2022

Possible Methods for Incunabula Digitization for Preservation and Analysis

Steven Andrews

Steven Andrews   Spring 2022

Image Books:

Possible Methods for Incunabula Digitization for Preservation and Analysis

Abstract: This paper explores different approaches that PSU's Special Collections Department might take toward digitizing its 1490 incunable containing Werner Rolewinck's *Fasciculus temporum* and second-edition *Malleus Maleficarum*. Digitization will reduce wear-and-tear of the volume and allow access to its contents by a wider range of users.

When we wish to study early print works (incunabula) the ideal option is having direct physical access to the originals. However, this is often impractical for a number of reasons, including: inconvenient physical locations, security concerns, and preservation issues. These factors all contrive to make digitization of incunabula important to researchers and laypersons alike as a supplement to, or even substitute for, physical access. Digitizations can incorporate imaging methods that reveal more detail than the human eye is capable of seeing and can allow for more efficient and productive research through searchability and interconnectivity. Throughout this paper we will use as an example the case of Portland State University's fifteenth-century incunable (hereafter referred to as the Codex), which contains the Malleus Maleficarum and the Fasciculus Temporum. It is important to keep in mind that much of what is covered applies in a broader context beyond the Codex and even beyond incunabula. This is

particularly true of the academic use of nonlinear imageboards that we will discuss shortly.

The ideal digitization of an incunable like the Codex is one that best conveys the context of the physical incunable, while also allowing the display of additional information. A digital version of an incunable can take many forms. In roughly increasing order of difficulty and expense, a partial list includes: image books, image books with manually created structural data, image front transcriptions, edited transcriptions and scholarly editions. (Rydberg-Cox, 2-3). An exception in the preceding list is image front transcriptions; they may take the place of the image books with manually created structural data in instances where the work is amenable to Optical Character Recognition (OCR). This is a rare to non-existent case in incunabula due to difficulties in automated recognition of the extensive abbreviations present in them (Rydberg-Cox, 1-3). Since the various levels of digitization listed above are often dependent on the versions preceding them on the list; we can describe the image book as a form of digitization with a relatively low bar for entry and many avenues for improvement. Image books can range from a simple set of photos of an incunable's contents to an extensively indexed, annotated, and searchable archive of detailed images of the incunable with appendixes for special features of interest: results of various tests conducted on the physical incunable itself, specialized imaging of various features, bibliographies of scholarly work relevant to the incunable, etc.

To see an example of a simple image book we can look to the Library of Congress. Their digitized works are composed of static images with no searchability or appendices, etc. (*Biblia Latina*; Durand). The important fact is that the incunabula are

digitized and available for scholars to study. Further enhancement can, if desired and affordable, be added later (Tam, 182). One of the most important considerations during the digitization process is maintaining the context of the work as much as possible. We can see an example of loss of context by comparing the way in which two incunabula from the Library of Congress were digitized. The *Rationale Divinorum Officiorum* was imaged with a single folio face showing at a time, while the Gutenberg *Biblia Latina* was imaged with the verso face of the preceding page and the recto face of the following page visible. This allows readers to encounter the *Biblia Latina* in the same manner readers of the physical incunable have (*Biblia Latina*; Durand).

When we approach the digitization of the Codex with the aim of maintaining context we can start by imaging in a similar manner to the *Biblia Latina*. To further preserve the context of the Codex, the Malleus and the Fasciculus should be treated as a whole work rather than two separate entities. As with imaging the full spread of the book to best emulate the experience of the original, maintaining the link between the two books in the Codex would result in a digitized version that more closely conveys the actual experience of the reader. The more context we can preserve around a work in a given digitization, the better it will function as: an engaging educational tool, and avenue for understanding the perspective of past peoples. Visual cues and layouts that help recreate the manner in which people have engaged with the actual physical artifact are valuable when trying to maintain audience engagement (Bentkowska-Kafel, 211-212). It can also serve as a useful tool to help historians empathize with the experiences of the people we study. From a preservation standpoint, the more a digitization preserves the context of the original incunable, the less the original will need to be handled, reducing

the wear caused by extensive handling. This type of wear is an unfortunate phenomenon we have observed first hand during our study of the Codex.

