Newsletter

Photo: GeoInfo Summit participants, May 19th

In This Issue

1. Feb 24th GeoInfo Summit – Harvey Thorleifson
    - CGI – François Robida
    - CGMW – Manuel Pubellier and Benjamin Sautter
    - Codata – Alena Rybkina
    - DDE – Junxuan Fan
    - Loop – Laurent Ailleres
    - OGC – Mickaël Beaufils
    - OneGeology – Matt Harrison
    - CGI renewal – Mark Rattenbury
2. May 19th Discussion – Harvey Thorleifson

Feb 24th GeoInfo Summit – Harvey Thorleifson

CGI is the International Union of Geological Sciences (IUGS) Commission for the Management and Application of Geoscience Information. The goal of CGI is to foster the interoperability and exchange of geoscience information, through community leadership, collaboration, education, as well as development and promotion of standards and best practices.

On 24 February 2022, at 10 AM UTC, CGI and partner agencies hosted a 3-hour online GeoInfo Summit attended by geoscience information leaders from around the world. CGI Chair Harvey

Thorleifson stated that geoscience information is of escalating importance to worldwide society, with increasing emphasis on standardized, machine-readable, data-rich knowledge.

He noted that respected international geoscience information organizations happily support each other, and we are coordinating our plans, for example as we prepare for upcoming high-level IUGS meetings. He was pleased that there had been unanimous support for this joint, informal meeting, which had not included ambitious pre-meeting effort.

There were talks from CGI, CGI Work Groups, CGMW, Codata, DDE, Loop, OGC, OneGeology, and CGI Planning. Speakers had been asked to present an overview of their background, status, and plans, as well as mention of leadership recruitment, government-university-industry coordination, ensuring sustainability, and defining coordinated and achievable goals.

Harvey expressed appreciation to the speakers whose presentations are summarized here. Subsequently, on May 19th, a GeoInfo Summit discussion session was held.

Please visit the **CGI web site** at **https://cgi-iugs.org/**, and if you aren't a member, please click the **Join Now** button!

CGI – François Robida

François, CGI Past Chair, described how CGI, which focuses on geoscience while benefiting from broader Open Geospatial Consortium (OGC) arrangements, emanated from a 2003 Edinburgh meeting of geological survey and industry people, when multiple geological data model initiatives in the context of new standards such as GML, WMS, and WFS resulted in a consensus to build a common model that would permit the interoperability that is essential for humans and machines to conduct data-driven science.

To overcome the tendency for everyone to reinvent their own standard, he asserted that we must work together on standards that will be built by recognized and respected, broadly knowledgeable scientists, data scientists, and technical experts, and that are needed, endorsed, and used by the community.

François stressed the need to manage scientists' tendency to over-specify complexity, to improve or extend existing standards, to reconcile standards in related fields such as multiple borehole standards, to import procedures from related fields such as 3D, to publicize our work broadly, and to validate our work through testbeds and flagships, through an open consensus-building process.

In 2022, he sees a need for renewed coordination among geoscience information organizations, in part by clarifying complementary roles, as well as with industry

and academia, as we did in 2003, although we now have a new context such as semantic web, linked data, big data, AI, and 3D, and a need for us to coordinate with ocean and atmosphere. In closing, François stated that we can do nothing without standards.

CGI Work Groups – Harvey Thorleifson

To praise superb work by CGI Work Groups and partner groups, Harvey presented a summary of the CGI webinar that had been held in June 2021, that is available on YouTube, and that was summarized in the August 2021 CGI Newsletter. He also mentioned the September 2021 technical seminar on geoscience ontology and knowledge graph that also is available on YouTube. In addition, he described the comprehensive, 15-hour DDE Geoscience Standards Training in October 2021, for which videos are available on the DDE web site.

He described how the OGC Geoscience Domain Working Group (DWG) partners with CGI and others to make geoscience information findable, accessible, interoperable, and reusable (FAIR). Key DWG goals are interoperability, data definitions, formats, and services for publishing, search, and exchange; thematic/semantic coherence with related domains; interfaces with city/infrastructure and risk domains; showing value added by interoperability; best practices; and standards for spatial and temporal features, metadata, and other information.

