2017 LTER Information Management Committee Annual Meeting Report

Hosted in conjunction with Earth Science Information Partners (ESIP) Summer Meeting

July 24, 2017; Bloomington, Indiana

http://im.lternet.edu/news/meetings/2017

Introduction

The 2017 Annual LTER Information Managers meeting was held at the Indiana University, Bloomington, hosted in conjunction with the Earth Science Information Partners (ESIP) Summer Meeting. LTER Information Managers, LTER Network Communications Office (NCO) representatives, Environmental Data Initiative (EDI) representatives and others gathered to discuss issues related to LTER Information Management, the working relationships between the newly formed organizations and information management in this new era. Three new LTER sites are also coming onboard and the new Information Managers attended this meeting are seeking input on their data management plans and other collaborative efforts.

Contents

Attendees

Suzanne Remillard (AND), Don Henshaw (AND), Jonathan Walsh (BES), **Tim Whiteaker** (BLE), Jason Downing (BNZ), Stevan Earl (CAP), Brian Herndon (CWT), Kristin Vanderbilt (FCE/EDI), Wade Sheldon (GCE), Adam Sapp (GCE), Mary Martin (HBR), Emery Boose (HFR), Sven Bohm (KBS), Yang Xia (KNZ), Eda Melendez (LUQ), **Renée F. Brown** (MCM), Gastil Gastil-Buhl (MCR), **Stace Beaulieu** (NES), **Rob Bochenek** (NGA), Corinna Gries (NTL/EDI), Hope Humphries (NWT), **Li Kui** (SBC), Margaret O'Brien (SBC/EDI), John Porter (VCR), Duane Costa (EDI), Mark Schildauer (NCO), Julien Brun (NCEAS), Pier Luigi Buttigieg (GUEST), Karen Baker (GUEST)
bold names are new site IMs attending their first annual IMC meeting

Attended Remotely: James Conners (CCE/PAL), Dan Bahauddin (CDR)

Unable to Attend: Jim Laundre (ARC), Ken Ramsey (JRN), Hap Garrit (PIE), vacant (SEV)

Agenda at a Glance

08:00	Welcome and review of agenda
08:15	LTER IMC-EDI updates and discussion
08:40	LTER IMC-NCO updates and discussion
09:20	Advertisements (30 seconds each) of ESIP Sessions led by IMC or EDI people
09:45	NSF call with Peter McCartney including pre and post discussion
10:45	Group Photo (Courtyard) and break
11:00	Working Group Updates
11:45	Business Meeting
12:00	Lunch
01:00	Lightning Talks (2 minutes each)
01:05	Breakout Sessions
02:50	Break
02:55	Breakout Report Backs
03:55	Break
04:00	New site IM's present data management plans
05:00	Wrap-up and Action Items
05:30	Meeting Adjourned
07:00	Group dinner at The Irish Lion

LTER Participation in ESIP Meeting

Tuesday, July 25 • 9:30am - 11:00am
Create EML Using R and Share the Script via the Information Management Code Repository
Corinna Gries

Tuesday, July 25 • 2:00pm - 3:30pm
Use Application Programming Interfaces of Data Repositories to Create Local Data
CatalogsDuane Costa • Stevan Earl • Gastil Gastil-Buhl • John

Tuesday, July 25 • 2:00pm - 3:30pm
Tools and Strategies for End-to-End Sensor Data Management
Matthew Bartos • Mike Botts • Mike Daniels • Wade Sheldon

Tuesday, July 25 • 4:00pm - 5:30pm
Developing an Information Management Code RepositoryColin Smith
Kristin Vanderbilt

Wednesday, July 26 • 4:00pm - 5:30pm
Annotating datasets with measurement classes from the Ecosystem Ontology (ECSO)
Margaret O'Brien • Mark Schildhauer

Wednesday, July 26 • 7:00pm - 9:00pm
Research as Art at the WonderLab
LTER IM Posters Include:
        BES Arts and Science Integration Program (Jonathan Walsh)
        DEIMS: An IMS for Scientific Research (Eda Melendez)
        Luquillo LTER - LCZO Collaboration in Data Collection and Management (Eda Melendez)

Thursday, July 27 • 9:00am - 10:30am
Sustainable Data Management, Return on Investment
Margaret O'Brien • Anne Wilson

Elections

Jonathan Walsh & Yang Xia are rotating off of IM-Exec members. Thank you for your service.
Two nominations for 2 IM-Exec positions. Suzanne Remillard & Jason Downing are unanimously
elected. Thank you for agreeing to serve.

General Meeting Summaries

General Group Discussion

LTER-NCO Discussion

- Mark Schildauer & Julien Brun are present representing NCO.
- Mark is retiring as director of computing at NCEAS, but is still working on some projects (Artic data center, SNAP science nature and people, partnership, SEEK controlled vocab and semantics; semantic ontology)
- Work on ontologies continues at NCEAS, ESIP, etc.
- Currently 5 synthesis working groups at NCEAS. These science synthesis groups are ultimate data congruency checker. Opportunity to learn.
- Matt Jones will work more closely with Marty on LTER cyberinfrastruture.
- Julien works on mentoring education using remote sensing and will take over managing working groups at NCEAS. He codes in R and Python and has GIS skills. He aids working groups in pulling data together. Working groups were able to spin up quickly due to LTER data resources. There are issues of channels of communications to EDI that need to be resolved.
- Guest: Pier Luigi Buttigieg – polar research in Germany; works on semantics for environmental; expert knowledge engineer (sweet semantics group)
- Gastil provided some topics from Marty. She presented a slide on NCO infrastructure with the overview slides
- Marty works with Communications and Educations reps
- Status of network databases:
    - o BiblioDB is on hold while NCO finds generic solution
    - o PersonnelDB and SiteDB are archived
    - o ClimHy is up in the air
- LTER website is contracted to a web design firm (NDIC). Marty is looking for input from sites.
- Databits – another conversation; what's the future? Databits might become part of larger blog at NCO. Or continue as journal at EDI.
- IM webpage may go away or migrate to IM.lternet.edu.
- Mailing lists are moving over to Google Groups and site specific lists going away

LTER-EDI Discussion

- No formal presentation prepared, but discussion based on notes on agenda.
- EDI is the now the name of the repository. EDI provides a default citation following ESIP recommendations. Includes name of repository. Recommend putting LTER site in the dataset title. Data can be downloaded from many sources. The dataset needs to carry its own identity.
- DataONE citation refers to source of the data. DataONE is really a search engine.
- DOI resolves to the owner which is EDI, but should say which repository the data was obtained
- EML creators (individuals) are currently shown in the citation.
- Not all repositories / search engines will use the same EML fields in citations. So it may be a good idea to put the site in the dataset title. A temporary solution.
- The EML creator may be an organization. This option is not currently used by DataONE for searches

- How do we get our projects on the EDI agenda? Talk to EDI.
- Use EDI RFC mechanism to get feedback on ideas.
- EDI should maintain a list of projects to work on. But it cannot handle proposals of any kind.
- Responses to RFC on SiteDB have not been shared.
- EDI advisory board will start up in October. Mary Martin is our representative. Face-to-face meeting in spring. Also includes Peter Arzberger, Aaron Ellison, Rebecca Costella, and Ian Foster.
- EDI welcomes input & comments from IMC. EDI does not want input from IMC to select someone to be on the advisory board, they think it should be their selection.
- Encourage general conversation now. Formalize lines of communication later.
- EDI will also assume responsibility for units.
- Upcoming RFC. What should be tackled first? ClimHy, units, etc.
- Oxford Press has expressed interest in using the LTER controlled vocabulary.
- Successful thematic standardization workshop led by Margaret; community data model; several databases have been converted to that data model. See ongoing blog on EDI website.
- EDI is looking for suggestions on next workshop.
- IM-Exec is accepting requests for water coolers up to 6 months in advance.
- EDI wants many to many conversations; no layers to communicate; we can formalize the pathways later; therefore, IMC should have planned discussions to coalesce to get our best ideas out and then take them to EDI.
- Julien asked how many IMs have been contacted by NCO working groups. No one raised a hand.
- EDI would like to hear what the next data package design pattern workshop.
- Site characteristics for climate stations across sites.
- In person training workshop in the fall 2017; topic isn't decided; what other training do we want?
- EDI training sessions on Tuesdays
- Collin is EDI information manager.

Databits

- EDI is willing to host DataBits (including past issues). Include larger environmental IM community.
- An alternative would be tagged blog entries on NCO web page. But blog is not an editorial process. Databits was treated as more than a community newsletter; an edited publication. We could bridge the Databits publication to the blog by posting some current published article.
- Databits would include editorial filtering. There were a few instances where this was required. Open blog would not include any filtering.
- An editor (and themed issues) would help to generate content (Karen Baker reiterated the importance of having an editor that contacts people; pushes, motivates; everyone is very busy.
- Few submissions recently. Move to 1 issue per year? Align with some activity?
- Editor is really a coordinator (not much editing).
- Keep in mind that Databits is a newsletter, not a journal. Used to include brief sitebytes.
- EDI could serve as coordinator. Publish as a blog instead of a regular issue.
- What is our niche? Somewhere between ecoinformatics journal and blog.
- We should encourage people to submit blogs to EDI, but look around for thematic ideas for a databits version. Retain the option of a formal issue centered on a particular activity.
- MOB suggested we rotate the editors. Eda volunteered to be editor. She is concerned about losing the LTER Databits terminology and resource. It's a way of maintaining our identity.

Call to NSF

NSF Questions from LTER IMC:

1. Funding timeline for NCO & EDI?
2. Is NSF planning a new publication reporting process?
3. Renewal RFP timeline?
4. How will site IM guidelines be used?

Peter McCartney VTC

- Introduce 5 new information managers.
- How can we express support for EDI & NCO?
- Role of EDI and similar data repositories is a high-level issue at NSF.
- Need a smooth and unambiguous accountability process. This could be handled by good documentation. E.g. a list of datasets with DOIs. Demonstrate progress by incremental changes.
- Clarify what is under review. List of datasets could be used in reports and proposals.
- Still some technical issues (e.g. with DOIs) to be solved.
- Upcoming changes in annual reports (e.g. using DOIs)?
- Ideally there should be a linkage between dataset DOI, NSF grant number, and publication DOI.
- Updates on RFP timeline for LTER renewals?
- NSF must allow at least 90 days between RFP and submission deadline.
- Data Management Plans are important in site visits and will be a part of the review process.
- Moving back to use of IM guidelines. Needs to be a living document.
- Recent IMC working group updated the site IM guidelines. Still not clear to EB how these will be used.
- The guidelines must be endorsed by the scientists as well as the IMs.
- Current discussion at NSF about how to support the LTER ASM meeting.
- NSF is moving away from funding standing meetings. Focus on outcomes and products that come from a productive meeting.
- Increase opportunities for participation by data managers, statisticians, scientists, etc.
- Peter will be at ESA this year.

Discussion

- NSF will apparently run with the site IM guidelines. Wade will put on agenda for EB.
- Dataset DOIs by themselves are a reductionist measure of success.
- Data citations may be an issue where there are page limits.
- Triad between datasets, funding, and publications
- What is the best way to link a publication to a dataset?
- DataONE is partnering with DataCite for citation solutions. DateCite uses an XML specification as well as DOI.
- DataONE citation does not include the DOI. Chicken & egg problem.
- How can we get support to continue this meeting? A supplement to EDI and NCO is a possibility.
- Next ASM will be at Asilomar (Monterey CA).

Working Group Updates

Information Management System Guidelines – Kristin Vanderbilt

- Work Completed with creation of a formal document

WiReD - Brian Herndon

- IM website. Currently in Drupal 6. Move to Drupal 7? Or move to WordPress?
- What are costs & benefits of moving to WordPress?
- What do we want this website to do?
- What content should go to EDI, LTER general website, archive?
- LTER document archive is incomplete for IM documents. Could IM documents be stored here?
- NCO has made it clear that they're not interested in maintaining a document library. These historic documents are important to us as a community and should be maintained. Our current IM meeting notes are stored in about 4 different places.
- IMC use Google Drive for active collaboration, but we need to have a place to put final products (like reports, notes, etc.)
- Current NCO website is being revamped by 3rd party in WordPress. One of three vendors was willing to move information from Drupal to WordPress.
- WordPress is also PHP/MySQL.
- The major task will be sifting the wheat from the chaff. Not discarding, but moving to lower levels in menus, etc.
- CMS is good for managing information but not for archiving.
- Another strategy would be to archive the current website and move items needed to new website.

