ISSN 2039-7941
anno 2022 numero 456

RAPPORTI TECNICI INGV

The INGV Mediterranean GNSS Archive (MGA)

456

ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA

REGISTRAZIONE AL TRIBUNALE DI ROMA N.174 | 2014, 23 LUGLIO

Rappresentante legale: Carlo DOGLIONI
Sede: Via di Vigna Murata, 605 | Roma

ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA

ISSN 2039-7941
anno **2022** numero **456**

RAPPORTI TECNICI INGV

The INGV Mediterranean GNSS Archive (MGA)

Sergio Bruni, Daniele Randazzo, Roberto Devoti, Enrico Serpelloni, Adriano Cavaliere, Leonardo Martelli, Grazia Pietrantonio, Nicola D'Agostino

INGV | Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Nazionale Terremoti

Accepted 24 May 2022 | *Accettato 24 maggio 2022*

How to cite | *Come citare* Bruni S., Randazzo D., Devoti R., Serpelloni E., Cavaliere A., Martelli L., Pietrantonio G., D'Agostino N., (2022). The INGV Mediterranean GNSS Archive (MGA). Rapp. Tec. INGV, 456: 1-22, https://doi.org/10.13127/rpt/456

Cover Distribution of the GNSS stations currently archived in the MGA system | *In copertina Distribuzione delle stazioni GNSS attualmente archiviate nel sistema MGA*

456

INDEX

Abstract

The Global Navigation Satellite System (GNSS) represents a primary data source in Solid Earth Sciences. In order to investigate the Earth's crustal deformation, time series of the estimated daily positions of the stations are routinely analyzed at the Istituto Nazionale di Geofisica e Vulcanologia (INGV) to investigate the deformation of the Earth's surface caused by tectonic and non-tectonic processes. The GNSS observations of the stations are processed using the three main scientific software: GAMIT/GLOBK, BERNESE, and GIPSY OASIS II. The accuracy and the strength of geodetic solutions often depend on the geometry and spatial density of the network, and the availability and quality of GNSS data. In many circumstances, GNSS networks are deployed for topographic purposes by private or public institutions, and a significant number of GNSS stations in large regions acquire continuous observations. It may happen that such networks do not collect and distribute data according to IGS standards, so it could be difficult to analyze this data using automated data-processing tools. For that reason, this data is often ignored or partially used by the scientific community, despite their potential usefulness in geodynamic studies. We have attempted troubleshooting this problem by establishing a centralized storage facility in order to collect all available GNSS data and standardize both formats and metadata information. Here we describe the processes and functions that manage this unified repository, called MGA (Mediterranean GNSS Archive), which regularly collects GNSS RINEX files from a large number of CORS (Continuously Operating Reference Station) located across a wide region of mainly the European and African plates. RINEX observation data and metadata information are provided to the analysts through an FTP server and dedicated web-services. The complete data set is stored in a PostgreSQL database in order to easily retrieve pieces of information and efficiently manage the archive content. The system implements many high-level services that include scripts to download files from remote archives and to detect new available data, web applications such as API (Application Program Interface) to interact with the system, and background services that interact with the database. During the development of this product, particular attention was paid to what has already been achieved by the EPOS Thematic Core Service WP10 (https://www.epos-eu.org/tcs/gnss-data-and-products), whose objective was: "*... to develop an open source platform with programmatic and web interfaces to store and disseminate raw data and metadata from GNSS stations operating in Europe*". Many ideas and tools presented here were inspired by that project.

Keywords GNSS; RINEX files; Mediterranean; Postgres

Introduction

The number of GNSS permanent stations has increased significantly around the world in recent years and still continues to increase. In Europe many private and public organizations collect, manage, and distribute GNSS data with little or no coordination efforts. The EPOS project, now in the pre-operational phase, will distribute the data made available by EU countries participating in the EPOS initiative, although networks still exist that deliver data without providing metadata at the accuracy level required for scientific purposes. In this technical report we outline the major features of a project started at the end of 2017 with the aim of creating a centralized GNSS data and metadata archive for the circum-Mediterranean region, which is made available to the INGV institutional activities, research projects, and a vast scientific community, even if not completely public. Since 2004, INGV has started to build a national network of GNSS stations for studying the active tectonic deformation in Italy, called RING (Rete Integrata Nazionale GNSS), which currently consists of about 220 stations (http://ring.gm.ingv.it). The area surrounding the Mediterranean Sea represents the boundary of the Eurasian, African, and Arabian plates, where geodynamic processes are rather complex and still under investigation. Among the ongoing processes in this region we recall the Glacial Isostatic Adjustment (GIA) concerning mainly northern Europe, tectonic deformation that could be related to intra-plate earthquakes, the Africa-Eurasia boundary zone, the Adriatic promontory kinematic behavior, the Alps and the Dinarides-Carpathian orogens surrounding the Pannonian basin that show and the Aegean-Anatolian driven mainly by the Hellenic trench and the North Anatolian Fault [Noquet, 2012]. GNSS networks provide accurate measurements of Earth's surface deformation but the data availability is limited to large research-oriented infrastructures. However, a growing number of regional GNSS infrastructures may represent a valuable data source even for geodynamic studies.

