

# 2019 LTER Information Management Committee Annual Meeting Report

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

July 15, 2019

Greater Tacoma Convention Center, Tacoma, WA



## Introduction

The 2019 Annual LTER Information Managers meeting was held at the Greater Tacoma Convention Center, Tacoma, WA, hosted in conjunction with Earth Science Information Partners (ESIP) Summer Meeting. LTER Information Managers, Environmental Data Initiative (EDI) representatives and others gathered to discuss issues related to LTER Information Management and specifically focused on engagement and collaboration. NSF Program Officer, Peter McCartney, joined the group in person for a discussion about NSF support for LTER sites.

<b>Contents</b>	<b>Page</b>
Attendees	2
Agenda	3
Elections	3
Working Group Reports/Updates	4
EDI	5
Semantics	10
DEIMS	11
WIRED	14
ECC	17
LTER Core Metabase	22
Zotero	27
Introduction to Breakout Group Topics	32
Breakout Group Assignments	33
Breakout 1: Engagement Charge	34
Breakout 2: Collaboration Charge	35
Breakout Outcomes (summarized from group discussions)	36
Breakout Groups Report Back and Group Discussion	38
Group 1	39
Group 2	43
Group 3	47
Group 4	49
IMC-NSF discussion topics	50
ESIP Session Speed Talks	51
Meeting Notes	
By Stevan Earl	52
By Suzanne Remillard	58

## **Attendees**

Gastil Buhl (MCR), Stevan Earl (CAP), Hsun-yi Hsieh (KBS), Li Kui (SBC), Suzanne Remillard (AND), Don Henshaw (AND), Kris Hall (SEV), Hap Garritt (PIE), Reuben Fresquez (MCM), Renée Brown (MCM), Yang Xia (KNZ), Jim Laundre (ARC), **An Nguyen (BLE)**, Saul Kelly Roman (MCM), John Porter (VCR), Dan Bahauddin (CDR), Corinna Gries (NTL), Emery Boose (HFR), Mary Martin (HBR), Margaret O'Brien (EDI), Darren James (JRN), **Greg Maurer (JRN)**, Jason Downing (BNZ), Sven Bohm (KBS), Mark Gahler (NTL), Adam Sapp (GCE), Stace Beaulieu (NES), James Connors (CCE/PAL), Chris Turner (NGA), Tim Whiteaker (BLE), Kristin Vanderbilt (FCE), Colin Smith (EDI)

***\*bold names are new site IMs attending their first annual IMC meeting***

Unable to Attend: Wade Sheldon (GCE), Sarah Elmendorf (NWT), Ed Johnson (CWT)

No active IM: LUQ

## **Agenda**

8:30 am -Introductions and Overview of IMC meeting  
9:00 am -Working Group Reports/Updates  
10:00 am -Morning Break  
10:30 am -Introduction to Breakout Group Topics  
10:45 am -Breakout 1: Engagement  
12:00 pm -Lunch on your own  
1:15 pm -Breakout 1: Engagement -Report Back and Group Discussion  
1:45 pm -Breakout 2: Collaboration  
3:00 pm -Group Photo  
3:15 pm -Afternoon Break  
3:45 pm -Breakout 2: Collaboration -Report Back and Group Discussion  
4:15 pm -Conversation with Peter McCartney, NSF  
4:45 pm -Business, Wrap-up and Action Items  
5:15 pm -ESIP Session Speed Talks  
5:30 pm -Adjourn  
7:00 pm -IMC Group Dinner at “The FORUM” (815 Pacific Ave A, Tacoma, WA 98402)

## **Elections**

No positions needed to be filled, so no elections were held.

# Working Group Updates

Environmental Data Initiative (EDI) - Corinna Gries

Semantics - Kristin Vanderbilt

Drupal Ecological Information Management System (DEIMS) - Corinna Gries

Website Improvement and Redesign (WIRED) - Suzanne Remillard

EML Congruency Checker (ECC) - Margaret O'Brien

Next Generation of Climate and Hydrology Data - Adam Sapp

Geographic Information Systems (GIS) - Adam Sapp

IM Orientation - Renée Brown

Core Metabase - Gastil Gastil-Buhl

Zotero - Tim Whiteaker



**EDI update**



# PASTA+ updates

Trusted CI, the NSF Cybersecurity Center of Excellence, recommendations for identity management services;

First-phase completion of MetaPype EML and online EML editor;

Software updates and upgrades of both the LTER and EDI DataONE Member Node software stacks;

Continued development of repository data package quality assurance/checking software;

Deployed additional reports and data package tracking through the EDI Dashboard;

Improvements to the EDI Data Portal search & discovery interface (e.g., duration-based temporal search);

Upgrades to PASTA operating systems, web servers, and use of LetsEncrypt SSL certificates.



# Outreach and Training

Newsletter, Twitter, Website, slack, e-mail

Webinars, town halls, data publishing workshops

Hackathon - developing data visualization package: a Shiny app that ingests an EML data package and produces plots and summary statistics for all variables as well as interactive graphs of the data

Data Help Desk at ESA and AGU

Summer Fellowship program



# Data Curation

Tools: EMLassemblyline, taxonomyCleanr, dataCleanr

IMCR Information Management Code Registry <http://imcr.ontosoft.org>

Data harmonization:

ecocomDP: <https://github.com/EDlorg/ecocomDP>

Clim-HydroDB 2.0: <https://github.com/lter/Clim-HydroDB-2.0>



# Questions

Who is using an Excel spreadsheet to gather metadata

Who is using a Word template to gather metadata

Who uses a database to manage metadata

# EDI Semantics Working Group

- Improve discoverability of EDI data
- 15 members - LTER, NEON, OBFS, Int'l Soil Carbon Network, NCEAS + Kristin & Margaret (co-chairs)
- Future Projects
  - Update to LTER Controlled Vocabulary
  - Controlled vocabulary for measurements
- ESIP session on Friday morning: "Location, Location, Location: Enabling Data Discovery by Place"

# DEIMS update

# DEIMS background

- ▶ DEIMS – build on Drupal CMS
- ▶ Basic web presence for LTER sites
  - ▶ Basic pages
  - ▶ Personnel
  - ▶ Bibliography
- ▶ EML metadata management
  - ▶ Complete information for EML data sets
  - ▶ Web interface for maintenance of database records

# DEIMS migration

- ▶ Drupal 8 is implementing object-oriented programming
- ▶ Content migration complete for NTL
  - ▶ Some changes
  - ▶ Fully documented in Github
- ▶ Drupal module 'Data Explorer' is done
- ▶ Planned:
  - ▶ Export of metadata into EMLassemblyline format
  - ▶ Module to access site datasets in EDI repository

# WIRED

## IM **W**ebSite Improvement & **RE**Design Group

**Origin:** Resulted from the 2016 IMC Annual Meeting in Santa Barbara

**Current Membership:**

Suzanne Remillard (AND), Marty Downs (LCO) and Margaret O'Brien (EDI)

**Original Objective:**

Migrate IM.LTERNET.EDU from Drupal 6 to Drupal 7

**Morphed Objective:**

Determine best path forward for storing and archiving LTER IMC related documents

http://im.lternet.edu/



# LTER Information Management

- Meetings
  - Annual IMC Meetings
  - Virtual Update
  - Training Workshops

Includes working documents and meeting notes

- Resources
  - IM Practices
  - IM Guidelines
  - Training Materials

Includes final products

Search this site:  Search

### Main Menu

- News
- Home
  - Terms of Reference
  - IMexec
    - EB Representatives
    - Meeting Notes
    - Reports
    - Roles
  - NISAC
  - Publications
- Projects
- Meetings
- Resources
- LTER Site Profiles
- Website Help

Includes working group notes, products, software, etc.

