Making sea level data FAIR

Andy Matthews, Permanent Service for Mean Sea Level (UK), antt@noc.ac.uk
Elizabeth Bradshaw, British Oceanographic Data Centre (UK), elizb@bodc.ac.uk
Mark Hebden, British Oceanographic Data Centre (UK), mahe@bodc.ac.uk

Sea level records are some of the longest ocean observations available, with the earliest continuous time series beginning in the 18th Century. The length of data available makes creating one complete findable, accessible, interoperable and reusable (FAIR) record a challenge.
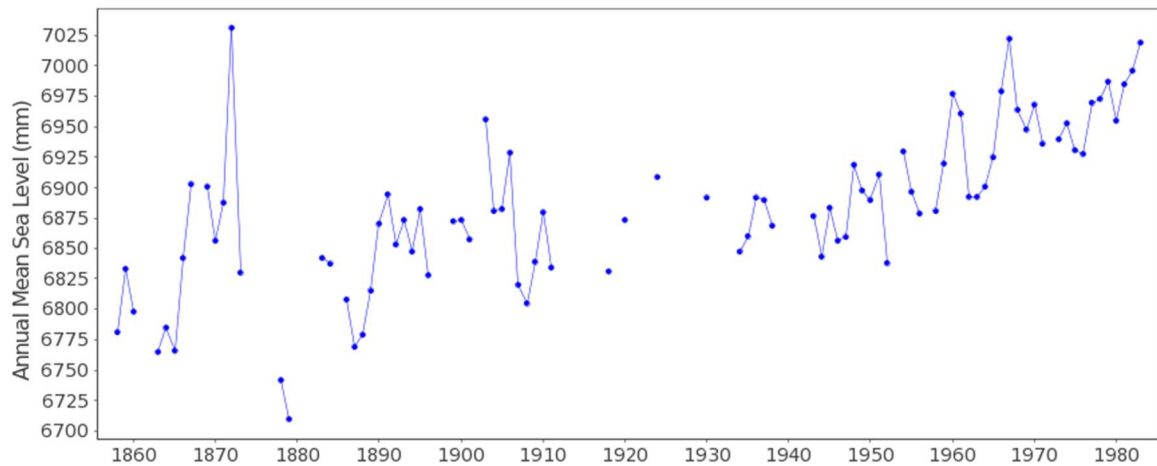

Figure 1: Annual mean sea level data from Liverpool, UK

Are the data findable?

We need to make the data findable and this can be done through creating standalone discovery metadata records, such as European Directory of Marine Environmental Data (EDMED) or NASA's Global Change Master Directory entries. We describe how in the future we will add the discovery metadata to the actual data such as in a netCDF file Attribute Convention for Data Discovery (ACDD).

We can also make sea level data more accessible by assigning persistent and unique identifiers such as Digital Object Identifiers (DOIs). The Global Extreme Sea Level Analysis (GESLA) dataset has been assigned a DOI, but we look at how we will tackle the issue of assigning DOIs to datasets that are growing such as the Permanent Service for Mean Sea Level (PSMSL) and Global Sea Level Observing System (GLOSS) datasets.

Sea level discovery metadata should make use of controlled vocabularies/ontologies/taxonomies and we describe those in common use and areas for development.

How do we make the data accessible?

Global sea level data are relatively accessible compared to other Essential Ocean Variables as they are deposited in the long established GLOSS international data centres e.g. this year PSMSL celebrates its 85th anniversary. Data in all the GLOSS data centres are freely available and organisations comply with the obligations for GLOSS members (Chapter 8: Obligations of GLOSS member states – http://www.gloss-sealevel.org/publications/documents/GLOSS_Implementation_Plan_2012.pdf). We are however looking to improve data flow between different time streams as some near real time data aren't making it into the PSMSL data bank. One of the possible reasons for this is the cost associated with quality controlling the data.

How can we make data interoperable?

Currently the GLOSS data centres each deliver data in their own format, but we are looking to use a common standard format such as CF netCDF and/or common data models to deliver data in one format to users.

We are also looking at increasing the granularity of our usage metadata. We are developing systems that will use Sensor Web Enablement (SWE) standards to help fully describe how we transform an observable property (such as the length of a piece of wire, the return time of a radar pulse, or the electric charge generated by a crystal under pressure) into a sea level measurement. This will improve the description of a time series where the sensor and platform changes many times.

We also need to ensure that we are using standard vocabularies for simple properties, such as time (ISO 8601) and country names (ISO 3166-1). For long time series, sometimes it turns out that these simple things actually aren't that simple. Is the country code that of the country where the sensor is located or where the organisation processing the data is based? There may be sites where the operating country has changed such as Hong Kong (from Britain to China).

How do we make sea level data reusable?

By storing sea level data in one of the global sea level data centres, we ensure that the data remain useable for the foreseeable future. By keeping comprehensive usage and lineage metadata alongside a dataset we will increase the reuse of the data, but also ensure that proper credit for the creation and preservation of a dataset is given.

Letting a user know what the quality of the data are and the level to which they have been screened will give confidence in the reuse of the data. Unique identifiers for data sets will help in the transparency and replicability of studies.