LTER IM Meeting 2018

* LTER IM Meeting
  + Around the Room
    - lots of new faces
  + NCO/EDI Reports
    - EDI
      * Mark Servilla
      * LTER and beyond - OBFS, LTREB etc.
      * EDI organization
        + Workforce Engagement

Meetings, webinars

* + - * + Infrastructure

data portal

upgrades

new dashboard

* + - * future work
        + refactored identity management - openID etc.
        + Python library (Metapype for EML)
      * questions
        + how much of EML can be mapped via SCHEMA.ORG

working on Google recommendations

working with EarthCube groups on better defining otherwise loose groupings

* + - * + Separate web interface for LTER Data Portal vs EDI Data Portal?

maintain both for legacy reasons

LTER Portal will sunset

does require effort

use same underlying PASTA

there are tools on EDI that are NOT available on LTER

* + - * + What do users notice that is different?

for putting in datasets

EDI supports citations

but rendering for data users is the same

* + - * + I wasn't aware of the citation stuff on EDI

bidrirectonal linking associated with SKOLEX

Google Data Search does some of that

* + - * + Citation linked to provenance

Have some primative elements for provenance

Are partnering with DataONE - they take lead

* + - * + Challenges engaging with OBFS etc? How can LTER help?

big brother/sister - exemplars

lots of them lack resources

were just at OBFS

did workshop for site managers

6 fellows (undergrads) went to 5 sites and did info. management projects

now everyone wants their own fellow

not getting lots of datasets - lack resources

may have one fellow work with multiiple sites

* + - * + Naming scheme for datasets?

scope: all using EDI scope

should LTER transition to that?

recognize that identifiers should be opaque - use DOIs instead

in portal use for filtering landsat and ecotrends

do maintain control over scopes - can't just add nilly-willy

how do we uniquely identify sites in metadata - not using scopes

* + - * + Authentication - use LTER credentials?

yes

but maybe migrating

LTER LDAP may move to different process

LTER identity in openID world?

not a problem

outsource identity management

then you need to map them to internal names, permissions and roles

* + - * + How do you want to retrieve enhancement requests?

info@

support@

edirepository.org

* + - NCO
      * Mark Schildhaur
        + also on DataONE leadership team
        + presentation largely by Julien Brun
      * 6 awarded LTER working groups
        + distribution of participation from sites
        + dominated by non-LTER participants
        + survey showed most members were extremely or moderately satisfied
      * breakdown by careeer
        + most are early career stage
        + pretty quantitatively competent - R etc.
      * personnel
        + Julien Bruin - scientific programmer main contact

0.25 FTE

* + - * .Technical Liason
        + great role for an IM

be looking for opportunities

* + - * collaboration environment
        + google drives and groups

why?

good integration with google docs and sheets

downside - searching is hard

free! - approved UC

box not a s broad

* + - * + high-end server - 384GB RAM
        + preserving working group projects - including code
        + git, github
      * Data Science Fellows
        + metajam R package

download datasets of interest from PASTA/ED via DataONE

integrate data with metadata in downloaded data object

* + - * + - Data Registry for logging datasets from Working Groups
        + many outside PASTA
      * IMs could help us with
        + finding associated data
        + getting latest dataset
        + discovering site data not in PASTA
      * Improving process for providing LTER Synthesis working groups
        + technical liason often  group leader (but could be IM)
        + as WG member
      * Data preservation
        + working on collaboration with EDI
        + large set of types of info - notes, manuscripts etc.
      * questions
        + can resources be used by non-working groups

e.g., High end server

don't have that process formally offered

but NOT off the table

feel free to drop us a note

does LTER?

no - but may be better for cloud

* + - * + applaud recommendation with regard to IMs in working groups

but it happened via happenstance

can we make it a more formal process?

perhaps has group issue statement

* + - * + WGs tend to pick tech help based on PI's previous contacts- how can we more effectively advertise

expertise list available as part of proposal process

hard to maintain

at what stage should IM be integrated

after proposal accepted?