LTER Sites' Website Guidelines - Mary Martin & Marty Downs

- Mary and Marty have made a first pass at the 2005 document. The WG will meet to discuss these changes.
- Marty has created a survey to determine how people are using websites.
- NSF thinking has changed about local data catalogs since the original document, so we need to keep these as best practices and guidelines for systems that identify the purpose and function.

PASTA EML Congruence Checks (ECC) - Margaret O'Brien

- Active since April 2017. Checker is gatekeeper to PASTA.
- Only integrity checks prevent uploading
- Currently 36 checks.
- New checks are more complex (i.e., datetime).
- They will meet in an afternoon breakout.

Afternoon Breakout Group Reports

Terms of Reference (ToR) – led by Don Henshaw

- Should NCO or EDI have ex-officio representation on IM-Exec?
- Should IMC have presence on the EDI advisory committee?
- EDI should select the person. As sites select their own advisory committees.
- Are IMs required to attend annual meetings? Sites should be encouraged to come.
- What will be the fate of IMC and IM-Exec meetings?
- IM-Exec might pursue funding for the annual IMC meeting. Dave Garrison was not encouraging.
- IM-Exec could also look for cost-effective venues.
- Sites could also contribute if they thought it was important.
- NSF has emphasized that meetings need deliverables to be funded.
- No extra cost for today's meeting as part of ESIP meeting.
- Real cost of this meeting includes our time.
- Expectations for maintaining working groups? Annual reports.
- IM-Exec terms? Recommend 3 years. EB terms are 3 years.
- Reduction in size of IM-Exec last year included ability to ask for help. Emery—report. Jason—meeting logistics. John—travel.

PASTA ECC new checks – led by Margaret O'Brien

- Date-time formats. See chart of all format strings currently in PASTA prepared by Sven.
- New checksum check.
- Schedule for rolling out new checks. Relatively quick turnaround. Every 6 months? Test on dev server.
- Some checks are for enabling better reporting.
- New version of EML by end of year?
- EML is in Github. See issues there.
- Currently no EML field to indicate a project is completed or ongoing. Closest fit is maintenance field.

Visioning for IM – led by John Porter

- Common tools, scalability, crossing disciplines.
- Education for users, each other.
- How much work would it take to fully annotate a dataset for machine use?
- Could crowd sourcing play a role?
- Current standardizations: DEIMS, repositories using EML, etc.
- An out-of-the-box IM system?
- Achieving agreement is one of the greatest challenges.
- Arctic Data Center is developing a web-based form for creating EML (based on Metacat framework).
- Compare ROpenSci community efforts.

New LTER Site IM's present their data management plans

Beaufort Sea Lagoon Ecosystems (BLE) – Tim Whiteaker

- Lagoon ecosystems along north coast of Alaska.
- Completely covered with ice in winter. Large input of fresh water in spring. Generally ice-free in summer.
- Freshwater inflows, coastal erosion, ice scour, ocean mixing, climate change.
- Sample water columns. Subsurface moorings. USGS stream gages. Climate stations. Camera towers.
- Use modeling esp. for winter processes.
- Looking for IM building blocks. Start simple.
- Project starts on 1 Aug 2017. Data collection starts Aug 2018. First dataset in 2019.
- PIs will handle QA/QC and local data management.
- IM will handle EML, data submission, website. Hire RA next year.
- Lots of data. Simulation modeling, camera imagery, underwater video. Maybe 50T/year.
- Data formats: CSV, netCDF, SHP?
- How to gather metadata from PIs?
- EML generation: Excel template, GCE toolbox, Morpho, R?
- How to organize data files?
- Website: static HTML?
- Integration with PASTA API: no local catalog.

Northern Gulf of Alaska (NGA) – Rob Bochenek

- Project starts 1 Oct 2017.
- Axiom is a small company that develops IM and CI technology for environmental, biological, and geo-scientific data.
- Leverage applications, systems, and hardware across partners.
- Goal: reach critical mass.
- Machine-produced data is reasonably standard (sensors, models, GIS).
- Human-produced data are more complex.
- Ontological database serves as broker to tie components together.
- Curation & publication path to NCI-DataOne.
- Server-side R and Python numeric workflows (Jupyter)
- Numerical workflows.
- Use Jupyter notebooks in data processing pathway.
- Issues: ISO -> EML conversion, getting data packages into PASTA.
- Pathway to DataONE exists. Is that sufficient?
- Different NSF program officers have different requirements: PASTA, Arctic Data Center, BCO-DMO.

Northeast U.S. Shelf (NES) Stace Beaulieu (No archived presentation)

- Understand how planktonic food webs change through time and space in response to changes in physical environment.
- Focal area extends south from Cape Cod. Larger area extends from North Carolina to Nova Scotia.
- Eastern boundary is OOI Pioneer Array.
- Several high-volume datasets: imagery, model output.
- Use Github as repository for workflows.
- Data in PASTA. Metadata in DataONE, BCO-DMO, NOAA NEIS.
- Metagenomic data in other databases such as Genbank. How to indicate availability?
- Data inputs: automated sensors, technicians, researchers, GIS & remote sensing.
- BCO-DMO is based on Drupal.
- Project starts 1 Sep 2017.
- Need to make a choice of local IM system.
- LTER network has lost its 2 Drupal champions.

Wrap-up and Action Items

Wrap-up
Notes should be loaded to Figshare for ESIP meeting. ESIP expects us to share our presentations.
Meeting report/annual report to EB are required products from this meeting.

Action items:
Revising the ToR
Databits will be published
IMEXEC will be formalized about announcing meetings by posting agenda a week in advance

Adhoc groups:
History of LTER IM book - Eda; will meet Wednesday Lunch
EML and cross-referencing pubs and dbs – Margaret
DIEMS has poster at science at art (Wed 7-9pm)
IMC dinner tonight at Irish Lion Pub across from Hyatt. Outside at 7 pm. Pay individually
ECC Tuesday dinner at Anatolia

Supporting Documents

Detailed Agenda

Agenda 2017 July 24 LTER Information Management Committee (IMC), co-located with ESIP

Time	Oak Room in IU Memorial Union building	Persons leading	Notes here on the right are not necessarily lined up with a time schedule.
8:00 AM	open the meeting, agenda, logistics	Gastil, Downing	Assign note-takers and facilitators to sessions

8:10 AM to ...	LTER-NCO discussion
	
	Updates from NCO
	Introduce Julien
	
	Questions from IMs or IMC for NCO
		Mark Sch, Julien	NCO outline topics
			
			status of network databases
			● BibDB
			● PersDB & SiteDB (archived, frozen, RFC period ended, next?)
			● ClimHyDB ... or is that EDI?
			
			whole-LTER website (not im.lternet)
			  Status:
			● Chose an Web Design firm to contract, named NDIC
			● See June 2017 Water Cooler presentation
			● Content is migrating now. Frozen July 15.
			  What involvement does NCO need from sites, thru site IMs?
			● recent request for input on logo
			● recent survey (not just web; communications too)
			● Mary Martin will update on this in her WG update.
			
			status of IM involvement in data synth WG
			● Mark Sch (and/or Julien) with Wade, to confer, outcome from April meeting.
			● Involvement of site IMs?
			
			DataBits, if it were to resume,
			● option: blog entries in larger site of NCO or EDI with just DataBits tag
			● NCO not offering to host and maintain a separate drupal or wp install just for DataBits but willing to receive tagged entries in existing blog.
			● Wider distribution / chance encounters by non-IMs if part of larger blog
			
			Migration of im.lternet.edu
			● WiReD working group
			● Some current content will be adopted by EDI
			● will/or not be migrated to new Drupal versions? By NCO?

... to 9:20 AM	LTER-EDI discussion
	
	Questions to ask EDI
	&
	EDI Questions for IMC
		Corinna, Margaret, Kristin, Duane, Colin	1. from Suzanne:
			a. EDI - the data citation fmt saying EDI not LTER
			  Response from EDI:
			    Using ESIP citation guidelines
			    Recommend Site Name included in dataset title
			b. EDI advisory group: how to bring up proj/issues?
			  (ie. a NISAC-like advisory committee)
			c. new congruency checks (ie checksum, intelRights)
			
			2. from Margaret:
			a. The conversation could take the form of 'here is what EDI is planning; does this fit with the IMC's plans?'
			b. particularly, we are thinking about things like databits. EDI has a newsletter, but some components could be left off if they overlap with DataBits content.
			C. Units EDI will handle. There will be an RFC from EDI.
			D. Vocab plan TBD. EDI will discuss with J.Porter this week. Possible EDI RFC.
			
			3. from Corinna:
			a. Report from the LTER, EDI, NCO, DataONE workshop in April.
			  EDI and NCO collaborative approaches to updating databases (request for comments - siteDB already done, next climDB?)
			c. We like your suggestion of how IMs can provide input to EDI
			d. Report from EcoCom design pattern workshop (upcoming VTC after ESIP and Margaret's upcoming blog post on the EDI website) Suzanne Remillard and Hope Humphries participated in that workshop and could report.
			e. Discuss priorities for next Data Package Design Pattern workshop
			f. Report on new congruence checker rule implementations
			g. Discuss in person training workshop (planned for fall 2017)

9:20 - 9:30 AM	speed-talk 30 second advertisement of ESIP Sessions being done by IMC and EDI people		N sites x 1 minute (30 sec setup + 30 sec ad) Skipped due to time over-run.

Time	Activity	Notes	Details
9:30 AM	short break		
9:45 - 10:45 AM	NSF call with Peter McCartney, including pre-discussion and post-discussion	facilitated by Wade	Is NSF planning a new way of reporting pubs for annual reports using DOIs that would cross reference funding codes?
10:45	Group Photo		Courtyard near Oak Room
10:50	short break		

11:00 - noon	WG updates (2* x 5 min each = 10 min total) Plan for next year's WGs
	
	Business Meeting ... and voting
		CLOSED to non-LTER people
		
		all other times are OPEN
			WG Updates from
			    WiReD () speaker: Brian H
			    *ECC ('checker') speaker: Margaret O (was already part of EDI section)
			    Site Website Guidelines speaker: Mary M
			
			Voting via Survey Monkey. 1 vote per site. Anonymous. Need ⅔ sites for quorum.
			
			See candidate statements on im.lternet.edu.
			
			IMExec 3 year term, 2year term (decided 3-year)
			DataBits Editor: Eda, Assistant Editor: 2 indicated interest

12:00	lunch one hour		
1 pm - 1:05	Lightning Talks 2 minutes each		J. Porter, W.Sheldon, L.Kui & M.O'Brien
1:05 - 2:00	Intro topics then split into small groups Option A: ToR updates Option B: Vision for IM Option C: PASTA ECC new checks	A: led by Don, Eda B: led by JP C: led by Margaret	A: outside in courtyard ToR - should IM Exec terms be 2 years or 3. How aligns with EDI and NCO grants ending 2 years from now? B: outdoors, downstairs C: in Oak Room
2:00 PM	short break		If report-back has slides, load on macbook during break please.
2:55 - 3:55 pm	Break-out report-backs		
3:55 PM	short break		

4:00 pm - 5 pm	1. Stace, Rob, Tim, new-site IMs present their Data Mgmt Plan and/or IM System (15 minutes each) 2. Wrap up. Action Items summary. Announce any newly scheduled ad hoc meetings during ESIP days. 3. Remind/update logistics for dinner (5 min)	1. Stace, Rob, Tim, 2. Gastil (wrap-up) 3. Jason	
7:00 PM	dinner at The Irish Lion	Downing	Include celebration of Mark Schildhauer's retirement.

IMC Annual Meeting Overview Slides

2017 LTER IMC @ ESIP

Coherent as a Team while Open to wider Community

Agenda Overview

- IMC-NCO

- IMC-EDI

- Commercial break - Ads for ESIP sessions

- IMC-NSF

- Photo op

- WG Updates & Business Meeting   *

- Lunch

- Lightning ⚡ Talks

- Break-outs

- Break-out report-backs

- New Site IMS presentations

- Wrap-up

- Dinner at The Irish Lion

Logistics

Note-taker volunteers each section of agenda

Scout for photo location VERY NEAR here. Lawn?

