Modernization of the EPN Historical Data Center

C. Bruyninx, J. Legrand, F. Bamahry, A. Miglio, A. Fabian

Royal Observatory of Belgium

EUREF Symposium, 01-03/06/2022, Zagreb, Croatia

Outline

- Introduction & Motivation

- Concept of modernized HDC

- Progress

- Conclusions

Introduction

- Monitoring availability and latency of hourly and daily RINEX 2/3 files at BEV and BKG regional data centers
  - Web pages with data latency plots
  - Emails to station managers when station data are missing or late
  - Help the user to retrieve EPN daily RINEX files (calendar web page on EPN CB website)

- Monitoring metadata in RINEX headers of daily RINEX 2/3 files
  - Ask station manager to correct inconsistencies between RINEX header and site log

- Monitoring of data quality of daily RINEX 2/3 files
  - Web pages with data quality plots
  - Emails to station managers when station data is degrading

Introduction- EPN Central Bureau monitoring

Requires to scan directory contents

- Monitoring availability and latency of hourly and daily RINEX 2/3 files at BEV and BKG regional data centers
  - Web pages with data latency plots
  - Emails to station managers when station data are missing or late
  - Help the user to retrieve EPN daily RINEX files (calendar web page on EPN CB website)

Requires download of the RINEX data

- Monitoring metadata in RINEX headers of daily RINEX 2/3 files
  - Ask station manager to correct inconsistencies between RINEX header and site log

- Monitoring of data quality of daily RINEX 2/3 files
  - Web pages with data quality plots
  - Emails to station managers when station data is degrading

EPN CB Historical Data Center (HDC)

- Contains the daily RINEX files used for EPN CB monitoring

- Published with a delay of 1 week (allows to iterate with station manager to correct errors)

- In support of EPN reprocessing

- RINEX header changed to be conform with site log
  - Very roughly : If a site log update includes a receiver change of 2 years ago, then all RINEX files of that station for the last 2 years are rewritten with an updated header
  - However, sometimes, RINEX files are touched even if it is not necessary
  - Manual process, so not error free

- Stored numerous flags indicating RINEX files that have a specific problem (e.g. early RINEX 3 files)
  - BUT, flags require a cleanup to be consistent from 1996 to now + decision of which flags should block publication of data file

Origin of daily RINEX data used at EPN CB

A. Daily RINEX data provided by the station manager at the moment a station is proposed to the EPN

Data from before the operational daily RINEX data upload to EPN RDC (BEV and BKG) has started

15% of the HDC data holdings !

B. Operational daily RINEX data downloaded from the EPN data centers

Download of the operational RINEX data (1)

Data flow principle in the EPN → Redundancy (data available at multiple data centers)

- But : EPN CB can not check all the data in all the data centers → need to make a selection based on the DC scans

- Major difficulty: different data centers do not necessary hold the same data

In practice:

✓ Scan all data centers to get file info (last update date + file size)

✓ Download the most recent RINEX file

✓ Perform checks

✓ Re-download RINEX file when newer RINEX file appears in a data center

→ Allows to take into account corrected RINEX files that were re-uploaded by the station managers

Download of the operational RINEX data (2)

<u>Early years</u>

Shopping all over EPN local data centers, regional data centers (BKG, BEV), IGS data centers

<u>Recent years</u>

EPN regional data centers BKG and BEV, and if data missing, after a couple of days, check other local EPN DC and global IGS DC

But, info on which DC the data were downloaded from was not stored in EPN CB database

Outline

- Introduction & Motivation

- Concept of modernized HDC

- Progress

- Conclusions

Concept of Modernized HDC (1)

Prepare the EPN CB HDC for certification as a trustworthy data repository (CoreTrustSeal, https://www.coretrustseal.org/) + FAIR data principles

→*Provide to the users the information on the origin of a RINEX file as well as the changes done to the file (= provenance information)*

Full transparency to the users that download data from HDC

- Only download data from BEV and BKG

- Need to provide users information on
    - Where did we get the RINEX files from?
    - Which changes did we make on the RINEX file? And when?
    - What was the original content of the RINEX file header?
    - What problems did we encounter when reading the RINEX file?

