IGS

Session 5:

Data Transfer and Data Centers

Session Description

Chairs: C. Noll, M. Scharber, C. Meertens

The Data Transfer and Data Centers session will present papers on current data center activities as well as various topics geared towards enhancing the availability and acquisition of IGS data and products. These topics include: strategies for reducing data file latency at IGS Data Centers, standardization of data file replacement notification methods, improving the efficiency of data file transfer among IGS Data Centers and defining a network topology for GPS data files for use in applications.

Session Summary

C. Noll, C. Meertens

The Data Transfer and Data Centers session presented papers on current data center activities as well as various topics geared towards enhancing the availability and acquisition of IGS data and products. Carey Noll opened the session with a review of the position paper drafted by Michael Scharber and Carey. Recommendations were given for data center involvement in real-time activities, archive of high-rate data, and handling of data file replacements.

The session continued with Yuki Hatanaka's (GSI) presentation on the GEONET in Japan. This network consists of over 900 permanent GPS receivers and will expand to 1200 receivers in the near future. The network is operated at 1 Hz and data transfer occurs in real-time. A quick analysis of the data is performed every three hours using IGS ultra-rapid orbits to monitor crustal deformations and provide reference stations for land surveying. The 1Hz data are archived for three weeks; after that time, the high-rate data are removed and 30-second data are permanently archived.

Ron Muellerschoen (JPL) discussed the Real-Time Working Group's recommendations for IGS Data Centers. He pointed out that the traditional role of the IGS DCs has been in the archive and distribution of files; real-time means moving from data distribution through files to data distribution through data streams. He recommends that IGS DCs initially become involved in real-time activities by subscribing and listening to data streams and thus cataloguing these streams and providing quality monitoring of their integrity.

Greg Anderson (UNAVCO, Inc.) gave a presentation on the Plate Boundary Observatory (PBO), which is part of the NSF-funded EarthScope initiative. This effort will initially involve nearly 900 continuously operating GPS receivers plus a network of strainmeters. The network will operate at 1 to 5 Hz. An Internet map server, ArcIMS (arcims.unavco.org), has been developed for geospatial referencing, mission planning, and user interface. The system uses Landsat, digital elevation models, and 1-meter aerial maps to help planners in site reconnaissance. UNAVCO and SOPAC will serve as data archives for PBO.

Heinz Habrich (BKG) gave an update on the new LAMP (Linux, Apache, MySQL, and Php) server installed at BKG. This data center serves as a regional data center for the IGS, the EUREF data center, and a German national data center; these user groups have thus influenced the design of the archive with project-specific directories. The system currently accumulates data streams that are then archived as hourly files. Future plans include serving as a GSAC wholesaler.

Yehuda Bock (SIO) discussed some of the recent IT developments at the Scripps Orbit and Permanent Array Center (SOPAC). These projects include enhancements to the GPS Seamless Archive (GSAC) web client wizard, command line client, and map server. GPS receivers in the Southern California Integrated GPS Network (SCIGN) have been upgraded to real-time with the cooperation of county governments. A tour of the SOPAC website highlighted the use of XML in several applications.

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

Michael Scharber
Scripps Orbit and Permanent Array Center
Cecil H. and Ida M. Green Institute of Geophysics and Planetary Physics
Scripps Institution of Oceanography
9500 Gilman Drive
La Jolla, CA 92093-0225 USA

Carey Noll
Crustal Dynamics Data Information System (CDDIS)
NASA Goddard Space Flight Center
Code 922
Greenbelt, MD 20771 USA

Abstract

The Data Transfer and Data Centers session will present papers on current data center activities as well as various topics geared towards enhancing the availability and acquisition of IGS data and products. These topics include: strategies for reducing data file latency at IGS Data Centers, standardization of data file replacement notification methods, improving the efficiency of data file transfer among IGS Data Centers, and defining a network topology for GPS data files for use in applications.

1. Introduction

Over the past decade challenges facing IGS Data Centers (DCs) have typically involved: reducing data file latency, improving the quality of service to the community, increasing the amount of data available to the public, maintaining high-availability servers, managing a systems infrastructure to support local analysts, and providing backup mechanisms for archived data, etc. The data and product archive and distribution requirements of the IGS DCs have expanded over the last ten years in response to new applications, pilot projects, and working groups established within the service. These new programs include the LEO, Real-Time, Ionosphere, and Troposphere working groups, and the TIGA and GLONASS Pilot Projects.

