British Columbia geoscience Spatial Data Infrastructure (BCgSDI): A progress report

Yao Cui[1, a], Jessica Norris[1], and Gabe Fortin[1]

[1] British Columbia Geological Survey, Ministry of Mining and Critical Minerals, Victoria, BC, V8W 9N3
[a] corresponding author: Yao.Cui@gov.bc.ca

Recommended citation: Cui, Y., Norris, J., and Fortin, G., 2026. British Columbia geoscience Spatial Data Infrastructure (BCgSDI): A progress report. In: Geological Fieldwork 2025, British Columbia Ministry of Mining and Critical Minerals, British Columbia Geological Survey Paper 2026-01, pp. 221-226.

Abstract

The British Columbia Geological Survey is modernizing its information systems by developing the British Columbia geoscience Spatial Data Infrastructure (BCgSDI). This modernization addresses the long-standing challenges of managing numerous disparate and legacy databases faced by many public geoscience organizations by introducing a phased transition from siloed applications, to point-to-point system integration and interoperability and, ultimately, towards a canonical data model. This canonical data model will enable the BCgSDI to be the single authoritative source of geoscience data independent of any individual application. Central to the BCgSDI is an end-to-end (from data source to release) approach that streamlines the full geoscience data lifecycle, from data acquisition and updating through validation, integration, and publication. The BCgSDI will be implemented with only a few technologies on a Cloud-native platform, to ensure scalability, while reducing the complexity and cost of developing and maintaining applications. The BCgSDI project can serve as an example of how to modernize a geoscience information infrastructure while maintaining business continuity and enabling future innovation.

Keywords: Spatial Data Infrastructure, SDI, geoscience, mineral resource, architecture, integration, canonical data model, BCgSDI, Cloud-native platform

1. Introduction

The British Columbia Geological Survey (BCGS) integrates research programs with historical data, providing geological and mineral resource information to guide societal decisions centered on the Earth sciences. Although the public-facing end of the Survey databases have been well-received, the back end is a patchwork of legacy systems, largely developed in-house to fulfill specific lines of business and services, which led to siloed applications lacking coherent data management. These systems, many of which are decades-old, have resulted in several enduring challenges including the duplication of efforts to maintain similar data in multiple systems, extra work to prevent data inconsistency for the same information, and added costs of maintaining disparate systems and obsolete technologies. In addition, the systems and applications are not immediately amenable to support artificial intelligence (AI) and large language models (LLM) for advanced data access and modelling.

