



# Zooplankton Meeting: Data and Information Infrastructure

Planning forward for data organization  
in support of zooplankton knowledge production  
April 2009

**Karen S. Baker**

**Scripps Institution of Oceanography, UCSD**  
Information Manager - LTER and CalCOFI  
Ocean Informatics CoDirector  
UCSD Science Studies Program Affiliate



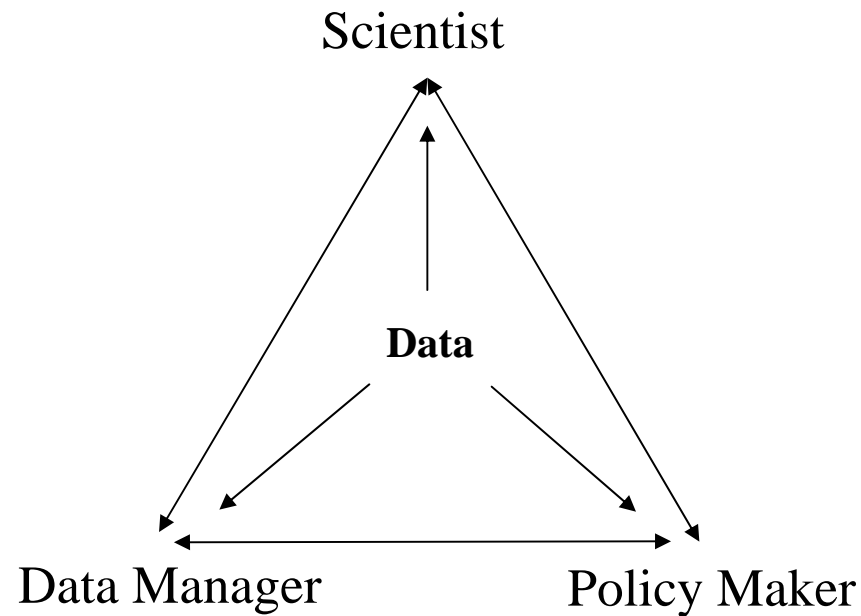
# Biological Data Are Complex

unruly

abiotic - biotic

well-structured - irregular

# Roles Associated with Data Work



part of all the jobs:  
expectation management

# Some Goals

## Management

- a) Comprehensive ways of making stock assessment
- b) Facilitation of collaboration
- c) Get data online

## Science

- a) Understanding the ocean ecosystem that determines stock assessment
- b) New types of research made possible by new data practices
- c) Response to expectations for data sharing

## Information Management

- a) Understanding the information ecosystem that supports both  
data analysis (science) and data curation (management)
- b) Useful systems and services that support scientific research and stock assessment
- c) Developing processes that enable knowledge-creation at all levels within multi-level networks

## **What we have in common:**

An interest in zooplankton data use

..... traditional science and network science

## **What is needed at the meeting:**

-A discussion of existing zooplankton data - their similarities and differences to inform the development and use of working standards for describing and structuring data collections for data re-use

## **Why this is needed:**

- Development of vocabularies, authoritative lists, and standards make data interpretable for use
- Provides an infrastructure for knowledge production by both
  - \* zooplankton researchers
  - \* other scientific researchers, and the public

# Data Sharing and Information Infrastructure Issues

## Guides for sharing

1. data policy
2. ethical guidelines

## Data holdings: what to share

2. data sources and data repositories
3. data levels and products

## Infrastructure elements: how to share

4. working standards: codes, vocabularies, and biases
5. network-ready datasets

# Data Sharing

it is wise to ask:

-what constitutes data?

-where are the data?

-how do you share data?

*data repository*

*research-based*

*resource-based*

*analysis lifecycle*

*curation lifecycle*

*network-science*

*information infrastructure*

*federation*

*continuing design*

*web-of-repositories*



*data hoarding*

*data production*

*scientific responsibility*

*knowledge production*

# Data Sharing

it is wise to ask:

-what constitutes data?

-where are the data?

-how do you share data?

data repository

network-science

research-based

information infrastructure

resource-based

federation

analysis lifecycle

continuing design

web-of-repositories

curation lifecycle



data hoarding ↓

data production ↓

↑ scientific responsibility

↑ knowledge production

# Data Sharing

it is wise to ask:

-what constitutes data?

-where are the data?

