Crosswalking or Jaywalking? The Visualization of Linked Scientific and Humanities Data

Fenella. G. France, Meghan Wilson, Chris Bolser, and Alberto Campagnolo, Library of Congress (USA)

Abstract

A critical aspect of shared data is using an easily accessible interface that is interoperable across a wide range of heritage institutions. An innovative approach to heritage science, where data is generated about the materiality of heritage materials, is linking this data back to a visual rendering of the heritage material to begin a process of linked data and integration between science and humanities. Using the International Image Interoperability Framework (IIIF), the shared canvas data model is being expanded for integrating linked scientific analyses to this digital surrogate. There are challenges with this approach for spectral imaging data due to the additional required layers of metadata in the spectral, spatial and temporal modes, which need to be consistent, and persistent, across sets of canvases.

Introduction

There is significant interest in creating new ways of accessing information about library, archive and museum collection materials. Since interactive display is a large component of how new users connect with materials, a truly interoperable interface that creates new knowledge will help garner interest in the deeper layers of information that are contained within heritage materials – the equivalent of document archaeology. These additional layers of information include scientific analytical information that adds to the knowledge of when and how an object was created, whether it has been added to or changed over time, and are the materials that it is composed of commensurate with the time period.

The underlying database structure is known within the Library of Congress as the Center for Library Analytical Scientific Samples – Digital (CLASS-D). This database has been created not only to contain bibliographic information about the objects housed in CLASS, but also to encompass scientific metadata, including reference information about the instruments used in testing, standard operating protocols for the scientific instrumentation, records of past and presents research projects carried out by the Preservation Research and Testing Division (PRTD), the results associated with those projects. Additionally, attachments include images of objects or data files from experiments, publications, and any other associated metadata files that could potentially be useful for clarification for future researchers accessing the data. This database will serve as a reference for the development and deployment of similar resources in other cultural heritage institutions, and support the dissemination and knowledge of preservation strategies to scientists around the world. The development of dynamic linking tables and forms that associate the related samples, instruments, projects, files, and results with one another in CLASS-D is crucial in demonstrating how the application of data science to preservation science can expand the latter field [1].

One of the challenges with interoperability is resisting creating a new database, user interface (UX) or infrastructure that does not allow for ease of interconnection and sharing between diverse and distant heritage institutions. These initiatives are always well-meaning, but often lead to unsupported non-sustainable siloed structures that cannot easily be integrated into an existing institutional IT structure. Engaging with the International Image Interoperability Framework (IIIF) had the function of starting the process working with an existing internationally recognized image sharing protocol that many institutions were using in some capacity. While the existing capabilities of IIIF for including visualization of scientific data do not fully meet the needs of the data, it has allowed for rich engagement with colleagues to begin the work, and look at what is needed to adapt the canvas for the inclusion of not just heritage, but potentially many other scientific data layers.

IIIF has a mission that the access to image-based resources is fundamental for the diffusion of cultural knowledge, as well as research and scholarly communications [2]. So much of what we do now is based upon the use of digital images as a surrogate for heritage and other research materials, including photos, books, newspapers, manuscripts, maps, music and archival materials. Through IIIF, an expanding community of the world's leading research libraries and image repositories has embarked on an effort to collaboratively produce an interoperable technology and community framework for image delivery. Since this is being widely used in many national and international repositories such as the National Gallery of Art, Yale University, Europeana, DARIAH, etc. this seemed to be a good framework to build upon, rather than trying to create something entirely new. IIIF states the following goals:

- *To give scholars an unprecedented level of uniform and rich access to image-based resources hosted around the world*
- *To define a set of common application programming interfaces that support interoperability between image repositories*
- *To develop, cultivate and document shared technologies, such as image servers and web clients, that provide a world-class user experience in viewing, comparing, manipulating and annotating images*

These goals align with the challenges and needs of linking and visualizing the connections between scientific and humanities data. At the IIIF international conference in June 2017 a group formed to discuss and expand the incorporation of scientific data into IIIF applications.

Visualization Components

IIIF requires a single image to be the linking component for any other layers. It was decided that the simplest image that best aligned with other IIIF images would be the full spectrum color rendition of the object (in this case we are generally referring to flat manuscript or singe page images, 3D for objects is also underway). The expansion of digitization of images is the inclusion of spectral imaging, adding further replicate image layers that render the view of the spectral and materials response of pigments,

inks and colorants on the page in relation to the specific spectral region or wavelength illumination under which the image was taken. This has proved more challenging than initially conceived since the new version of Mirador no longer supports a "slider" view allowing the viewer to click through and view all the associated views in a simple UX.

