Putting the FAIR principles into practice: the journey of a GNSS data repository

Webinar co-organized by Royal Observatory of Belgium and University of Ghent

October 11, 2022

belspo

ROYAL OBSERVATORY OF BELGIUM

UNIVERSITEIT GENT

ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

Welcome

Who we are

- Royal Observatory of Belgium (ROB) → maintaining a publicly available data repository containing GNSS data

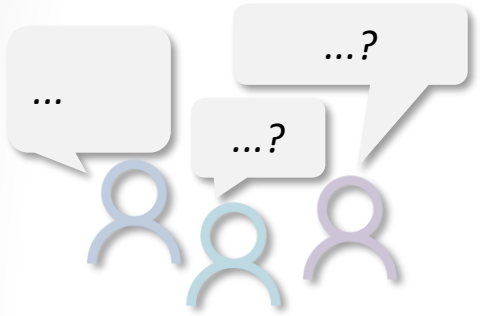
  C. Bruyninx, A. Fabian, J. Legrand, A. Miglio, A. Moyaert, D. Mesmaker, F. Bamahry, F. Bodranghien, E. Pottiaux

- Ghent University → open data specialists

  S. De Bodt, P. Oset Garcia, I. Van Nieuwerburgh

Demands from data users and managers

GNSS user community



- **Search for and download data** from GNSS multiple stations
- Get **data usage conditions**
- **Acknowledge (cite) data providers**

GNSS station managers

- **Need for better recognition of data provision, but no data citation** enabled…
- Track usage of data (funding agencies)
- **No data license information provided**

Demands from organizations

GNSS organizations

IGS INTERNATIONAL GNSS SERVICE

GGOS Global Geodetic Observing System

EUREF

Scientific community

EGU AGU EPOS EUROPEANPLATEOBSERVINGSYSTEM

EU policy actions

eosc

European Open Science Cloud

- **maximize interoperability and discoverability** of geodetic products and services…
- and more explicitly: **implement FAIR data principles**

EUREF 2021 Resolutions

Resolution No. 2.

The IAG Reference Frame Sub-commission for Europe (EUREF)

considering that major funding bodies, including the European Commission, promote and require the implementation of FAIR (Findable, Accessible, Interoperable, and Reusable) data principles

and recognising that FAIR data principles increase the value and the reuse of digital resources, by humans as well as machines

encourages the EUREF community to adopt these principles in all aspects of data management

Solution? Apply FAIR data principles?

Implementing FAIR data principles is challenging!

- FAIR data principles first introduced by Wilkinson in 2016 : **general principles and putting them into practice is not trivial**
- **Overwhelmed by papers and initiatives** about FAIR data principles
- Few hand-on papers
- Moving towards FAIR data requires

  "***substantial change in practice, technologies and their implementation***"

Asked help from open data experts at University Gent
and decided to ask funding from Belgian Science Policy Office → **FAIR-GNSS project**

FAIR-GNSS project

- Partners: **Royal Observatory of Belgium & Ghent University**
- Funding Agency: **Belgian Science Policy Office (BELSPO)**
- Duration: **15/12/2020 – 14/03/2023**
- Budget: 35 PM funded
- Goal: **Create an open data portal for European and Belgian GNSS reference station data collections, built upon FAIR guiding principles**

Agenda

Part 1 - GNSS data and the data repository

Part 2 - Introduction to FAIR data principles

Part 3 – The journey so far: FAIR for GNSS data

break

Part 3 – The journey so far: FAIR for GNSS data (cont.)

Part 4 – Lessons learned and next steps

Discussion - Q&A

Agenda

Part 1 - GNSS data and the data repository

Part 2 - Introduction to FAIR data principles

Part 3 – The journey so far: FAIR for GNSS data

break

Part 3 – The journey so far: FAIR for GNSS data (cont.)

Part 4 – Lessons learned and next steps

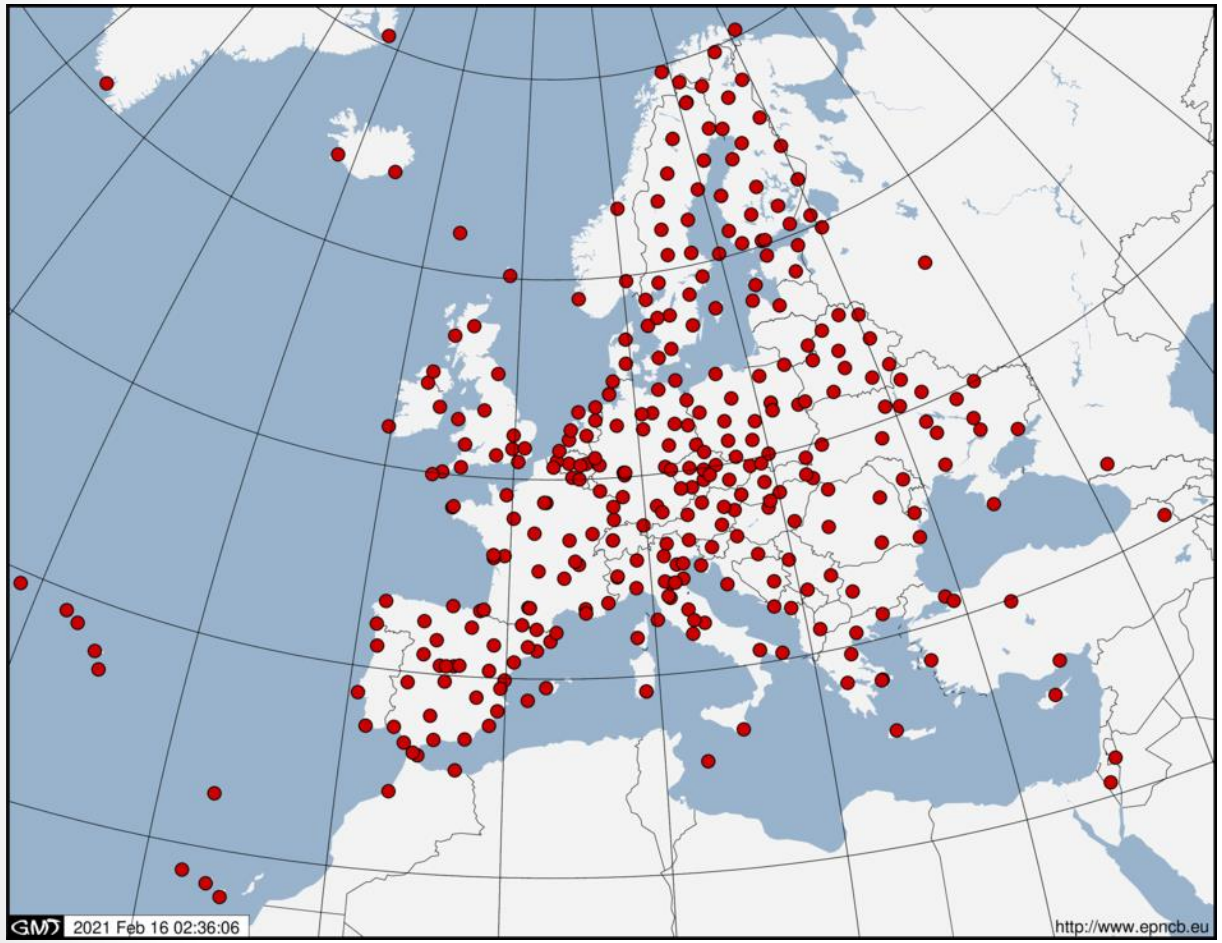
Discussion - Q&A

Part 1 - GNSS data and the data repository

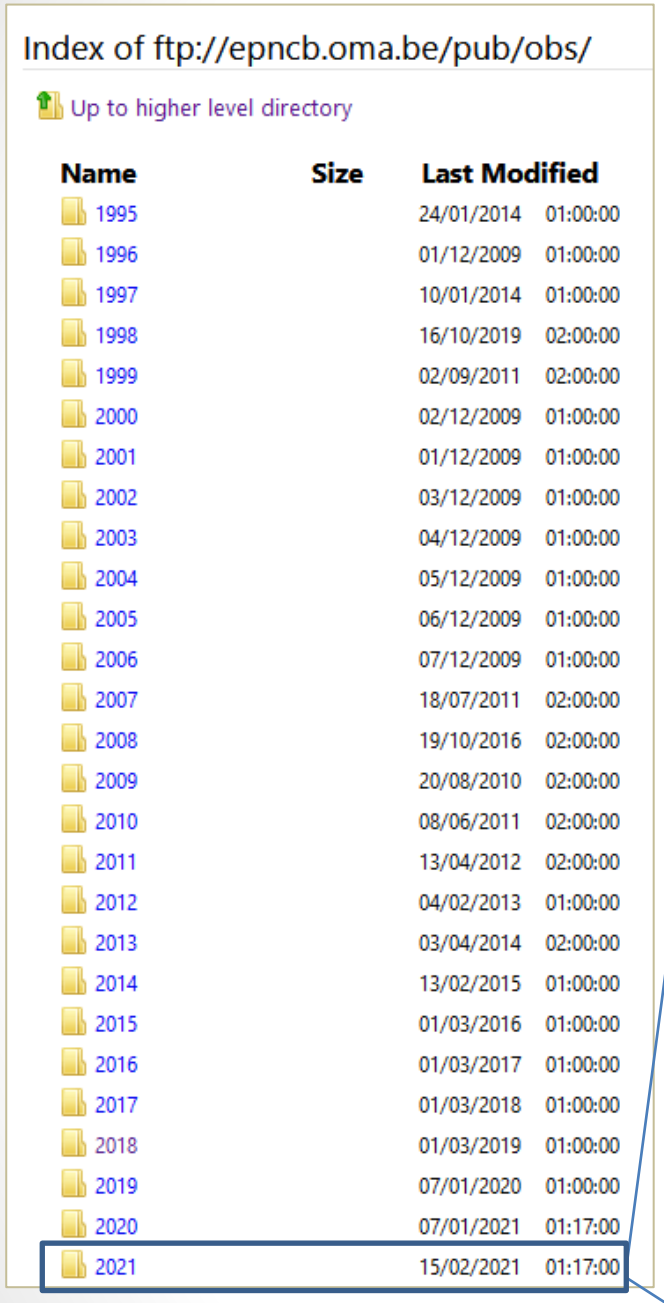
GNSS data in a nutshell

- **G**lobal **N**avigation **S**atellite **S**ystems, e.g. GPS, Galileo, GLONASS, BeiDou
- GNSS emit radio signals captured by **GNSS stations** installed at fixed locations on the Earth
- GNSS stations are equipped with high-precision GNSS instruments enabling mm-level positioning
- **GNSS data** used for multi-disciplinary applications: Maintenance of coordinate reference systems, monitoring of ground deformations, atmosphere,….

The GNSS data repository



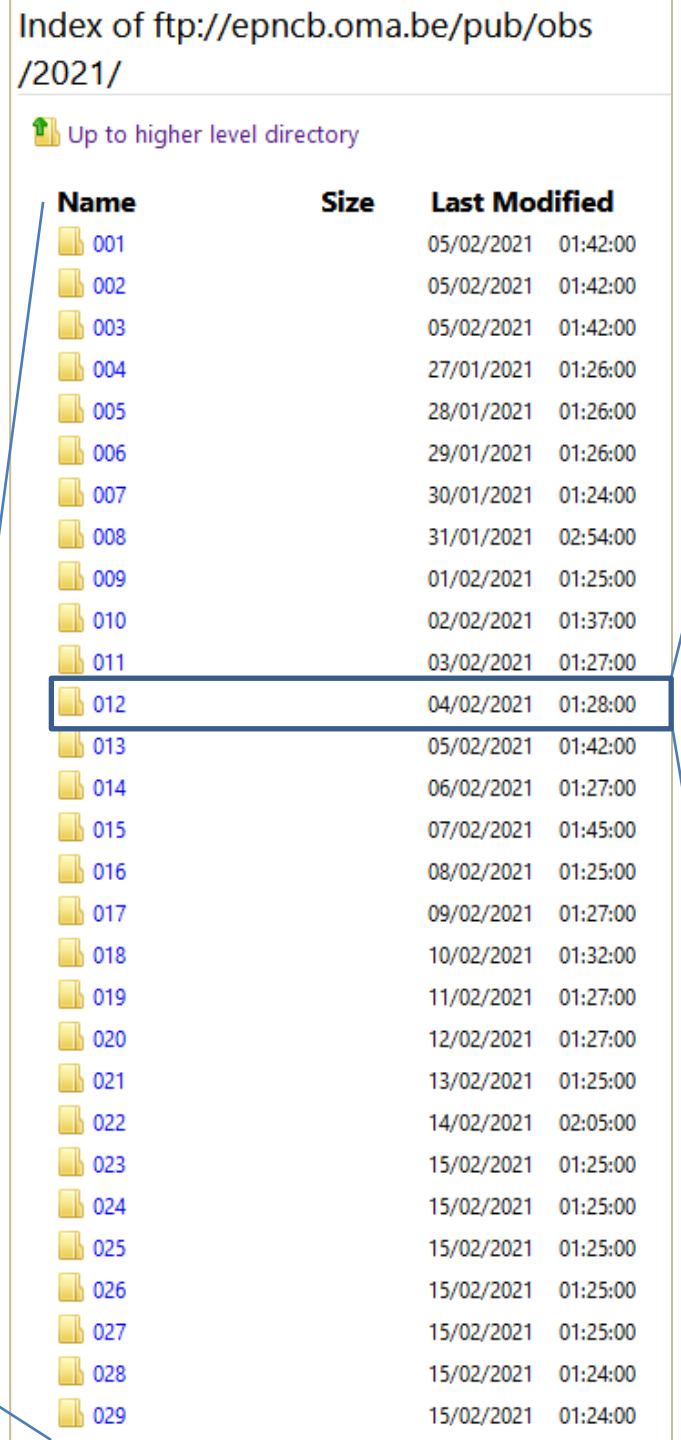
- **Historical data repository/center (HDC)** containing the DAILY observation data of 430+ GNSS stations (since 1996)
  1 file/day/station
- ROB = EUREF Permanent Network Central Bureau
- 100+ different agencies (data owners) are involved
- Voluntary contributions



Index of ftp://epncb.oma.be/pub/obs/

Up to higher level directory

Name	Size	Last Modified
1995		24/01/2014 01:00:00
1996		01/12/2009 01:00:00
1997		10/01/2014 01:00:00
1998		16/10/2019 02:00:00
1999		02/09/2011 02:00:00
2000		02/12/2009 01:00:00
2001		01/12/2009 01:00:00
2002		03/12/2009 01:00:00
2003		04/12/2009 01:00:00
2004		05/12/2009 01:00:00
2005		06/12/2009 01:00:00
2006		07/12/2009 01:00:00
2007		18/07/2011 02:00:00
2008		19/10/2016 02:00:00
2009		20/08/2010 02:00:00
2010		08/06/2011 02:00:00
2011		13/04/2012 02:00:00
2012		04/02/2013 01:00:00
2013		03/04/2014 02:00:00
2014		13/02/2015 01:00:00
2015		01/03/2016 01:00:00
2016		01/03/2017 01:00:00
2017		01/03/2018 01:00:00
2018		01/03/2019 01:00:00
2019		07/01/2020 01:00:00
2020		07/01/2021 01:17:00
2021		15/02/2021 01:17:00



Index of ftp://epncb.oma.be/pub/obs/2021/

Up to higher level directory

Name	Size	Last Modified
001		05/02/2021 01:42:00
002		05/02/2021 01:42:00
003		05/02/2021 01:42:00
004		27/01/2021 01:26:00
005		28/01/2021 01:26:00
006		29/01/2021 01:26:00
007		30/01/2021 01:24:00
008		31/01/2021 02:54:00
009		01/02/2021 01:25:00
010		02/02/2021 01:37:00
011		03/02/2021 01:27:00
012		04/02/2021 01:28:00
013		05/02/2021 01:42:00
014		06/02/2021 01:27:00
015		07/02/2021 01:45:00
016		08/02/2021 01:25:00
017		09/02/2021 01:27:00
018		10/02/2021 01:32:00
019		11/02/2021 01:27:00
020		12/02/2021 01:27:00
021		13/02/2021 01:25:00
022		14/02/2021 02:05:00
023		15/02/2021 01:25:00
024		15/02/2021 01:25:00
025		15/02/2021 01:25:00
026		15/02/2021 01:25:00
027		15/02/2021 01:25:00
028		15/02/2021 01:24:00
029		15/02/2021 01:24:00

Index of ftp://epncb.oma.be/pub/obs/2021/012/

Up to higher level directory

Name	Size	Last Modified
File: ACOR00ESP_R_20210120000_01D_30S_MO.crx.gz	1695 KB	19/01/2021 01:27:00
File: ADAR0120.21D.gz	1143 KB	19/01/2021 01:17:00
File: AJAC00FRA_R_20210120000_01D_30S_MO.crx.gz	3747 KB	19/01/2021 01:27:00
File: AJAC0120.21D.gz	2465 KB	19/01/2021 01:17:00
File: ALAC00ESP_R_20210120000_01D_30S_MO.crx.gz	1378 KB	19/01/2021 01:27:00
File: ALBA00ESP_R_20210120000_01D_30S_MO.crx.gz	1347 KB	19/01/2021 01:27:00
File: ALME00ESP_R_20210120000_01D_30S_MO.crx.gz	1722 KB	19/01/2021 01:27:00
File: ANK20120.21D.gz	753 KB	19/01/2021 01:17:00
File: ANKR00TUR_R_20210120000_01D_30S_MO.crx.gz	1803 KB	19/01/2021 01:27:00
File: ANKR0120.21D.gz	1355 KB	19/01/2021 01:17:00
File: AQUI00ITA_R_20210120000_01D_30S_MO.crx.gz	2063 KB	19/01/2021 01:27:00
File: ARGI00FRO_R_20210120000_01D_30S_MO.crx.gz	2305 KB	19/01/2021 01:27:00
File: ARGI0120.21D.gz	1006 KB	19/01/2021 01:17:00
File: ARIS0120.21D.gz	1152 KB	19/01/2021 01:17:00
File: ARJ600SWE_R_20210120000_01D_30S_MO.crx.gz	3111 KB	19/01/2021 01:27:00
File: AUBG00DEU_R_20210120000_01D_30S_MO.crx.gz	2616 KB	19/01/2021 01:27:00
File: AUBG0120.21D.gz	1262 KB	19/01/2021 01:17:00
File: AUT10120.21D.gz	343 KB	19/01/2021 01:17:00
File: AUTN00FRA_R_20210120000_01D_30S_MO.crx.gz	1808 KB	19/01/2021 01:27:00
File: AUTN0120.21D.gz	1614 KB	19/01/2021 01:17:00
File: AXPV00FRA_R_20210120000_01D_30S_MO.crx.gz	1997 KB	19/01/2021 01:27:00
File: AXPV0120.21D.gz	2039 KB	19/01/2021 01:17:00
File: BADH00DEU_R_20210120000_01D_30S_MO.crx.gz	2596 KB	19/01/2021 01:27:00
File: BAIA00ROU_R_20210120000_01D_30S_MO.crx.gz	2591 KB	19/01/2021 01:27:00
File: BAIA0120.21D.gz	637 KB	19/01/2021 01:17:00
File: BAUT00DEU_R_20210120000_01D_30S_MO.crx.gz	5600 KB	19/01/2021 01:27:00
File: BAUT0120.21D.gz	1597 KB	19/01/2021 01:17:00
File: BBYS00SVK_R_20210120000_01D_30S_MO.crx.gz	2080 KB	19/01/2021 01:28:00
File: BCLN00ESP_R_20210120000_01D_30S_MO.crx.gz	1300 KB	19/01/2021 01:28:00
File: BELL00ESP_R_20210120000_01D_30S_MO.crx.gz	2391 KB	19/01/2021 01:28:00
File: BOGE00POL_R_20210120000_01D_30S_MO.crx.gz	2582 KB	19/01/2021 01:28:00
File: BOGE0120.21D.gz	1981 KB	19/01/2021 01:17:00

Index of ftp://epncb.oma.be/pub/obs/

Up to higher level directory

Name	Size	Last Modified
1995		24/01/2014 01:00:00
1996		01/12/2009 01:00:00
1997		10/01/2014 01:00:00
1998		16/10/2019 02:00:00
2017		01/03/2018 01:00:00
2018		01/03/2019 01:00:00
2019		07/01/2020 01:00:00
2020		07/01/2021 01:17:00
2021		15/02/2021 01:17:00

Index of ftp://epncb.oma.be/pub/obs/2021/

Up to higher level directory

Name	Size	Last Modified
001		05/02/2021 01:42:00
002		05/02/2021 01:42:00
003		05/02/2021 01:42:00
004		27/01/2021 01:26:00
005		28/01/2021 01:26:00
006		29/01/2021 01:26:00
007		30/01/2021 01:24:00
008		31/01/2021 02:54:00
009		01/02/2021 01:25:00
010		02/02/2021 01:37:00
011		03/02/2021 01:27:00
012		04/02/2021 01:28:00
013		05/02/2021 01:42:00
014		06/02/2021 01:27:00
015		07/02/2021 01:45:00
016		08/02/2021 01:25:00
017		09/02/2021 01:27:00
018		10/02/2021 01:32:00
019		11/02/2021 01:27:00
020		12/02/2021 01:27:00
021		13/02/2021 01:25:00
022		14/02/2021 02:05:00
023		15/02/2021 01:25:00
024		15/02/2021 01:25:00
025		15/02/2021 01:25:00
026		15/02/2021 01:25:00
027		15/02/2021 01:25:00
028		15/02/2021 01:24:00
029		15/02/2021 01:24:00

Index of ftp://epncb.oma.be/pub/obs/2021/012/

Up to higher level directory

Name	Size	Last Modified
File: ACOR00ESP_R_20210120000_01D_30S_MO.crx.gz	1695 KB	19/01/2021 01:27:00
File: ADAR0120.21D.gz	1143 KB	19/01/2021 01:17:00
File: AJAC00FRA_R_20210120000_01D_30S_MO.crx.gz	3747 KB	19/01/2021 01:27:00
File: AJAC0120.21D.gz	2465 KB	19/01/2021 01:17:00
File: ALAC00ESP_R_20210120000_01D_30S_MO.crx.gz	1378 KB	19/01/2021 01:27:00
File: ALBA00ESP_R_20210120000_01D_30S_MO.crx.gz	1347 KB	19/01/2021 01:27:00
File: ALME00ESP_R_20210120000_01D_30S_MO.crx.gz	1722 KB	19/01/2021 01:27:00
File: ANK20120.21D.gz	753 KB	19/01/2021 01:17:00
File: ANKR00TUR_R_20210120000_01D_30S_MO.crx.gz	1803 KB	19/01/2021 01:27:00
File: ANKR0120.21D.gz	1355 KB	19/01/2021 01:17:00
File: AQUI00ITA_R_20210120000_01D_30S_MO.crx.gz	2063 KB	19/01/2021 01:27:00
File: ARGI00FRO_R_20210120000_01D_30S_MO.crx.gz	2305 KB	19/01/2021 01:27:00
File: ARGI0120.21D.gz	1006 KB	19/01/2021 01:17:00
File: ARIS0120.21D.gz	1152 KB	19/01/2021 01:17:00
File: ARJ600SWE_R_20210120000_01D_30S_MO.crx.gz	3111 KB	19/01/2021 01:27:00
File: AUBG00DEU_R_20210120000_01D_30S_MO.crx.gz	2616 KB	19/01/2021 01:27:00
File: AUBG0120.21D.gz	1262 KB	19/01/2021 01:17:00
File: AUT10120.21D.gz	343 KB	19/01/2021 01:17:00
File: AUTN00FRA_R_20210120000_01D_30S_MO.crx.gz	1808 KB	19/01/2021 01:27:00
File: AUTN0120.21D.gz	1614 KB	19/01/2021 01:17:00
File: AXPV00FRA_R_20210120000_01D_30S_MO.crx.gz	1997 KB	19/01/2021 01:27:00
File: AXPV0120.21D.gz	2039 KB	19/01/2021 01:17:00
File: BADH00DEU_R_20210120000_01D_30S_MO.crx.gz	2596 KB	19/01/2021 01:27:00
File: BAIA00ROU_R_20210120000_01D_30S_MO.crx.gz	2591 KB	19/01/2021 01:27:00
File: BAIA0120.21D.gz	637 KB	19/01/2021 01:17:00
File: BAUT00DEU_R_20210120000_01D_30S_MO.crx.gz	5600 KB	19/01/2021 01:27:00
File: BAUT0120.21D.gz	1597 KB	19/01/2021 01:17:00
File: BBYS00SVK_R_20210120000_01D_30S_MO.crx.gz	2080 KB	19/01/2021 01:28:00
File: BCLN00ESP_R_20210120000_01D_30S_MO.crx.gz	1300 KB	19/01/2021 01:28:00
File: BELL00ESP_R_20210120000_01D_30S_MO.crx.gz	2391 KB	19/01/2021 01:28:00
File: BOGE00POL_R_20210120000_01D_30S_MO.crx.gz	2582 KB	19/01/2021 01:28:00
File: BOGE0120.21D.gz	1981 KB	19/01/2021 01:17:00

Structure of the data repository:

- Similar to other GNSS data repositories within the International GNSS Service or some of its regional components
- File naming follows community-agreed international standards: 1 GNSS observation file/day/station (= simplification!!)
- RINEX format is standard format in use within the GNSS community
- Data downloadable via ftp or https

RINEX GNSS data files

Example of RINEX data file

Header

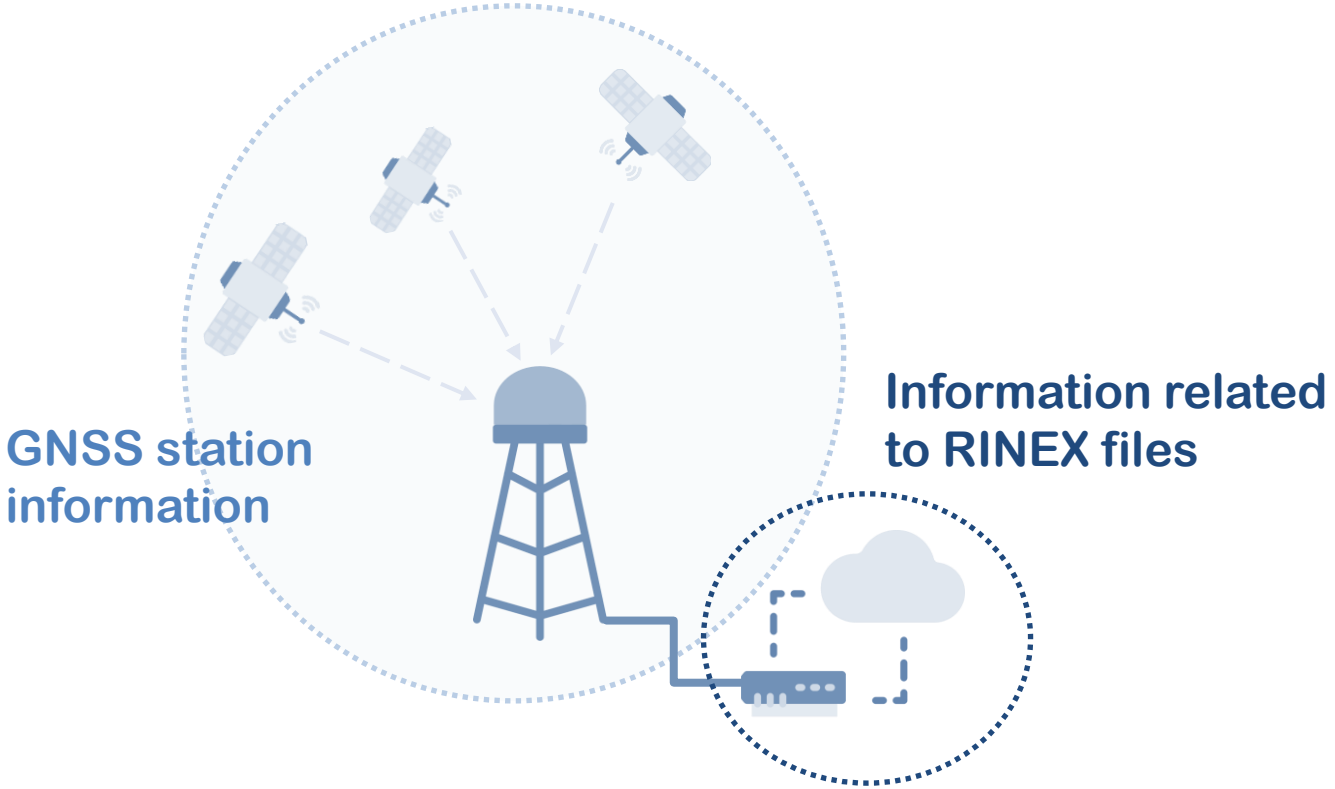
GNSS Observations

```
     3.04           OBSERVATION DATA    M (MIXED)           RINEX VERSION / TYPE
sbf2rin-13.4.5                          20210922 000349 UTC PGM / RUN BY / DATE
BRUX00BEL                                                   MARKER NAME
13101M010                                                   MARKER NUMBER
ROB                 ROB                                     OBSERVER / AGENCY
3057609             SEPT POLARX5TR      5.4.0               REC # / TYPE / VERS
00464               JAVRINGANT_DM   SCIS                    ANT # / TYPE
        0.4689        0.0000        0.0010                  ANTENNA: DELTA H/E/N
  4027881.8468   306998.2610  4919498.6524                  APPROX POSITION XYZ

....

> 2021 09 21 00 00 00.0000000  0 42
C05                40201851.771 5                 40201850.831 6  40201848.053 6
```

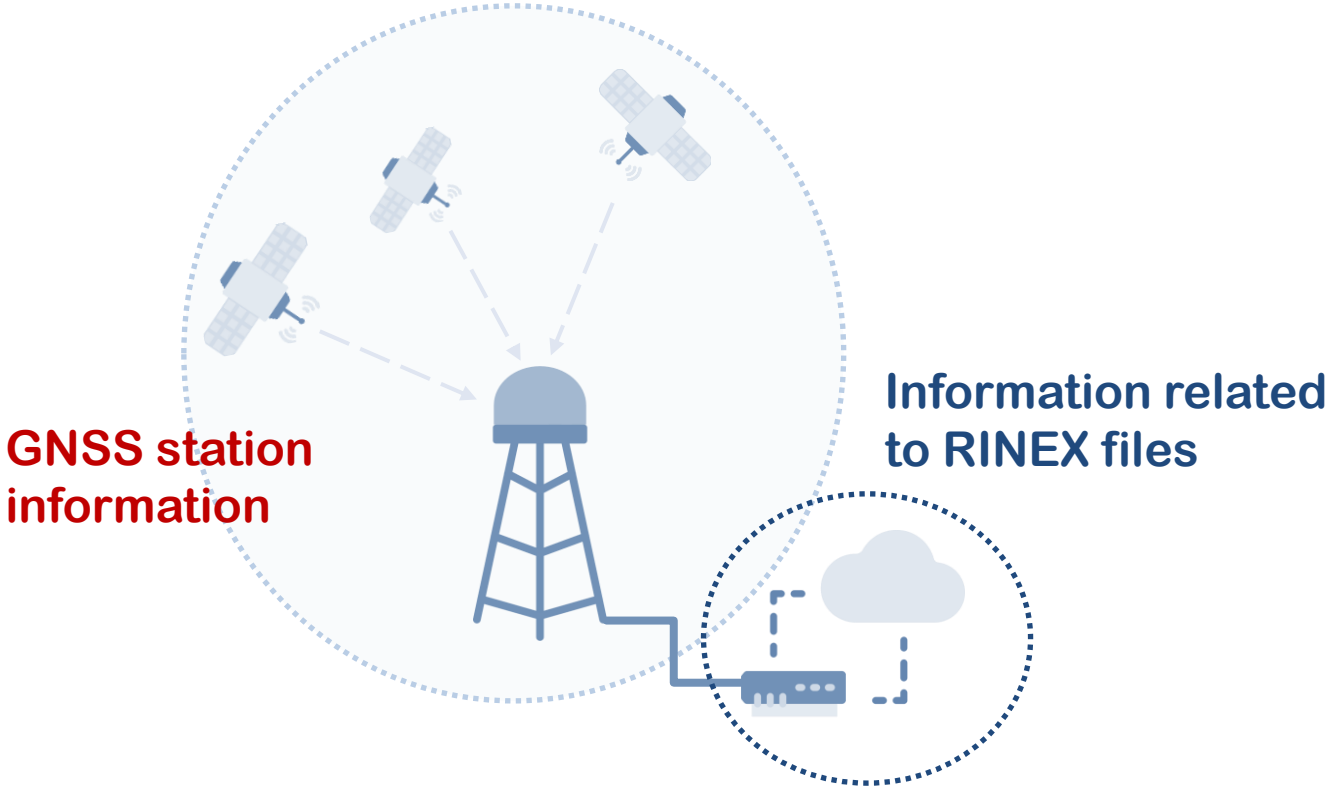
RINEX data file → RINEX header → description of the current station installation

→ GNSS observations

Additional information



Additional information



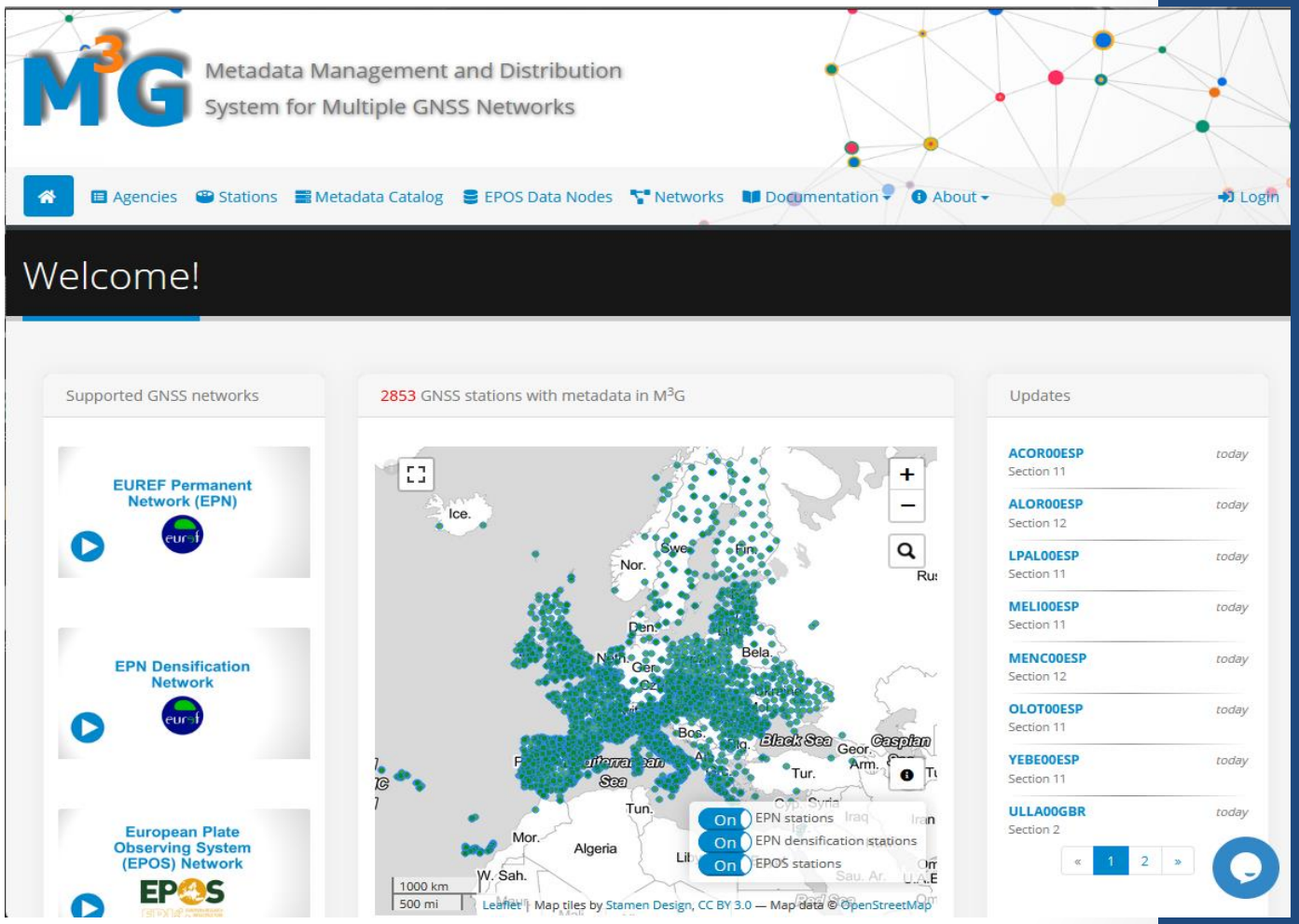
GNSS station descriptions

M³G: "Metadata management system for multiple GNSS networks"

