

## 4. Narrative

### Summary

The Jesuit scholar Roberto Busa, S.J. is widely considered to be the founder of humanities computing. He began collaborating with IBM in 1949 and in 1956 he established the first humanities computing center in Gallarate, near Milan, which from 1961-1967 was housed in a former textile factory. Our collaborative international project will contribute to the history of digital culture by reconstructing that center. Our detailed media archaeology and digital modeling aims to complicate the founding myth of humanities computing by providing a more detailed contextual history, in a form that will encourage further modeling, testing, and experimentation.

### Outcomes

Our cluster of convergent practices will produce the following outcomes:

1. The multimodal digitization and online publication of selected materials from the first years of the center's operation (1961-1967), now in the Busa Archive, at the Library of the Università Cattolica del Sacro Cuore in Milan, including correspondence, rare publications, flowcharts and diagrams, 80 documentary photographs, a short film, and 3D artifacts, such as punched cards, memory chips, and rolls of magnetic tape. All materials bear directly on the history of the center and its work.
2. A suite of interactive digital emulations of the punched-card machinery and methods used in the first experiments in literary data processing, including the production of the landmark *Index Thomisticus* (an 11-million word lemmatized concordance to the works of St. Thomas Aquinas).
3. Oral histories (digital audio files + transcripts with translations from Italian) with some surviving punched-card operators and others involved with Busa's landmark research project.
4. An immersive 3D model of the historic facility where the work took place—CAAL: Centro per L'Automazione dell'Analisi Letteraria (Center for the Automation of Literary Analysis)—a former textile factory the center occupied from 1961-1967. This virtual space will serve as a learning tool for media archaeology investigation, as well as an immersive interface connecting archival materials, emulations, and oral histories, making the whole cluster available for future research.

The combined result will be a multimodal environment—documents, media, tangible artifacts, virtual models, and emulations of the technologies and spaces. The process of modeling the space, machinery, and workflow of CAAL will allow us to address important questions about this moment, the birth of humanities computing, such as:

- What was the precise nature of the role played by human operators between the automated stages, sorting card decks, lemmatizing word lists, programming machines via plugboards, etc.?

- How were printed source texts for input vetted and prepared so that the operators could use them as they punched 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 is the evidence that the work of the center contributed to related technologies developed within IBM, such as Peter Luhn's influential KWIC (keyword in context) protocol for information retrieval?
- What can a virtual exploration of the architectural *Wissensräume* ("knowledge spaces") of Busa's center reveal about the ways that space has shaped the act of knowledge-making in humanities computing and the digital humanities?

Of course, some questions will arise only in the process of modeling and cross-checking archival materials and oral histories. Our collaboration aims to model this historic site (in both the architectural and institutional sense) in order to create a new knowledge site, a distributed open environment for further research and learning.

Additional outcomes will include a public presentation at USF after the funding period (spring 2019), a conference panel and demonstration involving all of the collaborators; a co-authored white paper for the NEH, and separate co-authored article to be submitted for publication in a peer-reviewed journal.

## **Method**

Although Busa's humanities computing center is our focus, we believe this approach will be useful in other cases, a way to conceive of historical study involving digital materials as a process of modeling artifacts and documents in meaningful relation to technology and infrastructure, combining a diverse array of convergent practices. We draw on theoretical approaches and methods associated with media archaeology (Parikka; Emerson; Rockwell and Sinclair), creative historical prototyping (U. Victoria Maker Lab; Sayers, et al.; Lipson, et al.), the archaeology of science (Haigh; Schiffer), and digital archaeology and cultural heritage, including issues of access and preservation (e.g., the London Charter). For the model-building, we will also build on best practices in game design. We expect the project to shed light on current debates in digital humanities about the influence of text-based analysis on today's practice. But our project will also call attention to possible alternative genealogies for digital humanities (Klein), spotlighting some forgotten or obsolete technologies (such as the Microfilm Rapid Selector), punched card machinery, and early stored-program computers with magnetic tape drives, while highlighting the hidden labor of the mostly young women who operated the machines and processed the cards (Terras and Nyhan).

