The Effect of Digital Technology on the Control of and Access to a Photographic Collection

Clare L. Birdsey, Andy Golding, and Ralph E. Jacobson, University of Westminster, Harrow Campus, UK

Abstract

This Digital Image Archive (DIA) project involves the production of a digital archive of a collection of Nineteenth Century photographic images and equipment. The DIA was designed to operate as a public access catalogue, to aid curators with information retrieval and to reduce the need to handle some original material. Much of the William Henry Fox Talbot collection at the Royal Photographic Society can no longer be exhibited due to damage caused by handling and light. Access to the collection is restricted because of preservation concerns over a growing percentage of the collection and insufficient resources for reproductions and supervision. These problems led to investigations into the possible advantages that digital technology might offer. Cross-referencing, administrative cataloguing and recording of preservation techniques are ideally suited to digital technology. The short life span of digital storage media and concerns over system compatibility requires the development of a complex long-term digital strategy that must be standardized if the archival material is to survive. Construction of the DIA involved investigations into software and hardware configurations concentrating on the ability to customize factors such as indexing from recognized standards and practices. This paper concentrates on the results of a survey into standards within digital indexing systems, which indicates that although there is growing awareness of standardization within digital cataloguing and multimedia, many organizations continue to devise independent systems.

Introduction

Providing archival material across locations and to wide audiences is a continuing concern for many organizations. Most collections contain a variety of media types, often relating to each other, all requiring the creation and storage of documentation. The quantity and diversity of information created is a factor that information management specialists have been concerned with for some time (Borko and Bernier, 1978). However, for researchers, the organization of information is only significant if there are efficient tools with which to retrieve it. The development of digital technology as a collections management tool and public access catalogue now enables most traditional retrieval mechanisms to be incorporated into one system. Combining new technology with a variety of different media is a complex task, which requires the production of a long-term digital strategy. Computer technology can appear to create a 'super' system that will solve all the problems encountered with paper methods (Geis, 1993, Jones, 1997). The reality can prove more time consuming and less beneficial than a re-organization of existing procedures. Traditional paper systems are difficult to manage but involve complex and integral structures created over a long period of time. The strengths and weaknesses of these methods should be observed and recorded before another system is implemented.

The production of standards and guidelines for libraries, museums and archives has been an important area of research since the nineteenth century (Borko and Bernier, 1978, Ritzenhaler, 1984). Computer orientated equivalents were created in the 1960s (Smith, 1990) and have been developed for different countries and organizations (ANSI-AIIM TR40:1995). This area has recently become more important to developers of digital systems (Blackaby and Sandore, 1997). If designed correctly a digital catalogue can operate as an accurate and reliable retrieval tool, however, an incompetent indexing system can prove more frustrating than paper methods.

Digital Image Archive Project (DIA)

The DIA project began in 1997 and was designed to assess two main subject areas. Firstly, image quality for the digitization of photographic material and its subsequent display, and secondly, the stages involved in disseminating the digital information to a wider audience. The observations noted in this paper stem from this second stage of research and concentrate heavily on the use of standardization in digital indexing systems.

Initial stages of the project involved identification of a photographic collection. The Royal Photo-

graphic Society (RPS) collection was selected as a suitable candidate for investigation. This collection was formed in 1853, with Queen Victoria and Prince Albert as Patrons. The collection holds over 120,000 photographic images from 1827 to the present day including experimental salt prints and daguerreotypes by early developers of the medium (The Royal Photographic Society, 1984). These and other materials are extremely fragile and access to many items is restricted. The RPS is a non-profit making charitable organization and resources for reproductions, supervision of original material and public access to the collection, is limited. Consequently, access to the collection is confined to bona-fide researchers and members of The Society. This lack of accessibility to the collection has been a concern of the curators for some time and efforts to reorganize the methods for information retrieval were already underway prior to commencement of the DIA project.