- Projects
  - NIS Proposals
  - Controlled Vocabulary WG
  - Drupal Ecological IMS (DEIMS) WG
  - EML Congruence Checker WG
  - GIS WG
  - SensorNIS
  - SiteDB
  - Unit Dictionary WG
  - Webservices WG
  - Software Developments
  - Inactive Projects
    - NIS Tiger Teams
    - NIMO Host Selection Process
    - IMS Guidelines
    - Data Package Reporting WG
    - Digital Asset Management
    - IM Website Improvement and Redesign (WIRED) WG
    - IM training WG

# WIRED Year 3 progress

- IMC Projects that were EDI relevant have been migrated to EDI GitHub repository
- Test of GitHub repositories for WG and meeting notes
  - Questions regarding the best place to store PDFs and searchability
  - High technological bar to migrate and add files to GitHub repository
  - Very time consuming to review and migrate documents
- In order to use the LTER document archive effectively, it needs attention in terms of tags and document type organization (not insurmountable, but time consuming)
- There doesn't seem to be a solution that isn't a lot of work; migration to Drupal 7 may be the path of least resistance
- NCO is exploring Constituent Relationship Management (CRM) tools for collaboration that may be useful (??)
- Need to assess the needs and desires of IMC membership for acquiring information and resources (annual IMC meeting breakout session)



# ECC - 2019

ECC Working Group  
2019





## ECC - Recent Activity

2018:

- dateTime format, dateTime congruence
- File header compared to attributeName (for manual examination)

2019:

- Entity size, Entity size congruence



# Future

## Definitely ...

- Support for EML 2.2

## Maybe ...

- Align checks with FAIR principles
- Alert-level (between Info and Warn)

THISWEEK:

More to talk about at these sessions:

Fair Metadata

<https://sched.co/PtOX>

Metadata Evaluation: Tools and Results

<https://sched.co/PtOm>



# New checks? you tell us.

---

## Congruence:

temporalCoverage

geographicCoverage

enumeratedDomainWithinBounds

pubDate

numericValuesWithinBounds

## Metadata:

nonPastaDoiResolves

linkChecker

fourDigitYears



# EML 2.2 New Features - Reminder

---

- Choice of TextType or markdown
- New elements for data papers
- Taxonomic classification with IDs
- Project tree includes more data background
- Annotations at 3 levels
- Units
  - Expanded list
  - Name strings standardized and aligned with LTER recommendations
  - conversion with udunits2 synonyms

# LTER core metabase

- A metadata relational database
- Built on Wade's GCE\_Metabase
- Team-redesigned by site IMs
- Focused on dataset EML support
- <https://github.com/lter/LTER-core-metabase>

DataSet	
DataSetID	int4[10]
Revision	int4[10]
Title	varchar[300]
PubDate	date[13]
Abstract	varchar[5000]
ShortName	varchar[200]
UpdateFrequency	varchar[50]
MaintenanceDescription	varchar[500]
< 1	4 rows 11 >

EntitySortOrder	
BeginDate	
EndDate	
UseOnlyYear	
< 1	1 row

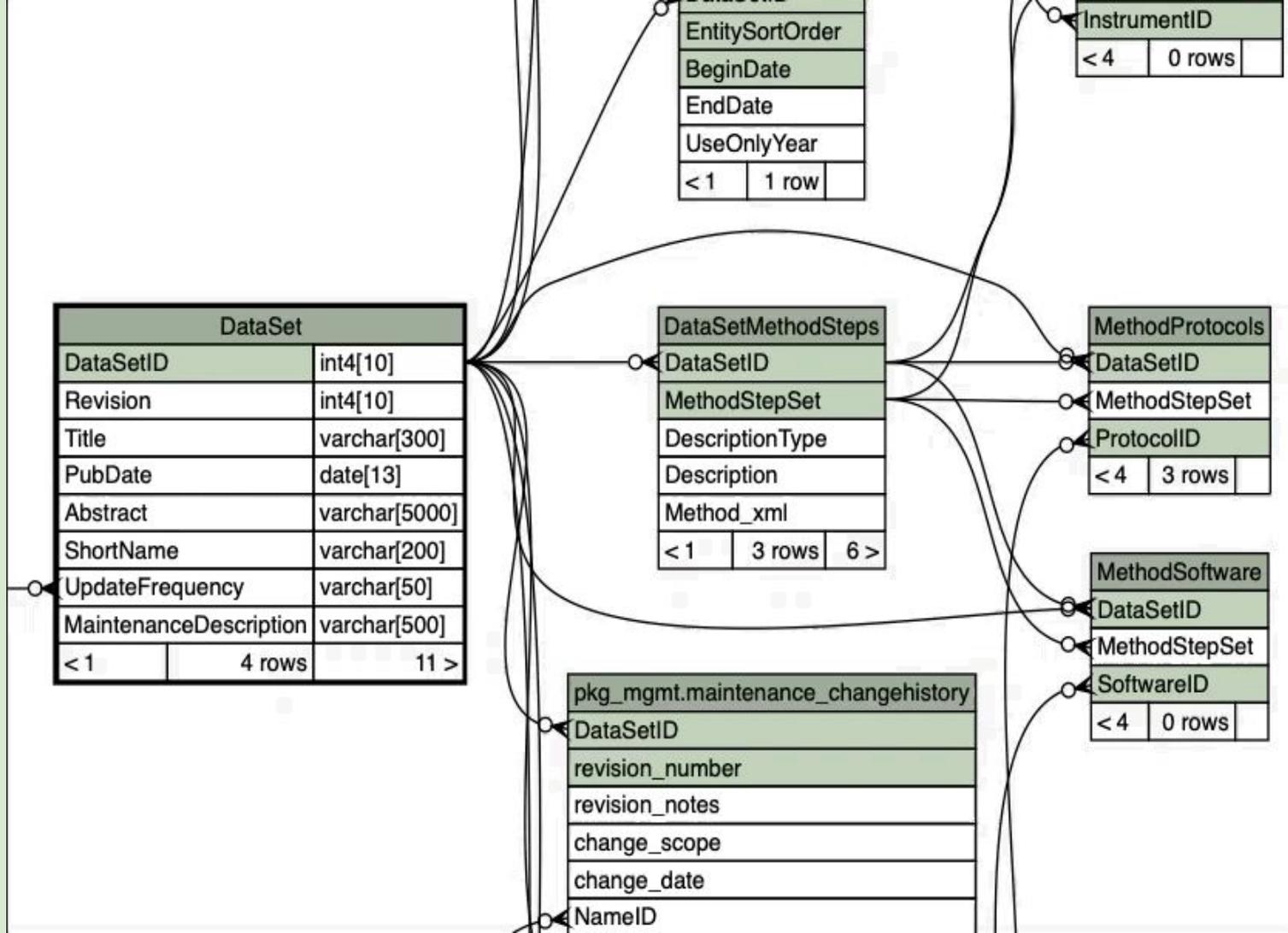
InstrumentID	
< 4	0 rows

DataSetMethodSteps	
DataSetID	
MethodStepSet	
DescriptionType	
Description	
Method_xml	
< 1	3 rows 6 >

