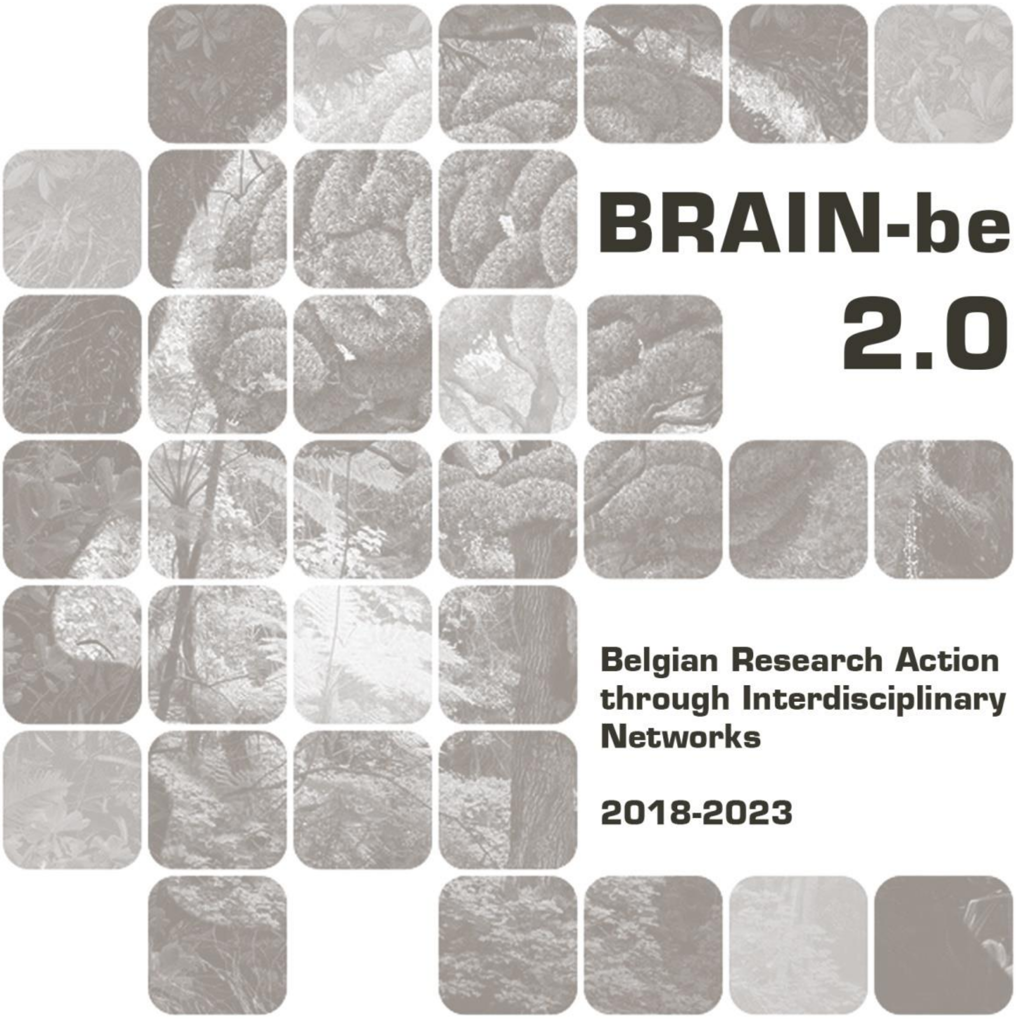

FAIR-GNSS

Open data portal for European and Belgian GNSS reference station data collections, built upon FAIR guiding principles

A. Miglio (ROB), C. Bruyninx (ROB), S. De Bodt (UGent), A. Fabian (ROB), J. Legrand (ROB), P. Oset Garcia (UGent), I. Van Nieuwerburgh (UGent)

Pillar 2: Heritage science

ROYAL OBSERVATORY OF BELGIUM

GHENT UNIVERSITY

belspo

.be

NETWORK PROJECT

FAIR-GNSS

Open data portal for European and Belgian GNSS reference station data collections, built upon FAIR guiding principles

Contract - B2/202/P2/FAIR-GNSS

FINAL REPORT

PROMOTORS: Carine Bruyninx – Royal Observatory of Belgium (ROB)
Inge Van Nieuwerburgh – Ghent University (UGent)

AUTHORS: Anna Miglio (ROB)
Carine Bruyninx (ROB)
Andras Fabian (ROB)
Juliette Legrand (ROB)
Stefanie De Bodt (UGent)
Paula Oset Garcia (UGent)
Inge Van Nieuwerburgh (UGent)

ROYAL OBSERVATORY OF BELGIUM

GHENT UNIVERSITY

belspo

.be

Published in 2023 by the Belgian Science Policy Office
WTCIII
Simon Bolivarlaan 30 bus 7
Boulevard Simon Bolivar 30 bte 7
B-1000 Brussels
Belgium
Tel: +32 (0)2 238 34 11
http://www.belspo.be
http://www.belspo.be/brain-be

Contact person: Georges Jamart
Tel: +32 (0)2 238 36 90

Miglio, A., Bruyninx, C., De Bodt, S., Fabian, A., Legrand, J., Oset Garcia, P., & Van Nieuwerburgh, I. ***Open data portal for European and Belgian GNSS reference station data collections, built upon FAIR guiding principles***. Final Report. Brussels: Belgian Science Policy Office 2023 – 26 p. (BRAIN-be 2.0 - (Belgian Research Action through Interdisciplinary Networks)

TABLE OF CONTENTS

ABSTRACT

The Royal Observatory of Belgium (ROB) maintains repositories containing decades of observation data from European (public EUREF repository) and Belgian stations permanently tracking Global Navigation Satellite Systems (GNSS, e.g., GPS or Galileo). They allow precisely measuring ground deformations, monitoring space weather, providing evidence of climate trends, input for numerical weather predictions, etc.

Although various communities already use ROB's public EUREF repository, the procedures to find and access the data were rather complex and non-machine-readable. Despite the fact that the GNSS data originate from a significant number of data providers, provenance information was lacking. Data licenses were only seldom available and no data citation procedures were in place to acknowledge GNSS data providers. Hence, ROB's GNSS data management procedures were in need of an urgent modernization.

Following FAIR data principles as guidelines, the FAIR-GNSS project aimed at: i) facilitating access and reuse of, and increasing trust in, GNSS data; ii) supporting the preservation of GNSS data; iii) contributing to the standardization of GNSS data citation; iv) creating an open data portal for European and Belgian GNSS data. To reach these objectives, the FAIR-GNSS approach focused first on the GNSS data and the accompanying metadata, then on the (meta)data access procedures, and finally on user services such as a data portal.

During FAIR-GNSS, two proposals for GNSS standardized metadata (for GNSS station-specific metadata and for RINEX file metadata) were developed and extended according to feedback and constant interaction with the scientific community. In particular, the proposal for station-specific metadata gained the approval of the GeodesyML task force of the International GNSS Service (IGS).

Tackling the complex issue of GNSS data citation, a proposal for a minimum set of standardized metadata elements accompanying the Digital Object Identifier (DOI) of a GNSS dataset was drafted and iterated with the community and the Global Geodetic Observing System (GGOS) Working Group on DOI. Based on this proposal, a software and a web interface to allow data owners to assign DOI to their data were implemented.

To follow the FAIR principles, the dataflow within ROB's GNSS data repositories was completely revised and improved. A new GNSS database was populated to include missing information on the provenance of the GNSS data, i.e. the RINEX files. This enabled the documentation of checks and corrections performed on the files, e.g. changes in the RINEX files header, correction of the sampling rate of the observations and the potential file issues.

In the conclusive phase of the FAIR-GNSS project, building upon the previous steps, APIs and an Open data portal (beta version) were designed and implemented to enable both humans and machines to access the curated GNSS (meta)data.

Interacting and maintaining engagement with the community during all these processes was key, but also very time consuming. In this regard, the webinar organized to share our experience in applying FAIR data principles and reach out to the community was well received and contributed to disseminate FAIR-GNSS project results.

Keywords: GNSS, Open data portal, FAIR data, Digital Object Identifiers

1. INTRODUCTION

The Royal Observatory of Belgium (ROB) maintains repositories containing decades of observation data from European (public EUREF[1] repository) and Belgian stations permanently tracking Global Navigation Satellite Systems (GNSS, e.g., GPS or Galileo). They allow to precisely measure ground deformations, monitor space weather, provide evidence of climate trends, generate input for numerical weather predictions, etc.

To make such scientific data more suitable for re-use, by both people and machines, funders and publishers worldwide increasingly demand the adoption of **FAIR data principles** [Wilkinson, 2016]. FAIR principles make data more Findable, Accessible, Interoperable, and Reusable (FAIR). However, they are general principles and putting them into practice is not trivial.