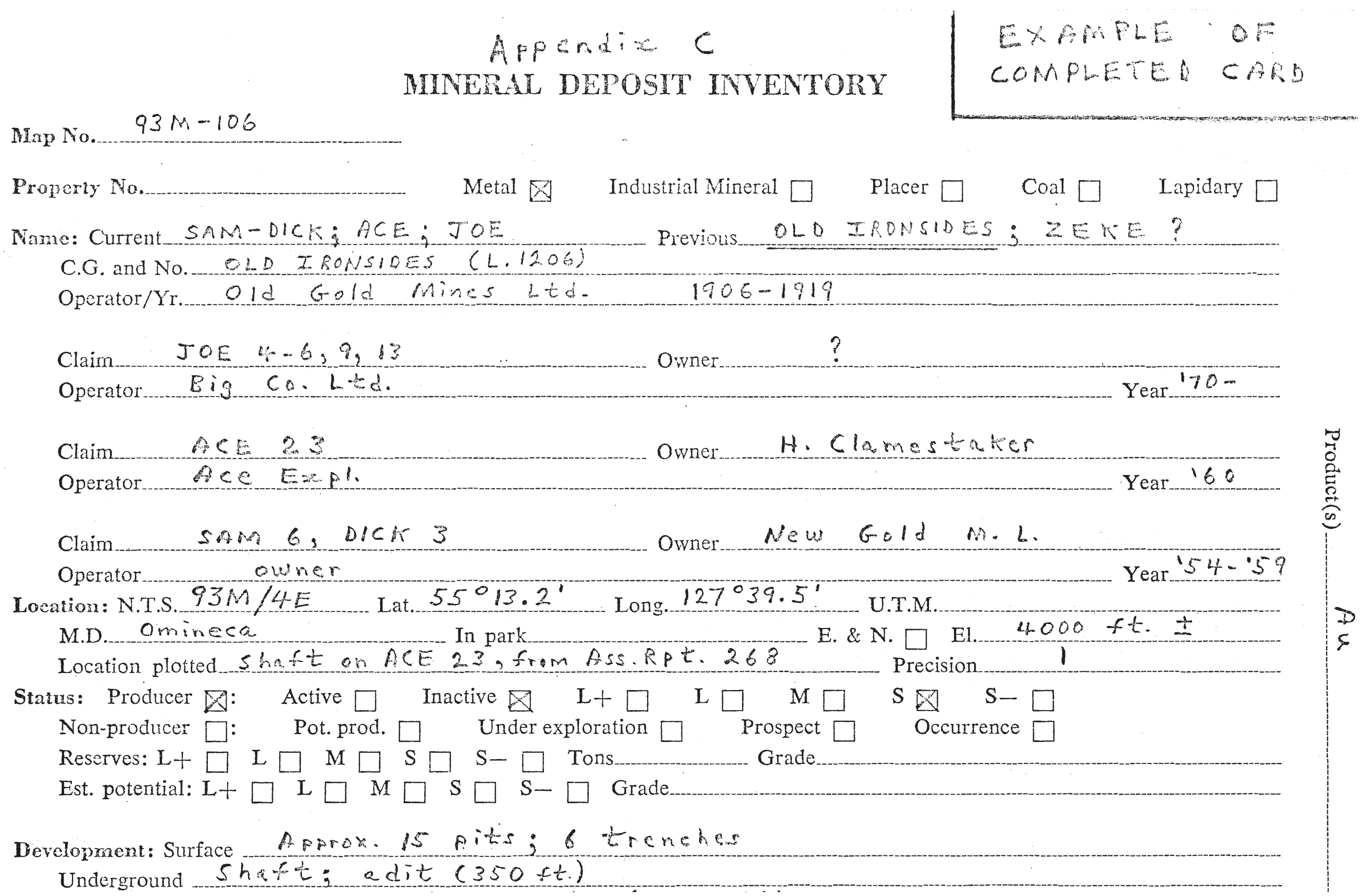
To eliminate current database-related impediments to delivering Survey programs, the BCGS is developing the British Columbia geoscience Spatial Data Infrastructure (BCgSDI). SDI was defined by the U.S. National Research Council (1993) as a framework of technologies, policies, and institutional arrangements enabling the creation, exchange, and use of geospatial data. In its modernization efforts, the Survey adopted SDI as a concept for a group of logically arranged information systems that are connected via a common interface, link data sources to products, and facilitate data update, sharing, and services (Federal Geographic Data Committee, 2024; Open Geospatial Consortium, 2025). The BCgSDI project aims to establish a coherent, integrated, and interoperable (i.e., easy to share with diverse clients) environment that supports the entire data lifecycle, from geoscience data sources to data products to applications and services. The primary focus is to build a foundation to manage geoscience data as authoritative reusable assets. In practical terms, the BCgSDI provides the database environment, processing workflows, and applications needed to capture, update, validate, integrate, and disseminate geoscience information. It supports both routine operational use and the creation of value-added data products and applications, such as web services and interactive tools, that serve mineral exploration, land-use planning, public safety, and research communities.

2. Context

The analogue to digital transformation in the last half-century created siloed information systems, largely because rapid advancements in computing technology outpaced enterprise-wide planning, system architecture design, and integration standards. Many public geoscience organizations face the same challenge of managing numerous disparate and legacy databases and applications.

Before the widespread adoption of electronic computing, geological and mineral resource information in British Columbia was published as hard-copy documents. Geologists collected field observations, compiled maps, and released comprehensive reports describing regional geology, mineral deposits, and mineral exploration activities in publications that served as complete records (e.g., Brothers, 1968). Although these reports were comprehensive, they were static, had inconsistent descriptors, and could not readily support comparative analysis. In 1967, British Columbia introduced a paper card system to record mineral deposit information (Fig. 1; Jones and McPeek, 1992). These cards had standardized entries such as location, commodities, mineralogy, claims, status, owners and operators of the property, development work and results, and references. By the early 1970s, these card records were transferred into a computerized database that eventually became MINFILE, the current BCGS online inventory of metallic mineral, industrial mineral, and coal occurrences in British Columbia. Marking a major transformation in information management (Montgomery et al., 1975) this enabled efficient querying and reporting to support advanced data analysis. However, the analogue to digital transformation occurred incrementally in response to emerging new technologies and commonly along independent organizational lines of business. To accommodate specific program needs, individual teams developed dedicated applications with their own back-end databases. Each system was optimized for a particular line of business (e.g., mineral occurrences, exploration tracking, and assessment work reporting) without common feature definitions, data models, computing platforms, and technologies. This approach marked the beginning of systems that were isolated in program-related silos, a practice that has largely persisted to the present not only at the BCGS but at most public geoscience organizations.

