STARS EU

Common Open Data Repository

Title	STARS EU COMMON OPEN DATA REPOSITORY
Version	V 0.2
Approval status	Final (approved by the Steering Committee on April 18th, 2024)
Date of issue	April 19th 2024
Author	Ernesto Pereda, PhD (University of La Laguna, WP leader)
Contributor(s)	Mr. José Manuel Erbez (Deputy director for open data an coordination, ULL Library)
Distribution list	
Abstract:	Open science is a critical element of EU policy and strategies, which improves dissemination, knowledge transfer to industry and society, and the transparency and reproducibility of research as part of our Dissemination and Impact, besides ensuring that the project outcomes and findings are publicly available on the STARS EU website. STARS EU research projects will progressively include data management plans. They will implement a dedicated repository compliant with the FAIR principles (M6), where all our research data will be available. The present document describes the steps carried out to implement such a repository.
Project Partners	1. Hanze University of Applied Sciences, Groningen (The Netherlands) 2. University West (Sweden) 3. Bremen University of Applied Sciences (Germany) 4. University of Franche-Comte (France) 5. Cracow University of Technology (Poland) 6. Silesian University of Opava (Czech Republic) 7. Polytechnic Institute of Braganca (Portugal) 8. University of La Laguna (Spain) 9. Alexander Moisiu University Durres (Albania)

CONTENTS

EXECUTIVE SUMMARY

This document describes the process that led to the creation of a common open data repository for the STARS EU alliance, explaining the standpoint from which each of the different partners started and the decisions that we made to reach the final product.

The common Open Data Repository (ODR) must include the open data from the alliance's research project in a way that is compliant with the FAIR principles. At the same time, it must be flexible enough to allow for a future upgrade to potentially include all the alliance's research production. Besides, it should be common to all alliance members but also respectful of the existing ODRs of those universities that already have their own.

Based on these premises, and after carefully analyzing the different solutions considered, we justify the adopted one, based on Zenodo's communities' tool, and comment on the planned future work to extend the scope of the ODR beyond a repository of open data to further promote and make publicly available all the scientific production and research data of STARS EU partners.

1. INTRODUCTION

The European Commission considers open science a policy priority and the standard working method in its research and innovation funding programs. The EU's Open Science policy aims to improve research quality, efficiency, and responsiveness by sharing knowledge and data as early as possible in the research process. In the STARS EU alliance, we are aware of the importance of making the results of our research publicly available to enhance knowledge transfer, contribute to regional development, and improve the dissemination of our activities. Therefore, we included in WP8 a specific task (8.3) to be delivered in M6 of the project, which involves setting up a standard open data repository for the alliance, where all the data from our research projects will be available, which is compliant with the FAIR principles. Such repositories should be familiar to all partners and different from the individual repositories that each university may have. They should ideally gradually encompass the research data already included in these repositories.

The FAIR principles, which stand for Findable, Accessible, Interoperable, and Reusable, were published in 2016 to provide guidelines for improving the management and stewardship of scientific data. These principles emphasize machine actionability, meaning the capacity of computational systems to find, access, interoperate, and reuse data with minimal human intervention.

- *Findable*: Data and metadata are assigned a globally unique and persistent identifier, described with rich metadata, and registered or indexed in a searchable resource.
Accessible: Data and metadata can be retrievable by their identifier using an open, free, and universally implementable standardized communications protocol.
- *Interoperable*: Data and metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- *Reusable*: Data and metadata are richly described with a plurality of accurate and relevant attributes, released with a clear and accessible data usage license, associated with detailed provenance, and meet domain-relevant community standards.

The FAIR principles aim to optimize data reuse, which aligns with the objectives of the EU's Open Science policy.

.

2. WHY A COMMON OPEN DATA REPOSITORY?

As indicated before and described in the Project, the common Open Data Repository (ODR) aims to create a public space where the research projects and initiatives of the STAR EU alliance can be shared according to the FAIR principles.

2.1 CURRENT SITUATION FROM EACH PARTNER

Before describing the solution finally adopted, it is important to explain each partner's different approaches to disseminate their research results. This was the starting point from which we had to create a new, common ODR.