It was decided that the William Henry Fox Talbot Collection (WHFT) within the RPS collection as a whole represented a manageable group for assessment. It consists of six hundred photographic images, twenty items of photographic equipment, hundreds of hand written letters and notes, and multiple forms of documentation. There are also paper prints from a computerized text-only database and 35mm transparencies of the images and equipment.

The management of the WHFT collection currently suffers these key limitations: access to some works of historical importance is restricted due risks of contamination; few reproductions are available for researchers and the public; staff cannot cope with the volume of requests for original material; and catalogue listings and databases do not always interrelate.

Investigations into construction of the DIA initially concentrated heavily on the practical considerations of hardware, software and storage media. At an early phase of the project it was decided that adherence to international standards and recognized practices would increase the potential life of the digital files (Birdsey et al, 1999). File formats, storage media and format, color calibration in scanning, and processing, were adopted in accordance with documented practice and standards (Mohlhenrich, 1993). These are important considerations in the digitization of documents and were used to enable future users to retrieve the stored data. Hindsight has shown that file formats, storage media, and technological capabilities are never constant factors with digital technology. It is for this reason that international standards committees and groups attempt to determine the recommendations for practice in this field (Cornell University Library, 1993).

During early research into the method of information retrieval from the DIA, it became evident that it was not a matter purely of hardware configurations and that the incorporation of digital technology alone would not improve access to the WHFT collection. Effective tools to retrieve the data once digitized proved to be vital (Cawkell, 1993). The emphasis turned toward customizing software to integrate cataloguing guidelines and thesauri. The existing database of the WHFT collection is custom designed and contains sixty one fields with some relating to obscure entries in publications and old catalogue numbers. Most of this information is essential and documents important details about the objects and related items. It is the poor method of organization that does not enable users to browse or search for information. Figure 1 displays an example of the relevance of the information recorded for each image:

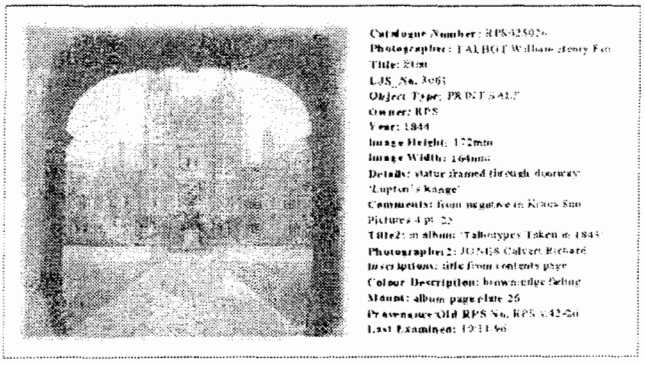

Figure 1. Data Fields containing information for the corresponding image.

The availability of the 35mm slides and database printouts does allow experienced researchers with 'exact' requests to retrieve information. However, the ability to cross-reference greatly enhances research possibilities, the curator of the RPS considers this vital to the effectiveness of the collection.

> This homogeneity in a collection of any kind is rare and very important for research. Individual items are important in their own right but it is the exploration of the connections between them that increases our knowledge of the history of photography. (The Royal Photographic Society, 1984a)

The next stage of the research project involved the creation of a successful information retrieval system that would convert the existing database to recognized standards or practices. A survey was carried out to determine the use of cataloguing guidelines and thesauri within prominent museums, archives and libraries, in the UK. The results would then be used to assist in the construction of an appropriate system for the DIA.

Assessing the Use of Standardization for Indexing Systems

The use of recognized practices and standards in digital technology is becoming more accepted (Getty, 1995) both for hardware and software specifications. Crucially, it is the need to share information with other organizations that drives this process. The availability of material in remote locations has introduced the need for compatibility and accessibility of digital records. With hardware specifications, the method of scanning and storage can be accurately noted and then followed by future users. However, when indexing a collection, the idiosyncrasies and inconsistencies of indexing methods often renders material difficult to retrieve (Hutchins, 1977, Shatford, 1986). Although digital technology provides a faster mechanism for searching through material, the fields and words used to describe it are the same as paper-based methods.

