doi: http://dx.doi.org/10.23925/1980-7651.2018v21;p56-62

Digitising Charles Babbage at the Science Museum, London: managing expectations, enabling access

Nicholas J. Wyatt[•]

Abstract

The papers of Charles Babbage at the Science Museum were historically catalogued in different ways and remained difficult to consult, but their digitisation in 2011 and subsequent incremental release online have increased access world-wide. This paper describes how such access has gradually fulfilled expectations of scholars world-wide, while catering for other audiences. It examines the challenges the Science Museum has faced and the role that archivists, historians and digital professionals have had in making digital access to the archives and related objects possible.

Keywords

Charles Babbage; Archives; Digitisation; Science Museum

• Head of Library and Archives, Science Museum, London, United Kingdom. ✉ nick.wyatt@sciencemuseum.ac.uk. The present paper derives from a presentation at symposium Doing History of Science in a Digital, Global, Networked Community: Tools and Services Linking Scholars and Scholarship, 25th International Congress of History of Science and Technology, Rio de Janeiro 22-29 July 2017.

Great Britain's Science Museum Group (SMG) holds internationally important historic collections for science, technology and medicine, including some significant archives. The Group consists of the Science Museum (London), the SMG National Collections Centre (Wroughton, near Swindon), the National Railway Museum (York), the National Science and Media Museum (Bradford) and the Museum of Science and Industry (Manchester). Its mission is to "engage people in a dialogue about the history, present and future of human ingenuity" and it is achieving this increasingly through digital means.

On October 2015, the SMG's Board of Trustees approved a new digital strategy,[1] which emphasised digital access to the Museum's collections. Prior to this, the Group's online delivery of digitised collections has been very poor. Digitised records were split across multiple microsites, such as *Ingenious* or *Making the Modern World*, many of which had weak content discovery, broken features, dated design, patchy data quality and poor user experience. The Museum developed plans for a new web portal to provide enhanced access to high-quality information and images for its archive and object collections from across the Group. One archive that it would prioritise was the papers of the mathematician Charles Babbage (1792-1871), held in the Science Museum's archive collections.

Babbage has been called the 'Father of Computing' and from 1820 until his death he worked on various designs for calculating and computing devices, his Difference Engines and his Analytical Engines. Charles Babbage's work was continued by his youngest son, Henry Prevost Babbage, who was the first to list the main series of drawings and notations, published in 1889 as *Babbage's Calculating Engines: being a collection of papers relating to them, their history and construction.*[2]

The Science Museum's Babbage archive contains almost the entire surviving technical material relating to his automatic calculating machines and the Museum also holds various models and components relating to his work. Most of the archive was loaned to the Science Museum, then part of the South Kensington Museum, in 1878 by H.P. Babbage and presented outright in 1905. Important additions were received from King's College, London in 1971. The papers now fill 11 plan press drawers and eight linear metres of shelving at the SMG National Collections Centre at Wroughton.

The Babbage papers contain three main types of material: Charles Babbage's notebooks, also known as scribbling books; his engineering drawings; and notations, which describe the way parts are intended to act and can be thought of as 'walk throughs' or 'traces' of micro-programs for various models or plans of the engines. There are separate smaller collections of Babbage's diplomas and letters to Charles Babbage, and his social diary for 1844. There are Babbage archives elsewhere, mostly at the British Library[3] and the

[1] "Digital Strategy 2015-2017", Science Museum Group, accessed 26 March, 2018, https://group.sciencemuseum.org.uk/policies-and-reports/digital-strategy/.

[2] Henry P. Babbage, *Babbage's Calculating Engines: Being a Collection of Papers Relating to Them, Their History and Construction* (London: Spon, 1889).

[3] The British Library holdings were microfilmed, and a list published as: *The History of Science and Technology. Series Three, The Papers of Charles Babbage, 1791-1871: a Listing and Guide to Part 1 of the Microfilm Collection* (Marlborough: Adam Matthew Publications, 1997). The archive is also indexed in "Explore Archives and Manuscript Catalogue", British Library, accessed 26 March, 2018,

University of Cambridge.[4] These mostly relate to his other interests, although Cambridge does hold one of Babbage's scribbling books.

