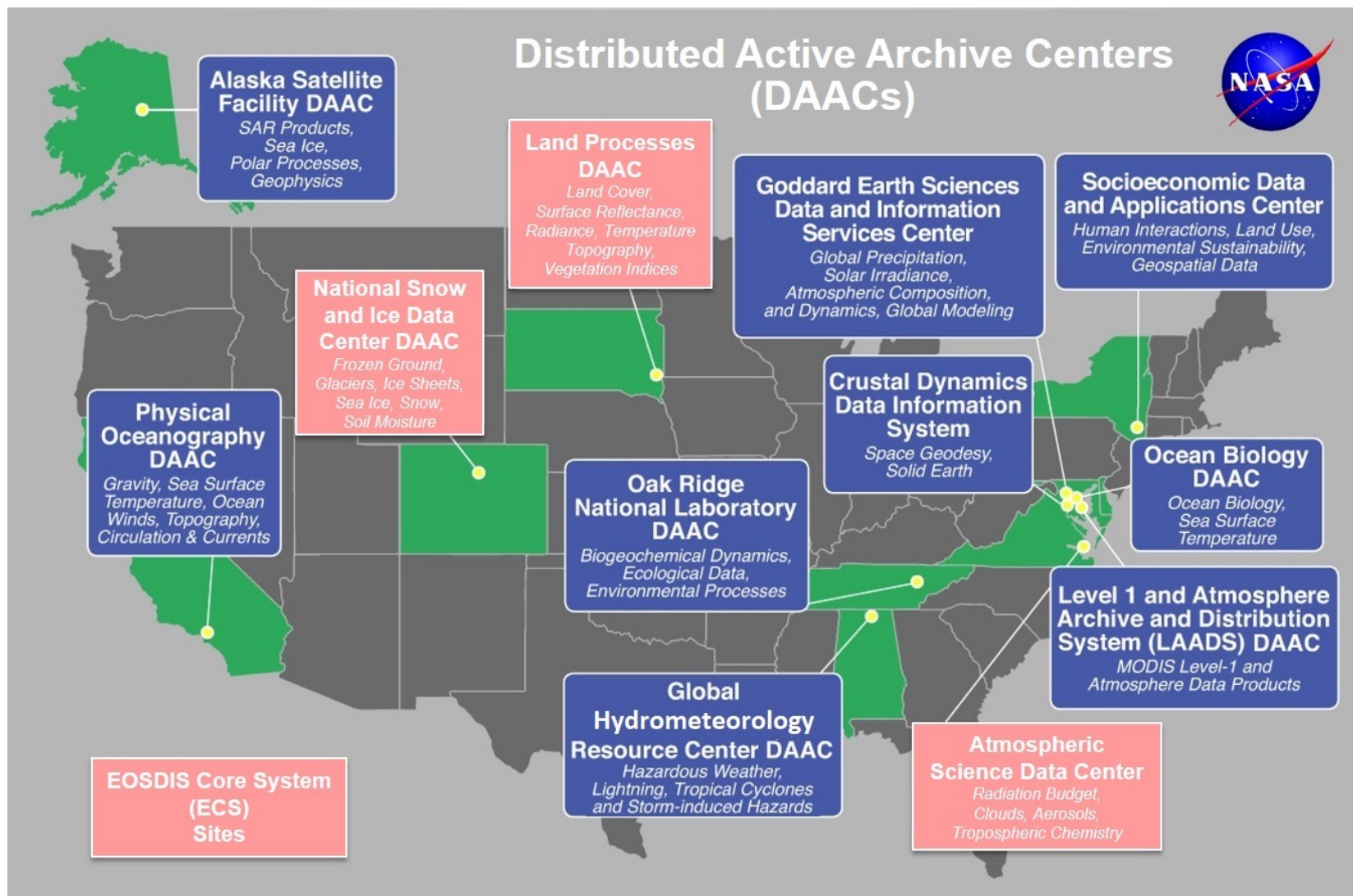


Tour de l'IGS

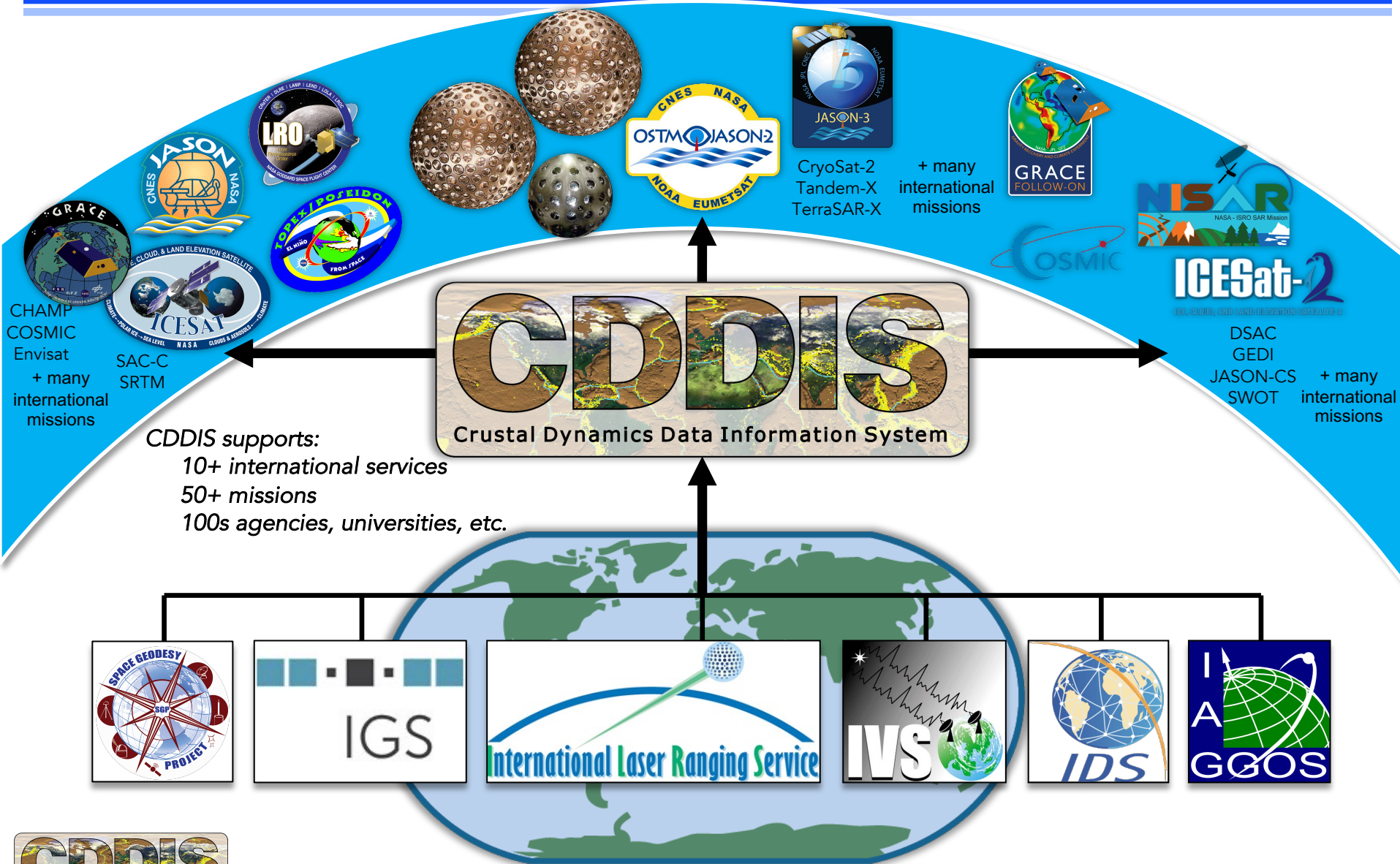
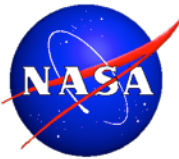
