Adaptation For Climate Change

Advanced Workshop 1 on JRA3, FREE-Data

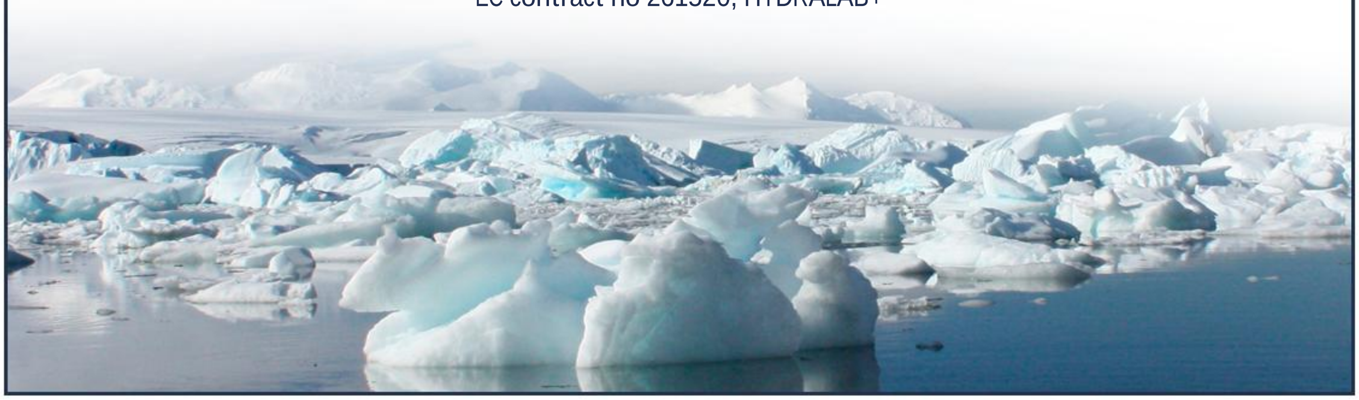
Deliverable Number 5.6

Status: Final

Version: 02

Date: 25 October 2016

EC contract no 261520, HYDRALAB+



Document Information

Title	Advanced Workshop 1 on JRA3, FREE-Data
Lead Author	James Sutherland (HR Wallingford)
Contributors	Mark Klein Breteler & Anna van Gils (Deltares)
Distribution	Public
Document Reference	D5.6

Document History

Date	Version	Prepared by	Organisation	Approved by	Status
17/10/2016	1	J. Sutherland	HR Wallingford		Draft
25/11/2016	2	J. Sutherland	HR Wallingford	F.C. Hamer	Final

Acknowledgement

The work described in this publication was supported by the European Community's Horizon 2020 Programme through the grant to the budget of the Integrated Infrastructure Initiative HYDRALAB+, Contract no. 654110.

Disclaimer

License

.

CONTENTS

Document Information 2
Document History 2
Acknowledgement 2
Disclaimer 2
License 2
Contents 3
1 Background to advanced workshop 1 on Free Data 4
2 Minutes of Advanced Workshop 1, on FREE Data 5
2.1 Introduction 6
2.2 Presentations 6
2.2.1 Data Management in HYDRALAB to date 6
2.2.2 Flexible research data management using the Hydra digital repository 7
2.2.3 Data requirements and adequacy assessments for the integrated use of marine data – an EMODENT practice for the Baltic Sea 8
2.2.4 OpenEarth 8
2.2.5 Interoperable mapping in marine and coastal science: 8
2.2.6 CEDEX physical model tests standardisation 9
2.3 Discussion and Conclusions 10

Tables

No table of figures entries found.

Figures

No table of figures entries found.

