The snapshot problem

Ian Spangler

*Abstract: This article proceeds from the premise that one’s geographical context should fundamentally inform how one engages with geospatial data preservation.*

# 1 Introduction

Unrendered geospatial datasets present unique challenges for anyone concerned with the long-term preservation of born-digital geographic information. Unlike most physical cartographic objects, it’s common for born-digital geospatial datasets to frequently change. Because of this, spatial data about tax parcels, building footprints, or zoning districts (to name just a few examples) can’t simply be accessioned and cataloged once; instead, they must be accessioned repetitively in order to accurately capture changes in the geographic phenomena that the data represent. If a dataset is updated continuously, then it should also be collected continuousl. The fact that this is not often the case in practice constitutes a significant problem for long-term preservation of geospatial data.

This *snapshot problem*, as I will refer to it, has arguably beleaguered map workers for over a century. The infamous correction slips pasted over extant volumes of Sanborn maps serve as an example of how cartographic laborers approached the snapshot problem as early as 1920.[^1] Major changes in subsequent decades including the computer revolution, rapidly changing digital media formats, and the rise of “ubiquitous computing” (Kinsley 2012) significantly influenced how librarians approached their preservation practices, and by 1991, Lai and Gillies (1991) predicted that librarians would encounter particular issues with “the periodic digital updates of spatial data.” They ask, “Which set of data constitute the ‘original’ cartographic material? Which data set should be retained or discarded? What happens when data are recorded continuously?” (Lai and Gillies 1991, 247) More than thirty years later, the data have only become more complicated (Snow 2023), and such questions around data authenticity, digital weeding, and repetivive acquisition remain largely unanswered. In what follows, I attempt an answer by reviewing the history of the snapshot problem and proposing some best practices for how spatial data librarians might continuously and repetitively collect geospatial data.

At its core, the snapshot problem stems from conflicting temporal horizons upon which data user communities need to act. This produces a gap in geospatial data archives that is most pronounced at the city and state levels. The interests of two kinds of geospatial data users are at odds: *preservationists*, who want regular snapshots of geospatial data so future users can reconstruct the past, and *practitioners*, who want up-to-date geospatial data so current users can make sense of the present. Practitioners are charged with making sense of the present *for the present*, while preservationists must anticipate how to make the present legible *for the future*. Persistent issues like limited funding, high personnel turnover, and popular geoportals that are ill-suited to preservation make it particularly difficult for the practitioners who make and maintain geospatial datasets to devote the necessary energy to then preserving and archiving them. Despite these challenges—in fact, *because* of them—I argue that geography and map library (G&ML) workers should foreground the snapshot problem in their geospatial data collecting workflows, rather than treating it as an add-on or afterthought.

There is no one-size-fits-all solution to the snapshot problem. Instead, drawing on the insights of the Civic Switchboard project (Thaler, Mattern, and Rapchak 2022) and related research on libraries as “civic infomediaries” (Robinson and Mather 2017), I argue that G&ML workers at city and public academic libraries can position themselves ideally to address the snapshot problem through place-based partnerships with local and regional municipal organizations. Collections development for digital geospatial data should acknowledge that “all data are local” (Loukissas 2019) by taking more geographically emplaced approach to preservation. I call this approach *situated preservation*. Practicing it requires thinking spatially about what kinds of knowledge geospatial data produce, understanding data as situated within specific means of production, and recognizing the constraints imposed by prevailing digital infrastructure for spatial data collecting. Rather than collections development models that follow a “just-in-time” or “just-in-case” approach (Ramsey 2023), situated preservation follows a just-in-*place* model—something that, as Kane (1997) suggests, combines just-in-time/case models, but also strategically leverages local social networks in order to develop meaningful archives through place-based collecting.

