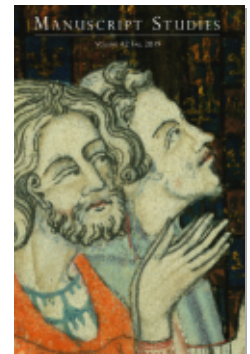*Labeculæ Vivæ : Building a Reference Library of Stains for Medieval and Early Modern Manuscripts*

Alberto Campagnolo, Erin Connelly, Heather Wacha

Labeculæ Vivæ: Building a Reference Library of Stains for Medieval and Early Modern Manuscripts

ALBERTO CAMPAGNOLO
Università degli Studi di Udine

ERIN CONNELLY
Schoenberg Institute for Manuscript Studies

HEATHER WACHA
University of Wisconsin, Madison

THE STAINS FOUND ON medieval manuscripts are signs that indicate a past life, more specifically the visible and invisible remains of human interaction over time. Reading these signals—in concert with conventional information such as script, collation, illumination, and size—can add to our understanding of their history and use. While recent work has been done on the uses of multispectral imaging in understanding the degradation and preservation of parchment, there is little preexisting scholarship on the presence and nature of stains in medieval texts.[1] Indeed,

---

1   On the uses of multispectral imaging, see A. Giacometti, A. Campagnolo, L. MacDonald, S. Mahony, S. Robson, T. Weyrich, M. M. Terras, and A. Gibson, "Visualising Macroscopic Degradation of Parchment and Writing via Multispectral Images," in *Care and Conservation of Manuscripts 15: Proceedings of the Fifteenth International Seminar Held at the University of*

the significance of stains has traditionally been underestimated. The *Labeculæ Vivæ* project focuses on those very manuscripts that are often overlooked due to heavy soiling and damage, effects that diminish their perceived quality and value. Led by a team of interdisciplinary postdoctoral scholars and collaborators, our project aimed to construct a library of stains to (1) provide an online database that will allow scholars, librarians, and conservators to better analyze materiality, provenance, use and preservation of manuscripts and early printed books; (2) document and disseminate a methodological approach for analyzing stains; and (3) provide a model for public-facing interdisciplinary collaboration. To our knowledge, this is the first interdisciplinary attempt to build a library of medieval and early-modern stains using the tools of medieval literature, medieval history, codicology and bibliography, multispectral imaging, chemical analysis, and data science. The project, supported by a microgrant from the Council on Library and Information Resources (CLIR), ran for one year, from August 2017 to September 2018, during which stains from about forty Western European manuscripts, ranging from the twelfth to sixteenth centuries, held in the University of Pennsylvania Libraries (library.upenn.edu), the Science History Institute (sciencehistory.org), the Library of Congress (loc.gov), the University of Wisconsin Special Collections, and the University of Iowa Special Collections (lib.uiowa.edu/sc), were imaged and subsequently analyzed.

Copenhagen, 2nd–4th April 2014 (Copenhagen: Museum Tusculanum Press; University of Copenhagen and the Royal Library of Denmark, 2016), 89–102; and A. Giacometti, A. Campagnolo, L. MacDonald, S. Mahony, S. Robson, T. Weyrich, M. M. Terras, and A. Gibson, "The Value of Critical Destruction: Evaluating Multispectral Image Processing Methods for the Analysis of Primary Historical Texts," *Digital Scholarship in the Humanities* 32, no. 1 (2017): 101–22, https://doi.org/10.1093/llc/fqv036. For the full data set, see A. Giacometti, A. Campagnolo, L. MacDonald, S. Mahony, S. Robson, T. Weyrich, M. M. Terras, and A. Gibson, UCL Multispectral Processed Images of Parchment Damage Dataset, UCL (University College London), London, 2015, http://doi.org/doi:10.14324/000.ds.1469099. D. Goltz, M. Attas, G. Young, E. Cloutis, and M. Bedynski, "Assessing Stains on Historical Documents Using Hyperspectral Imaging," *Journal of Cultural Heritage* 11, no. 1 (2010): 19–26, used imaging to investigate the visual properties of stains by quantifying stains in terms of optical density as means of assessing spatial distribution of the stain on the substrate.

