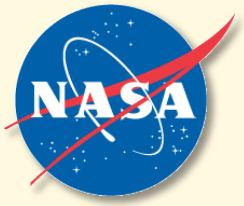
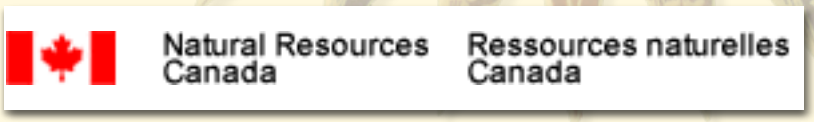
The Evolution of the IGS Flow of Data (and Products and Information) and Steps Ahead

C. Noll, NASA GSFC
M. Schmidt, NRCan
H. Habrich, BKG
B. Garayt, IGN





Evolution of the IGS Flow of Data and Steps Ahead

- Background
  - Past DC-Related Recommendations
  - Statistics on Data Availability
  - Statistics on Data Latency
- Issues
- Way forward
  - Data flow
  - RINEX construction
  - Compression

Past DC-Related Recommendations

- For DCs:
  - Provide statistics covering data availability, data latency, completeness of data files and the consistency of the records in the RINEX header and the site logs
  - Determine need for better harmonization of the IGS data center structure and contents
  - Investigate accumulation of data streams as a possible replacement for ftp file transfer of selected IGS data sets (i.e., high-rate 15-minute 1Hz files) with the provision that IGS data centers archive files of identical content
- For ACs:
  - Define requirements for data latency
  - Define requirements for data QC and validation at IGS data centers
- For others:
  - Update TEQC to accommodate new RINEX formats and future satellite systems

Statistics on Data Availability

- Daily status files available at CDDIS recently enhanced per IC request
- ftp://cddis.gsfc.nasa.gov/pub/gps/data/daily/*YYYY*/*ddd*/*YYddd*.status
- Software available to other DCs

IGS Tracking Network Status for 10-May-10 100510 10130 GPS Week 1583 Day 2 As of date: May 17 2010 10:27:21