The sheer number of new FAIR-related initiatives that emerged in the last couple of years (from FAIR-enabling recommendations to FAIR-assessment tools and metrics) offer guidance, but can become disorienting. Moreover, when moving towards FAIR data, one should consider the overall landscape, i.e. the complete **FAIR ecosystem**, which comprises policies, data management plans, identifiers, the use of standards, and the data repositories [EC, 2018].

ROB's GNSS data repositories were in need of a thorough modernization. Indeed, finding and accessing the data required rather complex and non-machine-readable procedures. In addition, GNSS data originate from a significant number of data providers but data provenance information was lacking and data licenses were hardly ever available. Finally, no data citation procedures to acknowledge GNSS data providers were available.

Moving towards FAIR data requires a community effort, agreement and support from local and international organizations. Focusing on the interest in FAIR data of the broad geophysical research community, both EGU and AGU conferences have nowadays sessions dedicated to FAIR data principles and Open Science. In addition, the European Plate Observing System (EPOS) ERIC[2] is implementing FAIR principles in view of linking to the European Open Science Cloud (EOSC) through ENVRI-FAIR[3].

In this respect, at the start of the project, the GNSS community had just initiated discussions about FAIR data and its implementation, and therefore there was no community-shared approach on how to align GNSS data with FAIR principles.

The FAIR-GNSS project responded to the challenges illustrated above, with the goal to create an Open Data Portal, built on FAIR data principles, and provide open access to the GNSS data in ROB's GNSS data repositories to both humans and machines.

2. STATE OF THE ART AND OBJECTIVES

The EUREF-HDC GNSS data repository is maintained by ROB [Bruyninx et al., 2019; EUREF-HDC, 2022] and contains the daily GNSS observation data (dating from 1996) provided to EUREF by more than a hundred agencies and originating from almost 500 stations located in Europe (see Figure 1). When

[1] EUREF is the IAG Reference Frame sub-Commission for Europe
[2] European Research Infrastructure Consortium
[3] ENVironmental Research Infrastr
uctures building Fair services Accessible for society, Innovation and Research

accessing the EUREF-HDC repository via File Transfer Protocol (FTP) or the Hypertext Transfer Protocol Secure (HTTPS), observation data files are typically grouped by year, day (number) and station name e.g., 2021/012/BRUX00BEL, as it is the case, for most of the GNSS data repositories within the International GNSS Service (IGS) [Johnston et al., 2017].

All these GNSS observation data files are stored in the RINEX format [Romero, 2021; IGS, 2018], a community-agreed text-based standard format used by all GNSS data repositories.

In addition to the EUREF-HDC repository, ROB provides public access to station site logs (the international standard for exchanging station descriptions, maintained and published by the IGS), site pictures, and data license of the station's GNSS dataset. This so-called **station-dependent information** is provided and maintained by the GNSS station managers via the ROB's M³G[4] system [Bruyninx et al. 2019].

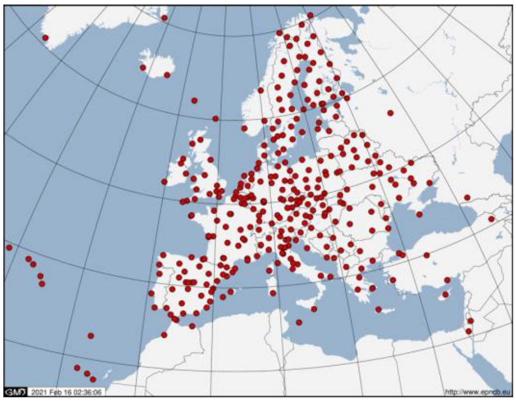

Figure 1: Distribution of the EUREF-HDC GNSS stations (as of Feb. 2021).

As EPN Central Bureau (CB), ROB also maintains **RINEX-dependent information** such as the results of crosschecks of the RINEX file headers versus site logs and daily data quality checks of the daily RINEX data in its repository. The resulting data quality indicators are available in graphical form from the EPN CB web site (https://epncb.oma.be).

At the start of the FAIR-GNSS project,

- the GNSS observation data in the HDC-repository were not downloadable together with the station-dependent information stored in M³G nor the RINEX-dependent information stored at the EPN CB (see Figure 2);
- site logs and pictures were available for all GNSS stations in the HDC, although only a minority of the stations had already inserted a data license for their station's dataset in M³G.

It is then clear that considerable challenges were posed by the need to exploit the available RINEX- and station-dependent information as well as the additional need of HDC users (including station managers) for searching for and downloading data of multiple stations (e.g., search for data based on the last file modification date and obtain information of the actual file modifications, etc.), getting data usage conditions (license), tracking the usage of the EUREF-HDC data.

[4] Metadata Management and Distribution System for Multiple GNSS Networks

The same applies to all GNSS data stored in the Belgian data repository. In addition, this GNSS data repository has never been opened to the public before.

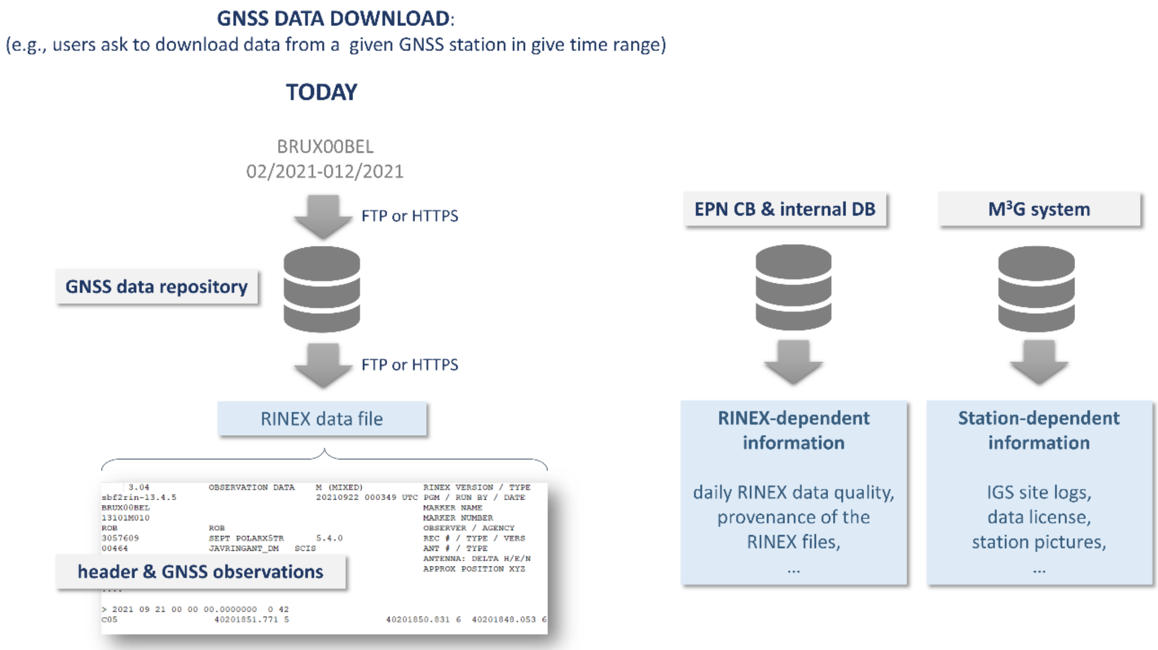

Figure 2: Schematic representation of how users were accessing GNSS data in the EUREF-HDC, at the start of the FAIR-GNSS project. Additional information is stored separately, grouped into RINEX- and station-dependent information.

In short, following FAIR data principles as guidelines, the FAIR-GNSS project aimed at these objectives:

a. facilitating access and reuse of, and increasing trust in, GNSS data;
b. supporting the preservation of GNSS data;
c. contributing to the standardization of GNSS data citation;
d. creating an open data portal for European and Belgian GNSS data.

3. METHODOLOGY

Considering that ROB’s GNSS data repositories are established infrastructures, FAIR-GNSS adopted an approach similar to Bailo *et al.* [Bailo et al., 2020]. They proposed a roadmap to make technologies already in place more compliant with FAIR principles by focusing first on the GNSS data and their **metadata** (additional information), then on the (meta)data access procedures, and finally on user services e.g., visualization tools or web user interfaces such as a data portal.

Therefore, focusing first on GNSS (meta)data, and as recommended by the EU action plan “Turning FAIR into reality [EC, 2018], FAIR-GNSS set as goal the creation of **FAIR Digital Objects** (FDO, data, software, protocols and other resources). FDOs are at the core of the uptake of FAIR data principles and are digital objects that are accompanied by metadata rich enough to enable them to be found, used and cited, and **persistent identifiers** (PIDs) to enable a long-lasting reference to the data.
Complementing this approach, [EC, 2018] also considers a **FAIR ecosystem** as an essential framework. By definition, a FAIR ecosystem encompasses data repositories (ideally CoreTrustSeal-certified [CTS, 2019]) as well as **standards** (such as **communication protocols** to access and exchange data).

