J. Impagliazzo, et al., eds. IFIP History of Nordic Computing 2 (Springer, 2009)

Organizing the History of Computing

'Lessons Learned' at the Charles Babbage Institute

Thomas J. Misa

Charles Babbage Institute, University of Minnesota, Minneapolis Minnesota 55455 USA
www.cbi.umn.edu

Abstract. This paper tries to distill some of the 'lessons learned' from the Charles Babbage Institute's quarter-century experience (1980-present) in organizing the history of computing. It draws on the author's (recent) experience as CBI director; conversations with Arthur Norberg, CBI's long-time founding director; and papers delivered at a special symposium appraising CBI's role in computing history, which appeared in the *IEEE Annals of the History of Computing* 29 no. 4 (October-December 2007).

Keywords: Charles Babbage Institute, oral history, archiving practices, electronic documentation, history of computing.

1 Introduction

When the Charles Babbage Institute (CBI) came to the University of Minnesota in 1980 (its origins were a few years earlier in California), no one knew what a research center in the history of computing should look like. Disciplinary research centers at the American Institute of Physics (founded 1961) and IEEE (also opened in 1980) were the closest peer institutions; there was also the Boston Computer Museum (f.1979) although the museum's relocation to the West Coast, let alone its present incarnation in Mountain View, California, was yet some years off.[1] The *Annals of the History of Computing*, then sponsored by the American Federation of Information Processing Societies (AFIPS), had started publishing in 1979. Arthur Norberg, with experience as a research historian at the Bancroft Library at U.C. Berkeley and at the National Science Foundation, arrived in Minnesota as CBI's first permanent director in 1981. He initially planned a full suite of activities, including doing oral histories, developing archival collections, conducting historical research, collecting economic data, engaging in policy work, and conducting educational and outreach activities. In fact, CBI has focused nearly all of its energies on the first three areas—oral histories, archival collections, and historical research. This paper will discuss each of these in turn, spotlighting choices that were made, assessing the consequences, and outlining challenges for the future.

[1] The UK's National Archive for the History of Computing (Manchester) was created in 1987; see [18].

J. Impagliazzo, T. Järvi, and P. Paju (Eds.): HiNC 2, IFIP AICT 303, pp. 1–12, 2009.

J. Impagliazzo, et al., eds. IFIP History of Nordic Computing 2 (Springer, 2009)

2 Developing Oral History Methods

Oral history as an identifiable field in the U.S. stretches back to the early days of folklore studies. In the past thirty years or so, the American Institute of Physics, the Smithsonian Institution, the Charles Babbage Institute, the Chemical Heritage Foundation, and several universities have developed oral history as a distinct method for conducting research in the recent history of science and technology. It is especially valuable when researchers have direct access to historical actors and/or where traditional documentary evidence is insufficient or unavailable. As Lillian Hoddeson recently emphasized, conducting oral histories requires the interplay of numerous skills: establishing trust with an interviewee, yet also being able at the right moment to ask probing, possibly difficult questions; relying on an interviewee's memory, yet also bringing documents that can help activate or sharpen memories (or establishing lines of questioning that can do the same).[2]

The "research grade" oral histories developed at CBI utilize a specific method and involve extensive interaction with the interviewee. They are nothing like a newspaper reporter fishing for a quotable anecdote or the popular free-form "reminiscence". CBI oral histories require extensive research beforehand by the interviewer, a substantial block of time for the tape-recorded interview itself (2-3 or more hours), and additional time for the subsequent process of transcription and editing. The interviewee has an opportunity to review the audiotape transcription, correcting mistakes and clarifying ambiguities. However, the goal is not to create a fluid, easy-to-read document ready for publishing in a popular magazine; rather, the goal is to preserve the nuances of the original conversation, while correcting obvious mistakes.[3] For better or worse, the edited transcript becomes the official version, and this is what is widely cited (we are recording 20,000 downloads a year from our online oral-history database). The audio recordings are retained, but in practice, researchers have rarely asked for them. The extensive time requirements for preparing, transcribing, editing, and (if permitted) web-publishing these oral histories make them an ideal research method for the study of certain key persons, who really need to be asked detailed questions about their activities, thoughts and motivations.

