

Final report

Ecological information management for LTER graduate students and researchers, August 14 – 16, 2012.

27 people applied for the workshop, representing 13 LTER sites with 2 applications from outside LTER. 5 people were unable to participate and everybody else was accommodated.

22 participants from 12 LTER sites, ranging from PhD students, post-docs to lab coordinators and faculty.

Trainers: James Brunt, Corinna Gries, Eda Melendez-Colom, John Porter, Theresa Valentine, Kristin Vanderbilt.

Website: <http://im.lternet.edu/node/1027>

Agenda:

- **Tuesday August 14, 2012, 8:30 am – 5:30 pm LTER Network Office**
 - Introductions
 - LTER information management and the data life cycle
 - Introduction to a data management plan
 - Data Collection (field sheets, QA/QC, mobile apps)
 - Digital Data Organization (Databases, Excel Spreadsheets, QA/QC, Authority information)
 - Metadata (content, format, creation, Morpho, DEIMS)
- **Wednesday August 15, 2012, 8:30 am – 5:30 pm LTER Network Office**
 - GIS data management
 - Data Security
 - Submitting data packages to the LTER NIS, other data repositories
 - Accessing and using data in the LTER NIS with the Kepler workflow system
 - Accessing and using data in the LTER NIS with R
 - Considerations and technologies for large volume streaming data management
- **Thursday August 16, 2012, 8:30 am – 10:00 am LTER Network Office**
 - Revisit the data management plan, discussion, questions, feedback

Training material developed:

The agenda above shows that the course was structured to cover all aspects of the data life cycle. The trainees collected data and this dataset was taken through the entire process. Each subject on the agenda is linked to a standard web page with presentations, exercise materials, and links outside material were developed and can be found here: http://im.lternet.edu/resources/training_material. The entire training course was video taped. Access to the videos may be provided at a later date.

Post training survey (11 responses)

Summary

Overall participants were very appreciative of the training, emphasizing that the overview of so many technologies involved in the data life cycle was very useful. Most agreed that further and advanced training would be very helpful, particularly in the area of security, metadata, database technologies, and the Kepler workflow system. Training for these technologies is not offered at most universities or not in a very useful way for ecologists (database technologies) while general R and GIS training is available elsewhere. Therefore, covering the LTER specific aspects for R and GIS is sufficient and was also highly appreciated. This particular group of participants felt that hands-on data collection and training in the use of Excel were not necessary. Several people thought that a general IM training should be conducted at each LTER site at a regular basis, some suggested making it mandatory.

Suggestions: offer a introductory beginners course and more advanced courses on specific topics. Have participants use their own data, give homework and pre-training work assignments.

On a scale of 1 – 5 (1 = excellent, 2 = good, 3 = neutral, 4 = poor, 5 = very poor)

| | |
|---|------------|
| How would you rate your overall satisfaction with this workshop? | 1.9 |
| How would you rate the overall organization and effectiveness of this training? | 2.0 |
| How would you rate the overall effectiveness of the workshop trainers? | 1.6 |
| How would you rate the overall effectiveness of the workshop trainers? | 1.3 |

Was there an appropriate mix of presentations and hands-on exercises?

Just right - 3

Not enough exercises -7

too many exercises - 1

Was the overall length of the training appropriate?

just right - 3

too short - 4

too long - 1

What did you learn that is immediately applicable to your work? Did the training provide you with knowledge that you can apply to your work when you get home? If so, please tell us about it.

Access will be very helpful to my research. I also enjoyed learning about the other programs that I can learn more in depth if needed in the future.

introduction to access was applicable

I'm a firm believer of documenting all of my data and metadata so this was a great workshop. I've already registered data and metadata for one of my chapters but I think the most useful part for me was learning a little about Access (I've just never had training on this program somehow), ArcMap, and learning about the fact that Ecological Archives also take just data papers. I knew most of the things taught during the workshop but the little pieces here and there were actually very helpful.

metadata development and workflows. I will be able to start implementing these tools, but I think it will take a while.

It definitely covers many aspects of information management and some of them a really handy such as using access and R programming. I will continue learn access and R when I get back home and try to use them in my on going and future studies.

The information on security from James Brunt, and information I got from one-on-one interactions with trainers are things that I can apply to my work when I get home. In particular, I will look into KeePass for managing my own passwords and may rethink some of my current back-up practices.

I learned that I need to make a lot more decisions about how I'm going to manage my data that's coming in.

