Bathy4All: Workflows for Multibeam Processing and Visualization

Vikram Unnithan[1], Peter Baumann[1], and Christian dos Santos Ferreira[2]
[1]Jacobs University Bremen, D-28759 Bremen
[2] University Bremen, MARUM - Center for Marine Environmental Sciences, D-28359 Bremen
24th July 2020

Abstract

Easily accessible, high resolution bathymetry data is required for all marine sciences. Workflows for efficient processing, analysis and visualisation of multibeam data using opensource tools and datacubes, which can be easily combined with satellite or climate data, are scarce. Bathy4all aims to design and implement workflows to process raw multibeam data present in data archives using the open source MB-System and rasdman to create bathymetry datacubes based on open standards and FAIR access. The combination of the best-of-breed preexisting components (MB-system, CUBE/CHRT, rasdaman) provide significant added value and progress beyond the state of the art. The deliverables include the integration of an advanced processing algorithm, a software demonstrator showcasing bathymetry datacubes, and a roadmap for inclusion of Water Column Data (WCD) and acoustic backscatter data. Stakeholders include not only marine sciences but also government agencies (e.g. BSH) and industry (e.g. maritime transport). The intelligent processing of bathymetry, building datacubes with standards-based APIs, inclusion of bathymetry in the EarthServer federation will promote the massive use of bathymetry by several earth sciences and remote sensing communities, in future including INSPIRE and thereby promoting NFDI4earth interests.

I. Introduction

Accurate and high-resolution bathymetry that is easily accessible is a key necessity for a wide spectrum of ocean research and sciences ranging from anthropogenic impacts, natural hazards, benthic habitats and ecosystems to maritime transport and security. Over the past decade, most German research vessels have been equipped with multibeam echosounders, which provide accurate seafloor topography and in some cases information within the water column. Multibeam bathymetry data, due to their large file sizes, are typically stored in archives as raw or gridded processed files. The interaction with these datasets in archives such as PANGAEA is at the basic level of data search and file extraction/download. Recent ongoing initiatives, for example the DAM Underway Data and the associated MaNIDA projects, are in the process of designing workflows to provide easier access to bathymetry using a combination of MB-System, CARIS and web-based marine data and mapping portal. The Bathy4all pilot project **aims** to explore ***possible alternatives and complementary workflow***s for multibeam data processing, analysis and dissemination by relying on opensource, FAIR data access and better integration and visualization.

The opensource solution MB-System is commonly used to process, analyze and visualize multibeam data, especially when users are attempting to reuse data store in the file-based

archives like, for example, PANGAEA. This pilot establishes a spatio-temporal datacube from existing bathymetry data, offered through a standards-based actionable datacube API, accessible through a variety of clients from map navigation (ex: Leaflet) over virtual globes (ex: NASA WorldWind) and Web GIS (ex: QGIS) to high-end analytics (ex: python). This unleashes multibeam data to a wider community. Additionally, the rasdaman engine used will allow combining the bathymetry datacube with the existing datacubes of CODE-DE, CREODIAS; Mundi DIAS, AWI, and HZG offering dozens of Petabyte of satellite, climate, and thematic data in the EarthServer federation.

II. Pilot description

Project components

The project will combine several existing tools into a novel bathymetry analytics system: MB-system for multibeam data processing and analysis, the CHRT hydrographic processing tool, and the rasdaman scalable datacube engine for organising, managing, and serving bathymetry data as datacubes. Automated ingestion procedures will allow continuous unsupervised updates and extension of the bathymetry datacube.

The open-source **MB-system** is the backbone of Bathy4all for multibeam data processing and analysis. MB-system will be then used to extract metadata, process and analyze sample datasets identified in the repositories. The **CUBE** (Combined Uncertainty and Bathymetric Estimator) with its Hierarchical Resolution Techniques (**CHRT**) algorithm will be integrated into MB-system. This algorithm is a statistical autofilter for bathymetry data that can be used to estimate which soundings are valid/invalid and to estimate the depth value at a given location. Hence, it can be used both to flag bad soundings and to generate grids (DEMs) based on sounding datasets. CHRT is the practical standard for many hydrographic agencies and commercial surveyors because it reduces processing time and effort. CHRT, which has a lightweight licensing scheme, is used in most commercial software packages for processing seafloor mapping data to provide an effective mechanism for quick automatic processing of multibeam data, and in addition provides an estimate of calculated depth uncertainties. In addition to bathymetry data, the possibility of visualization and processing of water column data (WCD) and backscatter data in MB-System will also be explored. **Rasdaman** (“raster data manager”) is an Array Database System for managing and querying massive multi-dimensional data (“datacubes”), such as spatio-temporal satellite image and climate data. Datacube queries are based on a high-level language for processing in the server where a multitude of highly effective optimizations boosts performance on such "Big Data", including just-in-time compilation of queries into machine code, multi-core parallelization, and distributed processing. The fully fledged, fully operational datacube engine supports the relevant datacube standards, such as OGC WCS/WCPS. This enables a variety of third-party clients to remotely operate on the datacubes such as map navigation, virtual globes, Web GIS, analytics and visualization. A versatile ETL tool which allows virtually any import, including pre-processing, to be configured in a straightforward way, including MB-System integration. Manifold interfaces in rasdaman allow for versatile standards-based access via Web services, but also via APIs in python, R, Java, C++, JavaScript and thereby providing the opportunity to combine bathymetry data cubes with relevant satellite, climate, and thematic other data cubes already available in the EarthServer Datacube Federation.