Though many of these challenges have been handled with some degree of success, none were dealt with in the face of real-time and/or high-rate GPS data of the breadth and magnitude available today. The constraints facing IGS DCs in a real-time environment, especially one where sampling intervals continue to increase in frequency, not only magnify existing DC challenges, they pose several new ones.

The increasing demand worldwide for GPS data at more frequent sampling intervals and decreased latency means IGS DCs have a clear responsibility to the global GPS community to make all reasonable attempts to facilitate the timely acquisition of GPS data and data products. For end users there are a multitude of fascinating and valuable applications of GPS data in low latency and/or high-rate contexts. Unfortunately, DCs providing these data face various challenges stemming from decreased tolerance when delays or outages of such services occur.

As the “window” of tolerance lessens, the strain on DCs tends to increase. This position paper is an attempt to highlight some of the challenges facing GPS data centers, in particular those associated with the IGS, as GPS data becomes: a) more popular, b) more voluminous, and c) more frequently produced/sampled. Most importantly, we outline a coordinated effort to streamline the IGS DC infrastructure of inter-archive, GPS data publication, retrieval, storage, and resubmission.

2. Discussion

The expansion of the IGS in real-time GPS analysis brings to the forefront a number of challenges for DCs and the infrastructure of Data Providers they depend on. The ways in which DCs choose to meet these challenges can, and will, impact a significant portion of their operations into the foreseeable future. In particular, it is still unclear exactly what role(s) they will play in the [re]distribution of real-time GPS data and data products. High-rate GPS data and data products also pose a number of significant hurdles to overcome, given the sheer volume of data involved and uncertainties surrounding the handling of these large data sets over time. Coordinating the identification of geodetic monuments (continuous GPS stations) in an unambiguous fashion throughout a variety of real-time, near real-time, and latent contexts is yet another issue that should be dealt with. So is the issue of how to most effectively deal with data resubmissions and recalls in the IGS.

Some of these issues are relatively new to IGS DCs. Some are not. All of them can be overcome with effective communication and active participation in IGS Data Center Working Group (DCWG) efforts. We have chosen four general topics to highlight in this position paper in order to catalyze useful dialog for the session as a whole. Recommendations, categorized as short term and long-term actions, are given following each topic.

2.1 IGS Data Centers and real-time GPS data/products

We begin the discussion of DC issues by introducing the topic of real-time GPS data and data products, and the role(s) DCs will play in the IGS real-time strategy.

At present, IGS DCs know very few details, if any, regarding the role(s) they will be expected to play in order to best support the IGS’s overall real-time strategies. We begin this discussion by asking the question: Will DCs be involved in the distribution of real-time data? If the answer is yes, then:

- Which centers?
- In what capacity?
- How many stations?
- Under what kinds of policies and constraints?
- Using what software and/or technologies?
- At what cost?
- How will users be able to access these data streams? From a single centralized source, or a set of sources?
- If the “streams” are distributed, how will they be indexed and published to the public?
- How will users subscribe/retrieve these data?
- How will individual streams be identifiable to the end user (or archive)?
- Will non-IGS data be allowed to join the same “pool”?
- Can the same backbone servicing real-time data streams facilitate the transmission of near real-time GPS data files, daily data files, products, site logs, and anything else exchanged among DCs?

If the answer to the above question is no, then:

- Why not? Thousands of people around the world already go to IGS DCs for most of their GPS-related needs. It seems a shame not to take advantage of this existing relationship and utilize these nodes for the centralized [re]distribution of real-time GPS data, for IGS and non-IGS stations alike. Without the involvement of these centers users of real-time GPS data will likely face an old problem, with a new twist – finding data of interest, determining its availability, format and fitness-for-use and acquiring it and massaging it for their own application.

These are all valid issues from a DC perspective. Assuming concise information on these topics is made available during this workshop, or shortly thereafter, it remains likely that numerous additional challenges will arise in the near future.