There are arguments that "... a bunch of TIFF images is no replacement for a leather-bound book" (Tam, 183-184). To ameliorate this issue, producing more refined image books with a wide variety of imaging types will allow the original context to be better preserved while still allowing the widest possible access. The more people viewing the digitized version of the book, the more opportunities there will be for the formation of interest in studying the book further. Portland State University Special Collections may be able to facilitate much broader engagement with their digitized works by making even simple scans of works available online (Tam, 181-183). For academic use, a high detail image book with comprehensive appendices would be invaluable. For the broader public, lower resolution images could be used in image books in order to reduce server load for the university, and speed up use of the image book for the end users. The application of watermarks to public images is a further security option. This separation should help prevent unscrupulous actors from misusing high detail scans to create forgeries. The focus of monitoring user's identities and activity could then be given to the much smaller academic side, saving money and time. If some of the specialized imaging, and materials analysis information was deemed of use in forgery production it could be omitted from the public version of the image book.

A new approach to the creation and presentation of image books takes inspiration from nonlinear imageboards like The Blank Atlas (The Blank Atlas). This approach could immediately provide access to images of incunabula and manuscripts, before any additional refinement. The Images could be uploaded with as little as a tag

identifying the work the image is taken from and a second tag with the image's folio number/page number/other position identifier of choice. Read-only access could be given to the general public, who could also be automatically restricted to lower image resolutions, and limited types of supplementary data. Special permission to edit tags, and view the restricted content, could be given to academics. This can entail as much or as little vetting of prospective academic users as seems fitting and practical. The academic user's ability to add tags would allow crowdsourcing of the identification of features of interest in the uploaded images. A technique that has been at times termed "folksonomy", this results in a non-hierarchical system of classification that can link information together and make it readily searchable (Kim, 76). A person studying manicules in a particular incunable, for example, could review the images and tag each one that contains a manicule. This would allow them to easily revisit the manicules as they continue their project. Later on someone else studying manicules more broadly could search for manicules and have all tagged instances in every incunabula in the system show up in their search results, allowing for easy comparisons between different works. Further tags could potentially be added and used to denote: printing dates, binding dates, authors, print shop(s) of origin, binding locations, paper repairs, fire damage, insect damage, water damage, presence of rubrification, etc. There are effectively limitless options for tags that scholars could add to help themselves in their research, and in helping themselves they would leave behind a more useful body of searchable image books. Images that contained a feature in need of specialized imaging could be tagged for this and studied by Special Collections to help target more detailed preservation and digitization efforts. The tag metadata would connect users to

the tags they add, thereby allowing administrators to trace malfeasance or single out users that make significant contributions. This record of activity could even be used to turn participation in the imageboard into a graded activity for a class. From a historiographical perspective the tagging records would provide an avenue for future researchers to investigate the use patterns of earlier historians, including the ability to link specific interests to individual historians by name. The comments on given images would provide fertile ground for discussion and cooperative investigation of the digitized works.

While moderation would be needed, a well-managed imageboard can become largely self-correcting with the larger community actively voting down contributions that are unhelpful or inappropriate (Britt, 5). These down-voted items can be reviewed when concerns about popular sentiment overriding academic rigor are raised. The most valuable aspect of this system is that it turns the digitization of an incunable into a living document, one that becomes more useful the more it is used. The people who engage with that digitization will in the course of their use refine a basic image book into an image book with manually created structural data, appended data from targeted and specialized imaging and bibliographies of relevant scholarship. This process could be managed directly by Special Collections to a highly variable degree based on interest and budget. The similarities between this approach and a wikipedia-style text-based digitization bear mentioning but as we are focusing on image books an image-centered format is more appropriate.

Further research is required to explore the advantages and limitations of the imageboard format. To that end, a limited prototype version of this form of digitized

incunable could reasonably be created by anyone with some familiarity with programming. It would be reasonable to expect that a future student could in the space of a term develop a limited test version in conjunction with Special Collections and begin exploring its uses and limitations.

Now that we have explored the digitization of incunabula we can touch on the use of imaging methods beyond simple visible light imaging that, as mentioned above, can help maintain more physical context for the digitized work and/or reveal features that are hidden to someone solely experiencing the physical book directly. We will focus on two of the wide variety of imaging options: LIDAR three-dimensional surface mapping and infrared reflectography. Both methods have the significant advantage of being non-destructive, and therefore are more forgiving of errors due to accidents, or inexperience on the part of new users.

