IGS Network Systems Workshop, Annapolis, MD November 2-5, 1998

Position Paper #5.0

DATA CENTER ISSUES

Loic Daniel
Jeffrey Dean
Myron McCallum
Carey Noll

Abstract

The IGS has been operational for nearly five years. Recent demands on the stations and data centers prompts the review of data flow and archiving methodologies. This position paper outlines the current structure and details problems to be addresses as well as questions for future implementation.

1. Introduction

The International GPS Service (IGS) was formed by the International Association of Geodesy (IAG) to provide GPS data and highly accurate ephemerides in a timely fashion to the global science community to aid in geophysical research. This service has been operational since January 1994. The GPS data flows from a global network of permanent GPS tracking sites through a hierarchy of data centers before they are available to the user at designated global and regional data centers (Noll, 1998). A majority of these data flow from the receiver to global data centers within 24 hours of the end of the observation day. Common data formats and compression software are utilized throughout the data flow to facilitate efficient data transfer. IGS analysis centers retrieve these data daily to produce IGS products (e.g., orbits, clock corrections, Earth rotation parameters, and station positions). These products are then forwarded to the global data centers by the analysts for access by the IGS Analysis Coordinator, for generation of the final IGS orbit product, and for access by the user community in general. The IGS, its data flow, and the archival and distribution at one of its data centers will be discussed.

2. Current Status of Data Centers and Data Flow

The flow of IGS data (including both GPS data and derived products) as well as general information can be divided into several levels (Gurtner and Neilan, 1995) as shown in Figure 1:

- Tracking Stations
- Data Centers (operational, regional, and global)
- Analysis Centers
- Analysis Center Coordinator
- Central Bureau (including the Central Bureau Information System, CBIS)

The components of the IGS dealing with flow of data and products will be discussed in more detail below.

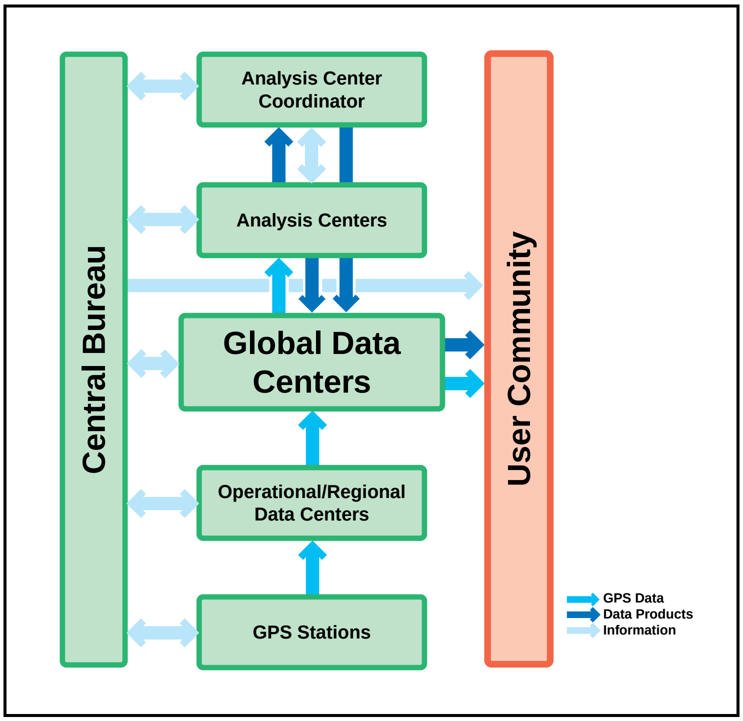

Figure 1. Flow of IGS Data and Products

2.1 Tracking Stations

The global network of GPS tracking stations are equipped with precision, dual-frequency, P-code receivers operating at a thirty-second sampling rate. The IGS currently supports nearly 200 globally distributed stations. These stations are continuously tracking and are accessible through phone lines, network, or satellite connections thus permitting rapid, automated download of data on a daily basis. The IGS has established a hierarchy of these 200 sites since not all sites are utilized by every analysis center (Gurtner and Neilan, 1995). A core set of over eighty sites are analyzed on a daily basis by most centers; these sites are called global sites. Sites used by one or two analysis centers for densification on a regional basis are termed regional sites. Finally, sites part of highly dense networks, such as one established in southern California to monitor earthquake deformation, are termed local sites. This classification of IGS sites determines how far in the data center hierarchy the data are archived. For example, global sites should flow to the global data center level, where regional sites are typically archived at a regional data center only.

2.2 Data Centers

During the IGS design phases, it was realized that a distributed data flow and archive scheme would be vital to the success of the service. Thus, the IGS has established a hierarchy of data centers to distribute data from the network of tracking stations: operational, regional, and global data centers. Operational data centers (ODCs) are responsible for the direct interface to the GPS receiver, connecting to the remote site daily and downloading and archiving the raw receiver data. The quality of these data are validated by checking the number of observations, number of observed satellites, date and time of the first and last record in the file. The data are then translated from raw receiver format to a common format and compressed. Both the observation and navigation files (and meteorological data, if available) are then transmitted to a regional or global data center within a few hours following the end of the observation day.

