HORIZON 2020
ICT - Information and Communication Technologies

Deliverable D3.5
Report on Open Data and Reproducibility Research

Project Acronym:	**EMPOWER**
Project Full Title:	**EMpowering transatlantic PlatfOrms for advanced WirEless Research**
Grant Agreement:	**824994**
Project Duration:	**42 months (Nov. 2018 - Apr. 2022)**
Due Date:	**30 April 2022 (M42)**
Submission Date:	**28 July 2022**
Dissemination Level:	**Public**

Disclaimer
The information, documentation and figures available in this deliverable, is written by the EMPOWER project consortium under EC grant agreement 824994 and does not necessarily reflect the views of the European Commission. The European Commission is not liable for any use that may be made of the information contained herein.

Executive summary

Repeated research is an important part of scientific methodology. Repeating or replicating research not only brings an independent perspective to the investigation, but also provides a basis for comparing different approaches and extending known results. The process of repeating others' research also allows us to learn, not only from their insights but also their experimental methodology. Replication is thus an important facet of scientific debate.

Up until recently repeating computer science research, while desirable in principle, was often difficult in practice. This is due to several well-identified factors such as the lack of incentives (e.g., in publication venues) that value such effort, lack of common research platforms and tools, and the difficulty of packaging the requisite digital artifacts in an easily shareable form and the mechanisms to combine and publish them. Change is needed.

A first set of nearer-term activities include a series of workshops educating students and involving faculties in the art of reproducible research, and identifying and developing a set of tools and best-practice-techniques for enabling reproducible research. We wanted to invite submissions that repeat results published in past papers and provide a brief analysis of the results as well as the ability to replicate them, similar to "Xen and the Art of Repeated Research" by Clark et al. The longer-term objective is to reach out to the main stakeholders, professional societies and conferences, with recommendations arguments and potential solutions for embedding reproducibility in the research life-cycle.

Because of the Covid situation, we had to back up to a different methodology, mostly using the EMPOWER NetworkingChannel as the forum for discussing this important topic. These will be organized virtually but additionally collocated with conferences and workshops when possible, such as our participation to the CNERT workshop on reproducibility at Infocom 2020 or at the IEEE future networks workshop in February 2022.

In addition, the newly established ESFRI SLICES initiative, now on the ESFRI 2021 roadmap, is developing a framework to cope with the full research life-cycle and is designing an understanding and framework for aligning with the best practices used in Open Science as well as an articulation with the EOSC initiative. This provides an interesting context to develop further this as a joint EU/US topic of mutual interest that should be pursued in the future.

Table of Contents

1. Introduction

Advanced platforms are meant to support the discovery process in digital infrastructures. Unfortunately, the COVID situation has demonstrated that providing evidence is a quite difficult and complex process. It is based on the very formal methodology, which is often founded on experimentally driven research.

A scientific instrument is not limited to a physical infrastructure but should also provide tools to support the full research life-cycle. In particular, it deals with open and FAIR data as well as reproducibility of experiments. The later has been a hotly debated topic but with little progress. It is also clearly stated as a requirement in the missions of the NSF funded PAWR projects as well as in the Horizon Europe projects (at least for the data management and open data part).

This is somehow the ambition of the European EOSC (European Open Science Cloud) initiative. Researchers and research stakeholders nowadays require that research data is made available for other researchers to examine, experiment and develop further. Additionally, preserving the data in conjunction with how conclusions from the data were drawn, accelerates the discovery process, enable easier reproducibility of the results and thus supports evidence. It is then necessary to develop policies and procedures for regulating the management and publication of research data in order to make them interoperable and widely available

In Europe, it is recommended to conform with the European Open Access policy, Open Research Data Pilot and FAIR principles in producing and managing research data.

We introduce and discuss some of the issues related to this important field of experimentally-driven research. It emphasizes the complexity of the overall process and advocate to add this topic as a joint EU-US activity for the future.

2. The full research life-cycle

Experimentally-driven research should be grounded on a solid methodology that is understood and implemented by other disciplines and we will certainly benefit from the experience of those fields that have already developed some know-how and tools. This is somehow the ambition of the European EOSC (European Open Science Cloud) initiative. As a consequence, it shows that one should not only target the deployment of the instrument/facility but as importantly, address the full research life-cycle, including open data, data management and reproducibility.

