Gartner

How to Address the Federal Government's Records Management Challenges

Debra Logan, Kenneth Chin

This note documents the systemic failure of the U.S. federal government to understand the nature of electronic record keeping. There is a way forward, but first we must change attitudes and perceptions.

Key Findings

- According to its own figures, only 1% to 3% of all documents turned over to the U.S. National Archives and Records Administration (NARA) are important enough to be kept forever (see Note 1).

- User-driven classification of records, especially of e-mail, has failed and will fail in virtually all cases.

- The technology missing when implementing records management (RM) is automated policy management.

- Business and IT must work together to create policy and automate it.

Recommendations

- Give up user-based classification. Embed classification information in document templates. Treat e-mail as a "separate issue" and default to the category of ephemera or require all e-mail to be sealed for a period of five years and retained for ten unless otherwise directed by law.

- Aggressively manage e-mail communications as a part of business processes, moving critical documents, containing vital business information, to RM and archiving systems. Document decisions there — not in e-mail.

- Deploy archiving solutions to decrease the amount of storage space and increase the probabilities of ever finding a particular piece of information again.

- Enforce all RM policies with automation and continuous auditing processes, not moral exhortations.

WHAT YOU NEED TO KNOW

Many federal government agencies are failing to manage their business information properly. Our research shows that the U.S. federal government has this in common with other national governments around the world. As the amount of electronic information has exploded, governments and businesses have lost control of it. Proper management of business information or governmental records requires aggressive culling. No one pays attention to drafts, duplicates, outdated files or any other category of useless information. Lulled into complacency by inexpensive storage and fully aware of our own shortcomings when it comes to the fallibility of human memory, government organizations have allowed individual bits and bytes to accumulate to alarming levels. We are awash in a sea of digital waste that is, even now, preventing us properly preserving and managing vital information.

Groups, such as the Citizens for Ethical Responsibility, that would like to see more transparency in government pressure legislators and executive managers to keep as much digital information as they can, especially when it comes to e-mail. Most of these hundreds of millions of petabytes contain nothing particularly interesting or significant. Through the end of 2009, cost cutting imperatives will force federal agencies to take action on the vast amounts of unclassified unstructured information that they retain. Desperation to reduce costs when faced with this mountain of digital detritus will result in millions of classification errors.

People inside and outside government believe that the truth is out there in the electronic record. That is misguided. The sheer volume of what is being kept prevents federal agencies from finding what they need for routine business, let alone what investigators or litigants think they might discover about how decisions were made in any given instance. A more rational approach, grounded in the disciplines of archival and RM practice is needed. Otherwise, we will not only lose what is completely irrelevant (most e-mail) but we will also lose some of what we really need. After intellectual and policy decisions are made, solutions must be automated.

This advice is targeted at federal agency heads, government policymakers, legislators, lawyers and judges who have the will and the vision to shape policy and practice. On the implementation level, it is targeted at records managers, information managers and CIOs who now have the burden of managing this task.

STRATEGIC PLANNING ASSUMPTION(S)

Through the end of 2009, cost cutting imperatives will force federal agencies to take action on the vast amounts of unclassified unstructured information that they retain.

ANALYSIS

The Extent of the Problem is Infinite

In the Richmond Journal of Law and Technology, Paul and Baron (see Note 2) characterize the information inflation problems that confront the legal system. The article cites many sources for the extent of the problem, the amount of information available, how fast it is increasing and the extent to which the interaction of people and information, via information technology, are giving rise to changes in civilization. The article deals primarily with challenges to the legal system. It is also illustrative of the challenges facing governments, the creators and enforcers of that system. Many points of the article are relevant to the discussion of government record keeping but here are a few of the most salient.

Gartner

- The information ecology that is emerging from the interaction of people, technology and techniques are giving rise to a worldwide change in the nature of civilization itself, which might be called "The Age of Information Complexity."

- Language itself is a living thing, subject to constant change and meaningful only in so far as it can be embedded in its context, which is historical, national, and personal, just to name a few of the overlapping mileu in which it exists and which must be understood to be interpreted. Any given human being can use language in a way that has never been used before — its possibilities are infinite.

- No form of present day (or future) search technology is adequate to find "all" or "nearly all" documents in a large collection and no form of categorization expertise or technology is sufficient to group and label individual objects in a way that will satisfy all information retrieval needs, either present or future.

Language gives rise to these conditions and to the characteristics of the information artifacts created with it. Finite computing systems cannot solve problems for which there are no boundaries. Syntax is what computers are good at. Semantics is what people are good at. The search for meaning and "truth," either as a computing problem, or an existential exercise in life, cannot be reduced to coded formulas. The search engine that can do this is not "just a few years away."

