

TECHNICAL MEMO

Center for Atmospheric Research

University of Massachusetts Lowell

600 Suffolk Street, Lowell, MA 01854

Prof. Paul Song, Director

Voice: 978-934-4900 Fax: 978-459-7915

umlcar.uml.edu

DATE: OCTOBER 6, 2011

Revision: 0.3

From: Ivan Galkin (x4912) Ivan_Galkin@uml.edu

DATA CENTER REPLICATION SOLUTIONS

EXECUTIVE SUMMARY

The Global Ionospheric Radio Observatory (GIRO, giro.uml.edu) operates a data center holding over 35 million records of HF vertical incidence sounding of the ionosphere at 65 Digisonde locations for 1987-2011 period. Through individual agreements with 42 Digisonde observatories, their real-time data are provided to GIRO in several minutes or less from the measurement. Two databases of the data center, Digital Ionogram DataBase (DIDBase) and plasma Drift dataBase (DriftBase) provide the research and application community with public remote access for interactive data analysis and processing using the UMLCAR-developed software tools (SAO Explorer and Drift Explorer) that connect to GIRO Data Center over the Internet.

Overseas connections of SAO Explorer and Drift Explorer to the GIRO data repositories in Lowell, MA, USA are experiencing periods of high latency due to the Internet bandwidth limitation and elevated data traffic access activities at the single point of GIRO access in Lowell. High latency of access to GIRO precludes seamless data analysis with online connection to GIRO databases, forcing remote users to manually download GIRO data to their local computer for offline analysis.

Establishing a mirror GIRO data center will allow interested organizations to cut down the latency in access to the historic and real-time GIRO data. The mirror data centers will be provided with content-replication software to automatically transfer newly acquired records of ionogram scaling across the network of mirrors. In this arrangement, multiple organizations will benefit from access to the near-real-time data of global ionospheric sounding.

DIDBASE Map of GIRO locations (July 2011)



TECHNICAL DESCRIPTION

Concept of Operation

The GIRO Data Center Replication system includes (a) snapshot copy of complete DIDBase and DriftBase contents at the time of the system initialization at UMLCAR, (b) partial peer-to-peer near-real-time replication of the GIRO ionogram scaling records (SAO and SAO.XML), and (c) yearly raw ionogram and drift data updates via mass media mailings. In this arrangement, the retrospective data archives are available in full at the mirror sites, real-time ionospheric characteristics are available throughout the mirror sites synchronously, and the voluminous real-time raw ionogram and drift data are still available only at Lowell node of GIRO until their yearly update is circulated via mail.

Hardware and Software

The hardware and software components of the replication system are listed in Table 1 and 2.

Table 1. Hardware components of GIRO Data Center Replication system

PLATFORM	COMPONENT	CAPACITY	CURRENT DEMAND	SUGGESTED HARDWARE, TBR
A	DIDBase/DriftBase Repository	7 TB	2.5 TB	Supermicro ICC ControlTower 7046T
B	Backup Platform	12 TB	2 x 2.5 TB	Supermicro ICC ControlTower 5037C

Table 2. Software components of GIRO Data Center Replication system

PLATFORM	COMPONENT	VERSION	DESCRIPTION	MANUFACTURER
A, B	Firebird	2.5	DBMS for DIDBase	Firebird (open source)
B	Firebird	2.5	Database Backup tools	Firebird (open source)
B	DIDB_FILL	1.4.19	Ionogram data ingestion daemon	UMLCAR
B	DRIFT_FILL	1.2.8	Drift data ingestion daemon	UMLCAR
B	DIDB_REPLICA	TBD	Scaling data replicator	UMLCAR
A,B	openSUSE	11.4	Operating system	OpenSUSE (open source)

Functionality

Once GIRO Data Center Replication system is installed at the hosting organization and configured for Internet access, local SAO Explorer and Drift Explorer workstations can be configured to use it instead of the master servers in Lowell. This will provide rapid access to retrospective GIRO data that were used to populate the mirror server at the time of its original configuration at UMLCAR.

The DIDB_REPLICA software is responsible for upload of records to all subscribed mirror sites. The record replication is scheduled by the database that senses new data ingestion event. Due to the current Internet bandwidth limitation, only scaled data (SAO and SAO.XML records) will be replicated overseas using automatic DIDB_REPLICA processor. The voluminous raw ionogram and drift records will be provided for offline ingestion.

CONTACT INFORMATION

Further information on status and availability of GIRO Data Center Replication Systems can be obtained from UMLCAR; contact person Ivan_Galkin@uml.edu.