GeoResJ 6 (2015) 9–16

Contents lists available at ScienceDirect

GeoResJ

journal homepage: www.elsevier.com/locate/GRJ

Sea level data archaeology and the Global Sea Level Observing System (GLOSS)

Elizabeth Bradshaw [a,*], Lesley Rickards [a], Thorkild Aarup [b]

[a] British Oceanographic Data Centre, Joseph Proudman Building, 6 Brownlow Street, Liverpool L3 5DA, UK
[b] Intergovernmental Oceanographic Commission of UNESCO, 7 Place Fontenoy, 75732 Paris Cedex 15, France

ARTICLE INFO

Article history:
Received 30 September 2014
Revised 29 January 2015
Accepted 9 February 2015
Available online 14 March 2015

Keywords:
Data rescue
Sea level
Tide gauge
Data archaeology

ABSTRACT

The Global Sea Level Observing System (GLOSS) Group of Experts (GE) data archaeology group is collating tools and producing guidelines for historic sea level data. They aim to aid the discovery, scanning, digitising and quality control of analogue tide gauge charts and sea level ledgers. Their goal is to improve the quality, quantity and availability of long-term sea level data series. This paper examines different tools for the automatic digitisation of tide gauge charts, the methods available for transcribing handwritten tide gauge ledgers and possible future developments that might speed up and partially automate these processes.

1. Introduction

At the 2013 Global Sea Level Observing System (GLOSS) Group of Experts (GE) meeting in Liverpool, UK, a data archaeology session was held. A number of actions were agreed upon [1] relating to uncovering data to digitise and sharing automatic digitisation software. GLOSS wants to provide its user community with guidance on finding, digitising, quality controlling and distributing their records.

Why put all this effort into recovering historical data? Recovering unique data and making them freely available is important, as they make a vital contribution to climate studies (sea level rise), oceanography (ocean currents, tides, surges), geodesy (national datum), geophysics and geology (coastal land movements) and a number of other disciplines. Long-term sea level data series are rare and the measurements unrepeatable. These data are used by the Intergovernmental Panel on Climate Change (IPCC), who published their Fifth Assessment Report in 2013 [2].

Global sea level reconstruction (e.g. Church and White [3]) combines satellite altimetry with tide gauge data. However, with only two decades of altimeter data, this reconstruction and others [4,5] rely on tide gauge information going back in time prior to 1993. In the PSMSL database the mean length of a revised local reference record is 37 years but at least 60 years of data are recommended for the analysis of long-term trends [6]. In addition, most of the time series are concentrated in the northern hemisphere (Europe, USA and Japan) [7] so introducing records from other, more isolated, locations would improve reconstructions. High-frequency data (typically defined as a sampling interval of one hour or less) can used to study high-frequency variability, storm surges [8] and the propagation of tsunamis. Unlike satellite altimetry data, actual tide gauge observations are used for coastal engineering and protection as well as to study coastal processes and changes in extreme sea levels and their return periods.

Data archaeology activities will help fill in the gaps in the global dataset, with records in the southern hemisphere being particularly valuable. Hunter et al. [9] used historical observations from a tide gauge at Port Arthur, Tasmania, to indicate an average rate of sea level rise from 1841 to 2002 and Testut et al. [10] used measurements taken from a tide staff on Saint Paul Island in the southern Indian Ocean to estimate a rate of relative sea level change from 1874 to 2009.

GLOSS is an international programme conducted under the auspices of the WMO-IOC Joint Technical Commission for Oceanography and Marine Meteorology. It was set up in 1985 to collect long-term tide gauge observations and to develop systems and standards "for ocean monitoring and flood warning purposes" and to "support the activities of the Permanent Service for Mean Sea Level" (PSMSL) [11].

The GLOSS Core Network of tide gauges, defined in 2010 (Fig. 1), consists of 290 stations around the world [12]. Most of these stations are operating currently but the distribution of long-term series is mainly concentrated in the northern hemisphere.

The PSMSL is the global data bank of long-term sea level change information from tide gauges. It contains monthly and annual

* Corresponding author.
E-mail address: elizb@bodc.ac.uk (E. Bradshaw).

