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

Case Study Proposal
Antarctic Treaty Searchable Database
Focus 1 - Government

Paul Arthur Berkman
The Ohio State University

November 2002

Description of the Case Study Subject

This case study will focus on applications of the *Antarctic Treaty Searchable Database* (http://webhost.nvi.net/aspire) to the organization, cost-effective management and knowledge-discovery of public-domain legal records that are created by governments in general. The *Antarctic Treaty Searchable Database* itself is part of a two-year National Science Foundation project through the National Science, Technology, Engineering, and Mathematics Education Digital Library (NSDL) with the goal of implementing:

> *...a comprehensive, sustainable, single-source digital collection of international environmental and ecosystem policy documents. This searchable digital collection will be derived (with increased granularity) from the 5-volume, 7000-page Marine Mammal Commission Compendium of Selected Treaties, International Agreements, and Other Relevant Documents on Marine Resources, Wildlife, and the Environment that has been compiled since 1993 with input from specialized agencies of the United Nations and United States as well as non-governmental organizations, universities, industry and the public*.

For background, the *Antarctic Treaty Searchable Database* (which was first implemented in 1999) contains all of the measures, decisions, resolutions, recommendations, conventions, protocol, articles, annexes and appendices as well as figures and tables that have been adopted at the ATCM between 1961 and 2001. As an indication of its utility, this searchable database is linked to websites for various national programs (e.g. Australian Antarctic Division, National Academy of Sciences) and international organizations (e.g Scientific Committee on Antarctic Research, International Association of Antarctic Tour Operators and Antarctic Southern Ocean Coalition). More importantly - this searchable database of permanent Antarctic Treaty documents has been linked to the websites for ATCM XXIV (St. Petersburg, Russia, 2001) and ATCM XXV (Warsaw, Poland, 2002) used by the 45 nations in the Antarctic Treaty System.

Historical evaluation of Antarctic environmental policies requires technologies for comprehensively searching through all of the measures that have been adopted at the Antarctic Treaty Consultative Meeting (ATCM) to discover conceptual relationships within and between

documents based on user- defined queries. In brief, the EvREsearch® technologies for *Information Management, Retrieval and Display Systems and Associated Methods* (Maynard 2001, United States Patent and Trademark Office No. 6,175,830) will provide automated solutions for breaking documents into information granules that are tagged directly (Fig. 1). Consequently, each information granule can be comprehensively searched for any term(s) that it contains. Moreover, for any search query, all of the relevant information granules are displayed in an expandable-collapsible hierarchy that comprehensively identifies relationships within and between the granules. Moreover, these objective assessments provide a dynamic framework for quantifying trends and relationships among policy documents that can be used to address user-defined questions.