MethodProtocols	
DataSetID	
MethodStepSet	
ProtocolID	
< 4	3 rows

MethodSoftware	
DataSetID	
MethodStepSet	
SoftwareID	
< 4	0 rows

pkg_mgmt.maintenance_changehistory	
DataSetID	
revision_number	
revision_notes	
change_scope	
change_date	
NameID	



... and that's  
not all!

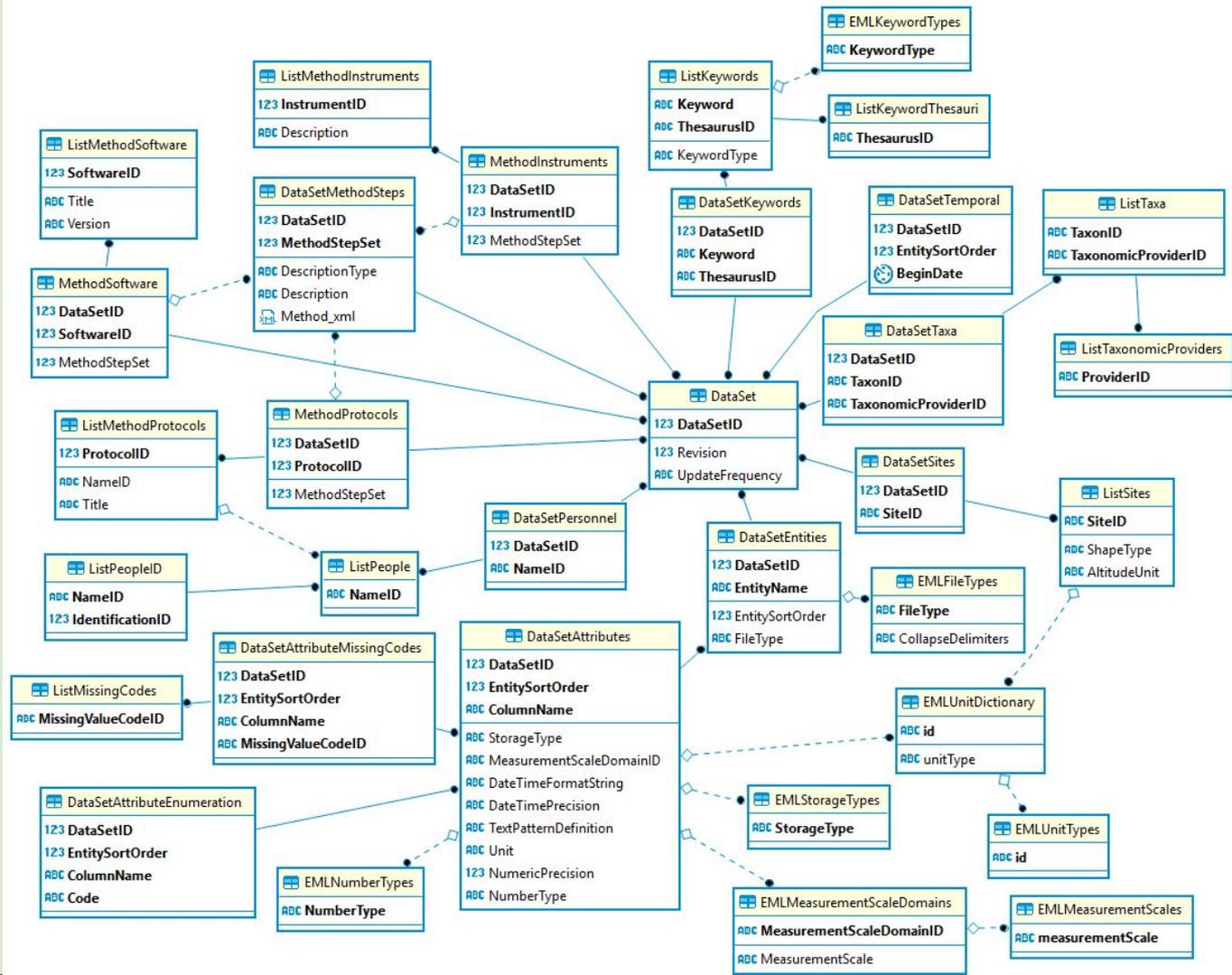


> `library(MetaEgress)`

Also on github

[https://github.com/Iter/  
LTER-core-metabase](https://github.com/Iter/LTER-core-metabase)

