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

Overview

Case Study 06:
Cybercartographic Atlas of Antarctica

Peter Gagné, Université Laval

August 2006

The Creator Context / Activity

Creator: The Geomatics Cartographic Research Centre (GCRC), a research unit in the Department of Geography and Environmental Studies at Carleton University (Ottawa, Canada).

Creator type: Scientific focus / Mixed sphere (university research groups & laboratory). The creator can be considered a mixed creating body because of its particular context. It is related to an organized research unit in a Canadian university.

Juridical context: Intellectual property laws apply. "These include license agreements, use rights to objects and data, and copyright." (FR 18) The creator also adheres to methodologies relating to scientific research and Carleton University rules, regulations and code of ethics. The Cybercartographic Atlas of Antarctica (CAA) Project is also subject to the rules and procedures governing SSHRC grant recipients.

Activity: Creation and maintenance of a dynamic, interactive, Internet-based, open source atlas portraying, exploring and communicating the complexities of the Antarctic continent for education, research and policy purposes.

The CAA seeks to inform and educate users about Antarctica and its relationship to the global environment. The development of the CAA involves the creation of content and the development of technology to render the creator's intent.

This is a nascent business practice—not because of the data that are used, but the way in which they are used. "It is not the intention of the CAA Project to collect substantive new data but to bring together selected existing datasets in a new multimedia form including experimental work with virtual reality and render these in a dynamic and engaging fashion." (FR 1) The final report also speaks of "the use of new map forms that provide dynamic visual representations." (FR 2) This notion is expanded upon with a definition of cyber cartography. "Cybercartography is a new theoretical construct in cartography that uses spatially referenced

information on a wide variety of topics of interest and use to society and expresses them within online digital mapping conventions and newly developed multisensory, multimedia, multimodal and interactive innovations designed to improve searching and learning outcomes for users." (FR 6)

Nature of the partnership
The GCRC currently directs the Cybercartography and New Economy Project, established in 2003, of which the CAA project forms a part. The CAA project was approved in 1999 by the Canadian Committee for Antarctic Research and re-endorsed as a Scientific Committee for Antarctic Research (SCAR) project in 2002. There are formal agreements of cooperation between the GCRC and international organizations and a memorandum of understanding with Natural Resources Canada.

The GCRC receives funding from the Social Sciences and Humanities Research Council of Canada (SSHRC) under the Initiative on the New Economy program. The CAA infrastructure is also funded by the Canada Foundation for Innovation.

The project claims to be a multidisciplinary collaboration. "In collaboration with experts from different fields of science, these data will be used to develop theme-specific modules for use by the general public to facilitate knowledge sharing in multi-disciplinary science…The project includes collaborators from a number of disciplines including Cartography, Geography, Psychology, International Studies, English, Cultural Mediation, Music Studies, Industrial Design and Computer Science." (FR 1)

These collaborators or partners seem to come from within the Carleton University community, as well as external partners. "The research and development of the CAA is being carried out in partnership with a number of research laboratories at Carleton University and in collaboration with an international team of Antarctic scientists and multimedia visualization experts who will share their expertise, laboratories, human resources and data with the project." (FR 4) "Other departments at Carleton University also collaborate informally by providing research opportunities to their students." (FR 4) For example, The CAA user interface (UI) was developed in collaboration with the Human Oriented Technology Laboratory as part of the iterative User Needs Analysis, interface design and human factors testing process. The final UI design was done by a student from the School of Industrial Design and its implementation is being carried out by the technology specialists.

To develop the learning modules incorporated into the CAA, "a partnership has been established with an Ottawa-area group of educators, Students on Ice. Human Oriented Technology Laboratory (HOTLAB) researchers are working with staff at Students on Ice to evaluate user needs." (FR 2)

The CAA Project involves the following partners:
- Argentina – CENPAT
- Australia – Data Centre and Atlas, Australian Antarctic Division
- Canada – Canadian Committee for Antarctic Research
- China – Chinese Antarctic Center of Survey and Mapping

- Germany – IPG, Universite Freiburg
- United Kingdom – British Antarctic Survey
- United States – Scientific Committee on Antarctic Research (SCAR); SCAR Geoscience Standing Scientific Group; Atlas of Antarctic Research and Joint Committee on Antarctic Data Management.