Updates from CDDIS



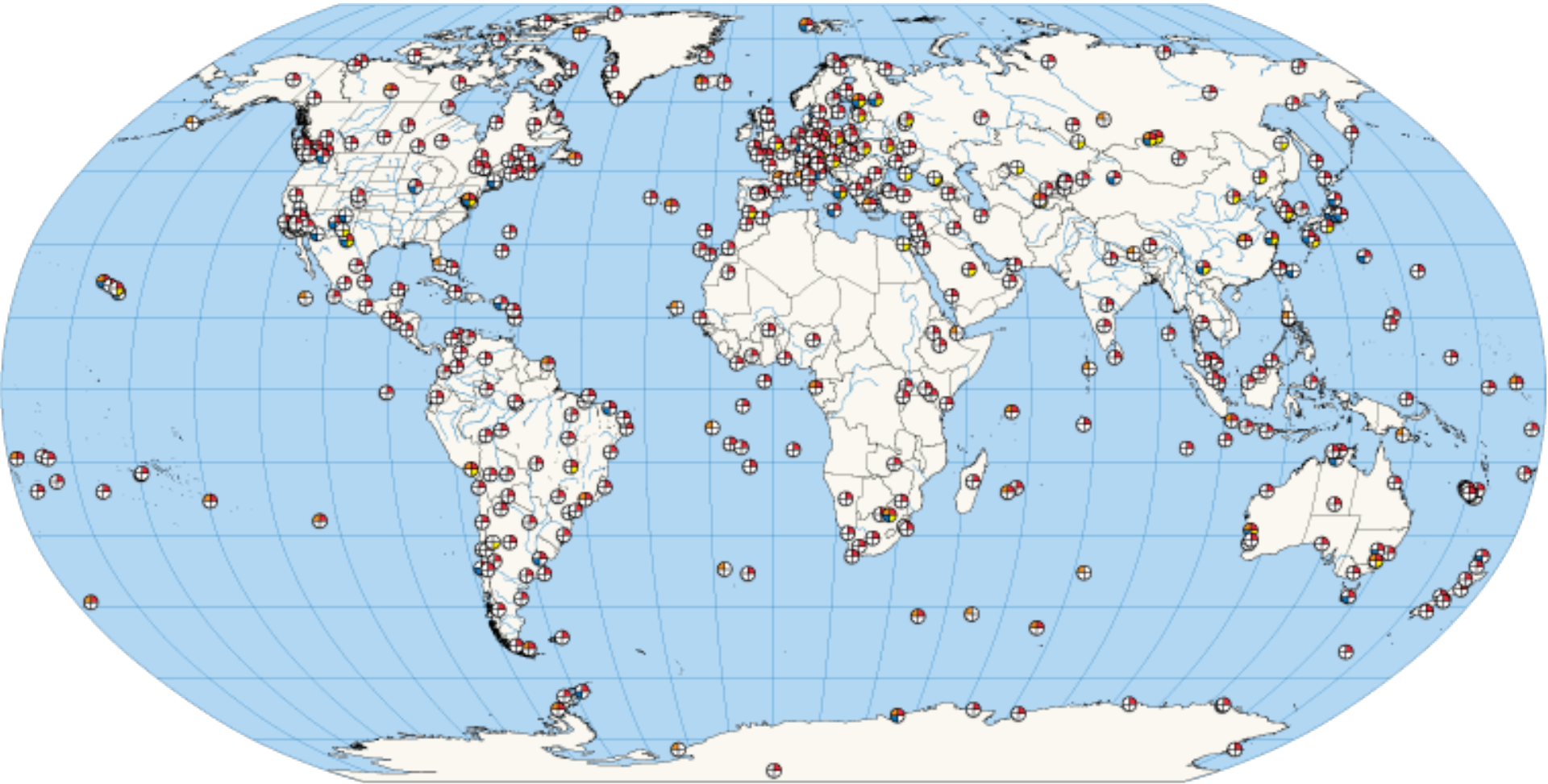
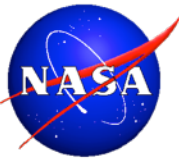
NASA Earthdata Archive System