http://dx.doi.org/10.1016/j.grj.2015.02.005
2214-2428/

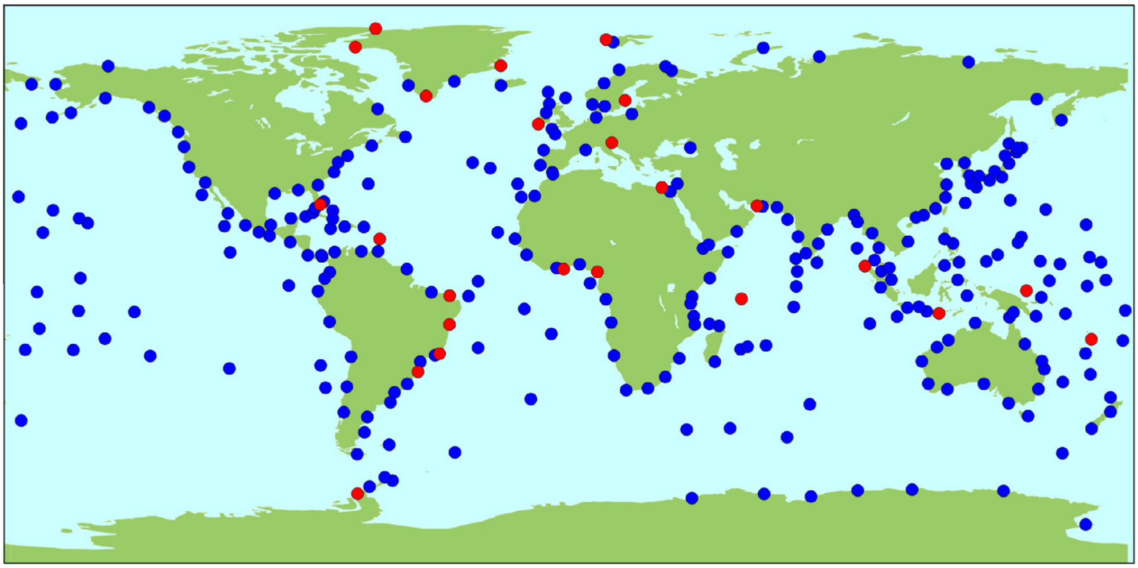

Fig. 1. The GLOSS Core Network of tide gauge stations as of 2010 (290 stations in total). New stations added in 2010 to this definition of the network are marked in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

mean sea level data. The higher frequency GLOSS data feeds into this database. There are no records at PSMSL longer than 100 years in the Arctic, Africa, South America or Antarctica (Fig. 2). GLOSS would like to focus its data archaeology efforts in these data-sparse regions to extend records back and help improve global climate change models.

2. Identifying data to rescue

Cartwright [13] gives an extensive account of the history of the observation of the tides and states that, between 1650 and 1825, "measurements… invariably consisted of visual observations". The first continuous analogue sea level record from an automatic tide gauge was made at Sheerness (Thames Estuary) in 1831. It wasn't until the mid-to-late 1980s that digital loggers were introduced to the UK Tide Gauge Network.

GLOSS has carried out a number of data archaeology activities in the past 10 years, including a questionnaire sent to member organisations to try to identify records in need of rescue. Caldwell [14] states, "Through this recovery effort and the work of individual agencies in digitising and quality controlling paper records, 91 tide gauge series were extended backwards by 1411 years of hourly data." Talke and Jay [15] carried out an extensive search of U.S. and Canadian archives for North American and Pacific Tidal Data and Pouvreau [16] provides a detailed description of data held in French archives. Fig. 3 illustrates the analogue

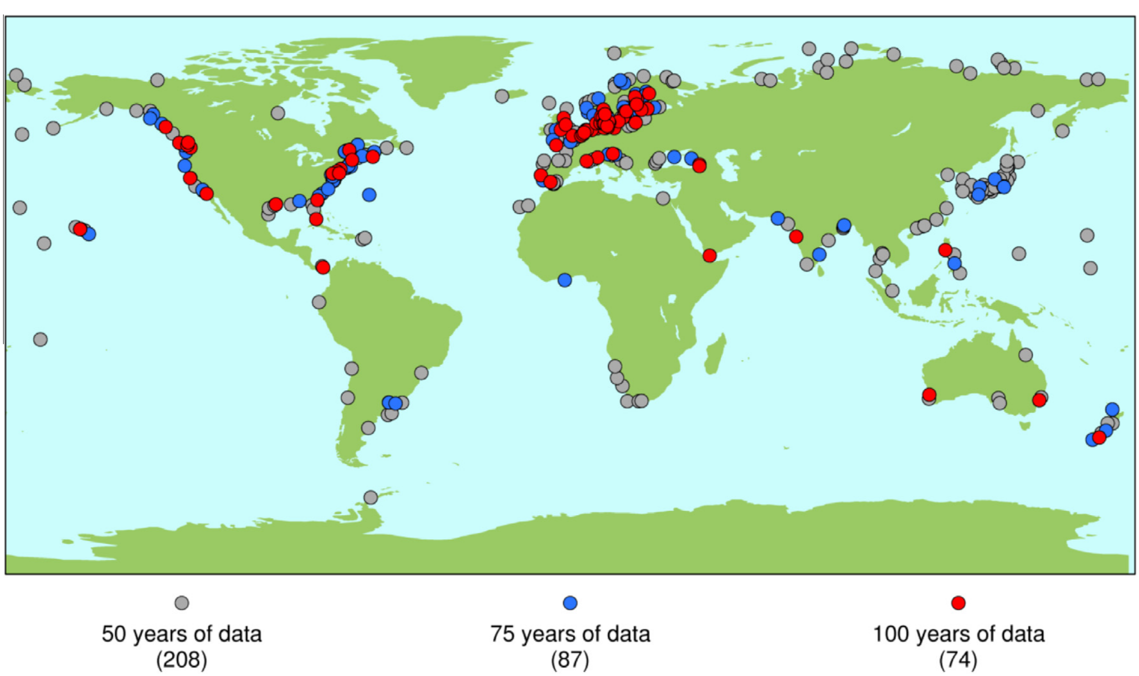

Fig. 2. Location of the longest records in the Permanent Service for Mean Sea Level.