-how do you share data?

data repository

network-science

research-based

information infrastructure

resource-based

federation

analysis lifecycle

continuing design

web-of-repositories

curation lifecycle



data hoarding

data production

scientific responsibility

knowledge production



# Science Studies

Social Informatics - Sociotechnical Configurations - Continuing Design

Comparative Interoperability Project:  
Interoperability Strategies for  
Scientific Cyberinfrastructure  
2002-ongoing  
<http://interoperability.ucsd.edu>



Geoffrey C. Bowker  
Science, Technology & Society  
Santa Clara University



Karen S. Baker  
Ocean Informatics & Science Studies  
SIO, UCSD



Helena Karasti  
Information Systems Design,  
Oulu University, Finland



David Ribes  
Sociology/Science Studies,  
Georgetown University



Florence Millerand  
Communications & HCI,  
Univ of Quebec, Montreal



Steve Jackson  
School of Information  
University of Michigan



Naomi Oreskes  
History of Science  
UCSD, Science Studies



Elena Aronova  
History of Science  
UCSD, Science Studies



Cyberinfrastructure  
geo-sciences

*Community Ontologies*



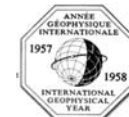
Network-Science  
ecological sciences

*DM & Metadata*



Ocean Informatics  
ocean sciences

*Learning Environments*



IGY Big Science  
geo-sciences

*World Data Centers*

# DataZoo: an information system for well-structured data

The screenshot shows the DataZoo website interface. At the top left, there are links for "Home" and "Log In". The main header features the "OI ocean informatics datazoo" logo, where "OI" is a large blue graphic of a wave and "I" is a large blue letter, followed by the text "ocean informatics datazoo". Below the header, there are three main sections: "Data", "Resources", and "Management". Each section has a blue button with an icon: a hard drive for "Data", a stack of papers for "Resources", and a wrench for "Management". Below each button is a light gray box with a description of the section's functionality. At the bottom of the page, there are logos for "SCRIPPS INSTITUTION OF OCEANOGRAPHY" and "UCSD".

Home | Log In

**OI** ocean informatics datazoo

**Data**  
  
Download, view, plot, and filter data.

**Resources**  
  
Resources including documentation, tools, and tutorials.

**Management**  
  
Data and metadata management interface.

<http://oceaninformatics.ucsd.edu/datazoo>



### Data Use Policy

The data available here are intended for scholarly use by the academic research community, with the express understanding that data users will properly acknowledge the originating investigator. Use or reproduction of any material herein for any commercial or redistribution purposes is prohibited without prior written permission from the responsible party.

By agreeing to this Data Use Policy, you are also agreeing to the respective data acknowledgment policies of each project.

### Local Users:

Log in with your IOD username and password for access to private datasets and data and metadata management interfaces.

Username:

Password:

### Public Users:

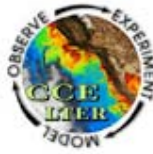
Enter your name and email address to gain access to public data. Your email address will not be distributed to any other sources.

Name:

E-Mail:

Use:





# California Current Ecosystem LTER



- Datasets
- Studies
- Prepared Views
- Integrate Data

Change Project

**cclter**

[Project Web Site](#)

- Data Use Policy
- Data Acknowledgement Policy



## - Datasets -

**Title / description** 
**Date begin**  (YYYY-MM-DD HH:MM:SS)

**Current contact** 
**Date end**  (YYYY-MM-DD HH:MM:SS)

**Duration**  (years)

Cruise
  Survey Cruise
  Process Cruise
  Season
  Timeseries
  Reference

chemical
  biological
  physical
  sampling
  societal
  signature

▼ keywords

Search

35 total

ID	Dataset Title	Type	Project	Current Contact	Reco
142	<a href="#">Bathymetry</a> 2-minute bathymetry for latitudes 15N to 45N, longitudes 105W to 135W, covering the current and historic CalCOFI and CCE sampling region	Reference	Public	National Geophysical Data Center	55318
17	<a href="#">California Current Ecosystem Data Summary - Annual Averages</a>	Cruise	CCELTER	Ralf Goericke	22



# CalCOFI - SIO

California Cooperative Oceanic Fisheries Investigations



Data



Resources



Management

Datasets

Studies

Prepared Views

Integrate Data

Change Project

calcofi-sio

[Project Web Site](#)

[Data Use Policy](#)

[Data Acknowledgement Policy](#)

?

## - Datasets -

Title / description

Date begin

(YYYY-MM-DD HH:MM:SS)

Current contact

Sydeman

Date end

(YYYY-MM-DD HH:MM:SS)

Duration

(years)



Cruise



Survey Cruise



Process Cruise



Season



Timeseries



Reference



chemical



biological



physical



sampling



societal



signature

▼ keywords


Search

5 total


ID	Dataset Title	Type	Project	Current Contact	Records
118	<a href="#">Bird and Mammal Census - Continuous Plankton Recorder surveys</a> Bird and mammal counts for continuous plankton recorder cruises on the great circle route from Vancouver, BC to Tokyo	Cruise	CalCOFI-SIO	William Sydeman	22983
117	<a href="#">Bird and Mammal Census - National Marine Fisheries</a> Bird and mammal counts for National Marine Fisheries cruises as part of the Rockfish Recruitment Survey	Cruise	CalCOFI-SIO	William Sydeman	16389
114	<a href="#">Bird and Mammal Census - ...</a>	Cruise	CalCOFI-SIO	William Sydeman	10255

# IchthyoDB: an information system for fish egg and larvae counts

[local login](#)



## IchthyoDB - Login



### User Login

**Users:** Please review the Data Use and Data Acknowledgement policies on this page. Fill out the information below to access the IchthyoDB dataset. Your contact information will not be distributed to any other sources.