NASA Missions & CDDIS



Data from Every Corner of the World



1500+ observing sites worldwide

CDDIS by the Numbers



■ Datasets/Products

- 244
- Standardized formats
- ~170 file types

■ 200+ data providers

- Large portion non-US
- Varied time delivery
- Multiple data/products

■ Files

- ~400+ million
- 55+ million per year (150,000+ per day)

■ Archive

- Same standard for ~40 years
- File size - 10 KB – 300+ GB
- ~80TB current
- ~200TB/year by 2023

■ Ingest - GNSS

- Daily: ~1800 files/day
- Hourly: ~37,000 files/day
- High-rate: ~120,000 files/day
- Products: ~950 files/day
- ~35GB per day

■ Distribution - GNSS

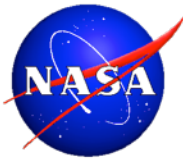
- 1.4+ billion files per year
- ~400TB distribution
- 300,000+ users

■ QC - GNSS

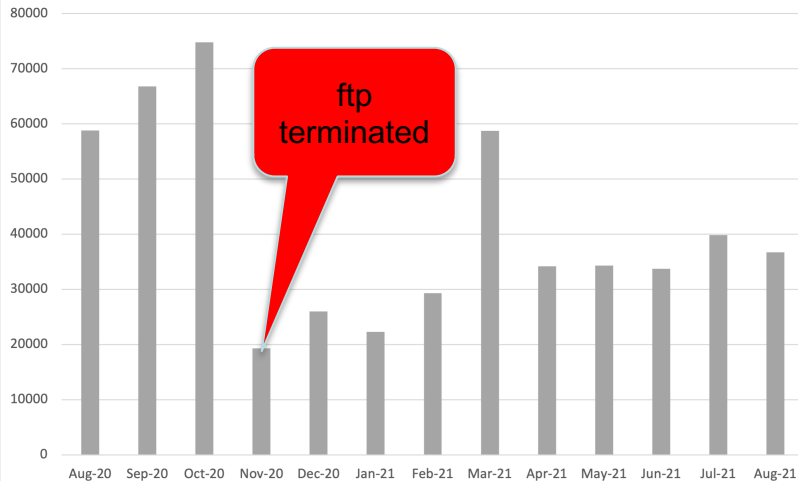
- ~0.05% rejected ~100 files
- ~8.5% or 17,000 files

Every file coming into CDDIS is opened and quality checked and metadata extracted – upwards of 50+ actions per file

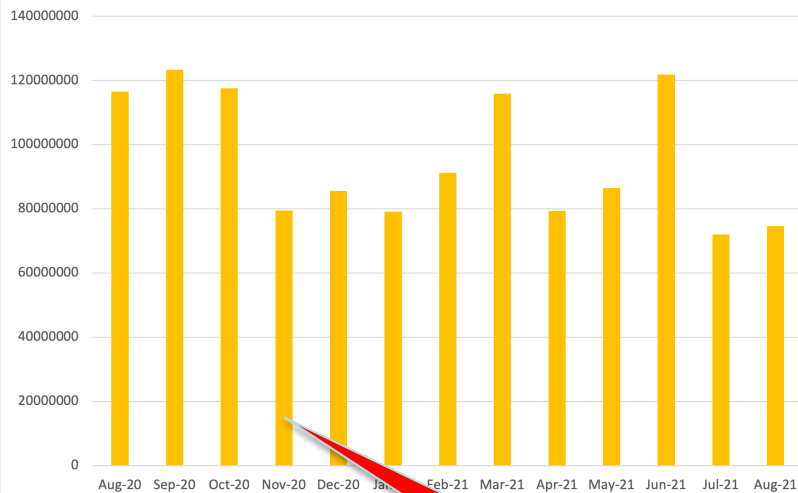
Post Anonymous FTP Comments



Distribution IPs (distinct per month)



Distribution Filecount (files per month)

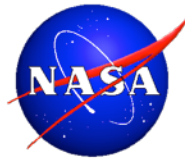


- ftp terminated 1 Nov 2020
- New “normal”
 - ~2.8M files/day - ~35% decrease
 - ~1.1TB per day – ~30% decrease
- Zombies
- Users not tracking scripts
- LOTS of unsupported software
- 50/50 split https & ftps
- Operational usage