1 BACKGROUND TO ADVANCED WORKSHOP 1 ON FREE DATA

'FREE Data' stands for *Facilitating the Re-use and Exchange of Experimental Data* and it is based on the premise that maximising the effective data exchange between hydraulic laboratories and among experimental models, numerical models and field case studies is key to increasing the impact of environmental hydraulics for understanding and managing climate change adaptation. The science of climate change adaptation will improve most efficiently when the best use is made of the most appropriate data source, which is driven by effective exchange between applications. According to the HYDRALAB+ description of work:

> *"The main aims of 'FREE Data' are to develop tools and protocols for the effective sharing of data that allows effective flux and exchange with numerical modelling and field case studies. As part of the work we will test and demonstrate these tools and protocols with a range of datasets, starting with relatively simple cases and building up to the most complex and most interdisciplinary datasets that will be generated in HYDRALAB+: those produced by work on eco-hydraulics.*
>
> *FREE Data has the following objectives:*
>
> - *To develop data standards and licenses to better facilitate the re-use of data;*
> - *To develop a novel and flexible data repository to facilitate the exchange of data;*
> - *To develop more effective knowledge sharing tools for community and stakeholder engagement;*
> - *To develop novel methods and protocols for interaction and effective data exchange between laboratory and numerical models and field observations."*

There are many ways of approaching this problem and many different organisations and communities developing their own approaches. HYDRALAB+ Advanced Workshop 1, on FREE Data was set up so that he HYDRALAB+ community could hear about some of the different approaches that have been developed or are in development. The minutes of this meeting are contained in Section 2, while Section 3 contains the presentations made during the event.

2 MINUTES OF ADVANCED WORKSHOP 1, ON FREE DATA

Hotel Mercure Gdansk Stare Miasto

Wednesday 21 September 2016, 09:00 – 12:30.

Attendees

Name	Institution	Name	Institution
James Sutherland	HR Wallingford	Malgorzata Bielecka	IBW Pan
Frans Hamer	Deltares	Peter Thorn	NOC (NERC)
Mark Klein Breteler	Deltares	Matthias Kudella	LUH
Anna van Gils*	Deltares	Paulo Jorge Rosa Santos	FEUP (U Porto)
Chris Awre*	U Hull	Javier López Lara	U Cantabria
Stuart McLelland	U Hull	Rita Carvalho	U Coimbra
Wietse van de Lageweg	U Hull	Ramon Gutierrez	CEDEX
Jens Murawski*	DMI	Estelle Morris	SAMUI
Miko Suominen	Aalto U	Jochen Aberle	NTNU
Karl-Ulrich Evers	HSVA	Agnieszke Herman	U Gdansk
Juana Fortes	LNEC	Agustin Sanchez-Arcilla	UPC
Gonzalo Malvarez*	U Pablo de Olavide	Iván Cácares	UPC
Shuting Yang	DMI	Maria Ionescu	GeoEcoMar
Björn Elsäßer	DHI	Frederic Moulin	CNRS
Thor U Petersen	DHI	Pierre-Yves Henry	NTNU
Elizabeth Bradshaw	BODC	Mindert de Vries	Deltares
Rosaria Ester Musumeci	U Catania	Anne Middelboe	DHI
Joël Sommeria	CNRS	Tom O'Donoghue	U Aberdeen
Grzegorz Różyński	IBW Pan		

* = Invited speaker. There were 39 people at the workshop (28 male, 11 female).

2.1 INTRODUCTION

The Coordinator of HYDRALAB+, Frans Hamer (Deltares) welcomed people to the workshop, particularly the invited participants, and emphasised the importance of data collection, storage and sharing to HYDRALAB.

2.2 PRESENTATIONS

2.2.1 Data Management in HYDRALAB to date

James Sutherland (HR Wallingford) presented information on the history of HYDRALAB, the types of facilities used and the efforts made to date on data management. From HYDRALAB-III, each project has consisted of:

- networking activities, which share knowledge,
- Joint Research Activities, which seek to improve the services offered by European hydraulic laboratories through the development of advanced instrumentation, methods and protocols.
- Transnational Access, where teams of researchers from different countries come to the rare and unusual facilities within the HYDRALAB network to perform experiments that they would not be able to undertake in their home countries.

