

Going Digital at the Wellcome Library: The Evolution of Digital Imaging and Innovation

Danae Dracht; Wayne State University; Detroit, Michigan/USA

Abstract

This report focuses specifically on practices within the Wellcome Library and Archives in London, United Kingdom, and what the Library is doing to promote and maintain its hybrid collections, both analog and digital. To remain a thriving and relevant information agency in a time when the value of brick-and-mortar institutions are constantly questioned, the Wellcome Library has worked to develop methods worth sharing to other institutions who are striving to preserve cultural and artistic heritage around the world. In light of its creativity and innovation, the Wellcome Library warrants further discussion, especially in countries outside the UK.

Introduction

“Pessimism of the intellect, optimism of the will.”

- Antonio Gramsci

Dr. Christopher Hilton, head archivist at the Wellcome Library, uses Antonio Gramsci's quote to refer to the massive digitization project at the Wellcome Library. “Nothing is finished yet,” says Hilton, “but we're willfully optimistic.”

The focus of this report is to describe imaging practices of an institution outside the scope of the United States and how this innovator in archival practices from the United Kingdom is approaching new developments in imaging technologies and the curation and management of images and born-digital materials. The Wellcome Library is a department of the Wellcome Trust, a biomedical research charity based in London, United Kingdom, working to support the public understanding of science. The Library's website explains, “We are one of the world's major resources for the study of medical history. We also offer a growing collection of material relating to contemporary medicine and biomedical science in society.”¹

Despite Hilton's claims that “nothing is finished yet,” much has been accomplished within the department in regard to digital preservation and access. The Wellcome Library has the financial resources to experiment and test innovative ideas for its Library and Archives. Ideas, it hopes, that can be shared. Hilton explains how in the past the Wellcome Library has learned from others and relied heavily on other institutions for the basis of its standards related to imaging practices. Now, in the twenty-first century, it is trying to give back.

A Brief History

The Wellcome Library's collections began with Henry Solomon Wellcome (1853-1936). To explore and discuss the Library's contribution to the field of digital preservation, it is important to first understand its history. Wellcome, born in Almond, Wisconsin, moved to London as a young man in 1878 to partner with Silas Burroughs as a Pharmaceutical salesman. Together, the two men founded the Burroughs Wellcome firm, which flourished even after Burroughs's death in 1895, when Wellcome owned and started managing the business independently. By 1897, he began systematically purchasing

materials for a new library.² He focused his energy on collecting items specifically related to medical history such as alchemy, witchcraft, anthropology and ethnography.³ Wellcome intended to use his independent wealth to amass his expanding medical history collection into a library and museum for public accessibility. While he did not live to see his vision properly through, by a series of continuous improvements, the Wellcome Library is now an online open-access platform featuring cover-to-cover books, video and audio, entire archive collections and manuscripts, paintings, prints, drawings, photographs, and ephemera.⁴

Although the twentieth century brought substantial changes for the Wellcome Library, the twenty-first century continues to bring significant transformations as Wellcome becomes a hybrid library featuring both the physical building with analogue information and its digital Doppelgänger in a technologically advanced environment. In 2007, the Library became part of the newly remodeled, “newly conceived” Wellcome Collection. Its website explains, “The Wellcome Collection seeks to explore the connections between medicine, life and art in the past, present and future; at its heart lies the curiosity that drove Henry Wellcome to amass his diverse collection.” The Medical Photographic Library was also newly conceived in 2007 and renamed Wellcome Images, complete with a new website and online database of digital images for public access.

Going Digital

What are some of the major resources in the collection?

