et de les comparer avec de précédents résultats. Les Archives nationales ont reconnu la nécessité de techniques et d'outils grâce auxquels la condition de l'objet pourrait être mesurée et enregistrée de façon objective.
Ce procédé quel qu'il soit doit satisfaire un ensemble de conditions pour être utilisable comme technique pratique afin de contrôler l'état de documents anciens et d'objets similaires. Tout d'abord, il doit être totalement sûr et permettre d'effectuer des mesures de façon récurrente sans aucun risque de provoquer ou d'accélérer les processus de détérioration. Ensuite, les résultats doivent être quantitatifs, objectivement réutilisables et donc comparables avec des mesures sur le même objet et des objets différents. La surface des objets n'étant en général pas très homogène, la mesure doit être effectuée avec une résolution suffisante afin que l'on puisse retrouver ultérieurement de façon fiable les zones étudiées au préalable. Enfin, l'outil doit être efficace sur la durée, de façon à pouvoir être utilisé comme méthode standard pour mesurer un nombre raisonnable d'objets retournés après une exposition, par exemple.

Nous avons développé un outil dit d'imagerie quantitative hyperspectrale destiné à une recherche véritablement fiable sur la durabilité du papier et l'écriture ; cet outil associe la résolution spatiale performante d'un appareil photo numérique et le grand nombre de bandes spectrales nécessaires à une courbe de réflectance haute résolution et à des mesures précises de couleur. L'outil doit être utilisé dans le cadre d'études d'application visant à identifier de façon fiable les différents types d'encres, à détecter en amont la corrosion de l'encre et la décoloration locale du substrat et à mesurer de façon objective le degré de dégradation du document qui en résulte.
Dans cet article, nous évoquons le principe d'utilisation de ce système dit d'imagerie hyperspectrale et nous présentons les premiers résultats expérimentaux.

Le principe d'utilisation de l'imageur quantitatif hyperspectral

Le terme d'« imagerie hyperspectrale » (HSI) fait référence à la constitution d'une série d'images numériques sur un grand nombre de longueurs d'onde optique différentes et bien définies, dans l'ultraviolet, le visible et le proche infrarouge. Si l'on applique des procédures de calibrage appropriées, la valeur de chaque pixel d'une image numérique de la série représente une mesure précise de la portion de lumière réfléchie par la zone étudiée à une longueur d'onde spécifique. Une mesure par imagerie hyperspectrale se compose d'un faisceau d'autant d'images de réflectance spectrale, qui contient une image pour chaque bande spectrale et souvent appelé « cube de données hyperspectrales » (voir fig. 1). Pour un ensemble donné de pixels, la séquence des valeurs de pixels dans les images du faisceau correspond donc au spectre complet de réflectance du document à cet endroit.

L'imageur quantitatif hyperspectral repose sur deux sources lumineuses réglables de longueurs d'onde identiques (TULIPS) et un appareil photo CCD (Charge Coupled Device), comme le montre de façon schématique la figure 2. Le document ancien ou autre objet étudié est éclairé par ces deux TULIPS avec un angle de 45°, ce qui assure une intensité lumineuse homogène et une imagerie non spéculaire de l'appareil photo avec un angle de 0°. Le dispositif complet est enfermé dans une chambre noire afin d'éviter toute source lumineuse extérieure parasite qui pourrait fausser la mesure.

Dans cette approche, le filtrage spectral nécessaire à une mesure hyperspectrale est appliqué à la lumière qui éclaire l'objet, contrairement à l'approche conventionnelle selon laquelle le filtrage spectral est appliqué à la lumière reflétée par l'objet vers le capteur CCD. Le grand avantage de cette nouvelle approche, c'est que pendant toute la durée de la mesure, l'objet est éclairé seulement par la petite portion du spectre lumineux réellement nécessaire à ce moment. Par conséquent, la quantité totale de lumière en contact avec l'objet est considérablement réduite, si bien qu'il n'existe virtuellement aucun risque que la lumière endommage les objets même très fragiles.

Pour une mesure hyperspectrale, les TULIPS sont réglées de façon synchronisée jusqu'à 70 bandes spectrales, allant de l'ultraviolet (365 nm), en passant par l'intégralité du visible jusqu'au champ spectral du proche infrarouge (1100 nm). Au-delà de la majeure partie de ce large champ spectral, la longueur d'onde centrale de lumière émise par les TULIPS peut être modulée en petites mesures spectrales de 10 nm, tandis que la largeur d'une bande spectrale s'étend de 10 à 16 nm.

Pour chaque bande spectrale, le document est photographié par un appareil numérique noir et blanc de qualité supérieure, sensible au-delà du champ spectral global couvert par les TULIPS. L'appareil photo a une résolution de 2048 x 2048 pixels (4 mega pixels) et un champ visuel de 125 x 125 mm, si bien que chaque pixel de l'appareil représente une surface de 60 µm x 60 µm sur le document (résolution supérieure à 400 dpi). Grâce à cette très haute résolution, l'outil peut produire des données de mesure fiables même pour la surface de petits éléments comme les lignes très fines de textes manuscrits ou imprimés.
Après avoir pris jusqu'à 70 images numériques du document, les données obtenues à partir des mesures de référence sont utilisées pour corriger chaque image, défauts de distribution d'intensité lumineuse, de transmission de l'objectif et de sensibilité du capteur CCD. Le résultat de cette procédure de calibrage est le cube de données hyperspectrales mentionné plus haut, qui contient pour chacun des quatre millions de points de l'objet, une courbe de réflectance constituée de valeurs allant jusqu'à 70 bandes spectrales différentes. Au total, lorsque toutes les bandes spectrales sont utilisées, l'imagerie hyperspectrale d'un document fournit environ 550 méga-octets de données de mesures au format numérique, disponibles ensuite pour une analyse quantitative, pour calculer les coordonnées de couleurs CIE du substrat et des pigments, par exemple.
Dans le chapitre suivant, nous allons décrire deux des mesures initiales que nous avons effectuées avec l'imageur quantitatif hyperspectral, en concentrant nos recherches respectivement sur les courbes spectrales des pigments et le jaunissement du papier.

Premières recherches expérimentales

1. L'image hyperspectrale d'une carte ancienne

La première application de ce procédé a consisté à mesurer par imagerie hyperspectrale la légende d'une carte hollandaise figurant les bords du lac de la ville de Hangchiu (aujourd'hui Hangzhou). La carte a été dessinée en 1667 par un cartographe de la Compagnie néerlandaise des Indes orientales pour être offerte à l'Empereur de Chine et c'est aujourd'hui une des pièces majeures de la collection des Archives nationales de La Haye (Pays-Bas).

A titre d'exemple, le haut de la figure 3 montre une image couleur traditionnelle de la zone que nous avons mesurée grâce au système d'imagerie hyperspectrale comprenant les 70 bandes spectrales de l'outil quantitatif. Sur cette image, cinq zones d'intérêt (ROI) sont identifiées dans l'encre manuscrite et en différents points des illustrations aux couleurs vives. Pour chacune de ces ROI, les courbes moyennes de réflectance de tous les pixels ont été calculées à partir du cube de mesure de données hyperspectrales. Les courbes de réflectance qui en résultent, comme le montre le schéma au-dessous de la carte sur la figure 3, ne révèlent pas seulement des différences significatives dans le champ spectral visible qui s'étend de 380 à 780 nm et permet la perception de la couleur par l'œil humain. Les courbes montrent également un comportement différent dans la zone du proche infrarouge au-delà de 780 nm, qui n'est pas accessible à l'œil nu ou à la photo couleur traditionnelle. En particulier, la courbe relative à la zone de pigment bleu montre une réflectance très inférieure dans l'infrarouge aux autres courbes qui augmentent toutes, bien que différemment, en allant vers l'infrarouge. Cela signifie que, contrairement aux autres zones, la zone de pigment bleu ne devient pas transparente dans l'infrarouge. A titre de comparaison, nous avons mesuré sur la toile le spectre de réflectance de neuf pigments bleus différents (azurite, bleu cobalt, manganese, indigo, bleu de Prusse, bleu outre-mer clair et foncé) qui étaient disponibles sous forme de préparation avec le liant Paraloid. Le spectre du pigment bleu sur la carte est très semblable au spectre de référence caractéristique de l'azurite, pigment minéral ; dans les deux cas, les différences résiduelles des deux spectres peuvent provenir des différences de matériaux utilisés pour le support et le substrat. Si l'on considère la période à laquelle la carte a été réalisée, l'azurite est en effet très probablement le pigment bleu utilisé sur la carte.

2. Le jaunissement du papier

Le jaunissement du papier est l'un des plus importants processus de détérioration. Il est donc très important de mesurer le niveau et le degré actuels de jaunissement pour faire des recherches sur l'influence de l'environnement lors du stockage et de l'exposition des documents.
Pour procéder à des recherches fondamentales sur le sujet, les Archives nationales stockent des feuilles de papier test dans différentes conditions pendant plusieurs années et mesurent de façon périodique l'avancée du jaunissement. Pour la première fois, l'imageur quantitatif hyperspectral a été utilisé pour mesurer et comparer les différences dans les zones de jaunissement pour des papiers stockés pendant

plusieurs années dans différentes conditions. Ci-après, nous examinons les résultats obtenus sur deux feuilles de papier acide (fabriqué à partir de pâte mécanique) stockées dans le même dépôt depuis 1997 aux Archives nationales : une feuille appartenant à un ouvrage relié, l'autre stockée dans une boîte spéciale en carton.

Le jaunissement, ou d'une façon générale une décoloration du papier observable à l'œil nu, correspond à une évolution dans la partie visible de la courbe de réflectance de l'emplacement correspondant sur le papier. La figure 4 montre, en deux endroits sur la feuille de papier de l'ouvrage relié, l'intégralité des courbes de réflectance mesurées en une seule fois grâce à l'outil quantitatif hyperspectral et comprenant non seulement la région spectrale visible mais aussi le proche infrarouge. Comparée à la première zone située au centre de la feuille, celle qui se trouve sur le bord montre clairement un jaunissement qui révèle les premières étapes du processus de détérioration. Les courbes de réflectance des deux zones sont assez planes et montrent des valeurs similaires dans le proche infrarouge, qui s'échelonnent de 780 à 1100 nm. Cependant, au-delà de la majeure partie du spectre visible (380-780 nm), la zone de papier la plus jaune montre des valeurs de réflectance nettement inférieures, en particulier dans la zone spectrale du vert et du bleu, pour des longueurs d'onde inférieures à 540 nm. Pour l'œil humain, cette diminution des valeurs de réflectance du bleu-vert se manifeste par un aspect plus foncé et jauni du papier, typique de la détérioration.

Les changements dans la couleur et l'éclat des papiers (et d'autres objets) visibles par le témoin lambda dans des conditions bien définies d'éclairage et d'observation sont exprimés précisément par le système des coordonnées de couleurs XYZ définies par la Commission internationale de l'éclairage (CIE). Les trois coordonnées de couleurs XYZ peuvent être calculées à partir des courbes de réflectance obtenues par mesure hyperspectrale.
A partir des coordonnées de couleurs XYZ, la Deutsche Industrienorm (DIN) 6167 définit l'indice dit de jaunissement, afin de faciliter la comparaison des papiers. L'indice de jaunissement est une valeur unique qui mesure en particulier le jaunissement du papier indépendamment de son éclat apparent. A partir du cube de données hyperspectrales de chacune des deux feuilles de papier (la feuille de l'ouvrage relié et la feuille volante), les indices de jaunissement ont été mesurés en différents endroits et les résultats ont été présentés sous la forme des deux images aux couleurs arbitraires de la figure 5. Ces images montrent que les surfaces de papier ayant un faible indice de jaunissement sont de couleur verte, tandis que celles ayant un indice de jaunissement élevé sont de couleur rouge. Les deux feuilles de papier ont en principe le même indice de jaunissement faible au centre de la feuille ; néanmoins, la surface située au bord de la feuille de l'ouvrage relié, d'une largeur de 20 mm environ, montre (Fig. 5B) une augmentation beaucoup plus importante de l'indice de jaunissement que le bord de la feuille volante (Fig. 5A).

Pour permettre une comparaison plus quantitative, on a déterminé pour chaque morceau de papier, la variation de l'indice de jaunissement (comme l'indique la figure 5) suivant une coupe transversale. Sur la figure 6, les courbes verte et rouge, qui correspondent respectivement à la feuille volante et à la feuille de l'ouvrage relié, fluctuent autour du même indice de jaunissement (environ 19) au sein de chacune des feuilles. Sur les deux morceaux de papier, on peut identifier une zone d'une largeur de 20 mm environ située sur le bord, où le jaunissement augmente considérablement. Cependant, cette augmentation est beaucoup plus importante sur la feuille de l'ouvrage relié ; elle atteint une valeur de 38 environ sur le bord, contre une valeur de 30 environ pour la feuille volante.

Une observation attentive des images de couleur arbitraire de la figure 5 révèle que d'importantes variations de l'indice de jaunissement se produisent aussi localement sur la partie interne des deux feuilles de papier, sous la forme de petits points rouges sur fond vert. Ces variations de l'indice de jaunissement mesuré sont causées par l'absence d'homogénéité du papier fabriqué à partir de pâte mécanique qui contient, clairement visibles, des morceaux de bois dont la taille peut atteindre plusieurs millimètres. Dans des cas semblables où le substrat n'est pas homogène, les mesures du jaunissement ou de la couleur avec un photo spectromètre traditionnel ne peuvent pas être réutilisées comme résultats représentatifs. En effet, avec ce genre d'appareils, les mesures peuvent être effectuées seulement à certains endroits, et chaque mesure produit des valeurs établies au-delà d'une petite surface souvent pas très bien définie. Contrairement aux appareils traditionnels qui mesurent un point unique, l'imageur hyperspectral effectue simultanément des mesures en des millions de points avec une haute résolution. C'est un grand avantage parce que les variations locales de la couleur du papier peuvent être mesurées précisément et prises en compte pour l'analyse des données, de façon à ce que les mesures puissent être réutilisables et comparables.

Conclusion

En bref, nous rendons compte du développement de l'outil d'imagerie hyperspectrale, utilisé dans le cadre de recherches sur la durabilité du papier et de l'écriture. La réflectance qui dépend de la longueur d'onde est mesurée à haute résolution spatiale en éclairant les documents avec une source lumineuse monochromatique dont la longueur d'onde est modulable, et en les photographiant avec un appareil photo numérique haute résolution. L'avantage sur d'autres procédés, c'est que la dose totale d'intensité lumineuse sur le document est minimale à tout moment et qu'il n'existe pas de risque de détérioration provoquée par la chaleur ou les UV.
Une première démonstration a consisté à mesurer par imagerie hyperspectrale la légende d'une précieuse carte du XVII[e] siècle. A partir des données hyperspectrales, les courbes moyennes de réflectance d'un nombre de ROI librement définies ont été établies. Ces courbes montrent des variations significatives non seulement dans la partie visible du spectre mais aussi dans le proche infrarouge.

Une seconde expérience a consisté à utiliser l'appareil pour mesurer et comparer la variation locale de l'indice de jaunissement des papiers-tests, feuille de l'ouvrage relié et feuille volante. Les courbes suivant des coupes transversales de l'indice de jaunissement et sa représentation par des images aux couleurs arbitraires montrent que pour les deux feuilles de papier, le jaunissement se produit sur une largeur de 20 mm environ, au bord de la feuille. L'augmentation de l'indice de jaunissement sur cette surface s'est révélée beaucoup plus importante pour la feuille de l'ouvrage relié que pour la feuille volante.

