The Preservation and Archiving of Geospatial Digital Data: Challenges and Opportunities for Cartographers

Tracey P. Lauriault, Peter L. Pulsifer and Fraser Taylor (GCRC) & Barbara L. Craig (Faculty of Information, Toronto),

S33 Cartography and Spatial Data
Chair / Président: Alan G Phipps (Windsor)
Room / Salle: 309 SA 13:00

Outline

1. Introduction
2. Contemporary Cartography
3. Data & Portals
4. Historical Traces
5. Maps and Archives
6. Rescue & Salvage of the Canada Land Inventory (CLI)
7. Multidisciplinary Archival Research
   - CS06 - Cybercartographic Atlas of Antarctica
   - GS10 - Preservation practices of scientific data portals
8. What is being done?
9. Conclusion

1. Introduction

- Examine
  - Emerging challenges
  - Possible strategies
  - Preservation and archiving:
    - Contemporary maps
    - Geospatial data
    - Associated technologies

- 2 Canada Social Sciences and Humanities Research Council (SSHRC) funded research projects,
  - International Research on Permanent Authentic Records in Electronic Systems (InterPARES), University of British Columbia
  - Cybercartography and the New Economy, Carleton University.

2. Contemporary Cartography

- Geo-enabling of content:
  - Photos
  - Video
  - GeoWikis
  - GeoRSS

  (e.g. US Library of Congress Flickr)

- NMOs
  - Maps are borne digital
  - Some digitization of maps
  - Web Mapping (e.g. Atlas of Canada)
  - Distributing framework data (e.g. GeoBase)
  - Interoperable services
  - Open Standards

- Cybercartography (e.g. Atlas of the Risk of Homelessness, Atlases of Indigenous Knowledge)

- Models and Simulations (e.g. Climate Change)

- Enterprise systems (e.g. VanMap)

- Geospatial Data
  - Myriad formats
  - Portals (e.g. GeoConnections Discovery Portal)
  - WEB Services
  - Live sensor feeds
  - Metadata standards
  - New Licenses (e.g. GeoGratis & GeoBase AGREEMENT FOR UNRESTRICTED USE, Science Commons)

3. Data and Portals

- Maps are data representations
- Ptolemaic census – structured information thousands of years old
- Cumulative data sets:
  - Trends
  - Frequencies
  - Patterns
  - Baseline for predictions
  - Raw material for unintended uses
  - Cost of reacquisition is high if possible!
- Public data for future generations
- Data are part of our digital heritage
- Data are critical national resources

- Geospatial, social and scientific knowledge are being distributed in portals
  - Search and retrieval of data
  - Data description
  - Display services
  - Platforms to share models and simulations
  - Collection and maintenance
  - Embodiment of data sharing
  - Metadata description
- Portals are contemporary & collective digital artifacts

(e.g. NRC Science Data Gateway, Data.gov)

4. Maps, Data, Technological Systems & Infrastructures Leave Historical Traces

- **Maps & Data:**
  - Are integral part of the record of a nation's history
  - They illustrate how we thought
  - Provide a distinct point of view
  - Context to understand and read social construction of places
  - Resources for others to build upon
  - Inform decisions (e.g. Hurricane Katrina - US FEMA preservation)

- Maps & data are often inseparable from the technologies that created them (e.g. Domesday, NASA space drawings)

- Cartography has its apparatus (e.g. Code, software, hardware, infrastructure)
  - These apparatus - technologies are also of historical significance
  - These are today's tools
  - These allow us to view today's content tomorrow
  - These are objects that represent our time

5. Maps and Archives

- **Maps, data, technologies and their infrastructures**
  - "are the products of society's need for information, and the abundance and circulation of documents reflects the importance placed in information in society"
  - "are the basis for and validation of the stories we tell ourselves, the story-telling narratives that give cohesion and meaning to individuals, groups, and societies" (Schwartz and Cook 2002).

- The function of an archive in a society "must deal with two intimately related, but separately conceived themes:
  - Knowledge and the shaping of archives and
  - Archives and the shaping of knowledge" (Schwartz and Cook 2002).

Contemporary cartographers should 'embrace' their role in preserving maps and data, subsequently shaping future knowledge & working with archivists to transform the archive so that it may ingest their artifacts.

6. Rescue and Salvage of the Canada Land Inventory

"where the information and the form of the record are so tenuously related, archivists must appraise, acquire, preserve, and control whole systems of information within which various physical media may exists" (Ahlgren & McDonald, 1981).

Consider the life cycle of information and include preservation as content is created!

