26 Will Today's Internet Maps be Available Tomorrow? The Preservation and Archiving of the Cybercartographic Atlas of Antarctica Through Action Research

Tracey P. Lauriault, D. R. Fraser Taylor, Peter L. Pulsifer

Geomatics and Cartographic Research Centre (GCRC),
Carleton University, Canada

Abstract

Preserving and archiving geomatics data and atlases is a growing and unresolved problem. The Geomatics and Cartographic Research Centre at Carleton University has an ongoing interest in the more effective archiving and preservation of geospatial digital data and has actively participated in a number of forums to address this issue. The Cybercartographic Atlas of Antarctica (CAA) is a collaborative, scientific and innovative open source, open standard and interoperable distributed atlas that includes multimedia, multimodal and multisensory data. The CAA was part of an action research archival Case Study of the International Research on Permanent Authentic Records in Electronic Systems (InterPARES 2) Project. Archival issues in geomatics, the action research Case Study process, and results are discussed and reported upon in this chapter. The chapter concludes with some recommendations and reflections on preserving and archiving Internet maps. The central argument is that today's Internet mapmakers and cartographers must include preservation strategies at the point of creation not after the fact to ensure that their digital artifacts will be available to tomorrow's users. If this is not done effectively much of our digital mapping heritage will continue to be lost.

26.1 Introduction

The use of complex cartographic digital data in a variety of forms is now commonplace through the Internet. Internet maps have improved as a result of an increasing variety of Web 2.0 tools such as blogs, WIKIs, VoIP and video conferencing, all of which have facilitated research and collaboration from disparate geographical locations often in real time. Data have become more accessible from science data portals, catalogues, libraries and gateways with specific metadata standards and

classification systems, that ease discovery and use. This situation is helping create an increasing volume of maps and Atlases on the Internet. But a major problem remains largely unresolved: the long-term preservation of Internet maps and atlases. This issue is addressed here using the Cybercartographic Atlas of Antarctica as a case study. This online atlas is multimodal, multimedia, interactive, dynamic, interoperable and multi-authored and uses open source standards and specifications. It thus poses considerable challenges, for preservation and archiving.

Currently, insufficient attention is being given by most data producers to effective preservation and archiving and, as a result, a great deal of digital information is being lost – much of it irretrievably. A major study of the key issues is being conducted by the International Research on Permanent Authentic Records in Electronic Systems (InterPARES 2) Project. This is an interdisciplinary research project that studies the life-cycle of records created in complex digital environments, from creation to permanent preservation. The focus is on records produced in the course of artistic, scientific and e-government activities. The Cybercartographic Atlas of Antarctica (CAA) is one of the scientific case studies for InterPARES 2. Interaction with the archival community and other content specialists from different disciplines through a participatory research process has been invaluable in finding solutions to what are common problems. This chapter will illustrate how a complex Internet atlas is being preserved and archived. It begins with a brief review of the changing technological and social context of cybercartography followed by a discussion of current data and map archival initiatives both in Canada and internationally. The chapter concludes with some recommendations and reflections on preserving and archiving Internet maps. The central argument is that today's Internet mapmakers and cartographers must include preservation strategies at the point of creation not after the fact to ensure that their digital artifacts will be available to tomorrow's users. If this is not done effectively much of our digital mapping heritage will continue to be lost.

26.2 A Brave New Map Making World

Before considering archival and preservation issues a brief consideration of the changing technological and societal context of cybercartographic atlases is required. Taylor (2003) introduced the theory and practice of cybercartography. Three years later the research supported by the Social Science and Humanities Research Council (SSHRC) of Canada conducted at the Geomatics and Cartographic Research Centre (GCRC) at Carleton University in Ottawa, Canada, has advanced both the theory and practice of cybercartography, which is more fully, described elsewhere (Taylor 2005, Taylor and Caquard 2006). The Cybercartographic Atlas of Antarctica (CAA) is one product of that research (Pulsifer 2006). One very significant research result has been the creation of the Nunaliit Cybercartographic Framework more fully

described in the chapter by Pulsifer *et al.*, in this volume. The open source, standards and specifications "aims to make it easy for anyone to build a cybercartographic atlas – telling stories and exploring the relationships between space, time, knowledge, and our senses" (Nunaliit 2006a). This allows non-specialists to create their own atlas modules and makes such atlases truly living entities.

Cybercartographic atlases are unique in their design and implementation and are developing concurrently with Web 2.0 technologies. Web 2.0 is "creating network effects through an architecture of participation" (O'Reilly 2005) and embodies "a second-generation of Internet-based services that let people collaborate and share information online in new ways" (Wikipedia 2006). The implications of these emerging technologies and services, such as Google Maps, for Internet cartography in general, and cybercartography in particular, are considerable and are illustrated in other chapters of this volume. Data collection devices are also in the process of converging: phones capture location, take pictures, record audio and send these media objects directly to a Web page. Many Web 2.0 principles and tools are used in the design of cybercartographic atlases and also in the organization of the research and production processes which involves a flexible structure of interdisciplinary research clusters that encourages collaboration (Lauriault and Taylor 2005). These innovations are becoming mainstream and today some national mapping agencies and geospatial data infrastructures are embracing open source approaches (Lee 2006, GeoConnections 2004) and these are interactive, dynamic national atlases, which are adopting open source (Atlas of Canada 2006) and interoperable standards (e. g. Open Geospatial Consortium). The nature of maps and mapping through the Internet has radically changed.

