# Current Status of and Backup Plans for Flow of IGS Data and Products

Carey Noll
*Code 922, NASA GSFC, Greenbelt MD 20771, USA*

**ABSTRACT –** *The IGS has been operational for nearly seven years. Recent changes in the data and products archived at the data centers prompts the review of the current IGS data flow and archiving methodologies. This presentation will outline the current structure at the IGS data centers, including the flow of both the daily and hourly data products, and will detail ideas for improvements to the data flow to ensure timely and consistent availability of IGS data and products.*

## 1 Introduction

The International GPS Service (IGS) has been an operational service within the IAG since 1994. During the planning phases for the IGS, it was realized that a distributed data flow and archive scheme would be vital to the success of the IGS. 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. 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.

## 2 Background

Operational or local data centers (OCs or LDCs) are responsible for the direct interface to the GPS receiver, reliably and regularly connecting to the remote site, hourly or daily, and downloading and archiving the raw receiver data. The quality of these data is validated and the data are then translated from raw receiver format to a common format (RINEX), decimated if necessary, and compressed. Both the daily observation and navigation files (and meteorological data, if available) are then transmitted to a regional or global data center ideally within an hour following the end of the observation day. For hourly data, files are transmitted from the operational data center within a few minutes of the end of the hour.

Regional data centers (RDCs) gather data from various operational data centers before transmitting them to the global data centers, thus reducing traffic over electronic networks. RDCs maintain an archive for users interested in stations of a particular region. Typically, data not used for global analyses are archived and available for on-line access at the RDC level.

The IGS global data centers (GDCs) are typically the principle GPS data source for the IGS analysis centers and the user community in general. The GPS data available through the global data centers consists of observation, navigation, and meteorological files, all in

RINEX format. The GDCs provide an on-line archive of daily and sometimes hourly data files as well as the products derived from these data, generated by the IGS analysis centers, associate analysis centers, and analysis coordinators. These data centers equalize holdings of global sites and derived products on a daily basis (at minimum). The three GDCs currently supporting the service provide the IGS with a level of redundancy, thus preventing a single point of failure should a data center become unavailable. Furthermore, three centers reduce the network traffic that could occur to a single geographical location.

## 3 Current flow of data and system status

Table 1 lists the major data centers currently supporting the IGS. Table 2 lists the types of data archived by the IGS data centers and the products created by the IGS analysis centers; Table 3 shows the types of data and products supported at the global data center archives.

Figure 1 depicts the current flow of thirty-second daily data files from the OCs to the GDCs. Only sites with complete site logs at the IGS Central Bureau are reflected in this figure. Both the CDDIS and SIO archive data from sites outside the IGS "umbrella"; these sites are not listed in the figure or table. Figure 2 shows the data flow for hourly data files; at this time, hourly data are archived and distributed by the GDCs at CDDIS and IGN and the RDC at BKG.

To aid in further understanding the flow of daily and hourly data files within the IGS, it is important to study the timeliness of the data products. Table 4 presents the data latency statistics for daily data archived at the CDDIS. Over fifty percent of the data are archived within two hours of the end of the observation day. Table 5 presents the current data latency statistics for hourly data at the CDDIS. As can be seen from this table, over fifty percent of the hourly data are available for download from the CDDIS within fifteen minutes. These statistics were derived by programs developed by the IGS Central Bureau and executed on a regular basis at many of the IGS data centers; the results for the CDDIS are updated frequently during the day and can be viewed at *ftp://cddisa.gsfc.nasa.gov/pub/reports/gpsdata/check_import.cddisa* (daily data) and *ftp://cddisa.gsfc.nasa.gov/pub/reports/gpsdata/check_hourly.cddisa* (hourly data).

## 4 Backup strategies

The mission of the IGS has been broadened to include the flow and analysis of hourly data sets, thus requiring a rapid turn-around of data from the station to the user. Furthermore, this access must be reliable, 24 hours per day, seven days per week. The IGS infrastructure must establish backup data flow paths to ensure that not only daily but also hourly data sets are accessible to users on an uninterrupted basis even if a key data center is unavailable. This scheme must be extended to include hourly high-rate data sets at those data centers supporting the efforts of future LEO missions.