Used by GNSS station managers to maintain description of their GNSS station:

- Location of the station
- Historical overview of equipment used in the station, changes in environment – *crucial information for reliable analysis of the GNSS data*
- Station responsible
- DOI, data license
- Station pictures

- Exported in community-agreed format: "Site log" or GeodesyML (under development)
  - *Does not contain all available information!*



https://gnss-metadata.eu

ROYAL OBSERVATORY OF BELGIUM

GHENT UNIVERSITY

GNSS station descriptions

M³G: "Metadata management system for multiple GNSS networks"

Used by GNSS station managers to maintain description of their GNSS station:

- Location of the station
- Historical overview of equipment used in the station, changes in environment – *crucial information for reliable analysis of the GNSS data*
- Station responsible
- DOI, data license
- Station pictures

- Exported in community-agreed format: "Site log" or GeodesyML (under development)
  - *Does not contain all available information!*

Example of GNSS site log

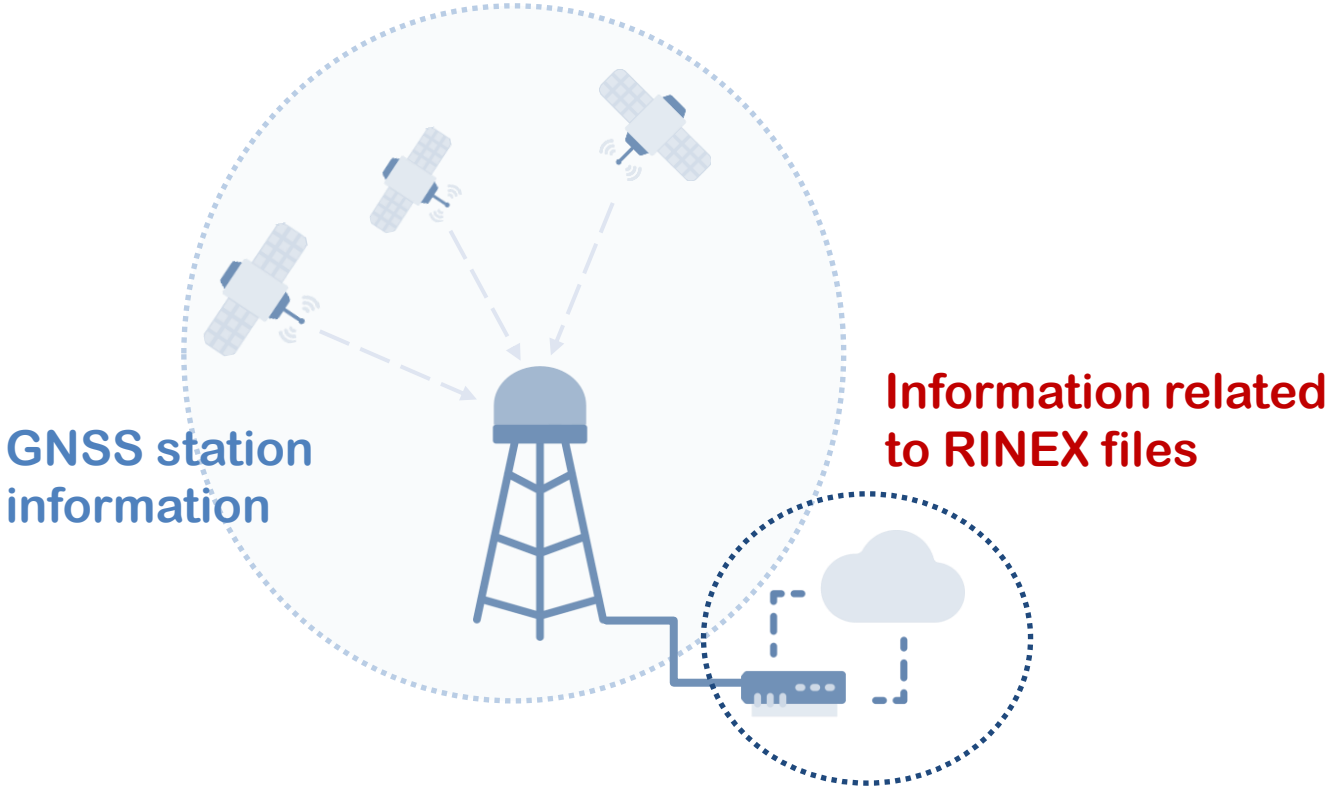
```
     BRUX00BEL Site Information Form (site log)
     International GNSS Service
     See Instructions at:
       https://files.igscb.org/pub/station/general/sitelog_instr.txt


0.   Form

     Prepared by (full name)  : GNSSatROB
     Date Prepared            : 2022-08-16
     Report Type              : UPDATE
     If Update:
      Previous Site Log       : brux00bel_20220405.log
      Modified/Added Sections : 6.12, 6.13


1.   Site Identification of the GNSS Monument

     Site Name                : Brussels
     Four Character ID        : BRUX
     Monument Inscription     :
     IERS DOMES Number        : 13101M010
     CDP Number               : (A4)
     Monument Description     : STEEL MAST
       Height of the Monument : 8 m
       Monument Foundation    : CONCRETE BLOCK
       Foundation Depth       : 3 m
     Marker Description       : CENTER OF HOLE IN STEEL PLATE
     Date Installed           : 2006-07-07
     Geologic Characteristic  : SAND
       Bedrock Type           : SEDIMENTARY
       Bedrock Condition      : FRESH
       Fracture Spacing       : 0 cm
       Fault zones nearby     : NO
```

ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

Additional information

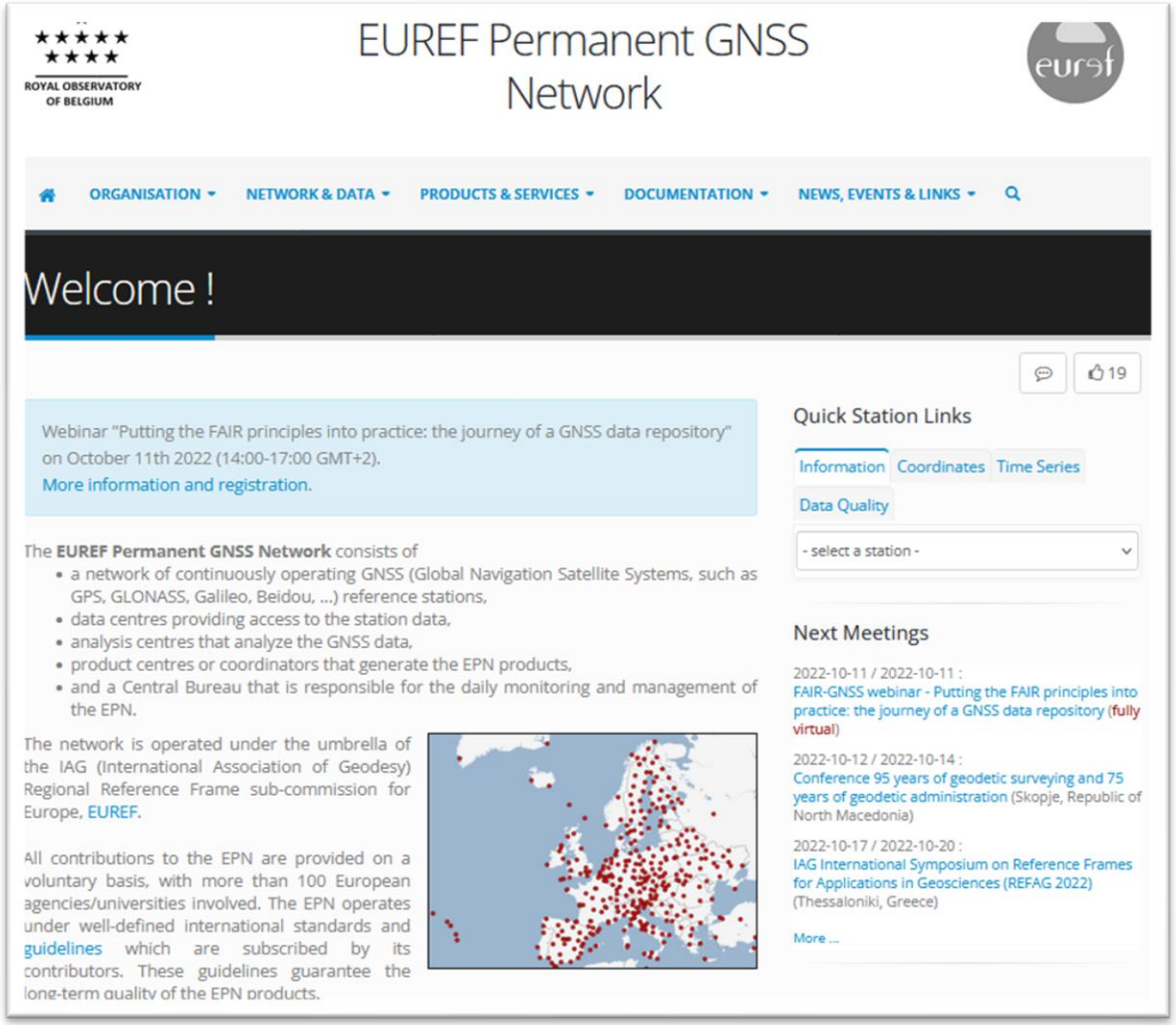


RINEX file-dependent extra information

Data repository maintained in the frame of ROB's commitment as Central Bureau of the EUREF Permanent GNSS Network (EPN) - https://epncb.oma.be

For each RINEX file:

- **Monitoring** data availability, **latency, and quality of daily GNSS data** files
- Verification of consistency of RINEX header with GNSS station description (site log)
- Publication on EPN CB web site
- Notifications to station managers



https://epncb.oma.be

RINEX file-dependent extra information

Station Configuration

RESPONSIBLE FOR METADATA	IGIK (Institute of Geodesy and Cartography)
RESPONSIBLE AGENCY	IGIK (Institute of Geodesy and Cartography)
CURRENT STATION CONFIGURATION	boge00pol_20210728.log
STATUS IN EPN	included since 05-05-2019
LAST AVAILABLE DATA	06-10-2022
LAST RECEIVER	LEICA GR30 (23-07-2021 to now)
SET TO TRACK	GPS+GLO+GAL+BDS+SBAS
LAST ANTENNA	LEIAR20 LEIM (27-09-2018 to now)
INDIVIDUAL CALIBRATION(S)	none
DATA ROUTINELY ANALYSED BY	BEV, BKG, WUT
INCLUSION IN OTHER NETWORK(S)	EPOS

Data Provided

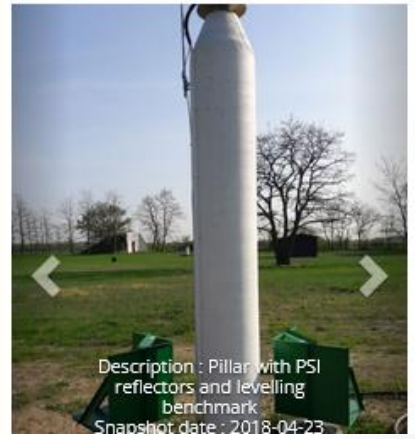
RINEX Data Quality

Position, Velocity & Time Series

Tropospheric Delays

Location

Pictures

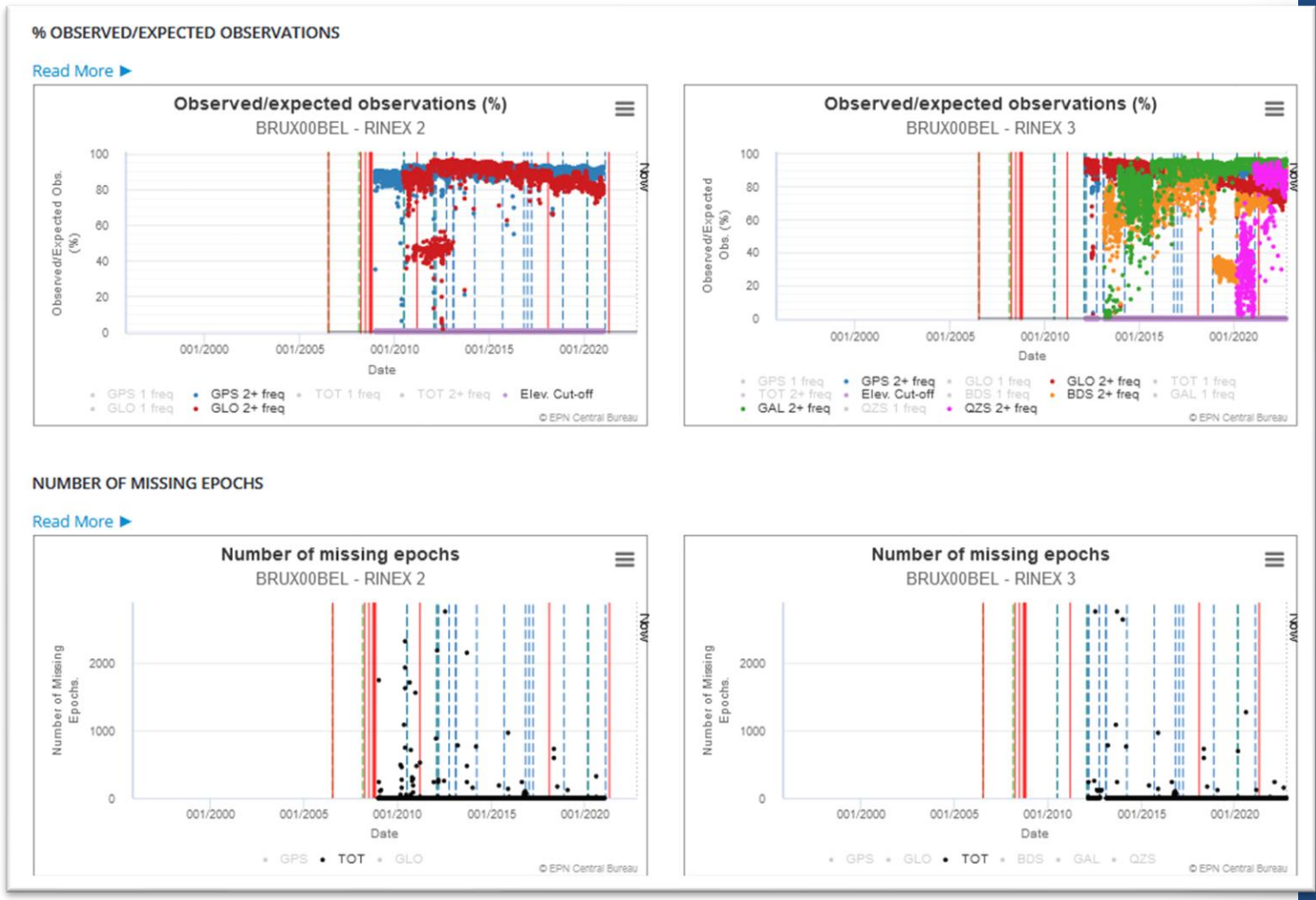


RINEX file-dependent extra information

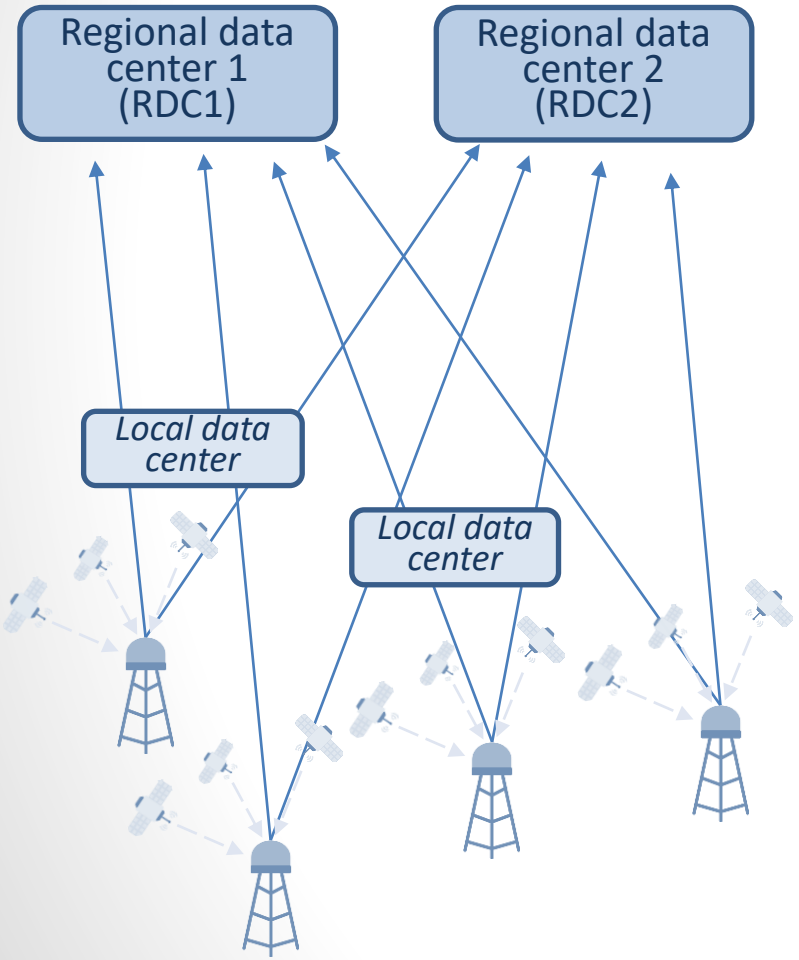
Data repository maintained in the frame of ROB's commitment as Central Bureau of the EUREF Permanent GNSS Network (EPN) - https://epncb.oma.be

For each RINEX file:

- **Monitoring** data availability, **latency, and quality of daily GNSS data** files
- Verification of consistency of RINEX header with GNSS station description (site log)
- Publication on EPN CB web site
- Notifications to station managers



Origin of the RINEX data files



Workflows of the EUREF Permanent GNSS network

EUREF data flow based on the principle of **redundancy** (no single point of failure)

GNSS stations send operationally their daily GNSS data to several data centers

All data should end up at the two EUREF regional data centers (RDC)

But, **no synchronization** between data centers

→ No guarantee that regional data centers have the same data holdings…

Historical data repository: 'as-complete-as-possible' archive of all GNSS data from EUREF stations by retrieving data from the EUREF data centers

(1file/day/station)

Origin of the RINEX data files

Data repository populated gradually since 1996 with

- **Operational data** retrieved from the EUREF data centers
  - To have complete archive: shopping around, scanning all the EUREF data centers
  - Most recent data (re)downloaded from EUREF data centers
  - **BUT, info on which data center the data were downloaded from was not stored!**
- **Data** directly provided by the station managers (15% of the data holdings)

- RINEX header is changed to be consistent with station description (site log)
  - **BUT, info on original RINEX header is not stored**

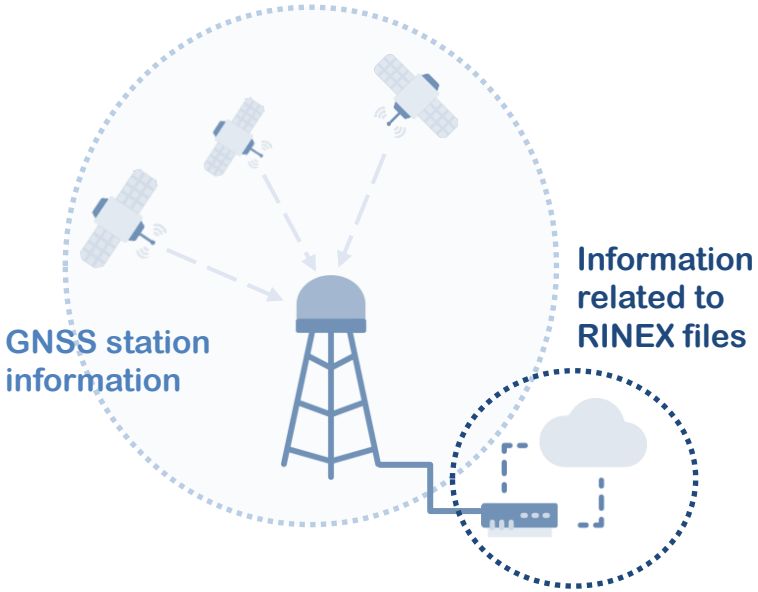
- Stored numerous flags indicating RINEX files that have a format issue
  - **BUT, flags require a cleanup to be consistent from 1996 to now + decision of which flags should block publication of data file**

Content of associated databases at the beginning of FAIR-GNSS

Description of each GNSS station (M³G)

- Site log or GeodesyML format (all stations)
- Site pictures (all stations)
- DOI, data license (almost no information available)

Provided and maintained by the station managers



RINEX-dependent information for each file in the repository (EPN CB)

- File name
- RINEX format
- File size
- Date of observations
- Creation date of file
- Last update time of file
- Data quality indicators

Computed and maintained by ROB

Download daily RINEX data from data repository

Unique station identifier
File naming conventions

TODAY

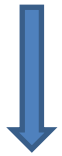
BRUX00BEL
02/2021-012/2021

RINEX data files

WTZR00DEU
001/2021-010/2021

RINEX data files

ACOR00ESP
005/2021-010/2021



RINEX data files

The additional available information is **not** exploited!

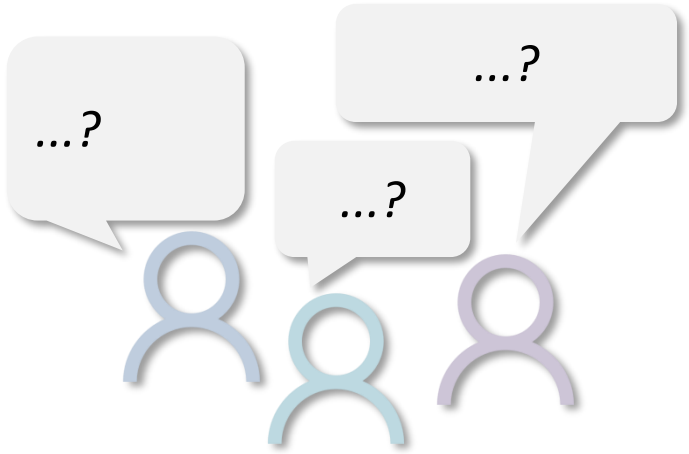
It is on a web site 'somewhere'…

User feedback on GNSS data repository

TRUST & RELIABILITY

"All stations come with RINEX header coordinates. How frequently are they updated and hence how reliable are they?"

"I am not able to download the RINEX data from your website and I am interested in the data from all the stations in a given country."



SEARCHING

"I need data from EUREF stations providing BeiDou-3 observations"

How can I cite usage of the data?

DATA CITATION

"Can I develop commercial service based on EUREF data?"

"Are data free to use?"

CONDITIONS OF USE

Download RINEX data

Downloading daily RINEX data
TODAY

BRUX00BEL
02/2021 - 12/2021

RINEX data files

Downloading daily RINEX data
FUTURE

BRUX00BEL
02/2021 - 12/2021

RINEX data files
+
Station-dependent information
+
RINEX-dependent information

Metadata should provide all info users need to know about downloaded data:
how to find the data, data license, where do the data come from, data quality metrics, data citation info

+ Search capabilities!

Download RINEX data

Downloading daily RINEX data

TODAY

BRUX00BEL

0

Downloading daily RINEX data

FUTURE

BRUX00BEL

Evolve towards application of FAIR data principles

+

Station-dependent information

+

RINEX-dependent information

Metadata should provide all info users need to know about downloaded data:

how to find the data, data license, where do the data come from, data quality metrics, data citation info

+ Search capabilities!

Goal of this webinar

To take participants through our journey (…so far)
towards applying FAIR data principles

with no ambition "to know it all"

…but hoping that our practical experience might be of some help
to those of you trying to do the same

Contents of the webinar

- **Part 1 - GNSS data and the data repository**
- **Part 2 - Introduction to FAIR data principles**
- **Part 3 - The journey so far: FAIR for GNSS data**
  - Where and how to start?
  - Gap analysis a.k.a. FAIR assessment
  - Metadata for GNSS data
  - Restructuring the data repository
  - Persistent identifiers for GNSS data
  - API preparatory work
- **Part 4 – Lessons learned and next steps**

Part 2 - Intro to FAIR data principles

Enabling data re-use

ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS

Webinar

October 11, 2022

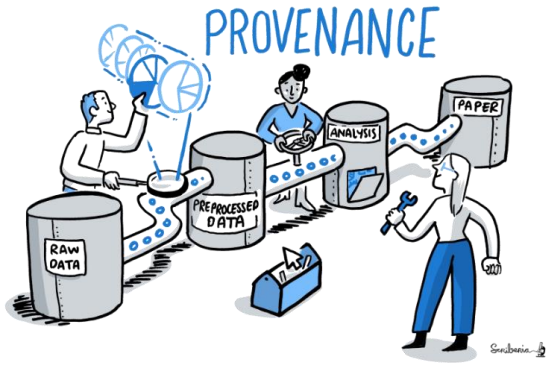
GHENT UNIVERSITY

Outline

- FAIR data principles basics
- Data stewards as a catalyst for FAIR implementation
- FAIR as an investment for the future: European initiatives, CoreTrustSeal & FAIR

Rationale for developing FAIR data principles

- Urgent **need to improve the infrastructure supporting the reuse of research data**
- Specific emphasis on enhancing the **ability of machines to automatically find and use the data**, in addition to **supporting its reuse by individuals**
- The FAIR principles are a guideline **to enhance the data reusability**

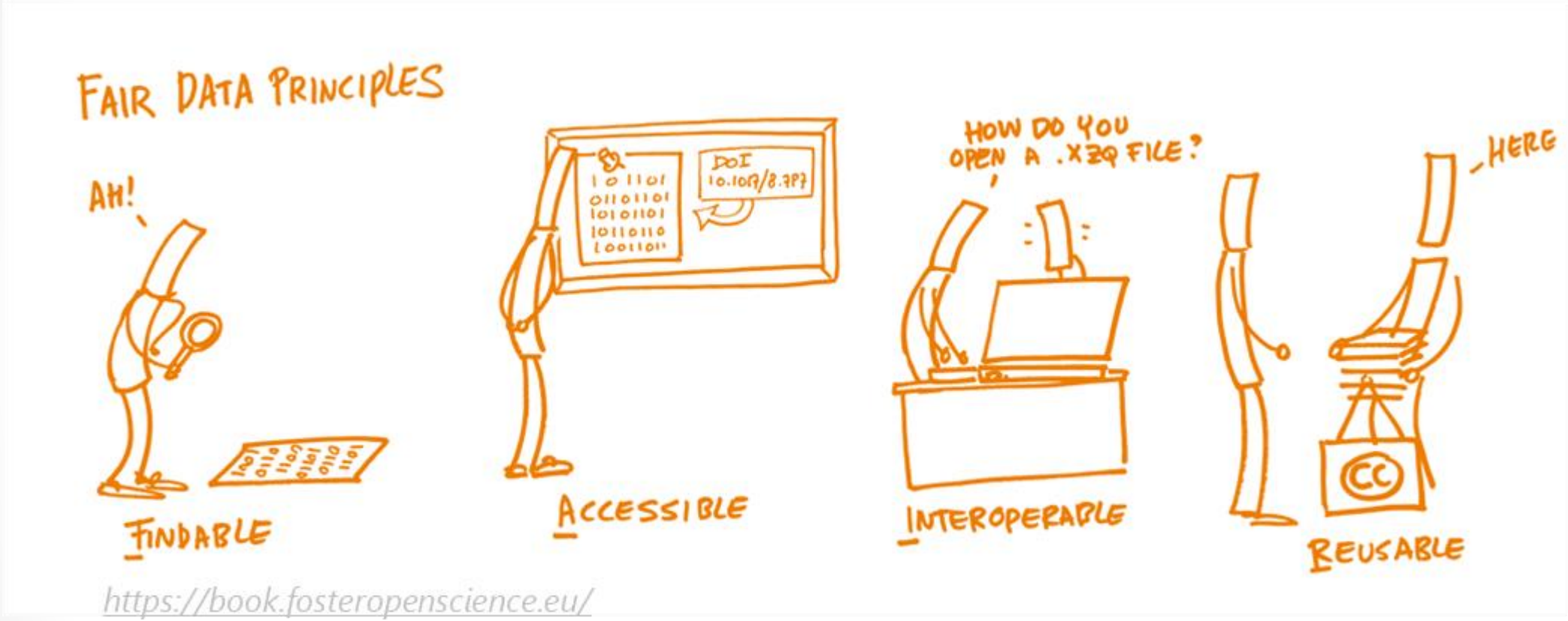


Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3,** 160018 (2016). https://doi.org/10.1038/sdata.2016.18

The ability of machines to automatically find and use the data

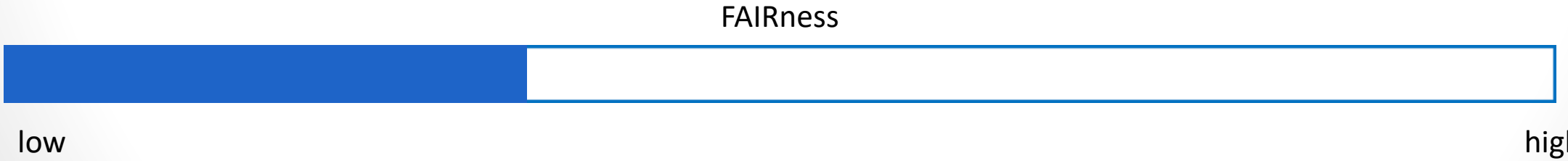
Machine-readability and actionability **to allow for the scope, scale and speed often needed in contemporary science**

- identify the type of object (with respect to both structure and intent)
- determine if it is useful within the context of the agent's current task by interrogating metadata and/or data elements
- determine if it is usable, with respect to license, consent, or other accessibility or use constraints
- take appropriate action, in much the same manner that a human would

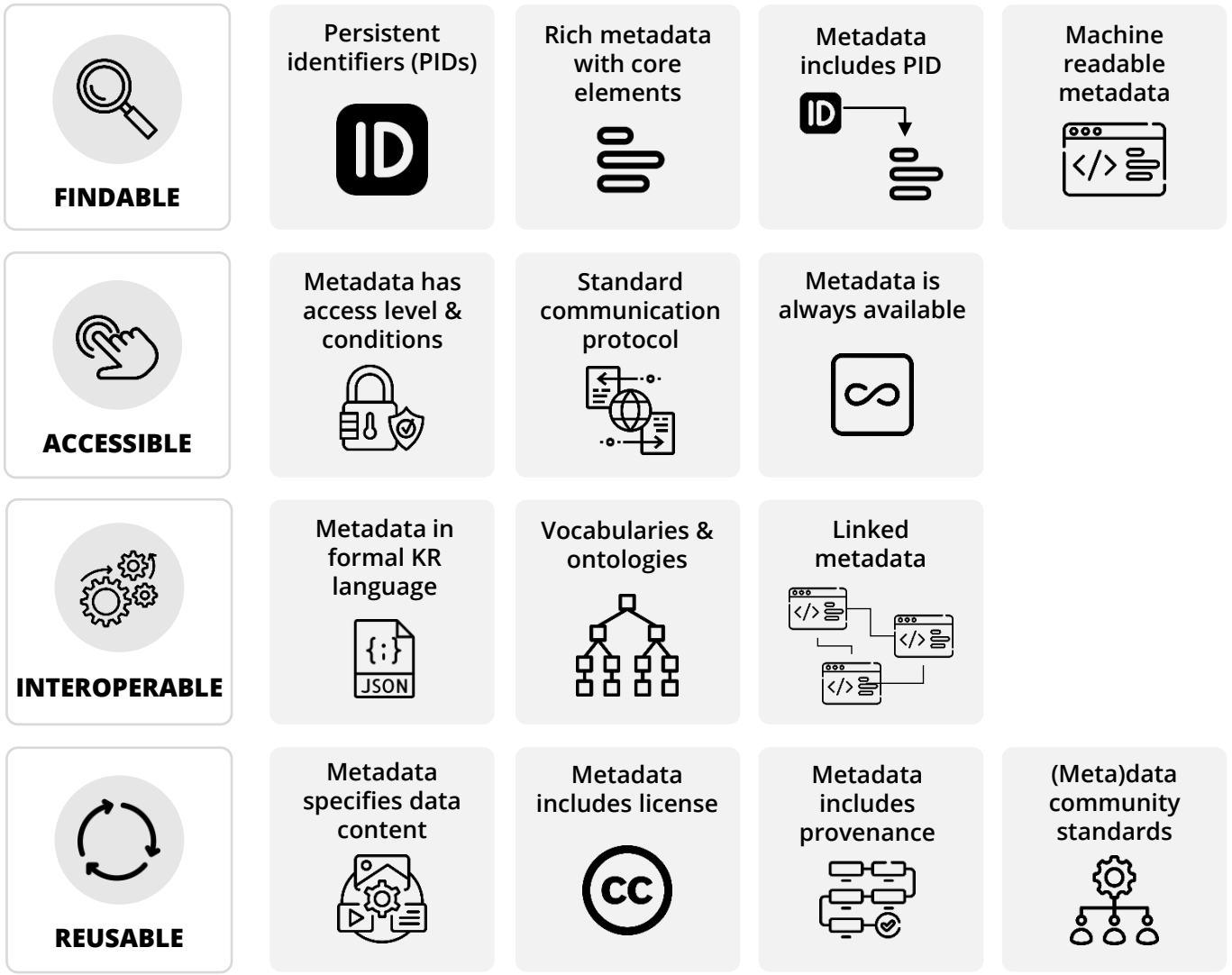

FAIR DATA PRINCIPLES
AH!
FINDABLE
DOI
10.1017/8.787
ACCESSIBLE
HOW DO YOU OPEN A .X3Q FILE?
INTEROPERABLE
HERE
CC
REUSABLE
https://book.fosteropenscience.eu/

FAIRness:
degree to which FAIR principles are applied

- FAIR principles provide “steps along a path” toward machine-actionability
- Adopting the FAIR principles, in whole or in part, leads the data along the continuum towards the optimal state
- Depends on the discipline and available tools to increase FAIRness
- Goal: FAIR by design → planning



FAIR data principles



Metadata describes the data

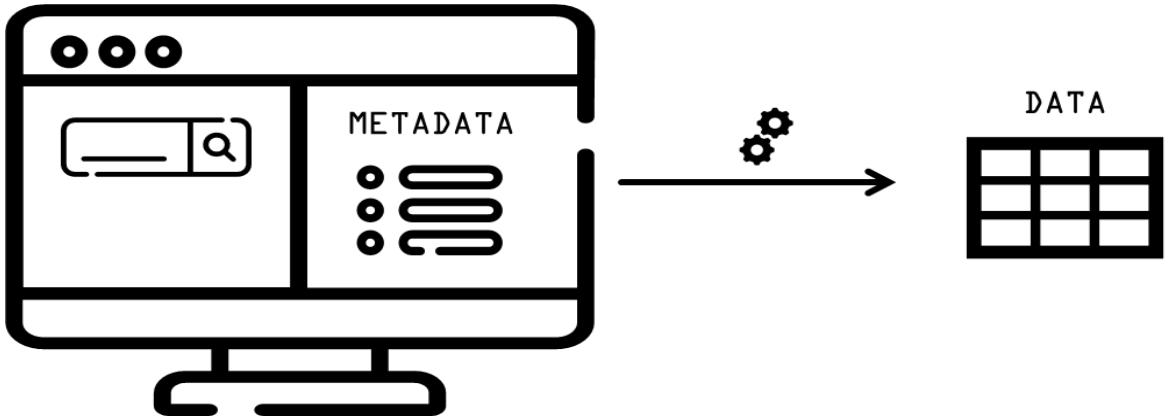
FINDABLE

Rich metadata with core elements

REUSABLE

Metadata specifies data content

Metadata: data about data, a machine-readable and structured form of documentation