The transdisciplinary, collaborative and partnership environment within which cybercartographic atlas technology, theory and content are being generated reflect the fast paced changes in the Internet over the last few years. Cybercartographic atlases are not only innovative, they are robust digital artifacts that combine elegant open source code with aesthetic cartography; they are also useable scientific artifacts that integrate vast amounts of data to represent complex social and scientific issues in new and meaningful ways. They help us understand and explain the world we live in and help formulate decisions on how to better manage our resources, understand our environment and meet societal needs over time. These atlases are important cartographic artifacts that build on existing technologies and which will continue to evolve as new content is added and as innovations in the Internet continue. This is an exciting period in the evolution of cartography and the Atlas of Antarctica in particular is a product of some of the most intensely collaborative science of our day. In the excitement of these new developments, attention tends to be focused on the creation of these new artifacts and on production processes as several chapters in this volume illustrate. However, if future generations are to understand these processes and be able to use their products other issues must be considered.

26.3 Archiving Internet Maps and Associated Data

The technological, institutional and organizational issues related to the long-term preservation of Internet maps, atlases and their associated data remain under-researched and largely unresolved. The basic digital data on which cybercarto-graphic and other forms of Internet mapping depend are rarely being effectively archived and preserved and as a result much is being lost, some of it permanently. John Roeder, a researcher on the InterPARES 2 project, discovered that one fifth of the data generated by the 1976 Viking exploration of Mars (Cook 1995, Harvey 2000), the entire 1960 U.S. Census (Waters and Garrett 1996), and the works of nearly half of composers (Longton 2005) and one-quarter of digital photographers (Bushey 2005) have been lost or threatened by technological obsolescence or inadequate preservation strategies. It has been argued that "in archiving terms the last quarter of the 20th century has some similarities to the dark ages. Only fragments or written descriptions of the digital maps produced exist. The originals have disappeared or can no longer be accessed" (Taylor, Lauriault and Pulsifer 2005). It has also been noted that "indeed digital technology is responsible for much of the loss as storage technology has given a false sense of security against loss and obsolescence" (Strong and Leach 2005). Internet cartographers are not alone in struggling with the challenges of trying to keep abreast with emerging technologies, meeting deadlines, collaborating and fulfilling milestones in tight budget environments. Archiving, when it does take place, is often an afterthought, an activity that happens when a project ends. Today, we have the ability to create more maps that include more information than ever before, yet we may have fewer maps and fewer atlases available to future generations than during the paper mapmaking era.

Even when the problems and challenges of archiving these data and digital artifacts are identified, the institutional environment is often not conducive to the systematic action required to address the problem. In Canada, for example, the Library and Archives Canada (LAC) is not ready to systematically archive research data, digital maps or complex cybercartographic atlases. LAC policies and guidelines for carto-graphic material primarily address paper maps. The LAC handbook for records and information management *Managing Cartographic, Architectural and Engineering Records in the Government of Canada* (2006a) makes only passing reference to digital maps such as "the National Archives acquires geomatic systems" (LAC 2006a) and "geomatic records include geomatic systems, discs, CD-ROMs and other cartographic material in electronic formats" (LAC 2006a). The Handbook refers the reader to the Canadian Committee on Archival Description *Rules for Archival Description Chapter 5* (CCAD 2001) for information pertaining to standards and practices for cartographic records. The *Rules* primarily address paper maps while general issues pertaining to digital databases and program description are covered

in *Chapter 9: Records in Electronic Form* (CCAD 2003). The LAC *Guidelines for Computer File Types, Interchange Formats and Information Standards* (2006b) does make reference to some geomatics specific file types. But adequate guidelines for the kind of digital cartography emerging on the Internet described earlier in this chapter and in other chapters in this book do not exist.

Currently, Library and Archives of Canada (LAC) does not have a digital data archive with the explicit mandate to ingest the results of Internet mapping or scientific endeavors. The Social Sciences and Humanities Research Council (SSHRC) is explicit that "all research data collected with the use of SSHRC funds must be preserved and made available for use by others within a reasonable period of time" (SSHRC 2002a). The same policy recommends researchers "ask your university library or data service if it can archive the data" (SSHRC 2002a) and if it cannot SSHRC provides a list of possible universities that can assist. The recommended data libraries are not archives, and do not have the technical nor human resources to archive Internet maps, their data nor cybercartographic atlases. Internationally there are some social science data archives such as the UK Data Archive (UKDA) but there does not seem to be any natural or physical science data nor digital maparchives in any national archive institution. The US National Science Foundation Cyberinfrastructure Project does however look very promising.

GeoConnections is the Government of Canada agency mandated to deliver the Canadian Geospatial Data Infrastructure (CGDI). GeoConnections conducted a study on *Archiving, Management and Preservation of Geospatial Data* (2005) which provided a well rounded analysis of preservation issues in the field of cartography such as: technological obsolescence; formats; storage technologies; temporal management; and metadata. The study also provides a list of technological preservation solutions with their associated advantages and disadvantages. It concluded with a list of proposed institutional and national actions. Phase II of the GeoConnections Program includes archiving as an information management strategy but these details are still under development.

