Date submitted: 14/07/2009

Rare map materials made digital: scanning and metadata to archiving and access

Kimberly C. Kowal
Lead Curator and Curator of Digital Mapping
The British Library, Map Library
kimberly.kowal@bl.uk
Tel: +44 (0)20 7412 7849

John Rhatigan
Project Manager, Vulnerable Collection Items Project
The British Library, British Collections
john.rhatigan@bl.uk
Tel: +44 (0)20 7412 7756

Meeting: **121. Geography and Map Libraries, Science and Technology Libraries** and **Division of Special Libraries**

WORLD LIBRARY AND INFORMATION CONGRESS: 75TH IFLA GENERAL CONFERENCE AND COUNCIL
23-27 August 2009, Milan, Italy

http://www.ifla.org/annual-conference/ifla75/index.htm

Abstract:

Creating access to cartographic images via digitization, and maintaining the integrity and authenticity of the digital objects and associated metadata, is a current preoccupation of map curators. Format-specific information about the complete digitisation process can aid in future planning of scanning projects to ensure reliable access, active management and long-term preservation. The recent experience of The British Library with a small project to scan maps held within rare books can contribute to this literature. Using a variety of skills and systems throughout the institution, a methodology was developed to manage each aspect of the process, from image capture and metadata creation to providing access and preserving material via digital archiving. The outcome is a unique application of existing library standards and systems, and illustrates an alternative to project-specific software programs or web pages for managing and providing access to digitised collection materials.

Introduction

Maps and other images contained within rare books can be overlooked in libraries, leaving significant materials undiscoverable and potentially vulnerable to loss. The British Library (BL) funded a project to select and digitise rare and valuable maps held within the rare printed books collection to mitigate this possibility. The project team developed a methodology to enable firm identification of particular copies within British Library holdings, a process that included identifying plates and maps in rare books, capturing security photography, and creating detailed, copy-specific descriptive metadata. The ability to exploit this information has huge implication for collection security and holds the key to the identification of disputed collection items.

Though the original project objectives were limited to security aims, it was decided during the early planning phase of the project that access to bibliographic information and images was an additional boon for researchers and curators, achievable at little further cost. The objectives were expanded, therefore, to include the creation of a catalog record in the library system, with standard MARC fields for cartographic material, and the use of the library system module for managing digital objects and associated metadata. A portion of the records were released to be made available via the library's public online catalogue with low-resolution copies of the images. The enhancement to the library's services, and user access to the collections, was an immeasurable and unforeseen benefit.

The resulting data and images were recorded using standards and methods that are well-documented, standardised and consistent, making them an attractive candidate for testing the library's emerging storage provision and archiving systems. The Library is currently using this body of data to test their emerging online tools for long-term preservation.

While not feasible for large projects, the targeted approach to digitisation may be suitable for smaller or more valuable collections that do not necessarily have a technological infrastructure beyond the library system, or that wish to integrate collections into a single resource discovery tool. In total, the project moved the library forward to achieving an end-to-end solution for digitized material, a method applicable to other small digitization projects.

Digitisation for security

The last twenty years has seen a significant number of published articles, guidelines, policies, and project reports devoted to digitising print-based material and providing access online. Discussion of the merits of digitisation as a security device, however, is more limited. In the map library community, a great deal of attention resulted from several high-profile thefts in the 1990s, and regular reports of such thefts throughout European and US national libraries and special collections continue in both the literature and online fora[1].

A number of relevant professional bodies have prepared general practical guidelines for library security of rare materials, and made recommendations re digital techniques. The MAGERT Task Force on Library Security for Cartographic Resources notes the increased use of digital imaging for security purposes, recognising its usefulness to document unique

[1] These are documented at length in library literature; see the bibliography of J. Kovarsky. Keeping it safe, keeping it available: theft prevention in special collections. *Library Student Journal*, July 2007. Online at http://www.librarystudentjournal.org as well as up-to-date online discussions, such as that led by J. Crowe, "Map Thefts". Online at http://www.mcwetboy.net/maproom/categories/map_thefts.php and T. Campbell, "Links relating to the theft of early books and maps". Online at http://maphistory.info/theftlinks.html.

attributes of an individual copy, thereby making it identifiable[2]. ACRL's Guidelines for the Security of Rare Books, Manuscripts, and Other Special Collections refers to the use of non-invasive marking and security technologies for paper to prove ownership[3]. According to Foley's findings from surveys of academic library special collections in 2005, however, adherence to such guidelines was "not the norm"[4]

This paper is meant to present a proactive application of security guidelines and policies as used in one library project, which also served as a means of improving user access.