Imaging and Analyzing Stains

There are various ways of gathering scientific data about the materiality of documents (and their stains); some require sampling and the destruction of matter, while others are non-destructive. Among the non-destructive techniques, the methods of choice are generally based on spectroscopy—that is, the analysis of the interaction between matter and electromagnetic radiation (of which light is a part). Specialized spectroscopic techniques—such as X-ray fluorescence (XRF), fiber optic reflectance spectroscopy (FORS), and Fourier-transform infrared spectroscopy (FTIR)—focusing on a point or a small area at one time reveal detailed information on the chemical composition of materials. Similarly, data from multispectral imaging, a versatile photography-based imaging technique that is typically applied to documents for the recovery of difficult-to-read information, can be utilized to map different materials on the whole area covered in the photographs. Narrowband multispectral imaging works by illuminating an object at specific wavelengths—from near-infrared (IR), through visible light, to benign ultraviolet (UV) radiations—and in capturing an image for each illumination, it results in a stack of registered photographs that are available for further analysis. Looking through the stack, one can notice how different materials react differently to each wavelength, and details that are not visible in natural light may begin to appear and be clearly noticeable under UV or IR illumination.

This project, thanks to the generous help of Michael (Mike) B. Toth from R. B. Toth Associates and William (Bill) Christens-Barry of Equipose Imaging, had access to state-of-the-art imaging equipment and software.[2] We used a Phase One IQ260 Achromatic camera, a 60 megapixel 16-bit monochrome digital back with a 8964 × 6716 pixel CCD array at 6.0 micron pixel size, with an iXR body and 80 mm lens producing 675 ppi resolution images (fig. 1). The narrowband special illumination necessary for multispectral imaging was provided by a third-generation LED light system designed by Bill

2   R. B. Toth Associates: http://rbtoth.com/spectral-imaging--partners.html. Equipose Imaging: https://web.archive.org/web/20171108182620/http://www.eqpi.net:80/eqpi/.
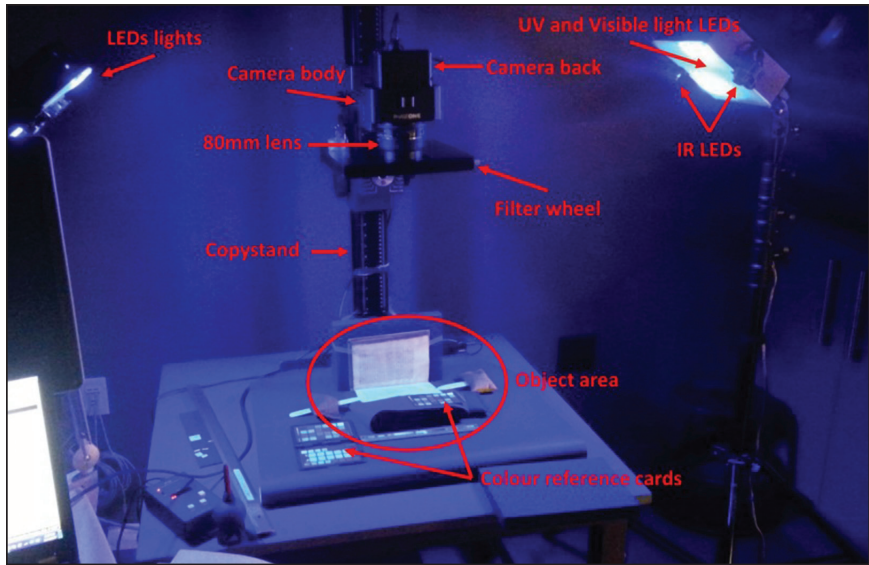
FIGURE 1. Narrowband multispectral system setting at University of Pennsylvania Libraries, November 2017.