guidance to proposers

not always thematized

hard to do outreach to potential proposers

especially early career

suggest leaving slots open for additional members after funding

would be fairly radical

competition for remaining places

* + - * + Non-LTER Synthesis WGs?

not many - only Arctic - in Arctic Data Center

also SNAP working groups

with TNC and WCS

Julien supports them - go into KNB

* + - * + When is next call for proposals?

check with Marty

* + - * I focused here on IM stuff - also do lots of education, meeting support
  + LTER Visioning
    - intro
      * internal organizational changes
      * rise of data science
        + what is our niche -  now more are doing things
      * would like to do something tangible - coding etc
        + but we need to identify problems that we can work on
      * time - LOTS we need to do day-to-day - but we need to work as a community
      * 6 groups
        + set of questions for each group - some common, some different
        + you don't need to answer all 3 questions

thoughtful, fruitful discussion

want writeup of conversations -whitepaper

* + - WG II
      * facilitator - Porter
      * Reporter Stacy
      * Discussion
        + tragedy of the commons

primary work needs to be for site

building for greater good 2nd priority

but can be synergy

often takes a long time to get things done

EDI leveraging

but EDI could go away

* + - * + WG structure

depended on US needing to do things

should WG do projects or act as adviser

I like model where we advise

* + - * + review of past working groups

issues of compensation

some sites could contribute more than others

givers and takers in the network

pays back in terms of reputation

try to incorporate a variety of skills

lots of administrative issues with paying

funded working group meetings/ hackathon

key features

champion

dedicated group

resources

meetings/hackathons

contracts for products

* + - * + GIS Working Group

still falls to individual sites to keep it going

can look at list of working groups - make review

* + - * + how are legacy products maintained
        + why would I want to give time to a working group?

hard to justify given site requirements

can do prototype but hard to maintain

can contribute to documents etc.

* + - * + What made things work - or not

needed champion - keeps it alive

some have fizzled

lack of resources

too ambitious topic for 1 week meeting

PDF - GIS - funded by LNO

* + - * + Wired working group

im site hard to navigate - but useful for us

how do we move forward

* + - * + EDI role

have EDI or NSF fund working groups - prototypes

repeated meetings

should be involved at the get-go

modifications to the system are easier if we can provide guidance and counsel

even use of APIs

can see how it ties into infrastructure

workflow development

site standard content

vs centralized workflows

issues

need to be able to detect changes

must be robust

PASTA stable repository

underutilized event notification

can cause things to happen when things are revised

DIEMS uses REST API for upload to EDI

* + - * + also things like best practices etc. independent of EDI

often links

* + - * + ClimDB

interface  - not just data

lots of infrastructure there

how to maintain

* + - * Questions
        + Better management of IMC WG

how to ID needed working groups?

could do surveys

use cases for your needs for IM commuity

can also poll science teams

"I can't find anything"

often aren't aware of what is needed

IMEXEC keeps track - could include working group coordinator

probably need to stick with organic development

need to id "perceived needs"

need a champion

funding/resources

need to provide a way to provide resources

funding

workshops

infrastructure

often resource needs are minimal - but also real

IMEXEC could help enunciate needs

having a post-doc has always been needed for our science projects

also interns

data science interns

best for discrete deliverables

roving software person - hard to manage time etc.

* + - * + EDI role

funding - even $10K per year for working group support would make a huge difference

can't fund proposals -NSF does that

EDI could support groups working on projects

how would requests be handled

facilitation

repository for data

facilitating documentation for getting data into the repository

* + - * + NCO role

may be better for IM working groups

harder through EDI - not LTER-centric

* + - * recommendations
        + seek funding to support working groups
        + involve EDI early and often
        + need mechanism for moving from ad hoc to officially recognized

with renewal cycle - 3 year time limit

IMEXEC /IMC needs to be involved

structured vs unstructured working group

scope of work

funding requirements

* + - * + improve communication with researcher community to help ID needs
        + elevate visibility of working groups to larger community

credit/reputation

* + - * + look  outward as well as inward

links to ESIP etc.