Growing awareness of the need for related cataloguing systems between digital and paper records, has led to the production of cataloguing standards and guidelines. US/UK MAchine-Readable Cataloguing (US\UKMARC), Anglo-American Cataloguing Rules (AACR2), Standard Generalized Markup Language (SGML), SPECTRUM: The Museum Documentation Association Standard (SPECTRUM), and standards by the British Standards Institution (BSI) and the International Organization for Standardization (ISO), all offer possibilities for bibliographic and photographic catalogues. These guidelines structure records into recognizable, documented formats to enable different users and systems to read the same information. Guidelines such as MARC incorporate codes that represent unique fields within the catalogue. Others, such as the BSI and ISO standards give advice on field choice but are less structured.

At the RPS, records for different collections are not always interconnected and the incorporation of new media is creating more formats. Approximately 20% of the collections are catalogued using catalogue cards, lists, or text databases. These have been created at various times and by several curators. The central reason for record inconsistency are that different catalogues refer to specific collections or object types and have been constructed at different times and by many curators. To combat the diversity of document formats, efforts have been made to adhere to the MDA standard SPECTRUM. The work of the MDA and the Museum and Galleries Commission (MGC) has raised the importance of standardization for many collections in the UK and they operate as a general source of reference for many organizations. An essential factor in the work of the MDA and the MGC is the development and implementation of standardized practice in archival documentation.

To review these issues and address the different cataloguing systems used, a survey was conducted to determine what standards and guidelines to apply to the DIA. Eighty organizations were contacted and fifty-four used for compilation of the results. The following list displays the types of organizations involved as percentages:

- Museums, 42%
- Libraries, 35%
- Institutes and Societies, 7%
- National Record Offices, 6%
- Archives, 4%
- Trusts and Centers, 4%
- Galleries, 2%

The survey assessed collection size, media types held, the existence of reproductions and the use of cataloguing software. However, the following areas were used to determine the DIA system: what types of catalogue(s) are used; what standards or guidelines, if any, do these catalogues conform to; what thesauri, if any, have been applied to images within the collection; and how effective this thesaurus is for information retrieval?

Results from the first set of questions revealed that 45% of all organizations used no cataloguing standards or guidelines. Although a large percentage of organizations preferred independent systems, guidelines for the DIA were not determined until a review of all results was complete. Figure 2 displays the results of catalogue choice.

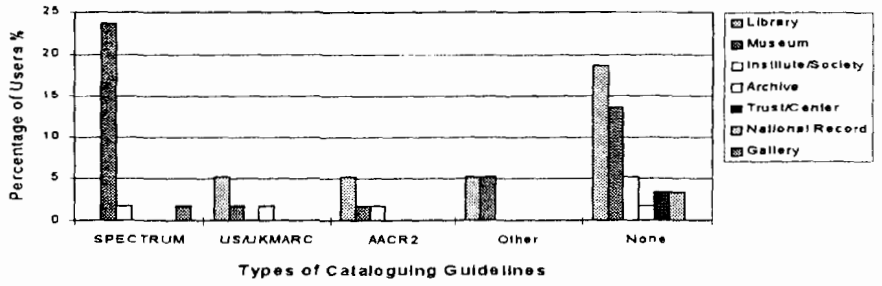

Figure 2. Cataloguing Guidelines Chosen by Different Types of Organizations

Of the 55% of organizations, who did use cataloguing standards or guidelines, 27% used SPECTRUM, of which museums formed 24%, institutes and societies 2% and galleries 2% respectively. The work of the MDA and MGC has improved awareness of standards in archival documentation and this was supported by the results. Museums and libraries appeared to use very different systems. With 10% of libraries using either MARC or AACR2 rules, and museums using SPECTRUM or a custom system.