In addition, [EC, 2018] encourages data repositories to become **Trustworthy Data Repositories**[5], which are required to make data FAIR and keep it FAIR over time [Lin et al., 2020 & FAIRsFAIR, 2020]. Hence, combining all the above, the general methodology used in FAIR-GNSS covers all project objectives and can be grouped into a preliminary step, three blocks and a conclusive step:

- Assess the level of FAIRness of ROB's GNSS data repositories in their present status and plan the necessary steps to bridge the gaps;
- Turn GNSS data into FDO by attaching metadata and PIDs (objectives a. and c.);
- Restructure the GNSS data repositories to become FAIR-enabling and Trustworthy (objectives a. and b.).;
- Use standard communications protocols to retrieve metadata and data (objective a.).;
- Design and implement an Open data portal built upon GNSS data as FDOs, stored in FAIR-enabling repositories and accessible via APIs (objective d.).

3.1 Gap analysis: FAIR assessment tools

At the initial stage of the project, we could only evaluate our "awareness" of FAIR data principles and the **online FAIRsFAIR's FAIR Aware questionnaire**[6] was perfectly suited for this purpose.

Then, to assess the level of FAIRness of the data hosted by ROB at the start of the project and identify gaps with respect to FAIR data principles, we choose to use the **FAIRsFAIR Data Object Assessment Metrics** [Devaraju, 2021], based on the RDA FAIR Data Maturity Model [RDA, 2020]. This model allows assessing the GNSS (meta)data at all stages of the research data lifecycle (from planning, collecting and processing/analysing data to sharing, preserving/archiving and re-using these data) and map them against the CoreTrustSeal requirements (see Figure 3). Hence, we decided to repeat this assessment at the end of the project to evaluate the progress made.

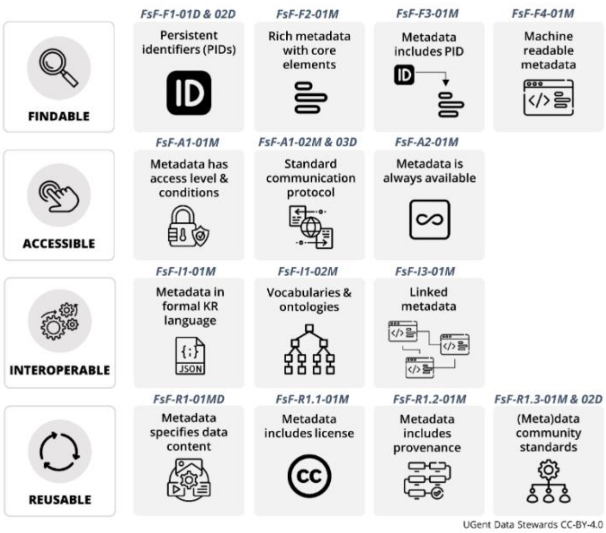

Figure 3: Infographics representing FAIR data principles and the correspondence with the FAIRsFAIR metrics identifier (in italics), with the following naming convention e.g., FsF-F2-01M stands for: FAIRsFAIR – FAIR principle FX – local id regarding metadata (M) or data (D).

[5] based on community-agreed practices, reliable infrastructures, sustainability & committed to data preservation
[6] https://www.fairsfair.eu/fair-aware

3.2 Turn data into a FAIR Digital Object: metadata and PIDs

Turning GNSS data into FDO required attaching rich metadata to the data and, for his purpose, FAIR-GNSS partners had to interact with the GNSS community to agree on the content of the **station-dependent** as well as **RINEX-dependent metadata**, and on how to express and structure this content.

Rich metadata had to also include information on data documentation, data license (to allow users to be aware of the conditions under which data can be accessed, used and shared), and data provenance (how the data were collected and eventually manipulated). To enable interoperability, the content of these metadata had to be expressed using **controlled vocabularies** adopted by the community, e.g., the list of standard names for GNSS receivers and antennas as published by the IGS. Finally, to structure the metadata in a suitable metadata schema, it was important to examine the most appropriate existing (open, generic and domain-specific) community **standards,** e.g. schema.org, CERIF, DCAT and, if necessary, extend this standard(s) to accommodate the specificities of the data.

Completing data with rich and structured metadata is key. Indeed, data will then be compliant with most of the defining FAIR principles (see Figure 3): (F1 and F2) provision of citable, user-oriented, rich and machine-readable metadata; (A1) metadata explicitly including information on access conditions; (I1) metadata expressed via commonly used controlled vocabularies, structured and described via a community-driven well-defined data model; provision of descriptive metadata (R1) as well as provenance metadata (R1.2) to make data reusable, use of license metadata (R1.1) and community standards (R1.3).

Finally, while choosing a metadata schema for GNSS data, FAIR-GNSS took the following into account:

- the perspective of the users accessing and using the data, as well as those managing and preserving the data in the repositories;
- compliance with FAIR principles;
- the European (and Belgian) initiatives mentioned previously;
- best practices at other GNSS data repositories all over the world.

A further step in the direction of turning data into FDO consisted of referencing data with a Globally unique and **Persistent identifier** (PID), which points to the data even if the data are moved to a different repository.

The assignment of PIDs covers all FAIR defining requirements (see Figure 3) and makes data i) Findable as PIDs unambiguously identify and locate a resource (F1), ii) Accessible as metadata should be retrievable by their identifier (A1), iii) Interoperable as dependencies between different datasets are included (I3), and iv) Reusable because of the PIDs role in data citation and, therefore, in dataset provenance (R1.2).

3.3 Restructure data repositories to become FAIR-enabling AND Trustworthy

In addition to finding community agreement on the relevant rich metadata, FAIR-GNSS would also have to ensure that these metadata are available for all the data stored in the repositories. Therefore, the data repositories and the related data flows needed to undergo a thorough restructuring. First, it was necessary to make an inventory of the data and metadata stored in the repositories and associated databases. Secondly, the repositories' databases needed to be redesigned to include the additionally required metadata. Once the new databases were created, the (meta)data had to be reorganized, completed with missing information and mapped onto the selected metadata schemas. This process required the implementation of software to automatize procedures such as data storage, data quality checks, databases population, etc.

In addition, to validate the readiness of ROB's data repositories to seek formal CoreTrustSeal certification, we also decided to assess the GNSS data (infra)structure with respect to the CoreTustSeal+FAIRenabling Capability Maturity (**CapMat**) template [L'Hours, 2022]. The CapMat levels ("Initial", "Managed", "Defined") allow repositories to assess and eventually provide public evidence of their current status and monitor progress towards being FAIR enabling Trustworthy digital repositories.

3.4 Use standard communications protocols to retrieve metadata and data

After focusing on the data, now accompanied by rich machine-readable metadata, stored in the repositories, FAIR-GNSS had to address the procedures to access the data over the internet. These procedures had to allow both humans and machines to search for, retrieve and download GNSS (meta)data and had therefore to focus on machine-actionable mechanisms.

Indeed, data access services are essential in providing FAIR access to the (meta)data i.e. allowing both humans and machines to query the (meta)data (F4), retrieve them via their PIDs (F1) and download them. FAIR data access implies the use of standardized open **communication protocols** and authentication procedures (A1), and the accessibility of metadata even when the corresponding data is no longer available (A2) (see Figure 3).

Therefore, FAIR-GNSS had to consider the development of a **web Application Programming Interface (API)**[7] that responded to the requirements for FAIR data access (well documented, able to interoperate with other web applications, and thus enable the reuse of the repository data). Web APIs allow computer applications to share and access machine-readable (meta)data or subsets of (meta)data, relying on standardized communication protocols such as HTTP and it's corresponding, secure version (HTTPS). The response of the API had to be structured and interoperable at the same time, by using a metadata schema and appropriate (and multiple) formats, e.g., XML, JSON, and the JavaScript Object Notation of Linked Data (JSON-LD).

Finally, FAIR-GNSS had to determine if the (very few) existing APIs for GNSS data were compatible with user needs and with the characteristics of its GNSS data and accompanying enriched metadata.

3.5 Open data portal

Building upon all the steps illustrated above (newly restructured data repositories, GNSS data as FDO, APIs, etc.), FAIR-GNSS had to create an Open data portal meant to function as a centralized platform to provide users with a singular location to search for, locate and download (meta)data. In addition, the open data portal must host the documentation related to the origin and quality of the data, the use of the interface itself and the APIs. As a preliminary step before the portal implementation, FAIR-GNSS had to perform a landscape study to examine what was recently implemented within other FAIR initiatives, at the European level as well as within the GNSS community.

