

Challenges, opportunities, and priorities: a vision for the future of the LTER Information Management Committee

background

LTER information management is in the midst of several seismic shifts. Internally, the LTER Information Management Committee (IMC) is adapting to a new organizational landscape that saw the end of the LNO, and the creation of new partners, the NCO and the EDI. It has now been three and two years, respectively, since these organizations were formed, and there seems to be a general feeling among information managers of settling into a comfortable relationship with these new partners. However, the relationships are still being forged, and each of these partners is still developing their own identities as they approach their first proposal renewals. Another shift, one that is predominately external, is the rapid evolution of the data sciences broadly. Whereas the IMC had been pioneering in ecoinformatics, contributing to major efforts, such as the development of EML and PASTA, among many, many others, we see now the ecoinformatics space advancing with and around us, and the niche of the IMC in this space and where it can best contribute are unclear. We see these changes also in the skills of new information managers, and in the skills of the scientists that the IMC supports. For example, while tools such as Excel will remain in heavy use for some time, LTER scientists will reach increasingly to platforms such as R and Python, and working in collaborative environments such as GitHub and GitLab. Many information managers are already witnessing the shift in skills and preferences among the graduate students at their sites. Meanwhile, the IMC contends with the usual types of changes that are always occurring in the background (e.g., new Network sites; grants ending; turnover among IMs; and shifting, sometimes opaque expectations of the NSF).

While the shifting LTER and ecoinformatics landscapes present numerous challenges to the IMC, they also present tremendous opportunities. Diligent work over many years among LTER information managers and others has contributed to the development of PASTA and the EDI. All LTER data are now housed in a central location with remarkable (and evolving) search, citation, and discovery capabilities. Consider that it was only a very short while ago when an investigator seeking data from three different LTER sites would have been required to visit three different LTER websites, each possibly employing a unique data catalog style and formatting. In addition to supporting this critical, centralized resource, the EDI is positioned to coordinate the implementation of new tools and approaches to advance ecoinformatics. Input from proficient information managers, and, increasingly, scientists will contribute to the formation of new ideas that they can then coordinate with the EDI to hone and implement.

This is an important time for the IMC, and crucial is to identify a community vision to guide how this group should navigate these new and evolving landscapes. To this end, a visioning exercise was conducted during the 2018 LTER All Scientists Meeting to identify communal goals and a potential road map as to how the IMC should address immediate and future challenges, capitalize on opportunities, and, generally, operate in a way to maximize the IMC's contribution

to LTER science and ecoinformatics. The visioning exercise centered on information managers and a select few friends from within and outside the Network working in small groups to consider the future of the IMC and a series of targeted questions. Notes from the small groups were summarized by the IMC to the salient points outlined in this document.

collaboration and integration

A persistent and resounding theme to come out of the discussions was recognition for the importance of collaborating with other groups. This theme was expressed in myriad ways but the general sentiment is that the IMC must not isolate itself in a LTER-exclusive space. There is too much individualization, even among individual LTER sites, that contributes to inefficiency and constrains interoperability. As one commenter suggested, "LTER systems should be interoperable with, e.g., USGS standards", and raised the question of whether development meetings should include other groups such as the USGS or NOAA.

One key to fostering collaboration will be for IMC members to think of themselves as part of a network rather than adopt a single-site mentality. Approaching information management challenges from the perspective of, "how can we best solve this problem as a group," will lead to greater harmony across the network and possibly beyond the network, which in turn promotes synthesis. The IMC can leverage EDI in this regard by utilizing cyberinfrastructure such as the EDI GitHub account for organizing technological initiatives, hackathons for rapid development of shared solutions, and the EDI website or other media for promoting these solutions.

working groups

Participants keyed in on the importance of working groups to the development and implementation of ideas throughout the history of the IMC, and that working groups will continue to serve this pivotal role. Given the outsized importance of working groups, efforts should be made to elevate their visibility beyond the LTER. This might be accomplished by strengthening partnerships with, for example, ESIP. The organic development of working groups from within the IMC will continue as needs and opportunities arise, but the IMC should work to improve communication with LTER

researchers to help identify those needs. The IMC should consider also a mechanism for moving working groups from ad hoc to officially recognized and structured entities with a scope of work and prescribed renewal cycle. This is another vision for working groups that could be addressed with strengthened partnerships to ESIP or another organization. Information management working groups are largely unfunded, and the IMC and, particularly, the IM Executive Committee should strive to identify a funding mechanism for working groups - as one participant noted, "\$10,000 per year would go a long way".