Christens-Barry that produces very specific and narrow bands of illumination, ranging from ultraviolet light (370 nm) to the near infrared (940 nm).[3]

Because of the nature of the project, we also utilized long-pass green and red filters to detect fluorescence energy: the filters remove the illumination wavelength, but let through longer fluorescence emission that can be recorded in the captured image, thus allowing the characteristic spectra of substrate, colorant, or contaminant substances to be more completely determined and analyzed. Capturing the photographs (and managing the metadata) is only the first step. For a deeper understanding of the data recorded and the variety of material responses to the different wavelengths, one needs to process the stack of images and analyze the data through statistical

---

3   We imaged at **370 nm (UV)**, 448 nm (deep blue), 476 nm (blue), 499 nm (cyan), 519 nm (green), 598 nm (amber), 636 nm (red), *740 nm (IR1), 850 nm (IR2),* and *940 nm (IR3).* (We have used **bold** to indicate UV, roman characters for visible light, and *italics* for infrared frequencies.)

algorithms capable of simplifying it and of finding patterns in it. This kind of analysis, thanks to color reference cards positioned in the scene, can also reconstruct color images, despite the fact that the camera is achromatic— that is, agnostic to color information.

One type of output that can prove particularly useful as an investigative tool to distinguish different components—that is, materials reacting in different ways under the different lights—is the result of principal component analysis (PCA), a statistical analysis technique for decomposing a set of data into its intrinsic variability, which preserves the maximum variability of the data in fewer dimensions, from which false-color images can be generated, where different components are assigned an arbitrary color to help in discerning similar and dissimilar light responses.[4] PCA works by analyzing the light response of each pixel throughout the full stack of images, and it is therefore essential that the images are well registered for meaningful results.

Working in a similar way to PCA—that is, by looking at the spectral response of single pixels (or groups of pixels) across the full stack—it is possible to plot spectral curves that are characteristic of the material (or groups of materials) present in the selected area (fig. 2).[5] These curves are particularly useful because their shape can be used to compare and discern various materials present in a document or collection, and they therefore allow the data to be analyzed by scientists and humanists alike, fostering communication and collaboration among different fields.

---

4   I. T. Jolliffe, *Principal Component Analysis,* 2nd ed. (New York: Springer, 2011).

5   See F. G. France, "Spectral Imaging: Capturing and Retrieving Information You Didn't Know Your Library Collections Contained," in *What Do We Lose When We Lose a Library?*, ed. L. Watteeuw and M. Collier (Leuven: KU Leuven University Library, 2016), 189–97; and F. G. France, "Visualizing Conservation Science: Communicating Data Through Imaging," *Conservation Perspectives* 32, no. 1 (2017): 16–18. For a step-by-step guide on how to generate spectral curves, see H. Wacha, "Stains as Spectral Curves," *Labeculae Vivae* (blog), 29 March 2018, https://web.archive.org/web/20190115100323/https://labeculaevivae.wordpress.com/2018/03/29/stains-as-spectral-curves/, accessed 14 May 2019.
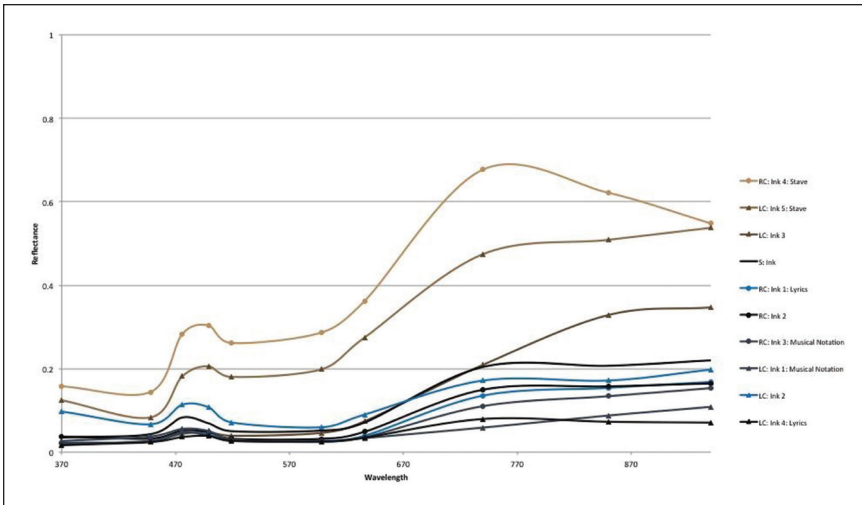
FIGURE 2. Spectral curves for ink stains on University of Wisconsin manuscript MS 170A, no. 8. We thank Leah Pope Parker for her work on the Wisconsin manuscripts.

