FAIRness of Micrometeorological Data and Responsible Research and Innovation: an Open Framework for Climate Research

Branislava Lalic,[1] Ivan Koci,[2] Mark Roantree[3]

[1]Faculty of Agriculture, University of Novi Sad, Novi Sad, Serbia
[2] Forecasting and Reporting Service for Plant Protection of the Republic of Serbia (PIS)

[3] Insight Centre for Data Analytics, School of Computing, Dublin City University, Ireland.

Abstract. During the XX century, the atmospheric science community became one of the major "big data" sources, and micrometeorological measurements are an important part of it. This trend is still going strong, supported by networks of so-called "unconventional" measurements and SMART technologies (citizen weather networks and meteorological instruments on different mobile devices, for example). The potential weakness of data coming from such diverse sources is the lack of **f**indability, **a**ccessibility, **i**nteroperability, and **r**eusability (FAIR). The opportunity for improvement is "to define a minimal set of related but independent and separable guiding principles and practices, which enable both machines and humans to find access, interoperate and reuse research data and metadata" (PwC EU Services, 2018). Enhanced FAIRness of micrometeorological data strongly supports almost all main pillars of Responsible Research and Innovation (RRI), particularly in STEM and STEM-related fields, leaving room for different levels of accessibility since FAIRness does not imply Open data status.
In a case study, an initial FAIRness assessment is performed for micrometeorological data measured in an orchard within the Forecasting and Reporting Service for Plant Protection of the Republic of Serbia (PIS) observational network.

1 Introduction

RRI concept. In an ideal World, Responsible Research and Innovation (RRI) projects and initiatives would be redundant. The collective consciousness of the research community will be enough to:

a) provide Open Access (OA) to research results (publications, data, models, methodologies),
b) pursue high-quality Science Education (SE) by committing our collective knowledge through science textbooks for all educational levels-from elementary to graduate schools;
c) develop equal opportunity research environments regardless of gender (Gender Equality), race, nationality, or religion;
d) provide whole Public Engagement (PE) in research and SE by maintaining communication with the general and specialized public and supporting Citizen Science (CS) initiatives;
e) overcome all ethical issues (Ethics) by constructive discussions with all interested parties;
f) participate in Governance (GOV) and convince policymakers that research results should be fully considered while designing strategic, long-term, and short-term plans.

However, a survey conducted in 2017 involving about 3100 researchers receiving H2020 funding revealed that most of them were unaware of the RRI concept (Novitsky, Bernstein, Blok, Braun, Chan, Lamers, Loeber, Meijer, Lindner & Griessler, 2020). This realization should be an alarm for the scientific community and society. Recent COVID19 pandemics confirmed that the scientific community could respond to today's greatest challenges. However, the growing anti-vaxxer campaign has confirmed our vulnerability and lack of RRI when conducting scientific research.

FAIR data concept. An important goal on the road map of future responsible research and innovation society is related to data management. Namely, during recent decades, the volume of data generated has been growing exponentially. “The total amount of data created, captured, copied, and consumed globally is forecast to increase rapidly, reaching 64.2 zettabytes in 2020. Over the next five years up to 2025, global data creation is projected to grow to more than 180 zettabytes. In 2020, the amount of data created and replicated reached a new high” (Statista, 2022). All forms of professional work, particularly in research, involve some degree of data management. The lack of data findability, accessibility, interoperability, and reusability (FAIR) costs Europe a minimum of €10.2bn per year - approximately 78% of the Horizon 2020 budget per year (PwC EU Services, 2018). Making data FAIR is essential for the development of responsible research and the innovation society.