To gather information about the current situation of each partner, we carried out a survey (see ANNEX I) with the corresponding responsibility of each institution. The results can be found in *Table 1*

University	open data repository?	If yes, provide the internet address.
Instituto Politécnico de Bragança (IPB)	Yes	https://dados.ipb.pt/
Université de Franche-Comté (UFC)	Yes	https://search-data.ubfc.fr/
Hanze UAS	Yes	https://dataverse.nl/dataverse/hanze
Silesian University in Opava (SUA)	No	
Cracow University of Technology (CUT)	Yes	https://rodbuk.pl/
University "Aleksandër Moisiu" Durrës (UAMD)	No	
University West (UW)	No	
University of La Laguna (ULL)	Yes	https://research-data.ull.es/research-data/
HS Bremen (HSB)	No	

Table 1: Individual open data repository from each partner

As seen there, four partners do not currently have a public open data repository. The other five have mixed situations: some have custom-made ones (e.g., IPB), others have common repositories with universities of their region (UFC), or they have outsourced the repository to external companies (e.g., ULL with Elsevier showcase).

Although the deliverable only entailed the implementation of "*a dedicated repository that is compliant with the FAIR principles (M6), where all our research data will be available*", we also wanted to find out the situation regarding the repositories of scientific production, so that we can use it at a later stage to enhance the availability of all the research results from the alliances. As it turned out, the situation for such repositories/scientific portals was more homogeneous, with all partners having some showcase for their publications, as shown in Table 2. Yet the format of each scientific portal was again very different among partners.

University	Repository of scientific production?	If yes, provide the internet address.
Instituto Politécnico de Bragança (IPB)	Yes	https://bibliotecadigital.ipb.pt/
Université de Franche-Comté (UFC)	Yes, but partial	https://univ-fcomte.hal.science/
Hanze UAS	Yes	https://research.hanze.nl/
Silesian University in Opava (SUA)	Yes	https://is.slu.cz/publikace/?vysledek=5521;lang=en
Cracow University of Technology (CUT)	Yes	https://repozytorium.biblos.pk.edu.pl/
University "Aleksandër Moisiu" Durrës (UAMD)	Yes	https://www.journal-uamd.org/index.php/IJRD/issue/view/23
University West (UW)	Yes	http://hv.diva-portal.org/smash/search.jsf?dswid=-5307
University of La Laguna (ULL)	Yes	https://portalciencia.ull.es/?lang=en
HS Bremen (HSB)	Yes	Publications: https://www.suub.uni-bremen.de/home-english/refworks-and-publishing/publishing-documents/ https://media.suub.uni-bremen.de/cris/ou/ou00002/publications.html?startother=0&locale=en Research & Transfer projects: https://www.hs-bremen.de/en/research/research-and-transfer-profile/research-projects/

Table 2: Individual portals for the scientific production

Finally, we also asked about an institutional policy for opening access, which, as seen in Table 3, mimicked the results for the open data repository.

University	Has a policy for open access?	If yes, provide the link for the current document.
Instituto Politécnico de Bragança (IPB)	Yes	https://drive.google.com/open?id=1N_jsRzlg5Zy2GjYhl2MyrSFcvu8SQnk-
Université de Franche-Comté (UFC)	Yes	https://drive.google.com/file/d/1TaoGsuRNn9KU5Jl2l7sgBsO98rfULLcC/view?usp=sharing
Hanze UAS	Yes	https://drive.google.com/file/d/13gjnLhN3qTKl2mtc2TfRP3ODJZg7Vstl/view?usp=drive_link
Silesian University in Opava (SUA)	No	
Cracow University of Technology (CUT)	Yes	https://drive.google.com/open?id=1hPrlQ9H6Sx2zab7eR715XxD6J6j5_2ye https://drive.google.com/open?id=1GImURNRuJbTnFvle1nM62xDgj6s-9hOz
University "Aleksandër Moisiu" Durrës (UAMD)	No	
University West (UW)	No	The work with an Open Science and Open Access policy is in progress.
University of La Laguna (ULL)	Yes	https://ull-es.libguides.com/c.php?g=688073&p=4921755
HS Bremen (HSB)	No	

Table 3: Open access policy for each partner

2.2 THE ALTERNATIVES CONSIDERED

Given the situation described above, we considered the following alternatives:

1.- To create a common ODR only for those partners who currently do not have one

We quickly discarded this first option, as we thought it did not comply with the project's requirements, nor was it aligned with the concept and vision of a European University Alliance to use the Latin expression, "*E pluribus unum"* (from many, one).