Researchers and research stakeholders nowadays require that research data is made available for other researchers to examine, experiment and develop further. Additionally, preserving the data in conjunction with how conclusions from the data were drawn, accelerates the discovery process, enable easier reproducibility of the results and thus supports evidence. It is then necessary to develop policies and procedures for regulating the management and publication of research data in order to make them interoperable and widely available

In Europe, it is recommended to conform with the European Open Access policy, Open Research Data Pilot and FAIR principles in producing and managing research data. This requires defining appropriate metadata (including compatible experiment description) on the data produced by or integrated into the infrastructure with the objective to ensure eventually data accessibility, reusability and interoperability with data produced by similar infrastructures/experiments for enabling complex experiments and multi-domain research.
Alignments with the relevant recommendations such as the ones published by EOSC FAIRs project, GO FAIR initiative and RDA for FAIR data management, and general European Open Access to research publications and Open Research Data Pilot policies, are of utmost importance.

The FAIR[1] (Findable, Accessible, Interoperable, and Reusable) Data Principles were developed to be used as guidelines for data producers and publishers, with regards to data management and stewardship. One important aspect that differentiates FAIR from any other related initiatives is that they move beyond the traditional data and they place specific emphasis on automatic computation, thus considering both human-driven and machine-driven data activities. Since their publication, FAIR became widely accepted and used.

To this end, advanced platforms should fully endorse and adopt the FAIR principles, in order to enable and foster the data-driven science and scientific data-sharing in this area.

[1] 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 [Last accessed 6 May 2022]

3. FAIR and open data

Understanding the data collected and processed within an advanced platform becomes essential to understand data usage from the target user groups. This should allow to develop an appropriate information model that will represent the data collected from the platform, experimental equipment and applications. We consider that the datasets generated by a platform, involving hardware and software infrastructure, can be roughly organized into five main categories:

- **Observational Data:** will be collected using methods such as surveys (e.g., online questionnaires) or recording of measurements (e.g., through sensors). The data will include mostly data related to signal or performance measurements, and network or service log data that allow for experiment evaluation and reproducibility.
- **Experimental Data:** where researchers introduce an intervention and study the effects of certain variables, trying to determine their impact.
- **Simulation Data:** is generated by using computer models that simulate the operation of a real-world process or system. These may use observational data.
- **Derived Data:** involves the analysis (e.g., cleaning, transformation, summarization, predictive modelling) of existing data, often coming from different datasets (e.g., the results of two experiments), to create a new dataset for a specific purpose.
- **Metadata:** concerns data that provides descriptors about all categories of data mentioned above. This information is essential in making the discovery of data easier and ensuring their interoperability.

In addition, an advanced open platform, will promote interoperability, thus non-proprietary, unencrypted, uncompressed, and commonly used by the research community formats should be adopted. In addition, the platform end users should have the ability to decide on a suitable license and attach it to their data.

It is important for the platform to derive a preliminary estimation of its dimensions (in number of single users as well as the data generated by the experiments in order to size the system needs in storage).

Data management raises several important concerns related to openness and privacy. It copes with the issues dealing with the governance of the data. Certainly, it will require to setup a data management framework to support the efficient and effective operation of the infrastructure and achieve the data dimension objectives. To accomplish this, the data management framework should clarify its own design goals, which are summarized below and presented in Figure 1:

- **Data Governance**: A systemic and effective Data Governance structure to support the data management operations through a hierarchical structure with appropriate roles (e.g., Data Manager, Data Protection Officer and Metadata administrator), implement all related policies and processes, and adopt standards and leading practices.
- **Data Architecture**: An agile Data Architecture that can perform efficiently to fulfill the infrastructure requirements, scales gracefully to accommodate for increased workloads, is flexible to integrate new processes and technologies, and is open to interact with other systems and infrastructures.
- **Data Quality**: Appropriate data transformation mechanisms to ensure Data Quality across multiple dimensions (e.g., accuracy, completeness, integrity), in order to improve data utility (e.g. further processing, analysis).
- **Metadata**: Appropriate metadata management mechanisms to facilitate collaboration between users by providing the means to share their data and also support FAIR data.
- **Interoperability**: Facilitate seamless interaction with other systems and infrastructures.
- **Analytics**: Deployment of statistical, machine learning and artificial intelligence techniques to draw valuable insights from data and appropriate visualisation techniques to interpret them.
- **Data Security:** Mechanisms to protect data from unauthorized access and protect its integrity.
- **Privacy**: Strict controls to manage the sharing of data, both internally and externally.