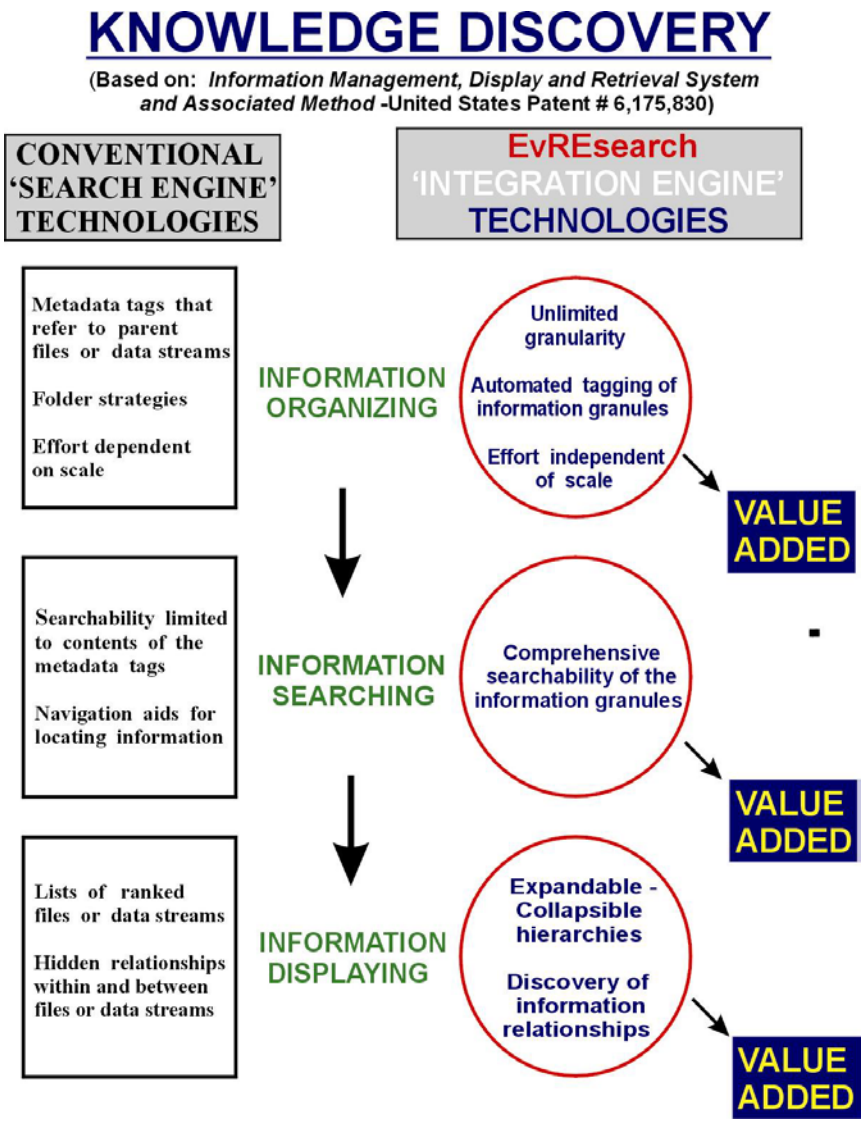

FIGURE 1: Value added by the recently-patented 'integration engine' from EvREsearch LTD in comparison to conventional 'search engine' technologies at different stages of information management and discovery.

Rationale for this Case Study

This case study is of interest to InterPARES 2 because it provides a practical demonstration of the design, implementation and application of a well-defined policy archive of digital records that are being used for governmental and non-governmental purposes. For example, Australian diplomats have requested information about the application of the *Antarctic Treaty Searchable Database* to the design of the newly-approved Antarctic Treaty Secretariat that will be hosted in Argentina. Similarly, the Dutch diplomats have requested information about the application of the *Antarctic Treaty Searchable Database* to identifying and designating "obsolete" policies within the Antarctic Treaty System. These examples illustrate the implications of archiving, repurposing and accessing digital records for decision-making and other 'e-government' activities. Through this case study, the *Antarctic Treaty Searchable Database* will be used to address the following research questions:

1. What are the (a) functional and (b) conceptual challenges of modeling the creation and preservation of digital records?

2. What are the benefits and limitations of accessing, searching and discovering knowledge from digital records that are tagged with current metadata schema?

3. What are the (a) technical and (b) administrative challenges of repurposing public-domain records?

4. How can the increased granularity of digital records facilitate the discovery of relationships within and between policy documents?

Case Study Methodologies

The focus of the methodologies for the *Antarctic Treaty Searchable Database* is to implement a case-study that involves "interactive, dynamic and experiential" records in several directions:

- re-purposing a body of policy documents in a manner that preserves their authenticity, accuracy and reliability;
- interacting with the national government agency that provides access to the policy documents;
- interacting with international institutions that are involved with persistence and access to the digital records for government and public purposes;
- modeling the process of creating and preserving a body of policy documents at higher levels of granularity (i.e. subsets) than originally produced, with the capacity to reconstruct the parent documents as well as new information resources; and
- modeling the process of accessing, organizing and interpreting relationships among policy documents from a searchable database that dynamically generates expandable-collapsible hierarchies for each search query (as opposed to lists that conventionally hide all relationships).