Name:

E-Mail:

Use:

I have read and agree to the Data Use and Data Acknowledgement policies

### Data Use Policy

The data available here are intended for scholarly use by the research community, with the express understanding that (1) all data users will properly acknowledge the originating investigator efforts, as indicated below, and (2) reprints of reports, articles, or other publications resulting from use of these data will be sent to Dr. Sam McClatchie, Southwest Fisheries Science Center (SWFSC), 8604 La Jolla Shores Dr., La Jolla, CA 92037-1508, U.S.A. Use or reproduction of any materials herein for any commercial purpose is prohibited without prior written permission from SWFSC. There are many caveats with these data, and we strongly advise developing a collaborative project with SWFSC staff (first contacts are Dr. Sam McClatchie (Sam.McClatchie@noaa.gov), Bill Watson (William.Watson@noaa.gov), and FRD Director Dr. Russ Vetter (Russ.Vetter@noaa.gov)).

### Data Acknowledgement Policy

The following citation should be included in any use of the data or figures from this database: "Data originate from the NOAA National Marine Fisheries Service, Southwest Fisheries Science Center, collected by NOAA CalCOFI survey teams. Ichthyoplankton data were identified by the SWFSC ichthyoplankton laboratory". The managing investigators should be contacted to determine whether co-authorship is warranted for published materials arising from use of these data.

[Search Interface](#) • [About this Project](#) • [Help and Documentation](#) • [References](#)

<http://oceaninformatics.ucsd.edu/ichthyoplankton>



# IchthyoDB - Egg and Larvae Counts

currently connected to the public database (logout)



## Tow Type Filters

- Oblique
  Vertical
  Surface
  Depth Stratified  
 Restrict to tows from CalCOFI cruises

## Time Period Filters

**Year**  to   invert  
**Month**  to   invert  
**Time**  to   invert

## Species Filters

Search All Fields

### Search Results

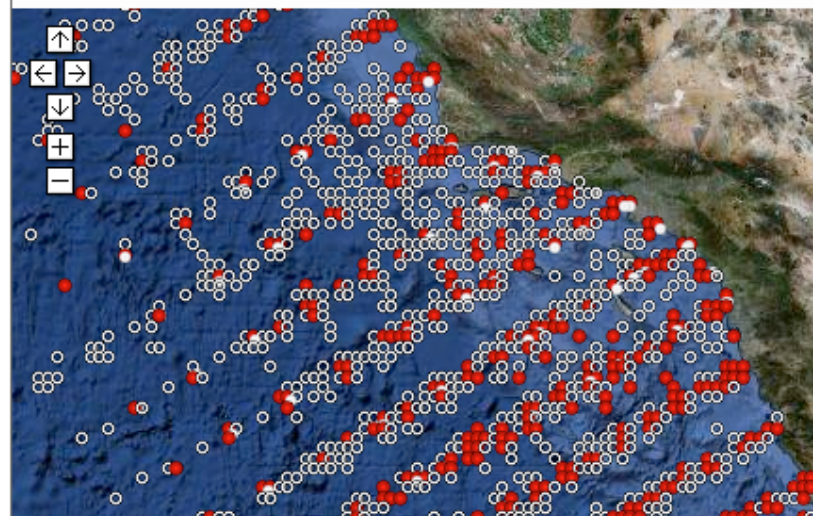
<input type="checkbox"/>	Code	Scientific Name	Common Name	Type	Occurrences
<input checked="" type="checkbox"/>	19	<i>Sardinops sagax</i>	Pacific sardine (pilchard)	Larvae	3911
<input checked="" type="checkbox"/>	19	<i>Sardinops sagax</i>	Pacific sardine (pilchard)	Eggs	2683
<input type="checkbox"/>	31	<i>Engraulis mordax</i>	Northern anchovy	Eggs	8154
<input type="checkbox"/>	31	<i>Engraulis mordax</i>	Northern anchovy	Larvae	17073
<input type="checkbox"/>	93	<i>Vinciguerra lucetia</i>	Panama lightfish	Larvae	12067
<input type="checkbox"/>	239	<i>Diogenichthys atlanticus</i>	Longfin lanternfish	Larvae	4174

## Geographic Filters

By Latitude/Longitude   
 By Line/Station

## Species Distribution

Limit to counts >=   Show zero tows   
 Show CalCOFI stations



# ZooDB: an information space for zooplankton collections

## Cooperative Zooplankton Dataspace

Mark Ohman Lab



Scripps Institution of Oceanography



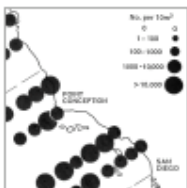
### Pelagic Invertebrates Collection

The Scripps Pelagic Invertebrates Collection is among the world's preeminent collections of marine zooplankton. It includes over 120,000 whole zooplankton samples containing some  $10^8$  specimens. In addition to worldwide geographic coverage, the Collection includes the remarkable CalCOFI zooplankton time series, which has surveyed the California Current since 1949.