The need for a public GNSS data repository has also been addressed by the European Plate Observing System (EPOS), where a specific working group, Thematic Core Service (EPOS-TCS-Work Process WP10) has been working on this topic since 2015. Currently, the general architecture is based on a network of regional archives based on GLASS nodes (Geodetic Linkage Advance System Software). Each local node hosts in his database the necessary information to distribute the metadata (from database) and the associated data (files stored in Data Centers). The Central European node is the main node exposed to the public, which collects data and metadata from different nodes and allows access to the whole database from the web. INGV contributes with a local node which provides data and metadata from the RING network.

The main difference between the MGA and EPOS archives is that the latter only collects data provided by certified Agencies that have formally agreed to provide standardized data and metadata. The contributing Agencies are also responsible for updating station metadata and submitting it to the EPOS database. The guidelines followed in MGA lead instead to archive data provided by heterogeneous networks (commercial, governmental, scientific or application-oriented), many of which do not provide station log-files, nor adopt standardized formats, but regularly provide GNSS data [Bruni et al., 2020]. Data consistency and metadata reconstruction processes are then accomplished by the archive manager, which ensures that data is provided in standardized RINEX (Receiver INdependent EXchange format) and systematically updates the metadata log-file for each station. Since no action is required by the data suppliers, an important requirement for the download operations is the autonomous recognition for the availability of new station data. Thus, the archive manager will regularly access the remote archives to detect new valuable GNSS data and add them to the archive database.

The data stored in the MGA repository are of two types: raw RINEX files containing the GNSS observations and the station metadata files containing auxiliary information useful for accurate data analysis. RINEX files contain raw observation data sampled at 30 seconds and are supplied in IGS-standard compressed formats (file with extension .Z and .gz). The metadata files keep track of instrumental hardware changes and other station auxiliary information relevant to data processing. The log file creation and update was handled by dedicated software and is described by a specific technical report [Randazzo et al. 2018] and [Randazzo et al. 2022]. Given the heterogeneous nature of the collected data, specific tools have been developed to identify and correct data discrepancies and make them reliable for the analysis.

At the time of writing, about 4100 stations are registered in the MGA system and more than 13 million RINEX files are stored. Figure 1 shows the location of the GNSS stations currently registered in the database. Figure 2 shows the time evolution of the number of RINEX files acquired per network.

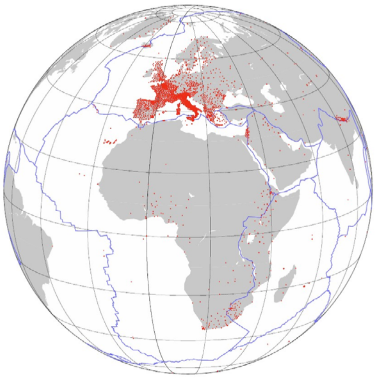


Figure 1 Distribution of the GNSS stations currently archived in the MGA system.

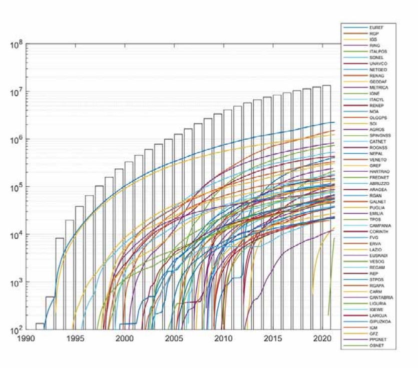


Figure 2 Number of RINEX files acquired per network. The vertical bars indicate the cumulative amount of files over time.

1. General description of the system

The main requirement of the MGA system is to collect all GNSS data in the European-Mediterranean region on a daily basis, store raw data and update metadata in a database, and detect the occurrence of new GNSS stations. The process responsible for the data download is called "gnss_downloader" (in short *Downloader*), which regularly scans remote repositories to download RINEX files of GNSS stations. The stations are conventionally grouped by "networks" that may reflect the ownership of the infrastructure but also other attributes of the stations (e.g. location, geographic region, research project, etc.). A specific configuration file defines the characteristics of the websites, from which the RINEX files are retrieved, and the basic parameters for scheduling and archiving processes. All RINEX files for a given day of the year (DOY) are indexed in the database and stored in the local file system with a standard relative path "ROOT/NETWORK/YEAR/DOY" as shown in Figure 3. The heterogeneous nature of the GNSS data sources (RINEX files) in terms of duration, sampling rate, version, compression, and so on, increases the complexity of the download process. Nevertheless, the *Downloader* is able to recognize and eventually reconstruct the standard data format. If an issue is raised during the download, an email is sent to the administrator without the process being interrupted.