I conclude the article with a case study that demonstrates situated preservation in action, drawing on my experiences starting a pilot project between the Leventhal Map & Education Center (LMEC), the Boston Public Library (BPL), and geospatial analysts at the City of Boston (COB). The case study highlights three types of considerations that G&ML workers should make when approaching the snapshot problem from a situated preservation standpoint: *place-based* considerations, which take into account the inherently geographical character of geospatial data preservation; *technical* considerations, which address issues of platformed infrastructure and programmatic snapshotting; and *personnel* considerations, which focus on funding, staffing, and collaboration across multiple institutions.

# 2 Methodology

The paper proceeds in three parts: a literature review (section III), a theoretical framework (section IV), and a case study (section V) based on my experiences as a G&ML worker. In assembling my literature review, I primarily searched Google and Google Scholar with variations of keywords including geospatial data preservation, spatial data librarianship, geoportals, geospatial data archives, and born-digital collecting. I used boolean search and phrase searching to to chain these keywords together (e.g., "geospatial data" AND "archive" AND "digital") and often bound my search results by publication date depending on the purpose of my search (e.g., I filtered for only articles published before 2000 when compiling articles about the first wave of geospatial data collecting).

As Hodder and Beckingham (2022) argue, modern search platforms do not “help us navigate the arrangement of the archive” in the traditional sense of the archive—instead, they “[allow] us to circumvent it.” Their description of searching through so-called “digital archives” (Hodder and Beckingham 2022, 1300) handily summarizes my workflow:

[W]e search by trial and error. We enter a term as a proxy for a broader theme. We refine it. We search again. If we are lucky, we might be able to tie our research question to a distinct and historically stable watchword that reveals dozens of new sources.

This iterative process led me to a set of about 300 articles and reports published between 1988 and 2024 in a variety of academic journals including *Computers, Environment, and Urban Systems*, *International Journal of Geographical Information Science*, and *Journal of Map & Geography Libraries* (the latter of which represented over 200 of the surveyed papers). This set of articles informed the following review and theoretical framework, while the case study into a geospatial data archiving partnership derives primarily from my own professional experiences as a G&ML worker. Empirically, it draws from meetings, conversations, and correspondences spanning about 16 months between autumn 2022 and summer 2024.

I discuss the partnership in greater detail in section V; for now, suffice to say that it only exists thanks to the work of prior (GIS) librarians at LMEC, the Boston Public Library’s (BPL) Digital Repository Services team, and support from our colleagues at Analyze Boston (<https://data.boston.gov/>).

# 3 The snapshot problem: Temporal challenges

To properly situate the snapshot problem—its characteristics, why it matters, and how we can go about resolving it—let me first address *why* geospatial data might be desirable to preserve at all. While geospatial data preservation is clearly a significant issue for a wide range of actors including geography and map library (G&ML) G&ML workers (Jobst 2010; Clark 2016; Bidney and Brown 2023), public agencies (Morris 2009; Committee 2014), and the private sector (Folger 2012), all of these actors are motivated by different forces when they think about how and why to preserve digital geospatial data. Then, I discuss the snapshot problem and its associated challenges, which include *digital infrastructure*, *social relationships*, *make vs. maintenance*, and *personnel*.

Geospatial data preservation, therefore, should not be understood as a value-neutral process, but rather—as critical archival studies have long observed (Ghaddar and Caswell 2019)—as a *contingent set of practices* that transcend mere “collection” to include political, economic, and cultural imperatives.

In lieu of an extended discussion of geospatial data preservation practices at the level of public agencies and private sector, suffice to say that G&ML collections development operates according to a different set of values. For public agencies, geospatial data preservation—both in terms of geographical fidelity and standard setting—is often a question of national security and economic development. Sine the early 1990s, U.S. state agencies have devoted a great deal of energy to . Meanwhile, for the private sector, . Because of its value in the contemporary platform economy (Zook and Spangler 2023), spatial data and their archives are particularly susceptible to being enrolled into processes of assetization (Birch and Ward 2024).While they certainly affect how G&ML workers approach spatial data collecting, this article is intended for an audience of library professionals and as such I want to focus on why and how spatial data librarians in particular choose to archive and preserve geospatial data.

