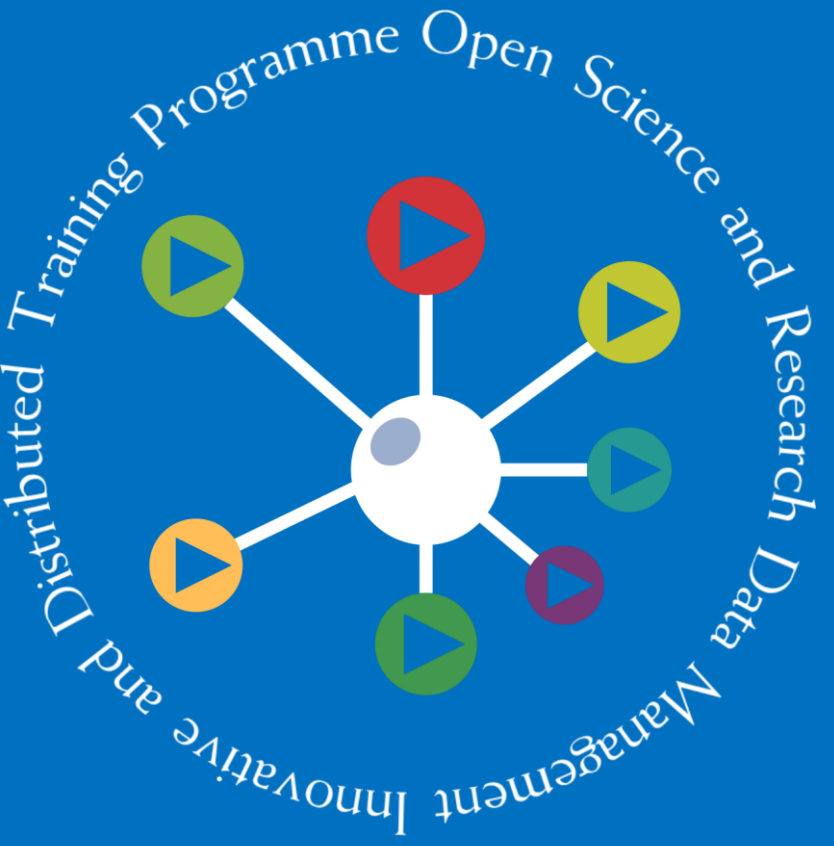
IO3: Distributed eLearning Training Platform for Internet of Things and Data Engineering



University POLITEHNICA of Bucharest
National Institute of Reserch and Development in Informatics (ICI Bucharest)

Scope

- Training Platform to share online digital training materials with Students and Early Stage Researchers
  - Host online Training Sessions and be able to collect feedback.
  - Work as a Living Lab for fostering hands-on activities on OS domains, like working with open and accessible data.

Activities

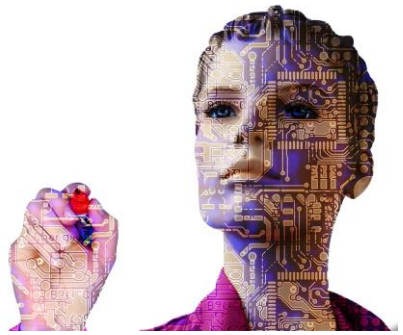
- O3/A1.Technical and Functional Specifications and architectural design (M1-M2)
- O3/A2. Implementation and first prototype evaluation (M3-M14)
- O3/A3. Update and product testing (M15-M30)

O3/A1.Technical and Functional Specifications and architectural design

- Analysis of the requirements for the distributed training platform
  - Investigate the possibility to host at each institution an instance of the platform
  - Interconnect the authentication credential technologies into one
  - Or, go towards a more centralised architecture for such a platform
- We need to analyse the requirements in terms of hosting training platform
  - Assess the possibility of developing add-ons on-top of already existing open source software (preference) like Moodle.
- UPB to conduct surveys and focus groups to find out what is really needed (e.g., professors, policy makers, research stakeholders) who would be able to use the online platform.

