https://doi.org/10.5194/essd-2023-334
Preprint. Discussion started: 22 September 2023

Enabling FAIR Certification for Micrometeorological Datasets

Mark Roantree[1], Stevan Savić[3], Michael Scriney[1], and Bransilava Lalić[2]

[1]Insight Centre for Data Analytics, School of Computing, Dublin City University, Ireland.
[2]Faculty of Agriculture, University of Novi Sad, Serbia.
[3]Faculty of Sciences, University of Novi Sad, Serbia.

Correspondence: Mark Roantree (mark.roantree@dcu.ie)

Abstract. The current state of weather-induced agricultural losses, water use for irrigation, the appearance of new invasive species and disease vectors, new environmental zoning of plant diseases and pests, deforestation, increased urbanization, rural-to-urban migration and increased urban energy consumption for cooling and heating, together impose a scientific demand for FAIR micrometeorological data. FAIR data and metadata should be easily discoverable by humans or machines, accessible under specific conditions or restrictions, conform to recognized formats and standards to be combined and exchanged, and licensed according to community norms, allowing users to know what kinds of reuse are permitted. However, the lack of FAIR data costs Europe a minimum of €10.2bn per year or approximately 78% of the Horizon annual 2020 budget. If data met the FAIR principle, it would improve data discovery and access, enable re-use, enhance understanding, especially across domains, reach as many people as possible, be cited more often, and open new routes to build cooperation. To support owners of micrometeorological data to make their data FAIR, the FAIR Micromet Portal was developed within the CA20108 COST Action to guide owners through FAIR principles, in a step-by-step manner, with the ultimate goal of making large volumes of data FAIR. This paper provides a detailed discussion on how this is achieved by validating micrometeorological data stored on the FAIR Micromet Portal against the full set of FAIR metrics.

1 Introduction

The deployment of sensing technologies have changed both the everyday life of human beings and the manner in which data can be acquired across a wide range of scientific and practical domains. Sensors are devices that detect changes in the health and performance of human beings, smart machines, buildings and cities, manufacturing, daily life and the measurement of the environment and climate. In healthcare, there has been considerable research in the usage of sensors with a detailed survey in (Rodolfo et. al. 2019); many examples can be found on the usage of sensors to drive performance in sport with recent examples of the adoption of machine learning (De Beéck et. al. 2018); smart cities (Santana et. al. 2017); and in climate, the capture of micrometeorological data using a range on ongoing sensors has been in place for a considerable number of years (World Meteorological Organisation 2021). A more general recent overview can be found in (Javaid et. al. 2021).

https://doi.org/10.5194/essd-2023-334
Preprint. Discussion started: 22 September 2023

In this research, we focus on the generation and acquisition of climate data and in accelerating its usage by adopting FAIR principles. The current state of weather-induced agricultural losses, water use for irrigation, the appearance of new invasive species and disease vectors (strongly depending on micrometeorological conditions), new environmental zoning of plant diseases and pests, deforestation, increased urbanization, rural-to-urban migration and increased urban energy consumption for cooling and heating impose scientific and societal demands for FAIR micrometeorological data. It is important to highlight the FAIR acronym for: Findability, Accessibility, Interoperability, and Reusability. This means that data and metadata should: be easily discoverable by humans or machines; accessible under specific conditions or restrictions; conform to recognized formats and standards to be combined and exchanged; and licensed according to community norms allowing users to know what kinds of reuse are permitted. While open data is the ultimate goal, it is important to have in mind that the FAIR concept implies open metadata *only*. Measurement results should be stored on a repository chosen by the data owner with a DOI and preferred licence, from closed to fully open with numerous options.

There is a distinct difference between open and FAIR data, related to the degree of accessibility and requirements for usability (Mons et. al. 2017). Open data is available without restriction while FAIR data may have specific conditions for access and usage. Open Government Data refers to the information collected, produced or paid for by the public bodies and made freely available for reuse for any purpose (Europa.eu 2023), with a licence specifying the terms of use. These principles for Open Data are described in detail in the Open Definition. Public sector information is information held by the public sector. The Directive on the re-use of public sector information provides a common legal framework for a European market for government-held data. It is built around the key pillars of the internal market: free flow of data, transparency and fair competition. It is important to note that not all of the public sector information is Open Data.

A recent report (PricewaterhouseCoopers 2018) on the cost of research that was non-FAIR compliant reached a conservative estimate of €10.2bn or 3% of all EU research expenditure given the lack of FAIR data. This report, targeted at research funders, data and related infrastructures and research organisations, identified the impact of research activities as most significant but also highlighted its impact on collaboration and innovation. In general, increased time and cost is repeatedly cited as the main negative impacts. This report provides detail on the different cost indicators: time, cost of storage, licence costs, research retraction, double funding, interdisciplinarity and potential economic growth. As the research presented in this paper emerges from EU Cost Action CA20108, an interdisciplinary action with climate researchers from different domains, data engineers and machine learning researchers, an interesting finding was that the cost of its effect on *interdisciplinary* research was difficult to estimate. However, findings indicate that: reproducibility is hampered if data is not FAIR; lack of access to, and the quality of data restricts inter-disciplinarity; and the benefit of accessing "disparate data from other disciplines" is lost for these teams.

This cost to research of non-FAIR data can potentially be exacerbated when dealing with high value data. For example, in the high value datasets identified by the EU (EUR-Lex 2022), the meteorological thematic category

https://doi.org/10.5194/essd-2023-334
Preprint. Discussion started: 22 September 2023

Open Access Earth System Science Data Discussions

includes datasets on observational data measured by weather stations, validated observations (climate data), weather alerts, radar data and numerical weather prediction (NWP) model data with the granularity and key metadata attributes listed in table 1.