Using spectral imaging, researchers accessing the data are provided with these new data layers that capture images of the documents in distinct wavebands of the visible and non-visible spectrum. Being able to provide this stack of images to researchers and scholars is critical for the evolution of new knowledge. Mapping on the layers of spectral images also allows integration of data from other non-invasive analytical techniques to map objects analytically. What is most desired is being able to annotate and include specific analytical data on each image layer, however this is still being pursued within the IIIF format.

The term "scriptospatial" is used to refer to a geographic information system (GIS) approach for documents, creating an interactive interface for scholars and scientists to engage with object data. Viewing digital cultural materials in multiple dimensions applies an archaeological approach, uncovering and interconnecting information strata of unique documents. Utilizing an object-oriented approach in conjunction with the data layer allows mapping of spatial and temporal data with increasing complexity for direct sharing and visualization of data. This linked interface will create a new digital cultural object (DCO) with interpretations to generate new knowledge.

This scriptospatial concept enhances cross-disciplinary research collaborations and analyses. These relationships support valuable and innovative creative approaches to data integration, while strengthening effective art and scientific collaborations. Scriptospatial mapping of data enables direct sharing and visualization of data to support analysis, with the capture of standardized instrumentation parameters and object metadata. The types of data captured and generated include characterization of pigments and inks on the object, retrieval of lost and obscured text, and illumination of creation methods. By exploiting Linked Data technologies, this information can be embedded and layered within IIIF-based representations of cultural objects, making it readily available, as annotations, via IIIF application programming interfaces (APIs).

Challenges

The difficulty is: sharing heritage science and humanities data of various formats in an easily accessible manner, and not having this separate from the collection item it relates to, i.e. disparate reports, interpretations, analyses etc. remains a huge challenge. Time and again, parts of research are published separately and not fully linked together. The data becomes separated from the actual object to a point where the connection becomes artificial and more prone to erroneous interpretations being created, especially when scholars don't have access to what should be linked data.

While the main focus of this paper is the visualization interface, the underlying CLASS-D UX was carefully crafted to assure ease of searchability, access, and retaining authority of records for accuracy and reproducibility. This UX championed dropdown selection criteria and standardized protocols and file nomenclature to assure the integrity of the underlying data.
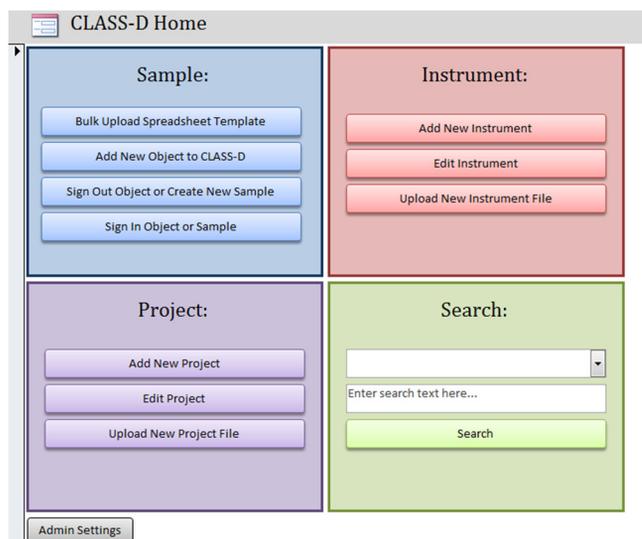

Figure 1. CLASS-D User Interface structuring access to data in entry and access modes

Why not have a visual interface that interconnects data, and a system that can also allow researchers to share, annotate, and build upon the rich layering of information and create knowledge while working in a collaborative manner? This would allow separate but related questions to be connected and associated to each other, such as:

- *How has this material changed over time?*
- *What are the original underdrawings and how can researchers share and relate these different layers to each other?*
- *Is this pigment commensurate with the time period or location that material was provenance to?*
- *Where are the other historical reports?*
- *What can new technologies and/or data mining add to the research?*

A further problem with linking scientific data from a heritage object is the three components of spatial, spectral and temporal. Spatial and spectral are more easily rectified on an image or canvas, but the temporal aspect provides additional challenges. For example, mapping and tracking changes over time for following treatment options and impact, layering event-driven data and historical changes such as maps, or the impact of environment (whether storage or exhibit) add to the dimensionality required, as well as the richness of the linked data [3]. An extension of the spectral and spatial is the conundrum of the rich point cloud data generated by 2D and 3D data sets, added to the need for higher resolution, non-static images [4].

