Data Management for 21st Century Ocean Research

Cyndy Chandler
Biological and Chemical Oceanography Data Management Office
Woods Hole Oceanographic Institution

British Oceanographic Data Centre
24 April 2015 Liverpool, England

Data Management

- Has always been essential to the scientific process
- Much has changed in my lifetime with respect to data management
- Those changes represent challenges and opportunities
- I have been managing field research data since I was 8 years old

The Data Harvest

Recently published report from RDA Europe (December 2014)

Research Data Alliance - Europe

The Data Revolution

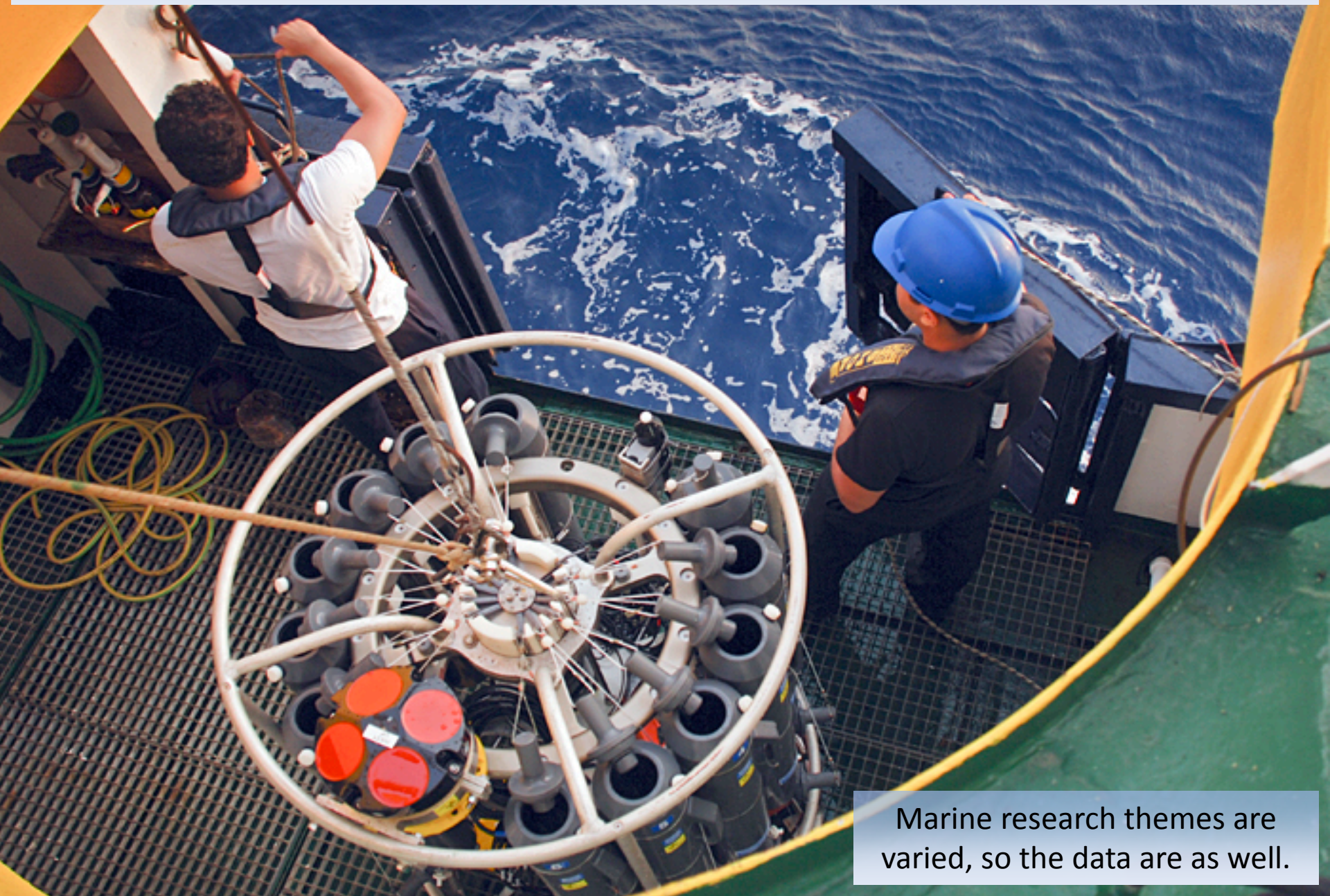
- "[The data revolution] isn't just about the volume of scientific data; rather, it reflects a fundamental change in the way science is conducted, who does it, who pays for it and who benefits from it. And most importantly, the rising capacity to share all [these] data – electronically, efficiently, across borders and disciplines – magnifies the impact."

The Data Harvest: How sharing research data can yield knowledge, jobs and growth (RDA Europe, December 2014)

Eight Global Sustainability Challenges

- Delivering water, energy, and food for all.
- Decoupling carbon emissions from economic growth.
- Safeguarding land, freshwater and marine natural assets.
- Building healthy, resilient and productive cities.
- Promoting sustainable rural futures.
- Improving human health by incorporating global change concerns.
- Encouraging sustainable consumption and production patterns.
- Improving governance and early warning systems to respond to complex future threats.

All photo images courtesy of Woods Hole Oceanographic Institution personnel or those who have sailed on one of our vessels.