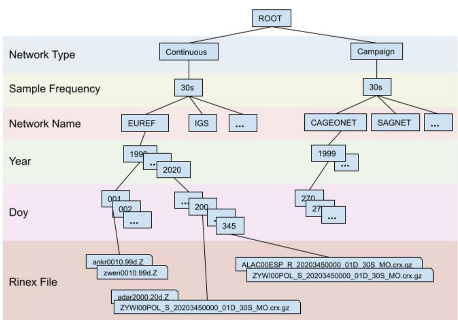


Figure 3 The MGA file system repository structure.

The front-end of the storage system is the so-called "*gnss_monitor*" (in short, *Monitor*), a web-based GUI that offers the user an intuitive and comfortable interaction with the MGA database. The *Monitor* displays the download statuses of each network and is capable of performing the main activities of promoting new stations to the download list and to report relevant messages to the user community.
In addition, a collection of software tools (gnss_tools) has been developed for use in low-level service processes and capable of modifying the database content. The interaction with the database is realized by two server services, implemented by the sub-projects "*gnss_api*" and "*fwss*" (Flask Web Service Server), which were developed separately but may be merged in the future. *Downloader*, *Monitor*, and the other system tools largely use these services to query or update the database.
To keep track of the development phases and facilitate the sharing of the code with the production server, the whole project was archived in the INGV GitLab repository in the group called "gnss", which is reachable via the link https://gitlab.rm.ingv.it / gnss. The entire system is represented in Figure 4, which shows the main processes and their interactions

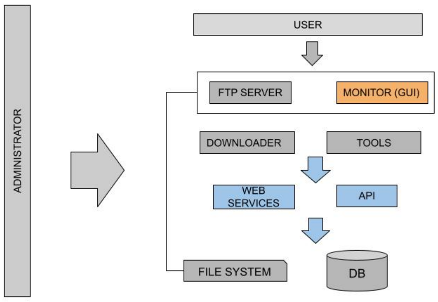


Figure 4 Schematic representation of the system. The arrows represent the interactions between components. The User interacts with the FTP Server and Monitor GUI. The Administrator interacts with any part of the system.

The following projects are also stored in the git group "gnss":

- "*gnss_datamodel*": the database data model, implemented using the ORM (Object Relational Model) SqlAlchemy and used by "*gnss_api*" and "*fwss*";
- "*gnss_checkrinex*": a Python module that inspects the RINEX files and extracts the information necessary to index it. It is used by "*gnss_downloader*" and "*gnss_indexer*";
- "*doc*": a repository with all documentation, including the following technical report.

2. Features of the Downloader

The *Downloader* is responsible for the scheduled download of the RINEX files and their merging into daily RINEX files in case of hourly or sub-daily acquisitions. Within the "download" activities, the *gnss_discover* script deals with the determination of the approximate position of the stations while the *gnss_indexer* performs the indexing of the downloaded file in the database.
The *Downloader* can be run mainly in two different modes, the first of which involves daily scheduling via crontab (the implicit parameters are defined in the configuration file), the latter is the interactive launch via command line from terminal with explicit input parameters. The *Downloader* can be configured differently for each network allowing you to set the following:

- remote entries (url), services (ftp, http, https, ...) and relative paths
- geographical area of interest
- "discovering new stations" mode
- priority of RINEX versions to be downloaded
- time window for downloading RINEX files
- scheduling time

A dedicated process is run every day for each GNSS network following the instructions in the configuration file. The latter is initialized with generic default parameters and for each network defines the access credentials, the website path, the type and priority of the expected formats. The expected data formats may be of RINEX 2.xx and/or RINEX 3.xx, whose downloading priority can be freely set. The *Downloader* will check the files in order to verify that both the format and the sampling frequency are consistent with those declared. Once connected to the web source, the list of available remote stations is compared with the local station list: the recognized RINEX files are downloaded to the repository folder while the remaining unknown data are processed by *gnss_discover* for the recognition of new stations. All downloaded RINEX files concerning stations registered in the DB are indexed by *gnss_indexer* and copied to the repository file system. The approximated position, calculated with Anubis, as described in the next section, software is also stored, together with other information present in the RINEX header section.
RINEX files may be retrieved from different web sources, establishing a many-to-many relationship between networks and websites or private web portals. In order to connect to the data sources different protocols may be used (ftp, ftps, http, https), some require authentication and some are designed to interact with the user. In this last case a dedicated "state machine" is needed to acquire the data. Sinec web sources may be subject to changes, the administrator must adapt the software to the new requirements.

