The Primary Source

Volume 23 | Issue 2 Article 1

2001

Individual Privacy, Institutional Accountability: The Challenge of Electronic Records

Patricia Galloway
University of Texas at Austin

Individual Privacy, Institutional Accountability: The Challenge of Electronic Records

by Dr. Patricia Galloway, The University of Texas at Austin

In the news:

Sale by businesses of customer databases led to demands for businesses to respect and protect individual privacy; banks have recently been ordered to notify customers about what data they collect and what they do with it, and to ask permission for same.

AP reported in 2000 that White House electronic document searches in response to various subpoenas were faulty because "some [email] message traffic from several computer systems was not stored in electronic archives. The previous administration required a lawsuit backed by the National Archives, American Historical Association, and Society of American Archivists before it handed over properly-scheduled electronic records to the National Archives, and then did so in the form of hundreds of hard drives removed from office computers.

Individual Privacy

Lawrence Lessig, a leading practitioner of so-called "cyberlaw," published in 1999 a very successful book: *Code and Other Laws of Cyberspace*. In it he argued in favor of an Internet cyberspace characterized by free speech almost everywhere, and pointed out that since *computer* code (programs) actually creates the architecture of the Internet, becoming in essence its *legal* code, then we should be concerned to be sure that unlegislated computer code does not create an Internet that violates Constitutional freedoms. Lessig maintains that open-source software that can be publicly examined and modified can help ensure that ordinary citizens have the power to protect their own privacy, aided also by alteration of laws limiting encryption algorithms. He believes that this individual power is necessary because the ease of regulating the virtual world of the Internet has made surveillance modes of undreamed-of intrusiveness inexpensive and doable. Or in Lessig's terms, the transparency of the Internet has made individual expression *perfectly regulable*.[1]

Like most private citizens I share Lessig's concerns for individual privacy; like most programmers I am only too aware of the ease with which my every online keystroke can be captured and made a valuable item of commerce without my permission. Yet millions of people are willing to trade some of their privacy away for convenience, and even I am reluctantly willing to bear with Amazon's assumption that I care what its database thinks I want to read. If pressed, most people would probably respond that their buying habits probably don't have much of a life in merchants' databases and probably fall victim to bit-rot as fashions shift. Perhaps they are right, although techniques of household profiling continue to proliferate, and there is nothing to prevent subpoena of a grocery store's electronic transaction records detailing a parent's beer-purchase habits for a child-custody lawsuit. So far really intrusive cases have not made it to public consciousness, but public response to the issue of identity theft enabled by electronic recordkeeping has made it clear that the value of personal privacy is one the public wishes to defend, especially privacy against government surveillance.[2]

---

[1] This essay was originally drafted in late fall of 2000, when there was no idea that there would soon be powerful arguments being made that intrusive surveillance modes be put in place; such legislation does not explain who is going to look at all that Internet traffic, because apparently there are systems in place that can analyze it automatically—and may already be doing so.

[2] In the wake of September 11, 2001, a poll revealed that 55% of those surveyed thought it was just fine for all the activity on the Internet to be surveiled.

Institutional Accountability

On the other hand, while we certainly don't want Big Brother looking over our shoulders, we do want to be able to regulate the actions of our own government, and have wished to do so since our country's beginning. Citizen oversight has been an issue for a long time and has increased in intensity since the end first of World War II and especially since the end of the Cold War. The 1966 Freedom of Information Act ushered in a host of state laws promising to give the citizen access to "public information" via the opening up of the actions and records of government to the "sunshine" of public scrutiny. In recent political campaigns the litany of public control of government has become a universal theme.

There is no arguing that government is most clearly accountable to the citizen through the records it makes of its actions: in fact, the memorializing of those actions in records is often required not only to document them, but to assure a citizen right. The huge usage of freedom of information and open records requests, not only by journalists but by citizens themselves, has underlined the importance of the records of governments and access to them. This same demand is beginning to be expressed with respect to electronic records as citizens become more conversant with computer technology, such that citizens are no longer satisfied to receive printouts of records that government keeps in searchable electronic form. The task of records managers and archivists who work with government records is to determine how they may give not only the same service to government administrators and citizens that they have in the past, but new services that the public well knows are enabled by electronic records.