O3/A2. Implementation and first prototype evaluation

- UPB, together with partners, develops a first prototype of the online training platform.
  - This will be put to test in the first set of training activities in the project
  - ... from where it will collect feedback from both Trainees and Trainers.



O3/A3. Update and product testing

- The platform will be continuously tested ...
  - Up-to-the-phase where we take the prototype
- The prototype will be able to host one or two training sessions in parallel with a limited number of online Trainees
  - Scalability up to 50 training sessions in parallel and hundreds of Trainees accessing in the same time the platform.
- The backbone runs in the Cloud (within the UPB's and NCI's Data Centers)
  - The possibility of synchronisation of curricula and activities between hosts.

Step 1... Let’s look at others

- We tried understanding how Open Science should be teached...
  - Disclaimer: couldn’t find similar Academia Programme
- The aim should be to
  - (a) teach students to conduct the entire research process for conducting a replication according to open science standards, and
  - (b) contribute to cumulative science by increasing the number of direct replications
- ... choosing suitable replication studies to guiding students through the process of conducting a replication, and processing results in a meta-analysis.

*JEKEL, Marc, et al. How to teach Open Science principles in the undergraduate curriculum—The Hagen cumulative science project. *Psychology Learning & Teaching*, 2020, 19.1: 91-106.

Barriers in learning?

- Within the time frame of a PhD, students who work empirically are typically asked to
  - (a) begin with a literature review in order to develop their research question,
  - (b) develop a research design,
  - (c) prepare the necessary materials for the empirical study,
  - (d) collect data from participants,
  - (e) analyze the data, and
  - (f) write up their thesis.
- The teaching goal of the thesis is to familiarize students with practical scientific work.
- Their skills for independently conducting research should be developed and tested:
  - These skills should include the ability to evaluate research questions critically, develop ideas about the type of evidence necessary to answer the research question, and conduct the respective experiment
  - In addition, recent evidence concerning the lack of transparency and openness of research call for focusing more strongly on teaching the transparent and comprehensive reporting of research.

Feeling Overwhelmed by Generating an Interesting and New Research Question

- When developing their first research question, students oftentimes find it difficult to obtain a comprehensive overview of the relevant research and methods due to the large amount of published papers in a given field.
  - This potential overload of information may often lead to negative feelings of being overwhelmed.
  - As a result, some students may develop research questions that have already been answered but that have been overlooked by the student's incomplete literature review.
- With a predetermined study design and statistical analyses of the original article and structured guidelines on how to proceed, conducting replications contributes to a balanced process between working independently and receiving orientation to successfully conduct a meaningful empirical study.

Focusing Only on (Overly) Simple Methods and Statistics

- Using replications motivates students to not think about research designs in the framework of the statistical procedure with which they feel most comfortable, but rather the appropriate framework for the research question at hand.
  - By re-analyzing the original data under supervision, students are introduced to statistical tests that could go beyond the methods they learned in class.
- Additionally, reproducing original analyses as well as replicating the experiment feels like “detective work”, which can be exciting for students.
  - The re-analysis also builds up the necessary confidence and expertise for the analysis of the replication data.

Lack of Orientation and Structure

- Reading scientific articles from the perspective of a scientist planning on replicating the presented work instead of a remote reader and consumer of the research helps PhD students to gain insights into the decision processes of the original authors
  - Furthermore, it allows them to reflect on their own methodological choices more systematically.
- Experiencing the scientific procedure as an observer and evaluator of the original work as well as an active researcher who is in the process of running the replication teaches the value of reproducible routines as well as the importance of open and transparent documentation
  - These insights help students develop a professional routine of making their work accessible to readers with little prior knowledge
- The replication process itself allows for a structured schedule with consecutive steps.
  - A replication project thereby lends more orientation and structure, which enables students to manage the process of writing a thesis more effectively.

