Symposium of the IAG Subcommission for Europe (EUREF)
Toledo, Spain, 4 - 7 June, 2003

Re-designed implementation of GOP EUREF data center

J. DOUSA AND P. SOUCEK[1]

1. Introduction

The Geodetic Observatory Pecny (GOP) EUREF local data center (Dousa, 2001) was officially included into the EUREF Permanent Network (EPN) in May 2002, but has been routinely active already from November 2000. Its original goal was the support of near real-time (NRT) analyses at GOP – a) the processing of selected European permanent GPS sites for exploitation in meteorological applications (Dousa, 2003a) and b) the processing of global solution for ultra-rapid orbit determination (Dousa, 2003b). The better distribution of near real-time data flow in Europe was primary motivation for setting up our data center. Later, expanding the GOP data center by additional services for external users, we have been motivated for re-implementation of the original system. More efficiency of new system was expected with respect to reducing the redundancy in download and data storage, while protecting the current data latency.

Additionally, some other features were requested in new implementation: modular basis of the system, flexibility in configuration, applicability in other similar tasks except the official data center, integrated database and monitoring system for statistical archive, the possibility of data stacking into the longer time units, support incoming data and some others.

2. Content and layers of the data center

The core layer of the data center consists of the data files stemming on two sources: a) retrieved by center itself through either mirroring or other downloading procedure or b) incoming by the external providers. Besides the GPS data itself, the products and information useful for near real-time GPS analysis are collected and provided. More than 100 European and global sites with hourly GNSS data are available in the GOP data center. Most of them are accessible in the anonymous area of the service, but some are separated into a protected area, such as the data, for example, from sites of Met Office UK within the COST-716 NRT demonstration campaign (Marel, 2003). The same layer includes also the products of IGS, IERS, EUREF (orbits, clocks, realizations of the reference frame, etc) as well as the products from GOP analysis center or other institutes (NRT orbits, satellite clocks, coordinates, troposphere parameters, satellite/station crux-files and others).

Here, we can distinct two types of data and product files – firstly, the periodical files with reference epoch coded in the name and secondly, the epoch-independent files, which are characteristic mostly by irregular data flow. While the files of the first type could be retrieved either by mirror, active download or incoming upload, the files of a second type are typically convenient for persistent mirroring.

Additional layer of the GOP data center is identified as the services of data sets locally generated (e.g. concatenated daily RINEX files, actual broadcast orbits, checking summaries, regularly updated extraction information lists etc.).

The upper layer represents the access to all the data, products and services available in the user area through ftp, http or ssh protocols either by anonymous or selective (e.g. password or identity protection) access.

3. Background of old and new implementations

Main structure of the original system was written in Bourne again shell, while the new system is prepared in Perl scripting language. Both systems additionally uses the external software tools for basic tasks of the service – a) *wget (or alternative)* for getting the data, b) *compress/uncompress, rnx2crx/crx2rnx, gzip/gunzip* for converting the files and c) *teqc* from UNAVCO (Estey and Meertens, 1999) for format and quality checking of RINEX data. Finally, the new implementation exploits more software packages for other specific tasks of the service, such as d) *gmt, gnuplot* and *grace* for visualization of monitoring and statistics, e) *MySQL* for maintaining the database and f) *PHP/CGI* scripts for web-page interface.

While the old system consists of different programs for various actions, the new implementation model is fully integrated into a unique program taking advantage of Perl modules. The main modules are derived from the basic data center activities as defined in Table 1. The modules can be started independently through the unique program and possibly in the parallel runs. It is clear, that many actions could be then implemented within an hourly schedule in a very frequent run, Fig. 1.

[1] Research Institute of Geodesy, Topography and Cartography,250 66 Zdiby 98, Czech Republic,
Department of Advanced Geodesy, Czech Technical University in Prague, Thakurova 7, Prague 6, Czech Republic, e-mail: dousa@fsv.cvut.cz

The old system's main activity was started in large intervals due to its sequentially dependent performance (download/mirror, download checking, linking and cleaning). Only incoming data were additionally implemented in a fashion rather close to the new system. More projects can be now easily integrated into the new system taking advantage of its flexible configuration files.

4. Differences of new and old implementations

Figure 2 shows the scheme diagrams of internal operation of old and new implementations. When discussing both approaches, we focus especially on the main consequences.

The original system (top diagram, Fig 2) was mirroring various sources in a parallel fashion. The acceptance and rejection masks were applied for the files – so far we call it *masked mirror*. A redundancy of more sources (source1, source2, etc.) for every file was necessary and causes a high extra loading traffic with corresponding local mass storage. In this way a stable performance and good data availability was achieved. The downloaded files were checked only for its size and name and they were linked into the output directory area – *outdata*. The first existence of every file in mirror-directories (individual for any source) is only used. The system couldn't separate download itself from linking the file to output directory as well as the incoming or mirror data handling could not be integrated.

