

LTSER Data Management Workshop Report

Andrea Previtali, Bryan Milstead, David Blankman, Inigo San Gil, and Kristin Vanderbilt

December 22, 2014

Introduction:

As part of the All Scientist of the Americas meeting (<http://www.ltsr-chile.cl/ilter-meeting-2014/>) a data management workshop was held on December 3rd, 2014 in Valdivia Chile. The main purpose of the workshop was to initiate a conversation among participants on how to better manage long-term ecological data. The workshop began with a presentation of the Drupal Ecological Information Management System (DEIMS), a content management systems designed to facilitate the stewardship and documentation of ecological datasets. This was followed by a question and answer period where participants discussed the challenges they face. We ended the workshop by collecting a set of recommendations from the participants.

Recommendations:

At the end of the workshop participants were asked for suggestions on how to improve information management in the LTSER sites in the Americas. The recommendations are:

- **Start Now:** it is clear that we need to initiate and/or improve information practices and it is imperative that we start without delay.
- **Develop Simple Guides to Data Management Best Practices:** many of the participants are just starting with data management and it would be helpful to have some simple guidelines for best practices.
- **Training:** in addition to developing guidelines there is a need for hands on training from basic data entry and management principles to more advanced information management and analysis procedures.
- **Data Sharing:** a clear concern expressed by many participants was the need to convince investigators of the need to share data. Data sharing policies can be developed to define who can access the data, when the data will be made available, and how to properly credit the researchers for their work. The ILTER data policy can be found here: <http://www.ilternet.edu/data-info-management/ILTER%20Network%20Data%20Access%20Policy-8.23.2008-1.pdf>
- **Funding to Support Information Management:** information management is an expensive and time consuming task. There was general agreement that additional funding is necessary to support data managers, training, and infrastructure.
- **Representative from the Americas on the ILTER Data Management Committee:** the Americas are not currently represented on the ILTER data management committee. It would be valuable to select a representative.

Links

Below we provide links to information on information management topics of interest to the participants. This is not a complete list. Please feel free to suggest additional materials and topics to include.

- **Formatting Data in Spreadsheets:** Much of the data entry done by the ILTSER sites is being done with spreadsheets such as Excel. Spreadsheets can be useful but also can lead to large problems if the data are not formatted correctly. The following paper provides information on how to use spreadsheets effectively:
 - <http://vita.had.co.nz/papers/tidy-data.pdf>

- **Training & Best Practices:** There are many internet sites devoted to this. As a start look at these websites:
 - The ILTER Data and Information Management Training Resources: <http://www.ilternet.edu/data-info-management/data-im-training-resources>
 - DataOne’s training resources: <https://www.dataone.org/education>
 - DataOne’s best practices suggestions: <https://www.dataone.org/best-practices>
 - Hook et al. 2010. Best practices for preparing environmental data sets to share and archive: <http://daac.ornl.gov/PI/BestPractices-2010.pdf>
 - Borer et al. 2009. Some Simple Guidelines for Effective Data Management. Bulletin of the Ecological Society of America 90:205–214: <http://dx.doi.org/10.1890/0012-9623-90.2.205>
 - Getting started in data science: http://treycasey.com/getting_started.html
 - How to become a data scientist: <http://blog.datacamp.com/how-to-become-a-data-scientist/>
- **Data Portals:** Where and how to archive and share data is an important decision. There are hundreds of data portals now available. Below is a mixed list of data portals and sites that catalog data portals.
 - <https://portal.ilternet.edu>
 - <https://www.dataone.org>
 - <http://databib.org/>
 - <http://www.re3data.org/>
 - <https://www.datacite.org/>
 - http://oad.simmons.edu/oadwiki/Data_repositories
 - <http://www.nature.com/sdata/data-policies/repositories>
 - <http://figshare.com/>
- **R** is an open source statistical computing language that can be used for data management, analysis, and GIS:
 - <http://cran.r-project.org/>
- **Rstudio** provides an interface to R and adds new capabilities such as an interface to gitHub. They also provide tutorials:
 - <http://www.rstudio.com/>
- **gitHub** is a code repository and version control system. Many data managers use gitHub to share information, code, and documents. gitHub allows multiple researchers to work on the same files and to share equally across computers:
 - <https://github.com/>
- **DEIMS** (Drupal Ecological Information Management System) was presented in the workshop. This is a content management system that facilitates the communication among researchers, data sharing, and metadata creation. DEIMS provides groups (labs, LTSER sites, ILTER groups) with a professional and modern Information Management system. A relatively large number of information technologists representing many sites banded together to address this common need. The idea behind working together is economy of scales. We all have a common problem, and few resources. By combining resources and strengths, we can have a solution that is good for everybody, but out of reach to do by one single group. DEIMS design was not exclusive, it is all free, open source, based on a leading, free, popular and open source Content Management System (Drupal - <http://drupal.org>). The following links will help you get started:
 - Poster presented at ILTER meeting <https://deims.ilternet.edu/deims-ilter-poster.pdf>
 - Project page at Drupal <https://www.drupal.org/project/deims>
 - Project at GitHub <https://github.com/ilter/deims>
 - DEIMS related blog posts <http://lno.ilternet.edu/search/node/DEIMS>
 - DEIMS presentation at Austin DrupalCon <https://austin2014.drupal.org/session/deims-drupal-profile-ecological-dat>