III. Relevance for the NFDI4Earth

Bathymetry is a "staple" ingredient that is needed by all marine and maritime branches of science and industry. Bathymetric information is required by the entire spectrum of stakeholders ranging from scientists all the way to public authorities and decision makers dealing with topics ranging from fisheries, offshore energy, carbon emissions and impacts of climate change.
However, unlike other Earth data categories, like satellite imagery and climate data, bathymetry data still are hard to obtain and to process, making them unavailable to non-experts. Additionally, while specialized tools for processing exist these largely are desktop-bound and not yet integrated with common Spatial Data Infrastructures. Consequently, seamless, smooth combination of bathymetry with other relevant data sources, such as Copernicus satellite data and INSPIRE authority data, is not possible today. Several challenges need to be solved in order to achieve the Holy Grail of unleashing the full potential for wide user communities. Bathy4all aims to find workflows to enhance the use of exiting raw bathymetric data by using the opensource MB-System. It thereby supports the **NFDI4earth commitment** to find opensource solutions and help FAIR access to data and associated products. It will provide an innovative solution or alternatives to workflows currently adopted by most hydrographic agencies routinely acquiring multibeam data.
The **datacube** model has rapidly gained acceptance as a cornerstone for analysis-ready data, together with the corresponding service model which is more powerful, but easier to use than existing API-based interfaces. Mature and widely adopted datacube standards show the way how datacube functionality can be presented to clients. Specifically, the OGC *Coverage Implementation Schema* (CIS) data model and the *Web Coverage Service* (WCS) service model suite define a standards framework adopted by the main open-source as well as proprietary tools, including MapServer, GeoServer, GDAL, QGIS, ArcGIS, and python/OWSlib. The EarthServer datacube federation has established an intercontinental network of Earth data centers where users get provided with a single, uniform information space acting like a single local data holding. Active work is being undertaken currently to link EarthServer with INSPIRE. Bathy4All will become a member of the EarthServer federation, thereby unleashing the data assets of all federations for access, analytics, visualization, and fusion.

IV. Deliverables

The project deliverables (also in Figure 2) are:
D1. Provide an overview of available and accessible raw multibeam data in German Data repositories. This document will also include a selection of bathymetry datasets to be used for testing algorithm and workflow development. D2. Software development, testing and implementation of CHRT client in MB-System. The software along with associated documentation will be available via GitHub and also from the MB-System websites. If the developed software passes the required tests, it might be included in future releases of the MB-System. D3. Workflows and automated interface scripts in MB-System to push processed multibeam data to rasdaman, along with the needed documentation will be provided via GitHub (as D2). D4. A software demonstrator will allow to retrieve bathymetry data from the Bathy4All datacube. Access will be possible without coding, through open standards APIs and a representative slate of tools. The

data access and visualisation pipelines, including the setting up of services within rasdaman and documentation will be provided via the project GitHub page (as D2, D3). D5. Final project report with brief analysis of designed workflows, strategies for possible improvements, recommendations for best practices and (at least 1) scientific publication. This final report will provide a possible **roadmap** for the marine community in terms of multibeam data, its future use and links to other data types and systems. The tentative title for the roadmap is "*Workflows for efficient access and utilization of multibeam data with opensource tools and FAIR data access*".

V. Work Plan and Requested funding

The **workplan** is broadly subdivided based on the two main software components used i.e. MB-System and rasdaman. The first part (WP1) involves the identification of suitable sample datasets for testing workflows and software development. It also includes the background study and research on available tools and data archives. WP2 requires more time as it deals with the software implementation and developed of a CHRT client within MB-System. All codes developed will be available via respective GitHub pages and will be considered for incorporation, after testing (M3) with a variety of chosen sample datasets, within the next beta MB-System releases. The next stage involves interfacing (WP4) between MB-System and rasdaman, such that processed multibeam data can be incorporated (WP5) into Earthserver federation and visualized and populated as a service (WP6). Scientific evaluation, documentation, reporting and publication (WP7) is an important part of the project. **Funding** is requested for 1 FTE for 1 year. The highly specialized nature of multibeam data and the software to be used and modified, necessitates an experienced post-doc, advanced researcher or developer working at least one full time equivalent (1FTE) in terms of time and effort.

Figure 1(below left): Bathy4All schematic architecture

Figure 2 (below right): Gantt chart providing an overview of the various deliverables (D) and workpackage (WP) as described in sections IV and V.