Metadata describes the data

FINDABLE

Rich metadata with core elements

REUSABLE

Metadata specifies data content

Metadata fields or elements	**Metadata values** Vocabularies, ontologies, ISO standards
Title	[redacted]
Creator	ORCID [redacted]
Description	[redacted]
Subject	[redacted]
Date	ISO-... [redacted]
Format	[redacted]
Rights	[redacted]
Related identifier	[redacted]

Metadata describes the provenance of the data

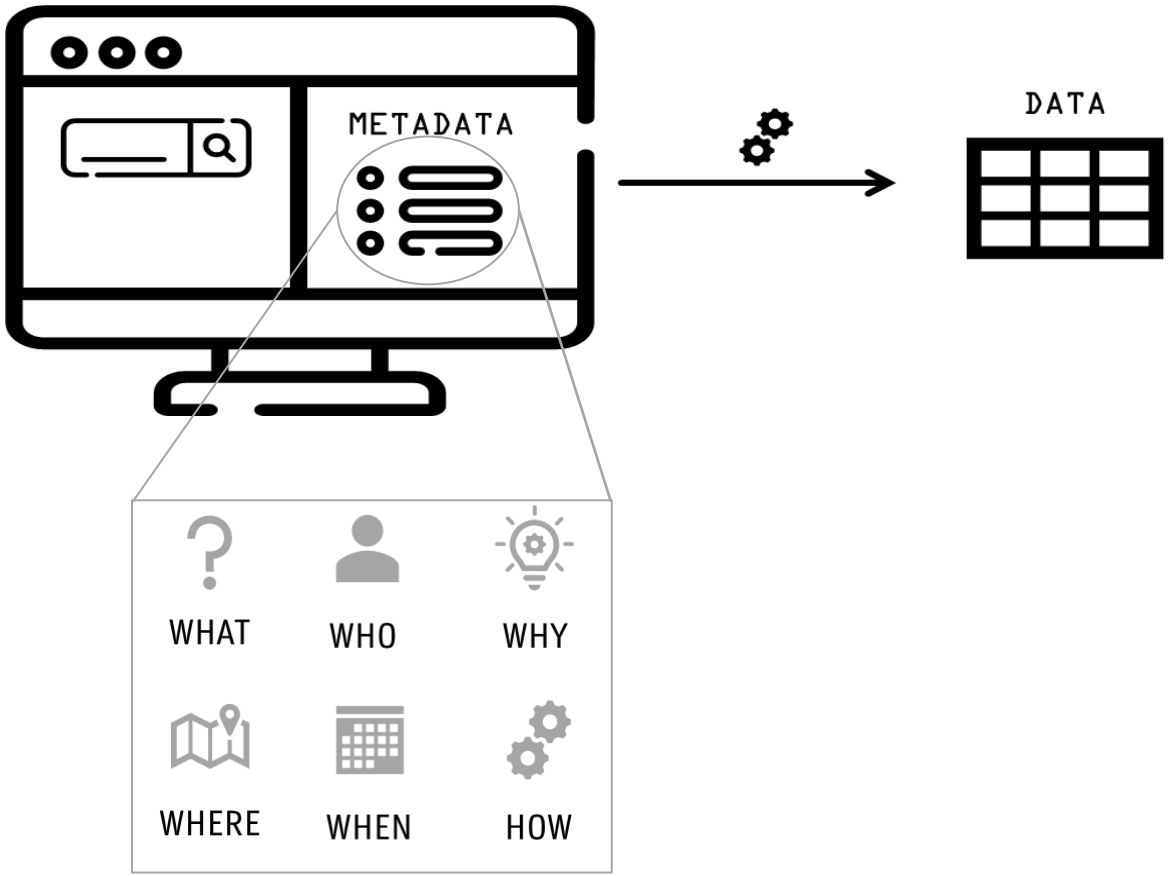
FINDABLE

Rich metadata with core elements

REUSABLE

Metadata includes provenance

Metadata: data about data, a machine-readable and structured form of documentation



Metadata standards: community standards to increase interoperability

FINDABLE

Rich metadata with core elements

REUSABLE

(Meta)data community standards

	Metadata fields or elements	**Metadata values** Vocabularies, ontologies, ISO standards
DataCite	Title	[redacted]
DataCite	Creator	ORCID
DataCite	Description	[redacted]
DataCite	Subject	[redacted]
DataCite	Date	ISO-...
DataCite	Format	[redacted]
DataCite	Rights	[redacted]
DataCite	Related identifier	[redacted]

ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

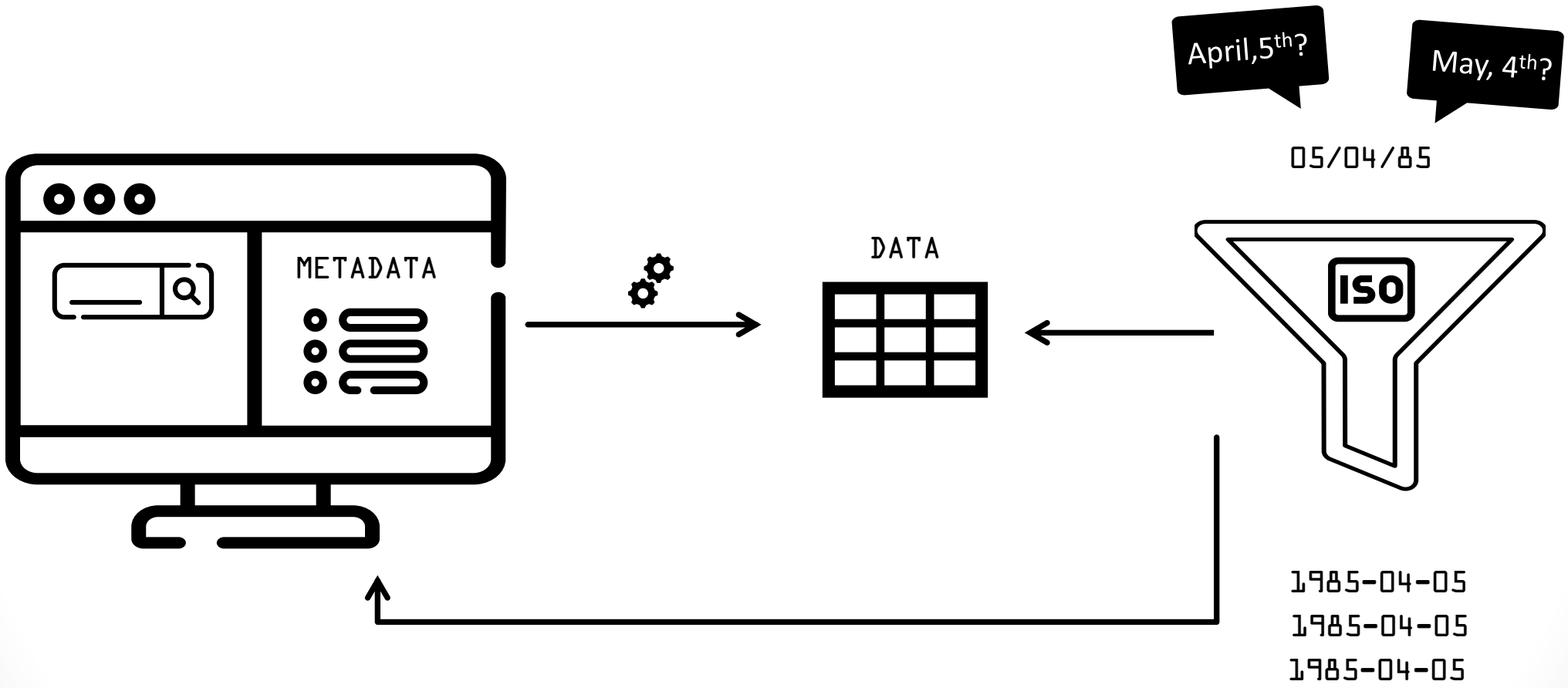
Data standards to increase interoperability and reusability

INTEROPERABLE

Vocabularies & ontologies

REUSABLE

(Meta)data community standards



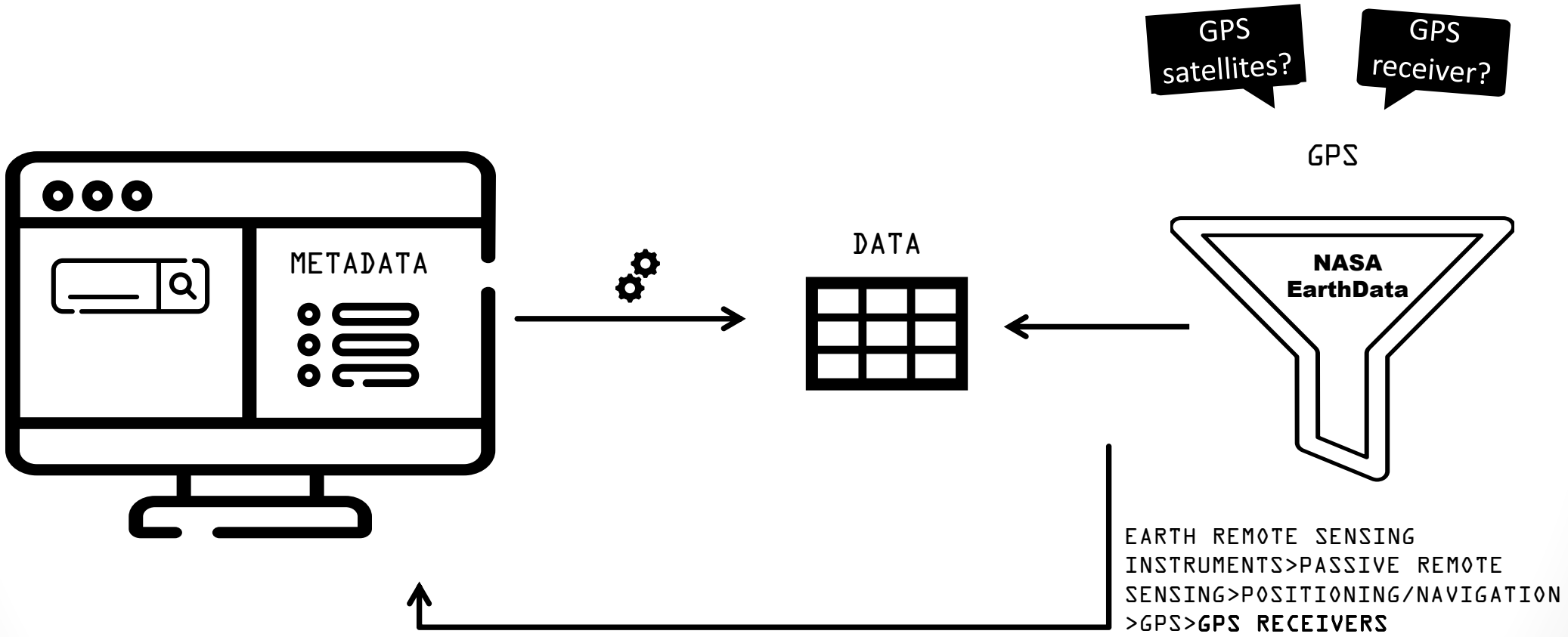
Data standards to increase interoperability and reusability

INTEROPERABLE

Vocabularies & ontologies

REUSABLE

(Meta)data community standards



Metadata standards: examples

FINDABLE

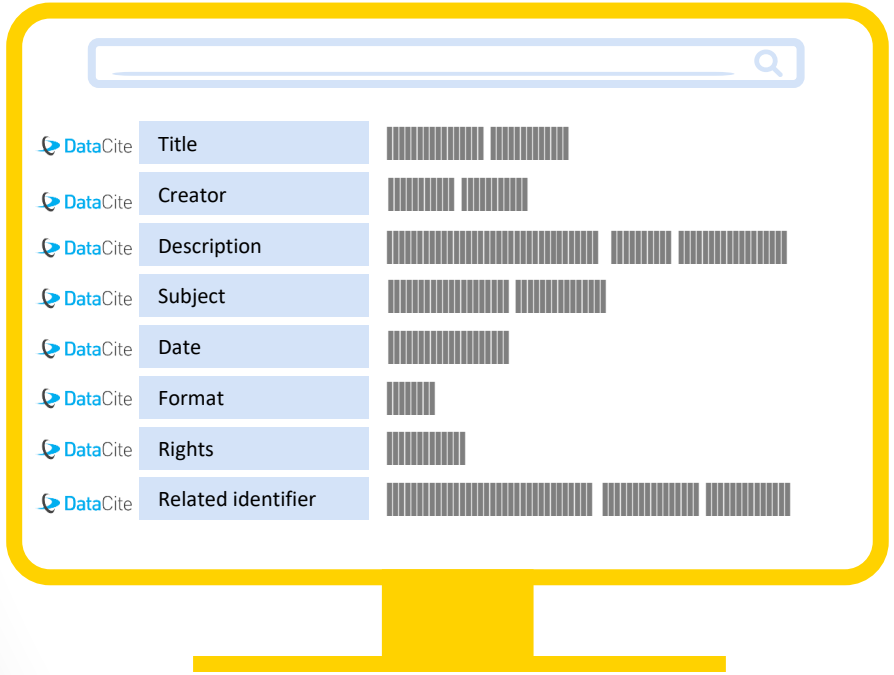
Rich metadata with core elements

REUSABLE

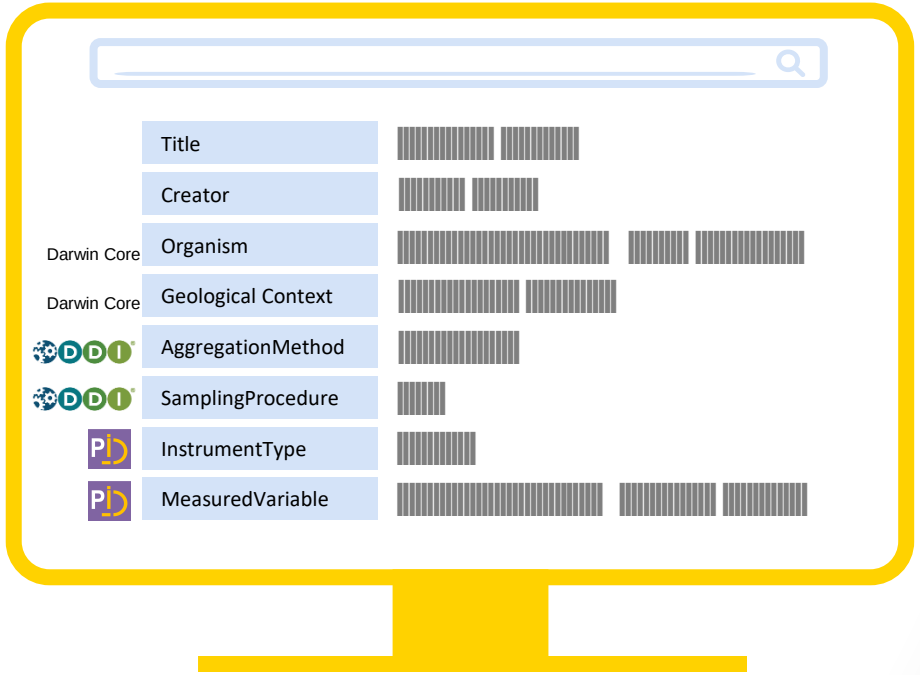
(Meta)data community standards

Domain-agnostic standards:

DataCite, Dublin Core, DCAT, etc.



Domain-specific standards:
Darwin Core, Data Documentation Initiative (DDI), Ecological Metadata Language, Persistent Identification of Instruments (PIDINST), etc.



ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

Data Catalog Vocabulary (DCAT)

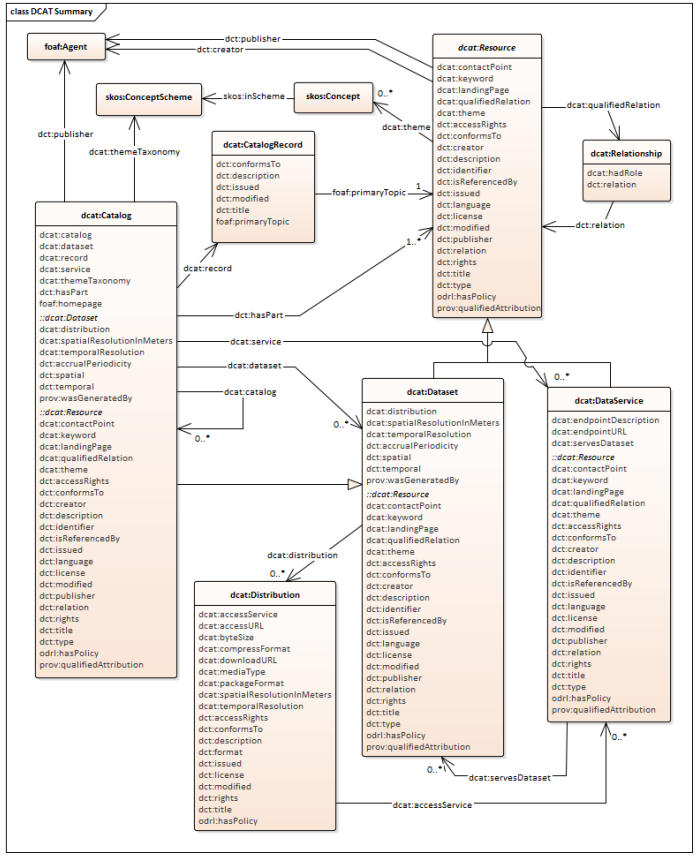
W3C vocabulary to facilitate interoperability between data catalogs

INTEROPERABLE

Vocabularies & ontologies

REUSABLE

(Meta)data community standards



https://www.w3.org/TR/vocab-dcat-2/

ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

DCAT Application Profile (DCAT-AP) to develop a suitable metadata schema

- Application profile = a set of metadata elements, policies and guidelines defined for a particular application
- Include relevant metadata elements, exclude others
- Combine metadata elements from multiple vocabularies
- Extend existing vocabularies or application profiles with domain-specific metadata elements
- Define mandatory, recommended and optional metadata elements

https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe/release/211

Exposing (meta)data

FINDABLE

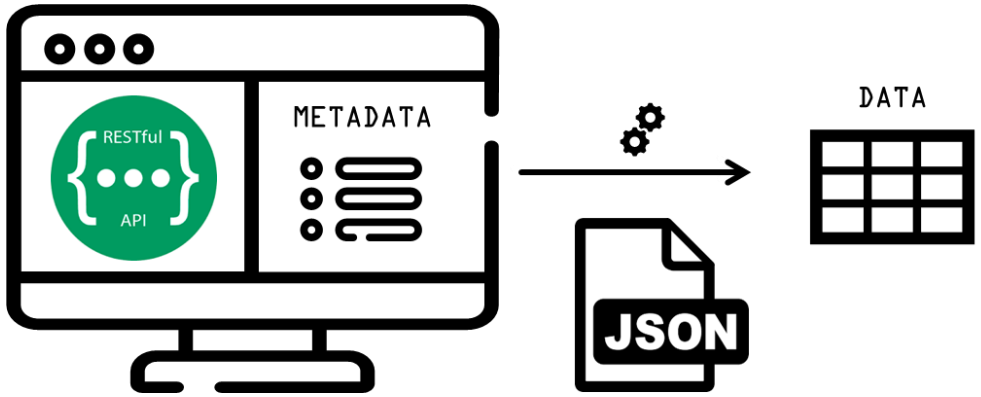
Machine readable metadata

ACCESSIBLE

Metadata has access level & conditions

Standard communication protocol

- **How/where can the (meta)data be accessed?**
- **How can data be transmitted?** Communication protocol (e.g. https, ftp)
- **How can the (meta)data be read?** Standard (meta)data formats, API
- **Who can access the data?** Authentication and authorization + access level metadata



Example of machine access: REST-API service with JSON response

Exposing metadata

Metadata is offered in such a way that it can be retrieved by machines, i.e. exposed or provided in a standard and machine-readable format.

- Metadata harvesting protocols (e.g., via OAI-PMH)
- Web services
- Embedded as structured data on a data page for use by web search engines
- Linked (open) data

OpenAIRE guidelines on access levels

4.1. DC Field

`dc:rights`

4.2. Usage

Mandatory

4.3. Usage Instruction

Use terms from the info:eu-repo-Access-Terms vocabulary . The values are:

- `info:eu-repo/semantics/closedAccess`
- `info:eu-repo/semantics/embargoedAccess`
- `info:eu-repo/semantics/restrictedAccess`
- `info:eu-repo/semantics/openAccess`

4.4. Example

```
1   <dc:rights>info:eu-repo/semantics/openAccess</dc:rights>
```

https://guidelines.openaire.eu/en/latest/literature/field_accesslevel.html
https://wiki.surfnet.nl/display/standards/info-eu-repo/#infoeurepo-AccessRights

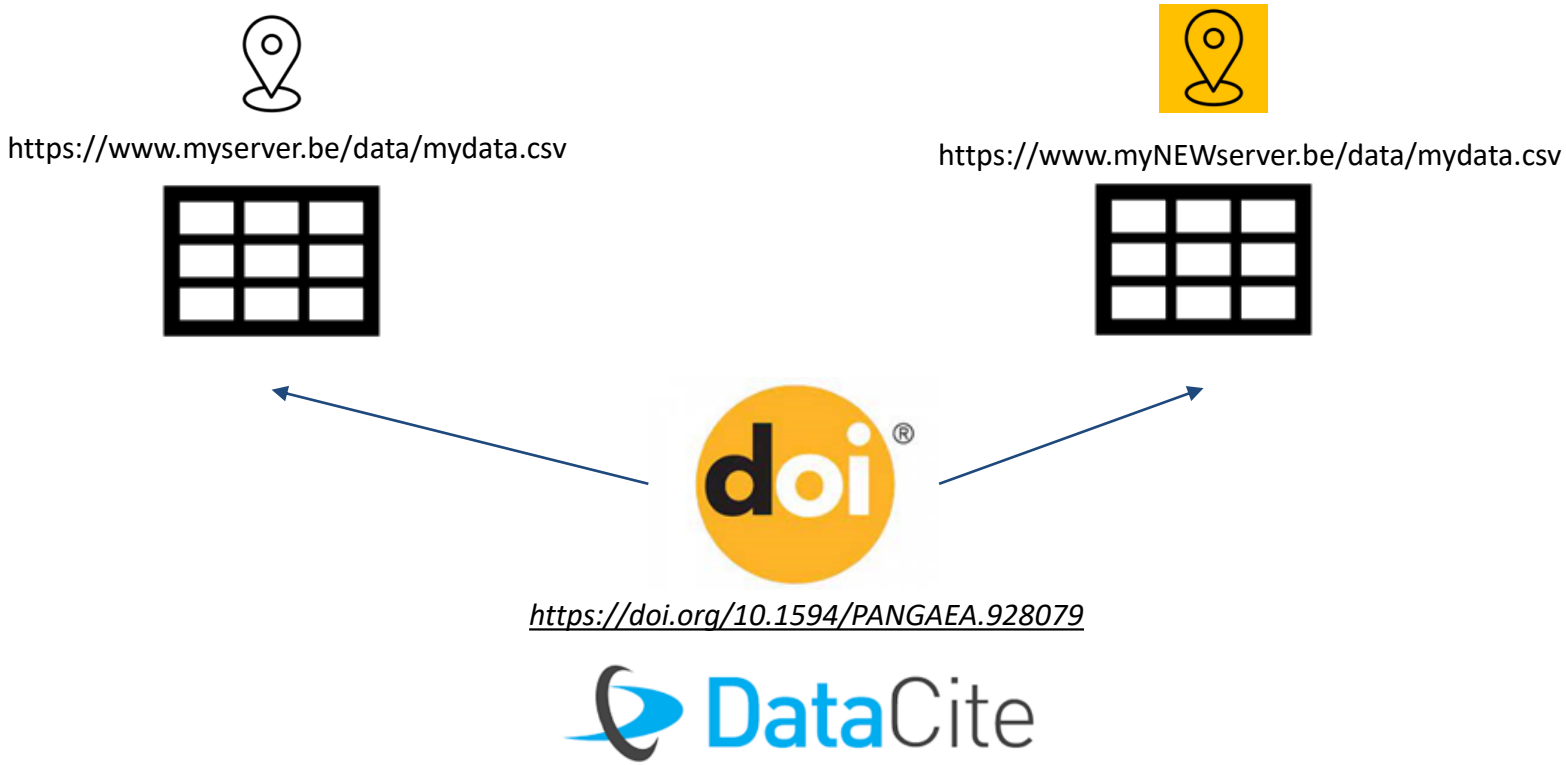
Globally unique and persistent identifiers

	Identifier	Globally unique identifier (guid, uuid)	Persistent identifier (PID)
What	a name that identifies a unique object gene identifier ENSG00000139618 book identifier ISBN 817525766-0	ensures that there are no two identical identifiers that point to different digital objects 123e4567-e89b-12d3-a456-426614174000	ensures that the digital object will remain findable over time and reduces the risk of broken links
How	can be (arbitrarily) assigned by anyone	by using a mechanism (e.g. using a registry service) for assigning identifiers	is determined by the commitment of the organisation that assigns and manages the identifier

Examples of globally unique, persistent identifiers

Identifier:	doi®	iD	R
Organisation:	DataCite	ORCID Connecting Research and Researchers	ROR
Scope:	Digital objects	Researchers	Research organisations

Digital Object Identifier (DOI)

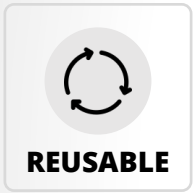


FINDABLE

Persistent identifiers (PIDs)

ID

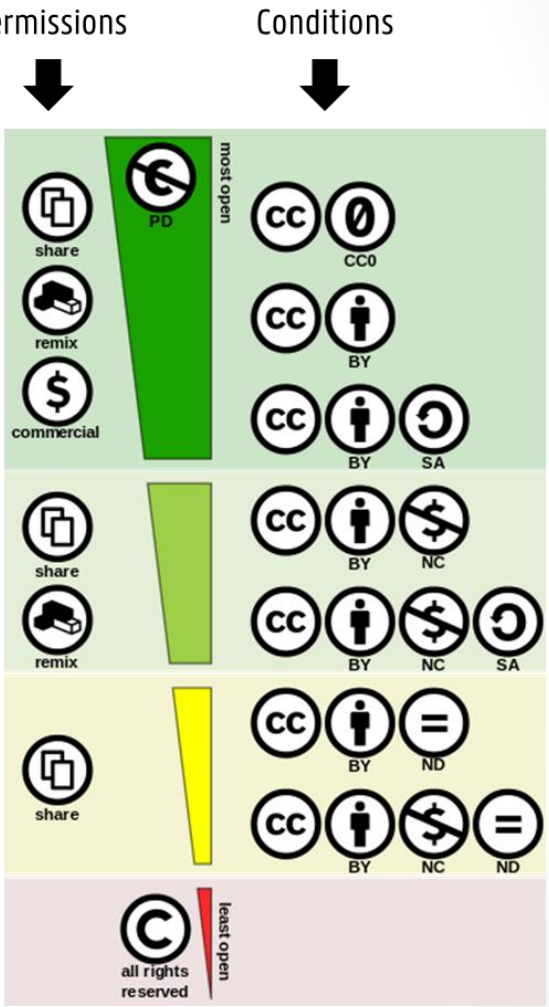
License as metadata to indicate how data can be re-used



Metadata includes license

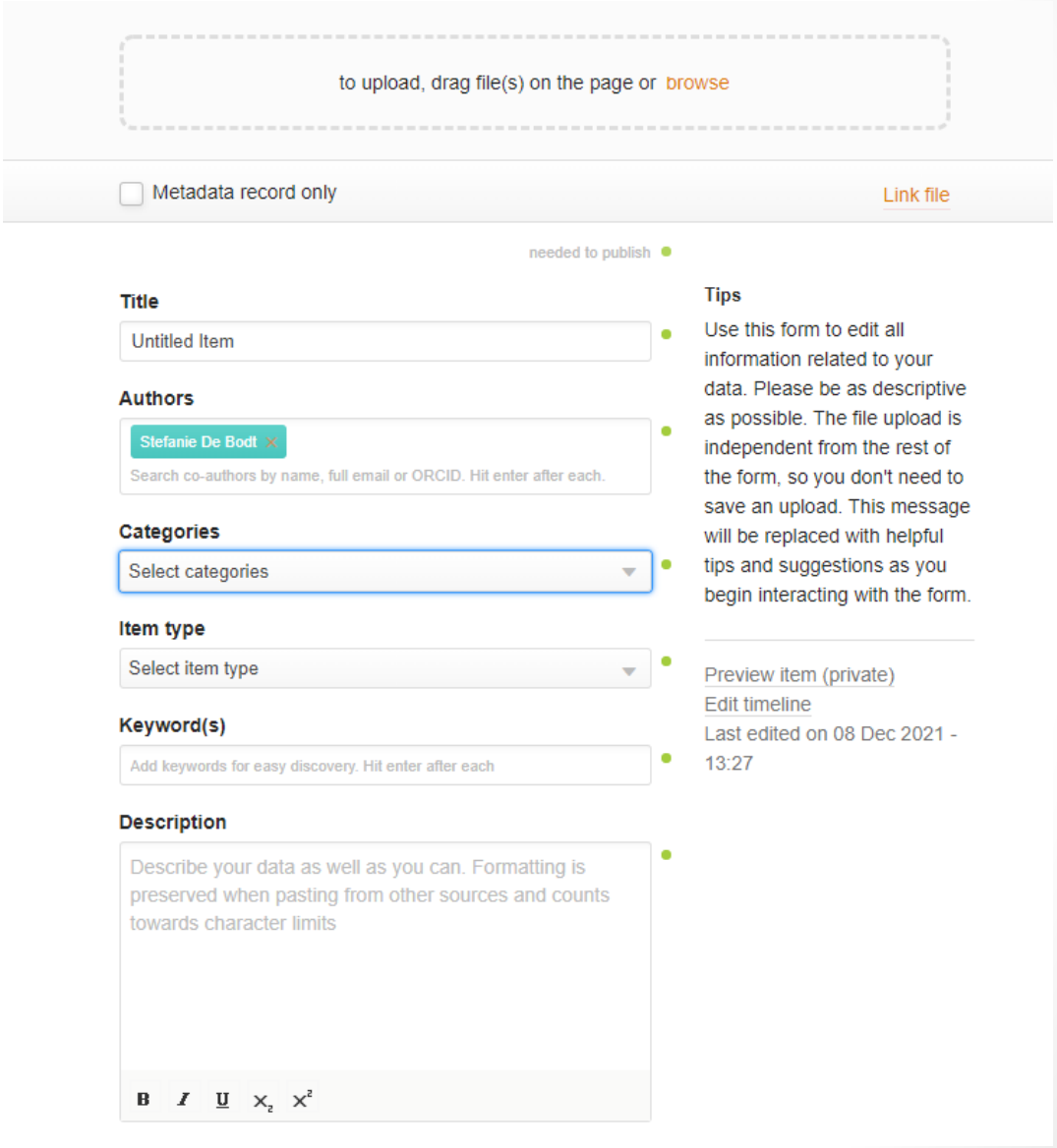
- Licenses are a means for owners of rights (in datasets) to **grant permission(s)**
- Define what can and cannot be done with data in a clear way
- Licenses are **not automatically** assigned
- **No license = no re-use possible**

- Creative Commons Licenses grant permissions under certain conditions:
  - **Attribution (BY)**: give proper credit to the original creator
  - **Copyleft (Share Alike)**: license the new (derived) creations under the identical terms (same license).
  - **Non-commerciality**: to prevent the new (derived) work to be commercially exploited.
  - **Non-derivatives**: no adaptations can be made.



Metadata in repositories

- Repositories: infrastructure to allow us to share & find data
- Provide **services to create & manage metadata**
- Typically, by filling a form with
  - Pre-defined fields
  - Lists of controlled terms to choose from (e.g. standard licenses)
- Metadata adheres to a specific metadata standard (interoperability)



Metadata schema in a landing page of a repository

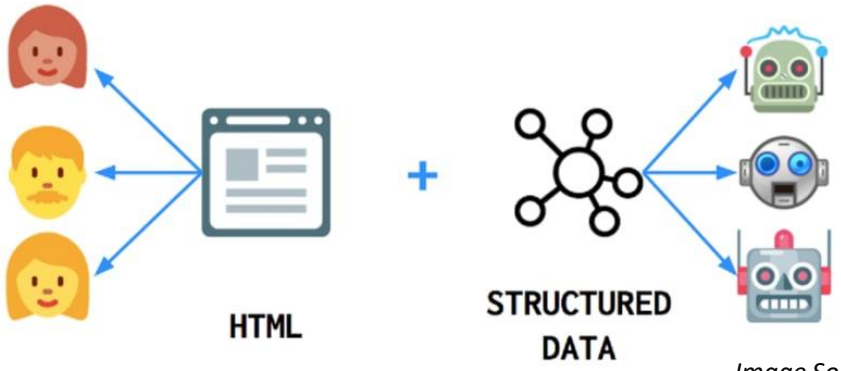


Image Source: ***Wordlift***

```
<!--BEGIN: Dublin Core description-->
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
<meta name="DC.title" content="EPMA data from tephra layer in IODP 374 Expedition Site U1524" />
<meta name="DC.creator" content="Di Roberto, Alessio" />
<meta name="DC.creator" content="Scateni, Bianca" />
<meta name="DC.creator" content="Di Vincenzo, Gianfranco" />
<meta name="DC.creator" content="Petrelli, Maurizio" />
<meta name="DC.creator" content="Fisauli, G" />
<meta name="DC.creator" content="Barker, S J" />
<meta name="DC.creator" content="Del Carlo, Paola" />
<meta name="DC.creator" content="Colleoni, Florence" />
<meta name="DC.creator" content="Kulhanek, D K" />
<meta name="DC.creator" content="McKay, Robert M" />
<meta name="DC.creator" content="de Santis, Laura" />
<meta name="DC.creator" content="The IODP Expedition 374 Scientific Party" />
<meta name="DC.publisher" content="PANGAEA" />
<meta name="DC.date" content="2021-07-05" scheme="DCTERMS.W3CDTF" />
<meta name="DC.type" content="Dataset" />
<meta name="DC.language" content="en" scheme="DCTERMS.RFC3066" />
<meta name="DCTERMS.license" scheme="DCTERMS.URI" content="https://creativecommons.org/licenses/by/4.0/" />
<meta name="DC.identifier" content="https://doi.org/10.1594/PANGAEA.933364" scheme="DCTERMS.URI" />
<meta name="DC.rights" scheme="DCTERMS.URI" content="info:eu-repo/semantics/openAccess" />
<meta name="DC.format" content="application/zip, 2 datasets" />
<meta name="DC.relation" content="Di Roberto, Alessio; Scateni, Bianca; Di Vincenzo, Gianfranco; Petrelli, Maurizio;
<!--END: Dublin Core description-->
```

Data steward support in FAIRification process

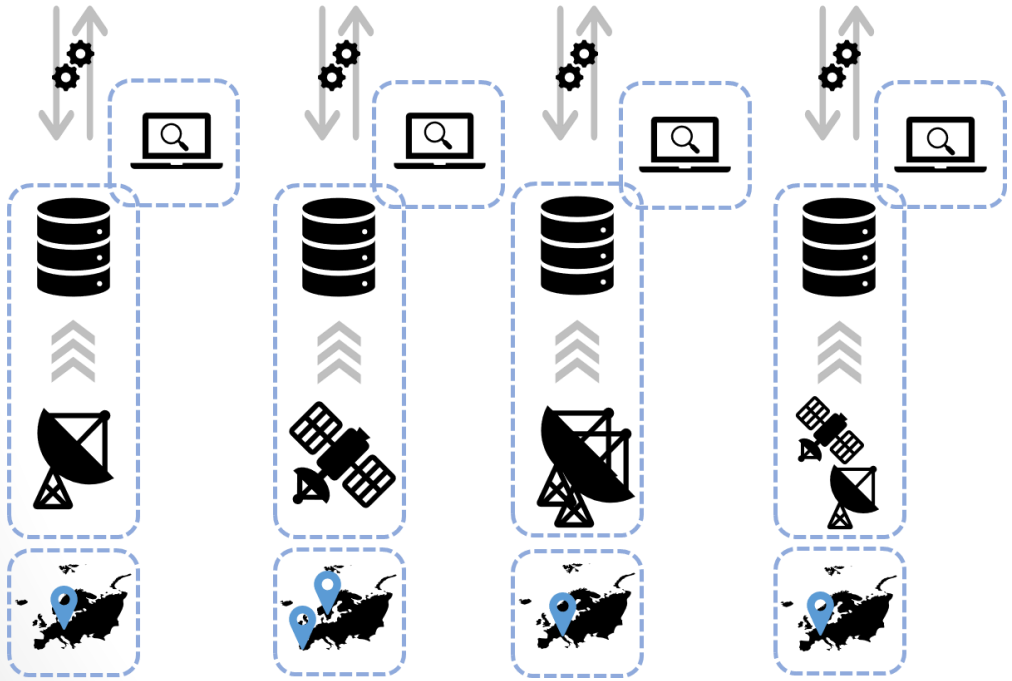
Explore	Design	Implement
Understand the FAIR terminology	Develop expertise and skills or involve skilled people	Engage with community to evolve to a “real” standard
Understand current gaps in data management	Set FAIRness goals to reach impact on reusability	Be pragmatic, implement what you can now, extend later
Which data is in scope?	Explore existing repositories, standards and terminology in the field	Technical implementation is not easy! Involve experts or use technology that is known
Which users and use cases are in scope?		Keep an eye on the evolution of certification and/or interoperability requirements

Trusted, FAIR-enabling and certified data repositories

- **TRUST principles**: a set of guiding principles to demonstrate digital repository trustworthiness
  - **T**ransparency, **R**esponsibility, **U**ser focus, **S**ustainability and **T**echnology
- **FAIR metrics**: FAIR assessment of Research Data Objects
  - e.g. FAIRsFAIR metrics
- **Certification**:
  - e.g Core Trust Seal certification
- **Trusted or trustworthy repositories**

Data portal for access to distributed data

- One-stop shop to query multiple, diverse data sources increasing visibility and re-use of data
- Uniform and stable data access
- Increased interoperability between datasets



- Distributed efforts for data import and curation
- Local, regional or data type specificities possible, avoiding one-fits-all compromises
- Data standardization and interoperability enabled by domain experts
- Distributed investment and operating cost

Contents of the webinar

- **PART 1 - GNSS data and the data repository**
- **PART 2 - Introduction to FAIR data principles**
- **PART 3 - The journey so far: FAIR for GNSS data**
  - Where and how to start?
  - Gap analysis a.k.a. FAIR assessment
  - Metadata for GNSS data
  - Restructuring the data repository
  - Persistent identifiers for GNSS data
  - API preparatory work
- **PART 4 – Lessons learned and next steps**

ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

PART 3 - The journey so far: FAIR for GNSS data

The FAIR-GNSS project's approach to FAIR data implementation

GHENT UNIVERSITY

Contents of the webinar

- **Part 1 - GNSS data and the data repository**
- **Part 2 - Introduction to FAIR data principles**
- **Part 3 - The journey so far: FAIR for GNSS data**
  - Where and how to start?
  - Gap analysis a.k.a. FAIR assessment
  - Metadata for GNSS data
  - Restructuring the data repository
  - Persistent identifiers for GNSS data
  - API preparatory work
- **Part 4 – Lessons learned and next steps**

A *possible* journey towards FAIR: where and how to start?