Project aims and workflow

Security Requirements

In the event of a loss of a collection item and a subsequent court case, the Library would need to satisfy the Court that it had taken all the necessary steps to prove ownership. There is a raft of UK legislation and European Directives relevant in this field, with the British Standards Guideline: *Code of Practice for Legal Admissibility and Evidential Weight of Information Stored Electronically (PD 0008)* denoting compliance to an auditable trail. The Code notes that the Civil Evidence Act 1995 is of major importance in this area, and evinces a shift from admissibility to the evidential value or weight of documents. The latter point requires an organisation to put in place procedures that would satisfy a court of the authenticity of the document in question. It is important to be able to demonstrate the sequence of events from the point of image capture, storage and future retrieval. Issues of admissibility and authenticity may, in a court of law, turn on the need to prove that the information was stored in a proper manner.

The project's minimum security aims were to establish enough information to prove that a disputed map was library property, identify if it was removed from a printed book, and if so, the specific volume it was associated with. Additionally, it would behove us to have proof of the authenticity of any data created, by demonstrating for instance that the captured image had not been changed since its creation, and a record of who produced the information and when. Our task therefore, was to develop a method to document, both textually and visually, the multi-faceted, copy-specific features of BL collection items, and to make this information a searchable means to access the security photography and the actual printed book.

Workflow

The main decision reached in the early stage of planning was to work with the BL's existing technology. The BL Integrated Library System (ILS) is the library system used by the BL for internal "back office" functions such as acquisitions and cataloging, and for the external library functions that they support, e.g. the Online Public Access Catalog (OPAC) and the requesting feature for delivery to the reading rooms. The Aleph 500 software package, produced by Ex Libris, was acquired by the BL in 2004. There were numerous reasons for selecting this system for the project: the standards for descriptive metadata, inbuilt resource discovery mechanisms, and preservation, all discussed later in this paper, along with pure

2 Map and Geography Round Table Task Force on Library Security for Cartographic Resources. Map Collection Security Guidelines, June 2007. Online at http://www.ala.org/ala/mgrps/rts/magert/MapSecurityGuidelines2007.pdf

3 Association of College & Research Libraries. Guidelines for the security of rare books, manuscripts, and other special collections. *College & Research Libraries News* 67, no 7 (2006): 426-33.

4 A.M. Foley. Can one man make a difference?: an analysis of the effects of the crimes of Gilbert Bland on rare book and special collections security measures and a review of the evolution of recommended security guidelines. Masters thesis, University of North Carolina Chapel Hill, 2005. Online at http://ils.unc.edu/MSpapers/3048.pdf.

practicality of using a system already supported and widely used. Use of the ILS determined the key control mechanism for workflow purposes; upon creation of a record, a unique, nine-digit ILS system number is generated, and this identification was adopted for all ensuing imagery and metadata related to the item.

The project team designed a Master Control Register to log and track the progress of each collection item, from the point of selection through to storage, using this unique identifier. Once an initial, "skeleton" record was created in the system, a photographic instruction sheet that included the System Number was completed to guide photographers in Imaging Services. The unique identifier was then used to assign filenames to all digital objects at the point of image capture. This was a major change to existing working practices in the photographic department, but paved the way for the project team to implement this key control mechanism, ensuring consistency throughout.

The main steps in the workflow were as follows (Figure 1):

- Selection of item by project staff
- Retrieval of item from storage
- Registration on master control register
- Creation of record in the system for each item (thereby generating a unique identifier)
- Preparation of photographic instruction sheet
- Photography
- Documentation of technical metadata
- Transfer of data to session folder (Raw/TIFF/JPEG files)
- QA data files and associated technical metadata
- Cataloging of map or view, linking new record to host volume record
- QA descriptive metadata
- Linking of access image to record for web delivery
- Storage of data files
- Return collection item to storage
- Close master control register