Currently, the Survey maintains more than 40 databases and applications, each designed for a specific business purpose. The particularly high-value systems include: MINFILE (mineral occurrences); ARIS (assessment report index system); Property File (historical documents); Publication Catalogue; BC Digital Geology; Surficial Geology; ExploreTrack (tracking mineral exploration and mining activities), and various geochemical and geochronological databases. Our operations commonly require access to data such as topographic base maps and mineral tenures managed by other government agencies. Most of the databases use different technologies, many of which would now be considered obsolete. These systems typically manage subsets of intrinsically related data independently, resulting in multiple copies of similar information stored in siloes lacking connections or interfaces (Fig. 2). Sharing data between systems commonly relies on manual steps or custom computer scripts, which are costly and risk introducing


Appendix C

MINERAL DEPOSIT INVENTORY

EXAMPLE OF COMPLETED CARD

Map No. 93M-106

Property No. Metal ☒ Industrial Mineral ☐ Placer ☐ Coal ☐ Lapidary ☐

Name: Current SAM-DICK; ACE; JOE Previous OLD IRONSIDES; ZEKE ?

C.G. and No. OLD IRONSIDES (L.1206)

Operator/Yr. Old Gold Mines Ltd. 1906-1919

Claim JOE 4-6, 9, 13 Owner ?

Operator Big Co. Ltd. Year '70-

Claim ACE 23 Owner H. Clamestaker

Operator Ace Expl. Year '60

Claim SAM 6, DICK 3 Owner New Gold M. L.

Operator owner Year '54-'59

Product(s) Au

Location: N.T.S. 93M/4E Lat. 55°13.2' Long. 127°39.5' U.T.M.

M.D. Omineca In park E. & N. ☐ El. 4000 ft. ±

Location plotted shaft on ACE 23, from Ass. Rpt. 268 Precision 1

Status: Producer ☒: Active ☐ Inactive ☒ L+ ☐ L ☐ M ☐ S ☒ S— ☐

Non-producer ☐: Pot. prod. ☐ Under exploration ☐ Prospect ☐ Occurrence ☐

Reserves: L+ ☐ L ☐ M ☐ S ☐ S— ☐ Tons Grade

Est. potential: L+ ☐ L ☐ M ☐ S ☐ S— ☐ Grade

Development: Surface Approx. 15 pits; 6 trenches

Underground Shaft; adit (350 ft.)

Fig. 1. Part of a standardized mineral deposit inventory card that was used by the BCGS in the late 1960s and early 1970s. The inventory cards were the template for what eventually became MINFILE.

