2022 LTER Information Management Committee Annual Meeting Report

*held at Asilomar in conjunction with the LTER All Scientists’ Meeting*

September 19, 2022

Participants (34, from top left): Jason Downing (BNZ), Margaret O’Brien (SBC, EDI), Miguel Leon (LUQ), Emery Boose (HFR), Tim Whiteaker (BLE), Kris Hall (SEV), Sage Lichtenwalner (PAL), Mary Marek-Spartz (MSP), Stace Beaulieu (NES), Adam Kennedy (AND), Suzanne Remillard (AND), Nina Laney (HBR), Hsun-yi Hseih (KBS), Risa McNellis (PIE), Renée Brown (MCM), Yang Xia (KNZ), An Nguyen (BLE), Mark Gahler (NTL), Mike Rugge (FCE), Dan Bahauddin (CDR), Sarah Elmendorf (NWT), Colin Smith (EDI), Corinna Gries (NTL, EDI), Gabriel Kamener (FCE), John Porter (VCE), Li Kui (SBC), Hillary Krumbholz (MCR), Marina Franz (CCE), Chris Turner (NGA), Jim Laundre (ARC), Adam Sapp (GCE), Stevan Earl (CAP), Stephanie Schmidt (AND), Greg Maurer (JRN). Not pictured: Julien Brun, Nick Lyon, and Angel Chen from the LNO.
Table of Contents

Table of Contents 2
Agenda 3
Reports and Updates 4
  Working Group Updates 4
    Unit Dictionary 4
    HyMet 4
    Zotero 4
    Non-Tabular Data 5
    Metabase 5
    Resources 5
  LTER Network Office Report 5
    Resources 6
Business Meeting 6
  Statements from IM Exec nominees 6
  Statement from EB Representative nominee 7
  Election outcomes 7
  Databits 7
Diversity, Equity, and Inclusion Activity 7
EDI Workshop 8
  Resources 9
Panel Discussions 9
  National Science Foundation 9
  National Ecological Observatory Network 11
Alternate group photo 12
Appendix A: Full-day meeting slides 13
Appendix B: LNO report 31
Appendix C: DEIJ and IM Activity report 38
Appendix D: EDI report and workshop slides 39
**Agenda**

<table>
<thead>
<tr>
<th>Time</th>
<th>Morning Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:30</td>
<td>Welcome</td>
</tr>
<tr>
<td>9:00</td>
<td>Working group highlights</td>
</tr>
<tr>
<td>9:20</td>
<td>LNO report</td>
</tr>
<tr>
<td>9:40</td>
<td>IM Exec election (1 member)</td>
</tr>
<tr>
<td>9:50</td>
<td>EB Rep election (1 member)</td>
</tr>
<tr>
<td>9:55</td>
<td>Volunteer for next Databits editor</td>
</tr>
<tr>
<td>10:00</td>
<td>Break</td>
</tr>
<tr>
<td>11:00</td>
<td>DEI activity</td>
</tr>
<tr>
<td>12:00</td>
<td>Lunch</td>
</tr>
<tr>
<td>13:30</td>
<td>EDI workshop</td>
</tr>
<tr>
<td>15:00</td>
<td>Break</td>
</tr>
<tr>
<td>15:30</td>
<td>NSF program managers</td>
</tr>
<tr>
<td>16:30</td>
<td>Group Photo</td>
</tr>
<tr>
<td>16:35</td>
<td>Brainstorming with NEON</td>
</tr>
<tr>
<td>17:00</td>
<td>Adjourn</td>
</tr>
<tr>
<td>18:00</td>
<td>IM dinner</td>
</tr>
<tr>
<td>19:30</td>
<td>Bonfire (reserved by Marty Downs)</td>
</tr>
</tbody>
</table>
Reports and Updates

Working Group Updates

Unit Dictionary

This working group formed during the last year with the goal of creating an updated replacement for the LTER units list. After establishing the scope of the group and explicitly specifying the need for a new system for managing units of measurement, is leaning toward adoption of QUDT (https://qudt.org/), which began within NASA but is now an independent 501c3 corp. Over the next year, this group hopes to make a final decision on which system to use for representing units, work with EDI and the EML development team to understand how the new system will work with their products, and start conversations with the maintainers of QUDT about its adoption and growth in the LTER community.

HyMet

The HyMet working group has continued its work to create a modern replacement for the Clim and Hydro DBs. Over the last year, they've finalized an R package, hyMetDP (https://github.com/EDIorg/hymetDP) to create standardized, ODM-CSV formatted data products from other sources (e.g. USGS), and to find existing standardized data packages. Over the next year this group will continue work on tools for plotting datasets formatted in the ODM-CSV standard, and will continue reaching out to LTER IMs to find representative datasets from LTER sites to be converted into the ODM-CSV/hyMet format.

