OGC Testbed-19 Call for Sponsors

Version 1.0 - 22 Nov 2022

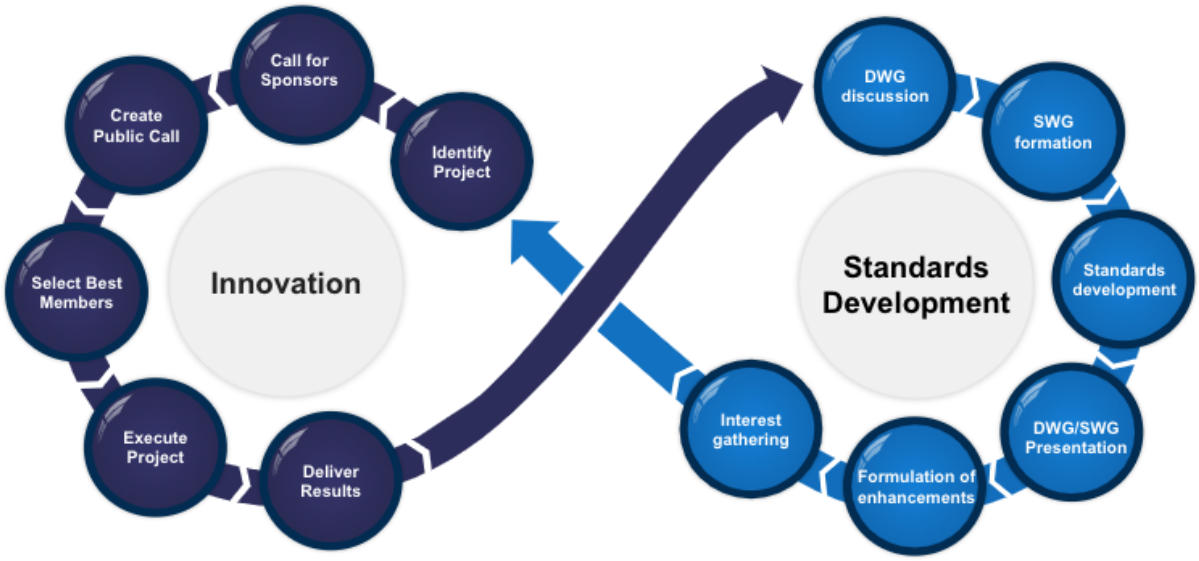
Chapter 1. OGC Testbed Introduction and Benefits

OGC Testbeds are OGC's largest Collaborative Solution and Innovation (COSI) Program initiatives The COSI Program is formerly known as the Innovation Program. Testbeds boost research and development to make location data and information more FAIR: Findable, Accessible, Interoperable, and Re-Usable. Additionally, testbeds provide a unique opportunity for sponsors to tackle location data and processing challenges together with the world's leading geospatial IT experts.

Testbed Sponsor benefits include:

- Outsource research to global consortium of geospatial experts
- Cost share with others - Higher ROI for each sponsor
- Reduce technology risk
- Stay on top of trends
- Show leadership
- Influence capabilities and tools

Challenges put forth by sponsoring organizations are refined and mapped to a set of work items that OGC member organizations will compete to address. The COSI team together with the OGC member sponsors will then select the most qualified organizations to join Testbed-19. In a collaborative effort, all Testbed participants, sponsors, and the OGC team work jointly on the goal to stepwise increase Technology Readiness Levels (TRL) of geospatial IT solutions, including software architecture, interface design, information and data models, as well as related standards and specifications.



Chapter 2. Synergistic Effects

OGC Testbeds provide a unique opportunity to explore technologies that appear disconnected at first glance. Combining these technologies in a single initiative and bringing several sponsors together allows us to create an interoperability environment that comes much closer to real-world situations. In consequence, Testbeds allow leveraging an outstanding quality of synergetic effects and address challenges that require collaboration among several sponsors and experts from member organizations.

Chapter 3. Technical Scope

Building on the success and outcomes of Testbed-18 and Testbed-17, as well as other ongoing COSI initiatives, OGC is now starting the preparation of Testbed-19.

The following technology areas have been identified in early conversations with sponsors, geospatial data experts, and OGC members. This list is not exhaustive, and we welcome additional topics.

