Canadian Foundation for Climate and Atmospheric Sciences (CFCAS)
Fondation canadienne pour les sciences du climat et de l'atmosphère (FCSCA)

Final Progress Report

Project Title: DRI Data Management

DRI Data Manager: Phillip Harder

1.0 Project Work

1.1 **Provide a summary description of a) the objectives of the study, b) the scientific findings and c) the project work undertaken.**

DRI data management had two components. One part was focused on acquiring and managing datasets for the benefits of the Investigators and directly contributing to the scientific program of the network. Data access was facilitated by the DRI website (for smaller datasets) and the Data Access Integration (DAI) group which was a partnership between GEC3, EC, DRI and Ouranos (for computationally demanding datasets) The other component was the data legacy which was the effort to organize and archive the datasets used and produced by the investigators. This involved the development of a data and metadata legacy system that would operate after the completion of the project

1.3 **Describe the tangible results or the measurable outputs generated by the project and how these results have been taken up by user groups for policy development or operational improvements.**

The tangible result of the data management activities is the data legacy. This has resulted in a query-able metadata database that connects to datasets generated/used by DRI. This publicly accessible distributed data system is available to anyone interested in finding drought/environmental data for the prairies. Tracking of the data legacy does indicate that users are making use of the data available but by whom and for what is unknown in the current implementation of the system.

1.4 **Explain any significant delays or departures from the research plan, or the rescheduling of activities, and how they were addressed.**

Primary delays in regards to the data management activities have related to the slow progress of data gathering. The initial formal call for data was made in the spring of 2009. As of Nov.25 2010 there are still various datasets that have not been submitted that have been pledged. The delay in the provision of datasets has set back the finalization of the data legacy. At this point these delays have not been an issue as time still existed. A deadline of the New Year will likely be made to ensure all available data will be included. These delays can be

attributed to data legacy archiving being a low priority for co-investigators relative to the amount of work required.

1.6 **Describe the participation of government (federal, provincial or local), university, industry or foreign researchers in the project.**

External Partner involvement has come in the form of coordinating with various groups to find a long term redundant archive for the data legacy. Discussions have taken place with the Water and Environmental Hub (WEHub) which are project of the Cybera network in Alberta. Wehub is interested in archiving water and environmental related datasets in their infrastructure. Discussions are preliminary and at the date of this writing no formal arrangement or data transfer has taken place.

The National Center for Atmospheric Research has expressed interest in setting up a long term web accessible archive for the DRI Data Legacy. Formal transfer has not taken place and is contingent upon the data legacy being finalized.

Many data providing groups have contributed to the data management and data legacy by providing and authorising the access and sharing of data. These constitute various agricultural, water and environmental agencies of the three Prairie Provinces, environment Canada among others.

2.0 Impact

2.1 Describe in broad terms how your work has contributed to the overall objectives of DRI and to our scientific understanding of drought.

DRI Data Management has contributed extensively to all of the various objectives of DRI. The science is based on data and the organization of the data within the network for the research as well as in the data legacy for public access has facilitated the scientific progress and sets a foundation for future research to carry forward.

2.2 Describe the significance / impact of the results in terms of some or all of the following areas:

Specific impacts/results are unknown but as it is publicly available governments, academia and others will be able to access datasets which may have future implications in terms of policy, research and or operations.

The Data Legacy has been linked with the Group on Earth Observations (GEO) activities and integrates some of the common data related aspects of their work as a result through the “Impacts of drought” sub task (Task WA-06-02b). A spin off has been the inclusion of the DRI Data Legacy as a decision support application (in response to a June 2009 call for proposals, benefitting societal benefit areas (Water, Agriculture, Weather and Climate)) that are of importance to GEO.

4.0 Reverse Impact Statement

4.1 Provide a reverse impact statement, describing what would have happened in terms of the project, the resulting science and the impacts on users/stakeholders, if the work had not been funded by CFCAS.

If DRI Data management had not been funded by CFCAS the data collection, management and dissemination during the project would not have been centralized leading to inefficiencies, replication of effort, and limited access to computationally demanding datasets. The CFCAS investment in DRI would not have gone as far as it has if data management had not been centralized. The DRI Data legacy would not exist which is a crucial summary and dissemination of the entire networks work. Future drought science in western Canada would not benefit from the CFCAS investments in DRI if the data utilized was not archived publicly; future drought research efforts would have to start at square one data wise.

6.0 Dissemination

6.1 Provide information on the dissemination of the research results (publications, including journal names and whether refereed), conference contributions, seminars, workshops or videos, websites or other methods of transferring the results.

Article for the CMOS Bulletin is in preparation.

6.2 Describe data management/sharing activities including organization of the metadata. Also, are the data being archived, and how will they be made available to other researchers?

The Data legacy provides the archival framework for the network. A selection of metadata attributes is stored in a searchable publicly accessible database which links to detailed metadata files and data files. This system is hosted on the McGill servers and (redundancy will be built into the system) and is accessible through the DRI website (www.drinetwork.ca). The Metadata follows the ISO 19115 standard. Data (due to the diversity in datasets) does not have a standard format (txt and xls are the primary formats utilized) but effort has been made to standardized the organization of the data. Datasets are available by downloading zip files from the DAI servers (redundancy will be developed)