- CLI was part of Canadian Geographic Information System (CGIS) 1963
- World's 1st GIS
- 2.6 million square kilometers of Canada
- Costs in the order of hundreds of millions of dollars in 70s
- Maps + computer programs
- 1980s 9 track tapes in boxes
- 1995 trans-organizational salvage & rescue group assembled
- 1998 Agricultural portion were restored
- GeoGratis now distributes – one of its most popular products

7. Multidisciplinary Archival Research

- InterPARES 2 was conceived to
  - "to ensure that the portion of society's recorded memory digitally produced in dynamic, experiential, and interactive systems in the course of artistic, scientific and e-government activities can be created in accurate and reliable form, and maintained and preserved in authentic form, both in the short and the long term, for the use of those who created it and of society at large, regardless of digital technology obsolescence and media fragility" (Luciana Duranti 2007).

  - CS06 - Case Study about the Cybercartographic Atlas of Antarctica (Lauriault and Hackett 2005)

  - GS10 - General Study examining the preservation practices of scientific data portals (Lauriault and Craig 2007; Lauriault, Craig, Pulsifer & Taylor 2008)

7.1 CS06 - Cybercartographic Atlas of Antarctica

- **Methodology**
  - IP2 questionnaire – 23 questions
  - 2 sets semi-structured interviews at 2 development stages
- **Observations**
  - Lack of persistent identifiers
  - Lack of detailed documentation – resource scarcity
  - Lack of preservation guidelines from funder
  - Documented open standards, OGC interoperability standards
  - Softwared development – w/source code & versioning system
  - Open source licensing – BSD license
  - Backward compatibility
  - XML easy translatability across time
  - Data from authoritative sources SCAR, Antarctic Treaty System
- **Conclusion**
  - CAA considered adequate to meet the challenges of technological obsolescence
  - Need to better package practices for archiving
  - Preserve at point of creation
  - Greatest roadblock is Canadian Institutional inertia
  - Need IR, TDR, data archive

7.2 GS10 - Preservation practices of data portals

- **Methodology**
  - Literature review + Empirical evidence:
    - IP2 Case Studies and
    - 32 scientific Data Portals
- **Observations**
  - Heterogeneity
    - Specific practices, methodologies, tools, technologies – cultures
  - Architecture of portals vary (e.g. enterprise, networked, protocols)
  - Collection types
    - Research Data Collections (e.g. NRC Gateway)
    - Resource or Community Data Collections (e.g. CIHI, Southern California Earthquake Centre)
    - Reference Data Collections (e.g. CGDI Discovery Portal)
- **Conclusions**
  - Archivists can build on existing portals – the data are already appraised, metadata are quite good, data sharing in place, licenses are explicit
  - Portal creators & data contributors – need to consider preservation and make it part of their mandate particularly since it is unlikely we will have a data archive any time soon!

8. What is being done?

- **European Union**
  - Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval (CASPAR),
  - Digital Repository Infrastructure Vision for European Research (DRIVER),
  - UK Data Archive (UKDA)
  - Data Archiving and Networked Services (DANS) in the Netherlands
- **US**
  - The Cyberinfrastructure Project
  - National Geospatial Digital Archive
- **International**
  - Open Geospatial Consortium (OGC) Data Preservation Working Group
  - CODATA Working Group on Archiving Scientific Data
  - Preserving Access to Digital Information (PADI)
- **Canada**
  - NRC CISTI – Data Management and Curation
  - Many earnest long winded reports with lonely recommendations

- Institution Repositories (IRs) + Trusted Digital Repositories (TDRs)

'blank spot in history' analogous to the Jedi Archive discovery by Obi-Wan Kenobi!

9. Conclusion

The map is not the territory, it creates a record of the territory and it occasions it. It is not just a recording: it constitutes the event. Fortunately, many of the 20th century's digital cartographers are still alive and therefore can be a part of the preservation process and can discuss the history of the mapmaking process of their digital artifacts, and they could also be representatives of technology that created these.

- Creators can collaboratively work with archivists, librarians, technology specialists to design cartographic artifacts that will stand the test of time and build them accordingly (Doorn and Tjalsma 2007; Kinniburgh 1981; Sleeman 2004; Schut 2000; Wilson and O'Neil 2009; Ahlgren and McDonald 1981).
- Build upon existing science and geospatial data portals where appraisal, cataloguing and issues of data quality, metadata and licensing have already been addressed.
- Transform these portals into and create trusted digital data repositories (TDRs)
- NEED National data & map preservation infrastructural strategies