One such challenge is how to facilitate the discovery, evaluation, and consumption of data from heterogeneous real-time sources (protocol, policy, data type) by end users. Assuming an effective strategy for handling real-time IGS data and data products is produced and distributed, there may be a significant burden for DCs to participate in the system. Some possible costs include:

- oversight, maintenance, policy familiarity
- hardware and network allocation
- community response, status, accountability
- redundancy, confusion (w.r.t. other real-time services/functions)
- security of computer systems and networks

Centralized real-time GPS data [re]distribution?

The existence of numerous real-time GPS solutions throughout the GPS community raises an interesting issue with regard to IGS DCs – each of which may be familiar with one or more real-time systems already. How will DCs with pre-existing real-time functions deal with the possible advocacy of a different real-time system on behalf of the IGS? Will they be urged to maintain additional system(s)? Or will it be possible to perform real-time services for the IGS while accommodating other projects at the same time, using the same software/strategy for both purposes?

From an end user perspective, having additional DCs to deal with is probably less desirable, excluding adventurous/bleeding-edge users with extra time on their hands and plenty of expertise. Is it desirable, resources and politics aside, to facilitate links in otherwise incompatible real-time GPS systems by "injecting" streams from many systems into a single, homogeneous system? Is this a strategy at all entertained by the IGS Real-Time Working Group (RTWG)? If not, why? If so, what do those strategies entail and how can IGS DCs help?

The fact is, new, real-time GPS systems are being developed each year. The IGS should recognize this trend and offer, or at least support, an effort to provide a homogeneous "circuit" GPS users can tap into to acquire a diverse set of data streams from a single source. In this model, participating IGS DCs could install and maintain software, which serves three primary roles:

a. receipt/acceptance of real-time "injection" events from a set of allowed sources
b. perform "forward" actions to pier DC real-time services of each injected piece of information
c. offer end users a publicly-available, and well-documented, mechanism for data consumption/ retrieval/subscription

Beyond the political aspects of real-time data redistribution, crucial points in this topic include:

- How to facilitate the "injection" of real-time GPS data from one real-time system to another, wholly different system
- How to identify real-time GPS data streams within the system in an unambiguous manner of use by an end user

The traditional role of IGS DCs has been in the distribution of GPS data and products in a file-oriented mode. Real-time data are distributed in a streaming mode. Therefore, a logical question to ask is what is the added value of the participation of IGS DCs in the distribution of data in streams? Initially, the data centers could become involved in more traditional roles in the real-time era: cataloguing of streamed data and monitoring of the data quality of the streams. DCs should therefore subscribe and listen to data streams to perform these functions. Additional involvement in handling real-time data streams could be addressed after experience is gained from these initial efforts.

Depending on the degree to which IGS DCs are involved in real-time GPS data [re]distribution, and the structure of that involvement, issues such as those described above will undoubtedly surface and receive the attention they deserve. They have been raised in this paper to help make DCs aware of looming concerns, if they are not already facing these issues.

Recommendations

a. Clarify the role(s) of IGS DCs with respect to real-time data and data products. See issues outlined above.
   1. Archive metadata and monitor quality of transmitted data streams. [SHORT TERM].
   2. Offer end users a publicly-available mechanism for data consumption/retrieval/subscription. [LONG TERM].
b. Participate actively in IGS RTWG discussions and policy-making. [SHORT TERM]
c. Address involvement with recommendations by the RTWG as feasible.

2.2 High-rate GPS data and their impact on IGS Data Centers

In this section we describe the storage aspects of high-rate GPS data, the role(s) expected of IGS DCs providing such data to the public, and how the provision of those services impacts the resources of the agencies providing them.

As data from an increasing number of GPS stations are made available in real-time, the natural impetus for using these data sets push the definition of "real-time" (at least for GPS analysis) closer and closer to instantaneous by increasing the frequency with which GPS observations are made. If these data are used (at their original interval) in the future, which is likely, DCs may face a rather significant problem – permanent storage, and public provision, of high-rate data, from hundreds (or thousands) of GPS stations. Given the hierarchical nature of DCs, top-level archives (Global DCs) and middle-tier archives (Regional DCs), may experience an explosion in the amount of data they are expected to manage. How the IGS chooses to deal with this explosion, from the standpoint of public retrieval (real-time, near real-time and historical) will have a significant impact on DC operations. To help put the significance of this trend into perspective, from a DC standpoint, we offer the Southern California Integrated GPS Network (SCIGN) and CDDIS real-time data archive as examples.

