



Developing a Vision for Improving the Discovery and Access of Bathymetric Data

A [webinar series](#) that was co-designed and hosted by the [GEBCO Technical SubCommittee for Ocean Mapping](#) (TSCOM), the [International Hydrographic Organization's Data Center for Digital Bathymetry](#) (IHO-DCDB), and the [Regional Center for the Atlantic and Indian Oceans of the Nippon Foundation - GEBCO Seabed 2030 Project](#). TSCOM provides technical advice towards the maintenance and improvements of GEBCO products and supporting data. The DCDB is the recognized IHO repository for all bathymetric data, with the intention of providing preservation, discovery and access of data. The Nippon Foundation - GEBCO Seabed 2030 Project has the goal of creating a complete map of the world ocean by 2030, and is actively assembling bathymetric data into the publicly available GEBCO world map.

Report written by:

Vicki Ferrini, Jennifer Jencks, Hayley Drennon, Tinah Martin, Sheila Caceres, Christiane Reiser, Erin Heffron, Federica Foglini, Alex Bastos

Executive Summary

This report summarizes the key points and outcomes of a four-part webinar series held 27 February - 01 March 2023. The webinars were designed as listening sessions focused on four complementary themes related to the management, sharing, assembly and access of bathymetry data.

During each webinar, a brief introduction was provided by the hosts Jennifer Jencks (Director of the IHO DCDB) and Dr Vicki Ferrini (Head, Atlantic and Indian Oceans Regional Center for the Nippon Foundation GEBCO Seabed 2030 Project and TSCOM member), to summarize the goals and rationale of the webinar series and the particular session. Each session was led by a community member experienced with aspects of a particular topic, and included 4-5 additional people who engaged in stimulating conversations providing perspectives on current barriers, opportunities, and future vision. These conversations highlighted efforts that can help achieve common goals, but also yielded input and commentary from the broader community.

More than 100 participants joined the live webinars, and recordings of the webinars have been viewed more than 500 times.

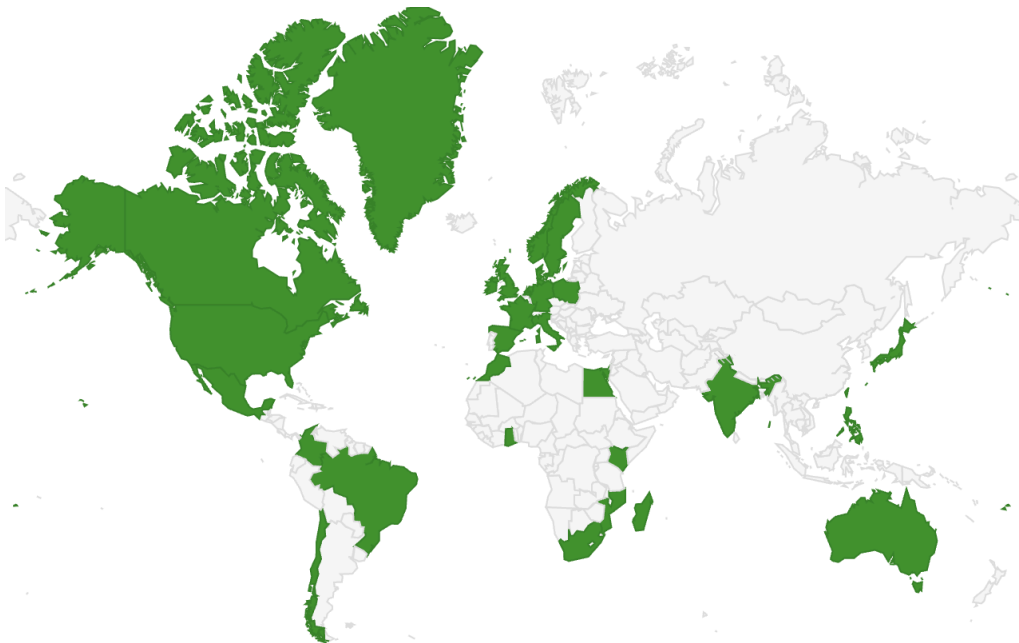


Figure 1: Map representing global participation in the webinar series

This webinar series was designed to connect members of the global mapping community through an inclusive exchange intended to foster the sharing of data, tools and knowledge. Throughout each session, participants provided input through live survey instruments along with questions to the panelists.