What I find in a review of the literature is that, in many G&ML contexts, geospatial data workers take a *service-oriented approach* rather than an *object-oriented approach*. Put another way, the administration of geographic information systems (or GIS) *as a service*—such as learning to use the software or teaching patrons how to use it—often precedent over geospatial data collections development. In cases where significant geospatial data collections development *has* happened at libraries, the snapshot problem persists: data snapshots are either not discussed in depth (Florance 2006), . For G&ML workers, geospatial data preservation is characterized by the remit of libraries more generally to make information accessible, but it also relitigates longstanding librarianship debates over “access versus ownership” (Kane 1997) in novel ways.

The preservation of born-digital geospatial data has been a concern of geography & map librarians since at least the early 1990s. This review focuses on the U.S. library context due to the author’s professional experience working in a U.S.-based map library, but future research would do well to review geospatial data preservation in a non-U.S. (and non-Western) geographical context.

When the U.S. Government Printing Office distributed the 1990 Census on CD-ROM to over a thousand depository libraries[[1]](#footnote-23) across the country, “the depository libraries… became guardians of a vast quantity of digital data [overnight]” (Lamont and Marley 1998, 143). These new formats for storage, combined with an increasing need for geographic information systems (GIS) to read, write, and manipulate the data in them, ushered many libraries into an age of GIS. Consequently, in 1992, the Association of Research Libraries (ARL) launched its first comprehensive engagement with GIS through its GIS Literacy Project March and Scarletto (2017). Formed in partnership with Esri—a software development company which maintains a strong grip on the enterprise GIS market to this day—ARL’s GIS Literacy Project funded librarians to visit Esri’s headquarters for free training on Esri’s GIS software (French 2000).

. In 1994, the Clinton Administration passed Executive Order 12906, *Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure*.

By the turn of the twenty-first century, GIS had become more widely used in libraries, with 89% of participants in the ARL project reporting that they offered GIS services in 1999 (Davie et al. 1999, 3). Some GIS-literate librarians treated the software as though it left less room for subjective interpretation than traditional papers maps. “GIS makes everyone think of maps in a whole new way,” said one librarian, reflecting on the ARL event. “Maps in a GIS are like the answers to questions,” unlike “traditional maps,” which the participant explained “have to be interpreted by the user” (Lang 1992, 882). While this . Other librarians who hadn’t yet learned the software expressed feeling “GIS induced guilt” as a result (Ferguson 2002, 94).

In those early days, libraries and librarians—as patron-facing institutions with a mission to provide “geospatial information [quickly and efficiently] to its users” (Larsgaard 1998, 168)—deferred rigorous planning for the long-term preservation of digital spatial data in favor of the more urgent for librarians to learn how to use, manipulate, and find that data using GIS. As Mary Lynette Larsgaard noted in 1998 (Larsgaard 1998, 165):

[D]oing reference work for a digital-data user can easily take ten times as long as doing hard-copy reference work. After all, users to come to a university library for print material already know how to read… For users of digital data, the library in effect often has to teach the users to ‘read’ and it is very time-consuming.

This sentiment

A tension emerged between

To summarize from Meagan Snow’s report on the topic of geospatial data preservation (Snow 2023), key challenges include *standards*, *state*, *scale*, *scope*, and *size*:

* **Standards**: While the two most common spatial data models (vector and raster) are relatively stable, those models can be encoded in a multitude of formats. Some of the most commonly encountered formats, such as .shp files or Esri geodatabases, are not easily readable without geographic information systems (GIS) software. By contrast, more readable plain text formats like .geojson lack inherent metadata.
* **State**: Information about the content and context states of spatial data is often difficult to capture. This includes the data’s processing history, how the data relate to specific projects in which they’re used, and what sorts of relationships exist between datasets.
* **Scale**: Because spatial data is regularly produced at varying levels of granularity (e.g., state, county, city), a single dataset may have multiple incarnations.
* **Size**: Spatial data—especially raster data such as satellite images, aerial photography, and georeferenced maps—can be significant in file size. In the context of geospatial data preservation, storage itself is fairly inexpensive, but the environmental costs of cloud storage make a “collect it all” philosophy neither tenable nor sustainable.
* **Scope**: The sheer volume of geospatial data produced every day makes collecting it all an infeasible prospect.

