

IOOS Biological Data Services Enrollment/Publication Process

Hassan Moustahfid (NOAA, US IOOS)

Philip Goldstein (USGS, OBIS-USA)

IOOS DMAC RAs Workshop Sep. 10-12 2013



Ocean Biogeographic Information System USA

Introduction to this Presentation:

Enroll data in the standard; publish data via services

- IOOS Core Biological Variables: Fish Species, Fish Abundance, Zooplankton Species, Zooplankton Abundance, Phytoplankton Species
- Applicable to any aquatic taxa
- Current and Future

Address the Topic: What is Enrollment?

... and answer in terms of the following ...

- Results of Enrollment
- Flavors of Enrollment
- Enrollment Steps
- Enrollment Techniques and Tools

Delivered Results of Enrollment – What is Accomplished When Enrollment is Complete?

- Dataset is on IOOS RNs web/database/webservices
- Dataset is on IOOS DMAC (Registry/Catalog/Viewer)
- Archive en route to NODC sync with OBIS-USA
- Metadata published and distributed
- Dataset included at Other Portal such as Data.gov

Enrollment Checklist

What is complete when enrollment is complete ...

- ✓ Data via IOOS RN and/or partner
- ✓ Referred by IOOS DMAC
- ✓ Complementary access via OBIS-USA and NODC
- ✓ NetCDF version of data
- ✓ Metadata Clearinghouse
- ✓ Data.gov

Value of Enrolling IOOS BDS Data: **the Wealth of Data**

Asset	Dataset Richness, Documentation, Quality and Standards Established
Value	Dataset publicly accessible and usable; Application opportunities multiplied
Security	Life cycle from origin → application → archive has been anchored
Investment	Enhanced relationship of data originator with IOOS and partners; also enhance data origination practices?
Community	Data originator's voice accessible to future development of standards and applications

Flavors of Enrollment

The “flavors” are in relation to ...

- Data Originator / Data Holder
- Associated IOOS Regional Node

The Flavors are:

1. Assisted Enrollment
2. Self-Enrollment
3. Train-the-enroller

Flavors of Enrollment (continued)

Assisted Enrollment

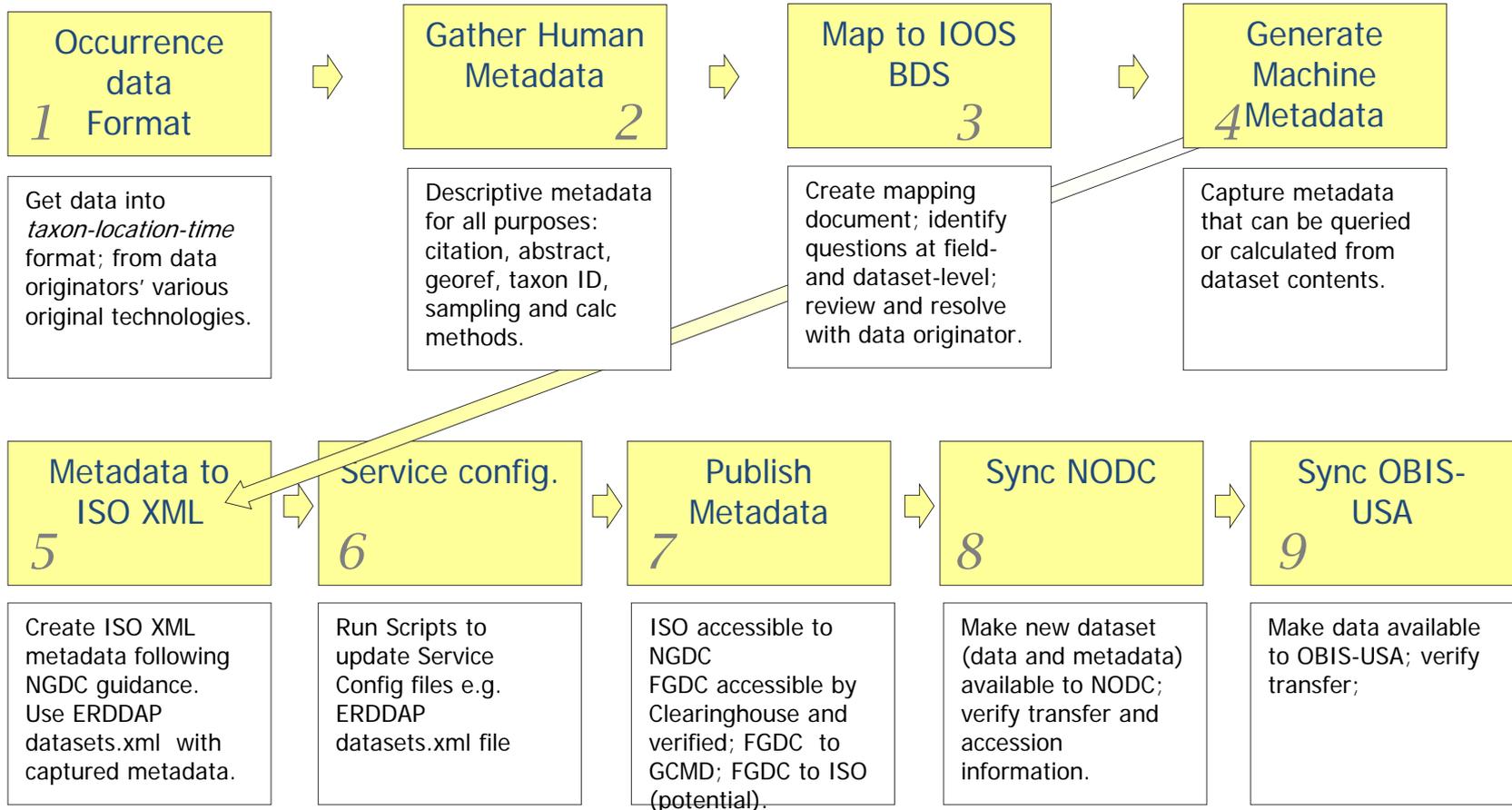
- IOOS RN assists data partner as opportunity or need presents itself