In addition to the partners contributing to the project, the CAA is seeking to be an active partner in the scientific community by sharing its results. "To improve access to data, the Geomatics and Cartographic Research Centre (GCRC) has recently agreed to help develop the Antarctic Data Directory System (ADDS), a project initiated by the Joint Committee on Antarctic Data Management (JCADM). This directory system can facilitate data sharing between the CAA Project and the global Antarctic science community." (FR 2)

Bureaucratic/Organizational Structure
The GCRC is governed by a director and includes 21 members and research associates who are leaders in Geographic Information Processing. The CAA project is part of a larger SSHRC Funded project entitled "Cybercartography and the New Economy" (CANE), which began in January 2003. There are thirteen collaborators and an Advisory Board for the CANE project, as well as a project manager and an assistant office administrator.

The CAA is carried out through a partnership of research laboratories at Carleton University, an international team of Antarctica scientists and multimedia visualization experts. The Scientific Committee for Antarctic Research (SCAR) are expert advisors. There are also numerous stakeholders, data providers and partners from industry, government and non-governmental organizations both in Canada and a number of other countries.

At any given time, there is a PhD candidate who is the lead researcher on the project and one postdoctoral fellow who participate in the management of the CAA. Approximately ten graduate students have contributed research; two students are involved with editing content and two technology specialists with the lead researcher direct production. "The project operates according to a flexible matrix structure composed of task/theme-based cluster groups." (FR 4)

Digital Entities Studied
- A Web-based, online multimedia cyber cartographic atlas that includes:
  - Content modules composed of a variety of digital entities in various media such as text, sound, interactive maps and moving image.
  - External data sources
- A body of records generated by the Human Oriented Technology Laboratory to support development of the navigation and user interface design of the CAA

Documentary Practices Observed
There is no records management or archives program, although the production environment, technologies and normative structures for document generation were selected and put in place with preservation in mind.

The fact that the project involves a nascent business practice has led to some problems with records management. "Given the innovative nature of the project, [the] policies, procedures and standards [to control records creation, maintenance, preservation and use] are in constant need of review and modification." (FR 18)

Records Creation and Maintenance

The CAA project has **processes and procedures** to be followed (e.g., the Author's Toolkit, Designer's Manual, training courses). "An Authors Toolkit is in production that will enable external, authorized content creators to add content directly to the online CAA." (FR 12) Subject specialists also adhere to the practices and procedures of their own disciplines. Individual content creators define their own processes and procedures (i.e., choose their own software), but then need to integrate their content into the CAA, which uses an XML schema. The intent is to design the CAA interface to enable content creators to upload their own content themselves.

"The CAA is being built to OGC [Open Geospatial Consortium] and ISO [International Standardization Organization] **specifications**." (FR 2) In addition, since content files are often initially developed in a software environment selected by the creator, in order for the content to enter the CAA, it must be normalized within the parameters of a special XML schema designed for the project, which consistently contains content, embeds metadata, and includes links to additional content generated in 'project-approved' file formats. There is also a formalized process to get geospatial data into a standard Open Geospatial Consortium framework. "Content modules are provided to technical specialists who run the content of an XML schema file through a compiler to convert it to language that meets the technical parameters set for the project." (FR 12)

As for the description of the digital entities, there are no unique and/or persistent **identifiers**, and there is no formal ID lookup system. The digital objects are identified by a unique combination of a file name and a location in the system. Some objects are identified in databases, with location information included with other metadata.

The **organization** and **aggregation** of digital entities depends on the team member and his or her functions within the project. The content creators determine how the content modules are organized in their personal work environments. The CAA's technical specialists organize the compiled modules within the structure/architecture of the CAA, which is based on technological issues that drive how the various entities are rendered. "At the moment the project does not have either the financial or human resource capacity to write a specification or technical manual." (FR 11) Content creators aggregate at the module level. Each module is an XML-tagged document validated against a schema developed by GCRC. The CAA's technical specialists aggregate by file type.