Ideally, interviewers bring contemporaneous documents and/or artifacts in order to encourage or stimulate the interviewee's memory. "Can you recall what you were thinking about in *this* memo . . . or in *this* diagram?" are such memory-invigorating questions. Hoddeson relates how an interview she did with physicist Richard Feynman, then 61, was going poorly owing to his recent surgery. He simply could not remember details of his wartime work, even when shown written documents he had drafted and personally signed. At last Hoddenson showed him a blueprint of the Oak Ridge (Tennessee) atomic facility, and this prompted one of Feynman's well-told "genius stories," in this instance about how he had spotted a crucial weakness in the complex system of pipes and pumps in the atomic facility's chemical plant (p. 193). Telling this story seemed literally to activate his memory, and soon additional stories and recollections were readily forthcoming. [8]

---

[2] See the AIP's guidelines <www.aip.org/history/oral_history/conducting.html> and [16] For connections to historical methods, see [8] and [17].

[3] For recent reflections on editing oral history manuscripts and preparing them for publication, see [9].

J. Impagliazzo, et al., eds. IFIP History of Nordic Computing 2 (Springer, 2009)

In addition to documents, key artifacts or even close colleagues can be helpful in jogging or jolting the interviewee's memory. Doing a group interview with colleagues has several obvious attractions, not least the pleasant social time and opportunity to reminisce with friends; yet it is important to keep in mind *why* one does the interview. If there are a specific set of research questions, then the interviewer must work extra hard to keep two or more interviewees on track and focused. Having a chance to sit around and record some of the "old stories" is frequently attractive to the interviewees, but then the interviewer needs to take special care. Researchers studying the complex phenomena of memory suggest that we do not really directly remember the original (ur–)memory; rather, we recall the last re-telling or recollection of that memory. If, somewhere down the years, the re-told story has wandered away from the original, then the oral historian's elaborate process of recording, transcribing, and validating of this more-or-less accurate storytelling as a "historical fact" can be seriously in error.

Controversies surrounding oral histories tend to crop up when someone asserts too much for the genre. The historian of modern biology Horace Freeland Judson, whose *The Eighth Day of Creation* (1979) depended heavily on oral histories with colorful pioneering figures in molecular biology, has been criticized in these terms [10]. (An additional liability is that researchers may consult his 52 oral histories, deposited at the American Philosophical Society, "only with the prior written permission" of Judson; moreover, several of them are "closed indefinitely to researchers."[4]) After significant time spent with pioneering figures in a field, it is certainly tempting to see their retrospective accounts in a favorable, even privileged light. A better approach recognizes that oral histories are one subjective source—no worse but certainly no better than any other single source. The pragmatic solution when faced with this seemingly intractable problem is to rely on the historian's fundamental methods of *source criticism* (paying careful attention to the perspective and potential bias of any source, as well as examining its reliability, authenticity, and distance from events) and *triangulation* (seeking to find corroboration among varied sources, whether documents, published accounts, artifacts or oral histories). It goes without saying that open access to all relevant documents—including, ideally, interviews and transcripts—is necessary for such triangulation.

It is important to emphasize that different modes of interviews, and other complementary research methods and sources, will be necessary to keep pace with changing questions and themes in the history of computing, especially the recent interest in the users of computing and the emerging awareness of the important roles played by non-pioneering figures. It certainly makes sense to invest many hours of intensive research and preparation time to be able to ask insightful questions of a pioneering figure. These figures can be the most valuable interviewees—and, sometimes, also the most troublesome. Often, people who have been in the public eye as a prominent scientist, engineer, or business executive have already experienced all the "usual questions" and have already formulated answers to them; consciously or not, they have erected or constructed a "mask", and it may not be easy to go beyond their public persona or identity. A careful oral historian can choose to avoid posing the usual questions, or to ask knowledgeable follow-up questions in the case of a pat reply. Hoddeson relates her mistake in interviewing a certain Bell Laboratories executive in his office, where over

[4] See <www.amphilsoc.org/library/mole/j/judson.htm>.

