



HAL
open science

Data sharing of in-situ measurements following GEO and FAIR principles in the solar energy sector

Philippe Blanc, Raphaël Jolivet, Lionel Ménard, Yves-Marie Saint-Drenan

► To cite this version:

Philippe Blanc, Raphaël Jolivet, Lionel Ménard, Yves-Marie Saint-Drenan. Data sharing of in-situ measurements following GEO and FAIR principles in the solar energy sector: An end-to-end implementation example in the solar energy domain ranging from data encoding up to search and discovery.. 2022. hal-03811628

HAL Id: hal-03811628

<https://hal.science/hal-03811628>

Preprint submitted on 14 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License

Data sharing of *in-situ* measurements following GEO and FAIR principles in the solar energy sector. An end-to-end implementation example in the solar energy domain ranging from data encoding up to search and discovery.

MINES Paris, PSL Research University, Center Observation, Impacts, Energie (OIE).

Authors: Philippe Blanc, Raphaël Jolivet, Lionel Ménard, Yves-Marie Saint-Drenan

Contact: lionel.menard@minesparis.psl.eu

ABSTRACT:

In-situ measurements are key for many sectors. In the Earth Observation (EO) domain they are of paramount importance often in conjunction with other EO data including numerical models and satellite, notably for calibration, data fusion and validation. Historically *in-situ* measurements mostly come from static local sensors including thermometer, anemometer, pyranometer, etc. Recently emerged via dynamic and mobile sensors in the scope of the Internet of Things (IoT) and crowdsourcing including citizen science. Whether static or dynamic, and considering any type of physical observations, they share the common concept of geo-location. Indeed, any *in-situ* measurements are bind with a geo-location expressed in latitude, longitude and altitude for any place on Earth. However, despite their importance, finding and accessing *in-situ* measurements is often difficult if not impossible. Moreover, when findable and accessible, they often lack of common standard procedures to properly identify and define their characteristics: provenance, IPR, quality / fitness for use as well as the characterization of the type and the content of the included measurements. This is often the case in the renewable energy sector and more precisely in the solar energy domain. In that context, this paper aims at showcasing a concrete example of data sharing of *in-situ* measurements following GEO and FAIR principles from various networks of ground-based stations providing high temporal resolution *in-situ* measurements in the solar domain for decades. The underlying question of data formatting and dissemination of solar resource data is one of the activities of the Task 16 of the PVPS program of the IEA¹.

Keywords: *In-situ Measurement, Solar Energy, Photovoltaic, Earth Observation, Data Sciences, Data Analytics, Web services, Standard, Interoperability architecture, Open Geospatial Consortium OGC, Group on Earth Observations, International Energy Agency, FAIR principles, GEO Data Management Principles, SciDoSol, PVPS, NetCDF, Climate and Forecast convention, Thredds Data Server*

Contents

1.	Introduction.....	2
2.	GEO and FAIR principles applied to data formatting and dissemination in solar energy.....	2
3.	Details in the FAIR implementation of data formatting and dissemination	5
3.1.	Workflow for normalizing and encoding data and CF metadata.....	5
3.2.	The Climate and Forecast (CF) compliant NetCDF format.....	5
3.3.	Enforcing conventions	5
3.4.	Gathering and enriching metadata.....	6
3.5.	The full workflow	7
4.	Deployment in the webservice SDI Thredds Data Server (TDS).	8
5.	Benefits of genericity and agnosticism.....	9
5.1.	Discovering with Panoply	9
5.2.	Example of data downloading and plotting within Matlab	12
5.3.	Example of QC in python (Jupyter Notebook).....	17
5.1.	Implementation of a visual plausibility control tool.....	21

¹ <https://iea-pvps.org/research-tasks/solar-resource-for-high-penetration-and-large-scale-applications/>

5.2.	Web-based application (The “wow effect”)	22
6.	Search and discovery with webservice-energy catalogue.....	25
7.	Conclusion and perspectives	28
8.	Annex: presentations done with PVPS T16	29
9.	Annex: libinsitu resources	29

1. Introduction

In-situ measurement are of paramount importance in the renewable energy sector and in particular in the solar energy domain, when considering pyranometric stations with dedicated sensors for surface solar irradiance (SSI). When combined with satellite and meteorological models supporting data fusion, validation and calibration, long time *in-situ* measurements have proven to provide informed decision making supporting the solar energy sector for transitioning toward a carbon-free energy supply by 2050². Despite their proven interest, *in-situ* measurements are spatially and temporally scarce and are difficult to find and access. They mainly come from ground-based measurement stations organized in networks operated by meteorological agencies, research institutions or industries. They are generally not homogeneously spread around the world, covering mostly the northern hemisphere. Within those networks, specific solar radiation measurements are not always available and when available, information on the data quality is generally missing. Dedicated solar radiation networks exist and among them the Baseline Surface Radiation Network (BSRN) providing high accuracy and high time resolution 1-min *in-situ* measurement from a network of 74 stations spread in contrasting climatic zones, covering a latitude range from 80°N to 90°S. Regarding networks of ground based stations covering the need of the solar energy domain, the general trend is a decrease in number and no new ones are currently planned. Indeed, measuring stations operation and maintenance is challenging and costly. Contrarily there is an increasing number of “virtual” measuring stations in the form of photovoltaic (PV) and concentrated solar power (CSP) plants that daily operate and collect valuable solar radiation data for internal business purpose. Unfortunately access to those data is mostly impossible for release constrains applied by PV and CSP operators. Initiatives to encourage those operators to openly release their solar radiation and meteorological parameters have been conducted (ConnectinGEO³). Despite a proven win/win approach for operators to openly shared their installations *in-situ* measurements, very few are currently available. Providing an open access to these untapped nuggets of measuring stations all over the word is definitely a way to further investigate in order to provide a much higher mesh in support to the above objectives of a carbon-free energy supply.

In addition to the scarcity of data described in the previous section, we have identified that finding and accessing *in-situ* measurements are often difficult if not impossible. Moreover, when findable and available, they often lack a common standard procedure to properly identify and characterize their provenances, IPR and quality / fitness for use as well as the characterization of the type and the content of the included measurements. This is often the case in the renewable energy sector and more precisely in the solar energy domain. In that context, this paper aims at showcasing a concrete example of data sharing of *in-situ* measurements following GEO and FAIR principles from various networks of ground-based stations providing high temporal resolution *in-situ* measurements in the solar energy domain for decades.

The following document will provide an example of this general approach to support a global and open access to *in-situ* measurements from existing research, academic and institutional networks while keeping on engaging discussion with the private sectors of PV and CSP operators for the release of their in-fields plants *in-situ* measurements. It is worth mentioning that the question of data formatting and dissemination of solar resource data is one of the activities of the Task 16 of the PVPS program of the IEA⁴.

2. GEO and FAIR principles applied to data formatting and dissemination in solar energy

² https://library.wmo.int/index.php?lvl=notice_display&id=22136 (2022 State of Climate Services: Energy (WMO-No. 1301) – Case Study 7)

³ <http://www.connectingeo.net/Index.htm>

⁴ <https://iea-pvps.org/research-tasks/solar-resource-for-high-penetration-and-large-scale-applications/>

In the solar energy domain, *in-situ* pyranometric measurements, related to incident solar radiation components, come from various sources including national meteorological office, institutions and solar energy related industries, operating at local, national, regional and international levels. This scattered landscape of data providers is not relying on common and standard procedures to cover the life cycle of *in-situ* pyranometric measurements. Indeed, they are generally repeating and propagating in-house practices that are often in place for decades, ensuring internal consistency and trust of *in-situ* measurements archives.

Generic guidelines for data sharing have been created in support to the EO community. International organizations advocating and supporting the use of EO data, including the intergovernmental Group on Earth Observation (GEO), have implemented Data Sharing Principles (DSP) to promote free, full, open and timely access to Earth observation datasets, products and services. Build upon these DSP, GEO has established Data Management Principles (DMP) to support discoverability, accessibility, usability, preservation and curation of data. Following the open data era, similar principles have been enacted and gathered around the Findable, Accessible, Interoperable, Reusable (FAIR) Principles⁵. Numerous examples of the implementation of such practices have been successfully conducted for satellite-based data.⁶ On the contrary very few have been implemented for *in-situ* measurements⁷. In the scope of this paper, special attention has been paid to follow GEO DMP and FAIR principles. In practice, from raw *in-situ* measurement data encoding up to the creation of metadata to support interoperable search and discovery the workflow has successfully leveraged each of the above-mentioned principles. By doing so, we want to demonstrate that following such principles allows mitigating the data trust issues described in introduction and thereby maximizing the usability of existing *in-situ* measurements in the solar energy sector.

Regarding GEO DMP and FAIR principles, a detailed comparison analysis has underpinned overlap between some of these principle as shown in the (Figure 1).



DSP-1: Open by Default

DSP-2: Available with minimal restrictions, at no more than the cost of reproduction and distribution

DSP-3: Available with minimum time delay.

FAIR does not explicitly address the DSPs

DMP-1: Metadata for Discovery

DMP-2: Online Access

DMP-3: Data Encoding

DMP-4: Data Documentation

DMP-5: Data Traceability

DMP-6: Data Quality-Control

DMP-7: Data Preservation

DMP-8: Data and Metadata Verification

DMP-9: Data Review and Reprocessing

DMP-10: Persistent and Resolvable Identifiers

Findable

Accessible

Interoperable

Reusable

Based on: GEOSS Data Sharing Principles Implementation Guidelines 2016-2025, GEOSS Data Management Principles: A Brief Overview, and The FAIR Data Principles.

Figure 1: Correspondence between data management plans (GEO DMP, left column) and the FAIR principles (right column)

⁵ <https://tinyurl.com/24bwrku8>

⁶ <https://ceos.org/news/open-data-part1/>

⁷ <https://tinyurl.com/manbtrsu>

For the sake of simplicity, we will use the four FAIR principles to illustrate in practice how we have implemented these principles all along the data workflow. Moreover, to follow a logical and natural path as per data workflow, we will illustrate these principles in the following order Reusable, Accessible, Interoperable, Findable (see Figure 2).













  	  	  	  
<p>NetCDF format for metadata and data encoding (self-content, self-described) conform to the CF-conventions</p>	<p>Accessible via the Thredds Data Server enabling data and metadata operations (Download, subset, search, discovery and access with OPeNDAP protocol,...)</p>	<p>Processes triggering the Thredds OPeNDAP server from remote applications including Notebooks, API (Python R,...) framework , Matlab, Web App Panoply,...</p>	<p>Findable via open and standard metadata in a catalogue (webservice-energy and GEO), D.O.I. and landing page</p>

Figure 2: Schematic representation of the match between the steps of the processing chain and the four components of the FAIR principle.

- **Reusable:** The choice was made to encode *in-situ* measurement in NetCDF⁸ format an open, self-content, self-described, compressed multi-dimensional data binary format. This binary encoding of the data is coming along metadata description following the Climate and Forecast (CF) metadata conventions⁹ enabling a formal description of data variable as well as their spatial and temporal properties. The combined NetCDF and CF metadata will be named hereinafter CF-NetCDF.
- **Accessible:** CF-NetCDF files have been deployed on a Thredds Data Server (TDS), a web application server providing metadata and data access supporting the following protocols: OPeNDAP (Open-source Project for a Network Data Access Protocol), OGC Web Map Service (WMS), Web Coverage Service (WCS) and HTTP (enabling download).
- **Interoperable:** Thanks to the OPeNDAP protocol supported by the TDS application server, interoperability is made possible. We have illustrated this interoperability concept with various examples empowering open access to the data via Jupyter Notebook, API (Python, R), Matlab, web-based custom applications and desktop application such as Panoply¹⁰.
- **Findable:** Making data searchable and discoverable is a key activity for a successful dissemination process. Metadata records have been created following coherent and complementary approaches. On the one hand manually generated metadata records following ISO 19139 standard have been created in the GEO community webservice-energy catalogue to gather *in-situ* measurement from various stations belonging to an identified network. On the other hand, an automatic metadata generation has been performed by the GEO DAB (Discovery and Access Broker) harvester directly from the TDS server onto each of the granular NetCDF station files. It worth notice that the webservice-energy catalogue based on the open source GeoNetwork tool enables automatic harvest following OGC CSW (Catalog Service for the Web) protocol. Since 2008 the webservice-energy catalogue is weekly harvested by the GEO DAB. It enables a such the widest possible dissemination of the *in-situ* measurements.

⁸ <https://www.unidata.ucar.edu/software/netcdf/>

⁹ <https://cfconventions.org/cf-conventions/cf-conventions.html>

¹⁰ <https://www.giss.nasa.gov/tools/panoply/>

In practice and in more details the following chapters explain the general methodology that has been followed to implement full FAIR compliant principles.

3. Details in the FAIR implementation of data formatting and dissemination

The general methodology described above has been brought into practice following an open and standard approach.

3.1. Workflow for normalizing and encoding data and CF metadata

We developed and deployed a workflow for encoding data from various networks into normalized CF-*NetCDF* files. We published the code and metadata as an open-source library in python: [libinsitu](#).

3.2. The Climate and Forecast (CF) compliant NetCDF format

We chose *NetCDF* as the main format for distributing *in-situ* measurement data.

NetCDF is a file format for structured numerical data. It shows several benefits:

- **Compact:** *NetCDF* is small compared to other formats. It is a binary format and supports lossless and loosely compression
- **Self-descriptive:** *NetCDF* embeds a powerful mechanism of metadata. This enables to document each variable (meaning, units, boundaries, etc, ...)
- **Wide support:** *NetCDF* is supported by a wide variety of geospatial tools (e.g. Panoply, HDFview, QGIS) and languages and API libraries (*Python*, *Matlab*, *CDO*, *GDAL*)
- **Efficient request and sub-setting:** The structure of *NetCDF* make it efficient to extract random parts of the data without reading the whole file. Querying and sub-setting *NetCDF* can be done on local files or remotely, over the OPeNDAP protocol.

