InterPARES 2 Project
International Research on Permanent Authentic Records in Electronic Systems

Diplomatic Analysis

Case Study 12:
Antarctic Treaty Searchable Database

Heather Daly, UBC

January 2006

INTRODUCTION

The Antarctic Treaty Searchable Database was conceived by Paul Berkman, a research professor at the University of California, Santa Barbara, as a means for providing improved public access to documents relating to Antarctic treaties and policies. Initially intended as a tool to help students studying the Antarctic, the database is now being linked to by various government and non-government organizations interested in the Antarctic. The database, instead of relating fields in tables, is developed based on the concept of information granularity,[1] and is thus rather unique in its structure and functional abilities.

The content of the database is received, both in hardcopy and digital formats from a variety of sources, including the U.S. Department of State, the Marine Mammal Commission, the Committee of Environmental Protection, the host nations of the Antarctic Treaty Consultative Meetings, and the Antarctic Treaty Secretariat. Documents received by the creator are assessed using rules he has developed himself. It appears that in most case, documents that are received from the above mentioned sources are not included in whole in the database.[2]

The creator of the Antarctic Treaty Searchable Database considers the entire database to be analogous to a record group. In addition, he considers each information granule within the database to be analogous to a record. The relationships that exist amongst these granules are related to records series. Though the entire database is likened to a record group, it is most frequently referred to as the primary subject of the case study, and as a record in its own right. Accordingly, the following text presents the results of a diplomatic analysis on the entire database as it is described in the case study Final Report.

The purpose of the diplomatic analysis is to assess the status of the identified digital entity as a record. Once the status of the digital entity has been determined, preservation strategies may be proposed by Domain 3.

[1] See the Case Study 12 Final Report for a compete discussion of this granularity.
[2] See pp. 11-12 of the Final Report for discussion of Berkman's rules for compiling database content.

IDENTIFICATION OF RECORD(S)

A record, as defined by the InterPARES glossary, is a document made or received and set aside in the course of a practical activity. A record must also possess all of the following five components, as established by InterPARES 1 research conclusions: fixed content and form, embedded action, archival bond, persons and contexts. The application of the definition of a record to the creator's digital entities is therefore analyzed according to the following parameters:

1. To be identified as a record, the digital entity must possess fixed content and form,[3] and be affixed to a stable medium (or physical carrier).

New content is added to the Antarctic Treaty Searchable Database as necessary, but the creator insists that when new content is added, "it still is possible to generate consistent and reproducible hierarchies across a prescribed set of years for any query" on the database.[4] Each new version of the database is saved and given a new edition number. Fixity of the entities within the database is achieved by the relational schema used to design and develop the database. The database does not use any proprietary software, and is therefore considered by the creator to be easily preserved in its current form over time. Thus, although the database is regularly updated, the versions that are set aside can be said to have fixed content and form. It is unclear in the report where the database is actually hosted, but it is reasonable to assume that it will be affixed to a server that acts as a physical carrier of the content. In addition, each edition is reproduced on webCDservers.

2. A record must also participate in an action, defined as the conscious exercise of will by an officer of the creator or by an external person, aimed to create, maintain, modify or extinguish situations. A record results as an unintended by-product or product of the action.

The Antarctic Treaty Searchable Database does not participate in an action as defined above. The database was conceived by Paul Berkman as a means to provide improved public access to documents related to Antarctic treaties and policies. As such, the database as it exists is not a by-product of Mr. Berkman's actions, but an end product. Its stated purpose is to publicize copies of various documents obtained through other organizations, and it must therefore be considered as a publication, or information resource, rather than as a record in itself. Although the treaties and policy documents are records of the actions in which they participate, the representations of the contents of those records ingested in the database do not participate in the same actions, and cannot be considered as records of them.

[3] The InterPARES1 Authenticity Task Force has defined fixed form as the following: 1) binary content of the record, including indicators of 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 set aside. (See ATF Research Methodology Statement, available at: http://www.interpares.org/documents/interpares_ResearchMethodologyStatement.pdf).

[4] Case Study 12 Final Report, p. 42.

3. A record must possess an archival bond, which is the relationship that links each record to the previous and subsequent record of the same action and, incrementally, to all the records which participate in the same activity. The archival bond is originary (i.e., it comes into existence when a record is made or received and set aside), necessary (i.e., it exists for every record), and determined (i.e., it is characterized by the purpose of the record).

Though the database is certainly linked conceptually to the original documents which are its sources, it cannot be said to possess an archival bond. First, there is no common activity in which the records participate, and thus no real relationship (based on common function and/or participation in the same activity) exists between them. Furthermore, the database was developed to stand on its own as an information source, and though information from copies of original records was used to provide content, the database can be understood without a direct and formal link to those original records. In other words, just as a book requires no other documents for its meaning to be understood or interpreted, the database has been created as a self-contained and fully comprehensible information source in its own right, and therefore neither possesses nor requires an archival bond.[5]

4. Record creation must involve at least three persons, whether or not they explicitly appear in the record itself. These persons are the author, addressee and writer; in the electronic environment, one must also take into account two additional necessary persons: the creator and the originator.

- The record's **author** is the physical or juridical person having the authority and capacity to issue the record or in whose name or by whose command the record has been issued.

  The authors of the database are Paul Berkman and George Morgan.[6] The project was undertaken on their initiative and under their direction.

- The **addressee** the physical or juridical person(s) to whom the record is directed or for whom the record is intended.

  The addressee of the database is the public.

- The **writer** is the physical or juridical person having the authority and capacity to articulate the content of the record.

  The writers of the database are Paul Berkman and George Morgan.

- The **creator** is the person in whose fonds the record exists.

  The creators of the database are Paul Berkman and George Morgan.