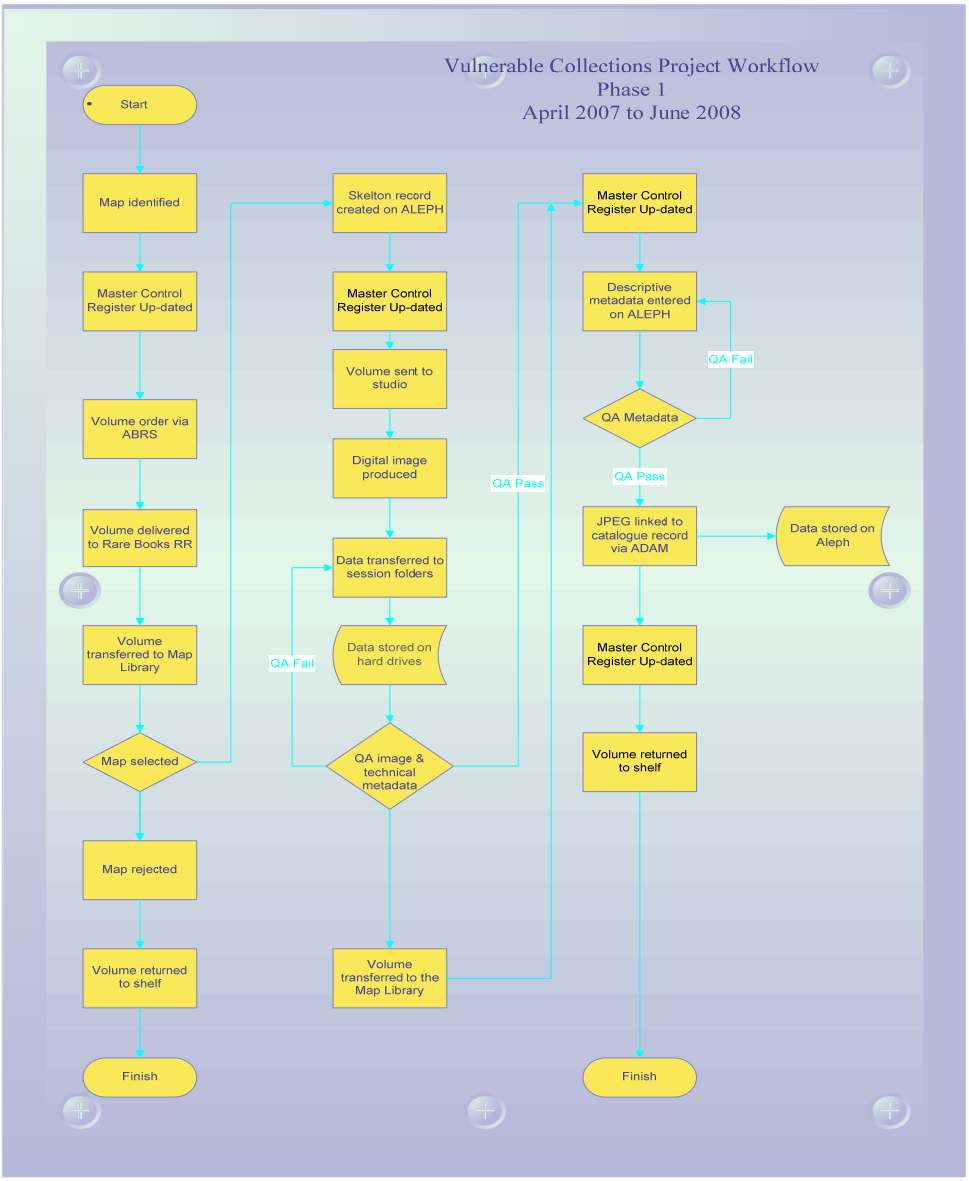

Figure 1: Project Workflow

Selection

Maps were selected for inclusion using a variety of cartographic research tools. Cartobibliographies, printed and manuscript catalogs and indices, union catalogs, and internal curatorial knowledge of the collections, were all employed to identify maps and views contained within other texts or disguised by collection-level records. Materials were reviewed at this stage to ensure they had appropriate ownership markings.

Security Photography

Security decisions

Our original approach explored the use of forensic imaging software to capture and embed a unique identifier in each digital image. This type of forensic software is common amongst law enforcement agencies (though it has a wide range of applications, including within the heritage sector) to overcome the problem of digital image integrity and ensure that evidential material is clearly identified and confirmed as the original material. In the event of a theft of a collection item, the software would provide a complete audit trail to support the chain of evidence.