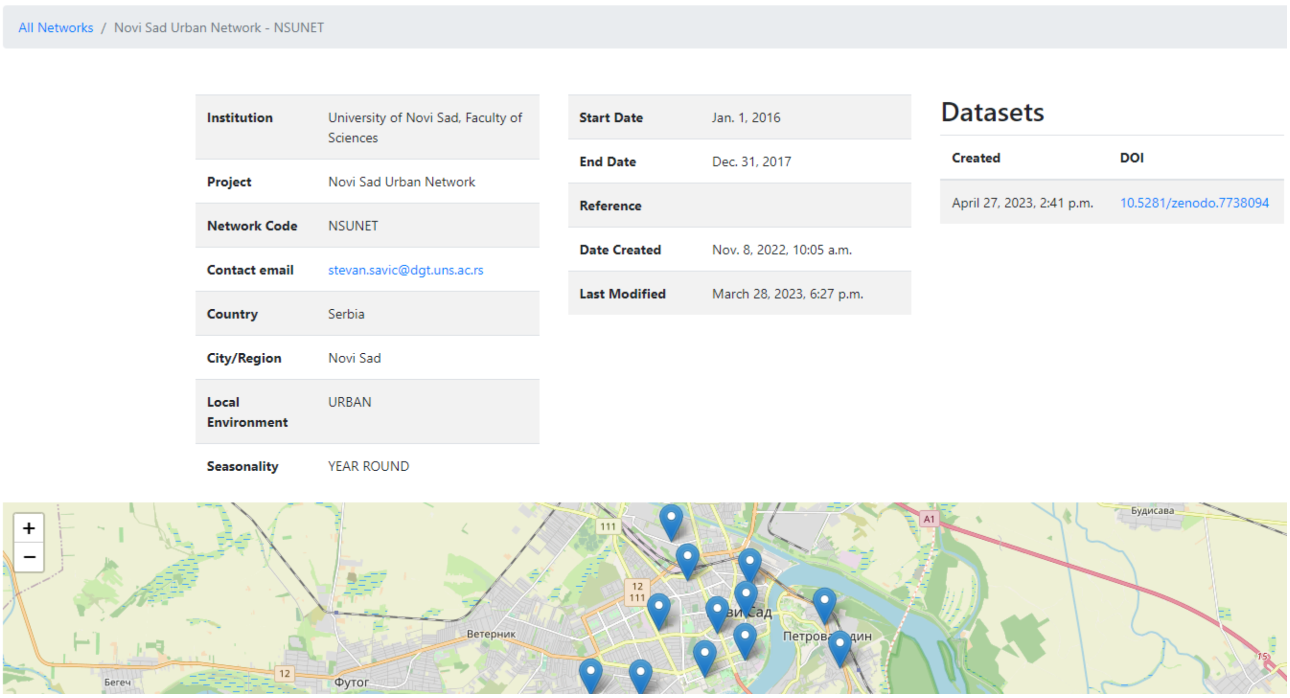


Figure 1. Network Metadata.
© OpenStreetMap contributors 2017. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

Motivation and Contribution.

Developing a FAIR repository for climate science data is a multidisciplinary effort involving climate scientists who generate data; data engineers with skills to create federated digital assets; and machine learning researchers who provide expertise on tasks such as gap filling, predictive modelling and exploiting deeper learning models for more complex machine learning tasks. One of the main goals of the CA20108 cost action (CA20108 Cost Action 2021) was to design and build a knowledge portal that is fully compliant with the FAIR principles for scientific data management. Recently, the FAIR Micromet Portal (FMP) was developed to capture metadata for FAIR datasets (Roantree et. al. 2023). Its purpose was to provide detailed metadata descriptions for shareable micrometeorological data using the WMO standard. While storing Network, Site and Sensor metadata locally, the system passes climate datasets to Zenodo, receives back the DOI and thus, creates a permanent link between the FMP and the storage platform Zenodo. In this way, the user exploits the (metadata) search functionality of the FMP to obtain both detailed descriptions and links to data on the Zenodo platform.

https://doi.org/10.5194/essd-2023-334
Preprint. Discussion started: 22 September 2023

Given that research clearly demonstrates that FAIR data has far greater impact than data which is not FAIR, it is therefore of considerable value to the research community if there is an automated or easy to use method for ensuring that data is FAIR. In essence, create a system where researchers can simply "file and forget", content in the knowledge that their data meets FAIR principles and is delivering a greater contribution to science. To achieve this, it is first necessary to understand and interpret the FAIR principles, to help validate any proposed solution. In this paper, we begin with an interpretation of the FAIR principles and articulate what is required to meet the threshold for each principle. We then use a case study from an urban micrometeorological network and perform a step through the creation of FAIR metadata through the FAIR Micromet Portal (FMP), first introduced in (Roantree et. al. 2023), as essentially a form of software wizard, a user interface that takes the researcher through a sequence of small steps, in the creation of FAIR metadata. In the remainder of the paper, the term FAIRNESS is used to refer to the FAIRNESS CA20108 Cost Action and FMP refers to the FAIRNESS Micromet Portal, a system devised to deliver micrometeorological data FAIR compliant.

Paper Structure. In §2, we provide a detailed discussion of the FAIR principals and out interpretation of these principles in order that appropriate metrics can be devised; In §3, we describe how the FAIR Micromet Portal meets FAIR requirements; In §4, a discussion is presented; and finally in §5, the paper finishes with conclusions.

2 Is My Data FAIR?

Since the original FAIR proposal (Wilkinson et. al. 2016), there have been a number of papers which sought to explore and interpret the different metrics (Berman & Crosas 2020); some authors suggested extensions and a tighter interpretation of the thresholds for compliance (D'Aquin et. al. 2023); while there has also been a recommendation of how FAIR systems should be implemented using the concept of a FAIR Data Point (Benhamed et. al. 2022). There have been a number of efforts at testing micrometerological data for FAIR compliance including our own case study (Lalić, Koci & Roantree 2022). What is clear is that some of the metrics and their proposed thresholds are inexact and open to interpretation. Thus, we begin with a discussion on the principles themselves, our interpretation of these and how metrics can be devised from the principles to determine FAIRness. For simplicity, our discussion refers to a globally unique identifier as a Digital Object Identifier (DOI) but accept any equivalent form of globally unique ID.