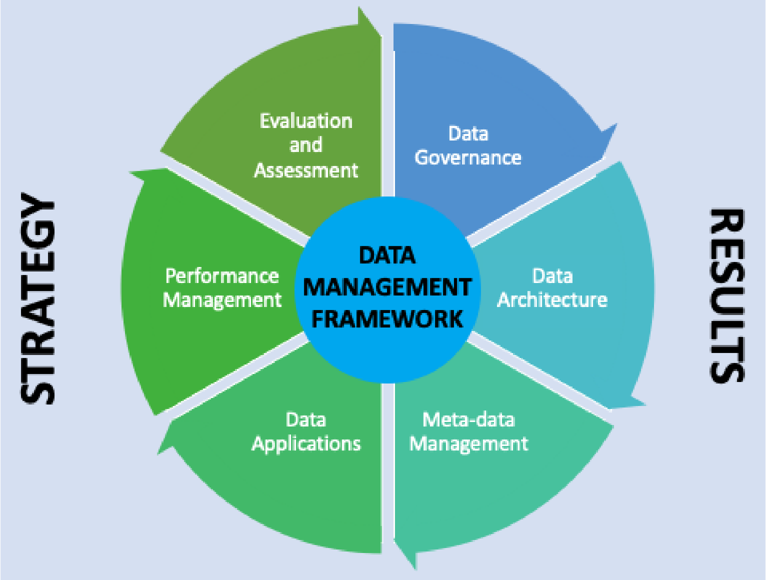


Figure 1: Data Management Framework

4. Interoperability with EOSC

EOSC has established itself as an important pillar in Europe for the implementation of Open science concepts by accelerating the adoption of the FAIR data practices among researchers in the European Union. Integration of the platform into European Research Infrastructure via EOSC will facilitate data sharing and reuse among the platform partners and the larger European researchers' community. This is particularly true for the European platforms but it could also provide a guideline for the cooperation on this topic with the US, as shown in recent talks from Kate Keahey[2].

Since aligning EU and US platforms will mobilize experimental research platform by jointly utilizing the geographically dispersed computing, storage and networking infrastructures, it is highly important that the different facilities interacting in the experimental workflow are interoperable with each other. Similarly, existing research needs to be accessible and directly pluggable to the platform's services and sites. It is thus necessary that resource description, availability, execution and data exchanges are effective. This can only be assured if a common interoperability framework is adopted across the platforms ecosystem so that different subsystems have a common understanding of resources, data/metadata and are on the same page with respect to the licensing, copyright and privacy requirements. The platforms infrastructure should be designed to ensure compatibility and integration with EOSC, and be ready to offer advanced ICT infrastructure services to other RIs and projects, with the special focus on the FAIR data management and exchange.

Therefore, it is of utmost importance to design the integration framework of the research platforms with EOSC in such a way that the data exchange between the platforms and EOSC is interoperable for scientific workflow management for data storage, processing and reuse.

Interoperability is an essential feature of EOSC ecosystem as a federation of services and data exchange is unthinkable without interoperability among different EOSC constituents. The meaningful exchange and consumption of digital objects is necessary to generate value from EOSC which can only be realized if different

[2] Chameleon Reproducibility Project, presentation of Isabel Brunkan, https://www.youtube.com/watch?v=zOLpNBurQD4, [Last accessed 6 May 2022]

components of the EOSC ecosystem (software/machines and humans) have a common understanding of how to interpret and exchange them, what are the legal restrictions, and what processes are involved in distribution, consumption and production of them. To facilitate this, an EOSC interoperability framework (EOSC-IF) should be defined as a generic framework for all the entities involved in the development and deployment of EOSC. The EOSC interoperability framework is a set of *not-so-specific* guidelines to ensure smooth integration of infrastructure services and seamless exchange of research data across the EOSC ecosystem. EOSC-IF is derived from the European Interoperability framework that defines interoperability of an information technology system by four key elements, i.e.., technical, semantic, organization and legal interoperability.

The FAIR principles, federated resource and user management and legal compliances pertaining to privacy, licensing and governance by European Commission are at the core of the EOSC-IF framework.

The interface should be built upon the foundations led by the European Interoperability Reference Architecture (EIRA), where interoperability is classified at four layers, namely: (i) technical, (ii) semantic, (iii) organizational; and (iv) legal.

FAIR Digital Object (FDO) is the core building block of EOSC-IF. Here, Digital Object refers to the kind of objects that allow binding all critical information about any entity. In EOSC, a digital object can be research data, software, scientific workflows, hardware designs, protocols, provenance logs, publications, presentations, etc., as well as all their metadata (for the complete object and for its constituents). Figure 2 shows a schematic of FDO. An FDO should conform to all the four layers of interoperability introduced earlier in this document by following the FAIR guidelines.