During the past year, the CDDIS global data center experienced two major computer failures. Unfortunately, the extended outage of the facility had dramatic effects on the IGS analysis centers and the user community in general since some sites only deposited their data at CDDIS. Furthermore, most IGS ACs and AACs send their data products to the CDDIS where they are made available to the other IGS data centers, the IGS ACs and AACs, and the user community. Not only were some data and products not immediately available, the CDDIS user community may have been unfamiliar with accessing the other data centers. Some redundant data flow paths already existed, but other sites had no data center to transmit their data to when the CDDIS was down. To prevent the unavailability of key IGS sites in the future, a strategy for the backup of the IGS data flow infrastructure must be developed and tested. This plan, once operational, can then be utilized in the case of planned system downtimes (e.g., for scheduled computer upgrades or network outages) or when unexpected failures occur.

The first step in the development of a backup plan for data (and product) flow is to identify the current network topology; this data flow is shown in Figure 1. This figure illustrates the fact that many operational and regional data centers deposit data to the CDDIS; fortunately, SIO also has access to many of these data centers and routinely downloads data from them as well for their archive. IGN mainly handles the flow of data in Europe with BKG and retrieves the remaining IGS global sites from either CDDIS or SIO.

The second step in the development of a backup plan is to identify possible alternate data flow paths should any of the three GDCs be inaccessible; the plan can then be extended to include other major IGS data centers (e.g., RDCs). It should be noted at this time that UNAVCO serves as a backup OC for JPL. During the past few months several successful tests have been conducted between UNAVCO, JPL, and the CDDIS where UNAVCO downloads data from selected JPL stations and forwards these data to the CDDIS. The CDDIS has installed procedures to peruse the UNAVCO data delivery area and archive any data not previously found in the JPL data delivery area. This process is now in place for both daily and hourly data. Each OC is designated a secondary data flow path to another data center. The proposed alternate data routes for the distribution of daily data files through the GDCs are illustrated in Figure 3; this process could be extended to the hourly data flow. For products, the current and proposed alternate data flow paths are shown in Figure 4.

In order to make the transition to another data center easier for OCs and RDCs, arrangements need to be made with each data center to minimize special instructions for depositing or retrieving data. The majority of the data handled by the CDDIS are delivered to password-protected accounts established on the computer facility for each data source. The CDDIS also retrieves data from a few data centers (e.g., AUSLIG, BKG, RDAAC, and SIO). SIO and IGN primarily retrieve data from all GPS data sources.

To ensure that data can be delivered and processed quickly in a backup situation, it is recommended that each GDC (and participating RDC) establish a general ftp area that

OCs and RDCs can utilize when they are required to post data to alternate data centers because their “prime” data center is inaccessible. For example, the directory */ftp/incoming/igsdaily* has been created at the CDDIS (host *cddisa.gsfc.nasa.gov*) to allow data centers to deposit daily GPS data via anonymous ftp. Similarly, procedures need to be established for IGS analysis centers to deposit weekly product files to alternate GDCs when their primary contact is unavailable. Automated programs could then be created at the data centers to peruse these anonymous ftp put-areas on a routine basis and process any data or products found there. These ftp areas should be secured so that files placed there can not be deleted or modified by anyone other than authorized data center personnel. A more secure method may be to allow anonymous ftp deposits through a “trusted” mode where only users from a list of pre-approved hosts could deposit files to the data center’s incoming area. Setting the protections on directories and subdirectories to prevent browsing and downloading could further enhance security with these anonymous ftp put areas by anonymous ftp users. The use of anonymous ftp to deposit data precludes the need for the data centers to establish and administer accounts for each data source and allows for easy monitoring for incoming data or product files; alternatively, a single username for depositing data could be utilized at all data centers. This method, however, would require the coordination and distribution of username and password information to a list of data center representatives but may be more secure than anonymous ftp.

