

Advancements in scientific data searching, sharing and retrieval

Ranjeet Devarakonda¹, Giri Palanisamy¹, Bruce E. Wilson¹

¹ Oak Ridge National Laboratory PO Box 2008 MS 6407, Oak Ridge, TN 37831 USA

In the recent years, there has been significant advancement in the areas of scientific data management and retrieval techniques, especially in terms of standards and protocols for archiving data. Oak Ridge National Laboratory Distributed Data Archive Center for biogeochemical dynamics (ORNL DAAC)ⁱ is making efforts in building advanced toolsets for these purposes. Mercuryⁱⁱ is a web-based metadata harvesting, data discovery and access system, built for researchers to search for, share and obtain biogeochemical data. Originally developed for single National Aeronautics and Space Administration (NASA) project, Mercury now used over fourteen different projects across three US federal agencies. Mercury renders various capabilities including metadata management, indexing, searching, data sharing, and also software reusability.

Mercury system harvests the structured data from several data provider servers around the world. The harvested files are indexed against Solrⁱⁱⁱ search API consistently, so that it can render various search capabilities such as simple, fielded, spatial and temporal searches across a span of projects ranging land, atmosphere, and ocean ecology. Mercury also provides implementation techniques for data sharing between data providers using OAI-PMH^{iv}. This chapter will talk about data harvesting of structured metadata, efficient ways of indexing and searching using Solr search API.

References:

-
- [1] Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC), Accessed May 2010, Available at: <http://daac.ornl.gov/>
- [2] Devarakonda, R, Palanisamy G, Wilson B, Green J. 2010. Mercury: reusable metadata management, data discovery and access system. Earth Science Informatics. 3:87-94
- [3] Apache Solr (Solr), Accessed May 2010, Available at: <http://lucene.apache.org/solr/>
- [4] Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), Accessed April 2010, Available at: <http://www.openarchives.org/pmh/>