Response by the Society of American Archivists to NARA Advanced Notice of Proposed Rulemaking

Prepared for SAA by its Electronic Records Section and its Government Records Section, and Endorsed by SAA Council

4 January 2002

Summary

Retaining the content, structure and context of records is a necessary component of responsible recordkeeping. When transferring records from one technical environment to another, one can potentially lose essential components of that content, context and structure. Appropriate management, retention and disposition of electronic records requires government agencies -- with guidance and legal authority from the National Archives and Records Administration -- to address the necessary elements and properties of the particular forms and formats of those records. We believe that Public Citizen's call for the retention of "the **entire** content, structure and context of the electronic original" is neither realistic nor helpful as a goal of good recordkeeping. Instead, agency policies, procedures and documentation -- including NARA-approved retention schedules, published disposition manuals, special transfer requirements, transfer forms, and other documentation -- should indicate the specific qualities of digital objects that must be retained in order to ensure their ongoing value as records.

While we support most of the requirements that NARA provides in 36 CFR 1234.22, we believe that they should be applied to all records produced or received with electronic information systems. We do not see "text documents" as a useful category for purposes of this regulation, since we do not believe that it is possible to identify any particular set of elements or properties that must be retained for all records that fall within its definition.

We also emphasize that retaining records in electronic formats will often strongly support the business needs of government, including provisions for public access. As Public Citizen argues, retention of records in electronic formats will often greatly promote both the convenience and quality of public access to those records. Recent case law provides a number of examples in which parties have been required to provide the electronic version of records as part of the evidence discovery process. While this does not imply that all records must be retained in electronic formats for their entire life, it does provide compelling support for the view that electronic source records can often contain important elements of evidence that may be lost in printed copies.

NARA's mission is ensuring "ready access to essential evidence." This is a primary objective of the archival profession as a whole. It is also a mission in which the federal government, citizens, and other potential researchers all have an interest. EFOIA represents an important set of business needs related to such access. Recent legislation

such as the Government Paperwork Elimination Act and Electronic Signatures in Global and National Commerce Act has also introduced a number of requirements to manage and provide access to government records in electronic formats. Of course, agencies must continually strike a balance among numerous business needs. This balance is recognized in the FOIA legislation, which states that each "agency shall make reasonable efforts to maintain its records in forms and formats" that can be both searched and reproduced electronically. We encourage NARA to continue expanding its efforts to promote the management and preservation of electronic records within the federal government.

Introduction

In response to the request for comments on proposed rulemaking published in the Federal Register on 10 October 2001 and identified as 3095-AB05, the Society of American Archivists (SAA) submits the following document.[1] The Society of American Archivists (SAA) is the oldest and largest association of professional archivists in North America. Representing more than 3,000 individuals and 400 institutions, the SAA is the authoritative voice in the United States on issues that affect the identification, preservation, and use of historical records.

Reconciling existing laws, regulations, and principles with the new demands of records created and received by new information technologies is a great challenge facing the archival profession. The SAA commends NARA on its recent and continuing efforts to provide further guidance on the management and preservation of electronic records. NARA's response to the Public Citizen petition and its solicitation of comments on that petition is indicative of NARA's willingness to engage with the larger archival community in discussion of these important issues.

General Comments on Public Citizen's Petition

Public Citizen's petition[1] rests on the claim that it is important to preserve records in their original form. The professional literature of archival science supports the general principle that at times it is appropriate to maintain records in their original form. One of the most compelling reasons is that originals have qualities that may not be reflected in

---

[1] Two constituent units of the SAA, the Electronic Records Section and the Government Records Section, are responsible for most of the comments included here. The following individuals from the two sections contributed texts and comments during the process of drafting this document: Paul Bergeron, City Clerk, Nashua, New Hampshire; Jelain Chubb, Missouri State Archives; Christopher Frey, University of Michigan; Geof Huth, New York State Archives; Craig Kelso, Missouri State Archives; Joe Laframboise, Kansas State Historical Society; Cal Lee, University of Michigan; Scott Leonard, Kansas State Historical Society; Pat Michaelis, Kansas State Historical Society; Marry-Ellyn Strauser, Missouri State Archives; Matt Veatch, Kansas State Historical Society; and David Wallace, University of Michigan. The Electronic Records Section and the Government Records Section both include members who are employees of the National Archives and Records Administration. Those members have abstained from the process of drafting this document.

facsimiles. These qualities can potentially serve a variety of business, legal, audit, operational and research needs. The need to retain originals, however, must be weighed against institutional priorities, available resources, and the limits of technical feasibility. It is thus important to provide very specific answers to the question, "What qualities of an original are useful or necessary to retain in their original form?"[2]

