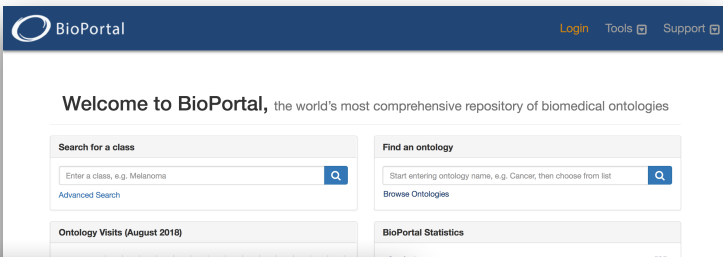


Building a Principled Interoperable Data System for Air Sensors

John Graybeal, Stanford University

Stanford Center for
Biomedical Informatics Research

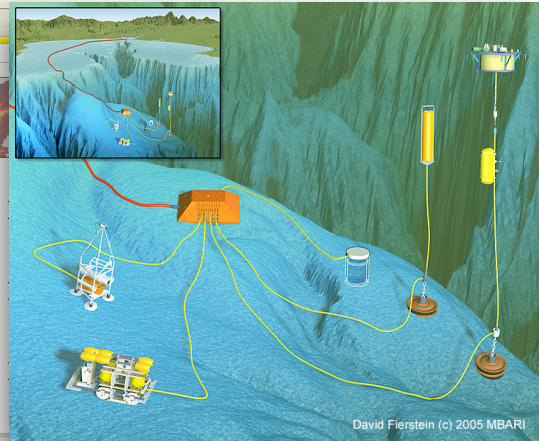
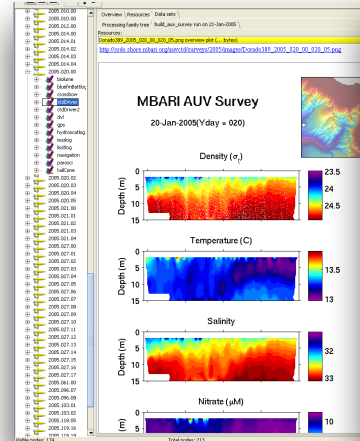
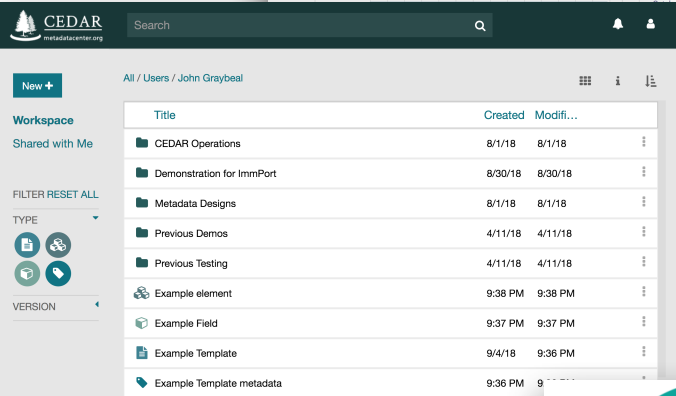
14 September 2018



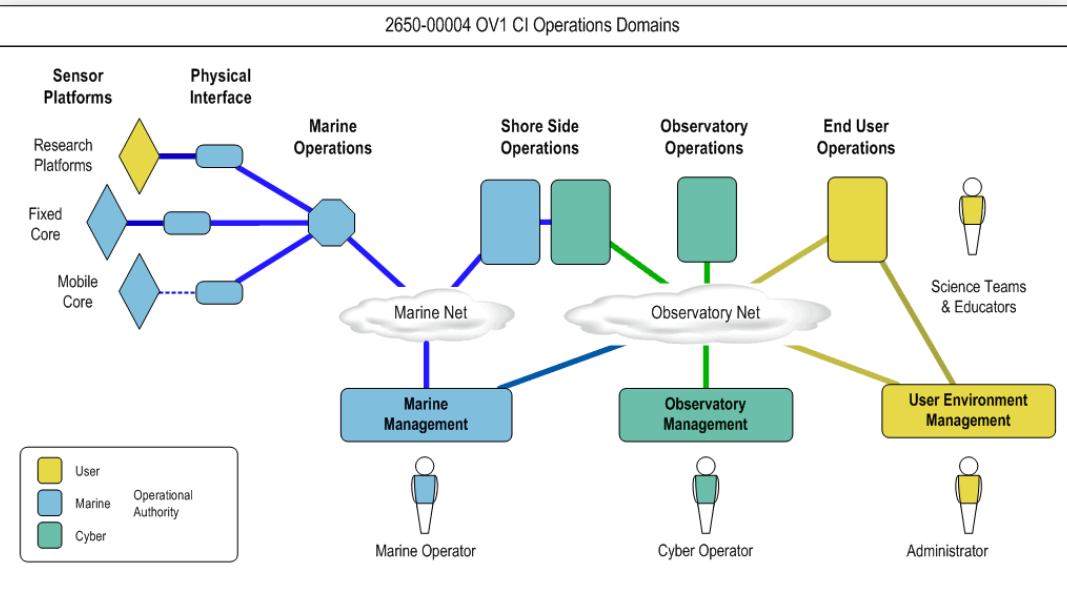
Stratospheric Observatory for Infrared Astronomy