2.1 Geographical region of interest and recognition of new stations

The *Downloader* has to know which station is worth downloading. For that purpose, a geographical Region of Interest (GRI) can be set as a rectangular box, valid for all networks, so that stations falling within the GRI are recognized as new potential stations. There is also the possibility to override the global GRI by defining a network-based GRI that defines the area of interest for each single network. This allows a flexible choice of GRI, tailored to any network, and permits downloading GNSS data based on the station location.
When the system recognizes a new station, it allows the user to check its RINEX files in order to decide whether to admit the new station to the list of active stations or to include it in a black list to avoid further downloads. RINEX files of potential new stations are downloaded to a temporary folder called "NEW_STATION/BUCKET"' and then analyzed by the *gnss_discover* routine, which computes the daily position for each file in the BUCKET by means of the Anubis software (https://www.pecny.cz/gop/index.php/gnss/sw/anubis). If the estimated station position falls within the GRI then the station name is entered in a dedicated table (station_ghost) and its status is set to NEW. If the position falls outside the GRI, the station is set to the DISCARDED status and its files are moved to the DISCARDED folder. Stations in NEW status are highlighted by the monitor that alerts the analyst showing the list of new stations for each network (this list is shown by clicking the green button "+" in the form of Figure 5). The station status can be changed to PROMOTED through the GUI interface, i.e., the station becomes part of the list of active stations and the newly promoted stations are notified by email to the workgroup. The station status can alternatively be set to IGNORED and thus no longer be taken into account in the download process. To facilitate both the choice of the status and the assignment of the 9-character marker, the software allows you to locate the station on the map and visualize any other station having the same 4-character marker (homonyms or synonyms) already present in the database.

2.2 Station and RINEX file naming conventions

We adopt the standard IGS naming conventions for the GNSS stations and RINEX filenames (https://www.igs.org/formats-and-standards/). Each GNSS station in the MGA database is identified by the extended 9-character string formed by the 4-char site ID, the 1-char monument number, the 1-char receiver number, and the 3-char ISO-alpha country code (e.g., MATE00ITA). This convention, if used correctly, allows each station to be identified with a unique name. IGS has long started to distribute the GNSS station files in RINEX3 format, which includes observations from all GNSS constellations and follows the long-name convention (e.g. https://files.igs.org/pub/data/format/RINEX305.pdf), making the oldest short-name convention obsolete (e.g. https://files.igs.org/pub/data/format/Addendum-RINEX211.pdf). Nevertheless, the versions 2 and 3 of RINEX files are both still accepted and currently available in many web archives. Recently, a further change was adopted by IGS, starting from 1 Dec. 2020 the compression method for the RINEX v.2 files changed from *.Z to *.gz using the gzip compression algorithm (https://lists.igs.org/pipermail/igsmail/2020/007990.html).
Currently, the MGA system recognizes, archives, and delivers all three types of RINEX files, i.e. RINEX2 (version 2.xx, UNIX-compressed and and gzip-compressed, e.g. nnnnddds.yyd.Z and nnnnddds.yyd.gz respectively) and RINEX3 (version 3.xx, gzip-compressed, NNNNMRCCC_T_YYYYDDDHHMM_DDU_DDU_DD.crx.gz). In addition, since the version 2.xx RINEX files have a 4-char name that identifies the station ID, we adopted the internal convention that in each network the stations must be uniquely identified by the four character site ID. This prevents having two stations of the same ID, but different locations, thus allowing to correctly identify the corresponding RINEX v.2 file in the remote archive.
The program must be able to recognize the RINEX data regardless of the file naming convention used in the remote archive: the station is uniquely identified by the approximate coordinates. For RINEX version 2.xx files already present in the file system, it was necessary to develop a specific code that would allow to refer the name to the 3.xx format and highlight cases of homonymy.

2.3 Metadata management

Once the RINEX files have been stored in the file system, the software extracts the metadata reported in the file header, making them available for subsequent data analyses. These metadata is the antenna/radome model, the receiver model, and the a-priori coordinates of the GNSS station.

3. The Monitor

The *Monitor* is a web user interface that displays the *Downloader* current status and allows users to explore the MGA archive. As shown in Figure 4, at the top left of the main page a small toolbar gives access to different contents of the database: MONITOR, NETWORKS, STATIONS, and REPORT.
The *Monitor* main page shows the list of networks and their corresponding downloading status. A graphical bar indicator to the right of each scheduled network highlights the status of the files downloaded in the last 15 days. Green color indicates normal behavior (downloaded files are equal or higher than the median over the last year), red color indicates download anomalies (low number of downloaded files, below the median). The number of RINEX files downloaded in the last two weeks are accessible by hovering the mouse pointer on the bar indicators.