J. Impagliazzo, et al., eds. IFIP History of Nordic Computing 2 (Springer, 2009)

the years he had held numerous interviews and where his "mask" was secure: she conducted a well-paced interview, and recorded all the "usual stories" that he'd already retold many times before.

Over the years, CBI researchers have done nearly 300 interviews, and, including those done by our collaborators and colleagues, our on-line database contains nearly 400 interviews. You can directly query the database by name or subject; for a suggestive listing see <www.cbi.umn.edu/oh/subject.phtml>; most of the interviews are available on-line as PDFs. Looking down the subject list at "X" there are entries for Xerox Data Systems and Xerox's Palo Alto Research Center that lead to a 52-page interview with Paul A. Strassmann, who was chief computer executive at Xerox. From his vantage point in 1989, Strausmann described the interactions between Xerox's mainframe-oriented Data Systems (XDS), and Xerox's established copier business, as well as the growth of Xerox Palo Alto Research Center (Xerox PARC) and its development of the Alto and Star computers. Staussman recalls Xerox's decision to embrace an integrated view of information technology and to distance itself from computers per se. A listing under "XML" leads to a 36-page interview with Don Chamberlin, who in 2001 was a researcher at IBM–San Jose. Chamberlin recounted his early life, his education at Harvey Mudd College and Stanford University, and his work on relational database technology. Together with Ray Boyce, Chamberlin developed the SQL database language. He also briefly discussed his recent research on XML query languages. While Chamberlin's interview is something like a career summary, Strassmann's interview is more focused on a specific phase of his career.

CBI's established model of oral histories, a particular model to be sure, has been so successful that only recently have we seriously considered other types and modes of interviews. At CBI, we have several different types of experiments for a research project on the National Science Foundation's FastLane system. Today, FastLane forms a comprehensive information infrastructure used in all phases of NSF's grant making, its core mission. To deal with the large number of legacy users at NSF as well as sponsored projects staff and researchers at universities, we plan shorter, targeted interviews with non-pioneering figures (combined with longer, traditional interviews with the key FastLane designers). We have a plan for a semi-automated transcription technique, using voice-recognition software; unfortunately, you cannot (yet) have your interviewee speak directly into a computer to produce a transcript, since the software needs to be 'trained' for a certain voice.

We also are developing a web-based interview platform for our research. This activity connects a web-based front end to a database for recording respondents' answers to semi-structured interview questions as well as their own unstructured anecdotes or stories (the capacity to do binary-format uploads will allow a degree of self-archiving, since users will be able to upload images, documents, and even binary files of their own spoken interview answers). We also are experimenting with a wiki to facilitate our interviewees participating in the writing of history.[5]

Our goal with these new methods is to gain access to people and sources that, practically speaking, were not available using our traditional, time-intensive oral history methods. We hope the new methods might better 'scale' with a larger numbers of

[5] Various experiments in Web-based history have been funded by the Sloan Foundation. For candid reports, see [6,7]. Early results were reported in [12].

J. Impagliazzo, et al., eds. IFIP History of Nordic Computing 2 (Springer, 2009)

interviewees. Again, this is a good example of the interplay of changing historical questions with new research methods and tools. Time-intensive oral histories made sense when our research questions concerned a limited number of pioneering figures. We need new research methods to understand a larger and more diverse group of individuals.

3 Evolving Archiving Practices

A second activity is archiving the papers of individuals, institutions, companies, and professional groups, which was a priority of CBI from the start. Here too Arthur Norberg and his colleagues (for most of its history CBI has had both professional historians and professional archivists on staff) faced some difficult choices. While the industry and professional groups advising CBI initially advocated a strategy of "saving everything", CBI's actual collections strategy was much more focused. Bill Aspray has recently described these debates, noting the several shortcomings in the save-everything collecting strategy. These involve numerous practical and intellectual problems: vexing problems of space, cost, appraisal, preservation, indexing, provenance, and (not least) effective cooperation with other peer organizations and institutions.[2]