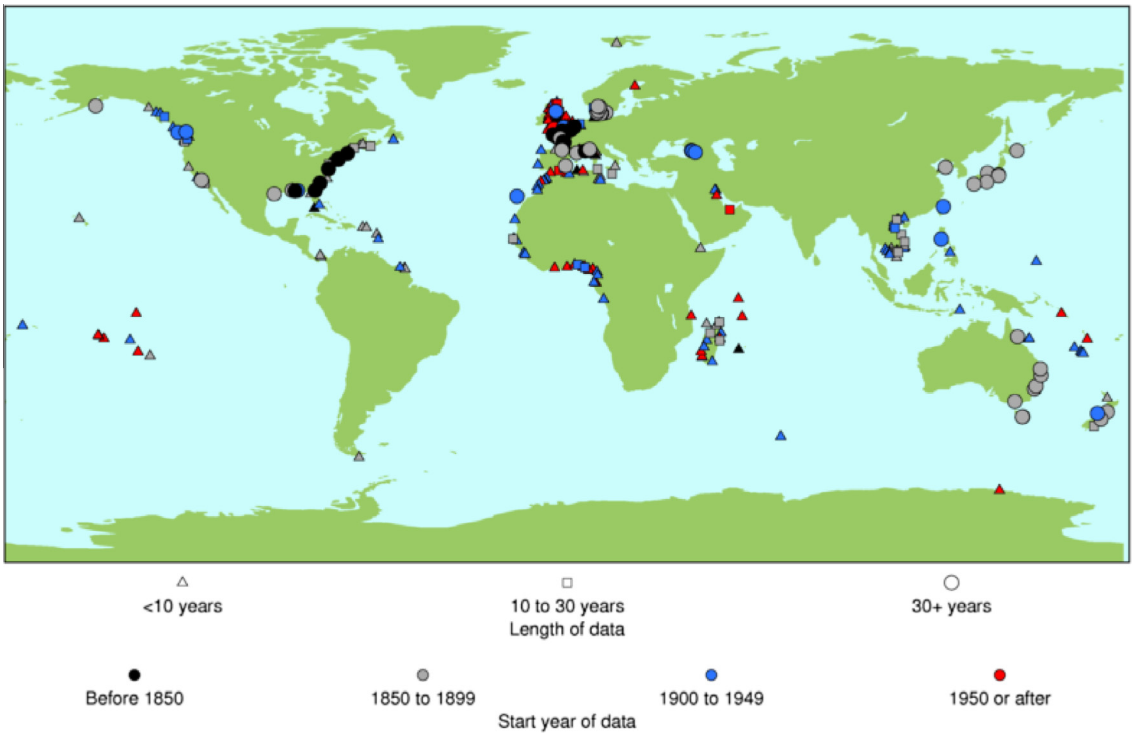

Fig. 3. Data uncovered by Caldwell [14], Talke and Jay [15] and Pouvreau [16]. The shape of the data point indicates the length of the record, the colour the earliest year of data found. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

records found by these three data archaeology activities [14–16] but many more records are lying hidden in unusual locations, yet to be discovered. GLOSS wants to encourage people to uncover these sources of data.

However, many records may not be held in organisational archives and may instead be in national libraries, archives and other collections (Fig. 4). Historical sea level data can be stored in many different formats such as journals, diaries, letters, newspapers and official port documents. Many archives and libraries make some discovery metadata available online, which can help to identify data that may be of interest, but there is often an issue with the granularity of records described and with the vocabularies used in their metadata. The terminology used to describe historical documents may not use obvious scientific metadata language. This lack of consistent online metadata means that a physical visit is usually required to see if a resource is relevant. This situation could be improved with the development of an ontology mapping related terms between different systems.

GLOSS will promote a Citizen Science approach to discovering long-term records by providing tools for volunteers to report data. There are an unknown number of records and locations and a specialist skill set is required to interrogate catalogues and records. Often people with an interest in family/local/maritime history will have the skills required. One proposal is that GLOSS develops a smartphone/tablet application and website allowing participants to report data sources they have discovered. Contributors would

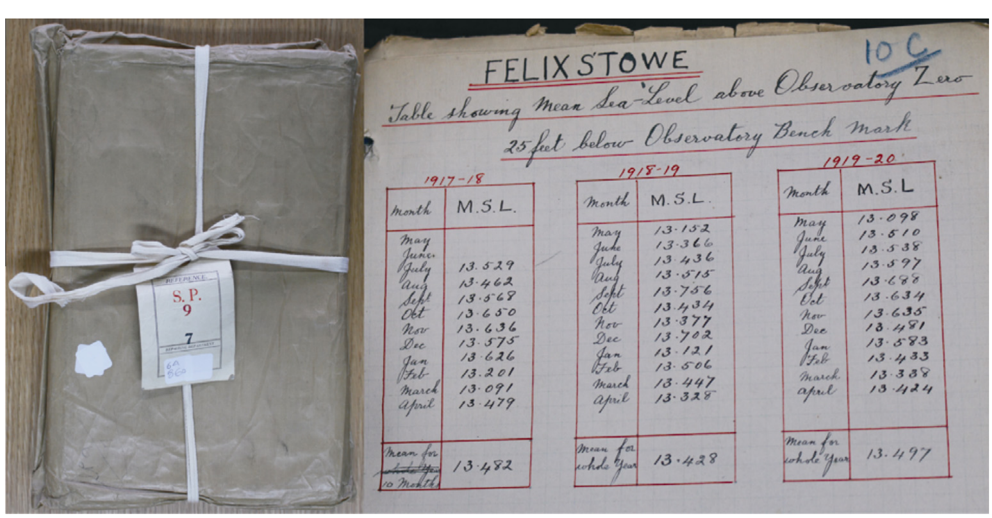

Fig. 4. Photographs taken of material found in the UK National Archives.

be asked to visit libraries and archives and look for material of possible interest, logging their findings via the website or app. Form input would be restricted to controlled metadata vocabularies, with an option to submit a photograph of the material if allowed.

Setting up a website or application is relatively straightforward, but to ensure that any project is a success, GLOSS would need to communicate with the volunteer community to ensure that they feel engaged. Nov [17] states "... citizen science designers and leaders should strive to increase volunteers' commitment to the project and its goals. This could be done by communicating the project's mission, achievements and its scientific contribution to the volunteers." Social media and blogs would provide a useful communication link and the GLOSS GE would work with members to create promotional material such as posters and leaflets to be sent to libraries and archives to encourage people to participate and hunt for hidden data. The Guide to Citizen Science, by Tweddle et al. [18] says that Citizen Science can be effective in "engaging people with how science works and for increasing their awareness of environmental issues and their local environment... One of the core strengths of the approach is that it can be used to present global issues – such as the impacts of climate change ... in a way that is locally relevant and meaningful".