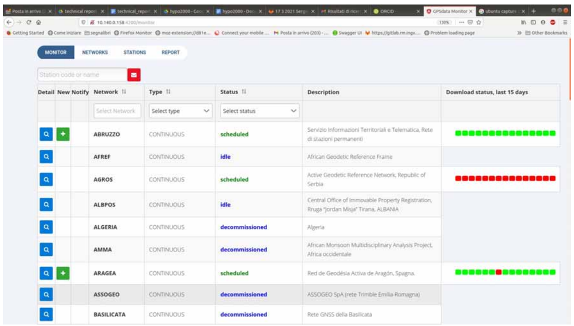


Figure 5 The *Monitor* home page.

The *Downloader* can be set as one of the following status: SCHEDULED, UNSCHEDULED, IDLE, RUNNING, and DECOMMISSIONED. If the *Downloader* is correctly scheduled for daily downloads then the monitor shows the SCHEDULED status, waiting for the scheduling time; if the *Downloader* is running and downloading files, the RUNNING status is shown, otherwise the status is UNSCHEDULED. The IDLE and DECOMMISSIONED status refer respectively to inactivity due to data temporarily unavailable for the network and to data no longer available due to network decommissioning. The download of RINEX files is performed by independent processes for each network, which can also run in parallel, and each process writes to its log file warnings and errors encountered during the execution. Log files are periodically rotated for writing with the aim of keeping their size as small as possible without loss of information. At each run, if at least one ERROR message appears in the log file, the *Downloader* sends an alert email to a list of users

3.1 Active stations list

By clicking on the network button, the list of active stations becomes visible. The Downloader will search for RINEX files on the web archives for all stations in this list. Selecting one or more stations from the list it is possible to perform the following actions: to display the data availability, to download a ".kmz" file to display on Google-Earth, to download an ASCII file containing the approximate coordinates (Earth fixed cartesian coordinates) and to plot the time series of the approximate station positions.

3.2 New and promoted stations

A "+" red button on the Monitor main page reports the presence of potentially new stations in the network. In Figure 5 the button is green, this means that there are new stations, but none of them are in the status NEW. By clicking this button a table with the list of identified new stations is open. Each station has to be promoted to the PROMOTED status in order to be regularly processed by the *Downloader*. To allow this evaluation process the new station is associated with one of the following status: NEW, IGNORED, DISCARDED, PROMOTED, and NOTIFIED. New potential stations are tagged as NEW and the corresponding RINEX files are temporarily saved in the NEW_STATIONS/STAGING folder, while the position, the first date of appearance, and the presence of homonyms in the database are displayed in the *Monitor*. The status of a station may be changed by clicking the "edit" button (pencil on the left), that allows to toggle the status (NEW, DISCARDED, IGNORED), to edit the monument and receiver number, and to set the country code. Once the 9-character ID of the station is completed, the station may be promoted to the "active station list" of the corresponding network. The promotion should be operated manually, by checking the stations and using the "PROMOTE" button. This last action completes the promotion process and the promoted-station RINEX files will be regularly downloaded from the remote archive. Newly promoted stations can be notified to interested users via email by clicking a button that appears in the *Monitor* main page.

Time series. A command line tool has been implemented to create the time series as a CSV file for a single station or for all stations of a network.

Moreover, the *Monitor* allows to:
- send “Free Notification” by email with service information to the user group;
- create map representation to show selected stations on Google map.

4. Base software infrastructure

4.1 Graphical User Interface (Client Side Language)

The Web GUI has been implemented using AngularJS, a JavaScript framework developed by Google. It is one of the largely used tools for Web application development.

4.2 Server Side Programming Language

The server side was coded in Python, a high-level programming language that has grown enormously in recent years because it is rich in tools and libraries, especially for scientific applications. The web services (“gnss_api” and “fwss”) have been implemented with REST (REpresentational State Transfer), which defines a set of architectural principles for system design. The Python services are based on the Flask framework while the connection with the “information” is managed by the ORM (Object Relational Model) included in the SqlAlchemy Python module.

4.3 Database Management System and the MGA database

To satisfy the characteristics mentioned in the previous paragraph, the use of a database is certainly of great help. The database has been designed to contain the metadata of millions of RINEX files and the database itself is used as a reference index for the information stored in the file system, with the advantage of being able to apply any selection criteria. The database contains the references associated with the GNSS stations, the path to download the RINEX files in the repository and all the metadata necessary for the geodetic analysis. PostgreSQL was chosen as the DBMS (Database Management System), as it is free and has good performance, extensibility (https://www.postgresql.org/docs/9.0/extend-how.html), scalability qualities, and is widely used by many companies that contribute to its innovation. As a database structure it was decided to use the one developed by the EPOS TCS WP10 team, which is the result of the work of a team of experts and is able to store all the metadata relating to the GNSS universe. The database tables are grouped by metadata i.e.:
- metadata T0: related to the infrastructure for storing RINEX data;
- metadata T1: related to the RINEX file;
- metadata T2: related to the stations;
- metadata T3: related to the data quality check.