In the SCIGN network, nearly all stations are occupied by Ashtech Z-XII3 receivers sampling at either 120-, 30-, 5- or 1-second intervals. Currently there are about 250 "live" stations. A typical daily raw GPS data file from the SCIGN network (at a 30-second sampling interval) is around 800 Kb in size. For 250 stations this amounts to around 200 Mb of space to store one copy of each raw data file, each and every day. In streaming mode, at a 1 Hz sampling interval, daily raw (MBEN) data files from these same receivers are typically 85 Mb in size.

If every SCIGN station were upgraded to a 1 Hz sampling interval, the resultant data files would require roughly 21 Gb/day to store a single copy of each. In one month those files would account for well over 0.5 Tb (~620 Gb). In one year that's nearly 7.5 Tb of space just to store one copy of each 1 Hz GPS raw data file produced daily from each SCIGN station (assuming 100% data recovery). If this still doesn't seem like a lot of data (to some it might not), keep in mind some archives may be required to store at least one, if not two, backup copies (online or offline), depending on their responsibilities. That would mean 15 Tb of additional space, per year, to store two copies of each SCIGN raw GPS data file! Of course, this scenario is partially hypothetical, and assumes permanent storage of all data files in their original state. However, the IGS might face a similar challenge.

The CDDIS has archived high-rate data in support of the IGS Low-Earth Orbiter (LEO) Working Group since mid-2001. Currently, high-rate data from approximately 55 sites are archived at the CDDIS each day. These data are submitted in files containing 15 minutes of data with a latency of 5-15 minutes on average. Observation, navigation, and a small amount of meteorological data in RINEX format are archived. The entire CDDIS archive of high-rate data for 2003 is approximately 140 Gb in size, implying an average of nearly 400 Mb per day of 1 Hz GPS data in RINEX format. The daily, 30-second RINEX data from these sites are also archived at the CDDIS in a separate directory structure.

If every IGS station were converted to high-rate sampling intervals, there could be a similar amount of data to deal with as that produced by SCIGN, if all SCIGN stations were sampled at 1 Hz. Scenarios such as these should raise a number of questions since it seems clear the trend in the GPS community is towards high-rate, and real-time.

One possible solution is to claim that permanent storage of high-rate GPS data is ridiculous and simply resample the files prior to permanent storage. Though it seems a shame to just throw the data away, doesn't it? Some questions this session should ponder:

a. Question: For any given high-rate IGS data set, who is responsible for the permanent storage of the data, if that is desirable?

   Response: Operational DCs are currently responsible for archiving their raw data for some period of time. Should this strategy be modified to include other agencies as well? What period of time is that?

b. Question: How long should high-rate data (raw, RINEX, etc.) be made available by DCs in its original state, if at all?

   Response: The IGS has always required DCs to store RINEX files. Should this strategy by modified to include raw data files as well?

c. Question: Should only a subset of high-rate IGS stations be subject to permanent storage of their data at the original sampling interval?

   Response: How is one to know what distribution of stations is most desirable, to most people, months or years from now? And, how will variations in data recovery, day-to-day, or month-to-month, affect this distribution?

d. Question: If top-tier archives cannot, or will not, store certain IGS data sets in their original state, where will the data reside and how will users find them in the future when comparable data sets are distributed across dozens of archives?

   Response: Propose "publication" to the GSAC for long-term preservation, discovery and acquisition.

Scenarios where agencies are thrust into an accelerated increase in the amount of data, and frequency, they're expected to manage call for policies to govern the "life" of the data – giving the agencies involved clear guidelines on how to manage these large data sets, in other words, where the data are generated and how often, where the data reside, in what state, and for how long, and how the data can be retrieved, if at all, how often, and by whom.

Answering these questions will help DCs devise a set of guidelines for managing high-rate GPS data and data products, aid in their resource planning considerably, and provide clarity for everyone involved – data providers, DCs, and end users alike.

One possible resource the IGS could certainly leverage to help "track" a diverse and distributed collection of high-rate GPS data files is the GPS Seamless Archive Center (GSAC – http://gsac.ucsd.edu, http://www.unavco.org/facility/data/gsac/gsac.html), a virtual archive composed of an assembly of agencies and investigators exchanging information about their respective GPS-related data holdings in a well defined, cohesive manner. Now operational, with an index of over 2 million GPS data files pertaining to over 10,000 different geodetic monuments, the GSAC is an ideal resource to help the IGS in this aspect.