### ZooDB - Zooplankton Database

The ZooDB zooplankton data module provides a queryable interface to extensive zooplankton data from the California Current System, 1951 through present. The plankton samples analyzed are mostly from [CalCOFI](#), a long-term sampling program that continues actively today. Most of the data are from nighttime samples from springtime cruises in two distinct geographical areas off California: Southern California and Central California.



10 m<sup>2</sup>.

### Brinton Euphausiid Plot Gallery

The Brinton euphausiid plot gallery presents geographical distributions of selected species, illustrated by Dr. Edward Brinton as a companion to Brinton and Townsend (2003). The selected plots are from 1969, the decade of the 1970s, and subsequent El Niño-associated intervals. Broad abundance intervals are expressed as numbers per



### BTEDB - Brinton and Townsend Euphausiid Database

The Brinton and Townsend euphausiid data module provides a queryable interface to extensive euphausiid data, 1951 through present, from the California Current System. The plankton samples analyzed are from [CalCOFI](#) (California Cooperative Oceanic Fisheries Investigations), a long-term ocean sampling program that continues actively today. Most of the data are from two distinct geographical areas off California: Southern California and Central California. There are also data from off Baja California prior to 1986.



### MEDB - Marinovic Euphausiid Database

The Marinovic euphausiid data module provides a queryable interface to extensive euphausiid data, 2001 through 2007, from the central sector of the California Current System. The plankton samples analyzed are from tows on CalCOFI lines 60 and 66.7, and sampling continues actively today. Most of the data are from summer (June or July) and fall (October or November) cruises off Central California.



### CEQuI - Comprehensive Euphausiid Query Interface

The comprehensive euphausiid data module provides an interface that queries across two research programs: the [Brinton and Townsend Euphausiid database](#) and the [Marinovic Euphausiid Database](#). A joint display of the three species in common is provided in table and graph form.



California Current Ecosystem LTER



NATIONAL OCEANIC AND  
ATMOSPHERIC ADMINISTRATION

<http://oceaninformatics.ucsd.edu/zooplankton>

**Moving from  
'Gimme your data'**

**to**

**'Let's re-envision data sharing  
and talk about the realities of data handling.'**

...all the while noticing and respecting the data differences

Focusing on data organization, coordination, and federation  
rather than data accumulation, hierarchies, & technologies