This approach, which would have placed all the focus on the digital image, still left the problem of how to match a disputed map to a digital image. This would be critical if a map were stolen by knowledgeable thief with the wherewithal to remove surface markings, bleach paper, or resize the items. The security device needed to go beyond the surface by embedding a unique identifier within the item or illuminating hidden identifiers.

Backlighting provided a partial solution to this dilemma. Paper made in the 16-17th centuries has within fabric-hidden identifiers – watermarks, paper blemishes and chain lines -resulting from the papermaking process. A consequence of paper manufacturing methods of the time was that the paper retained an impression of the mould on which it was made, including an imprint of the chain lines (and the smaller laid lines) and watermark. As both are handmade, they are to that extent unique. The passage of time adds new markers. Paper folds are visible to the naked eye, but paper stains and in some cases paper weakness (appearing "see through"), is not always visible. The unique "fingerprint" of the paper is revealed by a backlit image, and recorded via digital photography, providing the means for accurate identification of a collection item.[5] (Figure 2)

Figure 2: Direct light shot with detail of backlighting, revealing the paper's watermark.

5 The use of backlighting as a diagnostic tool to date handmade paper is not new. Curators have been studying watermarks for centuries to properly identify and catalog antiquarian materials. There has been substantial work integrating digital techniques for recording this as well, combining backlighting and software manipulation to date handmade paper, and extract watermarks for research purposes. For a commercial use, see Ian Christie's 'PaperPrintTM' method of digital imaging, which includes the identification of paper as a security element of his product. Interactive Access to Early Books. Online at http://www.earlybook.info

The importance of the host environment also informed how security photographic protocols developed. Rather than saving cropped digital images of the maps, the relationship of the map to the volume was recorded, illustrating, for instance whether it appeared within the text on the page, or printed separately and inserted, and whether it was folded or laid flat in the volume. Indeed, in the event of a theft of a collection item, the value of a full and complete digital record of the item is paramount.

Technical information

The digitisation was carried out on site in the BL imaging studio at St. Pancras. The Project Team identified the best method for digitising maps, managing quality assurance, and developing procedures for the backup and storage of files. Instructions were then fed to the studio, where up to six photographers could be working on the project at any one time.

The digital image was captured on Medium Format Mamiya 645AFD camera, with a Phase One P45/P45+ digital back and Phase One Capture One image processing software. Each was captured in one shot, i.e. direct digital capture, using an auto focus camera. The one-shot functionality allowed greater speed and turnover in production numbers, provided for instant approval of the digital image, and eliminated the need for film and processing costs. In accordance with Digital Library Federation recommendations[6], resolution was set to 300 dpi, with 24-bit sampling for colour material and 8-bit sampling for printed half-tones (single channel greyscale).

Initially, up to six shots were taken of each map: recto and verso, backlit, and the positioning of the map in the volume (which could sometimes consist of up to four shots in cases where the paper was folded) (Figure 3). Following review, in the interest of efficiency this was reduced to two shots, employing two techniques - the first using direct light showing the host environment and, the second, backlighting. Backlighting was achieved by placing a 1 mm thick electroluminescent sheet under the page.

Figure 3: Folded map in it lay in the host volume

[6] Digital Library Federation Benchmark Working Group (2001-2002), Benchmark for Faithful Digital Reproductions of Monographs and Serials. December 2002. Online at http://purl.oclc.org/DLF/benchrepro0212

File Format and Standards

An uncompressed master, “raw” file, captured directly from the camera sensor, formed the basis of all photography for preservation, commercial, and collection security purposes. From a collection security perspective, it was critical that this raw data (binary code) was retained as an archival master file. Any cropping or exposure adjustments were treated as instructions and did not alter the construction of the Raw file. In the case of Phase One Raw files from the P45 camera, the size of the uncompressed files was approximately 45MB. The Raw file was seen as strictly a data file and intermediary to producing other image formats.

The output from the Raw file (which cannot be edited) was TIFF v6 and JPEG file formats. Tagged Image File Format (TIFF) is the preferred medium for long-term preservation of digital objects, and is ideal for most black and white, greyscale and colour images. The archival TIFF is uncompressed to retain the maximum amount of image data. Our project strategy was to store a high-resolution digitised master as an uncompressed raster (bitmap) image.

A cropped, medium-compressed JPEG image file was generated from the Raw file to be used as an access image for web delivery. The maximum size of the JPEG was limited to not more than 400KB, (1500x1500 pixels).