In addition, an OGC Standard Working Group (SWG) manages GeoSciML (Geoscience Markup Language), an XML- and GML-based machine-readable format for geological maps. EarthResourceML (ERML) is a markup language for the delivery of mineral occurrence, deposit, mining, and resource data. Nineteen ERML vocabularies include mine status, commodity, and reserves; data delivery and implementation are well accommodated.

The CGI Geoscience Terminology Working Group maintains terms for earth features, properties, quantities, techniques, and processes that are agreed upon and made widely accessible, including to machines. Geoscience ontology research is being conducted in association with Loop. The Deep-time Digital Earth (DDE) Standards Task Group (DDE-STG) is facilitating adoption of CGI standards and terminology by DDE, the first IUGS-recognized big science programme that was launched in 2019.

CGMW – Manuel Pubellier and Benjamin Sautter

Manuel, with Benjamin Sautter, provided an update on the work of the Commission for the Geological Map of the World (CGMW), an international non-profit association governed by French law that is responsible for designing, coordinating, preparing, and publishing small-scale thematic earth science maps of the globe, continents, major regions, and oceans. CGMW was established as an outcome of the International Geological Congress (IGC), first held in 1878. The commission was conceived in 1910 at the 11th IGC in Stockholm, and established in 1913 at the 12th IGC in Toronto.

CGMW has several thematic sub-commissions, and an important role in preserving the legacy of

important paper maps produced over a century and more, while using current technology to produce the geological mapping databases that are needed for our future. The maps need to be freely accessible on an international basis, and at the same time, someone has to pay. CGMW is working, particularly with CGI and OneGeology, to take forward thinking on development of global seamless 3D databases at multiple levels of resolution.

An exciting and important current activity is a new 1:5M seamless geological map of the world, being produced in association with DDE, including much reconciliation aided by imagery and elevation databases, and updated bathymetry and seafloor geology, including mid-oceanic ridges, transforms, seafloor age, and hydrothermal vents.

Codata – Alena Rybkina

Alena described the role of Codata, the Committee on Data of the International Science Council, in building a foundation for a world of open data, by promoting open science and FAIR data. CODATA convenes a global expert community and provides a forum for international consensus building and agreements around a range of data science and data policy issues, from the fundamental physical constants to cross-domain data specifications.

Codata has national committees, scientific union members, institutional members, and partner organizations. Codata advances data policies, data science, data skills, and data documentation. While citing a report on Advancing Science as a Global Public Good, she stated that the major, pressing global scientific and human issues of the 21st century can only be addressed through research that works across disciplines to understand complex systems, and which uses a transdisciplinary approach. She noted that there are major challenges for many scientific domains that require work on data specifications, semantics, infrastructures, and more, as estimates imply 80% of our effort is spent on data wrangling, due to suboptimal data stewardship.

Nevertheless, much cross-domain collaboration is quickening, to address current global grand challenges. Alena then outlined the FAIR guiding principles for scientific data management and stewardship, and a document on Turning FAIR into reality. The Codata data documentation initiative is a cross-domain integration collaboration designed to interface with other standards and to help interoperability between different data types, standards, and formats.

DDE – Junxuan Fan

Junxuan Fan described DDE, for which the kick-off meeting was held at Beijing in February 2019, with representatives of 40 geoscience organizations in attendance, and with 12 founding organizations signing the accord to initiate the DDE program. DDE is a big science program dedicated to facilitating innovation in understanding the Earth's evolution and applications as well as the Sustainable Development Goals utilizing Big Data analytics, internet cloud computing, data mining, machine learning, and AI. The principal theme is

harmonizing global deep time geoscience data. The governance includes a governing council, executive committee, science council, science communities, working groups, and task groups. The primary goals are to establish an international consortium, linking isolated databases into an open, distributed system, making data FAIR, linking earth spheres, and promoting innovative earth system research, through data-driven knowledge discovery.

