**IGS**

# SESSION 9:
# DATA CENTER ISSUES

# Current Status of IGS Data Centers

Carey Noll
*NASA GSFC, Greenbelt MD, USA*

**Abstract:** The International GPS Service (IGS) has been operational for nearly ten years. This presentation will outline background information, current status, and recent developments at the IGS data centers. The overview will also include a review of data and product holdings at the various IGS data centers as well as statistics on data delivery. An introduction to the new IGS working group on data centers will be given.

## Background

The International GPS Service (IGS) has been an operational service within the IAG since 1994. The IGS has established a hierarchy of data centers to distribute data from the network of tracking stations: operational, regional, and global data centers. This scheme provides an efficient access and storage of GPS data, thus reducing traffic on the Internet, as well as a level of redundancy allowing for security of the data holdings.

## Recent Data Center Developments

The data and product types currently archived at the IGS Global Data centers are summarized in Table 1 below.

**Table 1. IGS Global Data Center Holdings**

| Data Type | CDDIS | IGN | SIO |
|---|---|---|---|
| **Data** | | | |
| GPS daily (D format)* | X | X | X |
| GPS daily (O format) | X | | X |
| GPS hourly (30-second)* | X | X | X |
| GPS hourly (high-rate) | X | | |
| GLONASS daily (D)$^\dagger$ format)$^\dagger$ | X | X | |
| GLONASS daily (O)$^\dagger$ format) | X | | |
| **Products** | | | |
| Orbits, etc.* | X | X | X |
| SINEX* | X | X | X |
| Troposphere$^\dagger$ | X | X | X |
| IONEX$^\dagger$ | X | X | |

Notes:  *  Official IGS data set/product
         †  Pilot project/working group data set/product

In 2001, approximately sixty percent of the daily GPS data files were available from the IGS global data centers within three hours; the same percentage of hourly, thirty-second files were available within fifteen minutes.

*Real-Time Issues.* Data center involvement in the archive and dissemination of real-time data will be dependent upon requirements developed by the IGS Real-Time Working Group and recommendations generated by this workshop. Early discussions imply that existing data centers would serve as a distribution or relay center, receiving real-time data from a network of stations and transmitting these data to interested analysis centers. Data center configuration information, such as storage capacity, network bandwidth, and redundant network connectivity needs to be determined for participating data centers. Redundancy of data flow paths is an obvious concern for the real-time activity.

*IGS Data Center Working Group.* At its last meeting in December, the IGS Governing Board recommend the formation of a working group to focus on data center issues. This working group will tackle many of the problems facing the IGS data centers as well as develop new ideas to aid users both internal and external to the IGS. The direction of the IGS has changed since its start in 1992 and many new working groups, projects, data sets, and products have been created and incorporated into the service since that time. Therefore, it is now an appropriate time to revisit the requirements of data centers within the IGS. The membership of this group will consist of contacts from the current IGS data centers as well as IGS colleagues with expertise in data archiving and data flow. Thus far, a draft charter has been developed and prospective members have been contacted; the charter will be presented at the IGS Governing Board meeting after this workshop for approval.

# GPS Seamless Archive Centers (GSAC); Streamlining Data/Metadata Exchange in the GPS Community

*Michael Scharber*
Scripps Orbit and Permanent Array Center

## Abstract

In an effort to help simplify and streamline both the discovery of, and access to, GPS-related data and metadata the GSAC, or GPS Seamless Archive Centers (a UNAVCO project) aims to build a cohesive, structured and replicated data/metadata exchange environment for all types of GPS users, data centers and researchers around the world. Through the formation of a structured set of cataloging mechanisms and participation of a network of GPS data providers, archive centers and top-level "brokers", called retailers, the GSAC offers users of GPS data the opportunity to span the entire GSAC-published data holdings of multiple physical data centers in a single "query" for information, or data collection session.

Providing a simplified and centralized point of access to the GSAC for users are the GSAC retailers - those agencies maintaining a GSAC database server (with a common GSAC relational schema) and providing an anonymously-available GSAC retailer "service" through a predefined http CGI protocol. The GSAC retailer service, in turn, provides a very simple API for client applications designed for data discovery, data collection, or both. The set of GSAC client applications, ultimately, offer various interfaces to the same superset of GPS-related data holdings and limited metadata contained in each and every GSAC retailer. From a single retailer client application a user can collect hundreds, even thousands, of files from dozens of archives, matching complex query parameters such as a window of time, a spatial bounding box, metadata constraints, or a combination of all three.

Also of use to data centers, the structured environment of the GSAC, combined with the various utilities available, provides convenient mechanisms for exchanging/mirroring data with other data centers, uncovering data corruption problems and metadata errors, and minimizing the time and network traffic required for these and similar operations.

Nearing the end of the "test" phase of the GSAC project several agencies are embarking on operational integration of GSAC utilities and services for the regular activities. In parallel, end user applications and web-based interfaces are being developed and tested for ease of use and functionality during this phase, after which users will have full access to the entire GSAC database of information.

With the anticipated participation of additional data centers the GPS community shall hopefully begin to reap the many benefits we hope the GSAC will provide. For more information please visit SOPAC's GSAC project page (http://gsac.ucsd.edu) for more information about the GSAC in general, SOPAC's participation in the GSAC and access to GSAC-related applications and contact information.

**IGS Workshop 2002**
**Ottawa, Canada**
**Data Centers, Ideas and Issues**

Loïc Daniel, IGN, France Edouard Gaulué, IGN, France

In this paper, I'll try to address some of the points that seem of importance to me, this will inevitably be biased by what we see from our position at IGN as a Global Data Center (GDC). I won't try to encompass all the aspects to be dealt with in this session or about data centers in general. This paper is intended mostly as a repository of present or potential problems that I have identified and an incentive to discuss and find ideas among the community.

## 1. Present situation

First I'll describe the raw characteristics of the data management at IGS data centers the way we do it presently, I'll try to summarize our activities and evaluate the impact in terms of computer and network loads.

### *Basic data management*

These are the tasks that should be operated on the fly, with a minimal additional delay induced at each step. The data moving operations rely upon a layered structure of Data Centers. The goal is to propagate observations of the stations and products from analysis centers to final places where they are easily available to everyone. Most if not all of the files should end at the GDCs.

Two types of data flows may be considered:

1) file transfers between data or analysis centers as part of the general scheme defined by IGS in order to ensure the best performance in data availability to analysis centers and users. This is the "IGS data flow". It represents the main part of the day to day activities of the data centers. The objective is to put the data and product files at places where they can be downloaded by users. This is the part of the data flow that can be controlled, optimized and supervised because all actions are triggered by an identified component of the data network and following a predefined time table.

2) file transfers initiated by users of the service, this is the "users data flow". Users can be IGS analysis centers and any other kind of user of the IGS. This is much less controlled by data centers, the files are provided for download on an ftp server and users get them as they want without registering or making a special agreement whatsoever. In some cases, a user will issue a request for offline data and the data center will have to restore data and provide access but this tends to be the exception since most data centers (at least the global ones) strive for putting online all the files created since the beginning of the service.