NetCDF comes along with a specific human-readable notation for *NetCDF* objects, data and metadata: Common Data Language (CDL). It enables to describe, with this tiny language, templates for *NetCDF* files to be creates, according to pre-assigned mandatory metadata and name, size, attributes of data variables. It can be used as text output to simply describe the structure of a given *NetCDF* file.

3.3. Enforcing conventions

The metadata attached to *NetCDF* is guided by CF conventions: a set of recommendations for Climate and Forecast data. We propose additional restrictions, in compliance with the CF conventions, to enhance the genericity and compatibility of the data.

Those recommendations are published in the form of:

- [A white paper](#)
- [A CF-NetCDF schema](#) (CDL file)

Our additional convention essentially enforces the following restrictions:

- **Uniform time scale:** The time must be in UTC and must have a uniform sampling from the beginning to the end of the measurements, without gaps or repetitions. Gaps are filled with NaN values.
- **Global metadata:** We enforce a minimal set of global metadata for describing the station (name, location, elevation, etc.)
- **Variable names and units:** We propose a set of supported variables, with their name and the preferred unit. The following variables are supported:

Name	Content	unit
GHI	Global horizontal irradiance	W/m ²

DHI	Diffuse horizontal irradiance	W/m ²
BNI	Beam (or direct) normal irradiance	W/m ²
T2	Air temperature	K
RH	Relative humidity	1 (ratio)
P	Air pressure	Pa
WS	Wind speed	m/s
WD	Wind direction	Degrees

Respecting those conventions enables to leverage the genericity of data and ease the development of new tools (Visualization, Quality-Check (QC) process, etc., ...)

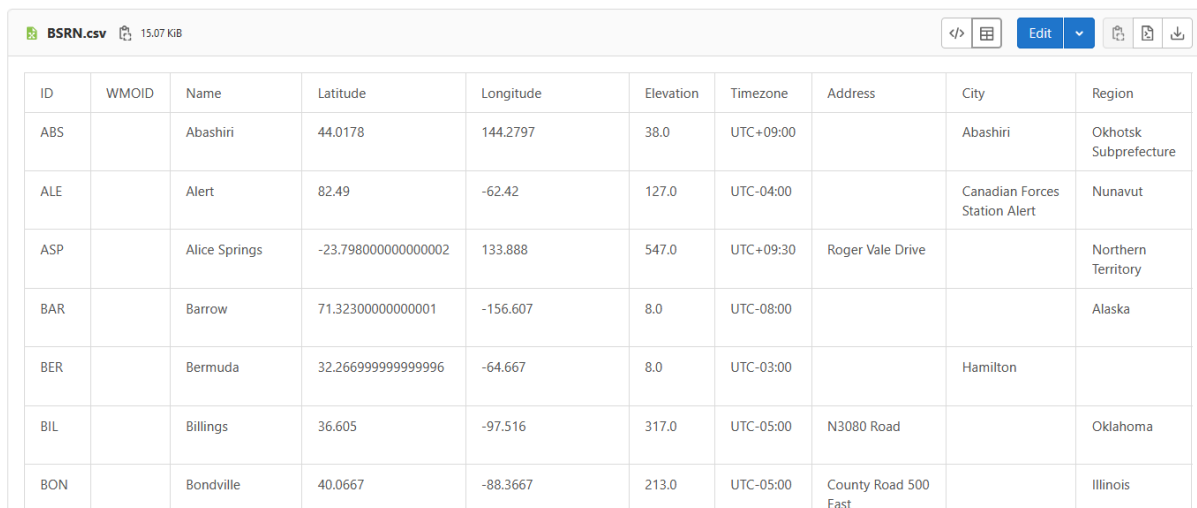
3.4. Gathering and enriching metadata

We gathered the metadata of various networks and individual stations and formatted it in consistent CSV files. This metadata is used in the workflow and embedded in the output NetCDF files.

We also enriched this metadata automatically with additional information, inferred from the geolocation:

- Local time zone
- Complete address, region & country
- Local climate

Those CSV files are published online (see screen copy in [Figure 3](#)), as part of *libinsitu*. We expect contributions from the community to verify and complete this information.



ID	WMOID	Name	Latitude	Longitude	Elevation	Timezone	Address	City	Region
ABS		Abashiri	44.0178	144.2797	38.0	UTC+09:00		Abashiri	Okhotsk Subprefecture
ALE		Alert	82.49	-62.42	127.0	UTC-04:00		Canadian Forces Station Alert	Nunavut
ASP		Alice Springs	-23.798000000000002	133.888	547.0	UTC+09:30	Roger Vale Drive		Northern Territory
BAR		Barrow	71.323000000000001	-156.607	8.0	UTC-08:00			Alaska
BER		Bermuda	32.266999999999996	-64.667	8.0	UTC-03:00		Hamilton	
BIL		Billings	36.605	-97.516	317.0	UTC-05:00	N3080 Road		Oklahoma
BON		Bondville	40.0667	-88.3667	213.0	UTC-05:00	County Road 500 East		Illinois

Figure 3: Extract of the metadata of stations of the BSRN network

3.5. The full workflow

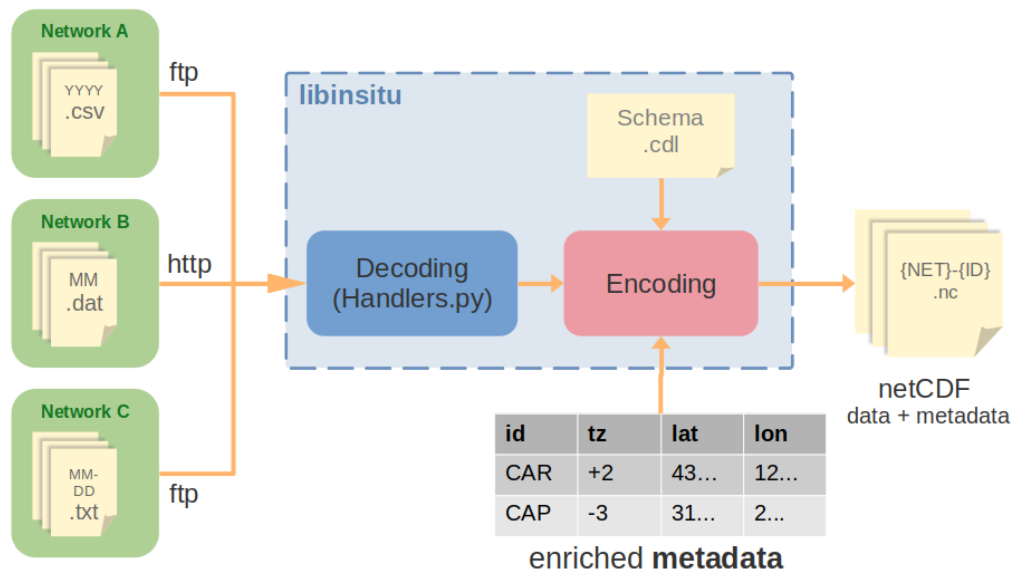


Figure 4: Full workflow for normalizing and encoding data and metadata

The full workflow consists in three phases:

- The data of each network is fetched from the source, either via FTP or HTTP
- The data is decoded and transformed into a generic internal structure. We implemented a specific handler for each network.
- The generic structure is encoded into *NetCDF* format, following the CDL schema. Network and station metadata is embedded into the result file.

The whole process is incremental. New data is appended to existing *NetCDF* files, along the time dimension. We currently support the following networks:

ID	Name	Number of stations	Available Data	Protocol	Input format
ABOM	Australian Bureau of Meteorology	22	GHI, DHI, BNI	http	csv
BSRN	Baseline Surface Radiation Network	77	GHI, DHI, BNI, T2, RH, P	ftp	dat
enerMENA	The enerMENA Meteorological Network	13	GHI, DHI, BNI, T2, RH, WS, WD, P	ftp	mesor
NREL_MIDC	Measurement and Instrumentation Data Center	12	GHI, DHI, BNI	http api	csv
SAURAN	Southern African Universities Radiometric Network	21	GHI, DHI, BNI	http api	csv
SOLRAD	NOAA's SOLar RADIation Measurement Network	9	GHI, DHI, BNI, T2, RH, WS, WD, P	http api	csv
SURFRAD	NOAA's Surface Radiation Budget Network	7	GHI, DHI, BNI, T2, RH, WS, WD, P	ftp	csv
IEA-PVPS	IEA PVPS Task 16 Network	115	GHI, DHI, BNI	n.a.	csv
SKYNET	International SKYNET Data Center	6	GHI, DHI, BNI	http	csv
ISE PV live	PV-Live dataset - Measurements of global horizontal and tilted solar irradiance	40	GHI	zenodo	csv
ESMAP	Energy Sector Management Assistance Program (ESMAP)	16	GHI, DHI, BNI, T2, RH, WS, WD, P	http	dat

The above processes enable the R (Reusable) of our RAIF implementation.

Once generated, the output *NetCDF* files are pushed to a TDS server, as explained in the following chapter.

4. Deployment in the webservice SDI Thredds Data Server (TDS).

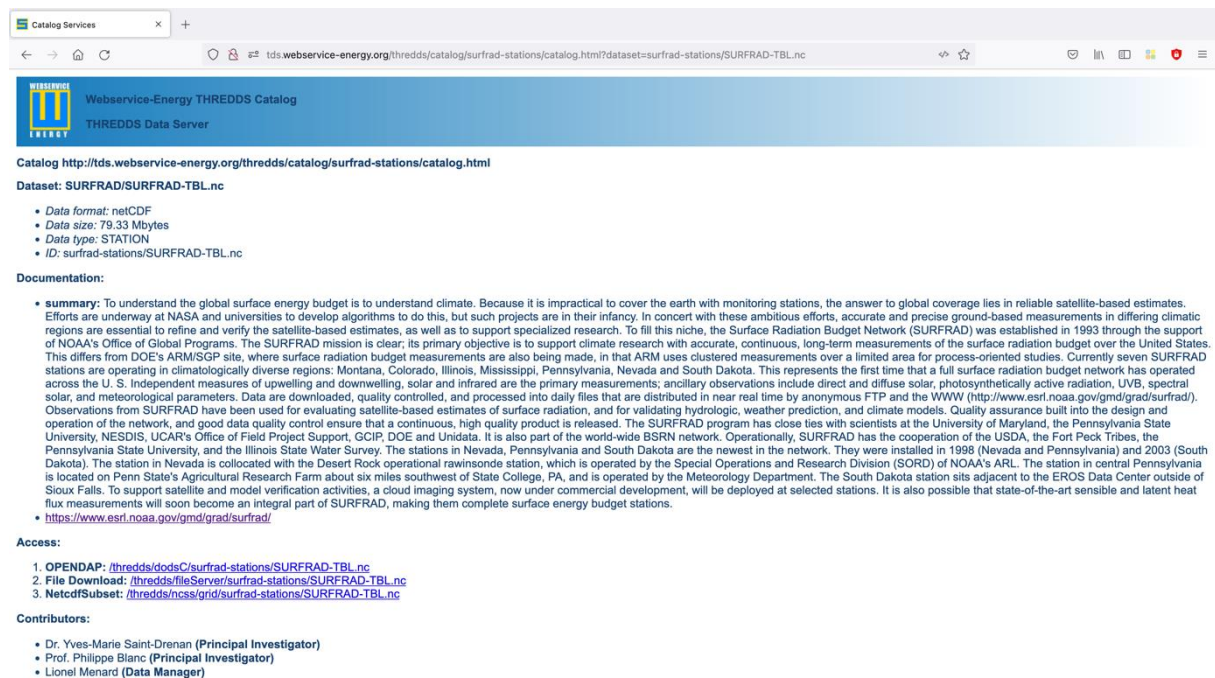
Thredds Data Server (TDS) is a web application server that provides access to metadata and data based on open standards and protocols including OPeNDAP, OGC WMS and WCS, HTTP and other remote data access protocols. TDS is in operation in the SDI (Spatial Data Infrastructure) webservice-energy.org for over a decade. It has proven to be an efficient and reliable application server to support metadata and data dissemination of NetCDF, CF compliant *in-situ* measurements.

Once properly encoded in NetCDF, CF format TDS provides out of the box, several key operations including:

- Download.
- Sub-setting in space and time and across all available data dimensions using NetCDF subset.
- Open and standard local and remote access based on the OpenDAP protocol.

On the top of these core operations TDS offer an internal modular XML based metadata framework allowing to document from a minimalistic up to a full description each *in-situ* measurement dataset.

As such this covers the A (Accessible) of our RAIF implementation.



