Towards common file formats and data standards for seafloor geodesy

Community Whitepaper for UNAVCO's "Future Directions for Seafloor Geodesy" Committee
September 2020 *(revision of July 13, 2021)*

Pierre Sakic[1], Valérie Ballu[2], Heidrun Kopp[3], Dietrich Lange[3], Jean-Yves Royer[4]

1: Department 1: Geodesy, Helmholtz-Zentrum Potsdam - GeoForschungsZentrum GFZ, Potsdam, Germany
2: Littoral Environnement et Sociétés, La Rochelle Université and CNRS, La Rochelle, France
3: FB4 Dynamik des Ozeanbodens, GEOMAR Helmholtz - Zentrum für Ozeanforschung, Kiel, Germany
4: Laboratoire Géosciences Océan, University of Brest and CNRS, Plouzané, France

DOI: *10.48440/GFZ.1.1.2021.001*

Introduction

Seafloor geodesy experiments have been expanding considerably in recent years. More and more research teams around the globe are conducting projects to monitor the tectonic or volcanic deformation of the seafloor. These experiments are commonly based on limited-duration experiments, but increasingly, permanent observatories are also being installed.

This dynamic development is very encouraging for the establishment of a strong community which arguably will lead to the emergence of a worldwide scientific and technical synergy. However, data and knowledge transfer between the different groups working on similar topics are still limited at the present time. This can be partly explained by the fact that the instruments, infrastructure, and processing software developed are custom-made and thus various file formats are used, although the fundamental observables are most of the time identical. One way to overcome this limitation is to set up exchange standards in the form of standardized file formats. These files would gather and store all the physical quantities observed and will prove useful for the processing simplification and, in the end, the extraction of the geodetic signal sought. Furthermore, uniformized formats would allow much more easily the comparison of software and processing methods between research groups, whether during tests or operational measurement campaigns. Standardized data will eventually provide a base for the activities of potential future national or international observation services. They would also make it possible to envisage the data dissemination similar to geodetic data recorded on land.

The example of the space geodesy/GNSS community

An excellent example of a strong geodetic scientific community is the *International GNSS Service* (IGS). Scientists and research engineers working on satellite precise positioning agreed to join forces three decades ago to pool their observations and know-how. To this end, data exchange standards have been developed. A best-practice example is the RINEX format (archiving raw observations independently of the GNSS receiver type used), the SP3 format (containing satellite positions & clock offsets), and the SINEX format, to store and exchange versatile data (station coordinates and velocities, normal equations....). The standards developed within the IGS framework are nowadays widespread and universally used, even outside the academic community.

Observable types

The physical observables used in seafloor geodesy are multiple. There are also different levels of observation, pre-processed, or fully processed data. One can make the analogy with the Data Processing Levels described by NASA for space remote sensing missions. They define four levels, from the raw data directly observed by the instrument and uncorrected for possible errors, biases, and drifts, to the final product.

As an example, we list and categorize the observables required for seafloor geodesy acoustic experiments:

Observed physical quantities / monitoring approach		**Level 1**	**Level 2**	**Level 3**	**Level 4**
		Raw observations	*Low-level exploitable quantities*	*Basic derived products*	*Final output products*
Sound Velocity	*CTD or XBT profile*	Conductivity, Temperature, Pressure	Sound speed profile		Precise positions, Velocities, Deformations (Strain)
	Direct SVP probe	Time travel measurement	Pointwise sound speed		
	Experimental/original device	e.g. time travel, pressure measurement...	e.g. integrated sound speed		
Acoustic travel-time	*Between seafloor beacons*	Raw hydroacoustic signal measurement (amplitude, phase...)	Two-way travel time	Distances	
	Interrogations from surface				
Beacon depths	*Long term observations*	Raw pressures	Pressures corrected from drift, atmospheric and oceanic signals.	Water depths time-series	
	Campaign-based measures				
Surface positioning	*Lever arm/Surface platform ties*	Angle and distance (topometry)	Vectors between the onboard instruments		
	Absolute positioning	Pseudo distance and carrier-phase at the GNSS antenna (RINEX)	Kinematic Positions in a global RF + local RF (if needed)	corrected position transferred to the acoustic head	
	Platform Attitude	Roll, Pitch & Yaw angles			

Beyond the observed data, it is also crucial to store the associated metadata (date and time of observation, nominal sampling, units...). If possible, the metadata should be stored in the same files as the observations.

Possible file formats

Several types of data formats can be considered. The most natural are those in plain text. The most trivial candidates are comma or tab-separated text files (CSV or TSV), which have the advantage of being easily importable with the most widely used scientific programming languages such as Python or R. Another interesting alternative in plain text would be the SINEX format, which is also commonly used in terrestrial and spatial geodesy. The SINEX has the advantage of being simple, robust, and versatile. Since most of the time, the data can be seen as time series, a modified version of a seismological format (e.g. SAC) might also be used.

Plain text files provide the advantage of being human-readable and are a feasible option at least for acoustic experiments where data volumes are manageable. Furthermore, file formats should be self-explanatory, *i.e.* contain explicit descriptions of the different fields, rows, and columns containing the data. However, these are very preliminary reflections which of course need to be discussed and refined by a wide community of users.

What would be the next step?

The main objective of this white paper is to highlight the current lack of cohesion in the possibilities of exchanging seafloor geodetic data. To improve on interchangeable file formats and standards, it would be advantageous to set up a working group on it. The following issues need to be addressed:

- centralize existing solutions for storing seafloor geodesy observations, whether defined by instrument manufacturers or research teams.
- list precisely which physical observable would be of interest to store, along with the associated metadata.
- define file structures to archive and read (by computer but also by a human) these data.
- encourage the use of pre-defined standards to enable feasible data-exchange in-between the stakeholders involved in seafloor geodesy.

Opening the opportunity to ensure long-term time-series continuity will require both our ability to reprocess the data, but also to install and describe permanent reference benchmarking on monitored sites. This paper suggests improvements in data and metadata management, however, the seafloor benchmarking issues need to be addressed for the future observer generations.