

---

# GEOH5: A Framework for Geoscience Data and Model Portability

John McGaughey<sup>1</sup>, Julien Brossoit<sup>1</sup>, Kris Davis<sup>1</sup>, Dominique Fournier<sup>1</sup>, Sébastien Hensgen<sup>1</sup>

<sup>1</sup>*Mira Geoscience Ltd, 309-310 Victoria Ave, Westmount (QC) H3Z 2M9 Canada.*

September 2023

## Abstract

We have developed a data structure called GEOH5 with the objective of general integration and storage of geological models, data, and metadata where dissemination, general access, and persistence are required. It answers the needs of modelers who require a structure that is compact, open, reasonably comprehensive in scope, and extensible. Although only a few years old, the GEOH5 data structure is already in use by thousands of users with increasing acceptance across the geosciences. This includes industry, academia, and geological survey organizations that are using GEOH5 as a documented, public, easy-to-use, vendor-neutral, and permanently accessible means of storing and disseminating models, data, and metadata.

GEOH5 is open source and free to use. It is based on open-source HDF5 technology because of its many advantages: wide acceptance across numerous data-intensive industries, self-describing behaviour through integration of data and metadata, fast I/O, excellent compression, file merging, cross-platform capability, unlimited data size, and access to libraries in a variety of programming languages. It provides both professionals and researchers with a robust means of handling large quantities of diverse data.

An open-source Python API called GEOH5Py facilitates reading from and writing to the GEOH5 data structure. A free, powerful GEOH5 reader called Geoscience ANALYST has been created to display the contents of GEOH5 files as tables, charts, documents, maps, cross-sections, and 3D visualizations. The combination of GEOH5, GEOH5Py, and Geoscience ANALYST provides a convenient and free mechanism for creating and sharing projects as well as immediately visualizing the results of Python modelling and data processing routines in the context of other data and model elements. Among other benefits, this allows researchers to focus on development of new methods rather than the creation of data structures, user interfaces, and visualization systems to support their work.

## Introduction

Barriers to interoperability, imposed by design or default by software vendors for commercial reasons, serve neither the interests of technology advancement nor the objectives of the data acquirers, interpreters, and researchers who need to disseminate their geoscientific data, metadata, and models. Geoscientists must often undertake complex and costly manual workarounds to share data and models among mutually non-interoperable systems, imposing costs as well as potential data loss and error introduction. The result is loss of productivity, poorer decision making, and dissatisfaction with proprietary systems.

We describe an open-format file structure, GEOH5 (Section 1), as a useful solution to the interoperability problem. We also describe an open-source Python API, GEOH5Py (Section 2), that provides a standard programmatic interface for reading from and writing to the GEOH5 format, and finally a powerful, free-to-use viewer of the content of GEOH5 files, Geoscience ANALYST (Section 3). The API and viewer are what make the GEOH5 file structure easy to use for geoscientists and promote its acceptance as a “standard”.

A useful analogy to GEOH5 is the ubiquitous Portable Document Format (PDF), an ISO standard that seeks to capture documents in a manner independent of application software, hardware, and operating system. In a broadly similar manner, GEOH5 provides an open, documented, extensible structure for storing and

---

sharing geoscientific models, data, and metadata. The structure is aligned with the FAIR guiding principles for making data Findable, Accessible, Interoperable, and Reusable (LIGHTSOM ET AL., 2022).

## 1. GEOH5: an open format for geoscience data and models

GEOH5 is a documented public, open, easy-to-use, vendor-neutral, and permanently accessible data exchange and storage format for the general geosciences. The power of GEOH5 lies in its capacity to handle various types of geological data—from point, curve, and surface data to drillholes, geophysical data, and 3D models. The format facilitates interoperability between different software, fostering a collaborative environment for geoscientists, researchers, analysts, and other stakeholders, including for public dissemination. It provides a unified format that bridges the gap between different software tools.

GEOH5 has its roots in the Hierarchical Data Format (HDF5), a universally accepted and widely used data model, library, and file format for storing and managing complex data. HDF5's attributes make it an obvious choice as a foundation for an open geoscience data standard: wide acceptance across numerous data-intensive industries, self-describing behaviour through integration of data and metadata, fast I/O, excellent compression, file merging, cross-platform capability, unlimited data size, and access to libraries in a variety of programming languages. It provides both professionals and researchers with a robust means of handling large quantities of diverse data. The content of GEOH5 files is readable and writeable by third-party software using scientific programming environments such as open-source HDFview, Python, MATLAB, Fortran, C, and C++. As an illustration of accessing GEOH5 content from C++, we provide GEOH5 importers and exporters as SKUA-GOCAD™ add-ons.

