# Context, Structure and Content:
# New criteria for appraising electronic records

Jim Suderman[1]

**Postscript**:  **Postscript:  Since this paper was delivered at the Annual Conference of the Association of Canadian Archivists in June, the final report of the InterPARES Authenticity Task Force, Authenticity Task Force Final Report, has become available.  Readers who find my paper interesting may wish to read this one as well, particularly as it relates to the verification of authenticity (see section 4.2.4.2) and preservation baseline requirement B.2 "Documentation of Reproduction Process and its Effects" (see section 3.2 of the Appendix).**

*Introduction*

Electronic records have unique characteristics.  This paper explores some of those in terms of appraisal, and proposes four criteria specific to electronic records.  It also discusses at what level of appraisal these criteria would best be applied.

The introduction of the macro-appraisal model clearly established appraisal as a hierarchical exercise.  That model is designed to isolate key areas where the best records are likely to be located by examining the structure and functions of the creating agency.  The appraisal of the actual record series follows the identification of those key areas.   Terry Cook cautions that

> **It is important to recall that there are several factors at this later stage which can refine or reverse a positive decision made by using the macro-appraisal model.**[2]

At this lower level archivists look at the different series to determine which of them best fulfil the values sought or identified at the macro-appraisal level.  As well various criteria are applied, such as completeness, uniqueness, barriers to accessibility, relationship with other archival holdings, etc.  While these criteria remain relevant and applicable to the appraisal of electronic series, the separability of content, structure and context - the three components of a record, along with other characteristics unique to electronic records, suggest that these "traditional" criteria are not sufficient in themselves to fully appraise electronic record series.  The four new criteria proposed in this paper apply at one level lower yet than what I have referred to as series level appraisal, thus two levels below macro/functional appraisal.  Paraphrasing Cook, one might say that there are several factors at this later stage that can refine or reverse a positive decision made at the series level.

There is a need to appraise at this lower level because the component facets of a record - context, content and structure - are independent of each other as well as of their medium in the electronic environment. With every migration, whether it occurs while the records are still operational, at point of transfer to an archives, or while in the custody of the archives, some aspects of the context, content and structure are affected.  The National Archives of Australia acknowledges this fact as follows:

> **In the electronic environment we consider that the 'original' means the content, structure and context of the original transaction but not all the attributes present in the original software or hardware platform. It is inevitable that some losses will occur at the point of migration from one version of the software to the next or one platform to the next but this is acceptable as long as the aspects of the record required for evidence are preserved.**[3]

It is therefore critical for archivists to determine what the essence of the record is in order to ensure that "the aspects of the record required for evidence are preserved."

This paper uses the definitions contained in the National Archives of Australia's publication *Keeping Electronic Records*. Content is defined as "That which conveys information e.g., the text, data, symbols, numerals, images, sound and vision."  Structure is "The appearance and arrangement of the content e.g., the relationships between fields, entities, language, style, fonts, page and paragraph breaks, links and other editorial devices."  And context is "The background information which enhances understanding of technical and business environments to which the records relate e.g., metadata, application software, logical business models, and the provenance (i.e., address, title, link to function or activity, agency, program or

[4]

section)."

In addition to the independence of the content, structure and context, records in the electronic environment have unique characteristics. Harold Naugler noted several of these and I have summarized them as follows:

- durability
- lifespan
- maintenance
- ease of editing, copying erasure and reformatting (manipulability)
- ease of manipulation, including the difficulty of tracing manipulation
- need for supporting documentation to describe the contents, arrangement, codes, and technical characteristics
- need for specialized personnel for the processing and maintenance of the records, introducing a new player in the normal clique of archivist, creator and user.[5]

From these unique characteristics I propose four appraisal criteria:
1. Durability,
2. Presentation/Rendering,
3. Manipulability, and
4. Technical Context.

*Durability*

Naugler refers to this characteristic primarily in terms of storage media. I define it here in terms of the durability of the native application's contribution to the record. Durability is an important consideration because changing the format of the record may have a fundamental impact on it, as will become clear through the illustrations used to clarify these criteria. Migrating records or data to an open format simply means a double conversion, as they will presumably have to be opened again in another application.

Illustration 1 provides a view of a record called Jim's Calendar. Much of the structure is established by Microsoft Outlook 98 - as testified by the buttons visible along the top. The data is ordered chronologically within the Calendar folder.