Regional data centers (RDCs) gather data from various operational data centers and maintain an archive for users interested in stations of a particular region. IGS regional data centers have been established in several areas, including Europe and Australia.

The IGS global data centers (GDCs) are ideally the principle GPS data source for the IGS analysis centers and the general user community. GDCs are tasked to provide an on-line archive of at least 100 days of GPS data in the common data format, including, at a minimum, the data from all global IGS sites. The GDCs are also required to provide an on-line archive of derived products, generated by the IGS analysis centers and associate analysis centers. These data centers equalize holdings of global sites and derived products on a daily basis (at minimum). The three GDCs provide the IGS with a level of redundancy, thus preventing a single point of failure should a data center become unavailable. Users can continue to reliably access data on a daily basis from one of the other two data centers. Furthermore, three centers reduce the network traffic that could occur to a single geographical location. Table 1 lists the data centers currently supporting the IGS.

2.3. Data and Product Holdings

All data centers archive data from the network in daily files by station and in RINEX format. These data consist of separate files of observation, broadcast ephemeris, and meteorological measurements. The current IGS network consists of nearly 200 sites; the number of sites archived at each GDC per day varies, depending upon each center's sponsor and user community obligations as well as hardware capabilities. In 1997, Hatanaka compression software was introduced to the IGS data centers and by 1998 became the operational method for storing and exchanging data within the IGS. This software, when combined with standard UNIX compression, reduces the size of the original daily RINEX file by at least a factor of eight and has thus reduced network traffic and time of download for the IGS community.

Table 1. Data Centers Supporting the IGS

Operational Data Centers	
ASI	Italian Space Agency
AUSLIG	Australian Land Information Group
CNES	Centre National d'Etudes Spatiales, France
DSN	Deep Space Network, USA
DUT	Delft University of Technology, The Netherlands
ESOC	European Space Agency (ESA) Space Operations Center, Germany
GFZ	GeoForschungsZentrum Germany
GSI	Geographical Survey Institute, Japan
ISR	Institute for Space Research, Austria
JPL	Jet Propulsion Laboratory, USA
KAO	Korean Astronomical Observatory
NIMA	National Image and Mapping Agency (formerly DMA), USA
NMA	Norwegian Mapping Authority
NOAA	National Oceanic and Atmospheric Administration, USA
NRCan	Natural Resources Canada
RDAAC	Russian Data Analysis and Archive Center
SIO	Scripps Institution of Oceanography, USA
Regional Data Centers	
AUSLIG	Australian Land Information Group
BKG	Bundesamt für Kartographic und Geodäsie, Germany
JPL	Jet Propulsion Laboratory, USA
NOAA/GODC	National Oceanic and Atmospheric Administration, USA
NRCan	Natural Resources Canada
Global Data Centers	
CDDIS	Crustal Dynamics Data Information System, NASA GSFC, USA
IGN	Institut Géographique National, France
SIO	Scripps Institution of Oceanography, USA

Due to the increased interest by the IGS user community for near real-time GPS data, hourly data files were introduced at the CDDIS in mid-1998. Currently, over thirty sites are providing hourly data files, with an approximate 15 minute delay, to the CDDIS. The data are retained in the individual hourly files for three days at which time they are deleted. At this time, no quality control checks are performed on the hourly files. The daily files containing data from the full 24-hour period are supplied by the data sources in the usual fashion.

Since the start of the IGS, the ACs have generated precise orbit files in SP3 format. These files are archived at all GDCs. Since GPS week 0723 (November 1993), the IGS Analysis Coordinator has provided the official IGS orbit to the GDCs and the CBIS. This orbit is typically available ten days after then end of the GPS week. A predicted (since March 1997) and rapid (since March 1996) orbit product are also generated by the Analysis Coordinator and available from the GDCs and the CBIS.

In January 1997, GFZ began supplying the GDCs with a combined troposphere product consisting of weekly zenith path delay (ZPD) estimates from the individual ACs. The product consists of weekly ZPD files, at a sampling rate of two hours, for about 100 globally distributed sites.

As of June, 1998, several IGS Analysis Centers are supplying daily, global ionosphere maps of total electron content (TEC) in the form of IONEX files to the GDCs. A daily IONEX file includes twelve two-hour snapshots of the TEC and optionally corresponding RMS information.

3. Current Data Flow and Timeliness of Data Delivery

Table 2 shows the current flow of IGS sites from station to global data center. The information in this table is based on data archived at the CDDIS since July 1998 (host cddisa.gsfc.nasa.gov). Only sites with logs available at the CBIS are reflected in this list. The table shows that ~45% of the data are delivered within one hour, 65% within three hours, and 75% within six hours.

4. Data Center Issues

4.1. QC of Current and Historical Data

Global Data Centers need to establish common quality control mechanisms to insure common quality data sets among themselves. To do this we need to define what procedures data centers must apply to the data before the data becomes public on their archive. The following describes what SOPAC has implemented (or implementing) in their archiving of RINEX data.

1) equivalent file sizes on remote server and local server after FTP transfers
2) strict compliance with the current RINEX format (RINEX vN)
3) header information checking against known site info (IGS log file) or data

NOTE: We already assume that the site has been established as an IGS site, i.e., a log file has been submitted to the CBIS and a DOMES number has been assigned.