The Ruling of SHI Othmer 1

As an example, we can highlight the analysis and investigative process for Othmer 1, a northwestern Italian manuscript, dated before the end of 1438, containing "Recipes and extracts on alchemy, medicine, metal-working, cosmetics, veterinary science, agriculture, wine-making, and other subjects." The manuscript, held at the Science History Institute in Philadelphia, is written on paper with iron-gall ink, with two columns per page, thirty lines each. The horizontal lines are ruled with a pale yellow-brown ink, at times barely visible, while only the inner vertical lines of the columns are very lightly traced with a plummet. The writing is placed on top of the ruling lines. The page is rubricated in red. Aside from the soiling from usage in the margins, and two insect holes in the middle of the page, two large stains are visible in the top part of the page. The upper one, near the head margin, is pale yellow-brown in color, while the lower one, in the gap between the two columns, has a central grayish core surrounded by a larger pale yellow-brown area.
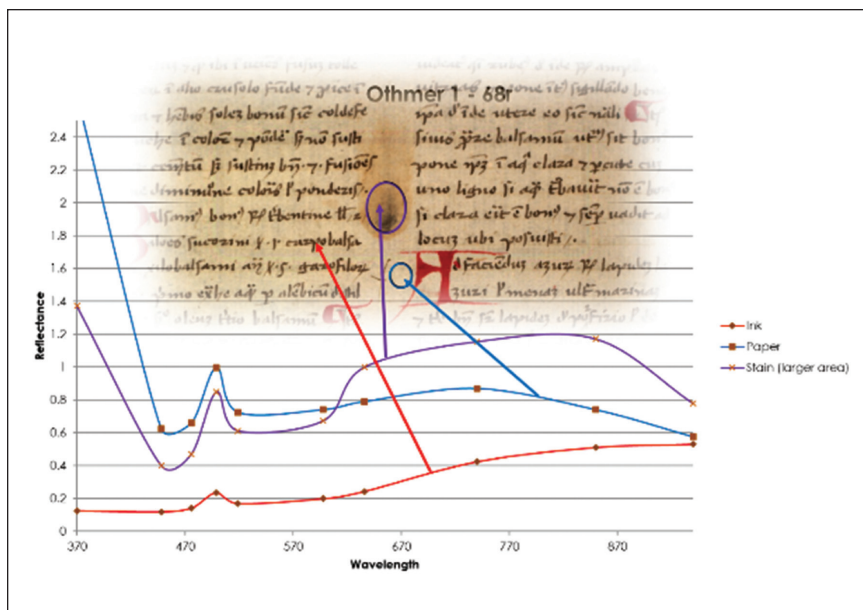
FIGURE 3. Plotting the spectral curves for the stain in the middle of the columns as a composite.

When making a first plot of the spectral curves of the various components of the page, we considered the larger bicolored stain as one component, conjecturing that the central core might be a more concentrated element of the same substance throughout the stain. This has a unique shape that was not reconcilable with anything else on the page, and followed only slightly, as is often the case, that of the paper substrate, since it did not cover it completely, masking its spectral response (fig. 3).