With the help of the data stewards
at University of Ghent
we started to understand:

- what FAIR principles are about
- related key concepts and terminology

A *possible* journey towards FAIR: where and how to start?

Start by looking at plenty of FAIR implementation initiatives…

…but applications depends upon the types of data, so how do we tackle the problem for our type of data?

Where we started

We identified some of the **most relevant initiatives,** including those **within the earth science and geodetic community…**

- *EU "Turning FAIR into reality" (2018) & "TRUST principles for digital repositories" (2020)*
- *"Perspectives on the Implementation of FAIR Principles in Solid Earth Research Infrastructure" Bailo et al. (2020)*
- *"From Conceptualization to Implementation: FAIR Assessment of Research Data Objects" Devaraju et al. (2021)*
- *"Ensuring Access to Precise Positioning by Improving Geodetic Standards" FrontierSI report (2019)*

How we started

We identified some of the **most relevant initiatives,** including those **within the earth science and geodetic community…**

- *EU "Turning FAIR into reality" (2018) & "TRUST principles for digital repositories" (2020)*
- *"Perspectives on the Implementation of FAIR Principles in Solid Earth Research Infrastructure" Bailo et al. (2020)*
- *"From Conceptualization to Implementation: FAIR Assessment of Research Data Objects" Devaraju et al. (2021)*
- *"Ensuring Access to Precise Positioning by Improving Geodetic Standards" FrontierSI report (2019)*

…we **extracted some useful and inspirational approaches:**

Data as **FAIR data objects** and data repository as part of the **FAIR ecosystem** and moving towards becoming a **trustworthy data repository**

Work plan focus on:

- **data & metadata** →
- data access →
- data visualization & products

FAIR data principles applied to, and assessed at, each stage of the **research data lifecycle**

Download GNSS data

Downloading daily RINEX data

TODAY

BRUX00BEL
02/2021 - 12/2021

RINEX data files

Downloading daily RINEX data

FUTURE

BRUX00BEL
02/2021 - 12/2021

RINEX data files
+
Station-dependent information
+
RINEX-dependent information

Metadata should provide all info users need to know about downloaded data:
data license, where do the data come from, data quality metrics, data citation info

+ Search capabilities!

ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

From FAIR guidelines…
to a practical approach for GNSS data

Downloading daily RINEX data
TODAY

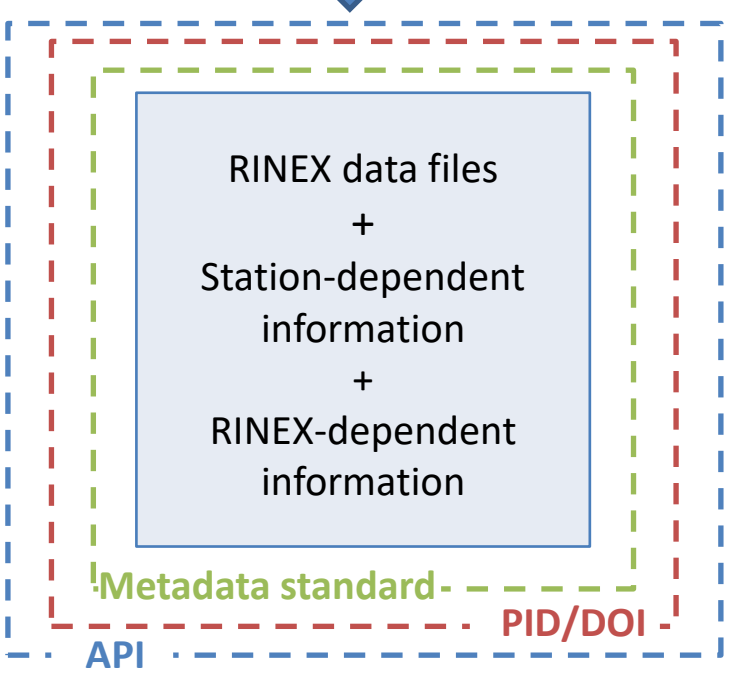
BRUX00BEL
02/2021 - 12/2021

RINEX data files

Focus first on GNSS data & metadata

Downloading daily RINEX data
FUTURE

BRUX00BEL
02/2021 - 12/2021

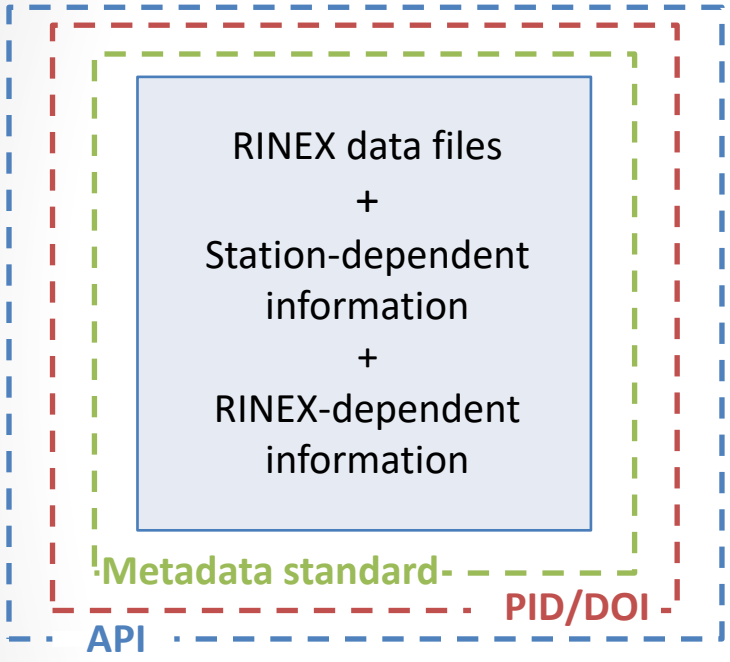


GNSS data as **FAIR Data Object:**

- Rich metadata including link to site log, license, provenance info, data quality metrics, …
- Citable GNSS data via PID i.e. DOI
- GNSS data are searchable and accessible and reusable by both humans and machines

FAIR-GNSS approach and FAIR

GNSS data
as **FAIR Data Object**



FINDABLE	Persistent identifiers (PIDs)	Rich metadata with core elements	Metadata includes PID	Machine readable metadata
ACCESSIBLE	Metadata has access level & conditions	Standard communication protocol		
INTEROPERABLE	Vocabularies & ontologies			
REUSABLE	Metadata specifies data content	Metadata includes license	Metadata includes provenance	(Meta)data community standards

Each stage of our approach brings us closer to being FAIR compliant

FAIR-GNSS approach: work plan

Assess where we stand with respect to FAIR principles

Gap-analysis

Identified what info is missing

Retrieve missing info & restructure databases accordingly

HDC repository restructuring ⇔ **Metadata for GNSS data**

Identify relevant metadata & suitable metadata schema(s)

Assign DOI to data, identify relevant info to include in the DOI metadata

DOI for GNSS data

API for HDC

Machine-operable access to GNSS data &metadata for searching/downloading

Data portal

Build on top of GNSS FDO objects to provide data access (to both humans and machines) to the HDC repository, with increased interoperability, visibility and potential reuse

FAIR-GNSS approach to FAIR data implementation Summary

We identified a methodology:

- Turn GNSS data into FAIR Data Object:
  - attach rich metadata (provenance info, license, link to site log,…)
  - make them citable by attaching PID i.e. DOI
  - make them searchable/accessible/reusable to both humans machines
- Focus first on data & metadata, then on data access
- GNSS data repository as part of the FAIR ecosystem: trustworthy & FAIR-enabling repository
- … last, but not least: assess level of FAIRness at multiple stages

Coming up next:
How do we stand w.r.t. FAIR data principles?
FAIR assessment of our data a.k.a. a gap analysis

Contents of the webinar

- **PART 1 - Introduction to FAIR data principles**
- **PART 2 - The journey so far: FAIR for GNSS data**
  - Where and how to start?
  - Gap analysis a.k.a. FAIR assessment
  - Metadata for GNSS data
  - Restructuring the data repository
  - Persistent identifiers for GNSS data
  - API preparatory work
- **PART 3 – Lessons learned and next steps**

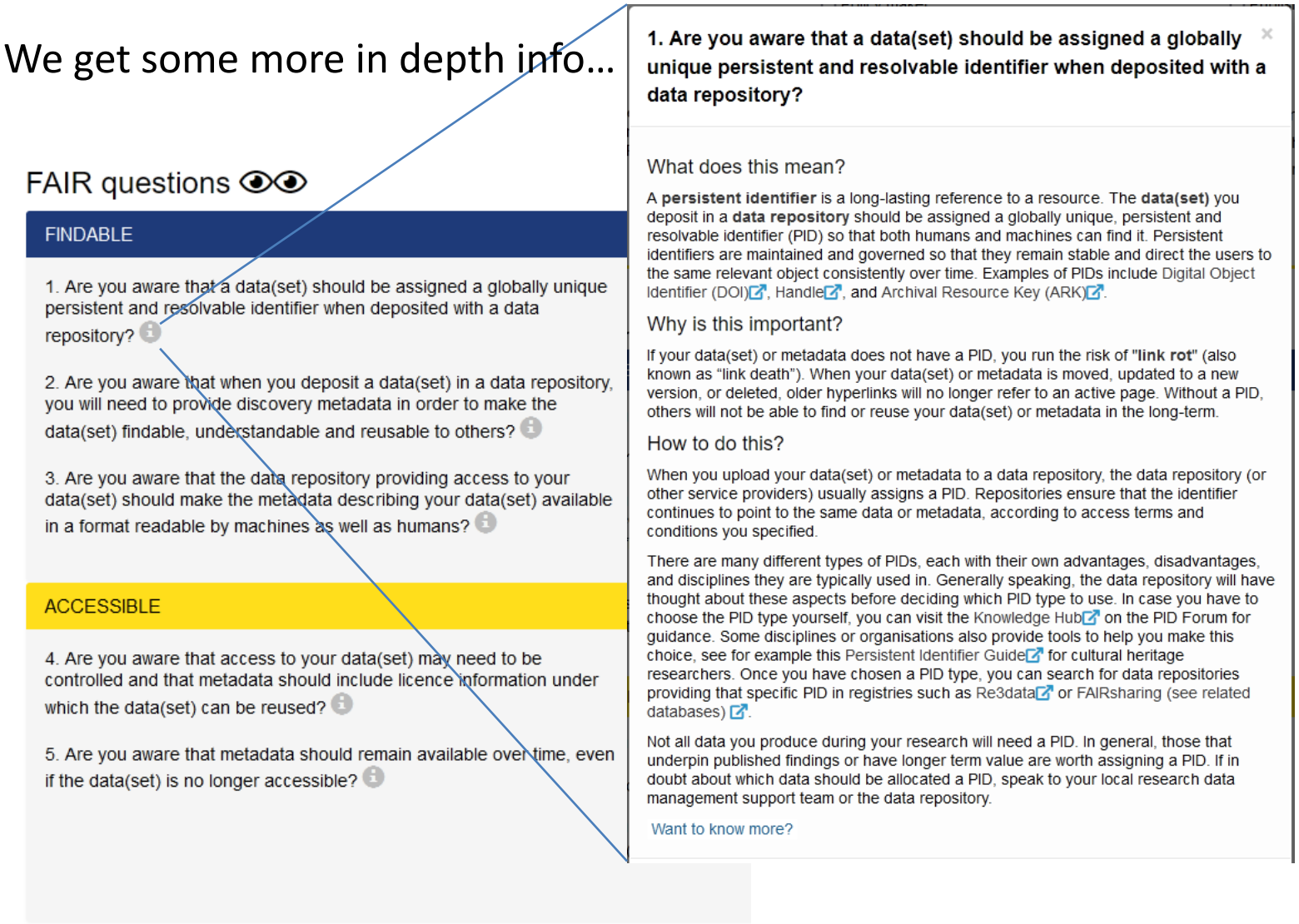
Gap analysis aka FAIR assessment

FAIR as a scale:
where do we stand?

plenty of FAIR assessment tools

Self-assess awareness with respect to FAIR: the FAIR Aware questionnaire

We get some more in depth info…



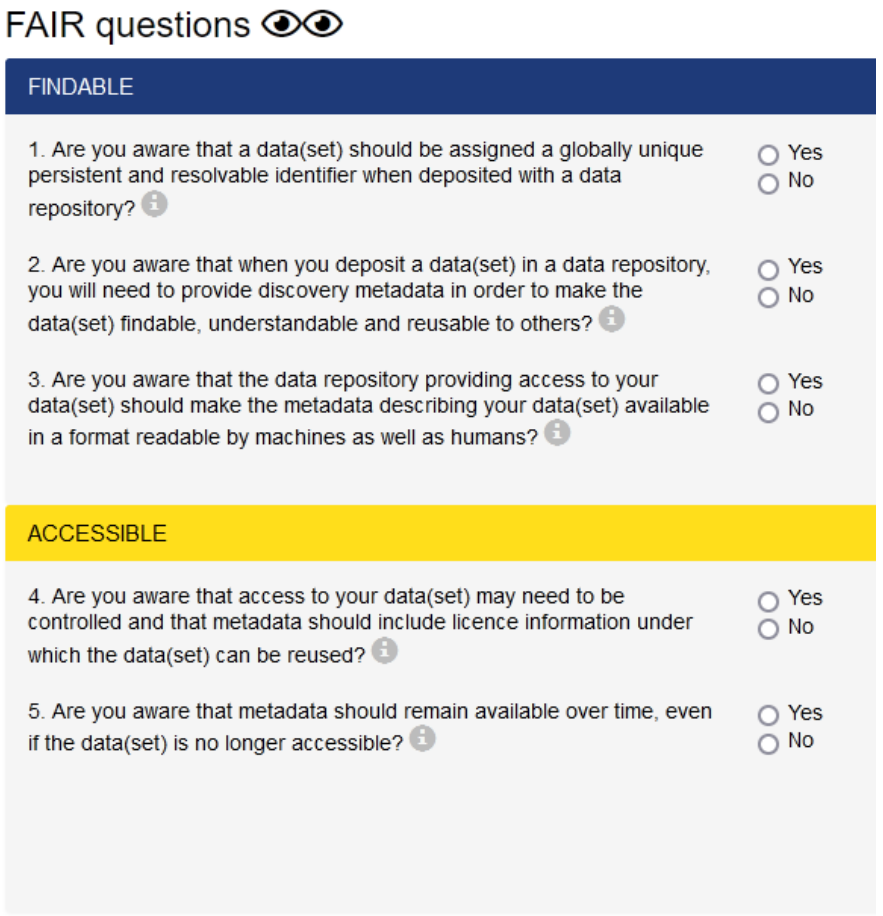
ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

Self-assess awareness with respect to FAIR: the FAIR Aware questionnaire

We get some more in depth info…



And we get some first results…

ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

Self-assess “compliance” with respect to FAIR: the FAIR Aware questionnaire

I INTEROPERABLE

6. the **metadata** describing your datasets **should use semantic vocabularies**

R REUSABLE

8. **metadata** describing your data **should follow** the specifications of **a community-endorsed standard**

7. **provenance information** about the collection and/or generation of data should be **included in the metadata**

10. **maintaining your dataset FAIR over time** requires professional data curation and preservation

..and we can already identify some first issues in our data wrt FAIR…

- **lack of community-agreed metadata standards**
- **lack of provenance information**
- **need for richer metadata**

Checklist to FAIR assess data in repositories: the FAIRsFAIR metrics

FAIRSFAIR Fostering Fair Data Practices in Europe **metrics:**

- a manual checklist
- assess the level of FAIRness of research data objects in trustworthy data repositories
- is a set of 15 criteria based on:
  - "*Turning FAIR into reality*" recommendations
  - Feedback from >30 stakeholders, including data providers and scientific communities
  - RDA RESEARCH DATA ALLIANCE FAIR Data Maturity Model Specification and Guidelines Recommendation of the FAIR Data Maturity Model WG
  - Work on FAIR certification of repositories, **aligned with CoreTrustSeal requirements** CORE TRUST SEAL

"From Conceptualization to Implementation: FAIR Assessment of Research Data Objects" Devaraju et al. (2021)

Checklist to FAIR assess data in repositories: the FAIRsFAIR metrics

FAIRSFAIR Fostering Fair Data Practices in Europe **metrics:**

- a manual checklist
- assess the level of FAIRness of research data objects in trustworthy data repositories
- is a set of 15 criteria based on:
  - *"Turning FAI*
  - Feedback fr
  - RDA RESEARCH DATA ALLIANCE FAIR Data M
  - Work on FAI

FAIRSFAIR object metric	RDA FAIR data maturity model	CoreTrustSeal Alignment
FsF-F1-01D Data is assigned a globally unique identifier.	RDA-F1-02D Data is identified by a globally unique identifier	R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation
FsF-F1-02D Data is assigned a persistent identifier.	RDA-F1-01D Data is identified by a persistent identifier RDA-A1-03D Data identifier resolves to a digital object	

ties

dation of the

nts CORE TRUST SEAL

"From Conceptualization to Implementation: FAIR Assessment of Research Data Objects" Devaraju et al. (2021)

FAIRsFAIR metrics and FAIR data principles

FINDABLE

FsF-F1-01D & 02D
Persistent identifiers (PIDs)

FsF-F2-01M
Rich metadata with core elements

FsF-F3-01M
Metadata includes PID

FsF-F4-01M
Machine readable metadata

ACCESSIBLE

FsF-A1-01M
Metadata has access level & conditions

FsF-A1-02M & 03D
Standard communication protocol

FsF-A2-01M
Metadata is always available

INTEROPERABLE

FsF-I1-01M
Metadata in formal KR language

FsF-I1-02M
Vocabularies & ontologies

FsF-I3-01M
Linked metadata

REUSABLE

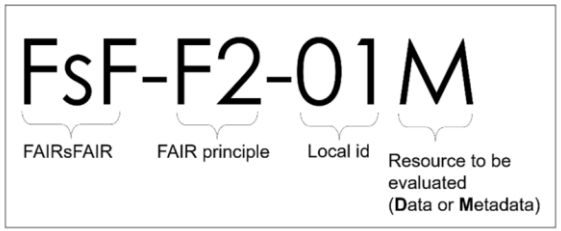
FsF-R1-01MD
Metadata specifies data content

FsF-R1.1-01M
Metadata includes license

FsF-R1.2-01M
Metadata includes provenance

FsF-R1.3-01M & 02D
(Meta)data community standards

Criteria naming convention:



Correspondence with CORE TRUST SEAL certification requirements

Apply FAIRsFAIR metrics to GNSS data: data inventory of the GNSS data stored in the HDC

- What kind of information about our data do we store?
- Where does it come from? From a file?
- Where is it stored? In which database?
- How reliable is this information?
- Could we extract additional relevant information about our data?
- …

Apply FAIRsFAIR metrics to GNSS data: data inventory of the GNSS data stored in the HDC

inventory attribute	description	inventory value	source	DB columns names	DB columns values data type	availability in the HDC repository
title	name given to the resource	RINEX file name	RINEX file	filename	varchar(60)	available
file type		RINEX version	RINEX file	rinex_format	char(2) 2: RINEX 2 3: RINEX 3 long name 3S: RINEX 3 short name	available
file type - versions	file format - version		RINEX file	rinex_version	varchar(5)	available
file size			DB	filesize	int	available
md5/sha1 checksum			DB (for some)	checksum_md5 checksum_sha1	char(32) char(40)	in progress
identifier	An unambiguous reference to the resource within a given context.	permanent ID, landing page	-	-	-	in progress
distribution	Link to (storage/download location of) the data file (hourly/daily RINEX file)	FTP link	retrievable from DB	url	varchar(200)	available
frequency	rate at which something recurs Type of RINEX (daily or hourly) -	only daily data	RINEX file (nominal)/DB	-		available
sampling rate	Typically 30 sec/Hz for the daily observation, but it can be 1 sec or 15 sec.	30 sec by default	RINEX file/DB			nominal available, real in progress
version	Changes to the file(s)		DB	fileversion	smallint incremented at each update of the file	foreseen in the future
creator	Creator: data owner	data owner in site log (S12) (ROR ID if available)	M^3G	-	-	available
contributor	A secondary contributor: point of contact	point of contact in site log (S11)	M^3G	-	-	available
publisher	Publisher	ROB	-	-	-	available
					varchar(30) Type changed in Header RINEX_VERSION RINEX_TYPE MARKER_NAME MARKER_NUMBER RECEIVER_SN RECEIVER_TYPE RECEIVER_VERS	

ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

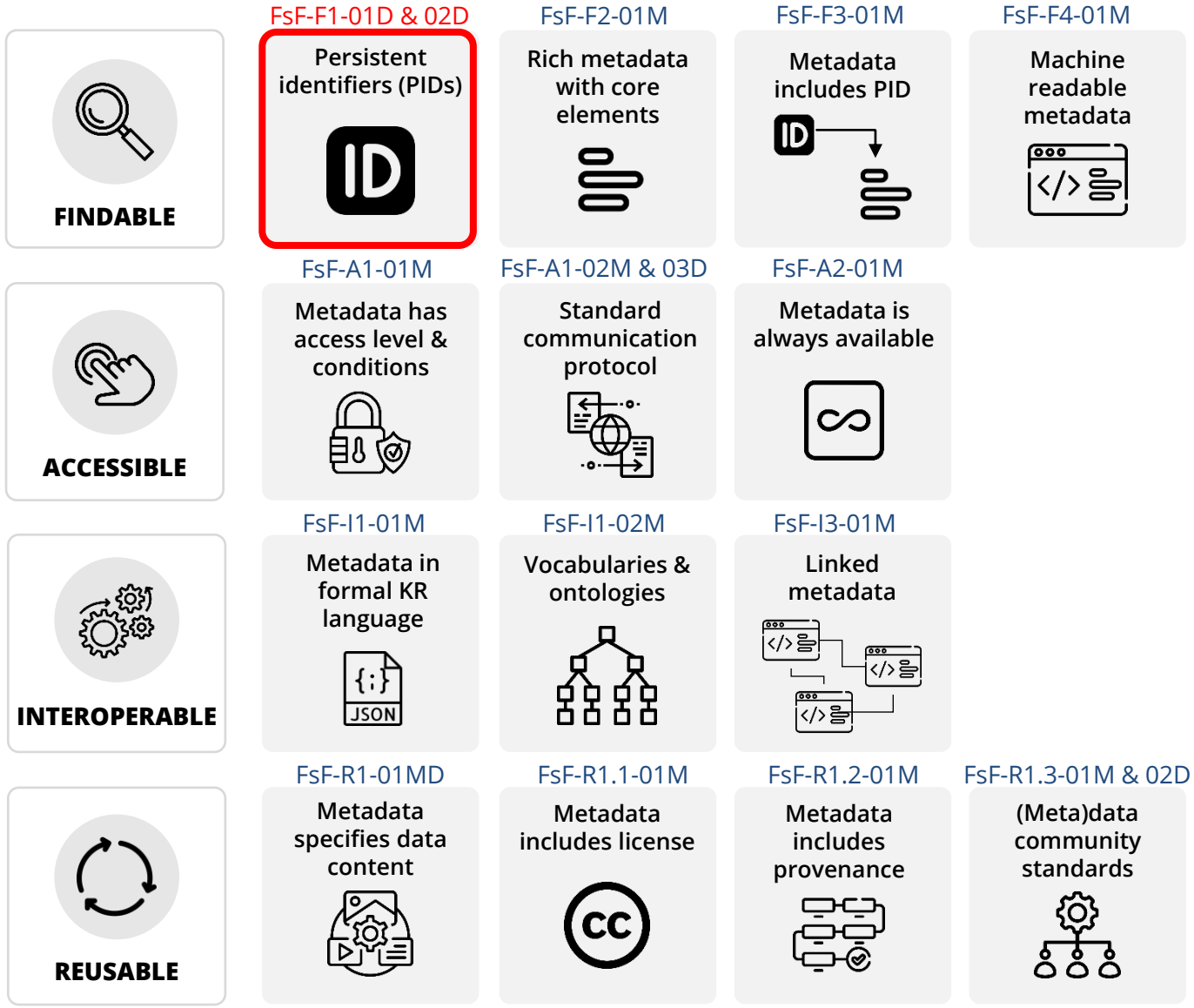
GHENT UNIVERSITY

Apply FAIRsFAIR metrics to GNSS data: data inventory of the GNSS data stored in the HDC

inventory attribute	description	inventory value	source	DB columns names	DB columns values data type	availability in the HDC repository
title	name given to the resource	RINEX file name	RINEX file	filename	varchar(60)	available
file type		RINEX version	RINEX file	rinex_format	char(2) 2: RINEX 2 3: RINEX 3 long name 3S: RINEX 3 short name	available
file type - versions	file format				varchar(5)	available
file size					int	available
md5/sha1 checksum					char(32) char(40)	in progress
identifier	An unambi within a giv				-	in progress
distribution	Link to (sto data file (h				varchar(200)	available
frequency	rate at whi RINEX (dail					available
sampling rate	Typically 3(but it can b					nominal available, real in progress
version	Changes to				smallint mented at each update of the file	foreseen in the future
creator	Creator: da	(ROR ID if available)			-	available
contributor	A secondary contributor: point of contact	point of contact in site log (S11)	M^3G	-	-	available
publisher	Publisher	ROB	-	-	-	available
					varchar(30) Type changed in Header RINEX_VERSION RINEX_TYPE MARKER_NAME MARKER_NUMBER RECEIVER_SN RECEIVER_TYPE RECEIVER_VERS	

- Essential task, but very time consuming!
- Effort from the whole team: trace all source of information

Apply FAIRsFAIR metrics to the HDC GNSS data



ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

Apply FAIRsFAIR metrics to the HDC GNSS data: an example

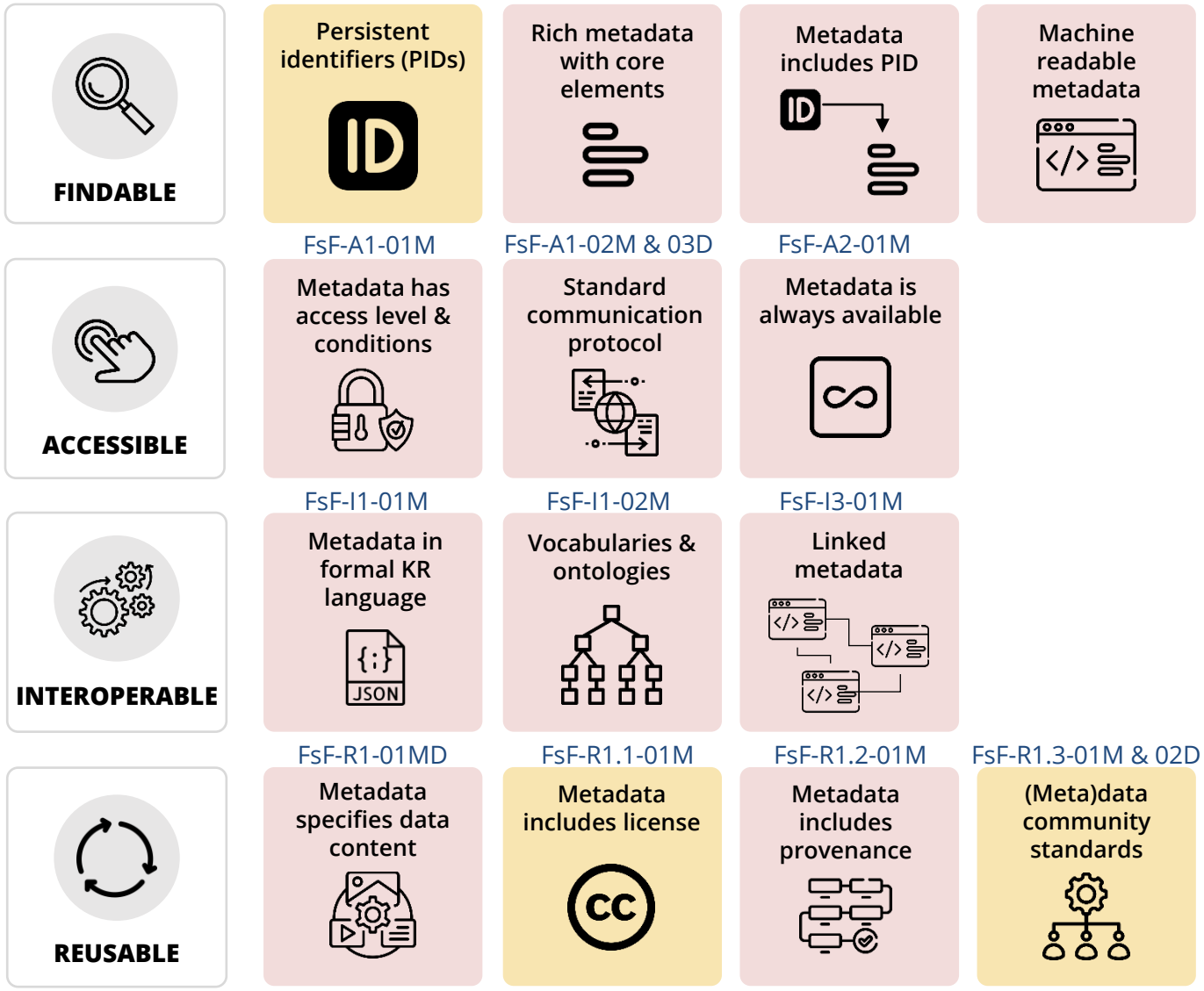
Metric		Assessment		Fulfilment	
Identifier	**Name**	**Requirement(s)**	**Method(s)**	**Requirements fulfillment**	**Description**
FsF-F1-01D	Data is assigned a globally unique identifier.	*a) Data identifier (IRI, URL)* *b) List of globally unique identifier schemes*	*Check if the identifier is specified based on a globally unique identifier scheme.*	Yes	*GNSS RINEX 3 data has a unique naming convention that is also accepted by the GNSS community. The RINEX 3 file name is a unique identifier for GNSS data. Depending on the Data repository every rinex data has a URI. Schema for the daily data: <data-repository>/<year>/<doy>/<conventional-rinex-name>.Z*
FsF-F1-02D	Data is assigned a persistent identifier.	*a) Data identifier (IRI, URL)* *b) Landing page of the identifier* *c) List of commonly accepted persistent identifiers for data*	*Check if the data identifier specified is based on a commonly accepted persistent identifier scheme, and it resolves to a landing page with metadata containing further information on how to access the data object. Note that this assessment method follows the current best practice to have a PID resolve to a landing page instead of its actual content.*	No	*Data has a URI identifier, but it dos not resolve to a landing page*

FAIRsFAIR metrics to the HDC GNSS data: first results overview

Not adopted

Partially adopted

Adopted

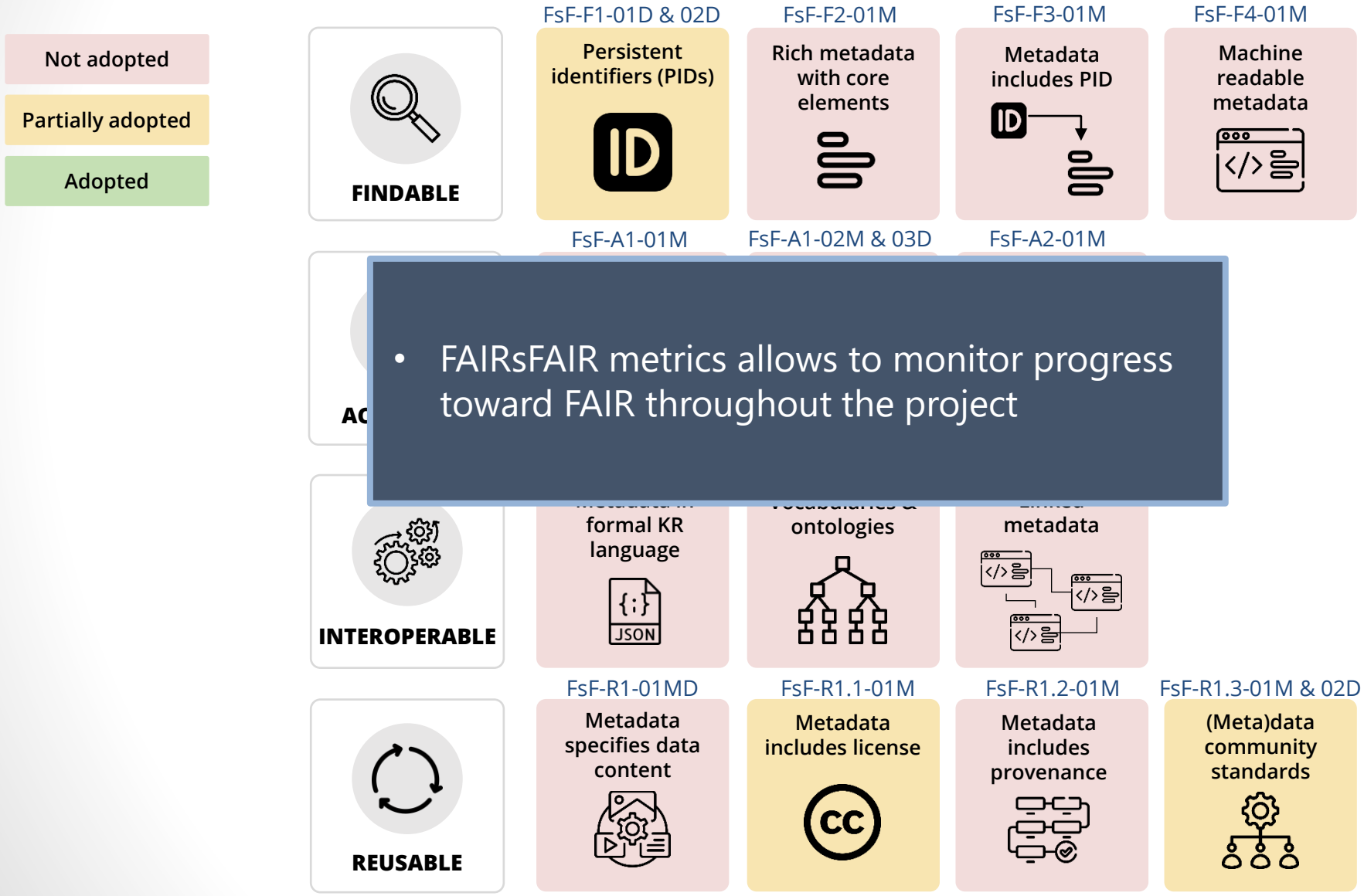


ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

FAIRsFAIR metrics to the HDC GNSS data: first results overview



- FAIRsFAIR metrics allows to monitor progress toward FAIR throughout the project

ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

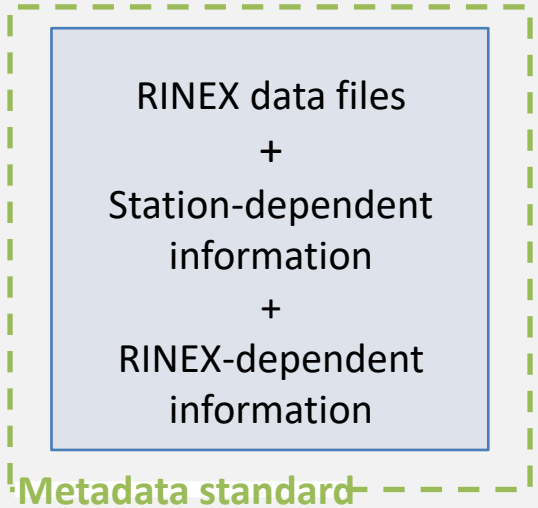
Gap analysis: where do we stand wrt FAIR? Summary

We identified the main issues to be addressed to move towards FAIR:

- Self-assessed awareness wrt FAIR via the FAIR Aware questionnaire
- Applied the FAIRsFAIR metrics to our GNSS data:
  - Performed a complete data inventory of the HDC repository as a preliminary step
  - First evaluation of the FAIRness of the GNSS data stored in the HDC: missing permanent identifiers, provenance information metadata, license ..