[5] It should be noted that metadata associated with copies of the original records are dissociated from those copies as they are ingested into the database, a practice which contributes to a loss of context in the archival sense, and a further erosion of the link between database representations and original documents.

- The **originator** is the person to whom the Internet account issuing or the server holding the record belongs.

  The originator of the database is EvREsearch.[7]

5. Finally, a record must possess an identifiable context, defined as the framework in which the action in which the record participates takes place. The types of context include juridical-administrative, provenancial, procedural, documentary, and technological.

- The **juridical-administrative context** is the legal and organizational system in which the creating body belongs.

  The Antarctic Treaty Searchable Database was conceived by Paul Berkman as a means of providing better access to Antarctic treaty and policy documents. The work was undertaken completely on his own initiative; no legal or government mandate exists for the database.

  Berkman and Morgan, the creators and authors of the database, presumably must answer in some form or another to the National Science Foundation as they have been funded by the NSF through the Division of Undergraduate Education and Office of Polar Programs, and through the National Science Digital Library program. In addition, the creators are subject to U.S. and international copyright and patent laws.

- The **provenancial context** refers to the creating body, its mandate, structure and functions.

  Paul Berkman is a professor in the Bren School of Environmental Science and Management at the University of California, Santa Barbara. George James Morgan, III operates EvREsearch in Columbus, Ohio. The development of the database was Berkman's idea, and grew out of his desire, and the desire of his students, for better access to Antarctic policy and treaty documents. With the help of Morgan, Berkman designed the database with the aim of facilitating knowledge discovery about policies and strategies in the Antarctic. This aim was achieved by creating a unique and highly searchable database that is accessible to anyone with an interest in viewing it.

- The **procedural context** comprises the business procedure in the course of which the record is created.

  The activity described in the Final Report is the creation of the database. Although this activity did not result in a record, it can nonetheless be broken down into its constitutive **procedural phases** as follows:[8]

[7] EvREsearch is the private company with which George Morgan is associated.
[8] The phases of procedure as dictated by Diplomatic Analysis; see Luciana Duranti, *Diplomatics: New Uses for an Old Science* (Lanham, Maryland and London: The Scarecrow Press in association with the Society of American Archivists and the Association of Canadian Archivists, 1998), 115.

a. **Initiative**: the introductory phase of any procedure is "constituted by those acts, written and/or oral, which start the mechanism of the procedure."[9]

   The initiative phase of the procedure of creating the database began with Berkman's desire for better access to treaty and policy documents. He contacted the U.S. Department of State, the depository for treaty documents, but they were not interested in pursuing his request. Therefore, Berkman decided to develop on his own a searchable database of the Antarctic treaty documents.

b. **Inquiry**: this preliminary phase "is constituted by the collection of the elements necessary to evaluate the situation."[10]

   During the inquiry phase of creating the database, parameters are established for collecting content for the database, and for choosing the appropriate level of granularity and the organization of hierarchical displays.

c. **Consultation**: this phase is "constituted by the collection of opinions and advice after all the relevant data has been assembled."[11]

   There is no formal consultation phase, though it can be assumed that consultation occurred between Berkman and Morgan throughout the creation activity.

d. **Deliberation**: this phase is "constituted by the final decision-making."[12]

   After compiling the chosen content, the next step is to implement the appropriate granularity of the documents, create categorical tags for each of the resulting granules, and index the collection of tagged granules. The granules are integrated, and relationships are assessed to determine whether the appropriate collections are included and the organization of the hierarchical levels meet the criteria of the user.

e. **Deliberation control**: this phase is "constituted by the control exercised by a physical or juridical person different from the author of the document embodying the transaction, on the substance of the deliberations and/or on its forms."[13]

   All controls are exercised by the authors of the database. Thus, no formal phase of deliberation control occurs.

[9] Ibid.
[10] Ibid.
[11] Ibid.
[12] Ibid.
[13] Ibid.

f. **Execution**: "the documents created in this phase are the originals of those embodying the transactions."[14] In other words, the execution phase results in the issuing of the first record capable of producing the consequences intended by its author.

   The execution phase of the procedure of creating the database occurs when the database is uploaded to the internet, and is therefore capable of being accessed and searched by the public.

- The **documentary context** is defined as the archival fonds to which a record belongs and its internal structure.

  The "Documentary Context" section of the CS12 Final Report outlines the internal structure of the database, rather than the fonds to which the database belongs. If the database is considered as a record group, this approach may be valid. However, if the database is to be considered as a record itself, which appears to be the intent in the Final Report, more information regarding the creator's recordkeeping system is required to answer this question.

- **The technological context** is defined as the characteristics of the technological components of an electronic computing system in which records are created.

  The database is available online, as well as through a webCDserver. The database uses the Digital Integration System (DIGIN), developed by EvREsearch. DIGIN operates independently from any specific hardware or software, and is written in PERL to provide a stable cross platform programming language that can read and write binary files, as well as to help process very large files. Documents to be incorporated into the database are in ASCII format to ensure that the entire system remains persistently interoperable and usable.

CONCLUSIONS

The above analysis reveals that the *Antarctic Treaty Searchable Database* does not meet all of the requirements of a record and, as such, cannot be considered one. Although the database provides innovative means of access to a variety of interesting and important documents and records, it must be recognized that the representation of those documents and records in the database do not have effects and contexts equivalent to the originals; in fact, the database consists of a collection of simple copies (i.e., transcriptions of the contents) of various treaty and policy documents, purposefully selected and gathered for dissemination. Conceived of as a tool for education and public information, and developed to stand alone as an information resource, the database must be treated as a publication, and preservation strategies developed accordingly.

[14] Ibid., 116.