Breaks and lunch are short

Raise topics today

More depth later this week

Use Application Programming Interfaces of Data Repositories to Create Local Data Catalogs
Duane Costa • Stevan Earl • Gastil Gastil-Buhl • John Porter

Tools and Strategies for End-to-End Sensor Data Management
Matthew Bartos • Mike Botts • Mike Daniels • Wade Sheldon

Networking Break

Examples - See Sched online

Sustainable Data Management, Return on Investment
Margaret O'Brien • Anne Wilson

Developing an Information Management Code Repository
Colin Smith • Kristin Vanderbilt

Networking Break

Create EML Using R and Share the Script via the Information Management Code Repository
Corinna Gries

Research as Art
ESRI

Wh-wh-which Website WG??!

- Whole-LTER website
    - www.lternet.edu
    - NCO, consulting with Guidelines WG
    - Marty
- Guidelines for LTER sites' local websites
    - *i.e.* http://bnz.lternet.edu, knz.lternet.edu, …
    - Guidelines WG
    - Mary & Marty
- The IMC's own website
    - http://im.lternet.edu/
    - WiReD WG
    - Brian

Interface between

LTER - NCO

Updates, Q&A

status of network databases
- ● BibDB
- ● PersDB & SiteDB (archived, frozen, RFC period ended, next?)
- ● ClimHyDB ... or is that EDI?

whole-LTER website (not im.lternet)
  Status:
- ● Chose an Web Design firm to contract, named NDIC
- ● See June 2017 Water Cooler presentation
- ● Content is migrating now. Frozen July 15.
  What involvement does NCO need from sites, thru site IMs?
- ● recent request for input on logo
- ● recent survey (not just web; communications too)
- ● Mary Martin will update on this in her WG update.

status of IM involvement in data synth WG
- ● Mark Sch (and/or Julien) with Wade, to confer, outcome from April meeting.
- ● Involvement of site IMs?

DataBits, if it were to resume,
- ● option: blog entries in larger site of NCO or EDI with just DataBits tag
- ● NCO not offering to host and maintain a separate drupal or wp install just for DataBits but willing to receive tagged entries in existing blog.
- ● Wider distribution / chance encounters by non-IMs if part of larger blog

Migration of im.lternet.edu
- ● WiReD working group
- ● Some current content will be adopted by EDI

Ongoing dialog

Ad-hoc discussions here
this week

Monthly IMC Water
Cooler vtc online

NSF Call

Questions to ask:

1 Funding timeline for NCO, EDI?

2 Is NSF planning new publication reporting process for proposals?

3 Renewal RFP timeline?

4 How will site IM system guidelines be used?

Post-Discussion points:

1

2

Working Group Updates

- Site Guidelines
  - Kristen thru IMS Guidelines, now Mary & Marty for Site Website Guidelines
- WiReD
  - Brian
- EML Congruence Checks
  - Margaret

IMC Business Meeting

Closed session

Candidate Statements (see speaker's notes)

Any other issues need votes?

    DataBits editor & assistant?

    ToR is a break-out later. Can RFC it after meeting.

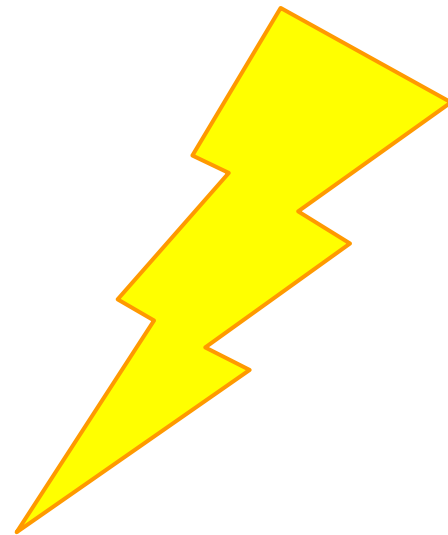# Lightning Talks

John Porter   R, EML and populating <taxonomicCoverage>

Li Kui & Margaret O'Brien   miniMetabase and R

Andrews (Don or Suzanne)   History or sitebyte (respectively)

Wade Sheldon   EML Packages from the GCE Data Toolbox
http://gce-lter.marsci.uga.edu/public/file_pickup/toolbox_demo/EML_GCE_Toolbox.pdf

Break-outs …. And report-backs

A. ToR Updates
    a. Don, Eda
    b. Note Taker:
B. IM Vision
    a. Lead by… John Porter
    b. Note taker: John Porter
C. PASTA EML Congruence Checker new checks
    a. Lead by Margaret.
    b. Note taker:

New LTER Sites' IM Systems

Stace Beaulieu at NES LTER Northeast U.S. Shelf

Tim Whiteaker at BLE LTER Beaufort Lagoon Ecosystem

Rob Bochenek at NGA LTER Northern Gulf of Alaska

Wrap-up & Logistics

Announce any new ad-hoc sessions this week.

Any travel reimbursement info to announce?

   Need food, shuttle receipts?

Dinner at The Irish Lion

"All those in favor of saying 'It's a wrap' instead of 'Meeting adjourned' say 'Aye.'"

Thank you for attending. Enjoy the ESIP week!

Lightning Talk (1) - Taxonomic Coverage – John Porter

Taxonomic Coverage

- Only about 600 of the ~6,000 EML-documented datasets include a taxonomic coverage
  - Taxonomic coverages may be as simple as a single species name, or an entire taxonomic tree
- There are now tools that can help translate simple species lists into taxonomicCoverage elements
  - The R EML package – handles formatting
  - The R Taxize package – handles taxonomic tree
- Why not combine them?

A Revised set_taxonomy() function

□ Bryce Mecum at NCEAS has been working on revisions to the R EML package that includes integrating Taxize

□ With that revised code, I've been able to write simple R code that:

- ■ Reads in a list of taxonomic names
- ■ Finds "accepted" names
- ■ Provides taxonomicCoverage elements including the hierarchy

The "snippet" can then be integrated into EML documents

```xml
<taxonomicClassification>
  <taxonRankName>kingdom</taxonRankName>
  <taxonRankValue>Animalia</taxonRankValue>
  <taxonomicClassification>
    <taxonRankName>subkingdom</taxonRankName>
    <taxonRankValue>Bilateria</taxonRankValue>
    <taxonomicClassification>
      <taxonRankName>infrakingdom</taxonRankName>
      <taxonRankValue>Protostomia</taxonRankValue>
      <taxonomicClassification>
        <taxonRankName>superphylum</taxonRankName>
        <taxonRankValue>Ecdysozoa</taxonRankValue>
        <taxonomicClassification>
          <taxonRankName>phylum</taxonRankName>
          <taxonRankValue>Arthropoda</taxonRankValue>
          <taxonomicClassification>
            <taxonRankName>subphylum</taxonRankName>
            <taxonRankValue>Crustacea</taxonRankValue>
            <taxonomicClassification>
              <taxonRankName>class</taxonRankName>
              <taxonRankValue>Malacostraca</taxonRankValue>
              <taxonomicClassification>
                <taxonRankName>subclass</taxonRankName>
                <taxonRankValue>Eumalacostraca</taxonRankValue>
                <taxonomicClassification>
                  <taxonRankName>superorder</taxonRankName>
                  <taxonRankValue>Eucarida</taxonRankValue>
                  <taxonomicClassification>
                    <taxonRankName>order</taxonRankName>
                    <taxonRankValue>Decapoda</taxonRankValue>
                    <taxonomicClassification>
                      <taxonRankName>suborder</taxonRankName>
                      <taxonRankValue>Pleocyemata</taxonRankValue>
                      <taxonomicClassification>
                        <taxonRankName>infraorder</taxonRankName>
                        <taxonRankValue>Brachyura</taxonRankValue>
                        <taxonomicClassification>
                          <taxonRankName>superfamily</taxonRankName>
                          <taxonRankValue>Ocypodoidea</taxonRankValue>
                          <taxonomicClassification>
                            <taxonRankName>family</taxonRankName>
                            <taxonRankValue>Ocypodidae</taxonRankValue>
                            <taxonomicClassification>
                              <taxonRankName>genus</taxonRankName>
                              <taxonRankValue>Uca</taxonRankValue>
                              <taxonomicClassification>
                                <taxonRankName>species</taxonRankName>
                                <taxonRankValue>Uca pugnax</taxonRankValue>
                              </taxonomicClassification>
                            </taxonomicClassification>
                          </taxonomicClassification>
                        </taxonomicClassification>
                      </taxonomicClassification>
                    </taxonomicClassification>
                  </taxonomicClassification>
                </taxonomicClassification>
              </taxonomicClassification>
            </taxonomicClassification>
          </taxonomicClassification>
        </taxonomicClassification>
      </taxonomicClassification>
    </taxonomicClassification>
  </taxonomicClassification>
</taxonomicClassification>
```

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Scope	Identifie	Revision	Title	beginDate	endDate	durationYears	entities	attributes	keywords	geographi	taxonomi	taxonomicSpecies
2	knb-lter-and	2719	6	Invertebrates of the Andrews Experime	6/1/1971	3/11/2002	30.7761807	1	16	7	2	5012	2513
3	knb-lter-and	2739	7	Spatial and temporal distribution and a	1/2/1994	10/15/2008	14.78439425	2	12	3	261	3274	2188
4	knb-lter-and	2720	8	Vascular plant list on the Andrews Expe	1/1/1958	1/1/1979	20.99931554	2	38	3	3	1108	666
5	knb-lter-and	3217	11	Plant succession and biomass dynamics	1/1/1962	7/31/2013	51.57837098	5	109	12	197	547	255
6	knb-lter-and	4346	6	Aquatic insect sampling in Lookout Cre	6/22/2001	9/20/2001	0.246406571	3	27	4	2	513	241
7	knb-lter-and	3136	5	Spatial and temporal distribution and a	1/2/1994	1/2/1996	1.998631075	1	9	3	2	315	214
8	knb-lter-and	2721	6	Bird species list for the Andrews Experi	1/1/1975	1/1/1995	20	1	24	4	2	303	164
9	knb-lter-and	2727	7	Riparian bryophyte list of the Andrews	7/1/1994	4/1/1995	0.750171116	1	8	5	2	247	149
10	knb-lter-fce	1159	2	Diatom Species Abundance Data from L	12/9/2006	3/28/2008	1.300479124	1	150	22	47	146	146
11	knb-lter-and	4041	11	LTER Intersite Fine Litter Decompositio	1/31/1990	6/12/2007	17.36071184	12	211	7	30	298	131
12	knb-lter-and	4547	5	Epiphytic macrolichens in relation to fo	9/23/1997	9/15/1999	1.976728268	4	44	13	119	208	127
13	knb-lter-and	4054	7	Plant biomass dynamics following loggi	7/9/1979	8/20/2014	35.1156742	6	67	3	3	321	119
14	knb-lter-and	2728	7	Epiphyte species list of Watershed 10,	1/1/1970	1/1/1972	1.998631075	1	13	3	2	180	104
15	knb-lter-and	3189	10	Stream cross-section profiles in the An	4/14/1978	9/6/2011	33.3963039	6	53	8	5	266	99
16	knb-lter-sbc	99	1	SBC LTER: Reef: Kelp Forest Community Dynamics: Kelp Forest Data to sup			-1	2	23	8	11	168	84
17	knb-lter-pie	175	5	Surveys of salt marsh breeding birds us	6/9/2004	7/10/2015	11.08281999	1	17	9	7	166	83
18	knb-lter-pie	349	1	Surveys of salt marsh wading birds and	5/25/2012	7/15/2014	2.138261465	1	21	10	2	166	83
19	knb-lter-and	3192	9	Fire history database of the western Ur	1/1/1993	1/1/1996	2.997946612	3	41	3	623	166	73
20	knb-lter-fce	1108	3	Macrophyte count data collected from	9/11/2006	9/17/2008	2.01779603	1	69	19	37	58	58
21	knb-lter-and	2732	7	Mycorrhizal belowground fungi species	6/3/1991	6/30/1994	3.074606434	2	20	3	10	141	56
22	knb-lter-hfr	222	3	Vegetation Cover in the Clearcut Site a	6/1/2010	7/12/2013	3.112936345	1	12	14	1	101	55
23	knb-lter-and	2733	6	Lichen abundance and biodiversity alor	1/1/1993	1/1/1993	0	1	11	7	1	81	55
24	knb-lter-fce	1164	6	Seasonal Electrofishing Data from Rook	11/1/2004	5/5/2014	9.505817933	1	76	28	19	54	54

