University of California
CDL
California Digital Library

An Overview of Climate-Related Data Issues

Within and Beyond the UC System

Steve Diggs: UCOP/CDL/UC3

PICES 2024 / W6

Honolulu, HI USA

2024-10-27

The University of California

Overview of the UC System

- 10 campuses
- 3 National Laboratories
- 5 Medical Centers

Population

- 294,309 total students
- 25,400 faculty
- 173,300 staff

Among the top 10 University systems in the world

University of California
CDL
California Digital Library

Climate Research Across the UC System

Climate Research Budget: $180 million

- Climate Grants: $80 million (2023)
- Climate Resilience Research: $100 million

Key Research Locations

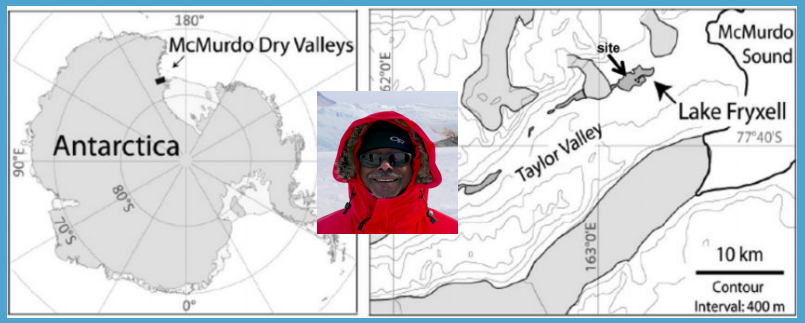
- UC Berkeley: Energy and Resources Group
- **UC San Diego: Scripps Institution of Oceanography**
- UC Davis: John Muir Institute of the Environment
- UCLA: Institute of the Environment and Sustainability
- UC Irvine: Center for Environmental Biology
- **UC Santa Barbara: Bren School of Environmental Science & Management**

Research Activities

- Climate Science and Impacts Research
- Sustainable Agricultural Practices
- Renewable Energy Innovations
- Adaptation and Solutions Development
- Policy Analysis, Advocacy, Climate Justice, and Equity Studies

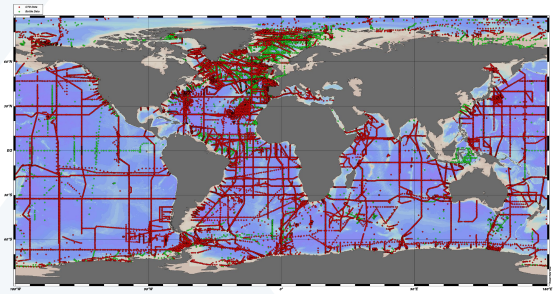
University of California CDL California Digital Library

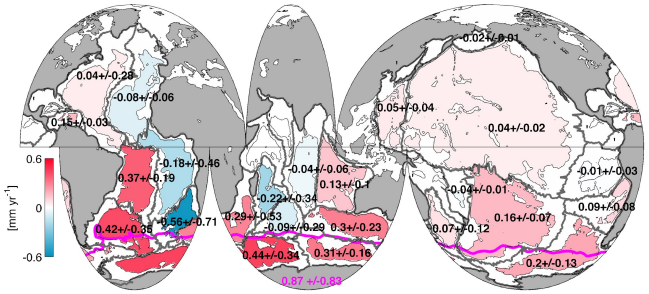
My career throughout the years



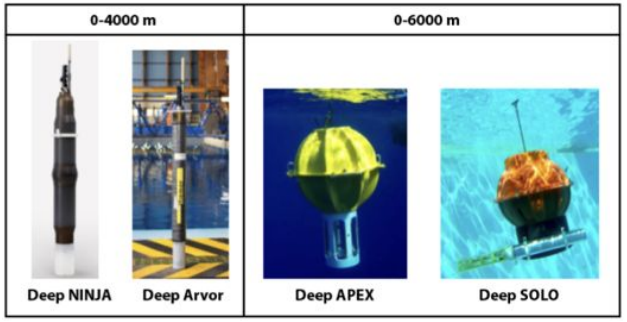
Technical Director / GO-SHIP Data Repository)

The **Global Ocean Ship-based Hydrographic Investigations Program (GO-SHIP)** coordinates a network of measurements that provide approximately decadal resolution of the changes in inventories of heat, freshwater, carbon, oxygen, nutrients and transient tracers, covering the ocean basins from coast to coast and full depth (top to bottom), with global measurements of the highest required accuracy to detect these changes.