These backup procedures should not only include data center instructions but user notification instructions. Furthermore, the IGS needs to determine how long the analysis community can tolerate a data center’s outage and subsequent delay in the availability of either daily or hourly data products. For example, when a data center outage occurs for more than one day (for centers involved in distribution of daily data only) or more than two hours (for centers involved in distribution of hourly data), the affected data center should send an advisory email to all data center contacts. If computer problems at the affected data center preclude a direct email, the data center contact should make arrangements with an alternate data center to post these alerts. This advisory message should include a description of the problem and an estimated time to recovery. At this time, the backup data flow procedures should begin. The data centers tasked to act in a backup capacity should notify the affected data center they are ready to proceed with alternate data flow procedures. Once all participants are notified and prepared, data should be transmitted to these alternate data centers. The data center or an alternate, advising the general user community of the outage and other sources for IGS data and products should send an email to IGSMail. Should the outage continue for several days, messages should be issued periodically to let the IGS community know the affected system’s current status and when users can expect operations may return to normal. Once the data center has resumed operation a follow-up email to data center contacts and IGSMail should be issued stating that “normal” operations can resume.

A future enhancement to this backup data flow plan would be to automate the switchover to alternate paths. This automation could be achieved when a specified data center is unreachable for a pre-determined amount of time, dependent upon the type of data to be transmitted.

Although the IGS Central Bureau Information System (CBIS) was not discussed previously in this paper, its importance in the daily operations of the IGS should not be overlooked. The CBIS provides information to both the IGS and the general user community about the service and the use of IGS orbits and data. The CBIS also maintains the various email exploders utilized to notify users and the members of the IGS infrastructure of general announcements, station, data center, and analysis center alerts, and routine reports. The IGS ftp site is mirrored at IGN, but a full web site mirror does not exist. Furthermore, no alternate email system is currently available for use in the event that the IGS ftp/web server computer or its network is inaccessible. The IGS Central Bureau should identify alternate servers for these important activities.

## 5 Recommendations

The following recommendations have been made to aid in the implementation of a viable plan for the distribution of data and products, both operationally and in a backup situation, through the IGS infrastructure.

- Ensure that data center information available through IGS Central Bureau (*ftp://igscb.jpl.nasa.gov/igscb/center/data/*) is complete and current for all IGS participants. Add backup data flow paths to this data center information.
- Establish and maintain email distribution lists for key data center contacts at the IGS Central Bureau. The use of distribution lists could be extended for all components of the IGS as has successfully been implemented within the ILRS (*http://ilrs.gsfc.nasa.gov/pointsof.html*).
- Complete a backup data flow plan for daily and hourly IGS data and IGS products. Ensure all data center (and analysis center) contacts have reviewed the plan and will implement its procedures.
- Data centers should create and monitor ftp put areas (anonymous or username/password) to facilitate flow of data and products in a backup situation.
- Schedule tests of the backup data flow paths on a regular basis to ensure their operational readiness.
- Identify backup facilities for the IGS Central Bureau Information System (CBIS) ftp and web sites. Create a backup system for the various mail exploders maintained by the CBIS, in particular IGSMail.
- Ensure the viability and commitment of all IGS data centers to the routine and backup operations of the service.

**Table 1.** Data Centers Supporting the IGS in 2000

| | | |
|---|---|---|
| **Operational/Local Data Centers** | | |
| ASI | * | Italian Space Agency |
| AUSLIG | | Australian Surveying and Land Information Group |
| AWI | | Alfred Wegener Institute for Polar and Marine Research, Germany |
| CASM | | Chinese Academy of Surveying and Mapping |
| CNES | | Centre National d'Etudes Spatiales, France |
| DGFI | | Deutsches Geodätisches ForschungsInstitut, Germany |
| 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 |
| NGI | | National Geography Institute, Korea |
| NIMA | | National Imagery and Mapping Agency, USA |
| NMA | * | Norwegian Mapping Authority |
| NOAA | * | National Oceanic and Atmospheric Administration, USA |
| NRCan | * | Natural Resources of Canada |
| PGC | * | Pacific Geoscience Center, Canada |
| RDAAC | | Regional GPS Data Acquisition and Analysis Center on Northern Eurasia, Russia |
| SIO | | Scripps Institution of Oceanography, USA |
| UNAVCO | | University NAVSTAR Consortium, USA |
| USGS | | United States Geological Survey |
| WTZR | * | Bundesamt für Kartographie und Geodäsie (Wettzell), Germany |
| **Regional Data Centers** | | |
| AUSLIG | † | Australian Surveying and Land Information Group |
| BKG | * | Bundesamt für Kartographie und Geodäsie, Germany |
| JPL | *† | Jet Propulsion Laboratory, USA |
| NOAA | | National Oceanic and Atmospheric Administration, USA |
| NRCan | | National Resources of Canada |
| RDAAC | | Regional GPS Data Acquisition and Analysis Center on Northern Eurasia, Russia |
| **Global Data Centers** | | |
| CDDIS | *† | Crustal Dynamics Data Information System, NASA GSFC, USA |
| IGN | * | Institut Géographique National, France |
| SIO | | Scripps Institution of Oceanography, USA |