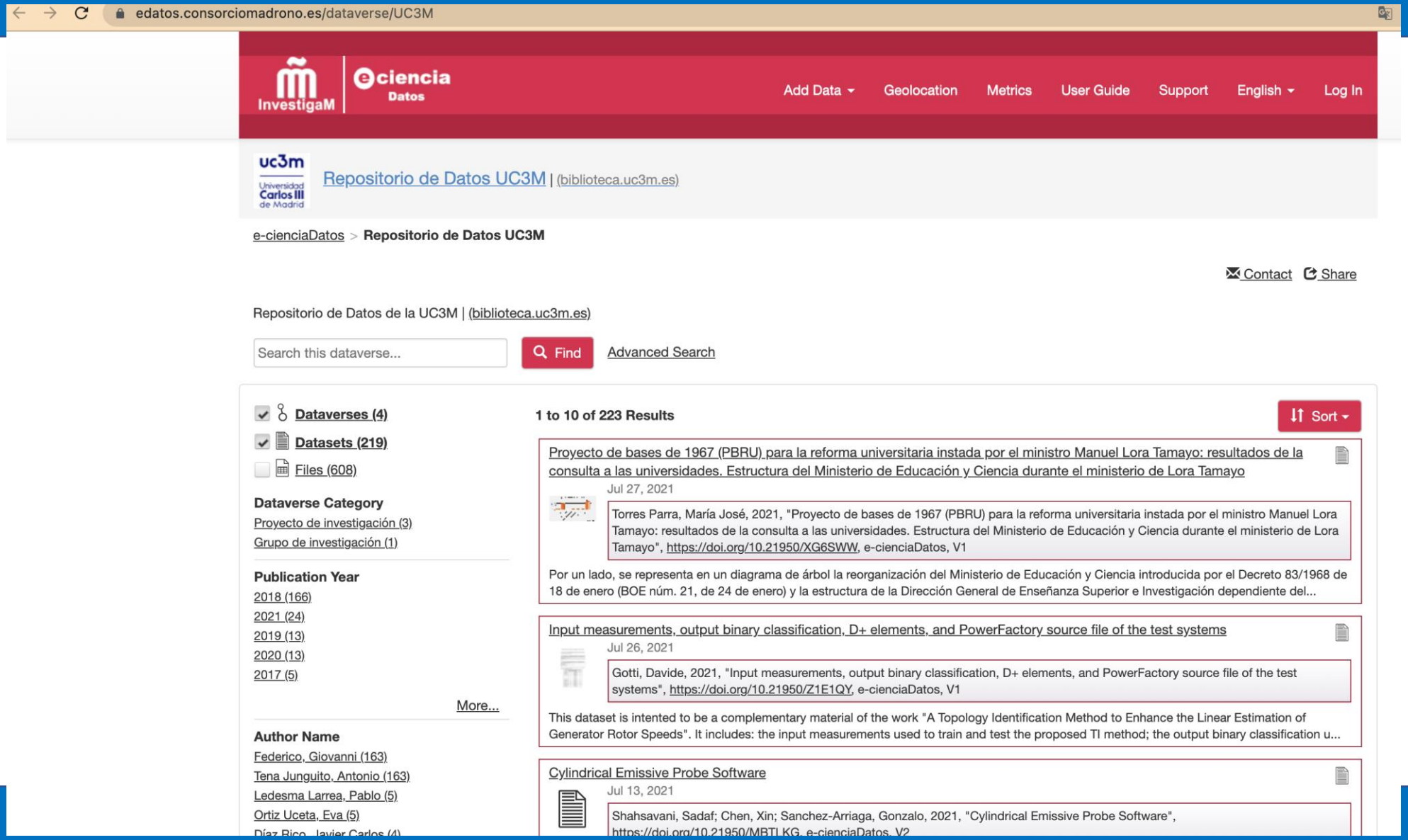
Low Sense of Purpose and Responsibility: Limited Scientific Thinking