Besides the general method of masked mirror, the new data center implementation (bottom diagram, Fig 2) applies active *pointed download* for all periodical data and products.

Table 1: Modules of new implementation derived from the basic data center tasks.

Module	Task specification
Checking local contents	Generating data request
Getting data	Active data download, mirror
Data checking and converting	Applied for downloaded data as well as incoming data
Additional services on the local data sets	Extracting information lists, concatenating data, data quality monitoring etc.
Maintaining services	DC performance monitoring, writing log records, directory cleaning etc.

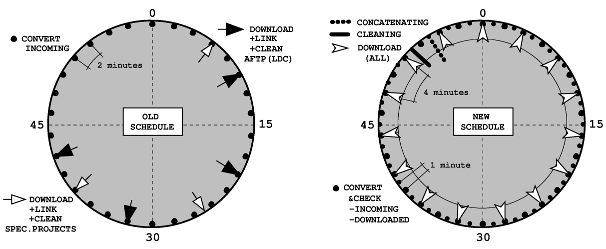

Figure 1: Old (left) and new (right) hourly schedule of basic action of NRT GOP data center.

The mirror is an optional method here, applied mostly for the non-peridical files only. 'Pointed' (download) means the unique requests for every missing (expected) file. This method reduces the data download and persistent data storage to its non-redundant minimum (see Fig 3). The files are always available in unified form in the official output area – *outdata*. They are provided there after format and quality checking, renaming and compression actions etc. As opposed to old implementation, the *outdata* area is a unique physical persistent storage for the data and is always checked for the content when a new request for download should be generated. Consequently, the *indata* directory is again only temporary area and internally, the data are handled here in the same way like in the *incoming* directory. The checking, converting, monitoring or other control procedures were thus unified for either type of data collection and started independently and more frequently from getting the data itself.

The final data/product unique repository (realized by the *outdata* area) together with the possibility of getting data from arbitrary period in past, it is good basis for generating derived files – e.g. concatenated daily data.

.

5. Special features of new implementation

The first special feature deals with protecting the data latency during a unique download, which strongly depends on the actual selection of the appropriate data source. We can simply describe it like a *source-history caching* system. It enables to benefit from the recent status of files downloaded from more possible sources. The caching database stores the information about date, latency, source and status of last getting request. Hence, it enables the *dynamic source adaptation* or maintaining the source priority at least.

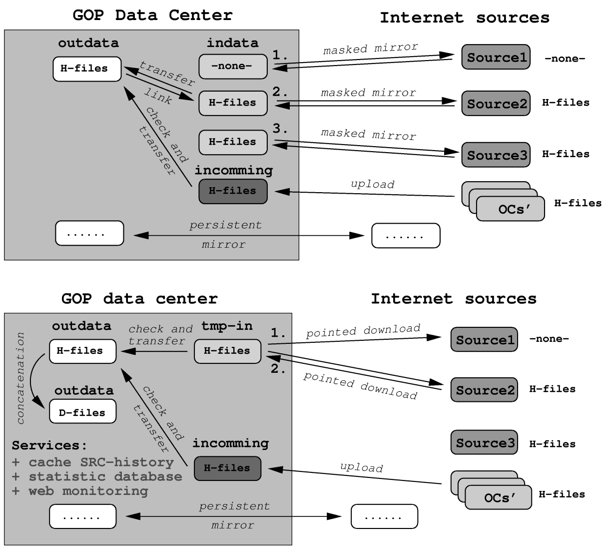

Figure 2: Old (top) and new (bottom) implementation schemes for GOP data center download.