WebService-Energy THREDDS Catalog
THREDDS Data Server

Catalog <http://tds.webservice-energy.org/thredds/catalog/surfrad-stations/catalog.html>

Dataset: SURFRAD/SURFRAD-TBL.nc

- Data format: netCDF
- Data size: 79.33 Mbytes
- Data type: STATION
- ID: surfrad-stations/SURFRAD-TBL.nc

Documentation:

- **summary:** To understand the global surface energy budget is to understand climate. Because it is impractical to cover the earth with monitoring stations, the answer to global coverage lies in reliable satellite-based estimates. Efforts are underway at NASA and universities to develop algorithms to do this, but such projects are in their infancy. In concert with these ambitious efforts, accurate and precise ground-based measurements in differing climatic regions are essential to refine and verify the satellite-based estimates, as well as to support specialized research. To fill this niche, the Surface Radiation Budget Network (SURFRAD) was established in 1993 through the support of NOAA's Office of Global Programs. The SURFRAD mission is clear; its primary objective is to support climate research with accurate, continuous, long-term measurements of the surface radiation budget over the United States. This differs from DOE's ARM/SGP site, where surface radiation budget measurements are also being made, in that ARM uses clustered measurements over a limited area for process-oriented studies. Currently seven SURFRAD stations are operating in climatologically diverse regions: Montana, Colorado, Illinois, Mississippi, Pennsylvania, Nevada and South Dakota. This represents the first time that a full surface radiation budget network has operated across the U. S. Independent measures of upwelling and downwelling, solar and infrared are the primary measurements; ancillary observations include direct and diffuse solar, photosynthetically active radiation, UVB, spectral solar, and meteorological parameters. Data are downloaded, quality controlled, and processed into daily files that are distributed in near real time by anonymous FTP and the WWW (<http://www.esrl.noaa.gov/gmd/grad/surfrad/>). Observations from SURFRAD have been used for evaluating satellite-based estimates of surface radiation, and for validating hydrologic, weather prediction, and climate models. Quality assurance built into the design and operation of the network, and good data quality control ensure that a continuous, high quality product is released. The SURFRAD program has close ties with scientists at the University of Maryland, the Pennsylvania State University, NESDIS, UCAR's Office of Field Project Support, GCIP, DOE and Unidata. It is also part of the world-wide BSRN network. Operationally, SURFRAD has the cooperation of the USDA, the Fort Peck Tribes, the Pennsylvania State University, and the Illinois State Water Survey. The stations in Nevada, Pennsylvania and South Dakota are the newest in the network. They were installed in 1998 (Nevada and Pennsylvania) and 2003 (South Dakota). The station in Nevada is collocated with the Desert Rock operational rawinsonde station, which is operated by the Special Operations and Research Division (SOR) of NOAA's ARL. The station in central Pennsylvania is located on Penn State's Agricultural Research Farm about six miles southwest of State College, PA, and is operated by the Meteorology Department. The South Dakota station sits adjacent to the EROS Data Center outside of Sioux Falls. To support satellite and model verification activities, a cloud imaging system, now under commercial development, will be deployed at selected stations. It is also possible that state-of-the-art sensible and latent heat flux measurements will soon become an integral part of SURFRAD, making them complete surface energy budget stations.
- <https://www.esrl.noaa.gov/gmd/grad/surfrad/>

Access:

1. OPeNDAP: thredds/dodsC/surfrad-stations/SURFRAD-TBL.nc
2. File Download: thredds/fileServer/surfrad-stations/SURFRAD-TBL.nc
3. NetCDFSubset: thredds/nccs/grad/surfrad-stations/SURFRAD-TBL.nc

Contributors:

- Dr. Yves-Marie Saint-Drenan (Principal Investigator)
- Prof. Philippe Blanc (Principal Investigator)
- Lionel Menard (Data Manager)

Figure 5: *In-situ* metadata and dataset basic operation display in TDS 1/2

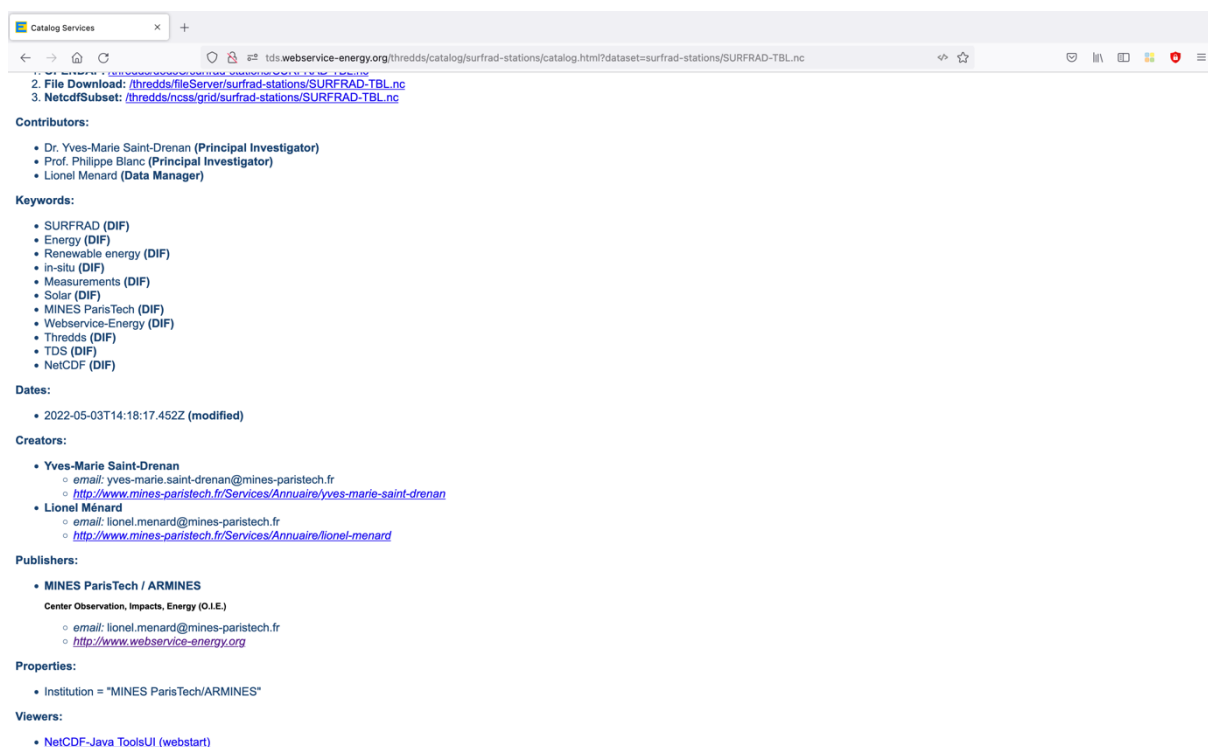


Figure 6: In-situ metadata and dataset basic operation display in TDS 2/2

5. Benefits of genericity and agnosticism

Once deployed and available on TDS as explained in section 3.2, *in-situ* measurements can benefit from extra possibilities supporting dataset dissemination and usage. Thanks to the standards natively supported by TDS a wide number of remote applications can seamlessly access *in-situ* measurements using standard protocols and APIs. The following examples illustrate access to *in-situ* measurements hosted and exposed on the TDS of the webservice-energy.org (<http://tds.webservice-energy.org>) SDI using remote applications such as such as Panoply a desktop NetCDF reader, script-based approaches like Jupyter Notebooks using Python language, Matlab solver, and finally using a dedicated web-based application.

5.1. Discovering with Panoply

Let's start with a desktop based application, namely Panoply. Panoply is a cross-platform application that read and plots geo-referenced and other arrays from NetCDF, HDF, GRIB, and other datasets. It is perfectly suited to read the *in-situ* measurement of time series of irradiation as we generate them following CF-Convention encoding as shown in the following screenshots.

All of these examples demonstrate the I (Interoperable) of our RAIF implementation.

The screenshot shows the Panoply interface with a table of variables on the left and their corresponding CF encoding on the right. The table lists variables such as BNI, DHI, elevation, GHI, latitude, longitude, P, RH, station_name, T2, and Time. The CF encoding on the right details the dimensions, units, and specific parameters for each variable, such as 'time' being in 'seconds since 1970-01-01 00:00:00' and 'GHI' being 'Short-wave downward (GLOBAL) radiation'.

Figure 7: Panoply – Reading a NetCDF-CF file

This screenshot displays the global attributes for the dataset in the Panoply interface. The left panel shows the same variable list as Figure 7. The right panel shows a detailed CF encoding block starting with 'global attributes:' and including fields like 'description', 'title', 'keywords', 'contact', 'institution', 'record', 'Network_ShortName', 'Network_Name', 'Network_Region', 'Network_Resolution', 'Network_TimeResolution', 'Network_References', 'Network_License', 'Network_Contact', 'Station_WMO_ID', 'Station_Name', 'Station_Latitude', 'Station_Longitude', 'Station_Elevation', 'Station_LocalTimeZone', 'Station_Address', 'Station_City', 'Station_Country', 'Station_SurfaceType', 'Station_TopographyType', 'Station_Rural_Urban', 'Station_KoeppeGeigerClimate', 'Station_OperationStatus', 'Station_TimeResolution', 'Station_DataEnd', 'Station_ContactPerson', 'Station_Institute', 'Station_URL', 'Station_CommissionDate', 'Station_DecommissionDate', 'Station_Comment', 'featureType', 'cdm_data_type', 'conventions', 'cdm_timeseries_variables', 'creation_date', 'Station_ID', and 'Station_DataBegin'.

Figure 8: Panoply – Displaying the CF encoding of global attributes

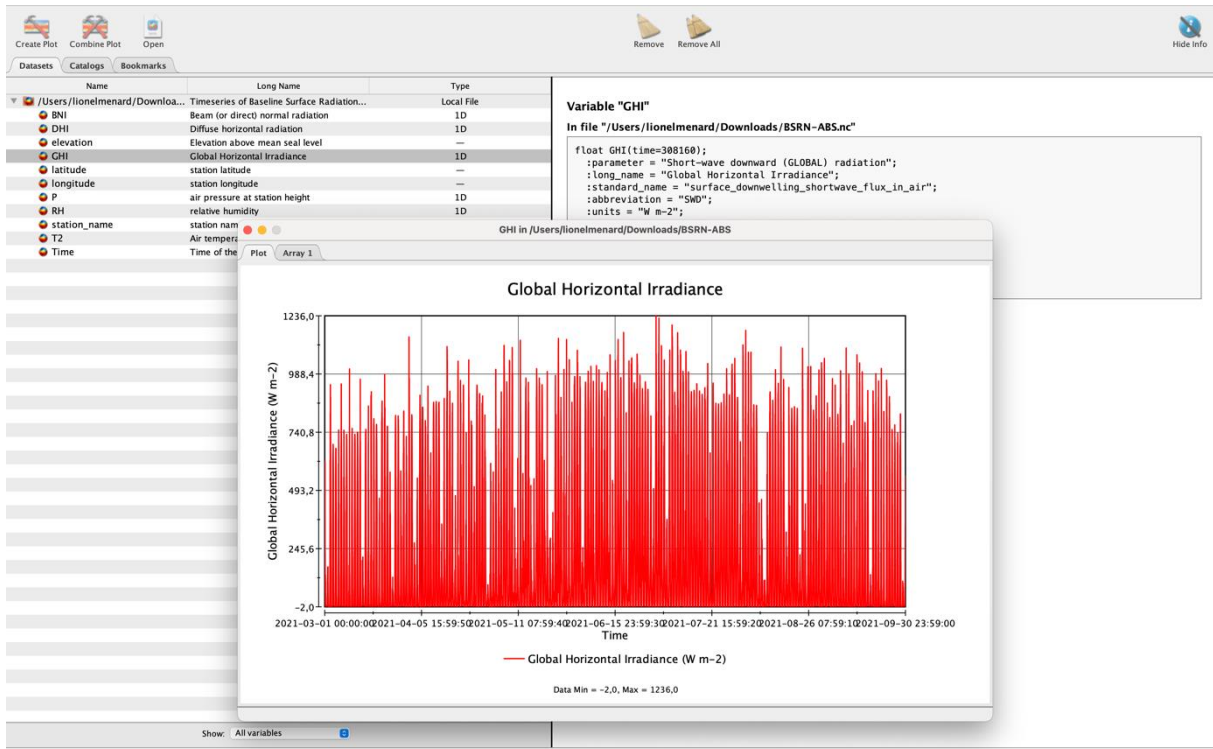


Figure 9: Panoply – Generating a graph of time series of Global Horizontal Irradiation (GHI)

5.2. Example of data downloading and plotting within Matlab

```
clear all
close all

file_nc = 'http://tds_bsrn_user:****@tds.webservice-
energy.org/thredds/dodsC/bsrn-stations/BSRN-CAR.nc';
```

Display the CF-NetCDF structure (CDL output)

```
ncdisp(file_nc)
```

```
Source:
  http://tds_bsrn_user:****@tds.webservice-energy.org/thredds/dodsC/bsrn-
  stations/BSRN-CAR.nc
Format:
  classic
Global Attributes:
  description = 'Archive of solar radiation networks worldwide
  provided by the Webservice-Energy initiative supported by MINES Paris PSL. Files are
  provided as NetCDF file format with the support of a Thredds Data Server.'
  title = 'Timeseries of Baseline Surface Radiation
Network (BSRN)'
  keywords = 'meteorology, station, time, Earth Science >
  Atmosphere > Atmospheric Radiation > Incoming Solar Radiation, Earth Science > Atmosphere >
  Atmospheric Temperature > Surface Temperature > Air Temperature, Earth Science > Atmosphere
  > Atmospheric Pressure > Sea Level Pressure'
  keywords_vocabulary = 'GCMD Science Keywords'
  keywords_vocabulary_url = 'https://gcmd.earthdata.nasa.gov/static/kms/'
  contact = 'Lionel MENARD, Raphael JOLIVET, Yves-Marie
SAINT-DRENAN, Philippe BLANC'
  institution = 'MINES Paris PSL'
  record = 'Basic measurements (global irradiance, direct
  irradiance, diffuse irradiance, air temperature, relative humidity, pressure)'
  Network_ShortName = 'BSRN'
  Network_Name = 'Baseline Surface Radiation Network'
  Network_Region = 'Global'
  Network_TimeResolution = '1 min'
  Network_Description = 'https://bsrn.awi.de/'
  Network_References = 'https://doi.org/10.5194/essd-10-1491-2018.'
  Network_License = 'https://bsrn.awi.de/data/conditions-of-data-
release/'
  Network_Contact = 'Dr. Amelie Driemel (https://www.awi.de/ueber-
  uns/organisation/mitarbeiter/amelie-driemel.html)'
  Station_WMO_ID = ''
  Station_Name = 'Carpentras'
  Station_Latitude = '44.083'
  Station_Longitude = '5.059'
  Station_Elevation = '100.0'
  Station_LocalTimeZone = 'UTC+01:00'
  Station_Address = '785, Chemin de l'Hermitage'
  Station_City = 'Carpentras, Carpentras'
  Station_Country = 'France'
  Station_SurfaceType = 'cultivated'
  Station_TopographyType = 'hilly'
  Station_Rural_Urban = 'rural'
```

```

Station_KoepenGeigerClimate = 'Csb'
Station_OperationStatus     = ''
Station_TimeResolution      = '1'
Station_DataEnd             = ''
Station_ContactPerson       = 'Thierry Duprat (thierry.duprat@meteo.fr)'
Station_Institute           = 'Meteo France'
Station_URL                  = ''
Station_CommissionDate      = ''
Station_DeCommissionDate    = '2018-12'
Station_Comment              = 'Closed in Dec 2018'
featureType                  = 'TimeSeries'
cdm_data_type                = 'TimeSeries'
Conventions                  = 'CF-1.6'
cdm_timeseries_variables    = 'station_name,latitude,longitude'
creation_date                = '2022-03-14T17:15:43.741057'
Station_ID                   = 'CAR'
Station_DataBegin           = '1996-08-01'
DODS.strlen                  = 10
DODS.dimName                 = 'ncshort'
DODS_EXTRA.Unlimited_Dimension = 'time'

Dimensions:
  time = 11790720 (UNLIMITED)
  maxStrlen64 = 64

Variables:
  station_name
    Size: 64x1
    Dimensions: maxStrlen64
    Datatype: char
    Attributes:
      long_name = 'station_name'
      cf_role = 'timeseries_id'
      _ChunkSizes = 10

  latitude
    Size: 1x1
    Dimensions:
    Datatype: single
    Attributes:
      long_name = 'station latitude'
      standard_name = 'latitude'
      units = 'degrees_north'
      _CoordinateAxisType = 'Lat'
      axis = 'Y'

  longitude
    Size: 1x1
    Dimensions:
    Datatype: single
    Attributes:
      long_name = 'station longitude'
      standard_name = 'longitude'
      units = 'degrees_east'
      _CoordinateAxisType = 'Lon'
      axis = 'X'

  elevation
    Size: 1x1
    Dimensions:
    Datatype: single
    Attributes:

```

```

        long_name      = 'Elevation above mean seal level'
        standard_name = 'height_above_mean_sea_level'
        units          = 'm'
        axis           = 'Z'
        comment        = ''

Time
  Size:      11790720x1
  Dimensions: time
  Datatype:  int32
  Attributes:
    parameter      = 'DATE/TIME'
    long_name      = 'Time of the end of period'
    standard_name  = 'time'
    abbreviation   = 'Date/Time'
    units          = 'seconds since 1970-01-01 00:00:00'
    time_origin    = '1970-01-01 00:00:00'
    resolution     = '1 min'
    time_zone      = 'UTC'
    axis           = 'T'
    calendar       = 'gregorian'
    _FillValue     = -1
    _Unsigned      = 'true'
    _ChunkSizes   = 1024

GHI
  Size:      11790720x1
  Dimensions: time
  Datatype:  single
  Attributes:
    parameter      = 'Short-wave downward (GLOBAL) radiation'
    long_name      = 'Global Horizontal Irradiance'
    standard_name  = 'surface_downwelling_shortwave_flux_in_air'
    abbreviation   = 'SWD'
    units          = 'W m-2'
    Range_LowerBoundary = -10
    Range_UpperBoundary = 3000
    Description    = 'global, mean'
    FirstData      = '1978-07-04T00:00'
    LastData       = '2018-12-31T23:59'
    _ChunkSizes   = 1024

P
  Size:      11790720x1
  Dimensions: time
  Datatype:  single
  Attributes:
    parameter      = 'Station pressure'
    long_name      = 'air pressure at station height'
    standard_name  = 'air_pressure'
    abbreviation   = 'PoPoPoPo'
    units          = 'Pa'
    Range_LowerBoundary = 0
    Range_UpperBoundary = 120000
    FirstData      = '1978-07-04T00:00'
    LastData       = '2018-12-31T23:59'

```

Get the data structure

```
nfo = ncinfo(file_nc)
```

```
nfo = struct with fields:
  Filename: 'http://tds_bsrn_user: [redacted]@tds.webservice-
energy.org/thredds/dodsC/bsrn-stations/BSRN-CAR.nc'
  Name: '/'
  Dimensions: [1x2 struct]
  Variables: [1x11 struct]
  Attributes: [1x48 struct]
  Groups: []
  Format: 'classic'
  Datatypes: []
```

Get time information (variable /Time with attribute)

```
nfo = ncinfo(file_nc);
nfo.Variables(5).Name
```

```
ans = 'Time'
```

```
nt = nfo.Variables(5).Dimensions.Length
```

```
nt = 11790720
```

```
nfo.Variables(5).Attributes
```

```
ans = 1x13 struct
```

Fields	Name	Value
1	'parameter'	'DATE/TIME'
2	'long_name'	'Time of the end of period'
3	'standard_name'	'time'
4	'abbreviation'	'Date/Time'
5	'units'	'seconds since 1970-01-01 00:00:00'
6	'time_origin'	'1970-01-01 00:00:00'
7	'resolution'	'1 min'
8	'time_zone'	'UTC'
9	'axis'	'T'
10	'calendar'	'gregorian'
11	'_FillValue'	-1
12	'_Unsigned'	'true'
13	'_ChunkSizes'	1024


```
strAttribute = nfo.Variables(5).Attributes(5);
disp(strAttribute)
```

```
Name: 'units'
Value: 'seconds since 1970-01-01 00:00:00'
```

```
% begin - end
time_begin = datetime(ncread(file_nc, '/Time', 1, 1), 'ConvertFrom', 'posixtime')
```

```
time_begin = datetime
01-Aug-1996
```

```
time_end = datetime(ncread(file_nc, '/Time', nt, 1), 'ConvertFrom', 'posixtime')
```

```
time_end = datetime
31-Dec-2018 23:59:00
```

```
% resolution
time_begin_dt =
datetime(ncread(file_nc, '/Time', 2, 1), 'ConvertFrom', 'posixtime');
dt = (time_begin_dt - time_begin)
```

```
dt = duration
00:01:00
```

```
vdn = datenum(time_begin) + [0:nt-1]*datenum(dt);
datestr(vdn(1))
```

```
ans = '01-Aug-1996'
```

```
datestr(vdn(end))
```

```
ans = '31-Dec-2018 23:59:00'
```

```
% Get index time for the day 2010-06-01T00:01:00 - 2010-06-02T00:00:00
```

```
idx = find((vdn >= datenum(2010, 07, 01, 0, 1, 0)) & (vdn <= datenum(2010, 07, 02, 0, 0, 0)));
vt = datetime(vdn(idx), 'ConvertFrom', 'datenum');
vt(1)
```

```
ans = datetime
01-Jul-2010 00:01:00
```

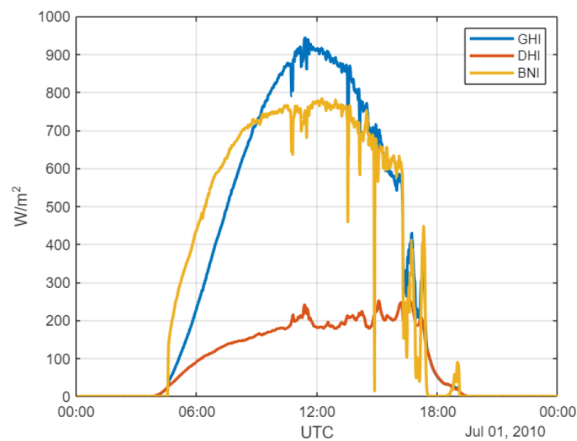
```
vt(end)
```

```
ans = datetime
02-Jul-2010
```

Get GHI, DHI and BNI for that specific day

```
vGHI = ncread(file_nc, '/GHI', idx(1), length(idx));
vDHI = ncread(file_nc, '/DHI', idx(1), length(idx));
vBNI = ncread(file_nc, '/BNI', idx(1), length(idx));

hp = plot(vt, vGHI, vt, vDHI, vt, vBNI, 'LineWidth', 2);
xlabel('UTC')
ylabel('W/m^2')
grid
legend('GHI', 'DHI', 'BNI')
```



5.3. Example of QC in python (Jupyter Notebook)

In order fully address open and interoperable approach, we have release free access to codes exemplifying the benefit of remote access thanks to TDS. The *in-situ* measurement that are handled in this application are directly extracted on-the-fly from TDS. Those codes written in Python have been deployed on a public GitHub (https://github.com/oie-mines-paristech/IEA_PVPS_T16_QC_pynb) and can be directly executed via a Binder. Binder is an online service for building and sharing reproducible and interactive computational environments from online repositories such as GitHub. In one single click you are provided with a fully functional Jupyter environment to interact with the proposed code repository. The following screenshots illustrate such workflow.

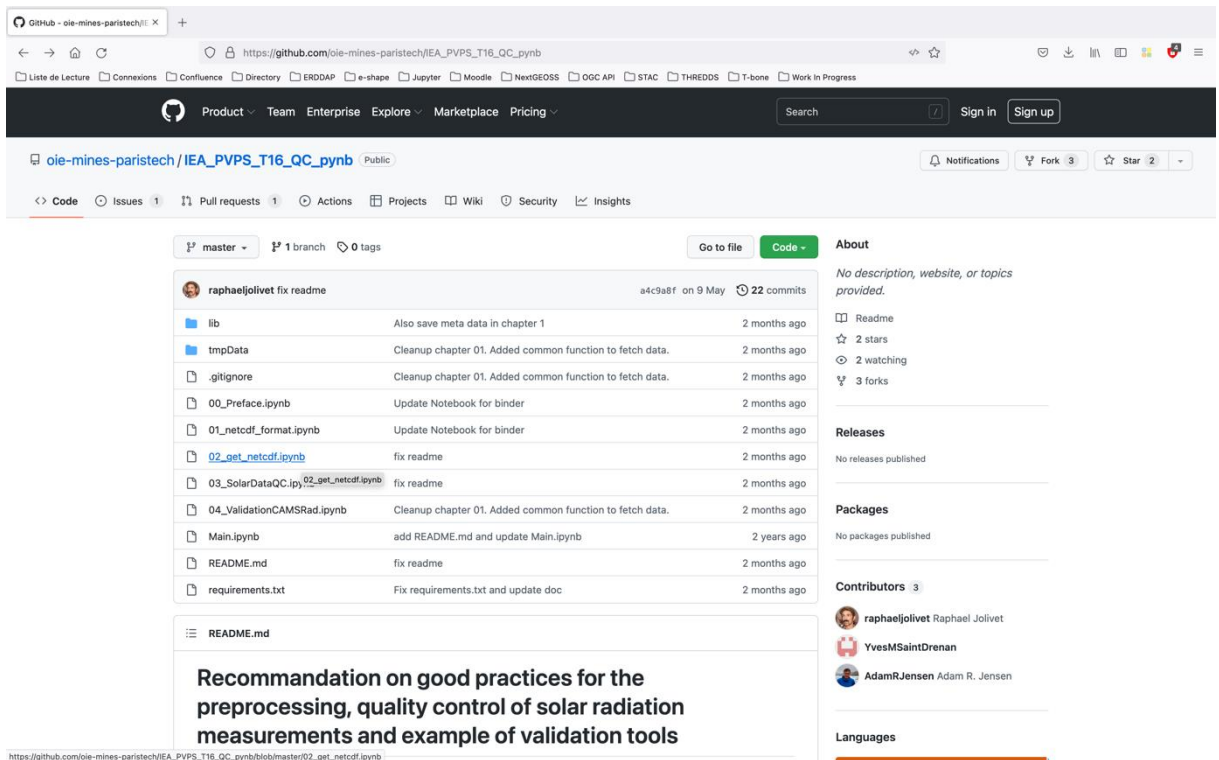


Figure 10: GitHub - Home of the repository application

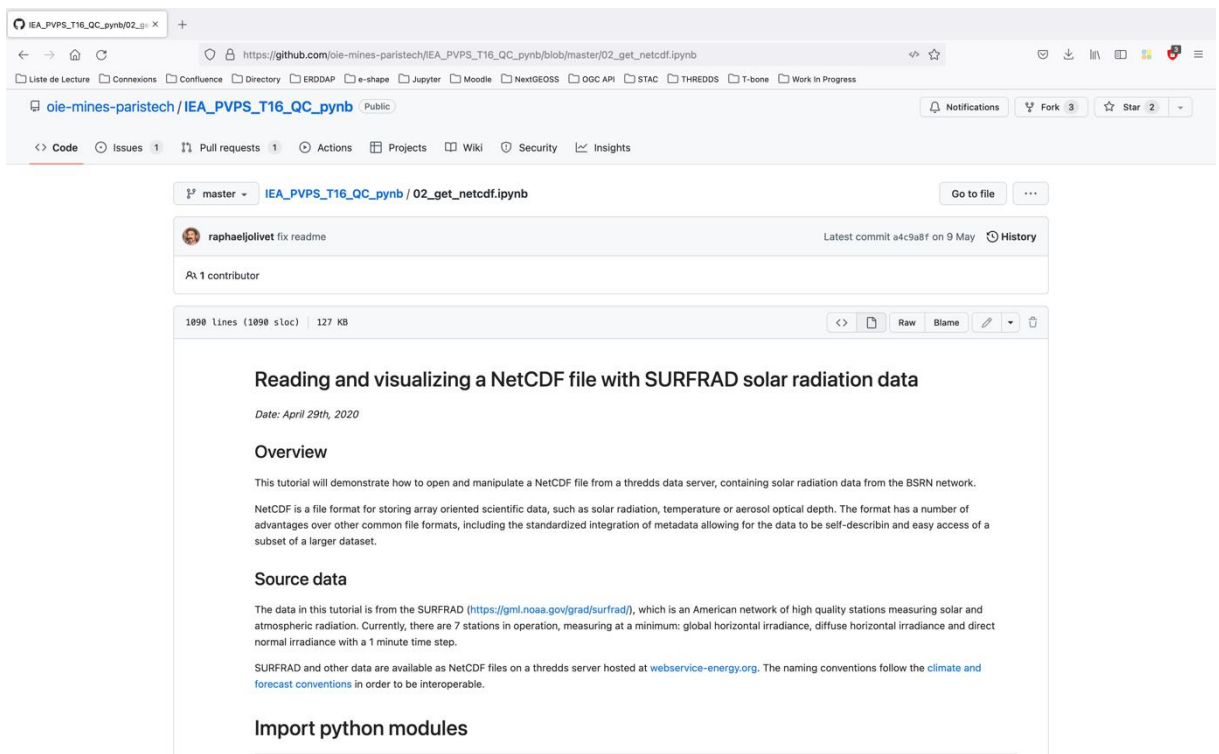


Figure 11: GitHub – Instruction about how to access to SURFRAD solar radiation data from TDS