In time, CBI evolved into a true physical archive, with extensive on-site storage, active collections development, professional archiving, and full-service accommodations for researchers. As far as I know, little thought was ever given to CBI serving as an archival clearinghouse such as the Oxford–Bath archives of contemporary science. The problem was not one of coordinating existing collections. Rather, there was a pressing need for some *place* that would collect, preserve, and make publicly accessible records and documentation on the history of computing (most university libraries in the U.S. are rather reluctant to take non-literary materials from beyond their own campuses), and a salutary awareness of that need on the part of the technical-professional community. The thought was that an important part of the history of computing had just occurred, and that it would likely be lost unless some institution such as CBI collected that history. Even today, CBI often seems to serve as a sort of "last chance" for computer history records, especially those held by individuals.

CBI's active collecting policy and practices took form with intellectual input from the technical community, Norberg, and Bruce Bruemmer, CBI's first professional archivist who served from 1984 to 1997. (Subsequently, CBI's full-time archivists have been Elisabeth Kaplan from 1999-2005 and Arvid Nelsen, hired in April 2007.[6]) For several years, CBI had the typical archival storage in the basement of an older building (in the University of Minnesota's Engineering Library). In 2000, CBI moved into a new, purpose-built special collections library. Andersen Library features office suites and reading rooms for the University's archives and special collections units as well as state-of-the-art climate-controlled archival storage space underground. The facility exploits the geological fact along the Mississippi River basin that layers of soft sandstone alternate with layers of hard limestone; the underground storage cavern was created when two 600-foot-long holes were bored through the sandstone layer,

[6] For a more detailed history, see [3].

J. Impagliazzo, et al., eds. IFIP History of Nordic Computing 2 (Springer, 2009)

and then reinforced with a concrete tube structure. Not only were construction costs quite reasonable, but also the ongoing energy costs are low since little temperature regulation needs to be done for the underground space (corrections to humidity are still needed). The facility was built with a special appropriation of Minnesota-state funding since the state needed large-scale remote storage when constructing or moving an entire library; accordingly, you can back a large U.S.-size "tractor-trailer" unit right into the loading area, offload boxes onto a gravity-slide conveyor ramp, and then store boxes on metal racks that stretch nearly two stories high.[7]

There are many irreplaceable gems in CBI's archives, but not every computer manual could—or should—be given the exacting archive-grade preservation required for the most valuable historical materials. Generally CBI followed the standards of the professional archiving community, which embraced a high degree of physical preservation, a detailed level of collections processing, extensive indexing of the collection, and broad dissemination of the resulting finding guides. Consequently, most of the archival material in CBI's 250 collections—including its two largest corporate collections (Burroughs and Control Data)—are processed down to the "folder" level. Researchers have access to these collections through a keyword search tool <http://discover.lib.umn.edu/findaid>. A recent update permits detailed and accurate searching across CBI's collections (as well as across the University's other archival collections as desired). Really, for the first time, we have gained practical access to the 731 boxes of the United States National Bureau of Standards sprawling collection of computer literature (1956-1978). CBI's U.S.-oriented collections are well known to computer historians, but recently we made a survey of our surprisingly wide international holdings (including founding documents on IFIP, documentation on ALGOL, information on Soviet bloc computing, and other topics, as well as the only WorldCat-listed copy of James Connolly's *History of Computing in Europe* [published by IBM World Trade Corporation in 1967]). [5, 20]

One challenge we are confronting today in archiving is the tradeoff between processing and accessibility. Put simply, with the accepted model of extensive processing, our collections have not typically been available to researchers until after we process them—sometimes for many years. Archivists in the U.S. have lately favored a model of "minimal" processing with an acceptance of describing collections, at least initially, at the overview box-level rather than at the more detailed folder level. Collections are more quickly made available to researchers, even though they may later undergo further processing (at the folder level). Another challenge we face is continuing to round out our existing collections from the "mainframe" era while starting or expanding more-recent collections documenting the personal computer, networking, graphics, office automation, internet, and mobile computing.