Coming up next, after the coffee break:

Put theory into practice: start by attaching metadata to our GNSS data



Coffee break
Reconvene @ 15:50 GMT+2

Contents of the webinar

- **Part 1 - GNSS data and the data repository**
- **Part 2 - Introduction to FAIR data principles**
- **Part 3 - The journey so far: FAIR for GNSS data**
  - Where and how to start?
  - Gap analysis a.k.a. FAIR assessment
  - Metadata for GNSS data
  - Restructuring the data repository
  - Persistent identifiers for GNSS data
  - API preparatory work
- **Part 4 – Lessons learned and next steps**

Metadata for GNSS data

FAIR-GNSS methodology

RINEX data files
+
Station-dependent information
+
RINEX-dependent information

Metadata standard

…start by answering some questions

What about metadata standards?

Which metadata standards? DCAT? DublinCore? DataCite?

Which metadata? For RINEX files, for station information…

What kind of metadata? Based on user needs for data analysis?

ROYAL OBSERVATORY OF BELGIUM

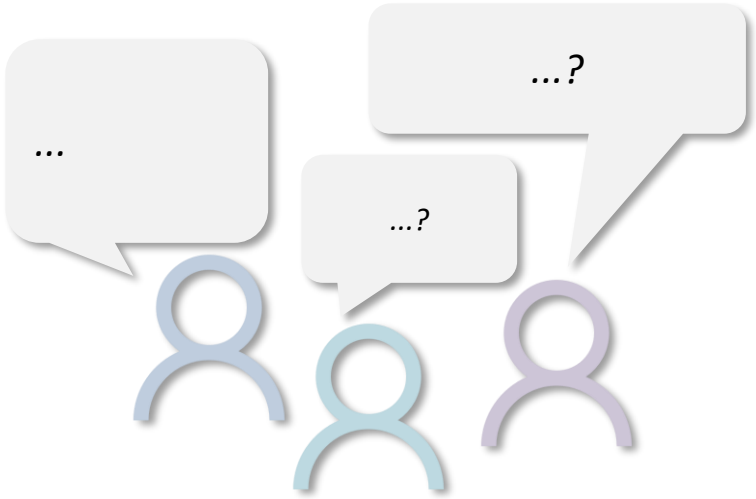
FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

Which information to provide as metadata?

Choice of **additional information to be provided as metadata** based on:

user requests when searching/downloading GNSS data



FAIR data principles
(data usage conditions, data provenance, standardization, …)

FINDABLE
- FsF-F1-01D & 02D: Persistent identifiers (PIDs)
- FsF-F2-01M: Rich metadata with core elements
- FsF-F4-01M: Machine readable metadata

ACCESSIBLE
- FsF-A1-01M: Metadata has access level & conditions

INTEROPERABLE
- FsF-I1-02M: Vocabularies & ontologies

REUSABLE
- FsF-R1.1-01M: Metadata includes license
- FsF-R1.2-01M: Metadata includes provenance
- FsF-R1.3-01M & 02D: (Meta)data community standards

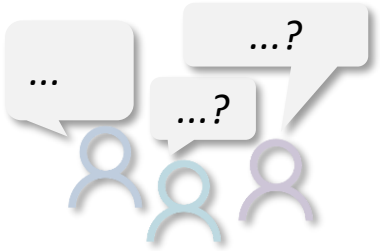
ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

Which information to provide as metadata?

Selected information based on user perspective



Station-dependent metadata
All info currently stored in the site log
DOI (Digital Object Identifier) - necessary to cite data owner when using the station data
License info - providing conditions of use
Provenance info (e.g. changes occurred to the site log after submission to M^{3}G)
Precise station coordinates (including their reference frame)
RINEX-dependent metadata
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)

RINEX data files
+
Station-dependent information
+
RINEX-dependent information

Metadata standard

RINEX data files
+
Station-dependent metadata: *station description, license, data citation info, …*
+
RINEX-dependent metadata: *provenance, link to site log, data quality, …*

Metadata standard

Station- & RINEX-dependent metadata

Station-dependent metadata:

Station description/site log, data license, data citation info (DOI), …

Optionally: GNSS network to which station contributes, provenance info, individual antenna calibration, ETRS89 coordinates, …

RINEX-dependent metadata (i.e. created on-the-fly upon data download):

Data identifier, repository, creation & publication date, data license, summary and keywords,

Optionally: quality check results, MD5, file size, …

Proposal for station-dependent metadata

RINEX data files
+
Station-dependent metadata
+
RINEX-dependent metadata

Metadata standard-

Proposal for RINEX-dependent metadata

Station- & RINEX-dependent metadata

Station-dependent metadata:

Station description/site log, data license, data citation info (DOI), …

Optionally: GNSS network to which station contributes, provenance info, individual antenna calibration, ETRS89 coordinates, …

RINEX-dependent metadata (i.e. created on-the-fly upon data download):

Data identifier, repository, creation & publication date, data license, summary and keywords,

Optionally: quality check results, MD5, file size, …

Proposal for station-dependent metadata

RINEX data files
+
Station-dependent metadata
+
RINEX-dependent metadata

Metadata standard-

Proposal for RINEX-dependent metadata

ROYAL OBSERVATORY OF BELGIUM

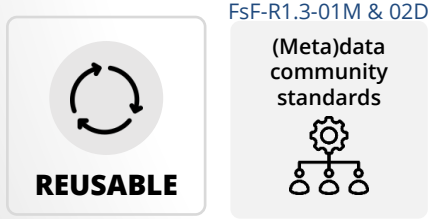
FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

Station-dependent metadata: which metadata standard?

Choice of **metadata schema(s)** based on:

- best practices within the scientific community and at other GNSS data repositories all over the world
- FAIR data principles
- user perspective when accessing and using the data
- data managers perspective when managing and preserving the data e.g. ROB



We observed that for **station-dependent metadata**
there exists
a **community-agreed standard format for station metadata**:
GeodesyML

Station-dependent metadata: GeodesyML

Australian Government
Geoscience Australia

GeodesyML

- **aligned with international standards** (ISO19115-1:2014 **and the Geography Markup Language (GML)** developed within the Open Geospatial Consortium (OGC)
- an **application schema of GML** for transfer of geodetic information
- first proposed in 2016
- **supported by the International GNSS Service (IGS)**
- in use within EPOS and the M^{3}G system

Boler et al. (2017) Progress toward a standard-based XML system for IGS network site log meta-data management and dissemination using GeodesyML.
Brown et al. (2016) Maximising interoperability and discoverability of geodetic products and services.

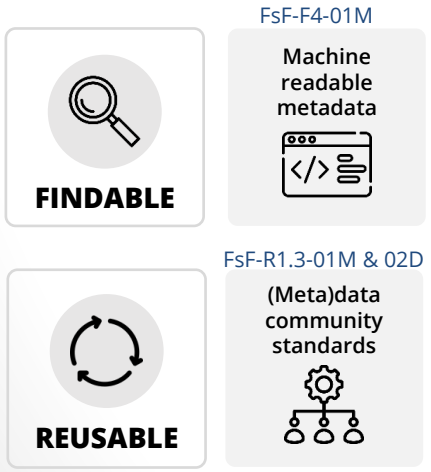
Station-dependent metadata: GeodesyML

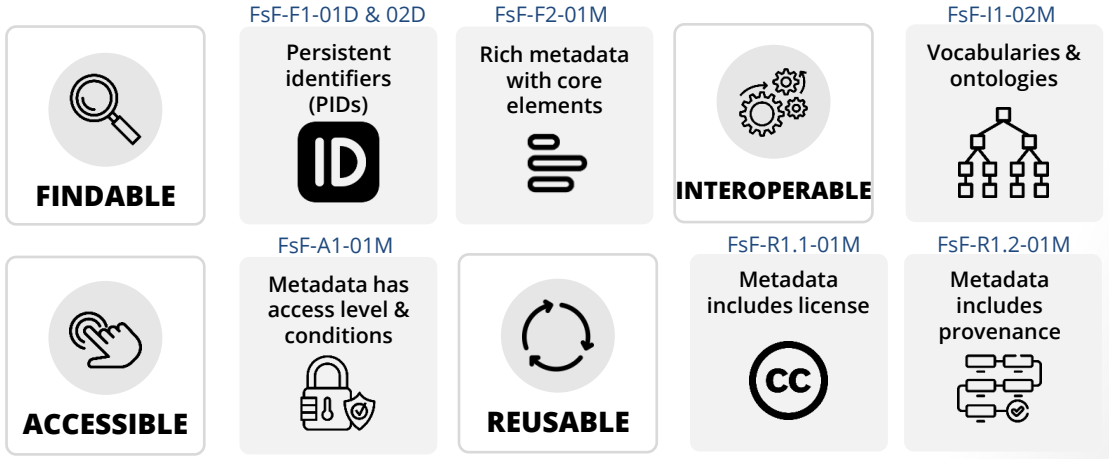
What is available in GeodesyML

- open source XML-based format
- aligned with international standards (ISO19115-1:2014 and the Geography Markup Language (GML)
- includes all information contained in the IGS site log, some additional info and no information is ever deleted (possibility to track changes)

What is missing in GeodesyML

- data license information
- provenance information
- use PIDs
- controlled terminology via code list files





ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

Station-dependent metadata: proposal to extend GeodesyML

Metadata element and description	GeodesyML element
Core metadata for GNSS station	
Site log content	`geo:siteLog`
License of use	`geo:license` to be added to GeodesyML,
Precise coordinates	`geo:Position`
DOI	`geo:SiteLogType/geo:moreInformation/geo:DOI`
Optional metadata for GNSS station	
File provenance information i.e. list of changes to site log	`geo:modifiedSection` to be added to GeodesyML,
Data curator (i.e. organization uploading, validating and distributing data) identifier	`geo:Metadata` to be added to GeodesyML,
Site pictures	`geo:Document`
Antenna calibration	`geo:Document`
GNSS network(s)	`geo:gnssnetwork` to be added to GeodesyML,

Station-dependent metadata: extended GeodesyML examples

controlled terminology/ code lists for:

- **license,**
- tectonic plates,
- receiver firmware,
- GNSS data centers, networks,
- types of changes in the site log,
- …

Code lists files to be maintained @ IGS

IGS

For example:

`geo:license` as a new property of the `geo:Document` GeodesyML class +
an associated **code list file**
`license-codelists.xml`

```
<codelistItem>
    <CodeListDictionary gml:id='GeodesyML_LicenseTypeCode'>
        <gml:description>File License information</gml:description>
        <gml:identifier codeSpace='urn:gnss-metadata.eu:gnss:license'>GeodesyML_LicenseTypeCode</gml:identifier>

            <codeEntry>
                <CodeDefinition gml:id='LicenseTypeCode_CC-BY-4.0'>
                                        <gml:description xlink:href="https://creativecommons.org/licenses/by/4.0/"/>
                    <gml:identifier codeSpace='urn:gnss-metadata.eu:gnss:license'>CC-BY-4.0</gml:identifier>
                                        <gml:name codeSpace='urn:gnss-metadata.eu:gnss:license'>Creative Commons Attr
                </CodeDefinition>
            </codeEntry>


                        <codeEntry>
                <CodeDefinition gml:id='LicenseTypeCode_CC0'>
                                        <gml:description xlink:href="https://creativecommons.org/publicdomain/zero/1.
                    <gml:identifier codeSpace='urn:gnss-metadata.eu:gnss:license'>CC0-1.0</gml:identifier>
                                        <gml:name codeSpace='urn:gnss-metadata.eu:gnss:license'>Creative Commons Zero
                </CodeDefinition>
            </codeEntry>

    </CodeListDictionary>
</codelistItem>
```

ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

Station-dependent metadata: extended GeodesyML proposal

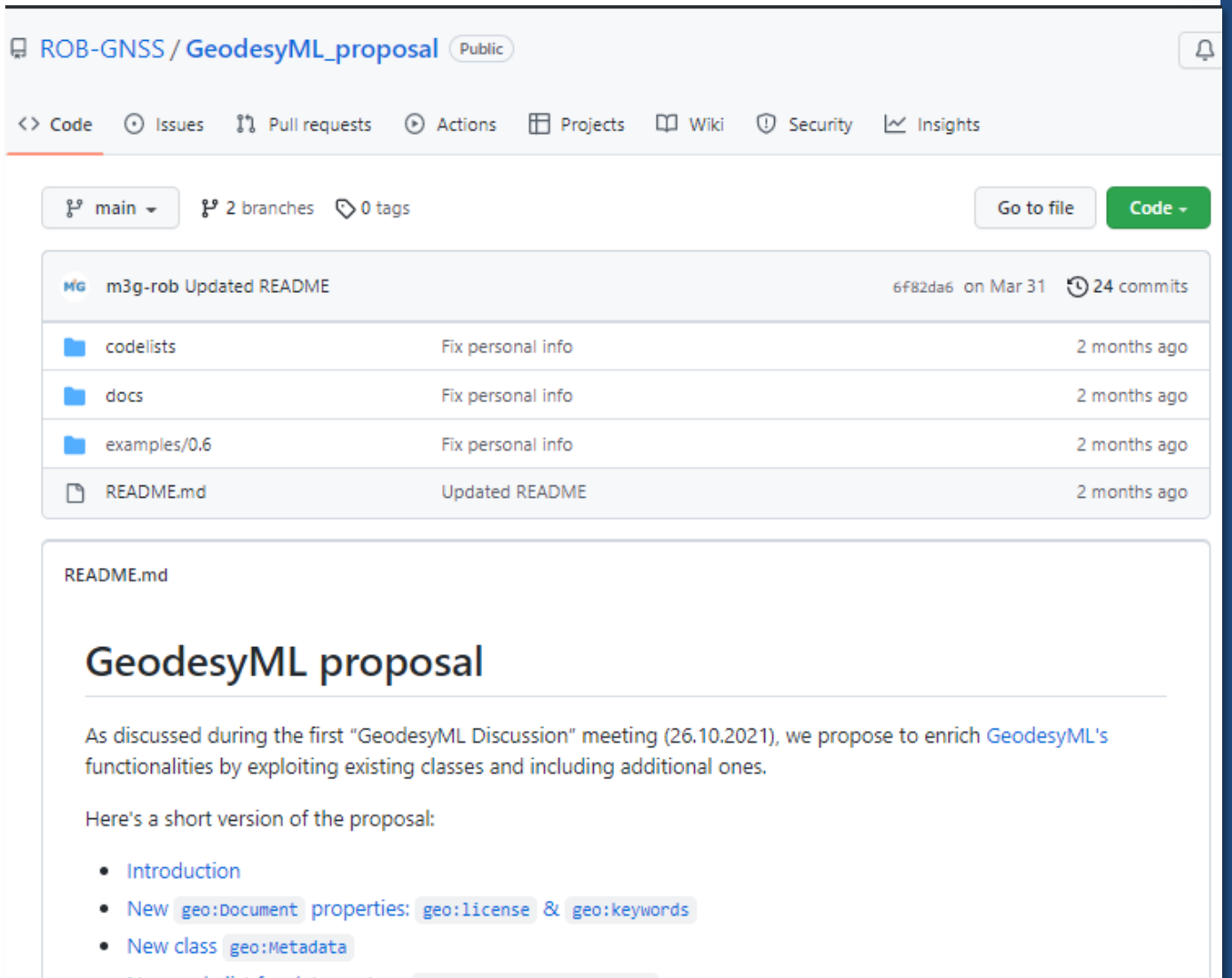
Richer metadata e.g.

- publisher,
- **track modifications**,
- **data license**,
- GNSS networks.
- **identifiers** (ROR or ORCID)

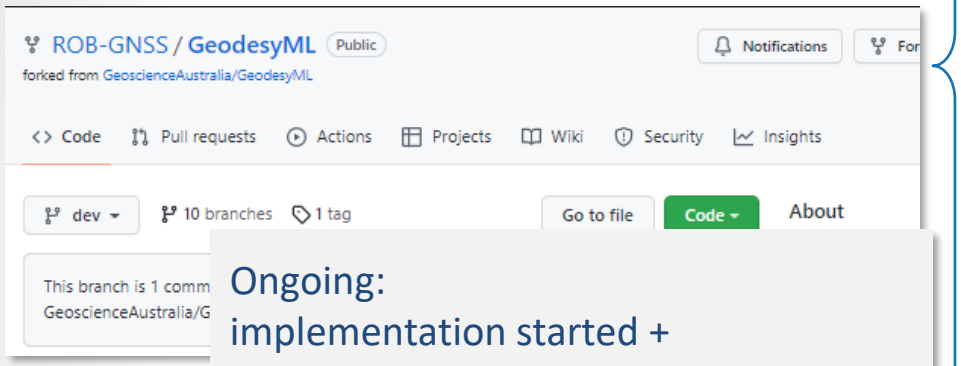
Controlled terminology/ code lists for:

- **license,** …
- tectonic plates,
- receiver firmware,
- GNSS data centers, networks,
- types of changes in the site log,
- …

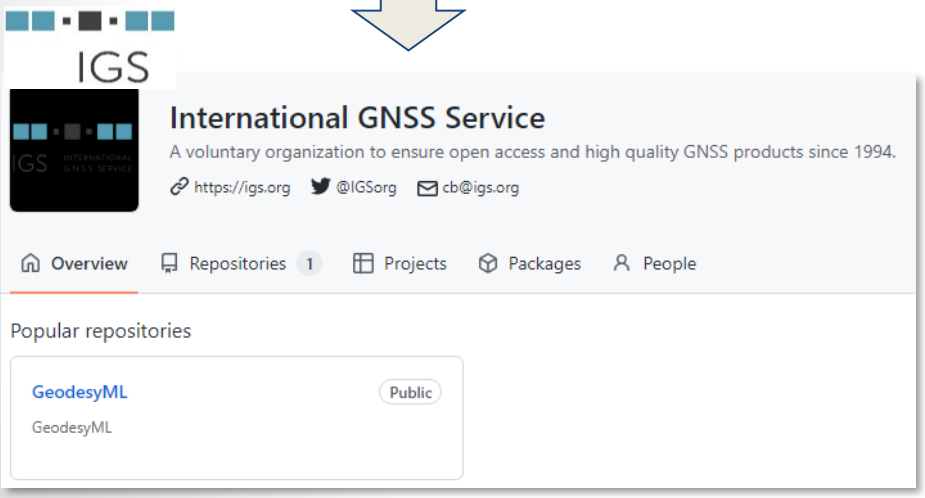
Get all this extra information when downloading GeodesyML files e.g. from M³G

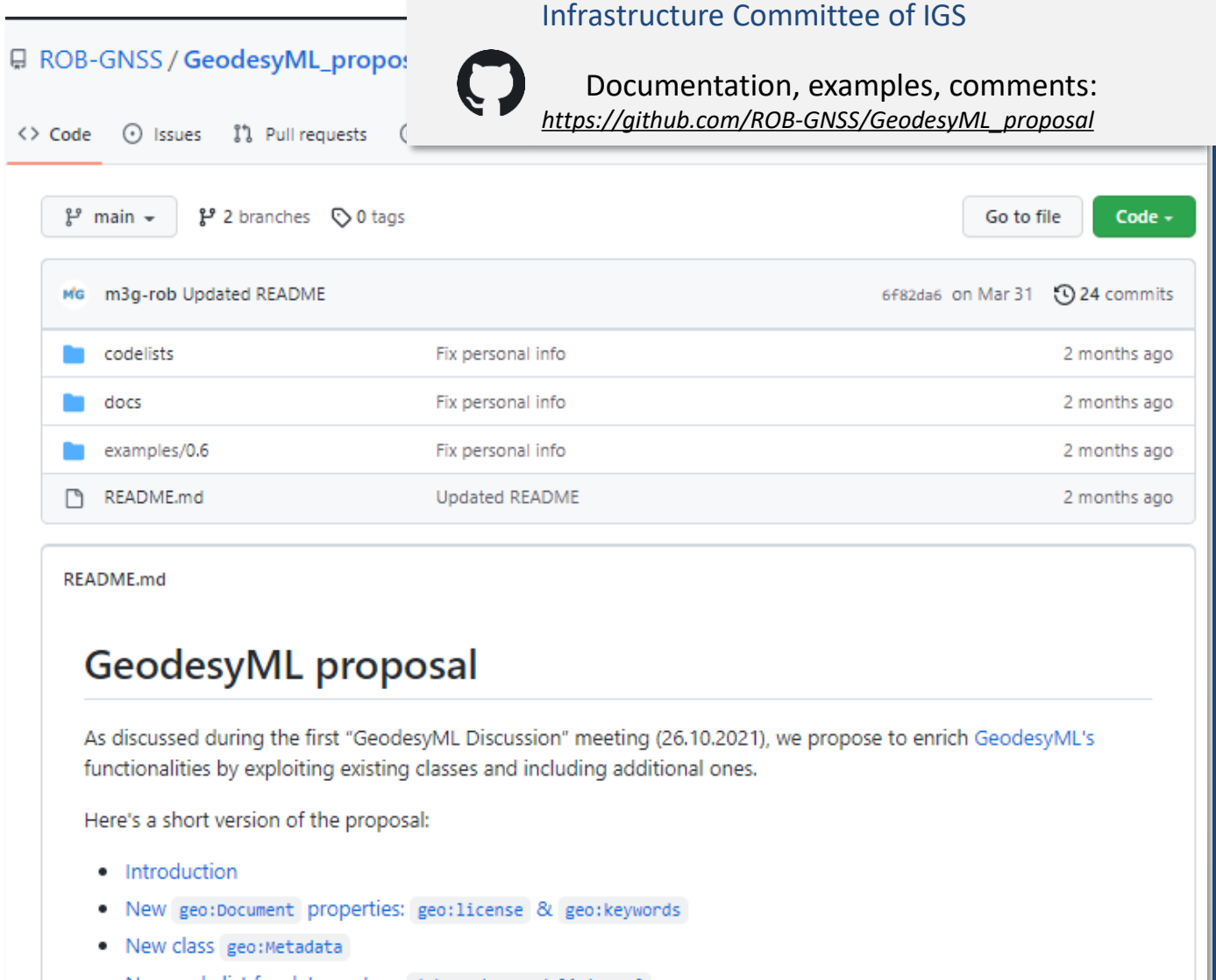


Station-dependent metadata: extended GeodesyML proposal @ GitHub



Ongoing:
implementation started +
new GeodesyML repo @IGS where our contributions will soon be merged





Discussed with the GeodesyML Task Force of the Infrastructure Committee of IGS

Documentation, examples, comments:
https://github.com/ROB-GNSS/GeodesyML_proposal

ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

Station- & RINEX-dependent metadata

Station-dependent metadata:

Station description/site log, data license, data citation info (DOI), …

Optionally: GNSS network to which station contributes, provenance info, individual antenna calibration, ETRS89 coordinates, …

RINEX-dependent metadata (i.e. created on-the-fly upon data download):

Data identifier, repository, creation & publication date, data license, summary and keywords,

Optionally: quality check results, MD5, file size, …

RINEX data files
+
Station-dependent metadata
+
RINEX-dependent metadata

Metadata standard

Proposal for station-dependent metadata

Proposal for RINEX-dependent metadata

RINEX-dependent metadata which metadata standard?

Choice of **metadata schema(s)** based on:

- user perspective when accessing and using the data, as well as those managing and preserving the data e.g. ROB
- best practices within the scientific community and at other GNSS data repositories all over the world
- FAIR data principles

We observed that for **RINEX-dependent metadata:**

- no metadata standard is available to describe RINEX data in GNSS data repositories

- identify existing metadata schema(s) covering the needs of the GNSS community to a maximal extent
- if needed, extend the schema of choice and develop an application profile

Explore existing metadata schemas

Examine existing metadata schemes (e.g. DataCite, schema.org or DCAT) used in Earth science community, and controlled vocabularies to populate metadata.

- Generic metadata schemas and vocabularies (*DCAT*, *DataCite*, *Dublin Core*, *Schema.org*) were considered

DCAT-AP field (Mandatory, Recommended, Optional)	description	DCAT-AP value	DataCite field (Mandatory, Recommended, Optional)	Comments
dct:title (M)	Title: a name given to the resource.	rdfs:Literal [1..n]	Title (M)	
dct:format (M)		dct:MediaTypeOrExtent [0..1]	ResourceType (M)	
dct:creator (M)	Creator: data owner (ROR ID if available)	foaf:Agent [0..1]	Creator (M)	
dct:publisher (M)	Publisher: ROB	foaf:Agent [0..1]	Publisher (M)	
dct:identifier (M)	An unambiguous reference to the resource within a given context.	rdfs:Literal [0..n]	Identifier (M)	DataCite requires this to be a DOI, whereas DCAT-AP does not have such requirement
dct:issued (M)	Timestamp for data & metadata - Date of submission of the resource	rdfs:Literal typed as xsd:date or xsd:dateTime[0..1]	PublicationYear (M) Date (R)	
dct:accessRights (M)	Access conditions to the files (open or not)	dct:RightsStatement[0..1]	Rights (O)	

Explore existing metadata schemas

Data Catalog Vocabulary (DCAT) - Version 2
W3C Recommendation 04 February 2020

DCAT Application Profile for data portals in Europe Version 2.1.0

metadata mapping

INSPIRE ➡ **DCAT** ➡ EUROPEAN DATA PORTAL

Focus on **DCAT** and DataCite existing generic metadata schemas
We mapped the content of the HDC database vs the existing metadata fields of the various schemas

metadata field	description	comments	mandatory	DCAT 3	DataCite field (Mandatory (6), Recommended, Optional)	data type
title/name	Title: a name given to the resource.	RINEX file name	mandatory	dct:title (M)	Title (M)	
file type		RINEX version		dct:format (M)	ResourceType (M)	
creator	Creator: data owner (ROR if available)	match the site log (S12)	mandatory	dct:creator (M)	Creator (M)	
publisher	Publisher: ROB		mandatory	dct:publisher (M)	Publisher (M)	
identifier	An unambiguous reference to the resource within a given context.	permanent ID, landing page	mandatory	dct:identifier (M)	Identifier (M)	station
dateSubmitted	Timestamp for data & metadata - Date of submission of the resource	dateLastSubmission	mandatory	dct:issued (M)	PublicationYear (M) Date (R)	observation (RINEX)
rights	Access conditions to the files (open or not)	always open	mandatory	dct:accessRights (M)	Rights (O)	observation (RINEX)
License	A legal document giving official permission to do something with the resource.	from M3G	mandatory	dct:license	-	station
md5 checksum		to be inserted in DB	mandatory	spdx:checksum	-	observation (RINEX)
Frequency	Type of RINEX (daily or hourly) - A rate at which something recurs	in the DB	optional	-	-	observation (RINEX)

ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

Explore existing metadata schemas: lesson learned

- Generic metadata schemas and vocabularies (DCAT, DataCite, Dublin Core, Schema.org) were considered, but none is capable to cover GNSS data specificities as it is.
- The most appropriate choice of vocabulary seemed to be the NASA Earth data GCDM Keywords , already in use within the community (e.g. GFZ data services).
- No specific field to store the links to the RINEX files in GeodesyML.
- Extend the **DCAT Application Profile** for open datasets and data portals (**DCAT-AP**) to include metadata attributes specific to RINEX files: GNSS-DCAT-AP.
- Considering Linked Open data (LOD) i.e. 5-☆ data, **DCAT-AP** is the schema of choice for its natural linked-data approach (JSON-LD as encoding format)

Data Catalog Vocabulary (DCAT) - Version 2
W3C Recommendation 04 February 2020

DCAT Application Profile for data portals in Europe
Version 2.1.0

metadata mapping

INSPIRE → **DCAT** → EUROPEAN DATA PORTAL

proposal for a DCAT-AP for GNSS data

RINEX-dependent metadata: proposal for a DCAT-AP for GNSS data

- type of RINEX file (e.g., compression format, frequency)
- RINEX file header: GNSS antenna and receiver type, …
- GeodesyML: GNSS antenna and receiver type, …
- software used to generate the RINEX file.

GNSS-DCAT-AP

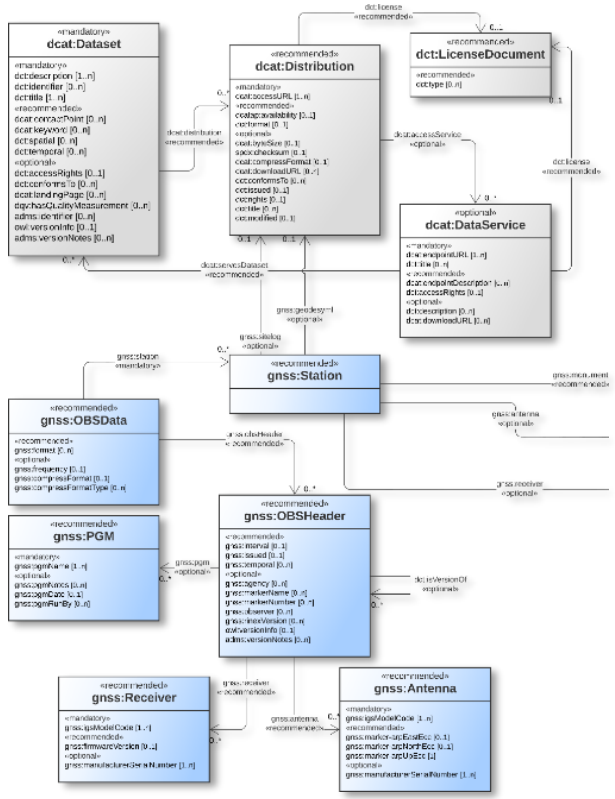
=

DCAT-AP classes

+

new RINEX-specific classes

…structured information in a **metadata schema** **GNSS-DCAT-AP**



RINEX-dependent metadata: GNSS-DCAT-AP proposal

GNSS-DCAT-AP
=
DCAT-AP classes
+
new RINEX-specific classes

- type of RINEX file (e.g., compression format, frequency)

…

GNSS-DCAT-AP recommended classes

Class name	URI	Mandatory/ Recommended/ Optional	Description
GNSS observation data	gnss:OBSData	Recommended	Domain specific vocabulary for RINEX observation files

properties

gnss:OBSData

Property	URI	Range & Cardinality	Mandatory/ Recommended/ Optional	Description
file format	gnss:format	rdfs:Literal [0..n]	Recommended	RINEX 2 / RINEX 3/ RINEX 4
frequency	gnss:frequency	dct:Frequency[0..1]	Optional	Daily/hourly RINEX files
CRX compression	gnss:compressFormat	xsd:boolean[0..1]	Optional	CRX compressed RINEX file (CRX RINEX)
CRX version	gnss:compressFormatType	rdfs:Literal [0..n]	Optional	Crinex Version (e.g. 3.0)

RINEX-dependent metadata: GNSS-DCAT-AP proposal

GNSS-DCAT-AP

=

DCAT-AP classes

+

new RINEX-specific classes

- type of RINEX file (e.g., compression format, frequency)
- RINEX file: GNSS antenna and receiver type, …
- software used to generate the RINEX file
- GeodesyML: GNSS antenna and receiver type, …

GNSS-DCAT-AP recommended classes

Class name	URI	Mandatory/ Recommended/ Optional	Description
GNSS station antenna	gnss:Antenna	Recommended	Domain specific vocabulary for the antenna associated with gnss:Station

GNSS-DCAT-AP optional classes

Class name	URI	Mandatory/ Recommended/ Optional	Description
GNSS station antenna from GeodesyML	geo:GNSSAntenna	Optional	Antenna info from GeodesyML installed on gnss:temporal (associated with gnss:Station)
GNSS station monument from GeodesyML	geo:Monument	Optional	Monument info from GeodesyML (associated with gnss:Station)
GNSS station receiver from GeodesyML	geo:GNSSReceiver	Optional	Receiver info from GeodesyML installed on gnss:temporal (associated with gnss:Station)

RINEX-dependent metadata: GNSS-DCAT-AP proposal

GNSS-DCAT-AP

=

DCAT-AP classes

+

new RINEX-specific classes

- Data license
- Quality metrics

DCAT-AP recommended classes

Class name	URI	Mandatory/ Recommended/ Optional	Description
License document	dct:LicenseDocument	Recommended	A legal document giving official permission to use the dataset.

dcat:Dataset

Property	URI	Range & Cardinality	Mandatory/ Recommended/ Optional	Description
has quality measurement	dqv:hasQualityMeasurement	dqv:QualityMeasurement [0..n]	Optional	A quality measurement performed on the Dataset (e.g. ratio of the number of GPS observations, on at least two frequencies, in the daily RINEX file with respect to the number of expected observations)

RINEX-dependent metadata: GNSS-DCAT-AP proposal

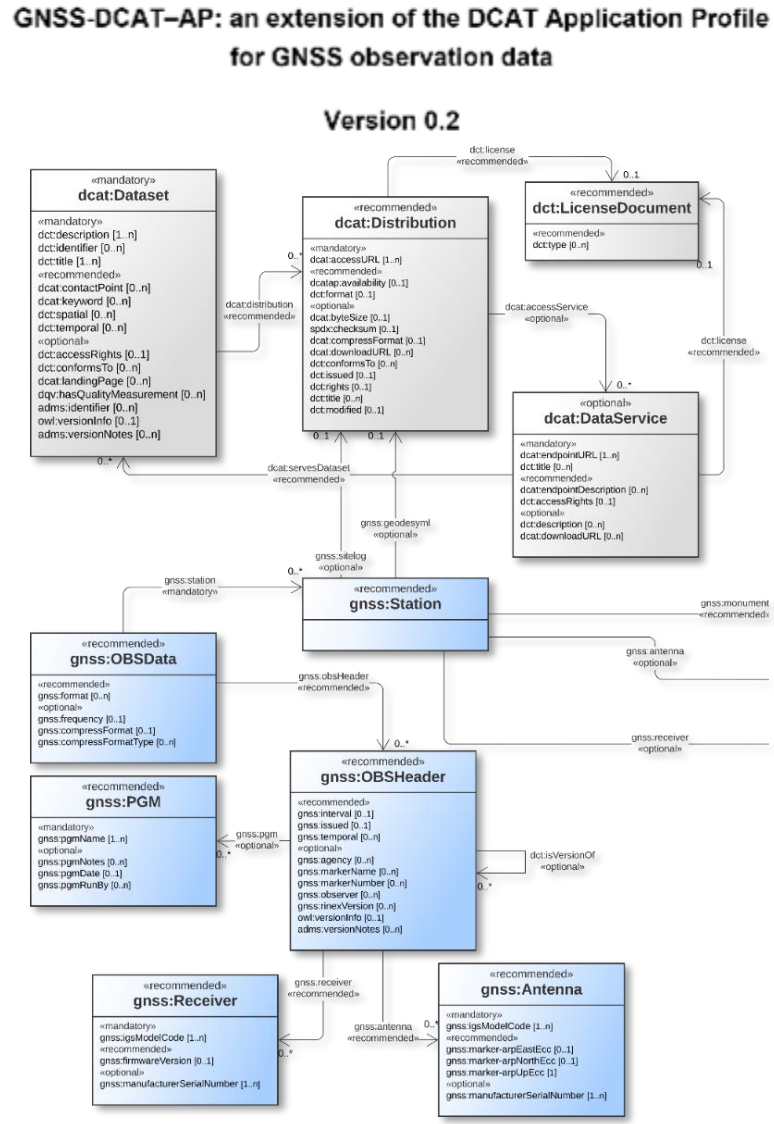
GNSS-DCAT-AP

=

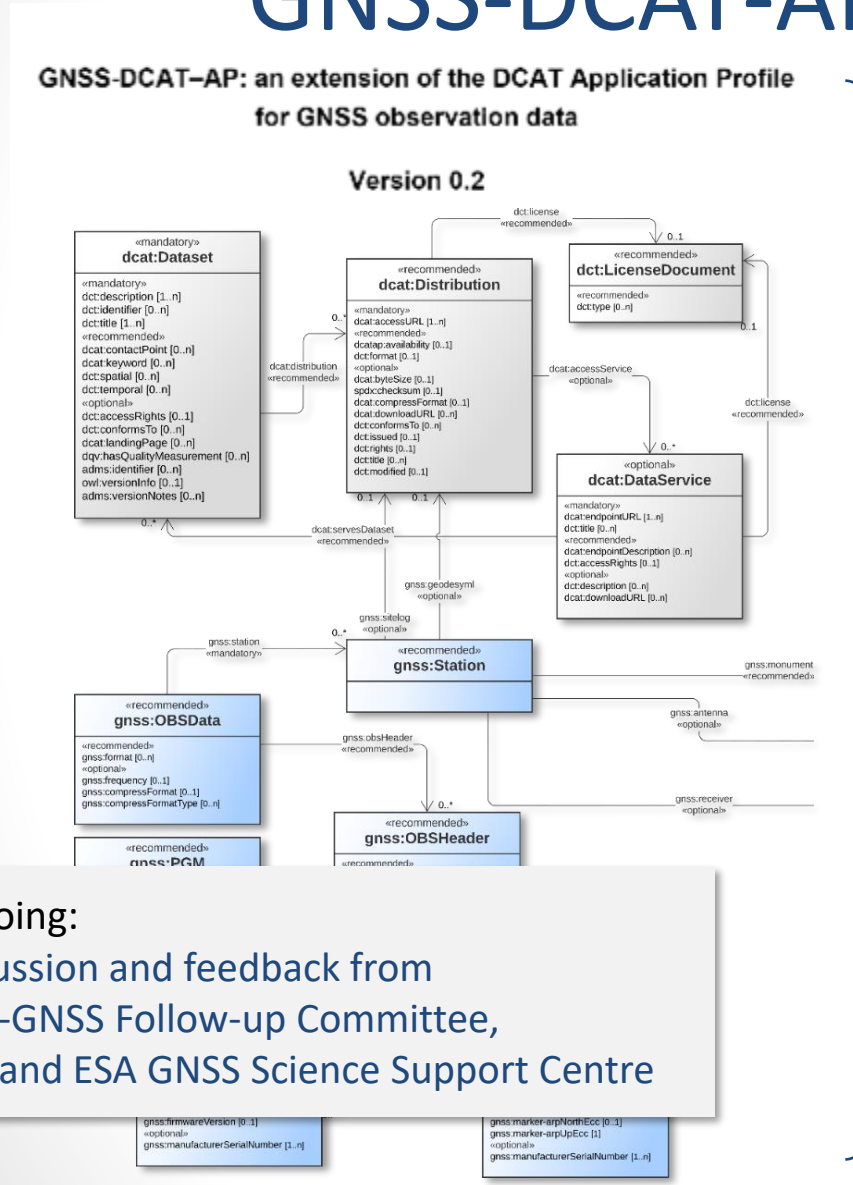
DCAT-AP classes

+

new RINEX-specific classes



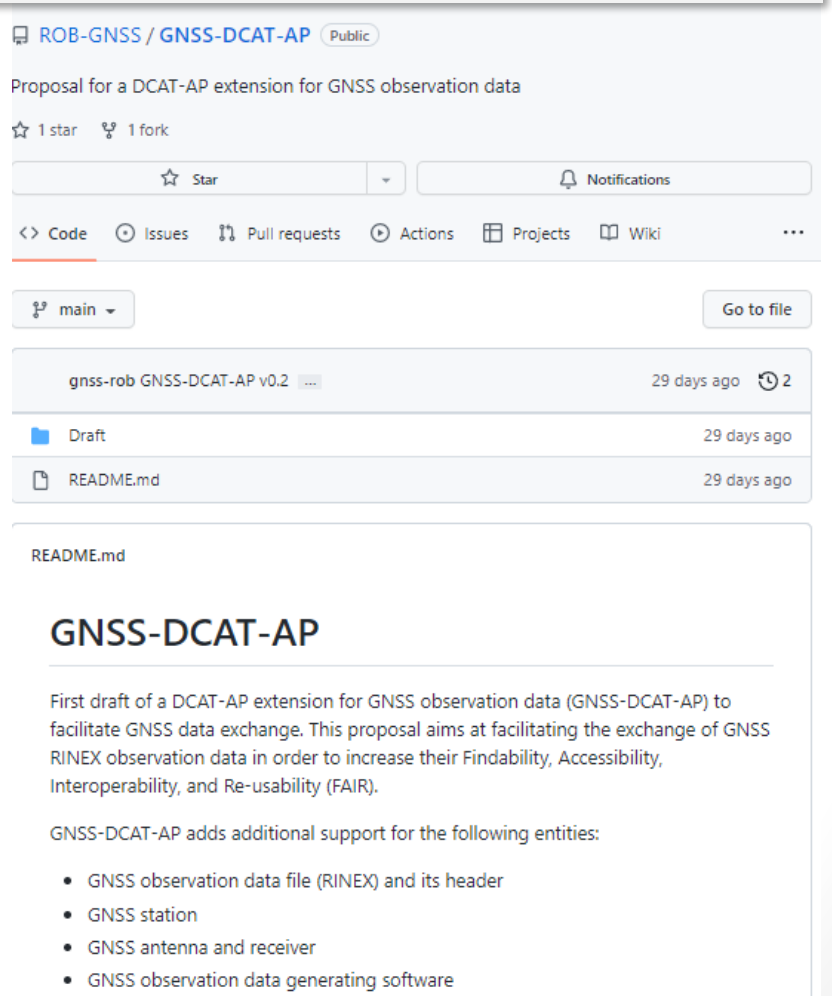
RINEX-dependent metadata: GNSS-DCAT-AP proposal on GitHub



Ongoing:
discussion and feedback from
FAIR-GNSS Follow-up Committee,
GFZ and ESA GNSS Science Support Centre

Documented and available on GitHub

Your feedback is very welcome!
https://github.com/ROB-GNSS/GNSS-DCAT-AP



Contents of the webinar

- **Part 1 - GNSS data and the data repository**
- **Part 2 - Introduction to FAIR data principles**
- **Part 3 - The journey so far: FAIR for GNSS data**
  - Where and how to start?
  - Gap analysis a.k.a. FAIR assessment
  - Metadata for GNSS data
  - Restructuring the data repository
  - Persistent identifiers for GNSS data
  - API preparatory work
- **Part 4 – Lessons learned and next steps**

FAIR GNSS data

Downloading daily RINEX data

TODAY

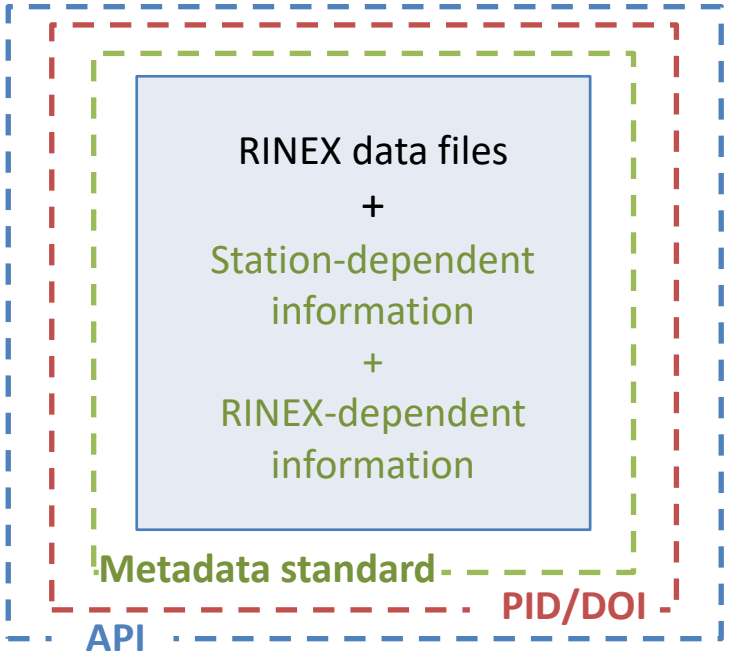
BRUX00BEL
02/2021-012/2021

RINEX data files

Downloading daily RINEX data

FUTURE

BRUX00BEL
02/2021-012/2021



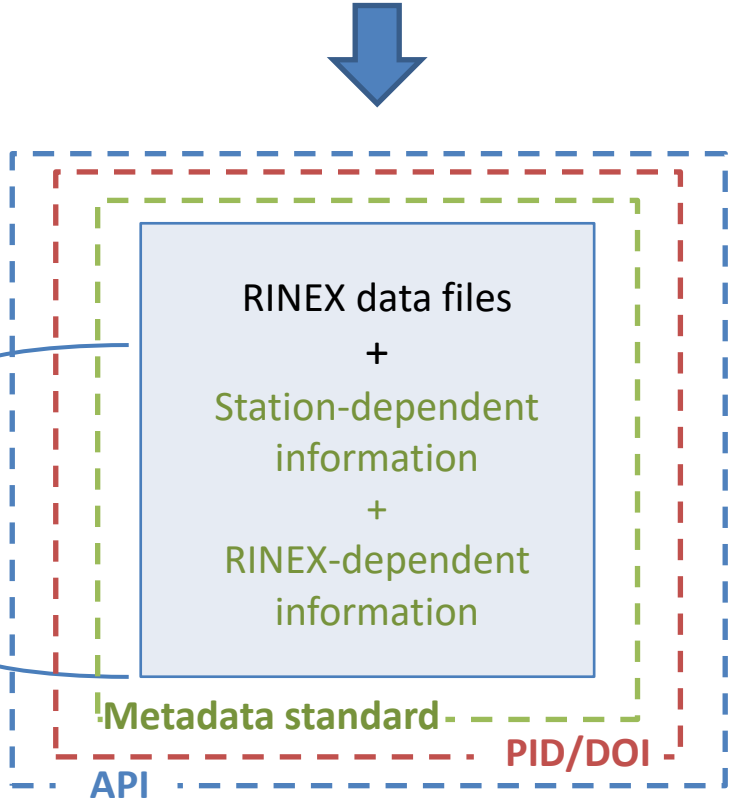
searchable by both humans and machines

FAIR GNSS data

Downloading daily RINEX data

FUTURE

BRUX00BEL
02/2021-012/2021



Our existing database does not contain all needed information

What content do we need to include in the new HDC database?

How to implement this for an already existing data repository?

Station-dependent metadata - GeodesyML

All information is collected by M^{3}G, stored in a database and exported in GeodesyML format (API + persistent link)

→ Persistent link to be provided together with the download of RINEX data of the BRUX00BEL station : https://gnss-metadata.eu/v1/sitelog/exportxml?id=BRUX00BEL

GeodesyML format

- ✓ Current version of M^{3}G: GeodesyML 0.4
- ✓ Next release: GeodesyML 0.5 (Nov. 2022?)
- ✓ Future release: GeodesyML *x.y* including all new functionalities discussed within GeodesyML taskforce of the International GNSS Service

Population of the M^{3}G database

- ✓ Station description: Mandatory for all GNSS stations (requirement)
- ✓ Data license (optional)
  - At start of FAIR-GNSS, few GNSS stations with data license
  - European Plate Observing System (EPOS): data license is mandatory (CC BY 4.0)
  - Today: 93% of GNSS stations in data repository inserted data license in M^{3}G
- ✓ Data citation information (optional)
  - Digital Object Identifier (DOI) → more later
  - Very few GNSS stations inserted DOI in M^{3}G!

RINEX file-dependent metadata

Data inventory learned us that:

Missing provenance information

- From where did the data files come (from which data center? Or provided directly by the station manager to us?)
- What changes have been made on the data (RINEX header)?
- Content of original data (RINEX header)

Error checking

- Inconsistent usage of flags
- More transparent rules enabling decision to publish a RINEX data file or not

Fields to be added to repository database

Decide what to download

For EUREF data centers:

- File name
- RINEX format
- File size
- Date of observations
- Creation date of file
- Last update time of file

Document origin of the data file

- Persistent link to original data file (Data center name + File name)
- File size at data center
- MD5 checksum at data center (*not available*)
- Last update time of file at data center
- Original RINEX header at data center

Document operations done on each data file

- RINEX header changed or not
- List of changes in RINEX header
- Date of change of RINEX header
- Station description used as reference for changing RINEX header: Station description from M^3G
- GDPR-related changes
- Final content RINEX header
- Warning/error flags

Document publicly made available data files to facilitate searching

- File name
- RINEX format
- File size
- Date of observations
- Creation date of file
- Last update time of file
- MD5 checksum
- File version (# updates in database)
- List of programs used to generate RINEX
- Content current (corrected) RINEX header

Associated (existing) database

- Station descriptions from M^3G
- Precise positions of each station
- GNSS data quality metrics

Fields to be added to repository database

Decide what to download

For EUREF data centers:

- File name
- RINEX format
- File size
- Date of observations
- Creation date of file
- Last update time of file

Document origin of the data file

- Persistent link to original data file (Data center name + File name)
- File size at data center
- MD5 checksum at data center (*not available*)
- Last update time of file at data center
- Original RINEX header at data center

Document operations done on each data file

- RINEX header changed or not
- List of changes in RINEX header
- Date of change of RINEX header
- Reference info used for changing RINEX header: Station description from M³G
- GDPR-related changes
- Final content RINEX header
- Warning/error flags

Document publicly made available data files to facilitate searching

- File name
- RINEX format
- File size
- Date of observations
- Creation date of file
- Last update time of file
- MD5 checksum
- File version (# updates in database)
- List of programs used to generate RINEX
- Content current (corrected) RINEX header content

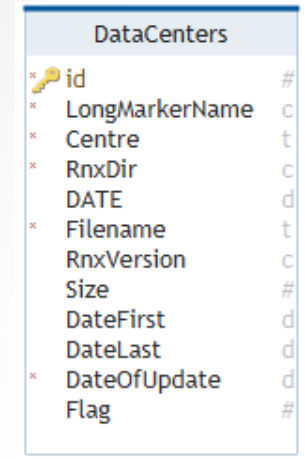
Associated (existing) database

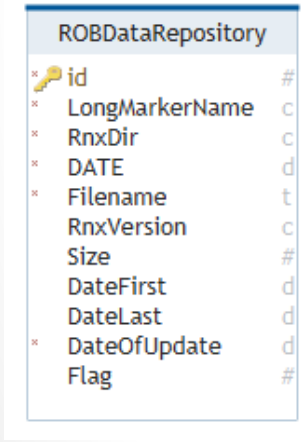
- Station descriptions from M³G
- Precise positions of each station
- GNSS data quality metrics

PROVENANCE
Persistent link to original data

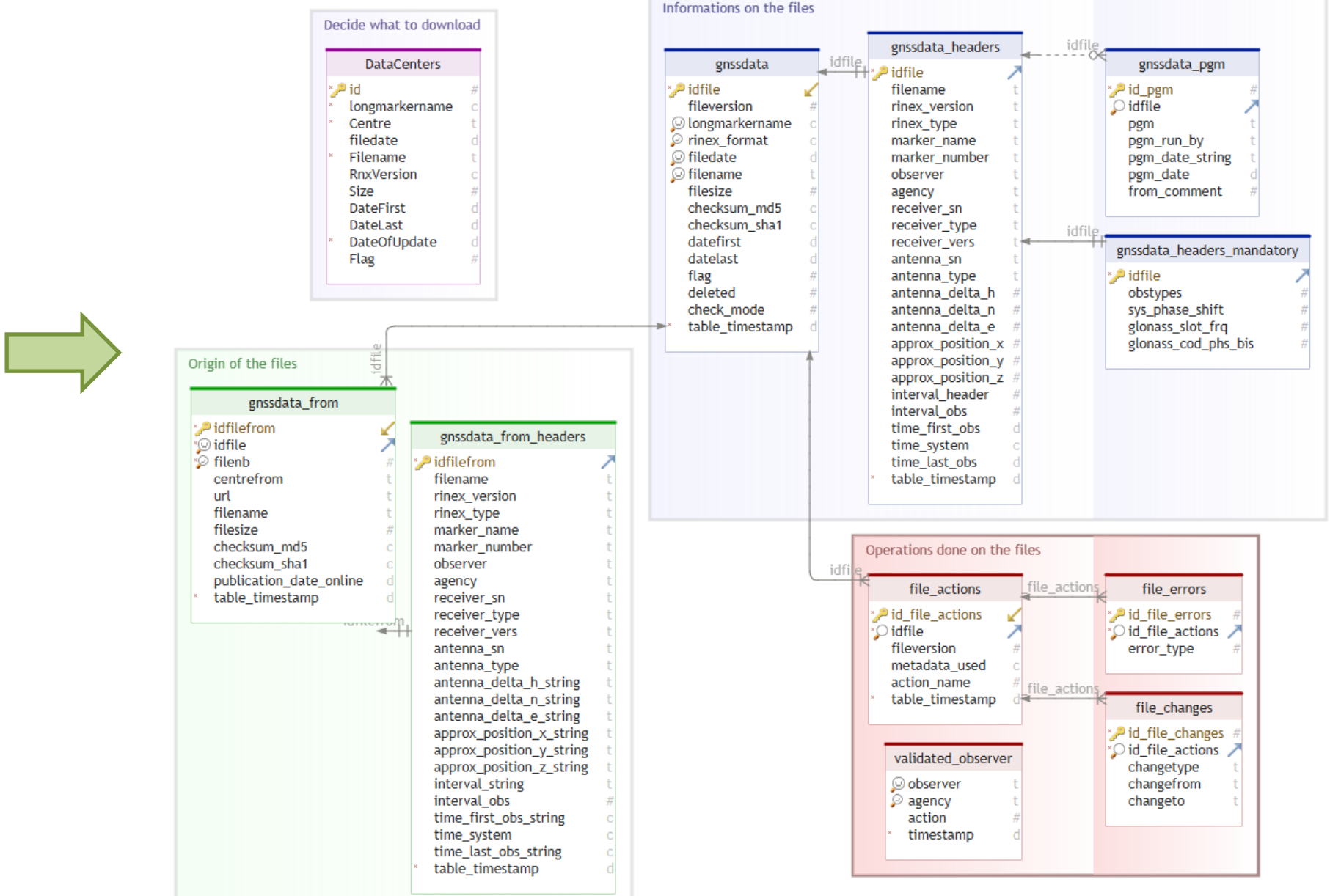
→ **Need for a completely new database**

OLD database





New database



ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

Implementation: Development of new backend

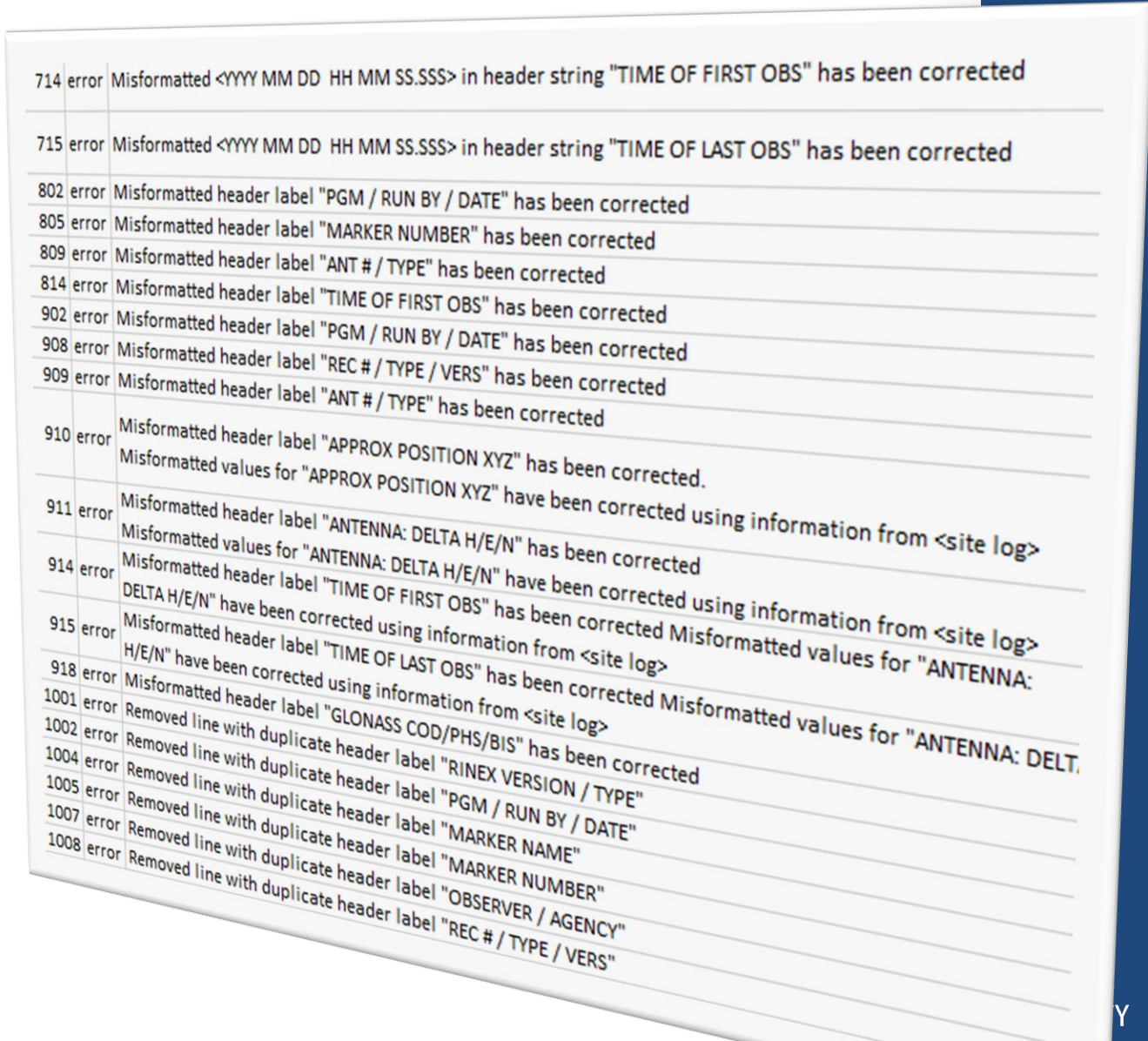
- Started in Summer of 2021
- Testing and development cycles
- Complete revision of software
  - To download RINEX data
  - To verify and correct RINEX headers
  - To verify and correct RINEX formatting errors and manage flags
  - To populate database
- Under internal testing on new incoming operational RINEX data from Nov. 2021

More rigorous checking of RINEX errors

ROYAL OBSERVATORY OF BELGIUM

Store results of RINEX verifications in internal database (and can be consulted later on)

- Incorrect/unknown RINEX format
- Missing mandatory header lines
- Misformatted header lines
- Duplicate header lines
- Inconsistent information with site log
- Sampling rate
- Date file name / date ‘FIRST OBSERVATION’ / timestamp of actual observations
- …



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

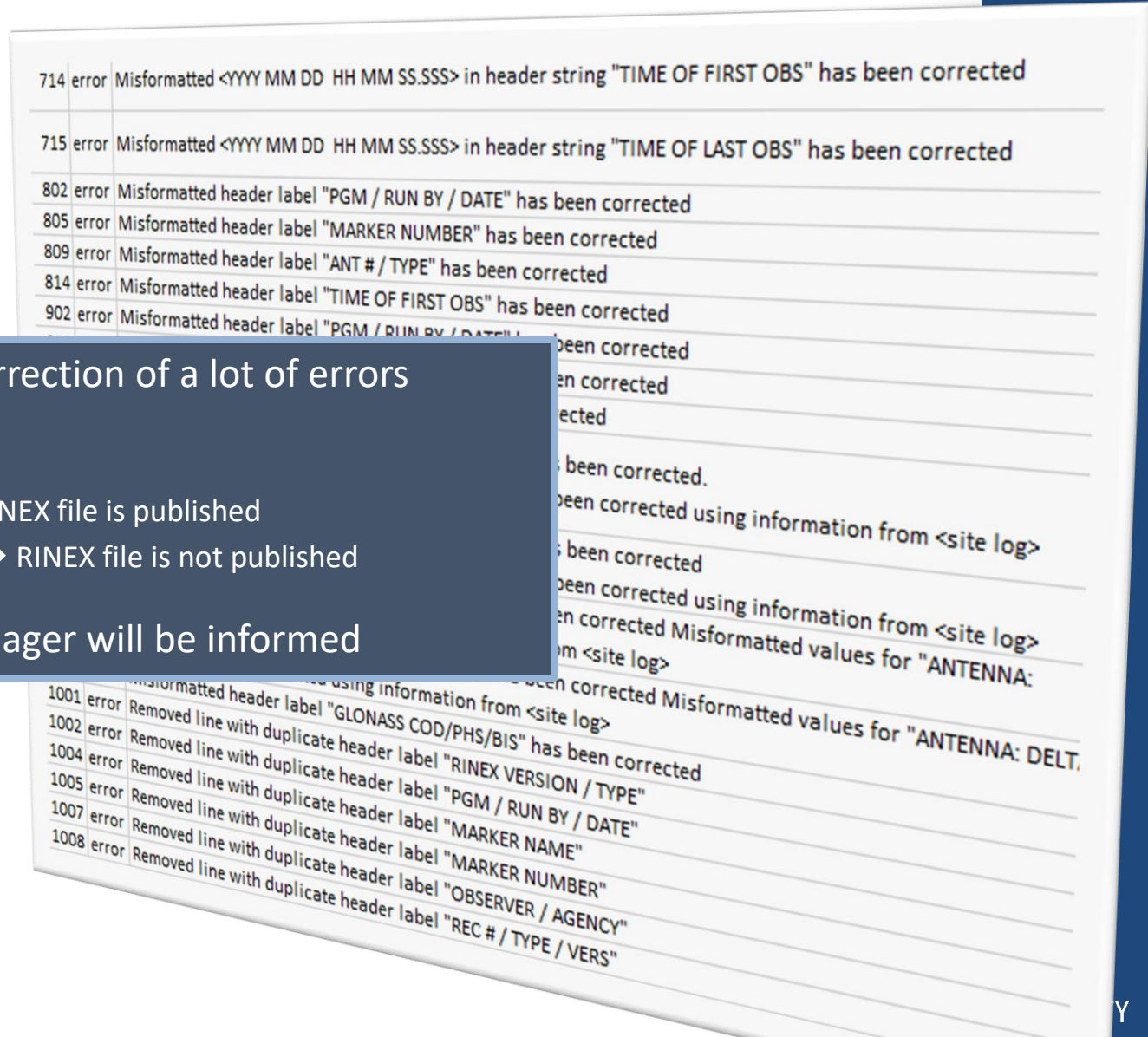
More rigorous checking of RINEX errors

ROYAL OBSERVATORY OF BELGIUM

Store results of RINEX verifications in internal database (and can be consulted later on)

- Incorrect/unknown RINEX format
- Missing mandatory header lines
- Misformatted header lines
- Duplicate header lines
- Inconsistent information
- Sampling rate
- Date file name / date 'FIRST OBS' / timestamp of actual observations
- …

- ✓ Improved automatic correction of a lot of errors
- ✓ Classification of errors
  - Error can be corrected → RINEX file is published
  - Error cannot be corrected → RINEX file is not published
- ✓ In all cases: station manager will be informed



Improved tracking of changes on RINEX files

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

First results

New provenance information directly allowed to detect and debug problems that were previously not seen

- Data files for same day and same station coming from one data center were flagged and coming from another from another data center were not flagged

- Workflows used at EUREF regional data centers unclear / different handling of GNSS data (publication or not)
- Meeting with data centers (04/2022) and a lot of email exchanges
- Data centers adapted their way of working

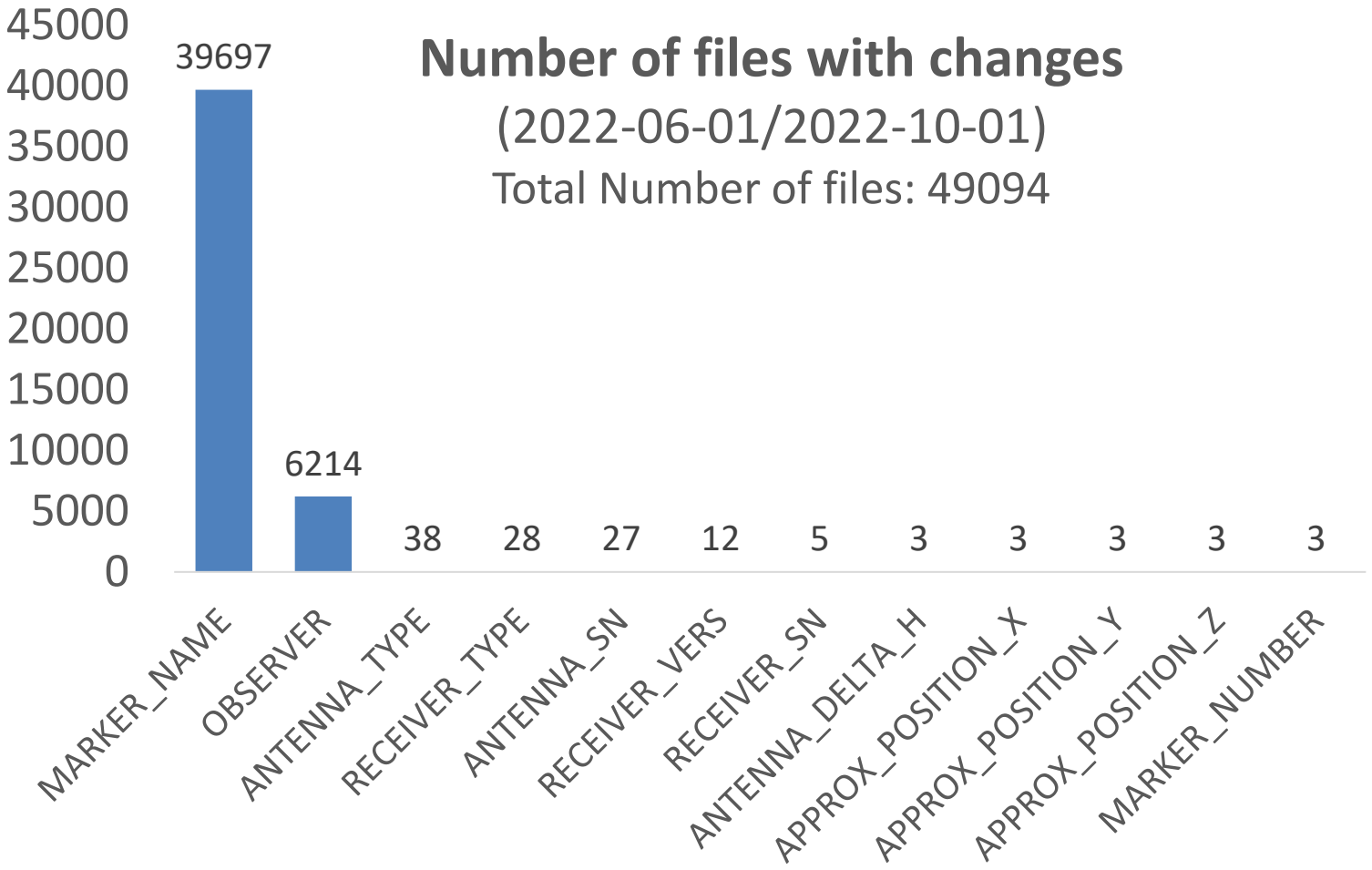
Restart filling database from scratch again for all data coming in from June 2022 on

For data prior to June 2022: populated new database with limited info from old database

Gathering provenance information from before June 2022 requires re-downloading all historical data from the original data centers!