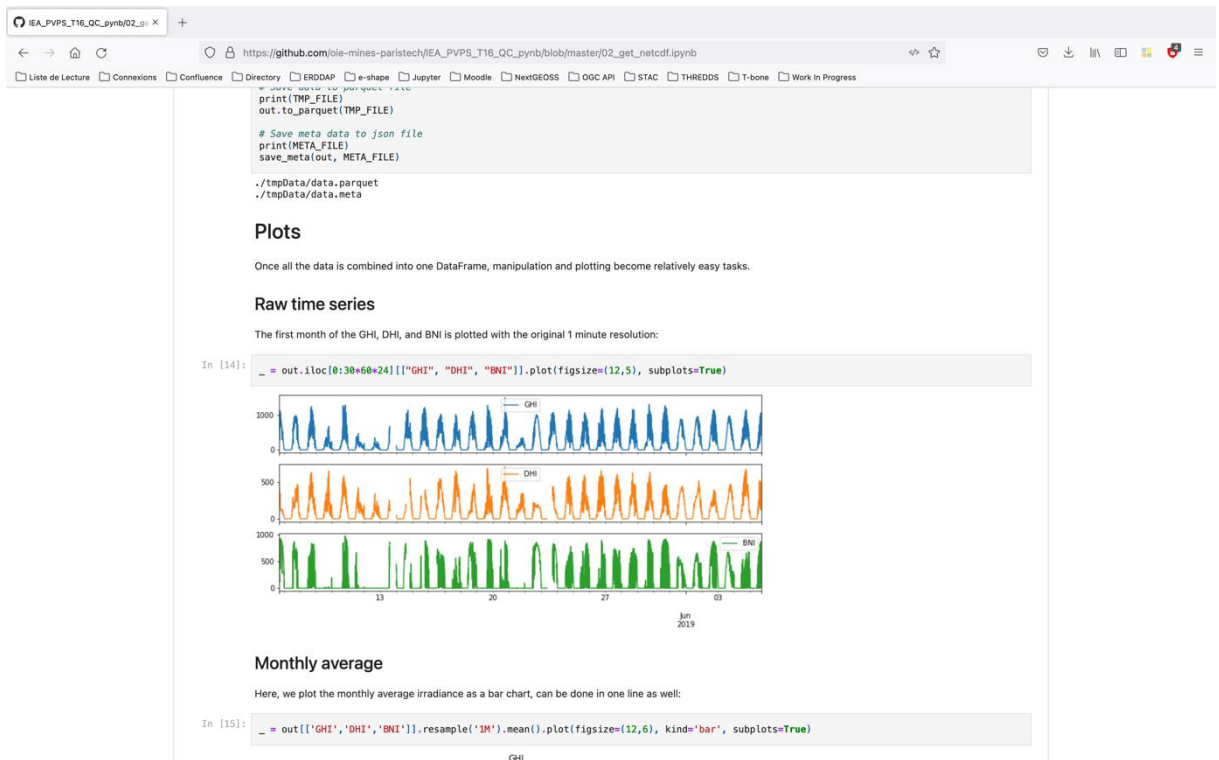


Figure 12: GitHub - Portrayal of the output from the Jupyter environment to support documentation

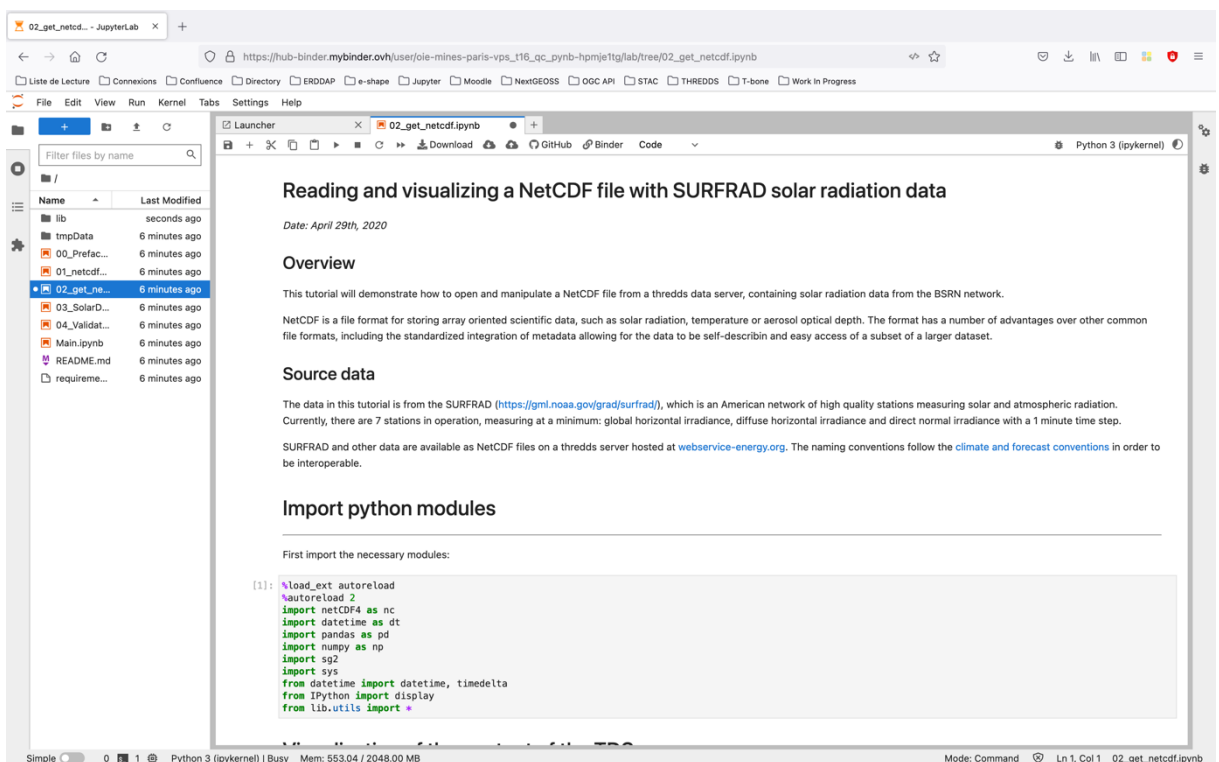


Figure 13: MyBinder – Jupyter environment generated from the GitHub repository

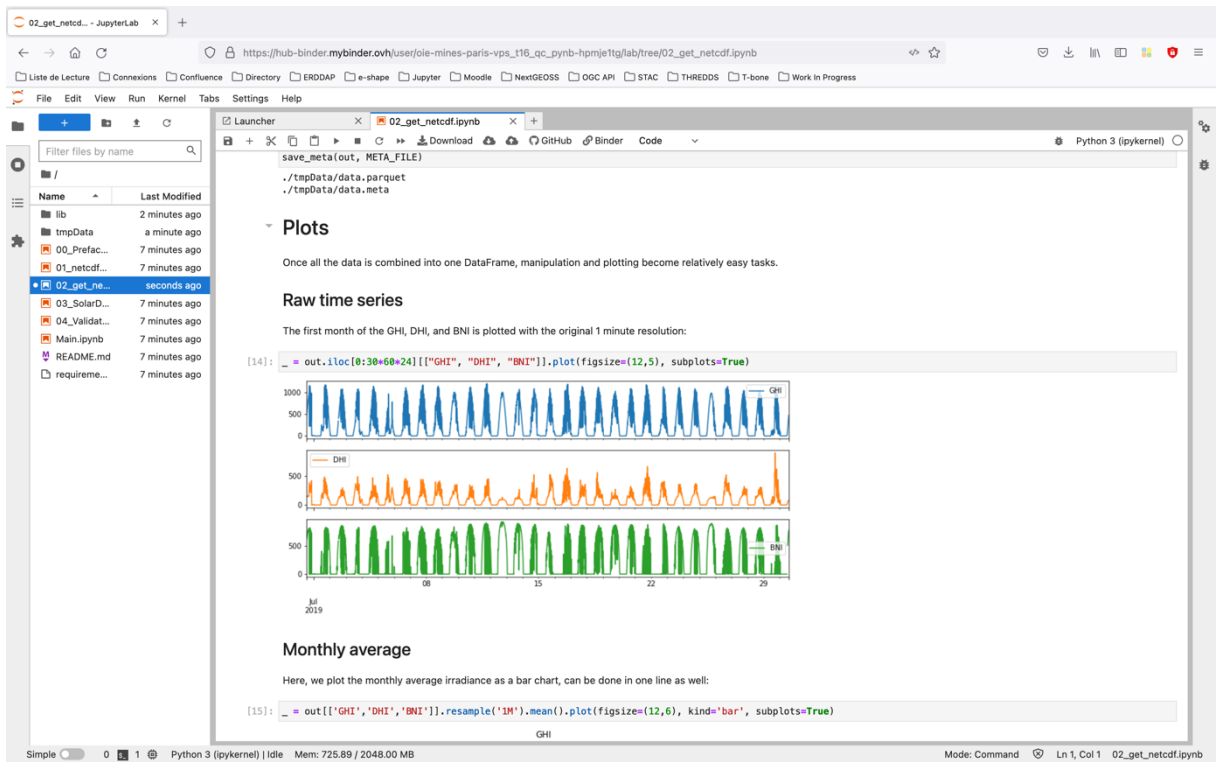


Figure 14: MyBinder - Generation of the plots from the Jupyter environment (1)

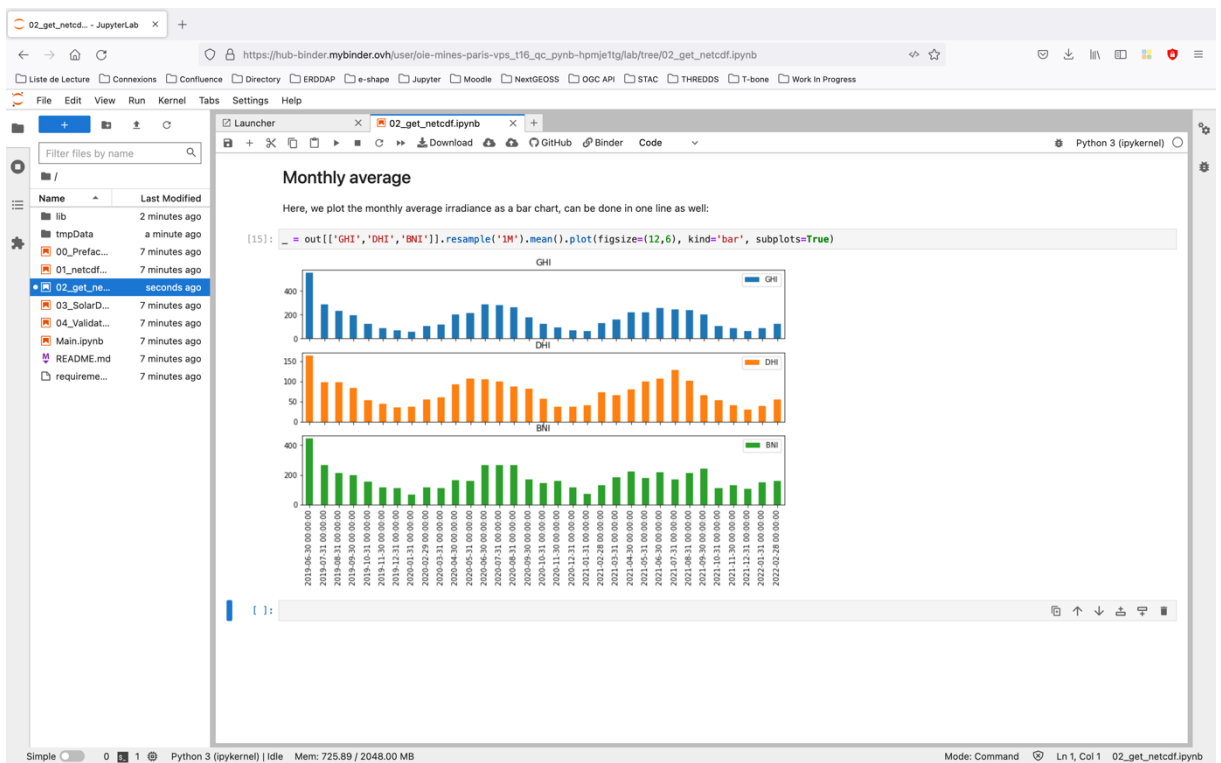


Figure 15: : MyBinder - Generation of the plots from the Jupyter environment (2)

5.1. Implementation of a visual plausibility control tool

The harmonization of the data has made possible to implement complex analysis to the *in-situ* measurement. An example is given in Figure 16, where a screen copy of the output of a visual plausibility control tool is given. This tool includes different visual inspection of the data compiled on a single page. The strength of this approach is that the origin of potential suspicious behavior of the measurements can be identified by observing the data in different representations. Without going into the details of this analysis, the example given in Figure 16 illustrates a GHI or DHI calibration problem (see lower plots of the left panel). This analysis allows detecting further faults or issues such as shadow, unshaded DHI sensor, temporal reference issues...

