InterPARES Project

International Research on Permanent Authentic Records in Electronic Systems

Applying Content Analysis to Case Study Data: A Preliminary Report

Prepared for the Authenticity Task Force by:
Ciaran Trace
Department of Information Studies
University of California, Los Angeles
12 June, 2001

Applying Content Analysis to Case Study Data

Introduction

Content analysis can be defined as "*an overall approach, a method, and an analytic strategy*" that "*entails the systematic examination of forms of communication to document patterns objectively*".[1] Content analysis is generally applied to narrative texts such as political speeches, transcribed interviews, and published literature, and seeks to elucidate, through close examination of the content and language of these texts, what can be learned about authors' or respondents' understanding of phenomena and terminology, as well as their beliefs. In using content analysis as a method, the objective is to get at aspects of meaning by examining the data qualitatively. In effect the method is used to examine how authors or respondents view and understand certain issues.

A key issue that surfaces in many inter-disciplinary research projects today relates to the different viewpoints, understandings, and terminological usage of researchers and research subjects. As a result, it can be difficult for researchers to understand the extent to which the issue that they are investigating is understood or conceptualized in the same way by research subjects, and by extension, to understand the full import of data gathered from those subjects. The aim of this preliminary report is to explore if, or to what extent, the data gathered through the Case Study Interview Protocol (CSIP) can be further exploited using content analysis techniques in order to enrich our understanding about how records creators, managers, and information systems personnel view the nature of records and the concept of authenticity. The CSIP was developed by the InterPARES Authenticity Task Force specifically to populate the Template for Analysis in order to help the Task Force address the following Domain I research questions:

1. What are the elements that all electronic records share?

---

[1] Catherine Marshall and Gretchen B. Rossman, *Designing Qualitative Research* (Thousand Oaks: Sage, 1995): 85.

2. What are the elements that allow us to differentiate between different types of electronic records?

3. Which of those elements will permit us to verify their authenticity over time?

4. Are these elements for verifying authenticity over time the same as those that permit us to verify their authenticity in time (i.e. at the point at which they are originally used)?

5. Can those elements be removed from where they are currently found to a place where they can more easily be preserved and still maintain the same validity?

This report explores the use of content analysis to address the following additional research questions posed by the US team as part of their participation in the InterPARES Project:

1. To what factors besides authenticity do records creators, custodians and systems personnel give weight when addressing the permanent preservation of electronic records?

2. Currently, when the long-term authenticity of records cannot be assured what metrics and/or heuristics are put in place, why, and by whom?

3. In what ways do records creators, custodians, and systems personnel conceptualize the nature and role of the electronic records and/or recordkeeping system being studied?

4. What are the variations in language used to describe records by record creators, custodians, and systems personnel?

5. What is the relative importance of business policy, pragmatism, and institutional culture of the organization in determining the fate of electronic records and recordkeeping systems?[2]

Data Collection

---

[2] See *Methods and Procedures for Analysis of Data Collected and Template Data Gathering Instrument and Raw Case Study Data*, (Los Angeles, CA: US InterPARES Team) December 13, 1999. Draft

To date thirty-two case studies have been or are in the process of being conducted for the InterPARES Project in a number of sectors and national settings. The protocol for conducting the case studies consists of structured interviews with information technology and managerial personnel about the electronic information/record-keeping systems with which they work, as well as the collection of documentation associated with these systems. The Case Study Interview Protocol (CSIP) used in the interviewing process was designed by the Authenticity Task Force to elicit specific, and as much as possible, comparable, information about aspects of systems and the records that they contain. The CSIP addressed 5 aspects of electronic records and recordkeeping: context, intrinsic elements of form, extrinsic elements of form, annotations, and medium & technological context. Most interviews were tape-recorded and were of approximately two hours duration. At this point in the InterPARES project, seven of the Case Study Interview Protocols have been transcribed.