En conclusion, les recherches initiales rapportées ici démontrent les avantages qui consistent à mesurer le spectre optique, donc les coordonnées de couleurs locales et le jaunissement, sur des documents anciens à haute résolution spatiale.
Nous pensons que l'outil quantitatif d'imagerie hyperspectrale a de grandes chances de devenir un instrument de recherche standard pour une analyse rapide et sûre des documents anciens.

Des identifiants pérennes pour les ressources numériques : l'expérience de la BnF


© DR

par Emmanuelle Bermès
Responsable fonctionnel de la Bibliothèque numérique sur le Web, Bibliothèque nationale de France

Introduction

La création de ressources numériques en ligne, qu'il s'agisse de numérisation, d'une édition ou tout simplement d'un billet ou d'un commentaire de blog, soulève la question de l'identification fiable et durable de ces données sur le réseau. Les bibliothèques sont depuis longtemps déjà confrontées aux problèmes de numérotation, des numéros attribués au moment de la publication, comme les ISBN et les ISSN, aux cotes qui permettent de disposer et retrouver les livres dans la collection.[1] La question des identifiants n'a donc rien d'une nouveauté, mais comme dans de nombreux autres cas, la transposition des pratiques dans le domaine du numérique, et en particulier celui du Web, pose un certain nombre de problèmes : l'impossibilité de faire des distinctions claires par supports ou par genres, la gestion de versions et de granularité…

Pour qui est confronté aux problématiques de gestion des collections numériques, ces difficultés sont courantes ; les questions qu'elles soulèvent ont déjà été posées, et parfois résolues, souvent en faisant reposer ces solutions, d'une part sur les métadonnées, d'autre part sur les identifiants. S'il est clair que les identifiants sont nécessaires à tous les niveaux d'organisation d'un système de gestion des objets numériques, la façon de construire et de gérer ces identifiants soulève encore bien des débats.

Les identifiants sur le Web : objectifs, fonctions, systèmes

La première question qui se pose concerne la finalité des identifiants. La terminologie reflète déjà deux axes possibles. Le terme de référence stable, ou lien permanent, évoque la capacité d'un tiers à nommer et retrouver la ressource : autrement dit, la citabilité. Quand on parle d'identifiant unique ou d'identifiant de ressource, on est plutôt dans une problématique d'unicité et de pérennité dans le cadre d'une utilisation pratique.

La citabilité peut être facilitée par le choix d'identifiants dits sémantiques ou signifiants, c'est-à-dire qui portent en eux-mêmes du sens. Ce sens repose en général sur les métadonnées de la ressource qu'ils décrivent, par exemple son titre. Ce type d'identifiants a été assez largement adopté sur le Web, notamment par les outils de gestion de plateformes de blogs. Cependant, ces identifiants signifiants, s'ils facilitent la compréhension et le référencement de la ressource, peuvent poser un certain nombre de problèmes. Tout d'abord, si la nature de la ressource change, le lien sémantique entre l'identifiant et la ressource peut être brisé. De plus, les identifiants signifiants sont profondément liés à la structure des documents qu'ils décrivent, ce qui peut rapidement, dans le cas d'une collection de masse, poser des problèmes de cohérence et de généricité. Enfin, d'une langue à l'autre, d'une époque à l'autre, la signification d'un mot ou d'un sigle peut changer ; elle peut être explicite à une époque et incompréhensible un siècle plus tard ; elle peut être anodine à un moment donné et devenir offensante en moins d'une décennie. Lorsqu'on travaille sur le très long terme et à l'échelle internationale, il peut donc être important de réfléchir à ces contingences et d'envisager le choix d'un système de nommage opaque.

Les identifiants opaques sont en principe générés par des machines à l'aide de logiciels. Parmi les normes existantes, nous pouvons citer UUID[2] (Universally Unique IDentifier), un identifiant construit par un algorithme normalisé sur la base d'informations techniques (quel ordinateur le génère) et temporelles (à quel instant précis on le génère). De tels identifiants soulèvent une tout autre difficulté : la nécessité de préserver un lien entre l'identifiant et la ressource décrite, puisque l'identifiant ne porte pas, en lui-même, d'informations sur le contenu de cette ressource. Par ailleurs, l'automatisation des UUID a permis à des logiciels de générer des identifiants très longs, ce qui les rend impossibles à exploiter pour des utilisateurs humains.

Alors que les identifiants signifiants facilitent la citabilité et donc l'accessibilité des ressources, les identifiants opaques sont plus pérennes et préférables dans un contexte où l'on vise la préservation à très long terme de ces ressources. On retrouve bien dans cette dualité un problème qui a été éternellement celui des bibliothèques : le dilemme entre communication et conservation. Nous voyons donc que ces deux contextes d'usage possibles, préservation/archivage et consultation/indexation, constituent deux problématiques différentes pour les identifiants, et que le choix d'un système permettant de résoudre de manière idéale les deux problèmes à la fois risque de se révéler utopique. Il faut donc réussir à concilier ces deux aspects en prenant en compte les fonctionnalités des identifiants à des fins d'archivage et à des fins de consultation.

Fonctionnalités des identifiants

La réflexion sur les fonctionnalités des identifiants n'est pas nouvelle et on la trouve déjà, par exemple, dans la RFC 1737 intitulée « Functional Requirements for Uniform Resource Names », datée de 1994[3]. Ce document définit une série de fonctionnalités qui sont encore celles qui sont questionnées aujourd'hui.[4]

La première de ces fonctionnalités est l'unicité, un identifiant étant supposé caractériser une ressource et une seule. En retour, la même ressource, même située à différents endroits, devrait avoir le même identifiant. On parle alors d'identifiants « globalement uniques », ce qui peut supposer une organisation plus ou moins centralisée à l'échelle internationale.

Dans un second temps, on recherche pour ces identifiants globalement uniques la pérennité, c'est-à-dire une forme de garantie qu'ils ne changeront pas et continueront à identifier la même ressource. La pérennité est la clef de la stabilité de la référence et la principale problématique de l'utilisation des identifiants. Cependant, il est désormais reconnu que la pérennité n'est en rien un problème technique, et qu'elle n'est garantie par aucun système connu d'identification. C'est la gouvernance qui assure la pérennité, c'est-à-dire la pérennité de l'institution qui donne les identifiants. Les institutions ou acteurs appelés à durer peuvent ainsi devenir des « autorités nommantes », à l'échelle d'une organisation locale, d'un pays, ou du monde. Ces autorités nommantes sont dès lors détentrices d'un pouvoir, celui de nommer, mais aussi d'une lourde responsabilité, celle de pérenniser leurs identifiants.

L'indépendance de l'autorité nommante doit dès lors être discutée : pour une institution, il peut être raisonnable de déclarer que le système le plus pérenne est celui qui lui impose le moins de contraintes, ou celui dont les contraintes correspondent le mieux aux besoins de l'institution. Ainsi un établissement comme la Bibliothèque nationale de France accorde-t'il une grande valeur à l'indépendance à la fois technique et budgétaire de son système d'identifiants, car cette indépendance lui garantit une liberté de mise en œuvre qui va être favorable, dans ce contexte, à la pérennité. En revanche, un petit éditeur pourra préférer un système très contraint, car cette contrainte va lui apporter un confort technique (en lui fournissant des outils) et une sécurité globale (en proposant par exemple un système de continuité si l'éditeur disparaît) qui sont indispensables à une véritable pérennité.

Par ailleurs, la pérennité comme l'unicité posent un problème d'échelle. Il est désormais admis que des identifiants peuvent être réaffectés, voire détruits sur le réseau. La pérennité ne se définit donc pas par « éternellement », mais par « suffisamment longtemps » à l'échelle des besoins de l'institution qui gère la ressource. Pour prendre un exemple, un identifiant comme <http://www.lemonde.fr/> est tout à fait stable mais son contenu change quotidiennement. Peut-on alors parler d'identifiant pérenne ? Une distinction est à faire entre des ressources « abstraites », éventuellement mouvantes, et des ressources « concrètes », stables et uniques, toutes deux ayant besoin d'être identifiées. De même, il est nécessaire de définir l'échelle d'unicité recherchée : certains identifiants sont uniques au sein d'un système et conviennent parfaitement pour un usage interne, bien circonscrit. Un tel

1. LUPOVICI, Catherine, « Le Digital Object Identifier : Le système du DOI », *BBF*, 1998, n°3, p. 49-54, http://bbf.enssib.fr

2. A Universally Unique IDentifier (UUID) URN Namespace, RFC n°4122, http://www.ietf.org/rfc/rfc4122.txt

3. http://www.w3.org/Addressing/rfc1737.txt

4. Voir notamment *Persistent Identifiers*. University College Cork, Ireland, June 17-18, 2004, http://www.erpanet.org/events/2004/cork/Cork%20Report.pdf, *DCC Workshop on Persistent Identifiers*. University of Glasgow, 30 June – 1 July 2005, http://www.dcc.ac.uk/events/pi-2005/, ou encore *NISO Identifier Roundtable*, National Library of Medicine, March 13-14, 2006, http://www.niso.org/news/events_workshops/IDdocs/ ID-06-report.pdf.

usage n'impose pas les mêmes exigences que la diffusion en réseau, qui requiert quant à elle des identifiants globalement uniques plus complexes à gérer.

D'un point de vue technique, les systèmes d'identifiants doivent se montrer aptes à répondre aux besoins spécifiques qui ont été définis, comme nous venons de le voir, par chaque producteur en fonction de ses propres moyens.

Ainsi, les identifiants doivent être applicables à n'importe quel niveau de la ressource : la ressource elle-même mais aussi la collection dont elle fait partie, les articles qu'elle rassemble, et pourquoi pas, le paragraphe de l'article (ou le commentaire du billet), et également différentes versions d'une même ressource. Comme dans toute gestion de collection numérique, un choix initial est nécessaire entre granularité logique et physique. Il faut donc définir les différents niveaux de granularité de l'information qui doivent être identifiés, et comment ils vont se décliner dans le système d'identification : le choix peut aller de l'attribution d'identifiants complètement indépendants à chaque niveau (mais il faut alors gérer des liens entre ces niveaux d'identifiants, grâce à une carte de structure) jusqu'à un système hiérarchisé qui reflète l'organisation de la collection.

Les identifiants peuvent être capables d'intégrer des modèles préexistants pour le fournisseur qui les utilise : par exemple, les ISBN et ISSN, les cotes d'une bibliothèque, un système de nommage préexistant utilisé pour les URL ou les fichiers. La pérennité repose en outre sur la capacité à s'adapter aux changements de l'environnement, et il est nécessaire de pouvoir étendre les identifiants et les adapter au fur et à mesure de l'apparition de nouvelles ressources, des évolutions du réseau, des standards du Web, des capacités des navigateurs.

Enfin, les identifiants peuvent être actionnables, c'est-à-dire qu'ils fournissent une méthode pour accéder à la ressource. Aujourd'hui, seul le protocole « http », le protocole du Web, est à même de répondre à cette exigence, car il est compréhensible pour un navigateur, l'outil que nous utilisons aujourd'hui pour parcourir le Web. Pour rendre des identifiants qui ne sont pas basés sur le protocole « http » actionnables, différentes méthodes ont été utilisées comme la mise en place de résolveurs, ou de greffons/plug-ins que l'utilisateur ajoute lui-même à son navigateur. Ces résolveurs et plug-ins utilisent eux-mêmes les technologies liées au protocole « http » pour rendre les identifiants actionnables.

Le contexte de l'identifiant pérenne doit permettre de savoir à quoi celui-ci correspond et d'accéder à la ressource elle-même. Soit on dispose d'autres informations sur la ressource, incluant éventuellement sa localisation, sous forme de métadonnées, soit on consulte une base de référence qui va donner l'adresse correspondant à l'identifiant, en passant par un service de résolution. Une solution n'exclut pas l'autre ; on peut avoir un identifiant associé à des métadonnées en plus d'un résolveur qui va donner l'URL correspondante.

Certains systèmes d'identification pérenne demandent ou recommandent la saisie de métadonnées conjointement avec l'enregistrement de la ressource. C'est le cas, par exemple, de DOI et de ARK. D'autres systèmes sont dédiés à l'échange de métadonnées, mais incluent ou nécessitent un système d'identification pérenne pour accomplir leur rôle, qui est de donner accès à la ressource elle-même. Parmi ceux-ci, on peut citer le système d'identification intégré dans le protocole OAI-PMH.[5]

Le rôle du résolveur d'identifiants est de faire correspondre au nom de la ressource son adresse réelle. Le résolveur peut être interne à l'institution qui donne les noms, ou externe et géré par une autorité indépendante.[6]
La combinaison du type de résolveur et du type de métadonnées associés à chaque système d'identification va être un facteur de choix déterminant. Ces deux éléments constituent le cœur du système, qu'il faut confronter avec les fonctionnalités attendues : par exemple, la possibilité de gérer plusieurs niveaux de granularité, la simplicité des mises à jour, la gestion des versions différentes d'une même ressource, etc.

Les systèmes d'identifiants pérennes aujourd'hui

Les systèmes d'identifiants pérennes actuels reposent sur l'existence d'une autorité nommante mondialement reconnue, qui dispose de la liberté et de l'indépendance nécessaires pour attribuer à des ressources des identifiants pérennes, uniques et adaptables, que les navigateurs interprètent soit directement soit à l'aide d'un résolveur.

5. http://www.openarchives.org/OAI/2.0/guidelines-oai-identifier.htm
6. Ex. : la Library of Congress dispose de son résolveur qui fonctionne pour les identifiants DOI et Handle. On dispose d'un identifiant comme : doi:10.1045/january2005-fox, et il suffit de le faire précéder de l'adresse du résolveur pour accéder à la ressource : http://hdl.loc.gov/doi:10.1045/january2005-fox. Il s'agit d'une simple fonction de résolution ; la ressource prise en exemple n'a aucun rapport avec la Library of Congress.

Comparer ou évaluer les différents systèmes est une tâche complexe, dès lors que, comme nous l'avons vu, aucun n'est supérieur aux autres en termes de pérennité ; ce sont plutôt les besoins des producteurs de ressources qui sont à prendre en compte dans le choix d'un système. La plupart des systèmes combinent les différentes fonctionnalités citées ci-dessus à différents niveaux. Il n'existe pas de critère de différenciation clair et net. Ils ont chacun leur façon de concevoir les choses, de les organiser.

Les identifiants pérennes ont une syntaxe commune qui est basée sur la spécification des URI du W3C.[7]
Cette syntaxe est composée de trois parties :
– un préfixe qui indique le contexte dans lequel l'identifiant est attribué (par ex. http:, ftp:, urn:, etc.) ;
– un élément qui permet de désigner l'autorité nommante qui a attribué l'identifiant au sein de ce système ; cette autorité peut être désignée par son nom, par un code attribué au sein du système, par un code ou un nom codé qu'elle détient par ailleurs (comme un nom de domaine), etc. ;
– enfin, le « nom » lui-même, c'est-à-dire une chaîne de caractères qui identifie la ressource de manière unique, au sein de ce système et pour cette autorité.

Le degré de normalisation de chacun de ces trois éléments est un élément clé de l'unicité des identifiants et de leur pérennité. Les préfixes sont enregistrés auprès de l'IANA,[8] ce qui garantit leur unicité : un préfixe qui ne serait pas enregistré risquerait d'être utilisé par différentes communautés ou à différentes époques sans que l'on s'en rende compte, ce qui menacerait l'unicité des identifiants. La désignation des autorités nommantes est propre à chaque système, mais l'utilisation d'un code numérique ou alphanumérique semble aujourd'hui la solution la plus fiable, attendu les risques de changements de noms liés aux institutions. Enfin, le nom lui-même est souvent laissé à la discrétion de l'autorité nommante, ce qui laisse présager de grandes disparités si l'on n'adopte pas rapidement des règles de bonnes pratiques. On retombe sur les questions de citabilité, de sémantique et d'opacité des noms évoquées en début d'article.