A folder was generated for each work session to manage and document the digital image capture and technical metadata. Each Work Session Folder stored all information related to a single map: the raw files with a record of any adjustments applied prior to processing (e.g. additional metadata, exposure and colour adjustment, cropping, skew adjustment etc), technical metadata, and processed images (Figure 4). From this, it could be quickly verified that the correct image had been captured and the metadata and file structure matched.

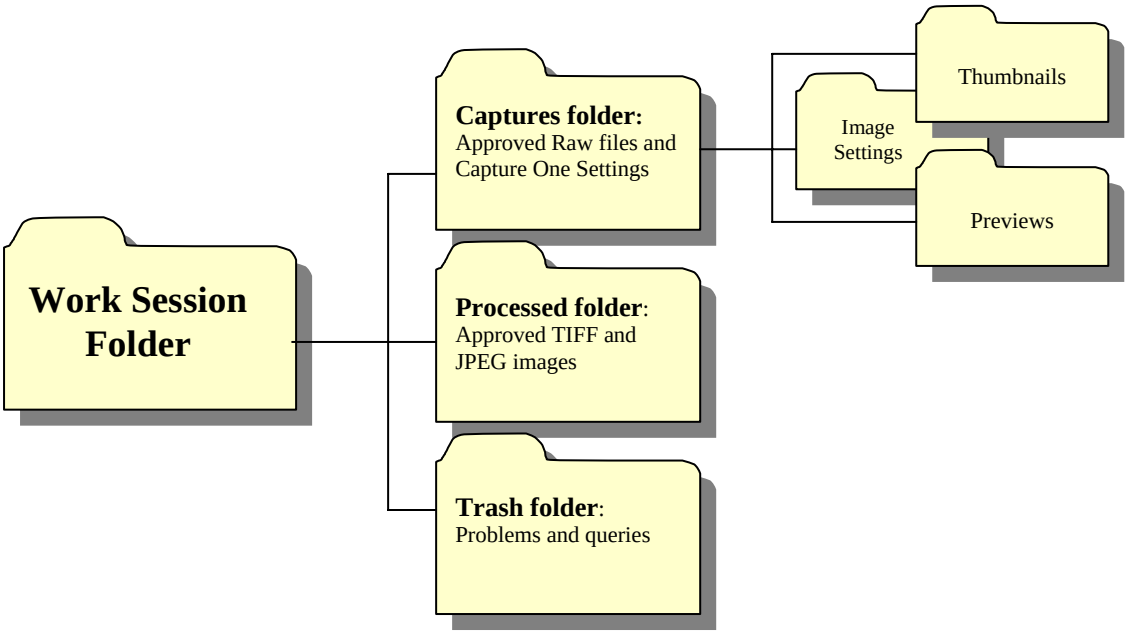

Figure 4: Model of the Work Session Folder

Quality control
To guarantee data integrity and consistency, quality control was integrated into the workflow at seven key stages in the project life cycle. Underpinning all the quality check reviews was verification that the unique identifier was consistent with the metadata (administrative, technical and descriptive) throughout the three file formats (RAW, TIFF & JPEG), and that linkages were correct. Any faults were flagged with the project manager and routed back to the photographer to correct or re-shoot.

The seven quality checkpoints were at:

1. Selection of map
2. Creation of bibliographic record in the system
3. Image capture
4. Completion of Work Session Folder
5. Derivation of alternate image file formats
6. Manipulation of administrative, technical, and descriptive metadata
7. Linking access images to bibliographic records in the system

Metadata

The decision to use the ILS System Number as the unique identifier simplified the management of descriptive metadata and allowed the newly-created records to be linked to existing records in the system representing the books in which the maps were held. The cataloguing module of the ILS supported both facets of descriptive metadata, ie bibliographical information presented according to internationally-accepted standards, as well as local notes describing copy-specific attributes of the individual maps. A complete treatment of the decisions taken in selecting cataloguing standards, and the policies and methods applied, is documented elsewhere[7].

System and standards

We wished to use existing resources in the library, both in terms of accepted bibliographic standards and established technology. The BL uses the Anglo-American Cataloguing Rules (AACR2) and MARC (Machine Readable Cataloging) format in its main system. Cataloging this rather specialised material: cartographic images within books and their digital manifestations, required extending capabilities of how we currently use the standards and system. It was therefore necessary to enlist support and advising from colleagues throughout the library for policy decisions and technical support to enable the ILS system to suit our project needs, in particular Systems Management, Bibliographic & Metadata standards, DataQuality & Authority Control, and Resource Discovery.