Recommendations

a. IGS operational data centers should archive raw receiver data indefinitely and provide access to these data upon request for data revision or scientific study on a limited basis. [SHORT TERM]
b. High-rate GPS data are valuable. If they are stored permanently in their original state, the data should be published to the GSAC to ease future discovery by end users and DCs alike. DCs should participate in the GSAC if they are not already. See http://gsac.ucsd.edu for more information. [SHORT TERM]
c. Devise guidelines concerning the "life" of high-rate GPS data and data products with respect to DCs. Specific points to address include:
    1. What role(s) will DCs play in the permanent storage of high-rate GPS data and data products, if any? [SHORT TERM]
    2. How long should DCs retain high-rate GPS data files in their original state, if at all? [SHORT TERM]
    3. What should DCs do with high-rate GPS data files after the period defined in (b.2.) has been reached, if anything? [SHORT TERM]
    4. Should high-rate GPS data files be mixed with other data files of the same type(s) at DCs, or stored in a separate location for users to retrieve? [SHORT TERM]

2.3 Improving IGS data file distribution and revision notification

In this section we address possible improvements in IGS data file distribution and strategies for notifying interested parties of data file revisions and advisories.

In the IGS community there are effectively two ways to acquire data files, by initiating a "pull" (shopping) or by receiving a "push" (subscription). Either you get the data yourself, or it is given to you. Typically, users access FTP/HTTP servers and download (or read) data files (or information) of interest. In this scenario the burden is on the user to "find" what they want, when they want it. When a user does find something he/she wants, and reads it, saves it or uses it in some way, he/she may or may not be interested in knowing if that information changes in the future. If he or she is interested, how will they know? When will they know? In this section we address the topic of information revision notification and propose a means of streamlining the distribution of data files among DCs.

In the IGS revision notification typically takes the form of an e-mail announcement to IGSMail <igsmail@igscb.jpl.nasa.gov> by the information provider (see Figure 1), but not necessarily on a file-by-file basis, and not in a standard format. It is the responsibility of DCs, primarily, and end users (to some degree) to:

1. Find and make use of channels employed by their information sources to learn of revision events
2. a) Understand the mechanism(s) used to issue those notifications, b) understand their format, and c) devise a strategy to read those messages
3. Evaluate revisions with respect to their specific needs

It is the provider's responsibility to:

1. Recognize when a revision or recall is warranted
2. Adhere to applicable policies and react appropriately

```
******************************************************************************
IGS Electronic Mail      10 Dec 12:41:20 PST 2003      Message Number 4753
******************************************************************************

Author: Stephen Delahunt

Four Character ID        : NRC1
IERS DOMES Number        : 40114M001
City or Town             : Ottawa
State or Province        : Ontario
Country                  : CANADA
Agency                   : NRCan/GEODETIC SURVEY OF CANADA

ADVISORY:
=========
A new site log has been submitted for NRC1.
Receiver 234-U (AOA SNR-12 ACT) was replaced with
Receiver 1130 (AOA BENCHMARK ACT) on Dec 09(doy 343), 2003 16:00 UTC.

NOTE:
=====
RINEX files with incorrect header information were replaced at CDDIS
today at 16:03 UTC.  The affected data spans the period:

        Dec 09 16:00 UTC through Dec 10, 16:00 UTC.
```

Figure 1. Example of an IGS data resubmission notification sent to IGSMail by NRCan/Geodetic Survey of Canada pertaining to the GPS station NRC1.

The worst case involves either the provider not issuing revision notifications, or the consumer not knowing how to a) find the channel(s) used to issue notifications, or b) use those channels to their benefit.

Exactly how DCs can improve upon the current atmosphere of revision notifications needs to be discussed. One option for consideration would be packaging revision notification messages in an eXtensible Markup Language (XML) structure[1].

[1] SOPAC has invested heavily in the development of XML schemata, stylesheets and utilities to be used in the exchange of GPS or Geodesy-related information. XML is a powerful, and popular, platform-independent means of describing and exchanging information, though it is obviously not the only choice in this scenario. For more information about SOPAC's XML-oriented projects please visit http://sopac.ucsd.edu/projects/xml.