These checks are necessary for data being collected hourly/daily, as well as for historical data.

The implementation of these for current data collection is straightforward. Files size checking (1) can be done by interrogating the remote server holding the data, or by the data file metadata being published and then checked (as proposed in the Seamless Archive implementation) by the center making the data originally available. Strict compliance with the RINEX vN data format (2) can be easily made by using UNAVCO's teqc utilities or any home grown RINEX file checker that is RINEX vN aware. Header information checking (3) is a little more tedious but necessary to maintain an archive with data consistent with its metadata. Header information must be compared against known site information. If discrepancies exist, archive operators would be informed.

In the case of a valid site change, log files should be published (by site operating agency) to note the change. If the discrepancy indicates that the site info is not valid, then it is the

Table 2. IGS Data Flow and Median Delay (by Station)

Station		Site Name	OC/LDC	RDC	GDC	Delay	No. Days
ALBH	*	Albert Head		NRCan	CDDIS	1	98
ALGO	*	Algonquin		NRCan	CDDIS	1	98
AMCT		Colorado Springs, CO		NRCan	CDDIS	1	96
ANKR	*	Ankara		BKG	IGN	34	69
AOA1		Westlake, CA		JPL	CDDIS	1	98
AOML		Key Biscayne, FL		NOAA	CDDIS	1	97
AREQ	*	Arequipa		JPL	CDDIS	6	95
ASC1	*	Ascension Island		JPL	CDDIS	1	98
AUCK	*	Auckland		JPL	CDDIS	1	97
AZU1		Azusa, CA		JPL	CDDIS	1	86
BAHR	*	Bahrain	NIMA		CDDIS	3	91
BAKO		Bakosurnatal			SIO	4	90
BARB	*	Barbados		NOAA	CDDIS	1	42
BARH		Bar Harbor, ME		NOAA	CDDIS	1	3
BILL		Temecula, CA			SIO		
BLYT		Blythe, CA			SIO		
BOGT	*	Bogota		JPL	CDDIS	2	94
BOR1		Borwiec	ISR	BKG	IGN	3	97
BRAN		Burbank, CA			SIO		
BRAZ		Brasilia		JPL	CDDIS	0	0
BRMU	*	Bermuda		NOAA	CDDIS	1	94
BRUS		Brussels		BKG	IGN	2	96
CAGL		Cagliari	ASI	BKG			
CARR		Parkfield, CA		JPL	CDDIS	1	98
CAS1	*	Casey		AUSLIG	CDDIS	7	77
CASA		Mammoth Lakes, CA		JPL	CDDIS	1	98
CAT1		Catalina Island, CA		JPL	CDDIS	1	98
CHAT	*	Chatham Island		JPL	CDDIS	9	94
CHIL		Chilao, CA			SIO		
CHUR		Churchill		NRCan	CDDIS	1	95
CICE		Ensenada		JPL	CDDIS	1	98
CIT1		Pasadena, CA		JPL	CDDIS	1	98
CLAR		Claremont, CA			SIO		
CMP9		Sylmar, CA			SIO		
COCO	*	Cocos Island		AUSLIG	CDDIS	8	80
COSO		Ridgecrest, CA			SIO		
CRFP		Yucaipa, CA			SIO		
CRO1	*	St. Croix		JPL	CDDIS	1	96
CSN1		Northridge, CA		JPL	CDDIS	13	30
DAM1		Sylmar, CA			SIO		
DAM2		Sylmar, CA			SIO		
DAV1	*	Davis		AUSLIG	CDDIS	4	95
DGAR	*	Diego Garcia		JPL	CDDIS	7	90
DHLG		Durmid Hill, CA			SIO		
DRAO	*	Penticton		NRCan	CDDIS	1	97
DUBO		Lac du Bonnet		NRCan	CDDIS	1	98
EBRE		L'Ebre		BKG	IGN	10	86
EISL	*	Easter Island		JPL	CDDIS	1	98
EPRT		Eastport, MD		NOAA	CDDIS	1	5
FAIR	*	Fairbanks, AK		JPL	CDDIS	1	91
FLIN		Flin Flon		NRCan	CDDIS	1	78
FORT	*	Fortaleza		NOAA	CDDIS	3	90
GALA	*	Galapagos Island		JPL	CDDIS	36	89
GLSV		Kiev			IGN	5	88
GODE		Greenbelt, MD		JPL	CDDIS	1	98
GOL2	*	Goldstone, CA	DSN	JPL	CDDIS	1	93
GOLD		Goldstone, CA	DSN	JPL	CDDIS	12	97
GOPE		Ondrejov	ISR	BKG			
GOUG		Gough Island	AWI		CDDIS	57	49
GRAS		Grasse	CNES		IGN	5	89
GRAZ		Graz	ISR	BKG	IGN	3	97
GUAM	*	Guam		JPL	CDDIS	1	98
HARK	*	Pretoria	CNES		IGN	5	93
HARV		Harvest Platform		JPL	CDDIS	4	92
HERS		Herstmonceux		BKG	IGN	2	95
HFLK		Innsbruck	ISR	BKG			
HNPT		Cambridge, MD		NOAA	CDDIS	2	97
HOB2	*	Hobart		AUSLIG	CDDIS	7	88
HOFN		Hofn		BKG	IGN	24	93
HOLB		Holberg		NRCan	CDDIS	1	98
HOLC		Pearblossom			SIO		
HOLP		Hollydale			SIO		
HRAO	*	Hartebeesthoek		JPL	CDDIS	1	83
IAVH		Rabat		JPL	CDDIS	86	58
IISC	*	Bangalore		JPL	CDDIS	2	98
IRKT	*	Irkutsk	RDAAC | DUT	(none) | BKG	CDDIS | IGN	21	95
JOZE		Jozefoslaw	ISR	BKG	IGN	11	32
JPLF		Pasadena, CA		JPL	CDDIS	1	95
JPLM		Pasadena, CA		JPL	CDDIS	1	98
KELY	*	Kellyville		NOAA	CDDIS	1	98
KERG	*	Kerguelen	CNES		IGN	5	90
KIRU		Kiruna	ESOC	(none) | BKG	CDDIS | IGN	3	94
KIT3	*	Kitab	GFZ	(none) | BKG	CDDIS | IGN	10	97