Tuesday 2:45pm



# Zotero Update

---

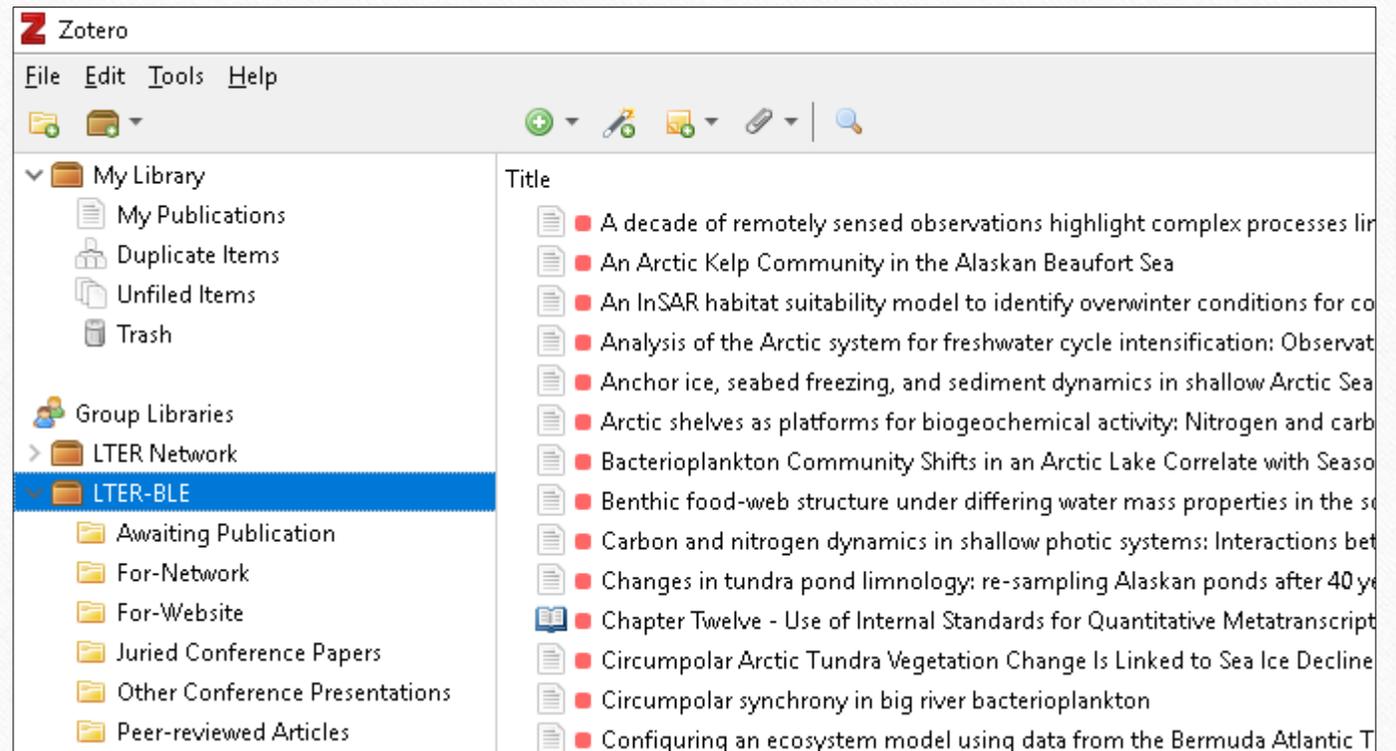
Tim Whiteaker

BLE-LTER

2019-07-15

# What Is Zotero

- Like EndNote, but free
- Syncs to cloud
- Has API



# We Made Best Practices For:

---

**Research.gov reporting,**

**and a bibliography for your website**

## Submit New Product(s)

Select the type of product you want to add to your report or upload multiple products using BibTex file. ?

Select Product:

Search our bibliography, or view [our library on Zotero](#).

Found 5 results for **arctic**

Harris,Carolynn M., Nathan D. McTigue, James W. McClelland, and Kenneth H. Dunton. "Do High Arctic Coastal Food Webs Rely on a Terrestrial Carbon Subsidy?" *Food Webs* 15 (June 1, 2018): e00081.

revised 2019-04-11



# Environmental Data Initiative

- Create , Package , Archive , Discover , Reuse -

WELCOME • ABOUT ▾ • DATA ▾ • NEWS ▾ • EVENTS ▾ • RESOURCES ▾

Welcome

**Find the Best Practices**

5 PHASES OF DATA  
PUBLISHING

DATA MANAGEMENT  
RESOURCES

The *Environmental Data Initiative* is an NSF-funded project accelerating

Tweet

# Stay Informed

---



Join the Zotero channel

<https://lter.slack.com>

# Introduction to Breakout Group Topics

Topics: Engagement and Communication

Two separate 75 minute breakout groups with 4 groups (8 people per group)

Predetermined groups (see list)

Five minute report back to the whole group

Please select a note taker and a report presenter for each group

# 2019 LTER Information Managers Committee Meeting

## Breakout Groups

### Group 1:

- Don Henshaw (AND)
- Tim Whiteaker (BLE)
- Dan Bahauddin (CDR)
- Adam Sapp (GCE)
- Greg Maurer (JRN)
- Renée Brown (MCM)
- Stace Beaulieu (NES)
- Hap Garritt (PIE)

### Group 2:

- Suzanne Remillard (AND)
- Jason Downing (BNZ)
- Margaret O'Brien (EDI)
- Mary Martin (HBR)
- Hsun-yi Hsieh (KBS)
- Reuben Fresquez (MCM)
- Chris Turner (NGA)
- Li Kui (SBC)

### Group 3:

- Jim Laundre (ARC)
- Stevan Earl (CAP)
- Colin Smith (EDI)
- Emery Boose (HFR)
- Sven Bohm (KBS)
- Saul Kelly Roman (MCM)
- Corinna Gries (NTL)
- Kris Hall (SEV)

### Group 4:

- An Nguyen (BLE)
- James Conners (CCE/PAL)
- Kristin Vanderbilt (FCE)
- Darren James (JRN)
- Yang Xia (KNZ)
- Gastil Buhl (MCR)
- Mark Gahler (NTL)
- John Porter (VCR)

## IMC 2019 Breakout 1 Topic: **Engagement**

### *Overview*

An LTER information manager provides essential services at the intersection of science and technology. While learning and staying abreast of technology is a challenge, engaging with our community of scientists is equally important to LTER science and presents unique challenges. With the goals of (1) enhancing our workflows, (2) improving our ability to converse and interact with researchers at our site, and (3) aiding newer information managers with these sometimes difficult aspects of the position, we would like to consider processes and approaches to improving engagement with our scientific community. Specifically, we would like to identify what has/has not worked well or is/is not efficient, and to identify ways in which we might all enhance engagement with our scientists.

We ask you to please consider:

- How can we stimulate more efficient and productive interaction with site researchers?
- What can we do to enhance the quality, extent, and ease-of-use of data and metadata that we receive from our investigators?
- How can we gauge the success of these interactions?

When addressing the above questions, you may find the following detail-oriented questions helpful to your discussion:

- How do IMs engage with site researchers? What works, and what doesn't?
- What are the perceived difficulties in these interactions, from both an IM and researcher perspective?
- How do researchers search for and use the data?
- How do researchers provide data?
- What quality control and analysis services do or could you, as an IM, provide? How do you make scientists aware of such services, and do they take advantage of such offerings?
- How do you convince a researcher that it is in their best interest to publish their data?
- How do you engage graduate students in the data process at your site?

### *Outcomes*

Desired outcomes from this discussion include (1) enlightenment, (2) ideas and materials that might contribute to a best-practices document, ideally a journal article, and (3) something tangible or conceptual that we might be able to share with our scientists that would help them from that side of the relationship.

## IMC 2019 Breakout 2 Topic: **Collaboration**

### *Overview*

The NCO maintains an archive of documents created by the LTER community. The purpose of the archive is to store and make available documents from LTER committees as well as various organization documents. The IMC created a Drupal website in 2011 as a place for collaboration and to archive IMC-specific documents (*e.g., meeting notes, working group information, committee and resource documents, etc.*). As more sophisticated collaboration tools became more common, such as Google Drive, working groups began to use these other platforms for collaboration. The IMC Drupal site is outdated, and the NCO will end support for it in the not-too-distant future. As a result, the IMC needs to identify new tools and platforms for collaboration and material archiving.

We ask you to please consider:

- What are the preferred methods to communicate, share resources, and facilitate data sharing and best practices? How do you do this at your own site?
- What tools exist for documentation of workflows and document storage, such as data collection protocols? How are you handling this at your own site?
- What are the desired platforms to preserve and share IMC resources like meeting notes, guides, and practices?
- What do new information managers need or want to know? How do you (a new IM) expect to go about finding it?
- The IMC Drupal site is and has been a helpful resource for people outside of the LTER so be thoughtful as to our audience(s).
- The NCO would like for the IMC to consider using a yet-to-be-identified platform (*e.g., Zoho*) that they are planning to adopt to support other LTER groups, such as the Education and Executive Committees. What needs does the IMC have beyond what such a system might support?