- Need to comply with EU GDPR

Concept of Modernized HDC (2)

- RINEX file not published if
  - No info on receiver or antenna in site log for a specific day
  - Day with a change of equipment or firmware change

- Provide access to RINEX files through ftp, https, API

- Need to be able to provide to users a list of all files that were changed during specific time frame

- Need to perform more rigorous checks on RINEX files and be able to correct automatically more errors

```
3.4   Receiver Type           : SEPT POLARX4TR
      Satellite System        : GPS+GLO+GAL
      Serial Number           : 3001376
      Firmware Version        : 2.3.3
      Elevation Cutoff Setting : 0 deg
      Date Installed          : 2012-01-31T14:00Z
      Date Removed            : 2012-03-05T14:00Z
      Temperature Stabiliz.   : 18.0 +/- 0.2
      Additional Information  :

3.5   Receiver Type           : SEPT POLARX4TR
      Satellite System        : GPS+GLO+GAL
      Serial Number           : 3001376
      Firmware Version        : 2.3-tst120216r34012
      Elevation Cutoff Setting : 0 deg
      Date Installed          : 2012-04-05T16:00Z
      Date Removed            : 2012-09-24T11:00Z
      Temperature Stabiliz.   : 18.0 +/- 0.2
      Additional Information  :
```

More rigorous checking of RINEX errors

- Store information in original RINEX header in internal database
  - Programs used to generate RINEX
  - Observer/agency
  - Marker name/number
  - Receiver/antenna type and serial number
  - …
- Store results of RINEX verifications in internal database (and can be consulted later on)
  - Incorrect/unknown RINEX format
  - Missing mandatory header lines
  - Misformatted header lines
  - Duplicate header lines
  - Inconsistent information with site log
  - Sampling rate
  - Date file name / date 'FIRST OBSERVATION' / timestamp of actual observations
  - …

More rigorous checking of RINEX errors

- Store information in original RINEX header in
  - Programs used to generate RINEX
  - Observer/agency
  - Marker name/number
  - Receiver/antenna type and serial number
  - ...

- Store results of RINEX verifications in internal
  - Incorrect/unknown RINEX format
  - Missing mandatory header lines
  - Misformatted header lines
  - Duplicate header lines
  - Inconsistent information with site log
  - Sampling rate
  - Date file name / date 'FIRST OBSERVATION' / times...
  - ...

714	error	Misformatted <YYYY MM DD HH MM SS.SSS> in header string "TIME OF FIRST OBS" has been corrected
715	error	Misformatted <YYYY MM DD HH MM SS.SSS> in header string "TIME OF LAST OBS" has been corrected
802	error	Misformatted header label "PGM / RUN BY / DATE" has been corrected
805	error	Misformatted header label "MARKER NUMBER" has been corrected
809	error	Misformatted header label "ANT # / TYPE" has been corrected
814	error	Misformatted header label "TIME OF FIRST OBS" has been corrected
902	error	Misformatted header label "PGM / RUN BY / DATE" has been corrected
908	error	Misformatted header label "REC # / TYPE / VERS" has been corrected
909	error	Misformatted header label "ANT # / TYPE" has been corrected
910	error	Misformatted header label "APPROX POSITION XYZ" has been corrected. Misformatted values for "APPROX POSITION XYZ" have been corrected using information from <site log>
911	error	Misformatted header label "ANTENNA: DELTA H/E/N" has been corrected Misformatted values for "ANTENNA: DELTA H/E/N" have been corrected using information from <site log>
914	error	Misformatted header label "TIME OF FIRST OBS" has been corrected using information from <site log> Misformatted values for "ANTENNA: DELTA H/E/N" have been corrected using information from <site log>
915	error	Misformatted header label "TIME OF LAST OBS" has been corrected Misformatted values for "ANTENNA: DELTA H/E/N" have been corrected using information from <site log>
918	error	Misformatted header label "GLONASS COD/PHS/BIS" has been corrected
1001	error	Removed line with duplicate header label "RINEX VERSION / TYPE"
1002	error	Removed line with duplicate header label "PGM / RUN BY / DATE"
1004	error	Removed line with duplicate header label "MARKER NAME"
1005	error	Removed line with duplicate header label "MARKER NUMBER"
1007	error	Removed line with duplicate header label "OBSERVER / AGENCY"
1008	error	Removed line with duplicate header label "REC # / TYPE / VERS"