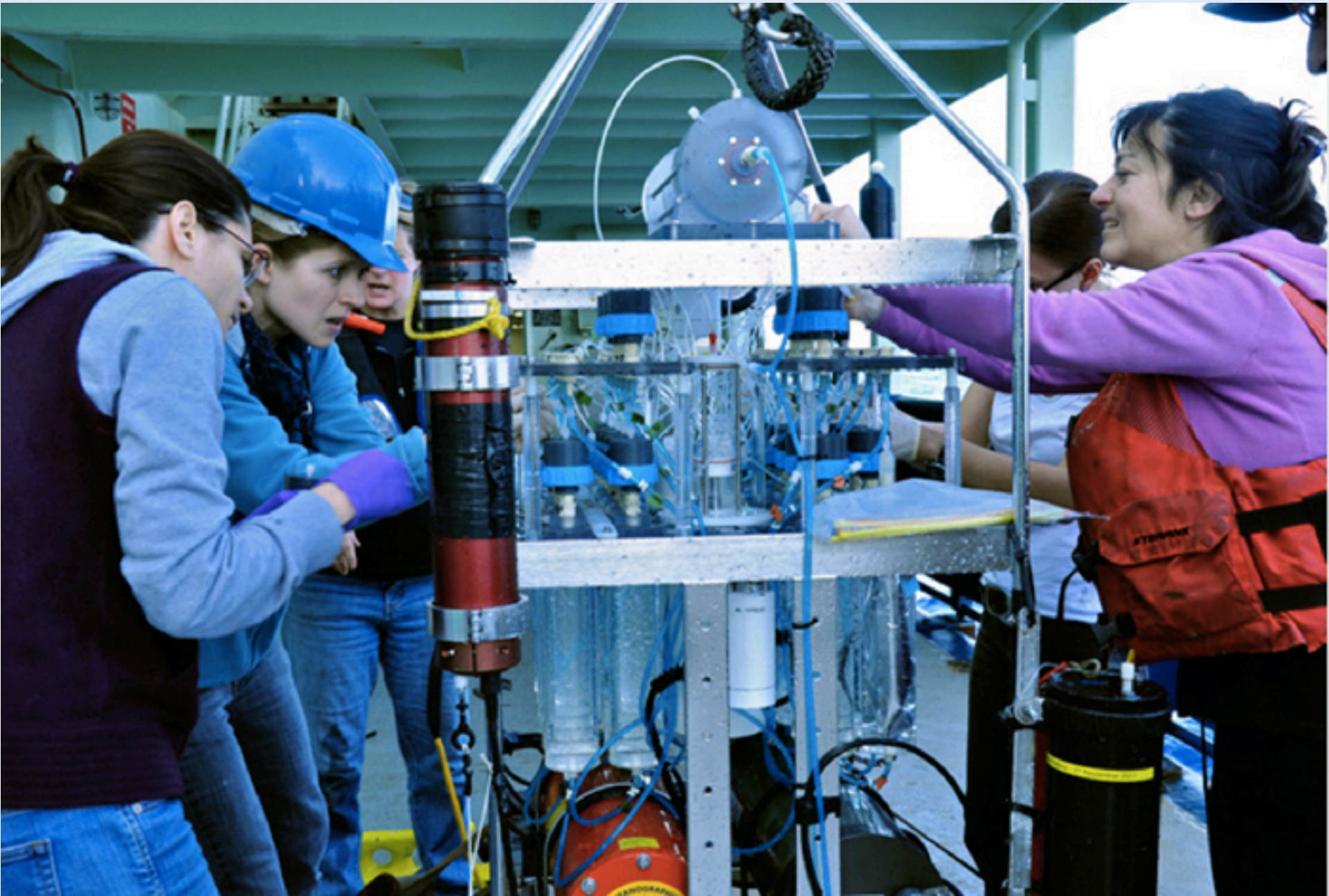
Marine research themes are varied, so the data are as well.

biological data

Cavolinia uncinata (photo by Karen Osborn, Smithsonian Institution, 2012)