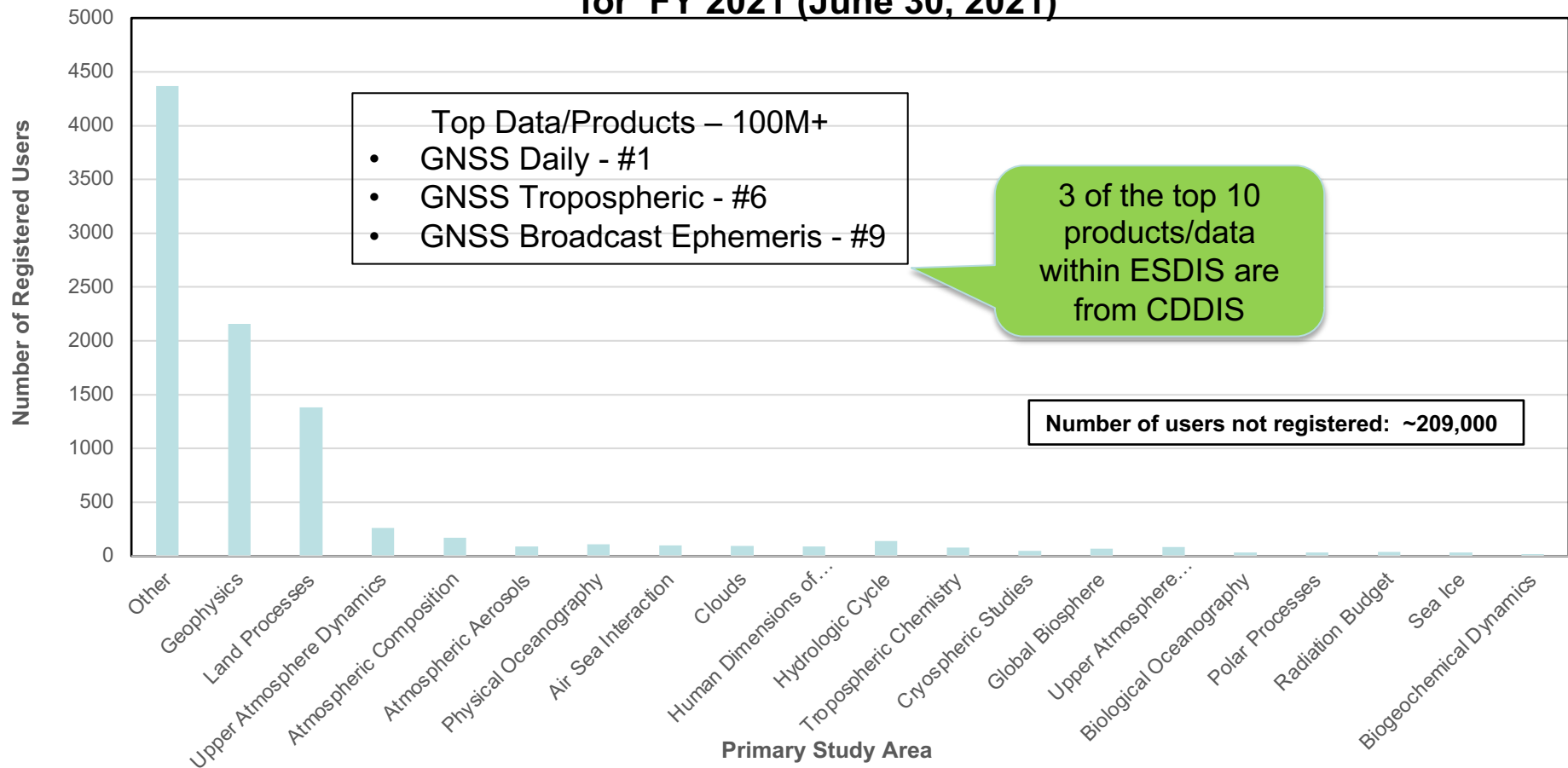


ftp
terminated

CDDIS Users Primary Research Area



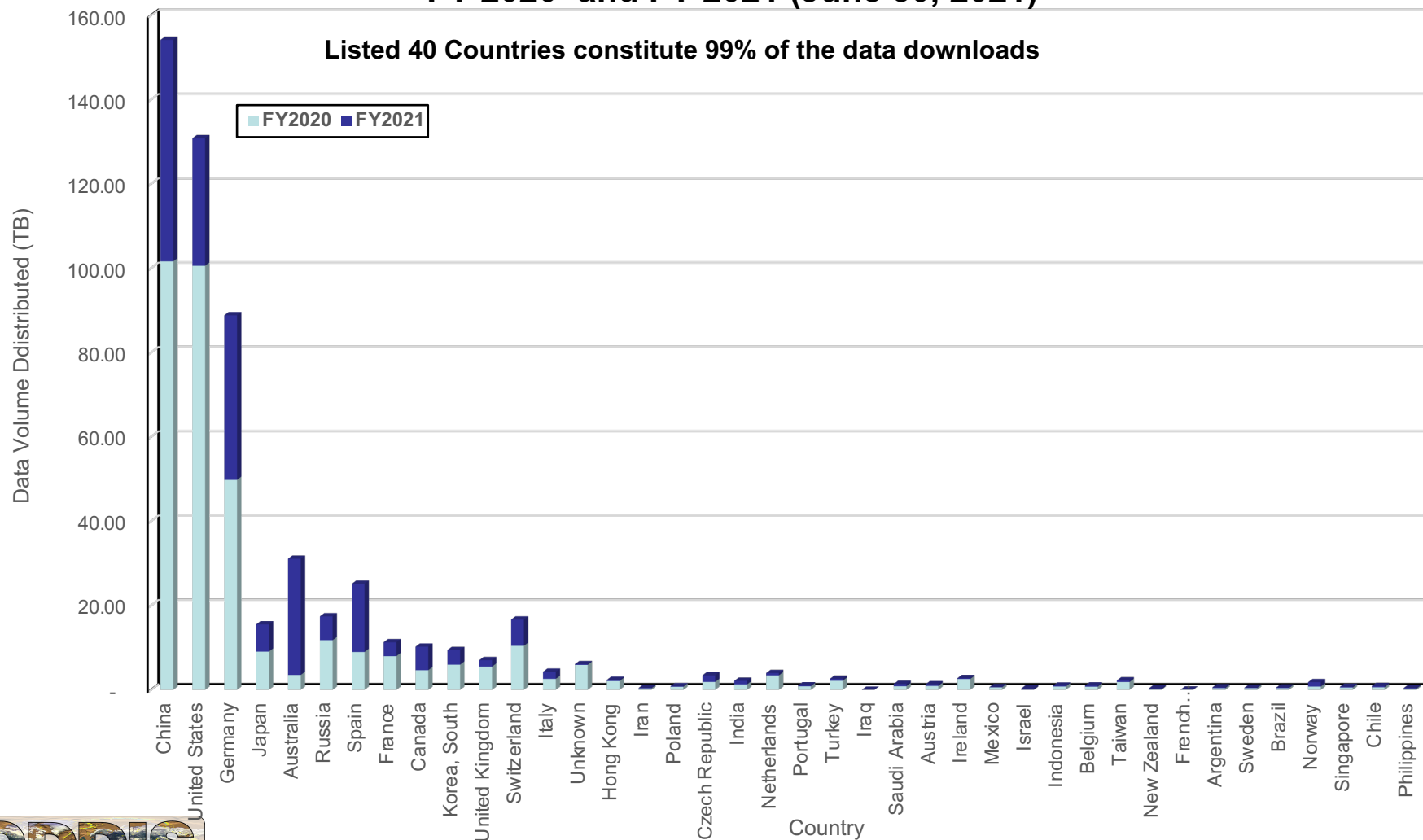
Number of Registered Users with Primary Study Area Downloading Data from CDDIS for FY 2021 (June 30, 2021)