[7] Set of rules and protocols that specifies how two (or more) software component/applications should interact over the internet.

4. SCIENTIFIC RESULTS AND RECOMMENDATIONS

4.1 Gap analysis

Following the approach outlined in the methodology section, we applied two FAIR assessment tools to identify the shortcomings and issues in our data repositories and the data stored therein.

Completing the FAIR Aware questionnaire for the EUREF-HDC at the beginning of the FAIR-GNSS project was the first step and highlighted some of the most relevant challenges in implementing FAIR principles for GNSS data, namely the need for maintaining data FAIR over time and for metadata to

- be expressed in a community-agreed terminology, or in other words, vocabularies to ensure data interoperability;
- be compliant with a community-endorsed standard;
- include provenance information about the collection/generation of data to allow data to be reusable.

As a second assessment tool, we applied the FAIRsFAIR metrics criteria (see Figure 3) and issued a first evaluation of the data and metadata stored in ROB's repositories in their initial status (see Figure 4). It was clear that GNSS data were almost not compliant with some FAIR principles. For example, as mentioned previously, some station managers provided data license for their GNSS data via the M^{3}G system. However, regarding the metadata, except for standards such as the IGS site log and GeodesyML[8] to store the metadata related to GNSS station, there was a complete lack of community-agreed standards for metadata describing the RINEX files.

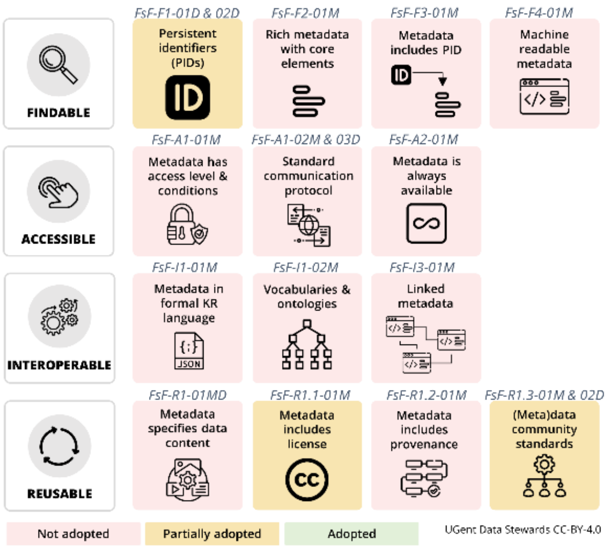

Figure 4: Progress towards FAIR GNSS data as assessed at the start of the FAIR-GNSS project (for the correspondence with the FAIRsFAIR metrics identifiers see Figure 3).

4.2 GNSS FAIR digital object: GNSS metadata

4.2.1 Station-dependent metadata

During the FAIR-GNSS project, the IGS Infrastructure Committee (IGS IC) was strongly willing to move from the (older) IGS station site log format to the GeodesyML format. Both the site log and the GeodesyML file contain the full history of the GNSS station information since the station was set up,

[8] Open-source XML-based format, aligned with international standards such as ISO19115-1:2014 and GML developed within the Open Geospatial Consortium (OGC)

but GeodesyML includes additional information and is easier extendable. Therefore, it was clear that GeodesyML should be used to exchange all station-specific metadata.

However, to comply with FAIR data principles, FAIR-GNSS drafted a proposal to add extra metadata elements to the GeodesyML format, such as data license, file provenance information, and a PID for the GNSS station dataset. Examples of PIDs are **Digital Object identifiers (DOIs)** which are strings (standardized by ISO) playing a key role in standardized data citation and currently used to uniquely identify journal articles, reports, and datasets. FAIR-GNSS discussed its proposal for extending the GeodesyML schema during frequent meetings within the GeodesyML task force of the IGS IC (see ANNEX 2). Because of these interactions, our proposal was updated to include the suggested modifications, and accepted for the upcoming new version of GeodesyML. The proposed (and accepted) changes that will be part of the upcoming version of GeodesyML are listed below and detailed on GitHub:

- provenance information about the station information including:
    - the organization that validated and distributed the station metadata,
    - the publication date of the metadata file,
    - the data license information attached to the file,
    - the URL pointing to the file;
- code lists file (to be maintained at the IGS level) containing standardized list of
    - data centres,
    - GNSS networks;
- information on the GNSS network (regional/global) to which the GNSS station belongs.

4.2.2 RINEX-dependent metadata

Concerning the RINEX-dependent metadata, as a first step, after discussing with GNSS data experts and users, we identified a list of metadata needed by users when searching for and then downloading GNSS observation data (see Table A1 in ANNEX 1).

In the quest of a schema for RINEX-dependent metadata, we first considered generic metadata schemas and vocabularies (DCAT, DataCite, Dublin Core, Schema.org). Finally, we decided to adopt the W3C Data Catalog metadata schema (DCAT), "*an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web*", as the schema of choice and extended the DCAT Application Profile for open datasets and data portals (DCAT-AP) to cover the characteristics and peculiarities of GNSS data. DCAT-AP is the standard for describing Public Sector Information (PSI) as open government data. It is a joint initiative of two Directorates-General of the European Commission, namely DG CONNECT and DG DIGIT, and the Publications Office of the EU. In addition, DCAT-AP is a DCAT extension suitable for Linked Open Data (LOD) and it covers the majority of metadata attributes we planned to associate to RINEX files, including geographic metadata in the form of GeoDCAT-AP. GeoDCAT-AP is in itself an extension of DCAT-AP that follows ISO 19115/19119 standards, and is mainly limited to the description of geospatial datasets in terms of their location metadata.

Our proposed schema, "GNSS-DCAT-AP", was made publicly available on GitHub. By definition, our GNSS-DCAT-AP includes all the mandatory DCAT-AP elements and includes now metadata attributes specific to RINEX files

- new (recommended) metadata classes to describe the specific characteristics of GNSS RINEX observation files such as the type of RINEX file (compression format, frequency, etc.), the content of the RINEX file header and the information regarding the GNSS station, the GNSS antenna and receiver associated with the station, etc.
- information on data provenance and on data license.

As DCAT-AP is independent from the serialization formats, we decided to make the metadata available in JSON and JSON-LD. JSON-LD is indeed recommended by W3C, used by Schema.org, and offers "context" such that machines can understand what a particular entity is about.

4.3 Persistent Identifiers

As mentioned in Section 3.2 of this report, an obvious choice for a PID for dataset is DOI.

In the GNSS community, where attributing DOIs to GNSS data is not common practice and the merit of the data providers is unrecognized, DOIs would allow data providers to receive credit through citation metrics and provide statistics on data usage to funders. In addition, through DOIs, users are aware of data access restrictions and cross-linking through various DOIs becomes possible (e.g. link publications and underlying data, datasets with other datasets via hierarchical relations, visualize relationships via PID graphs [Fenner, 2019]).

DOI for datasets are widely used within various scientific communities: the seismological community (FSDN recommendations for DOIs [Clark, 2014]), the GNSS community (UNAVCO minting DOIs for GNSS datasets, GFZ, CDDIS, ROB), the Earth science community (EPOS plan to attribute DOIs to all its datasets), and the geodetic community (GGOS WG on DOI for geodetic datasets).

The importance of interacting with the GNSS community is then apparent as well as the considerable time and effort that it requires.
Since the beginning of our FAIR-GNSS project, we interacted with the GGOS WG on DOI (see ANNEX 2) to raise awareness on the importance of assigning DOI to GNSS datasets, to address issues related to the complexity of assigning DOI to such data (e.g., data are dynamic), and move towards a community agreement. Indeed, GNSS datasets originate from a very large amount of GNSS stations, organized in global, regional, local networks where a station can contribute to different networks. It is therefore difficult to establish the most appropriate level of granularity of DOI for GNSS data. We discussed within the GGOS WG on DOI to propose a practical first approach to enable GNSS data citation, and finally agreed on the level of granularity: assign a DOI for the GNSS data originating from a given GNSS station.
In addition, the station data DOI may coexist with GNSS network DOIs[9], which in turn, could be then linked to the DOI of GNSS stations (via DataCite metadata properties such as relatedIdentifier, IsPartOf, HasPart).
To create (mint) a DOI, other than a DOI name (prefix and suffix) and a durable URL (landing page) to provide access to the data and a full bibliographic citation, one needs to fill in a DOI metadata schema (e.g., DataCite) with (mandatory, recommended and optional) metadata. As ROB can mint DOI via TU

[9] provide reference to the datasets of all GNSS stations belonging to the network as well as citation (and credit) for the group or institution operating and managing such a network

Delft as DOI allocation agency for DataCite, we used the DataCite DOI metadata schema to structure the information that accompany a DOI (e.g., publisher, etc.)