- Approximately 90% of student projects typically do not satisfy the high quality standards in the research field.
  - This is problematic because students might anticipate from the start that their work will never be read, discussed, or built upon.
- Such conditions exclude students from feeling responsible and being part of the collaborative effort of conducting scientific work.
  - Thus, it may also reduce the tendency for deeper, responsible, and more (self-)critical scientific thinking.
- Replication focuses on teaching students to appreciate research and the scientific method, develop their methodological and experimental skills, and motivate critical thinking by doing meaningful work.
  - If the replication attempt adheres to scientific standards, students' theses might even potentially result in their first scientific contribution in the form of a publication.

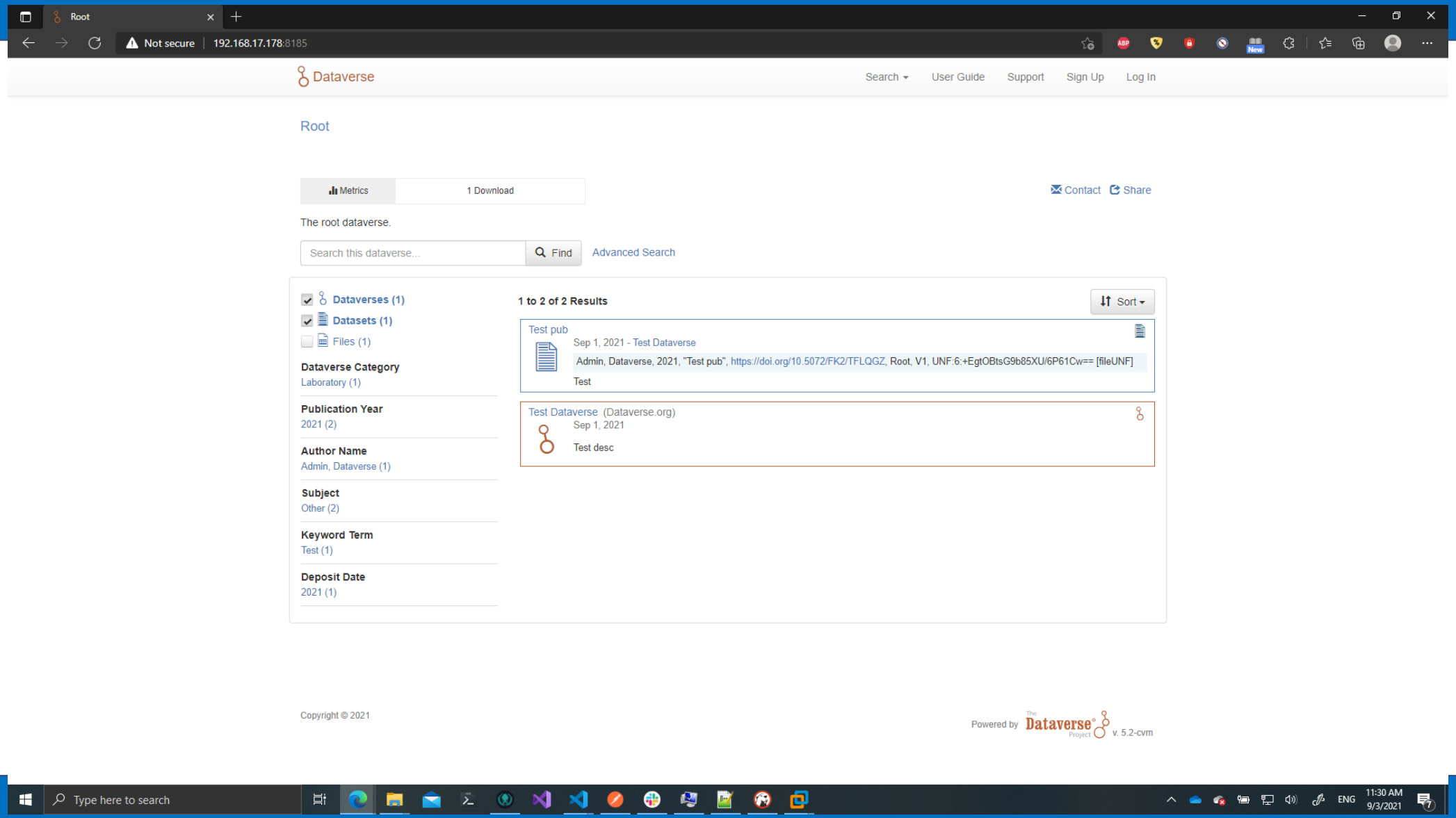
Procedural steps in Data Replication and Critical Thinking