It will be interesting to observe the interactions between archival theory and practices in the age of electronic and born digital records. There are certainly numerous collections of scanned documents available on-line today.[8] One solution, it may well

---

[7] Information on the Andersen Library building, including construction photographs, may be found at <andersen.lib.umn.edu/aboutandersen.html>.

[8] See (e.g.) the Alan Turing Archive for the History of Computing <www.alanturing.net/>; Mike Muuss' History of Computing Information <ftp.arl.mil/~mike/comphist/>; and J. A. N. Lee's materials <ei.cs.vt.edu/~history/>.

J. Impagliazzo, et al., eds. IFIP History of Nordic Computing 2 (Springer, 2009)

appear, to the practical problem of where to store large volumes of records is simply to digitize the whole lot and throw out the bulky paper originals. More than one person has observed to me that if Google can effectively search the entire world wide web, then certainly a decent search engine could retrieve scanned, digitized documents in an archive such as CBI. I think the problem may be more complex than it first appears. To begin, a digitized image of a document—a memo, research report, or drawing—is not searchable unless it has been OCR-ed and/or someone has assigned metadata to it (and both of these are time-intensive processes). There is the attractive hybrid, raw-OCR-plus-image model used by the "Making of America" project at Michigan and Cornell. This project gives its scanned books, journal articles, and other documentation from the 19th century a raw OCR (with around 99% but not 100% accuracy); the OCR-text results are entered into a searchable database; and these entries are linked to high-quality scanned images of the original pages. Yet even with substantial funding by the Mellon Foundation, and a decade of large-scale university effort at Cornell and Michigan starting in 1995, the total number of pages has reached just 3.8 million.[9]

At CBI, we face a numbers-and-resource-allocation question. CBI presently holds around 5,000 shelf feet of paper documents, or more than 5 million pages, in addition to a growing electronic archive. Our documentation is very diverse: loose handwritten notes, formal research reports, market surveys, laboratory notebooks, bound and unbound journals and books. Some of these "pages" are viewed regularly, at least as archival collections go, but many of them might be called up and examined only once in a very long time. Digitizing each and every of these pages would represent an enormous commitment in staff time, not only for the scanning itself, but more importantly for describing the documents, indexing them, and setting up and maintaining a database retrieval scheme. We estimate that such archival grade scanning costs us around $1 per page (although, like Google, we can scan second copies of book pages at lower cost). Corporate lawyers might have the resources to do this and to make large datasets of trial documents available and retrievable, even on-line; but it is not clear that this would be an effective use of archiving resources at CBI's (modest) scale of operations. If we made a full-scale commitment to digitizing our present collection, our collecting of new materials would probably grind to a halt.

Now, consider the status of a document not merely as a container for "information" but also as an artifact. CBI has a very large collection of computer manuals and other printed documentation, more than 500 boxes total, of which approximately half is processed and half unprocessed.[10] We are under some institutional pressure to reduce the size of the so-called unprocessed collections. Computer manuals are by no means unique; and there are catalogues of on-line computer manuals.[11] Yet a preliminary investigation of our users suggests that the computer manuals have received far more use than we had assumed. Then, too, taking (say) one of the three-ring notebook binders out of its box, and examining how it was used (colored tabs); even the smell of the

---

[9] See <quod.lib.umich.edu/m/moagrp/index.html> and <cdl.library.cornell.edu/moa/>. The current page count (June 2007) is at <quod.lib.umich.edu/m/moagrp/>.

[10] For the Computer Product Manuals Collection (1940s to 1980s) see <www.cbi.umn.edu /collections/inv/cbi00060.html> or <special.lib.umn.edu/findaid/xml/cbi00060.xml>.

[11] For the Manx catalogue, see <vt100.net/manx/>; a collection of documentation and software can be found at <www.bitsavers.org/>.

J. Impagliazzo, et al., eds. IFIP History of Nordic Computing 2 (Springer, 2009)

cheap plastic (no fine leather bindings here) tells you something about the social and cultural history of computing. If we digitized the "content" of the computer manuals and purged them as artifacts, we would certainly gain some open shelf-feet but we would also have lost the contextual information of the artifact itself, with clues to how and why programmers used it.