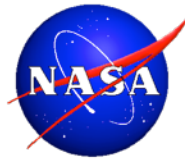
Country Distributions



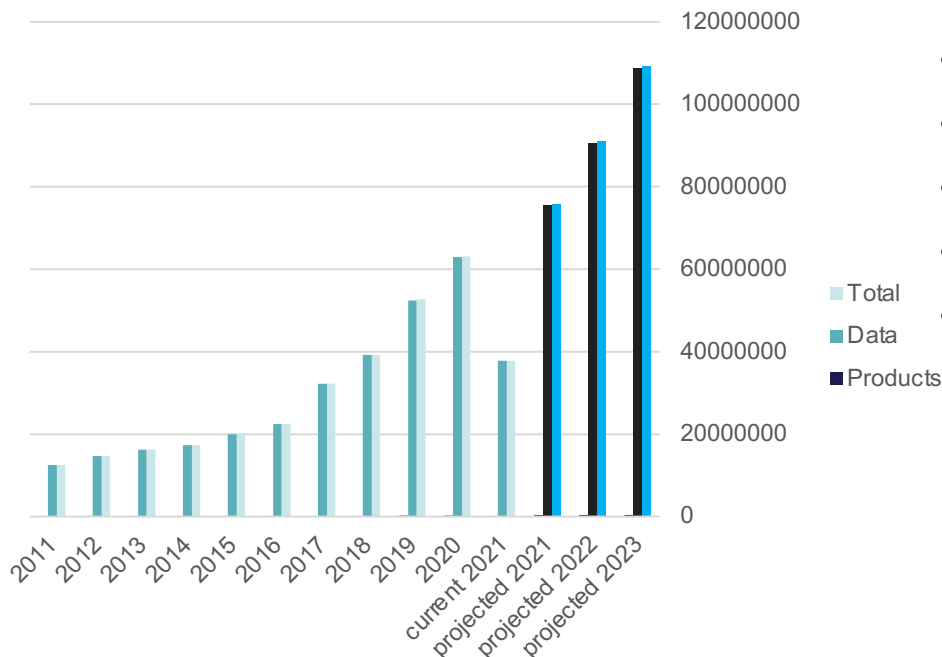
Top 40 Countries Downloading Data Product Volume from CDDIS for FY 2020 and FY 2021 (June 30, 2021)



GNSS Growth Since 2011

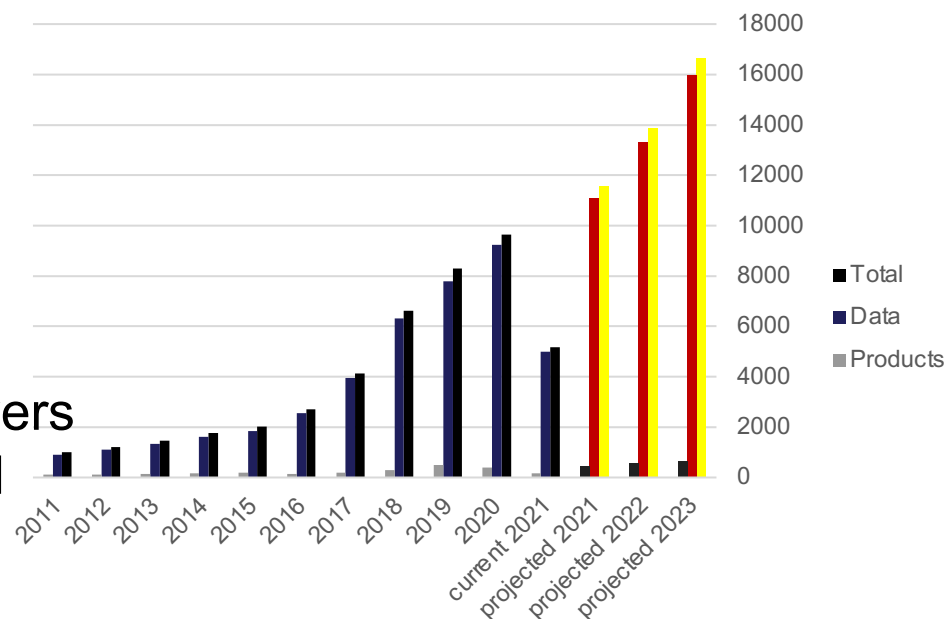


GNSS File Count



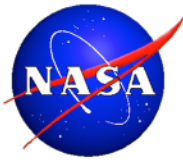
- ~20% growth per year in files
- ~100M files per year in 2023
- High-rate & Hourly are main drivers
- 1 high-rate site = 100,000+ files/year
- High-rate only ~50% of IGS sites
- RINEX V2 decreasing but V4 coming