Notes: * indicates data center currently transmitting and/or archiving hourly, 30-second GPS data from selected sites.
† indicates data center proposing to transmit and/or archive hourly, high-rate GPS data for LEO activities.

**Table 2.** Summary of IGS Data and Products

| | | |
|---|---|---|
| **GPS DATA** | | |
| **Daily** | Sampling: | 30 seconds |
| | Frequency: | Daily |
| | Number of stations: | ~240 at IGS CBIS |
| | Average delay: | 2 hours to GDC (average) |
| | Format/file types: | RINEX observation data (O) |
| | | RINEX observation data, Hatanaka compression (D) |
| | | RINEX meteorological data (M) |
| | | RINEX broadcast ephemeris data (N) |
| | | summary output from teqc (S) |
| **Hourly** | Sampling: | 30 seconds |
| | Frequency: | Hourly |
| | Number of stations: | ~70 regularly submitting data |
| | Average delay: | 5-15 minutes to GDC (average) |
| | Format/file types: | RINEX observation data, Hatanaka compression (D) |
| | | RINEX meteorological data (M) |
| | | RINEX broadcast ephemeris data (N) |
| **High Rate** (future) | Sampling: | ~10 seconds (TBD) |
| | Frequency: | Hourly |
| | Number of stations: | >30 planned |
| | Average delay: | TBD |
| | Format: | TBD |
| **IGS PRODUCTS** | | |
| **Orbit, clock, ERP** | Analysis Centers: | Seven |
| | Frequency: | Weekly submission of daily orbits, clocks; weekly ERP |
| | Format: | SP3 |
| | Combination: | Weekly precise (AIUB) |
| | | Daily predicted (AIUB) |
| | | Daily rapid (AIUB) |
| **Positions** | Analysis Centers: | Three GNAACs, three RNAACs |
| | Frequency: | Weekly solutions |
| | Format: | SINEX |
| | Combination: | Weekly (NRCan) |
| **Ionosphere** | Analysis Centers: | Five AACs |
| | Frequency: | Daily files |
| | Format: | IONEX |
| | Combination: | Not yet available |
| **Troposphere** | Analysis Centers: | Seven AACs (AAC solutions not archived at GDCs) |
| | Frequency: | Weekly files |
| | Format: | SINEX_TRO |
| | Combination: | Weekly (GFZ) |

**Table 3.** Global Data Center Holdings

| Data Type | CDDIS | IGN | SIO |
|---|---|---|---|
| **Data** | | | |
| Daily GPS (D format) | X | X | |
| Daily GPS (O format) | X | | X |
| Hourly GPS (30-second) | X | X | |
| Hourly GPS (high-rate) | Future | | |
| Daily GLONASS (D format) | X | X | |
| Daily GLONASS (O format) | X | | |
| **Products** | | | |
| Orbits, etc. | X | X | X |
| SINEX | X | X | X |
| Troposphere | X | X | X |
| IONEX | X | X | |

**Table 4.** Latency of Daily Data at the CDDIS (01-Feb-2000 through 30-Jun-2000)

| Delay (hours) | No. of Station Days | Percent |
|---|---|---|
| 1 | 10,488 | 43.36% |
| 2 | 3,137 | 12.97% |
| 3 | 1,911 | 7.90% |
| 4 | 1,225 | 5.06% |
| 6 | 1,191 | 4.92% |
| 8 | 635 | 2.63% |
| 9 or more | 5,601 | 23.16% |