2.1 Findable

There are six *Findable* principles presented at the top of Table 1, as F1 to F4, extended slightly from the original specification (Wilkinson et. al. 2016). We adjusted F1 to clearly articulate 2 distinct requirements: F1.1 requires that data has a DOI; F1.2 requires that metadata has a DOI. Our interpretation is that F1.1 is mandatory while (for reasons explained later), F1.2 is preferable but not mandatory. Our added stipulation to F2 (rich metadata description) is that different levels should be supported: (at least) a minimum level of metadata descriptions and

https://doi.org/10.5194/essd-2023-334
Preprint. Discussion started: 22 September 2023

Open Access Earth System Science Data Discussions

Table 1. FAIR Metrics and Interpretation

ID	Original Principal	Metric Interpretation
	Findable	
F1.1	Data assigned globally unique ID	Essential (directly or indirectly)
F1.2	Metadata assigned globally unique ID	Preferable (directly or indirectly)
F2	Data are described by rich metadata	Should support different levels
F3	Metadata explicitly includes the unique ID	Should be a metadata attribute
F4.1	Data are searchable	Directly or indirectly
F4.2	Metadata are searchable	Directly or indirectly
	Accessable	
A1.1	Data retrieved using standard protocol	Query Portal or high level language
A1.2	Metadata retrieved using standard protocol	Query Portal or high level language
A2	System is open, free, universally implementable	Should be open and free to read and search
A3	System includes authentication/authorisation	Requires authentication for *creating* metadata
A4	Metadata is accessible past the lifetime of data	Metadata and data are equally valuable
	Interoperable	
I1	Metadata uses a formal, accessible, shared, broad applied knowledge representation language	Accepts XML, JSON, CSV etc.
I2	Vocabularies follow FAIR principles	Standard metadata or ontology
I3	Metadata includes qualified refs to other (meta)data	Expose your API and metamodel
I4	System supports importing/exporting of metadata	Exportable as XML or CSV etc.
	Reusable	
R1	Metadata are richly designed with a plurality of accurate and relevant attributes	Facilitate rich metadata descriptions
R2	(Meta)data have clear data usage license	Articulate usage requirements
R3	(Meta)data associated with detailed provenance	Description of Data Generation & Manipulation
R4	(Meta)data meet domain relevant community standards	Must meet quality criteria

https://doi.org/10.5194/essd-2023-334
Preprint. Discussion started: 22 September 2023

advanced level(s) to support more sophisticated descriptions. Our interpretation for F3 is that the metadata should contain an attribute to capture the DOI. For F4, we again treat the two requirements separately: F4.1 and F4.2 advocate that data and metadata respectively are searchable. This may be facilitated directly through the system or indirectly using a separate (FAIR) system.

2.2 Accessible

Table 1 contains five *Accessible* principles extended slightly from the original specification where A1 is now articulated as A1.1 and A1.2 where we again distinguish between access to data and metadata. Our interpretation is that this requires either a high level query portal or standard protocol such as a RESTful API publishing data in standard formats (XML, JSON etc..). For A2, the term "universally implementable" is ambiguous so we interpret that to mean "easy to use" and assume A2 to require the system to be free to read and search. For A3, we assume that authentication/authorisation is required for the *creation* of FAIR data (otherwise it contradicts A2) but searching should require no such access to *open and free* data (A2). Our interpretation of A4 is that it is delivered though F1.2, which delivers an independent permanent DOI for metadata and interpret this principle as meaning: metadata and data are equally valuable.

2.3 Interoperable

Data interoperability is an important concept to understand as it is fundamental to the process of data integration. Its strategies are well understood now (Batini et. al. 1986) and in more recent times, methods have been devised to integrate data from both structured and semi-structured data (Scriney et. al. 2019), a crucial feature as not all micrometerological data will have a single fixed structure. In table 1, the three original *Interoperable* principles are extended here with a fourth principle (I4) to further enhance interoperability. In the original specification (Wilkinson et. al. 2016), I1 required than both data and metadata adopt the same formal representation but here, we restrict that formality to the metadata as one cannot make assumptions about the data which often requires a form of data *wrapper* to deliver that level of formalisation. Our interpretation is that metadata should be available in one of a small number of very popular standards, eg. XML, JSON, or CSV. I2 states a requirement for a common vocabulary which we interpret as the adoption of a standard metamodel or ontology. Considerable detail is required when integrating data from unrelated, heterogenous sources. In climate science, there are recent examples of single usage bespoke solutions, for example (Brambilla et. al. 2019), but integration generally requires a (meta) data model to describe data and in the FMP, the WMO guide provides the design and structure for metadata. I3 highlights an important feature of interoperability: the ability for 2 heterogeneous metamodels (or ontologies) to communicate which we interpret as exposing or publishing details of each FAIR Data Point (also highlighted in (Benhamed et. al. 2022)). We have added I4 as we believe that exporting metadata is a crucial feature in supporting I3, as we are in agreement with the FAIR analysis provided in (D'Aquin et. al. 2023).

https://doi.org/10.5194/essd-2023-334
Preprint. Discussion started: 22 September 2023

2.4 Reusable

The four *Reusable* principles are broadly in line with the original specification with some minor articulations. For R1, we restrict the rich description to metadata (and do not concern ourselves with data). For R2 and R3, we assume that usage requirements and a record of data creation and manipulation (for example, gap filling) are recorded. For R4, we assume a guarantee of minimum criteria (for example, key metadata attributes cannot be left blank).