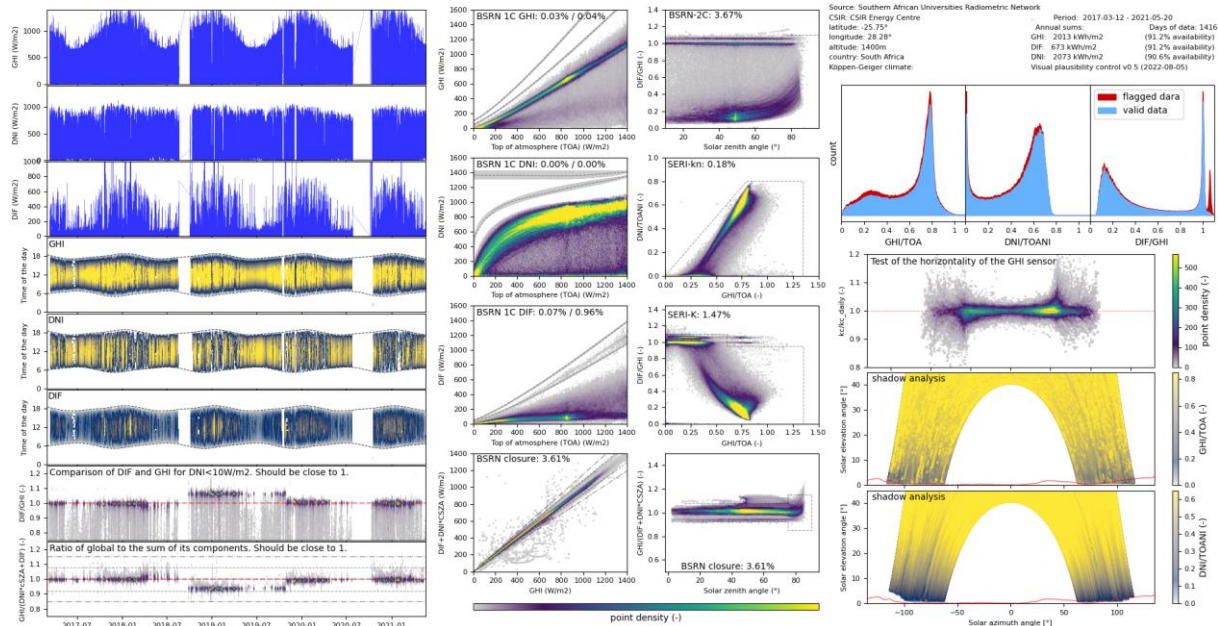


Figure 16: visual plausibility control tool applied to a station of the ESMAP network

Having data satisfying the FAIR principles has made the implementation of this complex analysis straightforward because once developed on the chosen data format the visual control procedure could be implemented easily on more than 100 stations. The resulting plots can be visualized on the web interface described in section 5.2.

5.2. Web-based application (The “wow effect”)

Recently, more and more development frameworks and APIs have arisen, and they’ve proven to be a very good complement to enable a wider and user-centric way for the dissemination of scientific data. In our environment as powered by the SDI hosting and providing open and interoperable access to our *in-situ* measurements, a dedicated web-based application has been developed. It is based on Flask. Flask is a micro web framework for creating APIs in Python. The application has taken the advantage of the open access to *in-situ* measurements provided by TDS in conjunction with the out-of-the-box DAP protocol to handle and display data and metadata. As a result, a very powerful yet simple to use web application has been released. All requests from this application are directly made on-the-fly on TDS via the DAP protocol. A first set of *in-situ* measurements networks have been hosted on TDS and are nicely displayed on a world map on the web applications. All operations including metadata and data display as well as export are made on the fly via the DAP protocol. It worth mentioning that due to access right restrictions some networks are password protected. The application is available here: <http://viewer.webservice-energy.org/in-situ/>

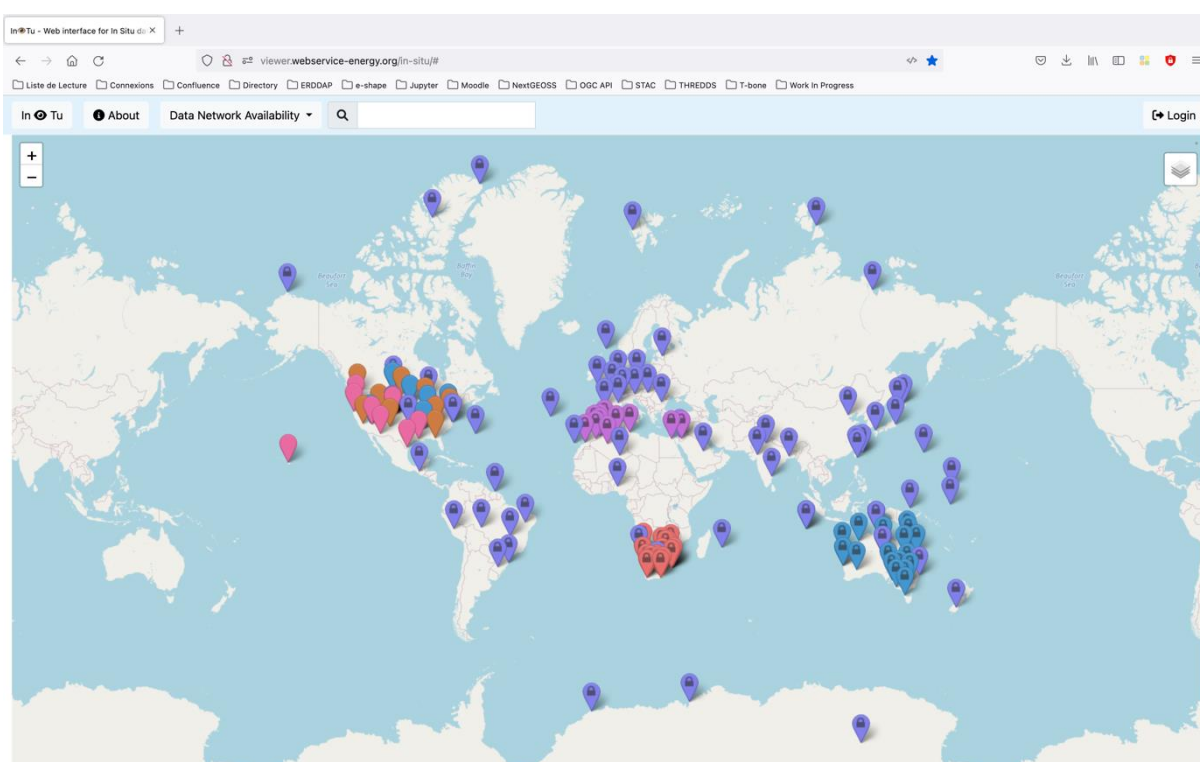


Figure 17: Viewer - Home page of the in-situ viewer

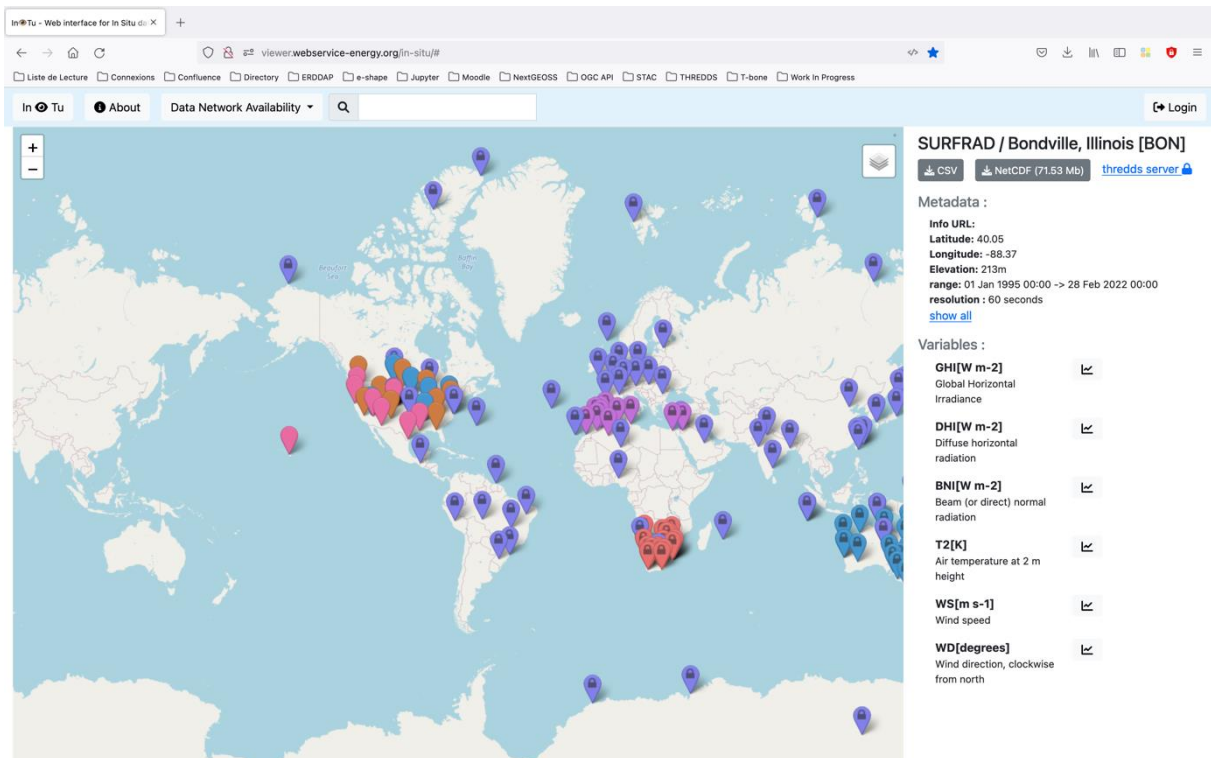


Figure 18: Viewer - General information of the SURFRAD Bondville, Illinois station extracted from the TDS NetCDF file

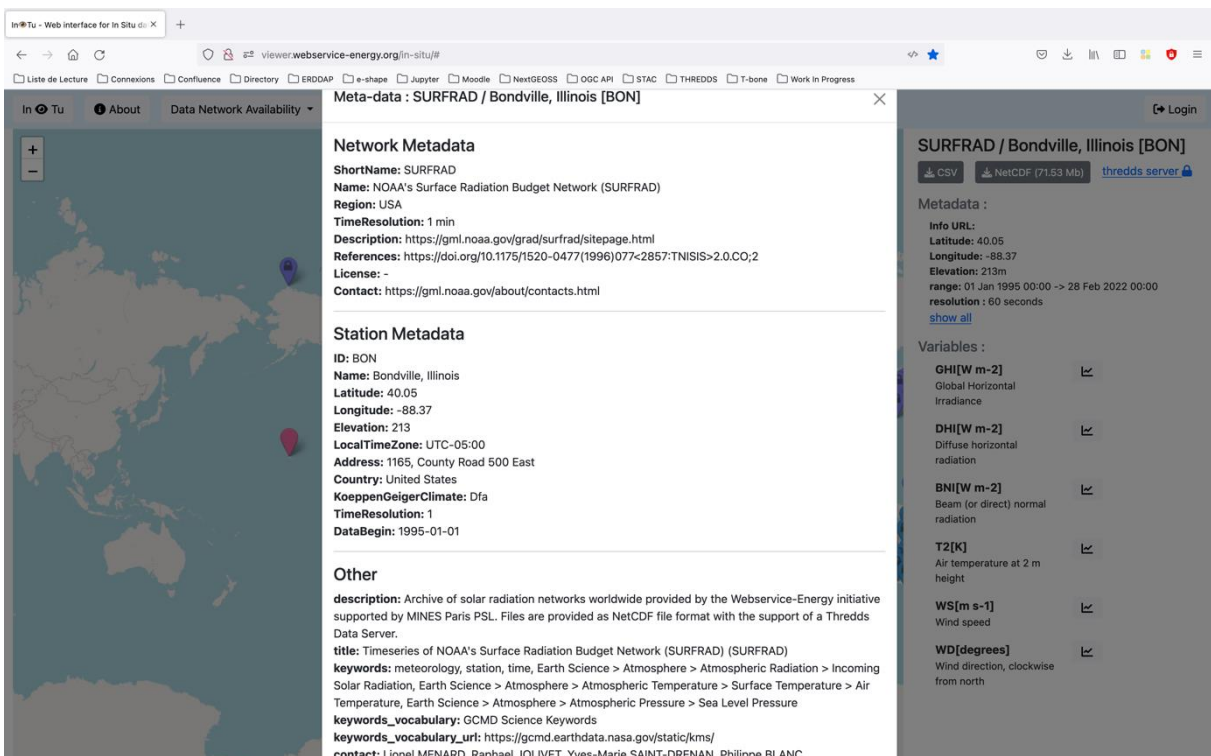


Figure 19: Viewer - Detailed information of the SURFRAD Bondville, Illinois station extracted from the TDS NetCDF file