A number of studies, reports and committees have made high level recommendations and provided strategies for improving the archiving of digital data in Canada. All of these recognize the poor state of Canada's digital data resources. The *SSHRC National Data Archive Consultation* (2002b) report discussed the preservation of data created in the course of state funded research projects, such as the Cybercartography and the New Economy project. The consultation identified important institutions, infrastructures, management frameworks and data creators and calls for the creation of a national research data archive. The report *Toward a National Digital Information Strategy: Mapping the Current Situation in Canada* (MacDonald and Shearer 2005) indicates that "the stewardship of digital information produced in Canada is disparate and uncoordinated" and "the area of digital preservation, which involves extremely complex processes at both the organizational

and technical levels, comprehensive strategies are not yet being employed. Many feel that much of the digital information being created today will be lost forever" (MacDonald and Shearer 2005). The Final Report of the *National Consultation on Access to Scientific Data* (Strong and Leach 2004) developed in partnership with the National Research Council Canada (NRC), the Canada Foundation for Innovation (CFI), Canadian Institutes of Health Research and Natural Sciences and Engineering Research Council (NSERC) expressed the concern that "the loss of data, both as national assets and definitive longitudinal baselines for the measurement of changes overtime" (2). This report also provides a comprehensive list of recommendations that include ethics, copyright, human resources and education, reward structures and resources to name a few, toward the creation of a national digital data strategy and archive. The CODATA Working Group on Archiving Scientific Data has been holding symposia and workshops on the topic and the Canadian National Committee for CODATA has been active in documenting and reporting scientific data activities. Currently, there are meetings toward a Canadian Digital Information Strategy with science as one of the themes and it is hoped that this will yield positive results. In December 2006, Library and Archives Canada hosted a National Summit on a Canadian Digital Information Strategy (CDIS). The challenges of the new Web 2.0 social computing environment, open access, the preservation of scientific data, interoperability and licensing among numerous other topics were discussed. A report is in its Draft form was released for public consultation in autumn of 2007. Finally, on a more technical side, the Open Geospatial Consortium (OGC) has developed a new Data Preservation Working Group that is currently seeking examples as part of their test bed process. To date none of the above reports, committees or recommendations have resulted in the creation of a national science or geomatics data archive nor have new policies yet been implemented. Archiving of scientific and geomatics data are technologically complex, however, the greatest obstacles are not technology, techniques or know how. The greatest obstacles are the lack of institutional will and the financial resources needed to implement what is already known, to develop transdisciplinary working groups and to finance research on issues that still need to be resolved. Unfortunately, the situation in Canada is not unique. Many nations and agencies have identified the problems: few, if any, have implemented the solutions suggested although the studies discussed in the Canadian context, the US Cyberinfrastructure project, the CDIS and the OGC Preservation WG are steps in the right direction.

26.4 The Antarctic Context

Antarctica is considered to be the continent of science, and "Antarctica plays a key role in many scientific questions – of which those related to global climate change

are probably the most prominent examples" (AntSDI 2006a). Antarctica is not a sovereign state, and the Antarctic continent and surrounding region is governed by a consensus based system known as the Antarctic Treaty System (ATS) (Pulsifer, Caquard and Taylor 2006). Antarctica does not have a central organization managing its geographic information or knowledge; however the facilitation, organization, management and development of information produced by the science community are carried out by the Experts Group on Geospatial Information – Geographic Information and the Joint Committee on Antarctica Data Management (JCADM) of the Scientific Committee on Antarctic Research (SCAR). The Cybercartographic Atlas of Antarctica (CAA) is a SCAR endorsed project. Data for the CAA are produced and maintained by a variety of organizations such as national mapping agencies, research institutions, scientific programs, research projects and individuals who deposit or make these discoverable in the Antarctica Spatial Data Infrastructure (AntSDI) which uses "the open standards based approach, the focus on enabling specifications and procedures and capacity building within the information community" (2006b). Scientific data are only archived if the creation and maintenance organizations have the mandate to do so. CAA researchers are active in these organizations and also participate in the International Polar Year (IPY) research activities. IPY (2007) is dedicated to intense interdisciplinary study of the Earth's Polar Regions and is co-sponsored by the International Council of Scientific Unions (ICSU) and the World Meteorological Organization (WMO). IPY participants have "developed specific recommendations on engaging archives, data discovery and access methods, standards and interoperability, and ways to ensure that all IPY data are captured and readily available" (IPY 2006a) and preservation is a principle in the IPY Data policy (IPY 2006b). Implementation of these policies has yet to be worked out. The CAA, SCAR and IPY are examples of global collaborative science in action.

The Cybercartographic Atlas of Antarctica (CAA) Project aims to develop an online atlas using primarily open source and interoperable technologies portraying, exploring and communicating the complexities of the Antarctic continent for education, research and policy purposes. The atlas is modular and interoperable (Pulsifer and Taylor 2005; Parush *et al.*, 2006). The CAA highlights the global importance of Antarctica as the continent of science and peace. Data from a number of international sources are incorporated in the CAA and are used to develop theme specific modules for use by the general public to facilitate knowledge sharing in multidisciplinary science (see Pulsifer et. al in this volume).

One of the authors of this chapter has been involved with digital archiving initiatives in Canada and internationally for a number of years and as a result of this experience the Cybercartography and the New Economy Project specifically included archiving as an integral part of the research. As a result of the involvement with the archival initiatives including some of the studies described earlier

the Cybercartographic Atlas of Antarctica (CAA) became a case study in the InterPARES 2 Project mentioned in the introduction. This was a deliberate attempt to seek help and guidance from the international archiving community with the preservation and archival challenges of cybercartography in particular and Internet mapping in general.