An underlying linked database of scientific data infrastructure has been created but the challenge of true interoperability and access to data between institutions and nations remains an enduring issue [5]. The European Research Infrastructure for Heritage Science (E-RIHS) has recognized this problem and convened a group to start to address the necessity for repositories and data tools to be available, but the outcome will be many years in the future. Working with existing sustainable tools and frameworks will allows colleagues to begin to collaborate more effectively and begin to highlight and potentially solve these problems.

Approach(s) / Results

Utilizing the existing IIIF Image and Presentation "Application Programming Interfaces" (API) to link, annotate, layer and add structured metadata is underway, using cultural heritage scientific, curatorial and scholarly data as a prototype. The IIIF community is moving to expand the capability for a broader range of Science, Technology, Engineering and Mathematics (STEM) data and datasets, but agree that starting with a smaller group of data to increase the functionality will be more productive.

The work is ongoing to investigate how a slider and layering component for a user to flip between various layers of spectral images – both individual wavebands and processed – and to ensure annotations can be added to a specific layer rather than just the original image on the canvas that all others images are layered from. Mapping specific materials across an image can be added as a layer to the spectral stack along with the elemental spectrum of the composition of that colorant (as shown in figure 1). Other layered analyses such as microscopy can be accessed from the left toolbar, with the inclusion of curatorial comment to the lower right. Clicking on the highlighted regions in the image exposes the different data, and allows the user to zoom in and magnify the higher specific resolution region (figure 2).

Another challenge relates to when there are data images inserted as annotations that it would be useful to be able to zoom into for detail. One of the major challenges with locating these other analyses or zoomed areas on the original canvas image is the gel-location or specific X,Y, coordinates of those analyses (top left corner, number of pixels and lower right corner coordinates). When we have high resolution areas from the spectral imaging, we have that specific pixel location, but for combining these with other scientific analyses, such as the microscopy from another instrument, the X,Y coordinates are not as easily created and registered on that base image since the resolution will be different. This challenge has been an ongoing discussion with geospatial colleagues, and as yet, it continues to prove perplexing to translate what can be done easily for google maps to a heritage material.
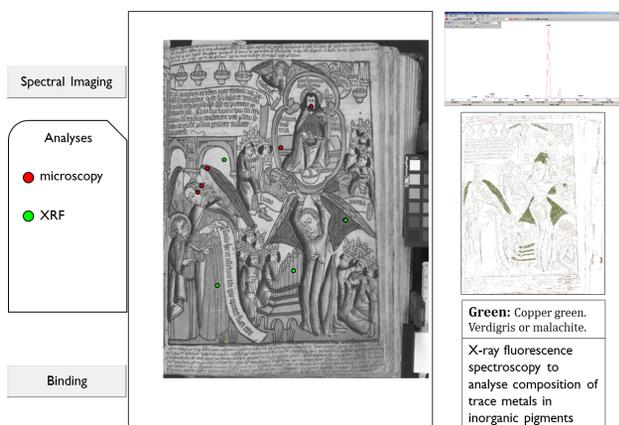
Figure 2. Creating spatially located X-Ray Fluorescence (XRF) data linked as annotation layers
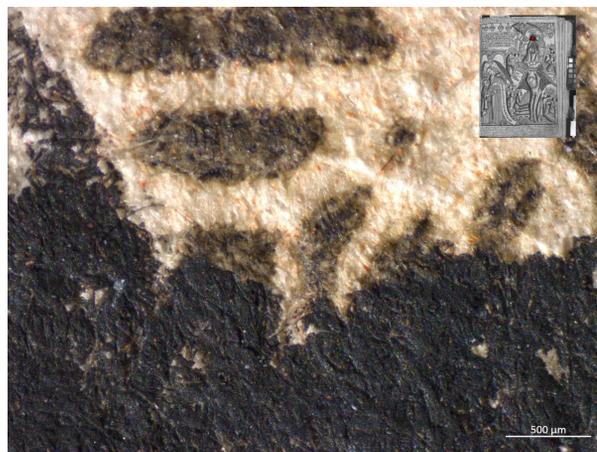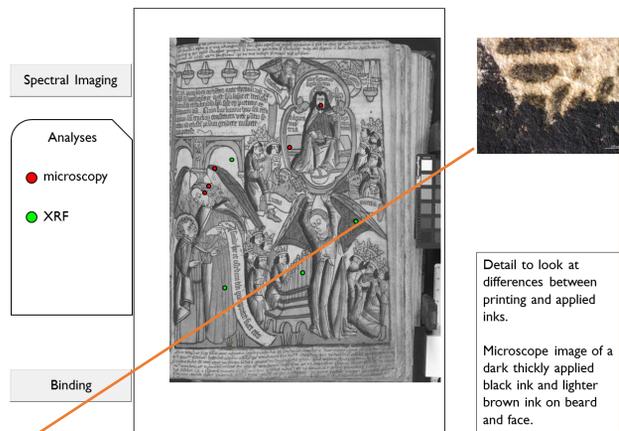
Figure 3. Creating spatially located Microscopy data magnified and expanded from annotated linked layer with location showing as red dot on original image