Records Creation and Preservation

The source documents for the *Antarctic Treaty Searchable Database* already exist and are provided by the United States Department of State, which is the Antarctic Treaty depository nation (http://www.state.gov/g/oes/rls/rpts/ant/). The periodically-updated *Handbook of the Antarctic Treaty System* from the Department of State (which is on its 9th version since 1961), is in 12 sections that are each accessible over the internet as a single locked "pdf" file:

Chapter I -- Foreword to the Ninth Edition
Chapter II -- The Antarctic Treaty System: Introduction
Chapter III -- Operation of the Antarctic Treaty System
Chapter IV -- Inspections Under Article VII of the Treaty
Chapter V -- Exchanges of Information, Including Data Management
Chapter VI -- Scientific Cooperation
Chapter VII -- Logistical and Operational Issues
Chapter VIII -- Tourism and Other Non-governmental Activities
Chapter IX -- Conservation of Antarctic Seals: CCAS
Chapter X -- Conservation of Antarctic Marine Living Resources: CCAMLR
Chapter XI -- Regulation of Antarctic Mineral Resource Activities: CRAMRA
Chapter XII -- Protection of the Antarctic Environment

In addition, the sections include short introductory notes from the Department of State "that do not necessarily represent the views of the Consultative Parties to the Antarctic Treaty."

The *Antarctic Treaty Searchable Database* (http://webhost.nvi.net/aspire) currently has over 630 unique information granules (as opposed to 12) that each contain a categorical tag describing its attributes and relative position in the database. These information granules represent all of the policy documents that were formally approved by the Antarctic Treaty Consultative Parties between 1961 and 2001, without the "short introductory notes" from the Department of State that only represent informal summaries from the United States. The database will be updated with documents that were approved at the XXV Antarctic Treaty Consultative Meeting (2002) in Warsaw, Poland.

Nonetheless, there currently is no formal government-sponsored international digital archive of Antarctic Treaty documents. However, there are several digital databases (including the *Handbook of the Antarctic Treaty System* and *Antarctic Treaty Searchable Database*) that exist on either CD-ROM or the internet. After overcoming financial, technical and logistic challenges - the Antarctic Treaty Secretariat in Argentina will serve as the digital archive for these international documents. Technological implementation of the *Antarctic Treaty Searchable Database* – from design to inception to updating – will be modeled with the practical objective of contributing to the Antarctic Treaty Secretariat.

Challenges of contributing to the persistent archive of an international organization (e.g. responses from foreign offices and national delegations) will be noted throughout this case study. Comparisons to other Antarctic Treaty database activities within the United States and other countries also will be noted. These analyses will be considered in light of other archival modeling strategies (e.g. Integrated Definition, IDEF, method as discussed in the Preservation Task Force report from InterPARES 1).

Persistent Archive Analyses

Strategies for implementing a persistent archive, which provides ongoing access to digital records independent of hardware and software architectures, effectively requires abstractions. In part, these abstractions may involve strategies for discerning and utilizing the inherent patterns that exist within files, datasets, databases and data streams. Approaches for working with these information resources directly, as opposed to indirectly through metadata, will be considered.

These comparisons will involve the four primary modules (a break module, an indexing module, a search module and an un-break module) that compose the EvREsearch™ 'integration engine' (Box 1).

BOX 1
EvREsearch™ 'Integration Engine' Modules
(based on USTPO Patent No. 6,175,830)

THE BREAK MODLE operates upon a set of expert rules that define patterns, attributes and subsets within the parent information resource(s). Based on these expert rules, the break module parses each information resource into subsets with categorical tags that are inserted with unique information content, format and organizational attributes for each of the resulting information granules.

THE INDEX MODULE parses through each of the information granules produced by the break module and creates a searchable database with a unique record for each information granule. This searchable database is a type of reverse index, where each record includes an address or location of the corresponding information granule (referenced within each categorical tag) and all strings (sequences of words, numbers or other symbols) it contains.