26.5 The International Research on Permanent Authentic Records in Electronic Systems (InterPARES 2) Project

The International Research on Permanent Authentic Records in Electronic Systems (InterPARES 2) is a collaborative project funded by the Social Sciences and Humanities Research Council's (SSHRC) of Canada, the US National Historical Publications and Records Commission (NHPRC) and the US National Science Foundation (NSF) at the University of British Columbia. The InterPARES 2 Project aims to develop and articulate the concepts, principles, criteria and methods to ensure the creation and maintenance of accurate and reliable records and the long-term preservation of authentic records in the context of activities that are conducted using experiential, interactive and dynamic computer technology. The InterPARES 2 Project is interdisciplinary since its goal and objectives can only be achieved through the contribution of several disciplines. The project espouses no a priori epistemological perspective or intellectual definitions and was conceived to work as a 'layered knowledge' environment. Working groups were established to identify appropriate research approaches in a multidisciplinary context with the expectation that this process would develop entirely new knowledge. Methods include surveys, case studies, modeling, prototyping, diplomatic, archival and text analysis.

The Cybercartographic Atlas of Antarctica (CAA) is a Case Study within the Scientific Records Focus and the Creation and Maintenance Domain of the InterPARES 2 project.

26.6 Participatory Action Research

The CAA Case Study fulfills two important functions. It is an action research process for the GCRC and it is a discipline specific case study for the InterPARES 2 Project. GCRC's research focus is on the application of geomatics to the understanding of socio-economic issues, the theory of cartography and cartographic education in an international context. GCRC has been involved in a variety of archiving initiatives but does not have professional archiving expertise. GCRC's participation in InterPARES 2 is an example of participatory action research where researchers

"deliberately aim to contribute to two domains, in the organization community in which they are a member" and "actively participate in the very phenomenon they [InterPARES 2] are studying in order to develop organizational competencies of individual actors, to develop learning capacities of the organization as a whole, and thereby to influence the course of events" (Locke 2001:14). Researchers at GCRC are stakeholders, study subjects and research participants. GCRC does not aim to develop a general law or approach to resolve a particular situation, nor is there a case from which a generalized theory can be developed, proved or disproved but GCRC wants to archive its Atlases. The InterPARES 2 Case Study process introduces archiving and related issues and by doing so educates all GCRC researchers and technology specialists about archiving issues and the importance of considering these in the development and implementation of cybercartographic atlases. It is a cyclical and multi-phased inquiry process where all concerned benefit from the interaction. The intent on both sides, GCRC and InterPARES 2 is the creation of a self-generating learning process, organizational change and self development (Locke 2001). The InterPARES 2 instrumental case study approach facilitates the examination of the dynamics of the creation and maintenances of the CAA, knowledge of archival issues in the field of geomatics, gains insight into this discipline and advances archival theory and practice. The CAA is an interesting Case Study since it is an example of how collaborative and distributed science projects operate.

26.7 The Cybercartographic Atlas of Antarctica InterPARES 2 Case Study

The primary information-gathering tool for this case study was the InterPARES 2 case study questionnaire, comprising 23 questions (InterPARES 2 2003). Two sets of semi-structured interviews at two different development stages of the CAA Project were conducted. The respondents were the primary investigator, a technical specialist, a researcher on the CAA and a researcher involved in other aspects of cybercartographic research. Respondents also reviewed the final responses to the 23 questions included in the report (Lauriault and Hackett 2005).

Two interviewers were involved in data gathering, Yvette Hackett, Library and Archives Canada and Tracey P. Lauriault, PhD Candidate, GCRC. The questionnaire and semi-structured interview process were approved by the Carleton University Ethics Committee and confidentiality and data storage methodologies were adhered to for interview audio files and transcription documents.

The report is structured to answer questions of interest to the archival community. Case Study researchers were asked to provide a description of the context of the objects under investigation. This included: the juridical and administrative context

of how the CAA is created; the provenancial elements of the project; procedural context related to the creation of the CAA; the documentary context around the creation of the Atlas and the technological context. The process of answering these questions helped GCRC make explicit some implicit, tacit and taken for granted assumptions in the production of the CAA.

Case study researchers then proceeded to ask the research subjects the 23 research questions developed by the InterPARES 2 research community (InterPARES 2 2003). These questions inquired about the development process and procedures, digital entities and how they relate to each other, authenticity issues, backup procedures, accuracy, reliability and authenticity and documentary processes (Lauriault and Hackett 2005). Yvette Hackett brought with her expertise from the field of digital archiving, was previously a researcher in the InterPARES 1 project and currently is an Institutional Researcher in InterPARES 2. Having an interviewer with this level of expertise was fortuitous since she was able to make the research questions more understandable to non-archivist respondents.

Respondents gained considerable insight from this process. Reflections on their production processes from an entirely different perspective identified both shortcomings and strengths in archival terms. For example the archival research enquiry revealed that: the CAA has no persistent identifiers, and there is no formal ID lookup system; digital objects are identified by a unique combination of a file name and a location in the system; some objects are identified in a database; some of the maps have embedded Geographic Markup Language (GML) to link to and describe related geo-referenced objects such as images or sounds; a multimedia metadata schema is being developed where some of the elements will be embedded within the information objects themselves and some will be linked to the object and these will become part of the Authors Toolkit, which includes a template of the XML schema which is completed by the content creators (Lauriault and Hackett 2005). This led to the realization by the Atlas production team that a focus on refining the Authors' Toolkit, developing a designers manual, recording training courses, and capturing some of the content in the CAAs forums and Wikis would assist with transferring knowledge about the technological processes used to create the CAA which will assist archivists with the activity of preserving the CAA in the form that the creators intended it to be which would not otherwise be possible.