DDE data-driven knowledge discovery is addressing: 1. Integrating a uniform high-resolution earth time system; 2. Origin and evolution of life and biodiversity; 3. How did sedimentary matter evolve and cycle?; 4. Reconstructing earth climate and atmosphere history from big data of multiple geochemical indices; 5. Global sea-level change through deep time; 6. Quantifying plate tectonics and deformation in 4D; 7. 4D architecture and evolution of deep-earth materials and dynamics; 8. Mineral evolution beyond 4D; 9. A globally shared big-data energy resource system for sustainable development; 10. Geophysical fields for prediction of seismic hazard.

Loop – Laurent Ailleres

Laurent described Loop, an integrated and interoperable platform enabling 3D stochastic geological modelling. The focus of Loop is on providing solutions for subsurface resource management, managing geological knowledge, 3D geological and geophysical modelling, as well as uncertainty risk mitigation. Loop knowledge management is using AI for knowledge extraction from literature, maps, and reports using geological ontology, with geological rules being encoded to ensure proper knowledge extraction.

Loop is a time-aware geological modelling platform, and the geological event management is capturing topological and time relationships between geological objects and structural events. Loop forward and inverse structural modelling is encoding structural geological rules in a time-aware context to account for folds, faults, shear zones, unconformities and intrusions. The modelling is based on probabilistic modelling and allows for the definition of an objective function for geology and quantification of uncertainty via posterior probabilities.

A milestone was reached recently with release of the GeoScience Ontology (GSO). In addition, Map2Loop data pre-processing and input now involves automation of data input for Loop modelling methods, geometrical analysis of geological maps, and topological analysis of the map. Secondary geological information automatically derived from maps includes normalized local formation thickness, and apparent fault throw. Loop Resources involves reducing the mining footprint with curvilinear geostatistics using structural frames.

Plans for the future include Loop Vision, an interoperable, integrative, automated and update-able workflow to model geology based on structural geological rules and geophysics. A platform will integrate geophysical modelling/constraints upstream with structural geological modelling/constraints. Loop provides a means to characterize and mitigate uncertainty that will guide further data acquisition. The framework accommodates lithology, alteration, mineralogy, metallurgy,

geotechnical parameters, and grades, allowing a better characterization of mining block and process optimization.

OGC – Mickaël Beaufils

Mickaël provided an overview of OGC and the GeoScience DWG, that partners with CGI and many other organizations in making geospatial information findable, accessible, interoperable, and reusable (FAIR). For findability, he cited persistent identifiers (PIDs), rich metadata, indexed data repositories, and PIDs in metadata; for accessibility, he cited standard communication protocols, open and free protocols, authentication where necessary, and metadata always available; for interoperability, he cited vocabularies, FAIR vocabularies, and linked metadata; and for reusability, he cited multiple attribute metadata, usage licenses, provenance, and community standards.

He outlined the OGC innovation continuum, in which an innovation cycle links to a standardization cycle. The innovation cycle involves project identification, call for sponsors, creation of a public call, selection of members, an executive project, and generation of results. This links to the standardization cycle, built on domain and standards working groups, involving DWG discussion, SWG formation, standards development, DWG/SWG presentations, formulation of enhancements, and interest gathering, which links back to the innovation cycle.

Key DWG goals are interoperability, data definitions, formats, and services for publishing, search, and exchange; thematic/semantic coherence with related domains; interfaces with city/infrastructure and risk domains; showing value added by interoperability; best practices; and standards for spatial and temporal features, metadata, and other information. Emphasis has been on the BhML borehole conceptual model, as well as (Geo)JSON(- LD), 3D-4D, BIM, geotechnics, ISO19156 for observations and measurements, OGC SensorThings API, OGC-API features, and SMART Data Loader. Much emphasis is now on geotech interoperability.

OneGeology – Matt Harrison

Matt spoke on OneGeology, which has 189 member organizations, and is dedicated to providing geoscience data globally. OneGeology technical principles are based on interoperability with emphasis on GeoSciML, data on distributed servers sent directly to the web client, participants delivering map data through a standard OGC web service, cookbooks and online help and buddy systems so all nations can serve their data, a central catalog of services, and a portal that can display and aggregate all of the maps.

According to the 2007 Brighton Accord, OneGeology is a geological survey initiative that will make public and internet-accessible the best available digital geological map data, with priority on interoperable, Internet-accessible, scientifically-attributed data and to make progress at levels appropriate to participants' capability.