Traditionally, each institution has used its own data acquisition, storage and analysis systems. There has been a shortage of meta-data collected, little standardisation and little consideration of data exchanges. Four initiatives have considered data exchange and data management:

1. Transnational Access Data Management Plans and Data Storage Reports have common formats, with free-text entries into the sections. These documents provide much of the information that is useful for discovery metadata, and normally describe the location of instruments and the conditions run for each test.
2. The HYDRALAB-III (2006-2010) Joint Research Activity on Composite Modelling (described as the balanced use of physical and numerical models) emphasised the need for protocols for data exchange.
3. The HYDRALAB-III Data Management Tools report (2009) made a number of recommendations, including that HYDRALAB build on existing technologies, develop a strategic view for data management, establish best practice for the documentation and management of data, adopt a standards-based approach to data management (including adopting the EC's CERIF data model for metadata) and identify a limited number of data formats for data exchange.
4. HYDRALAB-IV (2010-2014) shared meta-data about TA experiments using the UK Environmental Observation Framework (http:www.ukeof.org.uk) which is implementing data services based around the INSPIRE data standards for Environmental Monitoring Facilities. This involved mapping of the data in the HYDRALAB database to the UK-EOF schema, a bulk transfer of HYDRALAB data into the UK-EOF and accessing data from the UK-EOF catalogues through its application programme interface (api). This was partially successful, but was limited by the late delivery of the schema, and the incomplete mapping between HYDRALAB database and the UK-EOF schema.

The present HYDRALAB+ (2015-2019) has a Joint Research Activity on 'Facilitating the Reuse and Exchange of Environmental Data', known as FREE-Data, which is investigating:

- adopting metadata standards, chosen by participants;
- using unique identifiers for researchers, such as ORCID;
- Publishing data;
- The use of unique identifiers for datasets, such as digital object identifiers (DOIs);
- Using 'open' licenses for data and reports; and
- The potential development of common vocabularies for HYDRALAB terms

2.2.2 Flexible research data management using the Hydra digital repository

Chris Awre, Head of Information Services at the University of Hull, gave a presentation entitled "Flexible research data management using the Hydra digital repository at the University of Hull". Current research data management (RDM) initiatives at Hull are based on three main trends;

- The amount of data is growing;
- Data management is required across more disciplines; and
- there is an increasing perception of the value of data.

The Information Services team provide support for research data management throughout the data life-cycle. This includes providing guidance, training on data management http://libguides.hull.ac.uk/researchdata and access to templates for Data Management Plans, such as http://dmponline.dcc.ac.uk (which can be tailored for the HYDRALAB+ funding mechanism).

The University of Hull participates in Hydra, a multi-partner, open source initiative that can be applied to all areas of university research. Awre noted that it is based on 2 assumptions:

1. "No single system can provide the full range of repository-based solutions for a given institution's needs, yet sustainable solutions require a common repository infrastructure.
2. No single institution can resource the development of a full range of solutions on its own, yet each needs the flexibility to tailor solutions to local demands and workflows."

Therefore, they have denhanced Fedora as a digital repository system, with a range of customised 'Hydra heads' to suit different research communities' needs. Its original use at Hull was as a repository for theses, but it is most commonly used for open access journal articles. It's use is not compulsory for data, but researchers at Hull must put a record in Hydra to record data generated. In this way, data can be managed through a variety of systems, then metadata shared for discovery through a single point. The repository has disk space to archive data, provided by the university as an investment in the future.

2.2.3 Data requirements and adequacy assessments for the integrated use of marine data – an EMODENT practice for the Baltic Sea

Jens Murawski (DMI) gave a presentation on EMODNET (2009-2020), the European Marine Observation and Data Network, which is a long term marine data initiative from DG-MARE underpinning its Marine Knowledge 2020 strategy. This has seven data lots: bathymetry, biology, chemistry, geology, human activities, physics, sea bed habitat. However, the value of data is only realized when it is used and for that, it has to be fit for purpose. The EMODnet Baltic Sea Checkpoint is one of a series of regional projects set up "to assess the quality and adequacy of the current observation monitoring data … by testing the data against specific end-user challenges." The data adequacy assessments check data accessibility, completeness and coverage, resolution and precision. Limitations in several of the datasets have been identified, when used for particular challenges. This presentation showed that making data available is not enough, it must be understandable and sufficient to meet the user's needs.

2.2.4 OpenEarth

Ana van Gils (Deltares) presented information on the OpenEarth initiative, which was started by Deltares but is now international. Five levels of data are defined:

1. Raw data, collected by scientists;
2. Standard data – either NetCDF with CF conventions or a PostgreSQL implementation with PostGIS depending on the data type;
3. Tailored data – which is derived from one or more standard data sources to meet the needs of the professional user, e.g. significant wave height from a surface elevation time series, or a velocity field derived from a PIV experiment.
4. Graphics of data, using OGC standards; and
5. Catalogue of meta-data records.