# 4 Situated preservation: Geographical solutions

Site: What are the

Situation: Who made the data does not equal who preserves the data

Given these challenges, how

# 5 Case study: A civic data partnership for Boston, MA

1. City of Boston city council districts

## 5.1 Digital infrastructure issues

## 5.2 Code samples for programmatic snapshotting

# 6 Conclusion

“Even if copies of live datasets can be made through taking regular comprehensive deposits or change-only updates,” writes Chris Fleet, “it is often difficult to comprehensively archive relevant associated data, including relationships, linked data and styling information on how the dataset is presented” (Fleet 2023, 119).

**Footnotes**

[^1] See https://www.loc.gov/collections/sanborn-maps/about-this-collection/.

**Bibliography**

Bidney, Marcy, and Georgia Brown. 2023. “Considerations in Collecting Born-Digital Maps.” *Imago Mundi* 75 (1): 122–26. <https://doi.org/10.1080/03085694.2023.2231771>.

Birch, Kean, and Callum Ward. 2024. “Assetization and the ‘New Asset Geographies’.” *Dialogues in Human Geography* 14 (1): 9–29. <https://doi.org/10.1177/20438206221130807>.

Clark, John H. 2016. “The Long-Term Preservation of Digital Historical Geospatial Data: A Review of Issues and Methods.” *Journal of Map & Geography Libraries* 12 (2): 187–201. <https://doi.org/10.1080/15420353.2016.1185497>.

Committee, Federal Geographic Data. 2014. “National Geospatial Data Asset Management Plan: Portfolio Management Implementation Plan for the OMB Circular A-16 Supplemental Guidance (Steering Committee Approved: March 20, 2014).” Reston, VA: Federal Geographic Data Committee.

Davie, D. Kevin, James Fox, Barbara Preece, and Library Affairs. 1999. “Flyer 238: The ARL Geographic Information Systems Literacy Project.” Research. Washington, D.C.: Association of Research Libraries.

Ferguson, Anthony (Tony) W. 2002. “Back Talk—GIS Induced Guilt.” *Against the Grain* 14 (5). <https://doi.org/10.7771/2380-176X.3760>.

Fleet, Christopher. 2023. “Curating Born-Digital Maps.” *Imago Mundi* 75 (1): 118–22. <https://doi.org/10.1080/03085694.2023.2231753>.

Florance, Patrick. 2006. “GIS Collection Development Within an Academic Library.” *Library Trends* 55 (2): 222–34.

Folger, Peter. 2012. “Issues and Challenges for Federal Geospatial Information.” 7-5700. Congressional Research Service.

French, Mary. 2000. “The ARL GIS Literacy Project: Support for Government Data Services in the Digital Library.” *IASSIST Quarterly* 24 (1): 19. <https://doi.org/10.29173/iq579>.

Ghaddar, J. J., and Michelle Caswell. 2019. “‘To Go Beyond’: Towards a Decolonial Archival Praxis.” *Archival Science* 19 (2): 71–85. <https://doi.org/10.1007/s10502-019-09311-1>.

Hodder, Jake, and David Beckingham. 2022. “Digital Archives and Recombinant Historical Geographies.” *Progress in Human Geography* 46 (6): 1298–1310. <https://doi.org/10.1177/03091325221103603>.

Jobst, Markus, ed. 2010. *Preservation in Digital Cartography: Archiving Aspects*. Lecture Notes in Geoinformation and Cartography. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-12733-5>.