1. Identify feasible replication studies
   - Instructors identify studies for students. Feasible studies are empirical, include a statistical significant effect, use statistical methods, are not context or sample dependent, and do not require equipment that is not ready available
2. Match students with replication studies
   - Match students according to their preferences with available replication studies
3. Select central finding
   - Students identify the central staistically significant finding of the study as the replication target
4. Reanalyse original data
   - A reanalysis of the original data allows students to obtain a better understanding of the study
5. Do a power analysis
   - Students estimate planned sample size a priori based on the effect size reported in the article or derived from the reanalysis of the original data
6. Implement the study
   - Students implement the study in an experimental software
7. Analyse data
   - Students analyse the data to test whether the original effect can be replicated
8. Document results
   - Students publish their Open Data

So we start with data...



We installed and played with dataverse:



Test Dataverse (Dataverse.org)

publications_202109011044.tab

This file is part of "Test pub".

Version 1.0

File Citation

Admin, Dataverse, 2021, "publications_202109011044.tab", *Test pub*, Root, V1, UNF:6:+EgtOBtsG9b85XU/6P61Cw== [fileUNF

Cite Data File | Learn about Data Citation Standards.

Access File

Download Options
Original File Format (Comma Separated Values)
Tab-Delimited
RData Format
Variable Metadata
Data File Citation

Dataset Citation

Admin, Dataverse, 2021, "Test pub", https://doi.org/10.5072/FK2/TFLQGZ, Root, V1, UNF:6:+EgtOBtsG9b85XU/6P61Cw== [file

Cite Dataset | Learn about Data Citation Standards.

Metadata | Versions

Export Metadata

File Metadata

Preview

Download URL

Use the Download URL in a Wget command or a download manager to avoid interrupted downloads, time outs or other failures.

ID Type Select...

ID Number

URL Enter full URL, starting with http://

Notes

Depositor Admin, Dataverse

Deposit Date 2021-09-03

Files

For more information about supported file formats, please refer to the User Guide.

Upload with HTTP via your browser

Select files or drag and drop into the upload widget.

+ Select Files to Add

Drag and drop files here.