Below, we briefly outline possible solutions, in both scenarios, serving the best interest of the IGS overall. It is worth stating that a mixture of methods is desirable given the variety of users and contexts with which GPS data can be used.

Data file "shopping"

A typical data file shopper may or may not always know exactly:

- What they want before they shop
- Where they want to collect something they know they want
- When they want to collect what they want
- What it is they are collecting

As far as revision notifications are concerned, shoppers can be accommodated by making readily available to them a means with which they can "read about" a revision. Possible implementations include:

- E-mail notification to a publicly-available e-mail list server where every revision notice is posted (by providers), and users are urged to subscribe to
- Appending revision notifications (from providers) to a single, publicly-available, online log for web or ftp-based consumption
- Appending a version identifier to the revised file, in situ; likely to incur significant modifications to existing software
- Recording the revision of the file in the file itself by the data supplier, in situ

In situations where the "what" is either clearly defined, changes infrequently, and/or the source and consumer are fixed, and the "when" occurs regularly, "shopping" can saturate networks and waste resources. The consequences associated with repetitive data file shopping can be significant and varied.

Data file subscriptions

In contrast to data file "shopping", subscribers know what they want, know where to find it, and receive what they want on a regular, or at least source-determined, basis. Furthermore, the subscriber/provider relationship is typically more involved than a shopper/provider, primarily because the provider needs to "know" who their subscribers are, what they want, when they want it, and how they want to receive it.

Subscribers[2], once they have subscribed to a service, simply wait for the information they've subscribed to, and determine what to do with it when it arrives. Latency is minimized due to the event-driven nature of the strategy and the inherent network topology[3] of data flow – data providers controlling what is "provided" and when.

In this relationship providers may choose to issue notifications or warnings in the event certain provided information/services warrant revision or recall. In such cases the provider adopts policies to govern those notifications, giving the users a clearly defined means with which to implement revision

---

[2] In the IGS community the collection of "regular" users of IGS data files could very easily fall into the category of a "subscriber" to an IGS, subscription-based, data delivery system, if one were developed.

[3] Using LDM as an example, the network constructed will include a defined topology of source nodes and sink nodes – source nodes doing the "pushing", and sink nodes doing the "receiving". Any given node (DC or regular user of IGS data) could perform both functions if need be. The top-level archives (Global DCs) would, of course, receive everything (e.g., subscribe to everything) and push only to other Global DCs data files they are directly responsible for (e.g., a provider of).

notifications if they so choose. Though specifics on exactly how notifications would be handled in a subscription-based system need to be discussed, some possible events (along with who would benefit shown in parentheses) that could be triggered in a system using software such as the University Corporation for Atmospheric Research (UCAR) Local Data Manager (LDM – http://my.unidata.ucar.edu/content/software/ldm/) are:

- E-mail notification (subscribers)
- Submission of an information notice within the system itself (subscribers)
- Appending the notice to an online log for web or ftp-based consumption (shoppers)
- Noting the resubmission in a database (shoppers)
- Appending a version identifier to a publicly-archived file (shoppers)
- Recording the resubmission in the file itself (shoppers)

Implementation of a subscription-based system for IGS data files could reduce the latency of near-real time data distribution, by establishing a mechanism for notifying subscribers of revisions and recalls, and reducing the overhead[4] involved with inter-archive data distribution. Even if the only participants of such a system were DCs themselves, synchronization of the GPS data at those centers would be more robust using a system like LDM.

The main point is that a subscription-based strategy could simplify the process of actually identifying and publishing resubmissions, for future "shoppers" as well as immediate recipients on a subscriber list. The driving mechanism is event-based information handling. How events are actually dealt with is a topic DC members should discuss.

