GNSS Quality Control Improvements and Provider Performance Tracking at the Crustal Dynamics Data Information System (CDDIS)

Justine Woo (Justine.y.Woo@nasa.gov)
Sigma Space Corporation
Lanham, MD, USA

Rebecca Limbacher (Rebecca.i.Limbacher@nasa.gov)
Science Systems and Applications, Inc.(SSAI)
Lanham, MD, USA

Patrick Michael (Patrick.Michael@nasa.gov)
NASA Goddard Space Flight Center/Code 61A
Greenbelt, MD, USA

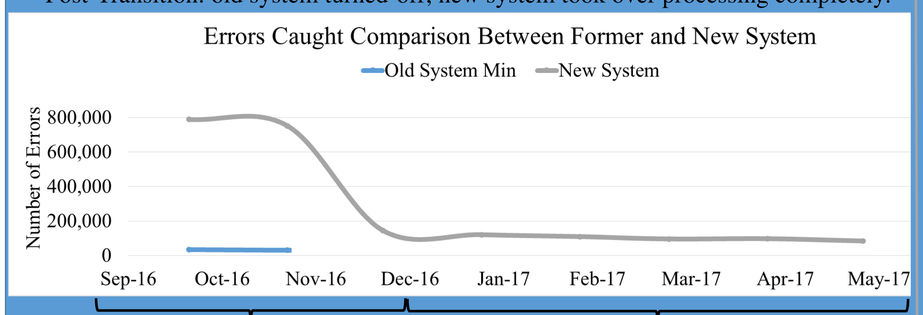
Carey Noll (Carey.Noll@nasa.gov)
NASA Goddard Space Flight Center/Code 61A
Greenbelt, MD, USA

Abstract: At the end of 2016, the Crustal Dynamics Data Information System (CDDIS) transitioned to a new hardware system and upgraded its processing software with the goal of increasing quality control measures while also improving the timeliness of GNSS data archiving. This new system increases our ability to consistently track errors and issues associated with data uploaded to the CDDIS and to perform post-ingest-and-archive checks on the files received. In this poster, we will review and contrast quality control measures employed at the CDDIS before and after the software upgrade, and present the methods used to address and eliminate these issues. The CDDIS aims to ensure the correctness and integrity of all GNSS data available to the IGS community; future goals and improvements in continuing this work will be discussed.

Pre- and Post-Transition Comparisons of Archive Operations

The CDDIS archive operations system processes all incoming files prior to moving them to the public archive. This was handled differently, pre- and post-transition

- Pre-transition period: old and new systems ran in parallel; the new system caught and moved errors to quarantine; old system populated existing tables and moved the file to archive.
- Post-Transition: old system turned-off; new system took over processing completely.



Pre-transition:

- Processing at 15-min intervals
- Empty file
- Unknown file
- Compression*
- Unaccepted data interval*
- RINEX naming scheme*
- Future file*
- File older than 5-years*
- Unknown RINEX version^
- Marker name matches file name^
- Uncompacted RINEX file sent^
- Date Matches File Name^
- MD5SUM and SHA512SUM

Post-Transition:

(Checks in addition to those pre-transition)

- Processing at 5-min intervals
- Mislabeled compacted RINEX file
- Invalid day of year
- Invalid RINEX V3 file
- Virus Scan

Additional Notes Post-Transition:

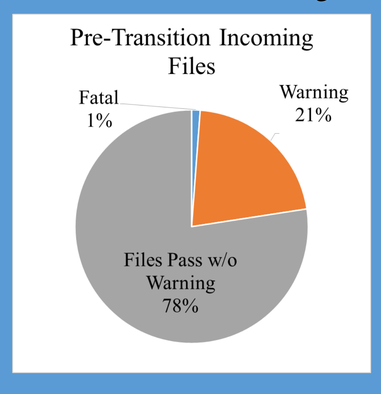
- All errors are recorded to the database allowing for error tracking
- Consistent checks regardless of the file type

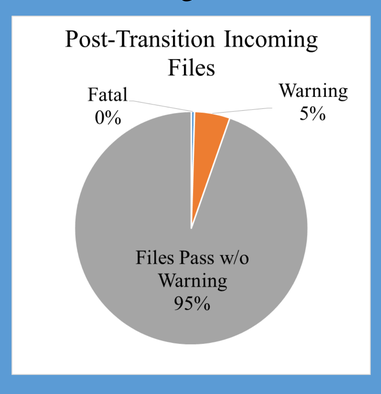
*Inconsistently checked by original software, consistently checked once implemented in new software
^Check performed by new software