Total number of station days: 24,188
Total number of sites 182

**Table 5.** Latency of Hourly Data at the CDDIS (24-Apr-2000 through 30-Jun-2000)

| Delay (minutes) | No. of Station Days | Percent |
|---|---|---|
| 5 | 10,714 | 13.34% |
| 10 | 16,715 | 20.82% |
| 15 | 14,907 | 18.57% |
| 20 | 12,890 | 16.05% |
| 25 | 7,971 | 9.93% |
| 30 | 2,766 | 3.44% |
| 35 | 2,295 | 2.86% |
| 40 | 2,582 | 3.22% |
| 45 or more | 9,451 | 11.77% |

Total number of station hours: 80,291
Total number of sites 64

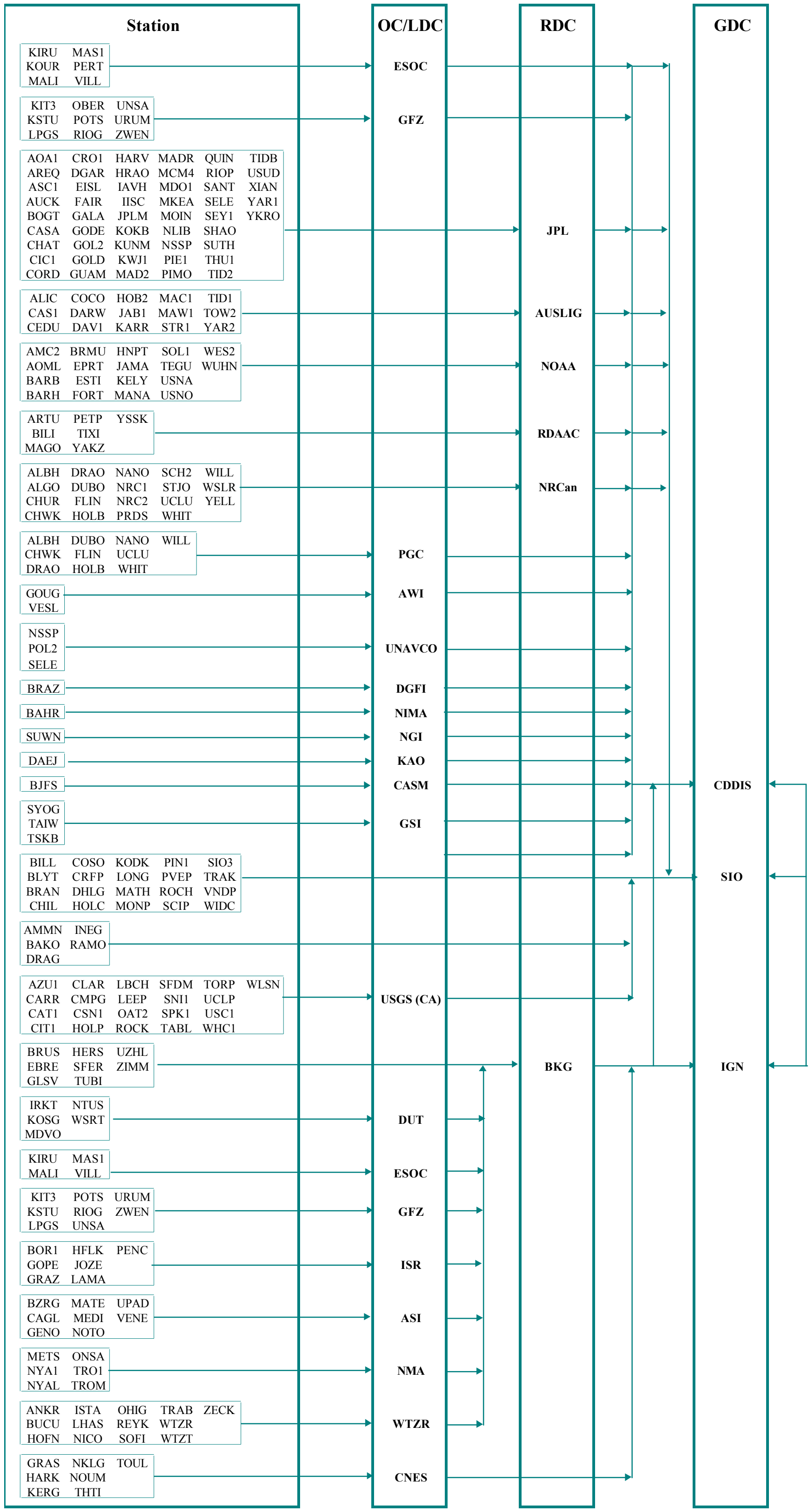
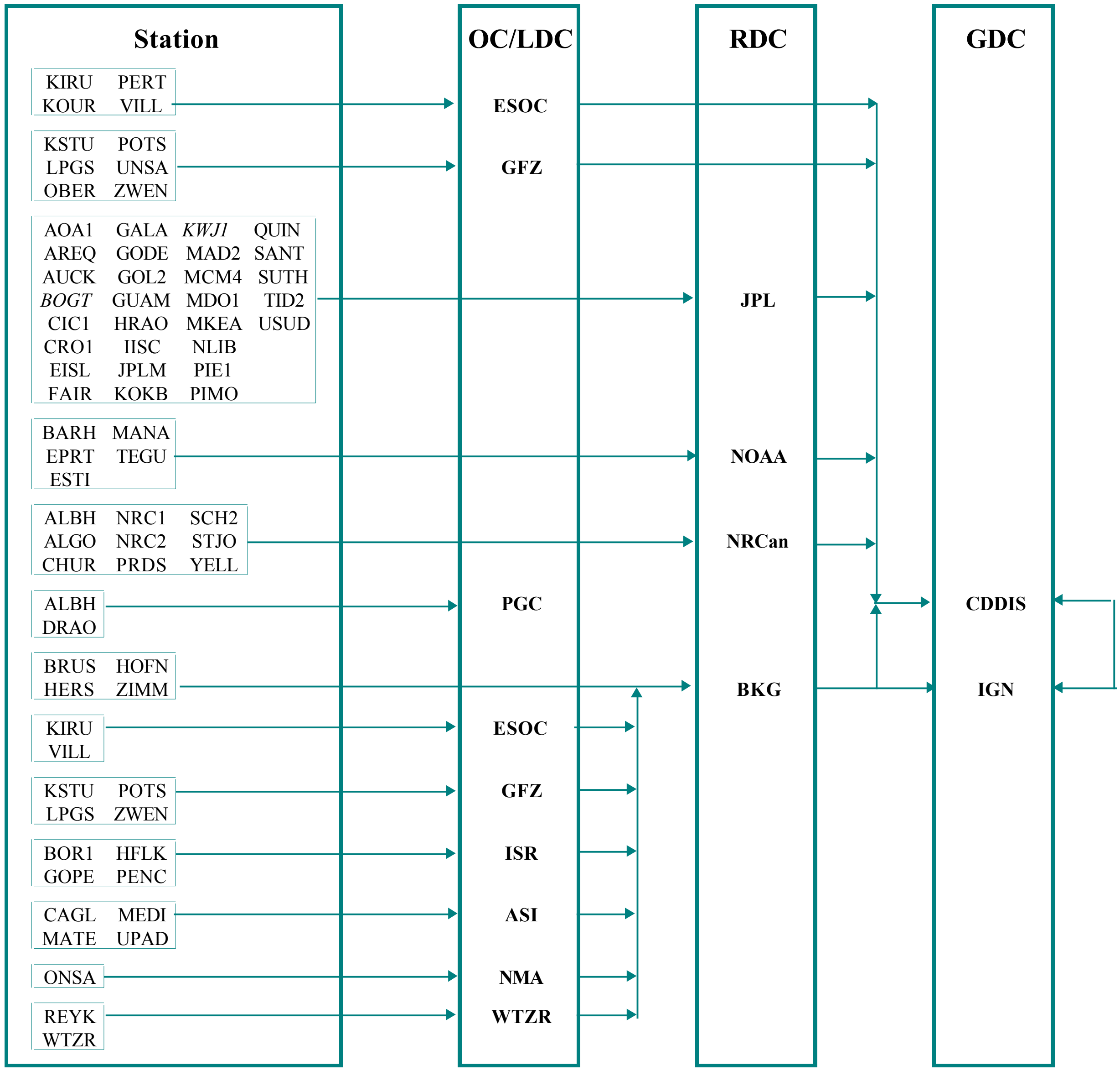