The approach to solving challenges with the temporal component is to potentially use the event and time stamp approach developed in the audio-visual materials IIIF initiative to add renderings of materials after time/changes through this model.
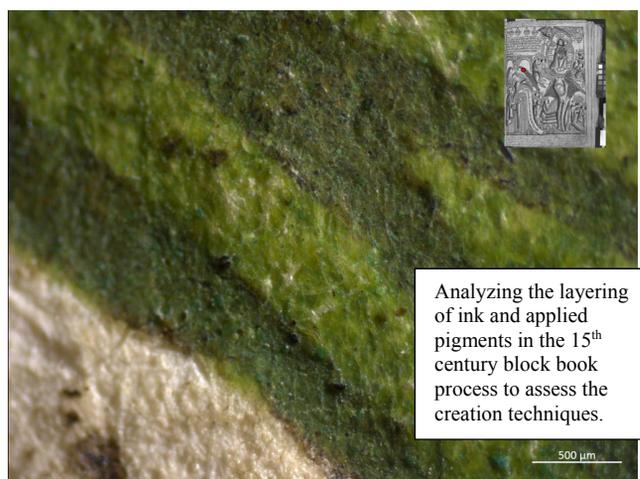

Figure 4. Adding spatially located Microscopy data with associated curatorial and scholarly interpretations

An interesting use case of a heritage institution using IIIF as a comparison tool for viewing is the Victoria and Albert Museum (V&A) Compariscope (http://www.vam.ac.uk/blog/digital-media/easy-image-alignment-with-iiif) where the video allows the user to move through and see comparisons of various aspects of images of a heritage object. The V&A are using this tool to present images or details of images for comparison, through overlaying the images and then providing an interactive control for contrasting each layer in place. This allows the user to look through an image to reveal an x-ray of the object, in this example looking at the construction of Balenciaga couture. The V&A note that their objective is not to make images perfectly align, as this would be a challenge with different imaging techniques, and this problem is similar to the geospatial registration issue we have discussed.

Figure 5. IIIF Compariscope showing the differing imaging views of Balenciaga couture.

An example of the problem of temporal challenges is: the authors have been working to link captured data from a range of historic materials in order to provide web-accessible access to information from fragile historic documents, including the 1507 Waldseemüller World Map and 1513 Ptolemy Geographia. Investigations revealed links to the same printing location allowing a virtual re-connection of previously disparate collection items. Developing an object-oriented approach to data access and sharing requires integration of spectral imaging with data from other sources in a variety of formats. This requires effective spatial metadata to allow linkages to specific locations within the images. This is necessary not only to register locations on the same section of a manuscript leaf in various spectral bands, but also to link other images and transcriptions with the spectral images. Based on geospatial mapping and layering of data used to identify points on satellite images, the same technologies, work processes and skills can be applied to spectral images of manuscripts and using the IIIF framework, expanding current applications to link and annotate layers of heritage and materiality data. However, in order to assess the potential of other online tools, a geographic information system the ArcGIS mapping tool, is being explored to address the above cartographic challenge [6]. While a story board can be created, again, the inclusion of an object as the "map" and central image of the visualization itself, is still not a function within the program.
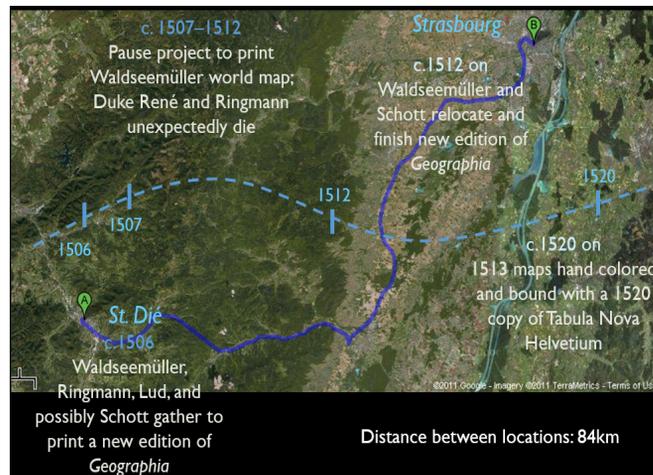