Until the late 20th century there had been little detailed analysis of Babbage's ideas and his influence on the development of modern digital computers was minimal. Historians took some interest; for example, John Michael Dubbey analysed Babbage's mathematical work for his PhD thesis in 1968 and a book published in 1978,[5] but drew mostly on Babbage's published works. When Anthony Hyman was researching Babbage's life he used the Science Museum's Babbage papers, some of which were photographed and reproduced in his published biography in 1982.[6] In 1989 the complete works of Charles Babbage were published in 11 volumes, but for the most part this covered his published works and not his archives.

The centenary of Babbage's birth saw two significant developments. The first marked the culmination of the first detailed analysis of this complex archive by Dr Allan G. Bromley, Professor of Computing Science at the University of Sydney and Visiting Research Fellow at the Science Museum, London. He arranged and described the archive and his catalogue was published in 1991 as *The Babbage Papers, A Cross-referenced List*.[7] Bromley divided the archive into major groups identified by letters of the alphabet which generally corresponded to Babbage's own groupings. Bromley also split the original A sequence of drawings to become A (Analytical engine 1833-1848), B (Difference Engine No. 2) and P (Analytical Engine 1857-1870). These three sequences and the notations (Group F) had originally been listed by H.P Babbage and Bromley used his numbering, but used the titles that appeared on the items themselves, whereas H.P. Babbage had used the titles in lists prepared by his father. Bromley's published list contained some typographic errors, so drawing A.1, for example, is dated 1843, when it should have been 1834, and he omitted some items.

Another Babbage scholar, C.J.D. Roberts also examined the Babbage Papers and listed and described the Difference Engine No. 1 K and U series, adding in archives from other institutions and private collections; this list is available on the web.[8]

The second significant development revealed a greater mathematical understanding of Babbage's work after the archive and remaining objects were given curatorial scrutiny within the Museum itself during research for a project to create the Difference Engine no. 2 (DE2). Its construction was to propel Charles Babbage into public consciousness, as Andrew Nahum describes in his history of displays in the Science Museum:

---

http://searcharchives.bl.uk/primo_library/libweb/action/search.do?menuitem=0&fromTop=true&fromPreferences=false&fromEshelf=false&vid=IAMS_VU2.

[4] The University of Cambridge holdings (Add. Ms. 8705) were microfilmed and published in 1991. They are also listed in its Archive catalogue:
https://janus.lib.cam.ac.uk/db/node.xsp?id=EAD%2FGBR%2F0012%2FMS%20Add.8705.

[5] John M.Dubbey, *The Mathematical Work of Charles Babbage*, Thesis (Ph.D), University of London, 1968; John M. Dubbey, *The Mathematical Work of Charles Babbage* (Cambridge: Cambridge University Press, 1978).

[6] Anthony Hyman, *Charles Babbage, Pioneer of the Computer* (Oxford: Oxford University Press, 1982)

[7] Alan Bromley, *The Babbage Papers in the Science Museum Library: A Cross-referenced List* (London: Science Museum, 1991).

[8] C.J.D. Robert, *"Charles Babbage's Difference Engine No. 1: Surviving Archives and Other Sources,"* accessed 26 March, 2018, https://sites.google.com/site/babbagedifferenceengine/de1archives.

> "Charles Babbage was an intriguing figure, little known beyond the circle of historians of early computing, but the exercise of replicating and building the originally uncompleted Difference Engine No. 2 for his centenary in 1991, led by Doron Swade at the Science Museum, has now unequivocally inserted Babbage into contemporary culture and popular thought."[9]

DE2 has remained on display in the Science Museum since then and is occasionally demonstrated for specialist groups. Only one other DE2 was made by the Museum and until recently it was displayed in the Computer History Museum in California. Swade documented DE2 and its construction and published several books on Babbage.[10]

Swade left the Museum in 2002 and continued his research on Babbage. By 2011 he had joined with others to form the Plan 28 Project,[11] set up to build a working prototype of the Analytical Engine. They contacted the Science Museum with a request to access digitised versions of the Babbage archive. After weighing up the options for funding, the Museum decided to go ahead with digitising the whole archive and associated items, not just the material associated with the analytical engine. It funded this work internally and appointed an external company to scan the collection on-site. Prior to its digitisation, the collection was completely re-catalogued by Cecilia Cassingham, the Museum's Archivist, using professional archival principles of arrangement and description on the newly-acquired archives management system, Adlib.[12] She continued to use the alphabetical arrangement described above, but to conform with archive standards slight changes were made so drawing A.1 became BAB/A/001, BAB being the archival code for the entire collection. The condition of the items was assessed, and some were too fragile or too large to capture and remain un-digitised.