GNSS Volume (GB)

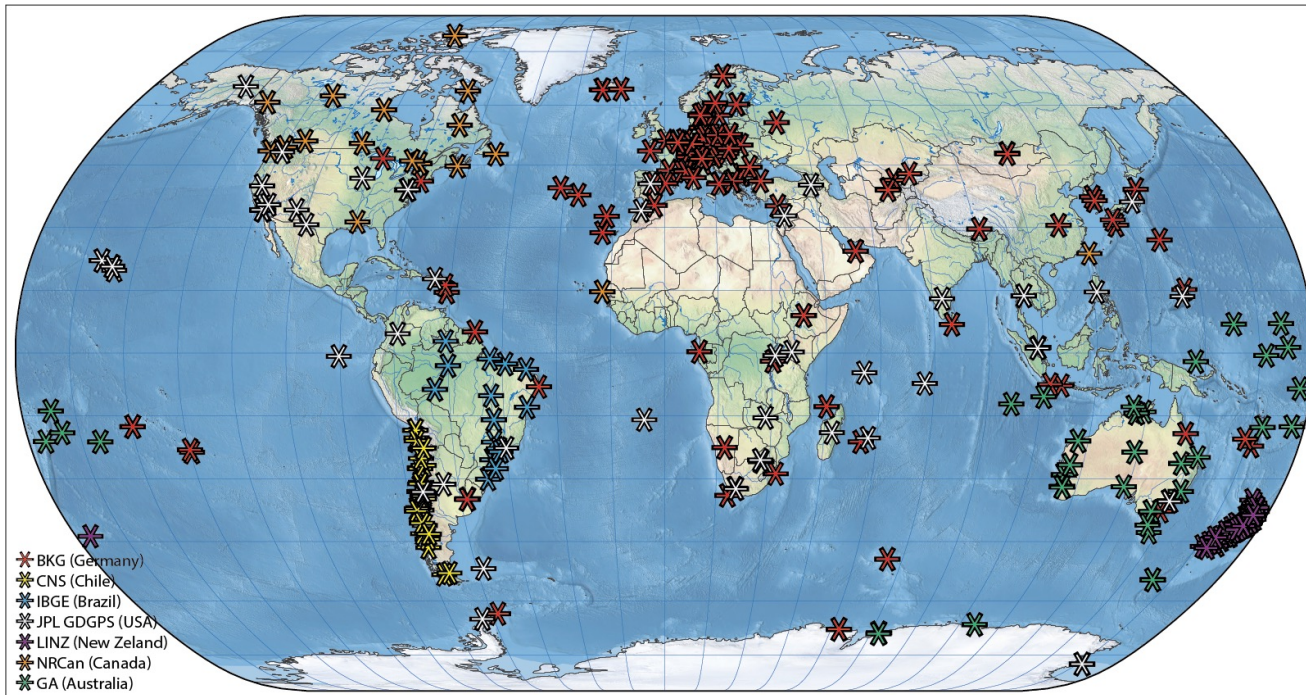


- ~20% growth per year in volume
- ~16TB per year in 2023
- High-rate, daily & products main drivers
- ~80TB of GNSS data by end of 2021

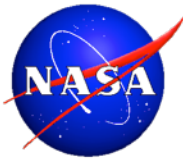
CDDIS - GNSS Real-Time



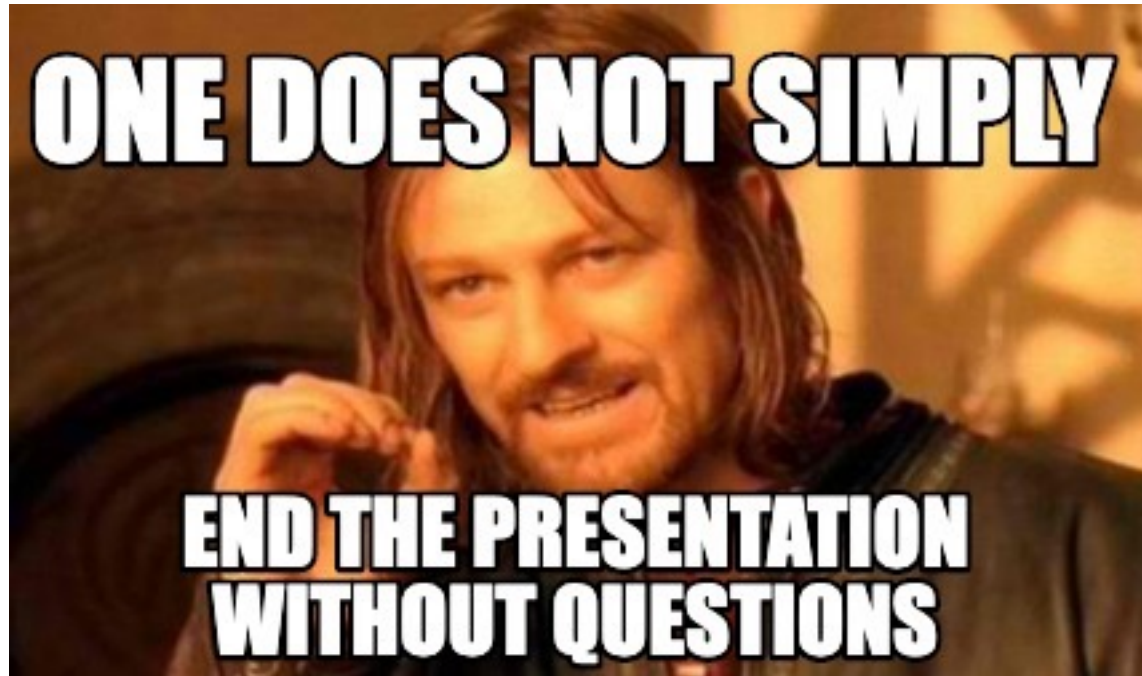
- All servers converted over to virtualization
- ~500GB daily transfer
- Record 15K simultaneous connections in May
- Average latency ~1.5 seconds
- 387 data streams – All the IGS streams plus exclusive NASA & Chile
- 30 products streams