This report discusses the potential utility of content analysis based upon a preliminary examination of one case study transcript. The case study examined was of a DataCAD® system of a small US architectural firm and one individual, a partner in the firm, was interviewed (hereafter "the respondent").[3] This case study, because of the depth of interaction between the respondent and the interviewers, provided an unexpectedly rich text for content analysis. In analyzing the case study, the aim was to identify themes relating to two aspects that are integral to InterPARES research: records and authenticity. The report does not attempt, however, to apply any quantitative techniques to the data such as analysis of frequencies of terms related to records or authenticity.

Data Analysis

In order to analyze this case study interview the CSIP transcript was read and re-read in order to code for emergent themes. In using qualitative techniques such as content analysis, it is important that the person analyzing the data allows themes to emerge naturally, rather than attempting to impose a preconceived set of themes on the data.

This approach ensures that any unanticipated themes are given the opportunity to emerge from the data and that no undue weight is given *a priori* to any preconceived themes. The transcript data was, therefore, analyzed thematically first, and only after themes were identified, were they examined in light of the questions already identified by the US team.

As stated earlier, the focus of the analysis was on understanding the respondent's views about two areas: records and authenticity. Throughout this report, selected extracts from the respondent's replies to the Case Study Interview Protocol questions are quoted in order to illustrate the main themes that emerged. These quotations, marked by italics and quotation marks, are referenced exactly as written in the transcript and as such reflect the flow and inconsistencies of the spoken word.

a. Records-related themes

The record-related themes that emerged from the data were reflections of three over-arching questions:

I.   How do organizations view their own records? How are these perceptions modified by outside influences?
II.   What factors influence the way records are created and formatted?
III.   How are electronic records and paper records similar and how are they different?

I. How do organizations view their own records? How are these perceptions modified by outside influences?

One theme that emerged in this section was that a distinction or demarcation is made within this organization when talking about the records of the business, i.e., that a

---

[3] Case Study #27, conducted by Philip Eppard, Kevin Glick and Rebecca Hatcher, Albany, NY, January 2001.

categorization of record types is recognized within this organization no matter how rudimentary such a categorization may be (although it is unclear how explicit this categorization is). Furthermore, this distinction is based on functional grounds. For example, the respondent references a natural categorization of records that he sees within the organization of two types of documents in use by the firm: *"the normal documents required to just run any business"* and documents that are *"instruments of [their] service"*. Whether such a distinction is of a purely functional nature or whether the distinction is also made in terms of how documents are valued is not overtly stated. However, using such a categorization, the respondent demarcates a group of documents that are unique to his particular business – that of an architectural firm. Such documents include contract documents, design documents, drawings and specifications. The notion of documents as "instruments of service" means that documents are viewed as instruments to get the job of the architectural firm done. However they do not represent the actual service that the business provides. The respondent clarifies this by stating that "*the product is really the finished building, it's not the documents themselves*". What is interesting is that the view of documents as "instruments of service" is not one that originated within the firm itself. The firm's insurance company has imposed such a view. The importance of such a classification of its documents is that it enables the firm, rather than their clients, to control ownership of the documents.

A second theme that emerged under this rubric was that electronic documents are seen as having long-term value for a mixture of pragmatic, financial and legal reasons. In this case study, the respondent privileges pragmatic and financial reasons over the legal reasons for keeping the records. Reasons cited by the respondent for the long-term retention of documents are their value as a seed for other projects and in case of legal actions (public liability), and also that the documents are seen as assets that can be resold or reused by future owners of the property. In the case of electronic records, the respondent also raises a very pragmatic reason for not discarding information. As the respondent states "*there's really no reason to discard the information, it doesn't take enough room to make it worthwhile to discard it.*"

Such a rationale for the long-term preservation of records can break down in practice. A case cited by the respondent shows how such a "*just as easy to keep them as not*" attitude can result in the loss of records. The case involved a change in software where the new software was not backward compatible. The company had to decide whether to translate the documents into a form that could be used by the new software or whether they would "*just forget them, and let them go*". The firm chose not to translate many of the records because it was no longer just as easy to keep them as not. An interesting point to note is that some of these electronic documents are still maintained, even though they are currently unreadable, in case they are ever needed. It is the space consideration that helps to keep electronic records around, a factor that seems to work against their paper counterparts in this firm.