The need to capture and organize descriptive metadata to accompany the digitized map images meant that in order to be effective, the system needed:

- Ingest the description of the map, its bibliographic source, and the individual copy condition
- Accommodate electronic searching and access to the records and potentially images, ideally linking the two, and
- Ensure institutional long-term maintenance, preservation, and technical support.

Use of the MARC 21 Format for Bibliographic Data elements was judged appropriate for many reasons. MARC is continuously growing to accommodate new technological advancements, and so was equipped to handle the necessary hybrid of print and digital information. It is sound for preservation. Its linked content standard, AACR2, is defined,

7 KC. Kowal and C. Martyn, "Descriptive Metadata for Digitization of Maps in Books: A British Library Project". *Library Resources & Technical Services* 53, no.2 (2009): 108-120.

negating the need to develop another, and is already well-known in the BL. The format will be supported by the library for system maintenance and staff skills / training in the years to come, and when it is abandoned, its high degree of interoperability will allow easy crosswalking.

Description of the map

It became evident fairly quickly that individually cataloging each map would be the most efficient means of identifying each and recording their location and context. Creating a new record in the library system for each map using MARC format and AACR2 standards captured the essential information for searching in the event of theft. The key data elements included in the project template were common to all MARC/AACR2 catalog records for cartographic materials: title, geographical area depicted, size of map image, cartographer, date and place of publication. Further bibliographical references provided additional information that could aid in identifying the map and documenting the significance of the piece by citing published references to it.

A single record represented the paper map, and an annotation and coding were added in reference to the existence of the electronic file[8]. For searching purposes and to avoid duplication of data, the records were tagged to represent both formats of the object: as a map or view (cartographic material), as well as a digital reproduction (Computer file/Electronic resource).

Bibliographic source

Bibliographic records of the pre-1700 books in which the maps were contained, with the associated technology for tracking and ordering the objects from storage, were already present in the ILS. Linking the newly-created bibliographic map records to the existent book records would firmly establish the "parent-child" relationship between the two items, while also providing the crucial location information and ordering functionality for the child (map) record (Figure 5).

Initially, the link was a tenuous one, built by manually entering the host and shelfmark information. To alleviate this time-consuming task, a macro was developed for inserting and populating a LKR field into each record in order to generate a hyperlink between the two automatically. As a result, both the parent and child records contain reciprocal links.

8 See Option 2: "Single record with a reference to the electronic item" in J. Weitz, Cataloging Electronic Resources: OCLC-MARC Coding Guidelines. Revised 2006. Online at http://www.oclc.org/support/documentation/worldcat/cataloging/electronicresources/