RRI and FAIR data. In the context of RRI, the FAIR data concept is commonly associated with Open Access to (research) data. Nevertheless, we would like to challenge the FAIR data concept concerning all RRI principles.

a) **Open Access**. Even though openness is considered a component of FAIR data, it is important to have in mind that it does not equate to open data (Mons et al., 2017) and vice versa. Namely, while open data may be findable and accessible, this does not guarantee that the same data are interoperable and reusable. In addition, findable and accessible data can be open according to specific conditions and restrictions or not open at all.
b) **Science Education**. SE is at a critical point. Anti-vaxxers' impact on reducing vaccination, the growing number of flat-Earth believers (it is not "knowing" it is "believing" since it is not based on facts), and climate change skepticism often using extreme examples, indicate that the SE paradigm must be changed. Enhanced FAIRness of data and methodologies with appropriate publicity and presence on social media can significantly contribute to this effort.
c) **Gender Equality**. Due to different social or professional circumstances, some researchers can have reduced physical approach to labs (work from home for different reasons, for example). In that case, a lot of research can be done if FAIR data are available.
d) **Public Engagement**. Citizen science and PE, as two sides of one "science & society coin" are strongly boosted by the FAIRness of data and methodologies as their integral part.
e) **Ethics**. Due to new legislation related to data privacy protection, all FAIR data must be clearly in line with appropriate ethics criteria. However, additional data anonymization will be a helpful addition to already FAIR data.
f) **Governance**. Hopefully, enhanced FAIRness of data would make them more acceptable for policy- and decision-makers.

The greatest challenges of the 21st century, such as climate change, natural hazards, biodiversity and ecosystem functions, food risks, deforestation, vector born (human, animal and plant) diseases, air quality, and urbanization, are either affected by or affect atmospheric conditions, particularly on the micro-scale.
Reliable and sufficient knowledge of environmental conditions or processes delivered from micrometeorological and microclimatological data play a central role in assessing and modeling trends and effects of climate change (CC) and adverse weather events on the environment and ecosystems on all spatial and temporal scales. Enormous efforts have already been made at the European level to centralize data from ground-based (synoptic scale) and satellite measurements, weather, and climate simulations and make them available for public use (COPERNICUS, ECMWF database, e-OBS, for example). These well-established data sources are broadly and successfully used in research, education, and economics. However, beyond specific initiatives (The European Eddy Fluxes Database Cluster, The Pan-Eurasian Experiment, for example), they are still missing one vital component – micrometeorological data, i.e., data addressing meteorological conditions of the microenvironment (few kilometers scale) open and available for various application potentials and user groups.
Micrometeorological data are usually collected as a part of scientific projects and observational networks developed for different purposes but they often "languish" in reports and institutional data silos. To address this shortfall, FAIRness of micrometeorological data should be enhanced. There are a number of existing projects and initiatives that aim to implement FAIR data principles.

- The GO FAIR initiative (https://www.go-fair.org/) is a bottom-up, stakeholder-driven and self-governed initiative that intends to offer an open and inclusive ecosystem for individuals, institutions and organizations through three implementation networks: GO CHANGE, GO TRAIN and GO BUILD.
- The FAIRsFAIR project (Koers et al., 2020) provies a detailed description of FAIR assessment frameworks and tools for datasets, repositories, and digital objects.
- FAIR data self assessment tool (https://ardc.edu.au/resources/aboutdata/fair-data/fair-self-assessment-tool/) developed by Australian Research Data Commons (ARDC) (Wilkinson et al., 2013; Wilkinson et al., 2016) is designed to make easier FAIR self-assessment and enhancement of datasets.
- In this paper, we will use micrometeorological data measured within the Forecasting and Reporting Service for Plant Protection of the Republic of Serbia (PIS) network (Lalic et al., 2020) to demonstrate some elements as proof of concept behind the FAIRness of micrometeorological data and ongoing CA20108 FAIRNESS Cost action (D Milošević, B Lalić, S Savić, B Bechtel, M Roantree, S Orlandini, 2022).

2 Methodology

2.1 PIS micrometeorological network

PIS micrometeorological network of automated weather stations (AWSs) was established in 2010, intending to provide information about meteorological conditions following the development of harmful organisms in agricultural production to support the work of plant protection specialists of PIS and producers. The network was initially designed, and AWSs were installed according to WMO recommendations for the special meteorological measurements, while all stations were purchased from licensed manufacturers with calibrated sensors. AWS locations were set within plant canopies using GPS. According to location settings, all AWSs can be classified as fixed (orchards, vineyards) and flexible - shifted among crop canopies during one season. For more detail about PIS micrometeorological data and metadata, please refer to Lalic et al. (2022).

