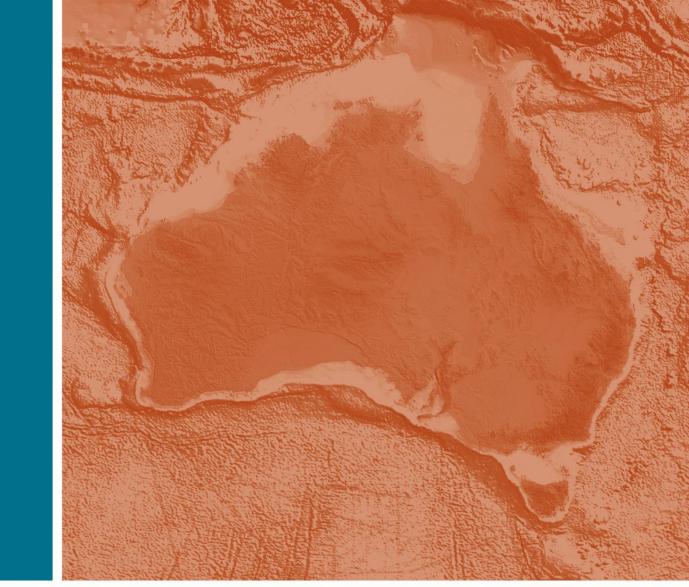**Australian Government**
Geoscience Australia

Discovering the Geoscience Australia GNSS Data Repository

Brandon Owen, Ted Zhou, Lazar Bodor, Frank Fu, Mohammad Choudhury, Jonathan Mettes and Ryan Ruddick

Abstract

Geoscience Australia (GA) maintains one of the most complete public repositories of Global Navigation Satellite System (GNSS) data from continuously operating reference stations across the Asia-Pacific region. The data sets found in this repository define Australia's national datum, contribute to the IGS and support numerous academic, government and industry projects. In a continuous effort to modernise our data centre workflows and procedures, we have developed a new event-driven GNSS data management system hosted in Amazon Web Services. This new data repository system allows us to move away from relying on file transfer protocol (FTP), letting users discover, interact and extract over 100 million RINEX files using secure HTTP APIs.

GNSS data repository overview

Until now GNSS data files were stored on Network Attached Storage (NAS) at GA. Data submission and retrieval was via FTP, which is insecure and cumbersome for users. For our new data repository we chose to architect it specifically for the Cloud, in this case Amazon Web Services (AWS). By combining Amazon S3, Amazon DynamoDB and Event Store we were able to create a highly available and scalable system.

Data submission and retrieval have been changed from FTP to secure HTTPS APIs. RINEX headers are now also validated and corrected against our Site Metadata Database using GeodesyML.
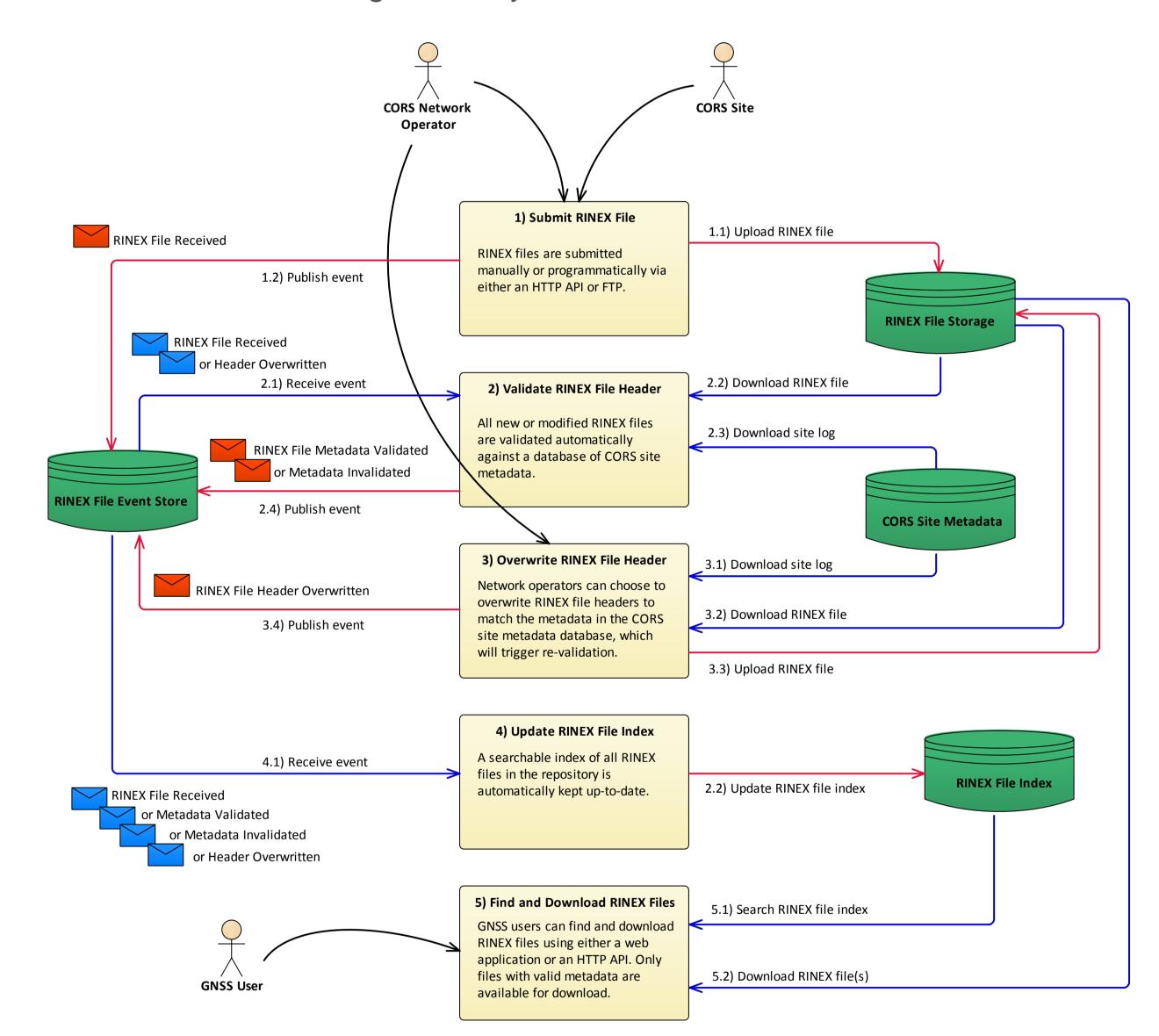