- **Analysis Ready Data (ARD)** - There is a clear interest from OGC members in developing a suite of standards to define an analysis-readiness framework and support a series of domain-specific specifications for creating ARD products from Earth Observations, both space-based and collected in other ways. This interest extends to addressing the challenges of ARD production and exploitation workflows, as well as what OGC standards such as API's can be used to implement them.

- **Agile Reference Architecture** -The development of an Agile Reference Architecture (ARA) to guide development and application of OGC standards is seen as a multi-year Testbed activity. The OGC Reference Model (ORM) was developed in 2011. As the technological landscape has significantly evolved and location information is truly everywhere, there is an increasing need to understand how the ORM should be adapted to address the challenges that stakeholders face today.

- **Open Science** - The goal of Open Science is to maximize scientific productivity and ingenuity through open sharing of resources, methods, and results across organizations involved in both basic and applied scientific research. This requires provision of robust interoperability mechanisms between and around the information systems that participating organizations depend on. Data discovery and access mechanisms, data and workflow description, and requests for processing capacities need to be harmonized.

- **Open Persistent Demonstrator and Community Building** - A common goal among earth observation organizations is to grow the user community for their data, products, and technologies. To realize this goal, it is necessary to have a persistent demonstration of current capabilities available together with training and educational materials that help the community to repeat and adapt scientific experiments, to build new workflows by adapting available material, and by clear guidelines on how to describe used materials, applied processing, and produced results.

- **OGC APIs** - OGC API standards are built upon the legacy of the OGC Web Service standards (WMS, WFS, WCS, WPS, etc.), but define resource-centric APIs that take advantage of modern web development practices. A multi-year Testbed activity is foreseen to continue the evolution, adoption, and implementation of these critical

standards. Specific goals include advancing a building block approach for custom API implementation and accessing a wider range of computational capabilities through API's

- **Cloud Native Geospatial Standards** - Cloud-native geospatial offers many benefits to location data users ranging from decreasing the burden on data providers, to drastically lowering the costs of managing that data. While significant progress has been made in this area, additional research is needed to continue the evolution of Cloud Native Geospatial Standards and identify gaps from current needs and traditional standards. Suggested targets have included cloud native metadata, vector tile, and point cloud access and additional topics are welcomed.

- **CDB Metadata Evolution -** Interoperability and transformation between different attributes schemes is always a challenge. One of the persistent and important concerns about the OGC CDB standard is how to evolve beyond the current use of CDB feature codes and attributes. A modern core OGC CDB application schema is a way to modernize the CDB 2.0 feature and attributes and consider an extension mechanism to the core using various profiles. The CDB 2.0 core conceptual model should not mandate any particular data dictionary or content, but rather provide the conceptual and logical metamodel for describing any ISO 19109 compliant application schema to the maximum practical extent. After that various profiles (e.g. simulation and modeling profile) can be developed and a mechanism will be designed to develop any extension profile for CDB for any particular data dictionary that complies with ISO 19109.

- **Coordinate Reference Systems Beyond Geospatial** - This task will cover 4D spatial temporal coordinate systems from seabed to space. Testbed 18 evaluated current standards with respect to the exact positioning of sensors at any location within the solar system and their corresponding data streams. The next step is to evaluate Implementation Specifications. Use cases have been identified and this task seeks additional sponsor participation as well as sample data that should be realistic but does not have to be authentic.

- **Data Centric Security** - Several items have been proposed to extend DCS within the geospatial realm:
    - Investigate OGC Encoding Standards and determine their streaming capabilities and recommend their use with and without DCS applied.
    - Define how to apply DCS to all OGC APIs, with a JSON or XML-based container.
    - Migrate from access tokens in Bearer HTTP header field to access tokens in DPoP HTTP header field demonstrating proof of possession (DPoP).
    - Investigate and define what vulnerabilities exist in the protocols, APIs and implementations that may enable an adversary to attack certain parts of the system.
    - Determine needed characteristics to OGC APIs and in particular the DCS extensions required to ensure the concept of ownership for OGC APIs and data

produced and stored within.