2.2 FAIRness validation methodology

We used a simplified methodology designed by Jones and Grootveld (2017) to discuss the extent to which the PIS micrometeorological data is FAIR and the measures required to improve the FAIRness of this data. The methodology is designed in the form of a checklist to simplify the initial steps toward enhanced data FAIRness. We selected data measured within an apple orchard at the Cenej location, in the Novi Sad region for this study.

3 Results and Discussion

In the framework of CA20108 FAIRNES Cost action activities, an extensive self assessment of PIS micrometeorological data is in progress. It includes following phases:
i) initial self assessment – intended to identify lack of basic features using Jones and Grootveld (2017) and ARDC self assessment tools;
ii) qualitative self assessment – intended to test quality of exsisting solutions and room for improvement. Some examples of good practices presented by Wilkinson et al. (2016) and Jacobsen et al. (2019) forms the starting point for the interpretation of results and implementation considerations;
iii) methodology pipeline design – identifies tools, methods and algorithms necessary to implement findings from phase i) and ii);
iv) execution and testing of methodology pipeline – one segment of PIS micrometeorological database will be used to test designed framework and to prepare fully FAIR data satisfiying European Open Science Cloud (EOSC) repository criteria.

Table 1. Findability – Data can be found by humans or machines

Ref	Metric	Test	Current	CA20108 Enhancement Plan
F1	A persistent identifier is assigned to data	True	AWS ID	DOI for every dataset
F2	There are rich metadata describing the data	Partial	Scientific paper (where available)	Metadata Description
F3	The metadata are online in a searchable resource	Partial	Scientific paper (where available)	Searchable Metadata Repository
F4	The metadata record specifies the persistent identifier	True	Paper DOI	Permanent Metadata Repository.

Table 2. Accessibility – humans or machines can gain access to data under specific conditions or restrictions

Ref	Metric	Test	Current	CA20108 Enhancement Plan
A1	Following the persistent ID will lead to data or associated metadata	True	AWS ID	DOI for every dataset & searchable metadata repository
A2	The protocol by which data can be retrieved follows recognized standards	Partial	http for most data with remaining data in csv format.	Online searchable repository
A3	The access procedure includes authentication and authorization steps	Partial	Mixed Username/password in open access catalogue	Registration required for Cost Action online repository.
A4	Metadata accessible where possible (even when data are not)	True	Paper DOI	Searchable Metadata Repository with access to data.

Table 3. Interoperability – data and metadata conform to recognized formats and standards in order to be combined and exchanged.

Ref	Metric	Test	Current	CA20108 Enhancement Plan
I1	Data is provided in commonly understood and preferably open formats	Partial	Partly JSON, partly (interactively) csv, Excel.	WMO GAMP with downloadable data & metadata in CSV format.
I2	The metadata provided follows relevant standards	True	WMO GAMP	WMO GAMP
I3	Controlled vocabularies, keywords, thesauri or ontologies are used where possible	Partial	According to best practice	WMO GAMP
I4	Qualified references and links are provided to other related data	Partial	Aggregated with other tools	CA20108 metamodel facilitates embedded links.

Table 4. Reusability – data and metadata are licensed, conforming to community norms and allowing users to know what kinds of reuse are permitted.

Ref	Metric	Test	Current	CA20108 Enhancement Plan
R1	Data are accurate, well described with many relevant attributes	Partial	Basic quality control checks.	Automated Quality Control checks including data imputation.
R2	The data have a clear and accessible data usage license	True	Data is free to use	Data is free to registered researchers.
R3	It is clear how, why and by whom the data have been created and processed	True	Fully documented.	Part of the Metadata specification.
R4	The data and metadata meet relevant domain standards	Partial	Both data and metadata meet standards for micrometeorological measuremets in a field/canopy layer.	WMO GAMP.