Reliability and authenticity are important concepts in the field of geomatics which has well defined professional practices in place. The CAA relies on data sources from authoritative and reputable scientific organizations and data sets are accompanied by good metadata. Reliability is therefore ensured by the quality of the base data used and by methods applied by content creators. The online CAA production environment is protected by security measures such as physical security and password protection. Some metadata will be embedded into information objects of the CAA and additional metadata will be linked to the objects at the module level.

GCRC researchers oversee the modules, providing additional checks and balances over the content and access to the CAA itself is restricted to the CAA's technical specialists. Further, an additional authenticity stamp is provided as the CAA will have its own domain name, and a trademark with branding ensuring that records are authentic which is important for archival purposes.

Once material is integrated into the CAA, only the CAA's technical specialists can add or modify online content. Also, 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 from Tigris.org - it is an open source content versioning system (CVS) for use with most popular operating system. The Subversion database is backed up regularly. Other digital objects which form part of the CAA are not captured by Subversion. The Authors Toolkit will eventually allow changes to associated metadata to be tracked as well.

The CAA creators were asked to describe which elements of the CAA they consider to be records for archival purposes. Atlas creators consider every digital object connected to the CAA to be a record. The processes used to create it and the activities to disseminate it are considered to be important. But, 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, 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. While not perfect, since there is no information at this time about how the digital objects are organized within Subversion, or within Confluence, the overall combination of practices and processes are considered to be quite robust and should ensure preservability.

The process fell a little short when it came to a backup system. The CAA researchers keep the digital entities under examination but these are not part of a recordkeeping system. They are currently kept in the production environment for the CAA, which has no recordkeeping features beyond version control and backup capability. The backup was done completely every 6 weeks, and changes only are backed up daily. The backup copy is kept offsite.

The CAA production process is considered adequate to meet the challenge of technological obsolescence. The 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. Migration should be easier since the technology evolves but does not become obsolete. Also, 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.

The CAA has good metadata practices in place and is developing a customized multimedia metadata standard (Zhou 2005). Ingested data are accompanied by FGDC and/or British Antarctic Survey DIF or Directory Interchange Format for the Antarctic Digital Database. The CAA also adheres to OGC interoperability specifications (2006) and the International Standards Organization 19115 Geomatics Standards (2003). The InterPARES 2 Description Cross Domain Team has mapped ISO 19115 into their MADRAS Archival Description Registry and Analysis System (InterPARES 2 USA 2006) for archival evaluation. Legally, morally and ethically the CAA includes standard intellectual property concerns while the terms of the Antarctic Treaty makes access to data much easier. Much of the data used in the creation of the CAA - a non commercial research product - can be used at no cost as part of the Antarctic Treaty System. The CAA also follows typical license agreements, use rights to objects and data, and copyright. The atlas itself includes use caveats and disclaimers (e.g. the CAA is intended for information, not navigation purposes) and the project must adhere to the requirements of the funding agency, Carleton University and the Nunaliit License (Nunaliit 2006 b).

Overall the case study revealed that many of the CAA development processes are adequate for preservation and archival purposes. The focus on interoperability, adherence to open source standards, the professional practice of the technology specialists and geomatics experts, metadata descriptions, schemas, high data quality standards and the innovative Nunaliit Cybercartographic Atlas Framework are all very strong features of the atlas in terms of its suitability for preservation and archiving (see Pulsifer et. al this volume for theoretical and technical details related to the production of the CAA). The answers given by CAA researchers provided InterPARES 2 with an increased understanding of cybercartography and collaborative science, practices and processes. InterPARES 2 researchers also learned that the fields of geomatics and producers of scientific data have very rigorous metadata descriptions, excellent standards, and professional data gathering and maintenance procedures that can be used as models for records created in the arts and in e-government. Accuracy, reliability, authenticity, good documentation and metadata are well established in both the geomatics and open source communities and these are considered to be good scientific practice.

The greatest challenges limiting the long-term preservation of the CAA at the moment are neither technological nor procedural. The greatest roadblock is simply the fact that no Canadian archival institution is currently in a position to ingest the CAA. The GCRC is having ongoing discussions with members of the Data Library at Carleton University, the Head Librarian and VP Research to attempt to archive the CAA as required by SSHRC. At the moment the Data Library does not have the

technological, policy and human resource capacity to archive the CAA. Discussions are also ongoing with the San Diego Centre for Supercomputing as they are part of the InterPARES 2 Research Team, InterPARES 2 Researchers, Library and Archives Canada, the National Research Council and GeoConnections. The CAA project is nearing completion, and no explicit transfer plan is yet in place. GCRC is fortunate to be located in the nation's capital as it has ready access to Canada's top officials in key organizations to assist with resolving this problem.

26.8 Interdisciplinary Debates

The discipline and practice of archiving has a long and distinguished history and well established traditions, theories, methods and practices. These are central to the InterPARES 2 Project. This research project has made explicit efforts to involve academics, professionals and practitioners from a variety of other disciplines specifically the arts, sciences and e-government. Not surprisingly each of these disciplines has its own traditions, which can conflict with those of archival science. The interdisciplinary research process is a challenging mutual learning process and often a contentious one. One of the most contentious issues, even within archival science is the definition of the term record. For InterPARES 2 a record is "a document made or received in the course of a practical activity as an instrument or a by-product of such activity, and set aside for action or reference" (InterPARES 2 Glossary 2006) and the establishment of these elements is based on the archival theory of diplomatic analysis. When an analysis of the CAA case study was carried out using the archival approach of diplomatics the archival researchers concluded that (Xie 2005):

> the online Atlas only possesses some of the elements … that are necessary to be considered as a record. The fact that both its content and documentary form are subject to continuous change and it does not possess an archival bond makes that the Atlas only partially satisfies the definition of record. This diplomatic analysis thus establishes the status of the Atlas as a publication and a potential record, rather than a record.