Regarding the standardization of DOI metadata for GNSS data, to make the information included in the DOI easily interpretable by machines, the GNSS community needed to agree on:

- the optional fields in the metadata schema (DataCite v4.4 [DataCite, 2021]) to be made mandatory,
- the controlled vocabularies to use to populate the various fields e.g., indicate a GNSS station by its 9-char id,
- where to source the information regarding the station to automatically populate the fields whenever possible.

The elaborated concrete proposal, currently under discussion with GGOS DOI WG, focused on

- harmonization within the community as a goal;
- FAIR principles as key guidelines;
- retrieving metadata from GeodesyML as much as possible;
- including PIDs, like ROR, ORCID, DOI in the DataCite schema and defining relation types
- developing recommendations (along the lines FDSN recommendations for seismic network DOIs (https://doi.org/10.7914/D11596)) for the content of specific DataCite fields that can be also used beyond GNSS data (e.g. repository = Publisher, agency = Creator, local partners = Contributors)

Based on this proposal, a software and a web interface, facilitating the minting of DOIs of GNSS data in ROB's data repositories, were developed and documented in ROB's M^{3}G portal. The next step will consist in contacting the GNSS station managers to ask their agreement that ROB's is minting DOIs for the datasets of their GNSS stations. This process has already started and will continue after the end of this project.

4.4 Re-organisation of data repositories & collection of required metadata

As illustrated in Section 3.3, to include the missing information missing in the EUREF-HDC repository, ROB had to completely restructure its repository and the corresponding data flows and codes.

First, we tackled the challenge to identify the best approach to structure and populate a new database for the EUREF-HDC containing new additional (RINEX- and station-dependent) metadata. Secondly, we implemented all software to populate this new database. In parallel, we applied the same criteria and procedures to the Belgian GNSS data repository.

Regarding the station-dependent metadata, all information, including an historical overview of the stations' setup, was (and is) collected by ROB's M^{3}G system. These data, stored in a database, are accessible via stable URLs and exported in the GeodesyML format. Additionally, the M^{3}G database needed to be completed with data license and data citation information as data license information was (and is) optional in the M^{3}G system. However, by encouraging station managers to attach a license to their data, finally we obtained data license information for the data of all stations in the Belgian repository and 93% of the GNSS stations in the EUREF-HDC repository.

Regarding the RINEX-dependent metadata, as illustrated previously, the gap analysis highlighted the need for information about the provenance of the data. In addition to this, there was the long-standing need to homogenize the use of flags in the error checking procedures of the EUREF-HDC, extend it to the Belgian data, and define more transparent criteria for publication of the RINEX files. Also, as RINEX data contain station operator names, this contact information was replaced by the non-personal contact of the agency, to be GDPR compliant.

The development of a new backend for both repositories began in 2021 and consisted of multiple development cycles, including the complete revision of the software to

- download RINEX data,
- verify and correct RINEX headers,
- verify and correct RINEX formatting errors and manage flags,
- populate the databases

as well as its documentation and testing.

A completely new database (including new fields, see Table A3 in ANNEX 1, in italics) was designed and populated (see Figure 5) . As from January 1st 2023, provenance information was collected for all RINEX data and also for new incoming RINEX files. Information on file changes and corrections are documented in the files themselves, while potential file issues are identified and included in the database.

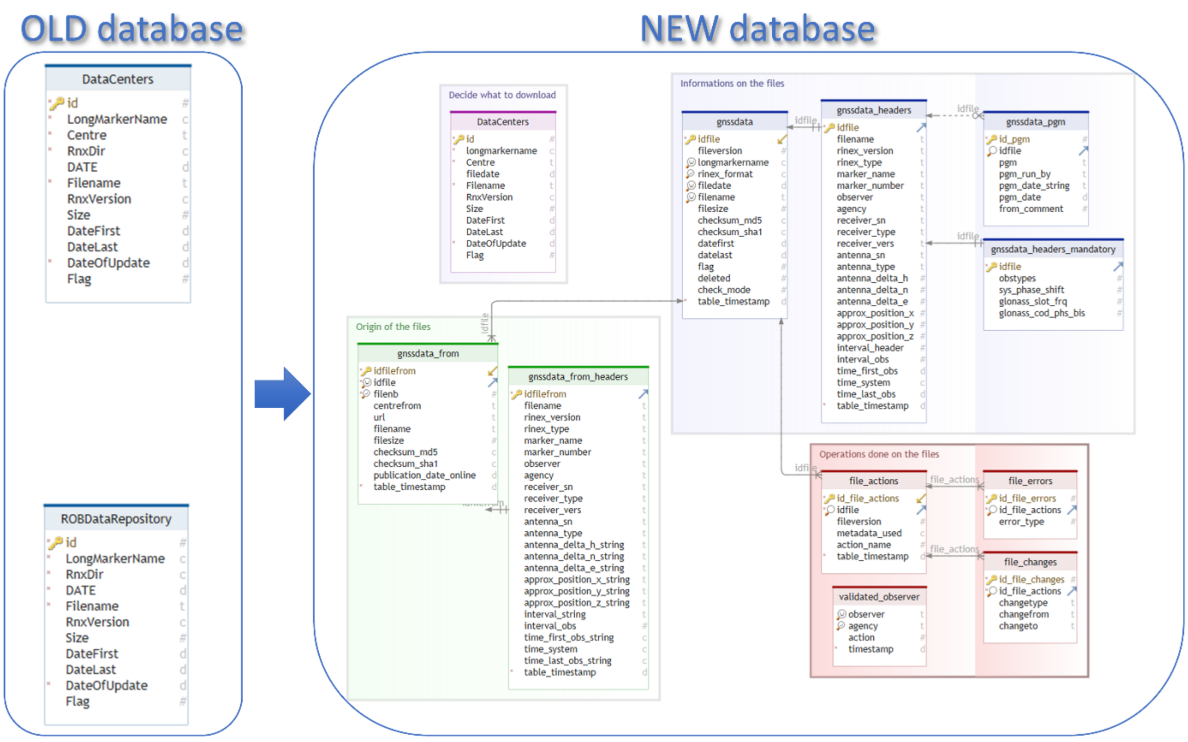

Figure 5: New database including new fields (see Table A3 in ANNEX 1)

To obtain these results, a close collaboration with the Regional EUREF Data Centers was necessary. Indeed, during the duration of this project, data centers were undergoing major restructuring procedures and we could not obtain straightforwardly the necessary persistent links to original data.

Additional information, provided to facilitate data search, was obtained from existing databases associated to the EUREF-HDC one (see Table A3 in ANNEX 1). This information was mapped onto the GNSS-DCAT-AP fields and provided to the users as structured metadata via API and the portal web interface (see sections 4.5 and 4.6).

At the beginning of the project, we applied the CapMat self-assessment template to the data repositories, which helps monitoring progress towards being FAIR enabling Trustworthy digital repositories and, at the same time, complements the FAIRsFAIR metrics. The great part of the 16 requirements of the CapMat template could be labelled as “Managed”, although a few were still labelled as “Initial” because of the ongoing implementation regarding access to the data and metadata and their documentation (e.g., criteria “13. Data discovery and identification”, “14. Data reuse”). Finally, following the restructuring of the repositories, we re-applied the CapMat and all the 16 requirements of the CapMat template could be labelled at least as “Managed”.

Regarding data preservation, in addition to the data cleaning and curation procedures mentioned above, BELSPOʼs Long-Term Preservation (LTP) platform was selected as the archiving platform for the long-term preservation of FAIR-GNSS data. Part of the Belgian data were copied on the LTP while the rest is awaiting the reorganization of the platform and final data archiving workflows.

4.5 API

As the last step in our approach, we focused on web APIs to access GNSS (meta)data stored in our repositories.

First, we considered the web APIs presently in use within the GNSS community and we noticed that, in general, there are (very) few non-Beta/operational APIs to download GNSS data, there is no common approach (e.g., depending on the request, metadata are output as JSON/CSV or simple text files with links to data files), and the (few) metadata attached to the downloaded data are not expressed in a standard format.

Then, we identified the key criteria for the design of FAIR API for GNSS (meta)data: i) make sure to provide all necessary metadata about the data, ii) use standards to ensure machine-readability and interoperability with other data, iii) allow to search by criteria based on user needs.
In this regard, the GNSS-DCAT-AP, our proposed metadata schema, provides rich GNSS metadata in a standardized structured form and therefore responds to criteria i) and ii). To address FAIR API design criterion iii), we prepared use cases for search criteria when retrieving station information and RINEX data (see Table A4 in ANNEX 1).