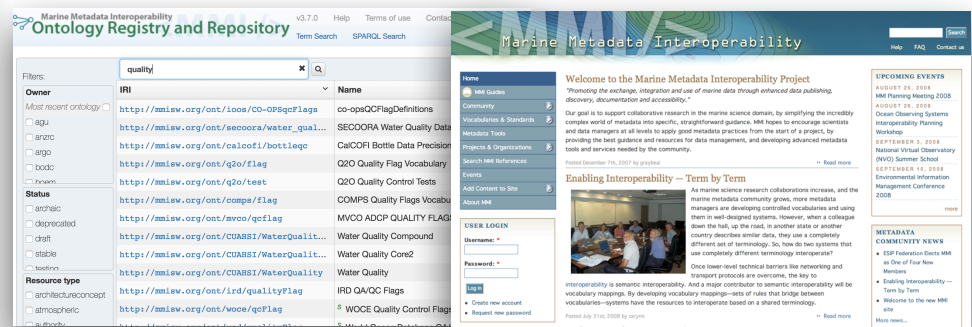
CEDAR



Ocean Observatories Initiative: CyberInfrastructure

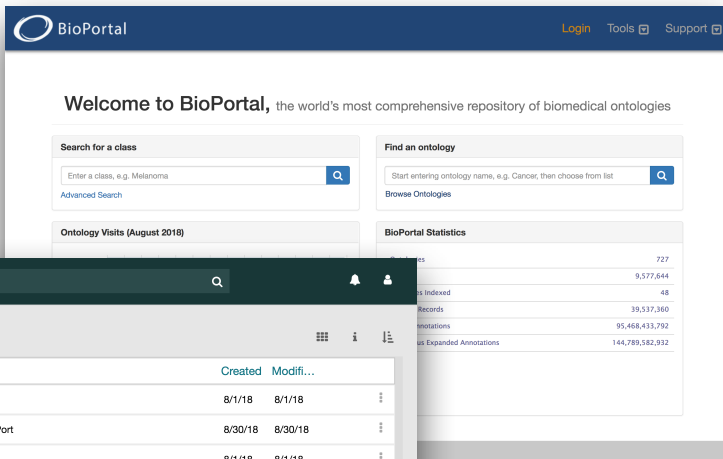


Shore Side Data System for the Monterey Ocean Observing System

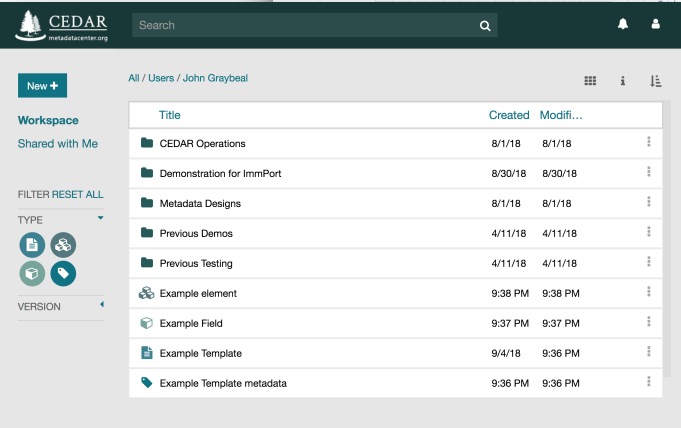


Marine Metadata Interoperability Ontology Repository Project & Community

BioPortal

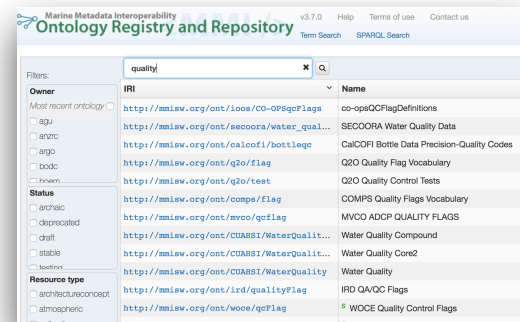


CEDAR



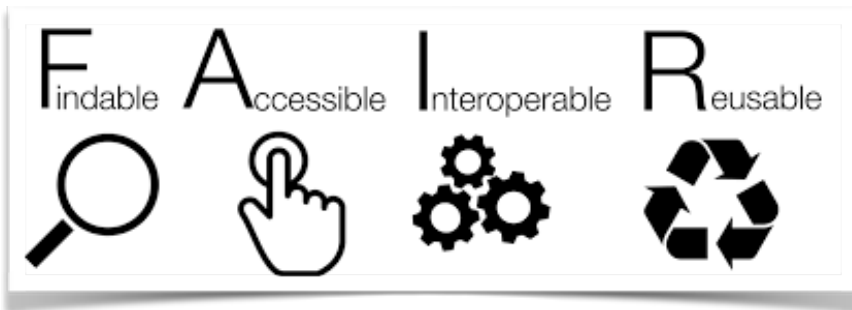
Building a Principled Interoperable Data System

Focusing on Metadata and Semantics



Marine Metadata Interoperability Ontology Repository

Let's talk about FAIR



I C N E
N C T U
D E E S
A S R A
B S O B
L I P L
E B E E
L R
E A
B
L
E

- To get FAIR data:
 - Maximize interoperability =
 - Maximize reuse =
 - Maximize value of data

