

Geoffrey Rockwell
geoffrey.rockwell@ualberta.ca
University of Alberta, Canada

Stéfan Sinclair
stefan.sinclair@mcgill.ca
McGill University, Canada

Melissa Terras
m.terras@ucl.ac.uk
University College London, England

How can digital methods be used to conceptualize historical research projects, including their teams, approaches, methods, and outputs? What methodologies can be used to synthesize and analyze archival records, workshop plans, photographic evidence, and oral histories? This co-authored paper describes an ongoing effort by a collaborative group* to understand and recover the work of Father Roberto Busa, commonly thought to be the “founding father” of Digital Humanities. Starting in 1949, Roberto Busa, S.J., began a landmark collaboration with IBM to build a lemmatized concordance to the works of St. Thomas Aquinas. In 1956 Busa founded the world’s first humanities computing center in Gallarate, Italy, located after 1961 in a former textile factory stocked with rows of IBM punched-card machines. This was CAAL, the *Centro per L’Automazione dell’Analisi Letteraria*—the Center for the Automation of Literary Analysis. There Busa and his mostly female student operators processed the monumental *Index Thomisticus*, a selection of the Dead Sea Scrolls, and other texts, from 1961-1967 (Terras and Nyhan, 2016; Jones 2016). However, there is much that is still unknown about Busa’s research approach and methods. We aim to recover what Busa and his operators did. By repurposing punched-card office machinery for *literary* data processing, Busa created an important pre-computing technology platform for humanities research, one which has become obscured over time. We aim to reverse engineer and reconstruct not just a particular technology (punched-card machines) but that first humanities computing center as a whole. By doing so, we will explore methods that can be useful for other historians as they look back upon site-specific projects and groups, using digital tools and methods to effectively interleave and investigate historical data sources.

Jeffrey Schnapp of Harvard’s metaLAB has remarked that “every cultural object is a network” (Schnapp, 2015). Reverse engineering involves taking apart a device or system not to replicate it but in order to better understand its design and purpose, its networked relations. The goal in this case is to break down and experimentally reconstruct the networked cultural objects—including specific machines, architecture, infrastructure, and human operators—that amount to the components of that Italian humanities computing center, and in that way to model a more capacious idea of “digitization” itself. We make use of a cluster of convergent practices:

1. The digitization of archives—paper-based documents with Dublin Core derived metadata, but

Reverse Engineering the First Humanities Computing Center

Steven E. Jones
stephenjones@usf.edu
University of South Florida, United States of America

Julianne Nyhan
j.nyhan@ucl.ac.uk
University College London, England

also 3D digitizations of physical artifacts such as punched cards, relay switches, etc.

2. A cultural-heritage virtual model of the architectural space, a 3D immersive environments of the center itself, created through basic photogrammetry and using Maya + the Unity engine, based on multiple archival images, as well as new scans of the building, still standing outside Milan (though much altered).
3. Emulations of forgotten or obsolete technologies, punched-card data processing systems as well as other “adjacent” technologies in the 1950s and 1960s.
4. Oral histories and audio-files of interviews with surviving punched-card operators, Busa’s secretaries, and others

The overall objective is to model this important early research center and its activities through a series of purpose-driven and interlinked emulations, 3D spaces, oral histories, and digitized documents and artifacts. We employ metadata to map archival materials and emulations onto the models in order to understand the material history of what is usually taken to be the first humanities computing center. In the process, we complicate the key terms themselves, including *first* and *computing*.)

Most of what has been known to date about Busa’s early literary data processing was derived from a handful of his own publications— first by Winter (1999); later, Jones (2016) drew on the Busa Archive to contextualize and extend his narrative account. Rockwell and Sinclair (2014) and Terras and Nyhan (2016), have continued to clarify the story in different ways. Actually modeling the machinery and workflow allows us to address specific questions about this important moment in the birth of linguistic data processing and humanities computing, such as:

- What were the precise roles played by human operators *between* the automated stages, sorting card decks, lemmatizing word lists, programming machines via plugboards, etc.? (How were these roles stratified and gendered?)
- What source texts were used for input and how were they prepared and marked up so that the operators could use them as the basis for what they punched on the cards?
- At what stage did IBM agree to print customized punched cards with what amounted to data fields unique to Busa’s projects? What was the nature of the data ontology behind these customizations?
- What is the evidence that the work of Busa’s center contributed to larger technology developments at IBM, such as Peter Luhn’s development of the influential KWIC (keyword in context) protocol for information retrieval?