For physical records, archivists have stipulated that records with "intrinsic value" are those records worth preserving in their original form. Intrinsic value was defined in a 1974 SAA glossary as embodying the "qualities and characteristics of permanently valuable records that make the records in their original physical form the only archivally acceptable form of the records."[3] In 1982, the U.S. National Archives (then NARS) issued a paper on intrinsic value, explaining that these "qualities or characteristics relate to the physical nature of the records, their prospective uses, and the information they contain."[4]

Although the "essential intangibility"[5] of electronic records makes it inappropriate to focus on their physical characteristics in the same way one would for paper documents, the question of which properties to preserve becomes even more important. As indicated by current NARA policies,[6] the long-term preservation of electronic records will generally require copying (also known as refreshing) and reformatting onto new physical storage media,[7] but it is important that one "not alter the character of the recorded information" in the process.[8] This requires one to preserve qualities at a higher level of abstraction than the physical medium. Though it is debatable whether or how the concept of an "original" is still relevant in this context,[9] there must still be some way of indicating the characteristics of an electronic source record that should be retained in any recordkeeping copies. This is supported by the revised definition of intrinsic value in the SAA Glossary from 1992, which refers to the "inherent worth of a document" rather than its original physical form.[10]

The identification of content, structure and context as essential characteristics of records is thus extremely important. If properly designed and managed, a recordkeeping system can retain these characteristics of records even though underlying hardware and software may change over time. As Public Citizen has pointed out, the need for a recordkeeping system to retain records content, structure and context for their required retention period was emphasized in both of the legal cases related to GRS 20. "In other words, as counsel for the Archivist put it at oral argument, if the information is part of a record under the RDA [Records Disposal Act], see § 3301, then it must be preserved."[11] This still leaves the difficult task of determining what specific pieces of information fit this description. We believe that Public Citizen's call for the retention of "the **entire** content, structure and context of the electronic original" [emphasis added] is not realistic. Instead, agency policies, procedures and documentation -- including NARA-approved retention schedules, published disposition manuals, "special transfer requirements agreed upon by NARA and the agency [to] be included in the disposition instructions,"[12] transfer forms, and other "documentation adequate to identify, service and interpret electronic records that have been designated for preservation by NARA"[13] -- should indicate the specific

qualities of digital objects that must be retained in order to ensure their ongoing value as records.

A related concept, introduced by the CEDARS project on digital preservation, is that of "significant properties." According to researchers on that project, "Whoever takes the decision that a particular digital object should be preserved will have to decide what properties are to be regarded as significant." Within the terminology of the Reference Model for an Open Archival Information System (OAIS), the "representation information" related to a digital object "must allow for the recreation of the significant properties of the original digital object."[14] Other recent work on digital preservation has reiterated and reinforced the important role of significant properties.[15]

For the management and preservation of federal records, this would imply the need to clarify the properties of "form or character" of electronic source records that either serve a sufficient business need to warrant preservation by the creating agency or have "sufficient administrative, legal, research, or other value to warrant their further preservation"[16] through transfer to the National Archives. It is worth reiterating that for electronic records, the properties of this form should be defined independent of underlying physical medium, whenever possible. This can allow for the creation of recordkeeping copies of an electronic record on various media, including new electronic platforms, and -- often but not always -- paper, microfilm or other analog media. Without the specification, capture, management, and verification of such properties, however, there is no guarantee that transfer across media will retain its value as a record of government business.

Public Citizen's Proposal 1

"The regulations should make explicit that recordkeeping systems that preserve electronic text documents must preserve the entire content, structure and context of the electronic original, a requirement that the Archivist's attorneys have stated is already part of GRS 20, although the text of GRS 20 contains no such language."

1A1. Is NARA's definition of electronic information system still adequate?

We would suggest the following wording:

> *Electronic information system.* A computer-based system that supports a set of one or more of the following functions: acquisition, creation, storage, processing, management and access to information.

This system will often include application programs, business applications, system software (including supporting databases and servers), desktop operating systems, network operating systems and servers. In order to clarify further the role of an electronic information system, we offer the following revised definitions:

> *Electronic recordkeeping system*. An electronic information system that supports the collection, organization and categorization of records to facilitate their preservation, retrieval, use and disposition.

This definition indicates that electronic recordkeeping systems are the subset of electronic information systems that provide recordkeeping functions. The system itself is electronic, but all of the records to which it is applied need not be electronic. If NARA would instead like the term "electronic recordkeeping system" to mean that subset of recordkeeping systems that manage electronic records, then we would suggest a wording that more explicitly states this relationship to specifically electronic records. The current 36 CFR 1220.14 definition, however, is ambiguous. A reader could also potentially interpret the wording of "system in which records..." to mean that an electronic recordkeeping system must be the electronic information system that physically stores electronic records. We do not believe that such a definition would be useful, since records management applications and other software that supports recordkeeping need not provide this function directly. The important point to emphasize about an electronic recordkeeping system is that provides important **functionality** that not all electronic information systems provide, regardless of what combination of hardware and software may support that functionality at any given point in time.