# Data Repository Types

## Similarities and Differences

### The 3 R's for Network Science

- Research-based

Examples: CalCOFI

- Resource-centered

Example: COPEPOD

- Reference-archive

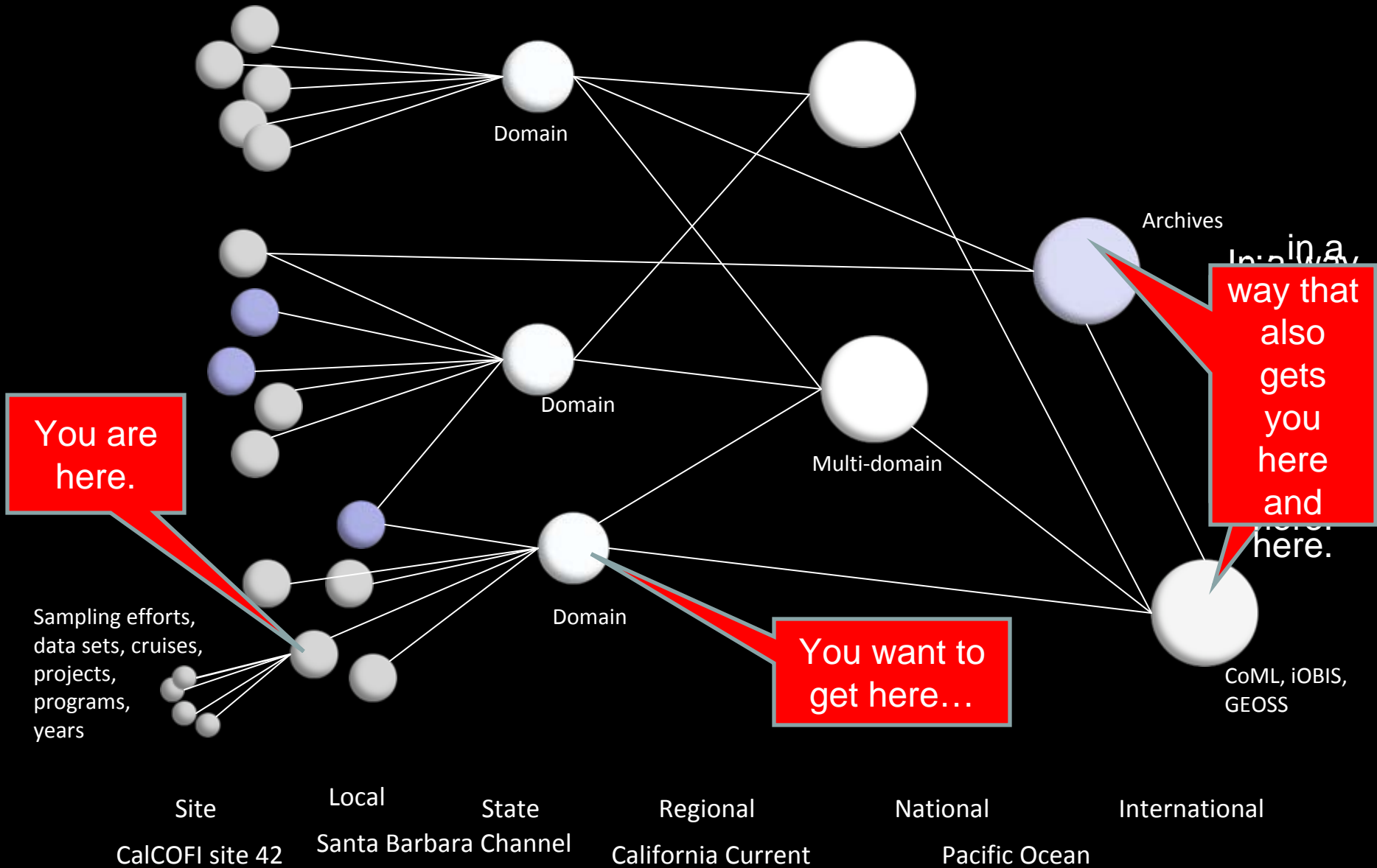
Example: NODC

And how they interface...

\*NSB, 2005. National Science Board: Long-Lived Data Collections: Enabling Research and Education in the 21<sup>st</sup> Century. NSF NSB-05-40

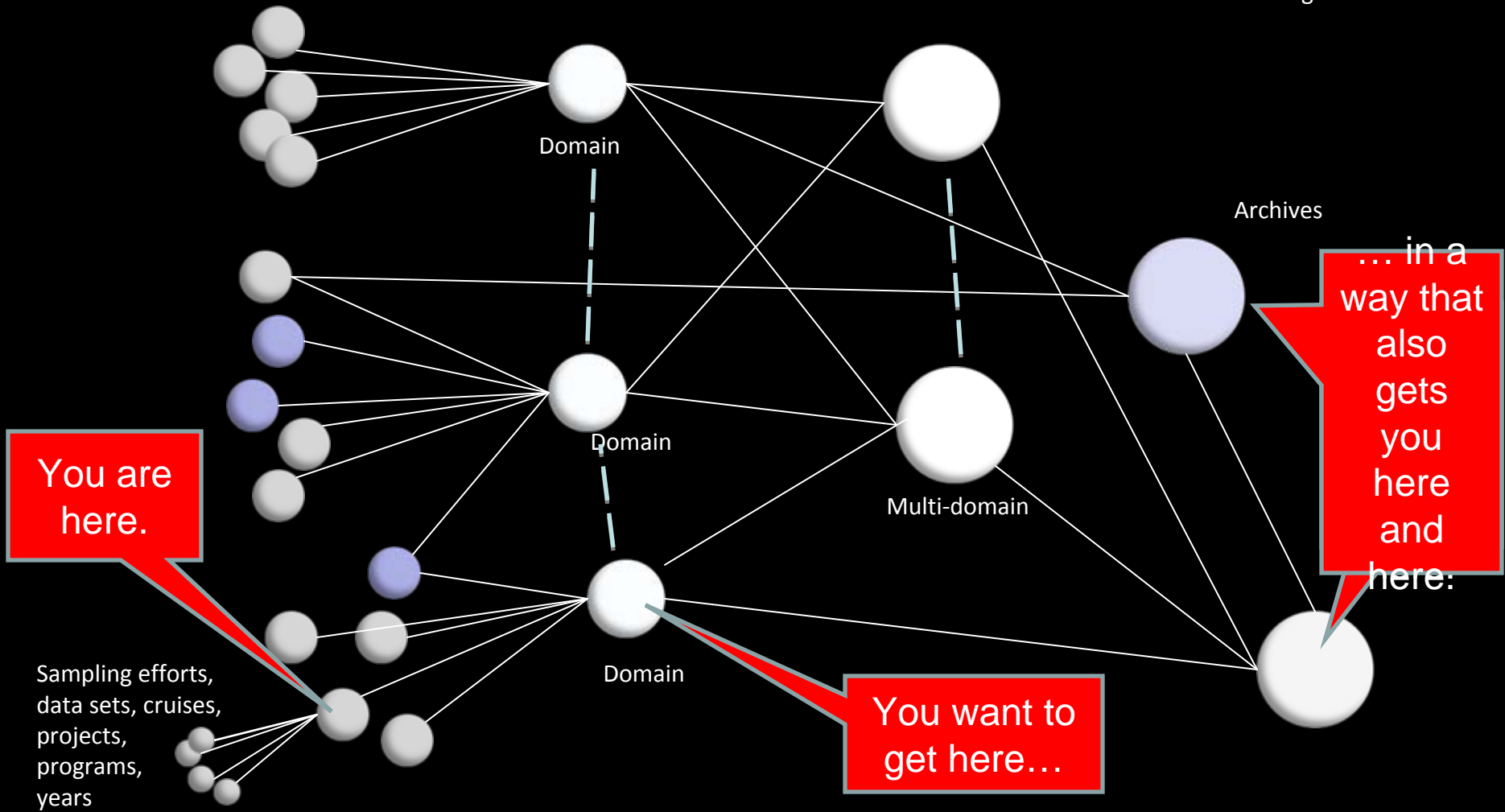
\*Baker, KS, and LYarmey, in press. Data Stewardship: Environmental Data Curation and a Web-of-Repositories. *International Journal of Digital Curation*.