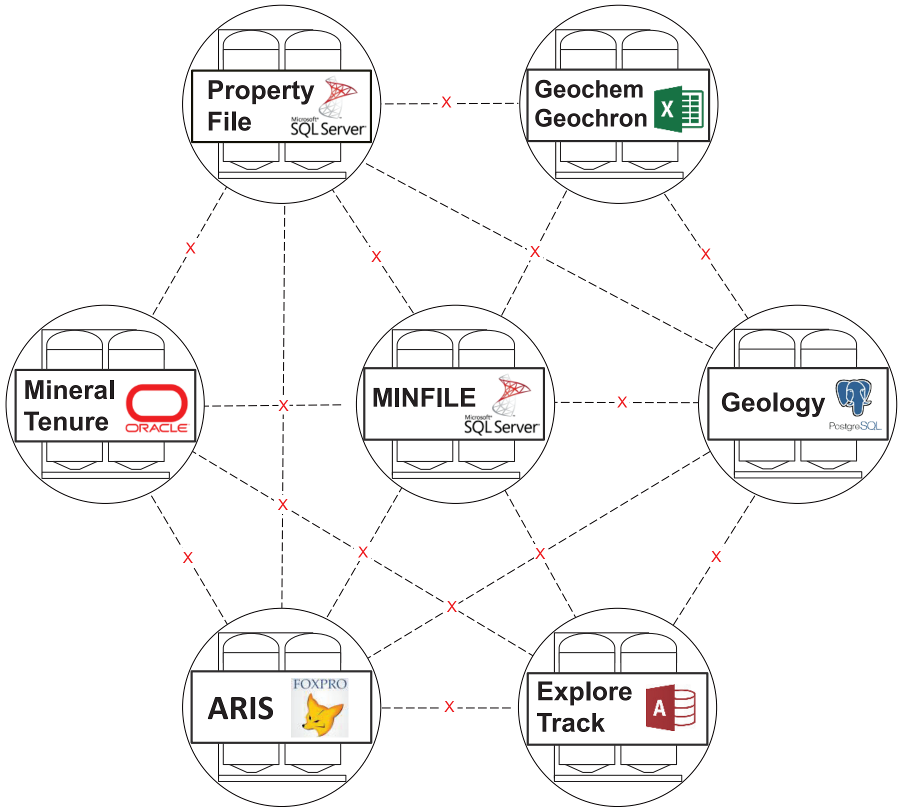


Fig. 2. Siloed legacy systems lacking coherent data management. Black dashed lines refer to intrinsic dependencies; red crosses indicate no connections or interfaces between databases.

errors. This fragmentation not only increases operational and maintenance efforts but also limits the ability to capture and update the data efficiently.

The siloed systems also impede managing and linking source data with derived data products after processing and interpretation. Field observations, laboratory analyses, instrumental measurements, and supporting documentation are not always traceable to derived datasets or published interpretations. As a result, it is difficult to track data provenance, apply consistent validation rules, or make comprehensive datasets readily available for modern analytics, including mineral potential modelling and emerging data-driven approaches.

Together, these issues increase the effort required to ensure data currency, consistency, and authority, while complicating data integration, analysis, and reuse across programs and clients. The Survey BCgSDI project is working toward resolving these issues.

3. Overview of the BCgSDI architecture

The BCgSDI system architecture is multi-tiered (Fig. 3). The lower level is for managing and updating the raw data in what we refer to as the geoscience operational database environment. The next level, which we refer to as the geoscience application database environment is for derived data products and application-level functions. The third level is for geospatial web services and data dissemination.

Serving as the foundation, the operational environment includes: 1) an observational database for capturing and compiling raw data that retains provenance information; 2) a staging database for data transformation (validation, standardization, review, approval, and integration); and 3) an archival database for preserving consistent and standardized results from the staging database (for details see Cui, 2011). Applications create derived data products from the operational environment and integrate them in the application database environment to support queries, reporting, business intelligence, data downloads, web interfaces, and interoperable data sharing to ensure the data and services can be broadly consumed.

Interoperable data sharing is achieved by adopting open standards such as the Open Geospatial Consortium (OGC) standards GeoSciML (GeoSciML Modeling Team, 2017), EarthResourceML (CGI-IUGS EarthResourceML Working Group, 2013), and OGC Web Map Service (WMS) and Web Feature Service (WFS). The BCgSDI will be extended for ready access to the British Columbia Geographic Warehouse (BCGW) and will include a document management system to access many historical and current technical documents. To reduce the complexity and cost of maintenance, the BCgSDI will be implemented in a Cloud-native platform using open-source software.