- **Geospatial Data Warehouses** - A data warehouse is an integrated, analysis ready data store created from two or more disparate data sources. They are typically used as data provision sources for business intelligence, simulation, and prediction applications such as machine learning models. Modern data warehouses often take the form of cloud services. These services, however, may have little or no capability for geospatial data analysis or provisioning. This task seeks to advance geospatial interoperability between existing cloud-based data warehouses and evaluate GeoParquet in particular as a possible standardized exchange format. GeoParquet defines a common, cloud-compatible way to encode and describe spatial data, but is currently still in an early phase of community development.

- **Discrete Global Grid System (DGGS):** A DGGS represents a spherical partitioning of the Earth's surface into a grid of cells. DGGS offers a new way for geospatial information to be stored, visualized, and analyzed. Based on a partitioning of the Earth's surface into a spherical grid, DGGS allows geospatial information to be represented in a way that more intuitively reflects relationships between data and the Earth's surface. With DGGS, providers and consumers of geospatial information can eliminate many of the uncertainties and distortions inherently present with traditional coordinate systems, and continue efforts towards digital earth concepts. To fully realize the benefits of DGGS, standard-compliant implementations are required to allow cell-id management across DGGS with varying structure and alignment. This task aims to begin the process to move towards an OGC Implementation of the DGGS API through the creation of an open-source based DGGS reference implementation and advance discrete analysis systems with operational DGGS capabilities suitable for large scale environmental modeling and analysis.

- **Digital Twins** - Digital twins integrate a variety of sensor systems with digital representations of the natural and/or built environment to create a digital representation of reality that is both holistic and targeted at specific applications. The goal of this task is to go beyond the often tightly integrated and isolated existing digital twin systems to generate new knowledge by exchanging as-built and real-time data between these systems. One of the core problems to address is efficient cross-domain data workflow for digital twins, as system boundaries are broken down and data is provisioned into new contexts. A new, cross-disciplinary perspective must be supported with best practices and guidance for the efficient and complete 2-way provision of data into and between digital twin systems. A specific focus will be on data models and OGC APIs to implement this perspective.

- **Federated Cloud Analytics** - Many systems have been or are currently being outsourced to the cloud. This leads to integration problems when data and processing capacities from several clouds are to be shared. This is not just a technical problem. There are different licenses, security measures, privacy models, governance approaches, and

business models to coordinate. It is a balancing act between transparency on the one hand and ease of use on the other. Users need to understand a specific constellation, but don't want to work their way through endless lists of licenses, credit card entry fields, or passwords. This task shall exercise a number of multi-cloud situations to fully understand the problem and to develop best practices and guidelines.

- **GeoPackage** - A GeoPackage is an open, standards-based, platform-independent, portable, self-describing, compact format for transferring geospatial information. The GeoPackage Conceptual Model Standard describes the model for encoding vector features, tile matrix sets of imagery and raster maps at various scales, attributes (non-spatial data), and extensions. For Testbed 19, there is interest to implement the GeoPackage standard using technology other than a SQLite database. Another GeoPackage challenges concerns efficient provisioning of incremental change sets for existing GeoPackaged data products

- **GeoDataCubes** - The OGC has explored and developed the concept of GeoDataCubes for some time. Currently, a new Standards Working Group (SWG) is being formed to make agreements and enhance interoperability between existing GeoDataCube (GDC) schemes and implementations. Testbed-19 can support this SWG in its efforts to create specifications such as a GeoDataCube API, architecture, and exchange formats by providing an experimental laboratory where rapid prototyping can address both technical challenges and criteria for decisions on GeoDataCube specification hurdles.

- **GeoDataCube API Implementation** - Significant research on this API has been completed in Testbed-17. The next step for GeoData Cube API realization will involve further prototype implementations to test operational and commercial models. Techniques for advanced visualization of GeoDataCubes should also be considered.

- **High Performance Computing** - Supercomputing and HPC capabilities are ever more advanced and available for use in different computing environments, as well as ever more necessary for large-scale processing and modeling of observational data. They continue to be operated, however, as unique and relatively closed resources that require users to develop specialized expertise as well as customize both their code and workflows. This Testbed-19 task, in collaboration with NSF-funded I-GUIDE investigators, will focus on integrating access to disparate supercomputing / HPC capabilities and facilities such as ANVIL into standards-based scientific workflow environments. The task will experiment with cross-platform implementation of connectors to HPC computing resources for open source geospatial analysis and modeling tools; advancing efforts to make HPC available as an open, platform-neutral component of distributed processing and modeling workflows for Open Science and other applications.