# the big picture



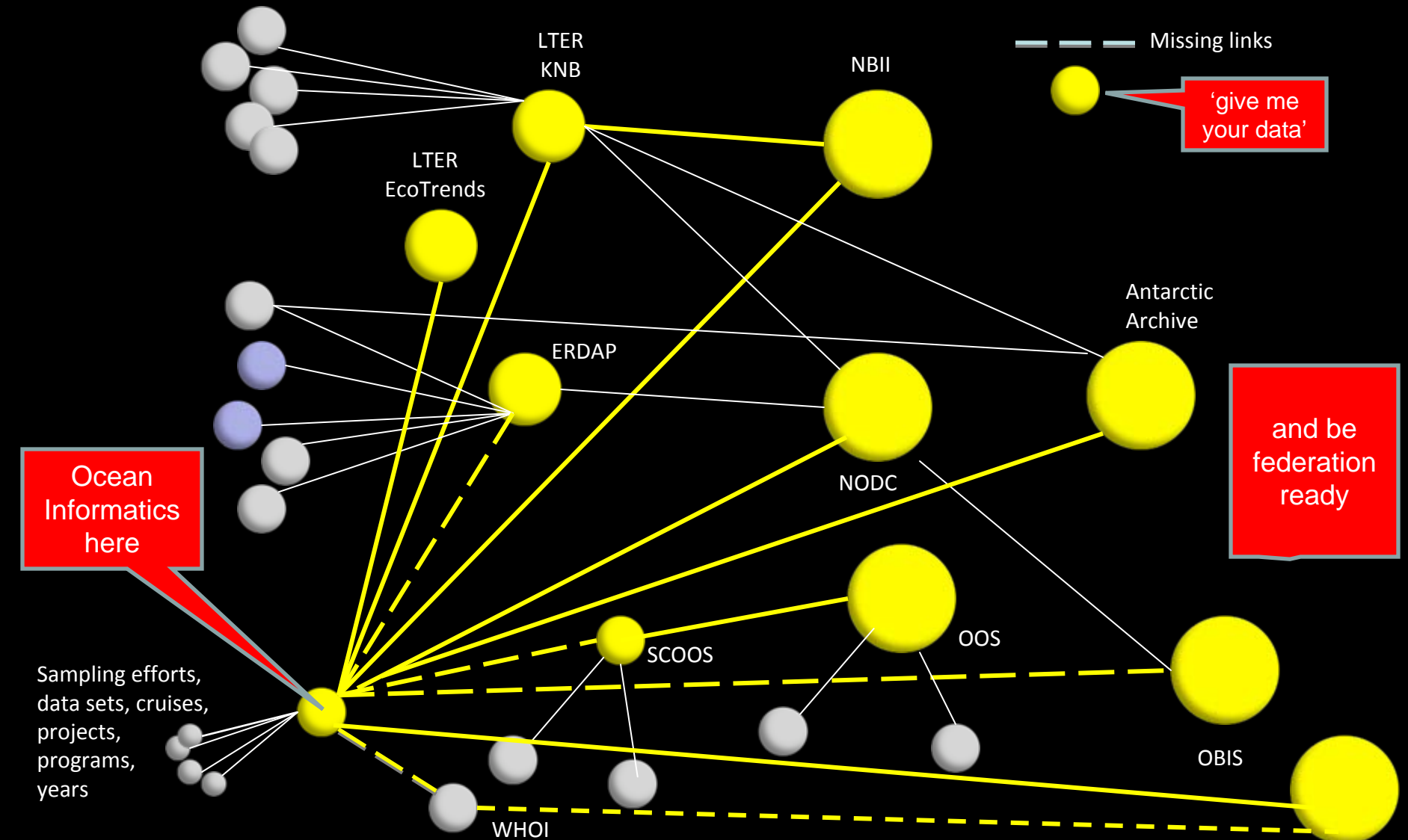
# A model of the big picture

--- Missing links



Spatial:	CalCOFI SCB & Extended	California Current	Pacific Ocean	Global
Repositories:	Site	Local		
Archives:	Regional	Themed	Multi-domain	Themed
Example:	CalCOFI	Ocean Informatics	PACOOS	EcoInformatics
			NCDC; NODC	
				GEOSS, IOBIS, ILTER