Un système d'identifiants pour la BnF

Le choix d'un système d'identifiants pérennes à la BnF a été réalisé en prenant en compte les différentes fonctionnalités évoquées ci-dessus et en les hiérarchisant par rapport aux besoins de l'établissement : d'une part, la visibilité de ses ressources sur le Web et la construction de services associés et, d'autre part, la mise en place du système de gestion cohérente des documents numériques au sein d'une archive compatible OAIS,[9] dite SI numérique, en vue de leur préservation sur le très long terme.

Le système d'identifiants pérennes devait donc répondre à la fois aux besoins du SI numérique et à ceux de la communication et de la consultation des documents, dans le respect de l'existant mais en préfigurant les futures extensions (intégration du dépôt légal de la Toile, dépôts de masters électroniques par des partenaires, numérisation de masse…). Le système devait fonctionner dans l'état actuel des techniques du Web, et favoriser l'accès aux documents et la citabilité. Il convenait d'éviter aux usagers la manipulation d'outils techniques spécifiques et de s'intégrer dans leurs pratiques documentaires sur le Web. Un des enjeux était d'assurer une meilleure visibilité et compréhension des contenus numériques pour les usagers et pour les robots d'indexation des moteurs de recherche.
La plus grande attention a été accordée à l'indépendance du système choisi, à la fois sur le plan budgétaire pour ne pas obliger l'établissement à s'engager dans un processus coûteux qu'il serait susceptible de vouloir abandonner un jour, et sur le plan technique afin que le système puisse être immédiatement intégré à l'architecture existante.

Avec ces contraintes, plusieurs solutions ont été explorées :
– Le nommage OAI associé à un système de résolveur tel que POI
  Dans la mesure où le protocole OAI est mis en œuvre à la BnF pour l'ensemble des documents numérisés, et prochainement pour tout le catalogue, la gestion d'identifiants OAI est un acquis. Il aurait donc pu être intéressant de s'appuyer sur ce système, en lui adjoignant un résolveur, pour gérer les identifiants pérennes de la BnF. Cependant, du point de vue des fonctionnalités, l'OAI ne répondait pas à l'ensemble des besoins. Les identifiants pointent vers des métadonnées et non vers les documents eux-mêmes. La granularité d'accès se situe donc obligatoirement au niveau de la notice bibliographique. De plus, l'utilisation d'OAI et du résolveur POI requiert un système complexe de redirections, ce qui n'est pas idéal du

7. RFC 3986 : http://www.ietf.org/rfc/rfc3986.txt
8. Internet Assigned Numbers Authority. La liste des préfixes enregistrés est accessible en ligne : http://www.iana.org/assignments/uri-schemes.html
9. L'Open Archival Information System, reconnu comme norme ISO 14721, est un modèle conceptuel pour la mise en place d'un système ouvert de conservation à long terme des objets numériques.

point de vue de l'accessibilité, de l'indexabilité et de la citabilité.

– La mise en place d'un système local semblable à celui de la NLA
  L'expérience de la Bibliothèque nationale d'Australie (NLA) a montré qu'un certain succès peut être rencontré dans l'élaboration d'un système local, garantissant l'indépendance vis-à-vis d'un domaine encore en manque de maturité, tout en prévoyant une compatibilité possible avec les systèmes internationaux dans l'optique d'une évolution future. C'est le travail de spécifications, d'organisation de la collection et de création des outils qui a semblé ici trop lourd et peu susceptible de relever le défi d'extensibilité posé par la prise en compte imminente de nouvelles ressources numériques.[10]
– Le système ARK : c'est le choix qui a été retenu finalement.

ARK (pour Archival Resource Key)[11] est un système d'identifiants pérennes créé et maintenu par la California Digital Library. Sa particularité réside dans l'ajout de deux éléments aux trois principaux éléments de syntaxe que nous avons vus ci-dessus : les autorités d'adressage (Name Mapping Autority Hostports – NMAH) et les qualifieurs.

Le NMAH est l'adresse du serveur qui va résoudre l'identifiant ARK. Comparativement au rôle joué par l'autorité nommante, on pourrait parler d'autorité d'adressage. Cette portion ne fait pas vraiment partie de l'identifiant, elle est optionnelle et peut être changée. De fait, il est possible d'avoir plusieurs autorités d'adressage pour un même identifiant, ce qui permet, par exemple, dans le cas de la BnF de respecter le contexte de visualisation des documents : un même document peut être ainsi indifféremment vu à l'adresse http://catalogue.bnf.fr/ark:/12148/bpt6k101412s et à l'adresse http://gallica.bnf.fr/ark:/12148/bpt6k101412s. Grâce à ce procédé, il devient possible de maintenir le même nom de domaine entre l'application d'accès (le catalogue ou Gallica) et le document lui-même. Pour Gallica, cette fonctionnalité est essentielle car elle garantit la possibilité pour les moteurs de recherche de ne pas être désorientés et d'indexer correctement les documents dans le domaine Gallica.

Le qualifieur est une chaîne de caractères qui permet de qualifier ce que l'on veut de l'objet. Il est optionnel.

10. Consulter à ce sujet le rapport demandé par la NLA pour le choix de son système d'identifiants en mai 2001 : Diana Dack, *Persistent identifiers. Rapport pour la National Library of Australia*, mai 2001, http://www.nla.gov.au/initiatives/persistence/PIcontents.html.
11. Consulter la spécification du format ARK : http://www.cdlib.org/inside/diglib/ark/arkspec.pdf

L'autorité nommante est libre de développer une hiérarchie et de rendre visibles des variantes : il devient ainsi possible de développer la granularité d'un document, de faire référence à des versions distinctes, ou d'appeler des services particuliers. Contrairement au nom lui-même, le qualifieur n'est pas soumis à une garantie de pérennité ; il peut donc être signifiant et également être associé à des services susceptibles de changer d'apparence ou de disparaître. On retrouve la dualité, évoquée plus haut, entre un identifiant abstrait qui est parfaitement pérenne et pointe vers un objet logique (l'identifiant ARK) et un identifiant concret qui désigne ou peut désigner une partie physiquement toujours identique de l'objet logique. Ainsi, l'identifiant ark:/12148/bpt6k85329c désigne l'objet logique qui correspond à la numérisation de « Wheler, George. Voyage de Dalmatie, de Grèce et du Levant » tandis que l'identifiant http://catalogue.bnf.fr/ark:/12148/bpt6k85329c/f4.pagination désigne l'objet physique correspondant à la quatrième image de cet objet logique, dans le contexte du catalogue et avec le service d'affichage de la pagination. De ces deux identifiants, seul le premier est pérenne ; le second contient des informations difficilement pérennisables, y compris des informations sémantiques (« catalogue », « pagination ») qui ont seulement vocation à rendre un service.

Les objets numériques appelés à intégrer le SI numérique de la BnF sont d'une grande diversité et se présentent avec leurs identifiants, précédemment attribués dans les processus de création ou de publication, qui sont de différentes sortes : des ISSN, des ISBN, des cotes de bibliothèques, des URI, des DOI, des identifiants automatiques de type UID, etc. Les processus de collecte ou de numérisation eux-mêmes sont variés : il existe trois chaînes de numérisation (une pour les livres, une pour les images et une pour l'audiovisuel) auxquelles on ajoutera le dépôt légal de la Toile et la production documentaire interne (records management). Chacun de ces processus dispose nécessairement de ses propres identifiants de production, qui logiquement sont tous différents puisqu'ils répondent à des besoins différents.
En particulier, ces différentes chaînes sont confrontées à des problèmes de granularité très variés et ont chacune leur façon de les résoudre.

Au moment où tous ces objets intègreront le SI numérique, il deviendra impossible de gérer ces différents identifiants ; ils seront donc enregistrés dans les métadonnées tandis que le système attribuera un nouvel identifiant ARK qui se trouvera dès lors au centre du système. Cet identifiant est chargé de faire le lien entre les métadonnées et les objets eux-mêmes, et grâce au système de qualifieurs d'ARK, il est également possible de gérer à ce niveau la granularité des objets. Rien n'empêche, lors de l'attribution du nouvel identifiant, de réutiliser dans la partie « nom » de ARK une partie de l'identifiant de production si cela semble pertinent. En l'occurrence, la BnF a fait le choix d'utiliser des identifiants opaques alphanumériques, plus pérennes et extensibles, qui intègrent en partie les anciens codes à barres qui servent à gérer la production de numérisation.

Du point de vue de l'accès, on dispose avec le système ARK d'un identifiant pérenne au niveau de l'objet et aux différents niveaux de granularité qui le composent, ce qui va servir aux applications et à la gestion du système lui-même, mais également aux usagers. Ceux-ci bénéficient désormais d'un système stable pour accéder aux objets et aux parties d'objets, ce qui permettra d'améliorer la citabilité. En outre, nos partenaires pourront construire sur le système leurs propres applications, par exemple des interfaces d'annotation ou d'enrichissement de contenu, sans avoir à faire une copie locale des fichiers source et ce, même s'ils veulent atteindre un niveau de granularité très profond. Utilisé également pour le catalogue BN Opale Plus, ARK permettra de rendre les notices adressables : il deviendra possible de faire un lien permanent sur une notice du catalogue, ce lien pouvant être utilisé dans des applications connexes comme l'entrepôt OAI, des catalogues collectifs ou des moteurs de recherche.

Glossaire et références

URL	Uniform Resource Locator	Chaîne de caractères permettant de localiser une ressource sur le Web. Cette chaîne est précédée du préfixe correspondant à la localisation des documents sur le Web : « http ». http://www.ietf.org/rfc/rfc1738.txt
URN	Uniform Resource Name	Chaîne de caractères permettant d'identifier (par son nom) une ressource sur le Web. Cette chaîne est précédée du préfixe « URN ». http://www.faqs.org/ftp/rfc/rfc2141.txt
URI	Uniform Resource Identifier	Chaîne de caractères permettant d'identifier une ressource sur le Web, par sa localisation ou par son nom. Cette chaîne est précédée d'un préfixe enregistré tel que : « http », « urn »… http://www.ietf.org/rfc/rfc3986.txt
DOI	Digital Object Identifier	Système d'identification des objets dans un environnement numérique, qui repose sur la suite logicielle « Handle » et est géré par l'International DOI Foundation. http://www.doi.org/
ARK	Archival Resource Key	Système d'identification créé et maintenu par la California Digital Library. http://www.cdlib.org/inside/diglib/ark/
OAI	Open Archive Initiative	Initiative à l'origine du protocole OAI-PMH, qui permet d'échanger des métadonnées sur le Web et intègre son propre système d'identification. http://www.openarchives.org/
POI	Purl-based Object Identifier	Système d'identification qui combine les identifiants attribués dans le cadre de l'OAI-PMH et le système de redirection PURL développé par OCLC. http://www.ukoln.ac.uk/distributed-systems/poi/

NB : nous signalons l'ouvrage récemment publié par le Consortium des bibliothèques européennes de recherche et la Commission européenne sur la conservation et l'accès :
Implementing Persistent Identifiers
Overview of concepts, guidelines and recommendations
de Hans-Werner Hilse et Jochen Kothe
Novembre 2006, 57 p.
ISBN-90-6984-508-3
Disponible en version PDF sur <www.cerl.org> et <www.knaw.nl/ecpa>

Persistent Identifiers for Digital Resources: The Experience of the National Library of France

by Emmanuelle Bermès
functionnally in charge
of the Digital library on the Web,
National Library of France

Introduction

Creating online resources, including digitization, electronic publishing or even just blog entries, raises the question of consistent identification of these resources on the Web. Libraries have long been facing identification issues, with numbers related to the publishing industry, like ISSNs and ISBNs, and library signatures used to organize and retrieve books in a collection.(1) The identifier issue is everything but new in libraries, but as often, the transposition of the practice towards the digital world and particularly the Web environment raises questions: it becomes impossible to dissociate genres or medias, it becomes necessary to manage versions and granularity…

When faced to digital assets management, these problems are current; these questions have already been raised and answered, often using both metadata and identifiers. While it is clear that identifiers are necessary at every organizational level of a digital repository, how to create and manage them still raises issues.

Identifiers on the Web: Goals, Requirements, Systems

The first question concerns the goals of identifiers. The terminology suggests two different points of view. Stable reference or permanent link refers to the capacity of someone to name and retrieve a resource, what we could call citability. When talking about persistent identifier or resource identifier, we are rather in a uniqueness and persistence perspective, with a practical use.
Citability can be facilitated through the use of semantic identifiers, which carry intelligible information in themselves. This information is based on the metadata of the resource, for instance its title. This kind of identifiers is widely used on the Web, in particular within weblog platforms and content management systems. But semantic identifiers, while facilitating resource retrieval and understanding, can raise different problems. First of all, if the resource changes of nature or content, the semantic link between the identifier and the resource might be broken. Moreover, semantic identifiers are deeply linked to the structure of the material they describe, which in the case of a critical mass of material might prevent them from being generic or coherent enough. Last but not least, through times and places the meaning of a word or acronym might be different: it can be obvious at a certain time and confusing one century later; it can become offending in less than ten years. When working on the long term and at an international scale, it might be important to consider these changes and to choose an opaque system.

Opaque identifiers are usually software generated. Existing standards include the UUID(2) (Universally Unique IDentifier), an identifier built upon technical information and date and time details. Such identifiers raise another kind of issue: they make it necessary to preserve the link between the resource and its identifier, because the identifier doesn't carry in itself any information about the content of the resource. The automation of identifiers generation also allowed the machines to create very long identifiers, thus making them impossible to reuse by human beings.

While semantic identifiers ensure the citability and accessibility of resources, opaque identifiers are more persistent and should preferably be used in a long-term preservation environment. This is again revealing a well-known issue of the library world: the dilemma between preservation and access. With two different contexts of use (archiving and preservation versus retrieval and access) it is impossible to find an identifier system able to solve all problems at once. Both aspects have to be taken into account and prioritized in order to achieve a system fitting one's own requirements.

1. LUPOVICI, Catherine, « Le Digital Object Identifier : le système du DOI », *BBF*, 1998, n°3, p. 49-54 http://bbf.enssib.fr

2. "A Universally Unique IDentifier (UUID) URN Namespace", RFC n°4122 http://www.ietf.org/rfc/rfc4122.txt

Identifiers Requirements

Functional requirements for identifiers are not a recent area of studies and we already find them, not very different from those we consider today, in the RFC 1737 titled "Functional Requirements for Uniform Resource Names", in 1994.(3)

The first requirement is uniqueness, as an identifier is supposed to qualify one resource, and only one. In return, the same resource, even located in different places, should have the same identifier. This is called 'global uniqueness' and requires some kind of centralized organization at an international level. Then these identifiers are supposed to be persistent, which means that we would want some guarantee that they won't change and that they will keep on identifying the same resource. Persistence is the key to stable reference and the major issue in identifiers design and use. Thus, it is now widely admitted that persistence is not a technical issue, and no identifier system is technically able to ensure it. Persistence is rather an institutional commitment, relying mainly on the governance of the system. Institutions or actors whose activities are supposed to last are able to become 'naming authorities' at a local, national or international level. These naming authorities become owners of the power of naming, but also of an important responsibility: to ensure the persistence of their names.