Site	Dly (H)	No. Exp.	No. Obs.	Pts. Del.	%	Avg. MP1	Avg. MP2	Pos. Diff	No. Slps	V	Receiver Type	Antenna Type	Ant. Height	Marker Name	Marker Number	Type	RINEX Version	Dly (M)
abmf	10	25780	25640	136	99	0.41	0.51	0.04	23	1	TRIMBLE NETR5	TRM55971.00 NONE	0.0000	ABMF	97103M001	M	2.11	630
abpo	1	26268	25800	0	98	0.35	0.37	0.09	23	1	ASHTECH UZ-12	ASH701945G_M SCIT	0.0083	ABPO	33302M001	G	2.11	14
ade1	20	25447	25133	0	98	0.43	0.41	0.04	4	1	ASHTECH Z-XII3	ASH700936B_M SNOW	0.0000	ade1	50109S001	G	2.11	1253
ade2	20	25447	25071	1	98	0.42	0.41	0.04	5	1	ASHTECH Z-XII3	ASH700936B_M SNOW	0.0000	ade2	50109S001	G	2.11	1253
adis	24	28500	22796	672	79	0.53	0.58	0.03	22	1	JPS LEGACY	TRM29659.00 NONE	0.0010	ADIS	31502M001	M	2	1490
aira	1	24806	22489	608	90	0.35	0.48	0.03	30	1	TRIMBLE 5700	TRM29659.00 DOME	0.0000	AIRA	21742S001	G	2.11	111
ajac	8	23674	23644	0	99	0.19	0.16	0.04	2	1	LEICA GRX1200GGPRO	LEIAT504GG NONE	0.0000	AJAC	10077M005	M	2	511
albh	1	25563	25235	0	98	0.22	0.26	0.03	3	1	AOA BENCHMARK ACT	AOAD/M_T SCIS	0.1000	albh WCDA-ACP 927	40129M003	G	2.11	8
algo	1	24839	24730	0	99	0.22	0.23	0.07	1	1	AOA BENCHMARK ACT	AOAD/M_T NONE	0.1000	ALGO CACS-ACP 8831	40104M002	G	2.11	11
alic	3	25127	25120	0	99	0.28	0.32	0.04	12	1	LEICA GRX1200GGPRO	AOAD/M_T NONE	0.0070	ALIC	50137M001	M	2.11	231
alrt	1	29982	28999	12	96	0.11	0.13	0.06	8	1	ASHTECH UZ-12	ASH701945C_M NONE	0.1000	ALRT	40162M001	G	2.11	112
amc2	1	24442	23963	0	98	0.33	0.34	0.06	7	1	ASHTECH Z-XII3T	AOAD/M_T NONE	0.0000	AMC2	40472S004	G	2.11	10
amu2	8	58175	58045	4	99	0.32	0.36	0.16	2	1	TRIMBLE NETRS	ASH700936D_M SCIS	0.0000	AMU2	66040M002	G	2.11	493
ankr	9	24955	24856	0	99	0.37	0.45	0.06	0	1	TPS E_GGD	TPSCR3_GGD CONE	0.0700	ANKR	20805M002	M	2.11	570
antc																		
areq	1	25340	25179	0	99	0.17	0.19	0.05	17	1	ASHTECH UZ-12	AOAD/M_T JPLA	0.0610	AREQ	42202M005	G	2.11	11
arev																		
artu	1	25479	24966	0	97	0.40	0.41	0.03	3	1	ASHTECH Z-XII3	ASH700936D_M DOME	0.0796	ARTU	12362M001	G	2.10	10
aspa	1	25738	24918	790	96	0.59	0.43	0.03	14	1	TRIMBLE NETR5	TRM55971.00 NONE	0.0000	ASPA	50503S006	M	2.11	84
auck	6	24755	24754	1	99	0.28	0.32	0.07	0	1	TRIMBLE NETRS	TRM41249.00 NONE	0.0550	AUCK	50209M001	G	2.11	373
aukt	6	24748	23920	342	96	0.35	0.41	0.04	20	1	TRIMBLE NETRS	TRM55971.00 NONE	0.0030	AUKT	50216M001	G	2.11	373

.
.
.

Statistics on Data Latency

- Monthly and yearly files summarize latency of hourly data at CDDIS
- Could be expanded to summarize daily and sub-hourly high-rate latency
- ftp://cddis.gsfc.nasa.gov/pub/gps/data/hourly/*YYYY*/
- Software available to other DCs

Hourly GNSS Archive Statistics -- 2010/04/01 to 2010/04/30

Mon.	00-04m	05-09m	19-29m	30-59m	01-24h	01-3d	3d-mis
abmf	----	----	89.31%	7.08%	3.61%	----	----
abpo	----	83.41%	3.20%	0.76%	5.02%	7.61%	7.61%
adis	----	----	83.63%	3.56%	12.81%	----	----
aira	20.72%	1.39%	----	76.77%	1.11%	----	----
albh	----	83.17%	1.13%	0.28%	7.50%	6.51%	6.51%
algo	86.46%	10.39%	0.63%	0.63%	1.89%	----	----
alic	----	----	7.39%	6.14%	78.24%	7.53%	7.53%
amc2	----	98.06%	0.83%	0.28%	0.83%	----	----
ankr	----	----	90.39%	5.15%	4.46%	----	----
areq	----	91.69%	3.66%	1.13%	3.52%	----	----
arev	----	86.83%	5.65%	2.96%	4.57%	----	----
artu	----	----	97.91%	0.14%	1.95%	----	----
aspa	67.23%	26.44%	1.27%	3.38%	1.69%	----	----
auck	----	----	0.70%	97.91%	1.39%	----	----
aukt	----	----	9.17%	31.67%	59.17%	----	----
.							
.							
.							
zeck	----	----	89.93%	5.87%	4.20%	----	----
zim2	98.88%	----	----	----	1.12%	----	----
zimj	0.85%	0.14%	98.45%	0.14%	0.42%	----	----
zimm	98.88%	----	----	----	1.12%	----	----
zwe2	----	----	85.79%	4.60%	8.77%	0.84%	0.84%
Avg.	20.23%	23.55%	39.57%	6.87%	9.00%	0.59%	0.59%
Mon.	00-04m	05-09m	19-29m	30-59m	01-24h	01-3d	3d-mis