Self Enrollment

- Partner works independently to the extent they desire
- Self-Enrollment is not intended to be 100%: IOOS stands by to help; IOOS always does a quality check
- Self Enrollment capability is influenced by:
 - Knowledge/skill resources available
 - Community practices, repeatability
 - Repeat enrollment by originator / RN

How?

Enrollment Process Steps 1 thru 9

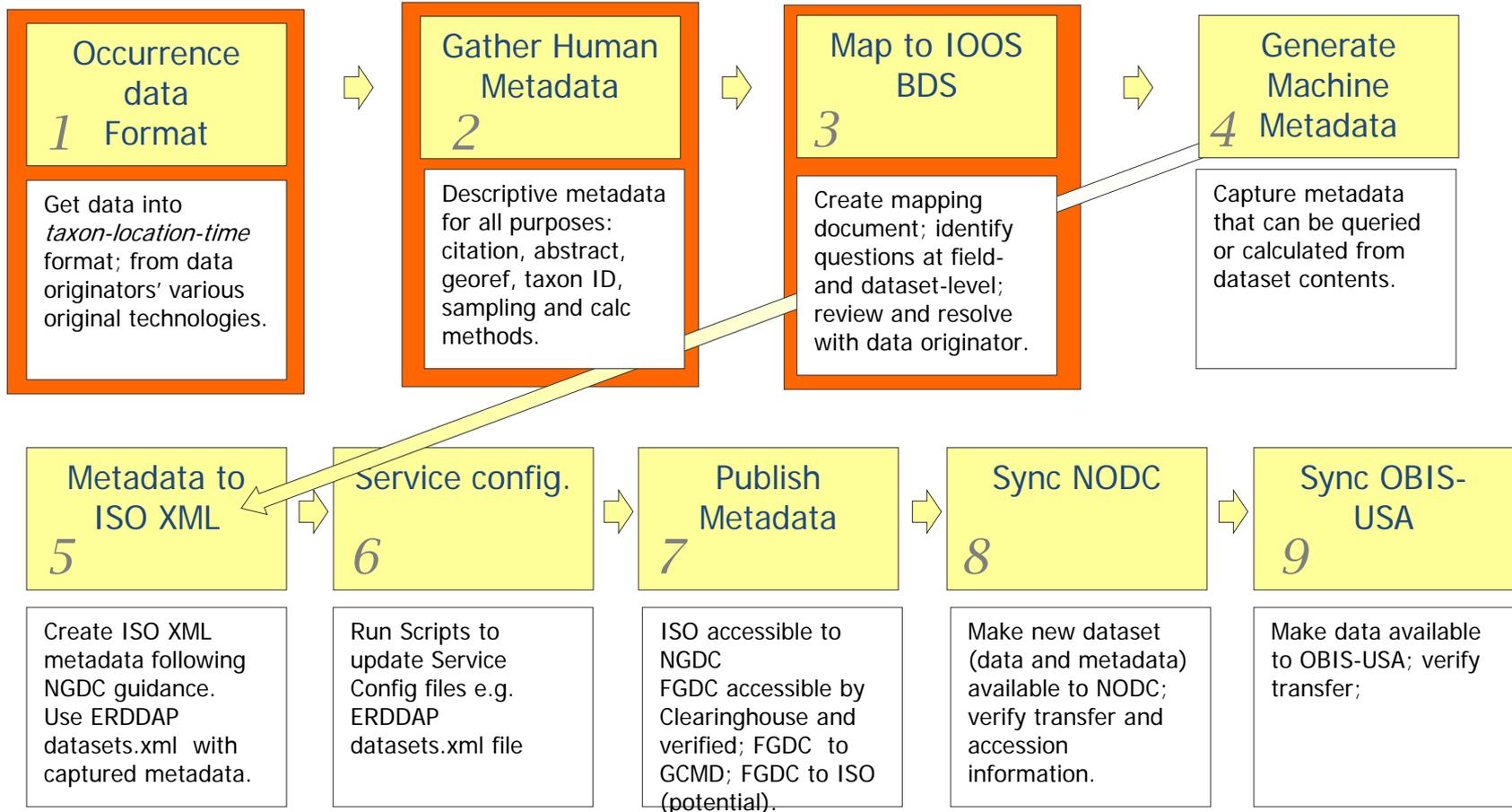


Enrollment Process Steps

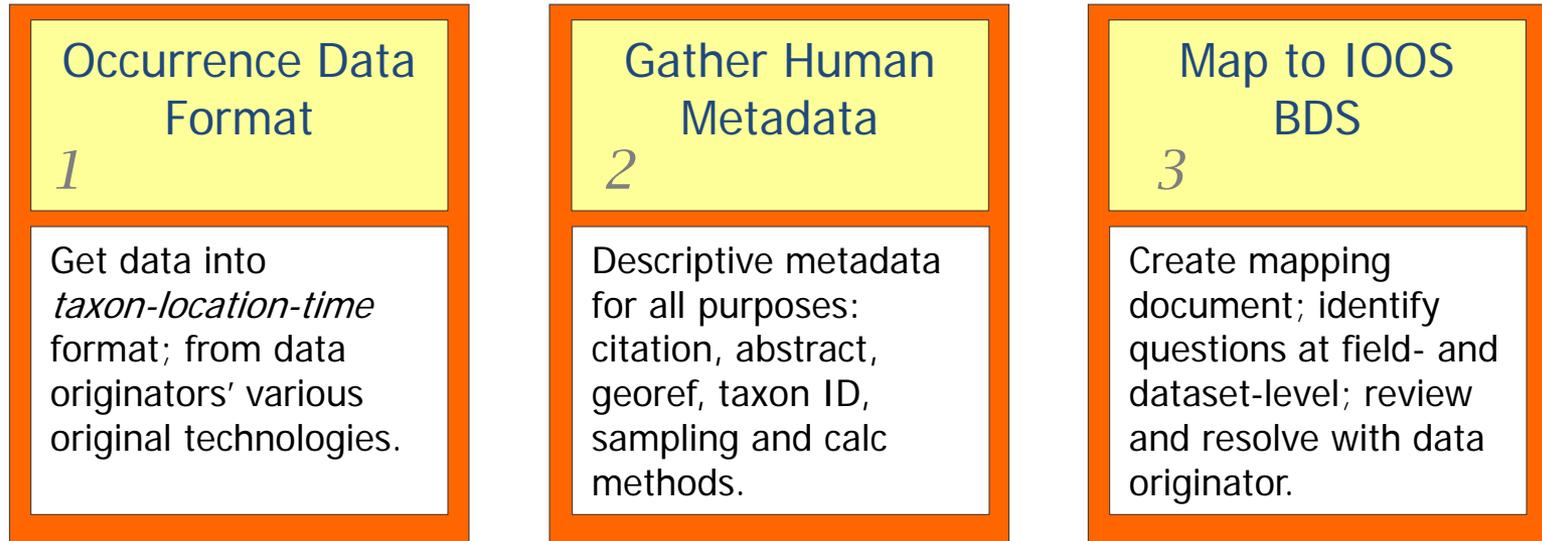
Where main emphasis is content expertise

Interaction:

- Highlighted steps indicate substantial involvement of data originator
- Here is where to balance cost and detail, and build enthusiasm
- Here is the excellent opportunity to build community
- Non-highlight steps are IOOS-internal