A PCA investigation, however, led us to notice the spectral response of the ruling ink, as opposed to the writing one. This also evidenced the pale yellow-brown components of the large stains, but not the grayish core. Further analysis showed that the central grayish core resembled the response of the writing ink (fig. 4). This led us to plot the spectral response a second time, and to separate the two components of the central stain in two different spectral curves: one for the inner darker spot, and one for the outer

FIGURE 4. False color image showing different ink components highlighted in different colors: ruling in yellow, writing ink in red, and rubrics in blue. The central stain shows similar color configurations: inner core in reddish tone and outside area in yellow.

light-colored area. The new curves followed the same respective patterns of the writing ink and of the ruling ink (fig. 5).

As is clear from the analysis, the stain must have occurred at the time of writing, and not afterwards. The stain thus reveals the work of an untidy scribe who spilled the ruling ink pot on the recto of folio 68, and then also
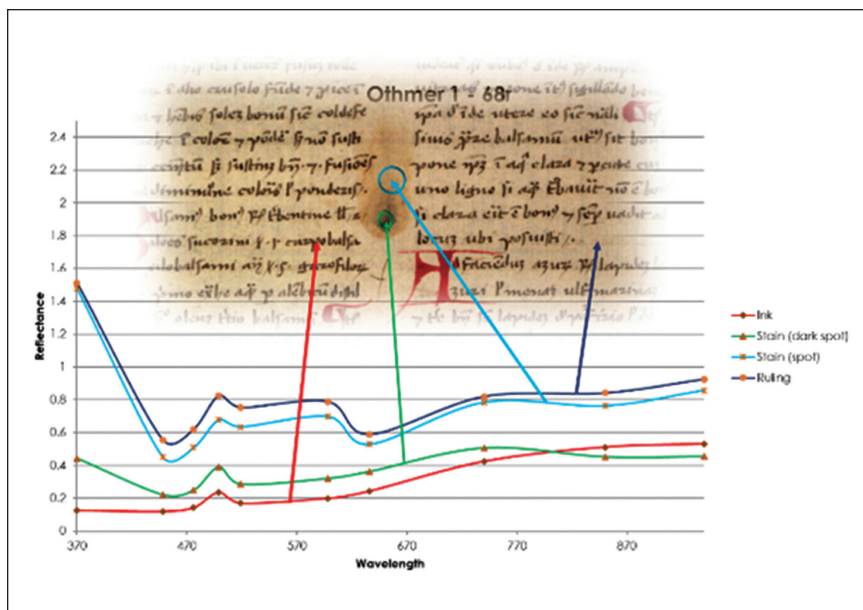
FIGURE 5. Plotting the spectral curves for each component of the central stain, revealing patterns similar to the writing and ruling inks.

dropped some writing ink on the same area, resulting in the composite stain that we can see today. This gives us clues regarding the kind of manuscript that Othmer 1 is: a working copy important for its content and not for its appearance.

Using Spectral Curves to Compare Wax Stains[6]

A University of Iowa Psalter manuscript, xMMs.Ps3, folio 86v, contains a distinct wax stain that has not flaked off the substrate. The resulting spectral curve provided a baseline from which similar curves could be measured

6   This section is excerpted from Library of Stains Data Visualizations, ed. Heather Wacha and Leah Pope Parker, University of Wisconsin–Madison, 2018, available at https://zenodo.org/record/2528377#.XEMz7M9TlR0.

and compared. Similar, but not identical, curves were identified in the following manuscripts:

A. One similar stain from the same folio: University of Iowa, xMMs. Ps3, 86v, Stain 2.
B. Similar stains on other folios from the same xMMs.Ps3 (Psalter) manuscript: fol. 82r, Stain 2, and fol. 130r, Stain 2.
C. Similar stains on other University of Iowa manuscripts: xMMs.Br2, fol. 14v, Stain 3, and xMMs.Se1, fol. *martyra*, Stain 1.
D. And in one University of Wisconsin manuscript: MS 80, fol. 110, Stain 1.