Over 7,700 images were created; the master files were captured as TIFF 6.0 uncompressed files, varying in size from 12 Mb to 310 Mb, depending on the size of the original document. Master sets were kept for preservation on a hard disk and on the Museum server with strict protocols restricting access. Lower-resolution user files were also created and Cassingham loaded these onto Adlib. Their copyright lies with the Science Museum, so Plan 28 received a sub-set of the images as JPEG images, with their use restricted by a license agreed with the Museum. Another project, funded by the Leverhulme Trust and based at Royal Holloway and Bedford New College, University of London, also gained access to licensed images for research analysing Babbage's notations. Initially, access to the images for others were very limited and they were not available in any form on the web, nor was Adlib accessible to anyone other than archivists.

---

[9] Andrew Nahum, "Exhibiting Science: Changing Concepts of Science Museum Displays," in *Science for the Nation: Perspectives on the History of the Science Museum*, ed. Peter J.T. Morris (London: Science Museum, 2013), 191.

[10] See, for example, Doron Swade, *Charles Babbage and his Calculating Engines* (London: Science Museum, 1991); Doron Swade, *Charles Babbage's Difference Engine No. 2: Technical Description* (London: Science Museum, 1996); Doron Swade, *The Cogwheel Brain: Charles Babbage and the Quest to Build the First Computer* (London: Little, Brown, 2000).

[11] "Plan 28: Building Charles Babbage's Analytical Engine," accessed 26 March, 2018, https://plan28.org.

[12] http://archives.sciencemuseumgroup.ac.uk/Details/archivescience/110000003.

There followed an increasing demand for wider access to the digitised archive. But who were the potential audiences for the catalogue and digitised images and what are their needs? Babbage scholars and researchers need access to all the metadata, higher-resolution images for detailed analysis and medium-resolution images for lectures, articles and books. Other historians of mathematics and science often only need access to medium resolution images to support research and to use in articles and books. The public often need mediated interpretation, selected images of highlights or basic biographical information; they do not want to wade through complex archival descriptions and hierarchies! Commercial organisations may need some mediation to understand the archives and must purchase copies for use in books or media productions.

The physical archive is made available to researchers at the main archive store at Wroughton and selected items have been transported to the Library's London home, the Dana Research Centre and Library for consultation by researchers there.[13] For several years some of the drawings, scribbling books and objects were on display in the Museum for the public to see in an exhibition devoted to Babbage. But for many researchers across the world, physical access has been impossible, so increased digital access has become vital to their work. For technical reasons, levels of access have improved only incrementally. The first major development came in May 2016 when the Adlib public portal was launched. This included the entire hierarchical description of the Babbage papers to item level as originally catalogued by Cecilia Cassingham (who had left by this time). Beata Bradford, the Archive Collections Manager, found that Adlib could not cope with multiple images per record, so she deleted images for the scribbling books from Adlib. The remaining images could then be displayed in the public portal, but with just one image per item. Furthermore, technical limitations prevented the display of higher resolution images which would have slowed down access considerably.

In parallel, medium-resolution images also became available on the website of the Science Museum's *Science and Society Picture Library*. This allowed commercial companies such as publishers or the media to purchase copies of the images and this remains the primary route for such use.

Both external research projects that had received licensed copies of the digitised archive have benefitted hugely from this access. The launch of the catalogue online enabled Tim Robinson from the Plan 28 Project, based in the USA, to produce a searchable database of all catalogued material with related content fully cross-referenced.[14] He has been helping the Museum archives team by pointing out discrepancies in cataloguing which is will inform changes to the online catalogue. Swade has continued to publish articles and book chapters,[15] benefitted by his access to the archive and has said that he could not have made

[13] "Dana Research Centre and Library," Science Museum, accessed 26 March, 2018, www.sciencemuseum.ac.uk/library.

[14] This is not publicly available, but it is described on the Plan28 website, Spring 2017 update, see note 11.

