Making sea level data FAIR

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

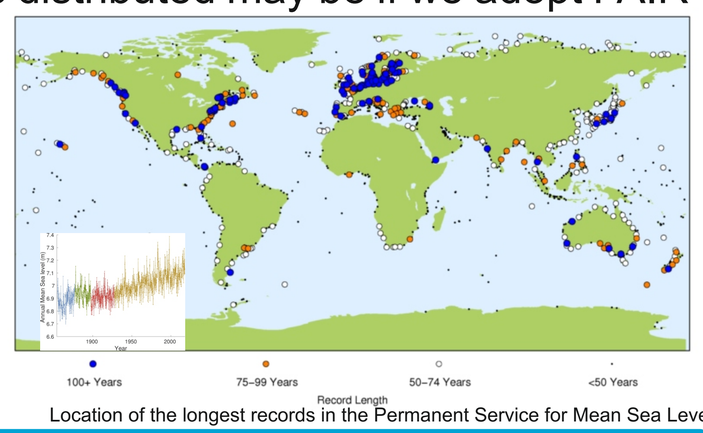
Sea level records are some of the longest ocean observations available, with the earliest continuous time series beginning in the 18th Century. The length of data available makes creating one complete findable, accessible, interoperable and reusable (FAIR) record a challenge. The current situation is:

- Sea level records are distributed by international data centres, under remit of IOC / WMO (e.g. PSMSL, BODC)
- Also available from data originators (e.g. NOAA, SHOM, JMA)
- Increasingly data is aggregated into large portals (e.g. EMODnet Physics, CMEMS, other sites we may be unaware of)
- There are well established best practices for how to measure sea level
- There are no commonly agreed standards or formats for distributing data
- Data isn't interoperable, so aggregated data can contain many flaws (e.g. lack of proper provenance information, or quality control indicators)
- During a 200 year record, measurement methods can change many times: knowing what happened when can be vital when interpreting the data
- Controlling the distribution of data isn't viable
- Controlling the data that is distributed may be if we adopt FAIR principles

Are the data findable?

Data is already findable, for example PSMSL is already included in the European Directory of Marine Environmental Data (EDMED) and NASA's Global Change Master Directory. However, discovery metadata can also be attached to the actual data file, such as in a netCDF file implementing the Attribute Convention for Data Discovery (ACDD). Sea level discovery metadata should make use of controlled vocabularies, ontologies and taxonomies e.g. the BODC Parameter Usage Vocabulary and the ICES Station Dictionary.

We can also make sea level data more accessible by assigning persistent and unique identifiers such as Digital Object Identifiers (DOIs). Data centres are in the process of assigning datasets DOIs, and we are investigating the feasibility of attaching identifiers to tide gauge locations and sensors.



Location of the longest records in the Permanent Service for Mean Sea Level

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 international data centres such as PSMSL. Data in all the GLOSS data centres are freely available and organisations comply with the obligations for GLOSS members.

How can we make data interoperable?

The first steps towards interoperability will involve data centres agreeing common data models and adopting standarrd formats such as CF netCDF. We are also looking at increasing the granularity of our usage metadata. We are developing systems that will use Sensor Web Enablement (SWE) standards (O&M, SensorML) 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.

We also need to ensure that we are using standard vocabularies. This has been done for simple properties, such as time (ISO 8601) and country names (ISO 3166-1), but we need to agree standards for more complicated cases, such as vertical datums (possibly the forthcoming Geodetic Register, ISO/PRF 19127), and methods of measuring sea level.

How do we make sea level data reusable?

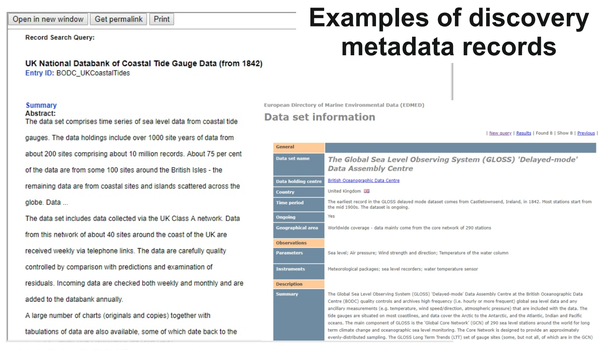
By storing sea level data in one of the global sea level data centres, we ensure that the data remain useable for the foreseeable future. By keeping comprehensive usage and lineage metadata alongside a dataset we will increase the reuse of the data, but also ensure that proper credit for the creation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

Is there a trade off between data curation and ease of use? We need to distribute metadata that accurately describes how and where sea level is measured but "how" and "where" change over time and are sometimes uncertain.

