

Between 20 Years and 2000: Sorting out Electronic Records in an Operational Context

Kenneth Thibodeau

I will address the InterPARES case studies from a pragmatic perspective. Specifically, what value do they offer to archival institutions?

One potential value is quite obvious to anyone with a pragmatic bias. If you want to do something about electronic records, you have to know what they are. You have to look at them both in depth and extensively. The individual InterPARES case studies are in-depth examinations of specific processes that (1) should produce and keep records and (2) rely significantly on computers; applications of computer technology. The aggregate of case studies around the world gives them the necessary breadth. The studies examine applications that include, are thought to include, or might include significant electronic records. Going into a case we do not know if the system actually keeps, or even could keep, electronic records. We could pre-screen cases according to this threshold, but not doing so puts us in a better position to learn what works and what does not work. Lessons of this sort will significantly enrich what the project eventually says about best practices.

Of course, on the whole we do need data – a lot of it – about real electronic records, and we need analysis of the data. The value of both is obvious, and in NARA's case direct. The collection based persistent object preservation method being developed for us by the National Partnership for Advanced Computational Infrastructure offers great promise for preserving electronic records in a way that is both immune to the vagaries of technological obsolescence and receptive to the benefits offered by technological advances. The method depends on archivists' ability to express abstractly the properties of electronic records that must be preserved. We must produce explicit models both of individual

records and of bodies of records. We must model all the different types of records that we will need to preserve and also any arbitrarily organized, arbitrarily complex body of records that we may receive. The InterPARES case studies will contribute substantially to addressing these needs both through the data about electronic records that is collected and in the norms concerning authenticity that will be produced by the analysis of the data.

But the collection and analysis of data about electronic records will need to continue for a long time. We are only on the threshold of the digital era. Innovations yet to come without doubt will be at least as great as those we have seen so far. Continuous change in information technology means that the challenges posed by electronic records are inherently dynamic. They will themselves change.

One of the most important premises of the information management architecture that underlies persistent object preservation is precisely that it must be independent of the infrastructure on which it is built. While we must rely on specific hardware and software, archival systems should be constructed in a manner that (1) enables any component of hardware or software to be replaced with minimum impact on the system, (2) allows such replacement with no significant impact on the records being preserved, and (3) enables the archival system to deliver the preserved records to other systems, even systems that have not been invented yet. The possibilities that this technology strategy offers of building inherently dynamic solutions to archival requirements only heightens our need to express archival requirements in a way that is not bound to the limitations of our knowledge of electronic records now, or at any other specific time in the future. As we build a rich store of data and analyses in the INTERPARES case studies, we must develop knowledge that is both greater than the sum of its parts and open ended. We need to expand our knowledge to new types of electronic records and new ways of organizing them that will emerge in the future. We also need to adjust archival concepts of the nature of records, of the archival bonds that link them, and of the evidence and information they provide.

Without abandoning well established principles and theories, we must be open to new realities and avoid distorting them through the constraints of anachronistic concepts and parameters.

Electronic records differ from traditional paper records in basic ways: in how they exist as durable objects, in how the parts of the record coalesce into the whole, and in the boundaries of the record. Traditionally, seen from an empirical perspective, a record is a specific aggregate of information, structured in a specific way, with that structure materialized in the inscription of the information on a durable medium. With an electronic record, there is no necessary connection between the structure of the record and its inscription on a storage medium. In fact, in the INTERPARES Preservation Task Force we are building a model of archival preservation on the recognition that it is literally impossible to preserve an electronic record. You can store the contents of the record, along with special bit strings that indicate how it should be structured and presented, but the sum of those bits is not itself the record. The stored bits always require the application of some software to be put into a state that is recognizably that of a record. Moreover, the aggregate of information that comprises the contents of a record may not be stored together. The assembly of the contents into the record may require sophisticated and complex processing. This processing may be transparent, and even unknowable, to the persons who create or use the record.

