Challenges to Standardization: A Case Study Using Coastal and Deep-Ocean Water Level Data

Aaron Sweeney¹, Kelly Striker², George Munov¹, Susan McLean¹
¹Cooperative Institute for Research in Environmental Sciences, Boulder, CO. ²National Centers for Environmental Information, National Oceanic and Atmospheric Administration, Boulder, CO, United States

Introduction

Sea levels recorded at coastal stations and inferred from deep-ocean pressure observations at the seafloor are submitted for archival and publication in multiple data sets and metadata forms. These forms include two forms of schema-level metadata and a custom binary format acquired and exchanged by metadata in a spreadsheet. The authors report on efforts to use existing standards to make this data more discoverable and more useful beyond their initial use in detecting tsunamis. An initial review of data formats for sea level data around the globe revealed heterogeneity in presentation and content. In the absence of a widely-used domain-specific standard, we adopted the general model for structuring data and metadata expressed by the Network Common Data Form (netCDF). netCDF has been endorsed by the Open Geospatial Consortium and has the advantages of small size when compared to equivalent plain text representation and provides a standard way of embedding metadata in the same file. We followed the orthogonal time-series profile of the Climate and Forecast discrete sampling geometries as the convention for structuring the data and describing metadata relevant for use. We adhered to the attribute convention for metadata for capturing metadata to support user search. By making it possible to structure data and metadata in a standard way, netCDF is supported by multiple software tools in providing programmatic cataloging, access, subsetting, and transformation to other formats. We will describe our successes and failures in using to existing standards and propose requirements for either augmenting existing conventions or developing new ones. Some of these enhancements are specific to sea level data, while others are applicable to time-series data in general.

Data collection not designed around re-use...

Coastal Sea Level Time-series Data, as sent to NCEI (Provider 1)¹

¹ No XML schema for data (non-standard content)
² Reference datum, platform description, georeference, sensor information, and units supplied in separate file

Ocean Bottom Pressure Time-series Data, as sent to NCEI (Provider 2)²

¹ No XML schema (non-standard content)
² No reference datum, platform, sensor, and georeference details supplied in file
³ Files created every 15 minutes
⁴ Data from all stations in single file

References

NCEI netCDF Templates v2.0: www.nodc.noaa.gov/data/format/netcdf/2.0 (includes CF and ACDD guidance)
NCEI Water Level Data (netCDF): www.nodc.noaa.gov/htbin/EnhancedCatalog/WaterLevel.html

netCDF promotes data discovery and re-use...

• Coastal and Deep-Ocean Water Level Data match the structure of netCDF orthogonal time-series
• netCDF metadata can be easily transformed to ISO-compliant XML, which drives discovery through data.noaa.gov and data.gov
• netCDF is easily converted to other formats (CSV, XML, WaterML2)
• Platform, sensor, georeference, and reference datum details, essential for data re-use, travel with data

Acknowledgements

The authors thank the NCEI netCDF Team, especially Ajay Krishnan and Matt Biddle, for providing a critical review of the resultant netCDF files.