> *Recordkeeping system*. A system in which records are collected, organized, and categorized to facilitate their preservation, retrieval, use, and disposition. The system can contain manual and automated components.

This wording eliminates the "manual or automated" distinction of the existing 36 CFR 1220.14 definition, which is potentially misleading. As indicated by the Electronic Work Group Report, a system consists of "people, machines, and methods organized to accomplish a set of specific functions." Given these necessary components, a recordkeeping system thus will never be fully automated. Conversely, in a contemporary government office environment, even a recordkeeping system that manages records that all reside on paper media will still include numerous automated (or at least electronic) components -- indexes, forms, directories, etc. -- in order to support the collection, organization and categorization of those records.

Should it explicitly include (or exclude) any types of office applications or other types of software such as the network operating system?

An electronic information system will tend to include one or more applications, and it will always rely on some components of one or more operating systems. Though a longer narrative explanation of electronic information systems could provide examples of such components, it would be both unnecessary and inappropriate to delineate them within the definition itself. The specific functional roles played by the layers of supporting technology will vary across technical environments at any given time, and they will vary even more dramatically across time. In order to implement the functionality of an electronic information system one needs a large number of low-level

activities -- processing of machine instructions, process management, remote procedure calls, load balancing, memory management, file management, directory management, input-output, storage allocation, data typing, compression, encryption, error correction, interfaces in between various software processes, access control, data caching, data locking, journaling, processing of user commands, desktop and windowing, etc. -- which can be carried out by any number of software or hardware components. It would be a mistake for the definition to attempt a coherent, exhaustive delineation of such components.

Is the definition of "text documents" sufficiently broad enough to cover documents produced by products other than word processing software, e.g. PowerPoint presentations or desktop publishing files?

We believe that "text document" is not a useful category for purposes of 36 CFR 1234. One can target specific structural elements for retention in the case of both relational database tables (field names, field lengths, foreign keys, primary keys, etc.) and email messages (required MIME header fields, multipart message boundaries, etc.). This is because each has a fairly "strictly prescribed form and format."[17, 18] The same cannot be said for those digital objects designated under the current definition of "text documents." In fact, the definition itself indicates that they have a "loosely prescribed form and format." Allowing disposition of electronic source records based on this definition could lead to the ironic and problematic situation in which the electronic records whose properties are least formally understood (and thus most difficult to identify or verify during transformation) can be destroyed after being transferred to a new medium.

We would suggest instead that NARA change the name of 1234.22 to "Creation and use of electronic records" and move it to the beginning of Subpart C, since it would appear that its provisions are appropriate for all types of electronic records. The general description of electronic records in 1234.1 indicates that they "include numeric, graphic, and text information," all of which are possible components of files created through commercially available office applications. Given the tight integration of many suites of office applications, anyone applying the definition of text documents to all such files would seem unable to exclude interactive content, macros, images, tables, spreadsheets, animation, hyperlinks and various multimedia components without knowing more about the specific "form and character" of the specific category of records.

If disposition specific to text documents is provided, we would encourage clarification of the distinction between this definition of text documents and that of "textual documents" in 36 CFR 1228.270(d)(2). The latter provides for only ASCII files (which may contain SGML markup), obviously excluding a large number of current office documents without further transformation. In that same light, we would also encourage clarification of the distinction between "data files" as defined in 1234 and 1228. In 36 CFR 1228.270(e)(3), for example, "textual documents with SGML tags shall include a table for interpreting the tags, when appropriate." Since software-dependent database files are often likely to be translated into SGML (or, more likely, XML) for purposes of long-term preservation

(either within an agency or when transferred to the National Archives), further explanation of the role of electronic recordkeeping copies may be helpful.

Should the definition of "text documents" be amended to include presentation and other specific files?

We recommend avoiding the delineation of specific examples within definitions of terms, since the list can often be inappropriately perceived as either exclusive or exhaustive.

1A2. If we determine that the section should be amended to reflect Public Citizen's proposed requirements, would coverage of the section be clearer if the term "electronic information systems" is replaced in §1234.22 by a delineation of specific applications that may produce original electronic text documents such as office suite application packages (e.g., Office 2000, Lotus Notes), or word processing or other office automation applications not integrated with the agency email or office suite?

Once again, we recommend avoiding the delineation of specific examples within definitions of terms, since the list can often be inappropriately perceived as either exclusive or exhaustive.

1B1. Are the definitions of "content," "structure," and "context" contained in the NARA GPEA guidance adequate for all types of records?

Since all electronic records have content, structure and context, this definition provides further support for moving the current 1234.22 to the beginning of Subsection G and applying it to all records produced or received with electronic information systems, rather than simply text documents.