Tables 1-4 present the first results of phase i) - initial self assessment. While none of the FAIR metrics can be categorised as fail, none are the 4 categories have passed all of the assessments. Of the 16 metrics, only 7 could be deemed as being met in full while the remained 9 metrics are deemed partially true. Thus, there is a clear case for improvement and the primary goal of the CA20108 Cost Action to to provide a Knowledge Portal and Data Repository where all datasets meet the 16 FAIR requirements in full. For the PIS case study, the immediate attention is devoted to metadata standardisation, keywords and vocabulary following CA20108 ongoing activities.

3.1 Discussion

In this section, we examine the FAIR metrics in detail and discuss what is required to ensure that micrometeorological experiments and datasets conform to the FAIR standard and in that respect, the work that is planned by the CA20108 Cost Action.

When considering the Findability metric, which states that data can be found by humans or machines, the metrics require that metadata is rich, searchable and permanent with a DOI for each dataset. The PIS requires richer metadata with a proper (metadata) search engine. The Accessibility metric requires that data can be retrieved under specific conditions or restrictions. It stipulates the use of username/password, a DOI for data retrieval and standard communication protocols. While it is accepted that data will not always be freely avaialble, it requires that metadata is always accessible. An online portal can deliver on all of these requirements and for the PIS case study, the lack of online repository meant that two of the metrics were only partially achieved.

The Interoperability metric is potententially the most difficult metric to achieve as it requires the careful specification of a climate metamodel. The requirement that data and metadata conform to recognized formats and standards for the purpose of data integration and exchange will necessitate the formal specification of metadata that conforms to a recognised international standard. For the CA20108 Cost Action, the metamodel specification will conform to WMO standards (WMO GAMP 2012) which ensures the first 3 requirments (in table 3) are met and in addition, the Cost Action metadata will facilitate links to external online sources. For the PIS case study, I4 is partially achieved as data is aggregated with other tools (pheromone traps, light traps, volumetric spore catchers, visual examinations of host plants and harmful organisms), processed and calculated data. The Reusability metric requires that the metadata specification is as exhaustive as possible (many relevant attributes), usage license, and data provenance. To maximize reusability, the Cost Action plans to include automated data quality checks and the inclusion of gap filling methods.

4 Summary

The FAIRNESS Cost action intends to enhance FAIRNESS by improving standardization and integration between databases and datasets of micrometeorological measurements that are part of research projects or local/regional observational networks established for special purposes (agrometeorology, urban microclimate monitoring). The PIS case study provides an example of an initial FAIRness self assessment and possible implementations. Additionally, the distinction between Open data and FAIR is addressed with a discussion of all 16 FAIR metrics in order to clarify these, sometimes, mismatching topics.

The actual return on investment in the implementation of FAIR data and RRI principles is almost impossible to assess due to unquantifiable elements such as 1) the value of improved research quality; 2) higher inclusiveness of low performing countries and individuals in STEM research; and 3) reduction of "crowding out" effect on high performing countries and other indirect positive spill-over effects in the longer term.

It is important to remember that FAIR data supports RRI but at the same time, RRI practitioners are making data FAIR. This is the positive feedback we are hoping to achieve. Therefore, our duty as professionals, whose work and results are directly or indirectly related to the general population and everyday life, is to embrace RRI principles, ensuring that this is the sole mechanism for conducting research.

Acknowledgements

The study was funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Grant No. 451-03-68/2022-14/200117) and H2020 - Co Change project (Grant No.873112). Micrometeorological measurements and phenological observations in crop canopies, orchards and vineyards by Forecasting and Reporting Service for Plant Protection of the Republic of Serbia (PIS) are financed by the Ministry of Agriculture, Forestry and Water Management of the Republic of Serbia and Provincial Secretariat for Agriculture, Water management and Forestry of the Vojvodina province, Republic of Serbia. Mark Roantree is funded by Science Foundation Ireland under grant number SFI/12/RC/2289-P2. The authors would like to acknowledge the CA20108 Cost Action - FAIRNESS.