Among available web APIs, we focused on the Representational State Transfer (REST) API architectural style. RESTful API provides a way for applications to communicate with each other over HTTP(S) and use the standard HTTP status codes to indicate the success or failure of the requests. In addition, HTTP RESTful APIs are widely used and easily described and documented with the open-source OpenAPI. Based on all the considerations above, we implemented and documented a HTTP RESTful web API for GNSS data to interact with the data catalogue of the newly restructured EUREF-HDC and Belgian repositories.

4.6 Open Data Portal

The developed Open Data Portal provides access (through a GUI and API) to the GNSS observation data within the two GNSS data repositories maintained by ROB:

- the Belgium repository containing the data of Belgian GNSS stations contributing to the European Plate Observing System
- the EUREF Historical EPN Data Centre (EUREF-HDC) containing GNSS data of the EUREF permanent GNSS network (EPN)

The provided GNSS observation data are accompanied by the full station descriptions and the standardized metadata, data usage conditions, and data citation information in an effort to make the data FAIR. The open data portal includes a web interface (see Figure 6) that

- provides direct access to the GNSS station information: users can display the stations on a map or consult and visualise the station info sheet including the station site log, site pictures, information on available data, data quality plots of the RINEX files , etc...
- allows to download daily 30-sec sampled GNSS data in the RINEX 2 or RINEX 3+ format using the web portal GUI or API. In addition to the RINEX data themselves, also the station-dependent and RINEX-dependent metadata, as developed in FAIR-GNSS, are available for download, in-line with FAIR data principles. The portal offers users the possibility to download the selected RINEX files as well as the corresponding metadata in various formats (simple text, JSON and JSON-LD) via simple scripts, accompanied by user documentation. In addition, the human-friendly web interface of the data portal allows users to search for the data in multiple ways (Figure 7), for example by station location (e.g., country, coordinates), by date range, last modification date, RINEX format versions, etc…
- provides a direct link to ROB's Metadata Management and Distribution System for Multiple GNSS networks (M^{3}G)

Although not part of the FAIR-GNSS project, the portal also provides a direct link to ROB's real-time broadcaster that distributes real-time GNSS data streams using the principle of internet radio.

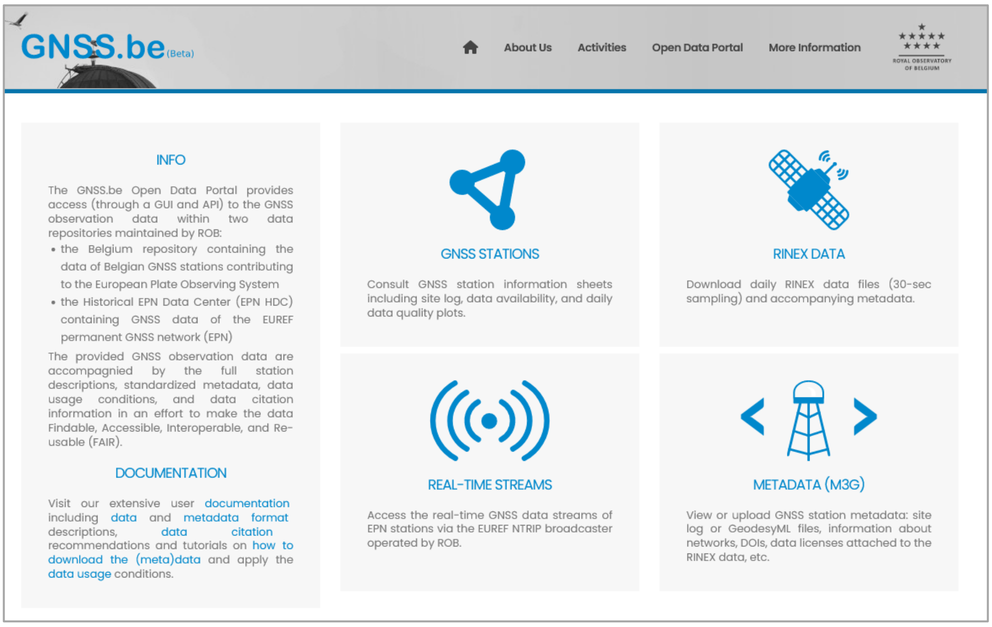

Figure 6: Screenshot of the main page of the open data portal.

The portal is still undergoing last internal testing. Once it has passed the internal testing (expected within Nov. 2023), the portal will be available from https://www.gnss.be/opendataportal.

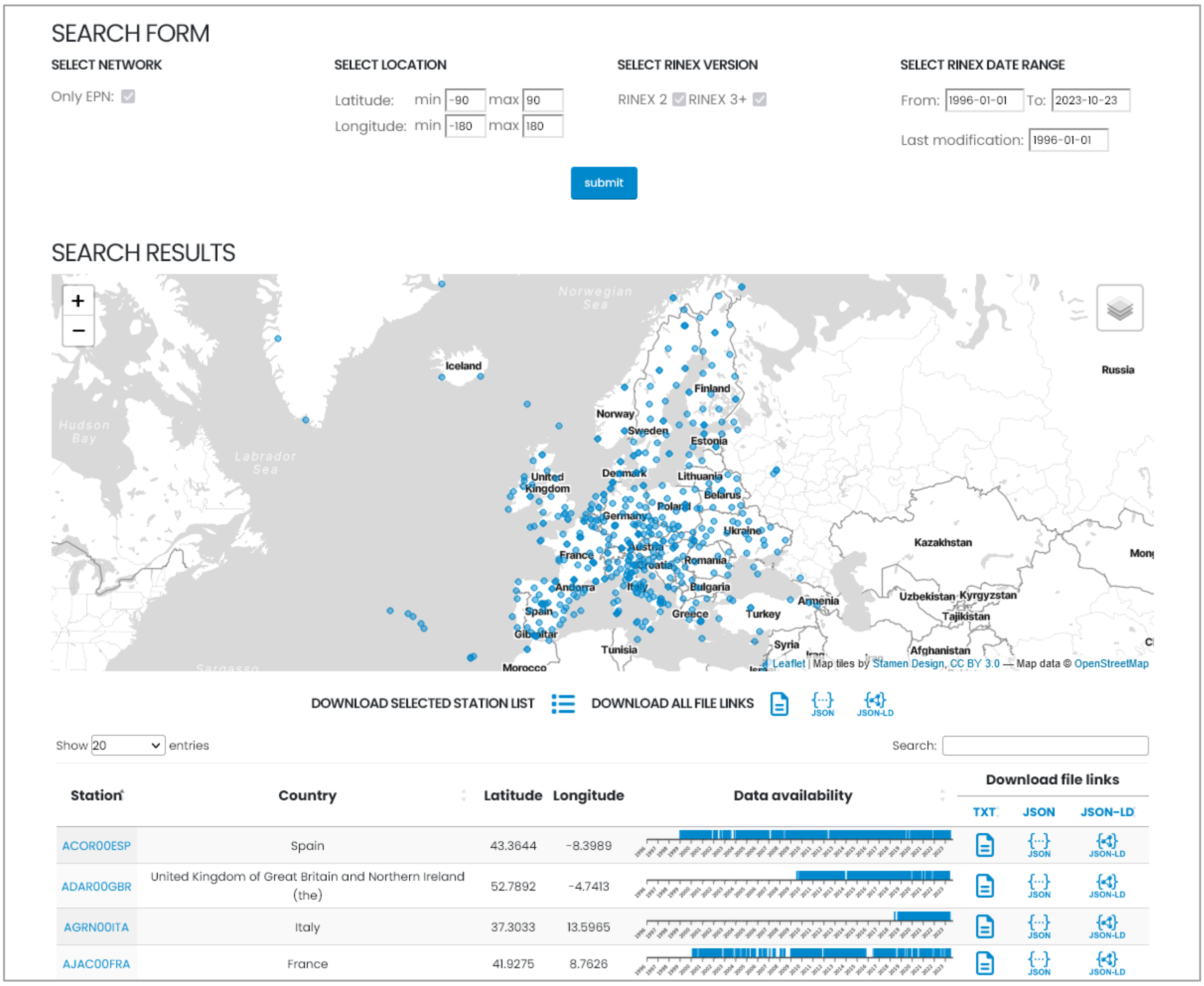

Figure 7: Screenshot of the data access page on the open data portal.

4.7 Summary and recommendations

In summary, we could address the challenges posed at the beginning of the project by turning ROB's GNSS data into FDOs that are downloadable (via web API and the portal web interface) together with the corresponding additional station- and RINEX-dependent information (Figure 8) (FAIR-GNSS objectives a. , see end of Section 2).

GNSS data as FDOs provide both human and machines with querying capabilities built upon rich standardized metadata, long-lasting and citable references to the data, information on license of use as well as documentation on data access procedures.

