LTER Information Management Annual Committee Report (2023-2024)

Reporting on activities from May 2023-May 2024

1. **Date of report**: May 27, 2024

2. **Name of Committee**: Information Management Committee (IMC)

3. **Name and site affiliation of the Committee Chair (or co-chairs or Executive Team)**: Sarah Elmendorf (NWT) and Mary Martin (HBR)

4. **List of current members**: Greg Maurer (JRN; EDI), Suzanne Remillard (AND), Renee Brown (MCM), Kate Morkeski (NES), Mary Marek-Spartz (MSP), Jim Laundre (ARC), Yang Xia (KNZ), Chris Turner (NGA), Stevan Earl (CAP), Miguel Leon (LUQ), Adam Sapp (GCE), Jason Downing (BNZ), Li Kui (SBC), Sarah Elmendorf (NWT), Hillary Krumbholz (MCR), John Porter (VCR), Stephanie Schmidt (AND), Sven Bohm (KBS), Mary Martin (HBR), Marina Frants (CCE/SIO), Sage Lichtenwalner (PAL), Nick Lyon (LNO), Marty Downs (LNO), Gabriel Kamener (FCE), Induja Mohandas (BLE), Corinna Gries (EDI), Mark Gahler (NTL), Dan Bahauddin (CDR), Ara Winter (BEMP/SEV), Risa McNellis (PIE), Mark Gahler (NTL), Emery Boose (HFR).

5. **How membership is determined (identified by sites, voluntary, appointed by the Executive Board, etc).**
   Membership and roles are outlined in the LTER IMC Bylaws (v.4 ratified 2021-07-27). Membership includes an Information Manager (IM) from each site that serves as the primary site IM contact. Additional members may be identified by individual sites, by the LTER Network Office (LNO), and by the Environmental Data Initiative (EDI) from among site or project personnel who are involved with information management. The Information Management Executive Committee (IM-Exec) is the steering committee for the IMC. IM-Exec members, including the IM-Exec Chair or co-Chairs and a representative to the LTER Network Executive Board (EB), are nominated and ratified by the IMC committee as outlined in the IMC Bylaws.
   IMexec is the executive board of the IMC. Current members are: Martin (HBR; 2024 co-chair and EB-rep), Elmendorf (NWT; co-chair 2025), Sapp (GCE; 2024), Turner (NGA; 2024), and Kui (SBC; 2026). Marty Downs (LNO), and Greg Maurer (JRN and EDI) participate in and provide reports during the monthly IM-Exec meetings.

6. **Meeting frequency**: Monthly in Virtual Watercooler (VWC), and annually (either in-person or virtual). See section 10 for future annual meeting plans.

7. **Major activities or accomplishments for the year**:

   - The 2023 IMC Annual Meeting was held on July 17, 2023 in Burlington VT, immediately preceding the Earth Systems Information Partners (ESIP) weeklong

meeting, with a theme of ‘Opening Doors to Open Science’. As always, this annual meeting provided a time for IMC members to complete necessary community maintenance and stewardship activities, connect and communicate with partners outside the IMC, and collectively work on IMC tasks. Highlights from the 2023 IMC meeting include:

- IMC members elected new members to the IMC Executive Committee and welcomed new editors for the Databits newsletter.
- Active IMC working groups each shared their work completed over the past year and what each hopes to accomplish over the current year.
- Presentations were given by partners essential to the IMC from the LTER Network Office, EDI, and the ILTER.
- IMs from several sites gave lightning talks to share new tools and techniques for information management tasks and challenges.
- Both the Units and EML Best Practices Working Groups hosted workshops aimed at providing more detail on the work they've done, demoing tools and processes for their work, and soliciting feedback from IMs on each working group’s progress and path forward.

- The IMC holds a monthly water cooler session to discuss pertinent issues related to the Network or informatics generally. Notable topics this reporting period included:
  - LTER Site GitHub Organization Recommendations (LNO)
  - Making use of new cloud platforms, data formats, and services
  - LNO presentation on Discourse site
  - LNO presentation on planned courses
  - Graduate student engagement
  - ESIIL and its big data solutions to LTER Researchers
  - IMs and PIs discussion on sample preservation
  - ezEML - overview and what’s new, with followup session for IMs interested in ezEML adoption

8. **Currently active subcommittees or working groups**:

- ***EML Best Practices***: A joint EDI/LTER working group formed to update the *Best Practices for Ecological Metadata Language (EML)*. This WG addressed elements now available in EML2.2, reorganized content for improved usability, and includes introductory material on general data package design. Sections with substantial changes were discussed in a work session at the 2023 IMC Annual Meeting with input from all IMs. The current draft was reviewed at the April IMC Virtual Water Cooler, where each IM reviewed 2 chapters each to provide fresh eyes on the content, feedback to the WG, and approval of revised content.

- ***HyMet (formerly ClimDB/HydroDB EDI/LTER)***: This working group focuses on improving the ways meteorology and hydrology datasets, which are collected by almost all LTER sites, are managed, published, and used. A primary goal is to create the next-generation version of the ClimDB/HydroDB system, which has been decommissioned. In the past two years the group has developed an R package to standardize LTER datasets to the Observation Data Model (ODM version 1.1, developed by CUAHSI). The group also began a pilot project publishing these datasets to EDI (9 datasets so far) and the Dendra platform (https://dendra.science). The working group plans to expand this pilot project to new LTER sites and data formats, and will apply for external support in the next year.