FAIR data == Interoperable data

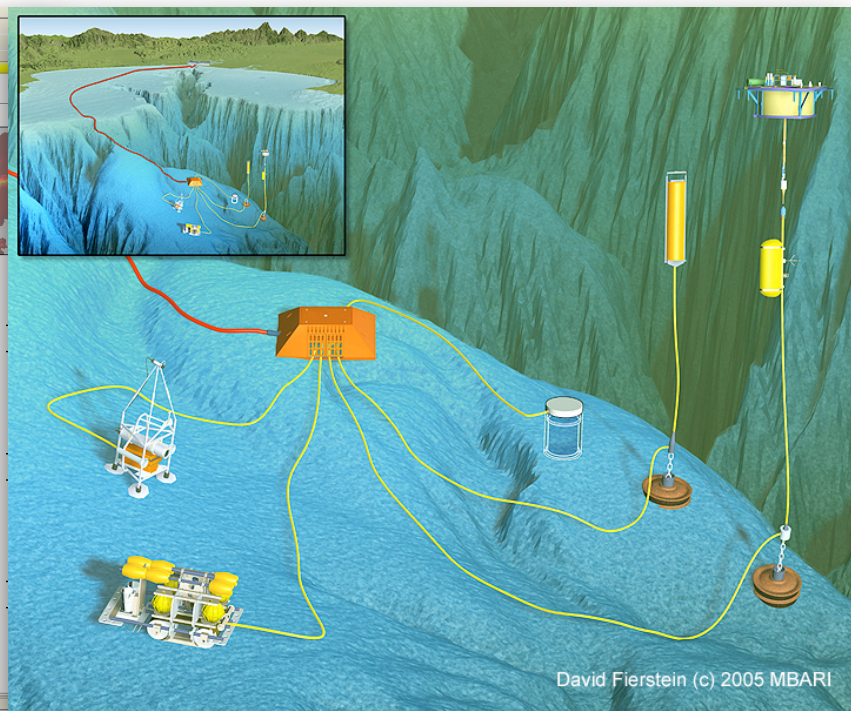
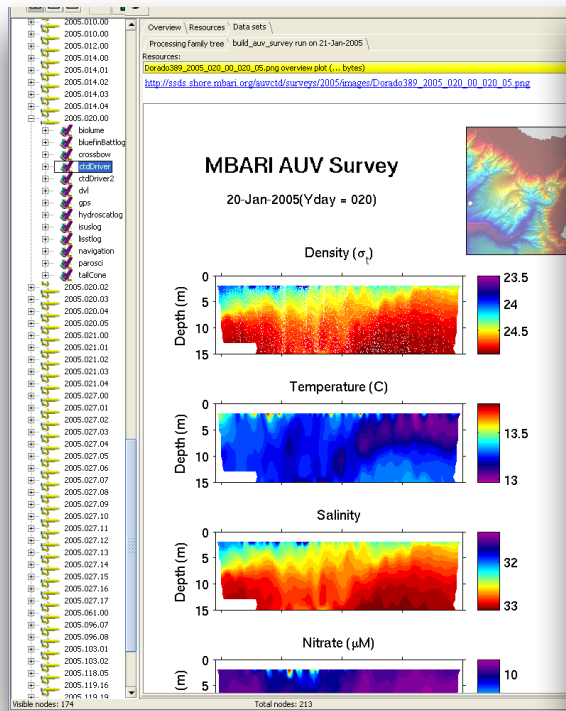
To maximize data value

- We take the many different variables from many different data sources, and systematically
 - harmonize them structurally (syntactically) and semantically,
 - describe them, and
 - publish them,
- so that they can be found and processed systematically.

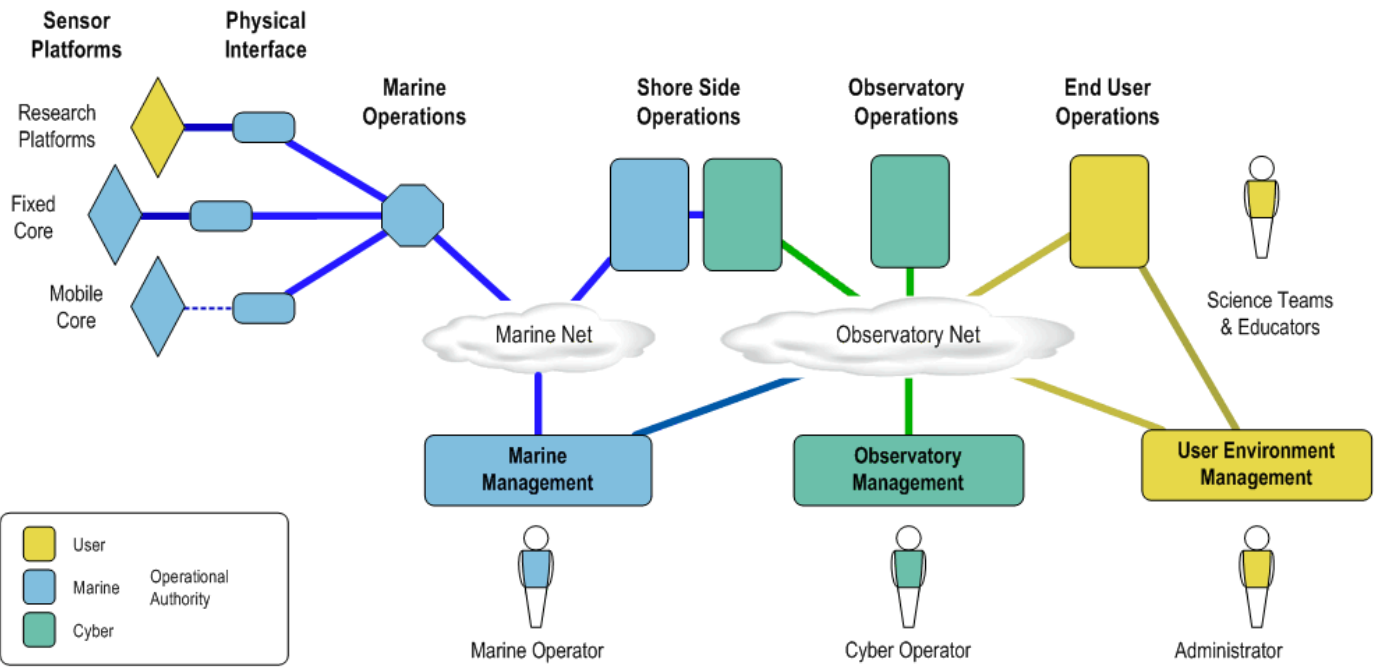
How to do this?