### 1.1 GEOH5 Data Structure

GEOH5 facilitates efficient data management and processing. Building upon the strengths of HDF5, GEOH5 introduces an effective structure to encapsulate geological data, including spatial and attribute information. The format employs a compact and intuitive tree structure, ensuring quick access to data and simplified data processing. This feature reduces the time spent on data retrieval and manipulation, significantly enhancing overall productivity.

The main structure of the GEOH5 format is shown in Figure 1, as displayed by the free HDFview program<sup>1</sup>. Groups, Objects and Data entities are stored in flat structures and indexed by a unique identifier as specified by the RFC 4122 standard<sup>2</sup>. Entities hold references to their own children for rapid navigation. At the top level, the Root container contains pointers to the full hierarchy of the file, providing the complete linkage between all entities and their dependents, ensuring a seamless and organized structure for efficient access and retrieval of information.

Groups are simple containers for other groups and objects. They are often used to assign special meanings to a collection of entities or to create specialized software functionality.

The current set of Objects implemented in GEOH5 supports a range of geological, geophysical, geotechnical, and mining data and model elements that can be attributed with properties: points, curves, surfaces, volumetric domains, drillholes, drillhole targets, rectilinear 2D grids, 3D grids, octree 3D grids, VP (vertical parameterization) grids, raster images, thin plates (to support electromagnetic modelling), airborne and ground EM transmitters and receivers, airborne and ground gravity and magnetic surveys, magnetotelluric surveys, tipper (ZTEM) surveys, microseismic events, ground deformation, plus various minesite data types.

Data are currently always stored as a 1D array, even in the case of single-value data. New data types can be created at will by software or users to describe object or group properties. Data of the same type can exist on any number of objects or groups of any type, and each instance can be associated with vertices, cells, or the Object/Group itself. Some data type identifiers can also be reserved as a means of identifying a specific kind of data. Data attributes include specification of the primitive type with optional descriptive metadata (e.g., units

---

<sup>1</sup> <https://www.hdfgroup.org/downloads/hdfview>

<sup>2</sup> <https://datatracker.ietf.org/doc/html/rfc4122>

and text description) and display parameters to be used by a viewer. Primitive types include float, integer, text, referenced or categorical, datetime, filename (which must correspond to a stored binary file as a data instance), and blob (which must correspond to a binary dataset as a data instance).