Figure 1: Data repository process diagram showing the events and interactions between a submitted RINEX file and the various components.

End-to-end performance, from submission & validation to download, is ~1.3 seconds under normal load of 300 files every 15 minutes and 1 file per second. Under a larger bulk load of 431,000 files across 20 minutes, the end-to-end performance is 32 seconds.

Each component of the repository produces logs that are stored in ElasticSearch. This allows us to holistically observe the system and identify issues with processes. This information can then be used to inform decisions around provisioning resources and troubleshooting issues at remote stations.

Event-driven model

The GNSS data repository is built on a highly-available event-optimised database. The Event Store database stores data as append-only series of immutable events over time. These stored events are all interactions the system makes with RINEX files, such as submission, normalisation, metadata validation and correction. Using immutable and append-only series of events enforces the integrity of the system, and provides a reproducible ledger of history for RINEX files.

GNSS Data

GNSS data files can now be searched and retrieved from GA's new GNSS data repository web interface located at: https://search.archive.geodesy.ga.gov.au/.
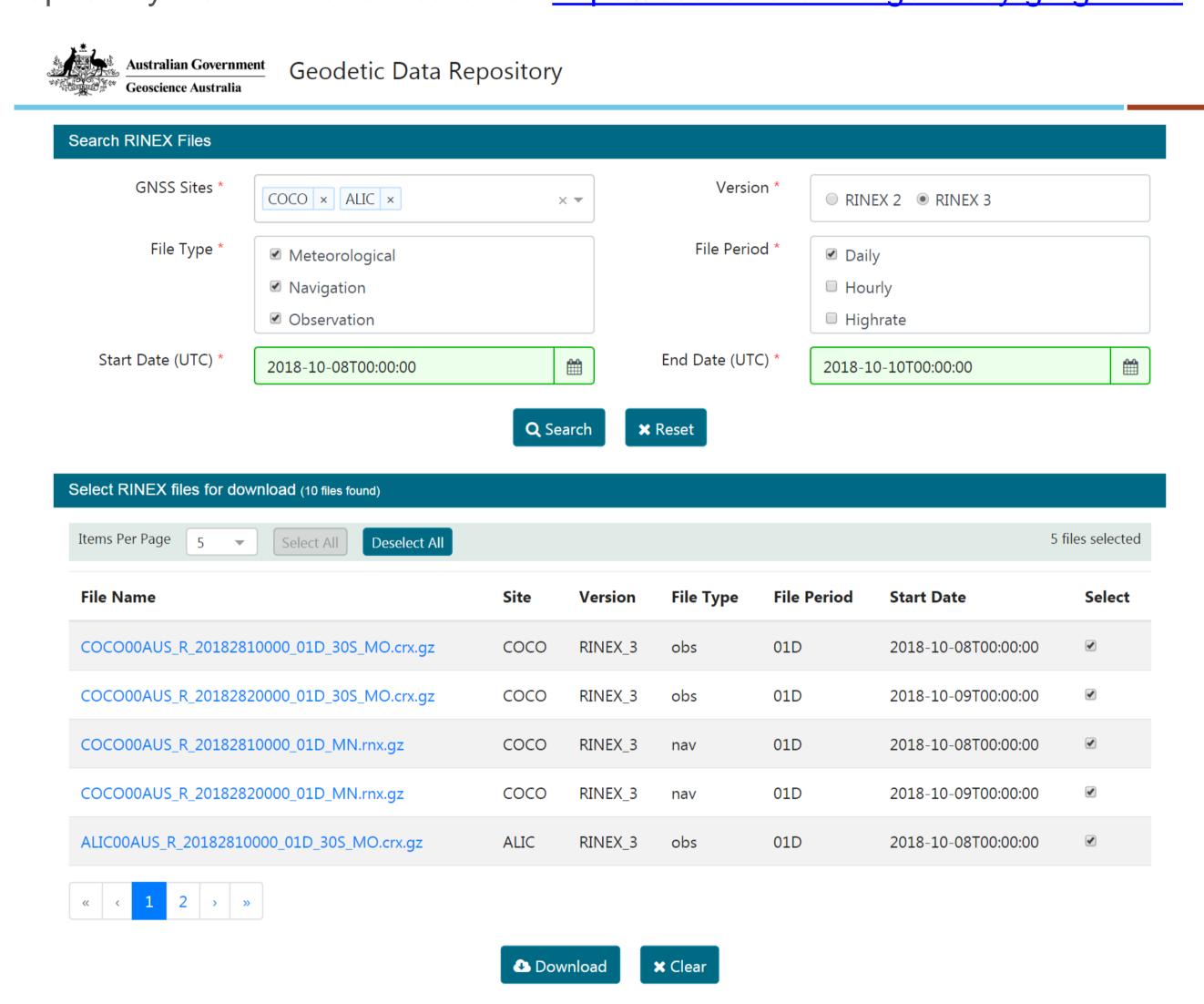

Figure 2: Web interface for searching and downloading GNSS data

```
# file download
curl 'https://api.archive.geodesy.ga.gov.au/rinexFiles/?stationid=COCO,ALIC&startdate=2018-1
0-08T00:00:00&enddate=2018-10-08T00:00:00&metadatastatus=metadata_valid&rinexversion=3&file
type=obs&fileperiod=01D'

# authenticate
authentication=$(curl -s --user "${OpenAM_ID}:${OpenAM_password}" \
  --data "grant_type=password&username=${user}&password=$(password}&scope=openid profile" \
  'https://prodgeodesy-openam.geodesy.ga.gov.au/openam/oauth2/access_token?realm=/')