Shore Side Data System

At Scale

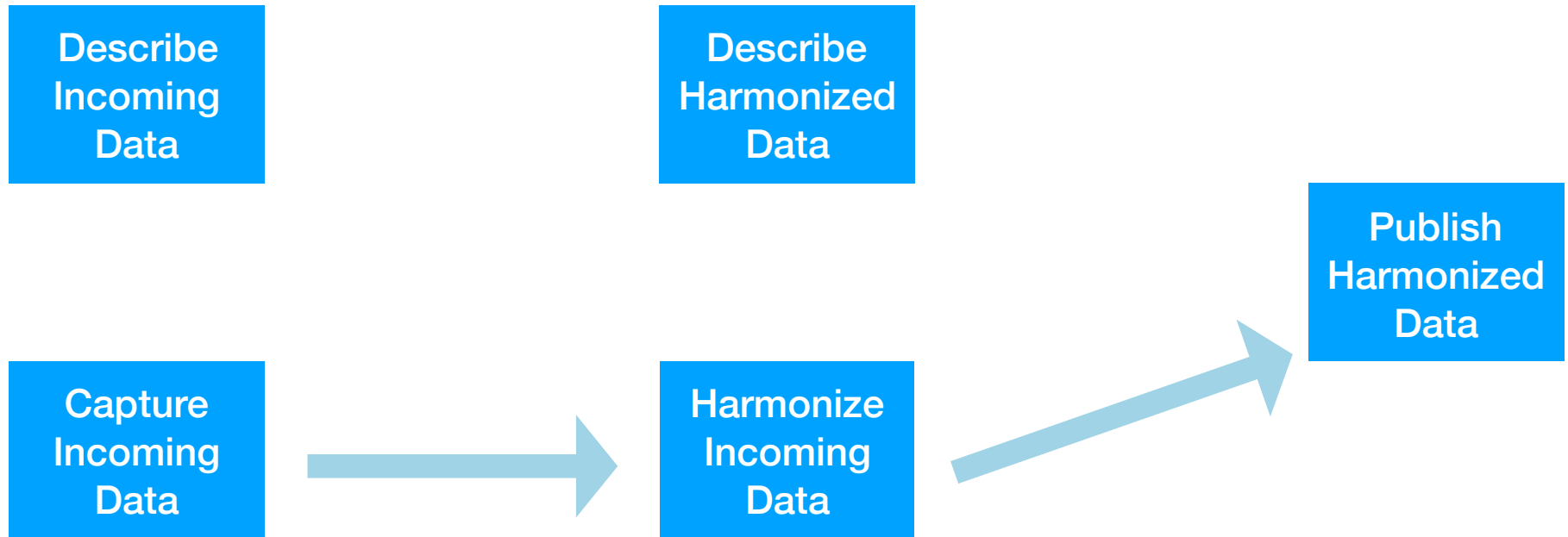


2650-00004 OV1 CI Operations Domains

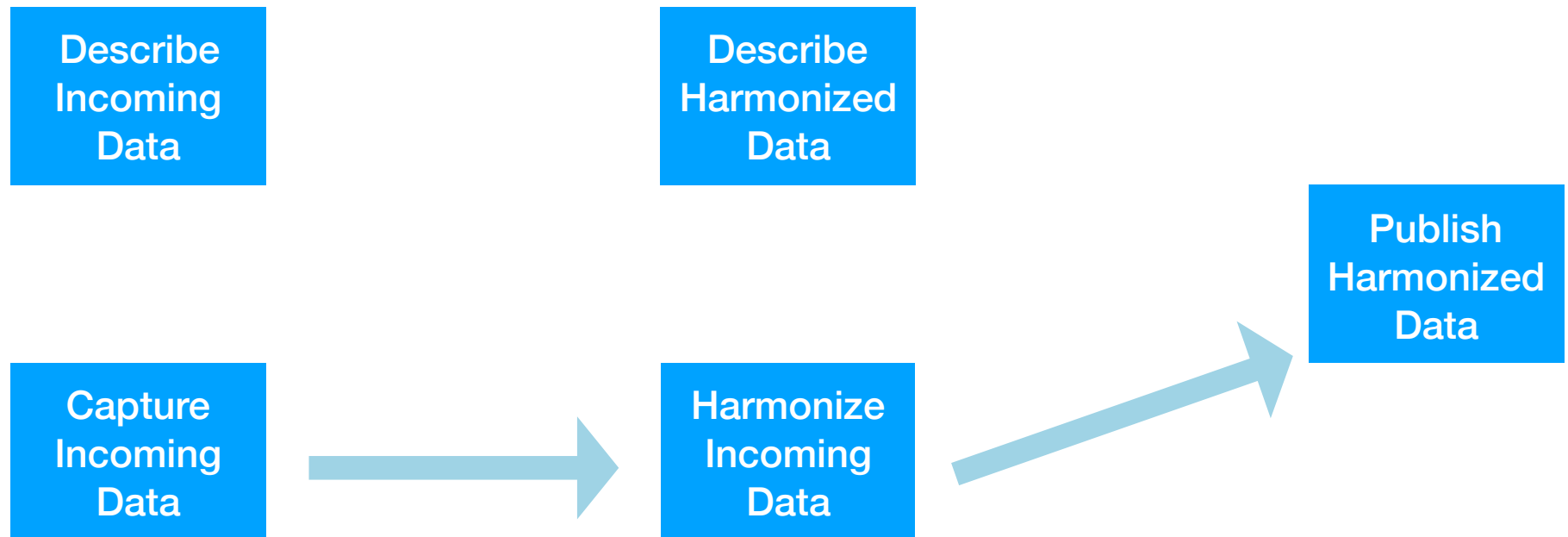


OOI Cyberinfrastructure

Data Flow Perspective

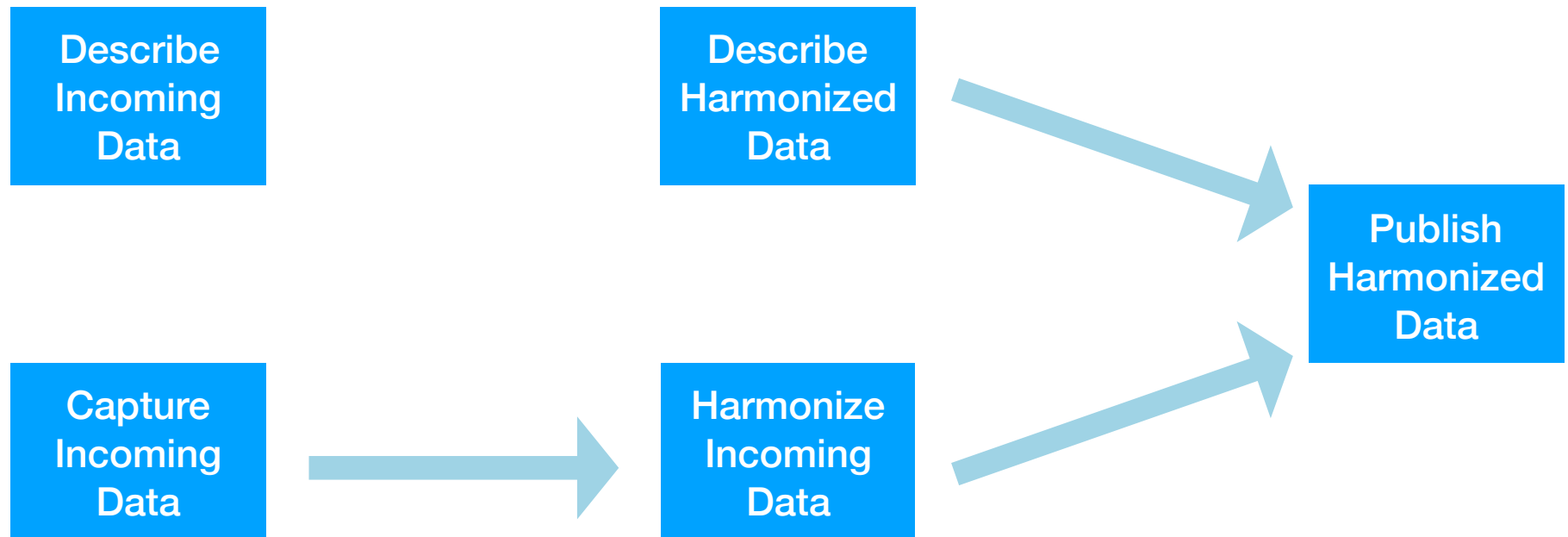


Data Flow Perspective