The raw data coming from the measurement devices is stored in subversion together with a description of the format and the scripts used to process the data. Conversion to standard formats allows other people to easily access and use the data. For example, the sand motor dataset is available on line at https://zandmotordata.nl/ using OpenEarth.

2.2.5 Interoperable mapping in marine and coastal science:

Professor Gonzalo Malvarez (Universidad Pablo de Olavide, Seville) presented his work on data interoperability from a range of European projects, including PEGASO, MEDINA and COASTGAP. The Medina Electronic Infrastructure and Pegaso Spatial Data Infrastructure both rely on OGC standards to cope with many, diverse sensors and users. They also rely on INSPIRE concepts and methods and had to be reliable. The Medina project liaised between mapping people and scientific laboratories. It used INSPIRE directive to standardize meta-data, although there was difficulty in convincing people to collect meta-data and share their data. The Medina E-Infrastructure (MEI) was a mapping tool (or spatial data infrastructure) to disseminate Medina products and enhance GEOSS (Group on Earth Observations System of Systems) use in marine monitoring. It incorporated INSPIRE, GEOSS and OGC standards. The main interface of the MEI was a map viewer, which had a variety of tools (such as query, measure distance, time slider, synchronization, split screen and zooming) for exploiting the results.

Recommended technologies for future developments include:

- the ISO 19156:2011 standard on Observations and Measurements (O&M) which defines a conceptual schema for observations (such as point time series) and for features involved in sampling.
- Sensor Observation Service (SOS) is a web service interface specification for discovery and access of sensor observations.
- Internet of Things. See report on "IoT Semantic Interoperability: Research Challenges, Best Practices, Recommendations and Next Steps" by the European research cluster on the internet of things. March, 2015
- WaterML 2.0 information model for the representation of water observations data.

Professor Alvarez's main message was to use existing standards and not to reinvent the wheel. You have to be able to accommodate the wide scope of data, which might be inconsistent. You must be prepared to devote adequate resources to data management.

2.2.6 CEDEX physical model tests standardisation

Ramon Guttierez Serret (CEDEX) presented the standardized way of recording physical model test data in CEDEX and illustrated this with three case studies:

- Barcelona Port South breakwater;
- Bilbao Port Punta Sollana breakwater; and
- Ship mooring of the new Container Terminal Port of Cádiz.

Raw data from every experiment is catalogued, classified and stored, with the information organized into a general test record card, test configuration record card, raw data files, and reports. CEDEX have definitions and protocols for each stage of the process and have effectively created their own ontology.

2.3 Discussion and Conclusions

The speakers and audience then discussed the presentations and related issues and concerns. The following conclusions can be drawn from this discussion:

- Flexibility is necessary. There is not one solution for all circumstances or all data types.
- Data will continue to be collected in a variety of formats, but these formats should be defined and a minimum level of meta-data is required for the data to be useful in the future.
- Journals are now encouraging authors to publish the data that went into a paper. Publishing data with a paper will become more common and is already mandatory for some journals. The journal may specify a list of acceptable data repositories. This data must be understandable, which will be helped by the use of documented (preferably standard) formats and meta-data.
- The publication of data-processing scripts alongside datasets makes the treatment of data and the production of derived parameters (that go into graphs and tables) transparent and open. Some data management systems, such as OpenEarth, already encourage the storage of data and processing scripts in version control systems, such as subversion, which allow the data-processing chain to be followed. This is likely to become more common in future.
- Good solutions are based on the use of standards. The adoption of internationally accepted standards (such as INSPIRE, OGC standards) greatly improves interoperability and removes the need to reinvent the wheel.
- Existing standards lack depth - they do not specify sufficient details to meet the specialist needs of the HYDRALAB community - but can be expanded to meet those needs, with the hope that the extra HYDRALAB features could be considered for inclusion in the standard in due course.
- There are tools and platforms that can be adopted and adapted and this is preferable to everyone writing their own. After all, no single institution can resource the development of a full range of solutions on its own.