The Email Classification Project

In deciding how Texan cyberspace will be regulated and what kind of electronic records will be retained for administrative and archival purposes, we have only the guidance of existing law and the ethical and practical standards of the archival discipline, but between them these two bodies of practice leave several questions unanswered. To make a start in achieving practical solutions to these questions, I am collaborating on a project to examine formal retention and archival practice with respect to email in the Texas Railroad Commission. My partners in this project are Susan Cisco, graduate of and adjunct professor of records management at the Graduate School of Library and Information Science at the University of Texas-Austin (GSLIS) and records manager for the Texas Railroad Commission; Mary Dee Harris, adjunct professor in the UT Department of Computer Science; and Martha Richardson, who has also studied and taught at GSLIS and is now an assistant to the director of the Texas Department of Information Resources. Email in a state agency, with its private feel and public function, is a particularly relevant electronic genre for attacking many questions pertaining to the privacy/accountability conundrum. And it is also especially apt because even at the national level, there is as yet no acceptable "best practice" for the handling of email that falls under the definition of a public record.

The Law

First we must ask what the law provides. Texas' public records law is not much different from that of other states including Mississippi--they all emerged during the sixties and seventies as part of a movement to open up government to public scrutiny. The law specifies that public records include:

[Government Code, Chapter 441 (Texas State Library and Archives Commission)]

441.006 [general powers and duties of TSLAC]: "...state records and other historical resources that document the history and culture of Texas as a province, colony, republic, or state."

441.031 [on records management division of TSLAC]: "'State record' means a document, book, paper, photograph, sound recording, or other material, regardless of physical form or characteristic, made or received by a state department or institution according to law or in connection with the transaction of official state business."

441.181 "'State record' means any written, photographic, machine-readable, or other recorded information created or received by or on behalf of a state agency or an elected state official that documents activities in the conduct of state business or use of public resources."

Also like most other states, Texas exempts from public access (but not necessarily from administrative or archival retention) certain records it keeps about individuals, such as medical records and student records, and others that it keeps about businesses, such as trade secrets. Thus apart from the few restrictions that may apply to public access to retained email records, there appears to be no legal restriction against the normal archival treatment of all email that flows on Texas government servers, and much legal support that favors it.

Electronic Mail in Texas State Government

In Texas as in other states, electronic mail is fast taking the place of many other kinds of communication, including paper-mediated letters and memos and telephonically-mediated telephone calls and voice mail. In fact, government offices are beginning to record a noticeable saving in long-distance costs, even without any explicit directives to replace telephone by email, just because of the better certainty of reaching one's correspondent by email and the reduction in work interruption in attempting to do so, especially across time zones. So there is no doubt that if we simply ignore email as an intractable problem, we will begin to lose a greater and greater part of the documentation of what our government does.[3]

Saving the Mail

But how to go about saving email is a non-trivial decision. The first problem would seem fairly obvious: where should we get it? Several studies have shown that email usage is so personal and ill-regulated in most workplaces that consistency of practice in saving email that qualifies as a record is a problem if we depend upon the individual worker to file and save it. The decision can as easily be made to capture email at the server level, before the recipient receives incoming mail and after the sender has sent outgoing mail. But if we choose to act at the server level, we shall have the harder problem of distinguishing between that which is a public record and that which is not.

Already at this stage we have the distinction between individual privacy and public accountability, in that we have to make decisions about what kind of privacy we will accord to incoming email. The law allows us to consider incoming email an action in a public forum, although it is clear that most members of the public do not so consider it, and it is also clear that email communications that touch on privacy-protected subjects present particular problems. Email that constitutes an official act, on the other hand, is all sent mail; yet there is sent mail that does not constitute an official act (baby shower invitations, for example). What all these considerations mean is that if we are to save email efficiently and consistently, we are going to have to recruit the computer to help us classify it.