The independence of the naming authority should therefore be discussed: for an institution, it might be reasonable to consider that a good identifier system should impose no or few constraints to that institution. The Bibliothèque nationale de France pays great attention to the independence, in terms of technique as well as costs, of its identifier system, because the independence is a guarantee of freedom of implementation and thus persistence in that specific context. On the contrary, a small publisher would benefit from the constraints, because they come along with technical assistance (a set of tools) and global security (for example with a continuity organization in case the publisher disappears).

Besides, persistence and uniqueness also raise a scalability issue. It is now admitted that identifiers may be re-affected or destroyed within the Internet. Persistence is not 'forever', but 'long enough' regarding the needs of the institution in charge of a resource. For instance, an identifier such as <http://www.lemonde.fr/> is stable but its content changes every day. Can it be considered persistent? We should distinguish 'abstract' resources that can change or move, and 'concrete' resources, unique and stable, both needing identifiers. It is also necessary to define the scale of uniqueness: some identifiers are unique inside a definite system which is convenient for internal use, while dissemination on the Web requires globally unique identifiers, harder to manage.

3. http://www.w3.org/Addressing/rfc1737.txt

From a technical point of view, identifier systems should be able to answer specific needs, as we just described them, for each producer considering his own means.

Identifiers should be applicable at every granularity level of a resource: the resource itself, a larger set or collection, various articles inside the resource, why not a line or a paragraph, and even different versions of the resource. Like with any digital collection management, an initial choice between logical and physical granularity has to be made. The different granularity levels must be defined, as well as the way to identify them: various solutions are possible, from the choice of completely independent identifiers at each level (then linked together with a structural map), up to a hierarchical system reflecting the organization of the collection.

Identifiers can integrate preexisting models used by the producer, like ISBNs, ISSNs, library signatures, ancient naming system used for URLs or for digital files. Persistence also relies on the ability to adapt to the changes in environment, and it is necessary to make it possible to extend the naming system and take into account new types of resources, network evolutions, Web standards or browsers capabilities.

Eventually, identifiers can be actionable: they give a method to retrieve the resource itself. Today, the 'http' protocol, used for the Web, is the only one to answer this requirement because it can be interpreted by the tools we use every day to browse the Web. Identifiers which don't rely on the 'http' protocol require different methods to make them actionable: resolvers, plug-ins, proxies. These resolvers use 'http' technologies to make identifiers actionable.

The context of persistent identifiers should make it possible to know more about the resource, to locate it and to retrieve it. Additional information, known as metadata, including the location of the resource, can be allocated to the identifier. A reference database can give the correspondence between an identifier and a resource location. Both solutions are not exclusive: an identifier can be associated with metadata,

and with a resolver which gives the corresponding URL.

Some identifier systems recommend providing metadata along with the resource naming: DOI and ARK for example. Other systems particularly dedicated to the exchange of metadata require an identifier system to fulfil their role, to give access to the primary resource. One of them is the OAI-PMH protocol.[4] The role of the resolver is to make the correspondence between the name of the resource and its location. The resolver can be inside the naming institution, or outside and managed by a third part authority.[5] The type of resolver and the quality of the metadata associated with an identifier system are essential in the choice of the system. They are at the core of the system and must be confronted with the major needs and requirements: granularity, updates, versions, etc.

Identifier Systems Today

Persistent identifiers systems currently rely on an international naming authority, which has the necessary power and independence in order to allocate identifiers that are persistent, unique, adaptable, and actionable through a resolver.

To compare or evaluate the different systems is a difficult task, considering that none is clearly better than the other regarding persistence; it is rather a question of producer's needs. Most systems combine various functionalities at different levels. There are no clear and definitive evaluation criteria. Each system has its own way of organizing and managing identifiers.

Persistent identifiers have a common syntax base on the URI specification from the W3C.[6] This syntax is composed of three parts:

– a scheme that indicates the context in which the identifier is attributed (ex. http:, ftp:, urn:, etc.);
– an element allowing to identify the naming authority that created the identifier inside the system; this element might be a code or encoded name that the authority already owns (like a domain name);
– the name itself, a string that identifies the resource inside the system and for this authority.

4. http://www.openarchives.org/OAI/2.0/guidelines-oai-identifier.htm
5. Ex.: the Library of Congress has its own resolver for DOIs and Handle identifiers. An identifier like: doi:10.1045/january2005-fox, once prefixed with the resolver's location http://hdl.loc.gov/doi:10.145/january2005-fox, gives access to the resource. It is a mere resolution service: the resource given in example has no relation with the Library of Congress.
6. RFC 3986: http://www.ietf.org/rfc/rfc3986.txt

The standardization level for these three elements is essential for the uniqueness of identifiers and for their persistence. Schemes are registered by the IANA,[7] as a guarantee of uniqueness: an unregistered scheme could be used twice by distinct communities or at different times without being noticed by anyone, thus threatening the identifiers' uniqueness. The designation of naming authorities is proper to each system, but the use of an alphanumeric code today seems the best solution for avoiding changes of names inside organizations. The structure of the name itself is often chosen by the naming authority, with possible lack of homogeneity if guidelines and good practices are not applied. We are still confronted to citability, semantics and opacity issues underlined previously.

An identifier System for The BnF

The choice of an identifier system for the BnF was made taking into account the functionalities we developed. Priority levels were determined regarding on one hand the issue of visibility of resource on the Web, and on the other hand the implantation of a long-term preservation and OAIS-compliant repository.
The system had to comply with both preservation and access needs, in the respect of previous IT systems but with the perspective of future extensions (integration of Web archiving, digital masters' deposit, mass digitization…). The system had to work within actual state of technology and enhance access and citability of the resources. We wanted to avoid complicated manipulations to users, and to integrate seamlessly their current practices of Web documentation. One of the issues was to ensure a better visibility and understanding of our digital material, for users but also for search engines.
The greatest care was granted to the independence of the chosen system, regarding the costs because we didn't want to involve the BnF in an expensive process that would later be given up, and also regarding the technical requirement so that the new system could be immediately integrated to the current architecture.

With these constraints, various solutions were explored.

– OAI identifiers with a resolver system like POI
  We already use the OAI-PMH protocol for digital material, and soon for the whole catalogue, so it is necessary to manage identifiers in this context. It could have been interesting to rely on this system, adding a resolver, to manage identifiers. Though, the OAI didn't fit all our requirements. Identifiers relate to metadata and not to resources themselves. The granularity level is necessarily the bibliographic record. Moreover, the use of OAI with a POI resolver requires a complex redirect system, a threat to accessibility, indexing and citability.
– Implementing a home-made system like at the NLA
  The National Library of Australia showed that a local system can be successful. It is independent in a domain that still lacks maturity, and takes into account future evolutions of international systems. The task of requirements, the digital management and the creation of tools stopped us from choosing this solution. The extensibility issue was not sufficiently addressed regarding imminent integration of new resources.[8]
– The ARK system was at least chosen.

7. Internet Assigned Numbers Authority. The list of registered prefixes is available online: http://www.iana.org/assignments/uri-schemes.html

ARK (standing for Archival Resource Key)[9] is an identifier system created and maintained by the California Digital Library. Its peculiarity relies on two elements added to the three major parts of the URI syntax: Name Mapping Authority Hostports (NMAH) and qualifiers.

The NMAH is the location of the server which resolves the ARK identifier. Compared to the naming authority concept, this is a location authority. This element is not really part of the identifier; it is optional and can be changed. Thus, it is possible to have various location authorities for one identifier, allowing for example in the case of the BnF to show the context of the resource: the same document can be consulted equally from <http://catalogue.bnf.fr/ark:/12148/bpt6k101412s> and <http://gallica.bnf.fr/ark:/12148/bpt6k101412s>. Thanks to this possibility it is possible to maintain the same domain name between the search interface (the catalogue BN Opale Plus or the Digital Library Gallica) and the document itself. For Gallica, this functionality is essential because it determines the possibility for search engines to correctly interpret and index the material in the Gallica domain.

The qualifier is a string that qualifies anything from the resource. It is optional. The naming authority is free to develop a hierarchy and make the variants visible. It is then possible to develop the granularity of a document, to reference different versions, or to call specific services. While the name is supposed to be persistent, the qualifier is not: it can embed semantics and be associated to services that might change, or disappear. This corresponds to the duality between an abstract identifier, persistent and linked to a logical object, and a concrete identifier which relates to a physically stable part of the logical object. For instance, the identifier ark: /12148/bpt6k85329c relates to the digitization of a logical object: "Wheler, George. Voyage de Dalmatie, de Grèce et du Levant", while the identifier <http://catalogue.bnf.fr/ark:/12148/bpt6k85329c/f4.pagination> relates to the fourth image of this object, in the catalogue context, and displayed with its pagination. Out of these two, only the first one is persistent. The second one embeds semantics (catalogue, pagination) which relate to a service and are not supposed to be persistent.

8. See the NLA report on identifiers: Diana Dack, *Persistent identifiers. Report for the National Library of Australia*, May 2001 http://www.nla.gov.au/initiatives/persistence/PIcontents.html
9. See the specification of the ARK scheme: http://www.cdlib.org/inside/ diglib/ark/arkspec.pdf

The digital objects that will be ingested in the digital repository of the BnF present a great diversity and come with a lot of already existing identifiers, attributed during the production or collect process: ISSNs, ISBNs, library signatures, DOIs, automated UID, etc. There are also several different digitization and collect processes: three digitization chains (one for books, one for still images and one for audiovisual material), the Web archiving, and the records management. Each one of these processes has its own necessary production identifiers, which are very different from one another, not surprisingly since they all have very different needs. In particular, the granularity issue is different for each process, and the strategies for solving it are also diverging.

When all this material is ingested in the repository, we cannot possibly deal with all these different identifiers, so they are carefully recorded in the metadata, while the repository attributes a new ARK identifier. ARK identifiers are at the core of the system because they provide the necessary link between metadata and digital objects, and with the qualifier system, it is also possible to manage the granularity levels. If it seems relevant, some elements of the original production identifier may be reused in the 'name' part of the ARK. As a matter of fact we decided to use opaque strings, more persistent and extensible, which integrate a part of the barcode previously used to manage the digitization process.

As a result, from the dissemination point of view, we have persistent identifiers at the object level and at different granularity levels that will be helpful for the repository management and for software applications and modules of the system, and additionally to end-users. Those will benefit a consistent naming system to have access to the objects and parts of objects, which will help for reference and citation issues. Moreover,

our partners will be able to build upon our repository their own applications, for example for annotation or content enrichment, without having to copy the source files and even if they want to consider a very deep granularity level. Used also for the BN Opale Plus catalogue, ARK makes the bibliographic records addressable: it is now possible to make a persistent link towards a catalogue record, this link being reused in related applications like the OAI repository, federated catalogues or search engines.

Glossary and References		
URL	Uniform Resource Locator	A compact string representation for a resource available via the Internet. URLs are used to 'locate' resources. The scheme for URLs is 'http'. http://www.ietf.org/rfc/rfc1738.txt
URN	Uniform Resource Name	Uniform Resource Names (URNs) are intended to serve as persistent, location-independent, resource identifiers. The scheme for URNs is 'urn'. http://www.faqs.org/ftp/rfc/rfc2141.txt
URI	Uniform Resource Identifier	A compact sequence of characters that identifies an abstract or physical resource. An URI begins with a registered URI scheme. http://www.ietf.org/rfc/rfc3986.txt
DOI	Digital Object Identifier	Framework for persistent identification of content objects in the digital environment. The system is managed by the International DOI Foundation, and based on the Handle software. http://www.doi.org/
ARK	Archival Resource Key	A naming scheme for persistent access to digital objects (including images, texts, data sets, and finding aids), currently being tested and implemented by the California Digital Library. http://www.cdlib.org/inside/diglib/ark/
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting	Provides an application-independent interoperability framework based on metadata harvesting. http://www.openarchives.org/
POI	Purl-based Object Identifier	A relatively persistent identifier for resources that are described by metadata 'items' in OAI-compliant repositories. http://www.ukoln.ac.uk/distributed-systems/poi/

NB: we wish to recommend the book recently published by the Consortium of European Research Libraries and the European Commission on Preservation and Access:
Implementing Persistent Identifiers
Overview of concepts, guidelines and recommendations
By Hans-Werner Hilse and Jochen Kothe
November 2006, 57 p.
ISBN-90-6984-508-3
Available as PDF at <www.cerl.org> and <www.knaw.nl/ecpa>

Identificadores perennes

La gestión de las colecciones digitales implica profundas reflexiones acerca de su composición, su descripción y además su identificación. En tiempos del acceso global a los recursos a través de las redes, el problema de la identificación única y perenne a los documentos digitales se plantea de manera aguda. Sin ser un tema nuevo, ni en las bibliotecas, ni en la Web, el problema de los identificadores perennes alcanza ahora un cierto nivel de madurez en los debates: aunque todos los actores de este campo se pongan de acuerdo sobre las principales funcionalidades que hay que poner en práctica, la selección de un sistema sigue siendo un problema. Este artículo presenta el recorrido de la Bibliothèque nationale de France en cuando a la selección y la puesta en funcionamiento del sistema de identificación ARK (Archival Resource Key), desde la evaluación de las funcionalidades hasta los ejemplos de funcionamiento actual.

Web Archiving at BnF

Introducing the BnF Web Archiving Team

- Catherine Lupovici has been the founder and Director of BnF Digital Library Department since 1999. She supervises the Web legal deposit strategy along with the BnF Institutional repository project. Catherine has been the International Internet Preservation Consortium (IIPC) Program Officer since 2005.
- Gildas Illien is a library curator with background in public management and communications. He served as librarian for the French Ministry of Foreign Affairs and lived in Canada, Scandinavia and Austria. Gildas came back to Paris to work on library construction and develop users' services before joining the team as Project Manager in 2005.
- Sara Aubry combines the skills of a digital archivist, a Web designer and a librarian computer scientist translator. A pioneer in Web archiving, she has been in charge of architecture and tool specifications since 2002. Sara also likes teaching the Internet to librarians and has been webmaster of the IIPC website since its creation.
- Clément Oury is a library curator with PhD background in early modern history. After a successful internship working on digital rights management, he was appointed at BnF in 2006 to organise the management and preservation of Web archives within the library institutional repository.
- France Lasfargues worked in the e-business field before discovering BnF where she was first involved in the management of cartographic collections and an use case for Gallica. She is now running quality assurance for the Web archives, and is also in charge of organizing communications with the public and coordinating our first use case project.
- Younès Hafri has a PhD in computer engineering and specialized in Web harvesting and related data management. He developed a Web crawler (Domino) before joining the library in 2004. Younès has developed more software and harvesting tools for BnF and for the consortium, including the BAT tools and BnF's curator tool.
- Bert Wendland began his career as system administrator in the Computing Centre of the Humboldt University in Berlin. In 2001, he joined the Electronic Publishing Group, a joint working group with the University Library. Bert participated in the German « nestor » project (Network of Expertise in Long-Term Storage of Digital Resources). He changed to BnF in 2006, where he is responsible for implementing and administrating the Web archiving technical infrastructures.