# the big picture...and where is PACOOS?



Ocean Informatics here

Missing links  
'give me your data'

and be federation ready

Sampling efforts, data sets, cruises, projects, programs, years

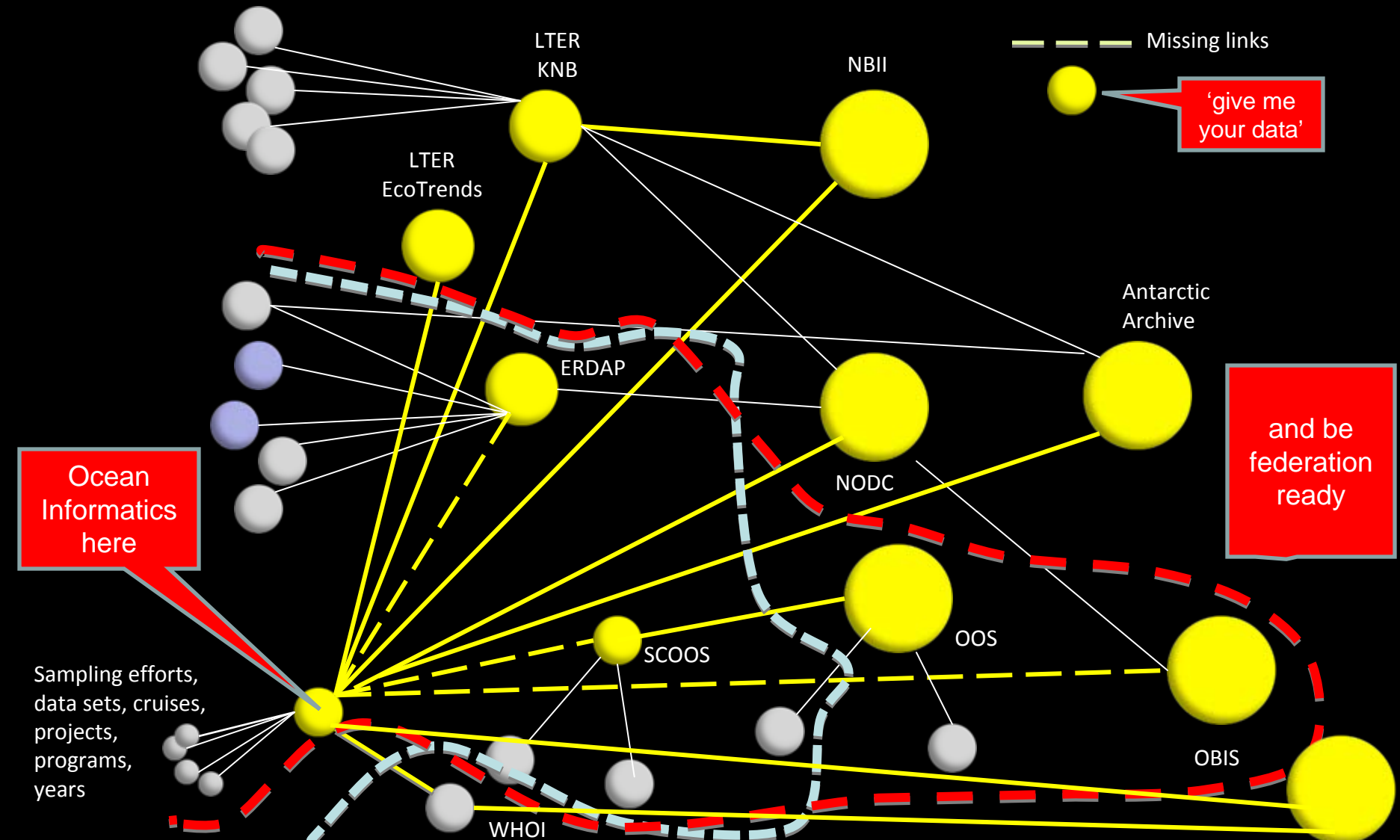
Spatial: CalCOFI SCB & Extended    California Current    Pacific Ocean    Global    ICES

Repositories: Site    Local

Archives:                    Regional    Themed                    MultiDomainThemed                    International