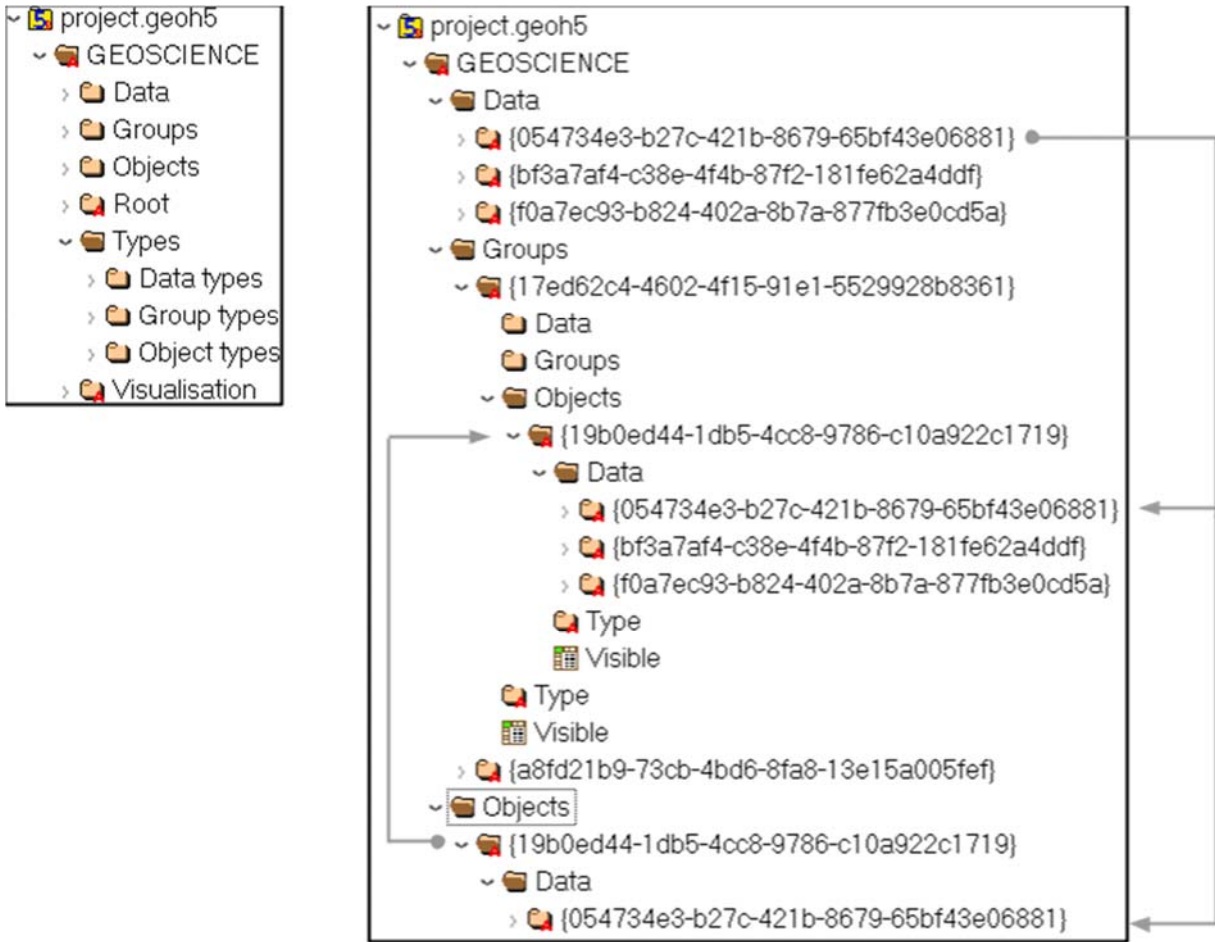


Figure 1: At left, main structure of the GEOH5 file format. At right, Data, Groups and Objects entities are stored in flat HDF5 containers, each indexed by a unique identifier. Pointers to the child entities are given for rapid navigation through the tree structure.

## 2. GEOH5Py: An open-source API

We created an open-source Python API to facilitate reading from and writing to GEOH5 format. With GEOH5Py, it is simple to build an application to read and write GEOH5, or to conveniently add GEOH5 to the import and export types supported by other software platforms. For example, we have used GEOH5Py to provide a conversion between the Open-Mining Format (OMF)<sup>3</sup> and GEOH5.

With the help of the API, users can easily create, modify, and remove objects and data programmatically. The main component is the Workspace class. It handles all read/write operations performed on GEOH5 with simple function calls, as demonstrated in Figure 2. This high-level interaction with the GEOH5 storage format allows practitioners to easily leverage the rich Python ecosystem to build their own custom processing routines. GEOH5Py itself relies on the open-source NumPy and H5py packages.

<sup>3</sup> <https://gmgroup.org/projects/data-exchange-for-mine-software>  
 GEOH5: Data and Model Portability  
 McGaughey et al.

Full documentation describing the GEOH5 format<sup>4</sup>, and its GEOH5Py API, are available online and updated with every release.