(Purkey and Johnson 2010)



GO-SHIP CTD data are inputs to the OWC algorithms that are used to estimate the time-varying correction of conductivity measurements from **Argo floats.**

CDL was explicitly created as a central unit within the UC system with a mission to coordinate, leverage, and amplify the experience, expertise, and capabilities of its partners.

CDL - California Digital Library

CDL provides transformative digital library services, grounded in campus partnerships and extended through external collaborations, that amplify the impact of the libraries, scholarship, and resources of the University of California.

Four active programs:

- Publishing & Special Collections
- Discovery & Delivery
- Collections & Licensing
- **UC Curation Center (UC3)**

UC3 - UC Curation Center

Digital curation:
Maintaining, preserving, and adding value to digital research data throughout its lifecycle

Areas of focus for UC3:

- Research data management
- Data publication
- Data metrics
- Persistent identifiers
- Digital preservation
- Data/software skills training for librarians

UC3
UC Curation Center

CDL's Role in Data Management

Responsibilities and Limitations

California Digital Library (CDL) Responsibilities:

- CDL supports research data infrastructure through collaborations with campus libraries and partnerships on global research data projects.

Limitations of CDL's Role:

- CDL does not engage directly with individual researchers at the campus level; instead, it provides tools, standards, and infrastructure that campus libraries can implement locally.

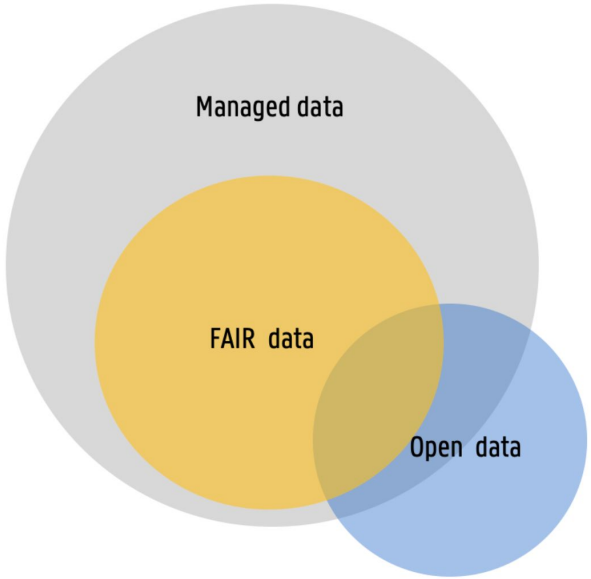
Collaborative Projects and Strategic Focus:

- CDL collaborates on initiatives addressing global research data challenges, focusing on scalable solutions to improve data curation and accessibility.
- Examples of current projects includes efforts to standardize data practices and improve data sharing frameworks in alignment with global trends in climate and ocean data management.

… and what about you?

A working definition of Open Science

- Open Science is a movement to make scientific research, data, and dissemination accessible to all levels of society At its core, Open Science aims to increase transparency, collaboration, and efficiency in the scientific process.
- A key component of Open Science is ***open research data***, which refers to the data underpinning scientific research results being made freely available with no restrictions on access. This allows anyone to analyze, verify, and reuse the data.



University of California
CDL
California Digital Library

slido

What job title best describes your position regarding collaborative efforts within PICES? (you may choose more than one)

ⓘ Click **Present with Slido** or install our Chrome extension to activate this poll while presenting.