From this data, we might surmise that the stains demonstrating similar spectral curves to the known wax curve (xMMs.Ps3, stain 3) may indeed be wax. However, when all possible wax curves are combined into the same chart, slight differences in the shape and intensity of the curves occur (fig. 6). While we cannot be certain of the reasons for these differences, several options emerge:

1. The spectral curve may be of tallow or a combination of wax and tallow.
2. The spectral curve may be of wax, but the wax perhaps contains some kind of variant such as an added pigment or color.
3. The spectral curve may indicate a difference in age from the known wax curve.

While it is not within the remit of the Library of Stains project to precisely identify each stain that was imaged, the data generated by this project provides a means of preliminary comparison and can lead to new hypotheses and new research directions in the study of medieval manuscripts. It was beyond the scope of this pilot study, but diverse statistical methods have been applied to multispectral imaging data to facilitate and quantify a variety of investigations. As digital protocols in this field advance toward standardization, these methodologies—for instance, chemometrics—are certainly an avenue of further research as means to quantify, with greater certainty, the relationship and significance between the data derived from different
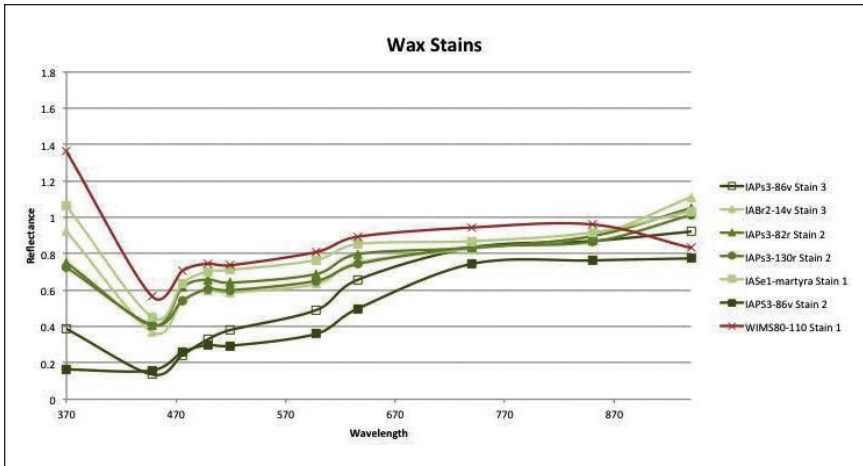
FIGURE 6. Plotting the spectral curves for wax stains in the University of Iowa and University of Wisconsin manuscripts.

manuscripts in different institutions using the same equipment and processing tools.[7]

The Library of Stains: An Online Resource for Everyone

Over the course of the project, we have collected about 220 Gb of data that is now hosted by the University of Pennsylvania.[8] The library is archived following a flat structure (figs. 7–9) similar to other prominent multispectral imaging projects, namely the Archimedes and the Syriac Galen palimpsests.[9]

---

7   On standardization, see F. G. France, M. Wilson, and C. Bolser, "Advances in Spectral Imaging Curve Analysis for Humanities Studies and Heritage Science," Archiving Conference, 2017, doi.org/10.2352/issn.2168-3204.2017.1.0.122; and M. B. Toth, "Developing Spectral Imaging: A Standardised Tool for Collaborative Conservation Support," in *Care and Conservation of Manuscripts 16: Proceedings of the Sixteenth International Seminar Held at the University of Copenhagen, 13th–15th April 2016* (Copenhagen: Museum Tusculanum Press, 2018), 309–22. On chemometrics, see F. G. France, "Visualizing Conservation Science: Communicating Data Through Imaging," *Conservation Perspectives* 32, no. 1 (2017): 16–18.

8   The repository (ark:/81431/p35d8nj0r) is maintained by https://ezid.cdlib.org/id/ark:/81431/p35d8nj0r. The citation URL is https://n2t.net/ark:/81431/p35d8nj0r.