References

Jones, S., Grootveld, M., 2017. How FAIR are your data? https://doi.org/10.5281/zenodo.5111307

Lalic, B., Marcic, M., Sremac, A.F., Eitzinger, J., Koci, I., Petric, T., Ljubojevic, M., Jezerkic, B., 2020. Landscape Phenology Modelling and Decision Support in Serbia. In: W. Mirschel, V.V. Terleev, K.-O. Wenkel (eds.), Landscape Modelling and Decision Support, Innovations in Landscape Research, Springer International Publishing, Cham, Switzerland, 567–593.

D Milošević, B Lalić, S Savić, B Bechtel, M Roantree, S Orlandini 2022. FAIRNESS Project-FAIR NEtwork of micrometeorological measurements, EGU22-971, Copernicus Meetings, March 2022.

Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L.O.B., Wilkinson, M.D., 2017. Cloudy, Increasingly FAIR; Revisiting the FAIR Data Guiding Principles for the European Open Science Cloud. Information Services & Use, 37, 49–56. DOI 10.3233/ISU-170824

Novitzky, P., Bernstein, M.J., Blok, V., Braun, R., Chan, T.T., Lamers, W., Loeber, A., Meijer, I., Lindner, R., Griessler, E., 2020. Improve Alignment of Research Policy and Societal Values. The EU Promotes Responsible Research and Innovation in Principle, But Implementation Leaves Much to be Desired. Science, 369, 6499, 39–41. DOI: 10.1126/science.abb3415

PwC EU Services, 2018. The cost of not having FAIR research data. DOI 10.2777/02999

Statista Research Department, 2022. Amount of data created, consumed, and stored 2010-2025. (https://www.statista.com/statistics/871513/worldwide-data-created/#:~:text=The%20total%20amount%20of%20data,replicated%20reached%20a%20new%20high.)

Koers, H., Gruenpeter, M., Herterich, P., Hooft, R., Jones, S., Parland-von Essen, J., Staiger, C., 2020. Assessment report on 'FAIRness of services' (1.0). Zenodo. https://doi.org/10.5281/zenodo.3688762

Wilkinson, M.D., Dumontier,M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., Hoen, P.A.C. 't, Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data, 3, 160018. https://doi.org/10.1038/sdata.2016.18

Wilkinson, M.D., Dumontier, M., Sansone, S.-A., da Silva Santos, L.A.B., Prieto, M., Batista, D., McQuilton, P., Kuhn, T., Rocca-Serra, P., Crosas, M., Schultes, E., Wlkinson, M.D., Dumontier, M., Sansone, S.A. et al., 2019. Evaluating FAIR maturity through a scalable, automated, community-governed framework. Sci Data, 6, 174. https://doi.org/10.1038/s41597-019-0184-5

Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., Courtot, M., Crosas, M, Dumontier, M., Evelo, C. T., Goble, C., Guizzardi, G., Hansen, K.K., Hasnain, A., Hettne, K., Heringa, J., Hooft, R.W.W., Imming, M., Jeffery, K.G., Kaliyaperumal, R., Kersloot, M.G., Kirkpatrick, C.R., Kuhn, T., Labastida, I., Magagna, B., McQuilton, P., Meyers, N., Montesanti, A., van Reisen, M., Rocca-Serra, P., Pergl, R., Sansone, S.-A., da Silva Santos, L. O. B., Schneider, J., Strawn, G., Thompson, M., Waagmeester, A., Weigel, T., Wilkinson, M. D., Willighagen, E. L., Wittenburg, P., Roos, M., Mons, B., Schultes, E., Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R. Schultes, E., 2020. FAIR principles: Interpretations and implementation considerations. Data Intelligence, 2, 10–29. doi: 10.1162/dint_r_00024

WMO GAMP 2012, Guide to Agricultural Meteorological Practices (GAMP) (WMO-No.134) Edition: 2010, WMO, 2012.