Metadata should tell a coherent story of how measurements are taken and have changed over the course of time, warn users of potential issues with the data and should use common standards to increase interoperability.

Examples of discovery metadata records



NASA GCMD record

GLOSS EDMED

GESLA DOI

Findable

www.gloss-sealevel.org

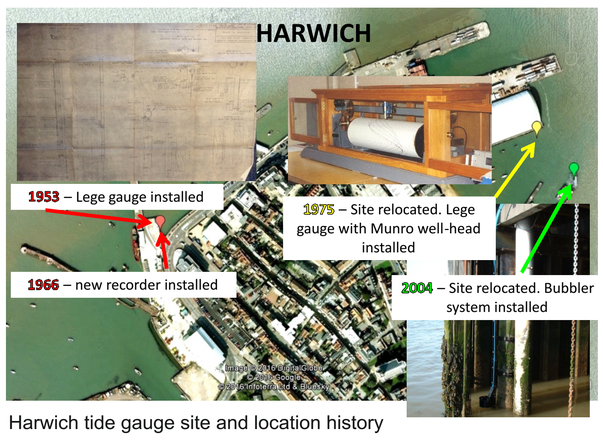
British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

GLOSS data centres

UNIVERSITY OF HAWAII SEA LEVEL CENTER

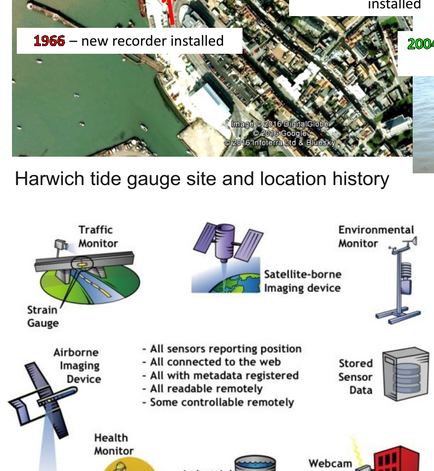
PSMSL

Accessible

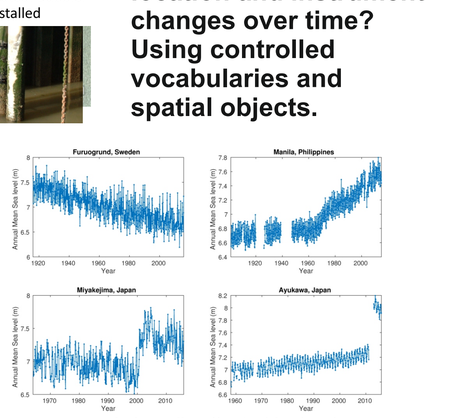


Harwich tide gauge site and location history

How do we deal with location and instrument changes over time? Using controlled vocabularies and spatial objects.

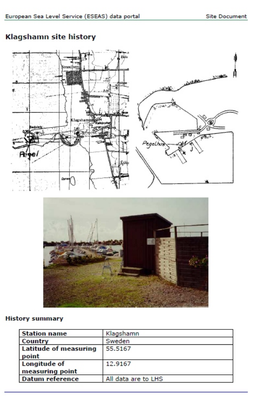


OGC Sensor Web Enablement overview (www.opengeospatial.org)



tide gauge records heavily affected by land movement

Interoperable



Example tide gauge site history document

ConceptID	Preferred label	Alt label	Definition	Modified
SDNPR001	metadata collator	MetadataCollator	Responsible for the compilation of metadata for one or more datasets and submission of that metadata to the appropriate SeaDataNet metadata repository.	8/18/2006 15:33:02
SDNPR002	programme operation responsibility	ProgrammeOperation	Responsible for the operation of a data collecting programme.	8/18/2006 15:33:02
SDNPR003	programme archive responsibility	ProgrammeArchive	Responsible for the archive centre handling distribution of delayed mode data from a collecting programme and the long term stewardship of its data.	8/18/2006 15:33:02
SDNPR004	programme realtime responsibility	ProgrammeRealTime	Responsible for the centre handling distribution of true and near real time data from a collecting programme.	8/18/2006 15:33:02
SDNPR005	contact point	ContactPoint	Person responsible for the provision of information in response to queries concerning the metadata or underlying data.	1/13/2011 15:45:16
SDNPR006	principal funder	PrincipalFunder	Person or organisation that funds the majority of an activity.	1/13/2011 15:45:36
SDNPR007	contributing funder	ContributingFunder	Person or organisation that contributes to the funding of an activity.	1/13/2011 15:45:52
SDNPR008	principal investigator	PrincipalInvestigator	Scientific lead of data collection within a programme	1/19/2011 13:01:07

Maintaining site histories and comprehensive metadata

Controlled vocabulary for contact and access roles

Reusable