*Illustration 1: Jim's Calendar, "normal" view*

*Illustration 2:  A sample entry from a paper-based calendar.*

Contrast this illustration with illustration 2, which is a scanned image from a paper journal or calendar.  The paper calendar, to my eyes, shows more directly the impact of its contributors.  It has an area for listing tasks and, at the bottom left, a place for recording expenses and reimbursements.  Different colors of ink and styles of writing suggest that different individuals contributed to the calendar.  Not all of these characteristics are apparent in the illustration of the electronic calendar.  There is a task list, but the information is presented in a uniform fashion. Unlike the paper calendar, MS Outlook contributes the ability to change the structure of the calendar from a presentation of the daily information to a presentation of information for the whole month (see illustration 3), an option not possible in the paper calendar.



*Illustration 3: Jim's Calendar presented by month.*

Having pointed out some of the differences between the same type of record, how is durability applied as an appraisal criterion?  In a test undertaken with this Calendar record, exporting the data from Outlook to an open format and reloading it back into Outlook worked very well.  Would it do so well if the data were reloaded into MS Outlook 2000?  Or 2005?  As soon as the structure of the data in the application is changed the option of reloading older data in its native format or even from open formats becomes less sure.  Emulating applications has been proposed as a means of overcoming this problem, but as the ICA's Committee on Electronic Records has warned, there are significant issues that affect emulation as a long-term strategy.[6]

By contrast, a simple text file created in WordPerfect 5.1 opens reasonably well in Microsoft Word 97.  This suggests to me that records created using WP5.1 have a higher durability than those created in Outlook.  The WordPerfect documents enjoyed the moderate lifespan of its native software (WP5.1), which is extended by the accessibility of the format using other, more recent applications.  Using a viewer application to look at the same text file may confer a high durability to records in this format. In this instance I conclude that the Outlook Calendar record has low durability because the application contributes a great deal to the structure of the record and that contribution cannot be carried out of the application with the content.

Why is durability important?  Is it not simply a technical matter concerning preservation?  I would argue that it is an appraisal concern because as a record's durability declines over the years, so will the contribution made by the original application to the record's structure.  As the record changes when it is accessed through another application so will the way the viewer understands the record.

If records have been preserved primarily for their informational value, durability might recede into insignificance.[7] But if the Calendar was preserved in part to convey some evidential aspect of how the creator went about his/her business, then loss of that original functionality may render the record no longer worth preserving.[8]  Which is to say that durability becomes a significant appraisal criterion where values other than informational value predominate.[9]

*Presentation/Rendering*

The manner in which records are presented or rendered is closely linked to durability.  This criterion, however, goes beyond durability to address which visual attributes give value to the records.  In our Calendar example, if it is decided that the essential attributes are simply the date and time, event details and knowing whose calendar it is, then there are no visual attributes other than that provided by the Gregorian calendar, that give value to the record.  The essential data, in this instance being very simple, could easily be sorted and presented in different ways, i.e., the presentation or structuring provided by the native format is not deemed critical. Other data, such as expenses might be deemed secondary and discarded. (Expenses is visible only in the paper calendar so far, but each appointment in the electronic calendar can also be "opened" to show this additional information). But what if one knew that the data was normally viewed by category, i.e., where appointments from other categories are not interspersed, (see illustration 4) rather than simply by chronology?  As illustration 4 shows, the structure of presentation is considerably different.  In this case the presentational structure of the native format may become a value or aspect that needs to be preserved.

*Illustration 4:  Jim's Calendar, categories view.*

Similarly, the content of Jim's Calendar is sound but the rendering is very different when it is extracted as tab separated values and viewed using Microsoft Notepad (see illustration 5).



*Illustration 5:  Jim's Calendar (portion) as
tab separated values and presented in MS Notepad.*

This particular format does not allow flexible rendering, this is the only presentation possible.  If this content is loaded into an application that can address the tab delimited data elements individually, such as a spreadsheet, then the data not only will look more comprehensible, but its data elements can be manipulated as well to allow for alternative renderings.