Zotero

The group added a section to the Zotero best practices for LTER sites (https://bit.ly/ZOTERO_BP) describing how to add a dataset as an item. Since Zotero doesn't have a dataset type, the procedure is to enter the item as a document type and add information in Zotero's Extra field indicating that this is a dataset, for example:

<table>
<thead>
<tr>
<th>Type:</th>
<th>dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version:</td>
<td>3</td>
</tr>
<tr>
<td>DOI:</td>
<td>10.6073/PASTA/CA34BE7554DDC67C9FA0F8DEA01F375B</td>
</tr>
<tr>
<td>Previous Version:</td>
<td>2 DOI:</td>
</tr>
<tr>
<td>Previous Version:</td>
<td>1 DOI:</td>
</tr>
</tbody>
</table>

The group demonstrated that one could use Zotero as a data catalog, which may be useful when a site has datasets archived at several repositories, especially if not all of the repositories are DataONE nodes.
Non-Tabular Data

Having achieved its goal of creating a set of best practices for developing ‘special case’ datasets ([https://portal.edirepository.org/nis/mapbrowse?packageid=edi.726.1](https://portal.edirepository.org/nis/mapbrowse?packageid=edi.726.1)) in 2021, this group had a relatively quiet year. This year, the best practices were updated several times, and they were promoted at several conference sessions, including at a session at ASM 2022 planned by members of this working group and aimed at making scientists more aware of these best practices. This group plans to spend the next year looking for opportunities to work with data managers outside the LTER network to promote the best practices, and updating the EML Best Practices document to include suggestions from these BPs.

Metabase

Work has been on hold recently with LTER Core Metabase. The main effort currently is EML-to-Tables, an R package for converting a set of EML files to tables that can be loaded into a database such as Metabase. This package would aid users in migrating from an existing metadata database or establishing a new database. Once the package is ready, the Metabase team plans to give a webinar on how to create and use Metabase.

Resources

Slides used for the main program of the IMC annual meeting, including working group updates, are in Appendix A.

LTER Network Office Report

Julien Brun from the LTER Network Office (LNO) delivered an update about recent developments there. There were a number of personnel updates to relay. The LNO has hired Nick Lyon and Angel Chen as data analysts with a focus on aiding synthesis groups. LNO has also hired Ingrid Slette as a postdoctoral fellow to explore impacts of compound precipitation extremes on ecosystem processes. Nick Lyon, Angel Chen, and Gabe De La Rosa (LNO’s communication liaison) were also present at the meeting to represent the LNO.

There are a number of funding and outreach programs that are active at and coordinated at the LNO. Synthesis science continues to be a big focus for LNO and much of this year’s update focused on current and upcoming synthesis group activities. LNO is hosting workshops to further educate synthesis groups about reproducible workflows, coding, collaboration, and communication, and they have published a [website detailing NCEAS scientific computing support](https://www.nceas.ucsb.edu/collaborative-scalability). There is currently an RFP for new synthesis groups for approximately $100,000 over a 2-3 year period, with proposals due in October 2022. In an assessment of benefits and barriers after an initial synthesis meeting, LNO reported the top benefit is hearing others’ ideas, while the top challenge is lack of researcher time. There was also some discussion of a new cross-site Research Experience for Teachers (RET) opportunity with a focus on biodiversity and climate change. The RET is at three sites with overlapping two-year teacher cohorts.
Julien also described some new tools being developed at LNO that are directly applicable to the work that LTER Network IMs do. The LTERHub is a website to connect people, discussions, and events related to the LTER network. Users can check whether they are registered at https://lternet.edu/directory. Julien reported on uptake of LTER Hub functions across the network. The LNO also created a new LTER Information Management Manual website, which is essentially a guide for getting started with IM in LTER. This site is a quick reference for tasks such as updating personnel lists, network publications, and writing EML, and replaces similar functionality from the now-defunct former LTER IM website.

Resources

Slides for the LNO update are in Appendix B.

Business Meeting

There were several elections and committee assignments made during the 2022 IMC meeting. Tim Whitaker ended his term on the IM Executive committee (IM Exec), and there were two nominees to take the resulting open position. We also elected a new IMC representative to the LTER Executive Board and assigned new members to the DataBits editorial team.

Statements from IM Exec nominees

Sara Elmendorf (NWT):

“I am interested in the role because I have benefited immensely from the LTER IM community, and feel that in order to continue this great tradition we all need to take a turn in the leadership positions that make this happen. I was reluctant to volunteer while we were ramping up for our renewal, but I am optimistic that the renewal craziness is on pause for the next few years.”

Mary Marek-Spartz (MSP):

“I greatly appreciate the nomination and I welcome the opportunity to run for a position on the LTER Network IM Exec committee. I am a new IM that joined the MSP LTER when it began in 2021. Prior to that, I worked on a long-running invasive plant biological control research project at the University of Minnesota, starting after undergrad as a Junior Scientist (2013) through the completion of my PhD in Entomology (2022). The experience gave me a great appreciation for the long and thorough investigation that goes into a weed biocontrol program, and the importance of building upon many years of information gathered on the intricate dynamic systems. I love spatial data, and I am always excited to talk about what free and open-source GIS can contribute to long-term ecology. I received my M.S. GIS in 2015, and continued my involvement with the Geo-Analytics program as an adjunct instructor where I focused on making programing principles and open-source software accessible to new GIS students. I have developed applications and packages in R, Python, and JavaScript for a range of purposes from interactive mapping to population modeling. On the IM Exec committee, I would love to promote the importance of developing non-tabular data protocols and to continue efforts to lower barriers of entry to data management and analysis tools for LTER researchers and all consumers of our data.”
Statement from EB Representative nominee

Mary Martin (HBR):

“I have been pinch-hitting as EB rep for the past few months, and would be pleased to continue serving in that role if elected this year. In both IMexec and EB, discussions and reporting have provided an interesting and informative window into network-wide operations. I anticipate that in the near future there will be EB discussions relevant to the White House Open Data Guidelines released last week. Having read those, I look forward to working with IMC and IMexec to bring the IMC perspective to EB, so that LTER responses to these new guidelines align well with our collective skillsets and resources.”