To contact us: name.lastname@bnf.fr

Introduction

Hosting the IIPC Steering Committee gives us the opportunity to make an update on BnF's organization and projects.
In 2006, we have been mostly focusing on building our own organization and internal dissemination inside the library. Provided we get appropriate funding, we hope in 2007 to participate more actively in the Web archiving international conferences, workshops and publications.
Much has been happening at BnF in 2005 and 2006 and we would like to report briefly on the most recent and striking milestones of our project.
After several years of discussion, France finally got a law for the Web legal deposit. Collections have become bigger too, thanks to our partnership with the Internet Archive.
As things were hence getting more real, we started installing the Web archiving activity more permanently within the library culture and organization.
So we set up a network of "Web-legal-deposit-friendly" librarians, launched a series of dedicated workshops and developed specific tools to encourage them to learn how to work with Web archives. One of the key issues of this long-term collaborative work is to manage to express policy in generic terms which can cope with scalability issues and harvesting specifications.
More projects are coming up, including harvesting of the next French presidential and general elections with « Heritrix », testing access tools with a use case, exploring more quality issues in conjunction with the captures of our national domain. Last but not least, time has come to build our institutional repository architecture while contributing to the international work being done in the fields of standardization, the WARC format and the Web archives metadata among other tasks.
We hope to continue exchanging with other institutions from abroad on all these projects and that the IIPC Consortium will keep on providing the framework we all appreciate for such collaboration.

Finally A Law for Web Legal Deposit in France

It took more than three years and a lot of patience, but it finally happened: the French Parliament passed a law which clarifies BnF's rights and obligations regarding the legal deposit of the Internet. The extension of legal deposit to the Web is now a reality with the vote followed by the official publication of the DADVSI law (DADVSI stands for *Droit d'auteur et droits voisins dans la société de l'information*), on August 3rd.(1) With this law, BnF's project can now shift from a rather uncomfortable twilight zone of experimentation to become a legal mission of the library.

1. Why the French keep talking about Web "legal deposit" rather than "Web archiving"

BnF's approach to Web archiving derives from the French legislation and tradition regarding legal deposit and the obligations endorsed by the National Library in this respect. Together with other institutions (especially INA, the National Institute for Radio and Television, which is responsible for preserving the audiovisual heritage of France), BnF is in charge of the legal deposit of all materials published in the country since the library's foundation as a legal deposit institution in 1537. Collections gathered through the national legal deposit have a particular status and the library is given specific rights and obligations regarding their management.

- Publishers have the duty to deposit copies of all published material at the library, which is responsible for collecting it extensively and preserving it with no limitation of time.
- The library has the duty to describe these collections: this task belongs to the Agence Bibliographique Nationale which publishes the Bibliographie nationale française.
- Access to these collections is possible, but restricted inside the library building and for researchers only.

The emergence of new technologies and publishing models (from printed documents to visual, photographic or musical material up to databases and software, more recently) has led to successive extensions of the legal deposit legislation and the integration by BnF of new types of documents fixed on a variety of formats and media in its collections.

Over the years and because of the multiple extensions of legal deposit to new materials, collecting resources has become a very large-scale, industrial process, reaching a critical size in many respects (over 1.7 million physical units were collected by BnF in 2004).

For some types of media (such as software and other multimedia material), the library no longer aims at collecting nor preserving all publications. Sampling has proved to be a useful way to maintain a reasonable scale of organization while keeping track of a wide variety of publications regardless of their "quality": the philosophy of legal deposit is indeed to keep a record of the "best" along with the "worst" as collections should be a mirror of society's global cultural production and evolution over centuries.

2. The exact scope of the Web legal deposit still needs to be clarified

The law is extending legal deposit to the Internet in the following terms:

> "Is also liable to legal deposit every sign, signal, writing, image, sound or every kind of messages communicated to the public by electronic channels" (Clause 39).

This broad definition leaves potential interpretation regarding the nature of the items which may be within the scope of legal deposit: the Internet shall not be restricted to websites, even though BnF's harvesting policy mainly focuses on this type of resource at the moment. Therefore, the law regards all types of "online electronic publications" constituting a set of signs, signals, images, sounds or every kind of messages, as long as they are disseminated through the public space of browsing. The public space of browsing may be understood as the set of websites linked to each other by at least one hyperlink.

All the people who produce or publish online material in order to communicate with the public by electronic channels are under the obligation of legal deposit. The law is to apply to those who somehow have a connection with the national territory – this has always been the case in the past for the other types of documents: unlike the National Library of Switzerland which collects "everything about Switzerland", the National Library of France collects "everything published (or even imported) in France".

The very nature of this technical and legal connection to the national territory and thus the exact scope of the electronic publications still remain unclear and need to be clarified in the decree which should be following the law in 2007. The electronic publications hosted under the French TLD (.fr) should, for sure, be in the scope, but legal deposit cannot be strictly restricted to the .fr. The possibility to crawl other domain names should be made explicit in order to comply with massive harvesting techniques and with the cultural reality of the "French" national production on the Web – a multifolded concept that seems equally hard to determine by harvesting robots, librarians and lawyers!

3. The law encourages bulk harvesting but keeps the possibility of individual e-deposits

> "Mandated institutions may collect material from the Internet by using automatic techniques or by setting specific agreements and deposit procedures together with the producers" (Clause 41 II).

BnF's "mixed" collecting strategy combines massive and bulk harvesting, selective and thematic crawls and individual e-deposits for resources which cannot be collected otherwise. Though priority is given to bulk automatic harvesting, this policy complies with the law, which leaves to the library many possibilities to balance between its three collecting modes in the future, depending on the evolution of the Web and of harvesting tools.

The law also stipulates that no obstacle such as login, password or other forms of access restriction should be opposed to the library by the producers. Of course, automatic harvesting is often confronted to such restrictions and this shall by no means be considered as a violation of the law since all producers have the right to protect their publications. But penalties shall apply to producers who refuse to help BnF collecting their resources when the library gets in touch with them in order to find a specific agreement in the case where automatic harvesting has failed. These penalties will not be applicable before 2009.

4. We shall not catalogue archived websites

The law is not explicit on this point, but the following clause is likely to be introduced in the forthcoming decree.

> "Online diffusion of descriptive information embedded in legal deposit electronic publications shall constitute their National Bibliography."

In other words, providing access and search tools to browse the archives is considered to be sufficient to advertise these very large-scale collections.

5. Restricted access to extensive resources

Due to copyright protection, the same restrictions should apply to Web archives access as for the rest of legal deposit material. The law is not explicit on this point but refers here again to the forthcoming decree which shall specify both selection and communication procedures that mandated institutions (BnF and INA) shall comply with. Hence BnF still needs to wait for this decree (hopefully in 2007) before giving access to its Web archives collections. Access is likely to be restricted to in-house browsing and analysis, by researchers only.

Collections Get Bigger: Reception of Large-scale Collections and The Partnership with the Internet Archive

1. Make it bulk to make it work

The initial assumption behind the BnF Web legal deposit project has always been that only massive harvesting of our national domain could provide collections matching the scale of the Web and comply with the philosophy of legal deposit, which is one of massive sampling rather than individual selection.

Introducing large-scale collections (see figures in the table below) within the library was a turning point in the project. The tangible presence of this huge, almost threatening, amount of data within their own workplace has been determinant in convincing BnF librarians that they needed to adjust to a new size and form of material. It led them to question drastically their approach to collection policy and description.

Except for small specific and event-based crawls (2002 and 2004 French elections, for instance), and individual e-deposits, the harvesting process is currently run for BnF by the Internet Archive (IA). BnF signed a 3-years agreement with IA in 2004 and both partners agreed to embark on a research project which has provided the library with most of the material it needs to settle its functional model.

The agreement with IA includes the delivery of snapshots of the French domain. A first snapshot was run in the Fall of 2004 and delivered in February 2005. A second one was run in the Fall of 2005 and delivered in March 2006. In 2005, IA also run for BnF its first focused crawl of ca. 3500 websites. Moreover, IA provided BnF with retrospective collections extracted from previous « Alexa » crawls: the library has thus acquired historical archives from 2001 to 2004 so far. Web archives, their indexes and the Wayback Machine software are currently delivered once a year on the popular high-storage-capacity red set of racks known as the Petabox (Capricorn Technologies), and stored in BnF's computing room. Archived Web resources can be accessed by librarians through BnF's professional network.

BnF and IA agreed this summer 2006 on a third delivery which should include a new .fr domain crawl, a focused crawl and more historical collections. For the

1. Text of the law (in French): http://www.legifrance.gouv.fr/WAspad/UnTexteDeJorf?numjo=MCCX0300082L. Dispositions specifically addressing Web legal deposit issues are in Title IV, Clauses 39 to 47.

first time, the 2006 snapshots will be indexed in full text, an exciting perspective for browsing and access tests. This year, the broad crawl should also be comprehensive of all hosts in the focused crawl as we plan to conduct a specific research on quality measurement by comparing archives of identical websites captured almost at the same time with the broad and the focused procedures.

2. More quality issues to address

We expect this research to provide the required material to keep on developing a comprehensive QA methodology using Heritrix reports. This also gives food for thought in order to define and possibly standardize useful metrics that are highly needed both to improve crawl monitoring and to qualify collections. Discussing collection quality with librarians (a sensitive issue) and with institutional repository managers will only be possible if common terms of reference can be found. For several years, our team has thus been trying to address, measure and formalize the issue of quality, especially the "depth" and the "completeness" of the archives.

Through this comparative approach, we also aim at running a cost/opportunity analysis with potential impact on work organization. The key issue is indeed to find the right balance between focused crawls (which require human selection and ongoing management but supposedly provide material of a better quality) and bulk crawls (which require limited human resources but might be more risky in terms of collection management). Ultimately, we would like to restrict the number of seeds collected through the focused process in order to maintain the library organization at a reasonable size and cost regardless of the increasing volume of our national domain.

Collections in figures	Delivered on Petabox 1 (February 2005)	Delivered on Petabox 2 (March 2006)	Total
Historical collections (Alexa extractions)	2 128 597 826 URL 65 007 980 Hosts 289 931 ARC	3 691 508 664 URL 64 540 256 Hosts 431 519 ARC	5 820 106 490 URL 129 548 236 Hosts 721 450 ARC
.fr broad domain crawls performed by IA with Heritrix	121 273 454 URL 504 000 Hosts 24 893 ARC	167 137 702 URL 1 227 755 Hosts 41 685 ARC	288 411 156 URL 1 731 755 Hosts 66 578 ARC
Focused crawls performed by IA with Heritrix		54 946 100 URL 2 024 924 Hosts 13 670 ARC	54 946 100 URL 2 024 924 Hosts 13 670 ARC
Total (URL)	**2 249 871 280**	**3 913 592 466**	**6 163 463 746**
Total (Hosts)	**65 512 003**	**67 792 935**	**133 304 915**
Total (ARC)	**314 824**	**486 874**	**801 698**

Collections acquired by BnF thanks to its partnership with Internet Archive

Re-inventing Policy in Light of Scalability: Librarian Networks at Work

2005 and 2006 have been years of intense dissemination inside the library. Web Archiving is not only about developing tools: in the long term, Web archives will replace many documents which are still at the core of the daily activity of many librarians. Many of these publications have started or even finished their migration toward the Internet. Moreover, BnF will soon give public access to its born digital collections from its reading rooms: there will be a need for assisting librarians who will be in charge of assisting visitors facing those new screens. In other words, the library will be strongly impacted by Web archiving at different levels.

With ca. 2500 people working at BnF and a capacity to adjust to change which cannot be very fast due to the size and history of the institution, it is crucial to prepare the staff to the upcoming shifts in collection management and public services. We know this long-term process is heavy, that it will ultimately impact the overall work organization, practices and culture. But we can learn from former technological revolutions which the library was able to face successfully: for instance, the arrival of audiovisual and multimedia materials within the scope of the library in the 80's, or the development of electronic resources and the beginnings of digitization in the 90's. Web archives are a new material, but the library is used to "adopt" new objects. Only should we be careful about the way they are being introduced in the organization, by involving staff as early as possible.

1. Developing Web skills among librarians

Our Web team contributes to the dissemination of information and skills within the library through the organization of various meetings and tutorials. Those exchanges are of great importance, for they encourage collaborative work in a completely new field and contribute to raise awareness and develop Web skills among librarians.

A network of 35 part-time subject librarians working in collection and legal deposit departments was set up in 2005 in order to contribute to theme or event-based projects requiring their expertise and their monitoring of websites. Each major collection department of the library has at least one "correspondent" in charge of following the Web archiving project and contributing to the definition of a collection policy statement in the field of his/her expertise. This network kept on growing in the past months, and we now have 70 people in different departments of the library who are somehow involved in Web collection management.

At this stage of the project, we intend to stabilize this group so that it does not get too big. Otherwise it will become difficult to provide assistance and tutorials to all of them. Besides, extending too far this form of organization could lead to misunderstandings about what collection policy is about when it comes to the Web legal deposit at a very large scale: content selection should be the exception, not the rule.

2. Inventing new forms of digital curation

In 2006, this collaborative work implied:

– improving our local curator tool (GDLWeb)[2] and extending its functionalities;
– running a series of tutorials and workshops gathering small focus groups of librarians;
– installing and developing a relationship and common terms of reference together with them;
– setting up the conceptual and technical framework aimed at getting librarians to express their Web archiving policy and expectations in generic terms which can cope with scalability issues and Web crawlers specifications.

In its latest (V.3) version, GDLWeb offers the following functions:

– accessing the BnF archives;
– proposing new seeds to crawl and giving crawling specifications (such as depth and frequency);
– change crawling specifications of existing seeds;
– checking archives quality on a technical-based, rather than collection-oriented analysis.

Through these projects, we hope to install a "Web archiving culture" inside the library and to formalize a general collection policy statement for BnF at large. It is only possible to take time in this continuing professional and cultural process because the large, automated snapshots massively enrich our Web collections at the same time. Without the broad crawls, we could not handle the identification and selection tasks as we are currently doing it. As mentioned earlier, the next step of our work plan will be to identify the critical balance and border between focused and bulk harvesting so that we find the optimal rationale between policy vision, harvesting cost and collection quality.

Awareness is thus being spread on the necessity to express collection development policies which should be consistent with technical harvesting requirements: a serious challenge for a long-standing institution like BnF.

Coming up Projects

1. Crawling the 2007 French elections

The forthcoming presidential and general elections should mark a decisive evolution in the use of the Internet by French political organizations and Web users at large. New strategies and tools have emerged (such as buying Google keywords, political mailing, blogs, podcasts, wikis and many other interactive and multimedia forms of expression) and spread within the public space of the Internet. Many experts and observers see this evolution as a possible turning point for the making of an "e-democracy" in the country. BnF is expected to keep traces of these digital

2 - Gestion du dépôt légal du Web : Web legal deposit management.

ephemera. The library already collected political websites in 2002 and 2004. But in 2007, the archiving task is likely to require more librarian brains, better harvesting capacity… and much more storage space.

A committee of 20 librarians and technicians has been set up, which will be in charge of selecting and harvesting the websites following a typology of publications which includes all political organizations websites and a significant sample of unofficial publications illustrating the manifold contributions and participation of various communities in the campaign. Harvesting frequencies will vary in time over a period of eight months, starting in October 2006. Priorities will need to be set in order to keep the harvesting process at a reasonable cost but on a 24 hours a day basis. On the technical side, this project will be an opportunity to test the latest version of Heritrix and to improve our capacities in crawl monitoring and data management and preservation

2. An use case to prepare smart(er) access

At the other end of the harvesting process, all heritage institutions involved in Web archiving are now willing to progress in the field of "smart" access. So is BnF, who decided to test tools and service situations related to Web archives access. This use case aims at providing a field of experimentation with a focus on users' critical reception and appropriation of new tools and new forms of research materials drawn from the Web archives.