[15] Such as: Doron Swade, "Photographing the Footsteps of Time: Space and Time in Charles Babbage's Calculating Engines," in *Space, Time, and the Limits of Human Understanding*, ed. Shyam Wuppuluri (London: Springer, 2017); Doron Swade, "Turing, Lovelace, and Babbage," in *The Turing Guide* (Oxford: Oxford University Press, 2017).

the progress he did without the licensed use of the archive material to hand and "I question whether without fast digital access the work could have been undertaken at all."[16]

The next major improvement in access came in December 2016 when the Museum's Collections Online portal was launched on the Museum's website.[17] This draws together archive metadata from Adlib, object metadata from MimsyXG (the non-public object database) and images from the SMG's Media Library (its digital asset management system, iBase Trinity), and presents them in a clean and crisp design. It took a few months for all the Babbage images to be added, as Bradford had to create Media Library records which linked the images to Adlib records via their reference code, so they could be harvested and displayed on Collections Online. A search for Charles Babbage now yields not just his archives, but also objects from the museum collection and pictorial material, such as portraits. The full Babbage archive hierarchy is shown but the landing screen hides much of the descriptive text, to ensure that the hierarchy is made more visible. The user can then explore the hierarchy to reveal individual items and, in most cases, images of them. Once an image has been selected the software allows for rotation and zooming (using higher-resolution images). This is especially important for some of Babbage's scribbling books, where the writing is both the right way and upside down on the same page. The medium-resolution images are made available under a Creative Commons Licence,[18] which allows for non-commercial use. There were problems initially with differing metadata and authority control standards and other technical issues, but these are gradually being resolved and feedback has been very positive.

Throughout this incremental release of metadata and images, staff have had to resolve many challenges, but they have also learned lessons. Differing past cataloguing practices have caused confusion, but it is likely that the naming conventions adopted in the most recent cataloguing of the archive will become standard. Images have had to be matched to correct metadata records, which was not always a straightforward process. Different metadata standards used for archives and objects need to be harmonised for ease of access, this is especially so for forms of names and their presentation online. Core audiences need to be identified, communicated with by email or in meetings and remain engaged. Throughout this period staff have tried to maintain close contact with researchers, and despite the occasional hiccough this has worked well. Scholars and researchers can be the most demanding of all, but they can provide useful insights and help with quality control, but their expectations do need to be managed carefully especially when improvements are only gradual.

Digital access to the Babbage metadata, its hierarchical relationships and to surrogates has certainly been a positive development. But there can be risks in making archives digitally available as this can lead to an increase in demand for the original archive. As Gillian Oliver stated in 2012,

---

[16] Doron Swade, e-mail message to author, June 25, 2017.

[17] "Search Our Collection", Science Museum Group, accessed 26 March, 2018, http://collection.sciencemuseum.org.uk/.

[18] Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0), Creative Commons, accessed 26 March, 2018, https://creativecommons.org/licenses/by-nc-sa/4.0/.

> "The act of digitization is likely to increase awareness of particular records exponentially and given that one of the motivators for digitization may be to make information more widely available to a very large group of users, archivists need to be aware of the potential of a corresponding increase in requests to consult originals. It is ironic that a preservation tool may in fact pose additional risks to the artefact, but awareness of this potential should be factored into digitization strategies."[19]

The main driver for the digitisation of the Babbage papers was to increase access, less so its preservation, and this has certainly happened. The Babbage papers have been the most accessed archive on Adlib online and on Collections Online it had 4,125 page-views from July 2017 to February 2018. Evidence for increased use of the physical original remains inconclusive and given the location of the store this is perhaps unsurprising as the Wroughton facility can be difficult to get to for many researchers. It also has less of the appeal of some archives digitised elsewhere, especially those of interest to family or local historians. There is some evidence that whole-scale digitisation can lead to a decrease in visits to reading rooms, such as in the National Library of Scotland's maps reading room. After it had digitised many of its maps, there was a 76% drop in maps issued to readers and a 61% drop in visitors; conversely there was a 267% increase in page views during 2017.[20]

Given time and resources more could be achieved, such as the digitisation of Babbage material that could not be scanned in 2011 and further improvements in cataloguing. It could be possible to collaborate with other repositories to create a more rounded digital resource containing a more diverse range of surviving archives, but this would require additional resources and institutional commitment. Significant progress has been made since 2011 and for many researchers around the world the digital liberation of the archive has revolutionised their understanding of the archive and of Babbage himself.

---

[19] Gillian Oliver, "The Digital Archive," in *Evaluating and Measuring the Value, Use and Impact of Digital Collections*, ed. Lorna M. Hughes. (London: Facet Publishing, 2012), 57.

[20] Laragh Quinney, "Evolution of a Map Service: Successful Digitisation and its Impact on Library Services" (presentation, Research Libraries UK Conference, London, March 15, 2018).