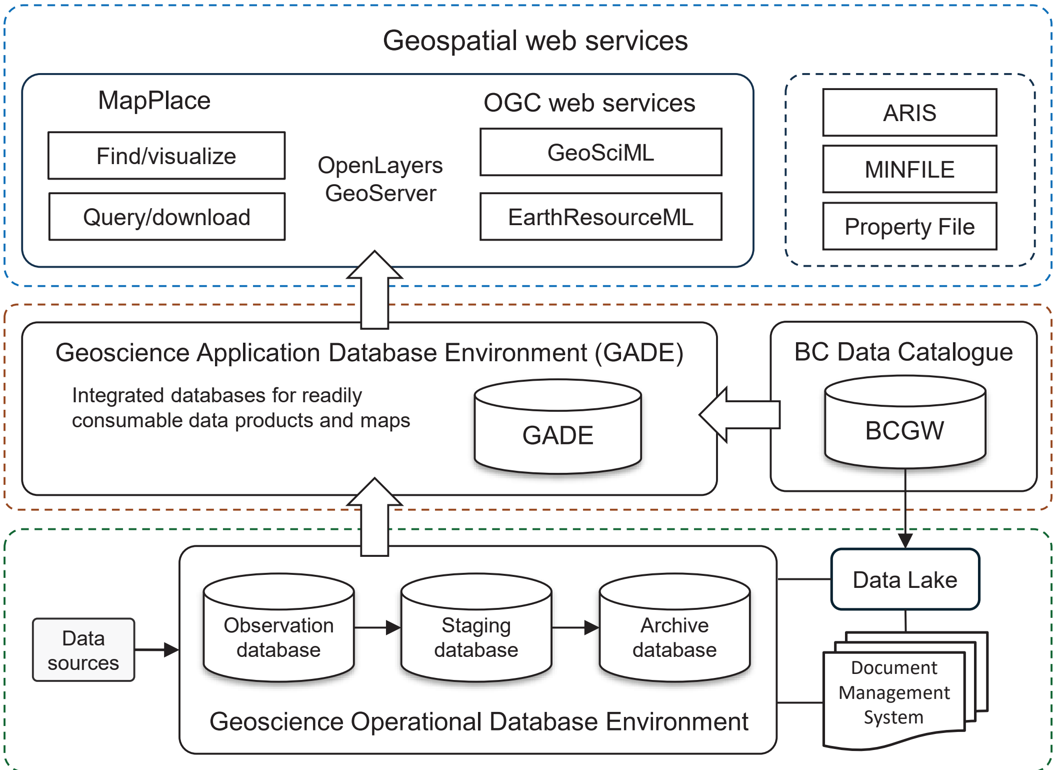


Fig. 3. High-level overview of the BCgSDI system architecture; BCGW refers to British Columbia Geographic Warehouse.

4. Progress

Modernization of the British Columbia information systems will take many years. Thus, we have adopted an incremental approach that ensures continuity of services while progressively improving data management practices. In the first phase, individual applications and their back-end databases are being upgraded to meet current security, performance, and supportability standards. This initial work, which is currently underway, focuses on high-value systems that anchor the broader integration including MINFILE, ARIS, Property File, Publication Catalogue, BC Digital Geology, Surficial Geology, ExploreTrack, and other essential geological and geochemical datasets. Where data need to be shared, we are introducing database to database (or point-to-point) interfaces to reliably exchange information (Fig. 4). This phase reduces the most urgent technical risks and delivers immediate improvements to usability and data currency, while allowing existing business processes to continue.

In the second phase of implementation, our focus will shift from connecting systems to consolidating and managing data more effectively. To reduce the complexity of point-to-point integrations and to improve efficiency and data consistency, we aspire to a canonical data model (Hohpe and Woolf, 2003), a standardized way of representing shared information in the BCgSDI. Modern practice treats the canonical data model as a universal translation layer that supports interoperability and governance across large enterprise environments (e.g., California Learning Resource Network, 2025). At the conceptual level, the canonical data model acts as the authoritative semantic layer defining key feature types and relationships between them. It establishes common and standardized data definitions, vocabularies, and data quality specifications, with each of the datasets recognized as the single authoritative source, independent of any individual application (Fig. 5). Importantly, although the canonical model provides a single source of truth for shared data definitions and datasets, it does not require a single physical database. Individual applications may maintain their own operational data stores, but they access and update shared information through the canonical layer. This ensures that data remains consistent and coherent across the system rather than fragmented into isolated silos. The BCgSDI will use the canonical data model concept to guide how data are structured, shared, and governed. In practice, synchronization mechanisms can be implemented to keep datasets aligned across environments and preserve a single authoritative source of record. In time, successful implementation of this phase will reduce duplication, enforce consistent definitions and validation rules, and streamline the full data lifecycle from collection and update through to publication and dissemination. It will also strengthen data exchange and data sharing with external partners and data providers, enabling analytic capabilities, including business intelligence and mineral potential modelling.