NARA defines "content" as "The information that a document is meant to convey. Words, phrases, numbers, or symbols comprising the actual text of the record that were produced by the record creator." We would suggest that the second sentence narrows the definition unnecessarily. By explicitly including the terms "words, phrases, number, or symbols," it implicitly excludes numerous other types of content (e.g. pictorial, aural). As with other items above, we would not recommend revising to attempt a fully inclusive list of possible content types. Rather, we would advise using only the first part of this definition, which is identical to the definition found in the current SAA Glossary.[19]

We find the definition of structure to be adequate.

The first part of the definition of context is consistent with our recommendation of a broad conception of a recordkeeping system: "The organizational, functional, and operational circumstances in which documents are created and/or received and used." It also emphasizes why we do not support Public Citizen's call for the preservation of the "entire content, structure and context" of records over time. In principle, the context of any given record is infinitely complex. Effective recordkeeping systems are necessary to

ensure the capture, management and preservation of an important portion of record context. The alternative is independent data and information elements that cannot adequately stand as evidence of government activities.

The second part of the definition of context emphasizes those components for which electronic systems are often very well suited: "The placement of records within a larger records classification system providing cross-references to other related records." Though both parts are important, we would suggest possibly combining them to form a much more compact definition:

> *Context*: The organizational, filing and classification systems in which documents are created, received and used.

This definition also eliminates "operational," which may be redundant. Though more succinct, we do recognize that this definition may be a bit less transparent and thus may not necessarily be preferable to the existing definition. We also believe that an explanation of context should include technical contextual elements (metadata, software documentation, etc.) as well as the business context. The current definition could be interpreted to include these elements, so it is not completely clear whether they should be added to the definition itself, though we would recommend that any supporting documentation contain some mention of them.

We would like to reemphasize that the retention of all context of electronic source records is not a desirable or realistic goal. Not all contextual information is of equal value. The organization of email within the inbox of its recipient is far less important than contextual information tied specifically to the filing and use of the message.

Do you agree with NARA's understanding of the terms "content," "structure," and "context" as they apply to text documents in the Federal Government? If not, what is your understanding of these terms?

Apart from a few minor suggestions provided above, we find the general definitions of content, structure and context to be adequate. We do take issue with the category of text documents. It is both appropriate and desirable for NARA to provide technical guidance on important properties of particular formats that are consistently applied within those formats (e.g. attachments to electronic mail messages), and it is further desirable to identify forms[20] of electronic documents that constitute specific records series with their own associated properties. Form of material (e.g. time sheet, birth certificate) will often indicate the appropriate retention period for a given record -- as codified through a NARA-approved retention schedule series or published agency disposition manual -- while the electronic format can often indicate specific properties that hold for all files created in that format (e.g. To and From fields in the MIME header of all electronic mail messages) and others that only hold on occasion (e.g. document property fields that are used for word processing documents on some occasions by not on others). The combination of the two can indicate not only what records should be retained, but also what properties should be retained in the creation of an electronic recordkeeping copy

from an electronic source record or the creation of an analog recordkeeping copy from an electronic source record. Primarily for reasons already stated, however, we do not believe that guidance on content, context and structure should be applied to text documents as a general category.

Do these concepts need to be defined in NARA regulations?

Yes. The concepts of content, structure and context are essential to understanding and managing records, especially electronic records. Their definitions should be included in 1234.2, and they also should be explicitly referenced in the section being addressed by NARA's ANPRM (current 1234.22) as recommended by Public Citizen.

1B2. What information about the content, structure, and context must be maintained as part of the record for the agency to conduct its business and for accountability purposes?

We believe that it is impossible to provide a specific answer to this question that would apply to all records. We can provide a few general base-line suggestions.

Content: For any essential data element of an electronic record[21] agencies should take appropriate measures to maintain the integrity of the bit/byte stream of that element if it is preserved electronically and ensure that the full contents of that data element get represented physically in some way if the recordkeeping copy is to be on an analog medium.

Structure: Agencies should preserve all the structure of a document that is necessary to make the content sufficiently unambiguous and meaningful as evidence.

Context: At the very least, agencies must create and preserve enough contextual documentation of a record to associate it with a given business function, activity and/or transaction.

Can we define the minimum metadata needed for text documents to provide adequate documentation, as we do for email messages (see 36 CFR 1234.24(a)(1)(a)(3))?

One could provide the minimum metadata to provide for adequate documentation of **all** records, but it is not possible to identify any specific set of additional metadata that should apply to text documents as a category.

Are the minimum metadata different for permanent and temporary records?

There will be additional metadata required for records that are preserved for long periods of time. This will include technical and administrative metadata, as well as documentation to help a user far into the future make better sense of the record. Most of this metadata will be added over time, however, and not at the point of creation. It is

likely that one will want to add some additional elements to a record that needs to be retained for longer than the expected life of the current hardware and software environment. This distinction, however, is purely based on technological requirements, rather than any division between temporary and permanent records based on appraised value.