Once new sources of historical data have been uncovered, they should be captured in such a way as to preserve the resource for future use, but also digitised to release the data for analysis. Woodworth [19] discovered tidal measurements from 1768–1793 in the personal journals of dockmaster and former privateer William Hutchinson in the Liverpool Public Library archives. The journals were used to construct the UK's longest near-continuous sea level record, at Liverpool from 1768 to the present day, and to calculate sea level trends.

3. Digitising charts

The most common form of analogue sea level records are paper charts generated by float tide gauges. As a float on the surface of the water rises it causes a weight (connected by a cable over a pulley) to descend. The angle of rotation of the pulley is directly proportional to the change in water level. A second pulley fastened to the same shaft drives a pen to trace on a chart (Fig. 5) [20].

Tide gauge charts are generated by a mechanical instrument driving a pen trace (e.g. Fig. 6). There are two main methods for obtaining digital data from these analogue charts. The first is to digitise the chart manually using a digitising table, which is labour intensive and time consuming. The second method, as employed by several GLOSS members, is to use software developed to extract data automatically from a digital image (scan or photograph) of the

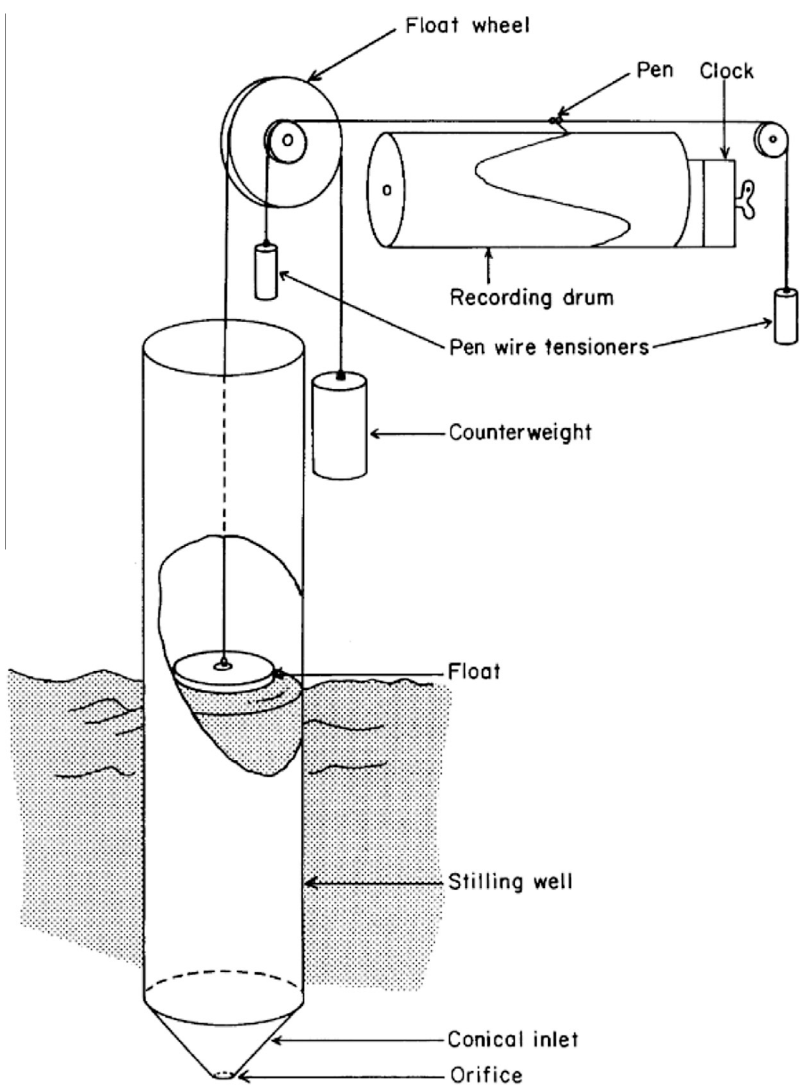

Fig. 5. Basic float tide gauge and chart recording drum [20].

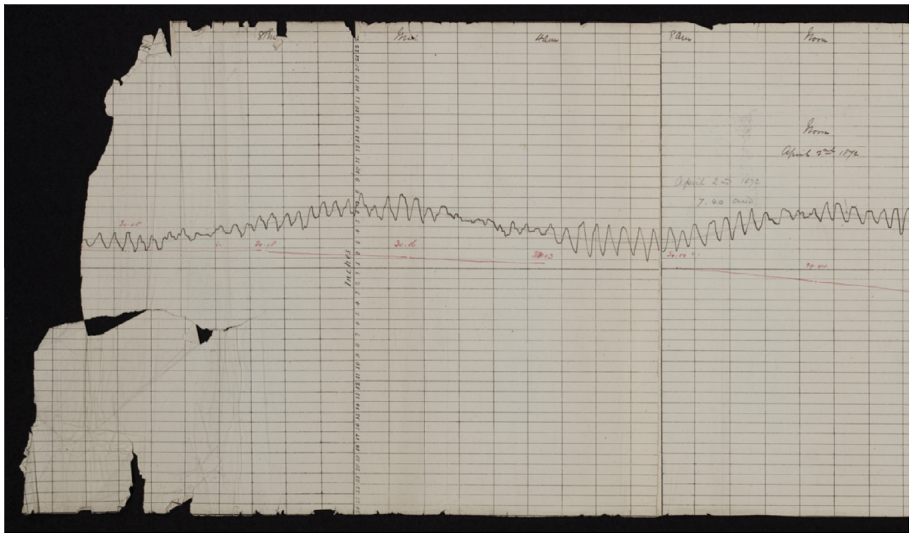

Fig. 6. Paper chart from Valetta Harbour, Malta, April 2nd 1872.