The Library initially used the Wellcome Images portal for its image-based content. However, when the Library began moving forward with its five-year transformation to a hybrid library with a digital constituent in 2009, the Wellcome Images portal could not sustain the necessary functions for opening a highly technical digital interface. The portal proved inadequate for serious researching, as users were unable to browse, read, or download full texts or records from the database. It also lacked the ability to zoom in on images or give more than fixed-resolution access to digitized materials in the collection.⁵

Therefore, a major resource for the collection, the Wellcome “Player,” was launched for public access in 2012. The Player was born out of two new applications: the Digital Delivery System (DDS) in operation with the updated Safety Deposit Box (SDB). Together with the Library catalog, the DDS and SDB gave the Player the ability to furnish both search and zoom options for digitized items.⁵ The new version of the SDB (SDB 4), implemented in 2011, had an important contribution to the Player. SDB 4 provided vital features for usability regarding the production and management of Wellcome's materials such as the ability to share data with other systems like the workflow tracking system Goobi and Axiell CALM, a graphical archival management system.⁶ The real concern at the root of each phase in the digitization project was the necessity to preserve born digital

information that was pouring into the Wellcome collections. See figure 1 as an example of the Wellcome Library's Player:

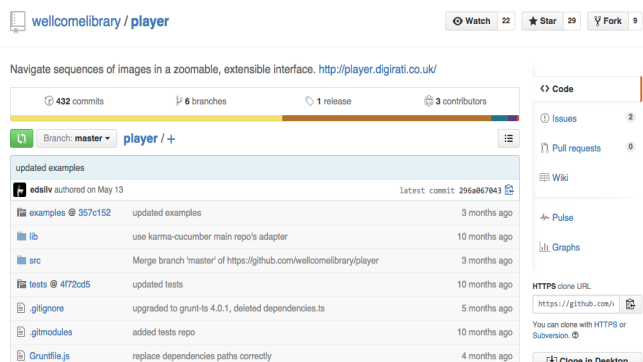


Figure 1. Wellcome Player

Codebreakers, a website launched in March of 2013, is another major resource for Wellcome and its digital repository. The site hosts approximately one million pages of documents and is the product of a £3.9 million investment by the Library dedicated to the digitization of history on DNA and molecular biology. Codebreakers accumulates documents from the Wellcome Archives and four other institutions including the Churchill College Cambridge. A review on Codebreakers explains, "The showpiece of a major redesign for the entire Wellcome Library site, Codebreakers splashily drives home the point that no library today is investing more in medical history than the Wellcome."⁷ Despite the advantages the site contributes for search results and overall user experience, Dr. Hilton, senior archivist at the Library, confirms with customer reviews highlighting a few user complications. For example, searchers need to be aware that the site is not a complete representation of everything in the collection. Also, navigation within the site can be taxing. After users follow a link displayed in the search results, going into Google and relocating Codebreakers again is easier than trying to return to the original results page. Figure 2 is taken from the Wellcome Library's website illustrating Codebreakers' digitized archives:

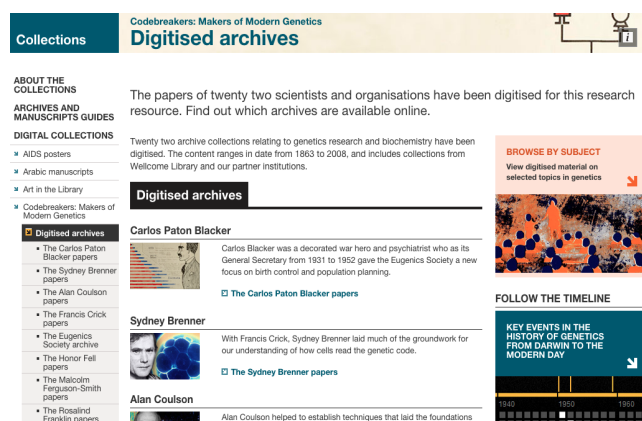


Figure 2. Codebreakers

The Wellcome Library also uses Preservica, a world-leader in digital preservation technology, as a major resource to conserve born-digital materials. Preservica works with the Library catalog, Sierra (and discovery layer Encore), its archive catalog CALM, its Wellcome Images database Bespoke, and the workflow system

Goobi to launch digitized materials into Wellcome's digital repository.⁸ In addition, the Library uses tools to characterize digital objects such as JHOVE, DROID, and PRONOM in coordination with SDB 4.⁶

Another key resource the Wellcome Library uses for its digital collection is its own storage system. The Library accesses the Wellcome Trust's existing corporate storage infrastructure, a storage area network (SAN) using fast hard disks, which are duplicated between sites to boost data security.⁶ Essentially, the major resources the Wellcome Library uses are commercial in nature, meaning that Wellcome partners with existing companies to manage their digital assets. Adrian Brown, who did a case study on the Wellcome Library, explains that the Library opted for a commercial, off-the-shelf solution with a collaborative approach involving the Digital Services staff and the Trust's in-house IT department, which "secured buy-in from IT and other departments from the outset."⁶ Collaboration of this nature, according to Brown, is essential to successfully execute a massive project like Wellcome's hybrid Library.