KOKB	*	Kokee Park, HI		JPL	CDDIS	1	90
KOSG	*	Kootwijk	DUT	BKG	IGN	2	98
KOUR	*	Kourou	ESOC		CDDIS	2	95
KSTU	*	Krasnoyarsk	GFZ	(none) | BKG	CDDIS | IGN	60	95
KUNM		Kunming		JPL	CDDIS	0	85
KWJ1	*	Kwajalein		JPL	CDDIS	1	98
LAMA		Olsztyn	ISR	BKG			
LBCH		Long Beach, CA		JPL	CDDIS	1	98
LEEP		Hollywood, CA			SIO		
LHAS	*	Lhasa		BKG	IGN	3	97
LONG		Irwindale, CA			SIO		
LPGS	*	La Plata	GFZ		CDDIS	22	96
MAC1	*	Macquarie Island		AUSLIG	CDDIS	4	94
MAD2	*	Madrid	DSN	JPL	CDDIS	1	97
MADR		Madrid	DSN	JPL	CDDIS	12	93
MAG0	*	Magadan	RDAAC | (none)		CDDIS | SIO	23	70
MALI	*	Malindi	ESOC		CDDIS	2	94
MAS1	*	Maspalomas	ESOC	(none) | BKG	CDDIS | IGN	1	97
MATE	*	Matera	ASI	BKG	IGN	4.5	95
MATH		Lake Mathews			SIO		
MCM4	*	McMurdo		JPL	CDDIS	1	98
MDO1	*	McDonald		JPL	CDDIS	1	98
MDVO		Mendeleevo	DUT	BKG	IGN	24	96
MEDI		Medicina	ASI	BKG			
METS	*	Metsahovi	NMA	BKG	IGN	8	82
MKEA	*	Mauna Kea, HI		JPL	CDDIS	1	98
MOIN		Limon		JPL	CDDIS	0	0
MONP		Monument Peak			SIO	4	97
NANO		Nanoose		NRCan	CDDIS	1	98
NICO		Nicosia		BKG	IGN	21	93
NLIB	*	North Liberty, IO		JPL	CDDIS	1	98
NOTO		Noto	ASI	BKG			
NOUM		Noumea			IGN	2	93
NRC1	*	Ottawa		NRCan	CDDIS	1	97
NSSP		Yerevan		JPL	CDDIS	35.5	89
NTUS	*	Singapore			IGN	3	80
NYA1		Ny Alesund		BKG	IGN	2	86
NYAL	*	Ny Alesund		BKG	IGN	2	92
OAT2		Oat Mountain, CA		JPL	CDDIS	1	98
OBER		Oberpfaffenhofen	GFZ	(none) | BKG	CDDIS | IGN	4	93
OHIG	*	O'Higgins		BKG	IGN	2	88
ONSA	*	Onsala		BKG	IGN	8	89
PENC		Penc		BKG			
PERT	*	Perth	ESOC		CDDIS	3	98
PETR	*	Petropavlovsk-Kamchatka	RDAAC | (none)		CDDIS | SIO	7	94
PIE1	*	Pie Town, NM		JPL	CDDIS	1	98
PIN1		Pinyon Flat, CA			SIO	4	97
PIN2		Pinyon Flat, CA			SIO		
POL2	*	Bishkek		JPL	CDDIS	22	93
POTS	*	Potsdam	GFZ	(none) | BKG	CDDIS | IGN	4	95
PRDS		Priddis		NRCan	CDDIS	1	97
PVEP		Palos Verdes, CA			SIO	4	97
QUIN		Quincy, CA		JPL	CDDIS	1	94
RAMO		Mitzpe Ramon			SIO	54	6
RCM6	*	Richmond, FL		NOAA	CDDIS	1	72
REYK	*	Reykjavik		BKG	IGN	3	69
ROCH		Pinemeadow, CA			SIO		
ROCK		Simi Valley, CA			SIO		
SANT	*	Santiago		JPL	CDDIS	1	97
SCH2		Schefferville		NRCan	CDDIS	1	98
SCIP		San Clemente Island, CA			SIO		
SELE		Almaty		JPL	CDDIS	88	86
SEY1		Seychelles		JPL	CDDIS	0	0
SFER		San Fernando		BKG			
SHAO	*	Shanghai		JPL	CDDIS	21	95
SIO3		Scripps			SIO	4	97
SNI1		San Nicholas Island		JPL	CDDIS	23	37
SOFI		Sofia		BKG			
SOL1		Solomons Island, MD		NOAA	CDDIS	1	98
SPK1		Saddle Peak, CA		JPL	CDDIS	1	98
STJO	*	St. John's		NRCan	CDDIS	1	98
SUTH	*	Sutherland		JPL	CDDIS	1	83
SUWN		Suwon-shi	NGI		CDDIS	1	96
TABL		Wrightwood, CA			SIO		
TAEJ	*	Taejon	KAO		CDDIS	1	96
TAIW		Taipei	GSI		CDDIS	0	0
THTI	*	Papeete	CNES		IGN	29	42
THU1	*	Thule		JPL	CDDIS	1	98
TID2	*	Tidbinbilla	DSN	JPL	CDDIS	1	98
TIDB	*	Tidbinbilla	DSN	JPL	CDDIS	12	95
TORP		Torrance			SIO		
TOUL		Toulouse	CNES		IGN		
TRAK		Irvine			SIO		
TRO1		Tromso	NMA	BKG	IGN	2	93
TROM	*	Tromso	NMA	BKG	IGN	2	60
TSKB	*	Tskuba	GSI		CDDIS	2	98
UCLP		Los Angeles, CA		JPL	CDDIS	1	80
UCLU		Uculet		NRCan	CDDIS	1	98
UPAD		Padova	ASI	BKG			
USC1		Los Angeles, CA		JPL	CDDIS	1	98
USNA		Annapolis, MD		NOAA	CDDIS	2	98
USNO	*	Washington, D.C.		NOAA	CDDIS	1	98
USUD	*	Usuda		JPL	CDDIS	1	98
VESL		Sanae	AWI		CDDIS	57	53
VILL	*	Villafranca	ESOC	(none) | BKG	CDDIS | IGN	3	97
VNDP		Vandenberg, CA			SIO	4	88
WES2	*	Westford, MA		NOAA	CDDIS	1	97
WHC1		Whittier, CA		JPL	CDDIS	1	98
WHIT	*	Whitehorse		NRCan	CDDIS	1	98
WIDC		Sky Valley, CA			SIO		
WILL		Williams Lake		NRCan	CDDIS	1	98
WLSN		Mt. Wilson, CA		JPL	CDDIS	9	95
WSLR		Whistler		NRCan	CDDIS	2	98
WSRT	*	Westerbork		BKG	IGN	2	97
WTZR	*	Wettzell		BKG	IGN	2	97
WTZT		Wettzell		BKG			
WUHN	*	Wuhan		NOAA	CDDIS	1	96
XIAN	*	Xi'an		JPL	CDDIS	16	82
YAKA		Yakutsk	RDAAC | (none)		CDDIS | SIO	7	96
YAKZ	*	Yakutsk	RDAAC | (none)		CDDIS | SIO	8	92
YAR1	*	Yaragadee		JPL	CDDIS	1	96
YELL	*	Yellowknife		NRCan	CDDIS	1	98
ZECK		Zelenchukskaya			IGN	41	91
ZIMM		Zimmerwald		BKG	IGN	2	96
ZWEN	*	Zveningorod	GFZ	(none) | BKG	CDDIS | IGN	22	92