The webinars emphasized the importance of community input for defining and prioritizing development related to data management. Curating metadata that meets the needs of multiple use-cases throughout the data stewardship lifecycle was a common theme. Variable requirements of different software tools and different users with respect to data access and products should be considered as part of data management efforts in order to avoid redundancy of effort and ensure that data can deliver optimum value to a broad community of stakeholders. Coming together as a community to share knowledge and tools in this evolving space is critical if we are to work as a global coalition to meet mutually beneficial goals.

Outcomes documented in this report will inform the development of technical roadmaps and best practice guidelines that can help address current needs and shape future collaborative development, to be drafted at a follow-on hybrid working meeting held 23-25 May 2023.

Webinar 1: Data Stewardship

The first webinar, “*Data Stewardship*”, held on February 27, 2023, focused on data sharing, obstacles to data sharing, solutions that have helped to address bottlenecks, metadata needs for sharing and attribution, long-term preservation, access and re-use of data.

Session Goals

- **Community Needs:** Understand data stewardship needs beyond the goals of the Nippon Foundation - GEBCO Seabed 2030 Project
- **Challenges:** Gather information to feed into the development of an inclusive approach to address cultural and technical barriers to data sharing
- **Proposed Solutions:** Identify necessary steps toward resolving community-identified barriers of data stewardship and suggest adapted solutions to those data stewardship barriers

Panelists

Christiane Reise, Panel Moderator, *IHO DCDB; GEBCO TSCOM; NOAA NCEI, USA*

Suzanne O’Hara, *Rolling Deck to Repository (R2R); Lamont-Doherty Earth Observatory (LDEO), USA*

Dr. Helen Snaith, *National Oceanography Centre (NOC), British Oceanographic Data Centre (BODC), Global Center of the Nippon Foundation - GEBCO Seabed 2030 Project, UK*

Sean Haughton, *INtegrated Mapping FOr the Sustainable Development of Ireland’s MARine Resource (INFOMAR), Ireland*

Tion Uriam, *Marine Division, Ministry of Transport, Information Technology and Communications, Kiribati*

Main Points

- Data stewardship extends from acquisition to long-term archiving in support of data use, preservation, discovery, access, and re-use by a broad and diverse user community.
- Data stewardship is a commitment to adhering to findable, accessible, interoperable, and reusable (FAIR) data standards.
- Prioritizing data management and sharing, with an elaborated framework and plan, saves technical and labor costs.
- Promoting a culture of data sharing requires building trust, acknowledging data contributors, and engaging with stakeholders to understand barriers and work collaboratively to address them.

- Continuous active engagement involving meetings, workshops and skills exchange will sustain the effort and constitute an incentive model for data sharing.
- A holistic approach, such as developing an MSDI (Marine Spatial Data Infrastructure), and including stakeholders in the data stewardship process is key.
- Harmonizing documentation standards, while difficult to maintain, it is very important to have an organized framework that would help in curating and understanding of the data.

Community Needs

- Identifying and keeping up-to-date with user needs and evolving technology
- Maintaining standards, taking into account resources: defining and implementing an efficient workflow from data acquisition, quality control, processing to a final delivery of the bathymetry product.
- Transferring knowledge through a coordinated collaboration, nationally and internationally
- Communication and updating the community on progress: not always synchronized, requires a sturdy organization and infrastructure.

Challenges

- Motivation to share data is often inhibited by a lack of trust and/or incentive
 - Fear of damage or misuse of data
 - Lack of QA/QC prior to data sharing can be a barrier due to concerns of misinterpretation or discrediting data source for poor performance
 - Institutional reluctance due to costs associated with preparing and sharing data
 - Absence of clear data sharing policy and/or pipeline to safely share data, on top of the political and financial aspect of acquiring data at nation level
 - A broad cultural shift is necessary
 - Sharing raw data that has not been quality assessed/controlled may be unappealing to some data providers, but does not need to inhibit the sharing of data. Raw data that is shared can be processed by the community and can be transformed into valuable data products.
 - Different data providers have different motivation for data sharing (e.g. academic vs industry vs government). Understanding the incentives of the different user communities and providing metrics that will impact their outcomes (e.g. promotion, revenue, advertising/recognition) may help increase contributions.
- Resources to share and document bathymetry data vary widely and are seldom adequate
 - The amount of time it takes to get the data ready and available through a secure pipeline which is not always widely accessible.
 - Bandwidth/resources of the data provider to ensure sufficient data organization, accurate metadata and proper attribution for the data.