THE SEARCH MODULE utilizes the search query with operators (e.g., Boolean logic) and terms that are in the same digital form as the contents of the information granules and/or the categorical tags. The search module then finds those database records that have matching terms and, depending upon the selected index schema, displays the search results in expandable-collapsible hierarchies using the information from the categorical tags within the information granules. Alternatively, the search results could be displayed as a conventional list that hides the relationships among the information granules.

THE UN-BREAK MODULE provides an additional tool for assembling information granules to reconstruct contiguous portions of the original information resource(s) as well as to construct new information resources.

Each module preferably is based on an *expert engine operating upon a set of expert rules that define the operation of the individual module.* These expert rules are generated by a person or persons who intimately understand the information resource(s) and can train the information management system. In particular, the flow through from pattern recognition strategies to the 'integration engine' manipulation of information resources to the persistent archiving will be modeled. This information flow will be considered in light of other archival modeling strategies (e.g., IDEF).

Knowledge Discovery and Modeling

Discovery of relationships within and between documents will be compared between 'integration engine' and conventional 'search engine' (Fig. 1) using the *Antarctic Treaty Searchable Database* (http://webhost.nvi.net/aspire). Results that are comprehensively described by expandable-collapsible hierarchies (Fig. 2) will the provide a basis for quantifying relationships between attributes (Figs. 3a,b), in contrast to lists that hide relationships other than the ranking among digital objects.

Proposed Case Study Team:

Dr. Paul Berkman, as the principal investigator on the NSDL grant and manager of the *Antarctic Treaty Searchable Database* project since 1999, will serve as the team leader for this case study. Proposed team members along with their contributions to this case study are:

Anne Gilliand-Swetland (University of California, Los Angeles)

Dr. Swetland and/or her graduate students will serve as the archivist for this case study.

Hans Hofman (Erpanet)

Dr. Hofman will contribute to the modeling aspects of this study, particularly with regard to knowledge discovery strategies associated with the analysis of relationships among tagged information granules.

Richard Marciano (San Diego Supercomputer Center)

Dr. Marciono, as a co-chair of the Policy Group, will assist with all aspects of implementing this case study in the context of managing and utilizing archived digital records in the policy domain.

Reagan Moore (San Diego Supercomputer Center)

Dr. Moore, who is a collaborator with Dr. Berkman on the NSDL project as well as lead investigator on another collaborative project to develop a prototype persistent archive for the National Archives and Records Administration, will provide advice and comments on the implementation of this case study.

Jim Suderman (Archives of Ontario)

Dr. Suderman, who is a member of the Policy Group and lead of another case study in the Policy Group, will interact in the implementation of this case study.

Fraser Taylor and Tracey Lauriault (Carleton College, Canada)

Dr. Taylor and Ms. Lauriault are developing the "Antarctic Cybercartographic Atlas" as an InterPARES 2 case study in the "Scientific Activities" Focus 2 group. Their involvement will provide linkages between Focus 2 and Focus 3 ("Governmental Activities"), which is where the *Antarctic Treaty Searchable Database* case study will be implemented. Moreover, synergy between these case studies will be generated by their conceptual overlap.

Bill Underwood (Georgia Technical Institute)

Dr. Underwood will contribute to the modeling aspects of this study, particularly with regard to knowledge discovery strategies associated with the analysis of relationships among tagged information granules.

FIGURE 2: Expandable collapsible hierarchy that was dynamically generated from the Antarctic Treaty Searchable Database (http://webhost.nvi.net/aspire) for the search term "minor or transitory."