chart. Various methods for converting a line on a digital image to a data value were reported in a paper on automated techniques for the digitisation of archived mareograms (charts), delivered to the 13th GLOSS Group of Experts (GLOSS GE) meeting [21]. NUNIEAU (Logiciel de NUmérisation des enregistrements graphiques de NIveaux d'EAU, Ullman et al. [22]) is the only publically available software for digitising tide gauge charts. Other organisations have developed their own software but have not made it available to external users. As well as software specifically developed for digitising tide gauge charts, there are several freely available applications designed to convert a pixel (e.g. line on an image) to a number value.

The data reduction section of the Intergovernmental Oceanographic Commission (IOC) Manual on Sea-Level Measurement and Interpretation [20] details several of the errors that can occur with tide gauge charts that cause problems with the digitisation of the record. In some cases, charts have been left on the rotating drum for over a month, which leads to an overlapping of the pen traces, making following the correct curve difficult. Also, in areas where the tidal curve is not sinusoidal (e.g. Fig. 7) selecting the correct curve to digitise requires some effort.

GLOSS should set up a comparison study to determine which programs are best suited to the task. This could be carried out by sending the same chart images to each organisation, asking everyone to extract the data using their own procedures. Comparing the results would allow GLOSS to provide recommendations to the user community and allow an estimate of the relative accuracy of each method of automatic digitisation.

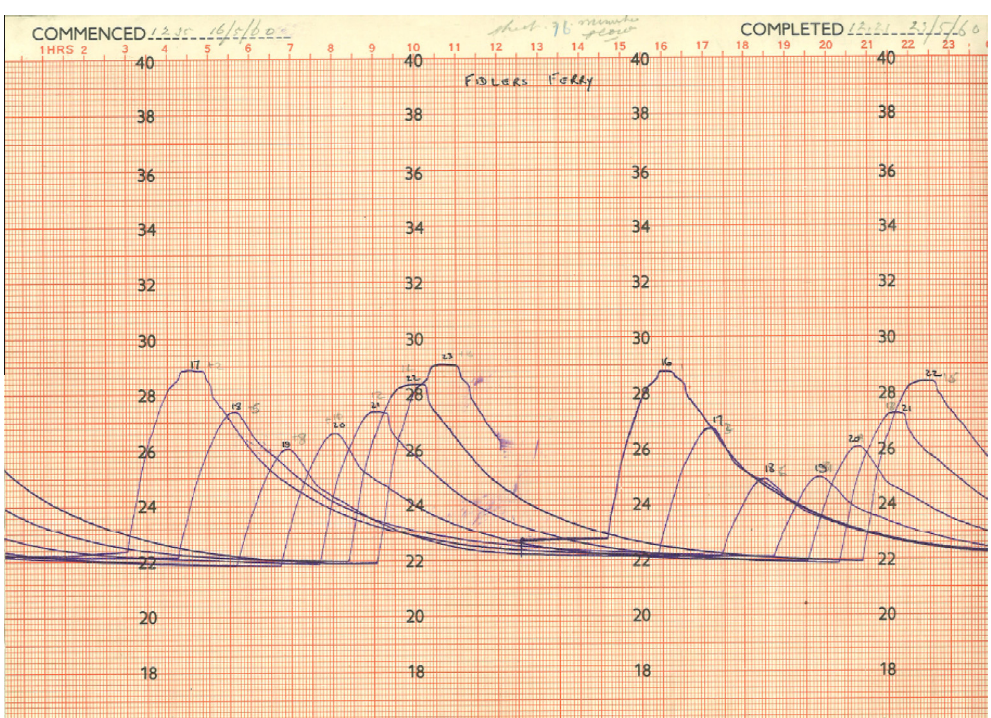

Fig. 7. Paper chart from Fiddlers Ferry, River Mersey, May 16th 1960–May 23rd 1960.

GLOSS will create a repository of software for scanning analogue charts but the ultimate aim is to create a web application for the GLOSS community where users could upload, calibrate and then digitise their images in their browser. This could then form the basis of a Citizen Science project where members of the public could help to digitise tidal records. They would be presented with an image containing a trace, which they could calibrate and then digitise. Dr. Richard Crouthamel gave a presentation at the 7th Atmospheric Circulation Reconstructions over the Earth (ACRE) workshop in Toronto, Canada, to promote the upcoming International Environmental Data Rescue Organization (IEDRO) project [23], Weather Wizards, where citizen scientists will be able to digitise pluviograms (rainfall charts). GLOSS will be in communication with IEDRO to see if the same system can be applied to tide gauge charts, reducing duplication of effort and resources.

4. Scanning ledgers

The other major source of analogue sea level data is handwritten ledgers. These are usually observations of high and low waters but sometimes contain higher-frequency data. The standard current method for digitising these data is to key the values manually. This has been performed by GLOSS countries, including France [24] and Spain [25] (Fig. 8).