Any digital object in the CAA must be described by the creator using **metadata** standards adopted by the project. The creators must submit instructions to explain the intended functionality and how the data are to be represented (colour, font, lines, etc.) Some of the maps will have embedded Geographic Markup Language to link to and describe related geo-referenced objects such as images or sounds. "Digital multimedia information objects (e.g., video clips, photos, audio, Web cams, etc.) [must] also be fully referenced and include metadata embedded

into the object and/or accompanying the object and/or referenced as a caption and acknowledged in the bibliography of each content module." (FR 3)

"A multimedia **metadata schema** is being developed. Some of the elements will be embedded within the information objects themselves and some will be linked to the object. This will become a part of the Authors Toolkit, which includes a template of the XML schema that is completed by the content creators." (FR 9) Metadata creation and use is based on the following standards:

- ISO 19115 Geospatial Metadata standard
- GCRC Multimedia Metadata standard (developed by the project)
- Federal Geographic Data Committee (FGDC) metadata standard
- British Antarctic Survey Directory Interchange Format for the Antarctic Digital Database

Changes to the code are captured in Subversion, a source repository system used by the project. "Subversion maintains all code, and all versions of that code are **tracked**....Subversion is…an open source content versioning system (CVS) for use with the most popular operating systems. The Subversion database is backed up regularly…The Authors Toolkit will eventually allow changes to associated metadata to be tracked as well." (FR 15)

"Short-term **maintenance** [of digital entities] during the production phase is primarily the responsibility of the technical specialists at the GCRC" (FR 15)

Recordkeeping and Preservation
The creator keeps the digital entities in the production environment, but there is no actual **recordkeeping** system or preservation strategy. However, The CANE project is mandated by its funding agency SSHRC to archive the research, which includes the CAA. The CAA falls within the Carleton University fonds, as part of the Department of Geography and Environmental Studies series and GCRC sub-series. It is possible that upon completion of the project the CAA will be transferred to SCAR.

A plan is in place to test and assess transferring the CAA to the Carleton University Data Library and the creator expected to develop **processes**, methods, tools and guidelines for archiving and to explore the possibility of expanding metadata tools for this purpose.

The primary investigator for the CANE project directs **preservation** activities and the lead researcher and two technology specialists oversee the production aspects of preservation. **Backup** procedures capture digital objects other than code (see above) at timed intervals. All digital entities are on a central server in the technology specialist's office and all are backed up every six weeks. The backup copy is kept on-site. There is also version control for the digital entities.

External users have no **access** to the production environment of the CAA, but can access the CAA via the World Wide Web. Access to raw data used to create the CAA content may be protected by copyright or use agreements, although most of the data is publicly available through the Antarctic Treaty System.

With regards to the question of **interoperability**, the CANE project (of which the CAA is a part) has "a commitment to an open source philosophy, interoperability and the adoption of and adherence to standards." (FR 4) It claims to be "to the extent currently possible, an open source and standards-based interoperable Internet product." (FR 6) Specifically, the CAA project "gives preference to open-standard products and formats, and to develop procedures, practices and tools that will minimize the loss of data to technological obsolescence." (FR 6)

With this philosophical orientation, there are some practical plans put down for the project. "Should the technology platform of the CAA change, the content of the Atlas would be re-built by re-accessing the XML content modules and processing them anew through a new compiler. Although this method will not protect all information objects included in the CAA…it should facilitate forward **migration** of the essential content, presentation information and intended functionality. Proprietary problems remain with some multimedia formats used..." (FR 16)

Accuracy, Authenticity and Reliability

<u>Accuracy</u>
"The accuracy and reliability of the content and how it is represented are reviewed according to typical academic and professional criteria of the disciplines and individuals involved." (FR 4) In addition, "data are acquired from authoritative sources and are peer-reviewed....Each would have been assessed against the Elements of Spatial Data Quality, which include…positional accuracy, attribute/thematic accuracy [and] semantic accuracy." (FR 14)

<u>Authenticity</u>
"Authenticity in geography is captured in standard metadata as data lineage." (FR 14) In addition, to prevent tampering with and corruption of data, "the online CAA production environment is protected by security measures such as physical security and password protection. Access to the CAA itself is restricted to the CAA's technical specialists." (FR 14) These measures lead the creator to be assured of the authenticity of its digital entities. "In the future, the formalization of module approval will further ensure the authenticity of the CAA's content, as will the implementation of the SCAR's proposed editorial function." (FR 14)

<u>Reliability</u>
"Within the geomatics profession, certain data management practices have been adopted which ensure quality, reliability and authenticity of geospatial datasets. Key elements in the metadata identify characteristics such as scale, accuracy, age, limitations on use and other important facts about the dataset." (FR 14)

Presumably in response to questions of reliability, the final report states, "The CAA will have its own domain name and a trademark with branding." (FR 14)