Example: CalCOFI    Ocean Informatics    PACOOS    EcoInformatics    NCDC; NODC                    GEOSS, IOBIS, ILTER

# the big picture...and where is PACOOS?



Spatial: CalCOFI SCB & Extended    California Current    Pacific Ocean    Global    ICES

Repositories: Site    Local

Archives:                    Regional    Themed    MultiDomainThemed    International

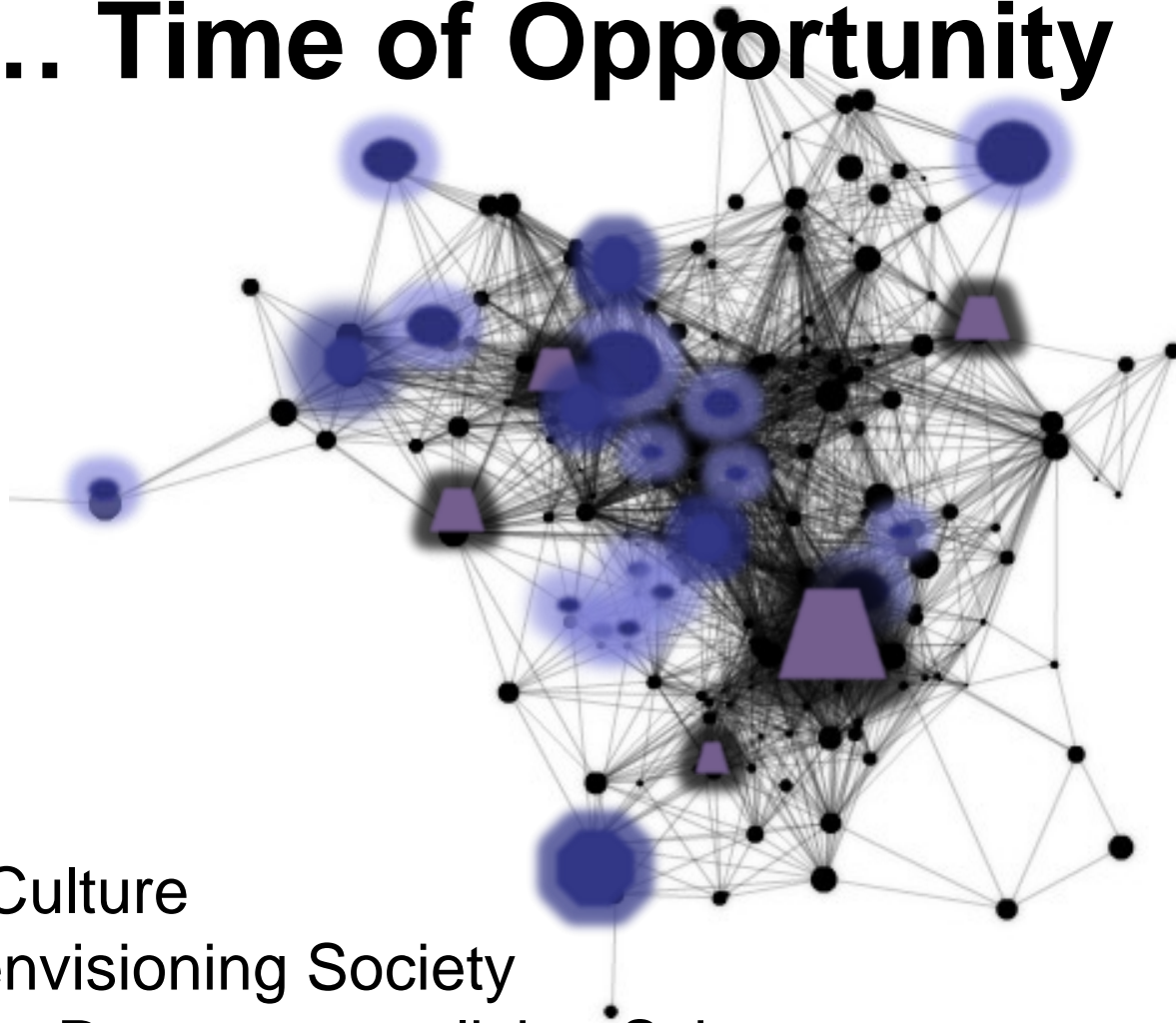
Example: CalCOFI    Ocean Informatics    PACOOS    EcoInformatics    NCCDC; NGEOSS, ICES, IOBIS, ILTER

Scientific Infrastructure:  
Standards-Making  
as a Recursive Process

Activity Lifecycles

# Time of Transition ... Time of Opportunity

- Federated
- Non-hierarchical
- Diverse
- Inclusive
- Distributed
- Coordinated
- Collaborative
- Flexible
- Sustainable



Re-imagining Culture  
by Re-envisioning Society  
by Re-conceptualizing Science  
by Re-framing Data Practices  
and Learning Environments