marine mammals

in situ biogeochemistry data



Sea floor bathymetry

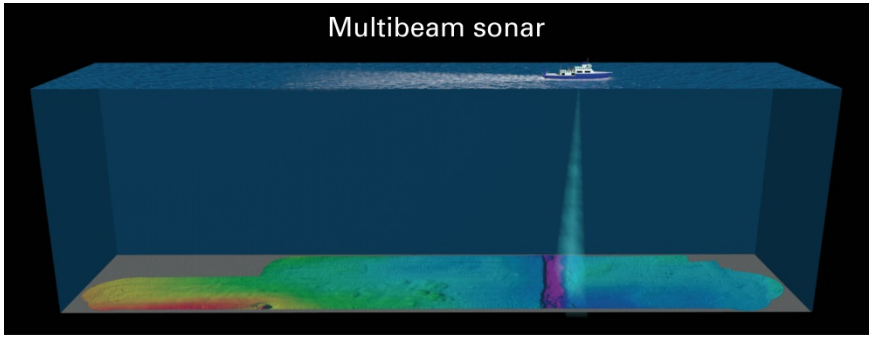
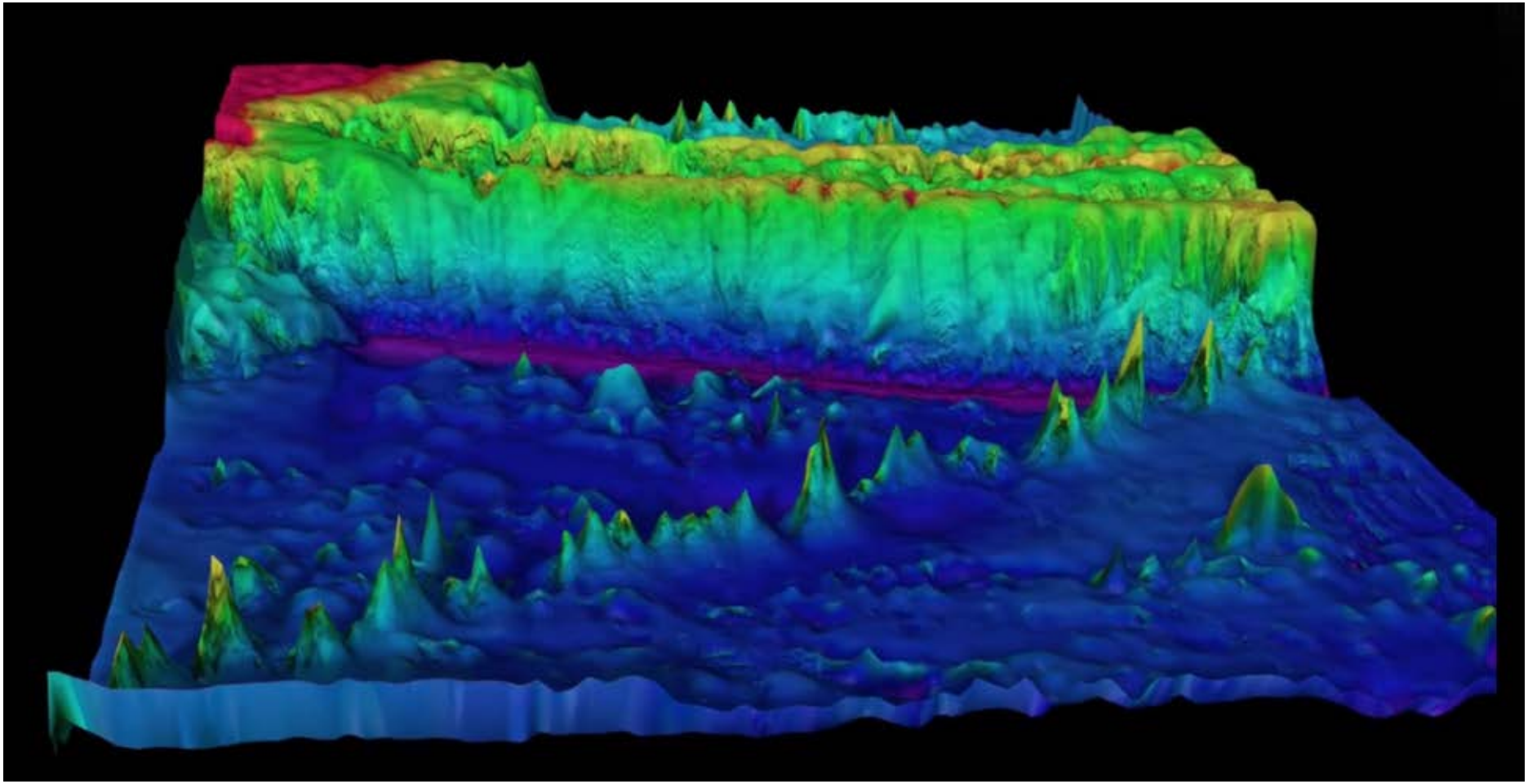
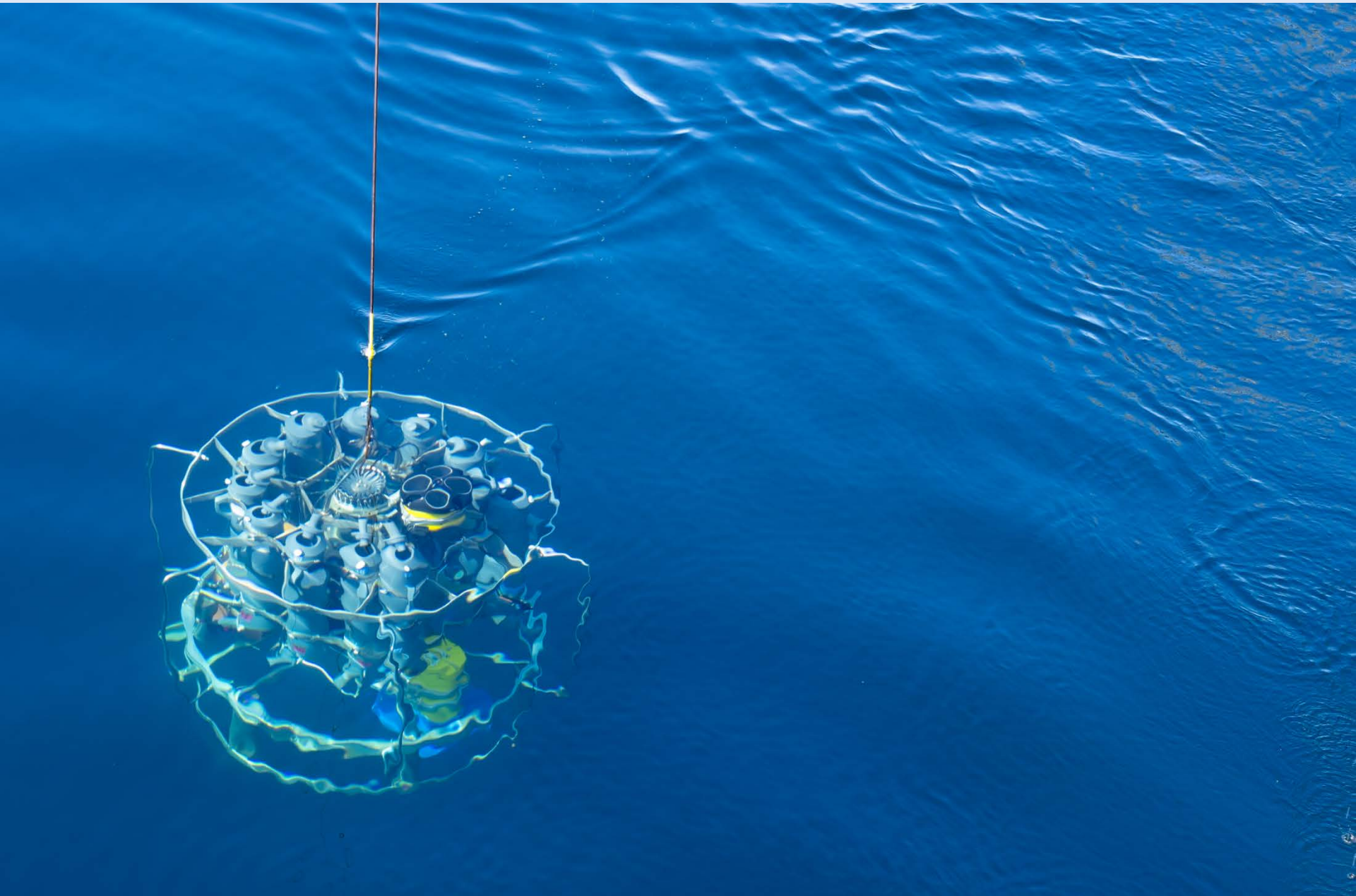