The next stage assessed the use of established thesauri constructed by notable institutions. Several options were available. Established thesauri such as, the Art and Architecture Thesaurus (AAT), by the Getty Information Institute, the Thesaurus of Graphic Material (TGM) by the Library of Congress, and the British Museum Materials Thesaurus. Guides to thesaurus construction were also considered, these included the Museum of London initiative to combine the AAT with custom terms (Holm, 1993), the MDA Terminology Working Group, and standards by the BSI and the ISO (BS6529:1984, ISO2788:1986).

Only a small percentage of organizations surveyed used a previously designed thesaurus for their collections. 4% of all organizations used either the AAT or the TGM. The use of independent or custom designed thesauri (10%) indicated that many organizations still prefer to design their own; reflecting the decision to create independent cataloguing systems. Over 25% of organizations that used an independently produced thesaurus stated that it was either 'good' or 'acceptable'. Further investigations revealed that established thesauri often form the basis for an independently produced thesaurus, listed as 'other' in the survey. This method is used in the Museum of London example. Throughout the types of organizations this hybrid approach, of using established thesauri and custom designed keywords, proved the most popular, with 15% considering it to be 'excellent' and 27% as 'good'. Figure 3 displays the thesaurus choice for each type of organization.

Results from the indexing survey show that standards and guidelines are followed but a high percentage of organizations created independent systems. The first series of results, cataloguing guidelines, was very positive. However, the use of SPECTRUM by growing numbers of organizations may be due to pressure by the MGC to conform to recognized practices, rather than because it provides an internationally recognizable format that will easily migrate into future systems.

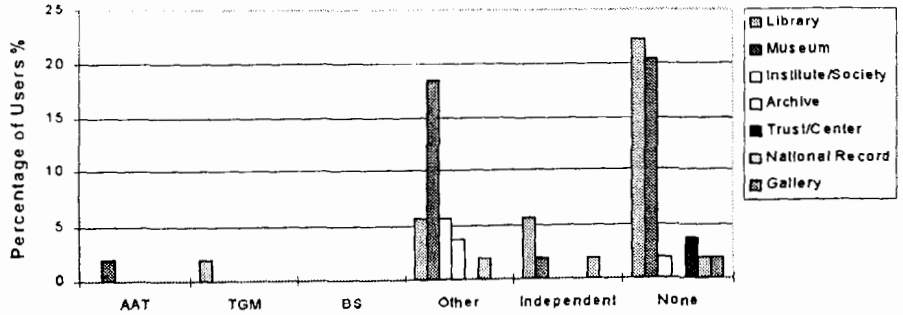

Figure 3. Thesauri Chosen by Different Types of Organizations

Although 45% of those surveyed claimed to use no cataloguing standards or guidelines, it has not been determined what percentage may have based their systems on several recognized methods. This area was not accurately determined and results for this section could have been improved. The cataloguing standard chosen for the DIA was SPECTRUM. The RPS has been attempting to conform to these requirements for some time and no evidence gained through this survey indicated that with the resources available, a change from this strategy would improve organization, or the future accessibility of the collection. The growing popularity of SPECTRUM, if only nationally, also proved a significant factor in the choice.

Results from the thesaurus section of the survey were less positive. Although there was widespread awareness of established thesauri, the percentage of use was very low. This may be attributable to the use of multiple thesauri, combining an established thesaurus with custom keywords. Although technically independent, many adhere to a hierarchical or parent-child structures (Will, 1992). At first, the choice of thesaurus seems diverse, however, many are constructed on the basis of similar or the same international standards (BS6529:1984, ISO2788:1986). Selection is therefore not confined to one thesaurus. A clear choice of thesaurus, for the DIA, could not be determined from the survey. What became apparent was the need to use multiple thesauri and attempt to constrain them within an overall structure. Consequently, the Library of Congress TGM was chosen. The need for descriptions of photographic processes and equipment, and artistic works, was met by the TGM. However it was felt that the latter category could require input from the AAT at a later date, which will then be confined to the TGM data structure.