Incoming Files Data Check

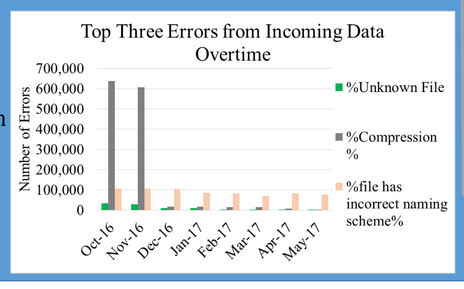
- Fatal Errors = errors in the file prevent the file from being archived
- Warnings = errors in the file are corrected and the file is archived
- Both fatal and warning errors are tracked in the database
- File metadata also extracted and tracked in the database

Comparison of Errors Detected from Incoming Files Pre- and Post-Transition Using the New Processing Software





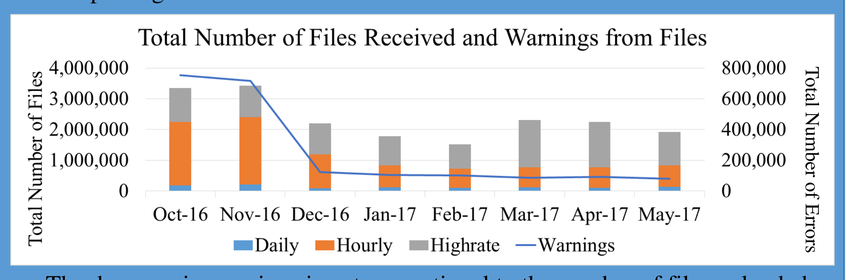
- The percentage of good files the CDDIS has been receiving increased from 78% to 95% after the transition
- Prior to the transition the greatest number of errors received were compression errors. The number of unknown files also decreased right after the transition



Drop in Incoming File Errors Post-Transition

Immediate Decrease Post-Transition:

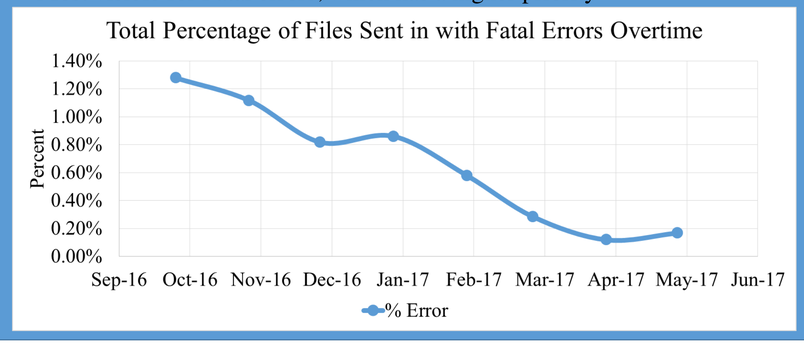
- Substantial drop in the number of files uploaded to CDDIS
- Percentage of incoming files generating warnings also dropped but to a steeper degree



- The decrease in warnings is not proportional to the number of files uploaded
- Less files with warnings sent in overall

Decrease Due to Efforts from CDDIS

- The reduction in the number of errors in incoming filers are the product of CDDIS tracking and contacting providers about their errors.
- Most visible for Fatal errors, which has a higher priority



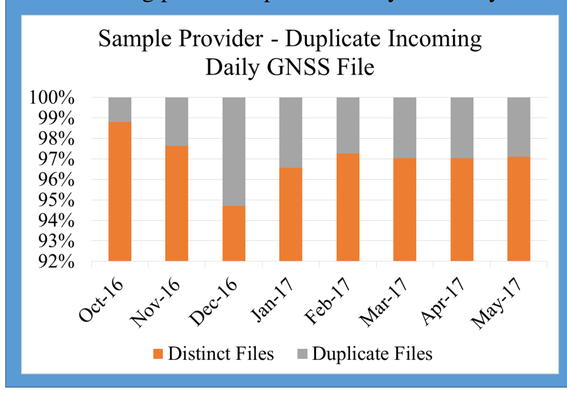
Post-Processing Checks

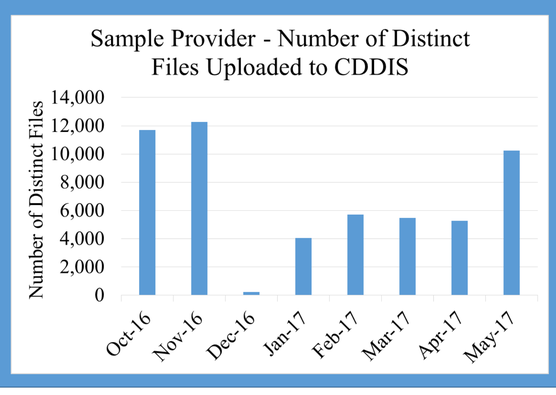
Weekly Error Reports

- Weekly error reports are generated from the database
- Focus on specific providers and the errors they are sending us allows us to address the issues caught more effectively by address the source of the problem
- Additional checks include checking for duplicate data
- Tracking provider upload activity/inactivity

Error Status: Contacted Jun 6 about large number of duplicate highrate files and uncompressed highrate data. No response yet.

