Electronically Generated Records and Twentieth Century History

Ronald W. Zweig

Department of Jewish History, Tel Aviv University, Ramat Aviv, Israel
e-mail: ron@ccsg.tau.ac.il

Abstract: The electronic generation of documents in modern offices will transform the nature of archives, and also the techniques of historical research. Although considerable attention has been directed to developing research methodologies for social and economic history using computerized numeric data, almost no attention has been paid to the impact of machine readable textual records on historical writing. This article considers the advantages and disadvantages for the historian of the shift from paper records to electronic documents, and suggests a number of approaches to historical research made possible by the new technology.

Key Words: archives, electronic records, electronic office systems, historical research, hypertext, HYTIME, linked documents, ODA, SGML, tagging, TEI, textual records, textual analysis

Growing numbers of social and economic historians use computers and data processing techniques in their research. They have generated their own datasets according to their particular interests, or have converted existing paper records into machine readable format. This discussion will consider a different category of source material: textual records that have been created on computers and stored electronically. It examines our future ability to access and use in research the national historical record being created today. The advent of electronic archives will create entirely unforeseen problems for diplomatic and political historians. Equally, however, it embodies new potentialities for information retrieval and analysis. This article will discuss both the problems and some of the possibilities created by the new technology.

Increasingly, governments are using computers to create documents. Electronically generated documents are usually converted into paper records, and it is the latter which are preserved for record purposes. However, this situation will change as personnel become more familiar with electronic records. According to the American National Academy of Public Administration, this preference for paper copies of electronic records will persist for another five to ten years. The more important or prestigious the government department is, the more likely it will be to have "state of the art" technology – integrated office systems and not just isolated word processors. Such systems store and distribute records electronically. As a result, a growing proportion of transactions in the decision-

The facts are really not at all like fish on the fishmonger's slab. They are like fish swimming about in a vast and sometimes inaccessible ocean; and what the historian catches will depend, partly on chance, but mainly on what part of the ocean he chooses to fish in and what tackle he chooses to use – these two factors being, of course, determined by the kind of fish he wants to catch.

Ronald Zweig is a graduate of the University of Sydney (Australia) and Cambridge University (UK). He is currently a Senior Lecturer in Jewish History, and Chairman of the Computing Committee, Faculty of Humanities, Tel Aviv University. He is the author of Britain and Palestine the Second World War; and German Reparations and the Jewish World; he is also editor of David Ben-Gurion: Political Leadership in Israel, as well as the journal Studies in Zionism.

making process are entirely electronic and only exist digitally. Already, entire offices work without paper backups, especially when dealing with documents generated for internal use. As archivists recognize, the “safety-net assurance that there are paper backups is increasingly untrue or even irrelevant.”

A recent Congressional Committee on Government Operations notes that by the end of this decade 75 per cent of all government transactions will be in electronic form.

In those departments where the perceived importance of the subjects dealt with is high – Prime Ministerial departments, Cabinet Offices, Foreign Offices, Defense, and the intelligence community – manual registries and conventional indexing of files are being replaced by computerized databases and information retrieval techniques, independently of the electronic generation of new records. Moreover, valuable historical records of current utility to governments, such as Cabinet protocols, are often made available retroactively in electronic format for the internal use of policy makers. That is, textual paper records created by conventional means in the 1940s and 1950s have been copied into machine readable format to facilitate current use.

If these electronic versions become available to researchers together with conventional paper records in accordance with the 30-year rule, then diplomatic and political historians would be well advised to start preparing themselves for the electronic archives of the future. This is true for our own research methodologies, as well as for those we teach our graduate students.

There is already a growing awareness that the electronic record itself, and not just its paper copy, is worth preserving. The National Archives of Canada has recently accessioned the complete electronic records of a government office (1.5 gigabytes for the 26 months that the office functioned – the equivalent of 2.66 million documents).

The growing number of electronic datasets accessioned by the National Archives of the United States, consisting mainly but not entirely of numeric data, indicates the changing nature of archival records. Some of the current holdings of the National Archives were created in World War II by the punchcard technology in use since the 1880s. However, most of the electronic records in the Center’s holdings have been created since the 1960s. In 1988, 167 datasets were sent to the National Archives by government bodies. In 1989, the number had grown to 745, and by FY 1991, 1632 datasets were received. The holdings of the Center for Electronic Records at the National Archives currently include over 10,000 distinct collections.

The National Historical Publications and Records Commission in the United States has described the management of information in electronic format as “the most significant and difficult challenge currently confronting the archival community.”

The archival profession has identified many technical problems associated with the accessioning and preservation of electronic records. First and foremost is the problem of standardization of document format, a question which is only now being addressed. The Office Document Architecture and Office Document Interchange Format (ODA/ODIF) provide for the transferability of text in a format that can be processed and used by radically different office systems. The standard itself is independent of hardware and software technology for storage, retrieval and interchange. Besides defining the means of recording the format of a document, the Office Document Architecture makes provision for a document profile that records information concerning the storage and management of the document (including title, subject, author, size, revision history and security attributes). There are other, competing standards that are also under consideration.

However, formulating a standard is only the first step. The largest problem is implementing it. A study conducted in Germany during 1988 examined fourteen commercially available document administration systems. Nine of these systems were part of complete electronic office systems, and others were purely archival. As the survey discovered, “none of the programs investigated provided documentation of the history of either the individual texts or the complete filing system.”

It will take years before one of these standards is adopted, implemented in commercially available software, and universally used. In all