How has the Wellcome Library used digital technology to promote and enhance its collections?

The Wellcome Library is one of the leading libraries in creating an open access policy that was initiated in 2005. Information that was once restricted by registered institutional and individual subscribers (for a considerable fee) is now freely available to the public. This was a novel business model in 2005. Two years later in 2007, the Wellcome Library began a project to digitize 480 historical Arabic manuscripts; this project turned into a £20 million endeavor in 2009 to create an entirely new digital infrastructure for the collections, which, as Dr. Hilton explains, was "one phase of mass digitization." This phase was a five-year project to store up to 30 million images. The Wellcome Library previously used uncompressed TIFF image files as their archival storage image format. Yet in order to store several million images, a better compromise was needed between total storage capability and the standard for image quality. JPEG2000 was chosen as an image preservation format since it was small in size yet provided "intelligent compression for preservation and intelligent decompression for access."⁹ JPEG2000 was well suited for balancing the desire for better preservation and access and meeting the needs for compressed image size, image quality, and decompression performance.

Dr. Henshaw claims that at the heart of the transformation from analogue to digital there are two key components: one element is mass digitization of unique collections and the other is the creation of an online experience "that transcends the generic offerings of mass distribution and offers a unique value."⁵ The precise way the Wellcome Collection successfully "transcends the generic offerings of mass distribution" follows a technical step-by-step process. Images are configured into the Library's online interface through a series of corresponding stages. For example, figure 3 illustrates how Preservica works together with the other systems within Wellcome's digital landscape:

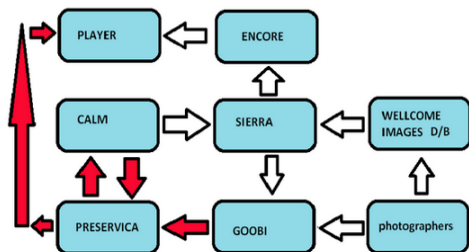


Figure 3. Preservica as part of a landscape

This process is a unique and original creation of the Wellcome Library. Despite the Library's preference for commercially created software to manage its systems, the method of transforming materials from the Wellcome Images database into the Preservica landscape to be accessible to the public is the Library's own model.

Furthermore, Dr. Kristy Henshaw explains in detail the "Seven Simple Steps of DDS [Digital Delivery System] Operation,"⁵ which is in conjunction with the Preservica landscape as it communicates with the Wellcome Player, and how it promotes the usability of the collections to users:

1. The user discovers a link to the digital content in the Library catalogue or anywhere else on the Web.
2. Clicking this link passes a command to the Player to open and to find a METS [Metadata Encoding and Transmission Standard] file with the filename corresponding to the system ID in the URL.
3. The Player opens the METS file, transforms the metadata to JSON [JavaScript Object Notation].
4. The Player reads the SDB [Safety Deposit Box] file IDs in the JSON file and requests individual files as appropriate from SDB.
5. For JPEG 2000 files, IIP Image server caches the JPEG 2000 files retrieved from SDB and creates and caches JPEG tile derivatives.
6. The Player checks the authentication status of the user against the access level of the item and its individual files.
7. The Player displays the content, or displays the appropriate login box/message.