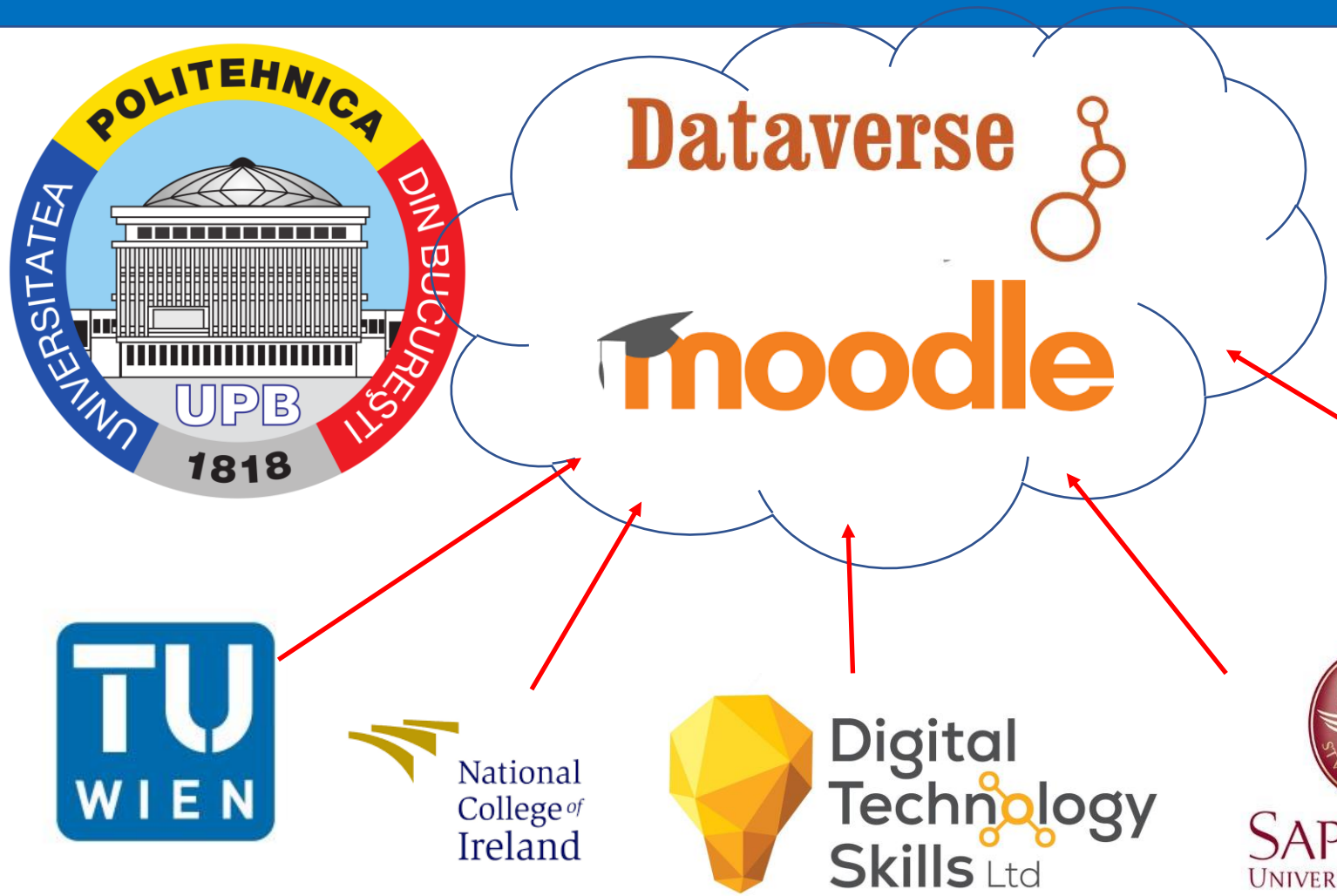
Metadata Tip: After adding the dataset, click the Edit Dataset button to add more metadata.

Save Dataset Cancel

Moodle...

- Moodle is an OS learning platform designed to provide educators, administrators and learners with a single robust, secure and integrated system to create personalised learning environments. You can download the software onto your own web server or ask one of our knowledgeable Moodle Partners to assist you.