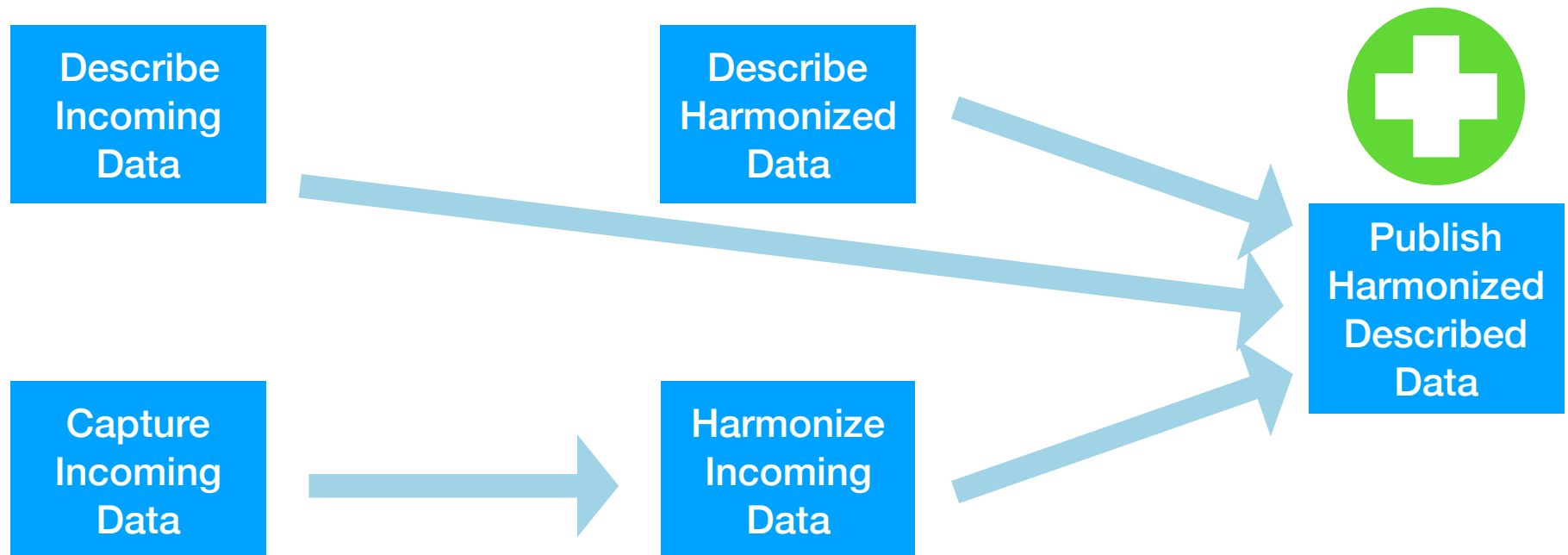
Not FAIR

Data Flow Perspective



FAIRer

Data Flow Perspective



FAIRest

Design Perspective



Describe data and transforms first

Introduction to the Semantic Vision

“Semantics”: Defined

Why not just manage your terms in lists and spreadsheets?