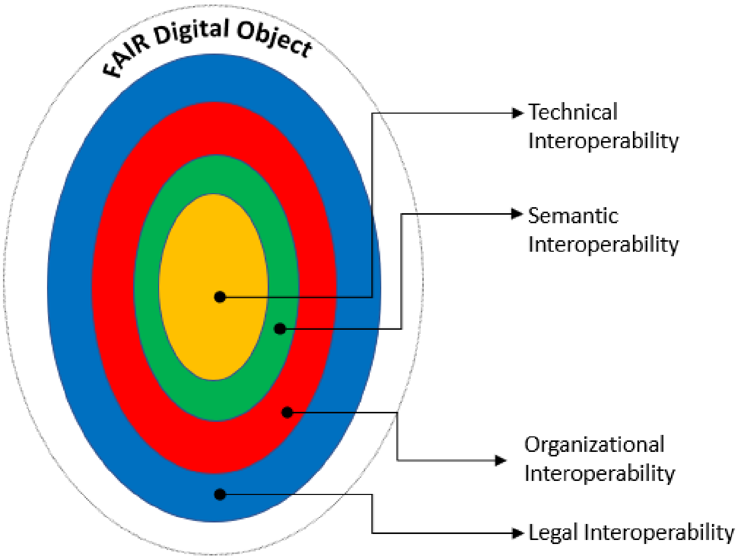

FAIR Digital Object
Technical Interoperability
Semantic Interoperability
Organizational Interoperability
Legal Interoperability

Figure 2: EOSC FAIR Digital Object illustration

FDO (FAIR Digital Object) and PID (Persistent Identifier) are key components of the EOSC architecture and supporting federated data infrastructure that ensures consistent implementation of the FAIR data principles. FAIR principles are realized through FAIR Digital object, and PID services and infrastructure provide facilities to access and manipulate FDO. The FDO Forum provides the following definition of the FDO[3]: “A FAIR digital object is a unit composed of data and/or metadata regulated by structures or schemes, and with an assigned globally unique and persistent identifier (PID), which is findable, accessible, interoperable and reusable both by humans and computers for the reliable interpretation and processing of the data represented by the object.”

[3] FAIR Digital Object Framework, https://www.go-fair.org/today/fair-digital-framework/, [Last accessed 6 May 2022]

According to FDO Forum, the FAIR Digital Object concept brings together FAIR guiding principles and Digital Object that supports interoperability across existing and evolving data regimes/frameworks using mechanisms of the structured machine-readable identifiers and principles of persistent binding for digital object of different types. It is important to mention that activities currently coordinated by FDO Forum are including essential results and current activities at the RDA (Research Data Alliance) on Data Types and Data Types Registries, Data Factories, others, and are aligned with EOSC technical development what is reflected in the EOSC Architecture and EOSC Interoperability Framework.

The following main requirements to FDO services and infrastructure defined in the FDO Framework:

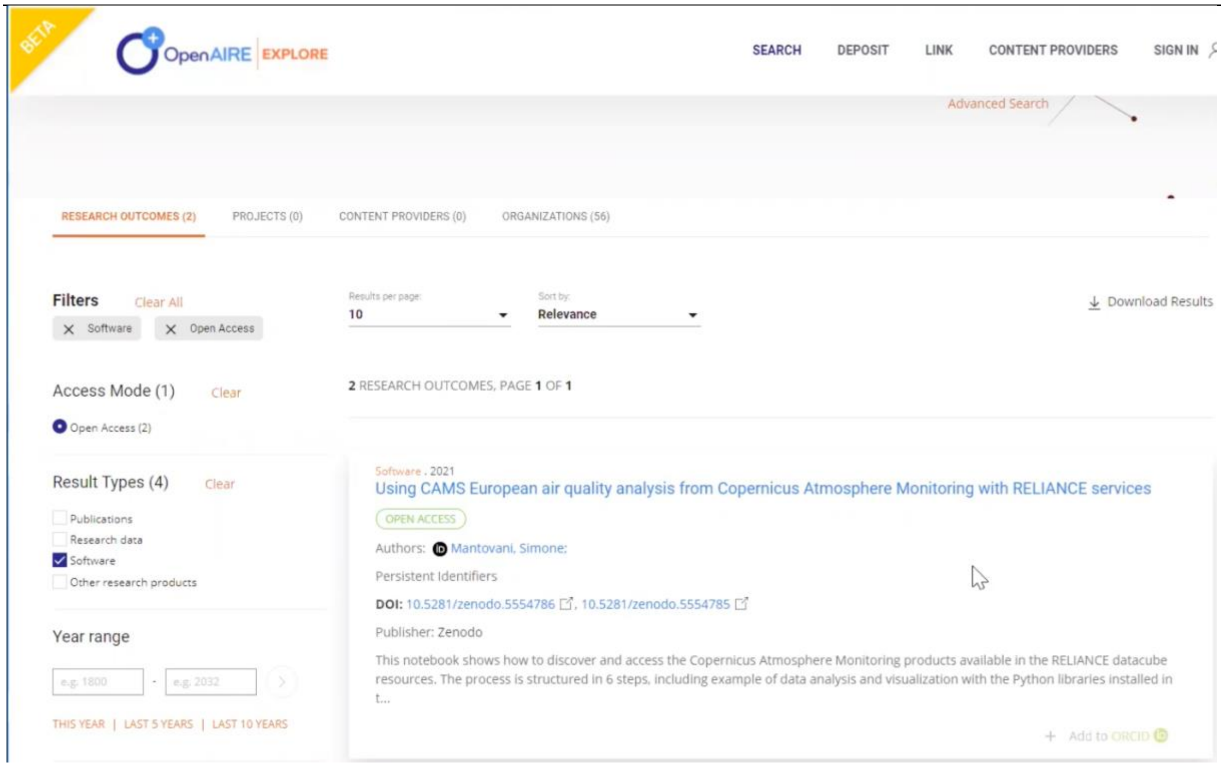
- General requirements include machine actionability, technology independence, persistent binding, abstraction and structured hierarchical encapsulation, compliance with standards and community policies;
- FDO is identified by PID; there are possible multiple PID frameworks defined by PDI scheme, namespaces, ontologies or controlled vocabularies;
- A PID resolves to a structured record (PID record) with attributes that are semantically defined within a (data) type ontology (which may be defined for different application or science domains);
- The structured PID record includes at least a reference to the location(s) where the FDO content and the type can be accessed, and also metadata describing FDO can be retrieved;
- Each FDO is accessed via API by specifying PID and additionally point. API must support basic operation with FDO: Create, Read, Update, Delete (referred to as CRUD), however subject to access control and policy;
- Metadata used to describe FDO properties should use standard semantics and registered schemes to allow machine readability and actionability.

5. Illustration of a full-research life cycle

In order to realize the vision of FAIR research, supporting the full research lifecycle, let’s consider a simple example borrowed from another field of research and illustrated by the Reliance[4] project. Reliance delivers a suite of innovative and interconnected services that extend EOSC’s capabilities to support the management of the research lifecycle within Earth Science Communities and Copernicus Users. Consider core services provided to the research community, that could be data, software publications, others. These core services are named after research objects that are for use by the experimenters and share by the experimenters. As an illustration gently provided by Anne Fouilloux[5], assume that you are doing some research related to the Copernicus air quality. Starting from OpenAIRE EXPLORE, we search for "Copernicus air quality" and you will find all the associated resources, mostly publications and only 2 software (Figure 3). The reason is that to be "classified" as "Software", we have to add specific metadata when publishing.

[4] Reliance project website, https://www.reliance-project.eu/, [Last accessed 6 May 2022]

[5] Demo showing what RELIANCE project has achieved on Open Science, FAIR and EOSC, https://zenodo.org/record/5906803#.Ym-fxPNBzBs, [Last accessed 6 May 2022]


BETA

OpenAIRE EXPLORE

SEARCH DEPOSIT LINK CONTENT PROVIDERS SIGN IN

Advanced Search

RESEARCH OUTCOMES (2) PROJECTS (0) CONTENT PROVIDERS (0) ORGANIZATIONS (56)

Filters Clear All

✕ Software ✕ Open Access

Results per page: 10

Sort by: Relevance

Download Results

Access Mode (1) Clear

● Open Access (2)

2 RESEARCH OUTCOMES, PAGE 1 OF 1

Result Types (4) Clear

□ Publications
□ Research data
☑ Software
□ Other research products

Year range

e.g. 1800 - e.g. 2032

THIS YEAR | LAST 5 YEARS | LAST 10 YEARS

Software . 2021

Using CAMS European air quality analysis from Copernicus Atmosphere Monitoring with RELIANCE services

OPEN ACCESS

Authors: Mantovani, Simone;