Election outcomes

Sarah Elmendorf was elected to fill the open seat on the IM Executive committee, and Mary Martin was elected as EB Rep.

Databits

Hsun-Yi Hsieh (KBS) and Mary Marek-Spartz (MSP) finished their first year as Databits co-editors. Hsun-Yi stepped down, with Mary continuing as an interim editor and Marina Frantz joining the editing team. In fall of 2022 Dah Bahauddin (CDR) will join Databits as an editor, and Mary will step down.

Diversity, Equity, and Inclusion Activity

Adam Sapp and Chris Turner led a DEI activity that began with an introduction to the following publications:


Attendees then brainstormed how to implement these ideas. The notes from these brainstorming sessions can be found in the DEI Report out document and a summary of these notes appears in Appendix C. Some key ideas and recurring themes from the activity include:

- Develop and showcase opportunities for underrepresented groups/individuals and for groups/individuals that are not at the top of the academic hierarchy (undergraduates, technicians, early-career)
• Start discussions/planning with large inclusive group and then scale back instead of starting with small group and scaling up
• Hold regular informal meetings or workshops (or even office hours) to introduce students to LTER data processes and standards. The word “informal” was used often, suggesting that formal academic settings may not foster open communication.

IM Exec will continue to coalesce these ideas into a DEI document for the IMC. We envision this document providing some measurable actions we can take as the IMC as well as some ideas each IM can employ at their individual site.

EDI Workshop

Corinna Gries and Colin Smith led an Environmental Data Initiative workshop which included three parts:

1. Updates from EDI
2. Demonstration of ezEML, a browser workflow for creating EML
3. Demonstration of ezCatalog, a quick way of creating an online data catalog, using GitHub pages, of your datasets published in EDI

The most important update is that EDI's funding has been renewed for a three year period, which is excellent news. NSF is expecting further progress towards securing outside funding streams for EDI in upcoming proposal rounds. There are some EDI staff members nearing retirement (or already retired), and the team is dealing with other personnel turnover. EDI will be looking to hire to fill some gaps in the coming year. Kyle Zollo-Venecek recently departed and EDI will be hiring a replacement this winter. There are also changes coming to the EDI Data Fellows program. The program may diminish in size and begin assigning fellows to EDI-specific data projects rather than allowing external sites to bid for fellows to work on their projects.

The EDI repository has also received a number of upgrades in the past year or so. Markdown and LaTeX support for metadata display on dataset landing pages and the detailed view is now available. The report generation tools on the EDI website have been upgraded and Corinna demonstrated ways to view graphs and statistics on data package downloads.

The EDI team has recently released the ezEML tool on their website. This is a web-based form for creating EML documents that has built-in consistency and completeness checks. As Corinna demonstrated, ezEML is user-friendly enough for researchers without information management training to use, but teams with information managers will want to provide some oversight to ensure consistency with other site datasets. EML documents can be created from scratch or by using templates to pre-populate EML elements. Entire EML documents, or individual elements, can be imported from published datasets already in the EDI repository, allowing further templating and metadata re-use capability. There are still some metadata elements that ezEML has incomplete support for, such as provenance and semantic annotations, but this functionality is in development for future releases of the tool.
Colin demonstrated ezCatalog, a JavaScript+GitHub solution for generating a data catalog. It is based on BLE LTER’s PASTA-JavaScript-Search-Client, with the addition of GitHub actions that enables users to generate a catalog hosted in GitHub pages for embedding or linking in a website. This tool is being used as a website data catalog at a small handful of LTER sites already.

Resources
Slides for the IMC annual meeting update and workshop are in Appendix D. EDI also presented at the town hall on Thursday (the latter has more screenshots).

Panel Discussions

National Science Foundation
Three NSF program managers (PMs) from the LTER working group, Doug Levey, Peter McCartney, and Cynthia Suchman, spent an hour in a panel discussion with the IMC in the afternoon. The discussion covered multiple topics centered on NSF support and expectations for IM-related activity in the LTER Network, EDI, and partner research networks. IMC had prepared some questions in advance and sent them to the PMs before the meeting. Some of these questions were motivated by the recent report from the Decadal Review Committee (DRC) that evaluated the LTER Network over the past few years. IMs wondered how they might contribute to, and procure resources for, increasing calls for synthesis-oriented science. IMs also wanted to know how NSF would respond to the DRC report, and what it might expect from an LTER network, and IMC, response. There were also questions about NSF support of, and expectations for, EDI. These and similar questions guided the discussion, and there were a number of important takeaways.