Further, the CAA

> has the potential to become a record. When the time comes to set the Atlas aside and in so doing, to stabilize its content and fix its documentary form, it becomes a record. This setting aside gives the "retired" Atlas an archival bond with other records that are organized in an identifiable documentary context. This setting aside also indicates that it ever participated in one action and thus now is filed with other records that were generated in the same action for further actions or reference. At this moment, the Atlas possesses all five components that are necessary for it to a record.

The five indispensable elements required for a digital entity to be a record are as follows: fixed content and form, embedded action, archival bond, persons and contexts (see InterPARES 2 Terminology Dbase and Glossary at http://interpares. org/ip2/ip2_terminology_db.cfm). In other words, since the CAA is a living document explicitly designed to allow for data to change and information to be added means that the CAA does not qualify as a record in archival terms. To become a record it must be something that is fixed in time and space to fit. To fit this definition of a record for archival purposes one possible approach is to create "records" by taking "snapshots" of the atlas over time. This is, however, a far from perfect solution and is both expensive and beyond the capacity of most institutions creating and using these products. Another approach is to modify and expand the archival definition of a record to reflect the nature of contemporary Internet digital media such as the CAA and related atlases. The existing debates over records in archival science needs to be broadened and include other disciplines where the term "record" has other definitions and connotations. If this is not done there may never be adequate records of our increasing participative, interactive digital era. Some of this information may at best be preserved but not systematically archived.

Another result of the interdisciplinary action research of InterPARES 2 has been the recognition of the importance of accuracy as part of the scientific record. To an archivist an authentic record does not have to be accurate but to a scientist this is a key element. While each scientific discipline differs, scientific data quality normally includes all or most of the following data quality elements: positional accuracy, attribute and thematic accuracy, semantic accuracy, and temporal information, reliability, lineage, completeness and logical consistency (Lauriault 2002; Ostman 1997). These are normally captured in metadata and geomatics researchers argue that digital data archivists must consider data quality if they are to ingest data from the sciences. Indeed the data quality of a record may be an important factor in the decision of what scientific data to archive and become a part of the appraisal process. Existing Archiving metadata could be expanded to include metadata standards from other disciplines and the InterPARES 2 Description Team is exploring this issue by analyzing the ISO 19115 metadata standard (US InterPARES 2 2006).

Interoperability is a problem among the rapidly increasing number of digital data bases and is a challenge of knowledge integration. The CAA was faced with the challenge of using information from different databases in different countries and to do this adopted an open source, open standards and interoperable approach. This decision was taken primarily for production reasons but has had beneficial effects in archiving and preservation terms as it helps overcome the problem of technological obsolescence. In the world of Internet cartography effective archiving and preservation requires ensuring interoperability in the future. Indeed it can be argued that interoperability is a key element in archiving all digital data and that an open source standards and specifications approach should be a facet of any archival strategy.

26.9 Conclusion

The participatory action research process between the InterPARES 2 Project and the Cybercartographic Atlas of Antarctica is thought to be a mutually beneficial research activity. InterPARES 2 researchers have qualitatively deconstructed the processes involved in the creation of the CAA, and in turn have been exposed to the implications of collaborative science endeavours. Key geomatics concepts such as data quality (Lauriault, Craig, Pulsifer and Taylor 2007) and interoperability are now influencing archival thinking and the existing archival concept of a record is being challenged. The Geomatics and Cartographic Research Centre (GCRC) through its ongoing interactions with InterPARES 2 had modified its production processes to take into account the needs of long-term preservation. It is important to note that this took place at the outset of the process creating the CAA. Not all problems have been resolved and challenges remain but the interaction has resulted in an interactive Atlas which can be preserved and used in the future and which is unlikely to become technologically obsolete. In turn the archival scientists in the InterPARES 2 project are now more comfortable with concepts of data quality and accuracy and of the need to include these in assessing the value of a scientific record. It has been argued that a major challenge for archivists is record. It has also been argued that a major challenge for archivists is not what to archive but what to leave out! As far as scientific records are concerned, the introduction of data quality and accuracy as an additional element may aid this decision making process as will the extension of existing archival metadata to included metadata standards from the sciences to assess data quality. The results of the Case Study will be published on the InterPARES 2 Internet Site for those who wish to read more and perhaps use the process to assist with the production of their own Internet atlases. Also, the InterPARES 2 project will publish its research results, which will include guidelines, requirements and models (InterPares.org).

   The major challenge still to be resolved to ensure that the Cybercartographic Atlas of Antarctica is available to future generations is to identify an institution that can ingest the Atlas in its entirety and commit to its long-term preservation. Preserving the growing number of digital artifacts of which the CAA is one example, and the digital data on which they draw is not primarily a technical problem. It requires national and international political will, increased financial resources and new institutions and policies.

Acknowledgements

References

Antarctic Spatial Data Infrastructure (AntSDI), 2006, *AntSDI Home Page* accessed September 2006 from http://antsdi.scar.org/about

Atlas of Canada, 2006, *Web Map Services,* accessed August 8, 2006 from http://atlas.nrcan.gc.ca/sites/english/dataservices/web_map_service.html

Bushey, J. and M. Braun, 2005, *Survey of Record-Keeping Practices of Photographers Using Digital Technology Final Report,* Accessed September 2006 from the Restricted Researchers Area of the InterPARES 2 Website.