- Lack of local infrastructure such as a repository/data center, to handle the data and make them available
- Documentation and guidance is needed at institutional levels, especially for countries with limited access to technology or capacity
- Proper citation and credit of data contributors constitute an incentive for data contribution.
- Metadata:
 - Foundational metadata should document the data for preservation (e.g. who, where, how was it collected?) and attribution
 - The provision, and preservation, of metadata is equally important to preserving the data itself.
 - Metadata is necessary for repositories to archive and to provide attribution while enabling processing and analysis by down-stream data consumers.
 - Globally standardized metadata might be challenging achieve but it is important to ensure that datasets are accompanied with basic information

Proposed Solutions

- Collaboration and Support
 - Collaborate with the researcher by providing the proper tools/documentations to allow for the contribution of standardized data; include a best practices document for bathymetry data acquisition
 - Attribution needs will vary by contributor and metadata and systems need to remain sensitive to those variable needs so that attribution itself is not a barrier to contribution. In some cases, the funder, vessel operator, individuals/institutions responsible for collecting or processing the data will be important to acknowledge, and in other cases, contributors may want to remain more private.
 - Provide technical support and capacity building for communities with limited access to technologies or pathways to contributing data
 - Solutions will emerge by engaging with communities not currently sharing bathymetry data to better understanding the limitations and challenges they face.
- Preservation and Data Sharing
 - Several upload/download tools are already available and mostly serve the community.
 - Data management plan should be established before data collection begins.
 - Preserving the raw data increases value and impact since the data can be used in a wide range of fields
 - Suggestions for increasing data contribution:

- Make clear instructions and terms of agreement for data sharing, applying or implementing common standards, metadata, straightforward delivery workflows
 - Maintain open systems and open data formats
 - Use a cloud based semi-automated upload system
 - Controlled vocabularies improve the flows of data and provide information that accommodates the several fields that benefit from bathymetry data collection
- Generating and maintaining harmonized metadata is key to making data shareable and accessible.
 - Cloud-based automated process, machine readable adaptable for GIS analysis. Overlooked but is hugely important. different standards makes curating data painstaking.
 - Data Packaging tools: using JSON files to manage the metadata could speed up the process
 - Harmonized metadata structure based on existing ones: some examples mentioned include IODE Quality Management based ISO 9001 Data Quality, assigning an internationally recognized Creative Commons data license

Webinar 2: Data Discovery & Identifying Data Gaps

The second webinar in the series, “*Data Discovery and Identifying Data Gaps*”, held on February 27, 2023, focused on the use and integration of web services for data discovery, access, and advanced geospatial applications related to data-driven decision-making and opportunistic data gap filling.

Session Goals

- **Understanding Data Discovery and Gap Identification:** how to improve data discovery and improve workflows for gap identification and analysis. This includes the steps, the manpower, and the tools necessary to make the data more accessible to the users.
- **Challenges:** Identify the challenges and opportunities, especially those the community can contribute to or even fully address with a coordinated effort.
- **Proposed Solutions:** Identify key actions to implement a coordinated data discovery and gap identification workflow.

Panelists

Erin Heffron, Panel Moderator, *independent contractor and GEBCO TSCOM member, USA*

Daniel Damaske, *MARUM - Center for Marine Environmental Sciences, University of Bremen and Data publisher for Earth & Environmental Science at PANGAEA, Germany*

Dr. Wetherbee Dorshow, *Earth Analytic, Inc, GEBCO TSCOM Member, USA*

Jesse Varner, *National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information International Hydrographic Office Data Center for Digital Bathymetry (IHO DCDB), USA*

Dr. Thierry Schmitt, *European Marine Observation Data Network (EMODnet), Service Hydrographique et Océanographique de la Marine (SHOM), France*

Main Points

- Data discovery includes searching for data of interest and identifying existing data to help reveal gaps in coverage that can inform additional data acquisition.
- The ability to perform advanced queries to get to the desired results quickly is an important part of data discovery - a user should not spend large amounts of their time in this initial data discovery stage.
- The user base is broad and has variable skills and tools which should be considered with the development of services, user interfaces, data formats and data products.