Recommendations

a. Establish guidelines for data file revisions within the IGS, including:
    1. What types of files are subject to revision notification [SHORT TERM]
    2. What types of revisions warrant notification [SHORT TERM]
    3. When notification should occur, and by whom [SHORT TERM]
    4. What mechanisms will be used to issue those notifications – there may be many [LONG TERM]

b. Define the structure and content of data file revision messages, possibly using XML Schema [LONG TERM]

c. Investigate using LDM, or comparable software, to establish an inter-archive data file distribution system with a defined network topology. [SHORT TERM]

d. Define and implement an automated, centrally managed subscription interface to the system mentioned in (c). [LONG TERM]

2.4 Improving GPS station/monument identification

The continued increase in the number of continuous GPS stations installed and activated each year, combined with the imminent arrival of integrative scientific settings such as EarthScope, will continue to burden GPS data centers (IGS included) with issues related to GPS station identification strategies

[4] An event-driven, subscription-based system like LDM would remove a significant amount of burden from DCs with respect to populating their archives for the IGS. They would no longer have to search for files at other DCs – files would automatically be distributed into the LDM network by the responsible agency. DCs would either distribute ("push") files into the LDM network, or simply receive files they aren't responsible for.

and file (or data stream) naming conventions. In this section we describe the concern, again from a DC perspective, that the common strategy of identifying GPS stations (to other agencies) using 4-character strings could be harmful (if not supported by a robust brokering service) in data distribution contexts involving new technologies, integrated science and/or data portals[5].

How could it be harmful? What happens when a DC or end user finds/collects two GPS data files from two different places, with two different 4-character identifiers, only to find the data files are actually identical[6]? Or worse, they determine that two files, collected from two different places, with the same 4-character identifier, pertain to two entirely different GPS stations?

New GPS data distribution contexts present, among other things, a valuable opportunity to investigate alternative GPS station (or monument) identification strategies – to avoid propagating station identifiers based on 4-character strings into modern technological solutions.

The IGS should use its reputation in the global GPS community to support an alternative GPS station identification scheme, and/or identifier brokering service, to relieve other sciences/applications/portals from the burden of accurately determining the source of a given GPS data file (or data stream) without opening the file (or analyzing the data stream). With a strategy like this in place, both existing and future geophysical data distribution contexts would benefit – at the very least, in reconciling ambiguities or uncertainties regarding station identification between one or more data files or data streams.

Recommendations

a. Investigate an alternative GPS station identifier scheme to better facilitate the exchange of GPS data files within and beyond the IGS, and GPS, communities. This scheme should address, at least, the following issues [LONG TERM]:
    1. avoidance of identifier clashes (two different stations/monuments with the same identifier)
    2. add value to the "name" of a data file or information stream by identifying "where" the data are associated with, among other things
    3. avoiding any dependency on individual agencies, past or present, apart from a centralized name-brokering service (if one is employed)
    4. avoiding any dependency on the "purpose" of the data, the sensor used or date/time the identifier was generated (e.g., no sequences)
b. Develop and make available software, or a centralized name-brokering service, that can be used, by anyone, to encode or decode station identifiers based on:
    1. current 4-character identifiers and conventions [SHORT TERM], and
    2. the scheme defined in (a) [LONG TERM].

3. Summary

DCs will face a number of challenges in the near future. These challenges can be overcome with a coordinated effort and dedicated approach. As the value of GPS for scientific, commercial, government, and military applications continues to increase so will the pressure on DCs to provide an intuitive

---

[5] These concerns do not aim to modify the RINEX file-naming convention or change the way any organization names their geodetic monuments. Issues in this section are raised in light of multi-disciplinary scientific research, centralized geophysical data portals or clearinghouses and integrated real-time data distribution systems – where there are no preexisting dependencies on a particular style of station or monument identification.

[6] If one DC employs an aliasing mechanism for data files it collects/receives from another organization, to avoid confusion with a set of files on its own system that pertain to a completely different geographic location, this scenario is very likely to occur.

interface for the GPS community and reliable service from which users can acquire the data and information they need. Real-time systems, high-rate data, and the sheer number of GPS stations in use today will make things interesting for DCs in the meantime. We have highlighted several key topics concerning DCs and their role in the GPS community. These topics are meant to stimulate discussion and proactive involvement in the IGS, and beyond. Many of these recommendations can be further discussed in the framework of the IGS DCWG. However, successful implementation of the resulting recommendations will require commitment on the part of all DCs.

4. Acknowledgments

The authors would like to thank Chuck Meertens, Lou Estey, and Myron McCallum (UNAVCO, Inc.) and Yehuda Bock (SOPAC) for providing valuable feedback on this paper.