Figure 4 illustrates this step-by-step process:

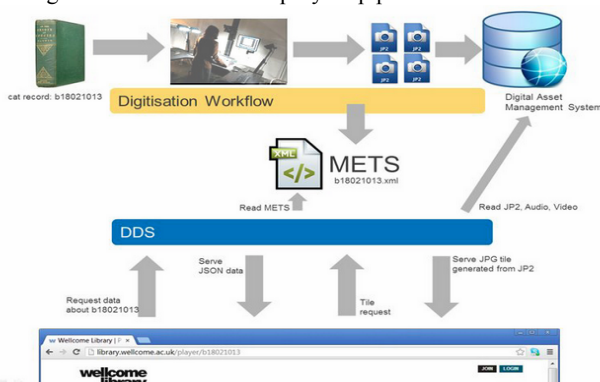


Figure 4. The Steps of DDS Operation

While digitizing is a part of the daily routine at the Wellcome Library, making this digital material available for the public is a

careful process, especially as it relates to allowing access of personal data to researchers either in the Library or abroad. In order to enhance and best manage its collections, a crucial aspect of the digitization process at Wellcome is deciding what materials are *suitable* for access. Careful consideration is given regarding privacy for those whose personal data is featured in the material. The Library has formulated its own policy composed of three phases to decide what should be accessible within its online repository: initial assessment, detailed assessment, and assessment prior to online publication.¹⁰

The first phase, the initial assessment, is applied to all materials when acquired or prior to being catalogued in order to determine how sensitive the materials are and in what way they should be presented to the public. In phase 1, the Library considers whether the material is personal data, or data that identifies a living individual and is in a relevant filing system. If not, the library then determines whether it has the individual's permission to make the information available and if the material is sensitive in a general sense. In other words, a decision is made on whether research access could cause substantial damage or distress to the close family of deceased individual. In Phase 1, the Library comes to one of four conclusions regarding access: (1) allow access according to the consent of terms; (2) assign access categories according to nature of material; (3) assign access categories according to ethical considerations beyond DPA requirements; or (4) allow access according to the consent of terms.¹⁰

The Library then proceeds to Phase 2, deciding which procedure is the most appropriate in a more detailed assessment of the materials, which usually falls into two categories:¹⁰

1. Material with permissions from the individual whose personal or sensitive personal data features in the material
2. Material without permissions from the individual whose personal or sensitive personal data features in the material

The second phase analyzes the suitability of materials in the Library only, and is not concerned with online publication, which is handled separately in Phase 3. Phase 2 adopts a different approach for appraisal depending on the nature of the materials. For example, the Library's policy explains that it uses (1) a risk assessment approach when it is not able to assess materials on an individual basis due to size or the mixed nature of the collection, such as archive collections or large groups of images, or (2) granular assessment when it is not feasible to evaluate material via "spot checks," or in instances where the degree of an item's sensitivity can be or must be individually assessed, such as film and audio recordings and individual images.¹⁰

The third phase of accessing data within the research collections is thorough as it deals with online, open access sharing outside of the United Kingdom. The Wellcome Library takes careful consideration when deciding what to share within its expanding collections. Henshaw explains that when approaching a massive digitization project, access levels and license codes are automatically granted depending on the specific project and later readjusted where necessary.⁵ However, when deciding whether personal information should be shared online, the Library relies on Phase 3. The policy explains: "We aim as far as possible to identify high-risk material prior to online publication by assessing all material due to be made available online. We will only make material available online to a general research audience if it is classed as open."⁵ Archives operating in smaller institutions working to digitize their collections are not exempt from the same

personal data access concerns and can consider the ramifications of Wellcome Library's policy to enhance their own collections.

Another way the Wellcome Library has used its own digital technology to promote and enhance its collections is by partnering with the Internet Archives as a part of the UK Medical Heritage Library digitization project. The project is massive in scale (45,000 pages digitized a day), with a special focus on the history of medicine and comparable science subjects of 19th century books from the collections of ten UK libraries including Wellcome. In 2014, the Wellcome Library created a large workspace by demolishing several walls on the top floor of 183 Euston Road in London, UK. The workplace was capable of housing over a dozen scanning units and thousands of books on shelves, trolleys, and crates.

In this newly furnished room, a team of fourteen worked toward the Internet Archive's goal to complete the UK Medical Heritage Library by unpacking, evaluating, cataloging, and digitizing medical history books and pamphlets. Unleashing commercial resources to support its digital repository and reaching out to other institutions to further its collections' development and access are just a few examples of how the Wellcome Library is using its technology to promote and enhance its massive wealth of information.

What are the Wellcome Library's ongoing digital projects?