Marine Biologist / Marine Ecologist

Physical Oceanographer / Climate Scientist

Fisheries Scientist / Fisheries Manager

Environmental Policy Analyst / Policy Officer

Statistical Officer / Data Scientist

Research Director / Chief Scientist

Government Minister / Deputy Minister

slido

How familiar are you with open science and open data?

ⓘ Click **Present with Slido** or install our Chrome extension to activate this poll while presenting.

Extremely familiar

Very familiar

Quite familiar

Moderately familiar

Slightly familiar

Not at all familiar

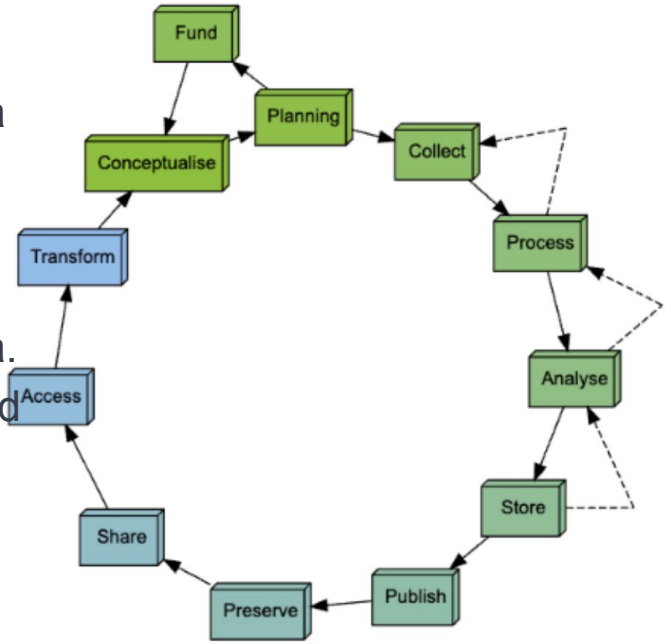
PICES: Reaching Consensus

Agreeing on terms, and solving Data Problems

The Data Lifecycle in Ocean and Climate Research

From Question to Data Use:

- Overview of the structured process that translates research questions into actionable insights.
- **Lifecycle Phases:**
  - **Question Formulation:** Define objectives and data requirements.
  - **Data Collection:** Gathering raw data through fieldwork, sensors, or sampling.
  - **Data Processing:** Applying QA/QC to validate data.
  - **Data Analysis and Modeling:** Deriving insights and creating summarized datasets.
  - **Data Archiving:** Storing data for future access and replication.
  - **Data Use and Visualization:** Translating data into actionable models and visual products.



The MaLDReTH
Research Data Lifecycle

Terminology at Each Phase: Establishing clear terms for each phase enhances collaboration and transparency.

University of California
CDL
California Digital Library

What is Data? Types and Categories in Ocean and Climate Research

Defining Data in Context:

- Data encompasses various types—quantitative (e.g., temperature readings), qualitative (e.g., observational notes), and metadata (data about data).

Types of Data:

- **Raw Data:** Collected directly from sources (e.g., sensor measurements, satellite images).
- **Processed Data:** Data cleaned and validated through quality control measures.
- **Derived Data:** Summarized data used to build models or answer specific research questions.

Examples include environmental samples, observational data, and modeled outputs.

Data Collection, Processing, and Quality Assurance

Raw Data Collection:

- Methods in ocean and climate science include sensor readings, water sampling, and satellite observations.
- Importance of standardizing collection methods to ensure compatibility and quality.

Data Processing and QA/QC:

- Quality control protocols to create reliable, processed datasets that are ready for analysis.

Terminology in Collection and QA/QC:

- Standardized language for sampling, processing, and validation enhances collaboration and reproducibility.

Analysis, Modeling, Archiving, and Visualization

Data Analysis and Modeling:

- Statistical and computational techniques to interpret processed data.
- Derived data products support decision-making and climate predictions.

Data Serving and Archiving:

- Importance of long-term storage for future access and replication.
- Use of consistent terminology in archiving for clear documentation.