* + - Breakout Group Reports
      * Group 1 - Martin - Earl - R. Brown
        + discussed EML and challenges of working with EML

concerned about complexity and time required

need better tools and workflows

avoid reinventing wheel

why are we managing web sites?

work more interactively with other groups

from site based to general

* + - * + synthesis

how to justify annual meeting when there is no funding

need justification

how to participate in synthesis working groups

* + - * + culture

some sites support, others don't

* + - * + harmonization

EcocommDp

issues with EML

* + - * + too entrenched

maybe consider abandoning EML

* + - * Group 2
        + recommendations
        + "secret sauce" for working groups
      * Group 3
        + discussion of climdb

follow example of EcommDP

connect with OBOE foundry to connect individual attributes to concepts

EDI could focus on training on ontologies & transfer of existing site based data into common model

scalable solution that would go beyond climate

go beyond LTER community

* + - * + GCE Toolbox, DEIMS

EDI role

good to have common tools for creating EML

R workflow

workshops

GCE toolbox

maintain, if not update

* + - * + interoperable outputs
        + broader vision

IMC needs to play role in collaboration with IMC

e.g., data congruency checker

* + - * Group 4
        + harmonization
        + how to work with EDI on semantic annotation
        + IMC can help

organize around nascent vocabularies

choose vocabularies for semantic work

* + - * + set priorities

inventory time series

use cases for science

start examing EcoTrends

* + - * + data harmonization

look for topical patterns

meteorological

ocean data

soil

forestry

* + - * Group 5
        + strategies to adapt to new technologies

struggled with idea of whether IMC should develop or just use new technologies

what is our main mission? preservation, curation

always new technologies coming online

strengthen relationship between EDI and sites

open source vs proprietary - when

open source unless proprietary works better

should we move more towards open source

status quo

need to keep looking at broader community

want to suit wide range of users

* + - * + self reliance

Matt Jones will retire, who will oversee EML

we have invested heavily

who will help maintain? push forward?

EDI or IM community? or numerous entities?

* + - * Group 6
        + sustain critical infrastructure

EML

GCE Toolbox

DEIMS

IMC will

bring together user communities

id funding sources

EML is beyond LTER -

can develop liason

* + - * + Essential Tools

EML

PASTA

VOCAB

need mechanism for identifying critical infrastructure

* + Renewals and Reviews
    - preparation, expectations, tasks
    - cohorts
      * get to know your cohorts - on LTER Network web site (but hard to find)
    - last cohort harmonized data inventorys
    - mid-term reviews
      * expected that IM gives presentation to panel
        + IM reviewer

often separate meeting

help deal with questions

may or may not be available in advance

if you are a reviewer, you can request early discussion

* + - * + Scientists

make concise - not so technical

* + - * + make use of guidelines

best practices for web sites

NSF guidelines

* + - * can advocate for early IM review
      * keys
        + Make sure data supporting presented research are online
        + never, ever, lie or mislead