- ***Unit Dictionary*** (EDI/LTER): The Unit Dictionary working group was formed with the goal of creating an updated replacement for the LTER unit registry, which had been deprecated and was in need of a major content update. Initial WG efforts examined extant models for units, and determined that the QUDT ontology (https://qudt.org) could be leveraged due to its existing alignment with environmental data, and the ability to submit units not currently in QUDT. In this reporting period work has focused on identifying commonly used LTER/EDI units not currently in QUDT and preparing a submission to request the addition of these units. The WG has developed tools to aid in the implementation of QUDT annotation in EML metadata. These include 1) an R tool that reads standard and custom units from metadata and generates updated files with QUDT annotation and/or xml code segments for insertion into EML and 2) a unit annotation web service that can be embedded in existing workflows (e.g., R or Python programs) to translate raw units to QUDT annotations. A number of sites have already adopted this code. Recommendations for inclusion of QUDT in EDI's EML editor (ezeml.edirepository.org) are also in the planning stages. A manuscript authored by WG members is currently in development.

- **Keywording:** (EDI/LTER): This group is working to improve LTER dataset discovery, specifically the precision and completeness of data set recall, and EDI search success by using standardized keywording and categorization. In other words, the group is applying metadata annotations with a constrained hierarchy of standardized terms to all datasets in EDI, and redesigning the LTER controlled vocabulary for EDI searches. The new metadata annotations and search vocabularies are also mapped to other scientific community ontologies (such as ENVO, NCBI) for the opportunity to develop a semantic search interface. The basic keyword improvements are intended to facilitate what is called a faceted search interface on EDI. This search interface will allow users to narrow a simple search further with more additional terms (e.g., an author search may be narrowed by a specific keyword). The anticipated outcomes are improved search results at EDI, and greater alignment between the terms that data providers and data seekers use to describe and discover datasets. An example faceted search tool is here: https://data-dev.microbiomedata.org/.

- **DataBits:** Marina Frants and Dan Bahauddin continue as the editorial team for DataBits. The editors solicit articles from members of the LTER community via announcements at the monthly virtual watercolor meetings, discussing article ideas with IMexec, and reach out to potential authors on an individual basis. An article discussing the scientific importance of biological specimen collections, with a focus on collections maintained by CCE-LTER and PAL-LTER, is currently in the editing stage and will be published shortly.

9. **Planned activities for the coming year**:
    a. Support of the LNO graduate synthesis course.
    b. Continued engagement on specimen/samples archiving best practices development together with site PIs.
    c. Publication and dissemination of the updated EML best practices, for use beyond the LTER IM community.
    d. Integrate the QUDT unit into the EML generation as an annotation node, enhancing the interpretation of the unit across LTER sites.

10. **Any recent or upcoming changes in leadership, purpose, or process**: At the 2023 Annual Meeting, Li Kui was elected to the IMexec, with a term ending in 2026. Pre-pandemic, the IMC Annual Meetings rotated on a 3 year cycle - 1) at the All Scientists Meeting (ASM), 2) at a selected LTER site, and 3) in conjunction with a national meeting aligned with information management (typically Earth Systems Information Partners [ESIP] or Ecological Society of America (ESA) Meeting). All LTER IMs are encouraged to attend the annual IMC meeting. During this reporting period, the annual meeting was held before the ESIP meeting in Burlington Vermont (July 2023). Beginning in 2024, to conserve funds and reduce carbon footprint, two meetings in the cycle will be virtual, with the third at the All Scientists Meeting. Our next annual IMC meeting is planned for 9-10 September 2024.

11. **Do you have specific questions, problems, or proposals for the LTER Science Council for actions that could improve the quality or quantity of research, education, engagement, or inclusion in the LTER Network? (Please include adequate background for a non-specialist.)**

    - Sample/specimen archiving. The LTER IM community needs guidance, support and resources in order to develop tools for accessing metadata about physical samples collected as part of the LTER program. Currently LTER samples are archived in a variety of locations that lack a centralized portal for search and discovery. There are likely efficiencies to be gained in collaborating on tools for discovery of LTER-affiliated samples. We believe the shift from site-specific systems for storage and search of LTER data to centralized data repositories (EDI, ADC, etc.) has greatly improved the useability, standardization, and

long-term preservation of LTER data. Is a similar shared system envisioned for LTER sample metadata? If so, who would design and build and with what resources?

- Amplify the IM community's messages on data citation, data management and timely data archiving. As information managers, we feel all these practices have improved tremendously over the decade. Students and researchers look to LTER PIs and leadership for modeling best practices.
- Support IM participation in Working Groups. Information managers bring extensive expertise in data integration, which is crucial for the success of synthesis efforts. More involvement in LTER synthesis working groups will help promote robust data management practices, enhance data quality, and facilitate seamless data integration across sites. We encourage the inclusion of an Information Manager (IM) in the leadership team of synthesis working groups.
- Help secure financial support for creating value-added datasets. There is substantial interest in producing additional datasets beyond the primary data that was traditionally archived by individual sites. These efforts include cross-site activities such as converting individual site's climate records to HyMet/CUAHSI standards, species composition data to ecocomDP formats, or creating more highly processed datasets, such as gap-filled data. The LTER IM community is eager to push these efforts forward, but anticipates that we will need to apply for funding for tasks that go beyond existing workloads.

Your report will be distributed to members of the Science Council and Executive Board before the May Meeting. The Committee's Executive Board representative will have about 5 minutes to present at the meeting and an additional 5 minutes for questions/discussion. We find that both the committees and Principal Investigators get more out of this face-to-face opportunity if the presentation focuses less on reporting activities and more on questions or issues that require discussion and input or action from the Science Council. Questions, issues or proposals that generate significant discussion at the Science Council Meeting may become the focus of deeper discussion at Executive Board or Lead PI Meetings throughout the coming year.