Do specific types of text documents require different minimum metadata?

If type is taken to mean a different form or format of records, then it would be possible to specify different elements for different types.

What relationship do you see between "content, structure, and context" and metadata requirements?

The line between data and metadata is not absolute. The important thing is to identify necessary elements and properties. Depending on the implementation of an agency, these may be retained as part of digital objects themselves or as external elements related to those objects.

Specifically addressing the Public Citizen proposed CFR wording, does compliance with the metadata and other requirements in this proposed § 1234.22(a)(5) meet the requirements for content, structure and context in its proposed § 1234.22(a)(1)?

These requirements are meeting different sets of objectives, and thus should both be included. Content, structure and context are necessary to preserve an authentic record. Whether their constituent elements are considered data or metadata will often come down to agency implementation. The requirements of 1234.22(a)(5) may be met, at least in part, by many of the same elements retained for purposes of 1234.22(a)(1), but (a)(5) will tend to require additional descriptive and administrative elements specific to its purposes. Retrieval of an electronic record, for example, will generally require things such as an identifier that is unique to the system and some sort of browsing or querying capability. The data elements necessary to perform such functions will most likely not reside within the electronic source record, but must instead be added as part of the recordkeeping system. Metadata for enabling retrieval, protection and disposition is neither necessary nor sufficient for the preserving the content, structure and context of the record itself.

1B3. Hidden Information: NARA's view is that hidden information (such as comments) in text records must be preserved as part of the record when the author intends to share the information with others, e.g., notes added to explain or comment on a draft report.

According to 36 CFR 1222.34, Identifying Federal records:

(c) Working files and similar materials. Working files, such as preliminary drafts and rough notes, and other similar materials shall be maintained for purposes of adequate and proper documentation if:

(1) They were circulated or made available to employees, other than the creator, for official purposes such as approval, comment, action, recommendation, follow-up, or to communicate with agency staff about agency business; and

(2) They contain unique information, such as substantive annotations or comments included therein, that adds to a proper understanding of the agency's formulation and execution of basic policies, decisions, actions, or responsibilities.

For documents that should be retained based on the conditions stated above, retaining annotations and comments would be necessary. It would seem that the second condition is provided specifically to indicate cases beyond those in which the author specifically "intends to share" the contents of the annotations or comments with others. The next paragraph of that same section would seem to support the idea that an electronic record that is used to carry out a given activity can have record status even if a paper version also exists.

(d) Record status of copies. The determination as to whether a particular document is a record does not depend upon whether it contains unique information. Multiple copies of the same document and documents containing duplicative information, including messages created or received on electronic mail systems, may each have record status depending on how they are used to transact agency business.

Though NARA goes on to explain in (f)(2) that extra copies of documents that are preserved "solely for convenience of reference" should be considered nonrecord material, it is by no means obvious that all (or even the majority of) use of electronic documents for which a paper version exists actually fall into this "convenience copy" category.

According to NARA's Introduction to its General Records Schedules, agencies are instructed, "When it is difficult to decide whether files are record or nonrecord materials, the records officer should treat them as records."[22]

We support the Electronic Records Work Group's use of the term "electronic source record," since the guidance cited above and the definition of record at 44 USC 3301 would seem to imply that a large proportion of electronic documents created by federal agencies are records. This does not mean that all such records should have long retention periods, nor does it mean that they all contain elements (such as comments or annotations) that would not be reflected in a paper printout. We would argue, however, that it does imply a need to address such issues at a much finer level of granularity than all "text documents." It is incumbent upon agencies, with guidance from NARA, to take measures to minimize the risk of destroying essential evidence through the deletion of electronic source records.

Is it essential or even misleading to require it when the document is viewed/printed from a system that does not indicate that there is hidden text?

Data stored on electronic media requires hardware and software to be discovered, accessed, read and understood. It is thus all "hidden" from us in a way that ink on paper is not.[23] A specific view of an electronic record using a specific application, given a specific set of user preferences on a specific computer platform is thus often not the appropriate standard for determining what to count as the "contents" of that record, nor, often, is printing that same record out onto paper using that same application, set of preferences and computer platform. A significant proportion of electronic formats (including those created by most common office suite applications) support multiple "views" of the same document. For some formats, one may be able to determine a general canonical view that can be assumed to reflect all significant properties of the "record version" of the document, but for a large number of others, such a universal determination would be both arbitrary and likely to neglect essential evidence of government business.

What types of documents besides word processing have hidden comments/text capability, e.g., spreadsheets with formulas?

Most electronic document formats include numerous forms of "hidden" content, as it is understood in the question above. Spreadsheets are no exception.[24] They can include formulae, macros, notes of various kinds, embedded complex objects, links to external data sources and often reflect dependencies between individual sheets within a given file. They also support a large number of formatting options (e.g. width and wrapping of cells, hiding of values, highlighting, generation of charts) and multiple views of any given sheet. The rendering of the spreadsheet on screen as well as the content and appearance of a printouts will vary considerably, based on factors such as operating system, spreadsheet application and version, user preferences, fonts supported, driver software and rendering hardware.