Additional questions will surely arise during the ongoing process of modeling and cross-checking archival materials and oral histories.

Although Busa’s humanities computing center is our focus, we believe this methodological approach would be useful in other instances, as a way to conceive of digitization as a process of modeling artifacts and documents in relation to technology and infrastructure. We draw on theoretical approaches and methods associated with media archaeology (Parikka, 2012; Emerson, 2014; Rockwell and Sinclair, 2014; Sinclair, 2016), creative historical prototyping (U Victoria Maker Lab; Sayers et al, 2016), the archaeology of science (Haigh, 2016; Schiffer, 2001), and on the methods and expertise of digital archaeology in the field of cultural heritage, including its attention to issues of access and preservation (Koller, 2009; London Charter, 2009).

The presentation at DH 2017 will include slides containing selections from the 800 historical photographs of Busa’s center, as well as other images, audio files, and demonstrations, including a prototype 3D virtual model of the center. The paper will explain the project’s practical aims and theoretical significance: for example, we address current debates in digital humanities about the influence of text-based analysis on today’s definitions and practices; or debates about possible alternative genealogies for DH (Klein, 2012). It will also spotlight the role of gendered labor in early humanities computing, and the entanglements of early humanities technology research with corporate and government funding. Our broader methodological purpose is to take up in practice what Jeffrey Schnapp has called the “defining design challenge of our epoch”—“to weave together information and space in a meaningful fashion” (Schnapp, 2015), and the methods will be of interest and use to others who are approaching multimodal archives and interpolating the information therein.

Acknowledgements

In addition to the coauthors of this paper, contributors to the overall project include Marco Passarotti and Paolo Senna (*Università Cattolica del Sacro Cuore*, Milan).

Bibliography

- Emerson, L. (2014). *Reading Writing Interfaces: From the Digital to the Bookbound*. Minneapolis: University of Minnesota Press
- Haigh, T., Priestley, M., and Rope, C. (2016). *Eniac In Action: Making & Remaking the Modern Computer*. Cambridge, MA: MIT Press.
- Jones, S.E. (2016). *Roberto Busa, S.J., and the Emergence of Humanities Computing: The Priest and the Punched Cards*. New York: Routledge.
- Klein, L. (2012). “Digital Origin Stories.” <http://lklein.com/2012/11/digital-origin-stories/>.
- Koller, D., Frischer, B., and Humphreys, G. (2009). “Research challenges for digital archives of 3D cultural heritage models.” *ACM Journal* (December 2009): DOI: 10.1145/1658346.1658347 <http://doi.acm.org/10.1145/1658346.1658347>.
- London Charter for the computer-based visualization of cultural heritage, 2.1 (February 2009): <http://www.london-charter.org>.
- Parikka, J. (2012). *What is Media Archaeology?* Cambridge, UK: Polity.
- Rockwell, G., and Sinclair, S. (2014). “Past Analytical: Towards an Archaeology of Text Analysis Tools.” *Digital Humanities 2014: Conference Abstracts*. Lausanne: EPFL and UNIL, pp. 359-60.

http://www.researchgate.net/publication/273449857_Towards_an_Archaeology_of_Text_Analysis_Tools.

- Sayers, J., Elliott, D., Kraus, K., Nowviskie, B., and Turkel, W. J.** (2016). "Between Bits and Atoms: Physical Computing and Desktop Fabrication in the Humanities." In Schreibman, S., Siemens, R., and Unsworth, J., Eds. *A New Companion to Digital Humanities*, 3-21. Wiley Blackwell.
- Schiffer, M. B., Ed.** (2001). *Anthropological Perspectives on Technology*. Salt Lake City: University of Utah Press.
- Schnapp, J.** (2015). "Aphorisms on the 21st Century Museum." <http://jeffreyschnapp.com/2015/01/26/aphorisms-on-the-21st-century-museum/>.
- Sinclair, S.** (2016) "Experiments with Punch Cards." <http://stefansinclair.name/punchcard/>.
- Terras, M. and Nyhan, J.** (2016). "Father Busa's Female Punched-Card Operators." In Gold, Matthew K., and Klein, L. Eds. *Debates in Digital Humanities 2016*. Minneapolis: University of Minnesota Press. <http://dhdebates.gc.cuny.edu/debates/text/57>.
- University of Victoria Maker Lab.** <http://maker.uvic.ca>.
- Winter, T. N.** (1999). "Roberto Busa, S.J., and the Invention of the Machine-Generated Concordance." *The Classical Bulletin*, 75,1: 3-20.