In A Nutshell

Need	Solution	Motivation
Unique Identifier	IRI (International Resource ID)	we want software to be very precise about what it references, anywhere in the internet
Descriptions	Triples: Subject-Predicate-Object	we want to make statements about things, and relate things to each other
Collections	Ontologies	we want to organize the things we say (like in web pages, term lists, or in models)
Agreed Rules	OWL, RDF, SKOS, ...	we want a set of well-defined basic relationships, so artifacts adhering to them can interoperate

Unique Identifiers: The IRI

- **why: software needs to be very precise about what it references, anywhere in the internet**
- *what:* IRI (Internationalized Resource Identifier): international version of URI, Uniform Resource Identifier; an identification string established by the W3C* that can be used on the web to uniquely identify a resource (may or may not be 'resolvable')



Recommend resolving it!

* W3C = World Wide Web Consortium

So what?

- ***What we have:*** identifiers, and ways to describe them and link them following standardized rules
- ***Why that is powerful:*** We have a consistent way for systems and computers to document concepts and work with them.
 - vocabularies and terms
 - taxonomies and relations
 - complex models of systems and domains
 - analytical (reasoning) systems that derive conclusions

There is Help

Guidelines

Tools

Standards

There is Help

Guidelines

Tools

Standards

Marine Metadata Interoperability
Ontology Registry and Repository

v3.7.0 Help Terms of use Contact us

Term Search SPARQL Search

quality

Filters:

Owner

Most recent ontology

agu

anzrc

argo

bodc

hoem

Status

archaic

deprecated

draft

stable

testing

Resource type

architectureconcept

atmospheric

authority

IRI	Name
http://mmisw.org/ont/ioos/CO-OPSqcFlags	co-opsQCFlagDefinitions
http://mmisw.org/ont/secoora/water_qual...	SECOORA Water Quality Data
http://mmisw.org/ont/calcofi/bottleqc	CalCOFI Bottle Data Precision-Quality Codes
http://mmisw.org/ont/q2o/flag	Q2O Quality Flag Vocabulary
http://mmisw.org/ont/q2o/test	Q2O Quality Control Tests
http://mmisw.org/ont/comps/flag	COMPS Quality Flags Vocabulary
http://mmisw.org/ont/mvco/qcflag	MVCO ADCP QUALITY FLAGS
http://mmisw.org/ont/CUAHSI/WaterQualit...	Water Quality Compound
http://mmisw.org/ont/CUAHSI/WaterQualit...	Water Quality Core2
http://mmisw.org/ont/CUAHSI/WaterQuality	Water Quality
http://mmisw.org/ont/ird/qualityFlag	IRD QA/QC Flags
http://mmisw.org/ont/woce/qcFlag	WOCE Quality Control Flags

There is Help

Guidelines

Tools

Standards



Login Tools Support

Welcome to BioPortal, the world's most comprehensive repository of biomedical ontologies

Search for a class

Enter a class, e.g. Melanoma



[Advanced Search](#)

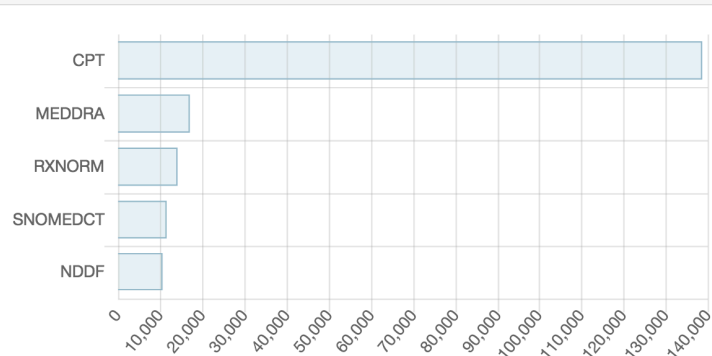
Find an ontology

Start entering ontology name, e.g. Cancer, then choose from list



[Browse Ontologies](#)

Ontology Visits (August 2018)



[More](#)

