



# InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems

## Domain 3 Research Questions

### Case Study 06: Cybercartographic Atlas of Antarctica

Sherry Xie, UBC

September 2006

1. **What types of entities does the diplomatic analysis identify in this case study? (i.e., records, publications, data, etc.)**

The diplomatic analysis (DA) of CS06 identifies the multimedia Cybercartographic Atlas of Antarctica (CAA), which is the digital entity studied by the case study, as a publication when it is presented to the public through the Internet.

- 1a. **If there are no records, should there be records? If not, why not?**

There are many records generated in the process of producing the CAA. Even the CAA could become a record when the time comes to set it aside with stable content. For instance, the Report mentions that the Carleton University Library is working with the CAA Project to attempt to archive the CAA at the end of the Project. In this regard, the Atlas is a potential record.

- 1b. **If there should be records, what kinds of records should be created to satisfy the creator's needs (as defined by an archivist)?**

Records have been created to support and document the production activities, which are, however, not the focus of this case study. They include contents and technological constructions for creating the Atlas (items listed at core research question 4a in the final report), operational records, and related research records (see answer to core research question 7 in the final report).

- 1c. **What characteristics of records (as defined by an archivist) are missing yet necessary to preserve these entities?**

As identified by the DA, the following characteristics are missing from the CAA:

- a fixed documentary form,
- a stable content, and
- an archival bond with other records generated in the same activity.

**2. Are the entities reliable? If not, why not?**

Yes. The CAA Project is led by the Geomatics and Cartographic Research Centre (GCRC), an organized research unit (ORU) in the Department of Geography and Environmental Studies at Carleton University, Ottawa, Canada. The CAA is a key deliverable of a larger research project entitled Cybercartography and the New Economy (CANE), a 4-year project commenced in January, 2003 and is funded by the Social Sciences and Humanities Research Council (SSHRC) of Canada under the Initiative on the New Economy (INE) program.

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.

The CAA is being built to the Open Geospatial Consortium (OGC) and ISO specifications. Data are acquired from authoritative sources and are peer-reviewed (e.g. British Antarctic Survey, Scientific Committee on Antarctic Research, scientific and academic journals and books, etc.).

The data reliability is therefore assumed based on the scientific authority of the CAA Project, advanced technologies available to it (e.g., visualization), and procedures followed for data incorporation.

**3. Are the entities accurate? If not, why not?**

The Atlas (and its various components) is accurate to the degree that professional standards and technologies permit. According to the Report, within the geomatics profession, certain data management practices have been adopted which ensure data quality and accuracy. Also, the cooperative disciplines and professions of the CAA Project have established procedures and practices to ensure data quality and accuracy. The Report also suggests that quality measures are dependent on the type of data and their function.

**4. To what degree can the entities be presumed to be authentic, and why?**

The entities are presumed to be authentic by the creator as a result of web server security (which is maintained by an external body) and corporate web page templates. Also, the exhibits must be authentic for accountability purposes. Because the components of every web page are linked to the government, they must not damage the image of the government; and because the exhibits are accessed by the public, they must not misrepresent the creator or its holdings.

## **Benchmark Requirements Supporting the Production of Authentic Copies of Electronic Records (these apply to the creator):**

### **1. Capture of identity and integrity metadata**

- Names of persons concurring in the formation of the record:
  - Author: CAA project team, which is established by the Geomatics and Cartographic Research Centre as a key object of the Cybercartography and New Information Economy project with responsibility to construct the Atlas.
  - Writer: CAA project team
  - Originator: There is no information found in the final report regarding whose or which server will be hosting the online Atlas.
  - Addressee: The general public.
- Name of action or matter: Construction of the Atlas.
- Date(s) of creation and transmission, that is:
  - *chronological date*: It seems currently the Atlas is still under construction. Presumably, its chronological date (including the dates for each new version when new data added to it) will be captured on the day of online publication.
  - *received date*: Not applicable.
  - *archival date*: Not applicable.
  - *transmission date(s)*: Not applicable.
- Expression of archival bond (e.g., classification code, file identifier):  
No archival bond identified between the would-be online publication and other records (i.e., digital components of the Atlas) generated in the process of creating the Atlas. The archived version of the Atlas, on the other hand, could acquire an archival bond at the moment when it is archived.
- Indication of attachments: Not applicable.
- Integrity of the Atlas:
  - Name of handling office: The CAA Project.
  - Name of office of primary responsibility: Same as the handling office.
  - Indication of types of annotations added to the record: Not applicable.
  - Indication of technical modifications: Not applicable.

### **2. Enforcement of access privileges**

Access rights are assigned within the Project based on job duties. According to the Report:

- “Content creators and CAA’s technical specialists have specific job competencies related to the technology or to their subject specializations.”
- “New material would generally be provided by content creators.”
- “Once something is integrated into the CAA, only the CAA’s technical specialists can add or modify online content.”