Persistent Identifiers

DOI: 10.5281/zenodo.5554786, 10.5281/zenodo.5554785

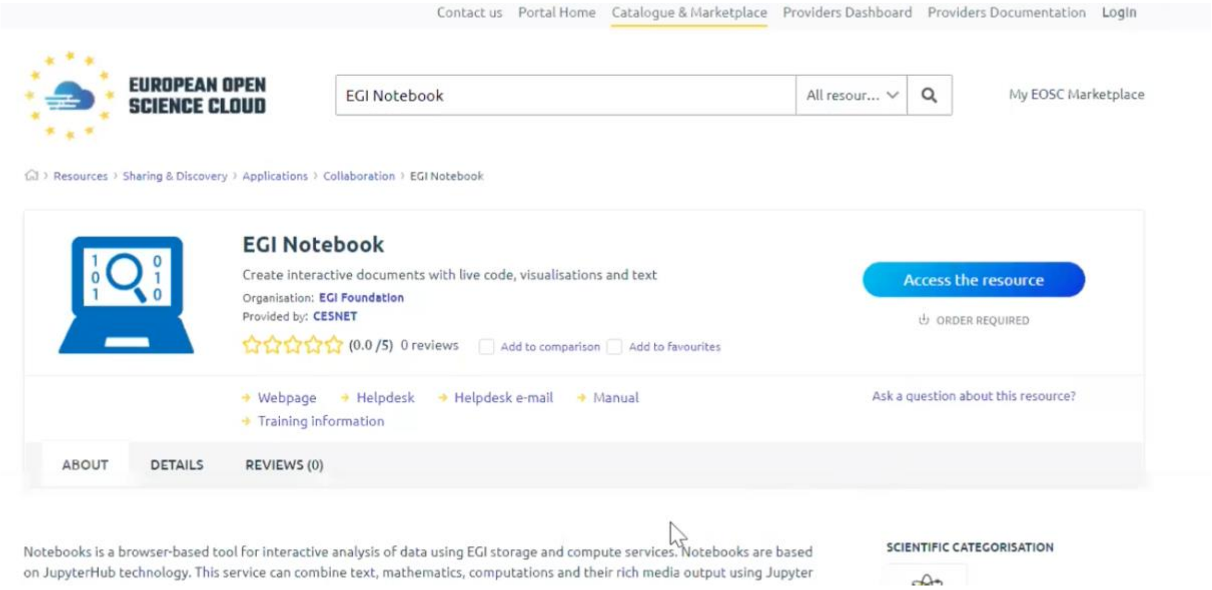
Publisher: Zenodo

This notebook shows how to discover and access the Copernicus Atmosphere Monitoring products available in the RELIANCE datacube resources. The process is structured in 6 steps, including example of data analysis and visualization with the Python libraries installed in t...

+ Add to ORCID

Figure 3: Search for "Copernicus air quality"

You are looking for a software, because someone has produced a software taking research data as input and producing a map of the air quality in a given region as an output. You find the software and with the software comes a set of additional metadata. The "Software" we found is an "EOSC Jupyter notebook" with a DOI and additional metadata so that OpenAIRE explore can "associate" it to a specific EOSC service, namely EGI Notebook (Figure 4). When we click on "EOSC Service: EGI Notebook", we are re-directed directly to the service that has been used to generate the original scientific results we found in OpenAIRE explore. The Jupyter notebook uses CAMS European air quality analysis from Copernicus Atmosphere Monitoring Service.


Contact us Portal Home Catalogue & Marketplace Providers Dashboard Providers Documentation Login

EUROPEAN OPEN SCIENCE CLOUD

EGI Notebook | All resour... | My EOSC Marketplace

Resources › Sharing & Discovery › Applications › Collaboration › EGI Notebook

EGI Notebook

Create interactive documents with live code, visualisations and text

Organisation: EGI Foundation

Provided by: CESNET

☆☆☆☆☆ (0.0 /5) 0 reviews □ Add to comparison □ Add to favourites

Access the resource

ORDER REQUIRED

Webpage Helpdesk Helpdesk e-mail Manual Training information

Ask a question about this resource?

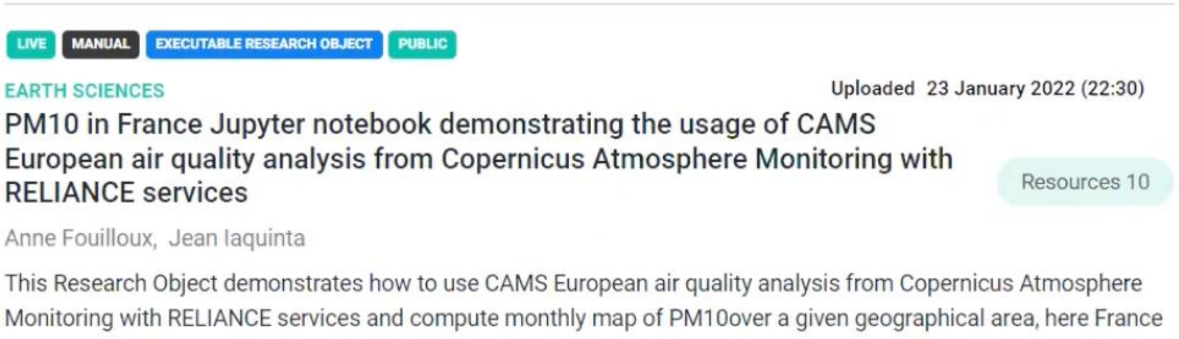
ABOUT DETAILS REVIEWS (0)