Basic intro to Access. Desire to find workflow programs. Desire to use programs that document your steps to use as metadata and insure repeatability.

The overall integration and utility of software programs in the data cycle. The introduction to specific software, especially the freeware, plus hands-on opportunities to walk through Morpho and Kepler, will be immediately useful.

I got a really good review and reminder on data management best practices that I will immediately apply to my data. I also got a better idea of how to use Access, and will definitely incorporate Access into managing some of my larger datasets and future datasets. I also got a great overview of other data management software and terminology, and although I most likely won't be using them soon, it really helped to have a better idea of the language.

I received a good introduction into various data management programs, and I am now excited to explore some of these programs in much greater detail. However, I don't feel confident in immediately applying much from this workshop to my work.

What aspects of this training did you find most useful (e.g., hands-on exercises, trainer expertise, materials covered, etc.)?

I enjoyed the hands on exercises and lectures though lectures could be a bit shorter with more detailed instructions/exercises to do independently. I enjoyed learning about the programs I never knew I existed.

lectures

I like that the presentations are online after the workshop so that we can have access to them. I really liked the hands-on exercises but I thought that there should be more time allotted because it was too much information in such a short amount of time that I feel introduced to some things (like ArcMap and Oxygen) but I know I have no idea what to do with those programs on my own. It was good to be introduced to them though.

hands on was most helpful. The presentations were helpful but could be shortened

I am interested in those software and technology that I didn't learn and use before. The exercises, trainer expertise and materials are very good, but some of the common used software such as using access to manipulate data and R programming may need more time and courses for us to figure out what we should learn and what we can use them for.

The trainers are clearly very knowledgeable and one-on-one interaction with them is quite useful. I think the presentations were very good overviews and gave us a good introduction to a variety of different softwares and best practices. The workshop gives us a good starting point and good resources from which we can continue further study on our own.

Exposure to so many different things was great--- being still early on in my graduate work, I now know directions I want go & programs I would like to incorporate. QA/QC and data safety best practices were GREAT info.

Trainer expertise. Wide range of programs and issues addressed.

Expertise of trainers, guided hands-on exercise.

I found the materials covered to be most useful.

I liked hands-on exercises, as well as the material that is now available online.

Were there other aspects or materials that might have been included or where more time should be allocated?

I thought the things covered in the 2nd day of training (Kepler, GIS programs, Oxygen, and R) were great but there wasn't enough time spent on each individual program. I think for big, complex programs like these, it would make much more sense to spend an entire day on each program so that we could leave the workshop with some working knowledge of what to do, how to do it, and how to troubleshoot a little ourselves. Kepler is so complicated with its numerous actors and the manual is not extremely helpful (in my opinion as many of the examples are actually out of date right now) so that the little bit we covered is just enough to get people in trouble but not really know how to use it. Similarly, I feel this is true for the other programs as well.

more time could be allocated to higher level tools such as databases and workflows

I would prefer to spend more time on R programming and access or other database management.

I would have been interested in learning more about MS Access and SQL Server. I understand the reasons behind focusing on non-proprietary softwares and it's good to know about them- but I am locked into using them for my project, and would like to learn more about how to use them effectively.

bigger picture, how things are interconnected

It would have been good to have had Intro homework to do before the workshop to get everybody closer to being on the same page and familiar with key basics. More emphasis on the big picture - why we use particular programs, why we perform particular steps in a particular program - what are the end goals?

I think that we could have dispensed with the data collection and use of Excel.

I think it would have been good to keep the example exercises simple, so that everyone could follow along. I would have liked to get an overview first and then perhaps spend a half day at the end doing a more involved hands on exercise with one of the software options to really get a better handle on something that I was unfamiliar with.

I would have liked to have much more time allocated to hands-on exercises with better developed protocols. While this may have been frustrating, I would have learned more. Also, I know that people bringing in their own data would be a huge headache, yet if this were the case, it would provide a much more tangible product.

If this training is conducted again in the future, do you suggest any changes in format, presented materials?

I would suggest a couple things:

- participants should be surveyed about their level of knowledge about each subject before the event to gauge everyone's expertise and needs more specifically
- presenters could talk/work a bit slower overall (many of us were lost during several presentations)
- hands-on exercises should be a bit less complex overall
- participants could be given detailed step by step instructions to work along with presenters, this would greatly aid keeping up. Otherwise if you miss one step you are lost.