The NSF PMs emphasized several times that the LTER budget is fixed, so initiatives to enhance LTER Network synthesis science and information management, like data harmonization efforts and cyberinfrastructure, will have to seek external support. As usual, this presents a challenge for IMs who have limited time and material resources and may need extra support, either from NSF or their site leadership, to begin writing or contributing to proposals. This very topic was re-visited later during ASM as IMs began crafting the IMC response to the Fourth Decadal Review of the LTER network. The NSF program officers named a number of specific programs that might be suitable for supporting IM-related projects in the LTER network. A few of these were:

1. Environmental Data Science Innovation & Inclusion Lab (ESIIIL), an NSF-funded data synthesis center based at University of Colorado Boulder. Though the center is new, starting some discussion with its directors (Jennifer Balch, who attended ASM) could lead to productive collaborations with the LTER network and IMC.
2. **Accelerating Research through International Network-to-Network Collaborations (AccelNet)**, which provides support for “grand-challenges” oriented scientific initiatives that require international and cross-network collaborations.
3. **Cyberinfrastructure for Sustained Scientific Innovation (CSSI)** is a program to fund emerging needs in cyberinfrastructure.
4. **NSF’s Division of Biological Infrastructure** is always on the lookout for extensible infrastructure and data initiatives that will have use-cases and benefits across NSF programs. Peter McCartney is probably a good contact for discussing these opportunities.

There was some specific discussion, spurred by a question from IMC, about how to improve the handling of biological specimens and physical collections in the LTER Network. The NSF PMs are well aware of this conundrum and noted that in 2015, NSF’s instructions in the LTER program solicitation started to make clear that samples are data, and that they should be archived and made accessible. Peter McCartney alluded to the “Collections in Support of Biological Research” program (CSBR, Reed Beaman as contact) as one potential avenue of supporting development of an LTER collections program. The group was also supportive of collaborations with NEON and other NSF programs (the DBI/OAC funded “Internet of Samples” program, and “Sampling Nature” RCN).

On a more general note about supporting new data management initiatives, the three program managers encouraged LTER sites, IMC, and EDI to reach out to them for suggestions and advice about how to win support for such programs. They even suggested putting together “one pager” documents outlining these initiatives and sending them to the LTER working group for feedback and recommendations on NSF programs to write proposals to. Doug, Cynthia, and Peter all expressed an interest in acting as a resource for LTER personnel (IMs included) to successfully navigate the complexities of NSF programs and proposal procedures.

On the topic of NSF, LTER, and IMC responses to the DRC report, and the recently released **OSTP memo** on research data access, NSF was somewhat reserved. The PMs scheduled a listening session for the following day at ASM and were hoping to hear more from investigators, network leadership, and IMC. The PMs mentioned that they don’t view the DRC report as something that will have immediate consequences, but it might lead to some longer-term shifts in priorities. For instance, NSF PMs didn’t share an opinion on whether the LTER Network, in general, was meeting its expectations for data management, or on whether the network should develop an new “Data Management Mission Statement” as suggested by the DRC. But they did suggest that the DRC report should at least generate some discussion on these topics. NSF responses to the OSTP memo are also still muted. We briefly discussed the LTER Networks relative responsibility to publish “primary” (long-term, core) versus “derived” (one-off, linked to articles) datasets, but there are still open questions on whether NSF or the LTER Network will need to reorder those priorities in response to the memo.

Finally, there was an open discussion on NSF support for EDI. The NSF grant supporting the EDI repository has been renewed as “sustaining funding”, but the reviews from this renewal
suggested that EDI would be expected to generate more of its own revenue in the future. We discussed this with the NSF PMs, but there are still questions on future support. In general, NSF appears willing to fund EDI for the near term, and they acknowledged that there is a fairly solid expectation of support for EDI even above NSF’s LTER Working Group, which is encouraging. Nevertheless, they did acknowledge that NSF does not like to be in the business of supporting infrastructure projects in perpetuity, so EDI needs to find ways to generate more of its own revenue. There are a number of possibilities here, including charging for higher-level data curation services (similar to DataONE’s efforts), consulting revenue, or finding new ways (unspecified) for the community to cover the responsibility and cost of running the repository. NSF alluded to some programs in Ocean Science that do this somewhat successfully. Whatever the way forward, ideas should be generated, in part, by the communities that EDI serves.

National Ecological Observatory Network

Christine Laney, a principal research scientist at NEON overseeing its eco-informatics program, joined the IMC meeting for a 20-minute discussion at the end of the day. Christine first briefly introduced the NEON program, its scientific goals, its primary data, and its data management program. In a sense, NEON is the LTER Network’s sibling network within the NSF-funded ecological sciences. Now that NEON is coming into its own, there are some opportunities to start learning more about each other and complementing each other’s different strengths. This applies in both the scientific domains and in the data management domain. After Christine’s introduction we had a discussion on some of these areas for collaboration.

Christine discussed and demonstrated a wide range of features in NEON’s data management systems. Of particular interest were the data catalogs, both for sensor network data and biological/physical collections. These catalogs are a cohesive and comprehensive portal to access data for all NEON sites, and this is in part made possible by the highly standardized suite of instrumentation and resulting data streams at each NEON site. The data management behind the scenes is well developed and might be an interesting area for investigation by LTER IMs. Many of the data management workflows (QA/QC, metadata aggregation, data publication) are running in R and python containers that are managed with the Pachyderm data pipeline platform. Overall there is a great deal of data science infrastructure and expertise at NEON that could be fruitful for IMs to interact with.

There are several areas for collaboration between LTER and NEON. The most probable and beneficial collaborations might focus on two priorities: 1) standardizing similar data collected at LTER and NEON sites into common formats, and 2) unifying ecological data discovery across the LTER and NEON networks. There is also interest amongst LTER sites in collaborating with NEON to archive physical samples and standardize sample collection methods. There are already some collaborations between EDI and NEON on data harmonization and publication of standardized data products. IMs were encouraged to attend a few sessions at the All Scientists’ Meeting to explore those topics in more detail.
Alternate group photo
Appendix A: Full-day meeting slides

Slides for the full meeting program are below (or Google Drive: 2022 IMC Annual Meeting). Slides from the LNO report and EDI’s report and workshop are in Appendix B and D, respectively.

2022 IMC Annual Meeting
Making your wildest data dreams come true*

Network: Asilomar Conference
Password: conference
but please don’t use all our bandwidth :)

*dreams not guaranteed to come true, be wild, or have anything to do with data

Welcome

Introductions
• My name is _____
• I’m at ___ LTER
• I’m excited about this ASM session: _____
• Ask me about ___

____
## IMC Annual Meeting Agenda

<table>
<thead>
<tr>
<th>Time</th>
<th>Morning Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:30</td>
<td>Welcome</td>
</tr>
<tr>
<td>9:00</td>
<td>Working group highlights</td>
</tr>
<tr>
<td>9:30</td>
<td>LNO report</td>
</tr>
<tr>
<td>10:00</td>
<td>Break</td>
</tr>
<tr>
<td>10:30</td>
<td>IM Exec election (1 member)</td>
</tr>
<tr>
<td>10:40</td>
<td>EB Rep election (1 member)</td>
</tr>
<tr>
<td>10:50</td>
<td>Volunteer for next Databits editor</td>
</tr>
<tr>
<td>11:00</td>
<td>DEI activity</td>
</tr>
<tr>
<td>11:45</td>
<td>Group photo</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Afternoon Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>12:00</td>
<td>Lunch</td>
</tr>
<tr>
<td>13:30</td>
<td>EDI workshop</td>
</tr>
<tr>
<td>15:00</td>
<td>Break</td>
</tr>
<tr>
<td>15:30</td>
<td>NSF program managers</td>
</tr>
<tr>
<td>16:30</td>
<td>Brainstorming with NEON</td>
</tr>
<tr>
<td>17:00</td>
<td>Adjourn</td>
</tr>
<tr>
<td>18:00</td>
<td>Dinner</td>
</tr>
<tr>
<td>19:00</td>
<td>IM bonfire</td>
</tr>
</tbody>
</table>

## Working Group Updates
Clim/Hydro DB 2.0 (HyMet) &

hymetDP: A Dataset Design Pattern for Hydrological and Meteorological Data

Project Overview

- Researchers needed standardized met and hydrologic data across sites
- ClimDB/HydroDB 1.0 - All LTER sites contributed standardized versions of their met and hydro data. Tools enabled data access and visualization.
- ClimDB/HydroDB 2.0 (hymetDP) - Based on CUAHSI ODM for hydrology and meteorology data
Create, Search, and Use hymetDP Data

hymetDP R Package (https://github.com/EDlorg/hyemetDP)

- Create from EDI data package or from other sources (e.g. USGS) “on-the-fly”

- Search an index of hymetDP data packages
- Read a hymetDP dataset and plot (WIP)

LTER Community Engagement

We’re reaching out to LTER IMs

1. Select 1 or 2 representative met/hydro datasets
2. Convert them to HyMet format
3. Upload to EDI Staging
4. Educate IMs on the process

Six hymetDP datasets on EDI staging

Search https://portal-s.edirepository.org using the “hymetDP” search term

If you are interested let us know
Other Community Engagements

We have also reached out to

1. USFS Experimental Forests & Watershed scientists and data managers
2. The Silicate Exports LTER synthesis working group
   a. This includes LTER and USGS scientists

to promote and educate about our efforts. Both are eager to participate, but will probably need an assist to start converting datasets.

Once more datasets are converted and features tested, we will re-engage.

There are still lots of open questions about how this will work!!!
Non-Tabular Data

The Non-Tabular Data WG is dead!

Long live the Non-Tabular Data WG!

- Data Package Design for Special Cases lives on and is updated periodically
- We’ve run conference sessions on this theme (including 1 at ASM)
- Next steps:
  - Update the 2017 EML Best Practices document?
  - Expand our scope & participants and try to fund more work in this area (FAIROS RCN)?
  - Any volunteers?