The GLOSS GE is exploring other methods for use in the future, as this process is time-intensive. Members of the GLOSS GE have been in contact with the OldWeather project [26], which is run in partnership with the ACRE initiative and has been very successful in digitising ships' log books. Though the layout of tide gauge ledgers may look similar, there are a number of differences between them and log books that would make using a similar approach challenging. Firstly, the ledgers' pages are more data dense than log book pages. With meteorological data, a citizen scientist might only have to capture six or so observations per page whereas there can be ten times as many data points per page in a tide gauge ledger (Fig. 8). Secondly, the data are very 'dry' compared to the ships' log books, being almost entirely numerical data with little 'human interest'. People are drawn into the story when digitising the logs, which contain comments about life on board, such as "The most acceptable occurrence was the issue of a double ration of whiskey, with which, hot water, and sugar, we tried to be cheerful, and make Christmas Eve rather less dreary than many of our days now seem." [27]. This human interest element is lacking in the tide gauge ledgers. This Citizen Science approach will need to be adapted to fit the purposes of the sea level community. (see Fig. 9)

The GLOSS GE is investigating developments in Handwritten Text Recognition (HTR) technology. Current projects to improve HTR tend to work with the written word and so require knowledge of sentence structure and word occurrence probabilities to reconstruct sentences [29]. This approach would not be directly applicable to sea level records, however tidal data by its very nature is periodic and predictable, so HTR technology could be adapted to take this into account and improve the automatic digitisation of handwritten tide gauge ledgers.

The British Oceanographic Data Centre (BODC) has plans to set up a 'big data' project using existing tide gauge data to create a climatology for a given location and time. This could then be used as a training dataset to aid the transcription of the ledgers. During a discussion at the 7th ACRE Workshop it was proposed that the sea level community work together with the meteorological community (and others, if interested) to develop a training dataset and algorithms for HTR technology that could cope with multiple types of scientific data, not just sea level records.

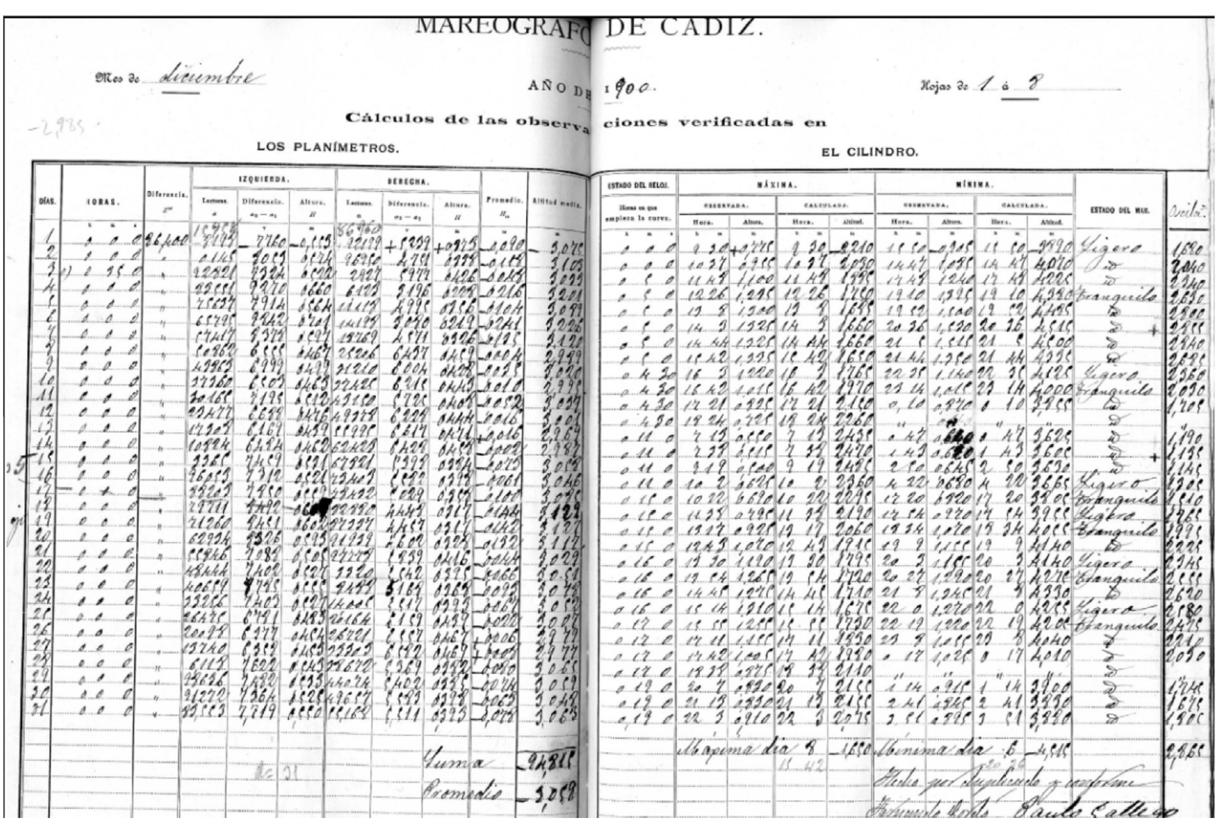

Fig. 8. Sea level observations from Cadiz, 1900, Instituto Geográfico Nacional archives at Madrid (In email correspondence from M. Marcos).