Stop waiting for the magic search (or other) technology to solve the problem of record keeping in government. It is, essentially, not a problem that technology will ever solve and the more we allow "information" to accumulate the more difficult the problem becomes. Focus on the part of the universe of information that is most important. We will never know for sure if we got the criteria for "importance" correct. There will be things, information artifacts and yes, truths, that will be lost forever. This is the way it has always been. Arguments about the interpretation of history will continue.

There are at least two reasonable possibilities. Declare all e-mail by its nature as ephemera and thus subject to a short retention period. Rapidly move it to offline storage and then to destruction (see Note 3). The other option would be to declare working documents of an administration or agency to be sealed communication for a period of four or five years. Some hindsight will add perspective to what we reasonably should be keeping and what we should not. Role-based retention should be practiced in any case.

This does not mean to say that people should not be investigated for suspected wrongdoing and prosecuted if a case is made. Proper attention to information retention is important. However, government's current approach to record keeping and information retention management makes justice less likely, not more. What the current approach does is cost money, waste time and inevitably make it more difficult to find the data to decide on matters of national, judicial and historical importance as well as to settle practical questions with a bearing on the day-to-day services government delivers to citizens.

Its time to decide what is rational, practical and doable. We will make mistakes. They will be better mistakes than we are currently making.

Why and How "Users" Will Not Solve the Problems of Records Management and E-Mail Archiving

The gold standard for retention decisions is contextually based human judgment. "Business users" in the course of a "business process" are best qualified to tag individual items for long or short term retention or deletion, but wait. What about the legal requirements? Surely lawyers and compliance officers must have a say. What if the artifact has relevance in multiple jurisdictions?

Gartner

What are the individual state and local government requirements? What if I need to prove something to my boss? In practice "user-based classification" fails for all these reasons and more. People are doing their best to carry out their responsibilities. Many do not feel qualified to determine what should be filed, where or for how long. The default option is to keep everything and hope that there will be someone else to sort it out down the line. It has simply got to stop. The federal government has itself acknowledged this. In a study by the GAO, it is stated that "the quality of record keeping varied considerably across agencies." The study concluded that, for many federal employees, the concept of a "record" and what should be scheduled and preserved was not clear (see Note 4).

There is an insurmountable problem that lies at the heart of the Federal Records Act: Individual employees must use their best judgment as to what constitutes a record. Sometimes that is clear and there is no difficulty in doing it. Newer forms of electronic information make it much more difficult if not impossible to make that judgment for each individual iteration of a document or each individual e-mail that crosses the desk of the average federal employee. The speed and volume of electronic communications mean that what we are asking people to do what cannot be done *by people.* Technology has created a problem that only technology can solve. Technology will not solve the problems that only humans can solve, namely the rational determination of what needs to be preserved and what does not. Only by a combination of human rationality and machine efficiency can the federal government fix its information management problem.

The Federal Records Act and other legislation not only require agencies to keep records, it also charges them to do so efficiently and economically. The Federal Records Program also requires the "simplification of the activities, systems, and processes of records creation, maintenance, and use." We must turn to technology, as it exists today, not as we hope it will be at some unspecified point in the future, as "good enough" to do some of this work for us.

What Technology Exists Today and What Agencies Should Be Using

The first priority for every federal government agency should be to get its e-mail house in order. As stated in The Code of Federal Regulations, Subchapter B, Part 1222.34 Identifying Federal Records, e-mail **may** be a record. But when is it and when is it not? Individual users should not be burdened with the task of identifying each and every e-mail as record or non-record. There is not enough time in the day. Assuming a volume of between 75 and 100 e-mails sent and 75 and 100 e-mails received by an average user in a day, further assume that it takes less than 30 seconds to classify each one. That is a total of 100 minutes per day classifying e-mail. Ask yourself, as a taxpayer, is this what we want federal government employees to spend their time on? No. This classification must be done by technology.

In addition to the time it takes to sort e-mail into records and non-records, e-mail is also a huge driver of incremental storage costs. Almost everyone in government and business relies on it as a defacto task management and document management system. People save years worth (and even now into decades worth) of the stuff. E-mail is a personal resource — not a corporate one. No one else but IT can look at it, except under extraordinary circumstances, such as when it is subpoenaed. Lawyers love to read e-mail. Why? Because it tells a story and provides a narrative for events on a daily basis. That is not a reason to keep it, but it does come in handy when investigators are trying to piece together, at a great deal of time and cost, events that are alleged to have happened. Yet e-mail is not the only way — or even the best way — to do that. It is the technology we love to hate and we all rely on it our day-to-day work. Here is what agencies should do to bring the e-mail problem under control, so that they can focus on value creating information management systems:

- If you do not already have one, create an e-mail retention policy document. Use the following Gartner research: "Toolkit Presentation: Creating an E-Mail Policy Document,"

Gartner

"Companies Differ on E-Mail Retention Strategies" and "How to Formulate an E-Mail Retention Strategy."