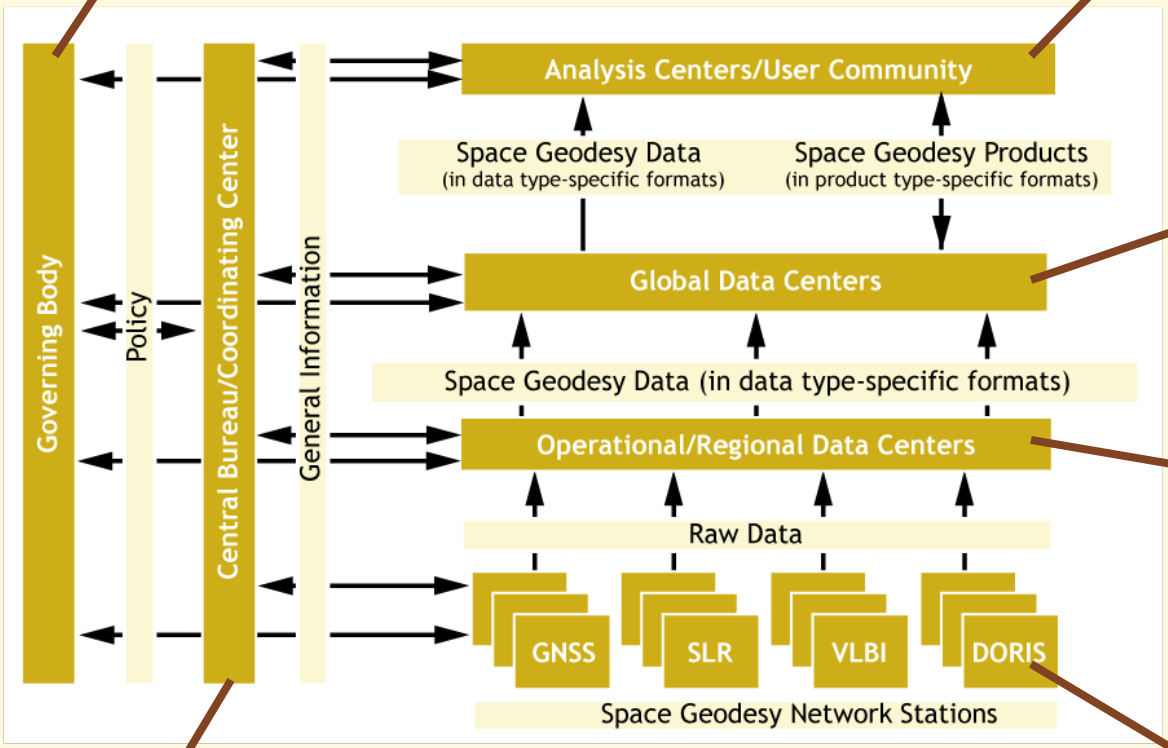
Steps Ahead

- Data flow
- Construction of RINEX files
- Compression
- Other?

Flow of Files within IGS

(Information, Data, Products)

- Governing Body
  - General oversight of service
  - Future direction



- Central Bureau
  - Management of service
  - Facilitate communications
  - Coordinate activities

- Analysis Centers and Coordinators
  - Provide products to users (e.g., station coordinates, precise satellite orbits, Earth orientation parameters, atmospheric products, etc.)

- Global Data Centers
  - Global data / product distribution
  - Both within and beyond the IGS

- Operational / Regional Data Centers
  - Interface to network stations
  - Perform QC and data conversion activities
  - Archive data for access to analysis centers and users

- Space Geodesy Network Stations
  - Continuously operational
  - Timely flow of data

Data Flow - Current Status

- Site guidelines: Data from IGS sites submitted to at a minimum of two DCs of which one is a GDC
  - DCs identified in site logs
- Current (actual) situation:
  - ODCs provide data to:
    - Either one or more RDC (usually one)

      and/or
    - One or more GDC
  - RDCs forward IGS sites to (one or more) GDC as required
  - RDCs/GDCs currently equalize selected data from selected sites
    - i.e., there is some equalization of data but there is no rigorous mirroring of primary data submissions or subsequent re-submissions
  - Result: inhomogeneous data set in DCs

Data Flow – Identified Problems & Goals

- Identified problems from GDC/AC perspective:
  - GDCs provide access to different sets of IGS sites
    - Users must “shop” multiple GDCs to retrieve required data
  - GDCs not synchronized and therefore not necessarily holding most current data
    - I.e., data holdings are not mirrored across GDCs
    - Inhomogeneous data set in GDCs and RDC’s (especially wrt replacement data sets)

- Goal: ensuring robust (24/7) AC/User access to data:
  - Primary data submissions must reach intended GDCs (and RDC’s)
  - Data resubmissions must reach intended GDCs (and RDC’s)
  - Continued data flow when designated GDC is unavailable (and by extension RDCs as well)

Data Flow - Recommendations

- GDC archive content
  - All GDCs archive data from **ALL** IGS stations as identified on the IGS network website
  - Advantages:
    - Ensures that data are consistent among GDCs and replacement data are distributed to all GDCs
    - Ensures users can easily get data from any GDC for any IGS site
    - Provides redundant data availability (also for resubmission)
- Data flow:
  - ODCs push primary data submission from their stations to ALL GDCs
  - ODCs push any/all subsequent resubmissions to ALL GDCs
  - ODCs issues advisory for ALL resubmissions
  - Advantages:
    - Implies simplified data flow
    - Ensures responsibility for data remains with ODC
    - Allows for publishing information about replacement data

Data Flow - Implementation

- How do we get there:
  - Prepare GDCs for new data flow paths (and additional storage required)
  - Prepare ODCs for data push to:
    - All GDCs
    - Appropriate RDCs
    - Others (as required on individual basis)
  - Implement ODC to GDC (and RDC) direct push

RINEX Construction Issues (1/2)

- Several methods are used to form daily RINEX file of 30-second sampled observations
  - Generated at station (receiver)
  - Created from concatenated hourly files
  - Created from accumulated either binary or RTCM high-rate RT data streams and filtered/decimated to 30-second sampling rate
- Different methods cause different results
  - Concatenated files are not necessarily equivalent to “true” 24 hr data files
  - RINEX files from RT data stream:
    - Number of epochs reduced due to loss of data in data stream
    - Increase in data gaps and/or cycle slips
  - RINEX files from RT RTCM:
    - Observation types in RINEX file: at most 4 observables transferred through RTCM, analysis-specific s/w formatting
    - Data field resolution: code observable from RTCM 3.0 less accurate than RINEX V2; HP-RTCM should address this
    - New RTCM format (HP-RTCM) should address accuracy concerns
  - Receiver features yield different # obs in epoch or different cycle slips

RINEX Construction Issues (2/2)

- Currently the IGS site guidelines state that daily/hourly files should NOT be created from streams
- However, for future consideration, creating daily/hourly RINEX files from streams:
  - Advantages:
    - Stream established directly from receiver to DC
    - Files available immediately following end of epoch (hour or day)
  - Disadvantages
    - Stream interruptions mean incomplete files
    - Inconsistent files at DCs if multiple DCs receive streams and generate RINEX
- Further research needed to address differences (see RINEX Construction Recommendations)

