

Ocean Data Management: Opportunities and Challenges

Karen Stocks

Director, Geological Data Center
Scripps Institution of Oceanography

Current Portfolio

Operational Data and Information Systems



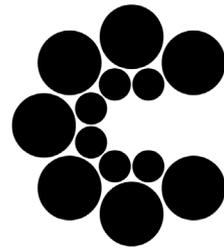
R2R

ROLLING DECK TO REPOSITORY



IODP
INTERNATIONAL OCEAN
DISCOVERY PROGRAM

Science Support Office



CCHDO

Technology/CI Development

DeCODER

Community



FARR

FAIR, AI Readiness & Reproducibility

iDOOS

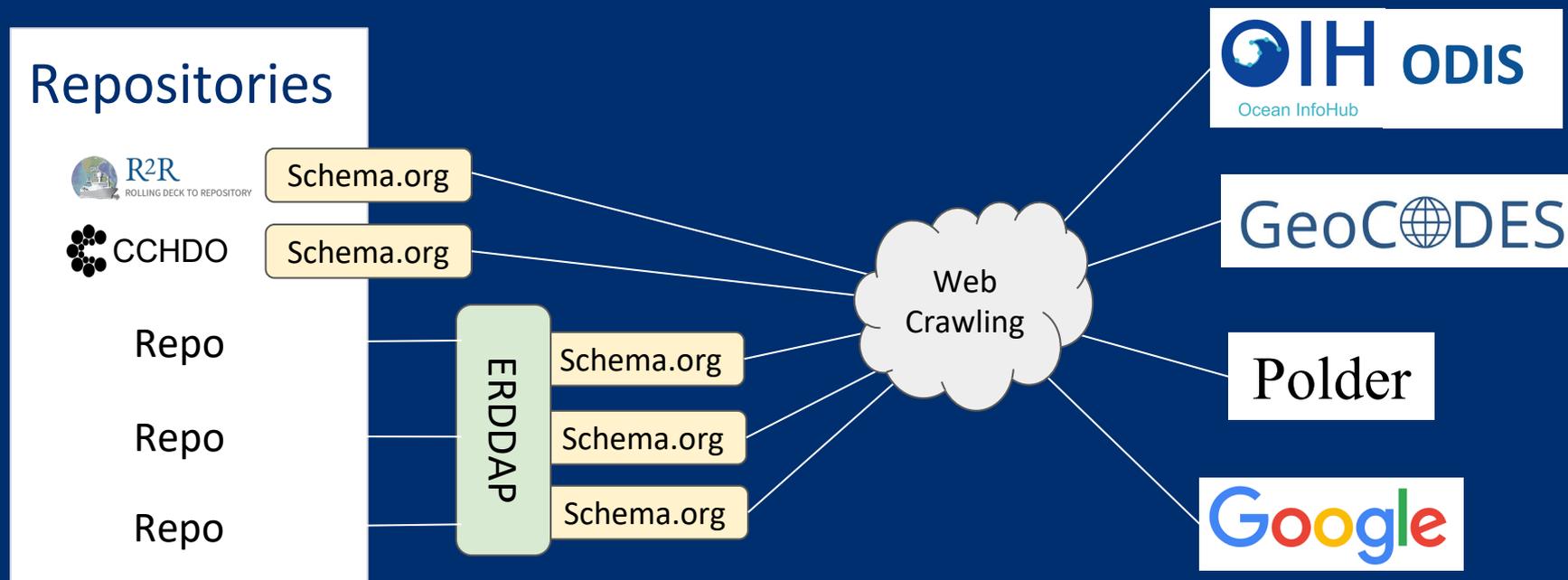
*EarthCube

Investments in data
repositories, standardization,
and FAIR data are enabling
advanced cyberinfrastructure



Schema.org (w/ERDDAP) for Discovery

Implementation of simple schema.org metadata markup on dataset landing pages can be crawled for cross-repository searching...by any search tool.



Cloud Resources and Computational Notebooks

 [Blog](#) [Forum](#) [Site](#) [Page](#) [About Pangeo](#) »

PANGEO

A community platform for Big Data geoscience

 contributors **82**  discourse **1.3k users**  Discord rate limited by upstream service [follow @pangeo_data](#)

This website contains general information about the Pangeo project. For news and updates about Pangeo, check out our [Medium blog](#) and our [Twitter feed](#). To engage with the Pangeo community, head over to our [Discourse forum](#) or browse our [GitHub repos](#).

TWITTER FEED

We use twitter to share announcements and retweet content of interest to our community.

[Tweets by pangeo_data](#)

DISCOURSE FORUM

Discourse is a message board where we discuss all aspects of the project in a friendly, inclusive way.

 Conflict-free Replicated Zarr
6 days ago

[pangeo.io](#)

PHYSICAL OCEANOGRAPHY

 [pangeo-gallery/physical-oceanography](#)

license **MIT**

 last commit **october 2021**

 Binderbot **no status**

 launch binder

DESCRIPTION

Examples of how to use Xarray, Dask, and Holoviews to load, process, and visualize cloud-based ocean data. Datasets include satellite observations (gridded sea-surface height) and ocean model output (POP, MOM6).