Especially since 2009, but as early as 2005 to present, digitization has been underway at the Wellcome Library and Archives. Digitization is "business as usual" for the Wellcome Library with a "create once, use often"¹¹ mentality. In an effort to foster an economically sustainable activity without, as Dr. Hilton admits, a foreseeable completion date, the Library considers the digitization process as nothing different or out of the ordinary within a normal workday. Wellcome is digitizing or "harvesting" approximately 800,000 pages per month and 65,000 titles from the Internet Archives. Also, since 96-97% of Wellcome's users do not come into the Library, Wellcome is seriously considering user motivation for visiting the website. In an effort to broaden its scope and targeted audience, as the Library transforms itself from analog to digital, it has recently started actively selecting and accumulating non-European materials as it continues to make full text data available and transforms metadata into linked data.

As the Library endeavors to become a world-class resource on the history of medicine—an open access point comparable to Wikipedia—it is reaching out to other collections by establishing collaborators in an effort to "give back" and position itself as a pertinent access location within other libraries. The Wellcome Library has crowdsourced its digitization projects to encourage community led and developed resources that strengthens a sense of community engagement. The Wellcome Library is striving to look beyond its own establishment's goals for ways to support and cooperate with other information agencies. For example, while the Wellcome Library originated the Player, the British Library adopted its design and improved and enhanced its features for its own digital resources. The Player was in turn readopted by the Wellcome Library in its newly revised form and appears on the Wellcome Library's site with the British Library's updated enhancement.

Also, in addition to implementing the Open Access Fund and setting the standard for freely available information, one example of a more recent project in 2014 and 2015 is the Library's site for researchers called London's Pulse: Medical Officer of Health

reports 1848-1972.¹¹ With only a year in operation, the new website has already had over 42,000 remote visits, about 12% of the Library's accumulative traffic.¹¹ As part of the Wellcome Library's ongoing projects, one of the preeminent goals for this site and others at wellcomelibrary.org is making materials available in other search engines such as a library catalog, and also the flexibility to appear on the Wellcome website in a variety of forms. Through the London's Pulse, the Library has succeeded in producing a supple, technologically advanced online resource that, according to reports, is meeting the needs of its intended audience. This resource is one of a long line of resources the Wellcome Library intends to put into action for researchers worldwide. Figure 6 illustrates the home page for the London's Pulse taken from the Wellcome Library's website (<http://wellcomelibrary.org/moh>):

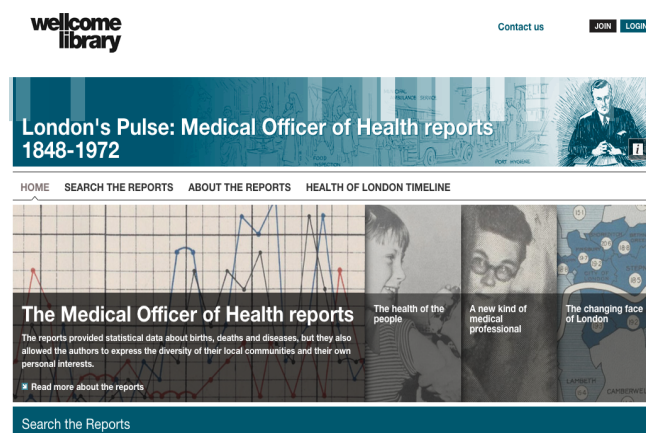


Figure 6. London's Pulse

In the fall of 2014, the Library started a digitization project in partnership with the Borthwick Institute for Archives, Metropolitan Archives, Dumfries and Galloway Council Archives, NHS Greater Glasgow and Clyde archives, and Royal College of Psychiatrists to preserve 19th and 20th century hospital records including patient notes, case notes, administrative documents, photographs, and registers. The joined collections, with help from the University of Glasgow Digitization Centre, are steadily being added to the Wellcome Library's new website with an estimated completion date in about two years. The Library hopes to have over 50 million digital pages published online by 2020.¹²