To meet the specific INGV requirements, the database structure was expanded simply by adding tables or fields to existing tables, in order to maintain compatibility with the services already developed: for example, it was possible to inherit the cartographic representation of the stations registered using the same GLASS GUI made by the WP10 team, as shown in Figure 6.

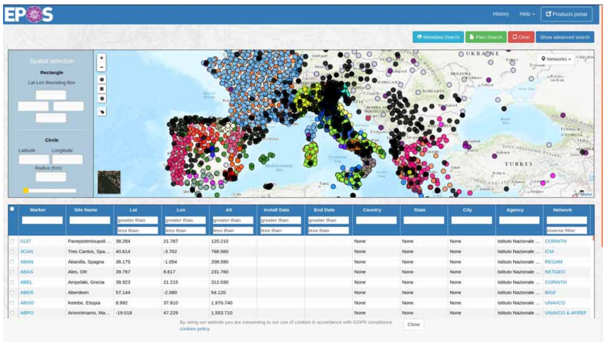


Figure 6 The GLASS GUI interface to represent the MGA registered stations.

5. Features and technological choices

5.1 Database

As mentioned above, appropriate tables or fields were added to the original database to store the metadata extracted from the header of the RINEX files, as this information is required for geodetic analysis. Specific files with the hardware/firmware changes over time are generated for each GNSS station. These files have a specific format depending on the software involved in the data analysis (i.e., GAMIT, BERNESE and GIPSY). This aspect has however been addressed in a complementary project, described in another technical report).
With the aim of avoiding homonymous among stations within the database, each station is associated with a name according to the new 9-character standard, consisting of the marker name (4 characters), monument number (1 digit), receiver number (1 digit), and ISO country code (3 characters). The union of these fields is called long_marker and represents the format adopted by the RINEX3 files. The long_marker identifies the station uniquely at a global level, unlike the RINEX2 files which instead are characterized by the 4-character marker and must therefore be unique within the same network.

5.2 User data access

The GNSS data can be accessed internally by INGV users. The use of GNSS data is limited and is allowed only at conditions dictated by network owners. Disclaimers in the Network sections of the Monitor give the indications and requests in order to publish relevant conclusions from these data. The MGA system allows downloading RINEX data files via ftp-anonymous protocol (ftp://gnssgiving.int.ingv.it). Data from different networks are stored in different directories according to the file system structure reported in Figure 3, continuous and discontinuous operating stations are grouped together respectively in CONTINUOUS and CAMPAIGN directories. The metadata are available in each network directory as standard log files in directories “log”, or in ready-to-use software dependent files, in the directories “stf”. All metadata are updated daily using original logfiles or RINEX headers as input data.

The users may exploit a number of access points of the following GNSS webserver (http://10.140.0.158:4300). For example, to download a list of stations contained in a circular area of radius R, centered in geographic coordinates (latitude and longitude):

```
endpoint: stationlist/circle
method: POST
json param: {
        "circle": {
                "lat": 42.4,
                "lon": 12.85,
                "radius":50000
        },
        "network_names": ["RING"]
}
```

Or alternatively, using the *curl* command:

```
curl -X POST http://10.140.0.158:4300/stationlist/circle -H 'Content-Type: application/json' \
-d '{"circle":{"lat": 42.4, "lon": 12.85,"radius":50000}, "network_names":["RING"]}
```

5.3 Metadata management

One of the goals of this software is to store the metadata contained in the RINEX files in special tables, in order to make them available for creating the station information files. This type of file, whose format is software-dependent (e.g., .stf for Gamit), is generated by a software produced by another INGV working group and is a standard format file that contains the history of changes to the station's GNSS devices (antenna, receiver, receiver firmware and so on). This file is obtained through a main query that selects only those records in which a change of the said parameters occurs. The use of the database is strategic, in addition to facilitating the selection, it greatly reduces the amount of calculations and processing times, less than 1 million of 12 million records are processed.

5.4 Database tools

Some specific tools have been implemented to correct some inconsistencies present in the pre-existing data and to periodically check the alignment between the file system content and the database metadata. A few have been integrated with the GUI and others can be called by command line.