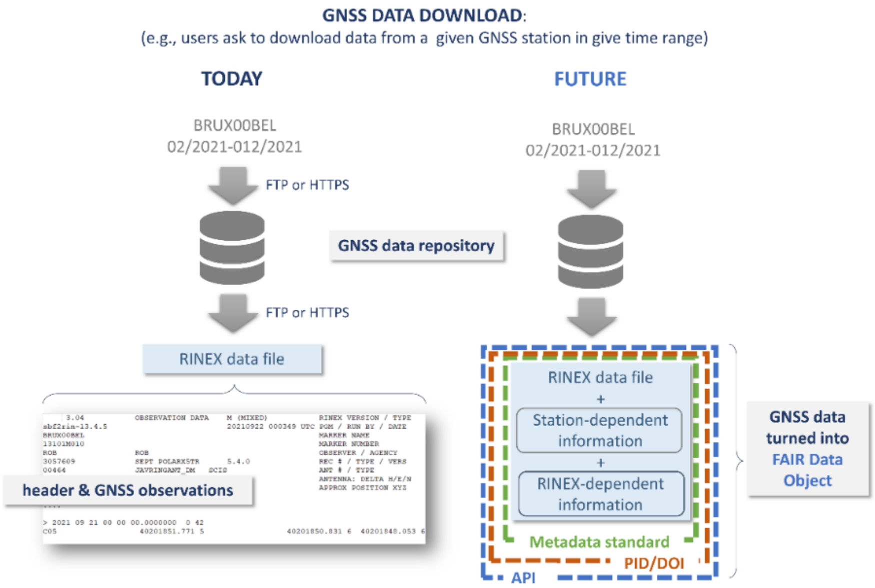

Figure 8: Schematic representation of how GNSS data in ROB's repositories, together with additional information, will be accessible as FDO.

FAIR-GNSS updated the management of ROB's GNSS data repositories and pioneered the application of FAIR data principles within the GNSS community. By doing so, we moved towards trustworthy GNSS research data management and data repositories (hence improving ROB's public services and data preservation, FAIR-GNSS objective b.), and encouraged the rest of the GNSS community to do the same.

The process of restructuring the ROB's GNSS data repositories and its databases was indeed time consuming, but it also enabled the identification of unexpected issues and enriched the understanding of various type of errors in the workflows and in the GNSS data itself.

Feedback and constant interaction with the scientific community required considerable effort and time, but were key to the realization of FAIR-GNSS objectives. It has to be mentioned that we strongly underestimated the time it takes to understand how to put FAIR principles into practice, to navigate among different metadata standards, to interact with the community, e.g. with data providers concerning DOI and data license, with various working groups to agree on standards. Due to the limited duration of the project, it was clearly not advisable to wait for the community to agree on the metadata standards we proposed.

As an example, in the context of assigning DOI to GNSS data, FAIR-GNSS effectively contributed to reach a baseline agreement within the scientific community on a proposal and, on this basis, we could implement software to mint DOI for GNSS datasets (FAIR-GNSS objective c.). Indeed, reaching a final community agreement on standardization of data citation is an extremely long and complex process. However, as aligning with FAIR principle can only be achieved by degrees, one has to proceed with a first implementation and keep on interacting with the community and adapt accordingly. The same applies to all FAIR-GNSS proposals for metadata standardization.

It should be apparent that the process of applying FAIR data principles i.e. implement a FAIR ecosystem requires investing resources (including considerable manpower), close collaboration with experts (e.g., data stewards) and with the scientific community, building new expertise on e.g., metadata, planning in the long term, and a continuous assessment.

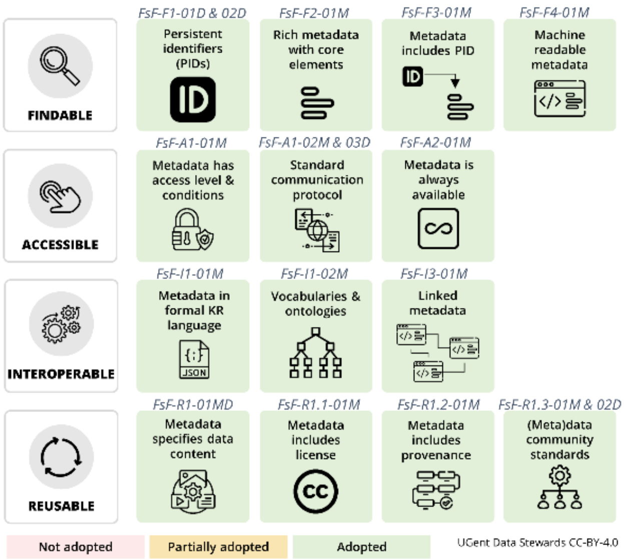

Figure 9: Progress towards a FAIR GNSS data as of the present day (for the correspondence with the FAIRsFAIR metrics identifiers see Figure 3).

Nonetheless, at each new development cycle, the level of FAIRness increases (see Figure 9), the repositories evolve towards becoming trustworthy data repositories, and the core service that ROB has been offering to the EUREF and Belgian GNSS communities improve.

FAIR-GNSS standardized metadata, APIs and newly developed data portal enable easier access to GNSS data, which is the basis for all (public and private) georeferencing applications. In addition, FAIR-GNSS enabled scientists to locate the data required for their research, often aiming at improving the monitoring and understanding of environmental phenomena (ground deformations, climate change).

5. DISSEMINATION AND VALORISATION

27 May 2021 - In the context of the 2021 EUREF Symposium, we organized a splinter meeting ("Towards FAIR GNSS data") to introduce the EPN community to the advantages of FAIR data principles (S. De Bodt "Introduction to FAIR data") and the challenges in applying them to GNSS data (C. Bruyninx "Towards FAIR GNSS data").

27 January 2022 – "Metadata for FAIR-GNSS" - Presentation at FRDN Knowledge Hub

27 May 2022 – Participation to the EGU 2022 General Assembly "Proposed metadata standards for FAIR access to GNSS data" (virtual presentation) within the EGU 2022 ESSI2.3 session. FAIR-GNSS results in terms of proposals for GNSS metadata standardisation and DOI for station daily GNSS data were presented at the EGU General Assembly 2022 (May 25-27 2022).

2 June 2022 – To continue raising awareness about FAIR data principles and illustrate the progress and reulsts of the FAIR-GNSS project to the GNSS community, we organized a splinter meeting during the 2022 EUREF Symposium "Towards FAIR GNSS data", which included an invited talk by K. Elger (chair of the GGOS WG on DOIs).

11 October 2022 - In collaboration with FAIR-GNSS partners, we organized a webinar ("Putting the FAIR principles into practice: the journey of a GNSS data repository"). The webinar attracted great interest: 200 registered participants (from 45 different countries, see Figure 10) including several members of Belgian FSIs.

Figure 10: Location of the registered participants of the FAIR-GNSS webinar.

The webinar main goal was to share our experience in applying the FAIR data principles, reach out to the scientific community, contribute to disseminate FAIR-GNSS results, get feedback on FAIR-GNSS proposals, and foster possible future collaborations experts in the field.

Success in adopting FAIR principles is also dependent on the engagement of the community. Consequently, we devoted a great deal of effort in interacting with the scientific community. This took more time than expected.

We were and are still actively involved in the GGOS DOI WG e.g., illustrating the FAIR-GNSS project, discussing its results, and contributing to the GGOS DOI WG presentations. We are also involved in the GeodesyML Task Force.

The use of GitHub repositories will continue to allow the community to give feedback, access the documentation on our proposals and encourage the active participation of different stakeholders.

A FAIR-GNSS project website (https://fair-gnss.oma.be) was created and kept up to date. It includes links to the GitHub repositories of the metadata proposals and to outreach material e.g., the webinar recording and slides.

6. PUBLICATIONS

6.1 Non-peer-reviewed publications

Bruyninx, C., De Bodt, S., Fabian, A., Legrand, J., Miglio, A., Oset Garcia, P., & Van Nieuwerburgh, I. (2023). ROB-GNSS/GNSS-DCAT-AP (Version V0.2). https://doi.org/10.5281/zenodo.7777568

Elger, K., Miglio, A., Bruyninx, C., Thaller, D., & GGOS DOI Working Group. (2022). Concepts for DOI minting for Geodetic Datasets. https://doi.org/10.5281/zenodo.7239189

Elger, K., Miglio, A., Bruyninx, C., & GGOS DOI Working Group. (2023). News from the GGOS DOI Working Group - Presentation slides. https://doi.org/10.5281/zenodo.7785923

Elger, K., Miglio, A., Bruyninx, C., & GGOS DOI Working Group. (2023). Why do Geodetic Data need DOIs? An introduction to data publications and the GGOS DOI Working Group. Zenodo. https://doi.org/10.5281/zenodo.7785720

6.2 Conference contributions

26-28 May 2021 - Organisation of a splinter meeting during the EUREF Symposium 2021: “Towards FAIR GNSS data” (virtual)

27 January 2022 – “Metadata for FAIR-GNSS” - Presentation at FRDN Knowledge Hub

27 May 2022 – "Proposed metadata standards for FAIR access to GNSS data" - Participation to the EGU 2022 General Assembly (virtual presentation) within the EGU 2022 ESSI2.3 session.

