

## Summary of LTER 2010 Cyberinfrastructure Needs and Assessment Survey

- Sites indicate that 10-20% of total site funding is devoted to information management (77%) with 15% of sites suggesting 5% of funding or less, with one site up to 50%. This compares similarly (roughly converted) with previous surveys with about 1-2 FTE positions for information management. Considering total site funding this suggests that many other things related to information management are not included in this estimate.
- Network-level information management (IM) activities comprise 10-20% of site effort (70%). This suggests that 1-2% of site funding is devoted to Network-level information management at sites. Considering the amount of cross-site effort related just to ClimDB, Hydro-DB, Site-DB, the All-site Bibliography, Personnel Directory as well as Working group efforts, IM meeting activities, All Scientists meeting and follow-on efforts as well as other IM activities and how LTER science is now performed, it suggests that this is an underestimate and IMs are rather conservative in their estimates.
- The “average” IM at a site generally is a biologist or ecologist (84% medium to high training in ecology/biology) rather than having background in computer science or database management (90-70% medium to low training respectively). Many have specific specialties in statistics, mathematics, GIS, modeling and others. Most have many years (18 and more) of job-related information management. Only 15% of IMs are a signatory PI on the LTER grant, a number unchanged since 2007.
- General Information management tasks, project management, web site content, and software development, metadata efforts comprise most of an IMs workload, rather than system administration, user-support, network infrastructure or other similar tasks.
- In 2005, 72% of the sites said that EML (at any EML level) existed for 70% or more of their total data holdings; this went up to 96% in 2007 stayed the same in 2010. In 2005, 36% of sites said that **complete** metadata existed for more than 50% of their data. In 2007 this was up to 89.4% and then reported at 76% in 2010. Although this suggests that less complete data exist, it is really saying that the LTER IMs have become EML experts, realizing what it takes to completely document a dataset, and also that more datasets are being put on line.
- Data were not harvested in the LTER Metacat in 2005, but in 2007, 85% of LTER sites had more than 50% of their total site data harvested to the Metacat. This is a similar number in 2010 but the amount harvested is closer to 90% of total site data. Direct links from the metadata to the data itself exist for 55% of the total metadata records in both years for 90% of the sites. For data that are not harvested, 44% of sites said up to 20%, but generally 2.5% to 10% of their data was purposely not harvested due to proprietary, confidentiality or other issues.

## Summary of LTER 2010 Cyberinfrastructure Needs and Assessment Survey

- Use of LTER site data by various groups was surprisingly similar in 2007 and 2010 with most use by site investigators and researchers, some but about equal use by government agencies (NSF, NASA etc) and the public sector, with less use by policy makers, congress and government and little use by lawyers and litigators.
- In 2005, 42% of LTER sites had some type of wireless data capability at the research site. This rose to 60% in 2007 and 85% in 2010.
- In 2005 sites generally used 10gb to about 1tb for site data storage with a similar figure in 2007, but in 2010 16% of LTER sites used 1tb or more, peaking at 50tb, including online backup.
- In 2005, 40% of sites had internet links slower than 100mb/s at the research site, dropping to 35% in 2007 and 20% in 2010. Connections greater than 100mb/s at the research site existed for 12% in 2005 with 32% in 2010. Links of greater than 10gb/s now exist for 22% of sites.
- Common database systems used by LTER sites include MySQL, SQL server and Oracle (36%, 28% and 12% respectively in 2005, 29%, 21% and 12% in 2007 and 54%, 29% and 19% in 2010) with more sites now implementing database systems rather than simple file systems and html web pages to manage site data.
- VIOP services such as Skype were commonly used by 36% of sites at their home institutions in 2005, increasing to 56% in 2007 and 92% in 2010. Video conferencing using Polycom, Tandberg or other h.323 systems were used by 36% of LTER sites in 2005, 64% in 2007 and 84% in 2010 . Use at sites varies and it is not clear why this number isn't reported at 100% now considering VTC use by the LTER Executive board, the IM group itself and others.
- An interesting note is that in 2005, a need for better internet connectivity at the site was not as much of an issue in 2007 but it was ranked as important for some sites in 2010. In a similar change, need for better software to manage data and metadata was not rated important in 2007, but was rated very important in 2010, particularly for metadata.
- Although a number of suggestions were given to improve both site-level and Network-level research, a common thread, of needs for common protocols, as well as more support for IM, including personnel were the most common responses. In a similar question, sites said that needs for more IM personnel were needed ((55% most important in 2007, 77% in 2010) followed by training for IM personnel.