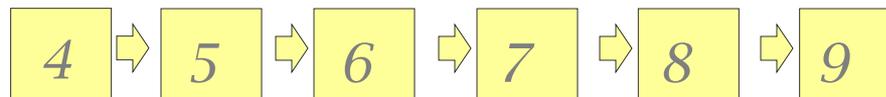


Enrollment Involving Data Partner, 1-2-3



Involving Data Partner

Internal to IOOS Technology / Organization



Step 1: Occurrence Data; Various Original Formats

Day	Month	Year	Latitude	Longitude	Scientific Name	Abundance
31	7	1991	30.43	-80.22	Pagrus pagrus	1
31	7	1991	30.43	-80.22	Ballistes capricus	3
31	7	1991	30.43	-80.22	Ballistes capricus	3
31	7	1991	30.43	-80.22	Ballistes capricus	3
31	7	1991	30.44	-80.20	Rhomboplites aurorubens	18
31	7	1991	30.44	-80.20	Rhomboplites aurorubens	18
31	7	1991	30.44	-80.20	Rhomboplites aurorubens	18
31	7	1991	30.44	-80.20	Rhomboplites aurorubens	18
31	7	1991	30.44	-80.20	Rhomboplites aurorubens	18
31	7	1991	30.44	-80.20	Rhomboplites aurorubens	18
31	7	1991	30.44	-80.20	Rhomboplites aurorubens	18
31	7	1991	30.44	-80.20	Rhomboplites aurorubens	18
31	7	1991	30.44	-80.20	Rhomboplites aurorubens	18
31	7	1991	30.44	-80.20	Pagrus pagrus	2
31	7	1991	30.44	-80.20	Pagrus pagrus	2
31	3	1992	32.51	-79.69	Centropomus striata	21
31	3	1992	32.51	-79.69	Centropomus striata	21
31	3	1992	32.51	-79.69	Centropomus striata	21

Data Sheet
10%

Matrix
KTE

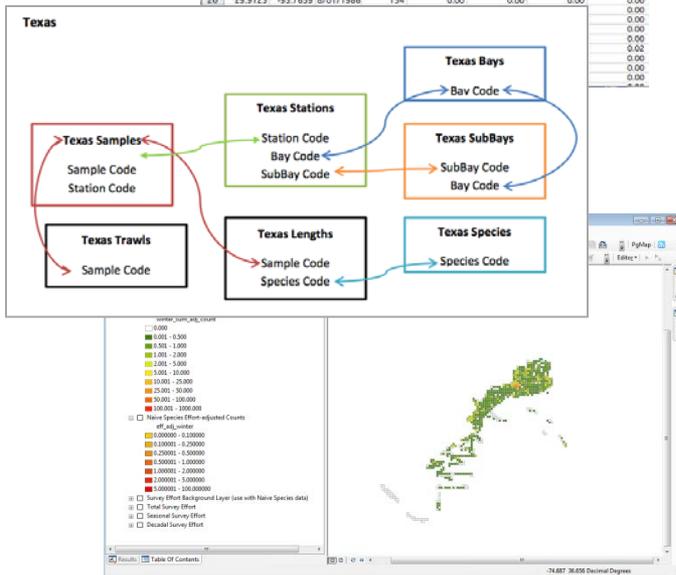
Relational Database
90%

GIS
KTE

Occurrence Data Format

1

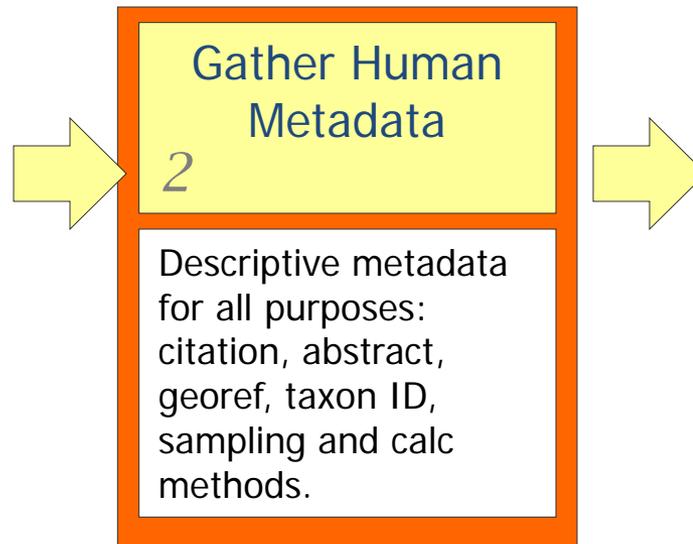
Get data into *taxon-location-time* format; from data originators' various original technologies.