The use of LIDAR to develop high resolution 3-D maps of the surface features of incunabula bindings, while apparently not a technique currently in use in incunabula research, may enable users to image surface features on the bindings and, with proper processing, identify the relative locations of particular features. (Pereira, 3-8). Surface mapping may eventually allow the 'generic' marks on incunabula bindings that were used fairly broadly across many binders to be identified with greater specificity as they likely have small idiosyncrasies that could be identified and used to identify their origin in a manner similar to fingerprinting. While many modern phones have LIDAR capabilities of varying sophistication, the best tools and methods for accurate high resolution surface mapping, using LIDAR or not, could be a promising avenue for future research. The more immediately tantalizing prospect is infrared reflectography of the

interior surfaces of the bindings. It is likely that the ink that bled into the inside of the binding and covers of the Codex will contrast better in the infrared spectrum, which may allow imaging of some portions of text that are too faded to be distinguishable in visible light (Miguel, 35). Unfortunately, the discovery of this possibility came rather late in the term and hurried attempts to make a trial imaging run with a low-resolution facial recognition infrared phone camera have run into hurdles from the continuing COVID-19 pandemic. Infrared reflectography seems to be less demanding on both technical expertise and budgetary levels than surface mapping. There is additional incentive to pursue this particular avenue as it has immediate research goals already tied to the possible identification of the bleed through text.

The possibility for long term scholarly benefits from the digitization method described in this paper may be worth investigating by Portland State University Special Collections. It appears to be breaking fresh ground in terms of developing academic organizational methods. Indeed, the implementation of this format could itself be fruitful ground for academic study. The use of crowdsourced tagging in an academic setting could open up a much broader body of historical works to both focused and comparative studies. While the imageboard format would present specialized imaging to great effect, the surface mapping and infrared reflectography are both worth pursuing regardless of how they wind up being organized.

Works Cited

Bentkowska-Kafel, Anna, and Lindsay MacDonald. *Digital Techniques for Documenting and Preserving Cultural Heritage*. Arc Humanities Press, 2018.

Biblia Latina. [Mainz, printer of the 42-line bible Johann Gutenberg between 1454 and 1455?, not after 1456, 1454] Image. Retrieved from the Library of Congress, <www.loc.gov/item/52002339/>.

Britt, Brian C. "Content Curation, Evaluation, and Refinement on a Nonlinearly Directed Imageboard: Lessons From Danbooru." *Social Media Society*, vol. 5, no. 4, 2019, p. 205630511988002.

Durand, Guillaume, et al. *Rationale Divinorum Officiorum*. [Mainz, Johann Fust and Peter Schoeffer, 6 Oct. 1459] Pdf. Retrieved from the Library of Congress, <www.loc.gov/item/47043536/>.

Kim, Heung-Nam, et al. "Collaborative Filtering Based on Collaborative Tagging for Enhancing the Quality of Recommendation." *Electronic Commerce Research and Applications*, vol. 9, no. 1, 2010, pp. 73–83.

Miguel, Catarina, et al. "Unveiling the Underprintings of a Late-Fifteenth-Early-Sixteenth Century Illuminated French Incunabulum by Infrared Reflectography." *Journal of Cultural Heritage*, vol. 40, 2019, pp. 34–42.

Pereira, Alvaro, et al. "Automatic Identification and Geometrical Modeling of Steel Rivets of Historical Structures from Lidar Data." *Remote Sensing*, vol. 13, no. 11, 2021, p. 2108.

Rydberg-Cox, Jeffrey. "Digitizing Latin Incunabula: Challenges, Methods, and Possibilities." *Digital Humanities Quarterly*, vol. 3, no. 1, 2009. ProQuest, http://stats.lib.pdx.edu/proxy.php?url=http://search.proquest.com/scholarly-journals/digitizing-latin-incunabula-challenges-methods/docview/2555208284/se-2.

Tam, Marcella. "Improving Access and 'Unhiding' the Special Collections." *The Serials Librarian*, vol. 73, no. 2, 2017, pp. 179–185.

The Blank Atlas, https://blankatlas.booru.org/index.php.