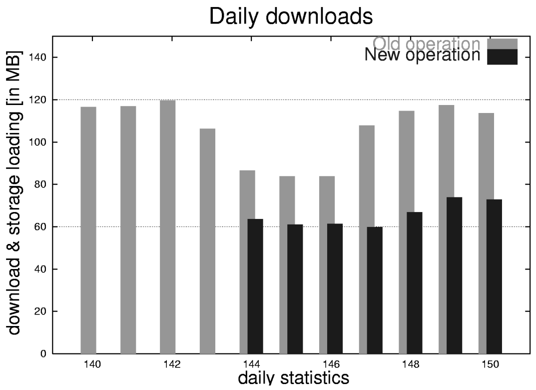

Figure 3: Decrease of redundancy in data storage and download during 10 days of comparative concurrent run for new and old system. The reduction of old system data statistics during days 144-146 was due to problems with downloads from CDDIS - many global sites were non-redundant during that period. About 120MB/day was standard data amount in the old implementation and only 60MB/day in new implementation (non-redundant minimum).

```
#-----------------------------------------------------------------------------------------
# NAME : TYP[;TYP] :  MASK[;MASK]               : SRC[;SRC]     : GRP : LOCAL-DIR         :
#-----------------------------------------------------------------------------------------

# hourly RINEX files
# ------------------
BRUS :RNX;HOURLY :brus*<doy><i>*;BRUS*<doy><I>* : GOP;CDDIS;BKG :AFTP :outdata/<yr><doy>/<I>:
GOPE :RNX;HOURLY :gope*<doy><i>*;GOPE*<doy><I>* : CDDIS;IGN;BKG :AFTP :outdata/<yr><doy>/<I>:
HURN :RNX;HOURLY :hurn*<doy><i>*;HURN*<doy><I>* : BKG;CDDIS;GOP :TOUGH:outdata/<yr><doy>/<I>:

# daily RINEX files
# -----------------
GOPE :RNX;DAILY  :gope*<doy>0*;GOPE*<doy>0*     : GOP;CDDIS;BKG :AFTP :outdata/<yr><doy>/0  :
GRAZ :RNX;DAILY  :graz*<doy>0*;GRAZ*<doy>0*     : GOP;CDDIS;BKG :AFTP :outdata/<yr><doy>/0  :

# orbits (SP3+ERP)
# ----------------
IGS  :SP3+ERP       :igs*<gw>*sp3*;igs*<gw>*erp*: GOP;CDDIS     :AFTP :orbits/<gw>          :
IGR  :SP3+ERP       :igr*<gw>*sp3*;igr*<gw>*erp*: GOP;CDDIS     :AFTP :orbits/<gw>          :
IGU  :SP3+ERP;ULTRA :igu*<gw>*sp3*;igu*<gw>*erp*: GOP;CDDIS     :AFTP :orbits/<gw>          :
# -----------------------------------------------------------------------------------------
```

Figure 4: Example of data configuration file for set of data. More configuration files could be generated and applied when starting the system. Beside the NAME and TYPE of datasets, usually defined in MASK of each line, sources (SRC) are set within an active order. GRP column defines the project group for the dataset access and finally LOCAL-DIR the final local repository. Standard wild symbols plus additional substituting date-masks are expanded to the actual operational setting.

```
#-----------------------------------------------------------------------------
# NAME:TYP[;subTYP]  :IP[;HOST]            :REMOTE-DIR                        :
#-----------------------------------------------------------------------------


GOP   :RNX            :pecny.asu.cas.cz     :gps/last10days<I>                 :
GOP   :RNX;HOURLY     :pecny.asu.cas.cz     :LDC/outdata/<yr><doy>/<I>         :

CDDIS :RNX;HOURLY     :cddisa.gsfc.nasa.gov :pub/gps/nrtdata/<yr><doy>/<hr> :
CDDIS :RNX;DAILY;GPS :cddisa.gsfc.nasa.gov :pub/gps/gpsdata/<yr><doy>/<yr>o:
CDDIS :RNX;DAILY;GLO :cddisa.gsfc.nasa.gov :pub/gps/gpsdata/<yr><doy>/<yr>g:
CDDIS :RNX;DAILY;NAV :cddisa.gsfc.nasa.gov :pub/gps/gpsdata/<yr><doy>/<yr>n:

CDDIS :SP3+ERP        :cddisa.gsfc.nasa.gov :pub/gps/products/<gw>             :
GOP   :SP3+ERP        :pecny.asu.cas.cz     :gps/orbits                        :
#-----------------------------------------------------------------------------
```

Figure 5: Sample of source definition file for mirror or download. One source can be applied for various types of data or products (e.g. when data are available in different directories). Like in the data definition file (Fig 4), the standard wild symbols as well as date-masks are substituted.

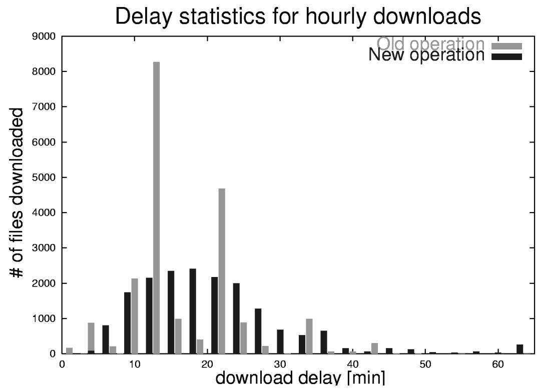

Figure 6: Data latency statistics for the old and new implementations show different pattern. It is clear, the old system was started in few high peaks (compare Fig. 1, left) and all the other data are 'incoming' or delayed due to a problem with slow access to any source. Scattered retrieves from the new implementation could be assessed also as an advantage, when the delays were not in total significantly increased and the sequential dependency for the file retrieves and maintaining is nearly reduced. So far we expect that the system is more resistant.