We will primarily explore access tools to the Web archives from a user's perspective. In addition, attention will be given to the emergence of new scholarly approaches, needs and research strategies that users will develop when confronted to Web archives for the first time. This test will take place one year before providing a much wider public access to the Web archives in BnF reading rooms. It will involve a sample of ca. 20 students and scholars in the fields of political and information sciences.

A thematic collection will be proposed as research material: the corpus of Web archives related to the French national, local and European elections collected in 2002 and 2004. The population for the use case is mostly attached to a single academic institution, the Institut d'études politiques de Paris. Two sets of users were defined: "testers" are graduate students attending a dedicated seminar on the political Web, who will be using the tools and exploring the corpus; "observers" are second level scholars in charge of providing a comprehensive study of the testers' practice.

The use case demonstrator will offer several possibilities to access the archives, from the well-known Wayback Machine to more sophisticated software allowing linguistic data mining and trend analysis. The interfaces will hence provide visualization, navigation, search and analysis functions. Evaluation will be organized in several steps over a period of 20 weeks. Qualitative data will be collected through interviews and focus groups. Survey protocols should focus on the following aspects: ergonomics, research strategies, cognitive aspects, content analysis, comparison of archival and online Web material use, integration of Web resources with other types of research material.

3. Building the institutional repository

Since 2005, BnF is engaged in a three years process aimed at building its institutional repository, also known as the "Fifth Tower" of the Tolbiac building. The need for a large scale digital repository has been increasing in the past years as digital resources of various types take a growing part in the library collections: digitized collections (the Gallica digital library along with other specialized heritage collections such as digitized newspapers or audiovisual materials) and born digital collections (such as electronic publications and, of course, the Web archives). All these resources need proper storage space and a common architecture capable of addressing long-term and large-scale data management and preservation issues.

In 2006, several working groups have been specifying their needs in order to set submission and preservation agreements between the BnF producers and the Archive. Discussions also aimed at formalizing all issues related to formats, metadata, rights and risk management. These specifications are made in light of the OAIS standards and the PAIMAS protocol. Because of their volume and their specific preservation needs, the Web archives will be a significant "client" of the future repository: a large scale storage infrastructure has already been acquired by the library and the Web archives might be the first to be "ingested".

L'archivage du Web à la BnF

En 2006, nous avons travaillé pour l'essentiel à construire notre propre structure et à la promouvoir en interne. A condition que nous obtenions les financements nécessaires, nous espérons, en 2007, participer de façon plus active aux conférences, ateliers et publications concernant l'archivage du Web, sur un plan international.

En 2005 et 2006, notre activité à la BnF a été importante et nous souhaitons évoquer brièvement les événements importants les plus récents qui ont marqué notre projet.

Après plusieurs années de discussions, la France a finalement obtenu une loi sur le dépôt légal du Web. Les collections se sont également développées, grâce à notre partenariat avec l'Internet Archive.

Dans la mesure où les choses se concrétisaient, nous avons commencé à asseoir de façon plus permanente notre activité dans la culture et l'organisation de la bibliothèque.

Nous avons ainsi constitué un réseau de bibliothécaires « favorables au dépôt légal du Web », lancé une série d'ateliers spécialisés et développé des outils spécifiques pour les encourager à se familiariser avec les archives du Web. L'une des questions-clés de ce travail de collaboration à long terme consiste à parvenir à communiquer sur notre politique au moyen de termes génériques pour évoquer des questions de scalabilité et les spécifications de collecte de données.

De nombreux projets sont à venir, qui consistent à collecter, grâce à « Heritrix », les informations consacrées aux prochaines élections présidentielles et législatives en France, à utiliser un cas d'étude pour tester des outils d'accès, à explorer davantage les questions de qualité conjointement aux captures de notre domaine national. Enfin, le temps est venu de construire l'architecture de notre propre dépôt institutionnel en contribuant à la recherche internationale qui a lieu, entre autres, dans le domaine de la normalisation, du format WARC et des métadonnées des archives du Web.

Nous espérons continuer à échanger avec d'autres institutions étrangères sur tous ces projets ; nous espérons également que le Consortium international sur la conservation de l'Internet continuera à constituer le cadre que nous apprécions tant de cette collaboration.

Archivo Web en la BnF

En 2006, nos hemos concentrado principalmente en la construcción de nuestra organización y diseminación interna dentro de la biblioteca. Si logramos obtener el financiamiento adecuado, esperamos participar en 2007 más activamente en las conferencias, talleres y publicaciones internacionales de archivo en la Web.

Han sucedido muchas cosas en la BnF en 2005 y 2006 y nos gustaría hacer un breve informe acerca de los hitos más recientes y significativos de nuestro proyecto.

Después de varios años de discusión, Francia finalmente logró una ley para el depóstio legal Web. Igualmente, las colecciones han crecido, gracias a nuestra sociedad con el Internet Archive.

Como las cosas se estaban volviendo más reales, comenzamos a incorporar la actividad de archivo en la Web de manera más permanente dentro de la cultura y organización de la biblioteca.

Por ello creamos una red de bibliotecarios « amigables con el depósito legal Web », lanzamos una serie de talleres dedicados y desarrollamos herramientas específicas para estimularlos a aprender a trabajar con los archivos Web. Uno de los aspectos claves de este trabajo de colaboración a largo plazo es lograr expresar la política en términos genéricos que permitan manejar los temas de la escalabilidad y las especificaciones de recolección.

Hay otros proyectos en curso, entre los que se incluye la recolección de las próximas elecciones presidenciales y generales en Francia con « Heritrix », que son herramientas de acceso de prueba con un caso de uso, para explorar más temas de calidad conjuntamente con las capturas de nuestro dominio nacional. Finalmente, ha llegado el momento de construir nuestra arquitectura de depósito institucional mientras contribuimos con el trabajo internacional que se está llevando a cabo en los campos de la normalización, el formato WARC y los metadatos de archivos Web entre otras tareas.

Esperamos continuar el intercambio con otras instituciones del exterior sobre todos estos proyectos y que el International Internet Preservation Consortium (IIPC) siga brindando el marco que todos valoramos para dicha colaboración.

The 2006 Hague Blue Shield Accord

28th September 2006

The representatives from ICBS and National Blue Shield Committees met in The Hague on September 27th and 28th 2006 to discuss and agree on the most effective way to support the new International Committee for the Protection of Cultural Property in the Event of Armed Conflict, established under the Second Protocol of the 1954 Hague Convention, and how best to respond to the generous offer of funding and facilities made by the Municipality of The Hague.
They agreed upon the creation of a new body – the Association of National Committees of the Blue Shield – and the following distribution of responsibilities.

The National Committees

- Define their own priorities within the mission of the Blue Shield.
- Are action orientated with practical projects, capacity building and raising the profile of emergency management (preparedness, response, recovery and evaluation/review).
- Will initiate, organize and maintain national, regional and local networks based on the networks of the five international NGOs including representatives of/observers from military authorities, emergency services, cultural organizations, civil emergency services, Red Cross and other relevant humanitarian organizations.
- Promote ratification and full national implementation by Government of the 1954 Hague Convention and its protocols.
- Need to be recognized by the ICBS.
- Use of the term National Committee of the Blue Shield (NCBS) is conditional on the continued recognition of the national committee by the ICBS.
- Must be set up in such a way as to conform to national legislation.
- Promote the aims of Blue Shield:
  – with national governments including identifying relevant resources;
  – fundraising;
  – national awareness raising and capacity building.

The Association of National Committees of the Blue Shield (ANCBS)

- Serves as communication centre, archive and resource base for ICBS and National Blue Shield committees and facilitate communication between the three levels of the Blue Shield movement in a systematic way.
- Provides a permanent postal address for ICBS and the ANCBS.
- Promotes awareness raising, capacity building, preparedness, response and recovery at national and international level.
- Promotes awareness of issues to decision makers and potential funders at all levels.
- Facilitates creation of national networks in areas where BS committees do not yet exist.
- Promotes bilateral and multilateral assistance systems between committees and emerging committees.
- Coordinates and disseminates information on international actions.
- Sets up a website containing access points to relevant databases (training, resources, specialists, conferences) with interactive links to other databases.
- Promotes Blue Shield brand by re-labelling projects wherever possible.
- Undertakes fundraising to sustain the costs of the secretariat and to make funds available for agreed projects.
- Supports proactively ICBS and National BS Committees in their activities.
- Assists ICBS in fulfilling its role under the Second Protocol.
- Promotes training for peacekeeping and other forces.
- Identifies 'champions' and other key supporters/influencers at international level who can be approached in a coordinated fashion by ICBS/ANCBS.
- Supports together with ICBS the creation of database of specialists who could be authorized by UNESCO and combatants.
- Process:
  – the ANCBS board will be comprised of representatives of the national committees and observers of the ICBS;
  – ANCBS is authorized to employ a secretariat.

International Committee of the Blue Shield Agreement on Purpose and Functions

Concluded by the representatives of IFLA (Peter Lor, Sjoerd Koopman, Christiane Baryla), ICA (Joan van Albada, David Leitch), ICOMOS (Gaia Jungeblodt), ICOM (John Zvereff, Cristina Menegazzi) and CCAAA (Crispin Jewitt, Kurt Deggeller) – The Hague, 25th September 2006 and amended 28th September 2006 (SC, JZ, KD, GJ, SK, JvA). This statement builds on the founding objectives of the ICBS of April 1996 and its Charter concluded in Strasbourg on 14th April 2000 - <http://www.ifla.org/VI/4/admin/nc-req.htm>.

Purpose

To promote the protection of cultural property (as defined in the Hague Convention) against threats of all kinds and intervene strategically with decision makers and relevant international organizations to prevent and to respond to natural and man-made disasters.

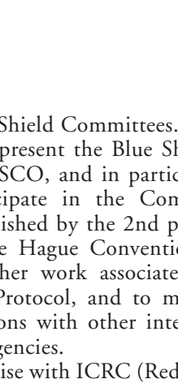
Functions

- To promote the ratification and implementation of the Hague Convention and its protocols, with the emphasis on advocating ICBS's philosophy and principles.
- To encourage the establishment of National Blue Shield Committees.
- To recognise/de-recognise National

Blue Shield Committees.
- To represent the Blue Shield at UNESCO, and in particular to participate in the Committee established by the 2nd protocol of the Hague Convention and in other work associated with this Protocol, and to maintain relations with other international agencies.
- To liaise with ICRC (Red Cross) and ICCROM, as other bodies referred to in the 2nd Protocol, and with other high-level bodies.
- In combination with the Association of Blue Shield National Committees, to ensure the running of an information clearing house on disaster situations and threats, with a view to international alerting.
- To issue statements on disaster situations and threats.
- To arbitrate in disputes of National Blue Shield committees.
- To provide expert advice and evidence to the International Criminal Court and other international tribunals in conjunction with national committees as required.

Governance
- The ICBS is constituted by the chief executive officers of the participating international organizations, CCAAA, ICA, ICOM, ICOMOS and IFLA, with a rotating chair.
- The chair provides a secretariat for core ICBS functions only, with support from the other organisations as appropriate.

Action and Timetable
A working committee will be composed (the president of ICBS serving as coordinator) from among the local organizers (incl. PCF/CER), the national committees and ICBS.
The committee will report back to all by 1st December on progress.
The committee will call a meeting in March/April 2007 for final discussion on ratification of articles and agree details regarding secretariat (housing, staffing, sustainable funding).
Those signing this Accord or adhering to it later are expected to provide input and feedback at all stages.



We wish to recommend two very interesting documents that have been recently published in the field of Web archiving.

Web Archiving
By Julien Masanès

2006, VII, 234 p., 28 illus., Hardcover
ISBN-10: 3-540-23338-5
ISBN-13: 978-3-540-23338-1
Price: 52,70€
To order this book: <http://www.springer.com/france/home?SGWID=7-102-22-72040423-0>

E-Journal Archiving Metes and Bounds: A Survey of The Lanscape
By Anne R. Kenney, Richard Entlich, Peter B. Hirtle, Nancy Y. McGovern and Ellie L. Buckley

September 2006, 120 p.
Published by Council on Library and Information Resources
ISBN 1-932326-26-X
ISBN 978-1-932326-26-0
Available online: <http://www.clir.org/pubs/abstract/pub138abst.html>

Announcement

International Conference on Newspapers Collection Management: Printed and Digital Challenges
April 3-5th, 2007
Santiago, Chile

This conference is to promote good practice in Newspaper Collection, Management and Preservation. It is being co-organized by the National Library of Chile, the IFLA Newspapers Section and IFLA-PAC.
The conference is divided into 6 sessions. Each session will last one and a half hours. Papers are to be for a maximum of 25 minutes. The themes are listed below:
- the digital future for newspapers in Chile (to include metadata, interoperability and publisher issues);
- the digital future for newspapers in American and European countries;
- the role of newspaper cataloguing in digitization;
- the role of newspaper preservation in digitization;
- copyright, publishers, legal matters for newspapers;
- education of librarians in newspaper collection management: new perspectives for the role of librarians.

For more information, please contact:
Ximena Cruzat, Director, National Library of Chile
<ximena.cruzat@bndechile.cl>
Hartmut Walravens, Staatsbibliothek zu Berlin
<Hartmut.Walravens@sbb.spk-berlin.de>
Christiane Baryla, IFLA-PAC Director
<christiane.baryla@bnf.fr>
Edmund King, British Library, Newspapers collection
<ed.king@bl.uk>

Reports

Preventive Conservation at The National Archives and Library of Ethiopia: A Mission Report
June 2006
By Caroline Laffont, Biologist, Bibliothèque nationale de France, and Anne Lama, Assistant in Preventive Conservation, Centre historique des Archives nationales, France

A mission on preventive conservation was led at the National Archives and Library of Ethiopia, in Addis Abäba, in June 2006. This mission was set up within the framework of the cooperation program signed in December 2005 between the Direction des Archives de France (DAF), the Bibliothèque nationale de France (BnF), and the National Archives and Library of Ethiopia (NALE).
This cooperation program was first established in 2001 and led by Mrs. Denise Ogilvie, Patrimonial Curator at the Centre historique des Archives nationales and Mrs. Anaïs Wion, Historian, who have implemented several projects[1] such as training sessions in Ethiopia or in France on manuscripts cataloguing or archives management.
Because the NALE was proceeding of its moving to a new building, it seemed important to undergo a mission specifically based on conservation and more specifically on preventive conservation. This mission had two main goals: a sanitary conditions assessment of collections before their moving and a training session on the main preventive conservation issues. This mission was carried out by Mrs. Caroline Laffont, Biologist at the BnF and Mrs. Anne Lama, Assistant in preventive conservation at the Centre historique des Archives nationales.

This training session aimed to develop awareness on why and how planning and carrying out policies and process to safeguard collections from deterioration. We tried to adapt our training, as much as possible, to the NALE collections specificities. Thus, we get onto on the dust removal, handling, the protective enclosures, the reformatting but also on the preventive measures to take for an exhibition and a disaster preparedness plan. The hands-on aspect of the training has been appreciated.
There were 17 trainees coming mostly from NALE but also from other institutions such as the Orthodox Church Library or the National Museum. Trainees were solely heads of departments or sections of archives and library, which have good educational skills in collections management and for some of them in conservation and preservation. They are graduated mainly from the University of Addis Abäba or School of Information Study for Africa (SISA); some of them have completed their education in India, Russia and France.
NALE envisage extending this training on preventive conservation to the whole staff. This training will be delivered by the trained heads of departments.