Notebooks is a browser-based tool for interactive analysis of data using EGI storage and compute services. Notebooks are based on JupyterHub technology. This service can combine text, mathematics, computations and their rich media output using Jupyter

SCIENTIFIC CATEGORISATION

Figure 4: EGI Notebook (Jupiter Notebook)

You now have access to the software that will execute exactly what has produced this research data. We can re-execute the Jupyter notebook but more importantly we can create derivative work. What you are willing to do, at first, is to reproduce the results. On the other hand, you would like to take your own data, use the same process and produce your own new results. The last step is that you go to the service, which is named Rohub[6] (Research Object Hub). And then you bundle your different resources, like the Jupyter notebook that you have used, the data that you have exploited, and the output that you have produced. You can now publish this research outcome, an "executable Research Object» as your own contribution made available to the community, defined as PM 10 in Figure 5.



EARTH SCIENCES

Uploaded 23 January 2022 (22:30)

PM10 in France Jupyter notebook demonstrating the usage of CAMS European air quality analysis from Copernicus Atmosphere Monitoring with RELIANCE services

Anne Fouilloux, Jean Iaquinta

This Research Object demonstrates how to use CAMS European air quality analysis from Copernicus Atmosphere Monitoring with RELIANCE services and compute monthly map of PM10over a given geographical area, here France

Figure 5: Executable research object

This full research-life cycle is really important, otherwise, the result that you produce cannot be published, because it simply cannot be reproduced.

There exists data initiative in our field, like the ACM good experiment level, it is nice and ambitious. But it does not yet fully align with the best practices in other fields of research.

[6] Research Object Hub (RoHub) website, https://reliance.rohub.org/ [Last accessed 6 May 2022]

6. Related events

The Covid situation forces us to transform some of the planned activities into virtual. Naturally, we used theNetworkingChannel to organize relevant activities in order for them to benefit from the large dissemination and audience of our channel.

The table below presents two events, fully dedicated to this topic.

EVENT NAME	RELATION TO EMPOWER	DATES	PLACE	OUTPUTS
Experiments in the Edge to Cloud Continuum	EMPOWER presentation by Serge Fdida	October 27, 2021	Virtual Means: theNetworkingChannel *Total registrations: 172* *Total attendees: 83* *Replay of the recordings: 132*	The increasing popularity of IoT devices allows us to communicate better, interact better, and ultimately build a new type of a scientific instrument that will allow us to explore our environment in ways that we could only dream about just a few years ago. This disruptive opportunity raises a new set of challenges: How should we manage the massive amounts of data and network traffic such instruments will eventually produce? What types of environments will be most suited to developing their full potential? What new security problems will arise? And finally: what are the best ways of leveraging intelligent edge to create new types of applications? In a research area that creates a new deployment structure, such questions are too often approached only theoretically for lack of a realistic testbed: a scientific instrument that keeps pace with the emergent requirements and allows researchers to deploy, measure, and study relevant scientific phenomena. To help create such instrument, the NSF-funded Chameleon testbed, originally created to provide a platform for datacenter research, has now been extended to support experiments in cloud to edge. In this presentation, we will first briefly describe the Chameleon testbed and then explain how it was extended to support edge to cloud experiments. We will introduce CHI@Edge, an extension to CHameleon Infrastructure (CHI), give a demonstration of how users can easily create an experiment spanning edge devices and significant cloud resources from one Jupyter notebook, and give examples of edge to cloud projects in research and education projects that our users created.
Network Datasets: what exists, and what are the problems?	Organized by EMPOWER	March 30, 2022	Virtual Means: *theNetworkingChannel* *Total registrations: 209*	Public datasets of network measurements have been created and made available for several decades. These datasets are interest to networking researchers (who are interested in workloads, topologies, and other characteristics of deployed networks