File Generation Differences

- At this time files created from streams are not equivalent to files created at stations and transferred by ftp
  - Adequate for real-time and near real-time applications?
  - Not adequate for long term archive and future analysis
- Different numbers of epochs (loss of data in data stream)
- Different RINEX observation types
  - Currently, 4 observables at most transferred through RTCM (HP-RTCM will resolve this)
  - Analysis software-specific formatted files (e.g., Bernese software 5.0 ignores C2 observable)
- Features of receivers may cause different numbers of observations within an epoch and different cycle slips
- Data field resolution (code observable from RTCM 3.0 less accurate than RINEX 2)

RINEX Construction - Recommendations

- Develop tool for comparison of RINEX files from various construction approaches, e.g., zero-baseline processing
  - Run the tool at the site and at the data center to recover receiver-specific and transport-specific issues
- Define minimum requirements for acceptance of an accumulated data stream of observations as a RINEX file in IGS data archives (work with IC)
  - Minimum number of epochs
  - Maximum tolerated epoch differences per satellite
  - Maximum tolerated numerical difference for each individual observation (after appropriate harmonization of compared RINEX files)
  - Data field resolution for each observation type
- Specification of observation types that are mandatory and others that are optional
- Agree on procedures to fill the gaps in the case data streams have been interrupted

Compression

- Currently used data compression (Unix compress) is inefficient and out of date
- Recommendation:
  - Change compression used in IGS infrastructure from UNIX compress (Z) to:
    - bzip2 or
    - gzip
- Discussion:
  - Both are widely distributed across multiple O/S
  - Better compression factor
  - Speed:
    - gzip faster than ‘Z’
    - Bzip2 slower than ‘Z’
  - Current leaning is toward bzip2 but need to complete testing/consultation
- Implementation scenarios:
  - Change compression throughout IGS infrastructure
  - Allow DCs to utilize more efficient compression on historic archives
- **Any changes must be coordinated with DCs, ACs, manufacturers, users, …**

Remaining Issues

- Long-term access to high-rate data
  - Can hourly files be created and archived from sub-hourly files?
  - Can files be “packaged” on a site/day basis
  - How long retained online?
- Data usage statistics
  - GDCs provide information (who, how much) on data downloads by ODC/station
- Continue to work with the IC and ACs to resolve these and other issues!

Data Center WG Meeting

- Tuesday, June 29, 16:30-17:30
- Topics:
  - Viability/Requirement for WG
  - Membership
  - Top Issues
    - Compression
    - High-rate data archive
    - RINEX formulation