- **Linked Data and Semantics** - Not least in the context of Digital Twins, the demand for unambiguous semantics and descriptions of the components, attributes, or other elements used in a workflow to create a product is increasing. This task will explore

interoperability patterns for semantic uplifting of existing Web APIs, exchange formats, and other elements. The goal is to allow full understanding of all essential elements that have been used for product creation. How to replace cryptic attribute names with machine and human readable data, how to ensure consistency across data products by clear workflow descriptions, or how to share information models on the Web with strong semantics? As a practical matter, much work is still needed to improve mechanisms for maintaining, discovering, and utilizing data links in distributed open computing environments such as the Web.

- **Deep Learning (DL) Training Data and Metadata** - Among various fields of exploration in artificial intelligence (AI) and DL, the availability of high-quality training datasets is an exciting area that holds great potential to make accurate predictions or perform a desired task. Training data is the initial dataset used to train machine learning algorithms and models. Training data is also known as training dataset (TDS), learning set, and training set. The goal of this topic is: 1) to discuss the cutting-edge topics of machine learning training data for the geospatial community; 2) to describe the spatial, temporal and thematic representativeness of TDS and their uncertainties; 3) to focus on sharing and reusability of TDS to increase the adaptation of TDS for geospatial analysis; such as ML training sets for a class of OGC Feature entities, activities, events, (etc) and the resultant AI recognition system for that OGC Feature class, introduction of a distinction within the class that defines a subclass, and inducing the resultant recognition module for that class updating such items as descriptors and metadata accordingly.

- **Moving Features** - OGC Moving Features specifies how to encode the result of a tracking algorithm in a simple and readable fashion and enables a one to one mapping of all the metadata. Additional research is suggested for space based use cases, handling FMV from satellites, improving tracking and accuracy, and adding dynamic shapes.

- **Model Driven Standards** - The OGC has experimented with model driven standards and is adopting automated transformation pipelines for model development, e.g. for version 3.0 of Observations, Measurements, and Samples and the MUDDI Conceptual Model. This work needs to be continued in order to allow rapid profile development for customer-specific solutions. How to efficiently build a standards based API e.g. to access GeoDataCubes by reusing existing standardized building blocks?

- **Persistent Demonstrator** - This task focuses on exploring a concept of “Persistent Demonstrators”, which aim to be a permanent way of demonstrating how OGC standards support various domain use cases (e.g. climate change, disasters). The task can explore how the persistent demonstrator concept can work. Work should also explore how temporal aspects can be incorporated.

- **Privacy and Data Ethics** - The OGC community can help address the pressing issue of ethics and geo understanding that it’s going to take sustained effort over a period of time to raise awareness and to make an impact. This initial task can experiment with

architectures (data quality, security, portrayal, etc.) based on specific use cases and in turn develop best practices and guidance based on the outcomes to contribute to a wider discussion on ethics.

- **Reproducibility** - Any scientific or computational workflow, including ML models and applications, must be repeatable for trust, reproducible for validity, and reusable for value. Testbed-18 investigated many challenges to reproducibility, including the difficulty of identifying what is "the same" or different in distributed geospatial datasets and algorithms, non-determinism in many computing tasks such as container formation, and deep integration of random seed operators into many / most ML modeling frameworks. Workflow archive mechanisms such as Wholetale.org do exist, but struggle to deal with distributed workflow systems. There is both interest in, and recognition of the need for further investigation of reproducibility in geospatial data workflows to support the benefits of open science and generally to increase trust in the outputs of increasingly complex, large-scale, and opaque workflows such as ML models.

- **Space-based Cloud Computing.** Exploring the use of OGC standards for edge computing on board spacecraft (e.g. a satellite). This can also include looking at communications between different networks (e.g. sensors orbiting planets communicating with LEO satellites, which communicate with aircraft, the ground, and instruments in/on the ocean).

- **Swarm and Mesh Networks** – Undertaking a proof of concept of OGC-driven interoperability amongst meshes of sensors (e.g. drone and satellite swarms). These networks require many standards to interoperate within their own networks and with external networks. Everything from data assimilation with high temporal and spatial resolution; data to models; asset coordination and recruitment of sensors; when, where, and how to find data are all examples of where standards are needed.