It can be argued that how the value is preserved is a technical consideration.  It may be preserved by keeping the records in their native format, if that has a high durability.  Alternatively, it may be preserved in the description of the record, leaving the secondary user (i.e., the researcher) to restore any particular means of presentation.  There will likely be other

ways as well.  What I am suggesting is that presentation should be an appraisal criterion simply because it would establish certain requirements for the long-term preservation of the record.  Any archival migration strategy would need to ensure that the visual characteristics of the records in their native format *that were appraised as giving value to the record* were accommodated in whatever new format or environment to which the records were migrated.  Failure to preserve these values might make the record not worth preserving.  The *Final Report* of the Victorian Electronic Records Strategy emphasizes this as follows:

> **From an archival perspective, it is important that both the content and structure are accurately captured.  The captured record should be identical in appearance to the original document as it was viewed by the creator of the record.**[10]

Thus, Jim's Calendar as it appears in illustration 5 does not meet the requirement articulated in the VERS *Final Report*, although this format has much to commend it purely in terms of preservation.

Like durability, this criterion does not have an analogy in the paper environment.  Record characteristics and functionalities are not, I think, as variable in paper form as they are in electronic form because in paper technology content and structure are normally closely linked.

*Manipulability*

Manipulability is greatest when the record resides in its native format and in its operational environment.  It is a curiously paradoxical value from an archival standpoint.  One might think that archivists would reject manipulability root and branch as a desirable value in records.  Yet this very value is consistently identified as desirable at least since 1984 when Naugler wrote:

> **For those machine-readable records containing information duplicated by textual records, the machine-readable records will, in the majority of cases, be appraised as having the better arrangement because they have greater manipulability.**[11]

Manipulability can apply to all three record components either singly or collectively.  In appraisal therefore, the archivist must decide for which components manipulability is archivally valuable.   In Illustration 5 the potential for manipulation is very different from what it is in its native format.  For example, I can replace all the R's with Q's, but cannot manipulate the data elements in the way that is possible in Outlook, its native format.  Thus if manipulability of the content in terms of discrete data elements is archivally valuable, then preserving the content as illustrated in Illustration 5 means the researcher must be alerted to this value, and presumably is expected to bring a resource to the record to restore that manipulability.

Manipulability, like the presentation/rendering criterion, is linked to durability.  Durability will be irrelevant where the manipulable characteristics of the native environment are not desired, i.e., when a very low manipulability helps preserve the value of the record.[12]

*Technical Context*

I address the technical context criterion last because it seems already near to adoption in Canada.  The most recent RAD draft chapter for the description of electronic records that I have seen (April 2000) contains a requirement for a description of the system in cases where it is deemed significant to an understanding of the unit being described.[13]  The specific descriptive elements listed in the chapter are not important here, but the recognition that system information can be significant to understanding records is.

This recognition helps define the criterion.  For example, how is one to determine whether the technological context is indeed significant to understanding the records?  And if it is, what aspects of that context are significant?  Knowing that Jim's Calendar was created in a networked environment would be important because it is possible that contributors other than the creator may have affected the content by entering, changing, or deleting information.  Knowing that MS Outlook was specifically designed to support that capability would alert an appraiser to learn who had such privileges, or to account for the impact of that fact on the record's value.  It might also be useful for the appraiser to know that Outlook allows appointments

to be made "in the past" or deleted without trace.

Illustration 6 provides yet another a view of Jim's Calendar, this time as comma separated values loaded into a Microsoft Excel spreadsheet.  An aspect of the technical context that it illustrates



*Illustration 6:  Jim's Calendar, comma separated values loaded into MS Excel.*

is that the data is in a different order than it appeared in the native format - still chronologically, but by date of entry rather than by date of event.  Date of entry information can be found within Outlook, but the aspect of the technical context illustrated here is that the date of entry information was not exported with the content of the record, except in the way the data was ordered during the extraction process.

*Conclusion*

It is on the basis of the unique characteristics of electronic records and the independence of content, context and structure, that this paper proposes new appraisal criteria.  Existing criteria remain useful and relevant, and for this reason it has been my goal to propose criteria where the analogy to existing criteria is weak or absent. In the case of durability, while it is true that paper records are reformatted onto microfilm or digital formats, this situation is more the exception than the rule.  Whereas in the electronic environment reformatting is rightly assumed to be a fact not only of a record's archival life, but of its operational life as well.  Thus durability is a temporal criteria, directly addressing the changing nature of electronic records over time.