Totals: 198 stations, 85 global stations, 162 archived at CDDIS

Notes: * indicates global stations
| notation indicates duplicate flow of data
Delay column indicates median hourly delay after end of UTC day to CDDIS
No. Days column indicates number of days reflected in statistics (01-Jul-98 through 06-Oct-98)
0 indicates no data received for time period
blank indicates data not archived at CDDIS

GDC's responsibility to attempt to get the operating agency of the site to correct the problem. If the problem cannot be remedied by the site operating center -either the data no longer exists or resources are unavailable - then it is the responsibility of the GDC to correct the problem in the data/metadata. This should only be done as a last resort to insure correct and quality data.

The re-checking of historical RINEX data has brought attention to file size differences, RINEX format non-compliance and incorrect header information. The same checks to current data are applied to historical data. Files size checking (1) is by far the most difficult task. If the data that was originally provided to the GDC are on-line, the check is as simple as comparing remote and local file sizes. If the data are off-line, the remote archive would be asked to make available to the GDC an archive index list; this list would contain file statistics for all off-line data. This list can then be used to do the file size comparisons. RINEX vN (2) compliance is the same as current data collection – pass the data through a RINEX format checker. Header checking is done the same for current historical data as for current data.

What does the GDC do when it finds corrupt data, which cannot be replaced by the site's operating agency? It is then the GDC's responsibility to correct the problems in the data, make comments of the changes in the header, and re-publish the data. The GDC's must be ultimately responsible for the quality of the data in their archive.