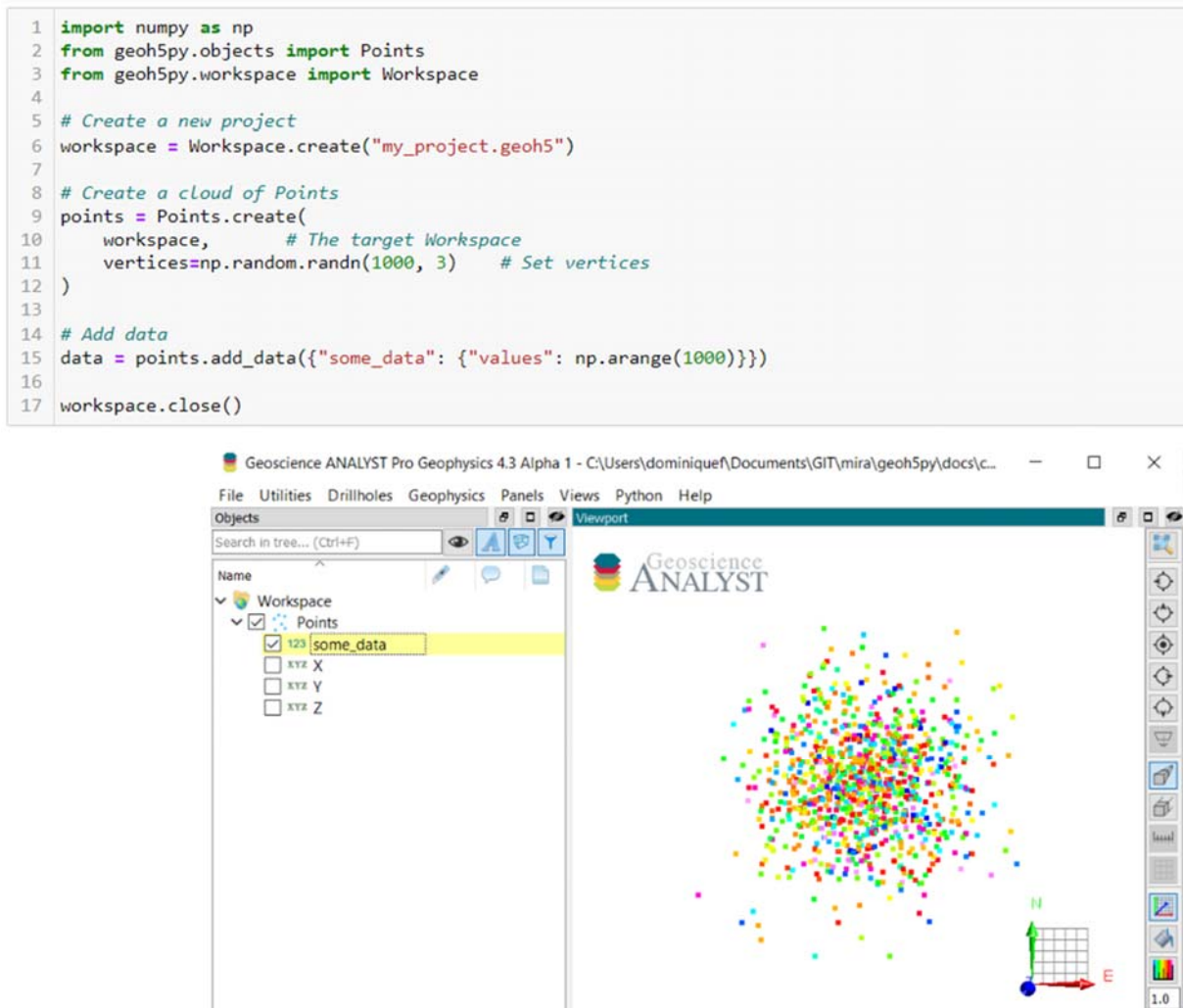


Figure 2: Example demonstrating the creation of a new GEOH5 file containing a Points object and associated data, with the file contents viewed by the Geoscience ANALYST reader.

### 3. Geoscience ANALYST: a free GEOH5 viewer

The utility of the freely downloadable<sup>5</sup> Geoscience ANALYST reader is a principal motivation for geoscientists to adopt GEOH5. It is a powerful viewer that displays GEOH5 file data and metadata in tables, charts, documents, maps, cross-sections, and 3D visualizations. In the PDF analogy to GEOH5, Geoscience ANALYST plays the role of the freely downloadable Adobe Acrobat reader—the existence of which is a principal motivation for users to adopt the PDF document standard. However, in contrast to the Acrobat reader, the Geoscience ANALYST reader can also import additional data and save them back to the GEOH5 file.

It is intended that Geoscience ANALYST preserves data it does not understand (and generally be very tolerant with regards to missing information) when loading and saving GEOH5 files. This will allow third parties to write this format easily, as well as to include additional information for their own purposes that is

<sup>4</sup> [geoh5py.readthedocs.io](https://geoh5py.readthedocs.io)

<sup>5</sup> <https://mirageoscience.com/mining-industry-software/geoscience-analyst>

not included in this formal specification. In the current implementation, Geoscience ANALYST automatically removes unnecessary information on save.

Geoscience ANALYST presents data object and property names in a conventional tree structure. Currently supported object types are points, curves, triangulated surfaces, drillholes, 2D (map) grids, 2D geophysical grids (curved in X-Y, vertically-oriented, and topographically-draped), multiple types of 3D grids (regular cell size, ‘tartan’ grid, octree grid, vertical prism), and rasters. It provides multiple, linked object and property visualization modes: 3D cameras, 2D map views, cross-sections, 2D data profiles, decay curves, drillhole monitoring, scatter plots, box-and-whisker plots, histograms, and tabular data displays. When one or more points are selected in any of the display panels, the same points are indicated in all open display panels.

The combination of GEOH5, GEOH5Py, and Geoscience ANALYST provides a dynamic environment for research and software prototyping for geoscientists because of how easily it connects open-source Python libraries, open-source GEOH5 and GEOH5Py, and a free and powerful 3D viewer into which a wide array of contextual data and models elements (such as drillholes, geophysical data, geological models) can be easily imported. Figure 3 demonstrates a simple example in which the output of a Python data processing code written in a Jupyter notebook is easily displayed in 3D in Geoscience ANALYST, using GEOH5 as the common data structure.

This capability has enabled us to create a repository of open-source geoscience applications called “geoapps”<sup>6</sup> that, with public additions, could become a central repository to interfaces and applications including geological and geophysical data processing, modelling, and inversion codes.

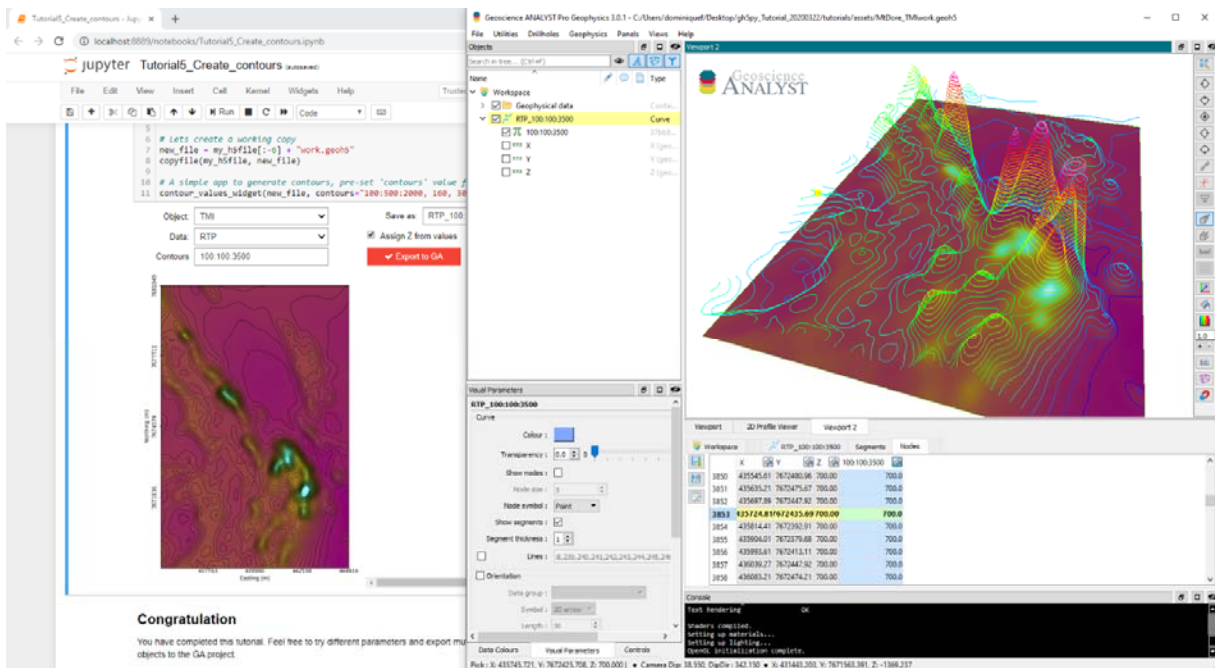


Figure 3. A Python processing routine in a Jupyter notebook (at left) provides its output as a new or updated GEOH5 format file, the 3D visualization of which is refreshed with a click in the Geoscience ANALYST application at right.

(We have also created paid versions of Geoscience ANALYST that fully encapsulate open-source and proprietary processing and modelling functions, and that permit users to add access to Python applications directly to the Geoscience ANALYST user interface—see Figure 4.)