ScientificName	Longitude	Latitude	Date
Oligochaeta	-75.017	38.75	7/21/80
Oligochaeta	-74.967	38.517	12/6/80
Oligochaeta	-74.967	38.517	12/6/80
Nemertina	-75	38.733	11/30/84
Nemertina	-75	38.733	11/30/84
Nemertina	-75	38.733	11/30/84
Polychaeta	-75.017	38.733	6/18/85
Squatina dumeril	-74.8	38.5	9/21/88
Rajiformes	-74.956	38.4175	7/25/95
Testudines	-74.833	38.5	5/18/78
Testudines	-74.833	38.083	5/18/78
Rhinoptera bonasus	-75.117	38.1	9/12/80
Rhinoptera bonasus	-75.067	38.05	9/12/80
Rhinoptera bonasus	-75.033	38.033	9/12/80
Elasmobranchii	-75.05	38.067	8/21/79

KTE "known to exist": Occurrence data in this form are known to exist, though not yet encountered in IOOS BDS enrollment experience to date.

Step 2: Gather Human Metadata



Some metadata can be generated from data (referred to as *machine metadata*); this is an internal technical task.

In contrast, *Human metadata* refers to essential descriptive details, vital to effective use of the data, that must be developed through human expertise and interaction.

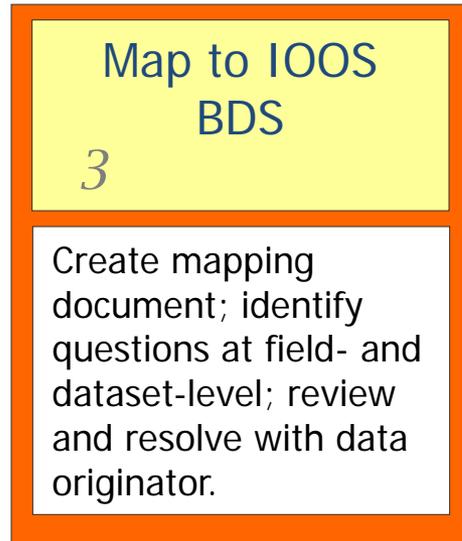
Human metadata is a key area to capture expertise from data originators; Data originator required; Enroller assists in every way feasible.

Human Metadata Development Captures Valuable Knowledge

- Quality dialog with data originator is required
- Connected to the IOOS BDS data standards process
- *Important Contents of Metadata:*
 - *Dataset citation & institutional context*
 - *Abstract*
 - *Thematic keywords*
 - *Georeference method, coordinate uncertainty*
 - *Taxonomic identification details*
 - *Sampling protocol / effort / conditions*
 - *Derived data calculation methods*
 - *Dataset history*
 - *Related resources*

These features enhance discovery, attribution, and use in applications; also influence future data capture.

Step 3: Map Data Contents to IOOS BDS



Map to IOOS
BDS

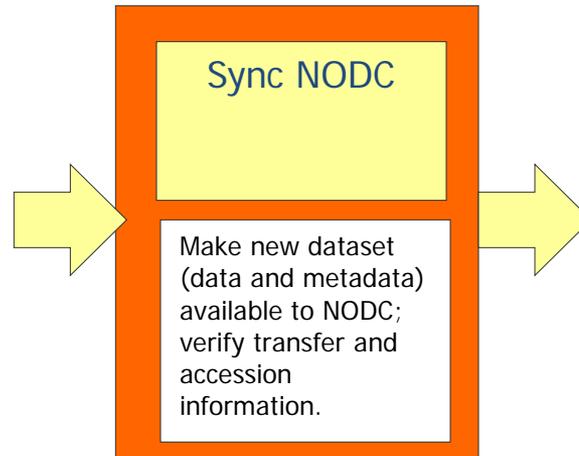
3

Create mapping document; identify questions at field- and dataset-level; review and resolve with data originator.

- Match partner's data contents with IOOS BDS terms
- Document exceptions, questions, any kind of follow-up
- Use document sharing to communicate and resolve items between partner and enroller
- Assist in metadata capture and development

Demonstrate an example of Enrollment Journal -
CAGES Louisiana data

Step 8: Sync with NODC

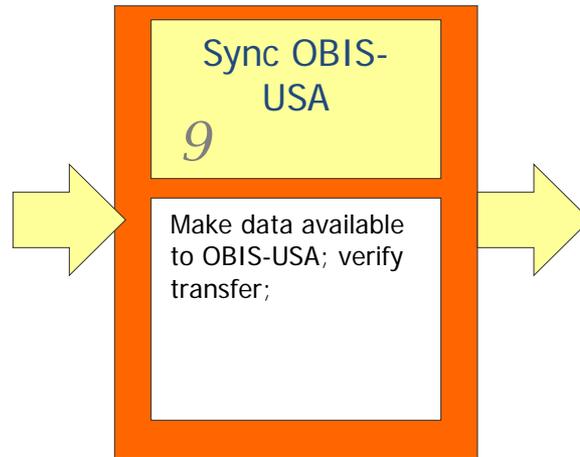


Make new dataset (data and metadata) available to NODC; verify transfer and accession information using the same process already in place between IOOS, RAs and NODC to archive observing data.