Illustration by Jack Cook, WHOI

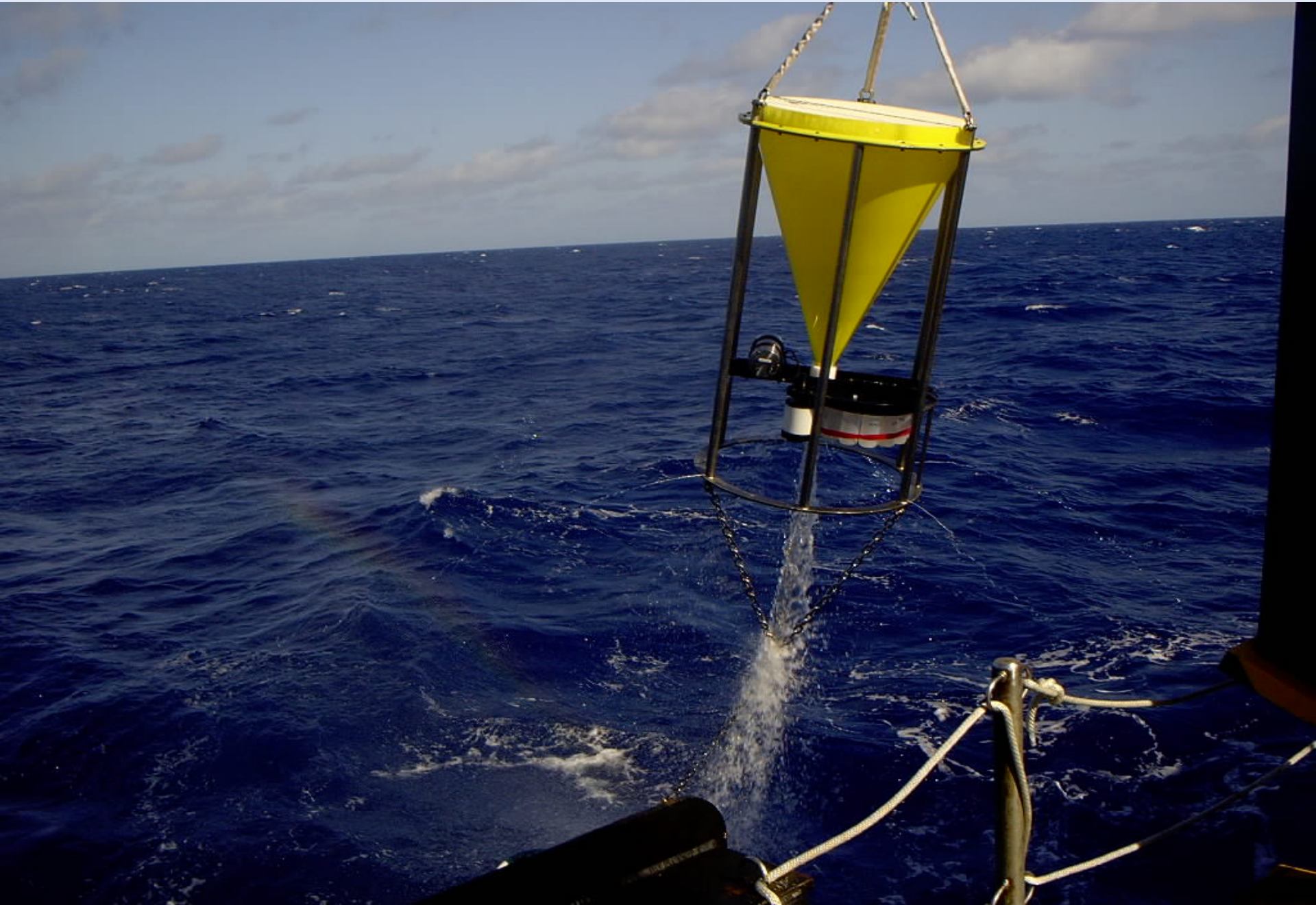


water column data

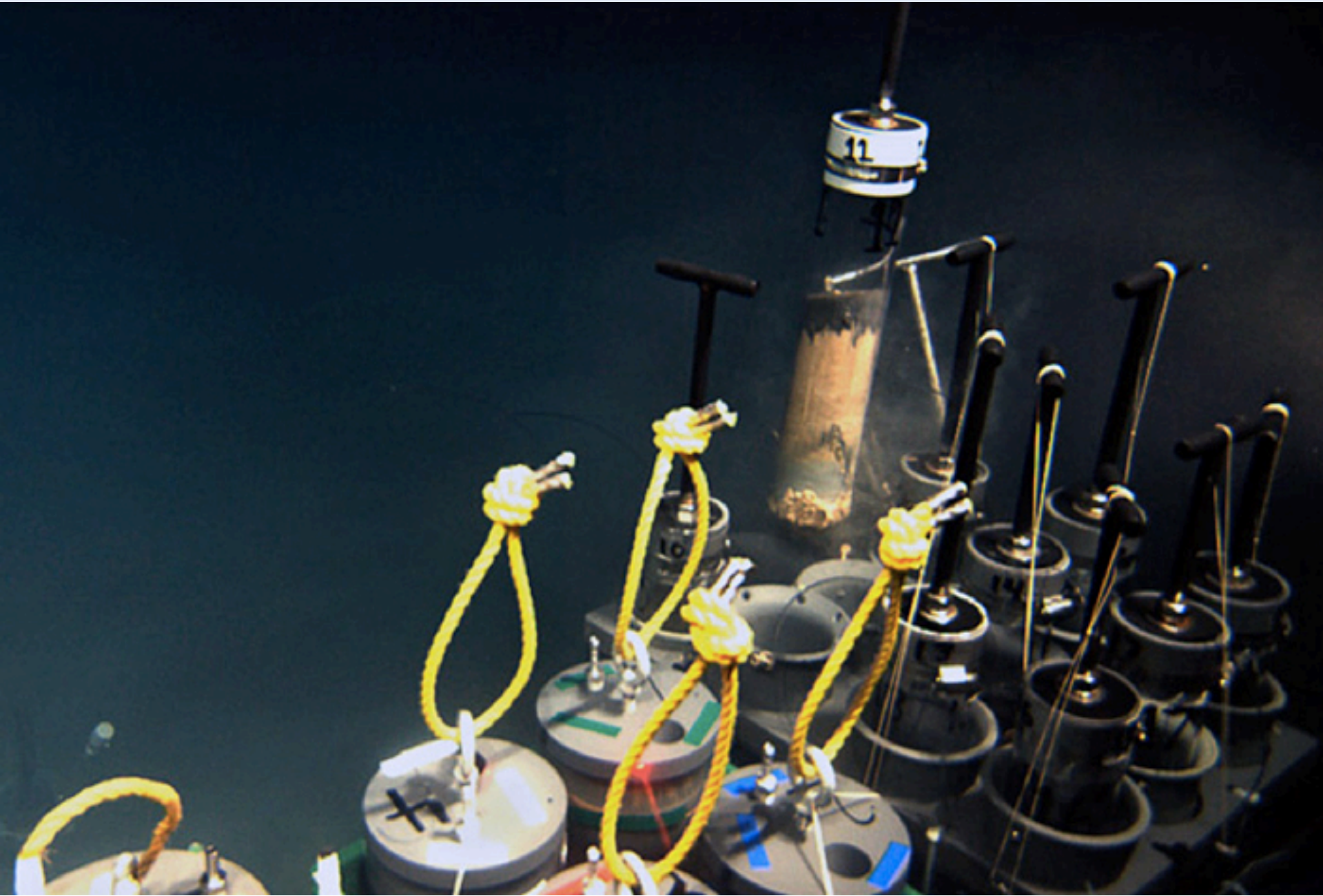


data from mooring deployments

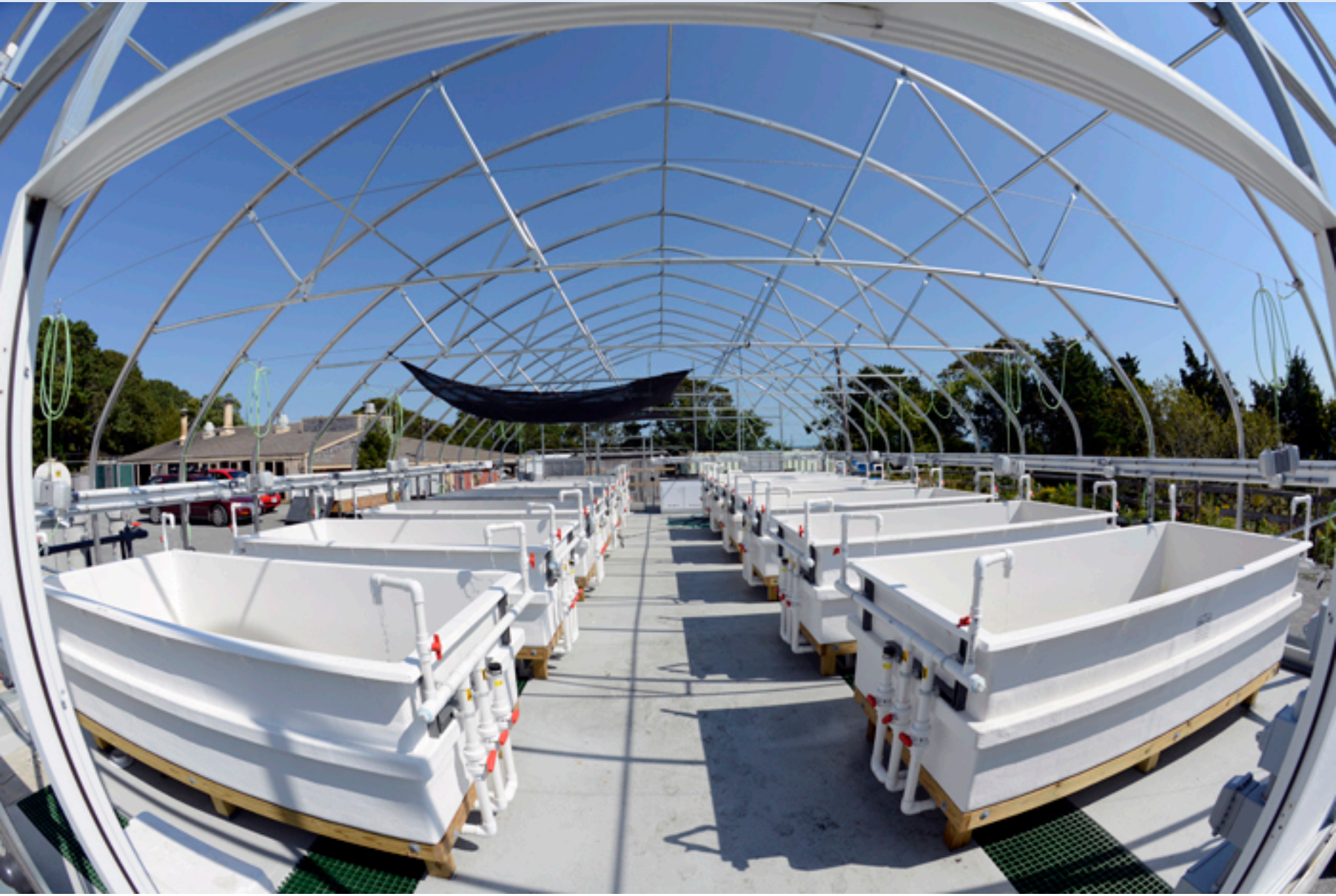
data from sediment traps



deep ocean ecosystem dynamics studies



laboratory and mesocosm experiments




When the data have been collected …
how will the data be managed and
will the data be truly available?

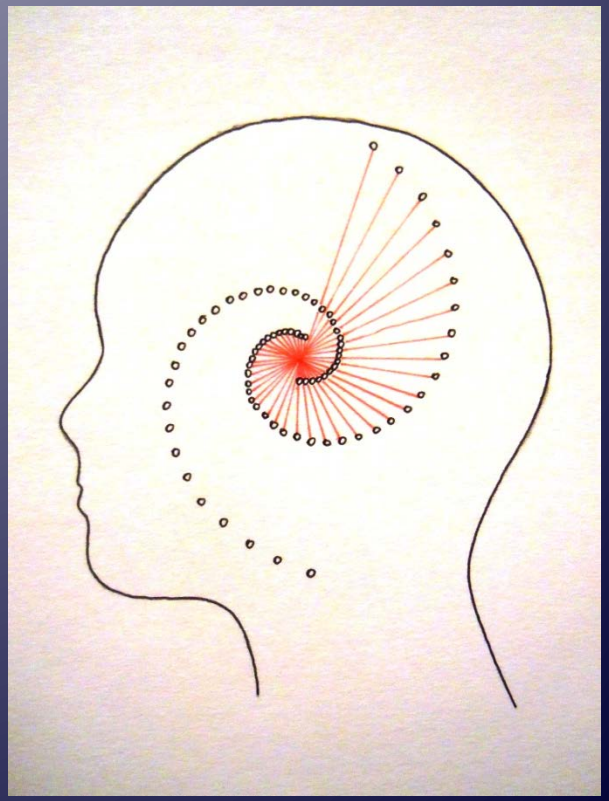
when the research is complete … what about the data?

Big Challenges & Big Opportunities

- Complex, large scale research questions
- Infrastructure (people, machines, systems) must be updated to support new research requirements
- Increased need for explicit declaration of machine–actionable information
- Standards-compliant metadata
- Documenting open-access data

Connectivity Challenges

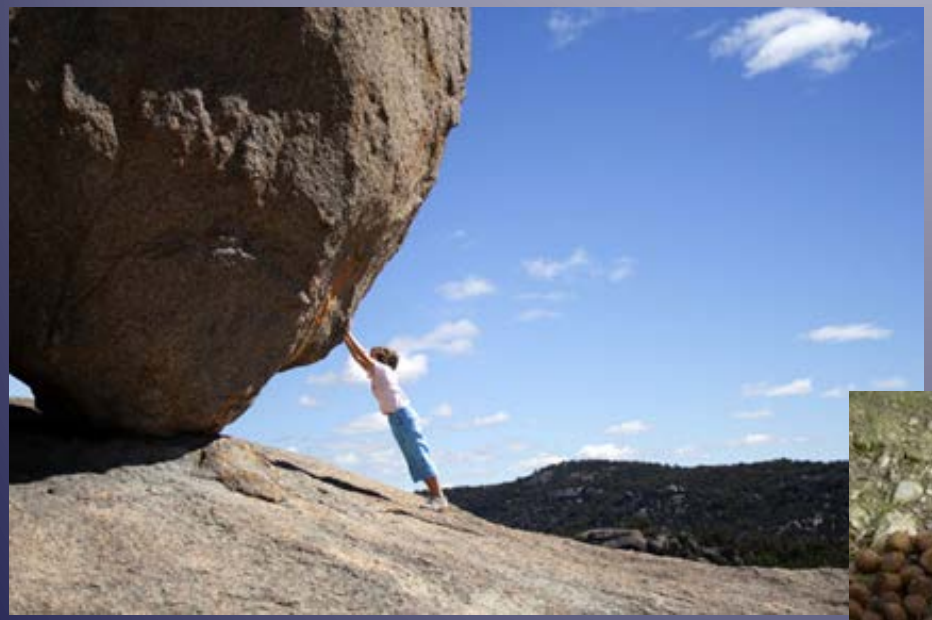
- Goals:
  - linking content at distributed repositories
  - improved interoperability (machine-to-machine)
- Technical strategies/solutions:
  - metadata content standards
  - controlled vocabularies
  - Brokering, Linked Data
- Not just technical
  - cultural conditions, behaviors
  - research data lifecycle