More rigorous checking of RINEX errors

- Store information in original RINEX header in internal database
  - Programs used to generate RINEX
  - Observer/agency
  - Marker name/number
  - Receiver/antenna ty
  - ...

- Store results of RINEX
  - Incorrect/unknown
  - Missing mandatory
  - Misformatted head
  - Duplicate header lines
  - Inconsistent information with site log
  - Sampling rate
  - Date file name / date 'FIRST OBSERVATION' / timestamp of actual observations
  - ...

✓ Improved automatic correction of a lot of errors

✓ Classification of errors
  Error can be corrected → RINEX file is published
  Error cannot be corrected → RINEX file is not published

✓ In all cases: station manager will be informed

Improved tracking of changes on RINEX files

```
     2.11                OBSERVATION DATA     M                         RINEX VERSION / TYPE
GN-RINEX 1.3           Geo++ GmbH            22-MAY-22 00:59           PGM / RUN BY / DATE
START OF RINEX header check by ROB          20220523 080524 UTC COMMENT
 with respect to                                                   COMMENT
 https://gnss-metadata.eu/v1/sitelog/exportlog?id=RAMO00ISR COMMENT
RINEX header corrections                                           COMMENT
 MARKER NAME    RAMO Mizp               -> RAMO                     COMMENT
 OBSERVER removed to comply with GDPR                              COMMENT
END OF RINEX header check by ROB                                   COMMENT
RXO2RXO 1.11w          Geo++ RX Converter  2022-05-23 06:34       COMMENT
RAMO                                                               MARKER NAME
20703S001                                                          MARKER NUMBER
                                                                   OBSERVER / AGENCY
                       Survey of Israel                            OBSERVER / AGENCY
02628                  JAVAD TRE_G3TH DELTA3.6.14                  REC # / TYPE / VERS
CR20014604             ASH701945B_M     SNOW                       ANT # / TYPE
  4514721.8549  3133507.8433  3228024.6792                         APPROX POSITION XYZ
        0.0030        0.0000        0.0000                         ANTENNA: DELTA H/E/N
      1     1                                                      WAVELENGTH FACT L1/2
      9    C1    L1    D1    P2    L2    D2    P1    S1    S2# / TYPES OF OBSERV
     30                                                            INTERVAL
   2022     5    22     0     0   0.0000000     GPS               TIME OF FIRST OBS
     18                                                            LEAP SECONDS
                                                                   END OF HEADER
```