- “Access to the production environment is currently limited to the CAA’s technical specialists.”
- “External users have no access to the production environment in which the CAA is created.”

3. **Protection against loss and corruption**

There are procedures and considerations in the Project regarding the protection of the digital entities generated for creating the Atlas. According to the Report:

- “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.”
- “Use of open source software will make the CAA more sustainable than if proprietary products were being used. If, for example, PostGIS becomes obsolete, its open source nature requires that future specifications and standards include earlier versions.”
- “Content modules can be re-compiled. The use of XML for the content modules should make the CAA easily translatable (via new compilers) into any future markup languages.”

4. **Protection against media and technology obsolescence**

No information found about procedures established against media deterioration. The Project’s philosophy of using (mostly) open source technologies is their consideration regarding technology obsolescence.

5. **Established documentary forms**

The documentary form for the Atlas is basically decided by the web technologies used by the Project to present the contents online. As stated in the DA, it is not fixed due to the ever changing technologies.

6. **Ability to authenticate records**

Not applicable.

7. **Procedures in place to identify the authoritative record**

Not applicable.

8. **Procedures in place to properly document removal and transfer of records from the creator’s originating system**

Not applicable.

Based on the above examination against the benchmark requirements, the authenticity of the Atlas can be assumed to a great extent.

**Baseline Requirements Supporting the Production of Authentic Copies of Electronic Records** (these apply to the preserver):

Not applicable.

**5. For what purpose(s) are the entities to be preserved?**

The Atlas is created to inform and educate users about Antarctica and its relationship to the global environment. Presumably, this is also the purpose of preserving it.

**6. Has the feasibility of preservation been explored?**

Yes. According to the Report, the digital entities are currently kept in the production environment for the CAA which has no useful recordkeeping features beyond version control and backup capability (see answer to core research question 13 in the final report). The backup is done completely every six weeks, and changes only are backed up daily. The backup copy is kept onsite.

Subversion, a source repository system used by the Project, captures the CAA code (see answer to core research question 13 in the final report). It also captures changes to the code. Subversion maintains all code, and all versions of that code are tracked. The Subversion database is backed up regularly.

Other digital objects which form part of the CAA are not captured by Subversion. Backup procedures capture digital objects other than code at timed intervals.

Of particular importance to the long-term viability of the CAA are the XML-tagged content modules created by the content creators. These are considered the “master” content element. They are processed via a compiler to make them web-ready. 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. While this method will not protect all information objects included in the CAA (i.e., sound, video, Flash, etc.), it should facilitate forward migration of the essential content, presentation information and intended functionality.

Proprietary problems remain with some multimedia formats used in the CAA.

**6a. If yes, what elements and components need to be preserved?**

The overall digital entity is a multimedia cybercartographic atlas, constructed using primarily open source web-based technology including some proprietary multimedia information objects. It is constructed from a wide variety of digital components. To preserve the Atlas needs to preserve all of its contents, technological specifications, and all other supporting records documenting the process of construction.

**7. Which preservation strategies might most usefully be applied, and what are their strengths and weaknesses, including costs and degree of technical difficulty?**

The solution developed by the SDSC (San Diego Supercomputer Center) for the preservation of VanMap (case study 24) may be suitable for the CAA as well.

Have no idea about the cost.

**7a. Which alternative preservation strategies might be applied? What are their strengths and weaknesses, including costs and degree of technical difficulty?**

B1.2. Encapsulation

For the archival version.

B1.4. Conversion

For the online version.

Have no idea about the cost.

**8. What additional information does the preserver need to know to facilitate appraisal and preservation?**

Appraisal is not applicable when the Atlas is under consideration. As for the contents and technical constructions generated for the Atlas, appraisal could be done based on the preserver's understanding about the nature of the activities that generate these digital objects. Other appraisal techniques, such as to remove duplicates and to select authoritative entities can be useful for the appraisal too.

More understanding about these digital entities could be useful for the preserver. Long-term preservation can only be done through a combination of intellectual and technological considerations, which, in turn, requires close collaboration between the creator and the preserver.

**8a. If required information is missing, where should it come from and how should it be made manifest?**

More communications between the creator and the preserver could be helpful.

**9. Are there any policies in place that affect preservation?**

No policies are mentioned in the final report that would affect preservation.

**9a. Are there any policies in place that present obstacles to preservation?**

Not applicable.

**9b. Are there any policies that would need to be put in place to facilitate appraisal and preservation?**

Well-established records management (RM) policies, RM program and recordkeeping system could facilitate the appraisal and preservation of the digital entities generated for the Atlas.