Document summaries: What elements of document summary information are commonly available from all major word processing applications?

Details vary from one application to another, but some examples from Microsoft Word 97 include:
- Title
- Subject
- Author
- Manager
- Company
- Category
- Keywords
- Comments
- Hyperlink Base

- Date and Time Created
- Date and Time Modified
- Date and Time Accessed
- Date and Time Printed
- Last Saved By
- Revision Number
- Total Editing Time
- Various document statistics and possible custom properties

It is also worth noting that many commercial email applications also store numerous elements that users will generally not see unless they make an additional effort to look for them. Many of these are created and retained automatically, and can often have important recordkeeping value. John Jessen, Managing Director of Evidence Associates, points out some of the noteworthy elements from a standard Microsoft Outlook email message:[25]

Attachment	Expires	Outlook Internal Version
BCC	Flag Status	Outlook Version
Billing Information	Follow Up Flag	Read
Categories	From	Receipt Requested
CC	Have Replies Sent To	Received
Changed By	Icon	Relevance
Contacts	Importance	Remote Status
Conversation	In Folder	Retrieval Time
Created	Internet Account	Sent
Defer Until	Junk E-mail Type	Size
Do Not AutoArchive	Message	Subject
Download State	Message Class	To
Due By	Modified	Tracking Status

Other summary information can be stored in locations such as templates, the file system of the operating system, directory services of a network operating system and supporting system files (e.g. the Windows Registry). We believe that a survey of the most popular office suite applications would probably reveal a number of properties that they have in common.

What other office applications that produce text documents have a similar feature?

Office suites tend to have many properties that can be applied to all file formats, with a few that are format-specific.

Is the document summary feature used in your agency and, if so, how widely? Does any agency require staff to complete the document summary routinely?

Since we are not representing a specific agency, we cannot answer this question as it is stated. Many anecdotal accounts would seem to indicate that conscious, extensive use of summary features on an agency-wide basis is rare. Basic principles of technological adoption would also suggest that users will not exert extra effort in using the features unless they have an incentive to do so. The only way to truly determine the extent of use would be collect empirical data at more than an anecdotal level.

It is important to note, however, that many features can be retained automatically. Electronic document management systems (used in some federal agencies) also often make use these native features of office suite applications.

Is a default normally used?

Without any incentive to do otherwise, it is likely that most users do not take any action to change document property values from the default. As noted previously, some are assigned automatically. Given the presence of an EDMS and/or records management application, users have considerably more incentive to use such features.

How does the agency use the information if they retain the document in a non-electronic recordkeeping system?

We would speculate that this fairly uncommon, though some of these properties can appear automatically as part of a print out, either as part of a cover sheet or along the margins of a document.

1C. If we determine that § 1234.22 should be amended to reflect Public Citizen's proposed requirements, should we retain the current paragraph (a)(3) for electronic recordkeeping systems only?

Yes. Since standard interchange formats can be important to numerous forms of electronic records, this provides yet another reason to move the existing 1234.22 to the beginning of Subpart G.

1D. Do you see any other issues that should be considered as we evaluate the Public Citizen Proposal 1?

We share Public Citizen's concern that agencies often do not comply with their stated policies to "print and delete" records. We support the rationale behind Public Citizen's proposed wording for 1234.22(a) to include "recordkeeping systems that maintain the official copy of text documents produced on electronic information systems" rather than only "electronic recordkeeping systems that maintain the official file copy." Given our concerns about "text documents" as a category, however, we would recommend wording such as the following: "recordkeeping systems that maintain the official copy of records that were produced or received with electronic information systems."

NARA's current wording could be interpreted to allow agencies to output records to hardcopy (paper, microform, etc.) without adhering to any of the other requirements of § 1234.22. Public Citizen's proposed revision explicitly requires agencies to provide for access, security, disposition and metadata standards for any records created electronically. We would also encourage NARA to include references to the following in its guidance on the disposition of electronic sources records: provisions for policy implementation auditing and periodic quality control verification of recordkeeping copy production.

Proposal 3

"Public Citizen states in its petition that records in electronic form have unique advantages, including wider and easier distribution, searching and indexing the records, and storage. Public Citizen further states that 'electronic records carry advantages for research, even if the records have not been maintained in a system that satisfies all of the attributes of an ideal electronic recordkeeping system.' Public Citizen argues that it is important to address the disposition of both text documents and data files whenever new information systems are developed."

3A1. Does (and should) electronic information "system" as used in this proposed paragraph include word processing applications?

Yes.

If so, does the word processing application technically "store" the text documents produced with the software?

No.