BioPortal Statistics

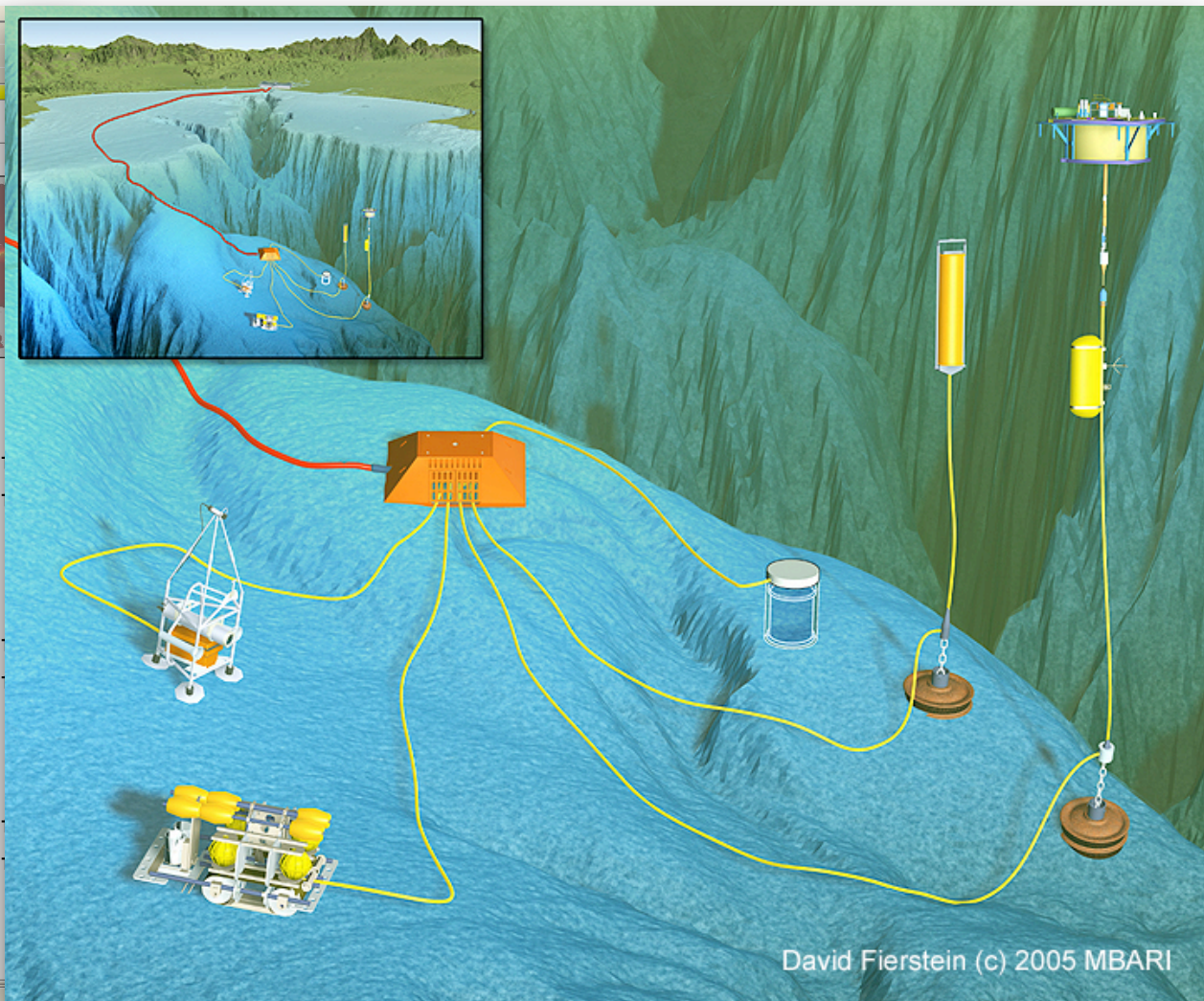
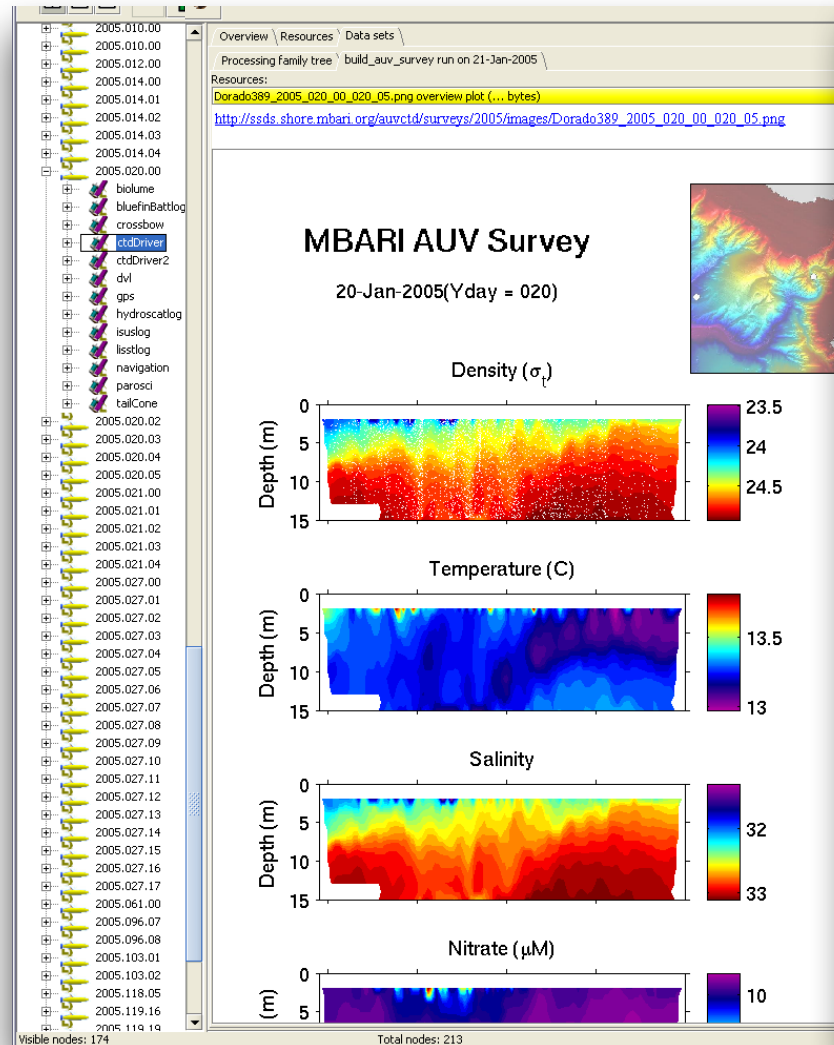
Ontologies	727
Classes	9,577,644
Resources Indexed	48
Indexed Records	39,537,360
Direct Annotations	95,468,433,792
Direct Plus Expanded Annotations	144,789,582,932

Introduction to the Metadata Vision

“You need to describe your stuff.”