Two important aspects of this coordination with NODC are:

1. Make sure that NODC has the same version, that is the same snapshot in the lifecycle of the data contents, that IOOS is serving
2. Where NODC needs metadata to accession an archive, and IOOS needs metadata to serve Biological Data Services, to the extent that these two uses of metadata are based on the same contents, IOOS and NODC can coordinate so as not to duplicate efforts.

Step 9 – Sync with OBIS-USA



OBIS-USA to carry the biogeographic component of dataset: less application-specific detail but discoverable in the national marine biogeographic dataset. OBIS-USA refers users to IOOS for live service-accessible detail and applications.

Practical integration with OBIS-USA enabled by:

1. Common data and metadata content and format standards
2. Common enrollment practices.

IPT and Darwin Core Archive (DwC-A)

What is a Darwin Core Archive?

- Created by IPT, Administered by IPT, Made Accessible by IPT, Consumed by IPT
- Zipped together, DwC-A consists of three files:

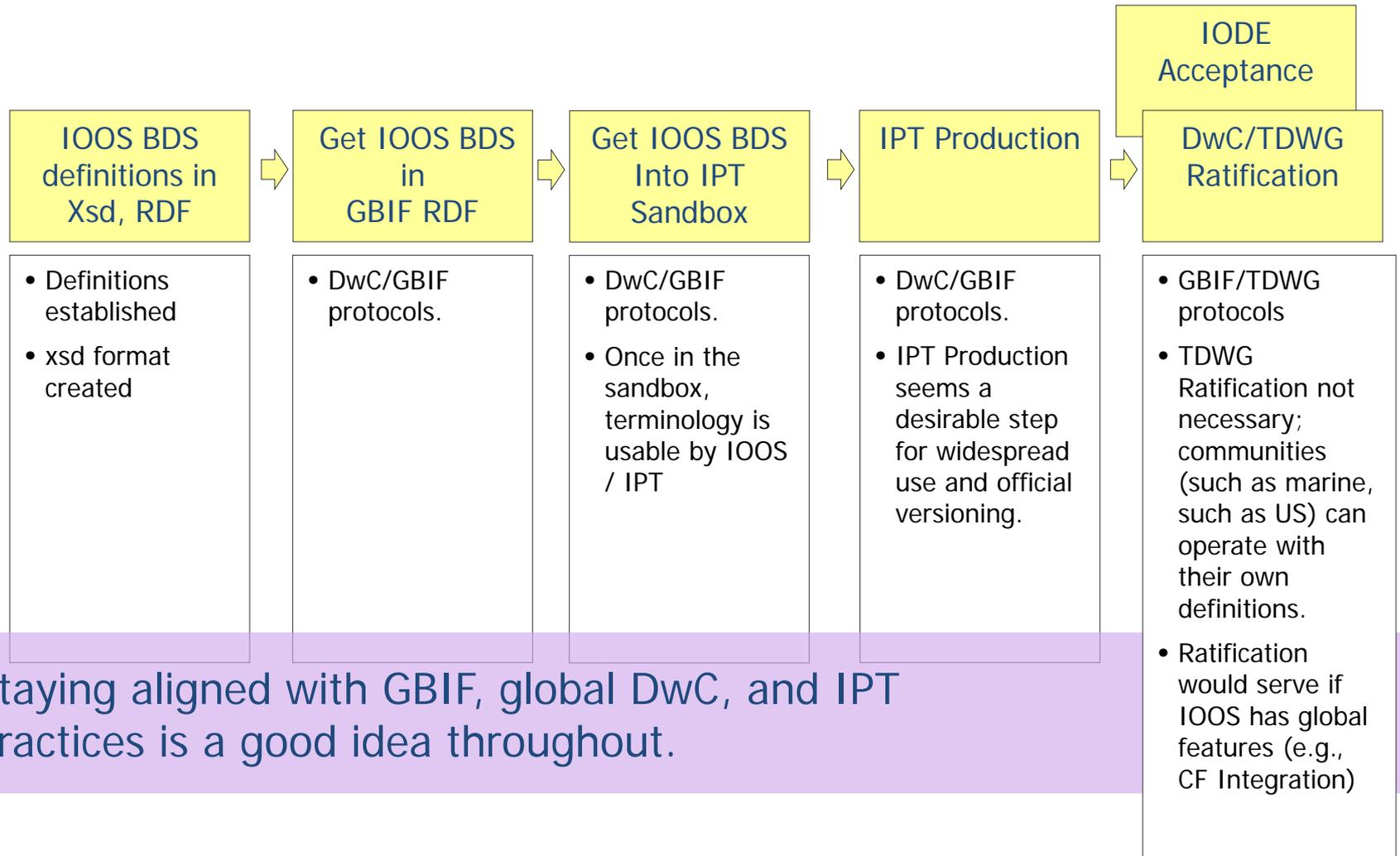
Biological
Observation
Records
(columns-and-
rows)