- **Sensor Integration**: A significant barrier to sensor integration is often the variety of standards, formats, and protocols employed in sensor systems. To build impactful sensor systems, such systems must embrace this diversity. This task focuses on the integration of sensor systems regardless of their technical constraints and deployment environment. There is a need for a semantic and operational framework of standards that facilitates sensor integration independently of technological restrictions.

- **Subsurface representation and visualization** - Accurate and complete representation of both the built and natural underground environment is critical for safety, for disaster resilience, and for creating more livable habitats. The nature of the subsurface, though, is that it is hard to observe and easy to ignore until it is too late. This task will develop open standard methods that utilize multiple sources of data and lines of evidence to interpret and map subsurface features such as geology, for use in 2D / 3D models and immersive visualizations. Such methods will also involve estimating and communicating uncertainties in those representations, leading to more appropriate use and improved public understanding. Another focus will be suitability of subsurface representations and

observations for incorporation into urban and environmental digital twins.

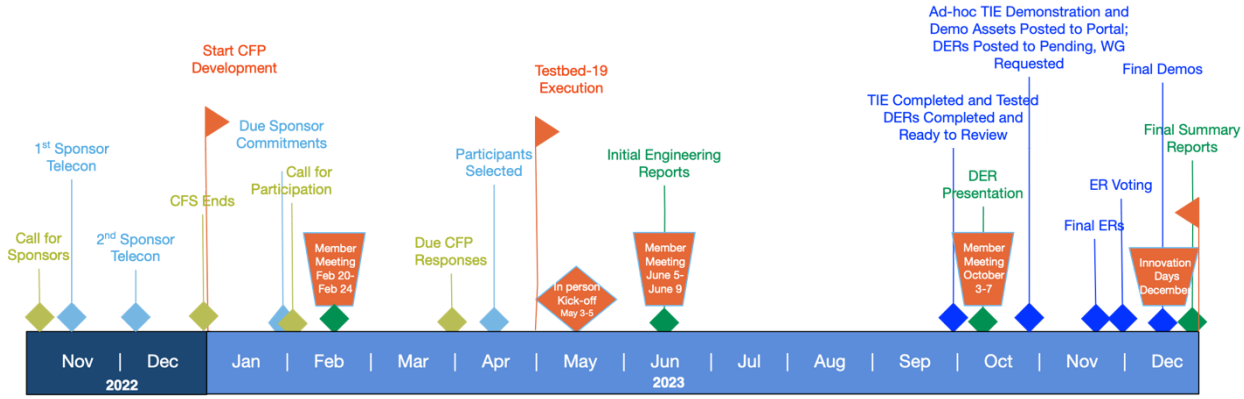
- **3D Visualization** - The combination of gaming engines, enhanced 3D video stream formats, and compression techniques with increasing amounts of real time, simulated, and historic data allows visualizations to be created that speak to many different types and communities of users. Photorealistic scenes rendered into an individual's direct environment open up new communication channels that were unavailable due to the high degree of interpretive skill required. This task investigates how new channels implemented through photorealistic visualization can be used for applications such as risk communication that can be compelling without simply frightening people (e.g. showing their houses appearing to float away). The task will also explore how analytics can be effectively communicated within 3D scenarios by symbolizing attributes of the visualized elements. Pure visualizations are thus developed into analysis-ready environments, where formerly "dumb meshes or scenes" become queryable real world representations.

- **Web APIs for the Aviation Community**

Chapter 5. Testbed Timeline

The development of Testbed topics is a collaborative process between OGC and interested sponsors. During the first phase (November, December, January), sponsors and the OGC team work together to refine use cases and requirements on the sponsor side and transform these into actual work items. Each work item will be assigned to an OGC member organization for implementation during the Testbed-19 execution phase. Once these discussions are complete, OGC will develop a Call for Participation that is planned to be released to the public in early February. After a 45-day response period, OGC together with the sponsors will select the best participants based on the received proposals to form the Testbed-19 team. The execution phase, the phase of system design, rapid prototyping, testing, and documentation will be executed from May to December 2023.

The following timeline highlights all major milestones for Testbed-19



Chapter 6. Call to Action

Interested in Sponsoring? Please contact Marge Cole and Trent Tinker via email at innovation@ogc.org or you may use the OGC COSI Program contact form found here. Please contact us ASAP as sponsor discussions are taking place in November and December.