9   For the Archimedes palimpsest, see http://archimedespalimpsest.net, and for the Syriac Galen palimpsest, see http://digitalgalen.net/. See also D. Emery, "The Galen Palimpsest and

The repository is archived under a single directory (the root) in such a way that all files either are core data or serve as support in the understanding and use of data by part of both humans and machines. The **root** directory contains the whole repository and is accompanied by a *site map* and a *checksum digest* for bulk download. Inside the top-level directory, there are two subdirectories and four files (see fig. 7): the files are an introductory *ReadMe* file, the *copyright declaration* (Creative Commons Attribution 4.0 International Rights: CC-BY-4.0), a *how-to file* explaining how to generate spectral curves from the data set utilizing the open-source software ImageJ, and a file containing *general metadata* on the project (creators, funding, copyright, language, etc.).[10] The subdirectories are the **Data** folder, containing all the data, with relative metadata as CSV files at each level (fig. 8), and the **General Metadata** folder, containing metadata, as CSV files, on the different sets of data that can be identified: the various holding institutions, the different analysis carried out on the items, and the substrate material (fig. 9). The main folder is the **Data** directory, as it contains all the analysis data files and relative metadata. Each document is identified by a unique ID based on the holding institution and its call number (e.g., UPenn_LJS24). The document-level metadata provides information about the holding institution, the catalog entry, and the number of sides that were imaged during the project. Each side is identified by a unique ID based on the document ID and appended leaf and side information (e.g., UPenn_LJS24–2r). Each side subfolder contains at least one subfolder containing the analysis data. Typically there is only one subfolder containing the multispectral imaging (**MSI**) data—that is, the stack of *TIFF files* and relative *metadata*—however, documents from the Library of Congress contain also **FORS**, **FTIR**, and **XRF** data. Following the hierarchy *Document > Side > AnalysisData*, one finds all the relevant data and metadata for each side that was analyzed for each manuscript or printed book.

The library is free to use, and set up in a way that it is hoped it may foster data reuse and facilitate further research and analysis.

---

the Modest Ambitions of the Digital Data Set," *Manuscript Studies: A Journal of the Schoenberg Institute for Manuscript Studies* 3 (2018): 83–109, https://doi.org/10.1353/mns.2018.0004.
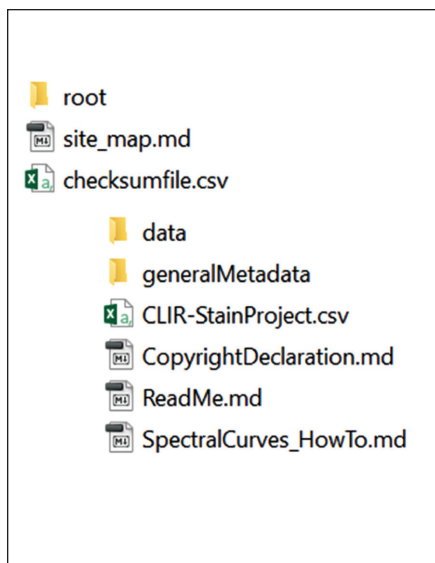10 ImageJ is available at https://imagej.nih.gov/ij/.

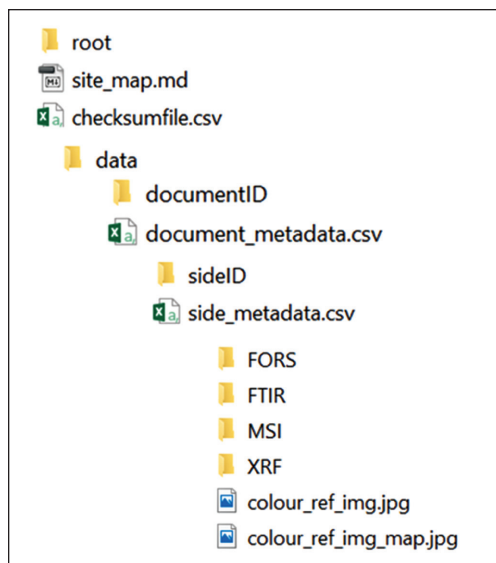FIGURE 7. The Library of Stains home directory structure.
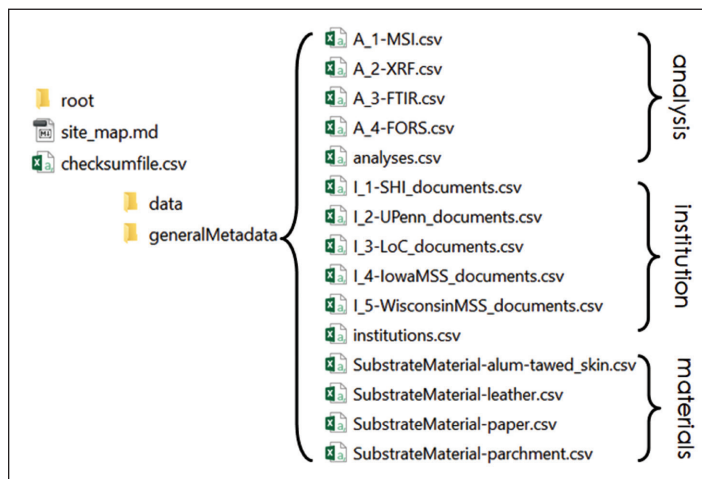