Presentation or rendering in the paper environment follows a culture centuries old and uses a technology that is equated with the term "fixed".  That there is no comparable tradition in the electronic environment is clear at least to Dick Brass, Vice President of Microsoft's eMerging Technologies group. He said in an interview that

**… his group is not "anti-paper," that they love it, they venerate it. 'We respect it, and we think the tragedy of computing to date is that we didn't sufficiently imitate it.'[14]**

The manipulability criterion is a counterpoint to the presentation/rendering one.  Manipulability provides for the greater or lesser manipulation of content, structure and context.  As the contrast between a paper calendar and an electronic one makes clear, "archiving" the paper calendar is an all or nothing exercise, whereas with the electronic calendar content, structure and technical context can be preserved wholly or only in part.  For example, in Jim's Calendar, content can be wholly preserved, while at the same time preserving no structure and only a part of the technical context.

Of the four proposed criteria the technical context criterion is perhaps the one most closely linked to existing criteria for paper records as it reflects some of processes involved in creating the record.[15]

Another goal of mine has been to propose criteria that address one or more of the facets of a record, i.e., the criteria may not deal with the record as a whole, but merely one of its components. Archivists routinely make decisions about the essential context to preserve records in the paper environment. This decision-making process is reflected in the functional appraisal model where functions are appraised and only those that are deemed significant in some way are documented. Thus, housekeeping records are routinely destroyed because their importance to the essential context of the more valuable records is considered to be low.

Because archivists have always approached context in terms of essence I believe this practice to be valid. Archivists seek to determine the essential contextual elements to preserve meaning and authenticity through arrangement, description and the custodial chain of ownership. Archivists now need to address with equal confidence appraising not only the record as a whole, but its individual components as well. What technical structure or structures are needed to preserve the values ascribed to a series of records? Not addressing this matter results in either preserving all structure(s), something which I have tried to illustrate as being difficult in the extreme, or a more random preservation, based on preservation priorities or convenience rather than appraisal considerations.

Angelika Menne-Haritz writes:

**Appraisal is a body of methods and techniques to destroy in order to preserve. Preservation as a professional task results from this concept. By destroying consciously and in a responsible way the remains are saved and can be appreciated in all dimensions of their value.[16]**

If this is true, then archivists have to determine what elements of a record can be destroyed that will still leave the essence of the record intact. It is essential to do this because we know that there are changes in content, structure and context even during the operational life of an electronic record. We know that converting records or data to new platforms, even open ones, has an impact on structure and context, and even on content.

As well, new appraisal criteria must be sensitive to the limitations of the technology contemporary with the records under review. Recordkeeping initiatives are underway to ensure that better electronic records will be created in the future. This too is an appraisal exercise, and it is valuable to recall the observation made by Lily Koltun with reference to electronic records:

**These are not in fact archives whose value is derived from their office of origin, but from the theorizing and selection principles of archivists who identify their source and scope, judge their value, select and preserve them prior to their creation and then "appraise" them once again post-creation.[17]**

So it is clear that archivists are defining what the archival record is even now. In her article "Are We Collecting the Right Stuff" Carolyn Heald wrote, "I fundamentally disagree with the notion that archives store information; we store artifacts in which information inheres."[18] Archival appraisal must have criteria that help illuminate the right virtual stuff.

*So far as possible I have adhered to the words of the text; but Icelandic is a highly idiomatic language, and Icelandic idiom is not English idiom. I have not hesitated therefore, in departing from the verbal idiom in order to preserve the sense.*
*G.H.Hight in his 1913 Translator's Introduction to*
*The Saga of Grettir the Strong*

Version 1.02, 2 November 2001

---

[1] Jim Suderman is the Coordinator of the Electronic Records Program at the Archives of Ontario in Toronto, Canada. Jim will be discussing Archives of Ontario electronic records implementation

issues in his upcoming paper "Implementing Custody of Electronic Records at the AO" for presentation at the Association of Canadian Archivists annual conference in Vancouver, British Columbia, 20-25 May 2002.

[2] Terry Cook. "Mind over Matter: Towards a New Theory of Archival Appraisal" in Barbara L. Craig, ed. *The Archival Imagination. Essays in Honour of Hugh A. Taylor* (Ottawa: Association of Canadian Archivists, 1992), 58.

[3] National Archives of Australia. *Management of Electronic Records*, Appendix 3 "Preserving Electronic Records through Migration."

[4] National Archives of Australia. *Keeping Electronic Records*, chapter 4 "Records - Their Creation and Management."