*non-exhaustive catalog of current archive & collaboration Tools:* IMC Website (Drupal), Databits, Google Docs (Team Drive(s)), Slack Teams (LTER, EDI, NCEAS), LTER email lists, Virtual Water Coolers, EDI Webinars, ESIP Clusters, GitHub (LTER, EDI, NCEAS), Wiki pages, Trello

### *Outcomes*

Desired outcomes from this discussion include: (1) foremost, a decision, (2) critically, to identify what needs to be done to realize the decision and how it will get done, and (3) an assessment of the IMC's needs in these regards that will aid the NCO in their selection of a new platform. Finally, we would like for IMs to walk away from this discussion about collaboration at the network level with ideas about how we might improve collaboration, documentation, and resource sharing at our sites.

# Breakout Outcomes

## Breakout 1: Engagement

Keys to success:

1. Buy-in from lead PI
2. Regular contact with researchers
3. Incentives and education for what IM can do for researchers
4. Educate the IM on the researcher's work
5. Shift website design and system administration to other personnel

Possible action items:

1. Create a flyer highlighting what IM can do for PIs. John Porter started this, though it is more about what sharing data can do for PIs. [Link to "why share data" document.](#)
2. Encourage funding agencies to require a list of archived datasets and DOIs in reports.

## Breakout 2: Collaboration

IM Online Collaboration Tool Requirements (not all of these need to be fulfilled by the same website):

1. Search and navigation on the home page.
2. Collaborative space for archiving and editing documents, including versioning, e.g., Google Docs.
3. Code versioning, e.g., GitHub.
4. A place for discussions.
5. Email lists.
6. Email interoperability.
7. Virtual meetings.
8. Task management (working groups and committees).
9. Track outcomes of working groups.
10. Calendar.

Action item: Morph the WIRED group into a group for migrating IM website content.

Group tasks:

1. Communicate our needs to the LNO.
2. Move appropriate content to Google Drive.
3. Draft the structure and links needed for a new IM website.

4. Use wget to cache the Drupal website (Sven will do this). Place this static copy online as a legacy reference.
5. Create zip files for each working group web page (including subpages and linked files) and provide these to working group leaders (Sven will do this).

Action Item: Use virtual water coolers to enhance collaboration by discussing common IM tools, solutions, and workflows. The first of these will be on August 19, 2019. Jason Downing is leading this.

# **Breakout Groups Report Back and Group Discussion**

### 1. Tim read the topic

#### 2. Round robin on successes and failures in this area

- \* Hap - local PIs are much easier to contact and interact with. Dispersed and non-local PIs or researchers are less frequent and more difficult/formal to contact.

- \* Dan - more local researchers. Maintain a database of what data collection efforts are going on. Some of that is populated by events around research presentations. Challenge comes when researchers change the plan and don't alert IMs. Sometime research doesn't go through any formal approval process.

- \* Renee - very distributed sites. 1 co-PI at her institution who is assigned to manage IM, and the IM personnel. Good rapport is important. Visiting the site helps provide some rapport with research personnel. Yearly PI meeting that includes IM and PM. Summer all-hands meeting helps, but students don't seem to get it in some cases, so PIs need to make IM important to students/postdocs. Deadlines for "core" datasets and PIs can have support held if they don't meet the deadline.

- \* General question: What is core data? What priority does it have over other data (like grad student data)?

- \* Adam - Wade sat on the executive committee for GCE and had a good idea of research activity. Annual meeting that includes students. 50/50 local/distributed. IM team has some one-on-one discussions with new students or researchers.

- \* Don - IM team processes & manages many of the core datasets (long-term met data, etc). Getting lead PI buyin for IM is important. What are the incentives? Grad student trainings are helpful as are metadata parties. IM rep on the executive committee. Grad student trainings are also helpful with pointers to helpful resources.

- \* Stace - IM is the data provider. Puts cleaned cruise data online for the scientists to use. Also educates scientists of what the value of these data area and how to access it. She is training students and technical staff in metadata and data publishing. PIs don't see the value in this yet - she is trying to build the value proposition for using EDI. Also hoping to develop materials to educate PIs and other researchers.

- \* Tim (BLE) - New site, there was an "all-hands" site visit during creation of the site. IM meets with exec committee (subset of PIs).

- \* Scheduling 1 on 1 time about IM with researchers helps lots.

#### 3. Summary (important points):

- \* Contact with PIs on a relatively regular basis

- \* Tracking projects and data products from start to finish so that metadata is being created along the way.

- \* Buyin at top level (lead PI) - communication in field, annual meetings, monthly executive meetings.

- \* Give incentives to PIs on why metadata and publishing is important - one of these could be showing them how easy it is to find and use their data.

- \* Scheduled 1 on 1 meetings

- \* Metadata trainings or "parties"

- \* Educating PIs and students

\* How do we get feedback from them? - ask a followup question after the data submission process.

## # Breakout session 2 notes (Collaboration, Group 1)

### 1. Tim read the topic

### 2. Identify and assess needs for an IMC collaboration platform

\* There are a couple broad needs: IMs need to find resources and collaborative resources for network-wide standards and protocols. They also need ways to document and archive how they apply these resources to their own individual sites.

\* Share files, resources for IMs (best practices, guidelines, meeting minutes, protocols, tools)

\* NCO seems to want a collaborative platform that is very personnel based.

\* Being able to search and discover these resources also important.

\* There is a long history of documentation associated with IMC development and working group activities (some are in a zotero library tag - lter-im) that is documented on the current IMC website. This content needs to be hosted somewhere and organized in a more useful way.

\* What should the relationship be to other websites or content sources (like EDI).

\* Google Drive and TeamDrive allow collaborative editing which is advantageous. They do not allow search and sort capability that we want - so, whatever CMS and web pages we develop for LTER IM would need to point to our documents in an intuitive way.

\* NCO mentions Zoho, what functionality does it have?

\* What other functionality do we need:

\* Website of some kind is critical

\* Calendar? – probably useful

\* Versioning? - on active documents or even historical documents

\* Forums?

\* Restricted access documents? - yes for proposal drafts or reviews or other stuff.

\* Searchbar and good navigation - ON THE FRONT PAGE

\* Other collaboration needs (team talk, code development) are covered by other web tools that could be linked from the eventual LTER IM website. For example, a web page listing the members of an IMC working group could link to the relevant Slack channel (they don't necessarily need to be a feature of a new website).

### 3. Tasks to meet our needs

\* Decide what tool? Maybe just use what NCO decides

\* Move content over?

\* Organize content in an intuitive way in a (shared) storage space

\* Decide what responsibility IMs actually have for maintaining websites. - Our job (consensus) should be to manage content (data and metadata only - this relates to using the NCO tool)

\* Decide what we need that is (Still) unique to IM?

\* Create a new working group to absorb WIRED WG since it seems a bit stalled

### 4. Summary of our decisions:

\* We need a website - whatever NCO uses will be best, needs search and navigation clearly visible on front page.