II. What factors influence the way records are created and formatted?

A theme that emerged in relation to this question, was that of the role that expectation plays in the way that records are created and formatted. "Expectation" (the coder's term) here relates to the notion that documents should conform to the needs of the various parties involved, whether this be the needs of the architectural firm itself, outside professional and governmental bodies, or clients of the architectural firm. This theme of "expectation" appears to hold equally true of paper and electronic records. The way records are created and formatted appear to be conditioned by three main factors:

i. Influence of professional standards and regulations (external expectations)
ii. Influence of custom (both internal and external expectations)
iii. Influence of the customer (external expectations)

In this case study, the creation, form and content of the documents are influenced by the interplay of these factors, not necessarily all to the same degree. Parts of the document that are influenced in this manner include the form, content, terminology, use of symbols and graphic conventions. The respondent states that the format of construction documents is usually the same *"in just about any architectural firm in the country"*.

Legal documents are also standard forms of agreement. Floor plans, as a document type, have an almost universal arrangement. The respondent states that this has *"grown up"* through *"custom"* into a professional standard. The following quote from the respondent demonstrates just how powerful an influence customs, and the associated expectation that documents will be a certain way, can have in his domain:

"It's very [difficult] to deviate from the custom because people in the field who eventually have to read these documents in order to build a building won't know what to do with the set unless they know what to expect from it. Minor changes can cause a problem...."

In this instance, the change that the firm made was to reposition a document called a schedule within the set of drawings (a schedule is a document that gives floor numbers, door types, size, and materials in a tabular form). The schedule was put first in the set of documents rather than leading with the floor plan as is customary in the profession. As a result, the people in the field could not find the schedules because they were used to finding it after the floor plan. This example is useful in demonstrating a fourth factor that influences the way in which documents are created, that is, that there is a necessary order and interrelatedness to the documents with which the architects work. This order and interrelatedness brings with it an associated need for the documents to reference each other in a particular manner. The respondent mentions, for example, that among architectural documents, the floor plan is seen as the *"homepage"*. It forms the *"basic key to which all information is tied"*.

III. How are electronic records and paper records similar and how are they different?

A key theme that emerged in this section was the acknowledged differences between paper and electronic records and between paper records and electronic systems. In particular the change from paper to electronic records has had implications in terms of how effectively the work of the firm can be carried out (seen in terms of *"speed"* and

"efficiency"). Also of import is the manner in which the processes of document creation and use, as well as the firm's procedures and workflow, have been changed (reflected in the loss of intermediaries, use of layering and the seamlessness of the design process). In this case study numerous differences between paper and electronic records are alluded to throughout the course of the interview. Changes that the respondent mentions that have an impact upon how effectively the work of the firm can be carried out include the speeding up of the process of document creation. Tied in with notions of speed and efficiency is the fact that the nature of the workflow at the firm has been affected by the introduction of electronic systems. The firm no longer uses traditional draftsmen. *"It's changed the, the mix in the office. Now we really only require architects, there is no draftsman position any longer here"*. This change in procedure is seen as an advantage to the architects in terms of the speed, efficiency and also the seamlessness of the process of document creation. By *"seamlessness"* the respondent refers to the act of communicating both in a technical sense and in the sense of communicating a thought or idea.

"You could say we were supposed to do the thinking and the designing and they were to take that information and put [it] in a document that could go out and be bid. So we would communicate with the draftsmen with sketches, by repeatedly marking up their work, redlining their work, ah, to make sure what they interpreted was what they had in mind. That's not necessary any longer, it's just as efficient, more efficient for us to put it on paper as we think about it. We tend to, we design at the computer now, and we design such that these documents form the basis of the contract documents later on. It's more seamless than it used to be".