  - “proposal to preservation”



Connecting ALL Research Products

- DATA (raw/original and final form)
- Original proposal, science plan
- Cruise reports
- Conference abstracts and presentations
- Traditional publications
- Social media
- Citizen science

Thoughts at this point?



This seems like
a Sisyphean task ...

... with ever increasing moments when you realize there is so much to do, that you should probably just nap instead.

U.S. NSF-FUNDED RESEARCH DATA

BCO-DMO

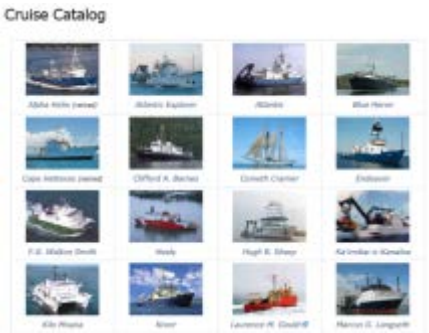
Biological & Chemical Oceanography Data Management Office

- funded by US NSF to provide data management support at no cost to researchers funded by the program managers who fund BCO-DMO
- data from current, hypothesis-driven research projects, and legacy data from large coordinated research programs
  (e.g. US GLOBEC and US JGOFS)

EXAMPLE: OPEN ACCESS CRUISE DATA

Rolling Deck to Repository (R2R)

rvdata.us



1. NSF funds a research cruise.

2. R2R serves the original underway data.

Biological & Chemical Oceanography Data Management Office

3. BCO-DMO manages the post cruise data.

4. meeting abstracts and

5. formal publications (data and peer-reviewed papers) complete the research data life cycle.

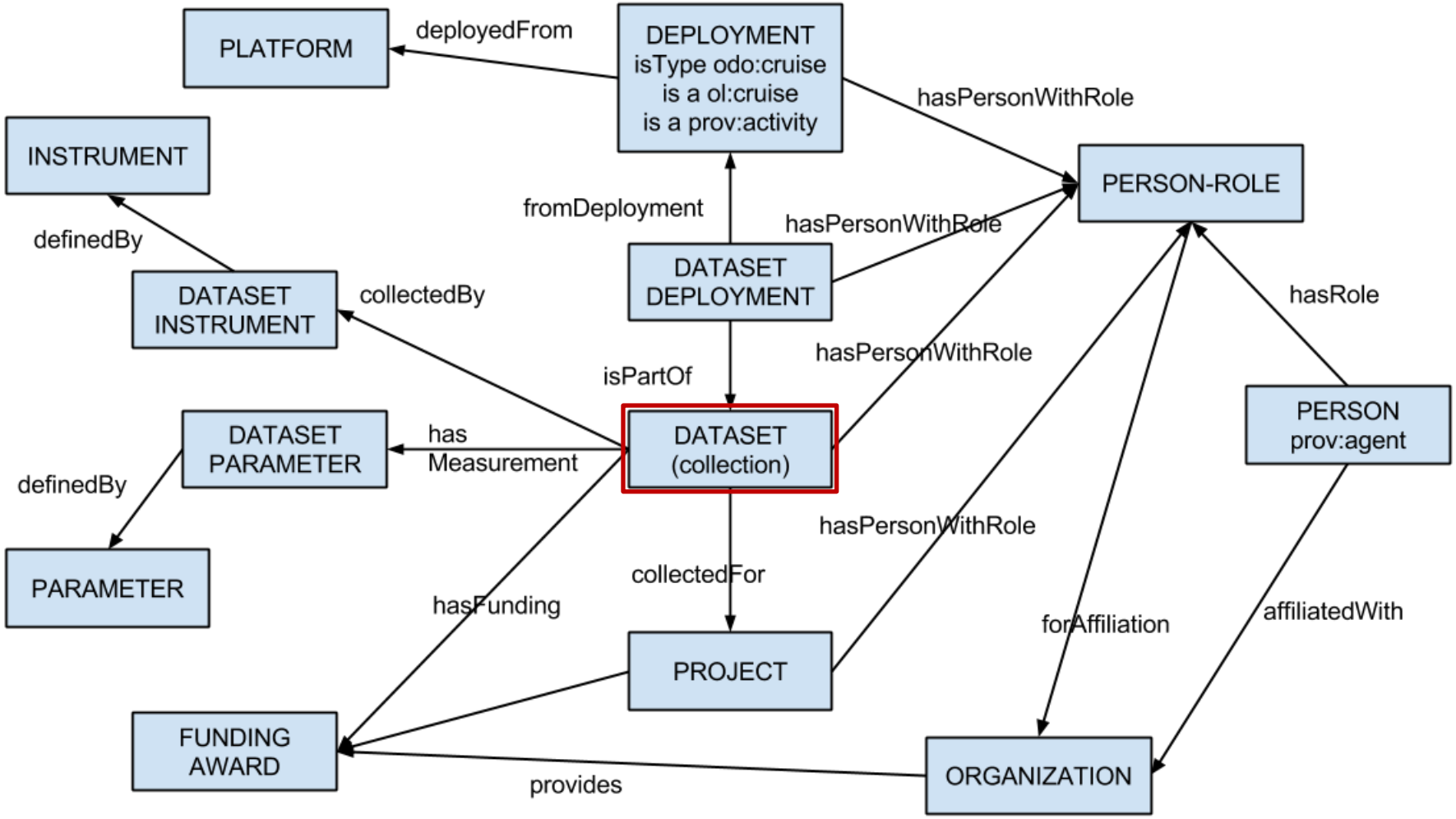
American Geophysical Union

MBLWHOI LIBRARY

Resources are discoverable at each repository, but what if we could connect them?

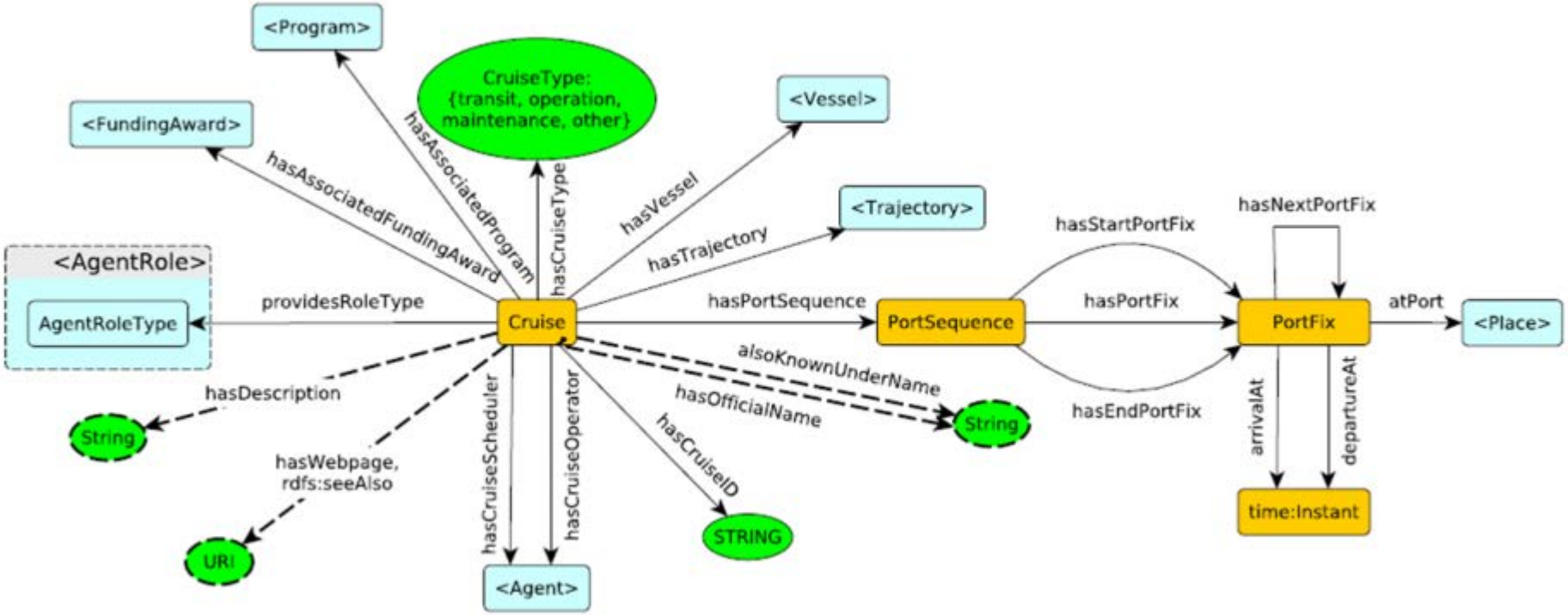
bco-dmo.org

MARINE DATA MAIN CONCEPTS



ONTOLOGY DESIGN PATTERNS

Create templates that describe the important aspects of each concept.



VOCABULARIES: CRUISE RELATED

R2R and BCO-DMO both:

define what cruise, platform (e.g. vessel) and instrument are; match them to NVS terms

Use ICES platform codes

"R/V OCEANUS" http://vocab.nerc.ac.uk/collection/C17/current/32OC/

NATURAL ENVIRONMENT RESEARCH COUNCIL

Terms from the NERC Vocabulary Server (NVS) are important for federating content from distributed systems (Leadbetter *et al*. 2013a).

VOCABULARY MATCHING

Search Cruise

NATURAL ENVIRONMENT RESEARCH COUNCIL

Vocabulary Server

Rolling Deck to Repository (R2R)

rvdata.us

AGU American Geophysical Union

BCO-DMO Biological & Chemical Oceanography Data Management Office

MBLWHOI LIBRARY

All of these repositories share a common definition of cruise (from NERC Vocabulary Server), published out with the data resources when they are expressed as Linked Data.

Leadbetter *et al*. 2013b

bco-dmo.org

An example

- A researcher reads a paper
  - We have already assumed they have found and are able to retrieve the paper

http://www.pnas.org/content/111/22/8089.full

Patrick Martin, Sonya T. Dyhrman, Michael W. Lomas, Nicole J. Poulton, and Benjamin A. S. Van Mooy (2014) "Accumulation and enhanced cycling of polyphosphate by Sargasso Sea plankton in response to low phosphorus" *PNAS 2014 111 (22) 8089-8094; published ahead of print April 21, 2014, doi:10.1073/pnas.1321719111*

An example

- A researcher reads a paper
  - We have already assumed they have found and are able to retrieve the paper

http://www.pnas.org/content/111/22/8089.full

Patrick Martin, Sonya T. Dyhrman, Michael W. Lomas, Nicole J. Poulton, and Benjamin A. S. Van Mooy (2014) "Accumulation and enhanced cycling of polyphosphate by Sargasso Sea plankton in response to low phosphorus" *PNAS 2014 111 (22) 8089-8094; published ahead of print April 21, 2014, doi:10.1073/pnas.1321719111*

An example

- A researcher reads a paper
  - We have already assumed they have found and are able to retrieve the paper

http://www.pnas.org/content/111/22/8089.full

Patrick Martin, Sonya T. Dyhrman, Michael W. Lomas, Nicole J. Poulton, and Benjamin A. S. Van Mooy (2014) "Accumulation and enhanced cycling of polyphosphate by Sargasso Sea plankton in response to low phosphorus" *PNAS 2014 111 (22) 8089-8094; published ahead of print April 21, 2014, doi:10.1073/pnas.1321719111*

An example

- A researcher reads a paper
  - We have already assumed they have found and are able to retrieve the paper

http://www.pnas.org/content/111/22/8089.full

Patrick Martin, Sonya T. Dyhrman, Michael W. Lomas, Nicole J. Poulton, and Benjamin A. S. Van Mooy (2014) "Accumulation and enhanced cycling of polyphosphate by Sargasso Sea plankton in response to low phosphorus" *PNAS 2014 111 (22) 8089-8094; published ahead of print April 21, 2014, doi:10.1073/pnas.1321719111*

An example

- A researcher reads a paper
  - We have already assumed they have found and are able to retrieve the paper

http://www.pnas.org/content/111/22/8089.full

Patrick Martin, Sonya T. Dyhrman, Michael W. Lomas, Nicole J. Poulton, and Benjamin A. S. Van Mooy (2014) "Accumulation and enhanced cycling of polyphosphate by Sargasso Sea plankton in response to low phosphorus" *PNAS 2014 111 (22) 8089-8094; published ahead of print April 21, 2014, doi:10.1073/pnas.1321719111*

there is a data supplement

Proceedings of the National Academy of Sciences of the United States of America

PNAS Celebrating 100 Years 1915-2015

> Current Issue > vol. 111 no. 22 > Patrick Martin, 8089–8094, doi: 10.1073/pnas.1321719111

CrossMark click for updates

Accumulation and enhanced cycling of polyphosphate by Sargasso Sea plankton in response to low phosphorus

Patrick Martin[a,1], Sonya T. Dyhrman[b,2], Michael W. Lomas[c], Nicole J. Poulton[c], and Benjamin A. S. Van Mooy[a,3]

Significance

Phosphorus is scarce in many subtropical ocean regions, and phytoplankton in these regions adjust their biochemical composition such that they require less of it. We show here that phytoplankton in

Footnotes

[1]Present address: Earth Observatory of Singapore, Nanyang Technological University, Singapore 639798.

See Commentary on page 7890.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1321719111/-/DCSupplemental.

there is a data supplement

Proceedings of the National Academy of Sciences of the United States of America

PNAS Celebrating 100 Years 1915-2015

> Current Issue > vol. 111 no. 22 > Patrick Martin, 8089–8094, doi: 10.1073/pnas.1321719111

CrossMark click for updates

Accumulation and enhanced cycling of polyphosphate by Sargasso Sea plankton in response to low phosphorus

Patrick Martin[a,1], Sonya T. Dyhrman[b,2], Michael W. Lomas[c], Nicole J. Poulton[c], and Benjamin A. S. Van Mooy[a,3]

Significance

Phosphorus is scarce in many subtropical ocean regions, and phytoplankton in these regions adjust their biochemical composition such that they require less of it. We show here that phytoplankton in

Footnotes

[1]Present address: Earth Observatory of Singapore, Nanyang Technological University, Singapore 639798.

See Commentary on page 7890.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1321719111/-/DCSupplemental.

DOI

This Issue

June 3, 2014
vol. 111 no. 22
Masthead (PDF)
Table of Contents

PREV ARTICLE NEXT ARTICLE

Article Tools

Article Alerts

Figures

Fig. 1. Browse All Figures

Other Articles

Known Facts

- Publication: PNAS, has a DOI, has data suppl.
- Person name (author): Benjamin Van Mooy
- Dates of activity: 2010 and 2012
- Location keywords: Sargasso Sea
- Cruise: on vessel Knorr
- Data keywords: plankton, polyphosphate, lipid

Known Facts

general knowledge

- Publication: PNAS, has a DOI, has data suppl.
- Person name (author): Benjamin Van Mooy
- Dates of activity: 2010 and 2012
- Location keywords: Sargasso Sea
- Cruise: on vessel Knorr
- Data keywords: plankton, polyphosphate, lipid

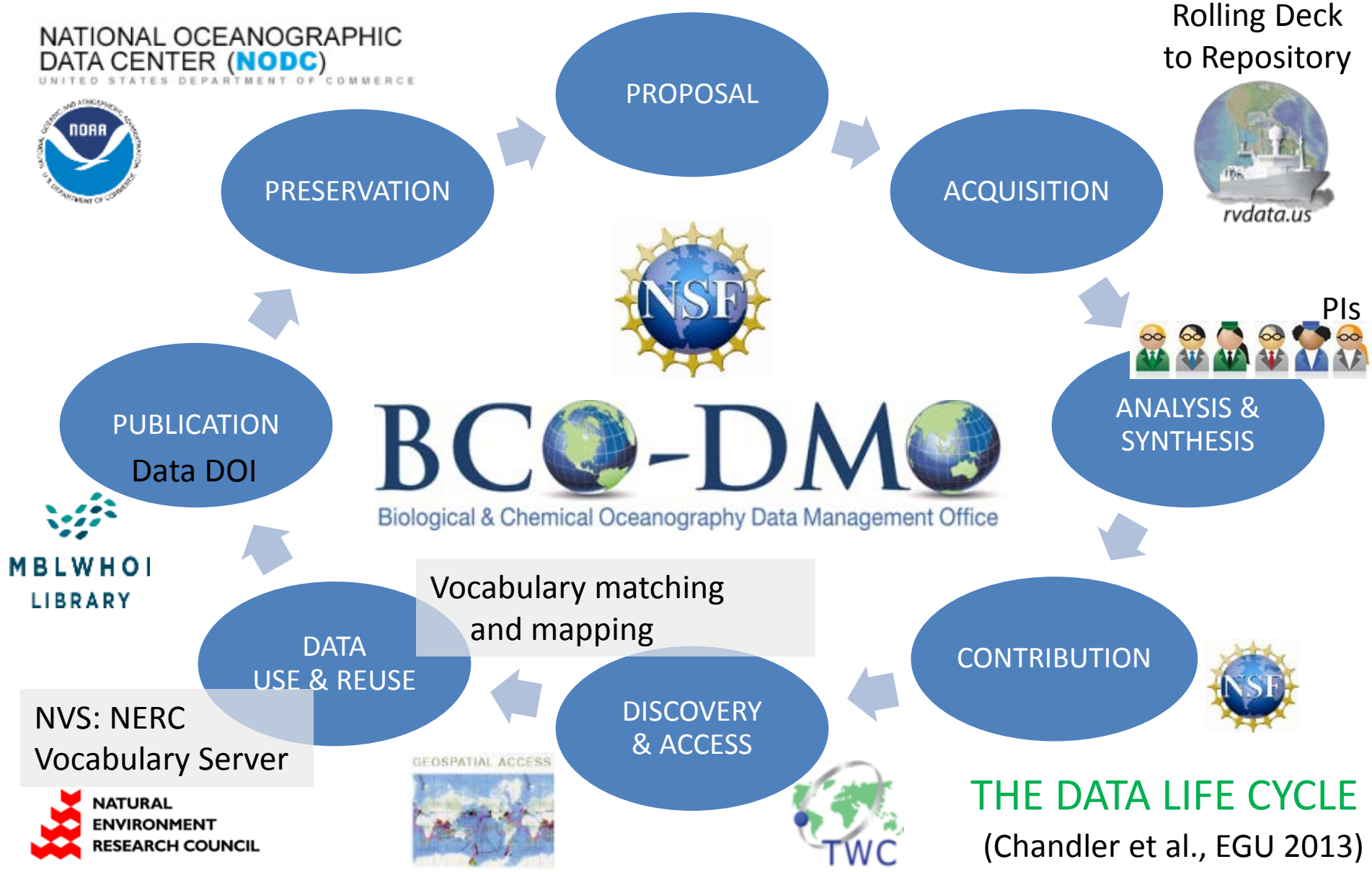
domain specific

Recommended Goal

- Each fact is explicitly declared and described by terms from (or linked to) community or global vocabularies
- Each term is identified by a globally unique Persistent IDentifier (PID)
- Each PID resolves to a semantic representation of that term, with relationships to other terms
- Published as standards-compliant & open access

BCO-DMO

Biological & Chemical Oceanography Data Management Office



THE DATA LIFE CYCLE

(Chandler et al., EGU 2013)

context matters

Semantic Web technologies can help

Modern data infrastructure requires

- information integration
- interoperability
- conceptual modeling
- intelligent search
- data-model intercomparison
- data publishing support

Semantic Web Technologies involve

- information integration
- interoperability
- conceptual modeling
- intelligent search
- data-model intercomparison
- data publishing support

inspired by Pascal Hitzler, WSU; Krzysztof Janowicz, UCSB (2013)

Challenge

- Very limited resources (funds, time, personnel)

Solution

- Strategic partnerships to develop, share & adopt common strategies and solutions
  - ❖ Regional
  - ❖ Domain-specific
  - ❖ International
  - ❖ Cross-domain

Research Data Alliance

RDA
RESEARCH DATA ALLIANCE
Research Data Sharing without barriers

- https://rd-alliance.org/
- supported by the European Commission, the National Science Foundation and other U.S. agencies, and the Australian Government; constructing the social and technical bridges that enable open sharing of data across technologies and between disciplines and nations with the ultimate goal of addressing the grand challenges of society.

RDA Vocabulary Services Interest Group (proposed Mar 2015)

- controlled vocabulary: representation; reuse; curation; linking … http://bit.ly/1GdY23p

Vocabulary Services Interest Group

RDA
RESEARCH DATA ALLIANCE

Group details

Status: Under community review

Chair(s): Stephan Zednik and Simon Cox

Vocabulary Services Interest Group members

Adam Shepherd
Adam Leadbetter
Adrian Burton
Anirudh Prabhu
Arthur Smith
Ben Evans
Cynthia Chandler
Cynthia Hudson Vitale
Fran Lightsom
Gema Bueno-de-la-Fuente
Herbert Schentz
Jane Frazier
John Watkins
Jonathan Yu
Kathleen Fontaine
Lesley Wyborn
Louise Darroch
Mark Donoghue
Martina Stockhause
Matthew Jones
Mike Brown
Reyna Jenkyns
Robert Groman
Simon Cox
Stephan Zednik
Stephen Richard
Xiaogang Ma

RDA Marine Data Harmonization Interest Group

- https://rd-alliance.org/

  -> Working and Interest Groups

OBJECTIVE: to promote the development of a common global framework for the management of marine data

Ocean Data Interoperability Platform

- ODIP (Ocean Data Interoperability Platform) EU, Australia, USA marine data interoperability
- Develop a framework to support effective sharing of data across scientific domains and international boundaries
- http://odip.org

Belmont Forum

- Belmont Forum: http://www.bfe-inf.org/
- established in 2009, brings together environmental and geoscience funding agencies from 15 nations and seeks to build a coalition of national resources to advance global environmental change research.

ICSU
WORLD DATA SYSTEM

Future Earth

- http://www.futureearth.info/
- funding coordinated through the Belmont Forum, will be the platform through which many global change research programs will be coordinated, and the broad research themes, including the Earth Sciences, will require advanced information architectures to enable trans-disciplinary data-information-knowledge transfer

- IOC UNESCO IODE
  International Oceanographic Data and Information Exchange http://iode.org
- community and capacity building
- Ocean Data Portal, OceanDocs, OceanTeacher, OceanTeacher, OBIS, Ocean Data Standards

"A scholar's positive contribution is measured by the sum of the original data that he contributes. Hypotheses come and go but data remain."

from: Advice to a Young Investigator (Nobel Laureate Santiago Ramón y Cajal, 1897)

Thank you, cchandler@whoi.edu (@cynDC42)
http://bco-dmo.org and http://iode.org

photo by Chris Linder (WHOI)