BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e37047
doi: 10.3897/biss.3.37047

OPEN ACCESS

Conference Abstract

Achieving FAIR Data Principles at the Environmental Data Initiative, the US-LTER Data Repository

Corinna Gries[‡], Mark Servilla[§], Margaret O'Brien[|], Kristin Vanderbilt[§], Colin Smith[‡], Duane Costa[§], Susanne Grossman-Clarke[‡]

‡ University of Wisconsin Madison, Madison, United States of America
§ University of New Mexico, Albuquerque, United States of America
| University of California Santa Barbara, Santa Barbara, United States of America

Corresponding author: Corinna Gries (cgries@wisc.edu)

Received: 10 Jun 2019 | Published: 18 Jun 2019

Citation: Gries C, Servilla M, O'Brien M, Vanderbilt K, Smith C, Costa D, Grossman-Clarke S (2019) Achieving FAIR Data Principles at the Environmental Data Initiative, the US-LTER Data Repository. Biodiversity Information Science and Standards 3: e37047. https://doi.org/10.3897/biss.3.37047

Abstract

The Environmental Data Initiative (EDI) is a continuation and expansion of the original United Stated Long-Term Ecological Research Program (US-LTER) data repository which went into production in 2013. Building on decades of data management experience in LTER, EDI is addressing the challenge of publishing a diverse corpus of research data (Servilla et al. 2016). EDI's accomplishments span all aspects of the data curation and publication lifecycle, including repository cyberinfrastructure, outreach and training, and enhancements to data documentation methodologies used by the environmental and ecological research communities. EDI is managing almost 43,000 unique data packages and their revisions from a community of nearly 2,300 individual data authors, most of which are contributed by LTER sites, and are openly accessible and documented with rich science metadata in the Ecological Metadata Language (EML) standard. Here we will present how EDI achieves FAIR data principles (Wilkinson et al. 2016, Stall et al. 2017), and report data use metrics as a measure of success.

The FAIR principles serve as benchmarks for EDI's operation and management: the data we curate are *Findable* because they reside in an open repository, with unique and

persistent digital object identifiers (DOIs) and standard metadata indexed as a searchable resource; they are *Accessible* through industry standard protocols and are, in most cases, under an open-access license (access control is available if required); *Interoperability* is achieved by archiving data in commonly used file formats, and both metadata and data are machine readable and accessible; rich, high quality science metadata, with automated congruence and completeness checking, render data fit for *Reuse* in multiple contexts and environments, along with easily generated data provenance to document their lineage.

The success of this approach is proven by the number and spatial and temporal extent of recent re-analyses and synthesis efforts of these data. Although formal data citations are not yet common practice, a Google Scholar search reveals over 400 journal articles crediting data re-use through an EDI DOI. However, despite improved data availability, researchers still report that the largest time investment in synthesis projects is discovering, cleaning and combining primary datasets until all data are completely understood and converted to a similar format. Starting with long-term biodiversity observation data EDI is addressing this issue by implementing a pre-harmonization of thematically similar data sets. Positioned between the data author's specific data format and larger biodiversity data stores or synthesis projects, this approach allows uniform access without the loss of ancillary information. This pre-harmonization step may be accomplished by data managers because the dataset still contains all original information without any aggregation or science question specific decisions for data omission or cleaning. The data are still distributed into distinct datasets allowing for asynchronous updating of long-term observations. The addition of specific and standardized metadata makes them easily discoverable.

Keywords

Long-Term Ecological Research, LTER, data repository, FAIR Data, environmental data, long-term data, data management

Presenting author

Corinna Gries

Presented at

Biodiversity_Next 2019

Acknowledgements

This work has been supported by the National Science Foundation grants DBI-1629233 and DBI-1565103.

Funding program

US National Science Foundation, Advances in Biological Infrastructure

Grant title

Environmental Data Initiative

Hosting institution

University of Wisconsin Madison, University of New Mexico, University of California Santa Barbara

Author contributions

The authors take on different roles in this project but are contributing to its success equally.

Conflicts of interest

None

References

- Servilla M, Brunt J, Costa D, McGann J, Waide R (2016) The con-tribution and reuse of LTER data in the Provenance Aware Synthesis Tracking Architecture(PASTA) data repository. Ecological Informatics 36: 247-258.
- Stall S, Robinson E, Wyborn L, Yarmey L, Parsons M, Lehnert K, Cutcher-Gershenfeld J, Nosek B, Hanson B (2017) Enabling FAIR Data Across the Earth and Space Sciences. Eos https://doi.org/10.1029/2017eo088425
- Wilkinson M, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, da Silva Santos LB, Bourne P, Bouwman J, Brookes A, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo C, Finkers R, Gonzalez-Beltran A, Gray AG, Groth P, Goble C, Grethe J, Heringa J, 't Hoen PC, Hooft R, Kuhn T, Kok R, Kok J, Lusher S, Martone M, Mons A, Packer A, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S, Schultes

E, Sengstag T, Slater T, Strawn G, Swertz M, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 https://doi.org/10.1038/sdata.2016.18