Original plan : before Fall 2022

- But data centers undergoing major restructuring and persistent links to original data could not be constructed


Number of files with changes
(2022-06-01/2022-10-01)
Total Number of files: 49094
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
39697
6214
38
28
27
12
5
3
3
3
3
3
MARKER_NAME
OBSERVER
ANTENNA_TYPE
RECEIVER_TYPE
ANTENNA_SN
RECEIVER_VERS
RECEIVER_SN
ANTENNA_DELTA_H
APPROX_POSITION_X
APPROX_POSITION_Y
APPROX_POSITION_Z
MARKER_NUMBER

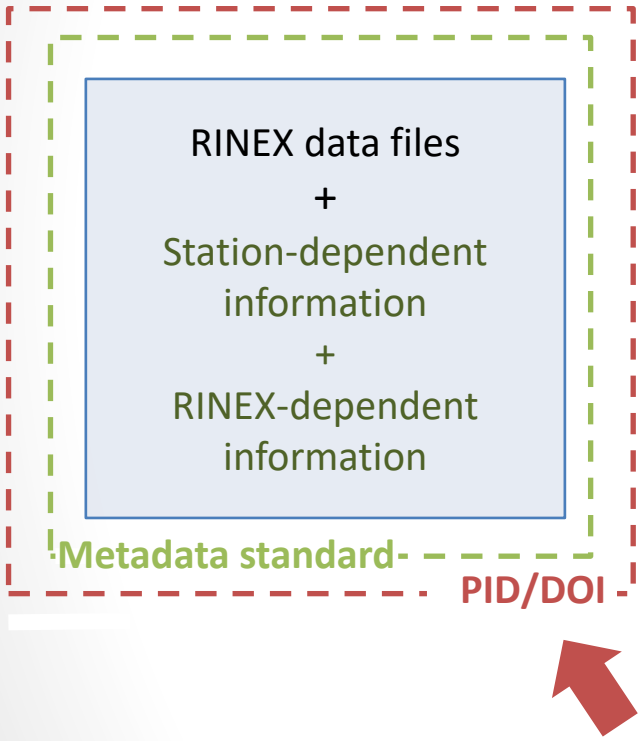
Conclusion

- Software development started in 2021
- Several development and testing cycles because new problems where discovered along the way
- Close collaboration with the Regional EUREF Data Centers was necessary
- Persistent links to original data at the Regional EUREF Data Centers are necessary to progress further
- Populating new repository database is a very time consuming process

Contents of the webinar

- **Part 1 - GNSS data and the data repository**
- **Part 2 - Introduction to FAIR data principles**
- **Part 3 - The journey so far: FAIR for GNSS data**
  - Where and how to start?
  - Gap analysis a.k.a. FAIR assessment
  - Metadata for GNSS data
  - Restructuring the data repository
  - Persistent identifiers for GNSS data
  - API preparatory work
- **Part 4 – Lessons learned and next steps**

Persistent Identifiers (PID) for GNSS data

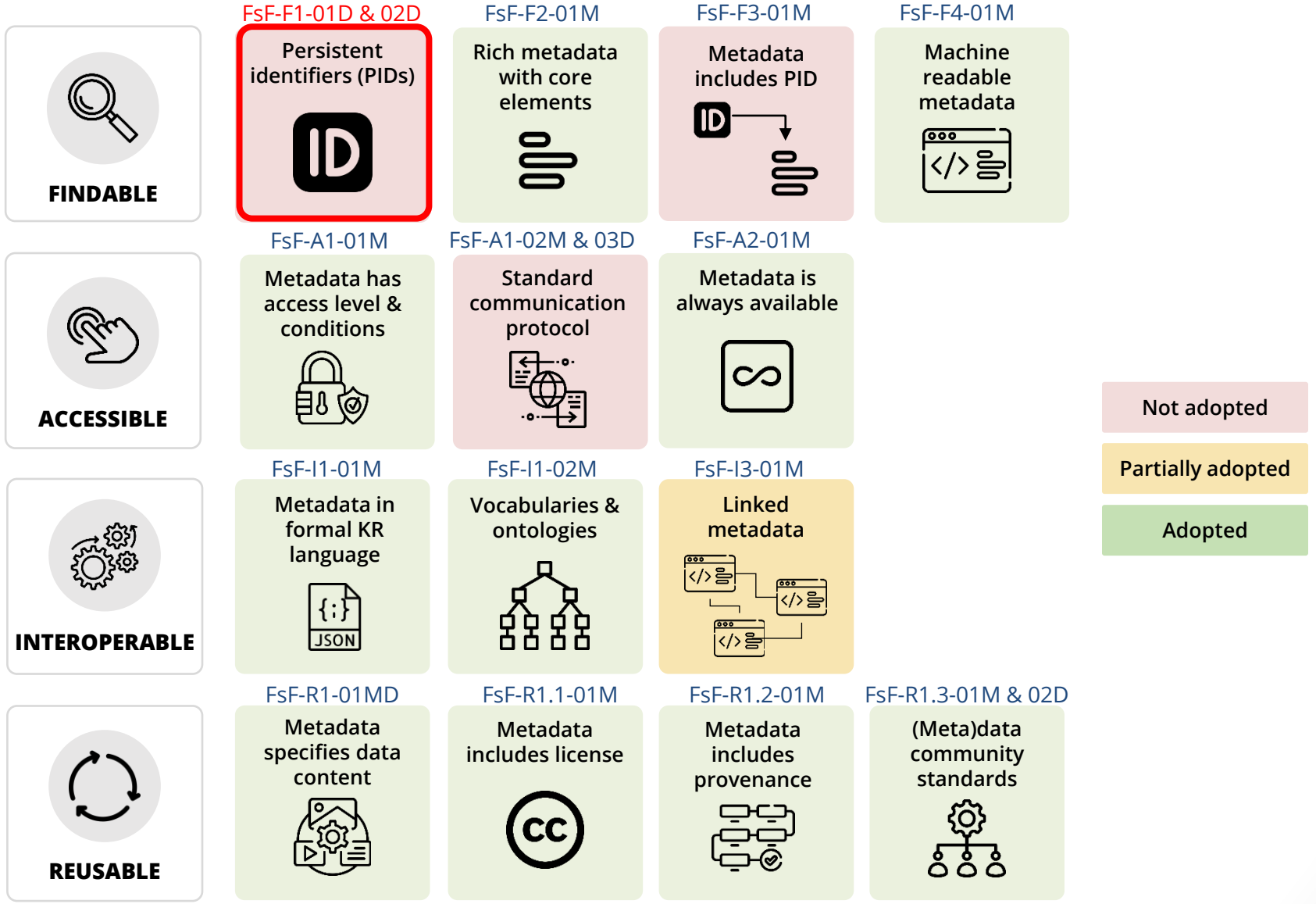


Introduction to FAIR data principles:

- A PID is a globally unique and persistent identifier pointing to a digital object
- Digital object = GNSS data
- Attributing a PID to the data is requested by the FAIR data principles with as main goal to make the data findable:
  - A http link whose address remains unchanged over time and which allows to find information on the data

DOI for GNSS data:

FAIRsFAIR metrics to monitor progress toward FAIR



ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

Digital Object Identifiers (DOI)

Digital Object Identifiers (DOI) are the obvious PID to be used for data sets (also used in seismology)

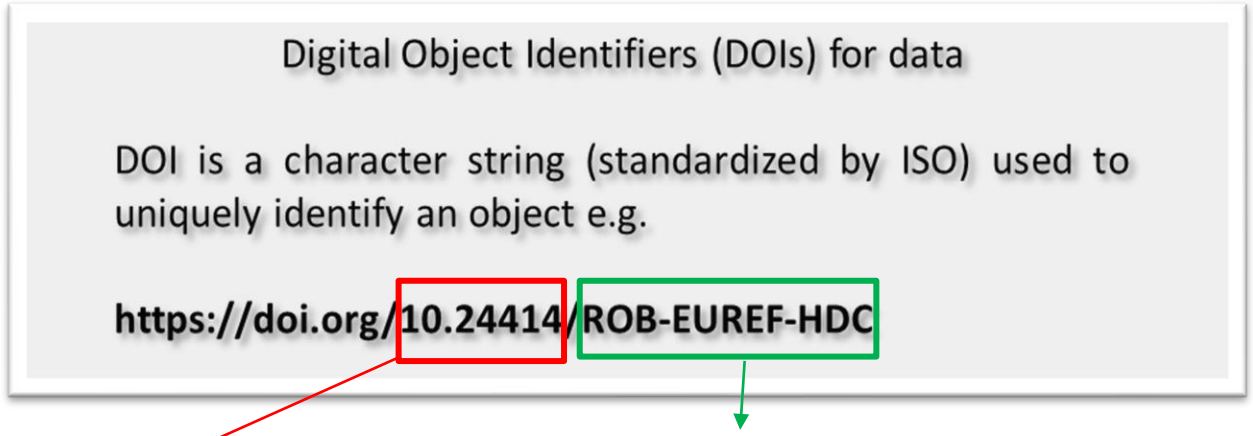
Digital Object Identifiers (DOIs) for data

DOI is a character string (standardized by ISO) used to uniquely identify an object e.g.

https://doi.org/10.24414/ROB-EUREF-HDC

Digital Object Identifiers (DOI)

Digital Object Identifiers (DOI) are the obvious PID to be used for data sets (also used in seismology)



How to get a DOI for data sets?

Allocation Agents, who are members of DataCite, assign DOIs to objects.

Clients, like universities, may sign a contract with an allocation agency to become **minting agencies** able to “register” or “mint”, DOIs.

E.g. ROB (= minting agency) mints DOIs thanks to a contract with an allocation agent (TU Delft).

Digital Object Identifiers (DOI)

Digital Object Identifiers (DOI) are the obvious PID to be used for data sets (also used in seismology)

Digital Object Identifiers (DOIs) for data

DOI is a character string (standardized by ISO) used to uniquely identify an object e.g.

https://doi.org/10.24414/ROB-EUREF-HDC

- a globally unique and long-lasting reference to a resource with a stable link to the resource (like research data)
- enable reliable long-term citation of the data
  - data providers receive credit through citation metrics
  - provide statistics on data usage for funders
- provide user information on data access restrictions
- enable cross-linking through various objects:
  - publications and underlying data
  - Link datasets with other datasets (hierarchical relations)
  - PID graphs

What is required to get a DOI?

The following are required to create a DOI:

- A DOI name (prefix/suffix)
- **A metadata file (an XML file) –** DataCite metadata scheme
- A stable long-term URL (also called the '**landing page**' that provides access to the objects)
    - The DOI will point to a "landing page" URL (using the information in the XML)
    - Landing page is maintained by the minting agency
    - All landing pages should include a minimal set of information including: links to access data (if available) or instructions how to access the data, a sample citation text
    - DOIs are typically "linked" via a global resolver (e.g. http://dx.doi.org/10.24414/datacenter.123xy) which redirects to this page.

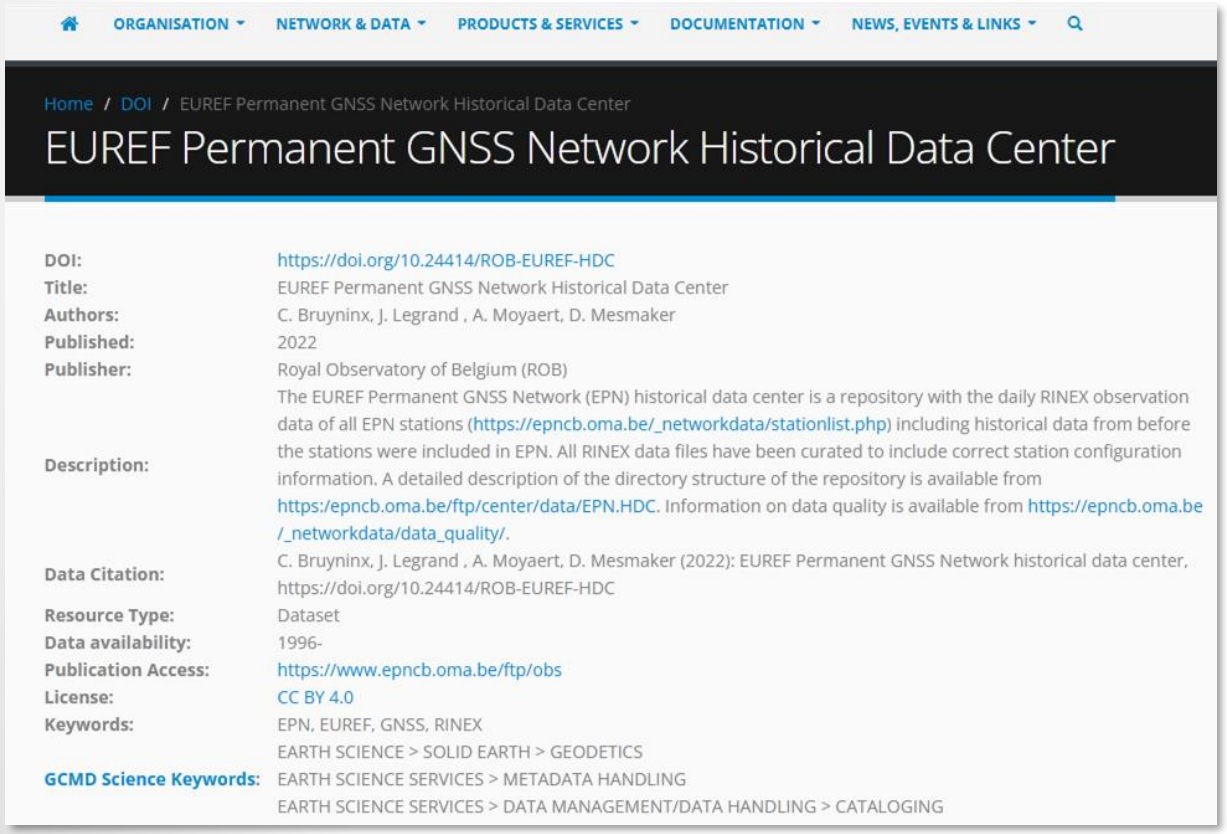
It is **not possible to delete a DOI** when it has been created.

URL and metadata can be updated at any time

DOI example

https://doi.org/10.24414/ROB-EUREF-HDC

DOI landing page



ORGANISATION NETWORK & DATA PRODUCTS & SERVICES DOCUMENTATION NEWS, EVENTS & LINKS

Home / DOI / EUREF Permanent GNSS Network Historical Data Center

EUREF Permanent GNSS Network Historical Data Center

DOI:	https://doi.org/10.24414/ROB-EUREF-HDC
Title:	EUREF Permanent GNSS Network Historical Data Center
Authors:	C. Bruyninx, J. Legrand , A. Moyaert, D. Mesmaker
Published:	2022
Publisher:	Royal Observatory of Belgium (ROB)
Description:	The EUREF Permanent GNSS Network (EPN) historical data center is a repository with the daily RINEX observation data of all EPN stations (https://epncb.oma.be/_networkdata/stationlist.php) including historical data from before the stations were included in EPN. All RINEX data files have been curated to include correct station configuration information. A detailed description of the directory structure of the repository is available from https:/epncb.oma.be/ftp/center/data/EPN.HDC. Information on data quality is available from https://epncb.oma.be/_networkdata/data_quality/.
Data Citation:	C. Bruyninx, J. Legrand , A. Moyaert, D. Mesmaker (2022): EUREF Permanent GNSS Network historical data center, https://doi.org/10.24414/ROB-EUREF-HDC
Resource Type:	Dataset
Data availability:	1996-
Publication Access:	https://www.epncb.oma.be/ftp/obs
License:	CC BY 4.0
Keywords:	EPN, EUREF, GNSS, RINEX
GCMD Science Keywords:	EARTH SCIENCE > SOLID EARTH > GEODETICS EARTH SCIENCE SERVICES > METADATA HANDLING EARTH SCIENCE SERVICES > DATA MANAGEMENT/DATA HANDLING > CATALOGING

XML file – DataCite schema

```
<resource xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://datacite.org/schema/kernel-4" xsi:schem
https://schema.datacite.org/meta/kernel-4/metadata.xsd">
  <identifier identifierType="DOI">10.24414/ROB-EUREF-HDC</identifier>
  <creators>
    <creator>
      <creatorName nameType="Personal">Bruyninx, Carine</creatorName>
      <givenName>Carine</givenName>
      <familyName>Bruyninx</familyName>
      <nameIdentifier nameIdentifierScheme="ORCID" schemeURI="https://orcid.org">https://orcid.org/0000-0001-6492-394
      <affiliation>Royal Observatory of Belgium</affiliation>
    </creator>
    <creator>
      <creatorName nameType="Personal">Legrand, Juliette</creatorName>
      <givenName>Juliette</givenName>
      <familyName>Legrand</familyName>
      <nameIdentifier nameIdentifierScheme="ORCID" schemeURI="https://orcid.org">https://orcid.org/0000-0002-9502-904
      <affiliation>Royal Observatory of Belgium</affiliation>
    </creator>
    <creator>
      <creatorName nameType="Personal">Moyaert, Ann</creatorName>
      <givenName>Ann</givenName>
      <familyName>Moyaert</familyName>
      <affiliation>Royal Observatory of Belgium</affiliation>
    </creator>
    <creator>
      <creatorName nameType="Personal">Mesmaker, Dominique</creatorName>
      <givenName>Dominique</givenName>
      <familyName>Mesmaker</familyName>
      <affiliation>Royal Observatory of Belgium</affiliation>
    </creator>
  </creators>
  <titles>
    <title xml:lang="en">EUREF Permanent GNSS Network Historical Data Center</title>
  </titles>
  <publisher xml:lang="en">Royal Observatory of Belgium</publisher>
  <publicationYear>2022</publicationYear>
  <subjects>
    <subject>EPN</subject>
    <subject>EUREF</subject>
    <subject>GNSS</subject>
    <subject>RINEX</subject>
    <subject xml:lang="en" schemeURI="https://gcmd.earthdata.nasa.gov/kms/concepts/concept_scheme/sciencekeywords/" s
    SCIENCE &gt; SOLID EARTH &gt; GEODETICS</subject>
    <subject xml:lang="en" schemeURI="https://gcmd.earthdata.nasa.gov/kms/concepts/concept_scheme/sciencekeywords/" s
    SCIENCE SERVICES &gt; METADATA HANDLING</subject>
    <subject xml:lang="en" schemeURI="https://gcmd.earthdata.nasa.gov/kms/concepts/concept_scheme/sciencekeywords/" s
    SCIENCE SERVICES &gt; DATA MANAGEMENT/DATA HANDLING &gt; CATALOGING</subject>
    <subject xml:lang="en" schemeURI="https://gcmd.earthdata.nasa.gov/kms/concepts/concept_scheme/sciencekeywords/" s
    SCIENCE SERVICES &gt; DATA MANAGEMENT/DATA HANDLING &gt; DATA ACCESS/RETRIEVAL</subject>
  </subjects>
  <resourceType resourceTypeGeneral="Dataset">Dataset</resourceType>
  <dates>
    <date dateType="Available">1996/</date>
  </dates>
  <rightsList>
    <rights xml:lang="en" schemeURI="https://spdx.org/licenses/" rightsIdentifierScheme="SPDX" rightsIdentifier="CC-B
    rightsURI="https://creativecommons.org/licenses/by/4.0/"/>
  </rightsList>
  <descriptions>
```

ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

DOI and the corresponding XML file and schema

Mandatory citation fields considering a typical citation format
`Creator (Year): Title. Publisher. ResourceType. DOIName`
http://schema.datacite.org/meta/kernel-4.4/

Creator:
The principal investigator(s), or the organization operating the network or producing the data or products. This is the entity who should receive "primary" credit for producing the data set.

Title:
A name or title (5-10 words are recommended) by which the resource is known. For example "EUREF Permanent GNSS Network Historical data Centre".

Publisher:

The institution (or data centre) responsible for making the data, i.e. DOI and landing page, permanently available. As a general rule, this will usually be the organization that mints the DOI.

Publication Year:
Year when the data is made publicly available. In the case of datasets, "publish" is understood to mean making the data available on a specific date to the community of researchers.

Use of DOI for datasets within the scientific community

- Seismological community: FDSN recommendations and DOI naming
- UNAVCO (…already minting DOIs for GNSS datasets), GFZ, CDDIS, ROB, …
- EPOS EUROPEANPLATEOBSERVINGSYSTEM foresees use of DataCite for DOI
- GGOS Global Geodetic Observing System GGOS working group on DOI: a common approach to assign DOI to geodetic data and products

Use of DOI for datasets within the scientific community

- Seismological community: FDSN recommendations and DOI naming
- UNAVCO (…already minting DOIs for GNSS datasets), GFZ, CDDIS, ROB, …
- EPOS EUROPEANPLATEOBSERVINGSYSTEM foresees use of DataCite for DOI
- GGOS Global Geodetic Observing System GGOS working group on DOI: a common approach to assign DOI to geodetic data and products

GGOS working Group: Digital Object Identifiers (DOIs) for Geodetic Data Sets

Created in Dec. 2019

ROB members joined the working group with the goal to

- Reach a community-agreement on the open questions
- Raise awareness of the geodetic community of the importance of assigning DOIs to GNSS datasets

One year later, FAIR-GNSS project started (Dec. 2020)

Level of granularity of DOI for GNSS?

Large amount of GNSS stations

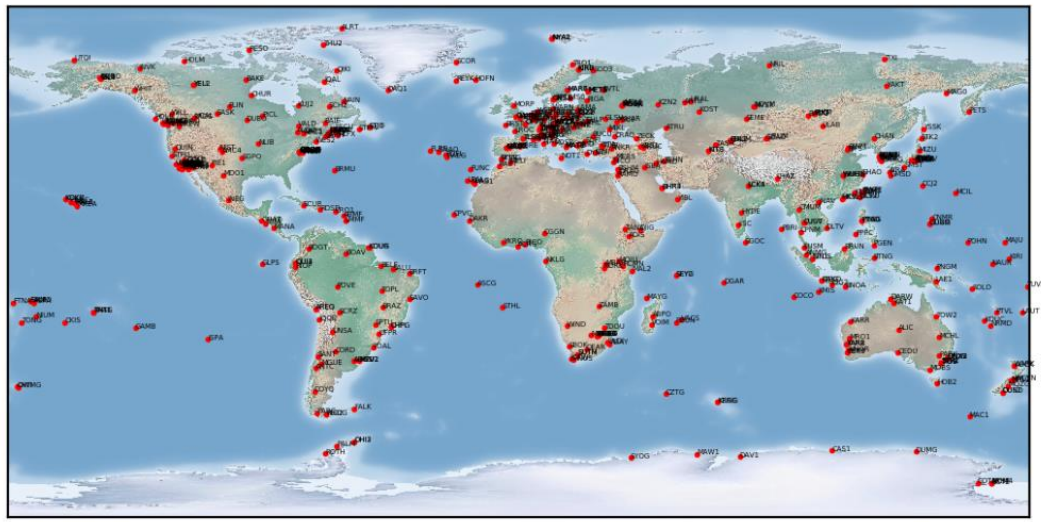
Organized in global, regional, local networks

Stations can contribute to different networks

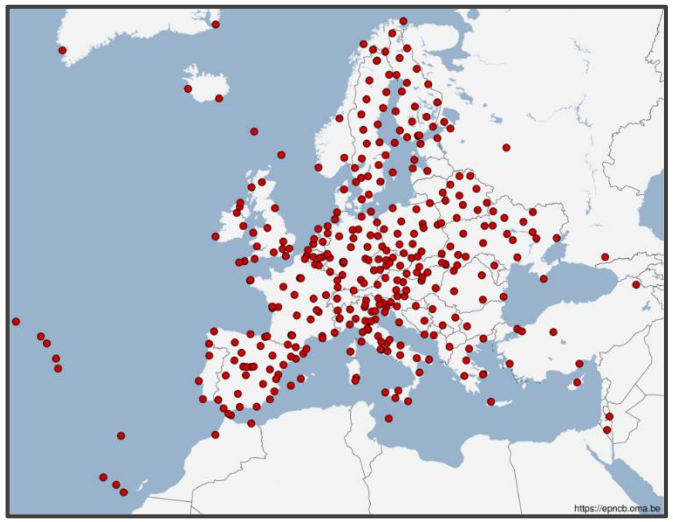
Assign one DOI per GNSS network?

Assign one DOI per GNSS station?

IGS network

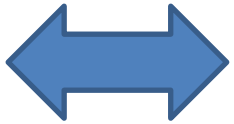


EUREF network



RENAG network

Agreement on level of DOI granularity



- GNSS-station DOI: Need to assign a **DOI** for the ongoing **GNSS data** measured for each **GNSS station**
- In addition, the station DOI may co-exist with GNSS network DOI
- The DOI of GNSS networks can be linked to the DOI of GNSS stations (`relatedIdentifier, IsPartOf, HasPart`)

DOI standardization for GNSS data?

Like all metadata schemas: DataCite metadata schema has mandatory and optional fields

To make the information provided by the DOI readable by machines, the GNSS community needs to agree on:

- Which optional fields should be mandatory in our community?
- The standard vocabularies to be used to populate the fields?
- From where do we get the information to fill in the different fields?

Data citation:

Creator (PublicationYear): Title. Publisher. (resourceTypeGeneral). Identifier

Mandatory

- Identifier
- Creator
- Title
- PublicationYear
- Publisher
- resourceType

Optional

- FundingReference
- Rights
- Format
- rightsURI
- rightsIdentifier

Information on funding, data format, data license, …

Optional

- relatedIdentifier
- relationType
  - IsPartOf
  - IsIdenticalTo
  - IsSourceOf
  - HasPart …

Allows to build relations between DOIs

ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

Standardization of DOI metadata

Exploit values from IGS site log/(extended) GeodesyML to semi-automatically populate DataCite schema

GeodesyML

```
1.Site Identification of the GNSS Monument
Site Name/
<geo:siteIdentification>
<geo:siteName>
```

```
12. Responsible Agency (if different from 11)
Agency/
<geo:siteOwner gml:id="siteOwner">..
<gmd:organisationName>
```

```
<geo:license codeSpace=
"urn:gnss-metadata.eu:gnss:license" codeList=
"https://gnss-
metadata.eu/GeodesyML_ext/codelists/
license-codelists.xml" codeListValue="CC-BY-
4.0"><![CDATA[CC-BY-4.0]]></geo:license>
```

resourceType (M)

title (M)

creator (M)

rights (O)

DataCite metadata schema v.4.4

```
Dataset
```

```
<title xml:lang="en">ELIS00ATA –
GNSS station at  Princess Elisabeth
Station Antarctica </title>

<creator>
<creatorName nameType="Organizational">
Royal Observatory of Belgium
</creatorName>
<nameIdentifier schemeURI="https://ror.org/"
nameIdentifierScheme="ROR">
https://ror.org/00hjks330</nameIdentifier>
</creator>

<rights xml:lang="en"
schemeURI="https://spdx.org/licenses/"
rightsIdentifierScheme="SPDX"
rightsIdentifier="CC-BY-4.0"
rightsURI="https://creativecommons.org/licen
ses/by/4.0/"/>
```

Standardization of DOI metadata

Exploit values from IGS site log/(extended) GeodesyML to semi-automatically populate DataCite schema

GeodesyML

```
1.Site Identification of the GNSS Monument
Site Name/
<geo:siteIdent
<geo:siteName>
```

```
12. Responsibl
Agency/
<geo:siteOwner
<gmd:organisationName>
```

```
<geo:license codeSpace=
"urn:gnss-metadata.eu:gnss:license" codeList=
"https://gnss-
metadata.eu/GeodesyML_ext/codelists/
license-codelists.xml" codeListValue="CC-BY-
4.0"><![CDATA[CC-BY-4.0]]></geo:license>
```

resourceType (M)

rights (O)

Further standardization of vocabularies under discussion within the GGOS and IGS

GGOS Global Geodetic Observing System

IGS INTERNATIONAL GNSS SERVICE

DataCite metadata schema v.4.4

```
Dataset
```

```
<title xml:lang="en">ELIS00ATA –
```

```
nameIdentifierScheme="ROR">
https://ror.org/00hjks330</nameIdentifier>
</creator>

<rights xml:lang="en"
schemeURI="https://spdx.org/licenses/"
rightsIdentifierScheme="SPDX"
rightsIdentifier="CC-BY-4.0"
rightsURI="https://creativecommons.org/licen
ses/by/4.0/"/>
```

ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

Standardization of DOI metadata

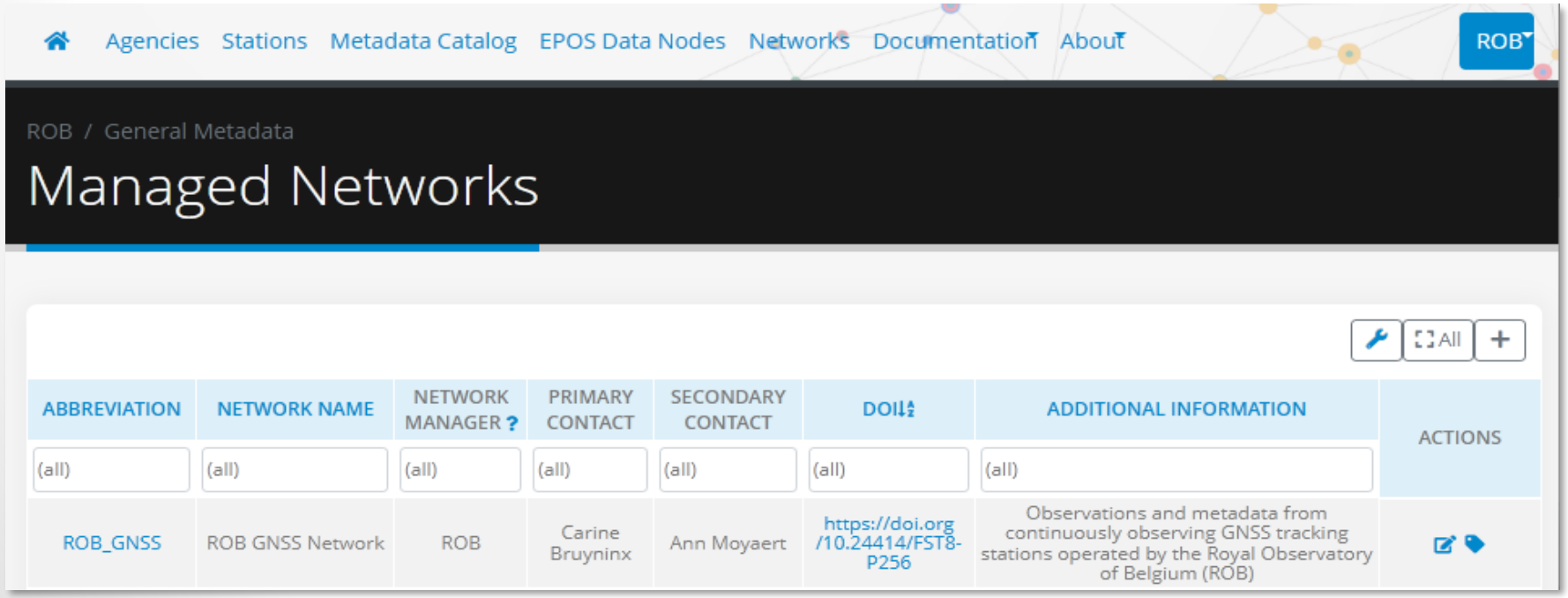
- FAIR-GNSS has elaborated concrete proposal, currently under discussion with GGOS DOI WG
- Harmonization within the community as a goal
- FAIR principles as key guidelines
- Retrieve as much as possible metadata from GeodesyML
- Include PIDs, like ROR, ORCID, DOI in DataCite metadata and in GeodesyML and define relation types
- Develop recommendations of content for specific DataCite fields that can be also used beyond GNSS data (e.g. repository = `Publisher`, agency = `Creator`, local partners = `Contributors`)
- Issue recommendations, similar to the FDSN recommendations for seismic network DOIs (https://doi.org/10.7914/D11596)?

Consequences for GNSS data repository

Today: Station managers can only insert DOIs for GNSS networks in M³G (6 so far..)

Upcoming version of M³G

- ✓ Station managers can insert DOI of the dataset of a GNSS station in M³G
- ✓ M³G contains information on which station belongs to which GNSS network and can potentially set up in the future the link between the GNSS network DOIs and the station DOIs



Consequences for GNSS data repository

Today: Station managers can only insert DOIs for GNSS network in M³G (6 so far …)

Upcoming version of M³G

- ✓ Station managers can insert DOI of the dataset of a GNSS station in M³G
- ✓ M³G contains information on which station belongs to which GNSS network and can potentially set up in the future the link between the GNSS network DOIs and the station DOIs

Raise awareness within GNSS community and station managers to

- ✓ Get a DOI for the dataset of their GNSS station
- ✓ Insert the DOI in M³G

Symposia
Splinter meetings
Webinar

Consequences for GNSS data repository

Today: Station managers can only insert DOIs for GNSS network in M³G (6 so far…)

Upcoming version of M³G

- ✓ Station managers can insert DOI of the dataset of a GNSS station in M³G
- ✓ M³G contains information on which station belongs to which GNSS network and can potentially set up in the future the link between the GNSS network DOIs and the station DOIs

Raise awareness within GNSS community and station managers to

- ✓ Get a DOI for the dataset of their GNSS station
- ✓ Insert the DOI in M³G

Symposia
Splinter meetings
Webinar

For GNSS station managers unable to mint DOIs:

ROB will offer to mint DOIs for GNSS data in our repository:

- ✓ Proposed DOI metadata scheme based on (extended) GeodesyML
- ✓ DOI metadata validated by the station managers

Users and DOIs for GNSS datasets

What about agencies using 100-1000 GNSS stations in their data analysis and wishing to cite the data?

- ✓ Does a station have a DOI ?
  - ✓ If the DOI metadata are standardized enough, then possible to develop DOI services that enable data users to inquire if a GNSS data set is associated with a DOI

- ✓ Impossible to cite 100-1000 GNSS station DOIs in a paper !
  - ✓ Try to develop a service to provide users of the repository on-demand a DOI that regroupes the specific set of GNSS stations they are using in a publication.
  - ✓ This DOI will link back to the individual station DOI (if existing) through the DOI metadata

- ✓ GNSS network DOI will still be available

Summary

Attributing DOIs to GNSS datasets is important because it allows to

- Cite GNSS data (crucial for crediting researchers and funders)
- Include information on data license

However, only few GNSS data with DOI

- No good practice of citing GNSS data using DOI in papers

GNSS data are complex (GNSS stations, local, regional, global networks)

- DataCite metadata scheme allows to deal with complexity of GNSS data

Similar to what was done within the seismological community, **GNSS community must agree** on

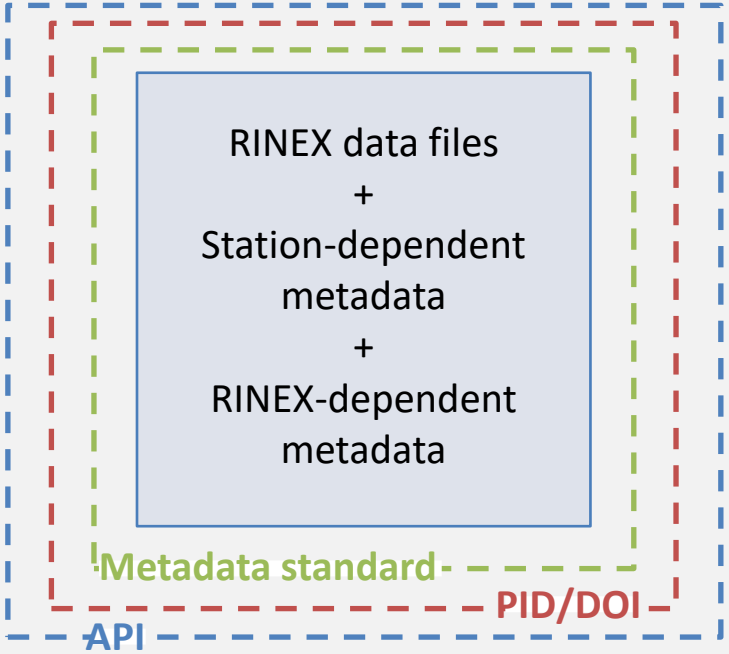
- Common standards for completing the DOI metadata (DOI/station + proposal for populating DOI metadata)

Discussion started in GGOS Working Group on DOI

Long-term goal ***Community effort***

Coming up next:

Preliminary work to implement API for GNSS data

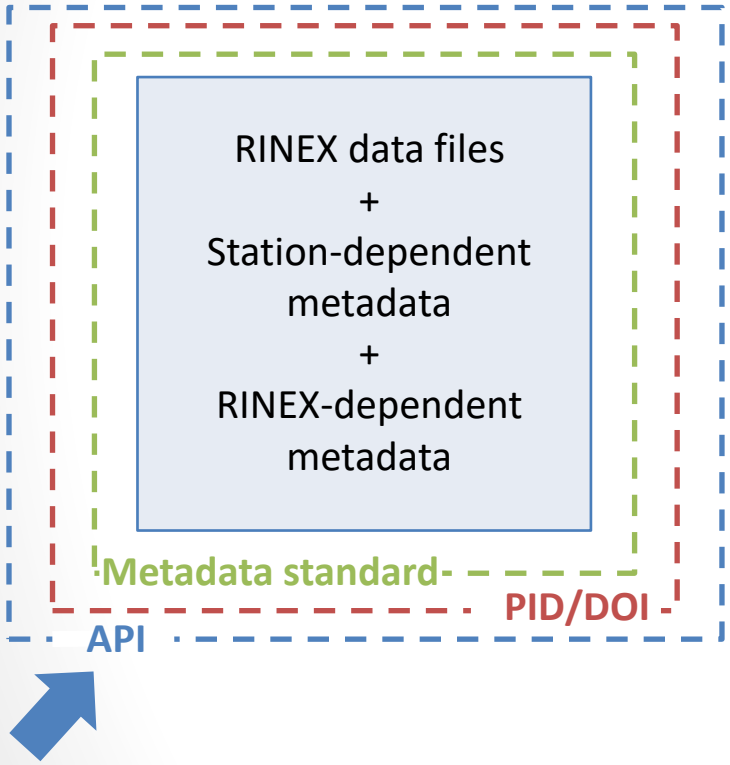


Contents of the webinar

- **Part 1 - GNSS data and the data repository**
- **Part 2 - Introduction to FAIR data principles**
- **Part 3 - The journey so far: FAIR for GNSS data**
  - Where and how to start?
  - Gap analysis a.k.a. FAIR assessment
  - Metadata for GNSS data
  - Restructuring the data repository
  - Persistent identifiers for GNSS data
  - API preparatory work
- **Part 4 – Lessons learned and next steps**

API preparatory work

FAIR-GNSS methodology



…start by answering some questions

From FAIR guidelines…
to a practical approach for GNSS data

Downloading daily RINEX data
TODAY

BRUX00BEL
02/2021 - 12/2021

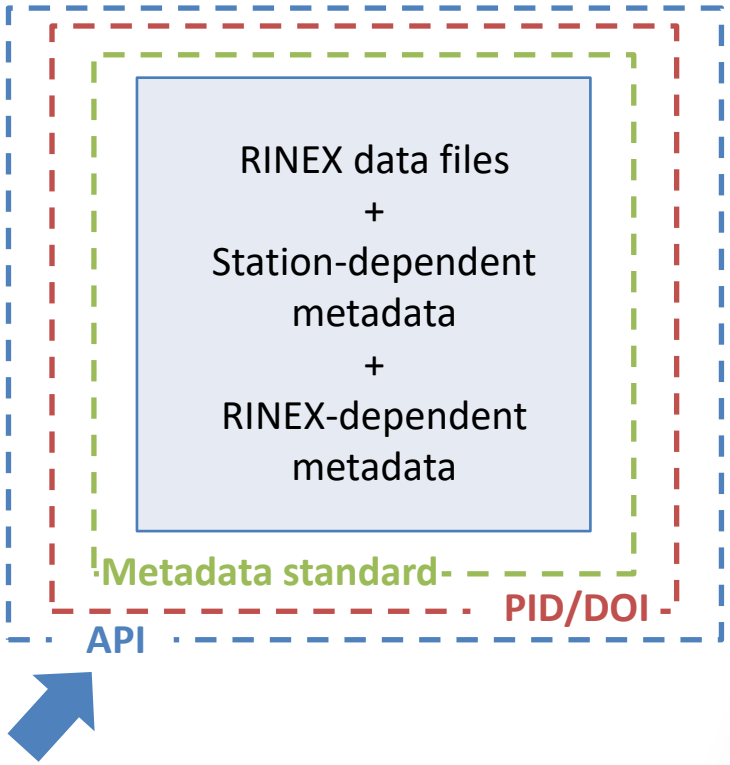
RINEX data files

Downloading daily RINEX data
FUTURE

BRUX00BEL
02/2021 - 12/2021

RINEX data files
+
Station-dependent information
+
RINEX-dependent information

+ Search capabilities!