Column List
(xml)

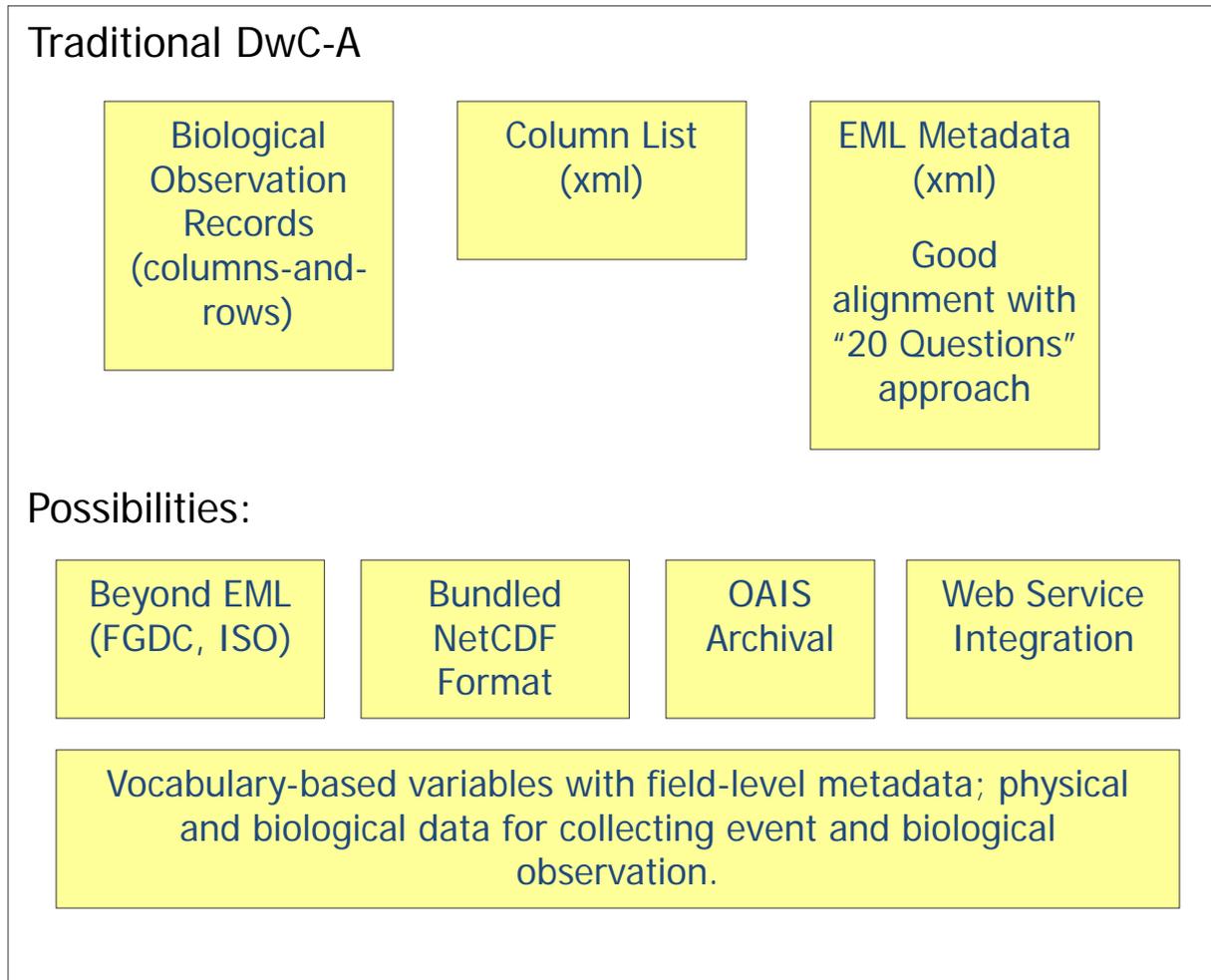
EML Metadata
(xml)