Do not spend a third day on rap-up/feedback. Some exercises (e.g. GIS) felt like they were too basic for anyone who has used the software, and too scattered/broad for anyone without experience.

I thought the first day of the workshop was perfect--I think it was a great overview of why data management is important and the programs used to document data and metadata aren't extremely complicated and the amount of time covering that information was just right. But I would suggest for the 2nd day (or maybe even more days than just 2 training days?) that maybe the focus can be just on one of the programs that was mentioned so that people can have a more in-depth understanding of one program than of 0 programs.

please remove the sections on GIS and Excel. If we worked in GIS, the focus should be on management of the data and metadata, not how to run a simple task like plotting points.

I may consider to give us introduction on the class, and we use some kind of handouts for specific tutorials and I can do exercises at home..

My main suggestion is to change how the hands on exercises are done. I like having hands-on exercises- but I found that once I got behind in an exercise, it was hard to catch up and get the full benefit of what was being covered. This was especially true for Kepler. One suggestion would be to have several presentations together and then have several hours to go through one hands-on exercise that integrates the different steps- going from raw data to processed data/a database/analysis- in one go, and have the opportunity to ask questions and work more at your own pace. I think the HOB0 field exercise took a bit too much time given the quality of the data returned and the amount of time we had in the workshop- so it might be better to use a standard/clean dataset that we can work with and see the full process with it. Also- I liked the lecture room- but I felt like it made it hard to interact with people sitting at other computer islands- so it might be nice to have a room where lectures are given that allows for more mingling; and a room for exercises that have the computer island set up.

more hands on, using our own datasets, more piece by piece transitions

I agree with the suggestions to attempt to work on our own datasets. I also agree with the idea of breaking it into Intro vs Advanced sessions and/or dividing up into more focused groups based on interests/skill level.

Deeper dive into publishing data, EML, XML, Kepler.

I would suggest that we not collect our own hobo data, but instead all use the same pre-prepared dataset in the exercises.

In the future I'd remove most of the introduction lecture. No time needs to be spent on selling the importance data management, why we came to the workshop. Also, generally keep more focused on manipulation of data without excursions into data gathering, GIS work...etc.

Please add any other comments you wish to convey

- overall I am extremely happy I attended and will look to participate in more workshops of this type, thank you for allowing me to participate!

It was often unclear what the end goal of many of the exercises was. Learning how to do some of the intermediary busy work can be useful, but many exercises (EML), used a great deal of time, and did not fully explain what/why we were trying to accomplish.

Hobo data section was extremely basic, and took way too much time for what most people got out of it.

A stipulation that students bring some of their own data to work on would drastically improve student interest in what was going on, and given the amount of time that had to be spent assisting students with problems one-on-one, having people working on different projects would not greatly increase the difficulty for instructors.

I thought that overall it was informative but I think the 2nd day could be changed in format and content a little bit. I also talked to other participants and quite a few of us feel that it is extremely unfortunate that the LTER does not require a mini-workshop for every LTER site to make sure the researchers and graduate students know this information. Good habits are formed and if we don't ever train our graduate students, post-docs, and other researchers, how can we expect good data management from future scientists? I think if it was mandatory for scientists in the LTER to go through a 1-day workshop (the part on data management, metadata, Morpho, etc), then the LTER would be increasing the quality of data management for many scientists as opposed to just the few who are interested.

Please invite me to the future workshops on sensor data management and advanced data management!
-John Crawford

Maybe we could skip the excel and data collecting part.

I thought the trainers did a great job. Based on the diversity of feedback that was given when we went around the room - I'd say that it might be good to have two workshops- one for introductory users that gives the broad overview of different softwares and concepts; and one that is more detailed and for more advanced users that are further along in their projects and want to work with their own datasets.

the instructors themselves were fabulous, personable and willing to help. although it may have been more geared towards advanced users, it was excellent for a 'younger' user like me to get exposure to so many possibilities

Information on these resources is not getting out to all individual LTER sites (I tend to ignore most network-wide emails).

Thank you very much!

The presenters went through examples way too fast. I was having a hard time keeping up with what the presenter was saying and trying to also repeat the steps on my computer when the software was unfamiliar. The presenter's should go through steps slowly and make sure everyone's on the same page. I understand that there would be people who would find that too slow if they were already familiar with the software, but that's why there should be an additional hands-on lab component later to give them more of a challenge.

Overall this was great however! I was really impressed by the expertise of the trainers and it was great to get a sense of what the information managers have to deal with at the LTERs!