4.2. Seamless Archive for Data Discovery

Storing GPS data at distributed archives provides users with benefits such as speedier local access, and regional or research-specific data support, but it can also increase confusion when supposedly identical files contain different header information or file content, regardless of how slight the differences. These problems will increase along with increases in regional networks. A cooperative arrangement which improves coordination of data holdings between GPS archives, combined with a basic but similar user interface, would provide data users with easier access to GPS data and metadata by allowing them to simply contact one center, instead of contacting all of them separately. We call this concept of an interoperable multi-archive system the "GPS Seamless Archive Center" or GSAC.

The strategy for a seamless GPS archive was developed by participants from ten archive centers at a workshop sponsored by the UNAVCO Facility on 11-12 November 1997 (Table 3). A summary of this workshop can be found at URL *http://www.unavco.ucar.edu/community/events/meetings/*. The primary output product from the workshop is the definition of tables used to identify specific data holdings, and mechanisms to access data from any data center in a standard manner. Although not specifically focused on the IGS data, we propose that identical methods would provide uniform access to any of the Operational, Local, Regional or Global Data Centers, and help resolve data identity, delivery and time-delay issues. Each participant in the GPS Seamless Archive Centers will maintain their individuality and bring their own strengths into play, yet provide the user community with a familiar and consistent data access look-and-feel to all archive holdings by providing standardized data access. The workshop concluded that a process and necessary software tools must be developed to define where data might exist (monument location tables), and for what time ranges. A summary description and some software tools can be found at the URL *http://www.unavco.ucar.edu/data/gsac/gsac.html*. Having this information in distributable tables means that historical information is as readily available as recent information – if a file changes, for example, due to discovery of an incomplete download which is later corrected, then the table can be updated when the file is replaced. The process of identifying data existence, data file holdings, and changes would be standardized. Regardless of the size of the data provider, identical tools and processes would be used, improving data quality and usability.

Table 3. GPS Seamless Archive Participants

GPS Seamless Archive Center (GSAC)	GSAC Metadata via FTP Access (as of Aug'98)
CDDIS (NASA GSFC)	ftp://cddisa.gsfc.nasa.gov/pub/GSAC
IRIS/DMC	N/A
JPL	
NCEDC/BARD (Berkley/Stanford/ USGS)	ftp://quake.geo.berkeley.edu/pub/GSAC
NGS/NOAA	
PANGA (Cent.Wash. Univ.)	
PANGA (Univ. Washington)	
PGC/WCDA	ftp://sikanni.pgc.nrcan.gc.ca/pub/GSAC
SCEC	ftp://ramsden.ucsd.edu/pub/GSAC
SOPAC	
UNAVCO	ftp://ftp.archive.unavco.ucar.edu/pub/GSAC
Univ. Texas-Austin	N/A

These tools can be operated by anyone wishing to participate in the GSAC process. GSAC specifications are used to create specific tables showing data holdings for each archive. The tables are stored in a well-defined manner to simplify information exchange. Some GSAC's will advertise their own data, but for data-user's convenience, other GSAC's will collect these "advertisements" and provide a common interface to search and retrieve data from any of the GSAC data areas. A data requester can go to any GSAC, search the holdings-tables of all GSAC archives, and make a data request to the archive with the desired data. An overview of the GSAC concept is given in Figure 2. A follow-on GSAC meeting will be held in early 1999 to define data request mechanisms that will allow users to obtain data from any GSAC, regardless of where the data may physically reside.

5. Future Directions

5.1. Y2K

How will the new millennium affect data archives? Issues to think about are: file directory structure, file naming, and data time stamping. There needs to be a way of