Such a simple thing as the appearance of an electronic record is not necessarily an attribute of the record itself. It may depend more on the hardware and software used to present the record, than on what is in the stored bits. The appearance of the bits depends on things like screen size and resolution and on the specific software used, and the appearance can be changed easily and often as a result of user options, such as window size, zooming, or switching between draft and page mode. I often sense that, while archivists may explicitly acknowledge such differences, in practice they are hindered by an

attachment to the notion of a record as a fixed inscription bounded on a two dimensional medium. That notion is increasingly inappropriate.

While a paper record can only be filed in one place, computer technology makes it possible for not only individual records, but even components of records to participate simultaneously in different arrangements. A record can be assigned to as many files as needed because the connection is logical not physical. Similarly, a name, address or any other data can be stored in one place, but be part of many different records. This serves the same basic purpose as designating a 'record copy' of paper documents: it facilitates ensuring the reliability of information, but at a finer level of granularity. Recognizing the expanded possibilities that digital technology creates for establishing and maintaining links among recorded information objects, restricting the types of links that are regarded as archivally significant to those that mimic the fixing of ink on paper and the sequential arrangement of paper files? The answer should come not from information technology, but from consideration of the characteristics of records as such and their evidential value.

Some documents on Web pages have a dynamic character: they only acquire a specific form or contents as a result of choices made by the reader. On the Web, relationships among records may be very fluid because the reader has options for navigating at will through an effectively unlimited chain of hypertext links. Of course, the fact that an individual may follow hyperlinks in the course of an activity is not sufficient to establish that the linked documents are records of that activity. But we must at least admit that the practice of consigning a record to a physical file is only a technique used to establish the archival bond between a document and other records. Although it may be the best practice when information technology is limited to paper, it is not a definition of the archival bond, or even a requirement for its existence. Hyperlinks are logical, rather than physical bonds. To limit the expression of the archival bond to practices that functionally replicate physical filing would risk arbitrarily truncating or even distorting the evidence provided by records.

Evidential value derives essentially from the relationship between the record and the activities in which it was produced and used. The question is not how are documents connected, but whether they are made or received in the course of an activity and whether they should be retained as evidence of that activity. – One might say that Web documents which only acquire a specific form or contents as a result of real-time choices made by the reader cannot be records because they are evanescent. But one could also simply say that such documents are made in the course of an activity. The question then is whether they should be preserved.

It is the natural accumulation of documents in the course of activity that makes them records. It is the organic connection between the records and the activities that gives them evidential value. Not the use of some particular method for expressing these connections in a durable manner. Records only have value in the context of an archival fonds. In the ideal, an archival fonds is integrated with the organization of these activities. As Robert-Henri Bautier of the Ecole Nationale des Chartes of France said 30 years ago, “An archival fonds is in effect the ensemble of all those items of any nature that any administrative body, any physical or juridical person has automatically and organically assembled for reasons of its own functions or activity.” (Bautier p. 22-23) The archival fonds is a comprehensive and organic unity which intentionally reflects the activities and procedures of its creator. (Nahuet 1998, p 99)

The traditional hierarchical arrangement of records is very well suited to the traditional hierarchical organization of an institution, itself a classic technique which facilitates and controls communications in the course of activities. Not only did the structure of the archival fonds reflect this hierarchy, but ordinarily the organizational units were collocated with their documentary resources. But in the environment of digital networks institutional hierarchy loses its function of communication. There is no longer any need to place information resources near the organizations which use them.

Individuals profit from the capabilities that PCs and digital networks give them to become more autonomous. The information they need to carry out activities can exist outside of the organization as well as within systems or collections that are in its custody. Paul Lasewicz describes the situation in private sector organizations:

“Unlike their paper-pushing counterparts from as late as the 1980s, the employees of the 21st-century corporation will be characterized by an ability to access the information they need, instantaneously, at their desktops. Instead of working through layers of corporate structure to obtain answers, future ‘knowledge workers’ will dive into electronic channels and grab the internal and external information they need in a few keystrokes. That world is here today. In the corporate environment the emergence of Intranets and ‘push’ technologies like Web-mounted news broadcasts have revolutionized the information environment.” (Lasewicz 1997, p. 63)

The changes go beyond individual levels. There are information systems that serve the needs of several organizations. In some cases, it may be possible to extract from such systems components and records which only serve the interests of a single organization, but in many cases the components are inextricable and the records may have primary value for several of the system’s clients.