NOTEBOOKS

- [Sea Surface Altimetry Data Analysis](#)
- [Along Track Altimetry Analysis](#)
- [MOM6/CESM Ocean Model Analysis](#)
- [MITgcm ECCOV4 Example](#)
- [Gulf Stream Currents](#)

[Contributor Guide](#)

[Gallery for CESM LENS on AWS](#)

[NCAR Gallery](#)

[Pangeo & Dask Gateway.](#)

[NASA CCMP Winds Gallery](#)

[Example Gallery](#)

[Physical Oceanography](#)

[Sea Surface Altimetry Data Analysis](#)

[Along Track Altimetry Analysis](#)

[MOM6/CESM Ocean Model Analysis](#)

[MITgcm ECCOV4 Example](#)

[Gulf Stream Currents](#)

[ESIP Gallery](#)

[Gallery for XGCM Examples and Usecases](#)

Highly standardized,
machine-readable data and
metadata are foundational



Increasing Data Facility Expectations

Persistent
Identifiers

Cybersecurity

Controlled
Vocabularies

Certification

AI/ML

Provenance

APIs

Use License

Machine-
Readable

Cloud

Schema.org

Challenge: Heterogeneous & Under Described Data

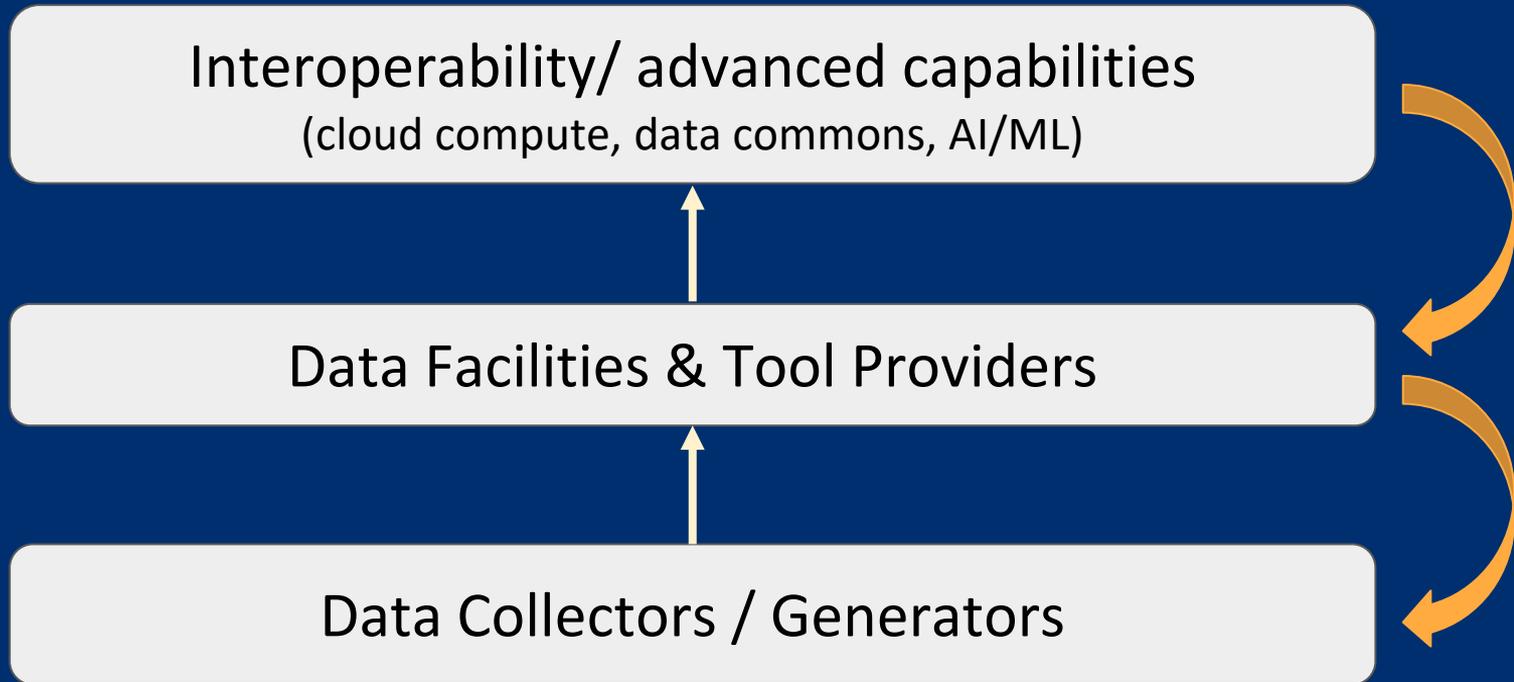
R2R has written 32 parsers for GNSS (GPS) navigation data format variants



FAIR has to start at
data collection



Investments needed at all levels



Potential Actions

Keep funding advanced CI development...with embedded geoscientists

Run a gap analysis: who is getting left behind?

Which domains/data types do not have repositories?

How FAIR are ocean data repositories?

What is needed?



Require and enforce Data Management Plans that ensure standardized, well described data and metadata

Build coding literacy into science education