EMLStats.py summarizes elements of documents not available through other sorts of searching (Duration, #entities, #attributes, #keywords, #coverage elements)

Lightning Talk (2) - Mini-metabase for generating EML – Li Kui

Lightning Talk (3) - EML Packages from the GCE Data Toolbox

Wade Sheldon

EML Packages from the GCE Data Toolbox
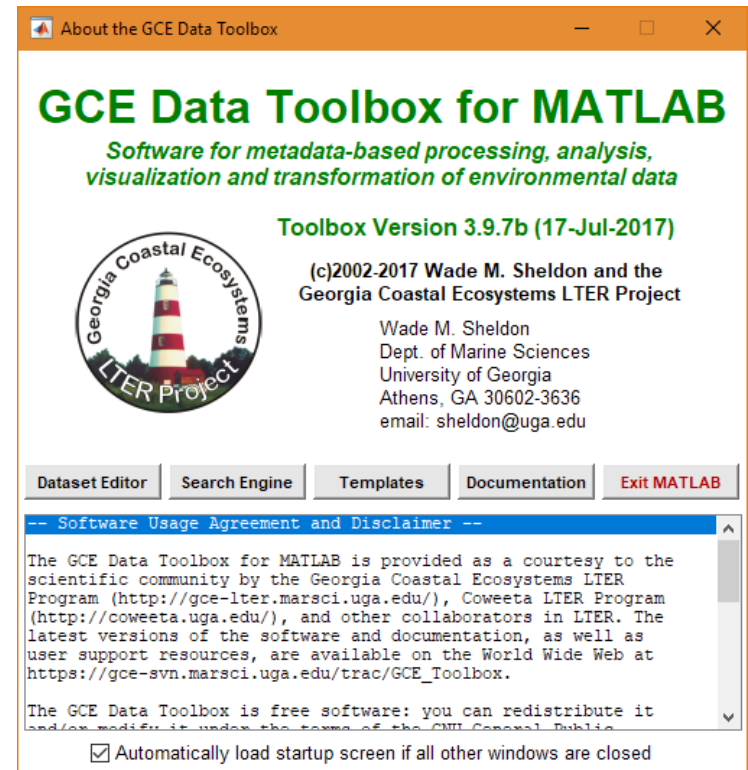
Wade Sheldon

Georgia Coastal Ecosystems LTER

University of Georgia

email: sheldon@uga.edu

Background

- Generalized MATLAB package for metadata-based data management

- Documentation, attribute, Q/C rule metadata managed in templates

- Metadata applied on import to document and Q/C data

- Ideal for sensor data, but useful for any tabular/SQL data source

EML Export – Metabase

- Toolbox can now register new data sets in a Metabase instance

- Top-level metadata uploaded via JDBC to create ID/Accession

- Entity files generated and metadata uploaded via JDBC

- Metabase catalog application/services generate PASTA EML

EML Export – Standalone

- Toolbox can also create EML packages for PASTA natively
- Generates complete EML docs with physical and attribute meta
- Can also generate dataTable EML fragments for adding to EML docs
- Q/C flags automatically documented as coded text columns
- Unit dictionary unit mapping or auto custom units with STMML

More Info

▸ GCE Data Toolbox website:
https://gce-svn.marsci.uga.edu/trac/GCE_Toolbox

▸ Support resources:
https://gce-svn.marsci.uga.edu/trac/GCE_Toolbox/wiki/Support

▸ Toolbox is FOSS (but requires MATLAB)

Breakout Group (1) – Terms of Reference (ToR) updates

LTER Information Management Committee Terms of Reference (ToR)

- Governance procedures have changed since ToR endorsed in 2011

- ToR overview, purpose, IMC membership, voting, elections, working groups, meeting rotation are still fairly consistent

- IMExec membership, Network roles, and budget and expenses have changed

IMC Meeting 24 July 2017, Bloomington, IN

IMExec membership

- 5 members plus 2 co-Chairs
- 3 year terms
- Ex-officio
  - IM rep to Exec Board (3 year term)
  - LNO Chief Information Officer (James Brunt)
  - LNO Information Manager
  - NISAC rep when discussion warranted
- Co-Chairs ex-officio to Science Council

Working Groups

- Ad hoc and formal working groups
- Formal status achieved by developing a ToR
- Active groups post roster of membership, roster of actions, file annual report
- Disbanded WGs provide final report
- IMExec assigns a liaison to each WG

Network Roles

- Elected representative to the EB
  - Wade Sheldon is current rep
- Membership in Network Information System Advisory Committee (NISAC)
  - 4 or 5 members and included an IM co-Chair
  - 3 year terms
  - Committee has been dissolved

Meetings and Budget

- Annual meetings on 3 year rotation
  - 1) IMC alone, 2) part of larger conference (i.e., ESIP or EIMC, 3) in conjunction with the All-Scientist Meeting (ASM)
- "Sites are required to send a representative"
  - LNO provides funding for IMC, IMExec winter meeting, and other working group meetings

Discussion / Questions

- Should we officially assign 5 members to IMExec? Or, limit the term length?

- Should we consider ex-officio representation on IMExec from NCO or EDI?

- Do we include a representative to the EDI Advisory Committee?

- Are IMs still "required" to attend the annual meeting?

- What is the general fate of IMC and IMExec meetings? Is there a continuing budget?

- What are the expectations for maintaining working groups?

LTER Information Management Committee Terms of Reference (ToR)

Should we officially assign 5 members to IMExec?

- YES

- Steven and Jonathan think 5 members works well… harder to coordinate 7 people

- 4 IM Exec members + 1 Chair… if a co-chair is needed, it would come from one of the other 4

- What about EB rep? Can we make that person also on IM Exec with the same 3 year term?

Should we consider ex-officio representation on IMExec from NCO or EDI? Do we include a representative to the EDI Advisory Committee?

- Currently Gastil takes a walk with Marty (NCO)
- How does NCO see their relationship with LTER Information Managers?
- How do we establish a more formal relationship? through ex-officio appointment with IM Exec? we should formalize this in ToR.
- How do EDI and LTER IM Exec communicate? let's formalize this also
- If EDI is offering an advisory board spot to an LTER IM, then we should decide the representation as a group, like we did with NISAC (then, IM Exec approves...)
- Does EDI actually want an LTER IM on their advisory board?

Are IMs still "required" to attend the annual meeting? What is general fate of IMC and IM Exec meetings? Is there a continuing budget?

- we strongly recommend/encourage IMs to attend IMC in person… otherwise remote in.
- but we are going to let go of the in-person IM Exec winter meeting, which historically was all about planning for IMC summer meeting
- IM Exec has monthly VTC meetings
- Re: funding issues
  - There is still funding for ASM….
  - IM Exec should add this as a responsibility --> pursue NSF supplements
  - Also, look for cheap options to meet; are sites willing to cost-share? (e.g., pay for lodging)

What are the expectations for maintaining working groups?
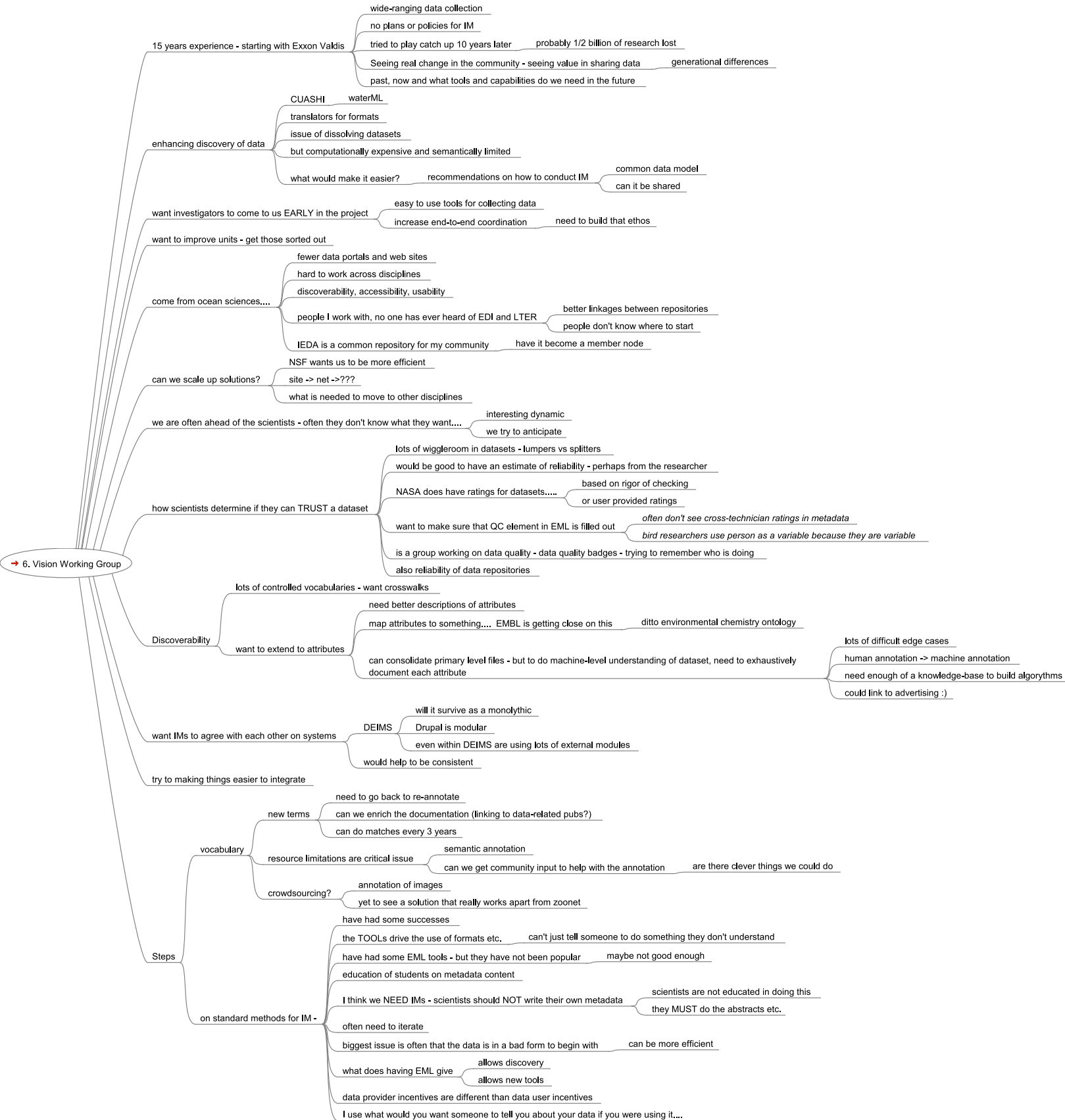
- we recommend report outs at annual IMC meeting as well as written report outs of working group duties and current membership on website
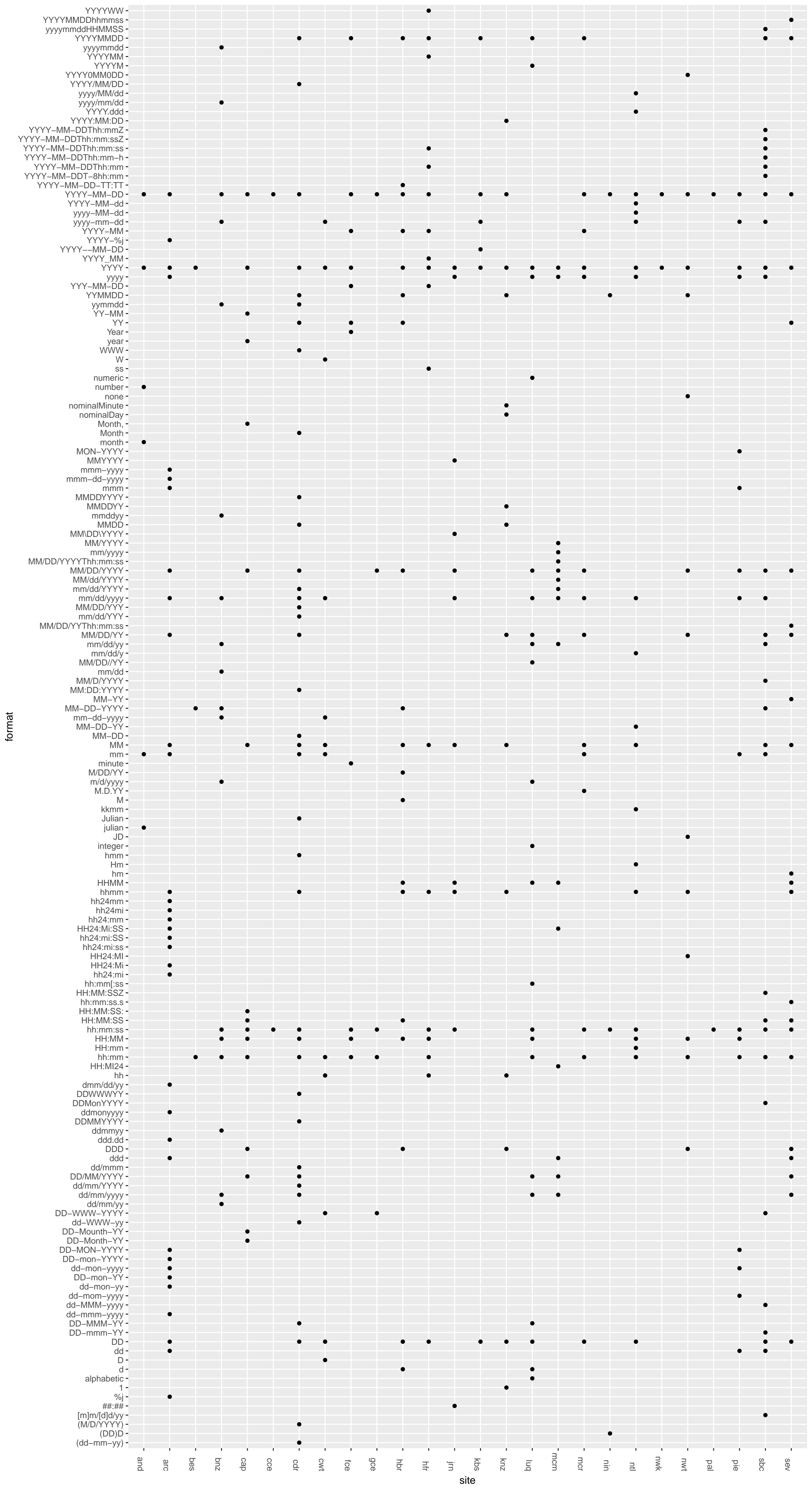
2 vs. 3 year terms on IM Exec?