Possibilities

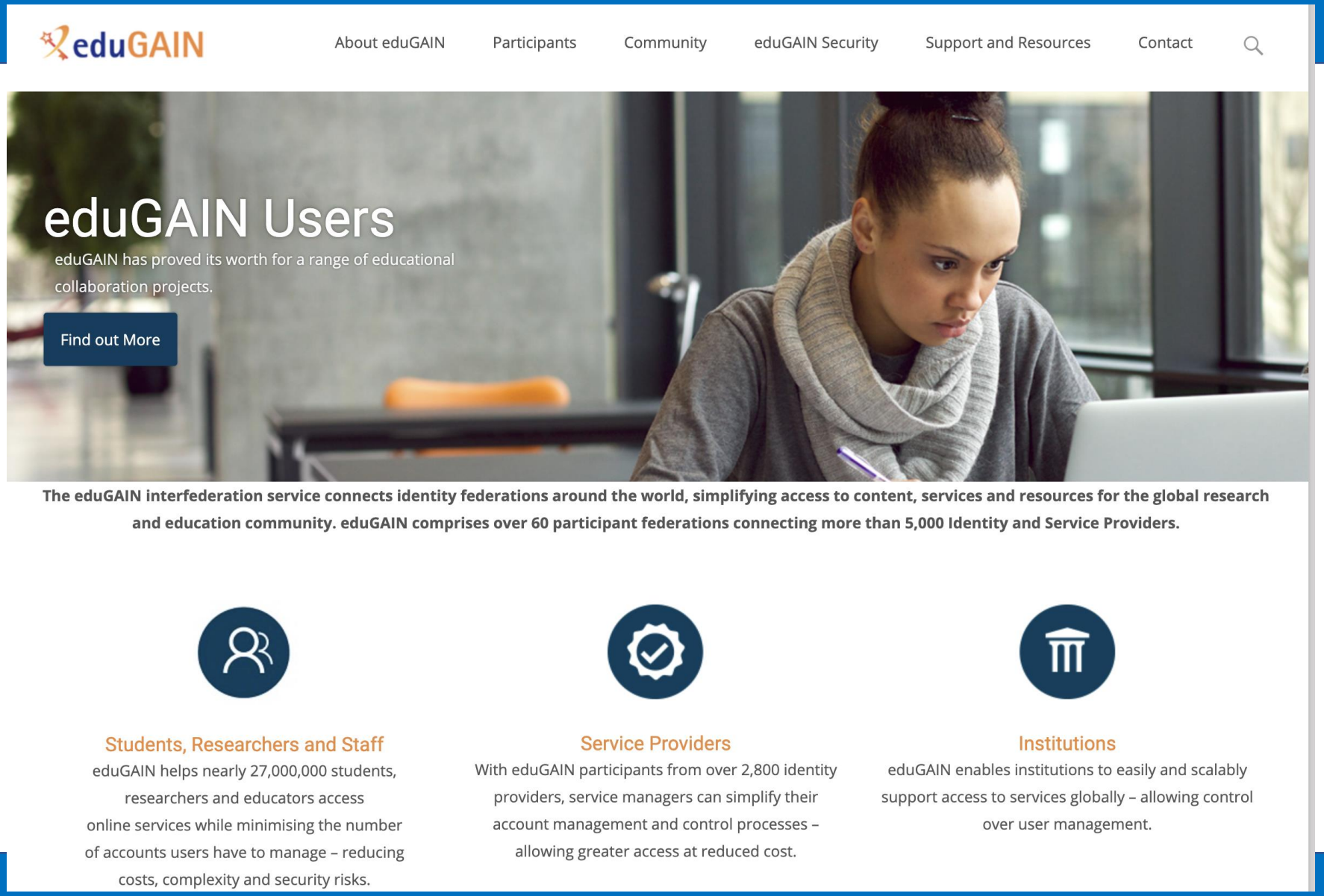


Easy to have

Can work for a comon curricula like the one for Reproductibility Science

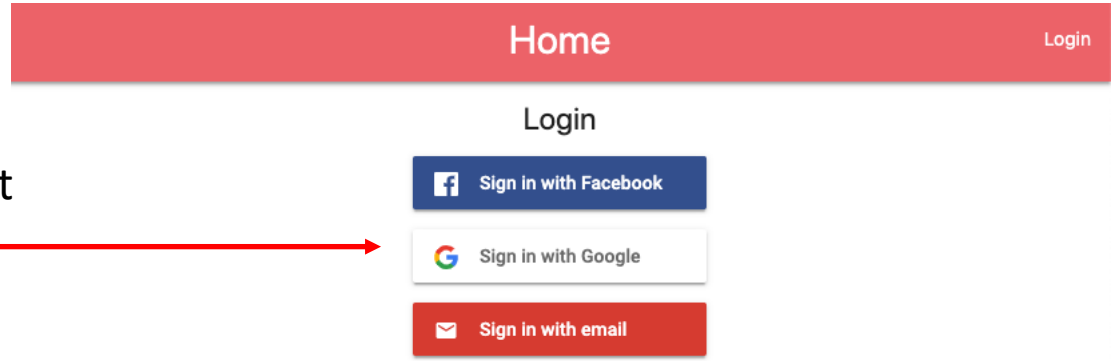
Not working when already existing materials need to be moved somewhere else (IPR / GDPR?)