be open about problems and proposed solutions

* + - * + don't get hostile - review can seem adversial

their JOB is to spot problems

don't get defensive

* + - preparation
      * get everything ready EARLY so IM is ready for PIs to use in THEIR preparation
    - Data management Plans
      * make sure you know what is in yours
        + look to see if it is at variance with IMC guidelines
        + can make changes
      * read the RFP VERY carefully
        + many don't!
        + it is like an exam question - it spells out what it need
      * within those constraints - make it INTERESTING
        + people have to read this stuff
        + make it as exciting as possible
    - discussion
      * NSF strips all hyperlinks
      * this is a chance to evangelize about IM - toot our horn
        + you are doing IM you are excited about some  aspect of IM
        + make sure you share that with the reviewer group
      * lack of release of reviews has been a problem
        + sites on probation are needing to do budgets prior to seeing reviews
        + hard to do
        + visit to NSF in person may help
        + tied to NSF requirements
      * sometimes site reviews are closely linked to proposal submission for sites on probation
  + EML 2.2 - O'Brien - EDI
    - backward compatible - relabel your eml and it WILL WORK - relax!
    - are new features
    - ready to go - except for documentation
    - see presentation - I'll do questions only
    - needs
      * update best practices
    - questions & discussion
      * in taxonId how do you resolve ITIS? not a URL
      * udunits C library and units
      * "standard" units are really predefined
      * EML upgrader web service?
        + help convert old units to new units
      * Still uses STMML - but with some additions
      * timeline?
        + can support schema quickly
        + markdown version may be an issue

rendering content may be harder and slower to implement

* + - * + congurency checks will be longer-term
        + will be grace period
      * KNB still has checker - will it work for 2.2
        + don't know
        + is blip on radar
      * What is expectation for legacy datasets?
        + what about OBFS and others
        + OLD versions remain immutable
      * arctic data center requires funding and ORCID
        + funding number
      * last time, sites pretty much updated everything from 2.0 to 2.1....
      * should have VTC to explore challenges for sites in updating and pursue common solutions
      * with versions - new metadata looks like new data
        + may confuse researchers
        + metadata revision vs data revision
        + can use data hashes to determine
        + but may want to label
        + could be added to additional metadata

emergent standard

* + - * citation fields - how in sync with EDI citations
        + same dataset from EDI looks different from citation from LTERNET.EDU
      * new annotation part is super exciting - allows us to semanticize EML
        + create RDF triple
        + vocabularies will provide precise sematics and links
        + tie to schema.org
  + Duane Costa
    - 10 cool features of EDI you've never used
    - preview your metadata
      * renders EML file from outside PASTA
      * can I link to?
      * no, need web form
    - Data Shelf
      * can save selected list of data packages
    - Journal Citations (EDI only)
      * associate journal article URL or DOI with data package
      * can add data citation on web
      * doesn't add to metadata - stored in external database
      * is in PASTA API
    - Expand search with LTER Controlled Vocabulary
      * searches on more specifc search
    - Search on project funding (EDI only)
    - map of search results - auto zoom
    - column sort search results
    - copy query URL and send to friend
    - evaluate upload options -
      * optional checkboxes for skip data upload (metadata only update) manual upload of data file
    - EDI Dashboard
      * shows how services are working
      * lists of private packages etc.
  + *Elections*
    - EXEC
      * R. Brown
      * T. Whitaker
    - co-Chair
      * Remillard
      * Earl
    - Databits editors
      * Porter
      * Bohm
  + Retirements
    - Brian may leave (CWT closing)
    - Wade retiring (but may still be involved)
    - Hope retiring from (NWT)
    - Don Henshaw  retiring (AND)
    - Eda Melendez  retiring (LUQ)
    - Jonathan Walsh (BES winding down)
    - Duane Costa retiring
  + SC Report
    - 40-year report due soon
      * answer criticisms from 30-year review
      * 2 page template from each site providing information
      * create metrics for what we have been doing
        + metrics for good data
      * LTER IM has been in forefront
        + open access for data
        + PASTA and EDI
      * quantifiable, cross site value
      * impact outside LTER
      * timeline - Oct 2019 full document due
      * SC will put together document that will be shared with NSF committee
  + Meeting with NSF LTER WG
    - reps
      * Jenn Burns - OPP - Antarctic - Rotator - oversee PAL and McM
      * Peter McCartney - biological informatics, field station improvements
        + manages EDI
      * John Shay - DEB - new to NSF
      * Colette St. Mary - rotator
      * Stephanie Hampton DEB division director