Finally, let me conclude this section with some thoughts on the "migration model" for electronic or born-digital records. A National Archives document categorizes the many on-going experiments in preserving digital records into five broad areas:[12]

(a) preserving the original technology used to create or store the records;
(b) emulating the original technology on new platforms;
(c) migrating the software necessary to retrieve, deliver, and use the records;
(d) migrating the records to up-to-date formats;
(e) converting records to standard forms.

The migration model focuses on (c) and (d) of these options. (The "Making of America" project mentioned above stores its scanned book and magazine pages in 600 dpi TIFF images, a stable and standard non-compressed form, as in (e); and then converts the huge TIFF files on-demand to compressed GIF images for delivery over the WWW.)

One certain liability of paper records is their physical bulk and the costs needed to store them securely over the long term. Yet one advantage is that once the paper records have been processed they are stable: during IBM's crisis in the 1990s it was forced to close its archives for a number of years, but, when the company's economic prospects improved, all that was needed was to unlock the IBM archive's doors and re-open the facility. Its paper records were still there waiting patiently. Electronic records may not so successfully survive such a temporary closure. Archiving experts are exploring the "migration model" to store electronic records over time. Records stored in yesterday's obsolete format need to be "migrated" to today's standard format to ensure continued accessibility; alternately, software capable of reading and interpreting the ancient data will need to be migrated onto up-to-date hardware. In the future, with the assumption of changing software and/or data formats, there will be the need for further migrations—a little-acknowledged downside to the seemingly relentless march of Moore's Law.[13]

It seems to me that the "migration model" entirely transforms electronic archiving: it was supposed to be a low-cost enterprise, but instead it might be one with surprisingly high and recurrent costs. Here are some considerations to migrating records themselves. First, since no piece of software can be perfect, the downstream or migrated records will have some minor but essentially undetectable errors; consequently, it seems very risky to dispose of the original records, and quite likely the same for the second and subsequent generations as well. So instead of having a single-generation electronic archive, we are likely to have multiple-generation archives, one for each of the migration generations. Over time a single collection of electronic records, which

---

[12] See Kenneth Thibodeau, "Preservation and Migration of Electronic Records: The State of the Issue" <www.archives.gov/era/papers/preservation.html> (accessed May 2007).

[13] Or worse, as the Thibodeau (supra) suggests: "the market has tended to exacerbate the problem of preserving electronic records. The pressures of competition have led the [IT] industry to obey Moore's law, replacing both hardware and software on a frequency of two years or less."

J. Impagliazzo, et al., eds. IFIP History of Nordic Computing 2 (Springer, 2009)

might have been stored in a fixed number of traditional archival boxes, will actually grow in size and complexity. Second, I think the migration model typically assumes a linear progression: you start with the "1st" generation of original records and migrate from there to the "2nd" generation and from the "2nd" to the "3rd", and so on. Remember each migration is more or less forced by soon-to-be obsolete formats (we are entirely assuming stable storage media and/or well-secured servers, another cost that should not be forgotten). If an archive is unable to conduct one of these generational migrations, it seems in great danger of having its collections slip off the standard migration pipeline. Hypothetically, what if an archive was forced to be dormant for ten years—or for any reason simply forced to skip one or more required migrations? Now, think about the next 75 or 150 years. Given these problems, some archival specialists reasonably advocate just printing out emails or other electronics documents and archiving them based on the well-tried paper models.

4 Conducting Historical Research

Third, CBI has conducted an active research program throughout the past quarter century. Indeed, you can chart CBI's longer-term research projects and find a close correspondence with shifting emphases in the oral histories conducted and even in the specific collections taken in. This, too, is part of the CBI model: our research projects have led to oral histories and networking in various communities, which in turn has frequently led to collection development in those communities, and, in the longer term, the creation of an entire infrastructure for research. For example, during the years when CBI did a NSF project on the "Computer as a Scientific Instrument" (1998-2001) CBI's oral histories focused in complementary fashion on NSF staffers and computer researchers active in the area; and similarly for another NSF-funded project on "Building a Future for Software History" (1999-2003). Several important archival collections in early networking and computer science came during the (sponsored) project that examined DARPA's influence on computer science, and which resulted in Norberg and Judy O'Neill's *Transforming Computer Technology: Information Processing for the Pentagon, 1962-1986* [14]. Norberg's own research on Minnesota's pioneering computer industry, including numerous oral histories and many individual and company collections, was published as *Computers and Commerce* [15]. Capitalizing in part on CBI's accumulated records, Jeffrey Yost, CBI's associate director, recently published a valuable overview titled *The Computer Industry* [19]. In short, there has been a continual interplay between research, oral history, and archiving activities, and it is difficult to imagine CBI today not having all of these functions.