Kane, Laura Townsend. 1997. “Access Vs. Ownership: Do We Have to Make a Choice?” *College & Research Libraries* 58 (1): 58–66. <https://doi.org/10.5860/crl.58.1.58>.

Kinsley, Sam. 2012. “Futures in the Making: Practices to Anticipate ‘Ubiquitous Computing’.” *Environment and Planning A: Economy and Space* 44 (7): 1554–69. <https://doi.org/10.1068/a45168>.

Kunkel, Rebecca. 2017. “Law Libraries and the Future of Public Access to Born-Digital Government Information.” *Law Library Journal* 109 (1): 67–82.

Lai, Pohchin, and Charles F. Gillies. 1991. “The Impact of Geographical Information Systems on the Role of Spatial Data Libraries.” *International Journal of Geographical Information Systems* 5 (2): 241–51. <https://doi.org/10.1080/02693799108927846>.

Lamont, Melissa, and Carol Marley. 1998. “Spatial Data and the Digital Library.” *Cartography and Geographic Information Systems* 25 (3): 143–49. <https://doi.org/10.1559/152304098782383142>.

Lang, Laura. 1992. “Mapping the Future of Map Librarianship.” *American Libraries* 23 (10): 880–83. <https://www.jstor.org/stable/25632729>.

Larsgaard, Mary Lynette. 1998. “Chapter 10: Availability of Geospatial Data Through Libraries in the United States.” In *Modern Cartography Series*, edited by Fraser Taylor, 3:159–84. Policy Issues in Modern Cartography. Academic Press. <https://doi.org/10.1016/S1363-0814(98)80011-7>.

Loukissas, Yanni A. 2019. *All Data Are Local: Thinking Critically in a Data-Driven Society*. Cambridge, Massachusetts: The MIT Press.

March, Gregory, and Edith Scarletto. 2017. “The Evolution of GIS Services Within North American Academic Libraries: Documenting Change Through the Decades (1995–2016).” *Journal of Map & Geography Libraries* 13 (2): 222–45. <https://doi.org/10.1080/15420353.2017.1313803>.

Morris, Steven P. 2009. “The North Carolina Geospatial Data Archiving Project: Challenges and Initial Outcomes.” *Journal of Map & Geography Libraries* 6 (1): 26–44. <https://doi.org/10.1080/15420350903432507>.

Ramsey, Melissa. 2023. “‘Just in Time’ Collection Development: Background and Current Challenges.” *Contemporary Issues in Collection Management*, August.

Robinson, Pamela, and Lisa Ward Mather. 2017. “Open Data Community Maturity: Libraries as Civic Infomediaries.” *URISA Journal* 28 (1): 31–39.

Schonfeld, Roger C. 2009. “Documents for a Digital Democracy: A Model for the Federal Depository Library Program in the 21st Century.” Association of Research Libraries.

Snow, Meagan A. 2023. “Preserving Geospatial Data.” Digital Preservation Coalition. <https://doi.org/10.7207/twr23-01>.

Thaler, Jane, Eleanor Mattern, and Marcia Rapchak. 2022. “Building Capacity for Library Participation in Open Civic Data Work with the Civic Data Education Series.” *Proceedings of the ALISE Annual Conference*, October. <https://doi.org/10.21900/j.alise.2022.1022>.

Zook, Matthew, and Ian Spangler. 2023. “A Crisis of Data? Transparency Practices and Infrastructures of Value in Data Broker Platforms.” *Annals of the American Association of Geographers* 113 (1): 110–28. <https://doi.org/10.1080/24694452.2022.2071201>.

1. That is, libraries which are legally obligated to receive certain government documents and make them accessible to the public. For an overview how the federal library depository program is adjusting to the proliferation of born-digital media, see Schonfeld (2009) and Kunkel (2017). [↑](#footnote-ref-23)