- Duplicates:
  - Hourly: 100 (3%)
  - Highrate: 3974 (99%)
- Warning Occurences:
  - Uncompressed files: 4431
  - Incorrect compression: 46
  - Sending o-files: 118





New Error Discovery

As the CDDIS has been making improvements, new types of errors have become known, but the struggle lies in how to address these issues. We believe input from the IGS community is vital. An instance of this is shown below, and we are working to find how to best address such items.

Table: Error Report on duplicate data comes back with negative values for the number of resent files

date	total files	unique files	unique name	duplicate files	resent files	% duplicate	% resent
2017-05-23	486	470	486	16	-16	3.2922	-3.2922
2017-05-24	500	484	500	16	-16	3.2	-3.2
2017-05-25	492	476	492	16	-16	3.252	-3.252
2017-05-26	488	472	488	16	-16	3.2787	-3.2787
2017-05-27	495	479	495	16	-16	3.2323	-3.2323
2017-05-28	501	485	501	16	-16	3.1936	-3.1936
2017-05-29	498	482	498	16	-16	3.2129	-3.2129

After working to track the cause of the issue, two types of anomalies were detected:

Different Site, Same File:

- Two files with different file names and marker names have identical content and checksums
- How should this situation be handled?
- Generic example*:

Filename	MD5 Checksum
<**site1**>1450.17n.Z	d41d8cd98f00b204e9800998ecf8427e
<**site2**>1450.17n.Z	d41d8cd98f00b204e9800998ecf8427e

Same Site, Different File Type:

- Two different types of navigation files with the same marker name have identical content and checksums
- How should this situation be handled?
- Generic example*:

Filename	MD5 Checksum
<site1>1450.17**g**.Z	c234727462d26866d48dc83236747fcc
<site1>1450.17**n**.Z	c234727462d26866d48dc83236747fcc

*The CDDIS also checks the sha512 checksum, but it was not included to conserve space

Addressing and Eliminating Errors

Provider reports* are generated and reviewed weekly.^ Fatal errors are addressed first, and then warnings are second.

Provider 1

Date	Daily Duplicates (%)	Hourly Duplicates (%)	High-rate Duplicates (%)	Unaccepted Data Interval	RINEX Naming Scheme	Empty File	Compression	Marker Name or Date Not Matching	Unknown File	Error Status
12/20/16	0	7	N/A	0	0	0	0	0	0	Not contacted
12/27/16	0	8	N/A	0	0	0	0	0	0	Not contacted
1/3/17	0	8	N/A	0	0	0	0	0	0	Not contacted
1/10/17	0	8	N/A	0	0	0	0	0	0	Not contacted
1/17/17	1	11	N/A	0	0	0	0	0	0	Not contacted
1/24/17	0	10	N/A	0	0	0	0	0	0	Not contacted
1/31/17	0	14	N/A	0	0	0	0	0	0	Contacted on Jan 31, fixing issues
2/7/17	0	15	N/A	0	0	0	0	0	0	Contacted on Jan 31, fixing issues
2/14/17	0	3	N/A	0	0	0	0	0	0	Contacted on Jan 31, fixing issues
2/21/17	0	0	N/A	0	0	0	0	0	0	No problems
2/28/17	0	0	N/A	0	0	0	0	0	0	No problems
3/8/17	0	0	N/A	0	0	0	0	0	0	No problems
3/14/17	0	0	N/A	0	0	0	0	0	0	No problems
3/21/17	0	0	N/A	0	0	0	0	0	0	No problems
3/28/17	0	0	N/A	0	0	0	0	0	0	No problems
4/4/17	0	0	N/A	0	0	0	0	0	0	No problems
4/11/17	0	0	N/A	0	0	0	0	0	0	No problems
4/18/17	0	0	N/A	0	0	0	0	0	0	No problems
4/25/17	0	0	N/A	0	0	0	0	0	0	No problems
5/2/17	0	0	N/A	0	0	0	0	0	0	No problems
5/9/17	0	0	N/A	0	0	0	0	0	0	No problems
5/16/17	0	0	N/A	0	0	0	0	0	0	No problems
5/23/17	0	0	N/A	0	0	0	0	0	0	No problems
5/30/17	0	0	N/A	0	0	0	0	0	0	No problems

Notes Provider 1 (Desired Outcome):

- Consistent and significant numbers of duplicate hourly files detected
- The database was queried to gather more information about specific file names and the times they were uploaded
- The data provider was contacted about the duplicate files and was provided with the information from the database
- The data provider reviewed their procedures and stopped uploading duplicate hourly files
- There have been no further errors since the problem was resolved

