IOOS Biological Data Services

Three Steps to Enrollment

Tune in  Turn On  Drop Out

May 28, 2014

Philip Goldstein (University of Colorado, OBIS-USA)
Hassan Moustahfid (NOAA US IOOS)
OBIS and IOOS Biological Data

Once Upon a Time in OBIS:

The message from Federal leadership about OBIS:

- Not enough information
- Errors and ambiguities
- Quality and completeness problems
- Can’t do enough with the data
- Can’t maintain support this way

Still around! What happened? …

1. Requirements-based and partner-based enhancements
2. OBIS-USA and IOOS joint development

Outcomes

- Presence-Absence-Abundance
- Integrate with Env Data & CF
- IOOS Biology in 3 RAs
- OBIS-USA from 3M to 28M
- IODE/OBIS enhancing too
Even Better: In 2015, encountering very rich data

Today we see many bio datasets like this:

- Taxonomic range, span trophic/functional groups
- Presence – Absence – Abundance (with effort)
- Intentional study locations relevant to research design
- Consistent sampling methods; time series
- Consistent env data and methods accompany bio data

What enables managing data this rich?

- Data content standards and data flow …
- … and the enrollment process, a consistent, repeatable process.
Enrollment is the process of developing data

• From original source …
• … to IOOS web service
• (and on to downstream services: OBIS, NCEI)
Enrollment Skills in Three Steps

1. Crosswalk data and metadata with DMAC standard
2. Put in common format for serving
3. Configure web services and IOOS Catalog

Skills:
- Love data
- Attention to detail
- Know the science agenda
- Communication
- Balance and adapt enrollment for local requirements

Skills:
- Data structures (table, RDBMS)
- Scripting, programming, for example, SQL, R, others

Skills:
- System admin and configuration (e.g., datasets.xml config file)
- Operations and testing
Enrollment Crosswalk Example
Florida NMS Fish Sampling Timeseries

- 18 years of data; via Sanctuaries MBON organization
- File ‘fk2004_dat1.csv’: 182,519 records (start with a single year)
Enrollment Crosswalk Example
Florida NMS Fish Sampling Timeseries

- Below, a look inside the contents of the fish data (legacy data)
Enrollment Crosswalk Example
Florida NMS Fish Sampling Timeseries

- Below, a glimpse of the fk2015_dat1.csv Enrollment Journal
- Analyze alignment, circulate, resolve questions, specify coding step
Configure web services and IOOS Catalog

Crosswalk data and metadata with DMAC standard

Put in common format for serving

Configure web services and IOOS Catalog

Enrollment Skills Alignment by Organization (proposed)
Assisted Enrollment – Outside Help

Data Originator  IOOS Bio Data Projects

Original data

Working Group: IOOS HQ, RAs, OBIS, other agencies, EDUs, Data Originators

Assisted Enrollment

USGS OBIS-USA

Original data

Assisted Enrollment

Hidden benefit of assisted enrollment:
- IOOS and OBIS globally learn new features

When enrollment requires outside help:
- In the IOOS projects, everyone helped, *ad hoc*.
- OBIS-USA does assisted enrollment regularly; lots of one-off data sources.
Self-Enrollment: Done in the Network

- Self-Enrollment: Skills reside in the network to get the job done.
- Repeat-Enrollment is the key: beat the learning curve.
- Self-Enrollment enables optimal adaptation of enrollment process for original programs’ science needs.

Original data

Data Originator

Joint Activity

Self-Enrollment

Get enrollment skill into the network
- Enable originators and RAs to self-enroll.
- Enable local decision-making on priorities.
- This is the goal of enrollment training.
Science Applications in the Network

- Facilitate joint science applications by originator and RA, within enrollment skill set.
- Incorporate application automation into enrollment cycle.
Enrollment Flavors: Legacy Data and New Data

Data Originator

Legacy Data

New Data

Data Originator / IOOS RA Joint Activity

- Keys to legacy data enrollment may be archived information or personal contact.
- Key to new data is coordination with research design.

Enrollment

Important roles for the enroller:
- Align and balance science and data activities.
- Represent management decision-making.
- Feedback to US, GOOS, IODE global practices.
Biological Data Enrollment
Training Plan (proposed)

Training model:

• The goal is to create independent enrollers
• Train-the-trainer
• Train by example – while enrolling actual data
• Format:
  • Hour sessions over time, telecon or webinar
  - or -
  • Site visit, e.g., two-day workshop
Biological Data Enrollment
Training Plan (proposed)

Training resources:
• Experienced enrollers as instructors
• Documents and tools
• Prior examples
• Reference implementation (proposed)
Biological Data Enrollment Training Plan (proposed)

Training preparation:

• Obtain dataset(s) to enroll
• Identify enroller(s) to be trained
• Verify technical choices / Prepare technical environment (web service installation, tools)
Biological Data Enrollment Training Plan (proposed)

**Training Agenda - Day 1**

1. Activity: Fill in the Enrollment Journal: Start by immersion; Jump into the first example dataset
2. Topic: intro to minimum data / rich data approaches
3. Topic: quality checks and how to respond
4. Topic: CF LLAT (latitude, longitude, altitude, and time)
5. Topic: Taxon validation
6. Topic: advanced information: absence, abundance, biological details, sampling details, tracking, env data
7. Topic: advanced min data / rich data: how to choose
Training Agenda - Day 2

1. Activity: Wrap up enrollment journal for first dataset example
2. Activity: Transform to servable format - e.g., R, SQL, others …
3. Activity: Populate ACDD metadata
4. Activity: Configure web service (e.g., ERDDAP)
5. Topic: Source data extraction: table, matrix, relational DBMS
6. Topic: metadata formats and methods
7. Topic: Discuss that might not have been represented in the first training example.
8. Topic: The global context of the OBIS/IOOS standard, and what it means to scientists, enrollers, and users.