FIGURE 8. Side-level metadata and files.

FIGURE 9. Data set navigation metadata that allows users to find all data relevant to a certain analysis technique, institution, or substrate material.

Visualizing a Library of Stains: Digital Mappa

Digital Mappa (DM) is a freely available online environment that allows users to easily create and display linked and annotated data. The Library of Stains project is one of several projects constructed for DM 1.0 and comprises the images and data visualizations for University of Iowa and University of Wisconsin manuscripts.[11] The recomposed RGB images for each folio, or side, have been uploaded into DM, and for each image, highlights have been drawn around the specific sections sampled for substrate, inks, pigments, and stain data. Each of these highlighted sections has then been linked to two annotations, one that provides the numerical pixel value data specific to that sample, and another that provides the spectral curve. In addition to individual data visualizations, the RGB images are also linked to three annotations that display the metadata, pixel values, and curves for the entire folio. Any preliminary analyses of the data for a given folio is noted in the "notes" section of the metadata annotation.

The visualizations have been organized in such a way as to provide two levels of analysis. First, curves of certain stains have been linked to similar curves from other manuscripts, suggesting that these stains may be similarly characterized. Second, the Library of Stains DM project has also provided in the project's front matter a series of "stain stories," similar to the wax story above. While these stories represent a hypothesis only, they are a first step toward suggesting a historical context in which the folio and stain was perhaps used or created.

Conclusions

This project aimed to reach a broad audience through multiple vectors of communication. The project was disseminated specifically to the scholarly community, with papers delivered at academic conferences in six countries.

---

11 Library of Stains, 1 November 2018, https://zenodo.org/record/2528377#.XEMz7M 9TlR0. The Library of Stains project will be ported to DM 2.0 once the software is ready and released.

It was also distributed to the wider public via the press offices of CLIR, the University of Pennsylvania, the University of Wisconsin, and the University of Iowa. An exhibition of this project took place at the Schoenberg Institute for Manuscript Studies, University of Pennsylvania (June 2018–August 2018), and another took place at the University of Wisconsin, Madison (February 2019–May 2019). The project's updates were featured in the Digital Medievalist social media and blog series and on the University of Wisconsin's Center for the History of Print and Digital Culture's website.[12]

A variety of social media platforms, including Facebook, Twitter (#Stain-Alive), and a blog hosted on the project website, were used to inform both public and scholarly communities.[13] With posts across these formats, the project attempted to engage and excite both academic and public audiences interested in the medieval world and the lived experiences of medieval scribes, scholars, and readers. Accessibility to primary sources through the database and Digital Mappa juxtaposed with descriptive metadata contextualizes the project, connects with public interest, and provides value in the form of education.

Acknowledgments

---

12 Digital Medievalist: www.digitalmedievalist.org; Center for the History of Print and Digital Culture: http://www.wiscprintdigital.org.
13 See the project blog at https://web.archive.org/web/20181130003118/https://labeculae vivae.wordpress.com/.