The Wellcome Library wants to ensure it can continue its digitizing practices through a sustainable and yet effective means. Therefore, the next phase of design and development within Wellcome's digital repository is to improve and expand on the Cloud and Wellcome's prototyping project. A few of the key areas Wellcome considers when developing Cloud technology are user experience and quality assurance. However, Robert Kiley, head of digital services, articulates a more precise and ambitious goal, which is "to position the Wellcome Library as a pioneer in the development of a new, interoperable, fully-scalable, digital library platform and, crucially, to make this platform available for other cultural heritage organisations to use."¹³ The results from Cloud technology are increasingly optimistic due to its startlingly low cost for a high volume of digital content. As expenses are cut and digitization becomes more affordable, the Wellcome Library anticipates that more information agencies with lower allocations can enter into a previously unthinkable realm of digital infrastructure and contribute to the history of medicine from all perspectives around the globe.

Conclusion

Alongside its dedication to develop preeminent methods worth sharing to other institutions preserving cultural and artistic heritage around the world, the team at the Wellcome Collection is willfully optimistic about the outcome of their digitization projects. Although the task may not be completed, the Library and its past accomplishments serve as an example for innovation in digital preservation. The challenge to digitize the Wellcome Collection alone is massive in scale. Partnering with other institutions such as the Internet Archives only heightens the volume of information yet to be digitally preserved and promoted through online access. By focusing on the historic details behind the Wellcome Library and Archives, recounting some of the major resources in the collection, considering the ways the Library has used digital technology to promote and enhance its collections, and unpacking the Library's ongoing digital projects, this research illustrates how the Library has proven to be a real-world example of an information agency exceeding the general standard of digitization. To find further information regarding the Wellcome Library's resources, its imaging practices, and its ongoing digitization projects, visit its blog (<http://blog.wellcomelibrary.org>) and explore its website (<http://wellcomelibrary.org>), which features multiple catalogues and collections. The Wellcome Library is a paradigm beyond the scope of the United States for how other information agencies are making massive steps in digital innovation. As an innovator in promoting open access and designing advanced platforms for digital resources, the Wellcome Library warrants further discussion, especially in countries outside the United Kingdom.

[13] R. Kiley, "Moving the Wellcome Library to the cloud." Wellcome Library Blog, 2015.

Author Bio

Danae Dracht is a graduate student studying archival administration and pursuing a master's degree in Library and Information Science at Wayne State University in Detroit, Michigan. Digital preservation and access has become a source of interest for her, especially after a summer studying in London, UK, where she gained insight on the digital imaging practices of other information institutions outside of the scope of the United States.

References

- [1] Wellcome Library. "What We Do," 2015. Retrieved from <http://wellcomelibrary.org/what-we-do/>
- [2] H.J.M. Symons, "The Wellcome Institute for the History of Medicine: a brief description." The Wellcome Trust, London, 1993.
- [3] "History of the Wellcome Library: a short history of the Wellcome Library from its foundation to the present day," 2015. Retrieved from <http://wellcomelibrary.org/what-we-do/history-of-wellcome-library>
- [4] "Digitisation at the Wellcome Library," 2015. Retrieved from <http://wellcomelibrary.org/what-we-do/digitisation>
- [5] C. Henshaw and R. Kiley, "The Wellcome Library, Digital." *Ariadne*, vol. 71, 2013.
- [6] A. Brown, "Practical digital preservation: a how-to guide for organizations of any size." Chicago, IL: Neal-Schuman, an imprint of the American Library Association, 2013.
- [7] N. Comfort, "Website Review: Codebreakers," *Bulletin of the history of medicine*, vol. 88, no. 4, pp. 738-740, 2014.
- [8] C. Hilton, "Preservica-Hybrid_Wellcome" [PowerPoint Slides], n.d.
- [9] R. Buckley and S. Tanner, "JPEG 2000 as a Preservation and Access Format for the Wellcome Trust Digital Library," King's Digital Consultancy Services King's College London, 2009.
- [10] "Access to personal data within our research collections," Wellcome Library, 2014. Retrieved from <http://wellcomelibrary.org/content/documents/policy-documents/access-to-personal-data.pdf>
- [11] L. I. Kaplish, "London's Pulse." *Online Searcher*, vol. 39, no. 3, pp. 10-14, 2015.
- [12] "Wellcome lead in digitization programme," CLIP update with gazette, 2014.