Units!
Working Group Charge

Determine best replacement backend

Mechanism to export unit and description for use in EML

Units element for EML

<standardUnit>millimolePerLiter</standardUnit>
<customUnit>microeinsteinPerMeterSquaredPerSecond</customUnit>

STMML for custom units

<stmml:unit id="microeinsteinPerMeterSquaredPerSecond"
name="microeinsteinPerMeterSquaredPerSecond" parentSI="joule" unitType="energy"/>

Kai Blumberg¹,², Simon Cox², Hajo Rijgersberg³, Chris Mungall⁴, James Overton⁵
Units of Measurement (UOM): Harmonizing Units of Measure Vocabularies on the Web. International Conference on Biomedical Ontology (ICBO)

unit backends

Existing Unit Vocabularies

UCUM Pattern for the unit symbols, widely used in biomedical. In FHIR (Fast Healthcare Interoperability Resources).

https://ucum.org/

Notable mention: CODATA working group ‘Digital Representation of Units of Measure’

https://units-of-measure.org/
Next steps for Working Group

Determine which identifier to key on: QUDT or UCUM?

Consult with eml-dev on proposed use of stmml
  id field holding unit URI, other stmml fields of interest
  ramifications of representing all units as "<customUnit>"

Public presence for WG materials
  e.g., documentation, BPs (git? wiki?)

Check in with EDI
  Limitations, their preferences

Start conversation with QUDT
  Assemble unit additions

---

Units Session
Wednesday 1030
Oak Shelter

---

Zotero for Data

Extra:

  Type: dataset
  Version: 2
  DOI: 10.6073/PASTA/F7204A847A1D71FCE18ED880363E62F8
  Previous Version: 1 DOI: 10.6073/pasta/8f9188936b08ef7cbe84ce3077487d6a

LTER Network Office Updates

Break Time!
### Election for IM Executive Committee Member

- The committee plans VWGs and the annual IMC meeting, prepares reports, and coordinates IMC activities
- Meet monthly via Zoom
- One opening, three-year term (2022-2025)
- Nominees
  - Sarah Elmendorf
  - Mary Marek-Spartz

### Election for Executive Board Representative

- The LTER Executive Board develops bylaws and strategies for the LTER Network
- Nominees must have served or currently serve on IM Exec

- The IMC EB Representative brings to this group the interests of and input from the IMC
- One opening, three-year term (2022-2025)

- Meet monthly-ish with EB, and monthly with IM Exec
Volunteers for Databits Editor

Diversity, Equity, and Inclusion Activity
Diversity, Equity, and Inclusion Activity

There are many cultural, economic, institutional, and social barriers in academic research.

Dismantling these barriers will require large-scale structural changes in both research institutions and society as a whole.

Ten simple rules to cultivate a sense of belonging in collaborative data science research teams.

The CARE principles for Indigenous Data Governance

10 Simple Rules to Cultivate Belonging in Collaborative Data Science Research Teams

Kaitlyn Gaynor, Therese Azevedo, Clarissa Boyajian, Julien Brun, Amber Budden, Allie Cole, Samantha Csik, Joe DeCesaro, Halina Do-Linh, Joan Dudney, Carmen Galaz García, Scout Leonard, Nicholas J. Lyon, Althea Marks, Julia Parish, Alexandra A. Phillips, Jai Ranganathan, Courtney Scarborough, Joshua Smith, Marcus Thomson, Camila Vargas Poulsen, Caitlin R. Fong
The CARE Principles for Indigenous Data Governance

<table>
<thead>
<tr>
<th>Collective Benefit</th>
<th>Authority to Control</th>
<th>Responsibility</th>
<th>Ethics</th>
</tr>
</thead>
<tbody>
<tr>
<td>For inclusive development and innovation</td>
<td>Recognizing rights and interests</td>
<td>For positive relationships</td>
<td>For minimizing harm and maximizing benefit</td>
</tr>
<tr>
<td>For improved governance and citizen engagement</td>
<td>Data for governance</td>
<td>For expanding capability and capacity</td>
<td>For justice</td>
</tr>
<tr>
<td>For equitable outcomes</td>
<td>Governance of data</td>
<td>For Indigenous languages and worldviews</td>
<td>For future use</td>
</tr>
</tbody>
</table>

Adapted from Carroll et. al. 2020

Data Principles

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Authority</td>
<td>Self-Determination</td>
<td>Inherent Sovereignty</td>
<td>OCARP</td>
<td>Open By Default</td>
<td>Findable</td>
<td>Sovereign</td>
</tr>
<tr>
<td>Relationships</td>
<td>Available and Accessible</td>
<td>Indigenous Knowledge</td>
<td>Indigenous Knowledge</td>
<td>Timely and Comprehensive</td>
<td>Accessible</td>
<td>Trusted</td>
</tr>
<tr>
<td>Obligations</td>
<td>Collective Rights and Interests</td>
<td>Ethics</td>
<td>Methodology and Approaches</td>
<td>Accessible and Usable</td>
<td>Interoperable</td>
<td>Reusable</td>
</tr>
<tr>
<td>Collective Benefit</td>
<td>Accountability</td>
<td>Intergenerational Collective Wellbeing</td>
<td>Evidence to Build Policy</td>
<td>Comparable and Interoperable</td>
<td>Reusable</td>
<td>Exchangeable</td>
</tr>
<tr>
<td>Reciprocity</td>
<td>Exercise Control</td>
<td>Relationships</td>
<td>Ethical Relationships</td>
<td>For Improved Governance &amp; Citizen Engagement</td>
<td>Actionable</td>
<td>Measurable</td>
</tr>
<tr>
<td>Guardianship</td>
<td></td>
<td>Data Governance</td>
<td></td>
<td>For Inclusive Development and Innovation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Adapted from Carroll et. al. 2020
Activity

Divide into 6 groups
Discuss the “Ten simple rules to cultivate belonging in collaborative data science research teams” and consider:

- How you can practice each of the simple steps at your site?
- How we can do so within the LTER Information Management community?
- How the CARE principles fit into these?
- How we can make all of this fun?