Name	Latitude	Longitude	Altitude (M)	Time Zone	Macroscale Environment	
s2-2	45.249166	19.837222	79.00000	UTC	Urban Street Canyon	View
s2-3	45.261388	19.848888	78.00000	UTC	Residential Area (Multi-Story Buildings)	View
s3-2	45.233333	19.809722	79.00000	UTC	Residential Area (Houses)	View
s5-2	45.25	19.816111	75.00000	UTC	Boulevard	View
s5-3	45.2625	19.826388	78.00000	UTC	Residential Area (Multi-Story Buildings)	View
s5-4	45.238055	19.832777	81.00000	UTC	Residential Area (Multi-Story Buildings)	View
s5-5	45.253055	19.8475	80.00000	UTC	Residential Area (Multi-Story Buildings)	View
s5-6	45.2425	19.847222	78.00000	UTC	Residential Area (Multi-Story Buildings)	View
s6-4	45.233611	19.791944	76.00000	UTC	Residential Area (Houses)	View
s6-8	45.251388	19.875555	76.00000	UTC	Residential Area (Houses)	View
s6-9	45.240555	19.881111	92.00000	UTC	Residential Area (Houses)	View
s8-1	45.272369	19.820833	77.00000	UTC	Industrial Area	View

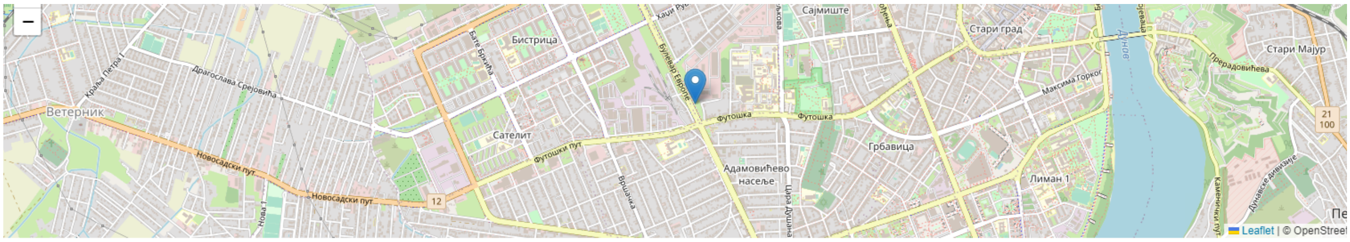
Figure 2. Network Metadata (Sites).

2.5 Summary

In previous work presented in (Lalić, Koci & Roantree 2022), a case study detailed the application of a FAIR test for micrometerological data using metrics very similar to those presented in Table 1. Conducted before the development of the FMP, it identified the *Interoperability* metric as the most difficult metric on which to achieve full conformance. Using a validation method of Yes, No, Partial for each test: **Findable** metrics scored 2 *Yes* and 2 *Partial* responses; **Accessible** metrics scored 2 *Yes* and 2 *Partial* responses; **Interoperable** metrics scored one *Yes* and 3 *Partial* responses; and **Reusable** metrics scored 2 *Yes* and 2 *Partial* responses. Thus, it is important to distinguish between data that has full FAIR compliance, partial FAIR compliance and not FAIR compliant. The outcome from this test

https://doi.org/10.5194/essd-2023-334
Preprint. Discussion started: 22 September 2023

was that full FAIR compliance is quite difficult to achieve although partial (and potentially high levels of) FAIR compliance are quite achievable.



Variable	Type	Description	Temporal Resolution	Start Date	End Date	Sensor Height	Values	Value Notes	Sky View Factor	Height/Width Values		
Air Temperature (Ventilated)	ChipCap 2 sensor, fully calibrated and developed by the General Electric Measurement & Control Company, and located in a ventilated radiation protection screen with dimensions of 200x240 mm. The accuracy of the temperature sensor was	Sensors measured the air temperature values every minute and every 10 minutes measured data was sent to the server (Uni of Novi Sad, Faculty of Sciences). 1-hour datasets were extracted for the period 2016-2017 and QC.	1 h	Jan. 1, 2016	Dec. 31, 2017	4 m	MEASURED/RAW	Dataset from this station is QC with 0 outliers and 0% of missing data. The whole procedure of QC is on this link: https://drive.google.com/drive/folders/1tvDpMMvN10JQni4o6X3SlqI5bJ5yls3r?usp=sharing		Boulevard width is about 108 m / Buildings heights between 6 and 30 m.		View

Figure 3. Site Metadata.

© OpenStreetMap contributors 2017. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

3 The FAIR Micromet Portal

In this section, we illustrate the capture of metadata for micrometerological data stored in the FMP. The procedure for uploading data, described in (Roantree et. al. 2023) is that metadata is recorded in the FMP while the actual data is transferred immediately to Zenodo, creating the DOI that is subsequently captured in the FMP, as seen in the top right of figure 1. The micrometerological data used in the earlier FAIR assessment provides a good case study for validating how the FMP conforms to FAIR principles while also ensuring that data meets complies with the more difficult metrics. Data was generated as part of the Novi Sad Urban Network (NSUNET) system where each site was equipped with multiple sensors and a variety of electronic and hardware devices (Milosevic et. al. 2022). The project's objective was to provide conditions for progressive urban climate research, for example, contributing

https://doi.org/10.5194/essd-2023-334
Preprint. Discussion started: 22 September 2023

to the thermal pattern differences of various urban surroundings. Metadata captured at the Network level is shown in figures 1 and 2.