3A2. Should we distinguish systems that only produce or use electronic records from those that store them?

We do not believe that this is necessary. An electronic information system will consist of numerous components. For purposes of this document, it is not relevant whether production and storage are performed by the same component. The important thing is that provisions are made for incorporating electronic source records into some form of recordkeeping system.

If an agency sends all its electronic records to a records management application (RMA), NARA believes there is no need to build disposition functionality into its word processing application or into a web tool that can search and retrieve documents from the RMA.

Appropriate disposition must be applied to all records. In an ideal recordkeeping situation, all documents created electronically would be under the control of the RMA. If there are documents outside the control of the RMA, however, their disposition must also

be addressed. They cannot simply be assumed to be nonrecords. It is not clear how one would build "disposition functionality into its word processing application." One would most likely have to instead build in disposition functionality into the agency's overall recordkeeping system, in order to address any documents that are not within the control of the RMA.

What do we do about systems used to produce electronic records that are only maintained in hard copy?

Disposition could still be considered at the design stage. An agency could identify that the electronic information system will not carry out the disposition actions itself and instead make provisions for doing so through other means.

3A3. *How should "enhancements to existing systems" be defined or qualified to indicate that new or different records are being created? NARA has a general policy that agencies must reschedule their records when an agency program is reorganized or otherwise changed in a way that results in the creation of new or different records (see 36 CFR 1228.26(a)(2)).*

NARA's policy appears to address this issue adequately.

3A4. *What activities does the term "produce" cover? Is there a clearer way to state these activities?*

"Produce" includes creation, revision, annotation and amendment. "Produce" seems to be the simplest way to cover this concept. We would also suggest adding the word "receive" to Public Citizen's list.

Do you see any other issues that should be considered as we evaluate Public Citizen proposal 3?

We agree with Public Citizen that is extremely important to address records scheduling and appraisal as early in the system development life cycle as possible, and this should hold for all electronic records rather than simply data files. We share NARA's concern over the wording proposed by Public Citizen, however, and feel that the provisions of both 1228.26(a)(2) and 1234.32(a) address this point sufficiently.

We would like to reiterate a definition provided by the Electronic Records Work Group:

> *Business needs*. An agency's need to conduct its business, maintain a record of its essential activities and decisions for its own use, support oversight and audit of those activities, and permit appropriate public access. Agencies have certain responsibilities under the Electronic Freedom of Information Act Amendments (EFOIA) to make records available in electronic format. Although NARA does not have the statutory authority to mandate how agencies comply with EFOIA,

agencies should be aware that public access is one of several business needs that they need to consider in scheduling their electronic source records.

As Public Citizen argues, retention of records in electronic formats will often greatly promote both the convenience and quality of public access to those records. Recent case law provides a number of examples in which parties have been required to provide the electronic version of records as part of the evidence discovery process.[26] While this does not imply that all records must be retained in electronic formats for their entire life, it does provide compelling support for the view that electronic source records can often contain important elements of evidence that may be lost in printed copies.

NARA's mission is ensuring "ready access to essential evidence." This is a primary objective of the archival profession as a whole. It is also a mission in which the federal government, citizens, and other potential researchers all have an interest. EFOIA represents an important set of business needs related to such access. Recent legislation such as the Government Paperwork Elimination Act[27] and Electronic Signatures in Global and National Commerce Act[28] has also introduced a number of requirements to manage and provide access to government records in electronic formats. Of course, agencies must continually strike a balance among numerous business needs. This balance is recognized in the FOIA legislation, which states that each "agency shall make reasonable efforts to maintain its records in forms and formats"[29] that can be both searched and reproduced electronically. We encourage NARA to continue expanding its efforts to promote the management and preservation of electronic records within the federal government.

---

[1] Public Citizen Litigation Group, "Public Citizen Petition to the Archivist to Amend GRS 20 and Electronic Records Regulations," (October 31, 2000). Available at http://www.publiccitizen.org/litigation/briefs/ERecords/articles.cfm?ID=672

[2] Task Force on the Artifact in Library Collections, "The Evidence in Hand: The Report of the Task Force on the Artifact in Library Collections," (Washington, DC: Council on Library and Information Resources, 2001). Available at http://www.clir.org/pubs/abstract/pub103abst.html

[3] Frank Evans, Donald Harrison and Edwin Thompson, "A Basic Glossary for Archivists, Manuscript Curators, and Records Managers," *American Archivist* 37 (1974): 424.

[4] Committee on Intrinsic Value, "Intrinsic Value in Archival Material," Staff Information Paper Number 21 (Washington, DC: National Archives and Records Service, 1982).

[5] James M. O'Toole, "On the Idea of Uniqueness," *American Archivist* 57 (1994): 632-58.

[6] 36 CFR 1234.30. Selection and maintenance of electronic records storage media.