The indexing systems, for the WHFT collection at the RPS now conform to one recognized standard and at least one recognized thesaurus. It is hoped that these efforts will enable future indexes to be more consistent with image classification. This will ultimately result in a clearer understanding of the collection by researchers attempting to retrieve information. An on-line database to the WHFT collection will be installed at the RPS this year and an assessment of the effectiveness of the indexing system chosen will then be reviewed and published.

Discussion

This project has been largely concerned with developing a new index for one non-standardized collection of photographic material. However, the integration of existing indexing systems, to produce a hybrid system that will enable successful information retrieval and reduce the cost of information management, is an important consideration for most organizations. The standards and guidelines applied to the WHFT collection should mean that for future, similar projects, integration would be readily achievable. Documenting archival objects using controlled index fields and the use of cataloguing standards or guidelines to constrain field choice should provide consistent, accurate records. Unfortunately, individuals make highly idiosyncratic choices when implementing cataloguing systems and thesauri for image description. Shatford (1986) elaborates here on the complexity of determining the subject of an image.

> There is no simple rule for determining the "principle or central subject" of a picture, but the indexer must make the attempt by asking the question: "What is this picture, viewed as a whole, Of?" Or, viewed as a whole, "What is this picture, About?" (Shatford, 1986)

Although the keywords chosen to describe an image may never be standardized, efforts have been made to determine thesaurus structures. Using a range of hierarchical thesauri should improve subject choice for each archival object. (Conclusions on the effectiveness of this method, for the WHFT collection, have not yet been drawn.) Ideally, an index created at the point of catalogue construction will be highly consistent. Unfortunately, the necessity of integrating existing indexing systems often creates overlaps in catalogue structure and keyword description. Standardizing multiple records can rarely be afforded by small, charitable organizations and applying standards and guidelines to single collections that will integrate over time, is one method that is often used.

Organizations considering indexing their collections face a bewildering number of options in terms of computer hardware and software. How does the developer choose amongst the vast range of inexpensive indexing software available? Software companies are aware of the use of standardization in these fields and there are examples of successful digitization projects, using collection management software, that adhere to standards and guidelines (Fitz, 1995). However, the availability of many digital products that can reorganize and improve accessibility to information does not necessarily justify their use. The ability to use

standards and guidelines within software applications will enable future migration and integration of records.

Ultimately, a review and assessment of existing projects will provide a knowledge base of options for future developers and there is a need for project managers to publish the results of digitization projects. The conclusions of this project have shown that insufficient use of previous knowledge can culminate in the production of impractical indexing tools. Furthermore, although standardization appears to offer improved methods for information retrieval, many organizations still favor independent systems. The use of such a variety of methods to complete the same task appears to be based on the idiosyncratic needs of each organization. Although beneficial at a local level, this can only result in more confusion for users of remote digital environments. Increased access to information can be gained through the use of a digital system. However, if insufficient research is carried out into the structure and format of record documentation, then information retrieval in these 'super systems' can prove less effective than the methods they are designed to replace.

Conclusions

On the basis of the results obtained from the indexing survey, the following conclusions are drawn.

- It was concluded that the straightforward reproduction of original material into a digital format and use of computer hardware would on its own not solve the problem of increasing access to material at the RPS. Assessment and review of previous projects revealed the additional benefit of standards in the indexing systems of the DIA.

- A central reason for ineffective information retrieval is the existence of documentation in multiple formats which often results in the use of disparate indexing systems for each area of a collection. The use of standardized methods for information management can integrate media formats and aid researchers in the retrieval of cross-referenced material.

- The importance of standardization in digital indexing systems has become more widely acknowledged. Although a large percentage of organizations studied consider independent systems to be necessary, the results indicate that a more consistent system can be created through the use of recognized practices and standards.

- The use of cataloguing standards and guidelines is only of use if they are capable of change and can be flexible. The development of future standards is dependent upon the ability to customize elements of an exiting system though retaining the overall structure; this maintains consistency of records. Understanding this may encourage collection managers not to develop their own system but to customize existing standards and guidelines.