Classifying the Mail

All email has basic metadata attached to it that indicates sender, recipient, date, and subject line. It would be nice if the subject line succinctly indicated the topic treated in the email, but that is not its purpose--the purpose of the subject line is to link conversations together by topic, since most email client software automatically repeats the incoming subject line when the recipient chooses to reply to a message, and many users simply allow the default topic to be repeated regardless of the topic treated in the reply.

We have accordingly found that if we are to provide subject access to email messages, we will need to process the content of the email message body in order to generate a secondary subject line. We have therefore experimented with content analysis on a corpus of email messages in order to construct a set of subject headings pertinent to the universe of discourse of that group of messages. In

---

[3] It seems that September 11 and the anthrax scare that has followed has also had the effect of increasing email traffic as it decreases faith in physical mail.

this part of the project we have sought the collaboration of Dr. Mary Dee Harris, a specialist in natural language processing, who experimented with this problem in one of her Computer Science classes in order to help us clarify the task. We expected to find that a limited number of semantic clusters would be defined by word collocations, and that indeed was the result of the experiment. Our next task is to decide how we can construct a thesaurus to help identify content and how we can combine content knowledge with known relationships between job assignments and records series in order to distinguish between official and unofficial communications.

Retaining and Archiving the Mail

As I have said, email is tending to replace in whole or in part paper correspondence or memos and telephone messages. Common archival practice to date has customarily ignored telephone messages and most memos, and has concentrated on correspondence created in the name of executive officials, on the rationale that only such documents have the legal power to effect action, which is the most significant part of what the archival public record seeks to document.

But electronically-mediated communications, beginning with the telephone, have drastically altered administrative structures, such that in the past twenty years we have seen hierarchical bureaucratic structures flattened and lateral communications across hierarchies proliferate. In other words, new media have enabled change in structure, yet archival practice has remained focused upon classical hierarchical bureaucracies. So our task in dealing with email will also include the necessity for recognizing its structural importance within organizations as well as its importance in crossing the boundary between the organization and its customers. For that reason we see the necessity for developing means of evaluating in-house email functions as indices of internal communication and efficiency.

Finally, we must be able to guarantee the "reliability" of the email we capture, which means that we need to be able to be sure that the agency user of email can be reliably identified. This task will require the careful collection of system administration records and the institution of passwording practices at each workstation, which we intend to test.

Providing Access to the Mail

The purpose of classifying and capturing email that can serve as a record of the activities of government is to provide access to it, for both administrative purposes and for citizen research. As I pointed out earlier, citizens are interested in having access to government records, and all evidence points to their desire to make that access more and more prompt. One of the astonishing features of any electronic records is that there is no practical reason why the citizen cannot have access to them virtually as they happen; if we feel that that is inappropriate, it becomes necessary to articulate why.

This is one of the aspects of the electronic record that interests me most: why and how do we choose *not* to provide instant access? What and for whom is the value of there being a delay between the issuance of an official record and public access to it? In the "float" between issuance and access there is a power differential of the same kind that is enjoyed by financiers engaging in "insider trading"; yet most of our conventions of negotiation in almost every political and diplomatic venue depend upon the ability to protect just such a power/knowledge differential. Studying email as electronic records permits us also to address this issue of "deliberative delay" and why and how we may need to understand it well enough to define the legal parameters to protect it explicitly. Doing so, we think, may mean the difference between preserving such records and withdrawing them entirely from public access.[4]

This essay merely scratches the surface of the issues we must tackle in learning how to deal with this significant category of electronic records. But as yet these problems remain unsolved and largely even unaddressed apart from the commercial application of automatic routing of incoming email to the

[4] George W. Bush's Executive Order 13233, which drastically curtails the access provided for under the Presidential Records Act, has recently foregrounded this problem.

appropriate answerer, so our field of opportunity is wide open and we look forward to making an important contribution to the management of public records in Texas, even if our work has no wider significance.

In August 2000, Pat Galloway left the Mississippi Department of Archives and History after twenty years to teach electronic records archivy at the Graduate School of Library and Information Science at the University of Texas-Austin. She can be contacted at galloway@gslis.utexas.edu