The archives staff, too, has conducted substantial field-shaping projects, in addition to the processing of collections, the handling of reference questions, and the provision of research assistance. *The High-Technology Company: A Historical Research and Archival Guide* (written by CBI staffers Bruce Bruemmer and Sheldon Hochheiser and published in 1989) was for some years the best and indeed only resource concerning this sector, and distributed by the Society of American Archivists [4]. Recently, CBI staff conducted a study "Documenting Internet2: A Collaborative Model for Developing Electronic Records Capacities in the Small Archival Repository."[14]

[14] A list of CBI's publications is at <www.cbi.umn.edu/research/staff_publications.pdf>.

J. Impagliazzo, et al., eds. IFIP History of Nordic Computing 2 (Springer, 2009)

Support for scholars working on their Ph.D. in the field has also been an important on-going activity through the CBI–Tomash Fellowship. Awarded annually since 1978, the list-to-date of 28 fellows reads something like a who's who in computing history.[15] Collectively, they have published highly respected works in all areas of the history of computing. In scholarly terms, the CBI–Tomash dissertations have resulted in the two best books on Silicon Valley, the definitive study of the internet, key studies of computing in organizations, the corporate and Cold War contexts, and several scientific disciplines, the key study of magnetic recording technology, as well as pioneering studies of computing in Chile and Italy, and comparative studies with Japan and England. The recent volume from MIT Press on the commercial internet has chapters written by no less than six former CBI–Tomash fellows, including those of both co-editors (William Aspray and Paul Ceruzzi) [1].

5 Conclusion

In conclusion, here are three items in my personal "wish list" of initiatives for the history of computing. One, I would like to develop a software tool for archiving email: it would permit the capture, sorting, editing, and archiving of emails, which currently arrive at our shop on 5.25 floppies, 3.5 floppies, as electronic files, and on CDs in a wide variety of formats. I think it is a good bet that emails will have high research value in the universe of "born digital" records. Emails frequently contain a heterogeneous mixture of content, even in the same message: personal or private comments, professional business, a bit of gossip here and there, as well as binary-code attachments such as images or Word documents or Excel spreadsheets. I believe that potential email donors would respond positively if we had an effective means for removing their private content, just as slips of paper with private comments were certainly removed from traditional paper archives, while leaving intact their professional, technical, or public content. Such privacy concerns are prominent in collections of emails we are working on now. Of course, at a certain moment, the edited content in such a database would need stability, to be locked down, so that subsequent users of the email database do not continue the editing process. Emails also contain a wealth of useful meta-data about how and when and where the message was sent. So for these reasons there are several obvious shortcomings with (say) creating PDF files from emails with the personal bits blacked out,[16] or just printing them out.

I also would like an institutional or collective means of "finding good homes" for archival collections that might not fit our collecting priorities but that still ought to be preserved somewhere (we have several of these in process right now). And, finally, I would like to think prospectively about how changes in our field's topics, themes, and

---

[15] The CBI-Tomash fellows are listed at <www.cbi.umn.edu/research/recipients.html>.

[16] For problems with improperly "redacted" or blacked-out text in PDF documents, see the well-publicized travails of the CIA, AT&T, and others: <cryptome.org/cia-iran.htm>, <it.slashdot.org/it/06/06/22/138210.shtml>, <news.com.com/2100-1028_3-6077353.html>. The NSA advises "any sensitive information must be removed from the document through deletion"; see [13] p. 4.