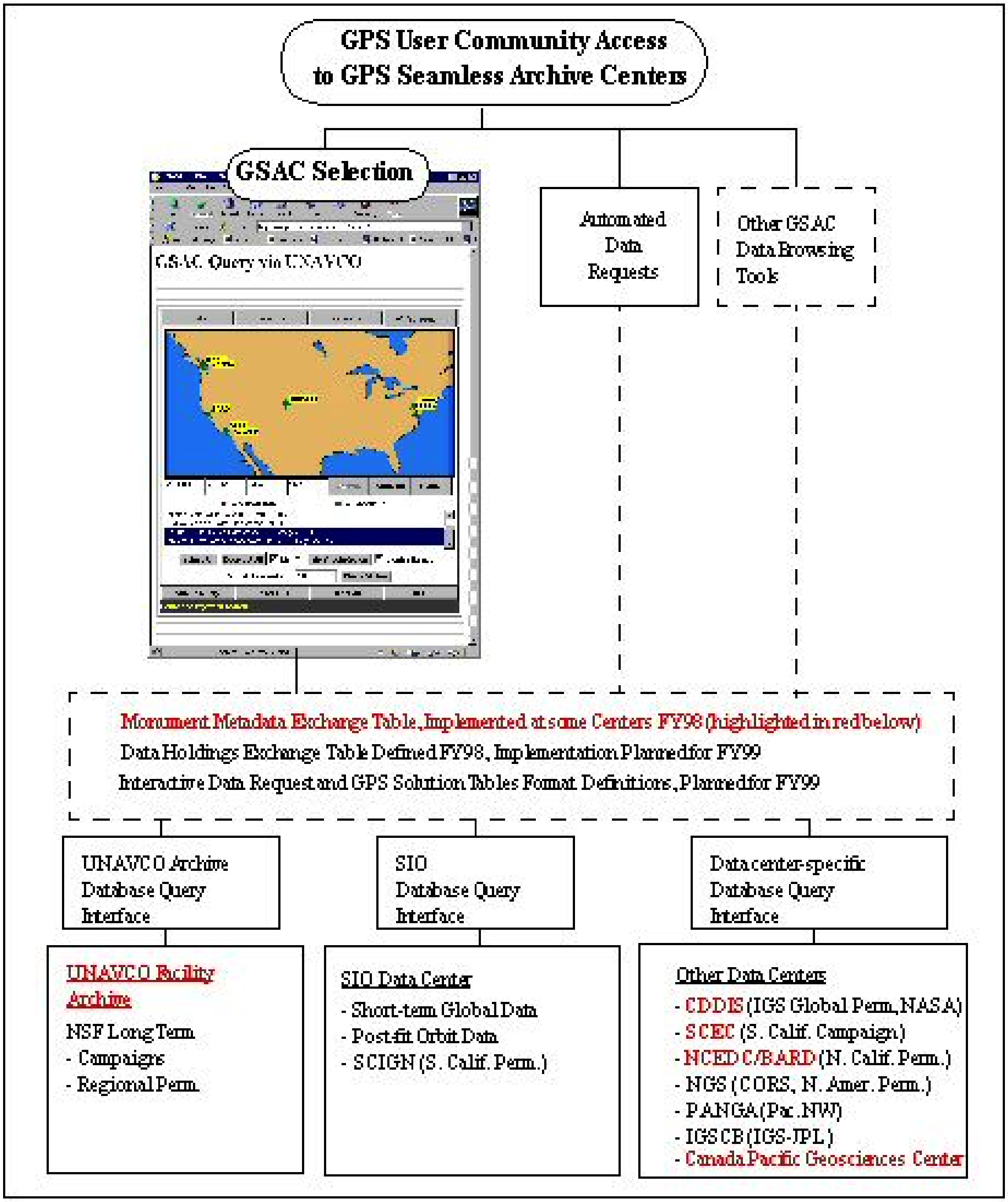

Figure 2. GPS Seamless Archive Center overview. Red indicates monument metadata exchange tables have been implemented

uniquely identifying data by date. The filename doesn't do this (we are limited to a two-digit year). Timestamps in the RINEX data are two-digit year. File directory structures between archives are dissimilar and usually do not have four-digit years.

If we address each issue individually we will find that modifying the file name and the timestamps is unnecessary. We must only address the file directory structure on each GDC. If the file directory tree contains data broken down by a four-digit year, then we have a means of differentiating truly by year. The data file naming convention becomes inconsequential. The timestamp of the data records in the RINEX file will cause no problem due to the required header field "TIME OF FIRST OBS" which is Y2K compliant. The change to Y2K compliant directory structures is simple, easy to implement, and will ultimately lead to less confusion in data storage. The same approach can apply to products and the GPS week 1000 issues, i.e., by making sure products data are stored by GPS weeks using four digits and not three.

5.2. Data Types

GDCs should archive a common set of data types. The list of data types needs to be established. This common list by no means limits the GDC from archiving other types, but rather, provides a minimal framework. This will lessen confusion among users when navigating multiple GDCs holdings.

5.3. Data Flow vs. Data Push

There is great concern in GDCs modifying the content of data distributed or collected from RDCs or ODCs. This can cause GDCs to have different data sets. In addition, it is difficult to notify the GDCs of the re-publishing of data, and then the re-distribution from GDC to GDC. To remedy this situation a defined data flow and method of archive population needs to be established.

It has been suggested that each RDC or ODC push data to one GDC. The GDC would then push the data to the remaining GDCs. This would create a well-defined path for data flow. In the case of corrupt data or metadata, the RDA would be asked to resubmit the data to the GDC, who would then push the re-submitted data to the remaining GDCs. In the event that the RDC or ODC can not correct the problem, the GDC responsible for that data (defined by the data flow) would correct the problem and submit the data to the other GDCs. This defined flow negates the need for archive interrogation for data existence and re-submissions, handles QC issues and data re-publishing efficiently, and can be used to help troubleshoot latency problems. This method will also insure that the same data set appears on each GDC. In addition, the GDCs should adopt Seamless Archive practices of insuring that data between the archives are equivalent (using md5 and file size checking).

5.4 Data centers specs and quality control

5.4.1 Specifications

The components of the IGS are diversely defined in their role and duties, the level of precision is not homogeneous. More precise and detailed specifications are needed for data centers, presently only little more than the role of each type of data center is defined. For each level of data center, a clear and comprehensive list of items to comply with is needed. The users and applications of the IGS are growing at a fast pace, the IGS service itself must evolve to reflect this. In order to be in phase with this and to be able to anticipate, a continuous process of reviewing, refining and modifying the specs should be established. Input should be searched from all parts of the IGS (data and analysis centers, users, etc.), even complaints should be encouraged. Analysis centers for example should be asked to provide a list of things they expect from data centers (whether or not they actually get them).

