IGS Data Centers: Status and Progress

Carey Noll, NASA GSFC/CDDIS
Fran Boler, UNAVCO
Heinz Habrich, BKG

IGS 20th Anniversary Workshop
Pasadena, CA
June 25, 2014
Then and now
Current status
  – MGEX support
  – Real-time activities
  – Metadata efforts
IGS Data Centers: Where we were and where we are

• Using CDDIS as an example:

  Data
  • 1994:
    – ~85 daily 30-second sites

  • 2014:
    – 500+ daily 30-second sites
    – 330+ hourly 30-second sites
    – 180+ high-rate 1-second sites
    – 140+ real-time sites
    – 110+ MGEX sites (daily, hourly, high-rate)

  Products
  • 1994:
    – Orbit and ERP from 7 ACs + combination
    – Weekly and daily rapid

  • 2014:
    – Orbit, ERP, reference frame, clock solutions from 9 ACs
    – Reference frame solutions from 5 AACs
    – Weekly, daily, rapid, ultra-rapid
    – Orbit, clock, and reference frame combination
    – ~600 files/week
    – Ionosphere
    – Troposphere
    – MGEX
    – RTS
    – repro1 and repro2
20 Years of IGS Data Archiving

Number of High-Rate Station Days
Number of Hourly Station Days
Number of Daily Station Days
Number of High-Rate Sites
Number of Hourly Sites
Number of Daily Sites

* Estimated values for full year
CDDIS archive is currently nearly 10 Tbytes in size; over 9 Tbytes of this archive, or 90%, is devoted to the storage of GNSS data and derived products.
User Access: Today

IGS Workshop 2014

IGS Data Centers: Status and Progress
IGS data centers currently supporting MGEX: CDDIS and IGN (Global DCs), BKG (regional DC)

All data supplied in RINEX V3 format

Separated from operational IGS data (archived in RINEX V3 format)
  - Under “campaign” directory structure at GDCs, e.g.,
    ftp://cddis.gsfc.nasa.gov/gnss/data/campaign/mgex

Files utilize same filenaming convention, e.g.,
ssssddd#.yyt.Z
  - Same filenames have caused confusion and errors in archive

Current “separation” not conducive for integration in IGS operations

Need to plan for integration of RINEX V3 into operational archive structure while ensuring clear “separation”
RINEX V3: Filenames

- Move to utilization of new filenaming convention as specified in RINEX V3.02 documentation
  - mmmmscccc_s_yyyyddhhmm_fff_tt.rnx/crx.gz
  - ALGO00CAN_R_20121600000_01D_GN.rnx.gz
  - KIRU00SWE_R_20141600000_01D_30S_MO.crx.gz

- New filenaming convention will require significant work for DCs (and ACs)

- Will remove some difficulties DCs have experienced in handling V2 and V3 data from same site

- Test:
  - CDDIS has started providing RINEX V3 data (from selected sites) in both old and new filename format
  - Purpose: Promote use of new filename format
  - Utilizes gzip for compression

- Need a “teqc-like” utility for data QC and metadata extraction
• Nearly 140 sites supply real-time streams to IGS RTS
• Some high-rate files created at the receiver, some from real-time streams
• DCs must ensure that files generated from these streams are sufficiently reliable; some suggestions:
  – tool for comparison of RINEX files from various construction approaches
  – minimum requirements for acceptance of an accumulated data stream of observations as a RINEX file in IGS data archives
  – mandatory/optional observation types to be included
  – procedures to fill the gaps in the case data streams have been interrupted
  – separate directories for distinguishing between files created from streams and by receivers?
Data Centers and Metadata

- Past recommendation: IGS DCs investigate implementation of improved metadata standards and search and data discovery tools to improve access to the diverse archive structures available through the IGS and to aid in the integration of these archives into GGOS and other global data systems.
- Metadata is essential for data discovery
- DCs extract metadata from incoming files for managing their archives
- Tools, e.g., teqc, also useful for metadata extraction (DCs need similar tool for RINEX V3!)
- DCs provide statistics on availability of data
- Information about usability of the data would also be helpful
IGS Statistics on Data Availability

- Daily status files available at CDDIS summarize data holdings and include key metadata
- Available for operational and MGEX daily data
- File location: ftp://cddis.gsfc.nasa.gov/gnss/data/daily
- Software available to other DCs
Data Centers and Metadata

- Increasingly, data centers are storing site, file and status metadata in databases (example, IGS Central Bureau Site Log Manager)
- With metadata stored in a database, GSAC (Geodesy Seamless Archive Centers) software then allows access via web services to useful metadata and to data
- GSAC in federated mode allows additional useful queries such as tracking of status of file delivery across multiple data centers
- CDDIS, SOPAC, and UNAVCO are providing GSAC services.
- Several European geodetic data centers are also trying GSAC (ROB, INGV, SEGAL, and others)
Data Centers and Metadata

• 20 years ago – text based site logs stored station metadata which was shared via file transfer (site log, SINEX, station.info, RINEX headers)
• 2014 – metadata is often stored and managed in databases but sharing is still largely by file transfer in same formats as 20 years ago
• Management of metadata in files is highly inefficient and error prone
• Migration to use of web services (GSAC or other), and away from metadata in files, should be considered as a way to efficiently transfer the best metadata available
• The format of the metadata output by web services is flexible i.e., can simulate familiar file formats
• Address issues relevant to effective operation of all IGS data centers

• Meeting:
  – When: Splinter Session 4
    Thursday, June 26
    16:00-17:30 (concurrent with Antenna and TIGA WG meetings)
  – Where: Noyes 147