Figure 6. Adding geospatial information relating to creators of the linked collection materials

The IIPImage server framework has been designed for access to web-based streamed remote data and the visualization of ultra and high-resolution scientific imagery [7]. IIP can incorporate multiple access protocols including the internet imaging protocol (IIP) zoomify, deepzoom and IIIF. Various institutions have examples of linked fused imaging and other analytical technique data, but to date has not been widely adopted. Five positive cultural heritage applications that are of interest include: being able to compare a range of imaging modalities, image blending, spectral imaging datasets, monitoring change (through image comparisons only) and linking images and data [8]. The Centre de Recherche et de Restauration des Musées de France (C2RMF) shows this tiling and zoom tool to good effect in one image of La Vierge au viole (1500-1520) that reveals minute details in the image as well as painting and construction techniques http://technologies.c2rmf.fr/iipimage/showcase/zoom/HD7, however none of their scientific data was available online.
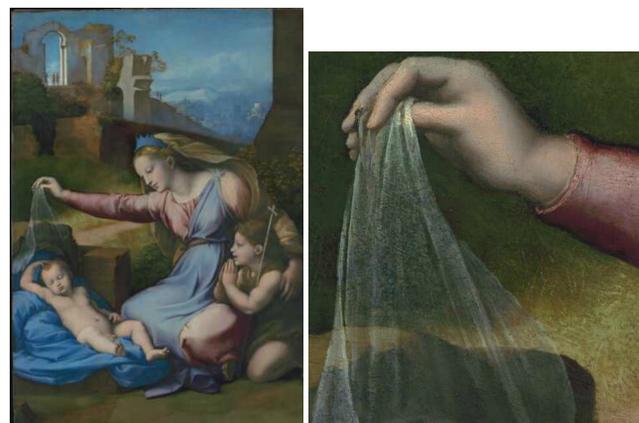

Figure 7. IIP Image Server zoom tool

An example of visualizing large scientific astronomy and biomedical image data is being examined to look at tiling and image display issues as well as the need for dynamic rather than static display for interactive analysis [9]. Extending beyond 16 bit precision scientific data to 32 bit or greater with current web technologies presented an additional recognized need for the future

as data sets continue to increase in precision, size and complexity. Adapting the above technologies to achieve and address these challenges is part of the ongoing developmental work.

Access and interoperability of data are critical elements for any imaging initiative with the establishment of standardized digital protocols for storing and accessing cultural heritage data being vital for interoperability between heritage institutions, and the preservation of international culture in libraries archives, galleries and museums. Discussions need to include those generating the data, users and potential users, and both information technology, data science and visualization specialists. A session at the Research Data Alliance (RDA) 11th plenary in Berlin, March 2018 will include consultation and dialogue with a range of colleagues from geospatial, heritage and other scientific disciplines to look at other approaches to the sharing and linking of disparate data types.

Conclusions

Increasing access through a spatial, spectral and temporal information-system approach to digital objects creates an interactive research interface for scholars and scientists to engage with the object and the full range of associated data. This framework links scientific and scholarly analyses to the creation of this new digital cultural object (DCO), with interpretations to generate new knowledge. Viewing digital cultural materials in multiple layers applies an archaeological approach toward uncovering and interconnecting information strata of historic and modern documents. This new innovative approach to engagement with digital collections reveals hidden depths of collections, engaging new users in a different world of information and object archaeology than that expect from normally flat digital renditions of cultural materials.