jwt=$(echo "$authentication" | jq -r .id_token)

# file submission
curl -i -XPUT -H "Content-type: application/octet-stream" \
  --data-binary @"${input_file}" -H "Authorization: Bearer ${jwt}" \
  "https://api.archive.geodesy.ga.gov.au/submit/$(basename ${input_file})"
```

Figure 3: Example commands for downloading and uploading files from GNSS data repository API

GNSS data files can also be uploaded and downloaded through the repository using the repository HTTP API at: https://api.archive.geodesy.ga.gov.au/.
This API enables convenient machine-to-machine automation of batch workflows via command line tools such as `curl` or any scripting languages such as `Shell`, `Perl` or `Python`.

The repository uses improved and widely supported `gzip` compression format for RINEX2 and RINEX3 files, instead of the prior `LZW` format.

Metadata Assurance

GA maintains a site metadata database with configurations for all its sites, using automated workflows built on top of GeodesyML to ensure it is up-to-date. Metadata inside any RINEX file submitted to the repository are validated against the site metadata database as a single point of truth. Inconsistencies are flagged and resolved by overwriting the RINEX file headers with the up-to-date receiver configurations.

Conclusions

GA's new GNSS data repository has allowed us to replace legacy storage solutions and linear workflows with a modern system architected specifically for the Cloud. This new repository provides a scalable template upon which we can build flexible and robust workflows to utilise our GNSS datasets.

The source code for the infrastructure of our GNSS Data Repository is Open Source and can be found at:
https://bitbucket.org/geoscienceaustralia/geodesy-archive/src/master

GEOSCIENCE AUSTRALIA
APPLYING GEOSCIENCE TO AUSTRALIA'S MOST IMPORTANT CHALLENGES

For Further Information: Ryan Ruddick
Email: gnss@ga.gov.au
Ph: +61 2 6249 9426 **Web:** www.ga.gov.au