J. Impagliazzo, et al., eds. IFIP History of Nordic Computing 2 (Springer, 2009)

research questions might shape and direct our archiving strategies and practices [11]. I welcome continuing opportunities to share and discuss our experiences.

References

[1] Aspray, W., Ceruzzi, P.E. (eds.): The Internet and American Business. MIT Press, Cambridge (2008)

[2] Aspray, W.: Leadership in Computing History: Arthur Norberg and the Charles Babbage Institute. IEEE Annals of the History of Computing 29(4), 16–26 (2007)

[3] Bruemmer, B., Kaplan, E.: Realizing the Concept: A History of the CBI Archives. IEEE Annals of the History of Computing 23(4), 29–38 (2001)

[4] Bruemmer, B., Hochheiser, S.: The High-Technology Company: A Historical Research and Archival Guide. Charles Babbage Institute, Minneapolis (1989), Available on-line, along with other research resources, http://www.cbi.umn.edu/hostedpublications/

[5] Connolly, J.: History of Computing in Europe. IBM World Trade Corporation (1967)

[6] Hessenbruch, A.: The Trials and Promise of a Web-History of Materials Research. In: Grandin, K., Wormbs, N., Widmalm, S. (eds.) The Science–Industry Nexus, pp. 397–413. Science History Publications, Sagamore Beach (2004)

[7] Hessenbruch, A.: 'The Mutt Historian': The Perils and Opportunities of Doing History of Science On-line. In: Doel, R.E., Söderqvist, T. (eds.) The Historiograhy of Contemporary Science, Technology, and Medicine, pp. 279–298. Routledge, New York (2006)

[8] Hoddeson, L.: The Conflict of Memories and Documents: Dilemmas and Pragmatics of Oral History. In: Doel, R.E., Söderqvist, T. (eds.) The Historiography of Contemporary Science, Technology, and Medicine, pp. 187–200. Routledge, New York (2006)

[9] Jones, R.: Blended Voices: Crafting a Narrative from Oral History Interviews. Oral History Review 31(1), 23–42 (2004)

[10] Judson, H.F.: The Eighth Day of Creation: Makers of the Revolution in Biology. Simon and Schuster, New York (1979)

[11] Misa, T.: Understanding How Computing Has Changed the World. IEEE Annals of the History of Computing 29(4), 52–63 (2007)

[12] Misa, T., Zepcevski, J.: Realizing user-centered computer history: Designing and using NSF's FastLane (1990-present). Paper to annual meeting of the Society for the History of Technology, Lisbon, October 12-14 (2008), http://netfiles.umn.edu/users/tmisa/www/papers/FastLane_SHOT-2008.pdf

[13] National Security Agency. Redacting with Confidence: How to Safely Publish Sanitized Reports Converted From Word to PDF (NSA Information Assurance Directorate Report # I333-015R-2005), http://www.fas.org/sgp/othergov/dod/nsa-redact.pdf

[14] Norberg, A., O'Neill, J.: Transforming Computer Technology: Information Processing for the Pentagon, pp. 1962–1986. Johns Hopkins University Press, Baltimore (1996)

[15] Norberg, A.: Computers and Commerce: A Study of Technology and Management at Eckert-Mauchly Computer Company, Engineering Research Associates, and Remington Rand, pp. 1946–1957. MIT Press, Cambridge (2005)

[16] Norberg, A.: How to Conduct and Preserve an Oral History, IT History Society (2008), http://www.ithistory.org/resources/norberg-article.pdf

J. Impagliazzo, et al., eds. IFIP History of Nordic Computing 2 (Springer, 2009)

[17] Thomson, A.: Four Paradigm Transformations in Oral History. Oral History Review 34(1), 49–70 (2006)

[18] Tweedale, G.: The National Archive for the History of Computing. Journal of the Society of Archivists 10(1), 1–8 (1989)

[19] Yost, J.R.: The Computer Industry. Greenwood Press, Westport (2005)

[20] Yost, J.R.: Exploring the Archives—Part One: International Records. CBI Newsletter 29(1), 8–11 (Spring 2007), http://www.cbi.umn.edu/about/nsl/v29n1graphics.pdf