- \* We can wait for NCO to decide this based on needs of other LTER parties, which should be broadly similar.

- \* We (IMC) should create a collaborative space for archiving and editing of document, including versioning - we favor Google Docs for this

- \* Create a new working group to move all content (IM resources) from D6 to the Google Drive above.

- \* Draft the structure and links needed for a new IM website.

- \* Once NCO decides on a CMS we will implement this framework on that new website.

- \* We should tell NCO what we need.

**2019 LTER Information Managers Committee Meeting Agenda**  
**Room 318 at the Greater Tacoma Convention Center, Tacoma, WA**  
**Monday July 15, 2019 8:30am - 5:30pm**

Group 2: breakout

Notes: Suzanne

Bullets and report out: Jason

Group participants: Suzanne Remillard (AND), Jason Downing (BNZ), Mary Martin (HBR), Hsun-yi Hsieh (KBS), Reuben Fresquez (MCM), Chris Turner (NGA), Li Kui (SBC), Margaret O'Brien (EDI) [collaboration session only]

**Topic 1: Engagement**

How do you engage with your researchers?

MM:

Quarterly and annual meetings

Best method is one on one

Investigators are categorized from proactive to least responsive

Lead PI helps with non-responders

Review or renewals help with response

JD:

Engage with researchers early

Code of conduct for participating

SR: Member of the executive committee at the site

HY: Engagement is bi-directional; researchers deposit data, but also want to access it.

MM: Dear colleague letter form NCO went just to IMC, but should go out to site exec as well.

This letter, which support funding for IM.

CT: Uses Trello for other programs to follow progress

Data and metadata training; mainly graduate students

Office hours; "metadata 411"; presentation then hands on hands

JD:

Regular meetings (zoom) for offsite groups

Join for a day of fieldwork

IM and PIs at LTER sites have close personal relationship

Elevator sales pitch about why data should be in EDI

MM: send report from EDI to PIs to show the value of checks

LK:

Sends excel version of metabase for researchers to fill out

SR: Prioritize datasets

LK: How do you deal with multiple repositories?

JD: Data does go into different repositories, but this is listed on site catalog

CT: What's the disadvantage to being inclusive when declaring pubs or databases 'LTER'?

MM: It's a grey line.

SR: We do generic QC for all our datasets. Our data that goes into EDI is really clean.

JD: Getting the concept that how you analyze data and how you archive is very different.

CT: Most of metadata 411 is talking to researchers about what a dataset it and what their dataset looks like.

How do people provide data?

CT: Tries to discourage

Tracking using Trello

Sets up folders on the cloud for researchers; but the folder organization isn't being adhered to.

JD: Tracks using a database; sends emails to individuals

HY: Do you see that researchers are more accepting of providing data?

SR: Seems like it is so. But they don't understand how time consuming providing data can be.

CT: Researchers understand the importance of providing data, but have problems with giving up their spreadsheet manipulation techniques.

## Report Bullets:

Spectrum of approaches to engagement:

Quarterly/Annual/Lab Meetings

One-on-one \*\*\*\*\*

Most include some structured format to exchange metadata with researchers  
(Excel, Word, webform...)

Spectrum of approaches to training:

Student centered

PI-centric

Parties

Email 411 support

Tools to track/prioritize submissions or package status

Post-Its → database → Trello(etc.)

Support of PI and leadership when tracking down and locating non-submitters

Emails from senior leadership

Ideas for "Code of Conduct"/contract for data submission

Important to sell users on the benefits to data release through the LTER

F:

A:

I: → analysis formats V archive format

R: → share code and process for producing the data

## Topic 2: Collaboration

What are our collaboration tools as a community? What are our requirements as a group? We need a list of our needs.

Our Needs: We need a place for notes, a place for code, email lists, email interoperability (integration with other tools), a way to hold virtual meetings (webinars and calls), task management (WG), track outcomes of WG (docs or code or website; knowledge base) and ability to make public or private, findability, calendar for scheduling calls.

WG should determine where they want to work, but provide final document

What do we want to keep private and those that we want to share?

What are the tasks that need to happen?

Talked a lot of about the mechanism; working group to spend a few concentrated days actually migrating information.

What resources do we need that we don't have?

i.e., Integration with task management to keep track of WG

Maybe different interfaces; public interface and web docs

Do we want to keep track of the project or the process? If it's just the project, then a final document may be enough. Seems like WG should decide how they want to operate and the IMC would only have access to the progress/final report.

Standing WG vs product oriented WG

An IMC presence on the NCO website and document archive needs to be better organized so that it is easier to find things.

A big desire to be able to easily locate information.

We need to use multiple tools for our work, but we need to be able to have these in one place.

Marty is looking into the lowest common denominator for all committees.

Best practices for conducting business (chat window in zoom, notes for VWC); some of these are in out ToR.

Instant messaging (like our slack channel); is it effective? It is for people using it.

GitLab used for storing code is fairly easy to use and can remain private.

## Report Bullets:

### Our Needs:

- a place for notes/files
- a place for code
- a place for discussions
- email lists
- email interoperability (integration with other tools)
- a way to hold virtual meetings (webinars and calls)
- task management (WG and committees)
- track outcomes of WG (docs or code or website; knowledge base)
- ability to make public or private
- Findability
- calendar for scheduling calls

### Working Group Specific Needs

- Decide where/how to as a WG
- Final Product/Annual Report

### How to:

- Task Oriented IMC Working Session(s)
  - Concentrated effort
  - Utilize collective knowledge database from senior IM's
  - Develop Recipes or workflows
  
- Develop additional Collaboration BP for LTER (or RTFB)

## **2019 LTER IMC Meeting**

15 July 2019

Tacoma WA

Notes by Sven

Summary by Emery

### **Breakout 1 (Engagement) Group #3**

How to engage with scientists

- Sites have different strategies for tracking research projects and data submissions.
- One common thread is that researchers now come to us with a request to archive their data and get a DOI for their publication.
- If we want scientists to take an interest in and value information management, we should do the same for researchers, e.g. by attending research seminars and talking with scientists about their work.
- Wherever possible we should offer presentations on data management, open science, and related topics at our sites.
- If funding agencies required a list of archived datasets and DOIs in annual reports, it would help to ensure dataset submission.

The job of the information manager

- Some tasks, including website design and system administration, are best done by others wherever possible.
- Over the years we are seeing a transition in required IM skills from database management to data science (with scripting languages such as R and Python).
- Should the task of creating EML (from metadata supplied by the site) be given to EDI?
- For how long will individual sites continue to maintain local information management systems?
- What would be good job description for an LTER information manager?
- If we have some flexibility in shaping our own jobs, what would we most like to do?

### **Breakout 2 (Collaboration) Group #3**

The IMC website

- The IMC website is currently in Drupal 6 at NCO. This presents a security risk.
- The website is mostly out of date.
- With a few exceptions (e.g. meeting notes), the website is not widely used.
- Upgrading the website to Drupal 7 or 8 would be a significant task. So far no one has volunteered to do this.
- We recommend endorsing NCO's choice of constituency management software.
- We also recommend using wget to create a static version of the website (html with linked files) that could be made available online somewhere (NCO?) for reference and would not create a security risk. Sven has offered to do this.