5. Conclusion

The BCgSDI modernization aims to transition BCGS information systems from isolated legacy databases and

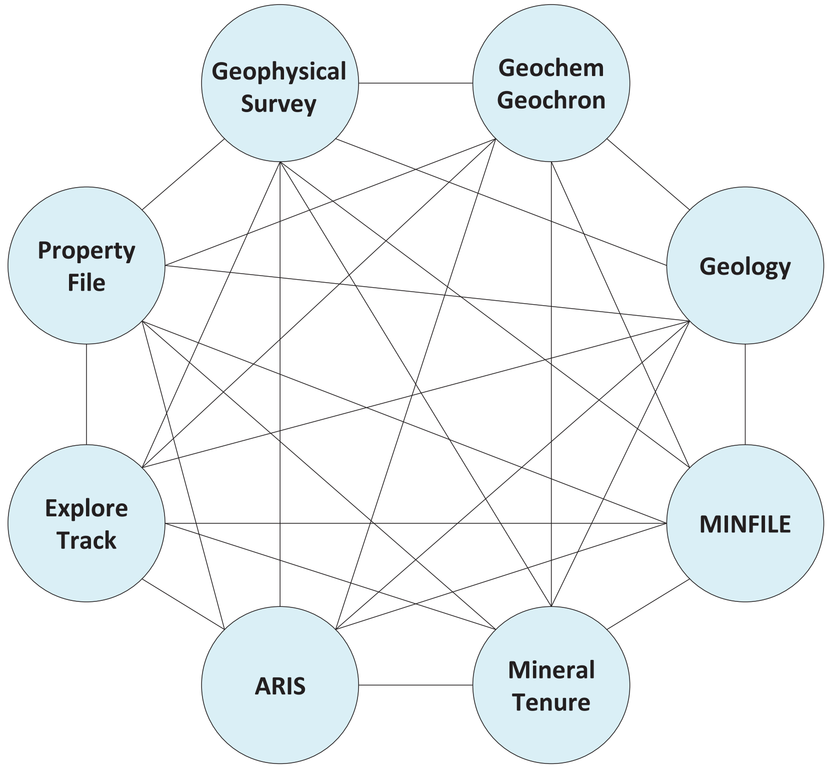


Fig. 4. Point-to-point integration showing individually deployed applications that interface with one another yet operate on separate back-end databases.

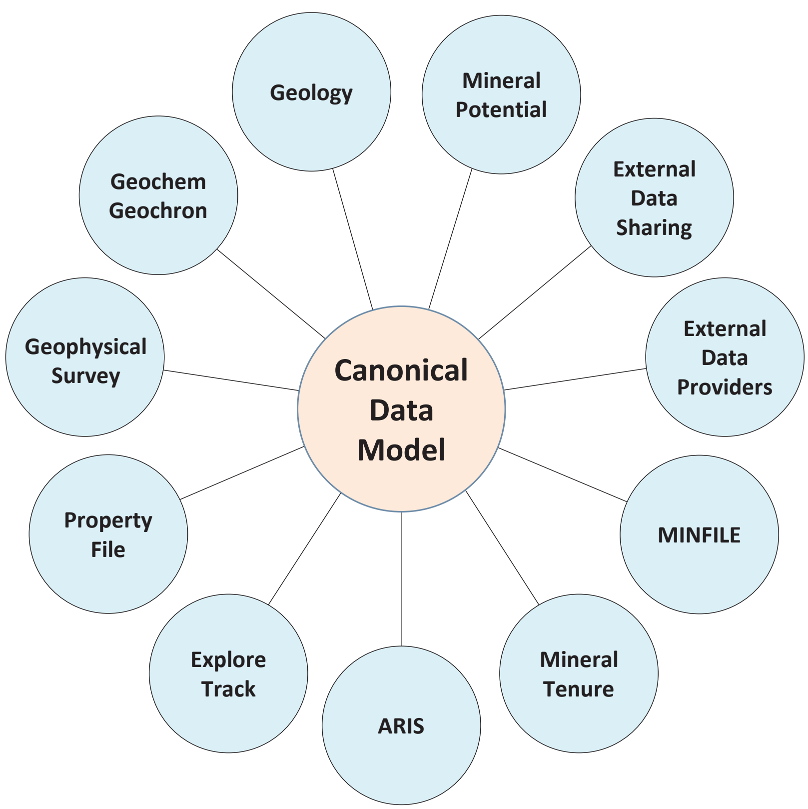


Fig. 5. Conceptual canonical data model as the single authoritative data source that can be used by all applications including data exchange and sharing with external partners.