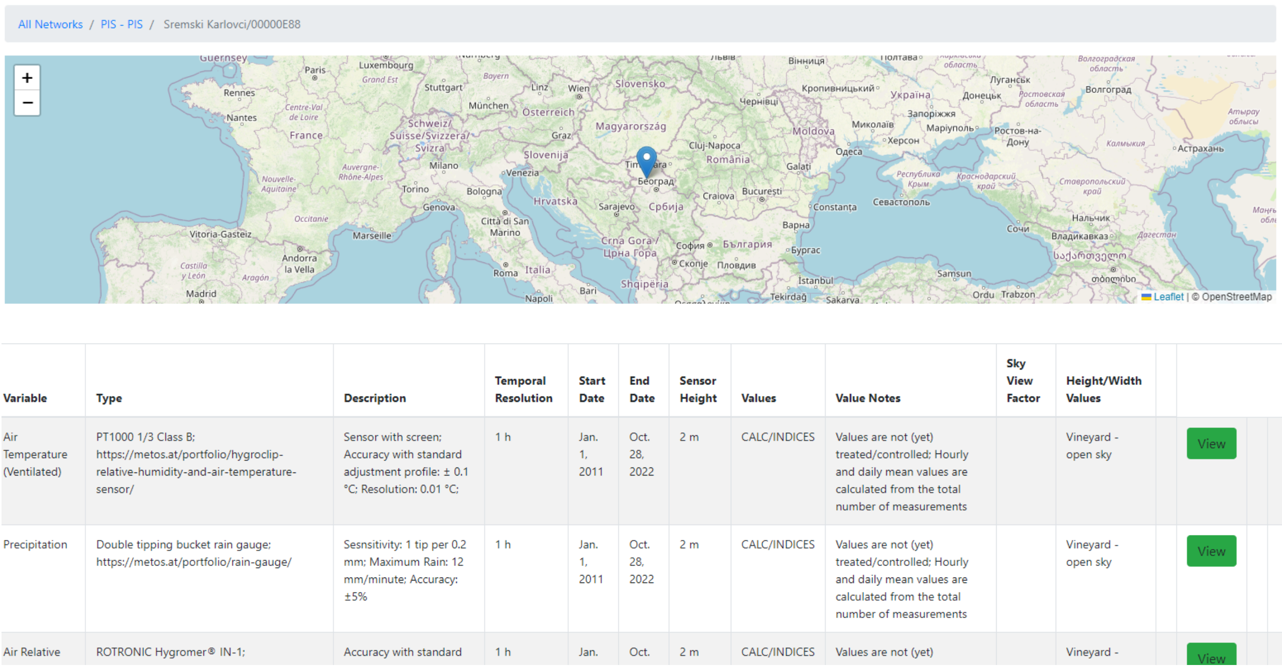


Figure 4. Sensor Metadata.

The **Network** is the regarded as the highest level of *abstraction* and thus, requires a relatively short metadata record. Figure 1 shows the capture of: Institution, Project, Network Code, Contact Email, County, City/Region, Local Environment, Seasonality, Start Date, End Date and Reference. These maps are generated by the FAIR Micromet Portal using two software libraries: the leaflet library is free to use through the BSD-2-Clause licence (leafletjs.com, 2023) and OpenStreetMap is open data, provided through the Open Data Commons Open Database License (OpenStreetMap, 2017). The system creates the attributes: Date Created and Last modified and DOI creation dates. The **Site** is also at a high level of *abstraction* requiring a relatively short metadata record. Figure 3 shows the capture of: Name, Latitude, Longitude, Altitude, Time Zone and Macroscale Environment. The **Sensor** requires the most descriptive metadata as it contains many parameters and settings and comes with a wide variety of functions which all require a description. Figure 4 shows the capture of: Variable, Type, Description, Temporal Resolution, start Date, end Date, Sensor Height, Values, Value Notes, Sky View Factor, Height/Width Values. For both Site and Sensor metadata, the system creates attributes for Date Created and Date Modified.

https://doi.org/10.5194/essd-2023-334
Preprint. Discussion started: 22 September 2023

Table 2. FMP Functionality and FAIR Conformance

Function	Description	Metric
Storage	WMO Metadata	F2,A4,I4
Search	Open Search By Attribute	F4.2,A1.1,A1.2,A2,R2
Data Model	WMO Standard	I1,I2,R4
Username/password	For Write Only	A3
Zenodo	Open Data Storage	F1.1,F4.1,A1.1,R2,(F1.2)*

4 Case Study Based Validation

In this section, we validate the FMP system against the FAIR principles using the NSUNET case study as this data is freely available on Zenodo (Savić et. al. 2023a) with a dataset description and statistical summary presented in (Savić et. al. 2023b). We begin with a discussion on FMP system functionality to understand how it automatically provides compliance with a large number of FAIR metrics. We then examine FMP metadata capture in detail to highlight how greater levels of FAIR conformance are obtained, followed by a discussion on NSUNET metadata and its conformance with FAIR principles.

4.1 FMP Metadata Structure

In table 2, we can see the different FAIR metrics that are automatically delivered by the FMP system, once the user provides a description for each attribute. By providing storage for metadata descriptions of climate data, the requirements for F2, A4 and I4 are met: data is described by rich metadata; metadata is accessible past the lifetime of data (Zenodo DOI) and importing/exporting of metadata is facilitated. The FMP's Search option means that both metadata and data (link to Zenodo) can be retrieved (F4.2, A1.1, A1.2, A2) with a clear policy usage (R2). By adopting the WMO description for climate data and measurements, I1 and I2 are achieved but so also is R4 (meeting a domain standard). The authentication screen ensures compliance with A3 while the adoption of Zenodo's technology ensures that F1.1, (globally unique IDs), F4.1 (searchable data), A1.1 (standard access), R2 (open data usage policy) are met. In addition, Zenodo could be used for metadata storage as for this case study validation, the FMP was used to export metadata for the entire NSUNET network which was stored on Zenodo and subsequently provided with a unique DOI (F1.2). In summary, conformance to 15 of the 19 FAIR principles is provided through FMP functionality.