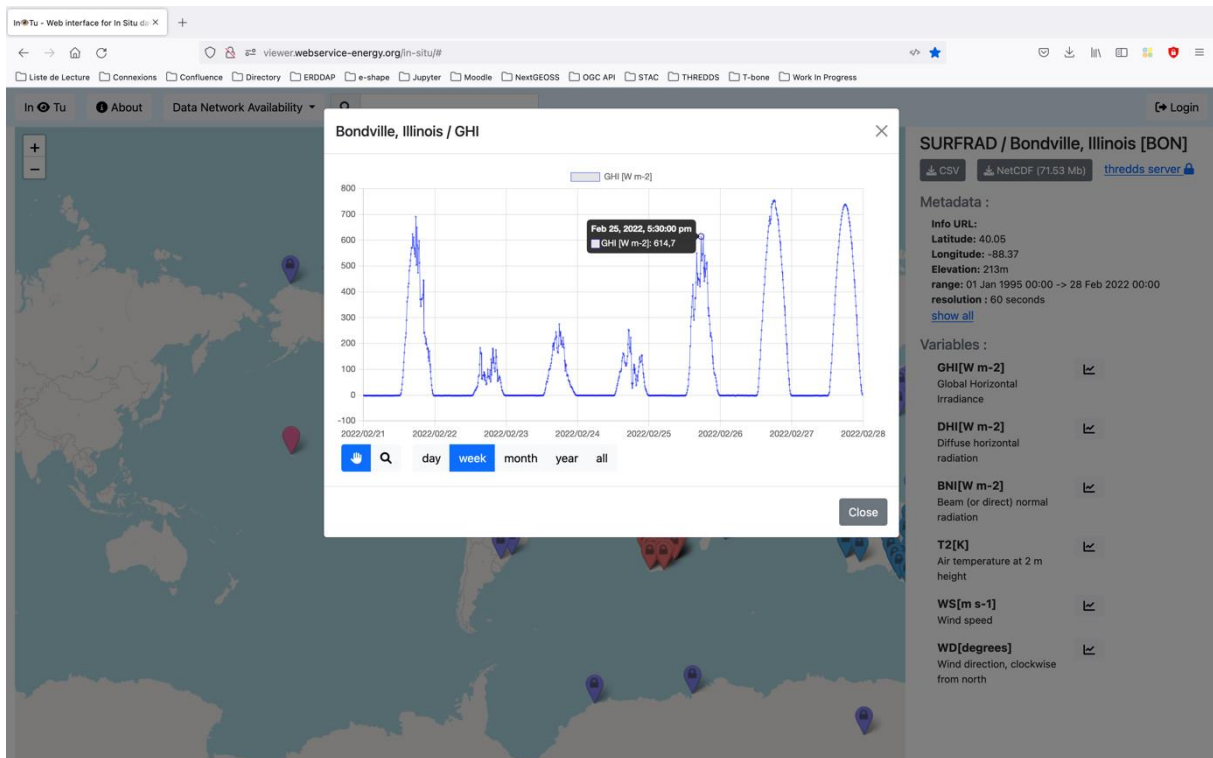


Figure 20: Viewer - Chart display allowing easy navigation over time-series and pop-up display of the values

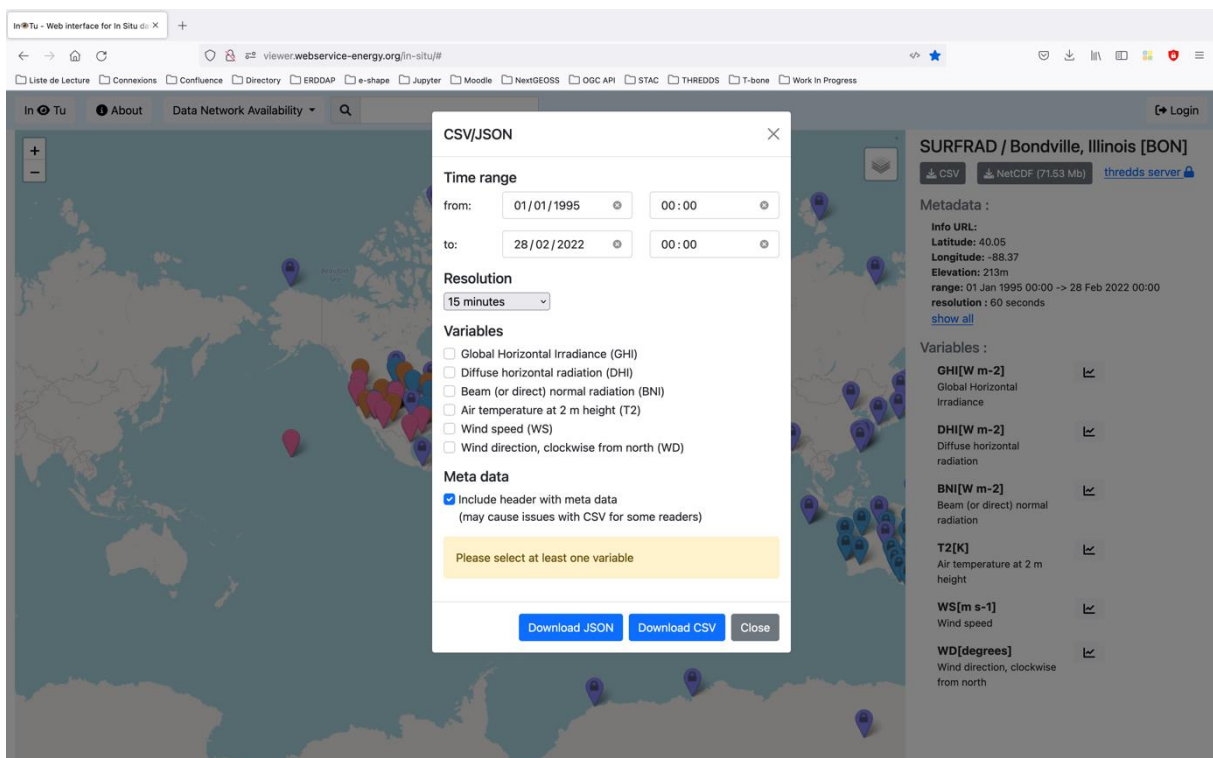


Figure 21: Viewer - Export function including CSV with additional information from metadata and variables filtering

6. Search and discovery with webservice-energy catalogue

Making the metadata, data and application searchable and discoverable is a key pillar of FAIR. To make it possible we have conducted a series of action to enable the widest dissemination possible while respecting an open and standard approach supporting interoperability.

We started by creating a dedicated metadata record in the catalogue component of our webservice-energy SDI. The metadata record is available here: <https://tinyurl.com/2s4thwsx>

It provides in a standard ISO 19139 metadata profile the needed information and key links to the application and the TDS.

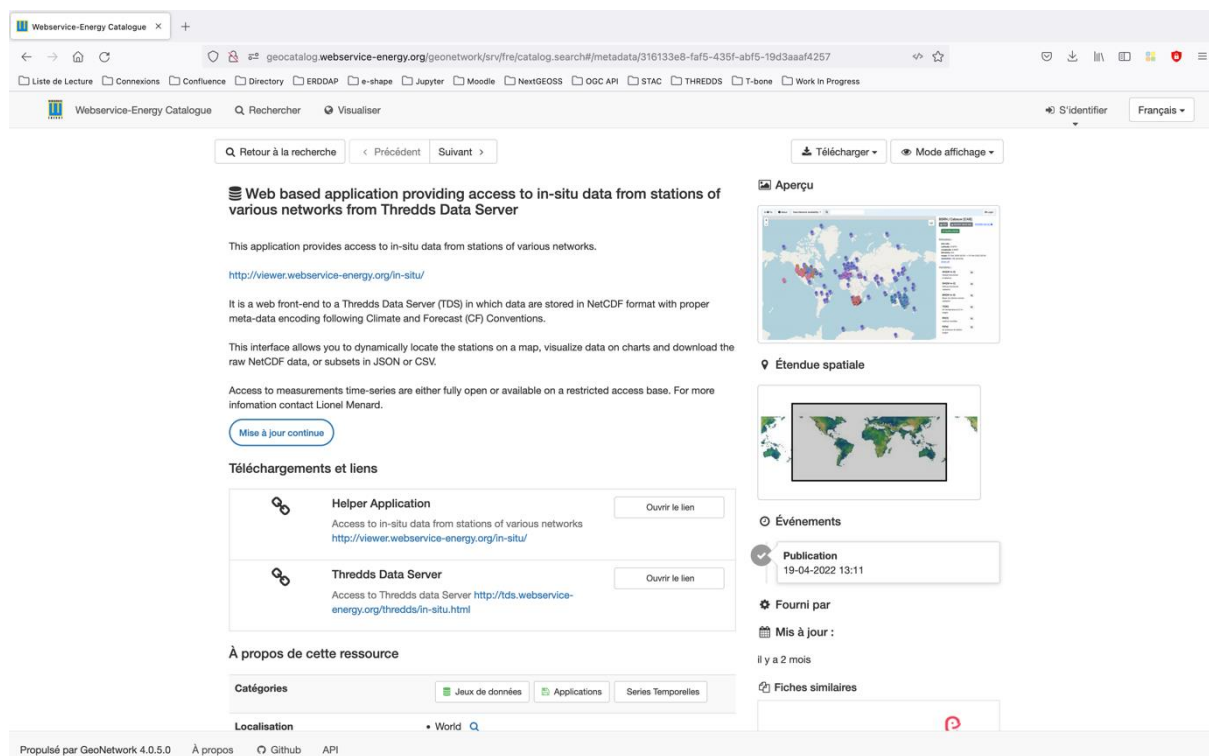


Figure 22: Catalogue - ISO 19139 metadata record providing necessary information and links to TDS and web application

The webservice-energy catalogue is part of a bigger network namely the GEO (Group on Earth Observation). GEO is gathering information from over 150 national and regional providers such as NASA and ESA; international organizations such as WMO and the commercial sector such as Digital Globe. As such the webservice-energy catalogue is weekly harvested thanks to the OGC CSW (Catalog Service for the Web) standard protocol by the GEO DAB (Discovery and Access Broker) one of the main component of the GEO infrastructure. The result of the harvesting process is made available without any further effort on the GEO Web Portal (GWP) ensuring a wider visibility of the resources coming from the webservice-energy to the GEO community and solar energy practitioners. The following screenshots illustrates the rendering of such harvesting when performing a search query on the GWP.

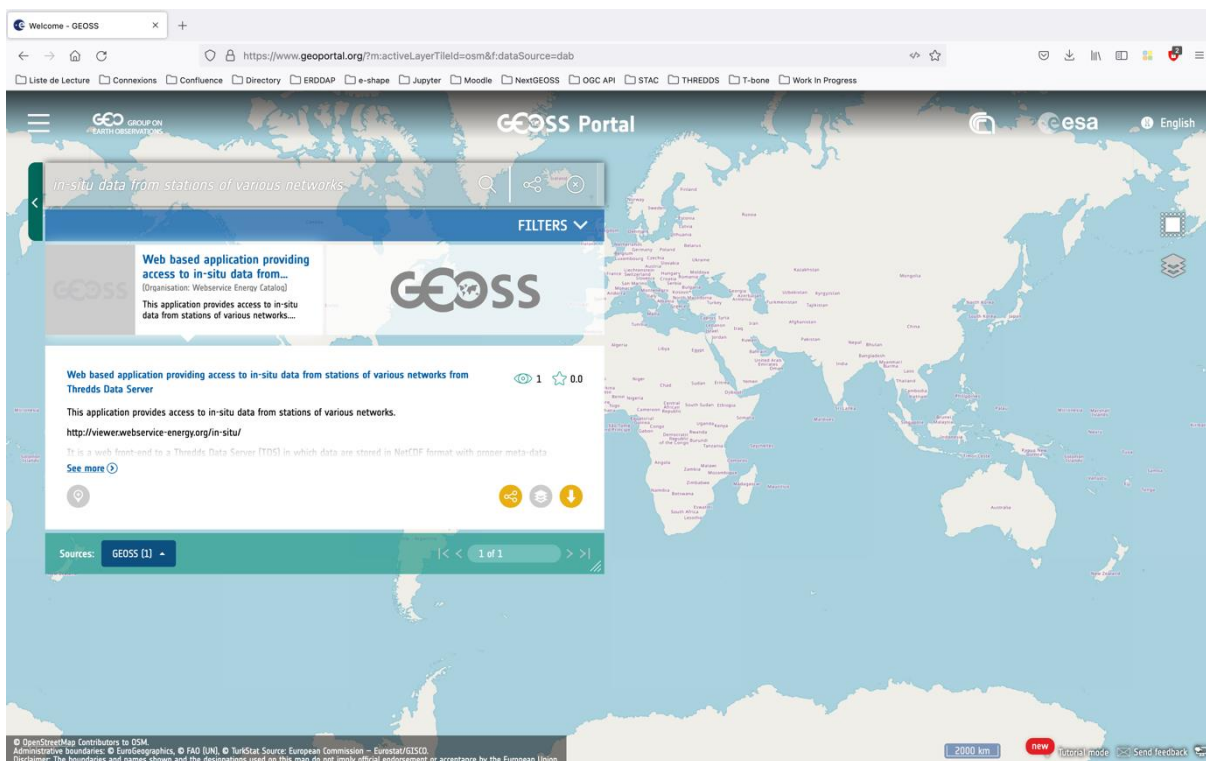


Figure 23: GWP - Keyword search for in-situ data in the GWP and display of the corresponding metadata record