### **1. Mixed-media digitization**

The selection of archival materials in this project is drawn from the historically significant period just after Busa had begun his collaboration with IBM, as he established CAAL's

infrastructure, methods, and procedures, roughly 1961-1967. Our collaborative team includes in Milan a Librarian-archivist as well as a computational linguist who is the Coordinator of Research and Secretary of the CIRCSE research center at Università Cattolica—both of whom share responsibility for the Busa Archive. A collateral goal is to begin to enhance the preservation of and access to this historically important set of materials, which includes extensive correspondence as well as photographs, a film, and other documentation and some physical artifacts, including original punched cards in various formats. Once procedures are in place at the Library, digitization of the Archive as a whole will continue beyond the funded period. During our funded period a limited set (approximately 25-30) of documents will be scanned, saved as archival TIFFs with derived JPEGs, then transcribed by workers already dedicated to such tasks for the Library (transcriptions will be checked and any necessary translations made by members of our team); metadata will be applied (following the Library's modified Dublin Core). The result will be a test-bed segment of the historical collection. Digital surrogates of the materials in the Archive will be hosted by the Library as part of the Busa Archive and made accessible via URIs and metadata, so that they can be integrated into our 3D model and emulations via the project website (hosted by USF). All online materials for the project will be made freely available under a Creative Commons CC-BY license (as it already the practice at the Busa Archive).

## **2. Software emulations**

Using the archival materials, including paper punched cards, unpublished flowcharts, diagrams, and photographs, as well as published (but rare) accounts by Busa and others, Rockwell and Sinclair will develop a series of online software emulators (via Javascript + HTML + image files) and detailed explanations of the actual punched-card workflow at CAAL, including specifics on markup, input, processing, and printing, as well as key examples of alternative technologies Busa considered or aspired to use. For example, the encoding schema actually used in the customized IBM punched cards has yet to be definitively determined, although Rockwell and Sinclair have begun to speculate on the possibilities (<http://stefansinclair.name/punchcard/>). Hitherto unknown flowcharts and other materials in the Archive, some produced at IBM, will allow for the emulation of suites of the punched card machines in various combinations, experiments in recovering actual and possible workflow design(s) at CAAL, including the crucial role played by human operators, who carried and loaded decks of punched cards, and lemmatized and sorted them by hand at key stages.

## **3. Oral histories**

The Busa Archive photographs show teams of student operators at work in the center. Like the vast majority of female workers in early data processing and computing, they have remained mostly anonymous. But recently Julianne Nyhan and Melissa Terras, with Marco Passarotti, have conducted interviews with some of them. Transcripts (with translations) are already being prepared for print publication, but selections of those transcripts (plain text or XML) and audio files (MP3) will be modified and integrated into the project, allowing Busa's operators to speak for themselves and offering potential counter-narratives to the official founding story. Additional interviews will be recorded for the project, for example with Busa's former secretary, Danila Cairati.

#### **4. 3D modeling**

Deploying the technology and staff at USF's Advanced Visualization Center (<http://avc.web.usf.edu>), Jones and Kaplan will construct an immersive 3D model of the center. The building stills stands. Parts of the façade and some characteristic architectural details remain from the 1960s, and these can be captured with new onsite photography. But the Busa Archive includes 80 medium-high resolution photographs of the building from the period, interior and exterior, including the operators at work, and images which one can zoom in on to verify the exact model of IBM machine being used, or the source texts that are being punched, or the methods for stacking, sorting, and transporting the punched cards. Our primary technique will be architectural photogrammetry, with the help of along empirical onsite measurements and Google Map-based coordinates for scaling and orienting the building.

The Archive contains a number of documents about the infrastructure, layout, equipment, staffing, and workflow of the center, which will assist in our model-building. With photogrammetry, and the use of Autodesk's Maya and the Unity 3D game engine, standard tools in cultural-heritage work, we'll construct an immersive annotated model of the physical space and the research that took place there. A pinned map will show its proximity to Busa's college and to other light-industry sites in the town. Users will be able to navigate the former-factory floor and even reconstruct possible configurations of the machines, in order to test hypotheses and fill in gaps in the collective knowledge of exactly how this pioneering work was done (and how it might have been done differently). A handful of representative 3D artifacts (sample EPROM chips, for example, and reels of magnetic tape for use with later machines), will be captured and 3D-modeled for inclusion.

The 3D model of the center will exist in various manifestations: both a simplified browser-based version (hosted at USF) and a more detailed and interactive gamelike version in Unity 3D will be freely downloadable via the project website and GitHub. We'll also make prototypes of more resource-intensive versions for local use, one using a large ultra-high resolution screen and polarized glasses, for example, and one configured for the Oculus Rift virtual reality headset (both platforms at the USF AVC). All materials will be interconnected: the immersive 3D space will be directly linked to the archival materials (via stable URIs), which will serve as annotations of the space, and users will be able to try the digital emulations, view flowcharts and machine specifications, read documents, or listen to oral histories by former machine operators, all from within the virtual space.

