Proposal

Project Title	2018-04: Adopt *gzip* as the preferred compression utility for all IGS data and products
Proposed By	Ryan Ruddick and Michael Moore, Geoscience Australia Ignacio Romero, Infrastructure Committee Chair

Overview

Problem	• To reduce file transmission time and storage costs it has been common practice to compress GNSS data and products. • The need for compression and the compression algorithm has not been strictly defined by the IGS, but the RINEX format version 2 recommends that the *compress* utility be used. • The *compress* utility applies the LZW algorithm. This algorithm is no longer the most efficient way to compress files, as such support and availability on modern systems is limited. • The RINEX format version 3 moves away from recommending the *compress* utility and recommends is compression is needed the use of either *gzip*, *bzip* and *zip*. • Within the IGS community *gzip* has already been adopted as the preferred option for RINEX version 3 files and the Analysis Centres (ACs) are planning to submit their Repro3 products using *gzip*. • Due to reduced support for *compress* many local and regional data centres have already begun adopting *gzip* as the preferred format.
Proposal	• This proposal is to formally adopt (or recognise) *gzip* as the preferred compression format for all IGS data and product files. • The proposal is that the changes should occur primarily within the IGS Data Centres. Station Operators (SOs) and ACs will be encouraged to change their submission processes over time. • It should be noted that the *gzip/gunzip* tool provided on most Unix/Linux systems is capable of decompressing *compress* (*.Z) files.

The following is a list of what is required from each of the IGS groups:

Station Operators (SOs)

SOs typically submit RINEX files either direct from a receiver, from commercial software (such as Leica Spider) or from a local data centre. Changes to these workflows could be difficult due to the number of SO's and costly for the SOs.

As such:

- SOs will continue to submit RINEX files using a valid compression format (*gzip, compress, bzip, zip*).
- SOs will be encouraged to move toward the submission of RINEX 3 for all data products. The lack of RINEX 3 data within the IGS network is holding back the ACs from obtaining multi-GNSS solutions.
- SOs will be encouraged to use *gzip* as the preferred format.

Data Centres (DCs)

DCs are the access points for all IGS data and products, as such it is important that there is consistency in the formats and standards used. There will need to be some level of coordination across the DCs to ensure that any change is made simultaneously to avoid sync issues. While it is recognised that the normalisation of historic datasets to *gzip* is a large effort, it is important to ensure consistency and interoperability.

As such:

- By 1 June 2020, DCs will change their ingestion workflows to accept data and product files in any of the valid compression formats (*gzip, compress, bzip, zip*) and normalise these files to *gzip*. From this data there will be no need for DCs to continue to provide *compress (*.Z)* files.
- LDCs and RDCs will be free to choose whether or not they normalise historic (prior to 1 June 2020) datasets to *gzip*.
- GDCs will coordinate a date to coordinate conversion of historic (prior to 1 June 2020) datasets to *gzip*.

Analysis Centres (ACs)

ACs are users the IGS data and providers of the products files. Through Repro3 there are already moves by ACs to provide products using *gzip* so the transition should not prove to difficult.

As such:

- By 1 June 2020, ACs will need to ensure that their software/scripts are able to download and ingest RINEX version 2 data files in *gzip*.

	• From 1 June 2020, ACs will be encouraged to submit products to the DCs using *gzip* as the preferred format. Infrastructure Committee (IC) The IC will: • Coordinate the transition by providing IGS-wide notifications and information. • Encourage the SOs to transition to RINEX version 3 data files. • There is no mention of compression format in the site guidelines so no changes will be required. RINEX Working Group • Rinex 3 data format already encourages *gzip* as the recommended compression method so no further change is necessary; ftp://ftp.igs.org/pub/data/format/rinex304.pdf • Rinex 2.11 data format in Section 4 recommends the use of the UNIX “compress” and the “.Z” extension. An addendum recommending *gzip* and “.gz” extension to the Rinex 2.11 will be issued and placed in the IGS website by the format definition; ftp://igs.org/pub/data/format/rinex211.txt
Resources	• To assist in testing software/scripts/workflow the Geoscience Australia (GA) Regional Data Centre will make available a repository of RINEX data using *gzip*. *ftp://data.gnss.ga.gov.au or API call to data.gnss.ga.gov.au*
Timeframe	Completion by 1 June 2020. Proposed Schedule 1 March 2020 – Proposal received by IC. 5 March 2020 – Discussion on proposal during IC teleconference. 13 March 2020 – Revised proposal distributed to wider community for comment. 30 March 2020 – Feedback reviewed and final proposal agreed upon. 1 April 2020 – Notification and timeline send to mailing lists. | Between these dates a there will need to be coordination and information led by the IC. |

	1 June 2020 – DCs provide all data and product files using *gzip* compression.