- Metadata is integral in helping users decide if the data is fit for their needs both in the case of grid and synthesis products, and with raw data (e.g., EMODnet DTM ‘quality index’ reference layer).
- Direct feedback from the community of users is important in driving and prioritizing enhancements of products, services, and interfaces.
- When possible, both raw and processed data are important to preserve. In some cases, data providers are able to make processed grids available (often at decimated resolution), but not raw source data. The limitation of this approach is that any data quality issues, which may only become evident when during integration with other products, cannot be easily addressed. Sharing a gridded data product without source soundings (either raw or processed) significantly limits future use of the data.

Challenges

- Metadata
 - It can be difficult to understand provenance, and to access data sources, once data have been integrated into compiled syntheses. This limits data access, use, and re-use.
 - Metadata is not sufficiently robust for supporting data discovery, use and synthesis (e.g., consider asking data providers for additional metadata (horizontal accuracy, vertical accuracy)).
 - It is not always obvious to the user when processed data is available which will lead to redundancy of effort. Improving inventories and search parameters to make it easy to be directed to processed data and/or data products is a recognized need.
 - Resolution needs and data quality standards vary. Some indication of this information would be useful to include in metadata to help users navigate archives and identify data of interest.
- Web Services need additional information and standardization in order to serve the multiple needs of users related to data coverage, data access and data discovery
 - Some types of web services, like WMS, allow for no further leveraging in online or desktop tools – they can’t be queried, used in raster analysis, or have their symbology changed.
 - Choosing to publish data utilizing other types of web services allows for more user control on visualization and for analysis tools to leverage the data.
 - Most of the audience use desktop GIS software to interact with data services and perform data discovery; there is a place, however, for tools like ArcGIS Online that allow for some level of GIS analysis without the complexity of a desktop application.

- Representation of data coverage (geometry type) on maps is insufficient for accurate assessment of data coverage and for data discovery/use and could be improved by standards and guidelines based on tangible use-cases.
 - Raster data is often faster to serve and visualize, and can hold a lot of additional embedded attribute information that allow for additional layers of information within one data source and clever use of raster analysis tools.
 - There may be some persistent issues with raster reprojection and geospatial disagreement between layers due to how they were generated.
 - Polygons allow for quick intersection and area analysis, but can be slow and complex to draw if not done carefully.
 - Specifying a geometry or data type might put an additional burden on the data provider, or service provider, but there is probably a need for some consensus.
 - Suggested as an important topic for discussion at the workshop.

- Knowing the users, understanding their needs, tracking the impact
 - Audience responses indicated that there is a missing educational component, in that they would benefit from tutorials or how-tos on the best ways to utilize web services and downloaded services.
 - Improving the reporting of how users access data and how often data are used (e.g. data citation) should help to improve user interfaces and services.

Proposed Solutions

- A geospatial overview of data should be easy for the user to access and use to address multiple problems. Provide basic information should include extent of coverage, data/device type/make, quality indicator, data density, how to access the data.
- Enhanced standards for metadata and data coverage geometries are needed to improve the value and impact of current services and would enable the development of new interfaces that can be used to address community needs, such as distinguishing mapped from unmapped areas of seafloor.
- Consider other ways to figure out user needs, such as user behavior in applications and/or routine outreach to community to ensure that services and metadata are adequate to meet their needs

Webinar 3: Data Access and Community Needs

The third webinar in the series, “*Data Access and Community Needs*”, held on February 28, explored how to improve the availability and accessibility of bathymetric data for multiple audiences.

Session Goals:

- **Experiences - Data Access And Accessibility:** Share experiences from multiple users. The panel was composed of professionals from different fields and areas of expertise that have different applications for bathymetric data.
- **Data Availability/Accessibility:** Discuss the known availability and accessibility challenges experienced as users of bathymetric data.
- **Envisioning The Future:** Synthesize different opinions and suggestions for the future.

