Challenges of Linking Digital Heritage Scientific Data with Scholarly Research: From Navigation to Politics

France, Fenella G.
frfr@loc.gov
FAIC, Library of Congress, USA

Toth, Michael B.
mbt.rbtoth@gmail.com
R. B. Toth Associates, USA

Hansen, Eric F.
ehan@loc.gov
FAIC, Library of Congress, USA

The Library of Congress has expanded its digital spectral imaging research of humanities artefacts that reflect the history of the United States, with the development of advanced imaging techniques that provide data for the studies of manuscripts that span the centuries: Portolan Charts – from 1320-1633, Jefferson's handwritten draft of the Declaration of Independence from 1776, and Herblock's political cartoons from 1929-2001. Using standardized digital imaging techniques, the Library of Congress Preservation Directorate is providing preservation scientists, conservators and humanities scholars with access to digital information on historic and fragile documents with conservation-safe, non-destructive technologies. This provides data for greater understanding of the original object, including revealing creation techniques, and identifying the origin of the substrate (paper, parchment) and media. The Library plans to host this data in standardized format for access as part of a broader preservation database of scientific reference materials of naturally-aged substrate, media (inks, colorants and pigments), treatment effects, environmental data and other document production and creation information. These recent advances in technology and digital access have paved the way for the improved utilization and interpretation of scientific analyses to contribute to scholarly interpretations of heritage materials.

1. Integration of Imaging and Data Management for Discovery

The Library of Congress has implemented digital spectral imaging of the following objects to collect preservation data, scholarly information and a cross-section of data on cultural heritage old and new:

- Portolan Navigational Charts:[1] Imaging is being used to non-destructively characterize a range of pigments, details of compass points, creation techniques and tools, and potential palimpsest information.
- The handwritten draft of the Declaration of Independence:[2] Hyperspectral imaging revealed layers of changes and different inks, and offered new insights into the original Jefferson text that was crossed out and overwritten.
- Herblock Political Cartoons:[3] A selection of the large original drawings were spectrally imaged to assess the condition of light sensitive inks, also revealing details of the drawings not previously discovered.

The application of digital hyperspectral imaging and associated non-destructive technical analyses to key cultural objects at the Library of Congress required the integration of complementary data to address a wide range of preservation and scholarly challenges. The advanced imaging data is incorporated into ongoing humanities studies of objects, generating digitally linked data sets in standardized format for Internet access. The non-destructive imaging capabilities allow researchers to characterize pigments and media on the artefact, retrieve hidden and lost text, and illuminate production methods of a range of cultural objects. Characterization of a range of materials has been enhanced through the development of a standardized spectral reference sample set, virtually eliminating the need for any sampling. Integration of the data from these technological advances with information from other preservation studies, allows greater scholarly access to the information available from fragile historic documents on parchment and paper.

Application of non-destructive imaging techniques allows the equivalent of optical

archaeology of these manuscripts and documents. The profound advantage of this technique is to provide a wealth of information and data – while minimizing and preventing further deterioration of fragile heritage items through handling and invasive analytical techniques. Integrating the range and volume of data collected from any one of these objects in a cohesive data set requires the development of a spatial "map" of the object with layers of information that relate to specific points and details on the document. This range of data can include materials characterization of pigments, colorants and other components, scholarly interpretations of text, organic and inorganic compound information, topographical layering of additions to the document, and evidence about equipment and tools used in the creation of the document. All this information adds to the interpretation of cultural objects and advances the role of non-destructive heritage science techniques in humanities research.

2. Imaging

The Preservation Research and Testing Division at the Library has developed its MegaVision-Equipoise Spectral Imaging System for preservation research into a range of United States' Top Treasures and international objects of cultural import from across the vast Library collection of nearly 145 million items, including: More than 32 million cataloged books and other print materials in 470 languages; over 62 million manuscripts; the largest rare book collection in North America; and the world's largest collection of legal materials, films, maps, sheet music and sound recordings. Collecting accurate standardized imaging data with this or any other imaging system requires integration and management of three critical factors: 1) the imaging system, 2) standardized metadata capture, and 3) efficient work processes.

The Megavision-Equipoise imaging system has been customized for cultural studies, including a focus on conservation safe lighting that minimizes light on the document while generating a high resolution image. The system collects a "cube" of standard registered images from the ultraviolet through the visible spectrum to the infrared (approximately 365nm – 1050nm) with a MegaVision 39 Megapixel monochrome camera and Equipoise LED EurekaLights. Integrated side-light panels add to the previous lighting system. The inclusion of a hyperspectral scientific reference sample collection has greatly aided the characterization of a range of materials. Standardized metadata with imaging and illumination information, as well as information about the document and its content is captured with the PhotoShoot™ software, which embeds metadata in the header of the image files, based upon EXIF and IPTC standards. The quality of the spectral imaging data relies on the adaptation of proven work processes to the requirements of each object. New imaging applications can then focus on acquiring scholarly information.[4] The process flow continues beyond the actual imaging, to include image analysis, potential post processing and validation of the data – an iterative loop requiring input from a range of personnel and expertise – requiring additional time, resources, and management.

3. Data Management

For the collection, coordination, access and presentation of this data, the Preservation Directorate is developing a comprehensive approach for access to digital files in a universally accessible format. This format will utilize an RFD framework for international collaboration and ease of access, a critical component being standardization of file formats from a range of instrumentation. The approach is to integrate scientific and intellectual scholarly information, including the interpretive data required for humanities researchers to utilize the information effectively. For example, this interpretive data could involve identification of a pigment that was not discovered until after the time period attributed to the document, or lost text that changed the interpretation of the document, and provided greater insight into the thought process of the creators. The range of data is being structured in a comprehensive format with the utilization of "scriptospatial" digital objects – essentially a geospatial information system (GIS) for documents. This data organization includes addressing the challenge of presented and linking data across a collection of items to show the scientific visualization and representation of changes across geographical locations, chronological time periods, changing

use of materials, and the development of new production techniques.