Document Source

This document was written in Rmarkdown, an open source text editing language, and is stored on gitHub.

The PDF version of this file can be downloaded here: https://github.com/FrayJorge/LTSER/blob/master/LTSER_DataMgmtWorkshopReport2014.pdf?raw=true

If you are a gitHub user you can download the Rmarkdown document and associated files by “forking” the repository: <https://github.com/FrayJorge/LTSER>.

Acknowledgments

Thanks to: Dr. Juan Armesto, Dr. Ricardo Rossi, Dr. Julio Gutiérrez, and Monerrat Lara for organizing the meeting and inviting us to participate.

Participants:

The workshop was organized by:

name	country	email
Andrea Previtali	Argentina	andrea.previtali@gmail.com
Bryan Milstead	USA	Milstead.bryan@epa.gov
David Blankman	Israel	dblankman1@gmail.com
Inigo San Gil	USA	isangil@lternet.edu
Kristin Vanderbilt	USA	vanderbi@sevilleta.unm.edu

The following people attended:

name	country	email
Alejandra I. Domic	Chile	alejandradowic@gmail.com
Álvaro G. Gutiérrez	Chile	bosqueciencia@gmail.com
Aurora Gaxiola	Chile	garambuli@gmail.com
Eva Dettweiler-Robinson	USA	evadr@unm.edu
Francisco Bonet	España	fjbonet@gmail.com
Gerardo Gutiérrez Orellana	Chile	gerajocker@hotmail.com
José D. Carriquiry	Mexico	carriquiry@uabc.edu.mx
José Luis Cortés Bugueño	Chile	jluis cortesbug@gmail.com
Julio Gutiérrez	Chile	jgutierrez@userena.cl
Luis Eduardo Calderon Aguilera	Mexico	leca@cicese.mx
Margarida Santos Reis	Portugal	mmreis@fc.ul.pt
Maria del Pilar Fernandez Murillo	Chile	mfernandez@userena.cl
Mauricio Cotera	Mexico	Mauricio_cotera@yahoo.com
Miguel Luna	Mexico	lluna.miguel@inifap.gob.mx

name	country	email
Ondina Angelica Riquelme Bugueño	Chile	oriquelme@userena.cl
Petr Znachor	Czech Republic	znachy@gmail.com
Priscila Molina Pérez	Chile	priscila.molina@wwf.cl
Rocio Sanhueza C.	Chile	rosancab@gmail.com
Simon Castillo	Chile	scb.139@gmail.com

Additional People to include on the mailing list for the workshop:

name	country	email
Aaron Ellison	USA	aellison@fas.harvard.edu
Juan Armesto	Chile	jarmesto@bio.puc.cl
Juan Ignacio Mauricio Isla Cortes	Chile	jimic12@yahoo.es
Ricardo Rossi	Chile	rozzi@unt.edu