Improved tracking of changes on RINEX files

```
     3.04                OBSERVATION DATA     M: MIXED              RINEX VERSION / TYPE
Spider V7.7.1.9072                          20220517 235956 UTC PGM / RUN BY / DATE
START OF RINEX header check by ROB          20220518 171524 UTC COMMENT
 with respect to                                              COMMENT
 https://gnss-metadata.eu/v1/sitelog/exportlog?id=STNB00PRT COMMENT
RINEX header corrections                                      COMMENT
 MARKER NAME    STNB                   -> STNB00PRT           COMMENT
 ANTENNA_SN                            -> 200160              COMMENT
 ANT TYPE       LEIAT504GG             -> LEIAT504GG     NONE COMMENT
END OF RINEX header check by ROB                              COMMENT
Spider V7.7.1.9072                          20220517 235956 UTC PGM / RUN BY / DATE
THIS RINEX FILE IS CREATED FROM RTCM V3.0 DATA               COMMENT
SNR is mapped to RINEX snr flag value [1-9]                  COMMENT
LX:      < 12dBHz -> 1; 12-17dBHz -> 2; 18-23dBHz -> 3       COMMENT
      24-29dBHz -> 4; 30-35dBHz -> 5; 36-41dBHz -> 6         COMMENT
      42-47dBHz -> 7; 48-53dBHz -> 8; >= 54dBHz -> 9         COMMENT
Product                                                       COMMENT
STNB00PRT                                                     MARKER NAME
83402M001                                                     MARKER NUMBER
                                                              OBSERVER / AGENCY
3055017              SEPT POLARX5          5.4.0             REC # / TYPE / VERS
200160               LEIAT504GG       NONE                  ANT # / TYPE
  5134903.4968 -1558816.5903  3436531.1590                  APPROX POSITION XYZ
        0.0000        0.0000        0.0000                  ANTENNA: DELTA H/E/N
G    8 C1C L1C D1C S1C C2W L2W D2W S2W                      SYS / # / OBS TYPES
R    8 C1C L1C D1C S1C C2C L2C D2C S2C                      SYS / # / OBS TYPES
E   12 C1C L1C D1C S1C C5Q L5Q D5Q S5Q C7Q L7Q D7Q S7Q      SYS / # / OBS TYPES
C   12 C2I L2I D2I S2I C6I L6I D6I S6I C7I L7I D7I S7I      SYS / # / OBS TYPES
DBHZ                                                          SIGNAL STRENGTH UNIT
   30.000                                                     INTERVAL
 2022     05     17     00     00    0.0000000       GPS      TIME OF FIRST OBS
```

Outline

- Introduction & Motivation

- Concept of modernized HDC

- Progress

- Conclusions

Development of new HDC backend

- Started in Summer of 2021

- Complete revision of software
  - To download RINEX data
  - To verify and correct RINEX headers

- Under internal testing on new incoming RINEX data at BEV and BKG since Nov. 2021

- Original plan:
  - before Fall 2022:
    - Re-download all data from BEV and BKG to fill EPN CB internal database with all provenance information
    - Apply rigorous RINEX checks

Major delays

- But, too many changes at RDC last months made it impossible to REbuild reliable database for all historical data

- New HDC needs persistent (=long-term and stable) links to data files in RDC :
  - ftp:// will be discontinued at some point
  - https:// should be used
  - But BEV, BKG archives are presently not complete for https (BKG history missing since outage, BEV ftp ≠ https)

- Different approach for handling RINEX files at BKG and BEV
  - BKG quarantines files with significant RINEX headers (presently under discussion)
  - BEV does not quarantine

- April 13, 2022: Meeting ROB, BEV & BKG to discuss harmonization and way forward

Conclusions & Outlook

- ROB is preparing to modernize the HDC taking into account the requirements to become a trustworthy data repository

- Progress is presently on hold due to ongoing changes at RDCs

- Need to wait for RDCs to stabilize and finish developments

Cite this presentation as:

C. Bruyninx, . Legrand, F. Bamahry, A. Miglio, A. Fabian (2022), Modernization of the EPN Historical Data Center,
Presented at EUREF symposium, 1-3 June, 2022, online from Zagreb, Republic of Croatia

Contact
Royal Observatory of Belgium
fair-gnss@oma.be
https://fair-gnss.oma.be/
Brussels
BELGIUM
Twitter: @be_GNSS

EUREF Symposium, 01-03/06/2022, Zagreb, Croatia

ROYAL
OBSERVATORY
OF BELGIUM

euref

EUREF2022