

Data Transfer and Data Center Session

Oral Papers

- ◆ 14:00-14:15 Michael Scharber, Carey Noll
Enhancing the IGS Data and Products Infrastructure – A Data Center Perspective
- ◆ 14:15-14:30 Yuki Hatanaka, Atsushi Yamagiwa, Masao Iwata, Shigeru Ootaki
Addition of Real Time Capability to the Japanese Dense GPS Network
- ◆ 14:30-14:45 Ron Muellerschoen, Mark Caissy
The Role of IGS Data Centers in the Context of Real-Time Data Streams
- ◆ 14:45-15:00 Greg Anderson, Kyle Bohnenstiehl, David Mencin, Mike Jackson
Building and Operating the Plate Boundary Observatory Using New Technologies
- ◆ 15:00-15:15 Heinz Habrich, Kurt Herzberger
New Server Concept at the BKG Data Center
- ◆ 15:15-15:30 Yehuda Bock, P. Feng, B. Gilmore, P. Jamason, D. Malveux, R. Nikolaidis, L. Prawirodirdjo, M. Scharber
SOPAC IT Developments – Tools for Research and Information Exchange
- ◆ 15:30-15:45 *Poster Summaries/Discussion/Wrap-Up*

Data Transfer and Data Center Session

Poster Papers

- ◆ Pil-Ho Park, Kwan-Dong Park, Jong-Uk Park, Hyung-Chul Lim and Jeong-Ho Joh
Proposing to Host the Fourth Global Data Center at Korea Astronomy Observatory
- ◆ Ignacio Romero, J. Dow, C. Garcia, J. Perez, E. Rojo, J. Feltens, H. Boomkamp
ESA/ESOC IGS Activities
- ◆ Stefan Schaer, Michael Meindl
Availability and Completeness of IGS/IGLOS Tracking Data
- ◆ Guenter Stangl, Peter Pesec, E. Cristea
The CERGOP2 Database - Information for Geodynamics in Central Europe
- ◆ Carey Noll, Maurice Dube
The IGS Global Data Center at the CDDIS: An Update



Enhancing the IGS Data and Products Infrastructure – A Data Center Perspective

Michael Scharber
Scripps Orbit and Permanent Array Center (SOPAC)
Scripps Institution of Oceanography (SIO)
La Jolla, CA USA

Carey Noll
Crustal Dynamics Data Information System (CDDIS)
NASA GSFC
Greenbelt, MD USA

Position Paper for the Data Transfer and Data Center Sessions
IGS Workshop 2004
Berne, Switzerland

March 02, 2004

Enhancing the IGS Data and Products Infrastructure - A Data Center Perspective

Overview

- ◆ Introduction
- ◆ Topics for Discussion
 - Real-time data and products
 - High-rate data
 - Improvements in data distribution and revision notification
 - GPS station/monument identification ideas
- ◆ Summary

Introduction

- ◆ Data Center Challenges:
 - Reduce latency of data
 - Increase size of on-line archives
 - Maintain high-level of availability of servers
 - Provide redundant archives
 - Expand archives to support pilot projects and working groups
- ◆ Recent influences:
 - Real-time data
 - High-rate data
- ◆ Implication:
 - Strain on data centers has increased

New Challenges for Data Centers

- ◆ Real-time GPS data and products
- ◆ High-rate GPS data archiving
- ◆ Improving IGS data file distribution and revision notification
- ◆ Improving GPS station/monument identification

IGS Data Centers and Real-Time GPS Data and Products

- ◆ Should IGS Data Centers distribute real-time data?
 - If so, what centers, how many stations, what s/w and h/w requirements, how to access data streams, etc.?
- ◆ Possible burden/cost/challenges for DCs:
 - Oversight, maintenance, policy familiarity
 - Hardware and network allocation
 - Community response, status, accountability
 - Redundancy, confusion with other real-time services
 - Computer system and networks security issues

Real-Time GPS Data and Products

IGS Data Center Participation

- ◆ Currently several, heterogeneous real-time solutions are used in GPS community
 - How do users find, evaluate and consume data from different sources?
- ◆ IGS DCs could provide a “homogeneous” circuit to diverse real-time data streams
 - Receive/accept data from allowed sources
 - Forward data to pier real-time services
 - Provide data consumption/retrieval/subscription service for users
- ◆ How do DCs:
 - Facilitate the injection of real-time data from one system to another?
 - Identify real-time streams within the system in an unambiguous manner?