<sup>6</sup> <https://pypi.org/project/geoapps>  
GEOH5: Data and Model Portability  
McGaughey et al.

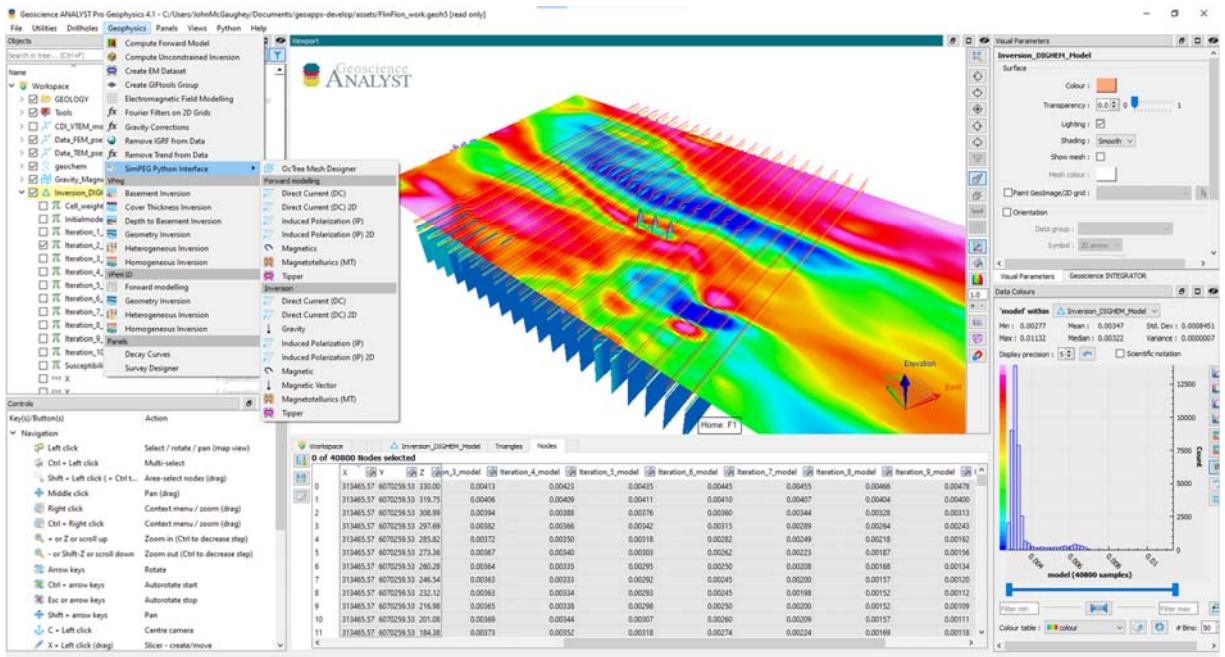


Figure 4. A version of Geoscience ANALYST illustrating the embedding of open-source “SimPEG”<sup>7</sup> Python geophysical inversion codes directly into the menu system.

## Conclusions

Although only a few years old, the GEOH5 data structure is already in use by many thousands of users with reasonably broad acceptance across the minerals industry. This includes geological survey organizations that are using GEOH5 as a convenient, compact, and permanently accessible means of disseminating models and data with embedded metadata. Anyone can build an application to read and write GEOH5, or conveniently add GEOH5 to the import and export types supported by modelling platforms.

In addition to portability, the freely available data structure, API, and visualization system provides significant benefits to open-source geoscience modelling initiatives, allowing modelling researchers to focus on modelling technology rather than the creation of data structures, user interfaces, and visualization systems to support their work. The Python API provides a convenient mechanism for immediately visualizing the results of Python modelling and data processing routines in the Geoscience ANALYST viewer at no cost, relieving Python application developers of the need to re-invent geoscience domain interfaces and visualization methods.

## References

LIGHTSOM, F.L., HUTCHISON, V.B., BISHOP, B., DEBREW, L.M., GOVONI, D.L., LATYSH, N., AND STALL, S. (2022).

Opportunities to improve alignment with the FAIR Principles for U.S. Geological Survey data: *U.S. Geological Survey Open-File Report 2022–1043*, 23p. <https://doi.org/10.3133/ofr20221043>

<sup>7</sup> <https://simpeg.xyz>