Backup Slides

Data Center Updates

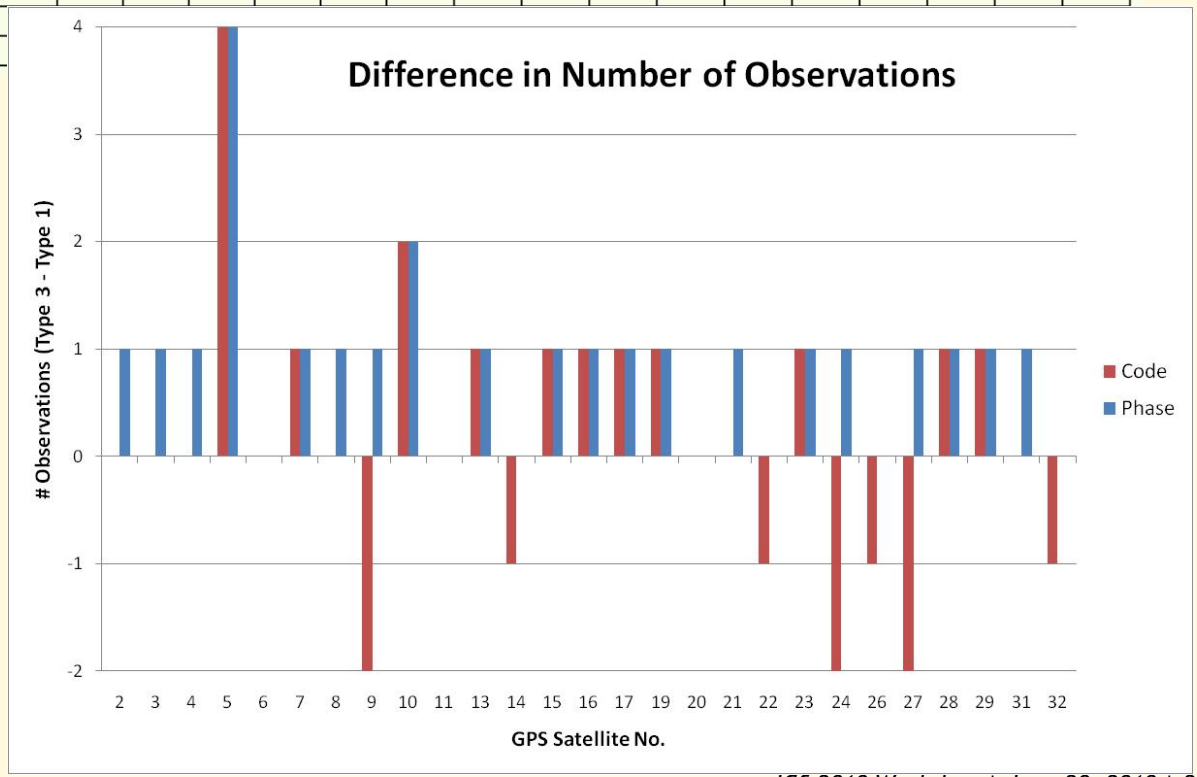
- CDDIS
  - Operational on new server system, same access for users, new access procedure for data and product upload
  - Working on revision to metadata and new data discovery capability
  - Archiving test data sets (software receiver, RINEX V3, L5)
- IGN
  - Developing new websites for the GDC and RF coordination
- KASI
  - GDC system upgraded for better reliability and backup capabilities
  - Three cluster servers for high availability of GDC system
  - Mirroring to backup server
- BKG
  - Archiving RINEX V3 data
  - Preparing to archive Galileo data

File Generation Differences (1/2)

- Single epochs are missing in from files created in data streams (statistics available at EUREF)

	174/2010 (23-06-2010)																						
	0h	1h	2h	3h	4h	5h	6h	7h	8h	9h	10h	11h	12h	13h	14h	15h	16h	17h	18h	19h	20h	21h	22h	23h
ACOR	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	96%	88.5%	88.6%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
ALAC	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	96.1%	88.6%	88.7%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
ALBA	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	96.1%	84.9%	88.6%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
ALME	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	96.1%	88.5%	88.7%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
AUT1	100%	100%	100%	100%	100%	95.5%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	95.6%
BELF	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
BELL	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	99.9%	100%	100%	100%	100%	100%	100%	100%
BOR1	100%	100%	100%	100%	100%	100%	100%																	
BORJ	100%	100%	99.1%	100%	100%	100%	100%																	

- Tests show that there are different numbers of observations/epoch in stream files than station files for some satellites



Compression (2/2)

- Implementation (proposed):
  - Sept 30, 2010: last date for post workshop comments to be received by IGS DCWG
  - Oct 30, 2010: complete synthesis of input and final recommendation by DCWG
    - Circulate to IC, DC's, AC, manufacturer's for last comment
  - Nov 30, 2010: complete feedback for final discussion and report completion at DCWG meeting at Fall AGU
  - Distribute recommendation (IGS, Manufacturers, etc.)
  - January 2011: commence implementation
  - NOTE: period of overlap (6mths – 1 yr) required to accommodate the necessary changes to multiple post processing software which retrieve data from IGS-DC's