Good
alignement
with "20
Questions"

IPT and IOOS BDS

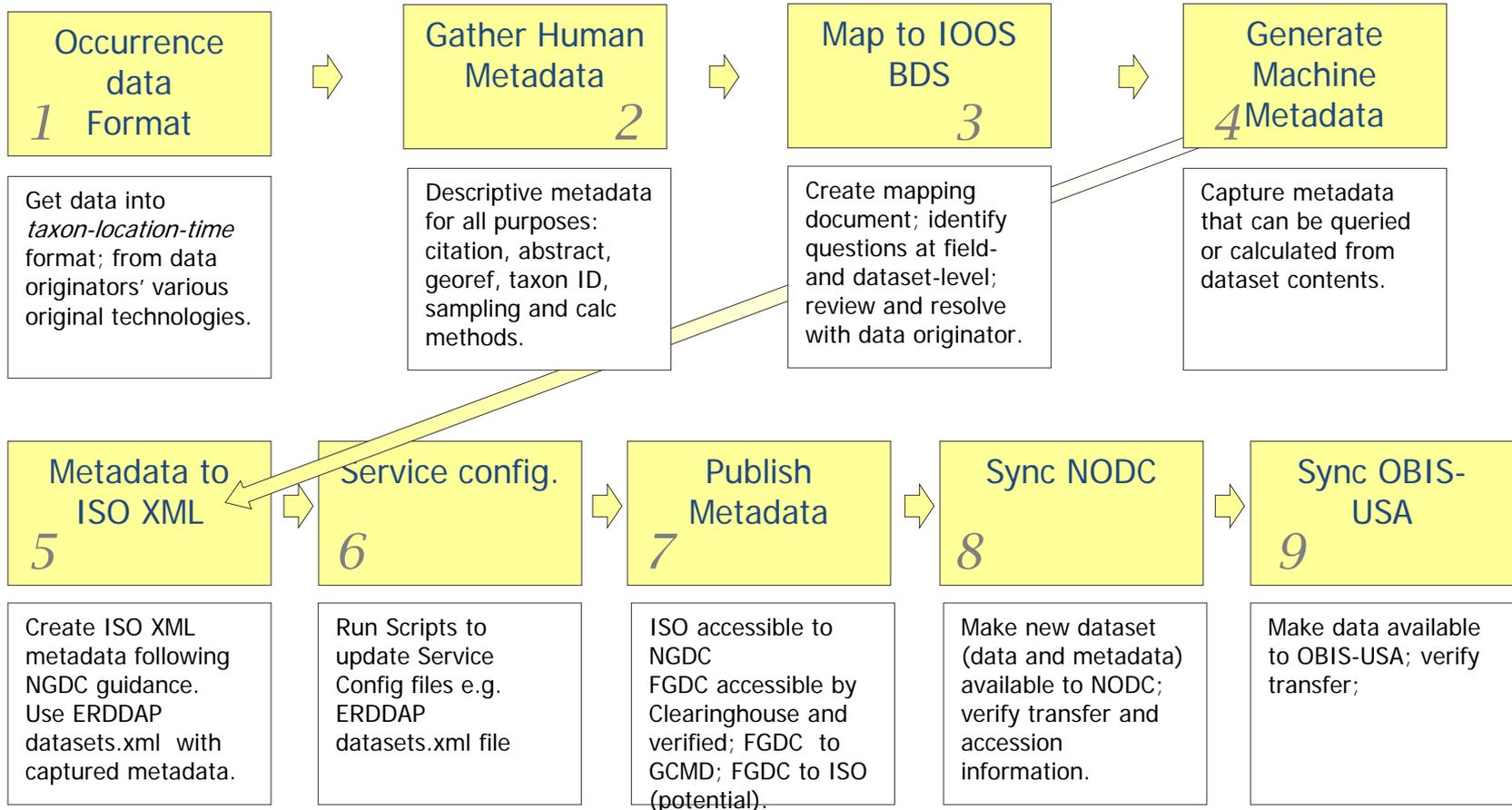


Darwin Core Archive – Future Possibilities



How?

Enrollment Process Steps 1 thru 9



Value of Enrolling IOOS BDS Data: **the Wealth of Data**

Asset	Dataset Richness, Documentation, Quality and Standards Established
Value	Dataset publicly accessible and usable; Application opportunities multiplied
Security	Life cycle from origin → application → archive has been anchored
Investment	Enhanced relationship of data originator with IOOS and partners; also enhance data origination practices?
Community	Data originator's voice accessible to future development of standards and applications