One of the challenges with many initiatives is not engaging colleagues in the early developments, and it seems to be more productive in the long run to share beta versions, before structures become less agile. A potentially useful way forward would be for many of us to move outside our comfort zone and think of these interoperable platforms as "living" rather than a static structure that tends to then get mired within institutional IT. These are very difficult actions to put in place, so this is simply posed as a point for discussion as we endeavor to collaborate closely and effectively.

Presenting an integrated depth of digital information can create a dialogue between researchers and colleagues from a range of disciplines. While the history and English connections may be expected, new commerce connections evolve from the knowledge of materials used (high/low cost, substandard or expensive paper), as well as engineering and artisan techniques that widen hour perception from the exposure of unexpected printing and construction techniques. We all need to truly engage with diverse audiences rather than focusing on information that might only be relevant to our field of knowledge. Providing layers of information and open data for mining, interpretation and future use, digital visualization interfaces allow and encourage collaboration across and between related as well as disparate disciplines, and provides true "added value" to existing data for reuse and greater interpretation.

References

[1] France, F.G., Online Scientific Reference Sample Collections and Shared Linked Data for Heritage Science and Related Disciplines", Proceedings of the Coalition of Networked Information (CNI), Albuquerque, NM, April (2017).

[2] International Image Interoperability Framework http://iiif.io/

[3] France, F.G., "Visualizing Conservation Science – Communicating Data through Imaging", Conservation Perspectives, Getty Conservation Institute, Vol. 32, No. 1, Spring (2017).

[4] Gámez Serna. C, Pillay, R., and Trémeau, A, "Data fusion of objects using techniques such as Laser Scanning, Structured Light and Photogrammetry for Cultural Heritage Applications", Proceedings of the International Workshop on Computational Color Imaging, Springer International Publishing, March 24, 2015, pp208-224.

[5] France, F.G., "Advances in Integrated Research Infrastructures for Science and Humanities Linked Data", Imaging Science and Technology, Riga, Latvia, May 2017.

[6] ArcGIS https://www.arcgis.com/features/index.html.

[7] IIPImage http://iipimage.sourceforge.net.

[8] Pillay, R., "IIPImage and an Analysis of JPEG2000 Encoding Parameters", Wellcome Trust, London, 10th November (2014) http://www.dpconline.org/docs/miscellaneous/events/1358-2014-nov-jp2k-ruven/file.

[9] Bertin, E, Pillay, R, and Marmo, C, "Web-based visualization of very large scientific astronomy imagery", Journal of Astronomy and Computing, Elsevier, Vol. 10, p43-53, (2015).

Author Biographies

Dr. France, Chief of the Preservation Research and Testing Division (PRTD) at the Library of Congress, researches spectral imaging techniques and addressing integration and access between scientific and scholarly data. An international specialist on environmental deterioration to cultural objects, her focus is connecting mechanical, chemical and optical properties from the impact of environment and treatments. Serving on standards and professional committees for cultural heritage she maintains collaborations with colleagues from academic, cultural, forensic and federal institutions. In February 2016 Dr. France was appointed a Council on Library and Information Resources (CLIR) Distinguished Presidential Fellow.

Meghan Wilson, a Preservation Science Specialist in the Preservation Research and Testing Division at the Library of Congress, has worked extensively with multispectral imaging technology, developing guidelines and workflows for technical operation of equipment and image quality control of capture and processing procedures. She specializes in advanced image processing techniques and is experienced in other non-invasive complimentary analyses. She has applied these skills to support cultural heritage institutions worldwide in the operation, training, quality control, documentation, and data management of spectral imaging systems.

Chris Bolser, a Preservation Technician in the Preservation Research and Testing Division at the Library of Congress, applies his forensics background to the analysis of cultural heritage materials. He is proficient in an array of advanced image processing techniques and has contributed to the development of standards and procedures of preservation imaging analyses.

Alberto Campagnolo is the CLIR/DLF/Mellon Fellow for Data Curation in Medieval Studies at the Preservation Research and Testing Division of the Library of Congress, Washington, DC (2016-2018). He trained as a book conservator (Spoleto, Italy, 2001), studied Conservation of Library Materials at Ca' Foscari University Venice (2006) and holds an MA in Digital Culture and Technology from King's College London (2009). He pursued a PhD in Digital Humanities from the University of the Arts, London (Ligatus Research Centre, 2015). Alberto has served on the Digital Medievalist board since 2014 and as Director since 2015.