what would you like to tell us

* + - * network level of IM is strong
        + some issues with individual site web sites
        + no comments on data accessibility
      * list of datasets and DOIs goes a long way towards simplifying a contentious part of review
        + gives reviewer simple way to see if sites are meeting requirements to share data
        + ease with which that was done may lead to requirement across BIO
        + list of DOIs issued by repository is big

in last RFP - pointed to IM intranet site - but not to other guidelines

* + - * how can we better prepare materials for proposals or renewals
      * what is best conduit?
      * response
        + DM Plan requirement

allows flexibility

varies by community

reviewers need to be familiar with that community

* + - * + outdated URL

you are providing documentation of community expectations

NSF provides criteria for reviewers - not LTER

reviewers should know about data policy document

guidelines are more about how you'll accomplish those goals

do you want to be reviewed on how

URL may not have covered all of those

consistency of mid-term reviews - some sites IM meets ahead of review for technical discussion with reviewer

* + - * seemed to be inconsistent across sites - not clear how that works
      * is a valuable process
      * response
        + IM reviewer needs to be full member of site review team
        + can the IM reviewer do technical review virtually

preferred way now

* + - * + timely question - are working on mid-term criteria now
        + want to keep IM as team member
        + program officer is supposed to be present during all parts of a review
      * + - been on both sides - really valuable to be able to talk tech
        + Virtual makes it easier to include multiple people and record
        + I enjoyed remote version more
        + ahead of time is good because there are so many details to deal with
        + IM needs to be at main review in person
        + response

we are leaning towards virtual

issue of cross-listing of data -based on requirements

* + - * especially issues for cross-funded data
      * response
        + NSF doesn't require specific repository for LTER
        + Antarctic has treaty obligation that data be available in Antartic Master Database

can be metadata pointer

* + - * + Want community to select the best ways/databases

has to be put somewhere

NOT self published

* + - * + what is issue? Expertise? Resources?

we produce EML, but if repository won't accept that, need to refactor

hard to crosswalk some standards

* + - * + When do data really need to LIVE in different places vs pointers to different places
        + some repositories may be better for different data repositories - vs requirements by funding sources
        + some repositories have resource limitations relative to accommodating different kinds of data

How do you see NCO and EDI institutional relationships going?

* + - * response
        + big changes in institutional structures
        + experimental phase
        + projects talking among one another
        + pretty good understanding of where project boundaries are
        + want EDI to remove need for sites to provide long-term stable repository
        + DataONE in final year of funding - aggregator function
        + coming up on 40 year review

LTER IM meetings

* + - * plea to find funding mechanism
      * response
        + merit of task force vs standing meetings
        + need to set dialog
        + need to ennunciate what tasks will be

Renewals - of EDI and NCO - changes on horizon

* + - * committed to making resources for IM available - got to have that
      * sustaining track for program for informatics in biology
        + things that don't meet development track at NSF
        + don't need lots of new things - but need to be sustained
        + can have very small, and very large systems
        + setting aside portion of portfolio for that

anyone can apply

channeling mature projects through that

partnership with EDI and NCO going well - stability is important

focused today on vision for IMC - our niche in new landscape - are there things YOU think we should do to help with synthesis

* + - * SC projects highlighted critical role that IM participants played
        + collaboration that allows this group to move forward
        + enabling technology
      * you know how leading edge is moving
      * what is your relationship with scientists and how you interact at your site?
      * when you look at what is happening at NSF
        + NEON coming online
        + Macrosystem Program
        + LTER - over 30 years now
        + GLEON, NPN specialized observing networks
        + sensors at field stations
        + be thinking beyond LTER community

need more like EIMC

* + - * group is well linked into PI's
        + IM in early, starting early
        + getting into other communities
        + bringing IM to front - at beginning is so important
      * Response from IM
        + going well - BUT we could do more
        + we see opportunties with synthesis working groups
        + EDI helps free up time for some archiving tasks
  + Wrap up
    - EML 2.2 session tomorrow AM
      * best practices working groups? How many?
      * EDI can put out RFP
    - get report writeups from visioning