4.2 NSUNET Case Study and Discussion

The full metadata record used in this evaluation is openly available (Savić et. al. 2023c) with extracts available in appendices A and B. In table 3, we show each metadata variable used to describe climate datasets and highlight the

https://doi.org/10.5194/essd-2023-334
Preprint. Discussion started: 22 September 2023

Open Access Earth System Science Data Discussions

Table 3. FMP Datasets and FAIR Conformance

Network Metadata	Metric	Site Metadata	Metric	Sensor Metadata	Metric
institution	F2,I3,R1	name	F2,I3,R1	Variable	F2,I3,R1
project	F2,I3,R1	latitude	F2,I3,R1	type	F2,I3,R1
network_code	F2,I3,R1	longitude	F2,I3,R1	description	F2,I3,R1
contact_email	F2,I3,R1	altitude	F2,I3,R1	temporal_resolution	F2,I3,R1
country	F2,I3,R1	time_zone	F2,I3,R1	start_date	F2,I3,R1
city_region	F2,I3,R1	macroscale_env	F2,I3,R1	end_date	F2,I3,R1
local_environment	F2,I3,R1			sensor_height	F2,I3,R1
seasonality	F2,I3,R1			values	F2,I3,R1
start_date	F2,I3,R1			value_notes	F2,I3,R1
end_date	F2,I3,R1			sky_view_factor	F2,I3,R1
reference	F2,I3,R1			height_width_values	F2,I3,R1
date_created	F2,I3,R1,R3	date_created	F2,I3,R1,R3	date_created	F2,I3,R1,R3
date_modified	F2,I3,R1,R3	date_modified	F2,I3,R1,R3	date_modified	F2,I3,R1,R3
DOI creation dates	F2,I3,R1			photo	F2,I3,R1
is_public	F2,I3,R1				
DOI	F2,F3,I3,R1	DOI	F2,F3,I3,R1		

four variables that were missing from the previous discussion on FMP functionality and compliance. The F3 metric is delivered by the DOI variable; I3 is covered by the many variables that contain references to other metadata; R1 is delivered by the totality of the metadata captured; and R3 provides some detail of provenance (dates, ownership and the totality of metadata).

While the validation supports the position that FMP usage ensures that scientific data is FAIR compliant, it relies on underlying technologies (eg. Zenodo) to cover some of the metrics. While this does not necessarily require that data be open (minimal data uploaded to Zenodo still provides lifetime DOIs), it highlights that there are *degrees of FAIR compliance*. What this suggests is that while two datasets may be FAIR compliant, one of these datasets may be easier for users to find, search and use. For this issue to be tackled, it will require more FAIR respositories, more scientists engaging with the process, and ultimately, a data flow process that becomes embedded as part of research data creation or acquisition and subsequent management.

The other issues that arises is whether or not FAIR data is actually interoperable. In other words, data can deliver on all 4 Interoperable metrics but still have low levels of interoperability. This is because true interoperability requires a fair higher level of data engineering, for example, to integrate two datasets from unrelated projects. Some efforts have used quite an extensive set of metadata descriptions to deliver data integration, for example, the research in (Frémand et. al. 2021) adopting the CF metamodel (Eaton et. al. 2022). This far stricter level of FAIR conformance (many mandatory metadata descriptions) leads to higher levels of interoperability but comes at a price. That price is

https://doi.org/10.5194/essd-2023-334
Preprint. Discussion started: 22 September 2023

related to resource overhead and acceptance of additional effort by the scientific community. This tradeoff requires a deeper discussion with scientists around the longer term benefits and significant impact of highly interoperable FAIR data. However, it may also require funding agencies to separately fund this additional overhead while simultaneously making FAIR data processing activities a mandatory part of the funding process.

5 Conclusions

In this paper, a motivation for making scientific data FAIR was presented with a particular emphasis on FAIR climate research data. Ideally, this process requires easy to use tools to enable scientists to attach a form of *FAIR Certification* to their data. Our approach was to demonstrate how the FAIR Micrometeorological Portal (FMP) can deliver this type of certification for climate research data. Of the extended set of FAIR metrics presented here, 15 of the 19 metrics, almost 80%, are delivered by the system itself, meaning that once data descriptions (FAIR metadata) have been recorded on the FMP, a high degree of FAIR compliance is guaranteed. As part of the validation for this research, it was shown how different elements of the WMO metadata address the remaining FAIR metrics using an urban network in Novi Sad comprisng twelve sites. The FMP system (fairmicromet.eu) is currently open to all researchers for metadata search and also for access/search to data available on the Zenodo open platform. For scientists seeking a FAIR Certification for their data, it currently requires membership of the Cost Action (cost.eu/actions/CA20108/) but it is planned to open the FMP to all climate researchers later in 2023.

Data availability. Metadata used for the case study evaluation is openly available at https://www.zenodo.org/record/8237900 (Savić et. al. 2023c).

Sample availability. Both the cleaned hourly air temperature datasets from the NSUNET system and the raw 10-minute interval data (original data) are openly available at https://www.zenodo.org/record/7738093 (Savić et. al. 2023a).

https://doi.org/10.5194/essd-2023-334
Preprint. Discussion started: 22 September 2023

Appendix A: NSUNET Site Metadata

Table A1. Sample Site Metadata

name	latitude	longitude	alt_m	date_created	macroscale_environment
s2-3	45.261388	19.848888	78	2023-03-18 17:49	Residential Area (Multi-Story Buildings)
s3-2	45.233333	19.809722	79	2023-03-18 17:50	Residential Area (Houses)
s5-2	45.25	19.816111	75	2023-03-18 17:52	Boulevard
s5-3	45.2625	19.826388	78	2023-03-18 17:54	Residential Area (Multi-Story Buildings)
s5-4	45.238055	19.832777	81	2023-03-18 17:55	Residential Area (Multi-Story Buildings)
s5-5	45.253055	19.8475	80	2023-03-18 17:58	Residential Area (Multi-Story Buildings)

https://doi.org/10.5194/essd-2023-334
Preprint. Discussion started: 22 September 2023