A preliminary list of the items that should be documented:

1) set of data and formats archive at data center
    - list of stations
    - list of products
    - list of other data
2) reporting
    - what should be reported, how (mail, web, ftp), time period, report layout, normalization
    - list of technical characteristics that should be published (extend the data center description form)
3) operational constraints
    - delays
    - number of concurrent accesses
    - Internet link rate and quality
    - quality checks

5.4.2 Quality Control

There are already some quality checks running about many things in the IGS, but as far as data centers are concerned there are no precise and routine evaluation procedures. Each data center has its own set of internal procedures. There are some raw checks done by the IGS CB but this is not enough; in order to get a clear view of what is going on and what should be improved or modified, the IGS needs to define and run a new set of evaluation procedures.

Each center has its own strong points and weak points, a monthly or weekly evaluation procedure should help improve the overall quality of service to users and to other components of the IGS. This should not be oriented as a good guy/bad guy discrimination process but as an objective and reliable reference for each center to help them refine their contribution.

5.5. Network

The IGS network should be considered as a component of the IGS itself. There should be a clear role in the IGS organization taking in charge the set up and evolution of the network. As is more and more the case in computer technology, the network itself is a major component of the overall structure. All centers participating in the IGS have a little part in the network but no one is in a place to take this in charge, even the global data centers. This is becoming more and more important as near real time activities are emerging. Some points in which more control and information are needed :

- data flow paths, minimizing duplicate paths, optimizing delays (as when a new station is integrated, there is a need to define the path and enforce it)
- backup paths and procedures (they are very rarely tested, in part because the network is reliable but the worst can always happen)
- assessing network performances and limits (also in a proactive way, similar to what is done in classic LAN network administration)
- users should be included in the picture. There is no point in having the best network for IGS internal management and not for users (e.g., a fair amount of users in Europe seems to access the service through a U.S. global center. This fact implies that something must be wrong somewhere)
- raise a flag when something is wrong (e.g., an unusually low data rate between selected centers, sub-optimized data path, etc.)
- plan for future experiences and evolutions (can the network deal with future plans, if not what should be done)

5.6. Knowing Our Community

We can only guess what is the IGS user community, we cannot get the big picture. This is too bad because we could improve the IGS by a better knowledge of our users, who they are, how many they are, what they use the IGS for, what they think of the service in general, etc.

The IGS has become involved in many fields as are its users; having more information about them is becoming a requirement to maintain and improve the service in the future.

Data centers already have a rough idea (the level of detail may vary, depending upon the techniques used for providing the data, anonymous or named ftp, for example), this can be used in a first step to establish an initial version of a user database.

In a second step, users should be encouraged to reference themselves in our user database; i.e., the CB could establish a page on the IGS web site to welcome comments and opinions. An IGS users mailing list could also be planned (maybe just think about it and see if it is feasible). This has to be coordinated with the PR activities presently done by the CB as part of its role.

6. Recommendations

- Stations should comply with established station guidelines. A recent IGS document authored by the IGS Infrastructure Committee on station guidelines will soon be available on the CBIS web site.
- Stations and data centers should review current data flow with the intent to improve timeliness.
- All stations used in routine AC analysis should deposit log files in the CBIS.
- Under the IGS "umbrella", data centers should only archive data from official IGS sites. Network and data center contacts should work with stations that submit data to ensure all required documentation is complete and available through the CBIS.
- Data centers should implement data holding verification routines to ensure equality in data holdings.
- Data centers should consider common directory structure to aid users in downloading data from multiple centers.
- SIO should further study procedures for quality-checking historic data.
- IGS data centers (at all levels) should participate in seamless archive activity.
- The various components of the IGS should establish and test backup data flow paths.
- A continuous process involving all parts of the IGS should be established to review and refine the standards of each component involved in the data flow activity. In a first step, the role and duties of the global and regional data centers have to be redefined with respect to the development of near real time activities in the IGS.
- The IGS CB should define and run, in a periodic way, a set of quality control routines to evaluate the quality of service and assess the compliance of each data center with respect to the IGS standards.
- An IGS network manager or a working group on network management within the IGS should be appointed. The main point is to be able to evaluate, plan, and adapt the network in a context of rapid changes and increasing activity. The attributions have to be clearly defined, they would encompass the whole IGS network (i.e., stations, data centers, analysis centers, and users).
- The IGS should obtain a clearer view of who our users are and what they need. Any part of the IGS involved with a user interface activity (CB, data centers, scientific meetings, etc.) contact and feedback from users should be queried. A user database, CBIS web page, and a mailing list should be set up to support this.

References

Gurtner, W. and R. Neilan. "Network Operations, Standards and Data Flow Issues" in *Proceedings of the IGS Workshop on the Densification of the ITRF through Regional GPS Networks*. JPL. 1995.

Noll, C. "Flow of GPS Data and Products for the IGS" in *Proceedings of the Workshop on Methods for Monitoring Sea Level*. JPL. 1998.