This two-way confident relationship, added to the fact the training was led on the first week, greatly facilitated the sanitary conditions assessment since goals and related constraints (necessary and unlimited access to premises and collections) were all understood, accepted, and made possible.
The sanitary conservation conditions assessment is the first assessment of this type led at the NALE. It was carried out in order to determine the necessary measures to be undertaken for moving collections into the new building. Thus, the whole collections (manuscripts, prints, records, photographs, microfilms) were examined to check out the different encountered degradations and more specifically, to check out if an active infestation (caused by insects or microorganisms) was occurring. Premises were also studied to define the risky points that should be controlled and maintained.

Two methods were applied according to the premises and collections characteristics:

1. Olgivie Denise, Wion Anaïs, *Co-operation with National Archives and Library of Ethiopia. Report and Cooperation Projects*, June 2001 and April-May 2003.

– a random sampling method to extrapolate the degradations, or materials, typology to the entire collections from a random-chosen documents sample. For that purpose, a data base has been implemented listing the most frequent degradations encountered by NALE collections. This data base construction is based on the AFNOR standard NF Z40-011 on "the physical statement of archives and library collections assessment methodology". As this random sampling method is not suitable in case of active biological deteriorations, we also used another method as follows:
– a visual inspection method for premises and collections combined with sampling on suspected marks of active biological deteriorations (microorganisms and insects). For the collections, the sampling was carried out with sterile and dry swabs or direct sampling which were analysed at the BnF laboratory in order to determine the state of micro-organisms viability and their deterioration ability for collections. For the premises, insects traps have been put in place in the old and new building in order to evaluate insects populations and risks of infestation.

Environmental conditions were also assessed. Climatic data loggers recording, temperature and relative humidity have been put at strategic places in repositories, but also outside in order to evaluate the building inertia and climatic fluctuations affecting archives and library collections. Even if the duration of the assessment was too short to have a global view (such an assessment would demand at least one year study to obtain a global view over the dry and rainy seasons), these collected results give information and conclusions useful for the preventive conservation policy implementation.

It appears that archives and library are in a good conservation conditions except for the dust accumulation and the lack protective enclosures and lack of storage space. The new building, built near the old one, within the Ministry of Culture and Tourism compound, will resolve this actual lack of space of storage but also of reading room and will improve security and safety of collections.

This assessment and the training session enable NALE to proceed to a moving planning, prioritizing actions that have to be carried out or materials necessary to acquire before the moving, but also in a medium and long-term period. The dust removal of collections was performed during the summer. At the present time, the moving should be done.

IFLA General Conference and Council, Seoul Pre-conference in Tokyo

August 16-17th, 2006
Report by Yukiko Saito,
PAC Director for Asia

A Pre-conference entitled "Preservation and Conservation in Asia" was held on August 16-17th at the National Diet Library (NDL), Tokyo, gathering about 200 participants each day. The meeting was sponsored by the IFLA Preservation and Conservation Section, Asia and Oceania Section and PAC Core Activity.

On August 16th, experts from the U.S., Thailand, India and Japan stated their activities and views under the theme "Preservation issues in Asia". The poor condition of Asian documents on palm-leaf, paper, or microfilm was described. Major problems in the Asian countries were identified, including lack of awareness among the general public as well as in governments, lack of conservation centers, lack of conservators, lack of funds. Useful suggestions for improving the current situation were also made, such as providing efficient training programs, establishing a regional center in Southeast Asia, and preparing an action plan in each country.

The theme of the next day was "Microfilming and digitization of documentary heritage in Asia". After the important role of the IFLA-PAC Core Activity was stressed by the new PAC Director, different case studies in China and Australia were presented. The advantages and disadvantages between microfilming and digitizing were discussed. It was made clear by many speakers that digitization was now replacing microfilming in the Asian region, partly because of the problems of microfilm storage. However, many problems remain when considering digitization as a preservation tool.

The full papers are available at NDL website: <http://www.ndl.go.jp/en/iflapac/preconference/program.html>

COSADOCA (COnsortium de SAuvetage du patrimoine DOcumentaire en cas de CAtastrophe) Disaster Preparedness in Libraries

20-21st September 2006
Switzerland

In Switzerland, the COSADOCA, gathering the Library of the University of Lausanne, the Cantonal Archives and the Library of EPFL (Ecole polytechnique fédérale de Lausanne), has organised its second practical workshop on the training place of the civil protection.

On 20th September, the organizers set up a fire simulation and on 21st September, a flood simulation. The exercice was implemented in cooperation with the local fire department and the local civil protection. During the two days, the librarians got to know how to put in practice the disaster plan (how to evacuate the documents, how to sort them and choose the correct treatment, emergency care like air-drying treatments).

For more information, please see at: <http://www.cosadoca.ch>

Emulation Expert Meeting

20th October, 2006
The Hague, The Netherlands
By Gregory Miura, Head of the electronic documents section, Audiovisual Department, Bibliothèque nationale de France, Member of the Emulation Expert Meeting

Emulation project – Background information

The Koninklijke Bibliotheek (KB) and the Nationaal Archief of the Netherlands are both facing the same challenges keeping all kinds of digital material accessible for the long term. Therefore, both organizations are closely working together in this field by developing an emulator for digital preservation. Both the KB and Nationaal Archief strongly believe that emulation is the only way to retain access to digital objects for which preservation of functionality is important. By using an emulator, these objects should be kept alive in their original environment.

In April 2005 the Nationaal Archief and KB started the emulation project and chose Tessella Support Services plc. to develop an emulator which should be flexible and durable for the long term. Key issue in this development is the conceptual model of the Modular Emulator, which was defined by the KB in cooperation with Jeff Rothenberg at the end of 2004. Jeff is also joining the project team during design and development of the emulator.

In 2005, the Nationaal Archief of the Netherlands and the Koninklijke Bibliotheek, National Library of the Netherlands, have started a project to develop a modular hardware emulator for digital preservation. In order to review the first results of our project and, more generally, to identify different approaches of emulation and to discuss related issues in the field of digital preservation, the Nationaal Archief and the National Library of the Netherlands have organised an Emulation Expert Meeting. This event took place on 20th October 2006 and was attended by 16 international experts in the area of emulation and/or digital preservation. At this meeting, we also discussed topics such as future user aspects, long-term platform independency and how to create a (distributed) service that will offer access through emulation.

At the end of the expert meeting the participants formulated and endorsed a statement concerning the use of emulation for digital preservation purposes:

"Emulation is a viable preservation strategy that has a number of unique advantages.

- It preserves and permits access to each digital artifact in its original form and format; it may be the only viable approach for preserving digital artifacts that have significant executable and/or interactive behavior.
- It can preserve digital artifacts of any form or format by saving the original software environments that were used to render those artifacts. A single emulator can preserve artifacts in a vast range of arbitrary formats without the need to understand those formats, and it can preserve huge corpuses without ever requiring conversion or any other processing of individual artifacts.
- It enables the future generation of surrogate versions of digital artifacts directly from their original forms, thereby avoiding the cumulative corruption that would result from generating each such future surrogate from the previous one.
- If all emulators are written to run on a stable, thoroughly-specified 'emulation virtual machine' (EVM) platform and that virtual machine can be implemented on any future computer, then all emulators can be run indefinitely.

In order to develop a practical, off-the-shelf preservation strategy based on emulation, a number of additional steps are required, including:

- create and demonstrate example emulators suitable for long-term preservation;
- develop fidelity criteria for each behavioral dimension of digital artifacts (e.g., display, sound, timing) and develop validation test suites, which evaluate these criteria and verify that the logical behavior of an emulator matches that of its target computer;
- research and develop device-independent input/ output mechanisms to allow unmodified programs to behave and interact appropriately with users on future computer platforms;
- develop methods for capturing and preserving contextual information describing the logical, physical, organizational and social environments in which digital artifacts were originally used, as well as documentation describing how they were used and what they were used for;
- develop methods for describing, managing and automatically interpreting information about the versions and configurations of software and hardware needed to render digital artifacts under emulation;
- define and develop a long-lived emulation environment to enable emulators to be run indefinitely. This environment could be equivalent to an emulation virtual machine (EVM) platform, though it may be implemented as a long-lived programming language along with a stable set of program library facilities. This environment should:
  – enable using old digital artifacts by running their original software under emulation on unforeseen future computers;
  – provide automatic configuration of emulators, software environments and applications to render old digital artifacts;
  – provide documentation, active user help and/or automatic reinterpretation of old interaction modes into future equivalents, to help future users to utilize old digital artifacts under emulation;
  – provide mechanisms to facilitate (or, ideally, automate) the future generation of surrogate versions of digital artifacts directly from their original forms.
- Develop network-based services for providing remote access to old digital objects via emulation, without requiring remote users to load and run an emulation environment on their local systems."

This statement is endorsed by all participants of the Emulation Expert Meeting:

– Geoffrey Brown, Indiana University, USA
– Raymond van Diessen, IBM Netherlands N.V., The Netherlands
– Hans Hofman, Nationaal Archief of the Netherlands, The Netherlands
– Jeffrey van der Hoeven, Koninklijke Bibliotheek, The Netherlands
– Vincent Joguin, ACONIT, France
– Bram Lohman, Tessella Support Services plc., The Netherlands
– Gregory Miura, Bibliothèque Nationale de France, France
– Bill Roberts, Tessella Support Services plc., The Netherlands
– Jeff Rothenberg, RAND corp., USA
– Jacqueline Slats, Nationaal Archief of the Netherlands, The Netherlands
– Tobias Steinke, Die Deutsche National Bibliothek, Germany
– Dirk von Suchodoletz, University of Freiburg, Germany
– Remco Verdegem, Nationaal Archief of the Netherlands, The Netherlands
– Randolph Welte, University of Freiburg, Germany
– Richard Wilkinson, Tessella Support Services plc., United Kingdom
– Hilde van Wijngaarden, Koninklijke Bibliotheek, The Netherlands

This statement together with a short overview and presentation slides will be made available on the project websites of Digitale Duurzaamheid (Digital Longevity) and the Koninklijke Bibliotheek:
http://www.digitaleduurzaamheid.nl/index.cfm?paginakeuze=327
http://www.kb.nl/hrd/dd/dd_projecten/projecten_emulatie-en.html

Open Scholarship 2006: New challenges for Open Access Repositories. Conference held at The University of Glasgow

18-20th October, 2006

After the OAI meetings at CERN in Geneva the Nordic Scholarly Communication conferences in Lund, Sweden, the Conference in Glasgow focused on the range of new challenges and opportunities faced by open access repositories. There were over 200 attendees from 24 countries.
Report by Jean-Louis Baraggioli, Director, Centre technique du Livre de l'Enseignement supérieur (CTLES).
Open access repositories are digital collections of research articles that have been placed there by their authors and made accessible without charge. They represent a valuable asset within educational institutions.
The issues presented at the conference covered a wide range of assets:
• setting up an institutional repository (IR) (at which cost?);
• responding to evolving university needs;
• investigating the collaborative understanding and tool development;
• watching and preventing the obsolescence of digital collections.

Describing and putting the subject in context, the speakers succeeded in giving a hint of the different projects in the UK and abroad. Many of the speakers approached the subject in very different ways and explored complementary issues.

An interesting feature was developed by Ingeborg Zimmerman (University Library of Zürich) with the presentation of ZORA (Zurich Open Repository and Archive) using BioMed Central's Open Repository service. ZORA, which is freely available online at <www.zora.unizh.ch>, provides researchers, students and staff with open access research literature.
The University of Zürich also has an open access near-mandate, expecting that researchers deposit a copy of all their published and refereed articles in ZORA, subject to copyright restrictions. Legal issues are of great concern among researchers and librarians and they were largely discussed during tutorials and workshops that took place the first day. Therefore, authors' attitude to copyright agreement is very important as it seems to reveal the general lack of awareness of what they have signed and its consequences. A very interesting contribution about The RoMEO (Rights Metadata for Open Archiving) and SHERPA projects were given through many experiences that developed model licence agreements. The misconception is that assigning copyright means that deposit is not permitted and many authors fear they will not get published if they do not sign or try to amend agreements. Usually, authors are interested but not really enthousiastic, and the challenge is to explain why they should bother depositing.

It seems difficult to deal with the references in the articles especially when we do not know anything about the references that appear in the process. Open archives ought to provide quality services for a global version because the problem is not only about access but about use. The consequence is that it changes scientific behaviours.
The collaboration among institutions appears as compulsory in order to choose among options and standards. Differences between institutional workflow proved challenging from acquisition and selection through to workflow and allocating costs.
It is important to understand how your partner works; the greater the understanding of differences and similarities, the higher the success ratio and the more realist national standards and approaches become.
The open access research literature is composed of free, online copies of peer reviewed journal articles and conference papers as well as technical reports, theses and working papers.

Websites
http://www.gla.ac.uk/enlighten
http://jiscstore.jot.com/SurveyPhase
http://www.library.manchester.ac.uk/projects/store
http://.jisc.ac.uk/whatwedo/programmes/programme_digital_repositories.aspx
http://rylibweb.man.ac.uk/spcoll/
http://clir.org (survey of the landscape of e-journal archiving)
www.zora.unizh.ch
www.sherpa.ac.uk/
www.thesealive.ac.uk
www.lib.gla.ac.uk/dedalus/

Newspaper Conference in Poland

19th-21st October 2006
By Else Delaunay, IFLA Newspapers Section

The Conference "Newspapers: Resources, Processing, Preservation, Digitization, Promotion/Information" was organised by the Poznan University Library, the Poznan School of Social Sciences, and the IFLA Newspapers Section. It took place in the new building of the Poznan School of Social Sciences.
The Polish organizers had managed to set up a very complete conference dealing with all aspects of newspaper work. Around 100 Polish librarians and scholars, as well as a few library professionals from Western Europe (France, Germany, United Kingdom), joined the conference. Some sponsors gave a short presentation of their achievements in preservation and digitization work.
28 papers were presented which gave a fine overview of the many varied activities going on in Poland at present, regarding newspaper history, cataloguing, preservation work and digitization projects. The organizers had succeeded in publishing almost all the papers beforehand as a volume in both Polish and English under the title:
Uniwersytet im. Adama Mickiewicza w Poznaniu.
GAZETY. Zasoby. Opracowanie. Ochrona. Digitalzacja. Promocja/ Informacja. Materialy z miedzynarodowej naukowej. Poznan, 19-21 pazdziernika 2006.
478p. ISBN 83-918728-6-6(1)

1. Available from/Disponible auprès de :
Poznan University Library
ul. Ratajczaka 38/40
61-816 Poznan - Poland
Tel: + 48 61 829 3800
Fax: + 48 61 829 3824
E-mail: library@amu.edu.pl

Adam Mickiewicz University, Poland. *NEWSPAPERS. Resources. Processing. Preservation. Digitization. Promotion/Information*. Conference Proceedings, Poznan, October 19-21 2006. 478p. ISBN 83-918728-6-6. This volume and a complete folder were handed over to each participant at the opening of the conference to facilitate understanding and discussions. Indeed a very successful conference summing up problems, achievements and projects in newspaper work in Poland and in general. A privileged moment also of reflection and exchange.

Conférence sur la Presse

19-21 octobre 2006
Poznan, Pologne
Par Else Delaunay, Section Journaux de l'IFLA

La conférence s'intitulait « Journaux : ressources, traitement, conservation, numérisation, promotion/ information ». Elle était organisée par la Bibliothèque universitaire de Poznan, la Section Journaux de l'IFLA et l'Ecole des sciences sociales de Poznan dont le nouveau bâtiment servait de cadre à la conférence.