**Figure 1.** Flow of daily IGS data files (by data center)




Note: Site in *italics* indicates former hourly site

**Figure 2.** Flow of hourly IGS data files (by data center)

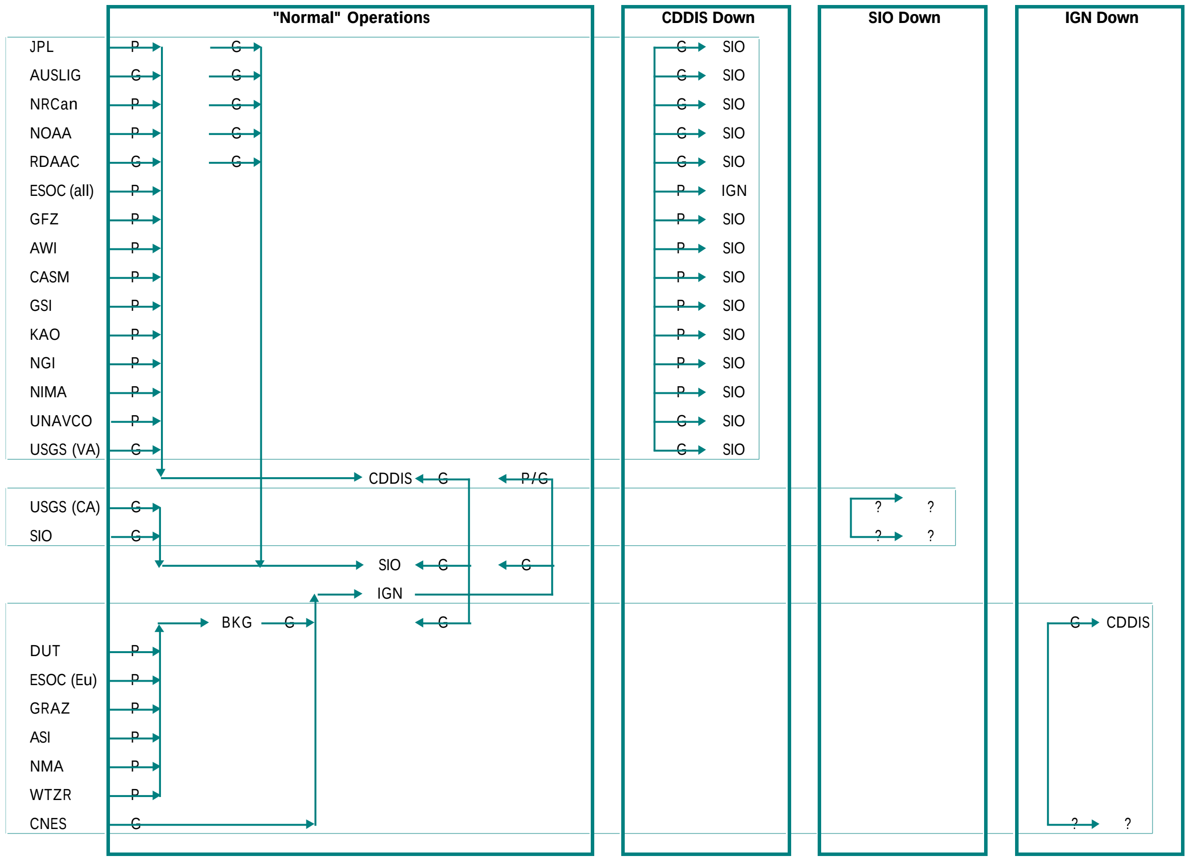



**Figure 3.** Backup flow of daily IGS data files

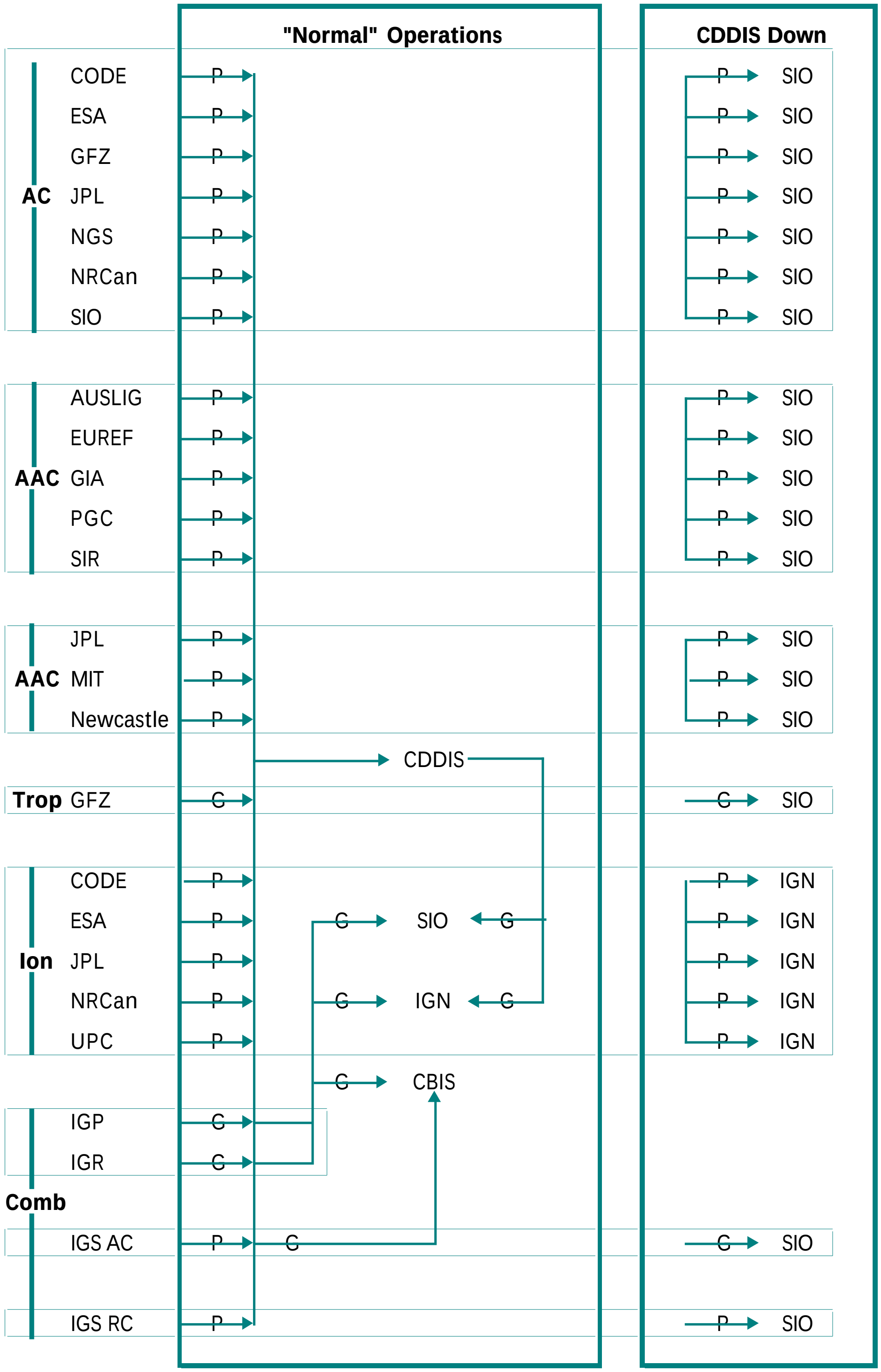



**Figure 4.** Backup Flow of IGS Product Files