[5] Harold Naugler. *The Archival appraisal of machine-readable records: a RAMP study with guidelines* (Paris: United Nations Educational, Scientific and Cultural Organization (UNESCO), 1984), para. 1.46, p. 14.

[6] See ICA Committee on Electronic Records. *Guide for Managing Electronic Records from an Archival Perspective* (Paris: ICA, 1997), 48. Emulation is defined as"One system is said to emulate another when it performs in exactly the same way, though perhaps not at the same speed." (Free On-Line Dictionary Of Computing).  Difficulties listed in the *Guide* include 1) an unwarranted assumption that it will be possible to run any operating system under an emulator indefinitely into the future, 2) the need to emulate only part of the native application to prevent creation or manipulation of the preserved records, and 3) emulation entails an ever expanding requirement for in-depth expertise in obsolete software.

[7] Informational Value: The value of records/archives for reference and research deriving from the information they contain as distinct from their evidential value.  Definition taken from Peter Walne, ed. *Dictionary of Archival Terminology* 2nd revised ed., ICA Handbooks Series Volume 7 (Munich: K.G. Saur, 1988).

[8] Evidential value:  The value of records/archives of an institution or organization in providing evidence of its origins, structure, functions, procedures and significant transactions as distinct from informational value. Walne, Peter, ed. *Dictionary of Archival Terminology* 2nd revised ed., ICA Handbooks Series Volume 7 (Munich: K.G. Saur, 1988).

[9] Durability also affects the archival functions of acquisition, description and preservation.  These functions are presumably simplest in cases where durability is either high, because the native format still endures (as is the case with simple text records preserved in their native format), or not important because the context and meaning provided by the native software is not deemed significant to understanding the record.

[10] *Victorian Electronic Records Strategy Final Report* (1998) section on Record Capture, sub-section entitled "Capture of Content and Structure", p. 18.

[11] Naugler, paragraph 4.12, p. 59. See also the Kansas State Historical Society. *Kansas Electronic Records Management Guidelines* [2000], section 7.2 deals with appraisal criteria, with sub-section 7.2.2 dealing specifically with manipulability.

[12] In this context it is interesting to note that the Public Record Office in the United Kingdom identifies three format types appropriate for electronic records in its publication *Management, Appraisal and Preservation of Electronic Records:*  1) transfer formats, 2) preservation formats, and 3) presentation formats. Public Record Office. *Management, Appraisal and Preservation of Electronic Records*. Section 4.16.

[13] *RAD* chapter 9 "Records in Electronic Form" revised version - draft for comment - April 2000, rule 9.7D2.  Some specific elements include system name and developer, hardware, operating system, and network configuration.

[14] Kim Honey.  "Beyond Paper, Part I", *The Globe and Mail*, Monday, March 5, 2001, p. R3.  Dick Brass is the Vice President of Microsoft's eMerging Technologies group.

[15] Michael Wettengel, Senior Archivist, Electronic Records Division of the Federal Archives in Germany identifies three contexts:  structural, functional and technical.  Michael Wettengel.  "Old Traditions and New Uncertainties. The German Archival Concept of a Record and Electronic Environments" in *The Concept of Record. Report from the Second Stockholm Conference on Archival Science and the Concept of Records, 30-31 May 1996* (Lund: Riksarkivet, 1998), pp. 139-40. Note that I have not addressed cost as an appraisal criterion.  This is due to the structure of resource allocation for archives, which can be generalized to say that it is optimized for the support of paper records.  Supporting archival records in electronic formats is seen by contrast to be costly.  At some point, and certainly in some institutions the process is already well underway, the structure of resource allocation will be modified to better support electronic records.  At this point then, it is impossible to set a meaningful cost criterion.

[16] Angelika Menne-Haritz. "What Can be Achieved with Archives?" *The Concept of Record. Report from the Second Stockholm Conference on Archival Science and the Concept of Record, 30-31 May 199 .p*. 15.

[17] Lily Koltun. "The Promise and Threat of Digital Options in an Archival Age" *Archivaria* 47 (Spring 1999), 123.  Koltun's article goes far beyond the scope of this paper as is illustrated by the sentence immediately preceding the one quoted:  "So now we have the full and staggering implication: that digital data represent the first medium collected by archives which can be totally dependent on the 'archiving function' for its birth, its definition of value, and its continued life."

[18] Carolyn Heald. "Are We Collecting The 'Right Stuff'?" *Archivaria* 40 (Fall, 1995), 182-188.