- Sven has also offered to create zip files for each working group web page (including subpages and linked files) and to provide these to working group leaders.
- IM-Exec could ensure that critical files for the IMC (e.g. current bylaws) are migrated to the NCO site.
- Individual working groups could ensure that critical files for their working groups are migrated to the NCO site.
- Other materials could be migrated from the current IMC website to NCO as the need arises. This would be more efficient than trying to migrate everything now.
- The WIRED working group could help to resolve any questions or issues that arise.

# takeaways

Easier to add and/or fix than to start or create from scratch

- IMs might need to put in upfront work: an incomplete and incorrect preliminary draft (on EDI staging or locally) is better than nothing. Showing it to people might invite them to jump in and edit the draft.

Start the conversation ASAP

- In the project lifecycles
- In grad student's tenure

1-on-1 (in person, in Zoom) is better than long emails or sharing written resources

- Grad student orientation; make students try using badly-documented data

Be proactive, situate yourself as a resource and a solution provider, stay away from shame and badgering.

Sweeten the deal as much as you can. Wake up calls (being put on probation) are good but rapport is better.

## IMC-NSF discussion topics:

Stevan

- Invite NSF to *briefly* share their thoughts with the IMC
- Stability and infrastructure:
  - Applaud recommended funding for the EDI, but also stress that more stability in the funding regime is desired (e.g., funded on 5-year non-competitive renewal intervals similar to the NCO model).
  - Related, what is the future of DataONE?

Renée

- Multi-part question regarding the efficacy of data repositories, and the branding of LTER data:
  - Prior RFPs dictated that LTER data do not necessarily have to be archived with EDI so long as they are archived and accessible (e.g., BCO-DMO, ADC, AMD). Has enough time elapsed such that we can evaluate whether that approach is effective? Note that EDI/PASTA+ is the gold-standard in terms of metadata content and data quality not matched by other systems.
  - Does data that are solely archived with other repositories satisfy NSF requirements, even if there are lower or different metadata standards?
  - Meanwhile, there is a greater IM burden at a subset of LTER sites (e.g., Arctic, Antarctic, and coastal/ocean sites) that are required by NSF to deposit data in domain-specific repositories in addition to EDI. What are ways to reduce this burden, while also ensuring all NSF & LTER requirements are met?

Suzanne

- So much of the focus has been on getting the data into a system. We are and have been quite successful at that. Should we now shift some focus to getting the data out, and, if so, how does the NSF suggest information managers can support and contribute to data reuse and synthesis?

# ESIP Session Speed Talks

Tuesday, 2:45 pm - A Metadata Database Built on Usage Patterns in the LTER Network (Tim, Margaret, Gastil)

Wednesday, 10:30 am - The Information Management Code Registry: Software Solutions for Information Management Needs (Colin, Kristin)

Wednesday, 1:30 pm - Getting Stuff Done with R, Python and Jupyter Notebooks (Gastil, Stace, John, Colin, Chris)

Friday, 10 am - Location, Location, Location: Enabling Data Discovery by Place (John, Kristin)

Friday, 10 am - Preparing climate and hydrological time series data for submission to CUAHSI (Corinna, Martin Seul, Adam, Margaret, Suzanne, Don)

**2019 LTER Information Managers Committee Meeting Agenda**  
**Room 318 at the Greater Tacoma Convention Center, Tacoma, WA**  
**Monday July 15, 2019 8:30am - 5:30pm**  
*rapporteur: Stevan*

working group updates

EDI, Corinna

- completed first phase of metapipe, Python package for building EML, building an on-line editor on top of it for folks that need or want a form-based system for generating EML.
- complete software upgrade for DataONE. Used to display every version of every dataset in DataONE, now only displaying the most recent but a slow purge of old data sets
- more reports on the EDI dashboard
- new ECC checks
- new search capability for time duration, e.g., I want all data sets that are longer than, for example, 10 Y - this is an initial approach
- complete upgrade of the pasta operating system, upgrade to new version of Java
- outreach & training: newsletter, twitter, website, slack, email, webinars, townhalls (EDI would like feedback from the group during the town halls); training workshops for fellows, hack-a-thon (data viz package currently in alpha stage); Kristin pushing data mgmt at ESA (actually on the ESA program); summer fellowship with nine fellows that is moving from an undergraduate program to a more expansive group; development of assembly line and related packages; advancing the IMCR; data homogenization: ecocomDP, Clim-HydroDB (we now have real long-term data for which we no longer need to necessarily substitute space for time), Colin has converted ~ 70 data sets to the ecocomDB model but contact Colin if you are interested to have other data sets converted; instructions will be coming for how to convert site climate data to the homogenized model.
- Corinna polled the audience as to how many people use Excel, Word to get metadata, and whether folks use database to manage the metadata. This is related to EML profiles.

Semanics, Kristin

add Kristin's slide for these notes

DEIMS, Corinna

Corinna has ported the NTL DEIMS system to Drupal 8, the code is available on GitHub and Corinna can work with folks to port their systems.

WiRED, Suzanne

add Suzanne's slides for these notes

ECC, Margaret

most recent checks (now up to 41)

- dateTime format, dateTime congruence
- file header compared to attributeName
- entity size, entity size congruence
- will have support for EML 2.2 but we will have to talk about how to adapt the rules for when the structure differs between the versions
- possible future implementations:
  - align checks to FAIR principles
  - considering an alert, which would be between an info and a warn
- to consider:
  - temporalCoverage
  - geographicCoverage
  - enumeratedDomainsWithinBounds
  - pubDate
  - numeriValuesWithinBounds
  - nonPastaDoiresolves
  - linkChecker
  - fourDigitYears

Clim-HydroDB, Adam

- Adam is working on GCE Toolbox functionality that will take climate data and convert to, I think, CUASHI format.
- migration of what is in ClimDb and add it to CUASHI
- new data to go into CUASHI using the identified format

GIS, Adam

following on Jamie's GIS working group at the ASM, Jamie has revived the GeoNIS website; not working currently but it should troll a site's data for GIS information that is then displayed on the map/portal. GCE starting to collect huge inventories of drone imagery - what to do about that? Still figuring out the status of this group and these efforts.

IM Orientation, Renee

Building on Gastil's Tea-With-the-Chair. Does not really have a name now but calling it IM Orientation by default. Google spreadsheet with questions of interest. Renee now leading this effort. Need participation from new and senior IMs. Tentatively planning regular VTCs on this topic - how do you collect metadata at your site, for example? Renee looking for feedback on how to move this group forward.