2 June 2022 - Organisation of a splinter meeting during the EUREF Symposium 2022: “Towards FAIR GNSS data” (virtual)

4 November 2022 – “Belgian metadata catalogue for permanently tracking GNSS stations in Europe” - Poster presented at BNCGG study day, Brussels

7. ACKNOWLEDGEMENTS

We thank BELSPO for funding FAIR-GNSS.
We are also grateful to the members of the follow-up committee for the fruitful discussions and their support: Eric Laureys, Belgium (BELSPO representative), Martin Lidberg, Sweden (president of EUREF), Elisabetta D’Anastasio, New Zealand (geodetic processing specialist at GNS Science), Kirsten Elger, Germany (chair of GGOS DOI WG) Eric Pottiaux, Belgium (senior researcher at ROB), Youssef Tawk, Switzerland (Business Development Manager at the GNSS division of Leica Geosystems), Michał Serdeń, Poland (Project Manager at TomTom), Karin Kollo, Estonia (Head of the Dpt. of Geodesy of the Estonian Land Board), and Tanja Milotic, Belgium (Project Coordinator LifeWatch at the Research Institute for Nature and Forest (INBO).
Finally, we would like to thank and M. Bradke (chair IGS infrastructure Committee) for feedback and constructive discussions.

REFERENCES

Wilkinson et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

European Commission, Directorate-General for Research and Innovation (2018) Turning FAIR into reality : final report and action plan from the European Commission expert group on FAIR data, Publications Office. https://data.europa.eu/doi/10.2777/1524

Bruyninx et al. (2019). GNSS metadata and data validation in the EUREF Permanent Network, GPS Solutions. 23. https://doi.org/10.1007/s10291-019-0880-9

Bruyninx et al. (2022). EUREF Permanent GNSS Network Historical Data Center [Data set]. Royal Observatory of Belgium. https://doi.org/10.24414/ROB-EUREF-HDC

Romero (2021) RINEX: The Receiver Independent Exchange Format Version 4.00, https://files.igs.org/pub/data/format/rinex_4.00.pdf (2021)

Bailo et al. (2020) Perspectives on the Implementation of FAIR Principles in Solid Earth Research Infrastructures. Front. Earth Sci., Sec. Solid Earth Geophysics. https://doi.org/10.3389/feart.2020.00003

CoreTrustSeal Standards and Certification Board (2019). CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022 (v02.00-2020-2022). Zenodo. https://doi.org/10.5281/zenodo.3638211

Lin et al. (2020) The TRUST Principles for digital repositories. Sci Data 7, 144. https://doi.org/10.1038/s41597-020-0486-7

FAIRsFAIR (2020) D4.2 Repository Certification Mechanism: a Recommendation on the Extended Requirements and Procedures (1.0). Zenodo. https://doi.org/10.5281/zenodo.5360937

Mons et al. (2017) Cloudy, Increasingly FAIR; Revisiting the FAIR Data Guiding Principles for the European Open Science Cloud. Information Services & Use 37, 1. https://doi.org/10.3233/ISU-170824

Devaraju et al. (2021) From Conceptualization to Implementation: FAIR Assessment of Research Data Objects. Data Science Journal, 20(1). http://doi.org/10.5334/dsj-2021-004

RDA FAIR Data Maturity Model Working Group (2020). FAIR Data Maturity Model: specification and guidelines. Research Data Alliance. https://doi.org/10.15497/RDA00050

Devaraju and Huber (2020) F-UJI - An Automated FAIR Data Assessment Tool (v1.0.0). Zenodo. https://doi.org/10.5281/zenodo.4063720

Fenner and Aryani (2019). Introducing the PID Graph (Version 1.0). https://doi.org/10.5438/JWVF-8A66

Clark et al. (2014), https://www.fdsn.org/pdf/V1.0-21Jul2014-DOIFDSN.pdf

DataCite Metadata Working Group. (2021). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs. Version 4.4. DataCite e.V. https://doi.org/10.14454/3w3z-sa82

ANNEXES

ANNEX 1 – TABLES

Table A1 – List of information users would like to get when downloading RINEX files

User requested metadata for RINEX files
Link (URL) to the RINEX file;
Link to the station-specific metadata file (via PID/DOI);
MD5/SHA1 checksum;
Access and re-use conditions to the files (open or not, data license, embargo period);
Sampling rate (typically 30-second for daily observations, but it can be 1-second or 15-second);
Timestamp for data & metadata;
Type of RINEX (daily or hourly, RINEX version);
Satellite constellations for which observations are included in the RINEX file;
Receiver & antenna types (as in GeodesyML or site log);
Changes made to the original RINEX file(s) (e.g. to correct an incorrect antenna name in the RINEX header);
Data quality metrics (e.g. as provided presently in graphical form by the EPN Central Bureau);

Table A2 - Information related to GNSS data available at the EUREF-HDC, grouped into RINEX- and station-dependent information.

RINEX-dependent information	**Station-dependent information**
File name	Station description (site log)
RINEX format version	Station pictures
File size	Data license
Date of observations	
Creation date of the file	
Date of last file update	
Data quality indicators	

Table A3 – Fields added to the new databases

<table>
<tr><th rowspan="2"></th><th colspan="4">Rationale for including new fields in the database</th></tr>
<tr><th>Decide
what to download</th><th>Document
origin of the
data file</th><th>Document
operations done
on each data file</th><th>Document
publicly made available data
files to facilitate searching</th></tr>
<tr><td rowspan="6">Fields existing at the beginning of the project</td><td>File name</td><td rowspan="6"></td><td rowspan="6"></td><td>File name</td></tr>
<tr><td>RINEX format</td><td>RINEX format</td></tr>
<tr><td>File size</td><td>File size</td></tr>
<tr><td>Date of observations</td><td>Date of observations</td></tr>
<tr><td>Creation date of file</td><td>Creation date of file</td></tr>
<tr><td>Last update time of file</td><td>Last update time of file</td></tr>
<tr><td rowspan="7">New additional fields</td><td rowspan="7"></td><td>Persistent link to original data file (Data center name + File name)</td><td>RINEX header changed or not</td><td>MD5 checksum</td></tr>
<tr><td>File size at data center</td><td>List of changes in RINEX header</td><td>File version (# updates in database)</td></tr>
<tr><td>MD5 checksum at data center (not available)</td><td>Date of change of RINEX header</td><td>List of programs used to generate RINEX</td></tr>
<tr><td>Last update time of file at data center</td><td>Reference info used for changing RINEX header: Station description from M3G</td><td>Content current (corrected) RINEX header content</td></tr>
<tr><td rowspan="3">Original RINEX header at data center</td><td>GDPR-related changes</td><td rowspan="3">From associated (existing) database:
• Station descriptions from M3G
• Precise positions of each station
• GNSS data quality metrics</td></tr>
<tr><td>Final content RINEX header</td></tr>
<tr><td>Warning/error flags</td></tr>
</table>

Table A4 – Use case criteria for APIs

Use cases – search criteria	
Retrieve station information	**Retrieve RINEX data**
A.2 – based on time range	B.1 – that have been changed since a specific date
A.3 – based on type of receiver/antenna/firmware version and date range	B.2 –based on time range
A.4 – for stations with an external clock A.5 – for stations with meteorological instruments	B.3 – based on a specific RINEX version
	B.4 – based on type of antenna/receiver
	B.5 – including specific signals
	B.6 – including specific satellite systems or numbers
Criteria based on GNSS user feedback	
– based on monument type	
– based on co-located geodetic instrumentation ("Local Survey Tie" and "Collocation Information")	
– based on responsible agency and/or operator	

ANNEX 2 – PARTICIPATION TO MEETING OF THE GGOS DOI WG AND THE GEODESYML TASK FORCE

GeodesyML Task Force meeting, online, 30/11/2021

GeodesyML Task Force meeting, online, 24/02/2022

GeodesyML Task Force meeting, online, 31/03/2022

GeodesyML Task Force meeting, online, 18/05/2022

GeodesyML Task Force meeting, online, 28/09/2022

GeodesyML Task Force meeting, online, 17/11/2022

GGOS DOI WG meeting, online, 10/02/2022

GGOS DOI WG meeting, online, 14/03/2022

GGOS DOI WG meeting, online, 12/05/2022