An intriguing issue raised by the respondent is the fact that in some instances the CAD system has to be dumbed down. This is necessary so that the documents that the system produces do not look too finished.

"In fact the system itself is made in the initial phases to look dumber than it is, because when we do the floor plans, the schematic floor plans, there's always a danger that the

clients will be afraid to make changes if it looks too finished. You can actually set the program to look sloppy, so that the lines don't meet when in fact they do meet. It's called overshoot. You just toggle overshoot on, and the drawing gets messy, it looks like it was done quickly. They're less afraid of changes that way, they don't think it's, we don't want them being shy about it, we want to get it right ".

The respondent does not state whether this is a problem that is unique to the use of the electronic system or whether this was also a problem with paper documents. One might speculate that this is primarily a problem that has developed with the introduction of electronic systems. It may be that by dumbing down the schematic floor plans, the architects are trying to create more of the look of an unfinished paper document ("messy") that their clients may be more used to and better able to work with at this initial phase of the project. In this instance, the ability of the electronic system to create more perfect documents actually works against the needs of the architects by interfering with their ability to have a dialogue with their clients about the design process.

The respondent brings up a particularly interesting difference between the paper and electronic documents in his firm when he raises the issue of the differences that exist in terms of the quality and completeness of the two formats. The respondent raises this issue specifically when talking about e-mailing documents. Deficiencies in the electronic record include a general lack of covering e-mail to contextualize where the e-mail was sent and to whom. The electronic record is also seen as being somewhat more ephemeral in that it is easier to get rid of than a paper record.

"To contrast the two, electronic versus paper, our records for the paper, via the transmittal, indicate what was sent and who, supposedly, received it. The electronic transmissions, [when] done with email, they tend to be sloppy. Very casual information sent with little documentation as to what sent, and when it was sent, unless those files are saved, and right now, that's become a problem for us getting them saved. We think they are, but we then find out they haven't been."

Although electronic documents have helped to change the work processes of the firm, this has not meant that the paper documents have become obsolete. As will also be discussed in the section relating to authenticity, there are still areas of the firm's business where paper documents play an important role. There are instances, for example, where paper documents are the preferred format because of lack of trust in the security of electronic documents and because electronic documents do not appear to have the same legal standing.

This case study also points to other similarities between paper and electronic documents, although not overtly discussed in this manner. A similarity can be seen in instances where both electronic systems and paper records are viewed as sources of inspiration for the work of the respondent. There is not a direct correlation here between electronic and paper documents but instead a point of commonality between a function of an electronic system and a paper record. In regard to electronic systems, the respondent values the ability to experiment with models by calling up ideas in three dimensions. The respondent talks about the ability of the system to "*feed*" him.

"*I was working on a rendering or a design for a, a think tank in the Berkshires. And, I'm struggling trying to get something that I like that works with the site. So right now I'm experimenting, I'm creating models which I don't, I'm not sure I'm picturing in my mind correctly, and I call it up on the screen, again with the CAD program, and look at it 3 dimensionally, and well something clicked this morning, I liked something that was, won't call it serendipity, it, but... So yes, the system does feed, feed me.*"

Paper documents can serve a similar function. In the instance cited in the case study, the respondent refers to how useful trade magazines can be to provide inspiration in the early stages of a design project.