Canadian Committee on Archival Description, 2001, *Rules for Archival Description Chapter 5 Cartographic Materials* accessed September 2006 from http://www.cdncouncilarchives.ca/rad_ch5.pdf

Canadian Committee on Archival Description, 2003, *Rules for Archival Description Chapter 9 Records in Electronic Form* accessed September 2006 from http://www.cdncouncilarchives.ca/RAD_chap9_revised_Aug2003.pdf

Canadian Digital Information Strategy (CDIS) accessed April 2007 from http://www.collectionscanada.ca/scin/index-e.html

CODATA Working Group on Archiving Scientific Data, 2006, *Home Page*, http://www.nrf.ac.za/codata/

CNC for CODATA, 2006, *Home Page*, http://www.codata.org/canada/

Cook, T., 1995, It's Ten O'Clock, Do You Know Where Your Data Are? *In Technology Review* 98: 48-53.

GCRC 2006. *Geomatics and Cartographic Research Centre Atlases Page.* Accessed September 2006 https://gcrc.carleton.ca/confluence/display/GCRCWEB/Atlases

GeoConnections, 2004, *A Developers' Guide to the CGDI: Developing and Publishing Geographic Information, Data and Associated Services* accessed September 2006 from http://www.geoconnections.org/publications/Technical_Manual/html_e/cgdiindex.html

GeoConnections, 2005, *Archiving, Management and Preservation of Geospatial Data,* Ottawa: Government of Canada, accessed September 2006 from http://www.geoconnections.org/programsCommittees/proCom_policy/keyDocs/geospatial_data_mgt_summary_report_20050208_E.html#4

Harvey, R., 2000, *An Amnesiac Society? Keeping Digital Data for Use in the Future,* Paper presented at the LIANZA 2000 Conference, New Zealand, 15-18 October 2000. Early version available at http://athene.csu.edu.au/~dharvey/TextFiles/AmnesiacSocietyRev.htm accessed 22 June 2006.

International Standards Organization, 2003, *Geographic information – Metadata,* accessed September 2006 from http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=26020

InterPARES 2, 2006, *International Research on Permanent Authentic Records in Electronic Systems Internet Site,* accessed September 2006 from http://inter-pares.org/

InterPARES 2, 2006, *International Research on Permanent Authentic Records in Electronic Systems Glossary,* accessed September 2006 from http://interpares.org/ip2/display_file.cfm?doc=ip2_glossary.pdf&CFID=137165&CFTOKEN=15668692

InterPARES 2, 2003, *23 Case Study Questions that the researchers should be able to answer at the completion of their investigation,* accessed September 2006 from http://interpares.org/display_file.cfm?doc=ip2_23_questions.pdf

InterPARES 2 USA, 2006, *Metadata and Archival Description Registry and Analysis System (MADRAS) beta version,* accessed September 2006 from http://www.gseis.ucla.edu/us-interpares/madras/guidelines.php

International Polar Year, 2006a, *International Polar Year Data Management Report Released,* accessed September 2006 from http://www.ipy.org/news/story.php?id=149

International Polar Year, 2006b, *International Polar Year 2007-2008 Data Policy May 2006,* accessed September 2006 from http://www.ipy.org/Subcommittees/final_ipy_data_policy.pdf

Library and Archives Canada (LAC), 2006a, *Managing Cartographic, Architectural and Engineering Records in the Government of Canada,* Ottawa: Government of Canada, http://www.collectionscanada.ca/information-management/002/007002-2050-e.html

Library and Archives Canada (LAC), 2006b, *Guidelines for Computer File Types, Interchange Formats and Information Standards,* Ottawa: Government of Canada, Ottawa: Government of Canada accessed September 29, 2006 from http://www.collectionscanada.ca/information-management/002/007002-3017-e.html

Lauriault, T. P.; B. L. Craig; P. L. Pulsifer and D. R. F. Taylor, Forthcoming 2007, Today's Data are Part of Tomorrow's Research: Archival Issues in the Sciences, Archivaria, vol. 64, Fall 2007.

Lauriault, T. P. and Y. Hackett, 2005, *CS06 Cybercartographic Atlas of Antarctica Case Study Final Report* accessed October 2006 from http://interpares.org/ip2/ip2_case_studies.cfm?study=5

Lauriault, T. P. and D. R. F. Taylor, 2005, Cybercartography and the New Economy: Collaborative Research in Action, in D. R. F. Taylor (ed.), *Cybercartography: Theory and Practice,* Amsterdam: Elsevier, 181-210.

Lauriault, T. P. 2002, *Section 6, Elements of Geospatial Data Quality,* accessed September 2006 from http://interpares.org/display_file.cfm?doc=ip2_geospatial_data_quality.pdf

Lee, H., 2005, Korea to Build Open Source National Map System, *Special to ZDNet Asia,* Tuesday , September 05 2006 02:46 PM accessed September 2006 from http://www.zdnetasia.com/news/software/0,39044164,39418984,00.htm

Locke, K., 2001, *Grounded Theory in Management Research,* Sage Series in Management Research, Thousand Oaks: Sage.

Longton, 2005, *Record Keeping Practices of Composers Survey Report,* accessed September 2006 from the Restricted Researchers Area of the InterPARES 2 Website.