Report out from groups (15 mins)

Outcomes

Synthesis Outcomes:

- Synthesize discussion into priorities and actionable outcomes for IM community, return this as a doc to IMC
- Group decides what to do with the outcomes decided upon in the discussion portion

Session outcomes:

- 1 or more concrete things we can do within the IM community to better incorporate the CARE Principles into our data policies and practices or to cultivate a sense of inclusiveness and belonging
Group Photo

EDI Workshop

EDI slides
Brainstorming NEON-LTER Collaborations

- standardize similar data collected at LTER and NEON sites into common formats?
- unify ecological data discovery across LTER and NEON?

Dinner at the dining hall
7p-9p Bonfire near Surf & Sand room
Appendix B: LNO report


New People and Activities

Ingrid Siette, LTER Postdoc
PhD Colorado State University
Impacts of compound precipitation extremes belowground

Nick Lyon, Data Analyst
Community ecologist turned data scientist; Former data scientist for the Herbivory Variability Network

Angel Chen, Data Analyst
B.S. UC Santa Barbara
Statistics degree; Former data curator for the Arctic Data Center

New support activities

- Short workshops and coaching for synthesis groups. Current topics:
  - Facilitation and virtual facilitation,
  - GitHub and collaborative analysis
  - Coding with the tidyverse
  - Science communication/visualization
- Seats in weeklong “Reproducible Research for Synthesis” short course for up to 3 participants per synthesis group
- Analytical sprints - discrete analytical tasks assigned to data analysts (3-4 weeks, full-time)
Synthesis Activities
For details on synthesis groups, see: https://ternet.edu/working-groups/

Enabling Synthesis Working Groups

Our Goal
Enable Participants to do science differently, more collaboratively and reproducibly
**Data Science Support**

**Our Goal**

*Working with participants to develop reproducible analyses*
- Iterate quickly
- Integrate new information easily
- Programming approach for reproducibility

*For them, their collaborators, and their future them*

[https://nceas.github.io/scicomp.github.io/](https://nceas.github.io/scicomp.github.io/)

---

**Participation in Synthesis**

*Count of Unique Synthesis Participants*

<table>
<thead>
<tr>
<th>Year</th>
<th>LTER</th>
<th>Non-LTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>2017</td>
<td>25</td>
<td>15</td>
</tr>
<tr>
<td>2020</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>2021</td>
<td>15</td>
<td>8</td>
</tr>
</tbody>
</table>

*Funding Cohort*

<table>
<thead>
<tr>
<th>Cohort</th>
<th>2016</th>
<th>2017</th>
<th>2020</th>
<th>2021</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTER</td>
<td>30</td>
<td>25</td>
<td>20</td>
<td>15</td>
</tr>
<tr>
<td>Non-LTER</td>
<td>10</td>
<td>15</td>
<td>10</td>
<td>8</td>
</tr>
</tbody>
</table>

*Figure 2. Participation of LTER-affiliated and non-LTER affiliated researchers in LTER Network Office synthesis working groups.*

---

*Figure 1. Count of participants in LTER synthesis groups with a primary affiliation to a site. Information managers (IM) are counted separately. Funding represents the year in which the working group was funded. Site order is from first to last funded, from top to bottom.*
New Synthesis Opportunities

2022 Request for Proposals

- ~$100 K covers travel and publication charges, 2-3 year duration
- Includes analytical, logistical, and facilitation support
- Announcement: June 2022; Due: October 2022

Possible small awards to capture energy and ideas emerging from ASM
- Could address both network coordination and scientific goals

Assessment

Expected Benefits at end of 1st Meeting

Expected Challenges at end of 1st Meeting
**LTERHub**: Find and connect with people, discussions, events

1,188 potentially active users as of 2022.05.16 …of which 562 have logged in

To check whether you are registered and under which email: [https://lternet.edu/directory](https://lternet.edu/directory)

To login or register a new user: [https://lternet.edu/lterhub](https://lternet.edu/lterhub)
Committee Discussion Topics

**Lead PIs**
- Coping with COVID
- Q&A with NSF Program Officers
- Joint Meeting with DEI Committee
- Sample Archives

**Information Managers**
- LTER IM Manual
- Site Unique Identifiers
- ClimHydro DB next steps
- Tools exchange
- EML generation
- DataONE Portals
- Best practices for non-tabular data
- Data use policy
- Static website generators

**Education/Engagement**
- Site Highlights
- ASM Planning
- Social Justice in Education
- Assessment
- LTER DataSampler
- DataNuggets/Data Classroom
- Young Voices of Science
- Science Education Resource Consortium (SERC)

**DEIJ**
- **Active Working Groups:**
  - Community-building
  - Field Safety
  - Resources
  - Fundraising
- **Topics:**
  - Planning for joint meeting with PIs
  - Climate assessments
  - Site’s struggles and successes
  - Guest speaker: Dr Gina Forrest on building support for DEIJ Initiatives
New Cross-Site RET on Biodiversity and Climate Change

Authentic Research Experiences for Teachers at LTERs (ARET@LTERs)

- 3 sites
- Strong partnership between site educators and investigators
- Overlapping 2-year teacher cohorts
- Shared data-focused experience in second year
- Recruitment from majority-minority districts

https://lter.github.io/im-manual/

Up next:
- Education/Outreach Manual?
- DEIJ Best Practices?
Appendix C: DEIJ and IM Activity report

1. *Lead from the front:* Self-educate, discuss, and reduce barriers to entry and retention in data science research.