- if we align EB member with IM Exec, then keep 3 year terms
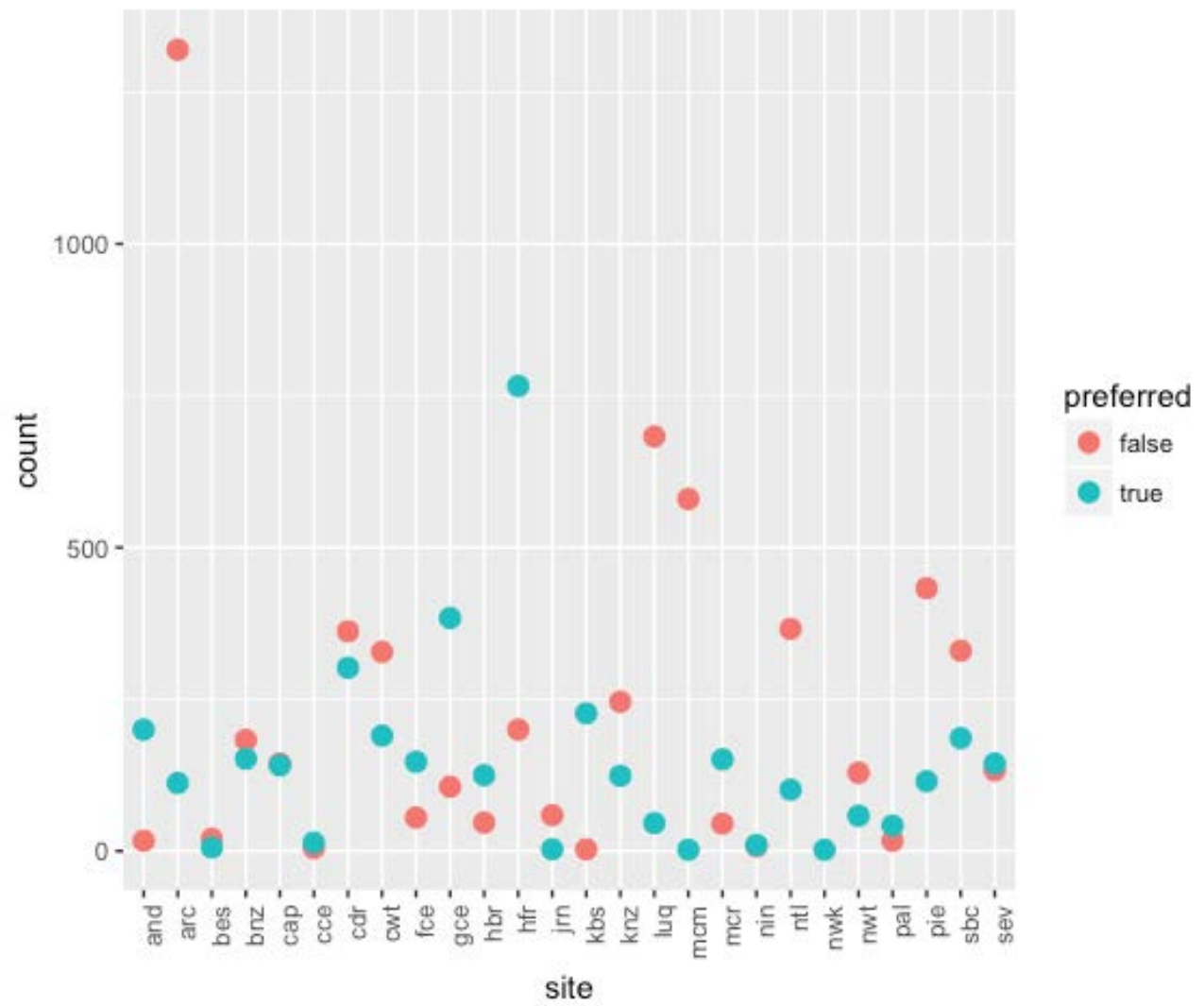- we recommend 3 year terms… Balance between continuity and allowing more folks to participate

Breakout Group (1) – Vision for IM

6. Vision Working Group

- **15 years experience - starting with Exxon Valdis**
  - wide-ranging data collection
  - no plans or policies for IM
  - tried to play catch up 10 years later
    - probably 1/2 billion of research lost
  - Seeing real change in the community - seeing value in sharing data
    - generational differences
  - past, now and what tools and capabilities do we need in the future

- **enhancing discovery of data**
  - CUASHI
    - waterML
  - translators for formats
  - issue of dissolving datasets
  - but computationally expensive and semantically limited
  - what would make it easier?
    - recommendations on how to conduct IM
      - common data model
      - can it be shared

- **want investigators to come to us EARLY in the project**
  - easy to use tools for collecting data
  - increase end-to-end coordination
    - need to build that ethos

- **want to improve units - get those sorted out**

- **come from ocean sciences....**
  - fewer data portals and web sites
  - hard to work across disciplines
  - discoverability, accessibility, usability
  - people I work with, no one has ever heard of EDI and LTER
    - better linkages between repositories
    - people don't know where to start
  - IEDA is a common repository for my community
    - have it become a member node

- **can we scale up solutions?**
  - NSF wants us to be more efficient
  - site -> net ->???
  - what is needed to move to other disciplines

- **we are often ahead of the scientists - often they don't know what they want....**
  - interesting dynamic
  - we try to anticipate

- **how scientists determine if they can TRUST a dataset**
  - lots of wiggleroom in datasets - lumpers vs splitters
  - would be good to have an estimate of reliability - perhaps from the researcher
  - NASA does have ratings for datasets.....
    - based on rigor of checking
    - or user provided ratings
  - want to make sure that QC element in EML is filled out
    - often don't see cross-technician ratings in metadata
    - bird researchers use person as a variable because they are variable
  - is a group working on data quality - data quality badges - trying to remember who is doing
  - also reliability of data repositories

- **Discoverability**
  - lots of controlled vocabularies - want crosswalks
  - want to extend to attributes
    - need better descriptions of attributes
    - map attributes to something.... EMBL is getting close on this
      - ditto environmental chemistry ontology
    - can consolidate primary level files - but to do machine-level understanding of dataset, need to exhaustively document each attribute
      - lots of difficult edge cases
      - human annotation -> machine annotation
      - need enough of a knowledge-base to build algorythms
      - could link to advertising :)

- **want IMs to agree with each other on systems**
  - DEIMS
    - will it survive as a monolythic
    - Drupal is modular
    - even within DEIMS are using lots of external modules
  - would help to be consistent

- **try to making things easier to integrate**

- **Steps**
  - vocabulary
    - new terms
      - need to go back to re-annotate
      - can we enrich the documentation (linking to data-related pubs?)
      - can do matches every 3 years
    - resource limitations are critical issue
      - semantic annotation
      - can we get community input to help with the annotation
        - are there clever things we could do
    - crowdsourcing?
      - annotation of images
      - yet to see a solution that really works apart from zoonet
  - on standard methods for IM -
    - have had some successes
    - the TOOLs drive the use of formats etc.
      - can't just tell someone to do something they don't understand
    - have had some EML tools - but they have not been popular
      - maybe not good enough
    - education of students on metadata content
    - I think we NEED IMs - scientists should NOT write their own metadata
      - scientists are not educated in doing this
      - they MUST do the abstracts etc.
    - often need to iterate
    - biggest issue is often that the data is in a bad form to begin with
      - can be more efficient
    - what does having EML give
      - allows discovery
      - allows new tools
    - data provider incentives are different than data user incentives
    - I use what would you want someone to tell you about your data if you were using it....