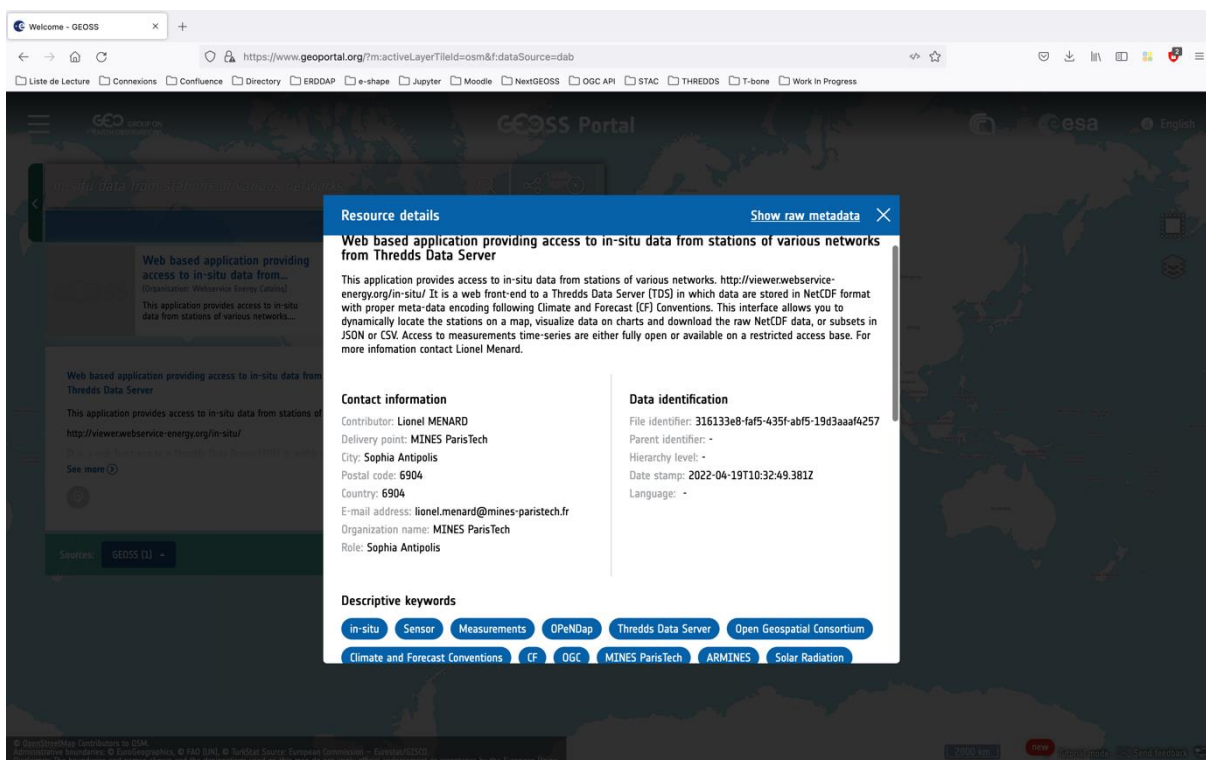


Figure 24: GWP - Detailed information of the metadata record demonstrating consistency regarding the original file

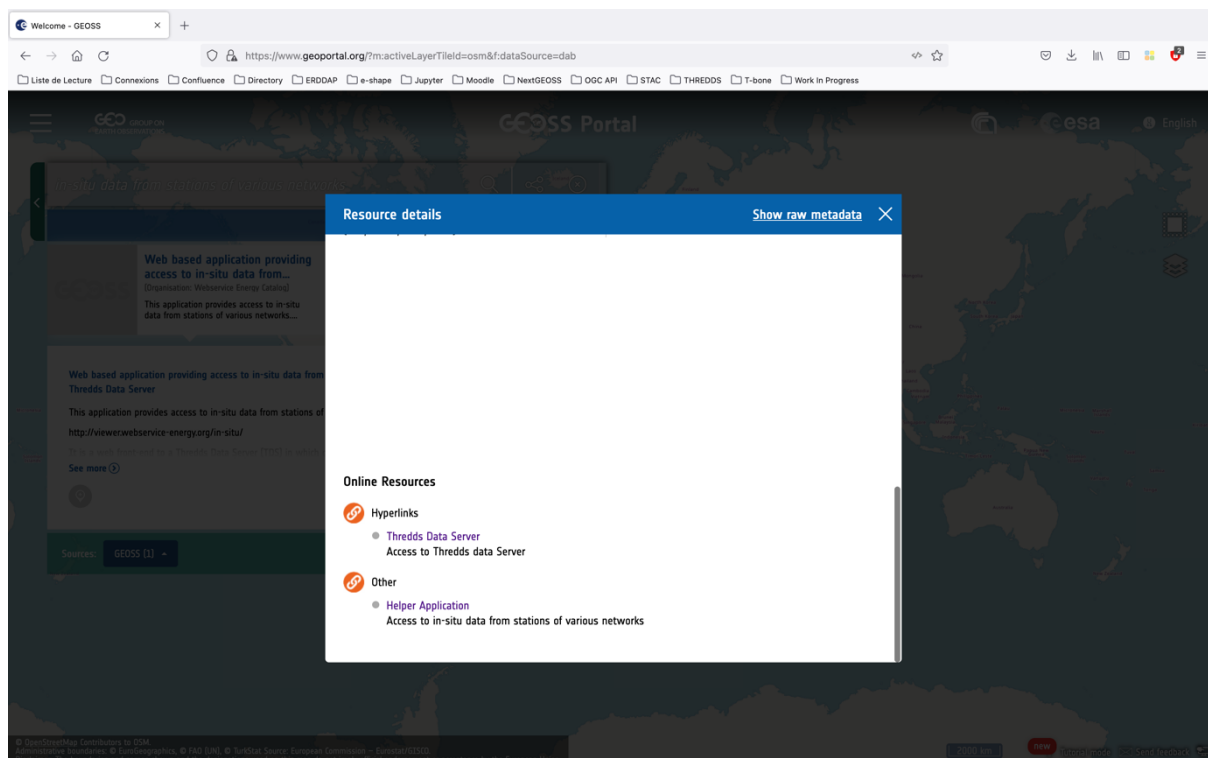


Figure 25: GWP - Consistency follow-up regarding the links to TDS and to the web application

In addition to the creation of the metadata record presenting the web-based application, extra efforts have been made to enable the dissemination of the granular NetCDF file from TDS on the GWP. A weekly harvest of webservice-energy TDS from the GEO DAB has been set up and consequently such files are searchable and discoverable on the GWP. This finalizes the effort to address the F (Findable) of our RAIF implementation.

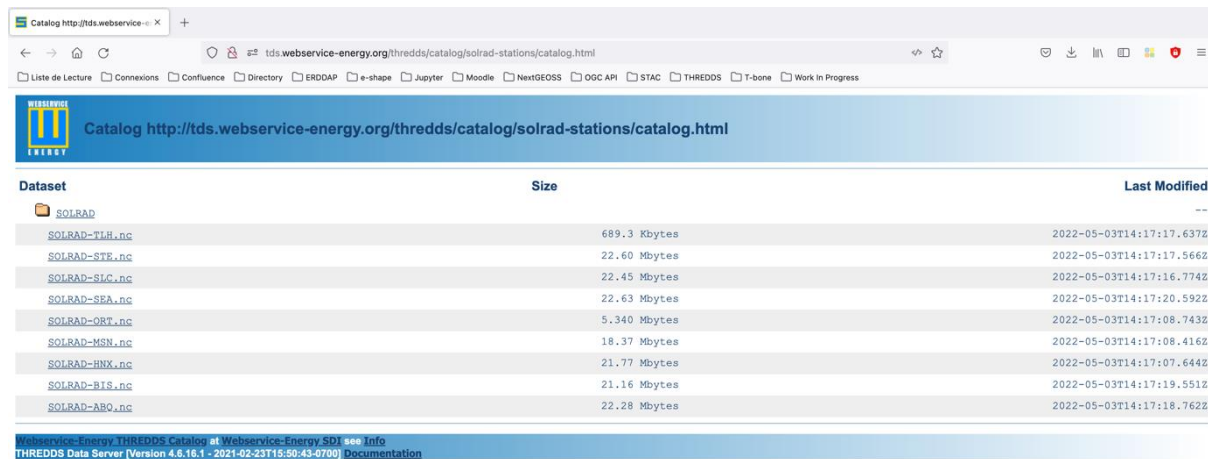


Figure 26: TDS - SOLRAD archive on webservice-energy TDS

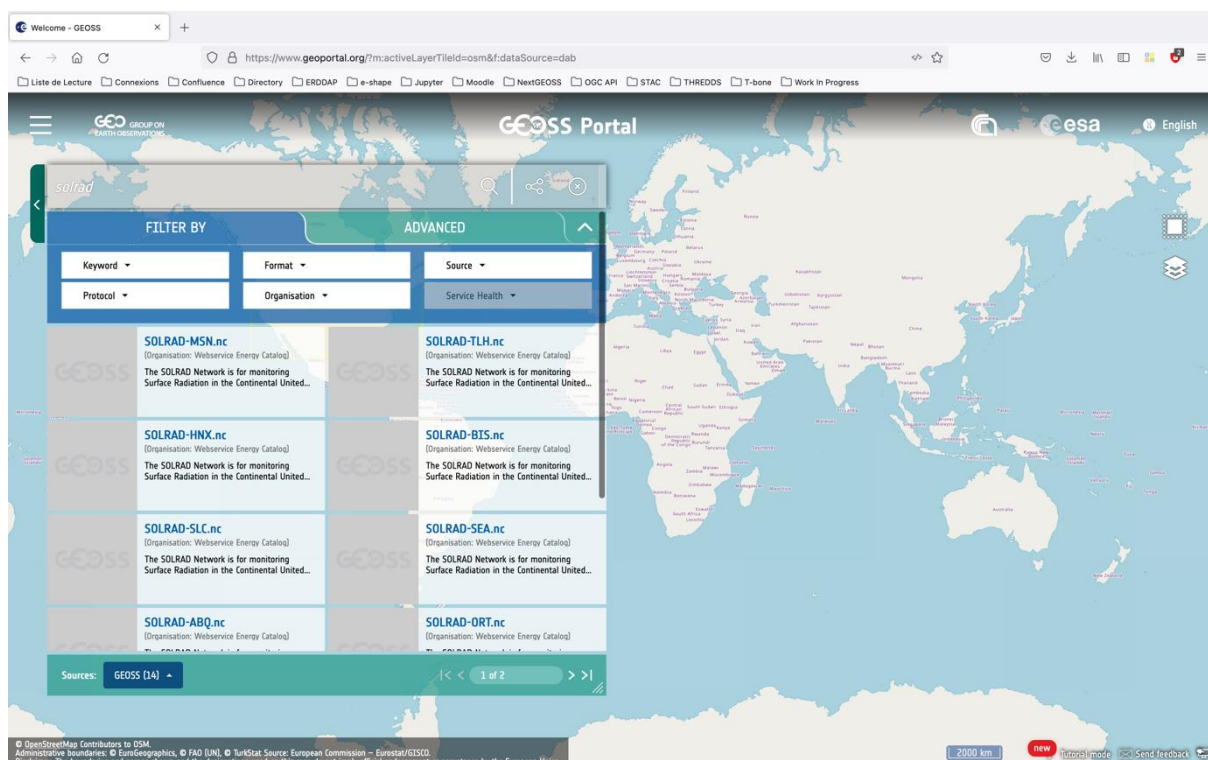


Figure 27: GWP – SOLRAD archive search results from the GWP

7. Conclusion and perspectives

This working paper demonstrates through a real case scenario a complete journey of moving from raw *in-situ* measurement coming from ground sensors up to full standard and interoperable datasets available on the web. The work carried out in this paper is easily replicable. Moreover all infrastructures and software components supporting the various RAIF actions are open and free of charge with no vendor locking issues. Beside some restrictions enforced by some networks policies to restrict access to their *in-situ* measurements, efforts are made to make access to those dataset on an open and free basis while applying the same thorough dissemination methodology independently of the policy status. We believe that this practical example could be mimic by many others EO communities dealing with *in-situ* measurements.

NetCDF encoding and CF metadata definition are key components in our dissemination workflow. Both have a potential direct impact regarding how the various downstream applications behave when it comes to consume such data encoded in NetCDF and supercharged with CF Convention metadata vocabularies. For CF as presented in this document, typical solar energy vocabularies has been added to support solar energy practitioner understanding of some key phenomenon (eg. GHI, DHI,...). Some dialogue will be necessary with CF community to allow such vocabularies to map existing definitions as currently implemented in CF.

Current works focus on the use of standard tool that benefit from the harmonized format of the data: quality control, visual analysis, inter-network validation and calibration. The present work has been shared with the solar energy and meteorological community to try to reach a consensus on the format and description of the data.

8. Annex: presentations done with PVPS T16

Lessons learned from an extensive visual inspection of solar radiation measurements, Yves-Marie Saint-Drenan (MinesParis-PSL), Adam Rasmus Jensen (DTU), BSRN workshop (June 2022 JRC, ISPRA)
<https://cloud.minesparis.psl.eu/index.php/s/8lHyHvNpv6CqdA3>

An operational processing chain for the collection, harmonization and visualization of *in-situ* measurements following FAIR principles, Philippe Blanc, Raphaël Jolivet, Lionel Ménard, Yves-Marie Saint-Drenan, Task 16 Expert Meeting #10 (April 2022, Sophia Antipolis)
<https://cloud.minesparis.psl.eu/index.php/s/pFYUSiOSwqLfm97>

Climate & Forecast, Netcdf, Thredds and ERDDAP for *in-situ* pyranometric measurements, Philippe Blanc, Raphaël Jolivet, Lionel Ménard, Yves-Marie Saint-Drenan, Task 16 Expert Meeting #8 (April 2021, Sophia Antipolis)
<https://cloud.minesparis.psl.eu/index.php/s/4jArA76SSbJczlB>

The perspective of Netcdf as a new data format for time series of SSI: the end of ASCII files? Lionel Menard, Philippe Blanc, Task 16 Expert Meeting #2 (March 2018, Paris)
<https://cloud.minesparis.psl.eu/index.php/s/2X6S5wIivQvY3vs>

9. Annex: libinsitu resources

The following resources are available online:

Source code of libinsitu

The code is available on our gitlab :
<https://git.sophia.mines-paristech.fr/oie/libinsitu>

The code includes a NetCDF schema (CDL) :
<https://git.sophia.mines-paristech.fr/oie/libinsitu/-/blob/main/libinsitu/res/base.cdl>

The code includes consolidated and enriched CSV metadata for several networks. We welcome any contribution on those files :
<https://git.sophia.mines-paristech.fr/oie/libinsitu/-/tree/main/libinsitu/res/station-info>

Documentation

The documentation of the library (both the command line utilities and Python API) is available online :
<https://libinsitu.readthedocs.io/en/latest/>

The documentation also includes the proposed conventions for the CF-NetCDF format:
<https://libinsitu.readthedocs.io/en/latest/conventions.html>

Mailing list

We invite contributors to join the public mailing-list solar-insitu@groupes.mines-paristech.fr to discuss the library or the NetCDF conventions :
<https://groupes.minesparis.psl.eu/www/info/solar-insitu>