Data Visualization and Use:

- Transforming data into charts, models, and maps to communicate insights to diverse audiences.

Unified Terminology for Analysis and Visualization:

- Standard terms help make data products accessible across disciplines.

Future Directions and Collaborative Next Steps

Establishing a Shared Framework and Terminology:

- Encourage cross-institutional efforts to create consistent, usable terms for data practices.

Supporting Long-Term Investment in Data Champions:

- Incentivize and empower individuals who promote data management best practices.

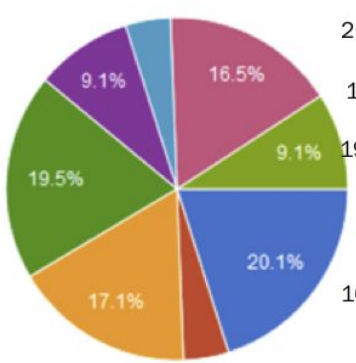
Envisioning the Future:

- A shared framework with unified terminology can create a scalable, sustainable foundation for global climate and ocean data resilience.

Scientific Data: Hard to Find, Hard to Use

Researcher Challenges with Data Use

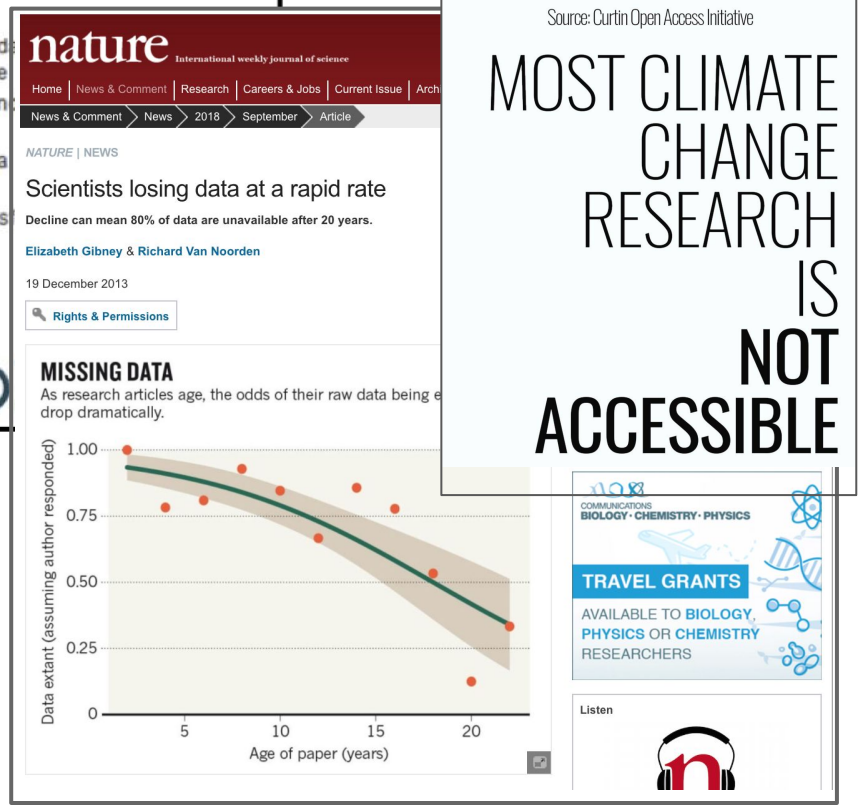
The top four issues accounted for 73% of respondents



20.1% Data complexity
Data volumes
17.1% Finding relevant existing d
- knowing what's out there
19.5% Lack of data standards an
exchange standards
Dealing with multiple data
types
Data access and file trans
16.5% Data management and
storage
Other

Data Management Skills Gap Analysis, April 7, 2017
http://bfe-inf.org/document/skills-gap-analysis

BELMONT FORUM



University of California
CDL
California Digital Library

CAN OTHERS DO WITH PUBLISHED AND OPEN EARTH DATA?

@scdiggs
steve.diggs@ucop.edu

Photo credit: John H. Moore