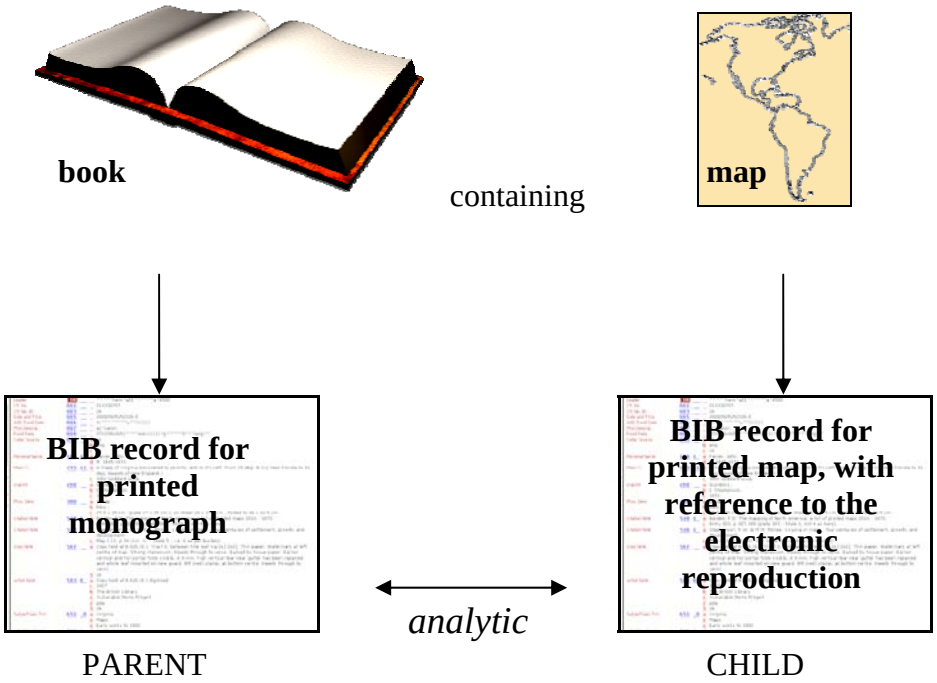

Figure 5: Relationship of records and physical objects

Item condition information

Record of the condition description of each individual map was considered to be vital to identify unequivocally the unique copy owned by the BL. Notes were included in the records with "copy specific" information, including printing anomalies; strength of impression; location of watermarks or inequalities in the paper; damage, such as stains, wormholes, or tears; or other markings including ink or bleed-through of text. The location of the map in the volume is also indicated.

In addition to the manual quality control check during cataloguing, automated batch evaluations ensured that the MARC fields used, with their standardised contents, were consistent. Batch changes were also conducted to rectify inconsistencies resulting from alterations in the project template.

Access

Use of BL standards and technology for the project expedited fulfilment of the library's strategic priorities relating to resource discovery and user access. It was judged that bringing this information in line with what the library currently offers researchers, ie web access to bibliographic records of holdings, was appropriate to support research needs and improve digital access. An enhancement of the library's system augmented the delivery of these goals by presenting a low-resolution image with the record.

Electronic searching

By selecting the ILS to store descriptive metadata, searching and access to the records throughout the BL was ensured. All of the standard fields are searchable. It also allowed the flexibility to either hide the records from public access or include them for general viewing. A sampling of records was randomly selected to be made visible via the online catalog.

Access to images

A module of the ILS, ADAM (Aleph Digital Asset Management), allows small-scale management of digital objects within the Aleph environment, and this priced add-on to Aleph

version 16.03 offered the ability to link scanned images directly to the metadata. A cropped, low-resolution jpeg of approximately 100-250 Kb was attached to each bibliographic record, and thumbnails created.

Attaching “access” images to the analytic bibliographic records served a number of functions. First, it enabled staff on the project to verify the correct filenaming assignment and image, and provided curators and Reading Room staff with an immediate and familiar tool to view maps identified as vulnerable. Secondly, rights management metadata and access control could be attached to the image. Following an initial period where the records and images were visible only in the staff GUI, partial access was extended to the web OPAC. For those records selected for public release, a thumbnail appears alongside each bibliographic record as a clickable link, which displaying the image in another window (Figure 6).