			Total attendees: 121 *Replay of the recordings: 151*	in/across the backbone, access, home and mobile networks), students who want to learn what “real” networks look, policy makers, and more. This NetworkingChannel event has two parts. The first part, we’ll identify and discuss public datasets of interest and their use. In the second part, we’ll identify some of the challenges of working with such datasets (including the difficulty of analyzing/comparing data longitudinally, the “ageing” of data) and the challenges of obtaining industry data (which may have significant commercial and proprietary value) and solutions to that challenge, such as benchmarks and workload models.
https://futurenetworks.ieee.org/conferences/fn-testbed-workshop	Talks by Brecht Vermeulen (iMEC), Serge Fdida (SU), Manu Gosain (PWAR office) …	7-8 February, 2022	Virtual	Testbeds are a critical part of network generation evolution and complement the development of standards. Use cases are facilitated with a number of innovations starting from the Physical layer with mmWave, mMIMO and spanning other layers of the network protocol stack. Given the increased complexity of the next generation of communication systems and skyrocketing development costs, the importance of publicly available testbeds is quickly becoming critical for researchers and developers to get access to state-of-the-art infrastructure in order to prototype and validate their novel ideas. Lessons learnt from various experiments help to ratify the standards. Testbeds also help as a catalyst for deployment of advanced wireless systems. Hence, it is essential to have experimental testbeds, where functional and end-to-end testing can be performed. This workshop explores diversity requirements as well as the challenges and enabling techniques towards the development of scalable testbed facilities for future networks to provide a standardized approach and best practices for collaboration. The workshop will feature two keynote and fifteen invited talks that will introduce a range of testbeds from academia and industry and will cover a range of issues related to experimental evaluation of existing and future wireless systems
IEEE Infocom 2020, CNERT workshop	CNERT workshop on reproducibility Conclusion	July 6, 2020	Virtual	The following panel was organized: Panelists: • Georg Carle, Technical University of Munich, Germany • Serge Fdida, Sorbonne Université, France • Kate Keahey, University of Chicago and Argonne National Laboratory, US

				• Deep Medhi, US National Science Foundation • Rob Ricci, University of Utah, US • Gwendal Simon, Huawei Technologies, France The moderator of the panel is one of our TPC-chairs: Michael Zink, University of Massachusetts at Amherst, US Questions addressed by the panel: • The community has been mainly focusing on reproducibility for network and systems research. Do you see other CISE research areas that should be supported in the future? • Do we need standards to facilitate reproducibility in the future? • What steps need to be taken to educate the next generation of CISE researchers to give them the ability to perform reproducible research? • Reproducibility takes more time and effort! What metrics for promotion and tenure (especially for junior faculty) have to be adjusted by university administrations and our research community to encourage reproducibility? • How long is the lifecycle of a reproducible experiment? What steps need to be taken to provide provenance for that lifecycle? • Specific test beds might be needed to reproduce results. What is the required lifecycle of such test beds to guarantee that results can be reproduced? • How will reproducible data be made broadly accessible and what policies for data management should be applied? • Are there requirements for reproducibility the funding agencies should enforce?

7. Conclusion

A scientific instrument, a test platform, is not limited to a testing infrastructure but should also onboard the full-research life cycle. This includes difficult issues such as open and fair data, data management, liaison with EOSC and reproducibility.

Some initiatives have developed in the field but with limited impact for obvious reasons related to the complexity of the problem and the lack of reward in engaging resources into this task. Nevertheless, aligning with the best practices as developed in other more mature domain of sciences provides incentives for dealing more seriously with this concern.

In particular, the FAIR and EOSC framework in Europe could provide a catalyst and methodology to progress in this domain. EMPOWER shed lights on this topic by organizing virtual events and adding this topic on the list of domains of mutual interests between the US and EU. Hopefully, this will achieve some tangible results in the future.

Annex 1: List of recommendations for further implementation

Our community is not yet sensitive to the full research life-cycle. This is unfortunate and this mindset is difficult to change. This is partly due to “publish or perish”, the competition for publication, even more important, the competition for being cited. As a consequence, the validation methodology of the published results is not always robust and often, reproducibility is hard if not possible.

There has been different attempt to change this situation, as for example, the ACM reproducibility badges. Several conferences and venues have implemented this framework. However, they do it once as it is a huge burden for the committee in charge. The only solution to scale it up is to transfer the burden to the authors. Yet, there is little incentive for the authors to deliver this information as the load is high and the reward low.

We suggest the following, knowing that some will be difficult to deploy:

- Test facilities, worldwide, should implement the full research life-cycle;
- As a consequence, they have to provide a repository to access fair data, as well as digital objects. They can follow the best practices developed by other communities and integrated in EOSC as described in D3.5;
- Authors should be incentivized. The only way to achieve this objective is through regulation. Therefore, we encourage and advise scientific societies such as ACM and IEEE to enforce a Data and Digital objects publication together with their published papers when this is relevant.