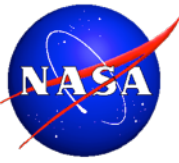
Upcoming Changes & Improvements



- Consolidation of high-rate data – 1 December 2021
- User Working Group (UWG) established
- Decrease latency to less than 30 seconds from file acceptance to availability in the archive
- CDDIS forum for user self help: <https://forum.earthdata.nasa.gov>
- New MEaSUREs data sets coming online: <http://sopac-csrc.ucsd.edu/index.php/measures-2/>
- Commercial GNSS data buys – Spire GNSS data
 - <https://earthdata.nasa.gov/esds/csdap/commercial-datasets>
- Cloud activities
 - DORIS in 2022
 - GNSS not until 2023

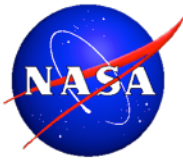


Any help and questions:
support-cddis@earthdata.nasa.gov

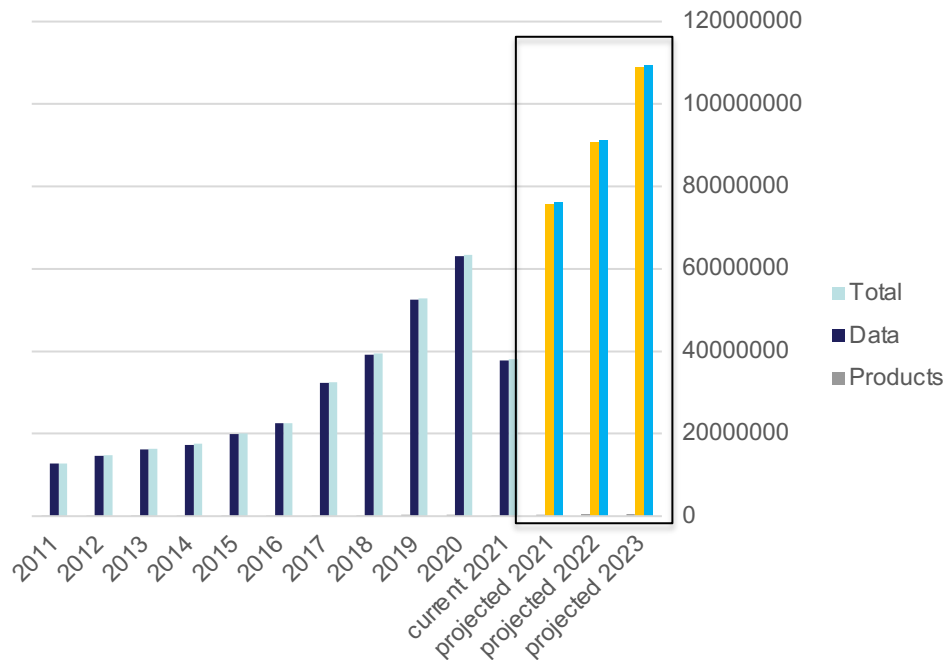


Backup

The Full Archive

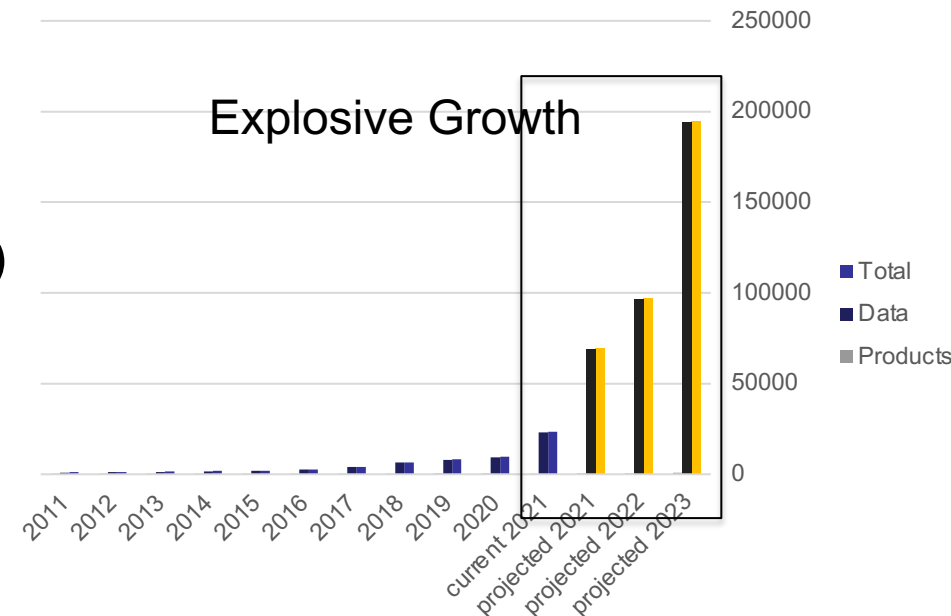


CDDIS File Count



- 800% growth since 2011
- 2011 = 50M, 2021 = ~400M
- 100M+ files per year by 2023

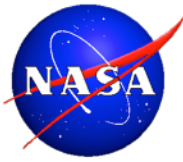
CDDIS Volume (GB)



- Archive will double in 2021 (~150TB)
- ~200TB per year in 2023
- New VLBI dataset - level 1A driver