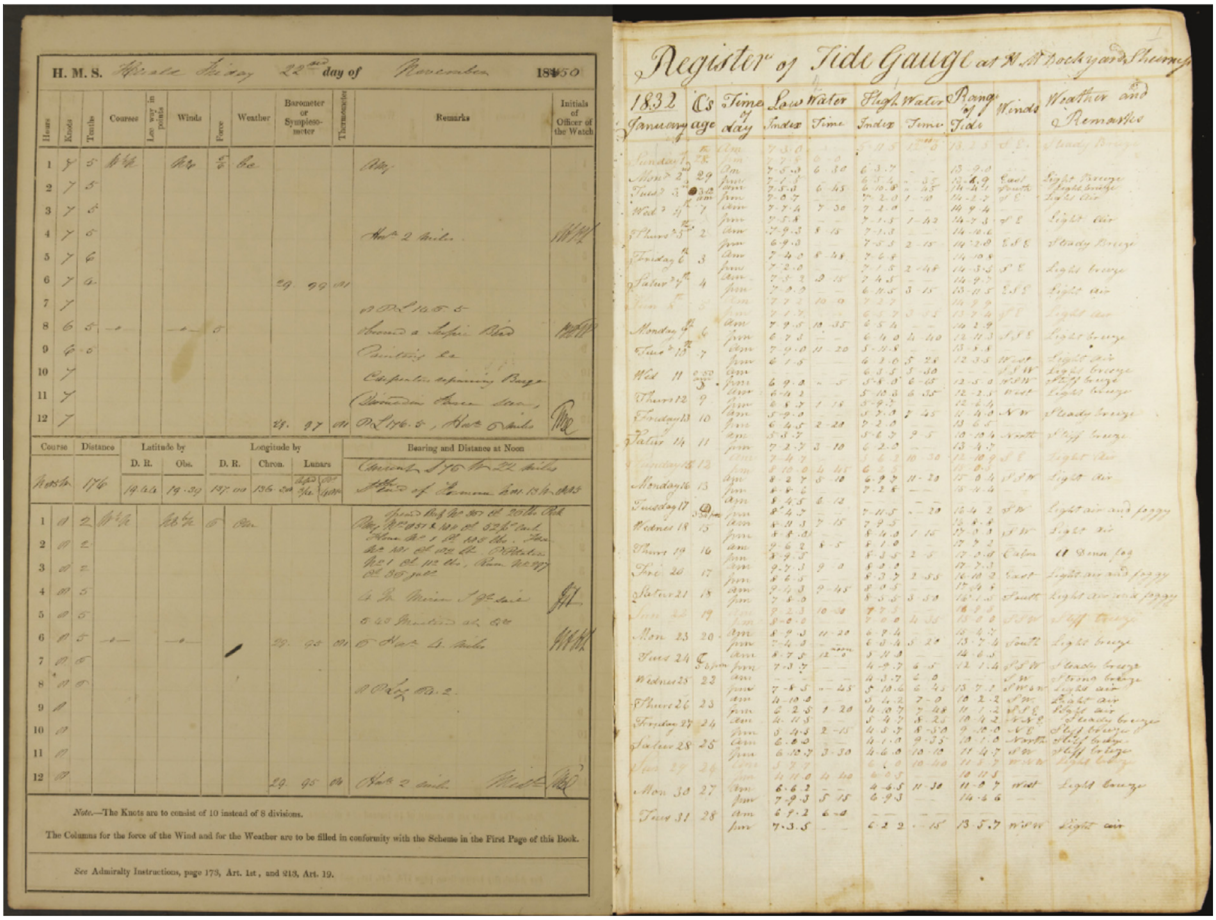

Fig. 9. Ships' log for HMS Herald, 22nd November 1850 (image on left, Meteorological Office at al. [28]) compared with Sheerness tide gauge register, January 1832 (image on right).

Any resulting data would need to be inspected for errors by volunteers, but this would be a less intensive process than entering the values manually.

5. Metadata

In order for any recovered data to be useful, it must, at the very least, be accompanied by basic metadata. GLOSS has produced a Manual on Quality Control of Sea Level Observations [30], which includes recommendations for the metadata supplied with sea level data. The metadata can be divided into three categories:

- information about the site (e.g. name, latitude and longitude)
- information about the instrument (e.g. type of gauge, manufacturer, calibration)
- information about the data processing (e.g. sampling schema, corrections applied, gaps in the record)

In order for the data to be incorporated into the PSMSL database, benchmark and datum history is required, which is often difficult to find with analogue data, but can be found in related records in hydrographic and geodetic agencies.

6. Conclusions

There are many challenges facing the sea level data archaeology community but it is hoped that improvements in technology, such as faster automated digitisation of tide gauge charts with minimal user input and automatic transcribing of handwritten ledgers, can overcome some of the obstacles. The GLOSS GE is aware of several organisations globally that are looking to start data archaeology projects and GLOSS would like to help coordinate efforts and share knowledge. The GLOSS website will act as a central location to view metadata relating to tide gauge records in archives, libraries, collections etc. and will provide a mechanism for users to report data they have uncovered. The GLOSS GE will ask its members to share software for digitising and quality controlling tide gauge data and will use the GLOSS website as a central location to access software and guidelines for quality controlling digitised data. GLOSS will organise a comparison study of the various different digitising programs. The GLOSS delayed-mode data archive centre at BODC will be the repository for high-frequency data (sampling interval of one hour or less) and the PSMSL will be the repository for any newly created mean sea level datasets.

Acknowledgements

The authors would like to acknowledge the GLOSS Group of Experts and user community for their input and assistance.

The National Oceanography Centre funded a three-month Citizen Science scoping study, which included a visit to the UK National Archives and the British Library. The Sheerness 1832 tide gauge register was scanned under the Jisc Content programme 2011-13 (Strand B: Mass Digitisation).

The authors would also like to thank Andy Matthews for producing the maps, Marta Marcos for supplying the Cadiz image, Adam Leadbetter, Justin Buck and Svetlana Jevrejeva for providing writing assistance and Paul McGarrigle for proofreading.

References

[1] GLOSS Group of Experts. Action List as agreed at GLOSS GE XIII. [online] UNESCO/Intergovernmental Oceanographic Commission. Available at: <www.ioc-unesco.org/components/com_oe/oe.php?task=download&id=23192&version=1.0&lang=1&format=1>; 2013 [accessed 26.09.2014].