Acknowledgements

The authors would like to thank The Royal Photographic Society (RPS) for their collaboration in this project, particularly Pam Roberts, Curator and Debbie Ireland, Picture Library Manager for their tireless effort and support.

References

ANSI-AIIM TR40:1995. *Suggested Index Fields for Documents in Electronic Image Management (EIM) Environments.* American National Standards Institute - Association for Information and Image Management International.

Birdsey, C.L. Golding, A. and Jacobson, R.E. "The Suitability of Employing Digital Technology for Accessing Photographic Collections". Paper presented at Symposium Creativity and Consumption, at the University of Luton, Luton, UK. University of Luton, Luton, UK, 1999.

Blackaby, J. and Sandore, B. "Building Integrated Museum Information Retrieval Systems: Practical Approaches to Data Organization and Access" in *Archives and Museum Informatics*, 11(2), 117–146. 1997.

Borko, H. and Bernier, C.L. *Indexing Concepts and Methods.* Academic Press, New York, USA, 1978.

BS6529:1984. *Examining Documents, Determining their Subjects and Selecting Index Terms.* British Standards Institution.

Cawkell, A.E. *Indexing Collections of Electronic Images: A Review.* Technical Report 15, The British Library Board, UK, 1993.

Cornell University Library. "New York State Program for the Conservation and Preservation of Library Research Materials" in *Preserving Archival Material through Digital Technology: Final Report*. Cornell University Library, Ithaca, NY, USA, 1993.

Fitzgerald, S. "Archives Cataloguing on Computer at the Royal Botanic Gardens, Kew: Using MARC, International Standards and Unicorn" in *Journal of the Society of Archivists*, 16(2), 179–191, 1995.

Geis, G.L. "Hype and hypertext" in *Performance & Instruction*, 2, 15–17, 1993.

Getty. Getty Information Institute and the International Committee for Documentation of the International Council of Museums (ICOM-CIDOC). *Developments in Museum and Cultural Heritage Information Standards*. 1995. Internet publication at: http://www.cidoc.icom.org/stand1.htm. Last updated 29th July 1996. Site consulted 1st May, 1999.

Holm, S. *Guidelines for Constructing a Museum Object Name Thesaurus*. Museum Documentation Association, Cambridge, UK, 1993. Internet publication at: http://www.open.gov.uk/mdocassn/holm.htm. Last updated 1998. Site consulted 13th April 1999.

Hutchins, W.J. "On the Problem of 'Aboutness' in Document Analysis" in *Journal of Informatics*, 1(1), 17–35, 1977.

ISO2788:1986. *Documentation: Guidelines for the Establishment and Development of Monolingual Thesauri*. International Organization for Standardization.

Jones, M. "Dazzled by Technology? Museums and the Information Revolution" in *MDA Information*, 2(4), 25–30, 1997.

Mohlhenrich, J. (Ed). *Preservation of Electronic Formats & Electronic Formats for Preservation*. Highsmith Press, Wisconsin, USA, 1993.

Piggott, M. *The Cataloguers Way Through AACR2. From Document Receipt to Document Retrieval*. The Library Association, London, UK, 1990.

Ritzenhaler, M.L. *Archives and Manuscripts*. Cambridge Press, UK, 1984.

The Royal Photographic Society. *The Royal Photographic Society Collection*. The Royal Photographic Society, Bath, UK, 1984.

Shatford, S. "Analyzing the Subject of a Picture: A Theoretical Approach". *Cataloging and Classification Quarterly*, 6(3), 39–62, 1986.

Smith, E. *The Librarian, the Scholar, and the Future of the Research Library*. Greenwood Press, New York, USA, 1990.

Will, L. "Thesaurus Principles and Practice. Paper presented at Symposium". *Thesauri for Museum Documentation Workshop*, held at the Science Museum, London, UK. Museum Documentation Association, Cambridge, UK, 1992.