Provider 2

Date	Daily Duplicates (%)	Hourly Duplicates (%)	High-rate Duplicates (%)	Unaccepted Data Interval	RINEX Naming Scheme	Empty File	Compression	Marker Name or Date Not Matching	Unknown File	Error Status
12/20/16	6	0	N/A	0	0	335	0	0	0	Not contacted
12/27/16	3	0	N/A	0	0	330	0	0	0	Not contacted
1/3/17	1	0	N/A	0	0	334	0	0	0	Not contacted
1/10/17	2	0	N/A	0	0	333	0	0	0	Not contacted
1/17/17	3	0	N/A	0	0	332	0	0	0	Contacted Jan 17
1/24/17	3	0	N/A	0	0	331	0	0	0	Contacted Jan 17
1/31/17	6	0	N/A	0	0	329	0	0	0	Contacted Jan 17
2/7/17	7	0	N/A	0	0	333	0	0	0	Contacted Feb 6, fixing issues
2/14/17	4	0	N/A	0	0	75	0	0	0	Contacted Feb 6, fixing issues
2/21/17	0	0	N/A	0	0	0	0	0	0	No problems
2/28/17	0	0	N/A	0	0	0	0	0	0	No problems
3/8/17	2	0	N/A	0	0	2	0	0	0	Not contacted
3/14/17	0	0	N/A	0	0	82	0	0	0	Not contacted
3/21/17	0	0	N/A	0	0	12	0	0	0	Not contacted
3/28/17	0	0	N/A	0	0	11	0	0	0	Contacted Mar 28, fixing issues
4/4/17	4	1	N/A	0	0	0	0	0	0	Contacted Mar 28, fixing issues
4/11/17	0	0	N/A	0	0	0	0	0	0	No problems
4/18/17	0	0	N/A	0	0	216	0	0	0	Not contacted
4/25/17	0	0	N/A	0	0	0	0	0	0	No problems
5/2/17	0	0	N/A	0	0	0	0	0	0	No problems
5/9/17	0	0	N/A	0	0	14	0	0	0	Not contacted
5/16/17	0	0	N/A	0	0	155	0	0	0	Contacted May 16, fixing issues
5/23/17	0	0	N/A	0	0	18	0	0	0	Contacted May 16, fixing issues
5/30/17	2	0	N/A	0	0	0	0	0	0	Contacted May 16, fixing issues

Provider 3

Date	Daily Duplicates (%)	Hourly Duplicates (%)	High-rate Duplicates (%)	Unaccepted Data Interval	RINEX Naming Scheme	Empty File	Compression	Marker Name or Date Not Matching	Unknown File	Error Status
12/20/16	N/A	0	N/A	0	0	0	1	0	0	Sent 1 file
12/27/16	N/A	0	N/A	0	0	0	1	0	0	Sent 1 file
1/3/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
1/10/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
1/17/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
1/24/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
1/31/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
2/7/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
2/14/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
2/21/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
2/28/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
3/8/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
3/14/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
3/21/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
3/28/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
4/4/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
4/11/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
4/18/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
4/25/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
5/2/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
5/9/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
5/16/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
5/23/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data
5/30/17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No data

Notes Provider 3 (No Data Uploaded on New System):

- Before the transition, the provider had been regularly uploading to the CDDIS daily, hourly, and high-rate data for 7 sites.
- Only a few files have been delivered since the transition
- The CDDIS has been in contact with the provider, and we are working to remedy the uploading obstacles faced by them
- Meanwhile, the CDDIS is trying to locate alternate sources to supply the missing data

*The reports displayed are edited to save space
^Usually on Tuesdays

Future Goals and Improvements

The CDDIS has been making strides to achieve the following goals:

- Further reduce the numbers of errors in incoming data
- Increase the data uploads from providers. While the CDDIS is able to retrieve data from providers, it is standard for the providers to upload data to the CDDIS archive.
- Compile a report of IGS data missing from the CDDIS archive.
- Further develop the quality checks and error tracking software as new issues are detected.
- Re-process all archived data using the new processing software.

To accomplish these goals the CDDIS has been developing preliminary plans to distribute performance statistics reports to the providers on a regular basis.

CDDIS will also be adding high-rate files created from real-time data to our archive in the future. Please see the CDDIS Operations Team poster "Real-Time Data and Product Performance Metrics at NASA"

More Information/Feedback

- Data and products are acquired as part of NASA's Earth Science Data Systems and archived and distributed by the Crustal Dynamics Data Information System (CDDIS)
- The staff welcomes feedback on the CDDIS and in particular the ideas expressed in this poster; contact Justine Woo (Justine.y.woo@nasa.gov) or Rebecca Limbacher (Rebecca.i.Limbacher@nasa.gov)
- CDDIS Helpdesk (cddis-help@lists.nasa.gov)
- CDDIS Operations (gsfc-cddis-ops@lists.nasa.gov)