

InterPARES Project

International Research on Permanent Authentic Records in Electronic Systems

Authenticity Task Force

**Research Methodology
Statement**

DRAFT

November 7, 2000

TABLE OF CONTENTS

INTRODUCTION.....	1
RESEARCH STAGES.....	1
TEMPLATE FOR ANALYSIS.....	1
GROUNDING THEORY.....	2
CASE STUDIES.....	2
CASE STUDY SELECTION CRITERIA.....	2
ELECTRONIC RECORDS TYPOLOGY.....	4

INTRODUCTION

The InterPARES research has been divided into four complementary domains of inquiry. The goal of the research in the first domain is to identify the elements of electronic records that must be preserved in order to ensure their authenticity over time. To accomplish that goal, the International Team has established the Authenticity Task Force. The specific questions the Authenticity Task Force is expected to address are:

- What are the elements that all electronic records share?
- What are the elements that allow us to differentiate between different types of electronic records?
- Of those elements, which will permit us to verify their authenticity over time?
- Are the 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 created and transmitted?
- Can the 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?

DOMAIN 1 RESEARCH STAGES

The work of the Authenticity Task Force is being accomplished in three steps. The first is to develop a template to guide the analysis of electronic records. The second is to carry out four rounds of case studies of electronic records and electronic systems in order to test the validity of the template. The third step is to establish a typology of electronic records based on authenticity requirements.

TEMPLATE FOR ANALYSIS

The *Template for Analysis* is a decomposition of an electronic record into its constituent elements. It is an idealized, or hypothetical, representation of an electronic record, based on the concepts and principles of diplomatics and archival science. The purpose for developing the *Template* is to identify and define all the known elements of an electronic record and to determine the role each plays in the verification of the record's authenticity.

The elements are drawn from the elements of an electronic record identified in the findings of the UBC Project, i.e., *The Protection of the Integrity of Electronic Records*. The elements have been refined and expanded by utilizing the InterPARES International Team's combined knowledge and experience with various types of electronic records and electronic systems. The *Template* will continue to be refined and adjusted with each round of case studies.

GROUNDING THEORY

To refine and test the effectiveness of the *Template for Analysis* as well as to construct the *Electronic Records Typology*, a form of grounded theory is being used. Grounded theory is a method for discovering concepts and hypotheses and developing theory directly from the data under observation¹ Cases are selected for study "according to their potential for helping to expand on or refine the concepts or theory that have already been developed. Data collection and analysis proceed together".²

CASE STUDIES

Four successive rounds of case studies will be conducted on electronic records, record-generating electronic systems as well as electronic systems that have the potential to generate records. The case studies will test the effectiveness of the *Template for Analysis* as a method for identifying and analyzing types of records according to requirements for authenticity across a range of technological environments and provide the basis on which to construct the *Electronic Records Typology*.

A *Case Study Interview Protocol (CSIP)* has been developed by the Authenticity Task Force to standardize the interview process for all InterPARES case studies as well as to provide data for populating the *Template Element Data Gathering Instrument (TEDGI)*. The Authenticity Task Force will use the *TEDGI* to carry out a diplomatic analysis of the case study data. After each round of case studies, the *CSIP*, *TEDGI* and *Template For Analysis* will be refined as necessary.

CASE STUDY SELECTION CRITERIA

Because a grounded theory approach is being used, theoretical, rather than statistical sampling is being applied in the selection of case studies. Glaser and Strauss describe the process of theoretical sampling as "a process of data collection for generating theory whereby the analyst jointly collects, codes, and analyzes his [sic] data and decides what data to collect next and where to find them, in order to develop his theory as it emerges."³

Accordingly, criteria for selection of case studies were developed for the first two rounds of case studies. Following the International Team's evaluation of these case studies and the analysis of the case study data, the criteria were adjusted.

¹ Barney G. Glaser and Anselm L. Strauss, *The Discovery of Grounded Theory: Strategies For Qualitative Research* (Chicago: Aldine Atherton, 1967), 6-7, 46.

² Steven J. Taylor and Robert Bogden, *Introduction to Qualitative Research Methods: the Search For Meanings*, 2nd ed. (New York: Wiley, 1984), 126.

³ Glaser and Strauss, 45.

Criteria for Selecting Round 1 and 2 Case Studies

Please note that not all the case studies exhibited all of the following criteria. However, each case study was expected to exhibit at least three of them. Research participants prepared a list of candidates and their brief description, including which criteria they met and why. In proposing candidate systems within the same archival institution, research participants were encouraged to ensure diversity in content and type of records (i.e., case studies representing a variety of systems proposed by the same institution.). The Authenticity Task Force then selected case studies that were representative of these criteria (and combinations of criteria) from the candidate lists. In selecting candidate systems from a range of institutions, the Task Force also attempted to include case studies on record-keeping systems performing similar functions (e.g., student registration systems in different universities).