2.- Externalize the common repository to a dedicated editorial company

In this case, we had different meetings with Elsevier, who oversees the open data from one of the partners (ULL, see *Table 1*), which includes both the legacy and the new data. Although it was an attractive possibility, technical difficulties precluded implementing this option before the deadline. Besides, most partners favoured using some of the free publicly available tools for this purpose.

3.- Implement a common ODR in an accessible platform

We contacted platforms like Open Aire (https://www.openaire.eu/) and Zenodo[1] (https://zenodo.org/). In both cases, we knew that different EU alliances had used them for similar purposes (for instance, in the case of OpenAIRE, CONEXUS, EUT+, Aurora, or Eutopia are using OpenAIRE CONNECT as a single-entry point to the alliance's research production).

As the team in Open AIRE informed us if all the partners have repositories already registered and indexed in OpenAIRE, setting up the first version would be easy and fast from a technical point of view. As this is not the case, the solution was not so simple.

Therefore, after a final discussion that took place during the annual meeting in UFC (Besanson, 21st March 2024), we decided to use the Communities facility (https://zenodo.org/communities), which is also the choice for the Open Repository for EU-funded research (Pilot), among other initiatives (https://zenodo.org/communities/eu/), for our common ODR, as described in the next section.

3. THE STARS EU COMMON ODR

The STARS EU common ODR was created in April 2024 in Zenodo as a community that will host the open data from all partners for the alliance's projects and gradually incorporate the legacy data from other projects of interest from each partner already publicly available in different repositories.

The URL for the ODR is https://zenodo.org/communities/stars-eu/. A screenshot can be seen in *Figure 1.*

The factors that finally weighed most in our decision to go for Zenodo were manifold. Here, we list some of the main ones:

- It is compliant with the FAIR principles (for instance, uploads are assigned a Digital Open Identifier to make them easily citable and trackable)

- It is free and completely open (its code itself is open source)

- All metadata data is openly available under a CCO license, and the content is accessible through open APIs.

---

[1] The name Zenodo derives from *Zenodotus*, the first librarian of the ancient library of Alexandria and father of the first recorded use of metadata, a landmark in library history (https://www.openaire.eu/zenodo-guide)

- It accepts any file format and both positive and negative results. Moreover, it does not impose format, size, access restrictions, or license requirements.

- Registering, either as an administrator to curate and manage data in a community or as a researcher to upload your data, is very easy and can be done using GitHub, ORCID, or OpenAIRE credentials.

- Plenty of information and tutorials are available on how to upload/curate the data (e.g., https://youtu.be/BPVSErzNtME). We think this feature will be especially useful for partners who currently have no experience with repositories.

- As part of OpenAIRE, it will be easier to integrate the community in OpenAIRE CONNECT to upgrade the common ODR and make it part of a comprehensive site that includes all alliance research results, not only the data.