Panelists

Dr. Alex Bastos, Panel Moderator, *Professor at the Universidade Federal do Espírito Santo in Brazil and Coordinator for the Brazilian Program of Marine Geology and Geophysics (PGGM), Brazil.*

Dr. Kerry Howell, *Professor of Deep-Sea Ecology at the University of Plymouth, UK*

Dr. Kelley Brumley, *Independent Contractor, Adjunct Professor at the University of Houston and Affiliate Faculty at the University of Alaska, Fairbanks, USA*

Dr. Gareth Davies, *Tsunami Hazard Scientist at Geoscience Australia, Australia.*

Dr. Vincent Lecours, *Professor of Geography (Applied Geomatics), Université du Québec à Chicoutimi, Canada*

Main Points:

- Multiple applications exist for bathymetric data and depending on the user's need, or expertise, the data could be interrogated and adapted to generate numerous products, all deriving from one data set.
- Educating both our community and the public is crucial. There is an intrinsic value in understanding our planet better. There is value in this type of information being available for everybody.
- Data hunting and sharing is complex as both require time-consuming processes of understanding and navigating repositories.

- Harmonizing crowd-sourced data standards would move the community forward through coordination and collaboration and would facilitate dissemination and training while creating more inclusion.
- Eliminating the culture of data hoarding by encouraging the sharing of proprietary data, especially in cases of publicly funded research, in whatever format is most convenient for those holding the data.

Challenges:

- Barriers to data discovery/use include:
 - Finding data and navigating the different databases can be difficult and time-consuming. Users often have to spend hours searching for data, downloading large datasets, and then reprocessing it into a format and product they can handle.
 - Data is often present in more than one repository creating redundancy and complexity for the user looking for data.
- Barriers to data sharing include:
 - Several entities can be reluctant to share bathymetric data because they require additional return on investment.
 - Even when data holders are interested in sharing data, the costs of preparing and uploading the data can be a barrier to sharing.
 - Sharing data is complex
 - It takes time, effort and following very strict rules to submit data to repositories in the right format and with the correct specifications.
 - For people that have data but sharing it it's not their main job, this task is usually left behind and considered the last priority.
 - No one has the resources to put the time and effort into making data findable and accessible. Even if they made their data available through a repository, it would not necessarily be in an easily reusable format.
 - Other issues mentioned were time/staff limitations, variability in data collection and lack of institutional support.

Proposed Solutions:

- Better standards are needed, but must recognize that multiple formats and standards will be needed to meet the needs of all users.
- Encourage Industry to collaborate with repositories and share data (e.g. business cases, buying data, tax incentives, charitable contributions).
- Share and promote tools that make it as easy as possible for people to share, access and navigate databases and data products.

- More resources are needed to get people trained on processing, cleaning, and sharing data in the best way possible.

Webinar 4: Data Processing, Transformation & Integration

The fourth, and final, webinar in the series “*Data Processing, Transformation & Integration*”, held March 1, focused on how we currently process, transform and integrate data, how to increase optimization of data processing, and how to efficiently share bathymetric products with the community.

Session Goals

- **User Community:** Understand user communities and gather their input to define and prioritize the products we make and deliver to the community.
- **Tools and Environments for Processing:** Identify differences in the tools and environments employed to handle high volumes of raw data and/or diverse data types, how we address the challenges presented by these data through workflows and distributed processing.
- **Data Integration:** Discuss how we integrate data from multiple sources and identify bottlenecks, challenges, and inefficiencies to optimize the integration pipelines as well as define how we ensure the products we are generating are useful in the long-term. This discussion will also help establish fit-for-purpose quality and data processing standards.
- **Opportunities for Data Acceleration:** Identify opportunities for collaboration to accelerate the creation of high quality openly accessible data products and envision the future of data processing, transformation, and integration.

Panelists

Hayley Caitlin Drennon (Panel Moderator), *Lamont-Doherty Earth Observatory (LDEO), Columbia University, Data Manager for Regional Center for the Atlantic and Indian Oceans*

Dr. David Sandwell, *SRTM, Scripps Institution of Oceanography, University of California - San Diego, USA*

Joshua Sixsmith, *AusSeabed and GMRT-AusSeabed, Geoscience Australia, Australia*

Christian Ferreira, *Center for Marine Sciences (MARUM) University of Bremen, Germany*

John Morton, *Global Multi-Resolution Topography (GMRT) Synthesis and Marine Geoscience Data System, Lamont-Doherty Earth Observatory (LDEO), Columbia University, USA*