The criteria for selection of round 1 and 2 case studies were as follows:

1. Systems that contain, generate, or have the potential or possibility of generating records.
2. Systems that have gone through one or more migrations.
3. Systems where migration(s) was (were) from one electronic system to another electronic system.
4. Systems for which several aspects of technological context (storage media, system software, application software, data format, schema) was changed, in the course of each migration.
5. Systems for which the pre-migration and the post-migration versions are available and are up and running.
6. Systems for which detailed documentation (design, implementation, migration, metadata) exists.
7. Systems with a diversity of information configurations (e.g., contain both text and images).

Case Study Evaluation Criteria for Round 1 and 2

Following the completion of the first and second round of case studies, the International Team identified and evaluated the status of the electronic systems as actual or potential recordkeeping systems. This identification and evaluation is an essential first step in establishing a typology based on the requirements for authenticity. To accomplish this step, the following questions were addressed for each case study:

1. Does the system contain electronic records?⁴ A record is identified on the following grounds:
 - a) it has a fixed form;⁵
 - b) it has a documentary form;
 - c) at least 3 persons (i.e., author, writer, addressee) are involved in its creation;
 - d) it participates in or supports an action either procedurally or as part of the decision making process;

⁴ A record is any document created (meaning made or received, and set aside either for action or reference) by a physical or juridical person in the course of practical activity as an instrument and by-product of it. An electronic record is a record created in electronic form.

⁵ Fixed form means: (1) the binary content of the record, including indicators of its documentary form, must be stored in a manner that ensures it remains complete and unaltered; and (2) technology must be maintained and procedures defined and enforced to ensure that the content is presented or rendered with the same documentary form it had when it was set aside.

- e) it has a stable content;
- f) it has an identifiable context;
- g) it has an archival bond with other records within or outside the system.

2. Does the system have the potential to create electronic records?
3. Should it be creating electronic records?
4. Is the system itself a record?
5. Where does the creator's presumption of authenticity come from? e.g., procedural controls internal to the electronic system? procedural controls external to the electronic system? technological controls? Specify the precise nature of these controls if they exist.

Criteria for Selecting Round 3 Case Studies

On the basis of the evaluation of the round 1 and 2 case studies and the analysis of the case study data, the criteria for selecting third round case studies have been adjusted. Round 3 case studies will focus on systems known to contain electronic records and will target specific kinds of electronic systems and electronic records. The criteria for selecting round 3 case studies are as follows:

Mandatory Criteria

1. The system contains, or is likely to contain, or should contain electronic records.⁶
2. The system is a live system, i.e., a system that is still being actively used by the creator to carrying out activities.

Desirable Criteria

1. Systems from different hierarchical levels in an organization; specifically, systems containing records that document management and strategic level activities;
2. Systems that contain supporting and narrative records;
3. Systems from the private sector
4. Financial management systems
5. Multimedia systems
6. Computer-aided design (CAD) systems

ELECTRONIC RECORDS TYPOLOGY

The purpose for developing the *Template for Analysis* and testing it through case studies is to define conceptual requirements for authenticity. Two levels of requirements are anticipated, the first level consisting of baseline or threshold requirements applicable to all electronic records and the second consisting of specific requirements associated with distinct types of electronic records. The *Electronic Records Typology* is being developed as an aid to the identification of the requirements.

⁶ The grounds for determining that the system contains records are identified in the *Case Study Evaluation Criteria for Round 1 and 2* (criterion 1) and on the creator's self-reporting that the system contains electronic records.

Ross defines typological analysis as “the structured arrangement of objects into homogeneous correlated groups.” A typology, in turn, is a “representational model” constructed by researchers in the course of typological analysis and serving to “describe a class of objects.”⁷ As Ross explains, there are two basic approaches to designing and implementing a typology. “The first is top-down and the second is bottom-up. In the former approach a researcher begins within the premise that a ‘group of entities’ ...forms a bounded set. Then the researcher attempts to select and define characteristics shared by the material and to determine whether objects/entities proposed as members of the group have the required attributes. In this approach the set becomes equivalent with the type. In the second approach the investigator starts with the objects and proceeds to describe the component elements. The elements are then grouped into attributes and the attributes subsequently grouped into restricted sets. These are shared component types that carry meaning.”⁸

Following the completion of the second round of case studies the Authenticity Task Force has adopted a top-down approach for the initial basic typology (i.e., for the highest level of categorization). The initial typology reflects the four basic categories of records identified by diplomatics. The categories are: *dispositive records* (records whose written form is required and constitutes the essence and substance of the action), *probative records* (records whose written form is required in order to prove that an action has taken place prior to its documentation), *supporting records* (records whose written form is discretionary and are created to provide support to an action and procedurally linked to it), and *narrative records* (records whose written form is discretionary and that do not participate procedurally in the action, but are created as part of the process of setting oneself at work). After the completion of the next round of case studies, it is likely to move to a bottom up approach for the development of sub-categories and, possibly, for the creation of additional primary categories.

⁷ Seamus Ross, “Dress Pins from Anglo-Saxon England: their production and typo-chronological development.” (D.Phil., University of Oxford, Christ Church and Institute of Archaeology, 1992), 9, 68.

⁸ *Ibid.*, 86.