The blurring of the limits of the body of documents readily available and used in the conduct of business, and the changes in the links that are created within this body and between the records and related activities and actors do not result only from technologically enabled possibilities. Long before computers starting showing up on the desktop, or in the palms of our hands, basic changes in the way business is conducted started weakening the close correspondence between an activity and the documents that are its by-products and

instruments. 20 years ago, Gerard and Christine Naud, of the National Archives of France, noted that in modern organizations,

“An activity is carried out simultaneously through several channels which diverge, then converge, each individual service, office or official being charged with only a part of the total process.

Given this, that which previously in archives could be designated as ‘the nature of the act’ today should often be characterized in a way that is at once both more supple and more precise. Most often, the notion of the ‘nature of the record’ only expresses its external form, rarely its reason for being.” (Naud, 1981, 218)

The weakening of the physical link and of the structural parallel between organizations and the records which they create and use does not invalidate the concept of the archival fonds. As Robert Nahuet has said, “Far from being a structure or an inert and simple skeleton, the archival fonds is the reflection of a network of communication and cooperation.” (Nahuet 1998, p. 100) At the threshold of the 21st century, structures are becoming more fluid; the network of communication, liberated from hierarchical chains, is extending every more widely; and networks of cooperation increasingly ignore organizational boundaries.

Such changes clearly have important impacts on the meaning of the records, the evidence and information they provide. The values of records derive from the organic relationships they have with the activities of their creators. “For the archivist, the context is that reality which gives meaning to the contents of the records and enables them to fulfill their function of proof and of evidence.” (Couture 1998, p. 15) To change the context of records is to change their meaning, their function of evidence, and with it the information that can validly be derived from the records.

Information technology changes the context of records, acting as a catalyst for changes in organizations, their procedures, and their relationships with persons and other organizations. In 1993 the Office of Technology Assessment of the Congress declared,

“Most Federal agencies now perform many key activities ... that could not be accomplished with paper systems.... Agencies such as the Internal Revenue Service (IRS), Social Security Administration (SSA), Bureau of the Census, and National Aeronautics and Space Administration would literally collapse without information technology.” (U.S. Congress 1993, p. 4-5)

From the start, the Clinton Administration has encouraged the use of information systems and of the Internet “to design a customer-driven electronic government that operates in ways that, 10 years ago, the most visionary planner could not have imagined.” (Gore, 1993, p. 113) At the end of 1999, Mr. Clinton reinforced this policy, charging agencies to undertake several initiatives to “use information technology to improve our society.” In this instruction, he described the important changes occurring thanks to this technology:

“The Internet and other information and communications technologies are changing the way we work, learn, communicate with each other and do business. These technologies are shaping our economy and our society in the same way that the steam engine and electricity defined the Industrial Age.... The Internet has the potential to enhance civil society as well as to boost commerce. Used creatively, the Internet and information technology can be a powerful tool for tackling some of our toughest social challenges as well as fostering economic growth....” (Clinton 1999)

The Internet puts citizens closer to the detailed activities of government and permits lighter, more supple structuring within government. Such changes are not confined to the U.S.

Similar trends are apparent in the private sector. (Lasewicz 1997) Today, business is conducted through digital technologies, in many important cases, such as those of worldwide financial transactions, in millions of daily transactions with no human intervention.

We have moved beyond the situation where people create records that happen to be electronic form. Information technology is no longer a tool; it is an infrastructure which is increasingly essential. The technology enables profound changes in the way business is conducted and in the roles that people and information systems play in these affairs. We must go beyond the evidence that records provide of individual transactions to discover evidence of how the technology is changing the very possibilities for conducting transactions. And we need to be open to further changes that, at best, we can only dimly foresee.

The wealth of the data being collected, the soundness of the methods being applied, and the richness and depth of the multi-disciplinary, international expertise brought to bear in the InterPARES case studies constitute a tremendous opportunity both to find ways to document such epochal changes in infrastructures and superstructures in the records we preserve, and to explore the implications of such changes on the nature of records, the structure of archival fonds, and the evidential and informational value of archival materials.