Sites are thinking about this already. The LTER K-12 Schoolyard and REU programs are meant to reach out to wider groups of people. We could also pursue programs that support underrepresented groups and offer trainings and workshops that are accessible to a wider audience (undergraduates, indigenous populations, people with physical limitations, etc.). The IMC should reach out to the LTER DEI committee to coordinate efforts. Collecting demographic information will be important for tracking progress. We must realize that these are big challenges and that there are no quick fixes.

2. *Seeing is believing:* Highlight the diversity of people, research, and accomplishments in data science.

Expanding authorship to non-PhD participants (e.g. undergraduate students, technicians) will allow these groups to gain the experience and highlight different paths available to someone in data science. Social media can be useful in sharing these success stories to a wider audience. However, we must be careful to not tokenize minority participants.

3. *Skin in the game:* Design research questions that are relevant to your research team members.

No group addressed this rule.

4. *Comfort through clarity:* Set clear expectations around coding practices and workplace conduct.

Use collaborations to set expectations and identify best practices. New students/technicians may not be aware of implicit norms and expectations; communally agreed upon expectations will allow people to participate in setting standards and ensure they know what is expected.

5. *All codes lead to Rome:* Embrace different modes of coding and communicating.

No group addressed this rule.

6. *People first:* Prioritize needs of team members in project scheduling and planning.

Start with everyone at the table! Including everyone at the beginning of discussions (and then pulling back if necessary) ensures that everyone’s voice is heard and also brings more diverse
ideas to the table. Remote/hybrid work is a way to be more inclusive and allow more people to be involved.

7. **Empowerment through ownership**: Create opportunities for ownership, leadership, and development among all team members.

Recognize/announce when a student/technician does something new or noteworthy. Ensure that the person responsible gets credit for their work they completed. Consult with students and share analysis to point out opportunities. Consult with PI to on what opportunities we can create and broadcast these (and other) opportunities as they become available.

8. **Open science**: Practice transparent and reproducible research within and outside of your research group.

Open collaboration from the start of a research project or publication. However, different views of authorship can make it difficult to always give appropriate credit (who should be included?).

9. **Safe learning spaces**: Create low-stakes environments to promote data science skills growth.

Have weekly office hours/workshops to introduce new participants to data packages/processing routines. Make time to meet with students one-on-one to discuss data submissions. Most importantly, make these meetings informal and low-stakes to encourage open participation.

10. **Have fun!**

Have fun, informal communication to make it easier to connect and share information and skills. Identify and acknowledge the ‘not fun’ aspects of data science and work to find alternatives. Remember, intensity is not a prerequisite – even for important work.

**Appendix D: EDI report and workshop slides**

PDF slides on Google Drive: [EDI_presentation_IM_meeting.pdf](#)
EDI Developments Supporting Site Data Management

EDI team

EDI updates

Funded for another 3 years
Kristin and Kyle left for other jobs
Susanne retires
Fellowship program change
Currently hiring to cover Kyle's tasks
Accomplishments

Supporting Markdown and LaTeX

Reporting tools
Reference Location or DOI | Reference Title | Reference Location or DOI | Reference Title | Reference Location or DOI | Reference Title
|---------------------------|-----------------|---------------------------|-----------------|---------------------------|-----------------|
LaTeX how to

[https://users.dickinson.edu/~richesod/latex/latexcheatsheet.pdf](https://users.dickinson.edu/~richesod/latex/latexcheatsheet.pdf)

Gotcha: don’t include LaTeX in markdown tags but separate para tags, works only on full metadata page for now

\[
\frac{D_{\mathrm{xy},t}}{L_{\mathrm{xy}}} = \frac{d_{\mathrm{Lj_t}}}{l_{\mathrm{Lj}}} \tag{1}
\]

\[
\frac{D_{\mathrm{xy},t}+D_{\mathrm{yz},t}}{(L_{\mathrm{xy}}+L_{\mathrm{yz}})} = \frac{d_{\mathrm{Lj_t}}}{l_{\mathrm{Lj}}} \tag{2}
\]

---

**Reporting**

Csv download of report
Online report paginated
ezEML

For unrelated data
For site related one-off data
For site long-term data management

ezEML demo and discussion

https://ezeml.edirepository.org/
ezCatalog demo and discussion

https://github.com/EDlorg/ezCatalog

Reporting with EDlutils?

https://github.com/ropensci/EDlutils
https://docs.ropensci.org/EDlutils/