Core Metabase, Gastil

- Not officially a working group. Working on a database schema/template that can be widely adopted.
- R package: MetaEgress

add Gastil's slide(s) to augment these notes

Zotero, Tim

add Tim's slides for these notes

Questions from this session:

- why were the Zotero best practices on the EDI GitHub?
- what to do about large data files, and especially large data files like drone output that have geographic details or other types of metadata (e.g., times)?
- if data go elsewhere, e.g., a genetic database, should EDI just link to it? yes, the place to search for data may not be the same place where the data are stored, e.g., Corinna suggesting that maybe PASTA is the search interface then link to wherever the data are located.
- Maybe we need a non-tabular data working group that could address drone data, large images, natural history collections. JP does not put all his image files in PASTA because they are huge and many but also because you need a search interface to view the data (so you know what to download).
- Whatever we do with these data, we should consider that it should be searchable by DataONE.
- We will form a working group on this issue (data types) that Corinna will lead

working group reports: engagement

group 1

- IM at a site begins with buy-in, which comes with incentives: funding, subcontracts; bottom-up vs. top-down approaches to buy-in; data on the website are useful data; provide statistical analyses; data publishing services; GCE: spending 30-minutes in advance will save hours in the long run
- communication is so important, IMs should go to the field to establish rapport; giving presentation and having informal conversation at site meetings: putting a face to your name; sites have regular meetings that are PI only that IMs should attend; have an IM orientation; some sites are working on website or brochure materials about services provided; one-on-one meetings with every single PI; IMs need to communicate better
- metadata parties to improve quality of the metadata, walk folks through the process; Q&A sessions about the data publication process - that is the IM should interview folks to identify the pain points

#### group 2

- spectrum of approaches to engaging people and all options should be used: meetings, one-on-one
- spectrum of approaches to training, most focused on students but some things like metadata parties work well for all including PIs
- no clear system for tracking what should be coming into the system
- support of the PI and leadership critical to getting people to provide their data is critical;
- code of conduct at the outset
- should have an elevator speech on the top of your head, these should be based on the FAIR principles
- think about the data holistically, not just by, for example, thesis chapters

#### group 3

- sites have different strategies but increasingly researchers are coming to us
- take an interest in the science - go to seminars, talk to scientists and students about their research
- NSF should require lists of data set DOIs in the annual reports
- should the role of the IM evolve, should we still be doing websites, server admin, etc.?
- skillset for IMs is changing from DBA to more data science and analytics
- should IMs be in the business of creating EML, maybe sites should provide the metadata and data to EDI and have them publish the data
- so, what is the future relationship of the IM and the repository
- what is the future role for the IM at a site, what would a job description look like going forward?

- what would IM like to do?

#### group 4

- it is okay to start by inspecting the data
- one-to-one is most effective but careful that it is not a bottleneck (versus working with a roomful of folks)
- expose students to data analysis: use data sets with good and poor metadata
- improve your own skills
- pitch yourself as a resource
- know the talking points
- support at the highest levels of the project is critical

#### working group reports: engagement

#### group 1

- need a centralized place for IM resources, platform for working groups to interact, search interface with tabbing with intuitive and prominent navigation, calendar, and forum could be helpful
- need a website and want to use whatever NCO is using - our job is to populate the content not maintain the infrastructure
- need to establish Google Team Drive structure
- decision: website (whatever NCO uses), and Google Team Drive for IMC

#### group 2

- need a place for notes and code, emails lists, virtual meetings, task mgmt, track outcomes of WGs, need public and private sections, each WG will have unique needs - they should decide where and how they work but all should produce annual updates and final products

#### group 3

- much of existing content is out of date
- create a flat version of the existing website as an archive

- create a zip file of each working group
- use whatever NCO chooses to use going forward
- individual items could be migrated, such as IM Exec moving the By-Laws

group 4

- archive the existing data but in a way that does not require additional resources
- CMS model has worked well so some sort of editable website could work well
- emphasize the value of something that is stable, not just long lasting but that is not onerous to update/upgrade
- how to identify when things are complete?

**2019 LTER Information Managers Committee Meeting Agenda**  
**Room 318 at the Greater Tacoma Convention Center, Tacoma, WA**  
**Monday July 15, 2019 8:30am - 5:30pm**

*Notes by Suzanne Remillard*

General Discussion:

The group agrees to adopt whatever CMS NCO decides to use.

Current website should be archived as is (Sven)

Can we create a structure for a new website (Renee, Greg)

If archived, how do we move forward until we know what NCO adopts? Google Team drive?

What we can do now:

Zip up and archive old website

Identify the IMC needs (from breakouts)

Develop structure of our space within NCO

Need to clarify with NCO about who can access the new web space. It can't be only one person.

NSF Discussion with Peter:

DEB

Cyberb-infrastructure building new tools

Keeping infrastructure of current tools working

Harnessing Data Revolution

Cross-site initiatives (10)

Core programs are shrinking

Entitlement programs are holding still

Initiative programs are growing

Stability and infrastructure:

DataOne future

Sustaining programs - hard to review; There aren't many reviews out there so it's hard to make decisions about what to continue to fund. How to assess the right things to do?

Cost recovery - pushing the programs to cover costs; but this is hard to do; needs to go into grants to cover costs for infrastructure.

Data aggregators - not primary repositories, but integrated resources where it is easier to find things.

Primary repositories - archived copies of data

Hard to get numbers on return on investment

Last review; 10 of 12 sites provided list of datasets with DOIs, which was really nice for the reviewers; this will likely be requested in the future.

Multiple repositories required for various data. Previous renewal request requires that data are in a recognized repository that meets standard and this won't change. It is fine to use any recognized repository. Also mentioned that they are also in a higher level data aggregator (like DataOne). It's fine in any appropriate repository that is persistent.

Is this multiple repository solution effective? Not all repositories are as of high quality.

Do we need LTER branding? Not as important.

Should we shift from concentrating on getting data into the system to getting data out? Hopefully, repositories now provide easy ways to get data in and free up IMs to help with analysis and data science. Repositories have search engines that provide data and should free up IMs time to do other things. Need to get people to use the repositories to search data.

Harmonizing data - is the benefit worth the outcome? This is a relative question. Most projects will do the harmonization once and then it's done. But if the harmonization benefits multiple projects, then it may be worth it. Is there a reasonable science driver behind it that will produce a useful result.

#### Databits

Should it be a PDF or just blogs on NCO or EDI website? It is not peer reviewed. But it is a nice, information newsletter. Some of the articles should be published in a peer reviewed journal. Do we want to continue?

Corinna suggests 3 different venues and 3 different goals.

Let's discuss at our September VWC. An and Sven agreed to be editors for the next edition and they could discuss different variations of the publication. Do we need a working group?

Somewhat controversial discussion and no clear path forward.

#### Next year's meeting

Restrained budget (but maybe not as we may not have used all of our funds intended for this meeting).

Is ESIP or a field site more expensive? Flights to field sites that are small more be smaller.

We should prep our sites to help fund us to go to a meeting.

ESIP would like to have LTER come next year, but our LTER sessions are sometimes scheduled on top of each other and on the last day, which is not idea.

Darren James suggested that in the future (they are on probation now), we consider holding a meeting at Jornada in conjunction with ARS. He thought that the USDA may be able to provide catering, which could help keep costs down.

Konza (Yang), Cedar Creek (Dan), and Sevilleta (Renee) all said their sites could easily accommodate a meeting.