Breakout Group (1) – PASTA ECC new checks (datetime)

Preferred vs non preferred date format strings

New LTER Site IM's present their data management plan
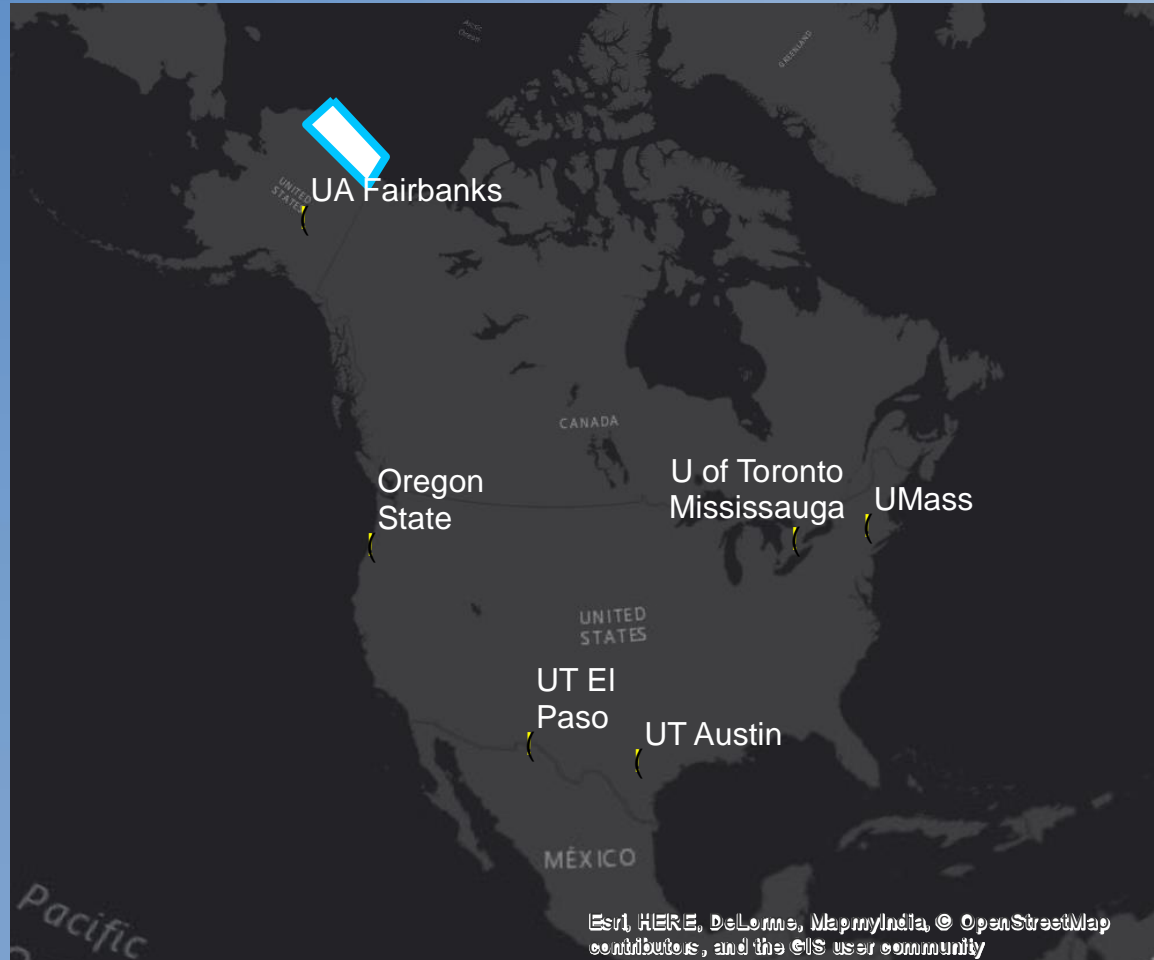
Beaufort Sea Lagoon Ecosystems (BLE) – Tim Whiteaker

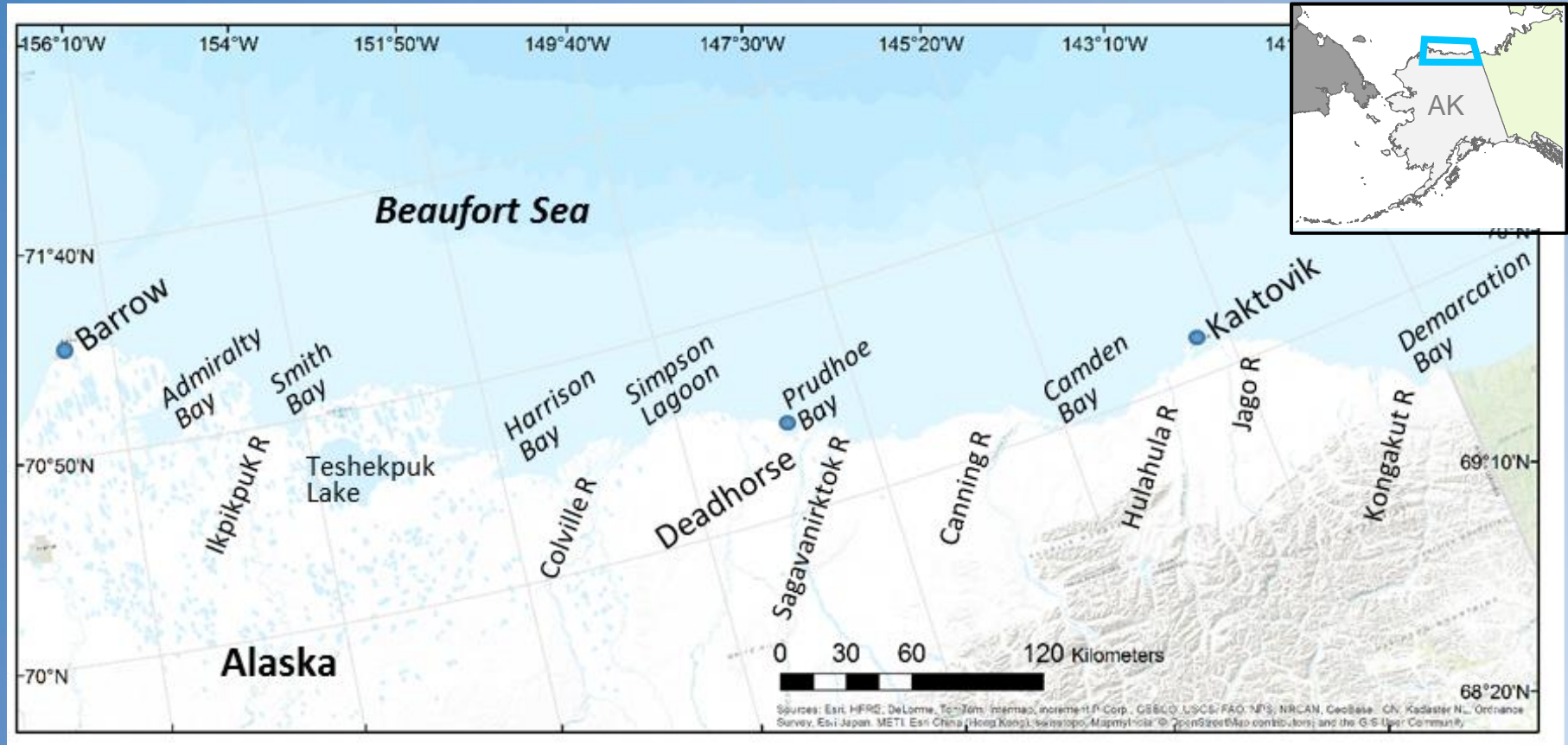# *Beaufort Sea Lagoon Ecosystems (BLE)*

Ken Dunton and Jim McClelland, UT-Austin
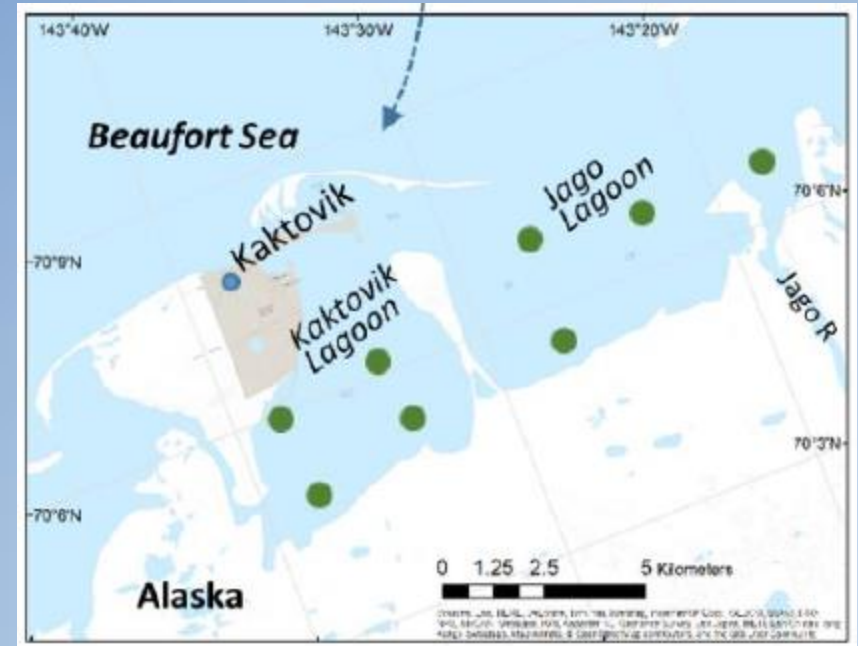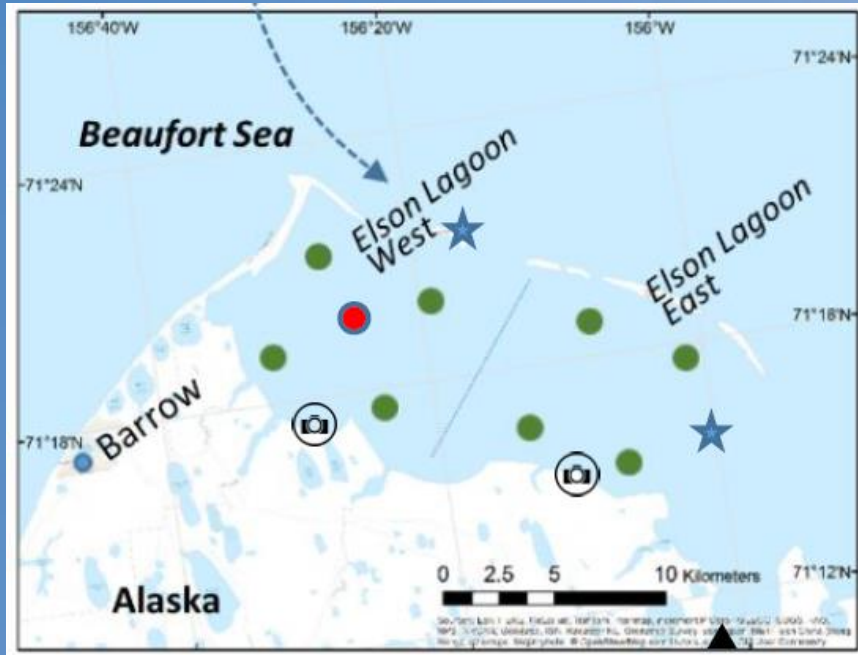IM – Tim Whiteaker

A Distributed Team

UA Fairbanks

Oregon State

U of Toronto Mississauga

UMass

UT El Paso

UT Austin

Targeting Lagoons Along the Beaufort Sea

- Three domains in the Western, Central, and Eastern Beaufort

Two lagoon systems within each domain

Ice-covered to Ice-free

Major Questions
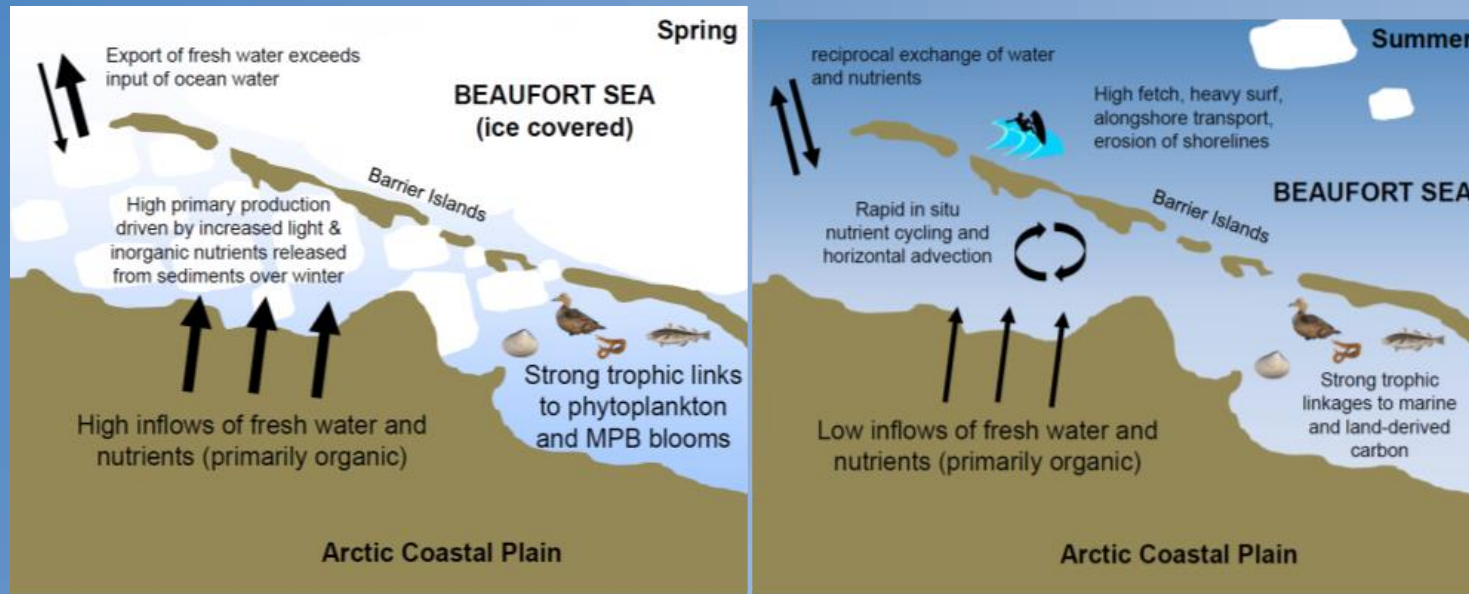
- How do interactions between freshwater inflows, coastal erosion, ice scour, and ocean mixing influence biogeochemical cycling and biotic community characteristics near the land-sea interface over seasonal to multi-decadal timeframes?