4. Conclusion

These hyperspectral imaging studies from the 1400s to the twenty-first century are revealing a range of scholarly and preservation information about seminal objects that represent specific aspects of their era. Linking non-destructive scientific imaging and digital technologies with humanities research augments the preservation of these often fragile cultural artefacts, while improving and increasing access for researchers, with extensive intellectual implications. This has created a powerful tool for probing the past; revealing levels of data and raising and answering further questions about the documents, questions that could not previously be contemplated due to missing and incomplete information. The combination of scientific and humanities research allows an enhanced dialogue between researchers, harnessing the strengths of researchers and scholars in each field, and creating a more effective interpretation of data.

Often the focus and import of this exchange of information and newly acquired data is on the movement of information from scientific analyses to the humanities. However it should be noted that the flow of data is not one-way. Humanities researchers and scholars provide knowledge of past eras and culturally related information that can prove of great benefit in the interpretation of scientific analyses, such as the knowledge of local and or cultural practices, and treatises on commonly used materials and practices. The comprehensive presentation of this data in a form that allows these two complementary streams of research to be linked and integrated greatly enhances this dialogue. This ongoing and iterative dialogue is a critical component for advancing and preserving cultural knowledge throughout the centuries – not only from the past, but also through preservation research into modern fugitive materials and media. This collaborative research can only be accomplished with the ability to capture standardized images, data and metadata from scientists and scholars and integrate them into a common data set, advancing the integration of heritage science and humanities research. The combination of technological advances and structured data access enhances accessibility to original scientific data files and images. Interpretation of this data is enhanced by ease of access to integrated data files. For humanities research, this provides access to linked data files, increasing the intellectual capacity to harness and share knowledge internationally through digital and technology advances.

References

Bouissac, P. (2006). 'Probing Pre-historic cultures: data, dates and narratives'. *Rock Art Research.* **Vol. 23**: 89-96.

Casini, A, et al. (1999). 'Image spectroscopy mapping technique for non-invasive analysis of paintings'. *Studies in Conservation.* **Vol. 44**: 39-48.

Christens–Barry, W., Boydston, K., France, F.G., Knox, K., Easton, R.L. and Toth, M.B. (2009). 'Camera system for multispectral imaging of documents: Digital Imaging Sensors and Applications'. *IS&T/SPIE 21st Annual Symposium Electronic Imaging, Science and Technology.* **Vol. 7249**: 1-10.

Ciula, A., Spence, P. and Viera, J.M. (2008). 'Expressing complex associations in medieval historic documents: the Henry III fine rolls project'. *Literary and Linguistic Computing.* **Vol. 23, No. 3**: 311-325.

Dawes, S.S. (2009). 'Governance in the digital age: A research and action framework for an uncertain future'. *Government Information Quarterly.* **Vol. 26**: 257-264.

Emery, D., France, F.G. and Toth, M.B. (2009). 'Management of spectral imaging archives for scientific preservation studies'. *Archiving 2009: Preservation Strategies and Imaging Technologies for Cultural Heritage Institutions and Memory Organizations.* VA: Society for Imaging Science and Technology, pp. 137-141.

France, F.G. (2007). 'Managing digital image repositories as key tools in the preservation of cultural objects'. *Imaging Science and Technology Conference.* Arlington, pp. 117-121.

France, F.G., Emery, D. and Toth, M.B. (2010). 'The Convergence of Information Technology, Data and Management in a Library Imaging Program'. *Library Quarterly special edition: Digital Convergence: Libraries, Archives, and Museums in the Information Age*. **Vol. 80, No. 1**: 33-59.

Grenacher, F. (1970). 'The Woodcut Map: A form-cutter of maps wanders through Europe in the first quarter of the sixteenth century'. *Imago Mundi*. **Vol. 24**: 31-41.

Jessop, M. (2008). 'Digital visualization as a scholarly activity'. *Literary and Linguistic Computing*. **Vol. 23, No.3**: 281-293.

Jessop, M. (2008). 'The inhibition of geographical information in digital humanities scholarship'. *Literary and Linguistic Computing*. **Vol. 23, No.1**: 39-50.

Nayar, S.K., Branzoi, V. and Boult, T.E. (2006). 'Programmable imaging: Towards a flexible camera'. *International Journal of Computer Vision*. **Vol. 70(1)** : 7-22.

Sculley, D. and Pasanek, B.M. (2009). 'Meaning and mining: the impact of implicit assumptions in data mining for the humanities'. *Literary and Linguistic Computing*. **Vol. 23, No.4**: 409-424.

Notes

1. Portolan charts are early nautical navigational maps based on realistic descriptions of harbours and coasts. These were first made in the 1300s in Italy, Portugal and Spain (*portolan* comes from Italian, portolano, meaning "related to ports or harbours"). The charts cover the period from 1320 to 1633, and were created on vellum or parchment.
2. The draft of the Declaration of Independence was handwritten in 1776 by Thomas Jefferson in iron gall ink on laid paper, with corrections and changes by Benjamin Franklin and John Adams.
3. Created by Herbert Block from 1929 to 2001, these cartoons represented the penmanship of a man who influenced public opinion in America throughout his 72-year career.
4. As new imaging applications are integrated into the process, rigorous attention to imaging details allows an efficient capture of data with careful document handling, standardized imaging procedures (including lighting, relative humidity and temperature control) and data management. These factors are supported with environmental management, security, contingency planning and IT infrastructure.