- Consider an e-mail or content archiving system that allows the retention, disposition and litigation management of all e-mail. The system should allow the archiving of other types of content as well. For specifications see "The Emergence of Integrated Content Archiving Requires an Enterprise Archiving Strategy."

- Eliminate the storage of personal e-mail archive files (.psts and .xxx Notes Equivalent) on common information systems, including desktops. Buy software that finds them and moves them or outright eliminates them. You do not even need to buy anything: Exchange allows you to do this and so do many e-mail archiving systems, if you already have it.

- Begin to socialize the idea of alternative solutions for doing the work that e-mail was meant to do, rather than making it the centerpiece of an information worker's IT toolkit. These include instant messaging for truly ephemeral conversations, collaboration tools such as wikis and blogs for internal use and discussion of issues, and more formal document management systems to create and store documents that have business value and retention rules.

- Create retention policies, which automates the rules, that make the best judgment about the status of the e-mail (record or non-record), and assign it to a retention folder, automatically. Use the functionality of the e-mail archiving system to periodically purge e-mail that is past its retention date.

- Formulate a policy for personal e-mail which includes allowing users to access their personal e-mail accounts from the office if that is acceptable under the charter of the agency. If they wish to keep personal e-mail there forever, they can. Remember that individual e-mail messages, if important enough, can be made into proper files and stored on the desktop or common storage devices.

Until and unless you have solved your e-mail challenges, e-mail should be your first priority. You will be surprised at how much easier it is for everyone to sleep after you do so.

The same method — automated policy management — can be used for different types of documents and files, especially Microsoft Office files, which are resident on servers. Following the clean up of drafts, duplicates, out-dated files and other non-record information, your agency can begin to concentrate on the document creation and review process, to prevent this problem from happening all over again. File archiving solutions can help with this.

RECOMMENDED READING

"The Emergence of Integrated Content Archiving Requires an Enterprise Archiving Strategy"

"Selecting Which Government Information Management Issues to Tackle"

"Obama Takes Quick Steps to Ensure Open and Transparent Government"

Note 1
Government Archiving

http://www.archives.gov/about/

Gartner

Of all documents and materials created in the course of business conducted by the U.S. federal government, only 1% to 3% are so important for legal or historical reasons that they are kept by us forever.

Note 2
Can the Legal System Adapt?

GovernCite as: George L. Paul and Jason R. Baron, "Information Inflation: Can the Legal System Adapt?," Volume 13, Richmond Journal of Law. and Technology (2007).

http://law.richmond.edu/jolt/v13i3/article10.pdf

Note 3
Librarians, Archivists and Ephemera

The category of ephemera is a difficult one and has been defined by librarians and archivists mostly by example. Libraries and archives do collect ephemera, but it is material that has survived by accident, not design. When it comes to e-mail, it is not the e-mail itself but the issues that are discussed that are ephemeral or of interest only to the parties involved and then only for a short period of time. The classic "Let's go to the bar" and "There's birthday cake in the lunchroom" sent to the company distribution list are examples. The problem is that it is difficult to separate these issues at the time of creation by any automated means. Yet, to preserve what matters and to be able to access e-mail as evidence or have any hope of classifying it, it is clear that at this point, we have to take radical action. For a librarian's view of ephemera and the difficulties associated with it, see Timothy G. Young, Evidence: Toward a Library Definition of Ephemera. See
http://www.ala.org/ala/mgrps/divs/acrl/publications/rbm/backissuesvol4no1/Young.pdf

Note 4
Testimony Details

Testimony before the Subcommittee on Information Policy, Census, and National Archives, House Committee on Oversight and Government Reform FEDERAL RECORDS For Release on Delivery Expected at 2 p.m. EDT Wednesday, April 23, 2008. Agencies Face Challenges in Managing E-Mail Statement of Linda Koontz, Director Information Management Issues.

Gartner

REGIONAL HEADQUARTERS

Corporate Headquarters
56 Top Gallant Road
Stamford, CT 06902-7700
U.S.A.
+1 203 964 0096

European Headquarters
Tamesis
The Glanty
Egham
Surrey, TW20 9AW
UNITED KINGDOM
+44 1784 431611

Asia/Pacific Headquarters
Gartner Australasia Pty. Ltd.
Level 9, 141 Walker Street
North Sydney
New South Wales 2060
AUSTRALIA
+61 2 9459 4600

Japan Headquarters
Gartner Japan Ltd.
Aobadai Hills, 6F
7-7, Aobadai, 4-chome
Meguro-ku, Tokyo 153-0042
JAPAN
+81 3 3481 3670

Latin America Headquarters
Gartner do Brazil
Av. das Nações Unidas, 12551
9° andar—World Trade Center
04578-903—São Paulo SP
BRAZIL
+55 11 3443 1509

Gartner