- Do temporal variations play a central role in determining trophic linkages, stability, and resilience of food webs?

A Coastal Research Nexus in the Arctic
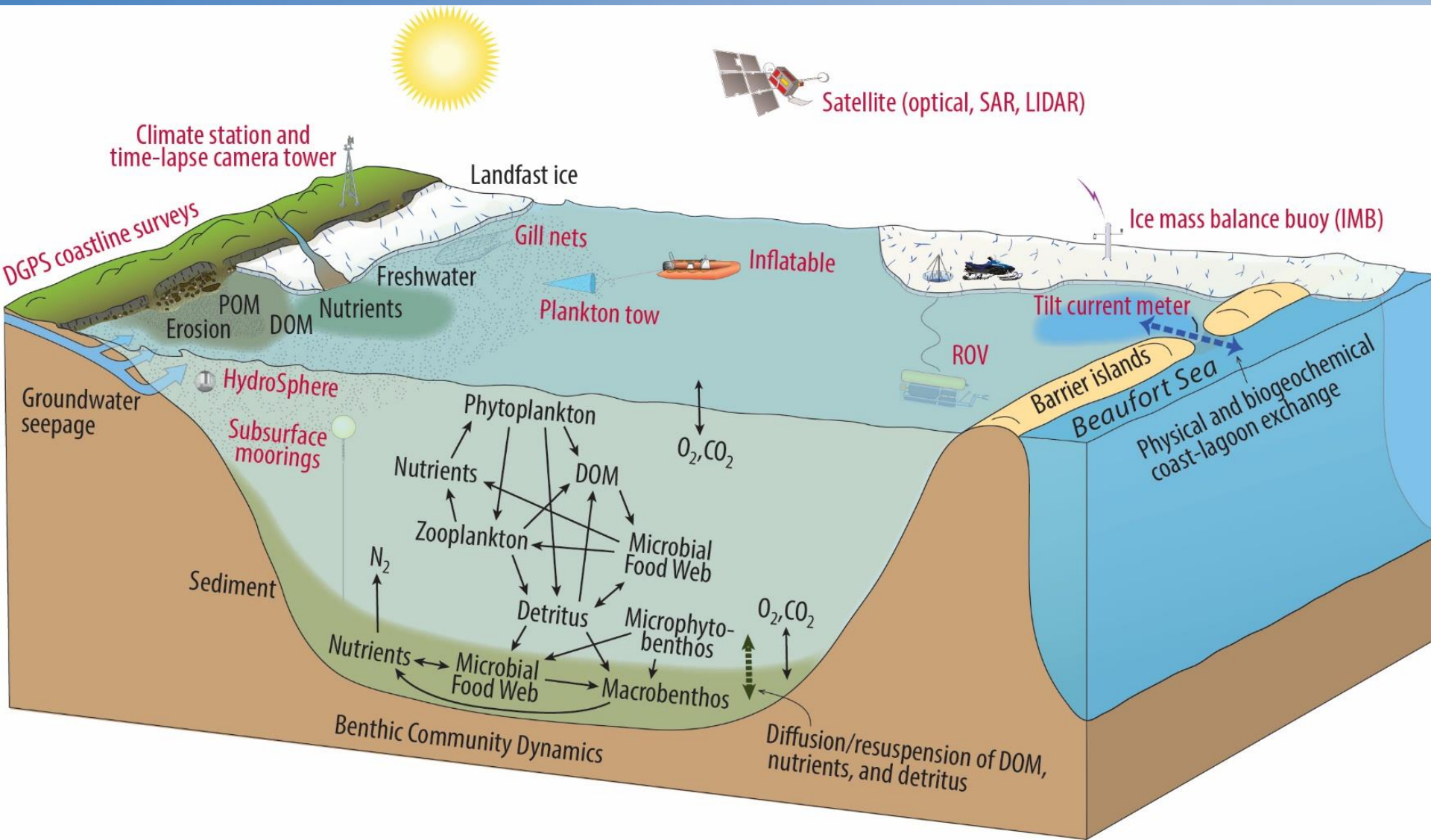
Information Management

"Playing in the Street" on Flickr by Sharon Drummond

Information Management

- Data collection starts August 2018
- PIs handle QA and local data mgmt
- PIs submit data for EDI archive to IM
- IM handles
  - EML
  - Data submission
  - Website

Where To Archive

- Local Archives
    - Model output
    - Images
    - Videos

    Metadata
    in PASTA?

- Uploaded to EDI
    - Sensor data
    - Sample results
    - Model result summary
    - Image/video analysis results
    - Digitized coastline
- Formats
    - CSV
    - netCDF
    - SHP?

Gathering Metadata

- Excel template?

- Maybe Cognito forms a la PAL and CCE

- Maybe ad hoc forms for each PI
  - set up prior to data collection?

EML

- Was originally thinking Excel template or GCE Toolbox

- Morpho seems nice if it will have long term support

- Currently thinking R since EDI supports it

File System

- How do you organize data files?

- Like PAL:
  - PI-1
    - Dataset-1
      - Date-1
      - Date-2
    - Dataset-2
      - Date
  - PI-2
    - Dataset
      - Date

Website – Static HTML?
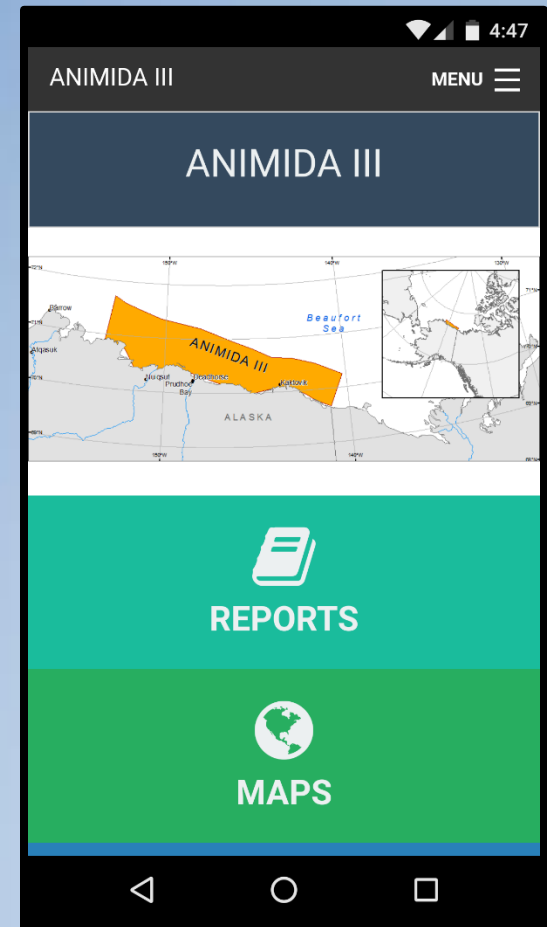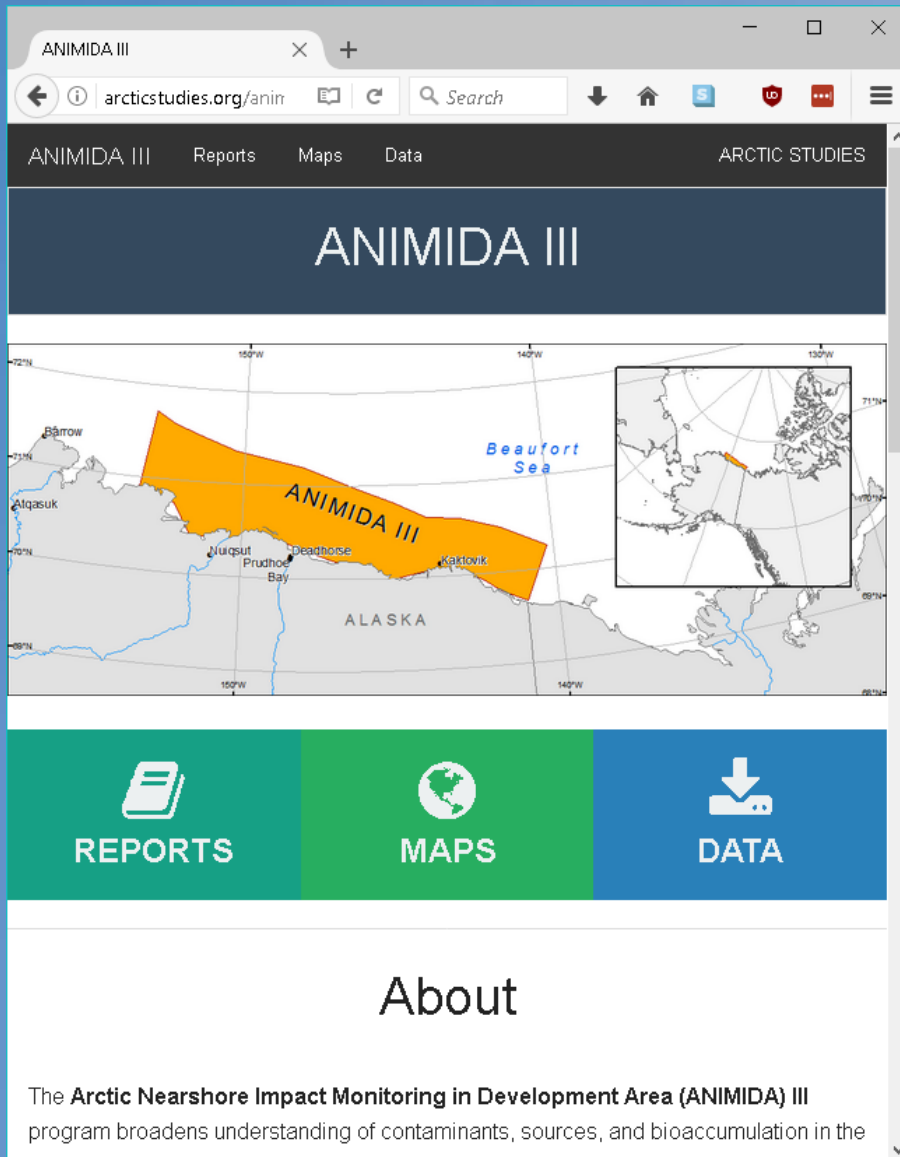
Integration with PASTA API

- No local catalog for me
- Can this be done with static HTML?
  - Access-Control-Allow-Origin: '*'
- DEIMS and PASTA? Datazoo?

Thanks!