Usually, the periodical files are collected only during a limited period. Another feature is therefore a possible use of new implementation for any period and data intervals, as well as nearly arbitrary group of the requests. Together with flexible configuration files (Figs 4 and 5), it enables various operational setting and thus further potential applications. Figs 4 and 5 show a configuration samples for data and source definition files. More projects can be easily integrated into a unique scheduler of routine operation as clear from Fig 1 and 'GRP' setting in Fig 4.

Finally, the statistics and monitoring system is necessary due to a high amount of data handled in the GOP data center. Besides the standard EUREF/IGS tables for latency of RINEX files, the new system integrates the database records for any data available in the data center. This supports the independent statistics visualization and monitoring system, achievable through the web-access. By coloring symbols, for example, the latency information can be combined together with the source of data stemming from the adaptive pointed download.

6. Latency limitations - special call for direct upload

The mean data latency (15-20min) in our data center is a consequence of wide spectrum of provided data – global, regional and still only few local (national) sites. Most of the files are transferred from the original source to other data centers before definitely mirrored or downloaded to the GOP data center. Only a very few files are directly incoming from the original source to the GOP center (about 7%). Such files are characteristics by its 2-8 minutes latency while all others by latency varying in 10-30 minutes (Fig 6), sometimes even longer.

Our '*Call for the first-hand data upload to GOP'* is motivated also by GOP activity within the TOUGH project (Vedel and Xiang-Yu, 2003). The GOP is responsible for near real-time analysis of network (consisting mostly of EPN sites) covering the central and east European countries. The aim of the analysis is the estimation of troposphere parameters for exploitation in meteorology. Including a data incoming hourly into the GOP data center, it is a platform for contribution to the TOUGH EU-project (2003-2005), which can be additional motivation for agencies operating the GPS in the specified area.

7. Conclusion

The Geodetic observatory Pecny data center serves as the provider of NRT GNSS data (hourly RINEX), products and information for NRT or post-processing analyses. After more than two years of data mirroring for our internal purposes as well as after new orientation to external users, the complete re-implementation of the original system was performed. The new system is written in the compact and modular way, with goal to support extending services within the data center. Among those, the data concatenation to a longer time units, providing a data and product checking, monitoring and statistics services etc.

Besides the higher configuration flexibility and possible extending services, the main advantage of the new implementation is a reduction of data flow and data storage. We achieved about 40% lower traffic with new download scheme as well as the demands for the local data storage. Although the data latency has a different pattern, there is no serious increase in the total latency. The only significant improvement in the latency could be achieved by data retrieving directly from their original sources. Unfortunately, the GOP data center is in most cases at the end of a chain of other data-centers.

The external appearance of the data center was preserved close to the old style, but the new system is definitely more flexible for new services within the EUREF or other user communities. Actually, the new implementation runs in parallel to the old official system due to its smooth later transition scheduled for the end of August 2003.

Acknowledgements

This project was supported by Grant Agency of the Czech Republic (103/00/P042) and by the Ministry of Education, Youth and Sports of the Czech Republic (LN00A005).

References

DOUSA, J.: *GOP data center for near real-time GPS analysis,* In: Reports on Geodesy, Proceedings of the EGS G9 Symposium 'Geodetic and geodynamic programmes of the CEI (Central European Initiative)', WUT, Insitute of Geodesy and Geodetic Astronomy, Warsaw, Poland, pp. 149-152, 2001.

DOUSA, J.: *GOP NRT tropospheric product for the COST-716 NRT campaign 2001*, Proceedings on the Symposium of the IAG Subcommission for Europe (EUREF), Ponta Delgada, Portugal, 5-8 June, 2002a.

DOUSA, J.: *The GOP analysis center: a global near real-time solution*, Proceedings on the Symposium of the IAG Subcommission for Europe (EUREF), Ponta Delgada, Portugal, 5 - 8 June, 2002b.

ESTEY, L.H. AND MEERTENS, C.M.: *TEQC: The multi-purpose toolkit for GPS/GLONASS data*, GPS solutions (publ. by John Wiley & Sons), Vol 3, No 1, pp. 42-49, 1999.

MAREL, H.V.D., ELGERED, G., BARLAG, S., DE HAAN S., OFFILER, D., GENDT, G., DOUSA, J., BROCKMANN, E., VESPE, F., PACIONE, R., RIUS. A., JOHANNSON, J. KRISTIANSEN, O.: *COST-716 near real-time demonstration project*, presented at Session G7 of the European Geophysical Society, 27th General Assembly, Nice, France, April 21-26, 2002.

VEDEL H., XIANG-YU, H. AND TOUGH TEAM.: *Targeting Optimal Use of GPS Humidity Measurements in Meteorology, (http://tough.dmi.dk/), 2003.*