Data Management Efforts for the Hanford 300 Area and Rifle IFRCs

R. Versteeg (PI), *Sky Research, Inc. Hanford IFRC Team (Principal Investigator: John Zachara) Rifle IFRC Team (Principal Investigator: Philip Long)*

Both the scope of the IFRCs (in terms of duration, number of participants, and diversity of disciplines involved in each IFRC) as well as the amount of data collected is substantially larger than previous geoscience research projects sponsored by DOE. This required an organized effort for data management and curation. Over the past several years, an approach for data management was developed and implemented for the Hanford 300 Area(www.300areaifrc.org) and Rifle IFRC(www.ifrcrifle.org) . As the two IFRCs share a common approach, software components and (where appropriate) data models this has allowed for synergistic development and resource utilization. The approach has two main elements: (1) implementing data management as a web application (resolving issues of distributed applications and data inconsistencies) and (2) using the maximum amount of existing software components (both commercial and open source). This extends both to the standard components of a web application (such as the database and the server), and to the software components for mapping, display, and data presentation. To this end, the web application has been implemented in Zend Framework, an open source, object-oriented web application framework implemented in PHP 5 with a MySQL database. The web application makes use of several Google components (Google Maps and Calendars), as well as several open-source packages for wiki (phpBB), graphing (jpGraph,Visit), and data analysis (R).

The data management effort for the IFRCs has several objectives. The first is to capture all data collected as part of the IFRC in such a way that this data is easily available and accessible, both for the core IFRC group and for future researchers. The second is to provide tools for basic data exploration and visualization. The approach to data collection included (1) iterative development of a comprehensive data inventory; (2) implementation of a structure allowing individual scientists to upload their data in a relatively unstructured format in so-called data packages (a loose free form assembly of data with some basic amount of documentation—typically identical to the scientist specific organization); given that the IFRC project PIs (John Zachara and Phil Long) strongly encourage data package assembly and submittal, compliance with this effort is high; and (3) full integration of selected parts of the data contained in the data package into the database.

As data management is an ongoing effort which occurred over the life of the IFRC, both in the continued integration of novel data, as well as in the expansion of features in the web interface and in the refactoring of the database and underlying implementation datamanagement capabilities have been evolving over the life of the IFRC. Current capabilities provide for sample tracking and management, on-demand graphing and contouring of different datasets, generation of time series animations of field and modeling data, cross analysis of different datasets, and download of user selectable subsets of the data. In addition, by aggregating components such as Google calendar, weather, twitter feeds, and webcams, the website provides a central point of access for field activities. At the SBR meeting, an overview and live demonstration of the system will be given.