Open Access Earth System Science Data Discussions

Appendix B: NSUNET Sensor Descriptions

Table B1. Sensor Metadata

Variable Name	**Metadata**
type	ChipCap 2 sensor, fully calibrated and developed by the General Electric Measurement & Control Company, and located in a ventilated radiation protection screen with dimensions of 200x240 mm. The accuracy of the temperature sensor was ±0.3 oC.
description	Sensors measured the air temperature values every minute and every 10 minutes measured data was sent to the server (Uni of Novi Sad, Faculty of Sciences). 1-hour datasets were extracted for the period 2016-2017 and QC.
temporal_resolution	1h
start_date	01/01/2016
end_date	31/12/2017
sensor_height	4.1m
values	MEASURED/RAW
value_notes	Dataset from this station is QC with 0 outliers and 0% of missing data. The whole procedure of QC is on this link: https://drive.google.com/drive/folders/1tvDpMMvN10JQni4o6X3SIql5bJ5yls3r?usp=sharing
sky_view_factor	
height_width_values	It is a small square 41 m with 50 m surrounded by multi-story residential buildings.
photo	
date_created	2023-03-18 18:18
date_modified	2023-04-04 15:57
site_id	96f4a0c9-118d-4673-880c-009de3084654

Author contributions. **Branislave Lalić**: Conceptualisation, Writing – review & editing. **Michael Scriney**: Data Curation, Software, Validation, Writing – review & editing. **Stevan Savić**: Data Curation, Writing – review & editing, Validartion. Mark Roantree: Conceptualisation, Methodology, Investigation, Writing – original draft preparation.

Competing interests. The authors declare that they have no conflict of interest.

https://doi.org/10.5194/essd-2023-334
Preprint. Discussion started: 22 September 2023

Earth System Science Data
Open Access Discussions

Acknowledgements. This work was conducted with the financial support of: the European Cooperation in Science and Technology (COST) Grant CA20108, Science Foundation Ireland under grant number SFI/12/RC/2289_P2 and Ministry of Education, Science and Technological Development of the Republic of Serbia, Grant No. 451-03-47/2023-01/200117.

https://doi.org/10.5194/essd-2023-334
Preprint. Discussion started: 22 September 2023

References

Antunes, Rodolfo S. and Seewald, Lucas A. and Rodrigues, Vinicius F. and Costa, Cristiano A. Da and Jr., Luiz Gonzaga and Righi, Rodrigo R. and Maier, Andreas and Eskofier, Björn and Ollenschläger, Malte and Naderi, Farzad and Fahrig, Rebecca and Bauer, Sebastian and Klein, Sigrun and Campanatti, Gelson: A Survey of Sensors in Healthcare Workflow Monitoring. ACM COMPUT SURV, 51(2), 1-37, 2018. https://doi.org/10.1145/3177852.

d'Aquin, Mathieu and Kirstein, Fabian and Oliveira, Daniela and Schimmler, Sonja and Urbanek, Sebastian, FAIREST: A Framework for Assessing Research Repositories, Data Intelligence 5(1), 202-241, 2023. https://doi.org/10.1162/dint_a_00159.

Brambilla, W., Conforti, A., Simeone, S., Carrara, P., Lanucara, S., and De Falco, G.: Data set of submerged sand deposits organised in an interoperable spatial data infrastructure (Western Sardinia, Mediterranean Sea), EARTH SYST SCI DATA 11, 515–527, https://doi.org/10.5194/essd-11-515-2019, 2019.

Batini, Carlo and Lenzerini, Maurizio and Navathe, Shamkant B. A comparative analysis of methodologies for database schema integration, ACM COMPUT SURV, 18(4), 3230364, 1986, https://doi.org/10.1145/27633.27634.

Ousamma Mohammed Benhamed, Kees Burger, Rajaram Kaliyaperumal, Luiz Olavo Bonino da Silva Santos, Marek Suchánek, Jan Slifka and Mark D Wilkinson. The FAIR Data Point: Interfaces and Tooling. Data Intelligence 5 (1): 184–201, 2022, https://doi.org/10.1162/dint_a_00161.

Berman, F., and Crosas, M. The Research Data Alliance: Benefits and Challenges of Building a Community Organization. Harvard Data Science Review 2(1), 2020, https://doi.org/10.1162/99608f92.5e126552.

Op De Beéck, Tim and Meert, Wannes and Schütte, Kurt and Vanwanseele, Benedicte and Davis, Jesse, Fatigue Prediction in Outdoor Runners Via Machine Learning and Sensor Fusion, 24th KDDM CONF PROC, 606-615, 2018, https://doi.org/10.1145/3219819.3219864.

Brian Eaton, Jonathan Gregory, Bob Drach, Karl Taylor, Steve Hankin, Jon Blower, John Caron, Rich Signell, Phil Bentley, Greg Rappa, Heinke Höck, Alison Pamment, Martin Juckes, Martin Raspaud, Randy Horne, Timothy Whiteaker, David Blodgett, Charlie Zender, Daniel Lee, David Hassell, Alan D. Snow, Tobias Kölling, Dave Allured, Aleksandar Jelenak, Anders Meier Soerensen, Lucile Gaultier, Sylvain Herlédan, Fernando Manzano, NetCDF Climate and Forecast (CF) Metadata Conventions, Version 1.11, 31 August 2022, https://cfconventions.org/cf-conventions/cf-conventions.pdf.

The Official Portal for European Data, https://data.europa.eu/en/dataeuropa-academy/what-open-data, (last access: May 12th 2023).

EUR-Lex. High-value datasets and the arrangements for their publication and re-use, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2023.019.01.0043.01.ENG, Document 32023R0138, December 2022.

FAIR Network of micrometeorological measurements, Cost Action CA20108, https://www.fairness-ca20108.eu/, 14 Oct 2021.