[7] Charles M. Dollar, *Authentic Electronic Records: Strategies for Long-Term Access* (Chicago, IL: Cohasset Associates, 1999).

[8] Kenneth Thibodeau, "To Be or Not to Be: Archives for Electronic Records," In *Archival Management of Electronic Records*, edited by David Bearman (Pittsburgh, Pa: Archives and Museum Informatics, 1991), 1-13.

[9] Abby Smith, ed. *Authenticity in a Digital Environment* (Washington, DC: Council on Library and Information Resources, 2000).

[10] Lewis J. Bellardo and Lynn Lady Bellardo, *A Glossary for Archivists, Manuscript Curators, and Records Managers*, *Archival Fundamentals Series* (Chicago: Society of American Archivists, 1992), 19.

[11] 2 F. Supp. 2d; see also 184 F. 3d at 910-11

[12] 36 CFR 1228.270(d)(4)

[13] 36 CFR 1228.270(e)

[14] David Holdsworth, and Derek M. Sergeant, *A Blueprint for Representation Information in the OAIS Model* (2000). Available at http://gps0.leeds.ac.uk/~ecldh/cedars/nasa2000/nasa2000.html.

[15] RLG-OCLC Digital Archive Attributes Working Group, "Attributes of Trusted Repositories for Digital Research Resources: Meeting the Needs of Research Resources," (Mountain View, CA: Research Libraries Group, 2001). Available at http://www.rlg.org/longterm/attributes01.pdf

[16] 44 USC 3303

[17] 36 CFR 1234.2, under "data file" definition.

[18] One could argue that relational database tables are more "strictly prescribed" than electronic mail messages, though this may not always be the case. Relational databases generally support large free-text fields, and binary-large objects (BLOBs). They also often contain stored procedures, triggers and pointers to external applications, all of which reach far beyond the bounds of the relational data model itself. Though electronic mail messages are generally described as "semi-structured," their allowable data elements are actually quite formally defined. In fact, both client and server electronic mail applications often store and manage messages within relational databases. As both database and electronic mail applications increasingly provide native XML support, the structured/semi-structured distinction between the two blurs even further. The important distinction for purposes of recordkeeping is the nature of creation and use. NARA quite rightly provides a handful of required data elements for electronic mail messages, since the way in which they are used tends to both require and yield those elements, regardless of the record series under which the messages are categorized.

[19] Bellardo and Bellardo, 8.

[20] For a description of the concepts of form and function (within the context of description and access, rather than records management), see David A. Bearman and Richard H. Lytle, "The Power of the Principle of Provenance," *Archivaria* 21 (1985): 14-27.

[21] For some simple formats, this will be a "body" element that is not further differentiated.

[22] National Archives and Records Administration, "Introduction to the General Records Schedules," (December 1998). Available at http://ardor.nara.gov/grs/grsintro.htm

[23] There is a great deal of digital data associated with a record that is hidden from the view of most users but can be revealed through the use of forensic tools. These include scratch files, "undo" text stored within files, various caches, history files, temporary files, and even data that has been deleted but can still be read off of the physical medium. While the proper destruction of such data can be a vital security concern in some environments, their ongoing retention is rarely a recordkeeping requirement, since they serve only to facilitate efficient operations (error recovery, crash protection, etc.) rather than contributing directly to the content, structure or context of a record as evidence of a given function, activity, transaction or event.

[24] For an analysis of many properties of concern in the preservation of spreadsheets, see Gregory W. Lawrence, William R. Kehoe, Oya Y. Rieger, William H. Walters, and Anne R. Kenney, "Risk Management of Digital Information: A File Format Investigation," (Washington, DC: Coalition on Library and Information Resources, 2000). Available at http://www.clir.org/pubs/abstract/pub93abst.html

25 John Jessen, "The Continuing Impact of Electronic Discovery," Paper presented at the National Conference on Managing Electronic Records (Chicago, IL, September 24-26 2001).

[26] See e.g. Greyhound Computer Corp., Inc v. IBM 3 Computer L. Serv. Rep. 138, 139 (D. Minn. 1971); Adams v. Dan River Mill, Inc. 54 F.R.D. 220 (W.D. Va. 1972); Minnesota v. Philip Morris Inc., No. CI-94-8565 (Dist. Ct. Minn.); National Union Electric Corp. v. Matsushita Electric Industrial Co., 494 F. Supp. 1257 (E.D. 1980); Williams v. Owens-Illinois, Inc., 665 F.2d 918 (C.A. 9, 1982); In re Air Crash Disaster, 130 F.R.D. 634 (E.D. Mich. 1989); State of New York and UDC-Love Canal Inc. v. Hooker Chemicals and Plastics Corp, Order, CIV-79-990 (W.D.N.Y. Nov. 30, 1989).

[27] Public Law 105-277, Division C, Title XVII

[28] Public Law 106-229

[29] 5 USC 552(a)(3)