#### **Environmental scan**

Some well-known digital archives dating back to the 1980s, such as the Perseus Project (<http://www.perseus.tufts.edu/hopper/#>), have combined various kinds of digitized materials in a single collection, often incorporating images and maps with textual documents, for example. After 2005, with the advent of Google Maps, GIS-based archaeological projects proliferated, and the problems surrounding the linking of metadata interactively to maps, images, and virtual environments have been much

discussed (see Koller, et al.). Our use of Maya and Unity reflects best practices in virtual archaeology, but it also reflects the world map + linked data-assets we associate with successful video games—which have a long tradition of representing “playable” technologies and machinery, as well as text annotations (for dialogue etc.), from within their immersive navigable environments. On our team, Kaplan has extensive experience in game design of this kind. A similar approach informs Angel David Nieves’s Soweto Historic GIS project (<http://www.angeldavidnieves.com/recent-unity-build-for-soweto/>), the interface for which links a 3D virtual space to maps, data, and scrollable annotations. We see valuable examples in this interface design, though our materials—including the emulations—will differ in significant ways, generically and in their relationship to the site in question.

Ultimately, there are few precedents for this kind of precisely focused project, using a confluence of digitization techniques across paper-based and artifactual archives, oral histories, and virtual space, in order to contribute to the history of a scientific or cultural institution. However, one exemplary effort just getting underway is the Vercelli Medieval Schoolroom Project under the direction of Gregory Heyworth (University of Rochester). It aims to recover a series of medieval texts as well as physical contexts, to construct a digital environment that models the physical space of a lost medieval school. Like this project, ours is organized around a physical site that was in effect constructed by the archival materials in question. In our case, more recent (mid twentieth-century) history is the focus, starting with the Busa Archive, instituted by the founder of the first humanities computing center, which is the physical and institutional space in question, the context of which is to be reconstructed.

### **Statement of Innovation**

Our project addresses important new historical questions, neglected until now, through an innovative methodology—applying a cluster of established approaches: the production of digitized archival materials in relation to a physical site; a cultural-heritage style virtualization of an architectural space; media archaeology and the reverse engineering of forgotten or obsolete technologies; and the production and curation of oral histories—in order to reconstruct the metaphorical and literal “site” of this extremely influential humanities research center. Ours is a deliberately eclectic and collaborative method for modeling history, an attempt to weave together archival data and virtualization of space.

### **Enhancing the Humanities**

Busa’s work was a historic milestone at the beginning of humanities computing in the postwar era, one strand of which fed into what we now call the digital humanities, an interdisciplinary field of increasing significance in the academy and the wider public arena. This project aims to recover, preserve, and make widely accessible foundational historical materials within a meaningful conceptual “architecture.” It will contribute to a deeper understanding of the role of technology in society by studying the origins of humanities computing at this groundbreaking moment. Technology extends to material and social infrastructure, and we aim to illuminate issues important to the humanities broadly considered, such as the gendered labor involved in this kind of work, the

specific pathways taken by Cold-War funding and corporate sponsorship for humanities research, and the emergence of new forms of interdisciplinarity across science and the humanities.

### **Larger goal**

Our larger intellectual purpose is to contribute a key chapter to the histories of science and technology, of pre-digital postwar computing, and of technology in the humanities. In addition, the project will offer a methodological case study of collaborative, distributed, multimodal digital modeling, a way of testing a more capacious concept of what “digitization” might entail. Jeffrey Schnapp of Harvard’s metaLAB has remarked that “every cultural object is a network,” and our goal is to map a set of networked cultural objects and documents which form the constellation within which CAAL makes sense. More generally, we take up what Schnapp has called the “defining design challenge of our epoch”—that is, “to weave together information and space in a meaningful fashion” (Schnapp). We aim for this project to serve as a test-case for the validity of that goal, and the panel presentation after the funded period, as well as the white paper and additional publication, will reflect on this aim in specific theoretical terms. The larger goal will persist beyond the funded period, as we continue to see the digitization of the Busa Archive and continue to enrich our multimodal project in many unforeseen ways.

### **Staff**

Steven E. Jones, drawing on research for his recent book on Busa, will manage the project as a whole and will work closely with Howard Kaplan and the AVC to build the virtual center.

Howard Kaplan, drawing on his experience in visualization, 3D modeling, and game design, will oversee staff at the AVC in creating the virtual center.

Geoffrey Rockwell, who has been working on the history of humanities computing, will oversee the creation of punched-card emulations, other software emulations of period machinery, and modeling workflow.

Paolo Senna and Marco Passarotti will oversee the digitization of selected materials in the Busa Archive in Milan, including ensuring appropriate imaging, metadata, management, and storage at the Università Cattolica Library.

Melissa Terras and Julian Nyhan, building on their research already in progress, will prepare appropriate selections of textual transcripts with translations and annotated audio files (MP3) of oral histories with Busa’s punched-card operators and other associates.

All staff members and unpaid advisers have conducted advanced research into Busa and CAAL and the history of humanities computing in general, and all team members have hands-on experience with the materials in the Busa Archive (see the biographies).