5.5 Homonymous stations

As previously described, the workflow designed for new stations avoids registering homonymous stations through the use of the nine-character marker. However, the development of this system started with data already existing within the archive and many stations were already defined, some of which with the same name. Then, once the pre-existing RINEX files have been indexed in the database, a specific tool, called *gnss_explorer*, was developed to bring these cases to the attention of analysts. This tool queries the database and generates a printout highlighting cases of suspected homonymy. By this tool, called *gnss_explorer*, the cases of homonymy have been deleted by renaming one of the two stations using the nine-character marker. An example of *gnss_explorer* output is shown in Figure 7.

marker	networks	NO.of diff.	net1	net2	common	% of diff.	ABS of difference between medians (x,y,z)
RECT00NGA (2498)	CANARY (10) SEGAL (73)	29	691	603	587	4.94	0.012 0.015 0.014
RECT00NGA (2498)	CANARY (10) AFREF (8)	2	691	749	676	0.30	0.012 0.006 0.0
RECT00NGA (2498)	SEGAL (73) AFREF (8)	34	603	749	579	5.87	0.0 0.021 0.014
PESH01ALB (3580)	METRICA (26) NOA (27)	0	113	1350	113	0.00	0.08 0.044 0.118
PESH01ALB (3580)	METRICA (26) ALBPOS (58)	0	113	332	0	0.00	0.094 0.056 0.113
PESH01ALB (3580)	NOA (27) ALBPOS (58)	0	1350	332	0	0.00	0.014 0.1 0.005
M0SE00ITA (1723)	EUREF (21) ITALPOS (30)	1672	5140	1894	1672	100.00	0.074 0.001 0.125
M0SE00ITA (1723)	EUREF (21) GEODAF (39)	220	5140	5129	4725	4.66	0.015 0.003 0.021
M0SE00ITA (1723)	ITALPOS (30) GEODAF (39)	1602	1894	5129	1602	100.00	0.089 0.002 0.146
PRAT00ITA (2339)	GEODAF (39) EUREF (21)	973	7283	7645	7242	13.44	0.001 0.001 0.012
PRAT00ITA (2339)	GEODAF (39) ITALPOS (30)	1654	7283	2032	1654	100.00	0.145 0.006 0.33

Figure 7 The output of *gnss_explorer.*

The *gnss_explorer* searches for all stations whose RINEX files are downloaded for multiple networks. For each of these stations it analyzes separately all networks two by two generating a specific output line. For example, the station RECT00NGA is downloaded from networks CANARY, SEGAL, and AFREF, so three lines appear in the output: CANARY-SEGAL, CANARY-AFREF, and SEGAL-AFREF. Looking at the output the analyst can establish whether the RINEX files belong to the same station or not. The fields reported on the output are:

- Marker of the station
- The names of the two networks
- Number of RINEX file that differs from those two networks
- Number of RINEX files downloaded from first network
- Number of RINEX files downloaded from the second network
- Number of common RINEX files between the two networks
- % of the different common RINEX files
- Absolute value of the difference between the two medians of the position x,y,z of the rinex files, common and not common of the two networks

If a station is recognized to be homonymous between two networks (same name but different positions), the analyst can use another tool, named *split_station*, to give each station a proper 9-character code that uniquely identifies them. *Completeness of the database.*

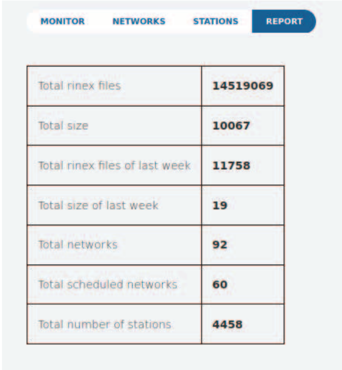
DB_completeness. To constantly check the correspondence between the files on the disk and the relative metadata indexed in the database, a specific script named *check_db_completeness* has been created, which discovers inconsistencies, such as non-indexed files or indexed files no longer present.

5.6 HW/SW requirements

The software system has been deployed on a virtual machine at INGV Rome, with the Ubuntu 20.04 OS, with 4 CPUs and 4GB of RAM. The disk space usage for the Mediterranean GNSS database is of the order of 10 TB at the moment of writing. The whole system has been mirrored on the ReCaS DataCenter at the University of Bari (https://www.recas-bari.it). Figure 8 shows a summary report offered by the Monitor GUI.

From this report we can infer that the increase of disk space is about 1 TB per year (~20GB per week).

Figure 8 Summary report from the Monitor GUI.



References

Bruni S., Randazzo D., Cavaliere A., Avallone A., Serpelloni E., Devoti R., D'Agostino N., Pietrantonio G., Riguzzi F., Cecere G., D'Ambrosio G., Falco L., Martelli L., (2020). *Attività sui dati, metadati e prodotti per le stazioni GNSS in area Mediterranea*, https://doi.org/10.13127/misc/57/22. In: AA. VV., (2020). Progetto "Sale Operative Integrate e Reti di monitoraggio del futuro: l'INGV 2.0". Report finale. Editors: L. Margheriti, F. Cirillo, F. Guglielmino, M. Moretti. Misc. INGV, 57: 1-186, https://doi.org/10.13127/misc/57

Nocquet J.-M., (2012). *Present-day kinematics of the Mediterranean: A comprehensive overview of GPS results*. http://dx.doi.org/10.1016/j.tecto.2012.03.037