New LTER Site IM's present their data management plan

Northern Gulf of Alaska (NGA) – Rob Bochenek

Northern Gulf of Alaska LTER

Rob Bochenek

Summary

- Introduction to Axiom
- Shared Cyber-Infrastructure Approach
- CI Data Flows
- CI Integration Architecture
- Data Integration Tools
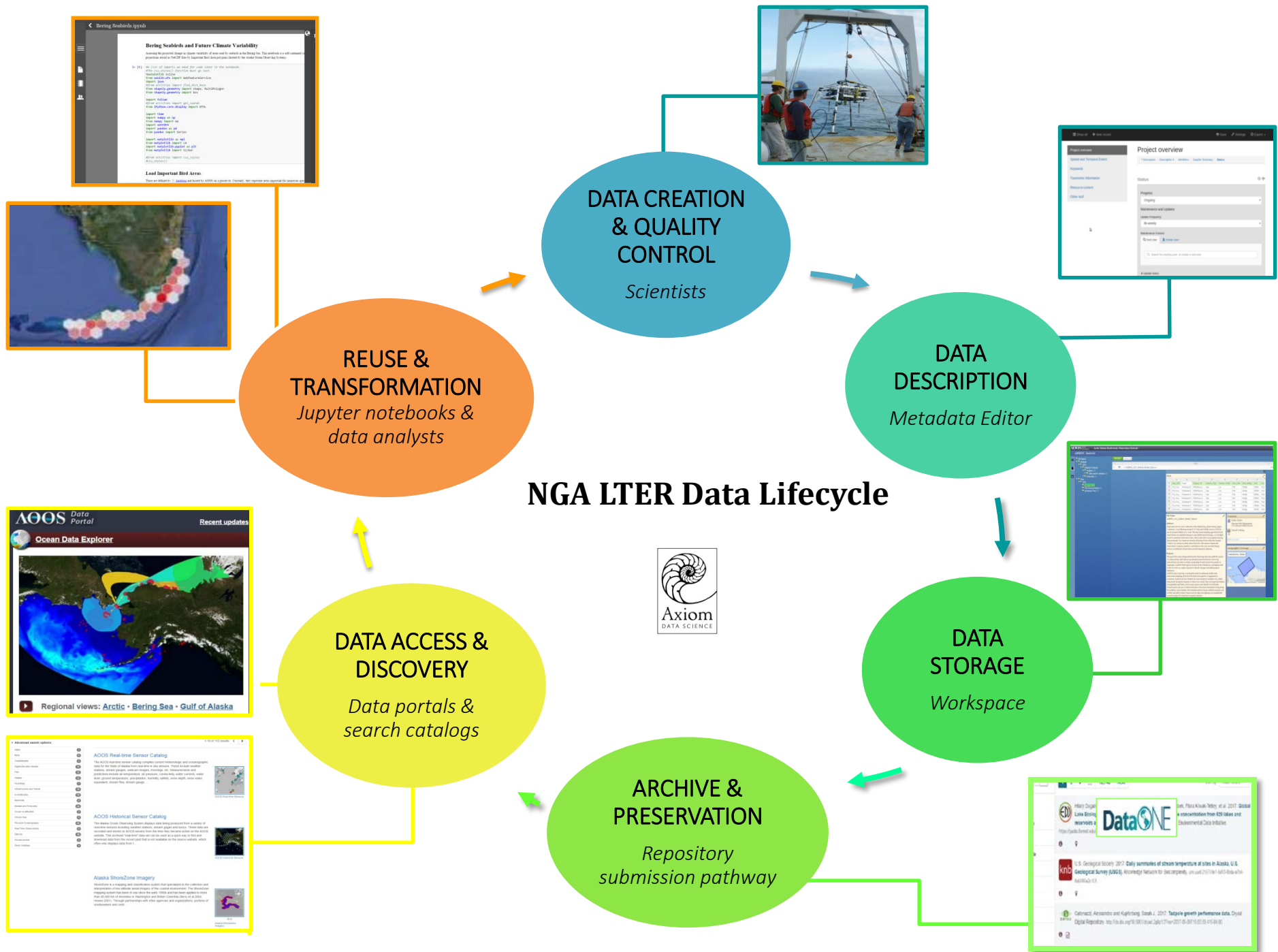- Demonstrate Data Management Tools

Who is Axiom

- Develop data management and cyber infrastructure technology for environmental, biological and geo-scientific data (numerical models, remote sensing, GIS, sensor networks and field campaigns)

- Small company (20 technical staff)

- AK Headquarters, OR and RI satellite offices

- Support several Federal, University and NGO Data Assembly Centers (IOOS, SECOORA, AOOS, BOEM, NAS, NSF, USGS Geomorphology Program, etc…)
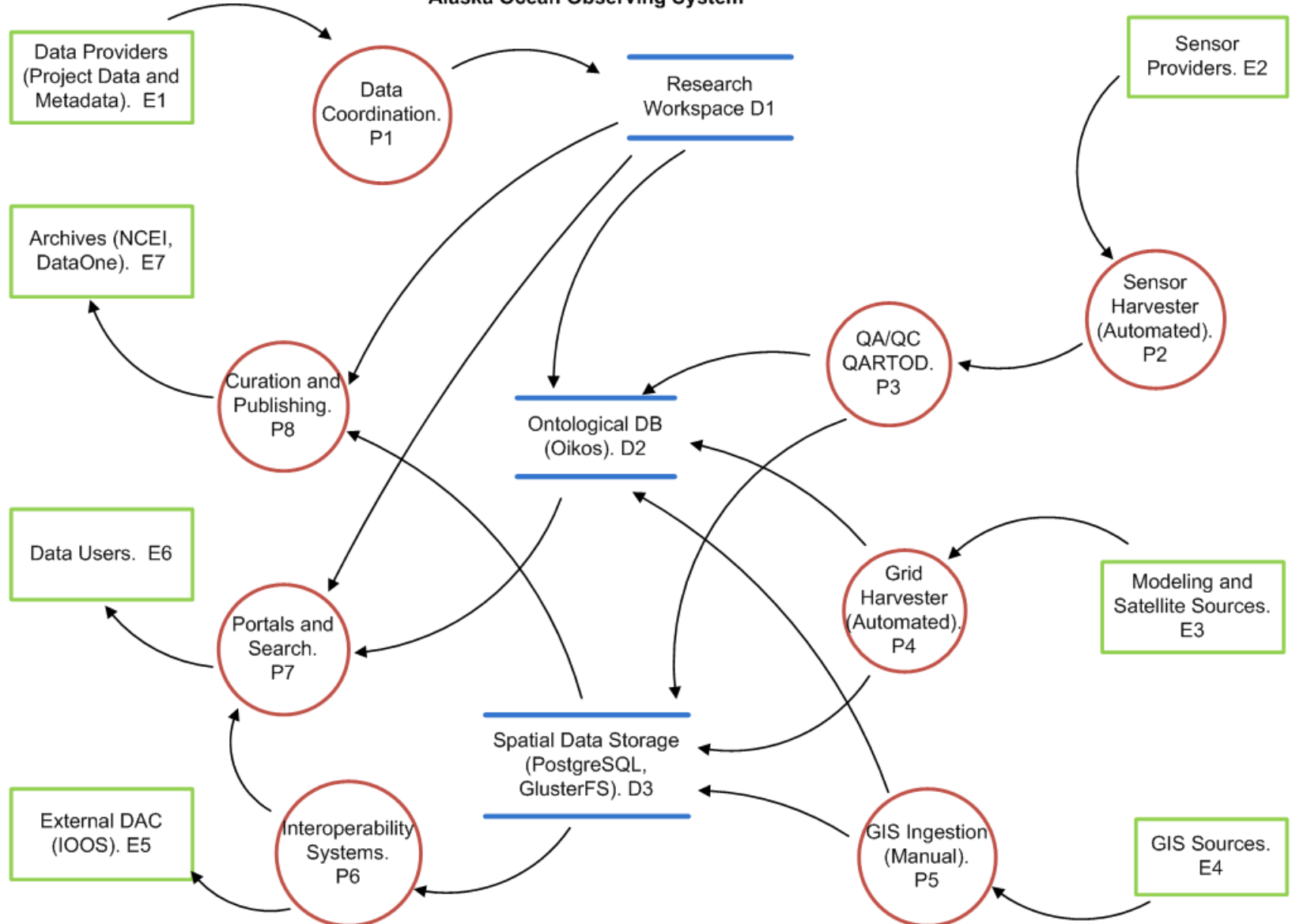
- Mission Driven

Shared Cyber-Infrastructure

- Leverage applications, systems and hardware across several partners
- Functional improvements can be funded by one partner and shared with the collective
- Reduces costs and affect of funding instability while accelerating progress
- As more partners participate momentum increases
- Standardize systems and interfaces across partners
- Reach critical mass: makes the impossible actually possible

NGA LTER Data Lifecycle

DATA CREATION & QUALITY CONTROL
Scientists

DATA DESCRIPTION
Metadata Editor

DATA STORAGE
Workspace

ARCHIVE & PRESERVATION
Repository submission pathway

DATA ACCESS & DISCOVERY
Data portals & search catalogs

REUSE & TRANSFORMATION
Jupyter notebooks & data analysts
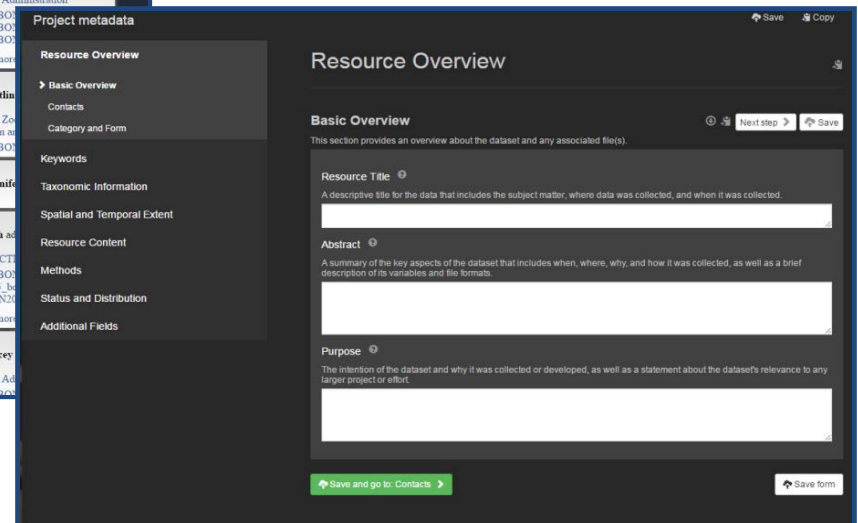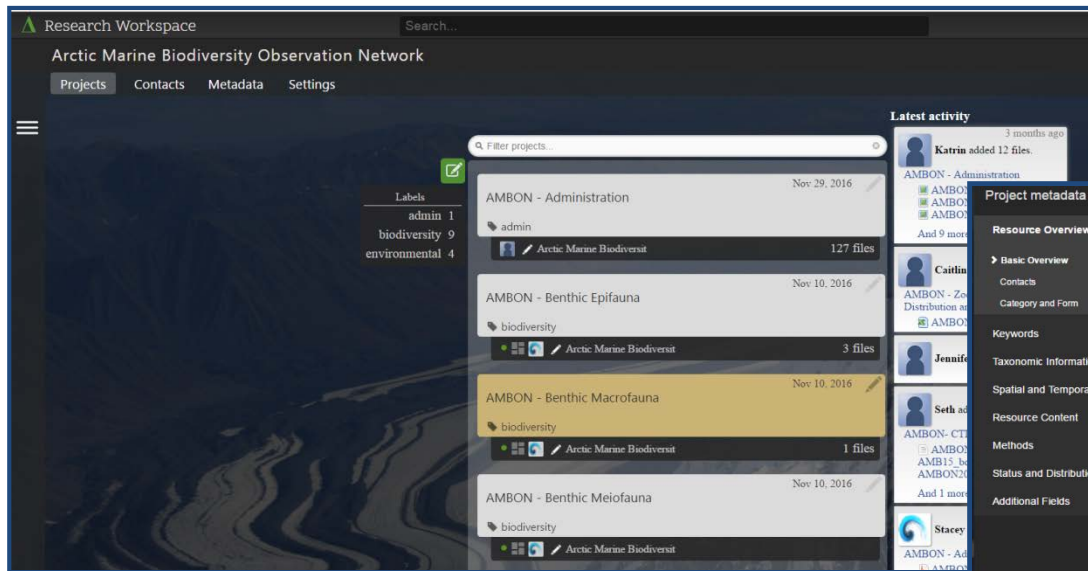
Data Flows

Level 1 (1.0) Data Flow Diagram
Alaska Ocean Observing System

Research Workspace

➢ Coordinate data exchange across networks

➢ Execute server side R and Python numeric workflows (Jupyter) on uplo
data AND any data in Oikos stack

➢ ISO 19110/19115-2 standards metadata editor

➢ Archive pathway to DataONE & Datacite DOI minting

Integration and Interoperability: Oikos Stack