"*Well, in the case of design, if you're stumbling, and you're not inspired, it helps to thumb through the ah, trade magazines, to just get yourself psyched about architecture*"

again. Um so that, in that particular instance, when you're in the early stages of design, yes, you leave the system to, for some inspiration."

b. Authenticity as a theme

The analysis of this case study suggests that there is recognition that authenticity is an issue when it comes to electronic records. A concern for the importance of the record being what it purports to be is largely driven by external forces. The firm's insurance company is restricting the movement of electronic documents outside of the firm because "*electronic documents are so easy to modify*". Documents, once they leave the architect's office, must usually be in paper format. This appears to have led to a tension within the firm between maintaining efficient business practices on the one hand and protecting the firm from litigation on the other. An example cited during the interview is the problems that arise in the firm's dealings with its contractors. The architectural firm requires its contractors to provide drawings of the work they carried out. The contractors often ask for the firm's documents so that they can modify them. The firm's insurance carrier does not want this happening. While internally there are procedures in place to limit access to the electronic systems, e.g., the system is password protected, there seems to be little expectation that any unauthorized changes would be made to the documents while they are still in-house.

Conclusion

Prior to discussing the implications of the case study content analysis, there are a number of caveats that must be raised. The first of these relates to the nature of the interview itself. The Case Study Interview Protocol was constructed with a particular purpose in mind. That is, it was highly structured in order to elicit specific information about aspects of systems and the records that they contain in order to determine authenticity requirements for the long-term preservation of electronic records. The interview protocol does not directly ask respondents to talk about how they view records or the concept of

authenticity. This analysis, therefore, relies upon the respondent elaborating on these topics when answering questions that relate to these issues. As such, there is a strong probability that the themes extracted about records and authenticity do not represent the respondent's complete views on these topics, an aspect that would have required a more open-ended interview format.

A second caveat is that the structured nature of the Case Study Interview Protocol itself meant that many of the questions elicited answers that focused purely on technical concerns. This is particularly true of certain sections (e.g., section 3 and section 5) of the CSIP. This type of technical information does not lend itself to in-depth qualitative content analysis. Finally, this report focuses on an analysis of only one case study. It is not known to what extent the themes extracted from this case study might be comparable to data gathered from other case studies that have yet to be analyzed. It is also likely that as more case studies are analyzed the themes extracted from this case study may undergo further refinement or augmentation.

It appears from a preliminary content analysis of the transcript of the CSIP for case study 27 that the use of this methodology can provide additional insight into the nature of records and authenticity as understood in organizational settings. The findings begin to tell us how organizations view their own records; the factors that influence how records are created and formatted; and how, and to what extent, electronic records and paper records are similar and different. The findings also begin to illuminate how the concept of authenticity is viewed by creating agencies. In particular the findings from this case study shed light on US research questions 1, 3, and 5 as outlined on page two of this report. In this particular case study, the data on records was richer than the data on authenticity. The "talk" regarding authenticity tended to focus more on technical answers to technical questions and as such the data was not as suitable for content analysis. It is quite possible that this finding will hold true for other case studies. This appears to be a reflection of the more open-ended questions that exist in the first section of the CSIP, where respondents are queried about the business activity and procedures, as opposed to the more technical, and closely defined, sections that follow.

In terms of the methodological implications of this report, although content analysis of the CSIP data yielded some interesting results, the study does highlight the need for less structured types of interviews, or the addition of unstructured interview components into the CSIP, in order to elicit more comprehensive and richer data about records and authenticity. The use of more semi-structured or open-ended interview techniques would provide a forum in which the respondent could elaborate more comprehensively on relevant issues. Such techniques would also allow the InterPARES project to gather data that would answer other questions of interest, such as that posed in US question 4. This question seeks to uncover the variations in language used to describe records by record creators, custodians and personnel. This question cannot be answered at the present time because there is no way of knowing to what extent the respondents' terminology, as captured in the CSIP transcripts, has been influenced by the structured nature of the language present in the CSIP itself.

While time-consuming and technically complex, content analysis offers a highly textured means to account for some of the more transparent or implicit issues that researchers encounter in understanding the data provided by those working with records and recordkeeping systems. As InterPARES moves into its second phase, and the closer examination of the nature and production of materials generated by sensory and performance systems, qualitative methods such as content analysis are likely to provide the researchers with considerable additional insight into the viewpoints of the creators and managers of such materials.