Les organisateurs polonais ont réussi à mettre sur pied une conférence très complète traitant de tous les aspects des journaux. Quelque cent universitaires et bibliothécaires représentant la plupart des grandes bibliothèques polonaises, ainsi que des professionnels de l'Europe occidentale (Allemagne, France, Royaume-Uni), ont pris part à ces trois journées. Quelques sponsors ont brièvement présenté leurs réalisations dans les domaines de la conservation et de la numérisation.
Les 28 intervenants ont présenté des exposés témoignant des nombreuses activités qui se poursuivent aujourd'hui en Pologne dans les domaines de l'histoire de la presse, du catalogage, de la conservation et de la numérisation. A l'ouverture de la conférence, a été remis à chaque participant un beau volume contenant la majorité des exposés en polonais et en anglais, une véritable prouesse de la part des organisateurs. Ce volume qui a facilité la compréhension et les discussions est intitulé :
Uniwersytet im. Adama Mickiewicza w Poznaniu. *GAZETY. Zasoby. Opracowanie. Ochrona. Digitalzacja. Promocjal Informacja*. Materialy z miedzynarodowej naukowej. Poznan, 19-21 pazdziernika 2006.
478p. ISBN 83-918728-6-6(1).
Adam Mickiewicz University, Poland. *NEWSPAPERS. Resources. Processing. Preservation. Digitization. Promotion/ Information*. Conference Proceedings, Poznan, October 19-21 2006. 478p. ISBN 83-918728-6-6.
Ce fut en effet une conférence très réussie résumant parfaitement les problèmes, les réalisations et les projets que suscitent les journaux, anciens ou récents, en Pologne et en général. Ce fut aussi un moment privilégié de rencontre et de réflexion.

CARDIN / CLAMED Hosts Disaster Information Management Workshop

November 7-9th, 2006
Cuba
By Beverley Lashley, Coordinator/Librarian, UWI Mona Library and Sheree Trottman, School of Continuing Studies, UWI Barbados

The Caribbean is susceptible to face both natural and man-made disasters, the latter we pay little attention to, although it has taken its toll on our social, physical and economic infrastructure. In our efforts as information generators and caretakers not only to ensure that our history is protected, but also to collect information which will assist in reducing the impact of disasters on our lives, thirty participants from six Caribbean countries – Barbados, Cuba, Jamaica, Montserrat, St. Lucia and Trinidad and Tobago, gathered in Cuba for three days from November 7-9th, 2006. Of importance was the opportunity presented to have an inter-cultural exchange of ideas that would foster future networking cooperation. The workshop entitled "Disaster Information Management" was hosted by The Latin America

Centre for Disaster Medicine (CLAMED) and executed in collaboration with the Caribbean Disaster Information Network (CARDIN), a project of the University of the West Indies Library at Mona Campus, Jamaica. Funding was made of sponsorship provided by the Ministry of Health, Cuba, the Social Science Research Council (SSRC), USA and the Pan American Health Organization.

The main objective of this workshop was to strengthen the institutional capacities of the disaster information units in the Caribbean countries, through the use of similar methodologies for the development of information services and the development of plans for disaster reduction in information units. The workshop topics included:

- developing a disaster plan;
- providing guidelines to salvage damaged material;
- restoring the physical properties of information materials (print, audio visual, electronic);
- causes of deterioration of materials;
- overview of preservation and conservation;
- role of knowledge management
- and becoming acquainted with lessons learnt from disasters in the Caribbean.

The IFLA (International Federation of Library Associations and Institutions) Core Activity on Preservation and Conservation (IFLA-PAC) donated copies of their "Disaster Preparedness and Planning: A Brief Manual" that was distributed to all the participants. Ten copies of the IFLA manual were also given to Dr. Guillermo Mesa, President, CLAMED, at the closing ceremony on behalf of IFLA for distribution to other agencies in Cuba.

Participants did not only receive new insights into the area of disaster management and its link to preservation and conservation of heritage, but also contributed to the dialogue by sharing their experiences. Perhaps the most valuable was the intercultural experience which proved that barriers can be overcome, be it language or other, when working towards a common goal.

The first sessions of the day focused on the background of disasters in the Caribbean and the existing information networks. Against the background of the changes that have affected the Caribbean, Dr. Mesa Ridel gave participants an overview of what the region faces and expects to face in the future. He focused on the need to implement a risk reduction element in every cycle of disaster management. He indicated that the environment that man lived in was affected by many phenomena both natural and man-made and the Caribbean was no exception. The Caribbean landscape has been changed by natural disasters such as hurricanes, threatened by earthquakes and more recently climate change. He informed participants that man-made disasters such as droughts, flooding, oil spills resulted in the displacement of human beings and a significant number of deaths. Statistics have revealed that from 1992-2003, a thirty-year period, the Caribbean has experienced 3.5 million deaths due to disasters. In his view, man-made disasters are accelerating the economic and financial problems faced in many countries as this leads to social inequities and international migration. The afternoon session laid the foundation for the visit to the National Archives of Cuba by giving participants an insight into the vulnerability of materials to disasters. Each participant was given a package of books about the National Archives of Cuba. The day ended with a visit to Casco Histórico, the ancient part of the city.

On the second day, the presentations focused on knowledge management as it relates to preservation and conservation, national information networks, bio-deterioration and disaster planning. A panel discussion provided the platform for sharing of practical experiences. Two different case studies were presented to show the impact of a disaster on libraries. The participants were able to visualize the effects of damage caused after Hurricane 'Ivan' in 2004 at the University of the West Indies Mona Library, and the National Library Services of Montserrat that has been continually plagued by the volcanic activity from the pictorial slides presented. The IFLA manuals were well received and used during the practical exercise on guidelines for preparing and creating a plan for disaster reduction in information units.

The final day of the workshop provided an insight into the type of work undertaken in preservation in Cuba, the benefits and challenges. An outstanding presentation by Dr. Splenger, focused on the need to preserve memory and old tradition especially the oral. Dr. Splenger gave an informative presentation on Cuba's preservation program. In giving a brief overview of the discovery of Cuba, he stated that the Caribbean has its own history and identity. The first mention of disasters was noted in Columbus's writings but drawings from the Indians who occupied the islands reflect the phenomenon. He cautioned participants that in any plan they develop for preservation, they must be careful not to exclude the community in the restoration process. The community is part of a great cultural heritage and possesses knowledge that may not be available to the public with regards to any particular project.

Another vital area that was examined was funding opportunities that are made available for preservation and conservation. This session did not only provide the guidelines on how to source funding but a listing of funding organizations available to Cubans. The final session was dedicated to a group practical exercise. The participants were divided into two groups with a mixture of English and Spanish speaking participants. This session was designed to encourage intercultural interaction and communication where they were required to complete the assigned. Both groups had to identify the hazards in and around CLAMED's building and to determine the likelihood of it being damaged or destroyed as a result of a disaster. At the end of the allotted time, the groups made their presentations.

The workshop culminated with a closing ceremony at which the guest speaker was Dr. Eliades Acosta, Director of the National Library of Cuba. In giving his reasons for accepting the invitation to address the gathering, Dr. Acosta expressed his amazement of how the Caribbean is one of the most integrated in culture but different in communication that is due to the language barriers. In defending our culture, Dr. Acosta stressed that education is important and users especially children should be taught that books are sacred and should be respected. The role of the information professional is important in this aspect and they must go outside their profession and bring awareness through campaigns in schools and the mass media. Whatever is the strength of that country in disaster management, it must also assist other countries in the region citing Cuba as being very strong in meteorology. In his closing address, Dr. Acosta reminded participants that the first thing that suffers during a disaster is our cultural heritage and new threats such as terrorism should be included in disaster management plans. "When you loose your memory, you loose your freedom and ultimately your history", he stated. He assured the participants that his organization/country is willing to cooperate, as importance should be given to a unified Caribbean.

Dr. Mesa closed the proceedings by thanking all those who contributed to making the workshop a success. He expressed his delight at the exchange of experiences that took place, the knowledge gained and the role played by the speakers who assisted in fulfilling the objectives of the workshop by imparting their knowledge in all aspects of preservation, conservation and disaster management. All participants were issued with a certificate of participation.

Contact:
Beverley Lashley, Coordinator/Librarian, CARDIN
UWI Mona Library, Jamaica
and Sheree Trottman, Women and Development Unit
School of Continuing Studies, UWI Barbados

PAC CORE ACTIVITY

USA and CANADA

LIBRARY OF CONGRESS
101 Independence Avenue, S. E.
Washington, D. C. 20540-4500 USA

Director: Dianne L. van der REYDEN
Tel: + 1 202 707 7423
Fax: + 1 202 707 3434
E-mail: dvan@loc.gov

PAC INTERNATIONAL FOCAL POINT AND REGIONAL CENTRE FOR WESTERN EUROPE, AFRICA AND MIDDLE EAST

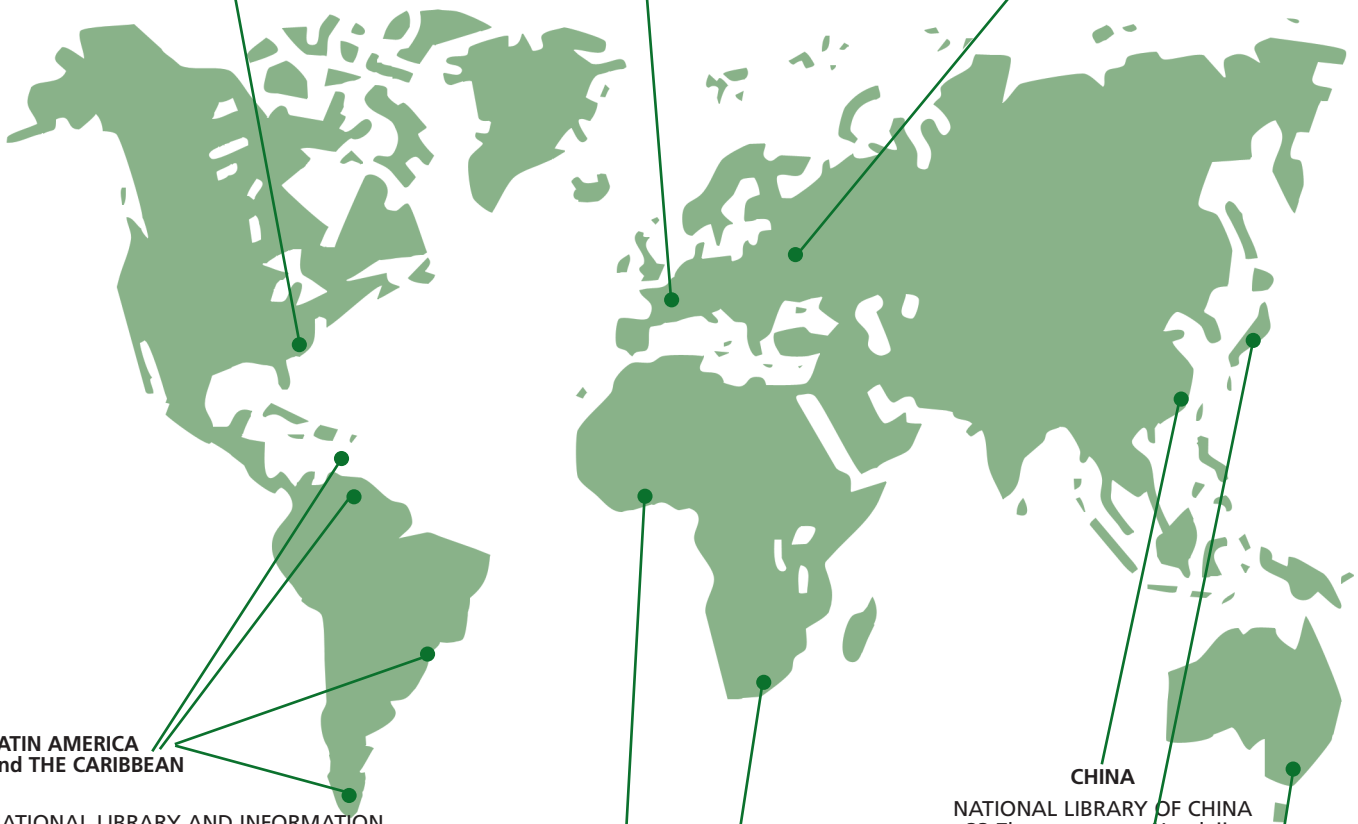
BIBLIOTHÈQUE NATIONALE DE FRANCE
Quai François-Mauriac
75706 Paris cedex 13 - France

Director: Christiane BARYLA
Tel: + 33 (0) 1 53 79 59 70
Fax: + 33 (0) 1 53 79 59 80
E-mail: christiane.baryla@bnf.fr

EASTERN EUROPE and THE CIS

LIBRARY FOR FOREIGN LITERATURE
Nikoloyamskaya str. 1
Moscow 109 189 - Russia

Director: Rosa SALNIKOVA
Tel: + 7 095 915 3621
Fax: + 7 095 915 3637
E-mail: rsalnikova@libfl.ru



LATIN AMERICA and THE CARIBBEAN

NATIONAL LIBRARY AND INFORMATION SYSTEM AUTHORITY (NALIS)
PO Box 547
Port of Spain - Trinidad and Tobago
Director: Annette WALLACE
Fax: + 868 625 6096

BIBLIOTECA NACIONAL
DE VENEZUELA
Apartado Postal 6525
Carmelitas Caracas 1010 - Venezuela
Director: Orietta PALENZUELA RUIZ
Tel. + 58 212 505 90 51
E-mail: dconsev@bnv.bib.ve

FUNDAÇÃO BIBLIOTECA NACIONAL DE BRASIL
Av. Rio Branco 219/39
20040-0008 Rio de Janeiro - RJ - Brasil
Tel: + 55 21 2220 1976
Fax: + 55 21 2544 8596

BIBLIOTECA NACIONAL DE CHILE
Av. Libertador Bernardo O'Higgins N° 651
Santiago - Chile
Director: Ximena CRUZAT A.
Tel: + 56-2 360 52 39
Fax: + 56-2 638 04 61
E-mail: ximena.cruzat@bndechile.cl

FRENCH-SPEAKING AFRICA

BIBLIOTHÈQUE NATIONALE DU BENIN
BP 401
Porto Novo - Bénin

Director: Francis Marie-José ZOGO
Te/faxl: + 229 22 25 85
E-mail: derosfr@uahoo.fr

SOUTHERN AFRICA

Preservation Unit
UCT LIBRARIES
University of Cape Town
Private Bag
Rondebosch 7701 - South Africa

Director: Johann MAREE
Tel: + 27 21 480 7137
Fax: + 27 21 480 7167
E-mail: jmaree@hiddingh.uct.ac.za

CHINA

NATIONAL LIBRARY OF CHINA
33 Zhongguancun Nandajie
Beijing 100081 - China

Director: Chen LI
Fax: + 86 10 6841 9271
E-mail: interco@nlc.gov.cn

ASIA

NATIONAL DIET LIBRARY
Acquisitions Department
10-1, Nagatacho 1-chome,
Chiyoda-ku, Tokyo, 100-8924 - Japan

Director: Yukiko SAITO
Tel: + 81 3 3581 2331
Fax: + 81 3 3592 0783
E-mail: pacasia@ndl.go.jp

OCEANIA and SOUTH EAST ASIA

NATIONAL LIBRARY
OF AUSTRALIA
Preservation Services Branch
Canberra Act 2600 - Australia

Director: Colin WEBB
Tel: + 61 2 6262 1662
Fax: + 61 2 6273 4535
E-mail: cwebb@nla.gov.au