Randazzo D., Serpelloni E., Cavaliere A., (2018). *PyGLog: a Python software for handling GNSS metadata and log-files*. Rapp. Tec. INGV, 396: 1-26, https://doi.org/10.13127/rpt/396

Randazzo D., Cavaliere A., Bruni S., Martelli L., Serpelloni E., Devoti R., (2022). *PyGLogDB: software to generate STATIONINFO files for GNSS data analysis*. Rapp. Tec. INGV, 446: 1-24, https://doi.org/10.13127/rpt/446

QUADERNI di GEOFISICA

ISSN 1590-2595 http://istituto.ingv.it/it/le-collane-editoriali-ingv/quaderni-di-geofisica.html/

I QUADERNI DI GEOFISICA (QUAD. GEOFIS.) accolgono lavori, sia in italiano che in inglese, che diano particolare risalto alla pubblicazione di dati, misure, osservazioni e loro elaborazioni anche preliminari che necessitano di rapida diffusione nella comunità scientifica nazionale ed internazionale. Per questo scopo la pubblicazione on-line è particolarmente utile e fornisce accesso immediato a tutti i possibili utenti. Un Editorial Board multidisciplinare ed un accurato processo di peer-review garantiscono i requisiti di qualità per la pubblicazione dei contributi. I QUADERNI DI GEOFISICA sono presenti in "Emerging Sources Citation Index" di Clarivate Analytics, e in "Open Access Journals" di Scopus.

QUADERNI DI GEOFISICA (QUAD. GEOFIS.) welcome contributions, in Italian and/or in English, with special emphasis on preliminary elaborations of data, measures, and observations that need rapid and widespread diffusion in the scientific community. The on-line publication is particularly useful for this purpose, and a multidisciplinary Editorial Board with an accurate peer-review process provides the quality standard for the publication of the manuscripts. QUADERNI DI GEOFISICA are present in "Emerging Sources Citation Index" of Clarivate Analytics, and in "Open Access Journals" of Scopus.

RAPPORTI TECNICI INGV

ISSN 2039-7941 http://istituto.ingv.it/it/le-collane-editoriali-ingv/rapporti-tecnici-ingv.html/

I RAPPORTI TECNICI INGV (RAPP. TEC. INGV) pubblicano contributi, sia in italiano che in inglese, di tipo tecnologico come manuali, software, applicazioni ed innovazioni di strumentazioni, tecniche di raccolta dati di rilevante interesse tecnico-scientifico. I RAPPORTI TECNICI INGV sono pubblicati esclusivamente on-line per garantire agli autori rapidità di diffusione e agli utenti accesso immediato ai dati pubblicati. Un Editorial Board multidisciplinare ed un accurato processo di peer-review garantiscono i requisiti di qualità per la pubblicazione dei contributi.

RAPPORTI TECNICI INGV (RAPP. TEC. INGV) publish technological contributions (in Italian and/or in English) such as manuals, software, applications and implementations of instruments, and techniques of data collection. RAPPORTI TECNICI INGV are published online to guarantee celerity of diffusion and a prompt access to published data. A multidisciplinary Editorial Board and an accurate peer-review process provide the quality standard for the publication of the contributions.

MISCELLANEA INGV

ISSN 2039-6651 http://istituto.ingv.it/it/le-collane-editoriali-ingv/miscellanea-ingv.html

MISCELLANEA INGV (MISC. INGV) favorisce la pubblicazione di contributi scientifici riguardanti le attività svolte dall'INGV. In particolare, MISCELLANEA INGV raccoglie reports di progetti scientifici, proceedings di convegni, manuali, monografie di rilevante interesse, raccolte di articoli, ecc. La pubblicazione è esclusivamente on-line, completamente gratuita e garantisce tempi rapidi e grande diffusione sul web. L'Editorial Board INGV, grazie al suo carattere multidisciplinare, assicura i requisiti di qualità per la pubblicazione dei contributi sottomessi.

MISCELLANEA INGV (MISC. INGV) favours the publication of scientific contributions regarding the main activities carried out at INGV. In particular, MISCELLANEA INGV gathers reports of scientific projects, proceedings of meetings, manuals, relevant monographs, collections of articles etc. The journal is published online to guarantee celerity of diffusion on the internet. A multidisciplinary Editorial Board and an accurate peer-review process provide the quality standard for the publication of the contributions.

Coordinamento editoriale
Francesca DI STEFANO
Istituto Nazionale di Geofisica e Vulcanologia

Progetto grafico
Barbara ANGIONI
Istituto Nazionale di Geofisica e Vulcanologia

Impaginazione
Barbara ANGIONI
Patrizia PANTANI
Massimiliano CASCONE
Istituto Nazionale di Geofisica e Vulcanologia

Istituto Nazionale di Geofisica e Vulcanologia
Via di Vigna Murata, 605
00143 Roma
tel. +39 06518601

www.ingv.it

ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA