

# Data Access Models

John Orcutt, Chair

Paul DiGiacomo

Takeshi Kawano

Milton Kampel

Benoit Pirenne

Iain Shepherd

# Topics

What are the restrictions on data access and how do these impact ocean research?

What exchange and institutional/national/regional/international data policies

What policy would best balance the interests of the researcher and society?

What is the balance between open data and intellectual property rights?

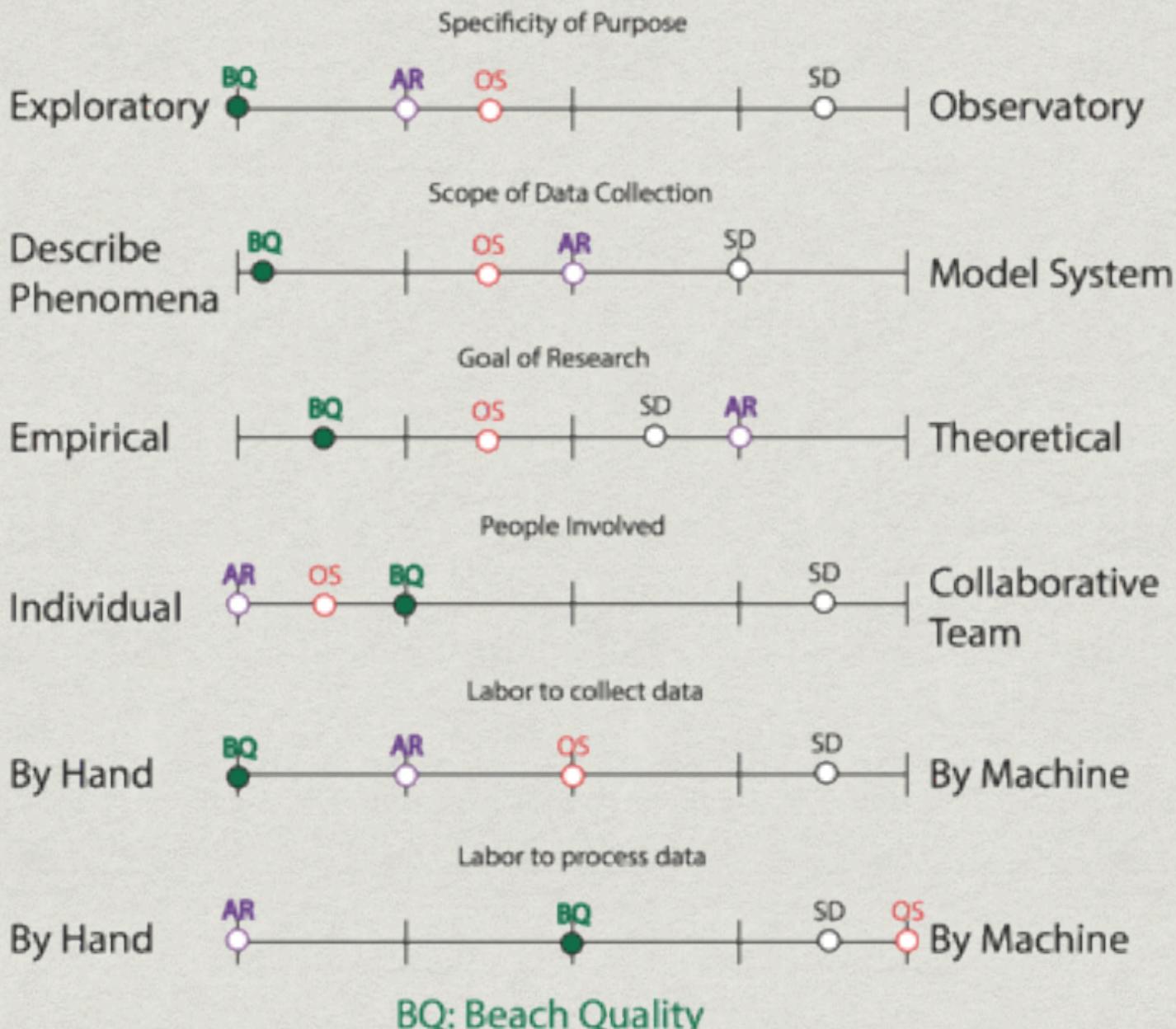
What are the roles of different organizational types in stimulating and funding ocean research?

What access models including IPR, business models for open data, etc.

# References

- rgman, Christine L. “The Conundrum of Sharing  
search Data.”
- eick, James. *The Information: A History; A Theory;*  
od.
- elsen, Michael. *Reinventing Discovery: The New En-  
trenched Science.*
- frey, John, and Urs Gasser, Interop: The Promise  
rils of Highly Interconnected Systems.

# Approaches to data collection



# Reasons for sharing data

Reproduce or verify research

Make results of publicly funded research available to the public

Enable others to ask new questions of extant data and

Advance and accelerate the state of research and innovation.

# ClimateGate

UK Joint Information Systems Committee (JISC), and the Programme Manager Hodson noted (Climate Research Unit Controversy. 2011)

University of East Anglia email publicity used to state that scientists were rigging the data to push the hypothesis of climate change

Climate science is by no means unique in the need for researchers to analyse complex data from a number of different sources. The aim of this investment is to improve the way research data is managed in UK universities. By showing how it can be made more open, these projects will help achieve proper recognition for the essential place of data creation

# Galaxy Zoo

Sloan Digital Sky Survey (SDSS)

> 200,000 volunteers

Classify galaxy images

Robotic telescope

New form of galaxy in addition to typical spiral and elliptical galaxies (*Green Pea Galaxy*)

# Reinventing Discovery: The New Era of Networked Science

Michael Nielsen

I wrote this book because I believe the reinvention of discovery is one of the great changes of our time. To futurists looking back a hundred years from now, there will be two eras of science: pre-network science, and networked science. We are living in a time of transition, the second era of science. But it's going to be a bumpy transition, and there is a possibility it will fail or fall short of its potential.

# Nielsen and the OOI

In September of 2009 an organization called the Ocean Observatories Initiative began laying a high-speed network for data and electricity on the floor of the Pacific Ocean. They're extending the internet to the ocean floor, with the eventual plan being to lay 1,000 kilometers (750 miles) of cable, from the shores of Oregon all the way up to British Columbia. This underwater internet will range more than 100 kilometers (60 miles) offshore. When it's complete, all manner of devices will be plugged into the network, from cameras to robot vehicles to genome-sequencing equipment. In fact, a volcano erupting underwater, and nearby genome-sequencing equipment switch on to take genetic snapshots of never-before-seen microbes vented during the eruption. Imagine a network of thermometers and other sensors mapping out the underwater environment, in much the same way the SDSS is mapping out the universe. But the Ocean Observatories Initiative is going even further than the SDSS, making their data completely downloadable right from the start, so anyone in the world can immediately download it and start looking for new patterns and asking new questions.

# Human Genome Project

## Bermuda Agreement (1996)

Under pressure from attendees, the funding agencies required that all scientists working on the human genome make the data quickly and openly available.

Genome data from other life forms continue to be held privately in spite of the great potential value to society. There are exceptions.

# Human Genome Project

The rationale for protecting data from external view  
comes largely from the academic rewards systems in  
which scientists are largely judged by their papers as  
well as the number and quality of subsequent  
citations to the work. Were the data open, their  
competitors could gain an unfair advantage. The  
ultimate academic goal is writing a scientific paper,  
which will ultimately lead to salary, grants and prizes

# Open Geospatial Consortium (OGC)

Not-for-profit corporation

478 members (as of last week)

- \* ESA, NASA, Microsoft, NOAA, ESRI, Google, U
- \* No data persistence, but open services to integ geospatial data

EU's created INSPIRE to encourage broad adopt of OGC within the community

# OGC Business Model

## Annual Dues

|                                   |          |
|-----------------------------------|----------|
| S/EU University                   | \$500    |
| S/EU Technical Committee Member   | \$11,000 |
| S/EU Principal Member             | \$55,000 |
| S/EW small company (<\$2,000,000) | \$2,200  |
| Individual University             | \$165    |

# NSF Example

Investigators are expected to share with other researchers, at no more than the incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants.

# NSF Ocean Sciences

PIs are required to submit, at no more than incremental cost and within a reasonable time frame (but no later than two(2) years after the data are collected), the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF/OCIO grants to the appropriate Data Center.

# National Data Centers

National Ocean Data Center (NODC)

National Climatic Data Center (NCDC)

National Geophysical Data Center (NGDC)

# NSF Data Management

Often data submission policies haven't been enforced

Proposals must now include no more than two pages labeled "Data Management Plan."

Should describe how the proposal will conform to the NSF policy on the dissemination and sharing of research results.

# Interoperability

INSTITUTIONAL

HUMAN

DATA

TECHNOLOGICAL

# 4 Layers

Technology - networks, switches and routers

Data - Receiving party must be able to understand the data received. Enclosures, for example.

Human - Common language (English). Need strong commitment to working together. Trust.

Institutional - Legal aspects, IPR, Commonality of laws in different countries.

|              | IOOS   | OOI   | Neptune & Venus | NDBC   |
|--------------|--------|-------|-----------------|--------|
| Geographical | Red    | Red   | Green           | Red    |
| Temporal     | Red    | Red   | Yellow          | Yellow |
| Data Type    | Yellow | Green | Green           | Green  |
| Biological   | Yellow | Green | Green           | Green  |

Interoperability matrix status for four observatories (US & Canada). The color code is associated with the following status:

# Licensing

Bayh-Dole Act in US (1980)

Deals with IPR arising from govt-funded research

University & inventor owns the IPR.

Govt retains a non-exclusive, non-transferable, irrevocable, paid-up license.

Follow-the-money idea - subcontracts

# BSD License

Berkeley Standard Distribution of Unix - eg OS X

Open software

Minimum restrictions on how data/software can be used & distributed

Copyright (c)<YEAR>,<OWNER>

All rights reserved.

# Hold Harmless

This software is provided by the copyright holder and contributors “AS IS” and any express or implied warranties...are disclaimed. In no event shall the copyright owner...be liable for any direct, indirect, incidental, special, exemplary, or consequential damages....

# GEOSS

Summary white paper, 2011

Legal options for the exchange of data through the  
GEOSS Data-CORE.

Does not include a copyright statement nor a *Hold  
Harmless* clause

# GEOSS Principles

**2), 3) AND 4) ARE PERMITTED  
BUT NOT REQUIRED**

The data are free of restrictions on reuse

User registration or login to access or use the data is permitted

Attribution of the data provider is permitted as a condition of use

Marginal cost recovery charges are permitted

# Creative Commons

The most modern of licenses

CC provides a web-based tool to choose among six licenses.

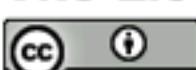
There are license catalogs for many countries, including members of the EU that recognize the national constraints as well as the EU requirements.

# Creative Commons

Three layers:

- \* Machine readable
- \* Human readable
- \* Legal code

# The Licenses



## Attribution

**CC BY**

This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.

[View License Deed](#) | [View Legal Code](#)



## Attribution-ShareAlike

**CC BY-SA**

This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. This license is often compared to "copyleft" free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects.

[View License Deed](#) | [View Legal Code](#)



## Attribution-NoDerivs

**CC BY-ND**

This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you.

[View License Deed](#) | [View Legal Code](#)



## Attribution-NonCommercial

**CC BY-NC**

This license lets others remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms.

[View License Deed](#) | [View Legal Code](#)



## Attribution-NonCommercial-ShareAlike

**CC BY-NC-SA**

This license lets others remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms.

[View License Deed](#) | [View Legal Code](#)



## Attribution-NonCommercial-NoDerivs

**CC BY-NC-ND**

This license is the most restrictive of our six main licenses, only allowing others to download your works and share them with others as long as they credit you, but they can't change them in any way or use them commercially.

[View License Deed](#) | [View Legal Code](#)

Scripps Wikipedia Readability UCSD Blink Home Google Clip to DEVONthink Read Later Google Calendar Apple Google Maps Online Pages

Google... Wireless... Kimpto... Kimpto... Wireless... Kimpto... Kimpto... Kimpto... Kimpto... Choose...

## License Features

Your choices on this panel will update the other panels on this page.

Allow modifications of your work?

Yes  No

Yes, as long as others share alike

Allow commercial uses of your work?

Yes  No

License Jurisdiction:

International

Selected License  
Attribution 3.0 Unported

This is a Free Culture License!

## Help others attribute you!

This part is optional, but filling it out will add machine-readable metadata to the suggested HTML!

Title of work

Attribute work to name

Attribute work to URL

Source work URL

More permissions URL

## Have a web page?

This work is licensed under a [Creative Commons Attribution 3.0 Unported License](#).

Copy this code to let your visitors know!

```
<a rel="license" href="http://creativecommons.org/licenses/by/3.0/deed.en_US"><img alt="Creative Commons License" style="border-
```