

# Metadata for Scientific Data: An Analysis Targeting Challenges and Opportunity in Our Global Information Ecology

Jane Greenberg<sup>1</sup>

<sup>1</sup>Professor, School of Information and Library Science (SILS), and Director, SILS/Metadata Research Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States, e-mail: [janeg@email.unc.edu](mailto:janeg@email.unc.edu)

The Extensive and continuous growth of digital data introduces unprecedented metadata challenges and opportunity (Greenberg, 2009). This paper reports on a framework analysis and a case study targeting the global metadata environment for digital data. The analysis examines *principles*, *domains*, and *structures* following the MODAL framework (Greenberg, 2005) for nine metadata schemes and 14 ontologies supporting data description, semantic analysis, and life-cycle management. The case study targets the Dryad data repository (<http://www.datadryad.org/repo/>) for data underlying published research in evolutionary biology and related disciplines.

Challenges discovered center on conceptualizing metadata and designing robust structures accommodating our global information ecology. Digital data initiatives frequently stray from core components, eliminating fundamental metadata features. Granular metadata descriptions provide rich detail, although efforts on this level can impede disciplinary and cross-domain interoperability, data discovery, access, and reuse.

Results identify a number of opportunities for effectively addressing what Lord, et al (2004) have termed the “data-deluge.” Several advances specific to metadata for scientific data include:

- *Biocuration*: A growing effort engaging professional scientists as data curators (Howe, 2008).
- *Automatic metadata generation*: Sophisticated and reliable algorithms and applications for harvesting, deriving, and extracting metadata.
- *A cultural shift among informaticians*: Metadata innovations relating to linked data and the Semantic Web.

The Dryad repository embraces these noted opportunities, and has motivated the formation of the Dublin Core Metadata Initiative—Science and Metadata (DCMI-SAM) community, a forum for individuals and organizations to exchange information and knowledge about metadata describing scientific data (<http://purl.org/dc/science>).

This paper will report further on the underlying research questions, methods, procedures, and results of the above noted analyses. The presentation will also contextualize outcomes, and highlight potential directions promoting global, interoperable metadata for managing, accessing, and using scientific data.