- Finally, many researchers from the alliance's different universities already have their data in Zenodo, facilitating their progressive inclusion in the community[2].
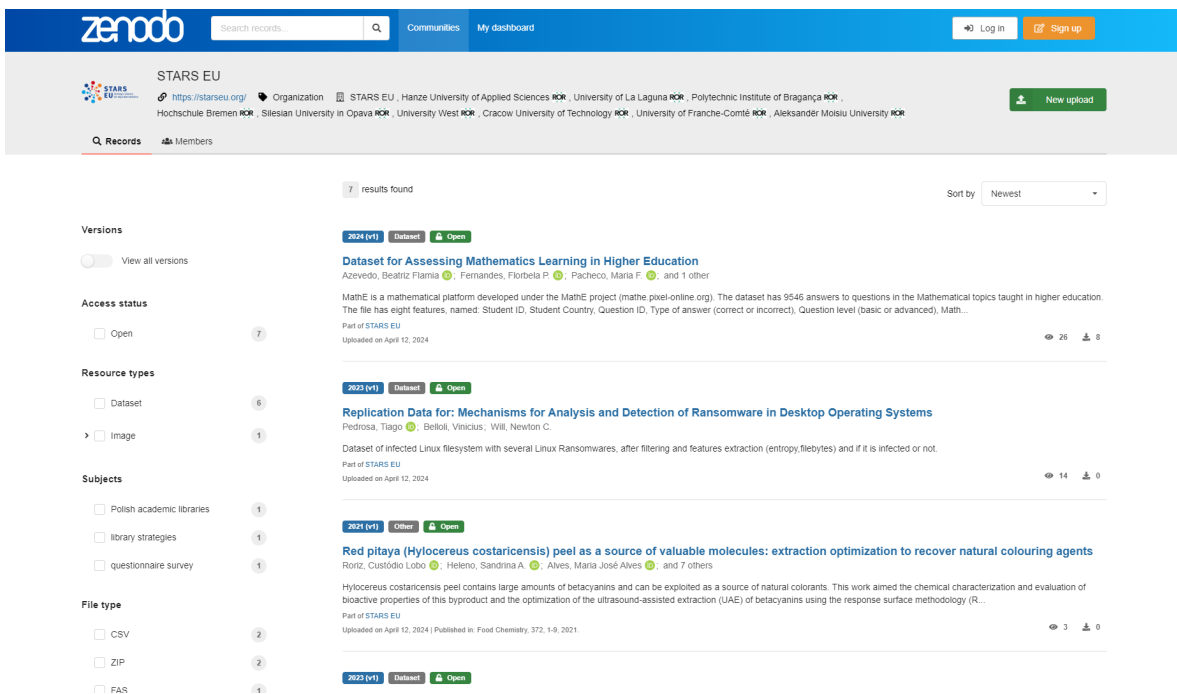

Figure 1: Screenshot of STARS EU common ODR as of April 17th 2024 in Zenodo communities

---

[2] See, e.g., https://zenodo.org/search?q=%22University%20of%20Franche-Comt%C3%A9%22&f=resource_type%3Adataset&l=list&p=1&s=10&sort=bestmatch for UFC,
Or
https://zenodo.org/search?q=%22University%20of%20La%20Laguna%22&f=resource_type%3Adataset&l=list&p=1&s=10&sort=bestmatch for ULL.

4. CONCLUSIONS AND FUTURE WORK

Diversity and complementarity are some of the hallmarks of the universities that conform to the STARS EU Alliance (https://starseu.org/), and this feature was apparent in the different situations of the open data repositories of each partner. We could leverage the experience of those universities with one institutional ODR at work to find the best solution that allowed us to comply with the deliverable before the deadline without sacrificing its quality. Moreover, we did it by keeping a unified brand and acting on the principle that, as an alliance, our ODR should be common to all partners while respecting the already existing individual repositories and should be kept as simple as possible to allow for those partners with less experience on the field to adopt this solution for their researchers / administrative staff efficiently.

In this vein, and as part of task 8.3 in WP8.3, a series of training sessions has been set for the partners. In these sessions, renowned experts in the field of Open Science and Open Data review the current situation, the different aspects, and the practicalities of open data sharing and open access in the international research community. This series is helpful for researchers and administrative staff from libraries alike.

As part of the future work, at present, we continue to work with both OpenAIRE and Elsevier to find the best way to upgrade the ODR to include the legacy data already present in each of the individual repositories and make it part of a comprehensive portal[3] including eventually all the scientific and data production, past, present, and future, of STARS EU alliance.

---

[3] Such as the one from EU-CONEXUS: https://eu-conexus.openaire.eu/

5. REFERENCES

(1) Open Science - European Commission - Research and Innovation. https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en.
(2) Open Science and Intellectual Property Rights - Research and innovation. https://research-and-innovation.ec.europa.eu/system/files/2022-04/ec_rtd_open-science-and-ip-report.pdf.