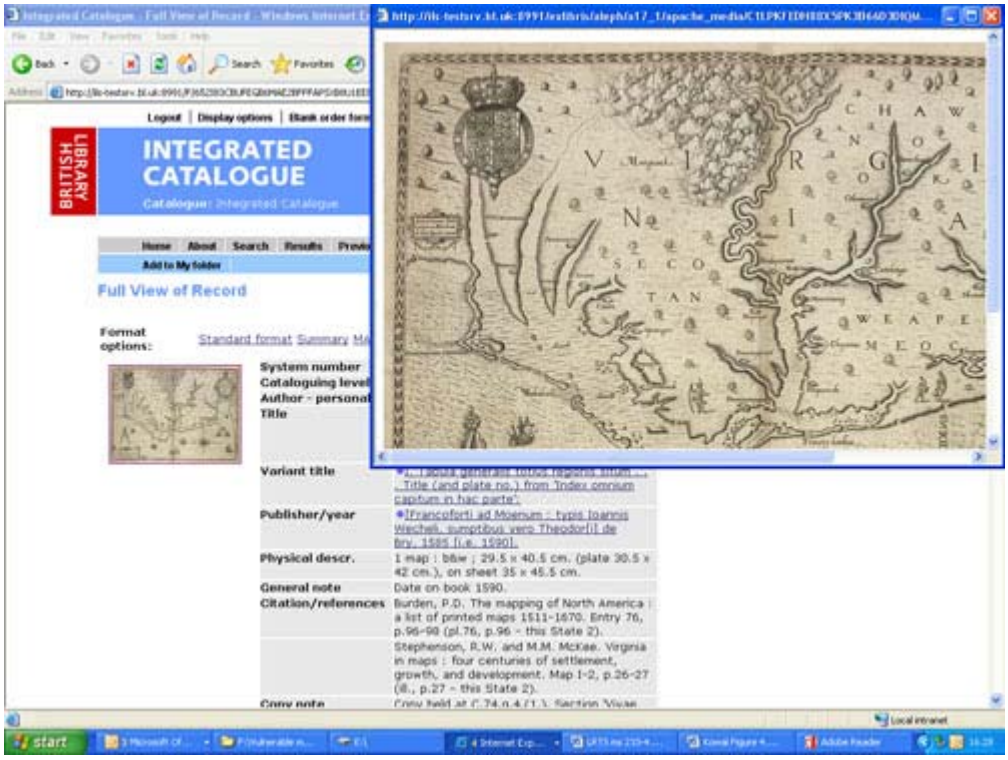

Figure 6: Display of project record in the Library's OPAC

There are numerous acknowledged limitations to this as a method of online image presentation. The ADAM technology is not meant to provide a digital surrogate of the map or act as a substitute for examining the original, and does not support interactive viewing features, such as zooming and panning. The resolution is not sufficient for detailed examination. Rather, the images serve to give an idea of the scale, generalisation, and extent of information included on the map to inform a decision as to whether the original needs to be consulted.

Rights management

Copyright information was attached to the jpeg file in ADAM. This specified that all images are copyrighted by The British Library Board, and produces a notice in the OPAC when the thumbnail is clicked on. Terms explicating BL’s privacy statement and general terms of use regarding copyright, legal disclaimer, and other policies and statements related to IPR must be accepted before images can be viewed.

Future improvements

There are a number of ways in which the user experience could be improved. Including spatial coordinates in the metadata would allow for geographical navigation in the future, offering tremendous benefits for searching. Such would expand the present reliance on one or two assigned textual Geographic Headings and enable map-based search interfaces and visualisation, or merging with other geospatial applications and mashups online[9]. The simple, though time consuming, task of inserting bounding coordinates at the time of cataloguing, using the Mathematical Data Area (MARC codes 255 and 034), would pre-empt the time when BL adopts geographic searching functionality. Due to time constraints, it was not possible to merge this into the workflow, but it would have expanded use of the records for the future.

Another, more obvious improvement would be an interactive viewing capability. Use of the ILS to display the maps required that the attached image be low resolution, and as a result text on the maps cannot be deciphered. The ability to zoom and pan online images has come to be expected by web users, and would require an additional application be added to the BL OPAC. Downloading of map images is another enhancement; while it is not appropriate for maps of this age to be georeferenced, there is still a clear demand to acquire map images as graphic files for use in geospatial applications and tools, and for use in research.

Archiving

In the long term, all the digital assets produced will be ingested into the BL's Digital Library. The Digital Library Programme is meant to provide an infrastructure to support digital assets throughout their life cycle, including preservation, resource discovery and access, digital policy and intellectual property management and administration features.

Until that time, a combination of existing resources serves as intermediate storage. Descriptive metadata is securely stored and backed-up on the library system, allowing searching and access to the records and low resolution digital images. The master digital images are stored on the BL's Imaging Services server, a SATABeast Storage System with some 24 Terabytes of space capacity. This was designed as holding space for the commercial arm of the BL to possibly generate revenue streams for internal and external customers of digital images.

The Library is starting to use using on-line tools and workflows to create and manage high-value material in a more streamlined fashion. The Fast Track to Safety Project is an initiative of the Digital Library Programme, set up to minimise the risks currently associated with intermediate storage of digitised collections. Currently, ingest flows using Ex Libris' product Digitool are being configured to merge descriptive metadata with digitised images to build digital objects. Future scenarios for access can have a big impact on ingest workflows and are being explored.

9 See M. Buckland et al, Geographic Search: Catalogs, Gazetteers, and Maps. *College & Research Libraries* 68 (2007): 376-387 and C. Fleet, Maps, mashups and metadata - geospatial data standards for access and retrieval. *WIDWISAWN* 7, 1 (2009).

Conclusion

There is still much debate about the role of digitisation as a security method. This project applied numerous established library practices and recommendations, related to ownership marking, backlighting to reveal paper materials, photography, and analytic and copy-specific cataloguing, to develop a solution to serve security aims. Insofar as that, the resulting methodology demonstrated discernment of the most appropriate techniques and an application that links directly bibliographic descriptions of both the original and the digital copies of the map, and the book in which the map appears; the digital image files captured; and the preservation strategy, making them widely searchable and visible while uniting them with the broader collections.