[2] IPCC. Climate Change 2013. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM, editors. The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press; 2013. p. 1535.

[3] Church JA, White NJ. Sea-level rise from the late 19th to the early 21st century. Surv Geophys 2011;32(4–5):585–602.

[4] Jevrejeva S et al. Recent global sea level acceleration started over 200 years ago? Geophys Res Lett 2008;35(8).

[5] Ray RD, Douglas BC. Experiments in reconstructing twentieth-century sea levels. Prog Oceanogr 2011;91(4):496–515.

[6] Douglas BC. Global sea rise: a redetermination. Surv Geophys 1997;18(2–3):279–92.

[7] Holgate SJ et al. New data systems and products at the permanent service for mean sea level. J Coastal Res 2012;29(3):493–504.

[8] Haigh I et al. Assessing changes in extreme sea levels: application to the English Channel, 1900–2006. Cont Shelf Res 2010;30(9):1042–55.

[9] Hunter J et al. The sea level at Port Arthur, Tasmania, from 1841 to the present. Geophys Res Lett 2003;30:7.

[10] Testut Laurent et al. Sea level at Saint Paul Island, southern Indian Ocean, from 1874 to the present. J Geophys Res 2010 [(1978–2012) 115.C12].

[11] Pugh D, Wyrtki K. Plan for a global sea level network. [online] Intergovernmental Oceanographic Commission, IOC Executive Council, Paris, 17th session. Available at: <http://www.psmsl.org/train_and_info/training/gloss/general/plan_for_gloss_1984.pdf>; 1984 [accessed 22.01.2015].

[12] Intergovernmental Oceanographic Commission. Global Sea Level Observing System (GLOSS) Implementation Plan – 2012. UNESCO/Intergovernmental Oceanographic Commission, Technical Series, No. 100 2012.

[13] Cartwright DE. Tides: a scientific history. Cambridge: Cambridge University Press; 1999.

[14] Caldwell P. Tide gauge data rescue. In: Duranti L, Shaffe E, editors. Proceedings of the memory of the world in the digital age: digitization and preservation. Vancouver; 2012. p. 134–49.

[15] Talke SA, Jay DA. Nineteenth century North American and Pacific tidal data: Lost or just forgotten? J Coastal Res 2013;29(6a):118–27.

[16] Pouvreau N. Trois cents ans de mesures marégraphiques en France: outils, méthodes et tendances des composantes du niveau de la mer au port de Brest. Université de La Rochelle; 2008 [Ph. D.].

[17] Nov O et al. Technology-mediated citizen science participation: a motivational model. In: Proceedings of the fifth international AAAI conference on weblogs and social media. Menlo Park, California: AAAI Press; 2011 [Barcelona, Catalonia, Spain, 17–21 July 2011].

[18] Tweddle, JC. et al. Guide to citizen science: developing, implementing and evaluating citizen science to study biodiversity and the environment in the UK. [online]. NERC/Centre for Ecology & Hydrology. Available at: <http://www.ceh.ac.uk/products/publications/documents/citizenscienceguide.pdf>; 2012 [accessed 26.09.2014].

[19] Woodworth PL. High waters at Liverpool since 1768: the UK's longest sea level record. Geophys Res Lett 1999;26(11):1589–92.

[20] Intergovernmental Oceanographic Commission. Manual on Sea-Level Measurement and Interpretation, vol. 1, Basic Procedures, Manuals Guides, vol. 14, UNESCO/Intergovernmental Oceanographic Commission, Paris 1985.

[21] Bradshaw, E. Identify automated techniques for the digitization of archived mareograms.[online] UNESCO/Intergovernmental Oceanographic Commission. Available at: <http://www.ioc-unesco.org/index.php?option=com_oe&task=viewDocumentRecord&docID=11924>; 2013 [accessed 26.09.2014].

[22] Ullmann A et al. Tool kit helps digitize tide gauge records. Eos, Trans Am Geophys Union 2005;86(38):p342.

[23] International Environmental Data Rescue Organization. Strip Chart Digitization Program [online] Available at: <http://iedro.org/activities/strip-chart-digitization-program/>; 2014 [accessed 26.09.2014].

[24] Wöppelmann G et al. Brest sea level record: a time series construction back to the early eighteenth century. Ocean Dyn 2006;56(5-6):487–97.

[25] Marcos M et al. The long sea level record at Cadiz (southern Spain) from 1880 to 2009. J Geophys Res 2011;116(C12):1978–2012.

[26] Brohan, P. oldWeather.org: Citizen Science for Climate Reconstruction. AGU Fall Meeting Abstracts. vol. 1. San Francisco, California, 3–7 December 2012; 2012.

[27] oldWeather Team. The new Jeannette opera house, Old Weather blog, [blog] 22 December. Available at: <http://blog.oldweather.org/2012/12/22/the-new-jeannette-opera-house>; 2012 [accessed 26.09.2014].

[28] Meteorological Office et al. CORRAL (UK Colonial Registers and Royal Navy Logbooks). [online]. NCAS British Atmospheric Data Centre. Available at: <http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__dataent_1239019538627371>; 2011 [accessed 26.09.2014].

[29] Sánchez JA et al. Handwritten text recognition for historical documents in the tranScriptorium project. In: Proceedings of the first international conference on digital access to textual cultural heritage. New York: ACM; 2014 [Madrid, Spain, 19–20 May 2014].

[30] Garcia MJ et al. Manual on quality control of sea level observations.[online] UNESCO/Intergovernmental Oceanographic Commission. Available at: <http://www.iode.org/components/com_oe/oe.php?task=download&id=14632&version=1.0&lang=1&format=1>; 2011 [accessed 23.01.2015].