advancing informatics

resource development

Perspectives on how the IMC can and should maintain its strong contribution to ecoinformatics is another area that highlighted the importance of collaboration. Information managers and the IMC collectively have profound knowledge and expertise of ecoinformatics. Working groups will capitalize on this knowledge to generate ideas and prototypes that would then be implemented by, for example, EDI with the idea-generating working group then contributing best practices. In this model, information managers contribute expertise to the development of new tools, and use their experience to check for fitness-of-purpose, but work with EDI to implement the ideas. This approach leverages IMC expertise, and the infrastructure and reach provided by the EDI. To maximize chances for success, the IMC should work with the EDI or other collaborators early (and often). There were numerous references in the conversations to the knowledge and expertise of the IMC, and that the most effective way to harness that strength is through the development of and contribution to best practices. However, in keeping with the concurrent theme of collaboration and integration, the IMC must consider the broader information management community when developing best practices. For example, best practices should not be limited to the LTER or even EDI but should consider DataONE and beyond. Pushing this idea further, LTER information managers and the IMC should become ambassadors for the Ecological Metadata Language, serving as liaison to other groups interested to adopt the specification.

semantic annotation and data harmonization

A subset of small groups was tasked with considering ways in which the IMC can contribute to the development of semantic annotation and data harmonization, which are likely next steps in the advancement of data discoverability and utility, respectively. This is another area of the conversations that focused on the expertise of the information manager and IMC collectively as improving the quality of metadata is a fundamental and first step toward advancing both objectives. One way to do that is to demonstrate the importance of the information managers, the IMC, and informatics to lead scientists in the context of encouraging investigators to work with information managers to provide the level of detail required to describe data well. This engagement has the added benefit of promoting information managers as a resource rather than merely enforcing a regulatory burden of providing data.

Very much related and in the context of solutions like ClimDB, the importance of controlled and common attributes was stressed. For the future, a common data model, which takes into account local priorities around a core set of attributes, is desirable. This would allow semantic technologies to be used to capture and stabilize the knowledge that links to the data across sites and systems. The EDI's role can be focused on training and facilitating transfer of site-based knowledge to ontologies and data model updates into a common solution. This solution can be scalable and applied to other solutions outside LTER. The process of creating

harmonized knowledge representations and data models aligned to these can be linked to synthesis group activities.

streamlining metadata generation and lifting new data providers

There is a perception that EML would be easier to write and easier to use; there is a perception that it is a black box (esp. coupled with DEIMS), which makes it difficult to work with it in novel or unique ways.

We can become more useful by streamlining these processes of generating EML so that we can reduce the workload and put our efforts on more valuable uses of our time. Generating the metadata needs to become a more trivial and mundane part of the process.

b. The EDI should maintain a framework or system to help sites, particularly new sites, build information systems following best practices. For example, allowing creating EML with R or providing a DB schema to new sites or sites that wish to change their systems. Recommendations and standard operating procedures (SOPs) would accelerate how sites plug in to existing systems in a harmonized way. This framework would also allow extensibility, allowing different technologies to be used, but with outputs validated to be interoperable. EDI can facilitate workshops to help sites exchange expertise and formulate best practices to make core tooling more robust in the long term. EDI can also use this mechanism to expose new solutions (recall the role of R when it was new).

critical infrastructure

Substantial effort has been invested into the development of tools that are of tremendous utility to many LTER sites. Tools such as the GCE Toolbox and DEIMS are in use by many sites to address some or many of the site's information management needs. In some cases, one or more such tools are mission critical to the site's information management system. To ensure the long-term viability of these resources, the IMC and IM Executive Committee should work with the EDI to develop coordination activities of existing tools or working groups that keep core solutions alive.

other sections, a conclusion?

...this is a nice statement: IMC members should always remember the core responsibilities of making data discoverable, accessible, and usable.