applications to a sustainable and service-oriented Cloud-based platform. Although the initial goal is to complete point-to-point system integration, the long-term vision is to achieve a single authoritative source of data by adopting a canonical data model. By streamlining the full data lifecycle from data sources to derived products and enabling interoperable access, we are improving the efficiency, reliability, timeliness, and usability of provincial geoscience information. The BCgSDI architecture is designed to be scalable, supporting advanced analytics and emerging technology such as large language models and artificial intelligence. Beyond the BCGS, the BCgSDI modernization approach provides a practical template for other agencies seeking to update their information systems. Ultimately, the BCgSDI architecture enhances the publication of analytic-ready datasets and maps, strengthens discoverability through standardized metadata and services, and reduces friction in collaborative, cross-jurisdictional geoscience initiatives.

Acknowledgments

This work has benefitted from the support and inspiration of many people. We appreciate Tina Qin and David Bayne for their contributions to improving and advancing the business case, and Dylan Dawson, Brian Gillespie, and Abigail Michel for strengthening the system architecture and shaping the decision on the Cloud-native platform. Yao Cui is grateful to Michael Ross, David Skea, Martin Davis, and Paul Ramsey, whose guidance, shared expertise, and thoughtful conversations have elevated my understanding of geospatial data systems and inspired key elements of this BCgSDI initiative. We thank Adrian Hickin and Marc-André Brideau for comments that helped improve this paper.

References cited

Brothers, D.L., 1968. Annual Report of the Mineral Industry of the Province for the year 1967. Department of Mines and Petroleum Resources, Province of British Columbia, 290 p.

California Learning Resource Network, 2025. What is a Canonical Data Model? <https://www.clrn.org/what-is-a-canonical-data-model/> (last accessed February 2026).

CGI IUGS EarthResourceML Working Group, 2013. EarthResourceML v2.0: An XML based data transfer standard for mineral occurrences, mines, and mining activities. IUGS Commission for the Management and Application of Geoscience Information (CGI). <https://earthresourceml.org> (last accessed February 2026).

Cui, Y., 2011. Improving the efficiency in the maintenance of the provincial geological database. British Columbia Ministry of Forests, Mines and Lands, British Columbia Geological Survey Fieldwork Paper 2011-1, pp. 159-168.

Federal Geographic Data Committee, 2024. Building the geospatial future together-The NSDI strategic plan 2025-2035: Reston, Virginia, Federal Geographic Data Committee, 10 p. <https://www.fgdc.gov/nsdi-plan/NSDI2035>

GeoSciML Modeling Team, (Eds.), 2017. OGC Geoscience Markup Language 4.1 (GeoSciML). Wayland, MA: Open Geospatial Consortium, 247 p. (OGC 16-008). <http://www.opengis.net/doc/geosciml/4.1> (last accessed February 2026).

Hohpe, G., and Woolf, B., 2003. Canonical data model. In: Enterprise Integration Patterns. <https://www.enterpriseintegrationpatterns.com/patterns/messaging/CanonicalDataModel.html> (last accessed February 2026).

Jones, L.D., and McPeek, C.B., 1992. MINIFILE-A mineral deposit information system for British Columbia. British Columbia Ministry of Energy, Mines and Petroleum Resources, British Columbia Geological Survey Information Circular 1992-02, 26 p.

Montgomery, J.H., Sinclair, A.J., Wynne-Edwards, H.R., Fox, A.C.L., and Giroux, G.H., 1975. Data capture in the construction of MINDEP's computer-based, mineral deposits files. Canadian Journal of Earth Sciences, 12, 698-703.

Open Geospatial Consortium, 2025. National Spatial Data Infrastructure (NSDI) Modernization Project. <https://www.ogc.org/initiatives/national-spatial-data-infrastructure-modernization-project> (last accessed February 2026).

U.S. National Research Council, 1993. Toward a coordinated Spatial Data Infrastructure for the Nation. Report of the Mapping Sciences Committee, National Academy Press, Washington, DC, 1993, 171 p. <https://openlibrary.org/books/OL10357307M/Toward_a_Coordinated_Spatial_Data_Infrastructure_for_the_Nation> (last accessed February 2026).