Real-Time GPS Data and Products

Recommendations

- ◆ Clarify the role(s) of IGS DCs w.r.t. real-time data and products
- ◆ Participate actively in the IGS RTWG discussions and policy-making
- ◆ Address involvement with recommendations by the RTWG

IGS Data Centers and High-Rate GPS Data

- ◆ Many IGS stations operate at higher frequencies and data are decimated for “final” archive and distribution
- ◆ Will users want data from more sites at 1Hz?
- ◆ Is archive of “raw data” at operational data centers sufficient? For how long?
- ◆ SCIGN example:
 - 1 site at 1 Hz = 85 Mb/day
 - 250 sites = 21 Gb/day => 7.5 Tb/year!

High-Rate GPS Data

DC Archive Questions

- ◆ Who is responsible for permanent storage of high-rate data in its original format?
 - ODCs
 - How long?
- ◆ How long should high-rate data be made available by DCs in its original format?
 - GDCs and RDCs archive in RINEX
 - Should raw data for high-rate data be used?
- ◆ If GDCs do not archive data in original format, how will users find these data?
 - Use GSAC
- ◆ Should only a subset of high-rate sites be archived at GDCs?
 - Site selection

High-Rate GPS Data

Recommendations

- ◆ Publish high-rate data to the GSAC to ease future discovery
 - IGS DCs should participate in the GSAC activity
- ◆ Establish guidelines for DCs concerning the life of high-rate GPS data
 - What role should DCs play in permanent storage of high-rate data?
 - How long should DCs retain high-rate data in their original state?
 - After that time, what should DCs do with these data?

IGS Data File Distribution and Revision Notification

- ◆ Data revision is a problem
 - For DC
 - For user
- ◆ Data archive (and user)
 - Must learn of data revisions
 - Develop procedures to retrieve/archive revised data
- ◆ Data provider
 - Know when a revision is required
 - Follow prescribed procedures to notify DCs and users
- ◆ Users acquire data by:
 - Initiating a “pull” (shopping)
 - Receiving a “push” (subscription)

Data File Revision Notification

“Shopping” Scenario

- ◆ Shopper may or may not know exactly what, when, where about the data they require
- ◆ Shoppers need to be able to easily read about revisions
- ◆ Possible solutions:
 - E-mail notification to listserv
 - Maintain publicly-available revision log
 - Append version identifier to data filename
 - Insert version identifier in data file

Data File Revision Notification

“Subscribing” Scenario

- ◆ Subscribers know what, where, when they want data on a regular, often automated, basis
- ◆ DCs control what and when data are provided
- ◆ UCAR’s Local Data Manager (LDM) could be used to manage subscription-type data exchanges
- ◆ Revision notifications can be incorporated into this scenario and would simplify process of re-publishing re-submissions
- ◆ Possible solutions:
 - E-mail notification to listserv
 - Maintain publicly-available revision log
 - Submission of notification within software system (e.g., LDM)

Data File Revision Notification

Recommendations

- ◆ Establish guidelines for data files revisions within the IGS
 - Types of files subject to revision notification
 - Types of revisions requiring notification
 - When notifications should occur and by whom
 - What mechanisms should be used to issue notifications
- ◆ Define structure and content of data file revision messages
 - Investigate XML for future implementation
- ◆ Investigate software (e.g., LDM) to establish an inter-archive data file distribution system with a defined network topology

Improving GPS Station/Monument Identification

- ◆ Continued increase in number of continuous GPS stations, both within IGS and in other programs (e.g., EarthScope, etc.)
- ◆ Could burden DCs (not only IGS) with GPS station identification strategies
- ◆ Four-character identifier could cause conflicts for users getting data both within and outside of the IGS
- ◆ IGS could influence alternative identification scheme or brokering service

GPS Station/Monument Identification

Recommendations

- ◆ Investigate alternative GPS station identifier scheme
 - Avoid identifier clashes
 - Add value to name of data file by identifying where the data are associated, etc.
 - Avoid dependency on agencies
 - Avoid dependency on purpose of the data
- ◆ Develop software or implement a centralized name-brokering service
 - Decode/encode station IDs

Summary

- ◆ DCs need to coordinate to overcome future challenges
- ◆ Encourage participation in IGS Data Center Working Group to further these ideas
- ◆ Meeting tonight 6:00-7:30 p.m., Room B77