Frémand, A. C. and Fretwell, P. and Bodart, J. A. and Pritchard, H. D. and Aitken, A. and Bamber, J. L. and Bell, R. and Bianchi, C. and Bingham, R. G. and Blankenship, D. D. and Casassa, G. and Catania, G. and Christianson, K. and Conway, H. and Corr, H. F. J. and Cui, X. and Damaske, D. and Damm, V. and Drews, R. and Eagles, G. and Eisen, O. and Eisermann, H. and Ferraccioli, F. and Field, E. and Forsberg, R. and Franke, S. and Fujita, S. and Gim, Y. and Goel, V. and Gogineni, S. P. and Greenbaum, J. and Hills, B. and Hindmarsh, R. C. A. and Hoffman, A. O. and Holmlund, P. and Holschuh, N. and Holt, J. W. and Horlings, A. N. and Humbert, A. and Jacobel, R. W. and Jansen, D. and Jenkins,

A. and Jokat, W. and Jordan, T. and King, E. and Kohler, J. and Krabill, W. and Kusk Gillespie, M. and Langley, K. and Lee, J. and Leitchenkov, G. and Leuschen, C. and Luyendyk, B. and MacGregor, J. and MacKie, E. and Matsuoka, K. and Morlighem, M. and Mouginot, J. and Nitsche, F. O. and Nogi, Y. and Nost, O. A. and Paden, J. and Pattyn, F. and Popov, S. V. and Rignot, E. and Rippin, D. M. and Rivera, A. and Roberts, J. and Ross, N. and Ruppel, A. and Schroeder, D. M. and Siegert, M. J. and Smith, A. M. and Steinhage, D. and Studinger, M. and Sun, B. and Tabacco, I. and Tinto, K. and Urbini, S. and Vaughan, D. and Welch, B. C. and Wilson, D. S. and Young, D. A. and Zirizzotti, A., Antarctic Bedmap data: Findable, Accessible, Interoperable, and Reusable (FAIR) sharing of 60 years of ice bed, surface, and thickness data, EARTH SYST SCI DATA 15, 2695–2710, https://doi.org/10.5194/essd-15-2695-2023, 2023.

Javaid M., Haleem A., Rab S., Singh R., Suman R., Sensors for daily life: A review, Sensors International, 2, 1-10, https://doi.org/10.1016/j.sintl.2021.100121.

Lalić B., Koci I., and Roantree M. FAIRness of micrometeorological data and responsible research and innovation: an open framework for climate research, 2nd AGROECOINFO CONF PROC, 75-80, 2022, https://doras.dcu.ie/28012/1/Lalic.Koci.Roantree-Open.Framework.for.Climate.Research.pdf.

Leaflet: A JavaScript library for interactive maps. Available at: https://leafletjs.com/, last accessed 14/09/2023.

Milošević, D., Savić, S., Kresoja, M. et al. Analysis of air temperature dynamics in the "local climate zones" of Novi Sad (Serbia) based on long-term database from an urban meteorological network, Int J Biometeorol 66, 371–384, 2022, https://doi.org/10.1007/s00484-020-02058-w.

Mons, B., Neylon C., Velterop J., Dumontier M., Bonino L,. Wilkinson M, Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud, Information Services & Use 37, 49–56, 2017. https://content.iospress.com/articles/information-services-and-use/isu824.

Muller R., Curry J., Groom D., Jacobsen R., Perlmutter S., Rohde R., Rosenfeld A., Wickham C., Wurtele J., Decadal variations in the global atmospheric land temperatures, JGR Atmospheres, 118, 5280–5286, 2013, https://doi.org/10.1002/jgrd.50458.

OpenStreetMap Contributors, 2017. Available at: https://www.openstreetmap.org.

Pastorello G., Papale D., H. Chu, C. Trotta, D. A. Agarwal, E. Canfora, D. D. Baldocchi and M. S. Torn 17 April 2017 A New Data Set to Keep a Sharper Eye on Land-Air Exchanges. EOS Earth & Space Science News. 98, 2017. DOI: 10.1029/2017EO071597.

PwC EU Services, Cost of Not Having FAIR Research Data: Cost-Benefit Analysis for FAIR Research Data, March 2018. http://publications.europa.eu/resource/cellar/d375368c-1a0a-11e9-8d04-01aa75ed71a1.0001.01/DOC_1.

Roantree M., Lalić B., Savić S., Milošević D., Scriney M., Constructing a Searchable Knowledge Repository for FAIR Climate Data, Report number EGU23-7786, Copernicus Meetings 2023. Full version available at: https://arxiv.org/pdf/2304.05944.pdf.

Santana E., Chaves A., and Gerosa M., Kon, F., Milojicic D., Software Platforms for Smart Cities: Concepts, Requirements, Challenges, and a Unified Reference Architecture, ACM COMPUT SURV, 50(6), 1-37, 2018, https://doi.org/10.1145/3124391

Savić S., Secerov I., Dunjic J., and Milosevic D. Hourly Air Temperature Datasets from city of Novi Sad - NSUNET system, 2023. https://www.zenodo.org/record/7738093.

https://doi.org/10.5194/essd-2023-334
Preprint. Discussion started: 22 September 2023

Stevan Savić, Ivan Šećerov, Branislava Lalić, Dongyun Nie, Mark Roantree, Air Temperature and Relative Humidity Datasets from an Urban Meteorological Network in the City Area of Novi Sad (Serbia), Data in Brief, 49, 1-11, 2023. https://doi.org/10.1016/j.dib.2023.109425.

Savić S., Secerov I., Dunjic J., and Milosevic D. Metadata of the urban meteorological network in Novi Sad (Serbia) - NSUNET system, 2023, https://www.zenodo.org/record/8237900.

Scriney M., McCarthy S., McCarren A., Cappellari P., Roantree M., Automating Data Mart Construction from Semi-structured Data Sources, COMPUT J, 62(3), 394–413, 2019, https://doi.org/10.1093/comjnl/bxy064.

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.

World Economic Forum. The Global Risks Report (14th Ed), 2019, https://www3.weforum.org/docs/WEF_Global_Risks_Report_2019.pdf.

World Meteorological Organization, Guide to Agricultural Meteorological Practices (GAMP), WMO-No.134, 2010, https://library.wmo.int/doc_num.php?explnum_id=3996.

World Meteorological Organization. Guide to Instruments and Methods of Observation: Volume I Measurement of Meteorological Variables, 2021, library.wmo.int/doc_num.php?explnum_id=11612.