Main Points:

- The user community is varied, supporting government, academia, and industry, however, most of the services have been shaped by the needs of the scientific community.
- Formats should be efficient, with respect to optimizing file size, performance, and enabling re-processing, while reflecting the needs of the users, which requires defining what an efficient format is, both internally and to the users (e.g. GSF swath format vs NetCDF or TIF raster format vs ASCII/CSV data format)
- Accessible and equitable data necessitates the creation of user-friendly, open source tools and software packages to limit the barriers to distributable processing and data visualization or download.
- Metadata that describe data provenance, processing, transformation and attribution are important for decision-making with respect to data integration and determining the “best/authoritative version” which should be housed in a known repository (e.g. NCEI/DCDB).
- Data inclusivity is critical because of the diversity of data format, structure, and provenance. Each group employs multiple tools to address the challenges of process/transfer/integration differently, but most rely on open source packages customized to meet the user and data manager needs.
- Processing effort and code base knowledge should be sustainably developed and distributable to the broader community through to help address human capacity limitations
- The global bathymetry community needs the combination of open access data, tools and algorithms, supported by open communication and collaboration

Challenges/Barriers

- User Community
 - Because of the diverse community, data should be efficient while reflecting the needs of the users, from data storage and management to the user downloading the data.
 - Common grid formats create frequent challenges to users
 - NetCDF, can often not be read by modern software
 - ASCII/CSV delivery are not standardized and often require translation using command line tools which are less familiar to a novice user
 - Different software packages are required to process different data
 - Swath data: MB-System is freely available and is uniquely able to handle legacy and large amounts of data, while commercial software (e.g. QPS, CARIS) have more modern user interfaces and are more approachable to many users
 - ASCII data can be processed/handled with multiple software packages (freely-available and commercially available), but these formats omit

additional data streams (e.g. vessel/vehicle attitude) that are sometimes necessary to deliver high quality data products.

- Tools and Environments for Processing
 - Many current open source and cloud based environments have both a steep learning curve and data streaming bottlenecks, making them inaccessible to those who do not have a strong coding background.
 - Lack of sustainability: we need to ensure our software is supported in the future, both in programming support, and in code adaptability/flexibility to handle both legacy and modern formats.
 - In areas where no data exists, or is currently inaccessible, we have no ability to confirm the validity of the bathymetry which makes editing challenging.
- Data Integration
 - Bottlenecks to data integration greatly reduces the speed of integration
 - Personnel cost, both in terms of financial cost and time, is the prevalent bottleneck
 - Software limitations both in terms of processing capacity and inability to handle a variety of raw input data formats
 - A lack of common community standards. (e.g. IHO standards are inadequate for deep water)

Proposed Solutions and Looking Forward

- Format Standardization
 - Formats should be efficient, with respect to optimizing file size, performance, and enabling re-processing, while reflecting the needs of the users, which requires defining what an efficient format is, both internally and to the users (e.g. GSF swath format vs NetCDF or TIF raster format vs ASCII/CSV data format)
 - Data acquirers should be encouraged to collect and produce well documented and easily accessible formats
- Open Source/Cloud Based Environments
 - Open source (processing software, computing environments) and cloud based environments have the potential to meet user and developer needs through data streaming, grid building, and remote processing
 - Containerization of these packages could help reduce scripting and installation load on the user
- Development: Generation of User-Defined Tools
 - Because of the diversity of users, both cross-sector and within the scientific community, the future of data delivery would necessitate integrating more customizability in on-the-fly grid generation based on user defined requests.
 - API's/Workflows/Tools should be well documented and distributed in order to equip the user with the ability to modify code to adapt to their format.

- Processing Standardization
 - When possible we should compare data to underlying bathymetry, as it is being collected, to create bathymetric data contextualization, corroborating the processing individual datasets require.
 - The standards we use for QA/QC should be fit-for-purpose (e.g. Data processing designed for research versus safety of navigation)
- Opportunities for Data Acceleration
 - Open communication and collaboration through meetings and working groups ensure that our standards are still meeting the needs of today's data users and processors
 - Making data more accessible means generating robust, easily accessible open-source tools that are able to reduce user effort and pull from a variety of existing sources/projects/initiatives.