Focus on access to GNSS data & metadata

Access to (meta)data and FAIR

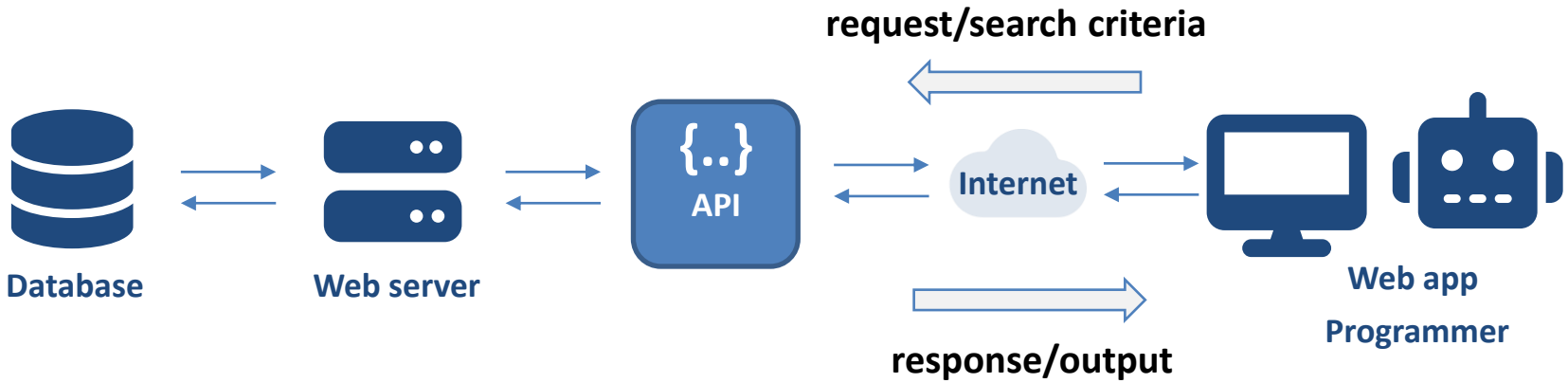
FAIR principles to enhance the **ability of machines to automatically find and access/use the (meta)data** aka **machine-readability and actionability**

- Standard communication protocol (e.g. https, ftp) to transmit the data
- Standard metadata formats & API to enable machines to find, access and interpret (meta)data
- Authentication and authorization + access level metadata

Focus first on

APIs (Application Programming Interface): a set of functions and procedures to allow applications to access repository data

API to access to GNSS (meta)data



search and download data via API

- allow to search criteria based on user needs
- provide all necessary metadata about the data/station ← **rich metadata**
- use of standards to ensure machine-readability and interoperability with other data ← **GNSS-DCAT-AP**

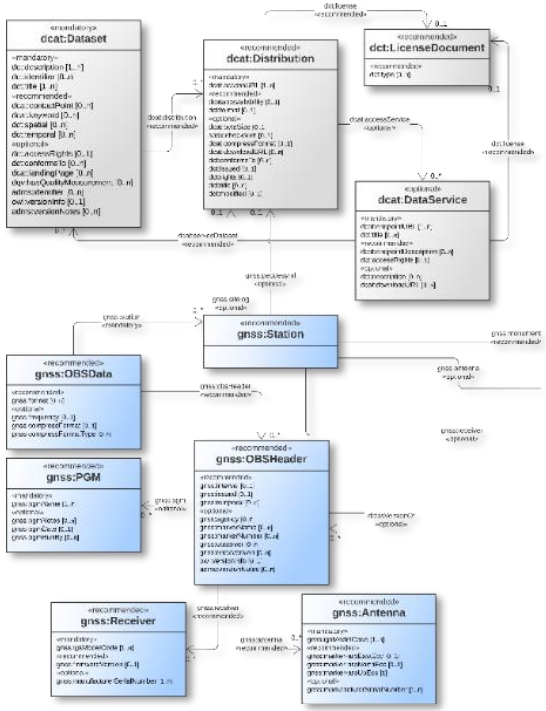
Metadata to enable machines to access GNSS data

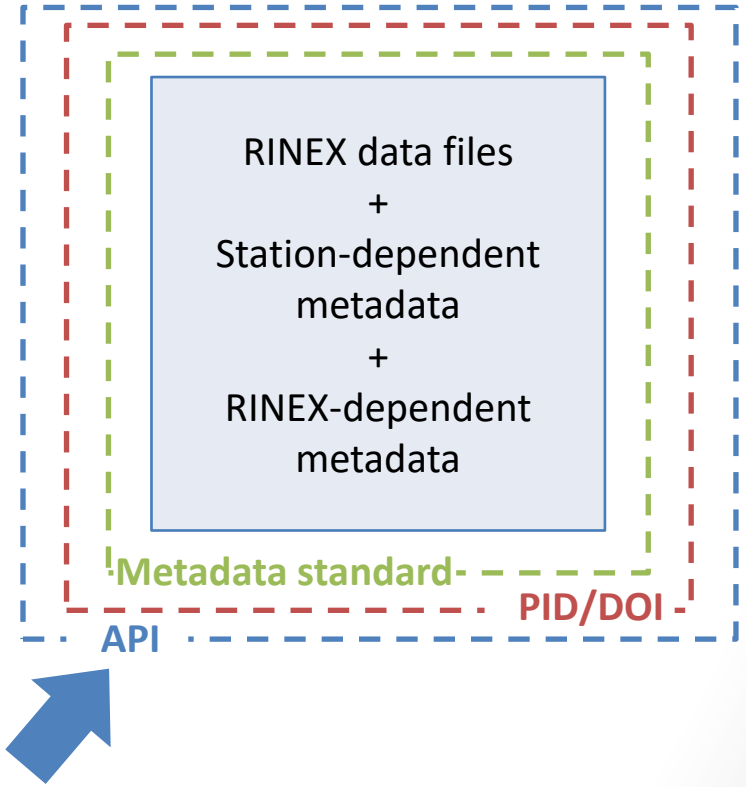
Metadata provide a **machine-readable and structured** form of documentation

rich metadata:

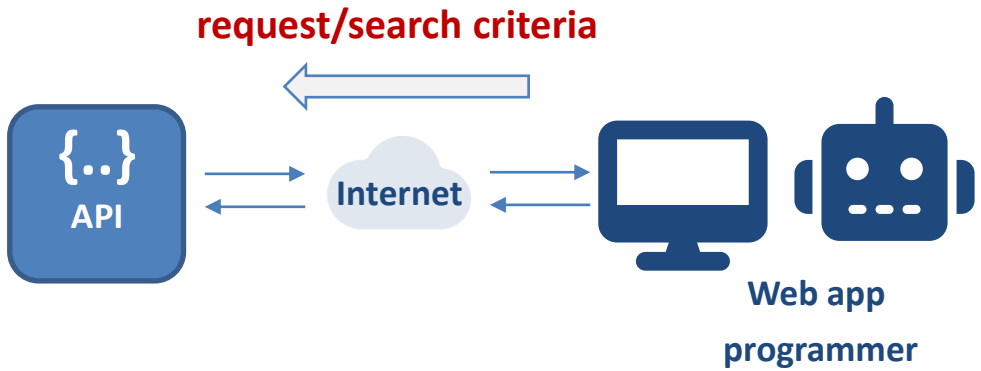
- provide all the information a user needs to know when **searching/downloading data**
- easy discovery and harvesting of specific data by machines (**standardized vocabulary and formats**)
- **interoperability** with other data sets
- several metadata **serialization formats** can be provided (user can select the most suitable)

structured information **metadata schema: GNSS-DCAT-AP**…





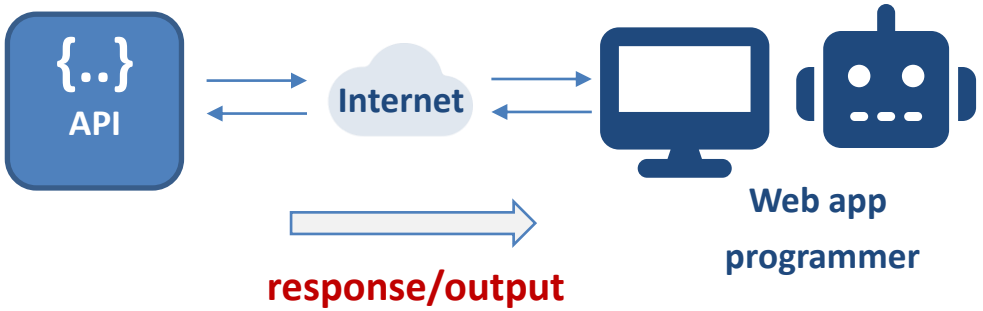
API to access to GNSS (meta)data: search for (meta)data



request/search criteria

- Type of GNSS data
- Time range
- …

API to access to GNSS (meta)data: retrieve the results of the search



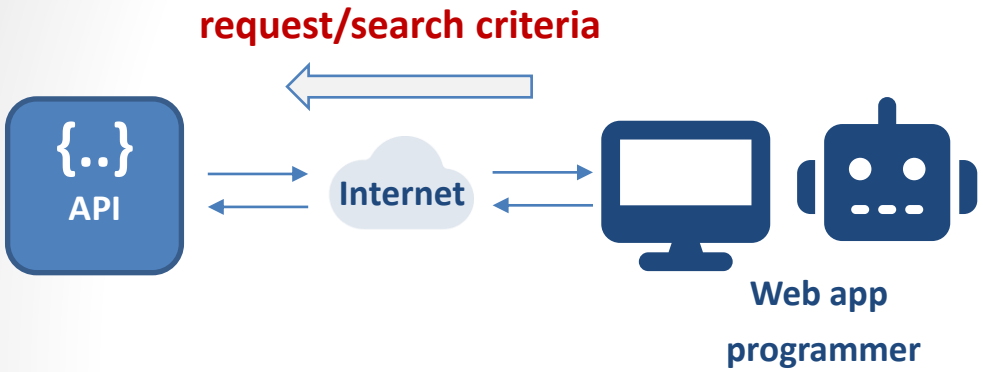
response/output

Depending on the request, different possible outputs:

- JSON/CSV etc. files with **metadata** & links to **data** files
- Simple text with links to **data** files
- …

API to access to GNSS (meta)data: started exploring some examples

UNAVCO Unified Web Services



request/search criteria

- Type of GNSS data
- Time range
- …

```
{
  "dataTypes": [
    "hatanaka",
    "obs"
  ],
  "items": [
    {
      "endDate": "2020-01-31",
      "items": {
        "410": "SC01",
        "1190": "P123"
      },
      "startDate": "2020-01-01"
    }
```

POST /gps/file-metadata/fetch-products/beta — GNSS/GPS Fetch Products Metadata

Beta APIs

Implementation Notes

Retrieves summary information, file paths, or file paths with file metadata of products given a list of station_id's between a start and (optional) end date.
POST method service that accepts a json object with the following parameters:
PARAMETERS:
"view": (string)
summary (summary returns a summary of file information - how many files returned etc..),
list (list returns a list of the ftp file paths),
metadata (metadata returns the list of ftp file paths along with other file metadata)
"output": (string)
output type that the user would like returned. current options:
if view is summary: "application/json"
if view is list: "text/plain"
if view is metadata: "application/json", or "text/csv"
"dataTypes": (list)
list of file types. example: ["hatanaka", "nav", "obs", "qc", "met"] or any combination. e.g. ["hatanaka", "qc"]
"products": (list)
list of product types - currently only accepts 'position' product type. example: ["position"]
"prettyPrint" (optional): (boolean)
true or false for json outputs
"items": (list of objects)
items is a list of objects with secondary parameters: startDate, endDate, and items - there can be several of these objects with different date ranges and items
"startDate": desired start date. required. "2019-01-01" or "2019-01-01T00:00:00.000Z",
"endDate": desired end date. can be null. example: "2020-01-01" or "2020-01-01T00:00:00.000Z",
"items": an object of site id's and accompanying site code (four character id): {"1190": "P123", "4733":"P123", "410":"SC01"}

Parameters

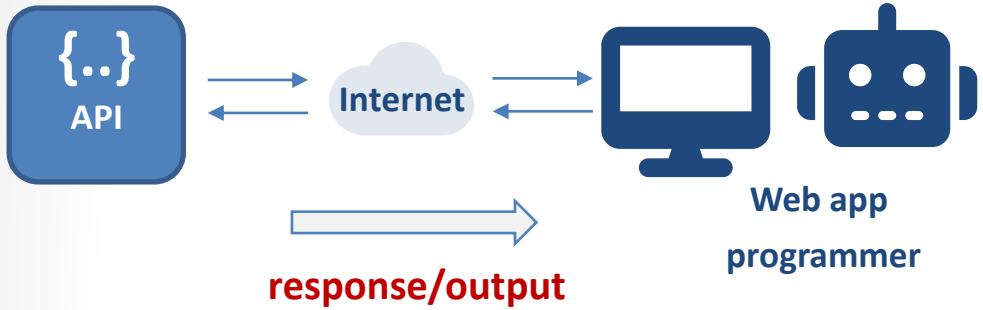
Parameter	Value	Description	Parameter Type	Data Type
information		data used to access the rinex files	body	Model | Example Value

Parameter content type: application/json

{
  "dataTypes": [
    "hatanaka",
    "obs"
  ],
  "items": [
    {

API to access to GNSS (meta)data: started exploring some examples

Beta REST-APIs



response/output

Depending on the request, different possible outputs:

- JSON/CSV etc. file with **metadata** & links to **data** files
- Simple text with links to **data** files
- …

UNAVCO Unified Web Services

Curl

```
curl -X POST --header 'Content-Type: application/json' --header 'Accept: application/octet-stream' -d '{ \
  "dataTypes": [ \
    "hatanaka", \
    "obs" \
  ], \
  "items": [ \
    { \
      "endDate": "2020-01-31", \
      "items": { \
        "410": "SC01", \
        "1190": "P123" \
      }, \
      "startDate": "2020-01-01" \
    } \
  ], \
  "output": "text/plain", \
  "prettyPrint": false, \
  "products": [ \
    "position" \
  ], \
  "view": "list" \
}' 'https://web-services.unavco.c
```

Request URL

```
https://web-services.unavc
```

Response Body

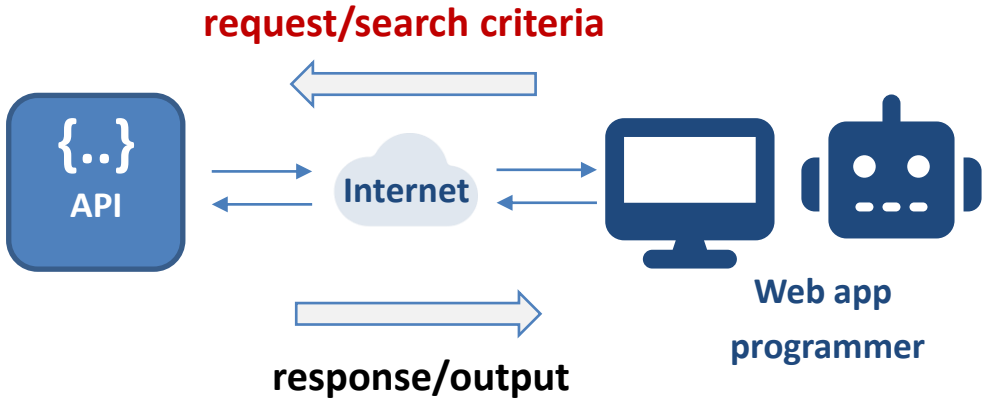

```
{
  "overall_file_size": 175258988,
  "overall_number_files": 140,
  "items": [
    {
      "site_id": 1190,
      "data_type": "hatanaka-compression of RINEX obs",
      "site_code": "P123",
      "number_files": 30,
      "site_file_size": 20746900,
      "dates_requested": "2020-01-01 to 2020-01-31",
      "sample_interval": 15
    },
    {
      "site_id": 410,
      "data_type": "hatanaka-compression of RINEX obs",
      "site_code": "SC01",
      "number_files": 30,
      "site_file_size": 78738071,
      "dates_requested": "2020-01-01 to 2020-01-31",
      "sample_interval": 15
    },
    {
      "site_id": 410,
```

API to access to GNSS (meta)data: lesson learned

We noticed that, so far and in general

- few non-Beta/operational APIs to download GNSS data
- no common approach
- different formats
- few metadata available and not in standard formats→ not FAIR compliant

Preparatory work for API: use cases for search criteria



Retrieve station information

A.2 *– based on time range*
A.3 *– based on type of receiver/antenna/firmware version and date range*
A.4 *– for stations with an external clock*
A.5 *– for stations with meteorological instruments*

Retrieve RINEX data

B.1 *– that have been changed since a specific date*
B.2 *–based on time range*
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*

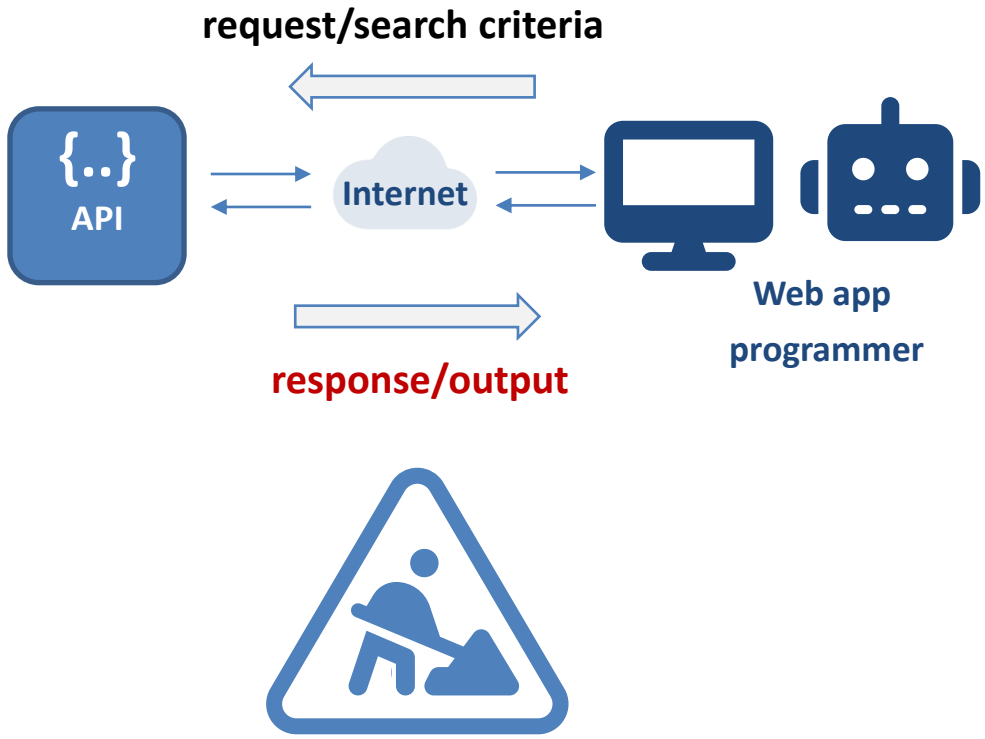
GNSS station information

First feedback from GNSS users:

retrieve data based also on

- *monument type*
- *co-located geodetic instrumentation ("Local Survey Tie" and "Collocation Information")*
- *responsible agency and/or operator*

Preparatory work for API, next step: use cases for output format



…complete the API design phase

Contents of the webinar

- **Part 1 - GNSS data and the data repository**
- **Part 2 - Introduction to FAIR data principles**
- **Part 3 - The journey so far: FAIR for GNSS data**
  - Where and how to start?
  - Gap analysis a.k.a. FAIR assessment
  - Metadata for GNSS data
  - Restructuring the data repository
  - Persistent identifiers for GNSS data
  - API preparatory work
- **Part 4 – Lessons learned and next steps**

Part 4 - Lessons learned and next steps

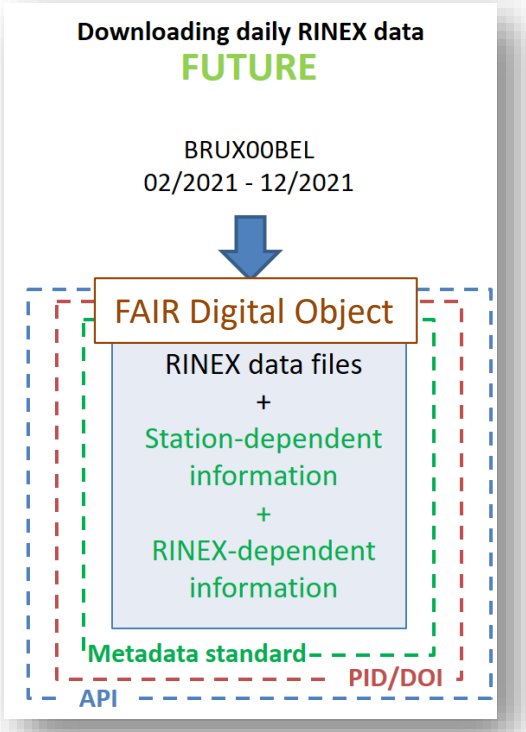
Summary

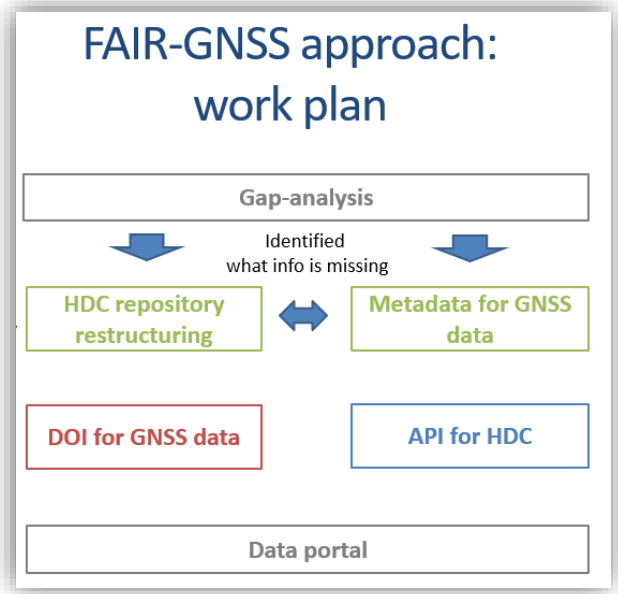
Understand FAIR data principles

→ Help from University of Ghent

How to apply FAIR data principles in practice to GNSS data in our repository?

→ Methodology





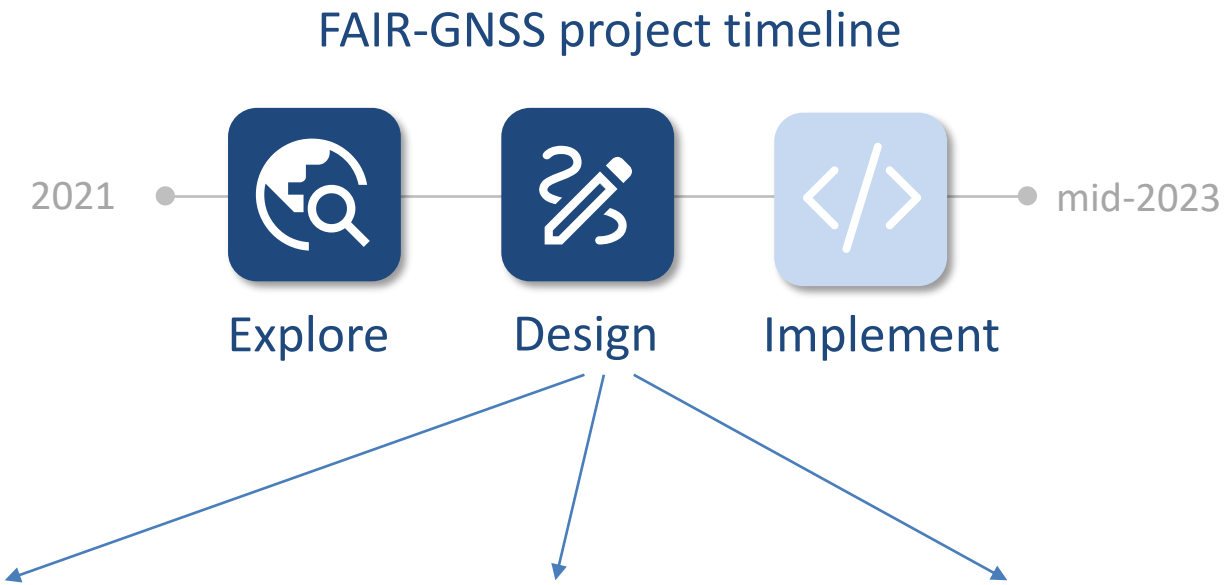
- *Missing provenance information*
- *Almost no DOI in use*
- *Need for standardized metadata*

ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

Summary



Metadata standards

- Station-dependent metadata:
  → Extended GeodesyML proposal
- RINEX-dependent metadata
  → GNSS-DCAT-AP proposal

New repository database

Including all information to generate rich metadata as required by the FAIR data principles

Persistent identifiers

DOI for GNSS data

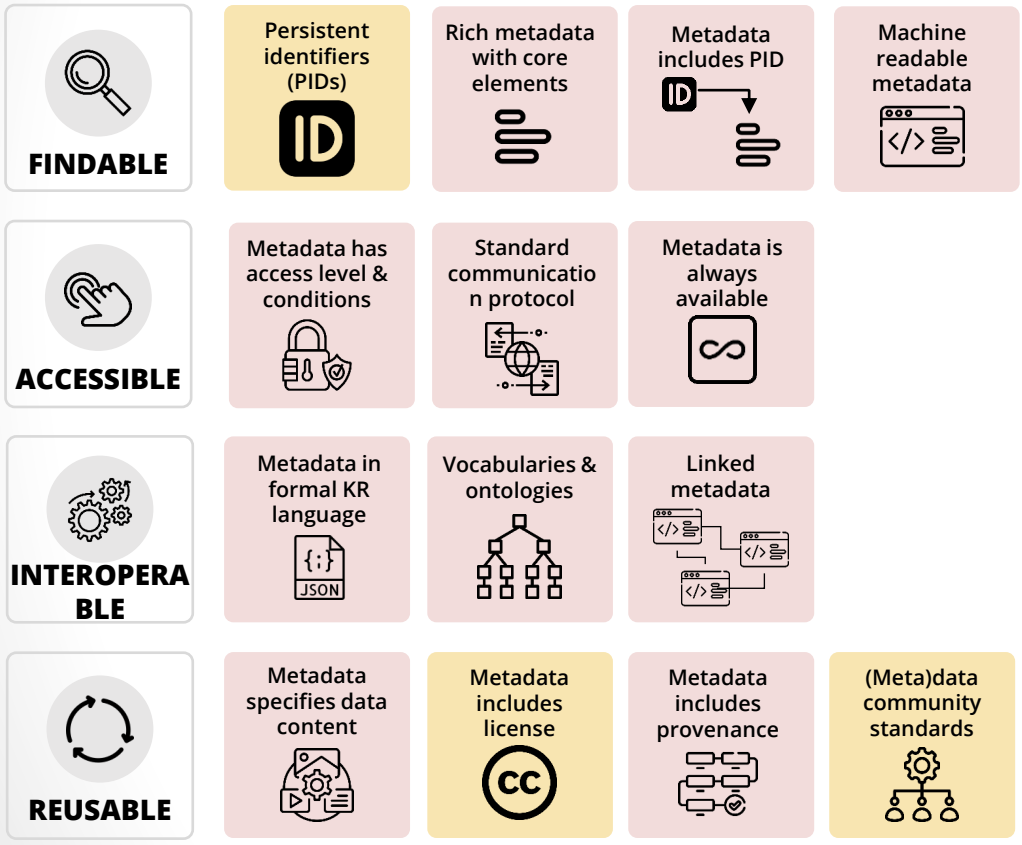
→ Standardization of GNSS DOI proposal

API

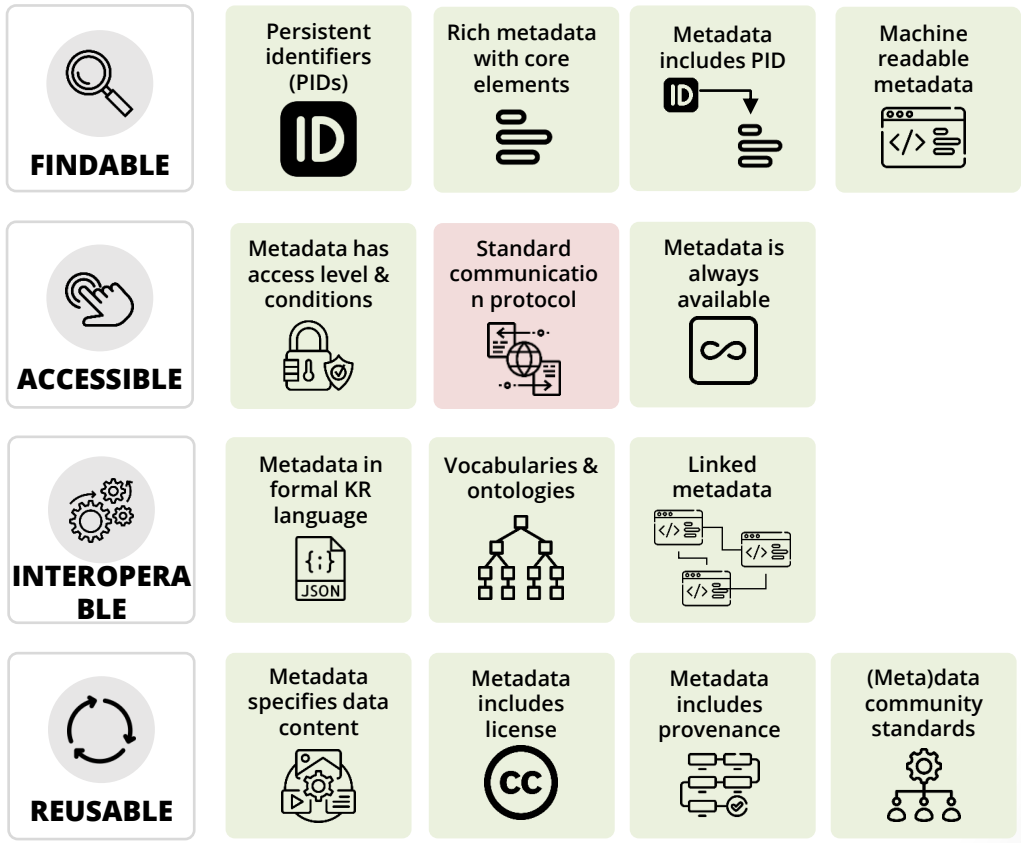
Open data portal

Progress towards FAIR: **Design**

at the beginning of the FAIR-GNSS project



…today



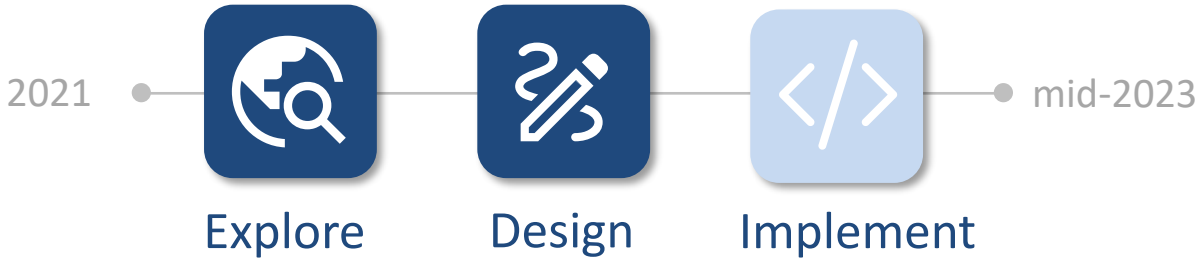
ROYAL OBSERVATORY OF BELGIUM

FAIR-GNSS Webinar October 11, 2022

GHENT UNIVERSITY

Summary

FAIR-GNSS project timeline



Populate repository database

- Software implemented
- Now running

Metadata standards

- Station-dependent metadata: extended GeodesyML to be implemented in M^{3}G
- RINEX-dependent metadata: GNSS-DCAT-AP to be implemented from scratch

Persistent identifiers

- Insertion of DOI for GNSS stations: almost ready in M^{3}G
- DOI minting service: to be implemented

API

- Use cases available
- To be implemented from scratch

Open data portal

- To be implemented (not from scratch)

Summary

FAIR-GNSS project timeline

Limited duration project → not possible to wait for community agreements on metadata standards

Need to start implementing using current FAIR-GNSS proposals

If community will request metadata changes later on, ROB will adapt its code accordingly

Multiple development cycles and each time we get a bit closer to FAIR

But, repository databases must be filled first (rest is about exporting)

Populate repository database

- Software implemented
- Now running

- extended GeodesyML to be implemented in M³G
- RINEX-dependent metadata: GNSS-DCAT-AP to be implemented from scratch

stations: almost ready in M³G

- DOI minting service: to be implemented

API

- Use cases available
- To be implemented from scratch

Open data portal

- To be implemented (not from scratch)

Lessons learned

2-year project
24PM for ROB (1 person…)
11PM for University of Ghent

Took time to really understand FAIR data principles (lost in all different metadata standards)

Much more people involved
- Almost whole GNSS team at ROB had to get involved in this adventure
- Underestimated the time to
  - Interacting with the community
    - Data providers (data license, DOI)
    - Working groups: agree on standards for metadata, known to be a very slow process
  - Restructure the repository

A lot of implementation work is still waiting
- Project will be extended until Sept. 2023
- And we will continue after the end of the project
- Improve core service that ROB has been offering to the EUREF community
- With the goal to evolve step by step towards becoming a trustworthy data repository

Your feedback is welcome!

Proposal for station-dependent metadata

https://github.com/ROB-GNSS/GeodesyML_proposal

Proposal for RINEX-dependent metadata

https://github.com/ROB-GNSS/GNSS-DCAT-AP

Follow IGS for the upcoming version of the extended GeodesyML:
https://github.com/International-GNSS-Service/GeodesyML

Contact us

FAIR-GNSS

fair-gnss @observatory.be

ROYAL OBSERVATORY OF BELGIUM

@be_GNSS

GHENT UNIVERSITY

@UGentRDM