MacDonald, J. and K. Shearer, 2005, *Toward a National Digital Information Strategy: Mapping the Current Situation in Canada,* Ottawa: Library and Archives Canada accessed September 2006 from http://www.collectionscanada.ca/cdis/012033-700-e.html

McLellan, E., 2004, CS24 *City of Vancouver Geographic Information System (VanMap) Interim Report,* accessed September 2006 from the Restricted Researchers Area of InterPARES 2 http://interpares.org/rws/login.cfm

National Science Foundation Office of Cyberinfrastructure accessed April 2007 from http://www.nsf.gov/dir/index.jsp?org=oci

Nunaliit, 2006a, *The Nunaliit Cybercartographic Framework Internet Site,* accessed September 2006 from http://nunaliit.org/index.html

Nunaliit, 2006b, *The Nunaliit Cybercartographic Framework License Page,* accessed September 2006 from http://nunaliit.org/license.html

Open Geospatial Consortium (OGC), 2006, *OGC Specifications (Standards),* accessed September 2006 from http://www.opengeospatial.org/standards

Open Geospatial Consortium (OGC) Data Preservation Working Group (WG), accessed April 2007 at http://www.opengeospatial.org/projects/groups/preservwg

O'Reilly, T., 2005, Web 2.0: Compact Definition? *O'Reilly Radar* accessed September 2006 from http://radar.oreilly.com/archives/2005/10/web_20_compact_definition.html

Ostman, A.,1997, The Specifications and Evaluation of Spatial Data Quality in *Proceedings from the 18th ICA/ACI International Conference,* Stockholm Sweden, 23-27, Volume 2.

Parush, A., P. L. Pulsifer, K. Philps and G. Dunn, 2006, Understanding Through Structure: The Challenges of Informational and Navigation Architecture *in* Taylor,

D. R. F. and S. Caquard (eds.) *Cybercartography. Special Issue of Cartographica on Cybercartography,* 41 (1), 21-34.

Pulsifer, P. L., S. Caquard and D.R. F. Taylor, (2006) Toward a New Generation of Community Atlases - The Cybercartographic Atlas of Antarctica in Cartwright, W., M. Peterson and G. Gartner (eds.) Chapter 14 of *Multimedia Cartography, Second Edition,* Springer-Verlag.

Pulsifer, P. L. & D. R. F. Taylor, 2005, The Cartographer as Mediator: Cartographic Representation from Shared Geographic Information in D. R. F. Taylor (Ed.), Chapter 7 of *Cybercartography: Theory and Practice,* Amsterdam: Elsevier pp. 149-179.

Pulsifer, P. L., A. Hayes, J.-P. Fise and D. R. F. Taylor (2007) An Open Source Development Framework in Support of Cartographic Integration in Peterson M. P. (ed.) *International Perspectives on Maps and the Internet.* Springer Verlag.

Pulsifer, P. L., 2006, *The Geomatics and Cartographic Research Centre Cybercartographic Atlas of Antarctica Project Description page,* accessed September 2006 from https://gcrc.carleton.ca/confluence/display/GCRCWEB/CAA+Project+Description

Social Science and Humanitis Research Council, 2002a, SSHRC *Research Data Archiving Policy,* Ottawa: Government of Canada, accessed September 2006 from http://www.sshrc.ca/web/about/policies/edata_e.asp

Social Science and Humanitis Research Council, 2002b, *Final Report of the SSHRC National Consultation on Research Data Archiving, Building Infrastructure for Access to and Preservation of Research Data* accessed from http://www.sshrc-crsh.gc.ca/web/about/publications/da_finalreport_e.pdf#search=%22report%20sshrc%20National%20Consultation%20on%20Research%20Data%20Archiving%2C%20Management%20and%20Access%20Systems%22

Strong, F. D. and P. B. Leach, 2005, *The Final Report of the National Consultation on Access to Scientific Data,* Ottawa: Government of Canada, accessed September 2006 from http://ncasrd-cnadrs.scitech.gc.ca/NCASRDReport_e.pdf

Taylor, D. R. F., 2003, The Concept of Cybercartography in M. P. Peterson (ed), *Maps and the Internet Chapter 26,* Amsterdam: Elsevier, pp. 413-418.

Taylor, D. R. F., 2005, (Ed.) *Cybercartography: Theory and Practice.* Vol. 4, Modern Cartography Series, Amsterdam: Elsevier, pp. 574.

Taylor, D. R. F. and S. Caquard, 2006, Special Issue of *Cartographica* on Cybercartography 41(1), April.

Taylor, D. R. F., Lauriault, T. P. and Pulsifer, P. L. 2005, *Preserving and Adding Value to Scientific Data: The Cybercartographic Atlas of Antarctica* presented by D. R. Fraser Taylor to PV2005: Ensuring Long-term Preservation and Adding Value to Scientific and Technical Data, The Royal Society, Edinburgh, Wednesday 23 November.

United Kingdom Data Archive (UKDA), 2006, *UKDA About Page*, accessed September 2006 from http://www.data-archive.ac.uk/about/about.asp

Waters, D. and J. Garrett, 1996, *Preserving Digital Information, Report of the Task Force on Archiving of Digital Information,* Washington, D.C.: CLIR.

Wikipedia, 2006, *Web 2.0 Definition* accessed September 2006 from http://en.wikipedia.org/wiki/Web_2.0

Xie, S., 2006, *Diplomatic Analysis of CS06 Cybercartographic Atlas of Antarctica,* Researchers restricted area of the InterPARES 2 Internet Site accessed September 2006 from http://interpares.org/rws/login.cfm

Zhou, Y., 2005, Thesis M.A. *Profiling and Visualizing Metadata for Multimedia in a Geospatial Portal. A Case Study for the Cybercartography and the New Economy Project.* Dept. Geography and Environmental Studies, Geography, Carleton University, Ottawa.