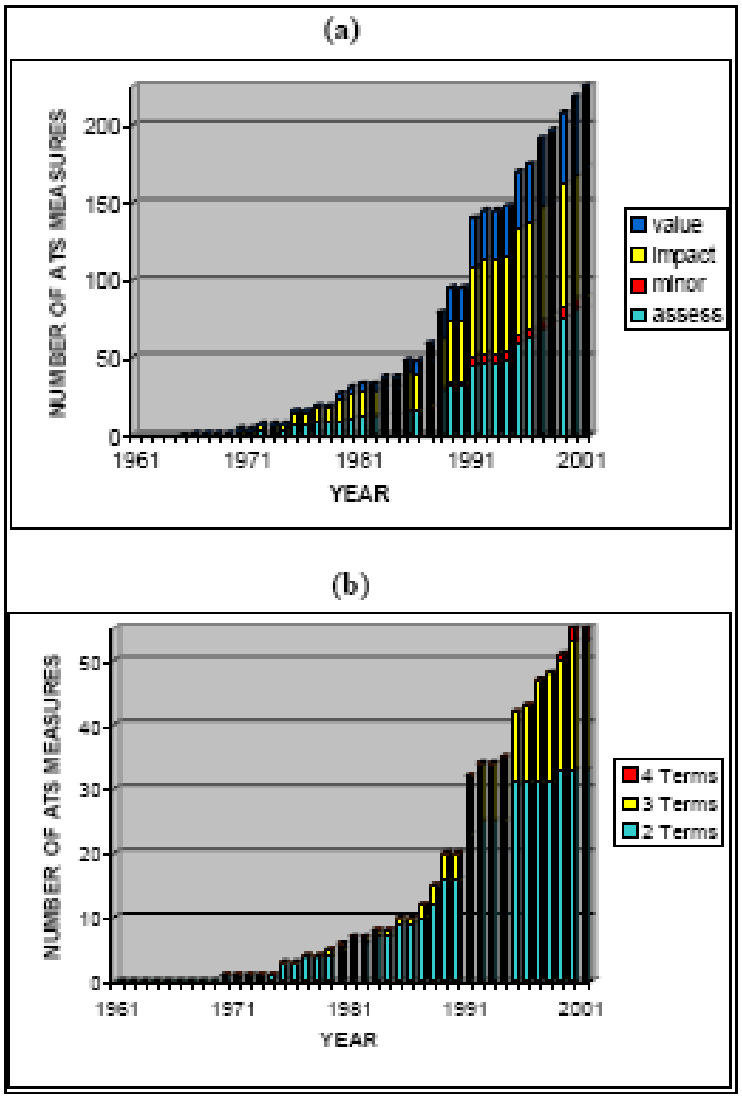

FIGURE 3: Cumulative annual evolution of measures that that were searched with 'integration engine' technologies in the *Antarctic Treaty Searchable Database* (http://webhost.nvi.net/aspire), based on documents that contain: **(a)** the terms of "impact", "assess*" (* is the wildcard search character), "minor or transitory" and "value;" and **(b)** 2, 3 and 4 of these search terms.

Timeline of Case Study:

Winter-Spring 2003	(1) Organize the case study team to assess objectives and methodologies for this project. (2) Begin noting responses to the *Antarctic Treaty Searchable Database* from diplomats and other individuals who are involved with the design of the Antarctic Treaty Secretariat. (3) Begin analyzing information management and knowledge discovery features of the *Antarctic Treaty Searchable Database* in relation to conventional 'search engine' approaches.
Summer-Fall 2003	(1) Update the *Antarctic Treaty Searchable Database* with new documents. (2) Measure time and resource requirements for updating the *Antarctic Treaty Searchable Database* in relation to persistent archival strategies. (3) Prepare report for the Fall 2003 meeting of InterPARES 2.
Winter-Spring 2004	(1) Evaluate progress during 2003 and reconsider methodologies. (2) Based on progress, continue with the assessment of persistent archival strategies. (3) Prepare report on metadata schema in view of strategies for automatically increasing granularity of information resources.
Summer-Fall 2004	(1) Review overall development of Antarctic Treaty Secretariat. (2) Review overall implementation of the NSDL environmental and ecosystem policy collection in light of information management strategies associated with the implementation of the *Antarctic Treaty Searchable Database*. (3) Prepare final report of the case study for presentation at the Fall 2004 meeting of InterPARES 2.