–Author claim, without offering citations.

Describing Stuff is Hard



But There Is Help

Guidelines

Tools

Standards

But There Is Help

Guidelines

Tools

Standards

← Drone Minimum Information

▼ Drone Minimum Information

▼ Project Information

Project Name*

Project Funder

Research Question

Project Investigator

▼ Drone Platform Information

Platform Type*

Manufacturer

Model Name

Model Number

Sensor

▼ Drone Sensor Information

Sensor Type*

Sensor Model

Firmware Version

CANCEL SAVE

But There Is Help

Guidelines

Tools

Standards

The screenshot displays the X-DOMES SensorML web application. At the top left is the X-DOMES logo with the tagline "Cross-Domain Observational Metadata for Earth". To the right of the logo is the text "SensorML" and a "GitHub" link. Below the logo, there is a section for "SensorML XML/RNG Input" with a radio button set to "From List". This section includes two input fields: "Local ID: Enter XML ID" and "Unique ID: Enter URI". Below these fields is an "Add section +" button. A modal window titled "Add section" is open in the center, listing various metadata sections with a green plus sign next to each: Textual Description, Display Name, Keywords, Deployment Identifiers, Identification, Classification, Physical Characteristics, Characteristics, Capabilities, Contacts, History, Process Definition URI, Type Of (Parent Process), Configuration, Features of Interest, Inputs, Outputs, and Parameters. To the right of the modal window, there are "Save.." and "Open.." buttons. At the bottom of the modal window is a "Save" button.

Let's Get (a bit) Real

ASCII

Binary

Raw Data 1

Raw Data 2

Raw Data n

ASCII

XSV
(CSV / TSV / ...)

Binary

NetCDF / HDF
(CF Conventions)

Raw Data 1



Structured
Data 1

Raw Data 2



Structured
Data 2

Raw Data n



Structured
Data n

ASCII

XSV
(CSV / TSV / ...)

Controlled Terms
(RDF / SKOS / OWL)

Binary

NetCDF / HDF
(CF Conventions)

CF
(+Controlled Terms)

Raw Data 1



Structured
Data 1



Harmonized
Data 1

Raw Data 2



Structured
Data 2



Harmonized
Data 2

Raw Data n



Structured
Data n



Harmonized
Data n

ASCII

XSV
(CSV / TSV / ...)

Controlled Terms
(RDF / SKOS / OWL)

Whole Description
(XML / JSON-LD)

Binary

NetCDF / HDF
(CF Conventions)

CF
(+Controlled Terms)

NetCDF + CF

Raw Data 1

Structured
Data 1

Harmonized
Data 1

Described
Data 1

Raw Data 2

Structured
Data 2

Harmonized
Data 2

Described
Data 2

Raw Data n

Structured
Data n

Harmonized
Data n

Described
Data n



ASCII

XSV
(CSV / TSV / ...)

Controlled Terms
(RDF / SKOS / OWL)

Whole Description
(XML / JSON-LD)

Binary

NetCDF / HDF
(CF Conventions)

CF
(+Controlled Terms)

NetCDF + CF

Raw Data 1

Structured
Data 1

Harmonized
Data 1

Described
Data 1

Raw Data 2

Structured
Data 2

Harmonized
Data 2

Described
Data 2

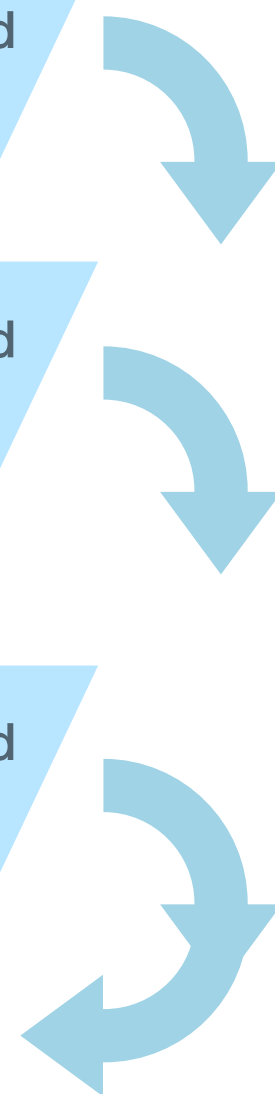
Raw Data n

Structured
Data n

Harmonized
Data n

Described
Data n

Group of Described Data Sets



Publish to Google's Data Set Index

- Create first-class web pages for each data set
- Include schema.org markup within the page
 - see ESIP (esipfed.org) for current practices and code
- Google will eventually index your pages in its Data Set Search (Beta!)

References

- CEDAR: cedar.metadatacenter.org
- BioPortal: bioportal.bioontology.org
- MMI ORR: mmisw.org/ont (MMI: coming back soon to marinemetadata.org!)
- ESIP COR: cor.esipfed.org/ont
- XDOMES: xdomes.org
- Searches:
 - science metadata guidelines, science metadata standards
 - semantic guidelines, semantic standards, globally unique identifiers
 - NetCDF, NetCDF CF, CF vocabulary