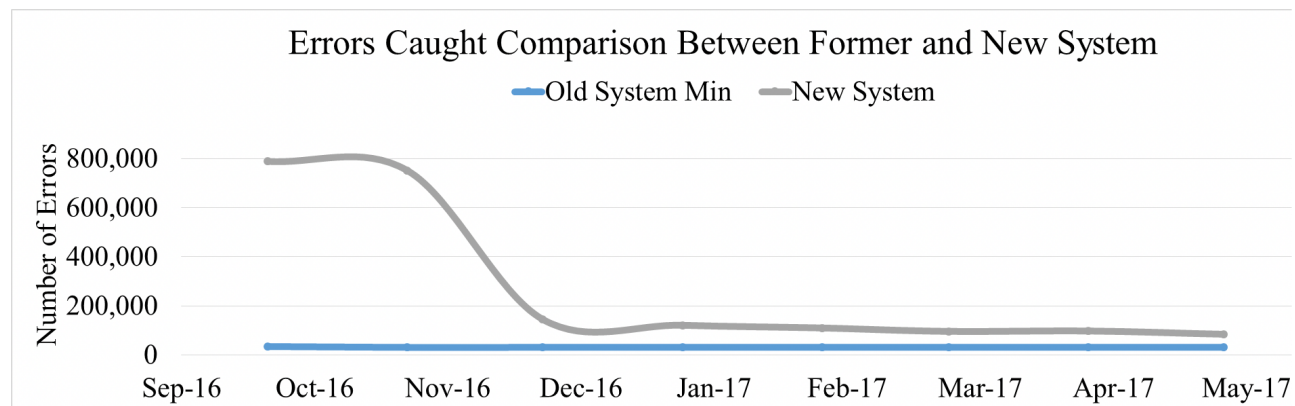


GNSS Quality Control (QC) History



- Prior 2016 very little ingest QC
 - Minimal name checking
 - No real checks for empty files, compression, etc
 - Essential just accept file and put into the archive
- Parallel Operations – Legacy and New
 - 22% of all incoming files had errors - ~27K files per day

- 2016 all new ingest software developed
 - All files are opened & checked with a core QC package set
 - 5-minute processing time – typically less than 120 seconds
 - Metrics collection on ALL incoming files
 - Less than 0.01% file rejects from providers
 - https://web.cddis.eosdis.nasa.gov/docs/2017/GNSS_QC_Poster_20170703.pdf

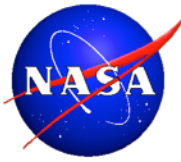


GNSS QC Package & Error Stats



- Basic file checks
 - Antivirus
 - Empty file
 - Unknown file name
 - File size
- File name checks
 - Marker name matches filename
 - RINEX naming scheme
 - Minute check (high-rate only)
 - Correct case for RINEX type
- Date/time checks
 - Future dates
 - Older than 2 years
 - Date matches filename
 - Invalid day of year
 - Unaccepted data interval
- Compression issues
 - Unexpected end of file errors
 - Corrupt input errors
 - Unknown or unexpected compression type
- Header content checks
 - RINEX version check
- Current errors stats
 - Fatal errors
 - ~0.005% or ~100 files per day
 - Warning
 - ~8.5% or ~17,000 files per day
 - ~92% of files without warning
 - below the CDDIS 98% threshold we like to see

Error Stats

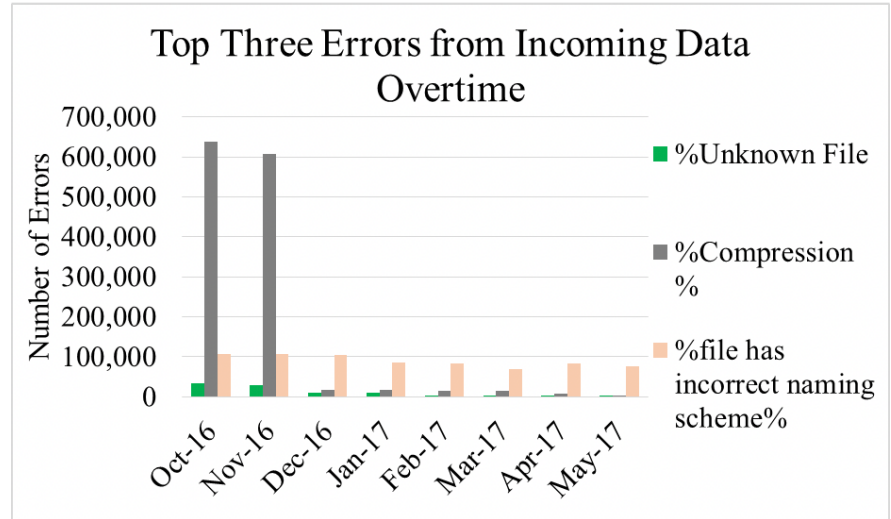


■ Historical - three main categories

- Unknown file types
- Compression issues
- Incorrect naming schemes

■ Current errors status

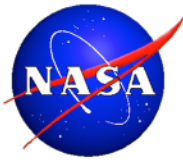
- Same 3 main errors
- Fatal errors
 - ~0.05% or ~100 files per day
- Warning
 - ~8.5% or ~17,000 files per day
- ~92% of files without warning
 - Below the CDDIS 98% threshold we like to see



■ Provider reports

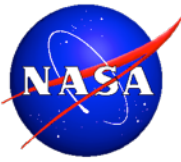
- Full stats on every provider is available
- Could put results on the web

File Outcomes



- Only 3 outcomes for every file
- Accepted – passes all tests and is placed in the archive
- Warnings
 - Correct in real-time if possible
 - Compression
 - Case issues
 - Placed in quarantine if unable to fix
- Fatal errors – predetermined fatal errors
 - Placed in quarantine
 - Contact provider

Not Mission Focused



IGS INTERNATIONAL
GNSS SERVICE