OneGeology is now migrating its focus to the application, technology and opportunities arising from developing a multiscale suite of digital twin earth system models. These will enable the better prediction and understanding of subsurface processes and their properties, and how these influence the world around us. The existence and use of an earth system digital twin could, for example, allow users to sample a volume of rock and run numerical simulations to understand the impacts of various inputs and changes to that system, such as alternative energy storage technologies.

These digital twin models will be an ensemble of 4D geospatial data, models, and visualizations. One of the key elements will be the cyber–physical interaction, whereby sensor perturbation monitored in the real world will ultimately be reflected within the virtual twin model in near-real time. A realistic model of the natural environment will provide new virtual opportunities to test and improve our understanding of these systems and to model and predict their behaviors into the future.

CGI renewal – Mark Rattenbury

As an illustration of planning and renewal, Mark, a CGI Councilor, discussed how CGI has initiated draft steps toward renewal. He described how the Commission for the Management and Application of Geoscience Information (CGI) is a Commission of the International Union of Geological Sciences (IUGS), and is an international scientific, technical non-governmental, non-political, and non-profit-making organization.

Mark then outlined how CGI aims to: help in exchanging knowledge on geoscience information and systems; support dissemination of best practice in geoscience information; encourage the development of geoscience information standards; play a full part in the international geoscience information community and represent IUGS on geoscience information matters; bring together individuals and organizations with an interest in the management and application of geoscience information; and provide open access to the standards and outputs it delivers, for the benefit of all.

He then noted that CGI: has been very successful and influential since its inception; has worked widely across the international community; has led with development of the GeoSciML and EarthResourceML data; has led model standards for information transfer with publishing of internationally agreed geoscience vocabularies; and continues to support these and other initiatives such as OGC's Geoscience Domain Working Group and IUGS's Deep-time Digital Earth.

He further described how CGI collaborates on focused projects; provides continuity between peaks in project activities; and offers trusted, stable and enduring governance capability to the geoscience information community.

He then suggested that CGI should: assess the maturity of current geoscience information standards; identify opportunities that lie ahead in the geoscience information standards domain; signal what could be achieved in the next four years; signal what could be achieved in

the next ten years; and identify collaborators and resources that could help achieve the goals.

May 19th Discussion – Harvey Thorleifson

The 90-minute GeoInfo Summit discussion session on May 19th at 10 UTC was a great success. Partner organization leaders noted that technical challenges cannot be addressed by a single organization. We need to address societal challenges, to attract young people, and to work efficiently. Leaders stressed the importance of concrete products. We emphasized the need for a shared vision that we all can rally around.

We asserted that we need to be needed, by making sure that our activity is relevant, in part by utilizing new capabilities. We stressed sustainability of our activity, and efficiency through coordination. We stated that standards are maintained if used, and we need to find a problem that calls for what we do. We thus saw the need for a community that uses what we do, with standards reaching critical mass.
We then talked about how to go forward. We noted that in-person meetings allow serendipity, social interaction, and meeting people. We then focused on a global solid earth digital twin as a focal point, perhaps to be a highlight of upcoming IUGS meetings. We saw that we need a workshop that reaches conclusions. We understand that we need a goal, for initial delivery at the IGC in 2024.

It was noted that a global solid earth digital twin would be inclusive of everyone, under a 'Save the Planet' sort of banner that attracts young people, for very legitimate reasons. We discussed the concurrent digital twin of the oceans, and there was mention of soil/geology coordination. We recognized that we should not promise something by a date, or stress that major funding should come first – but rather we need a roadmap for what each organization needs to do.

In closing, Harvey Thorleifson noted that we are operating as a multi-agency network, so each agency would have to respect its governance. He suggested that we plan for additional multi-agency, online meetings in September, to take the next step in bringing the solid earth digital twin concept into focus, for example in coordination with DDE.

SAVE THE DATE:

On September 20-21, 2022, for 90 minutes each day at 10 UTC, we will hold **GeoInfo Summit Follow-Up**, to be attended by leaders of global geoscience information organizations.

If you wish to attend, please contact the **CGI Chair** at thorleif@umn.edu.