Or... In terms of access?



Have you seen this?

This is based on OpenID Connect



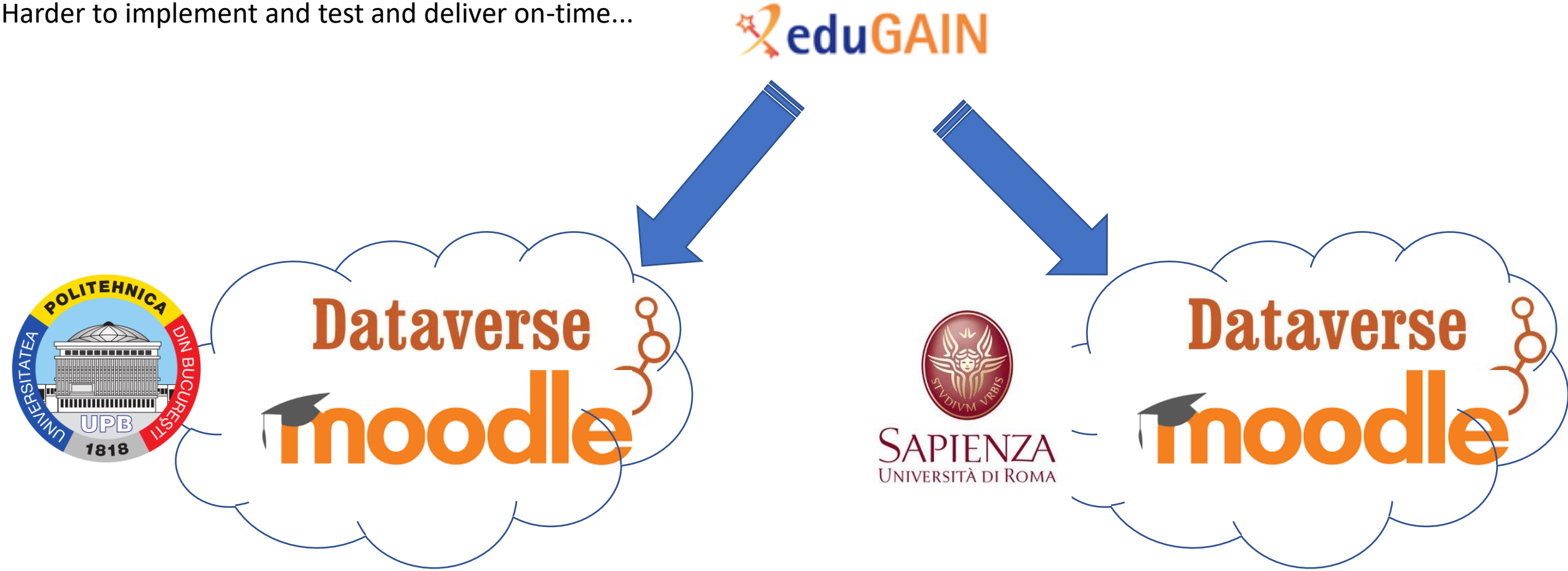
...eduGAIN adds another layer above OpenID Connect to allow for educational federations of ID providers

eduGAIN

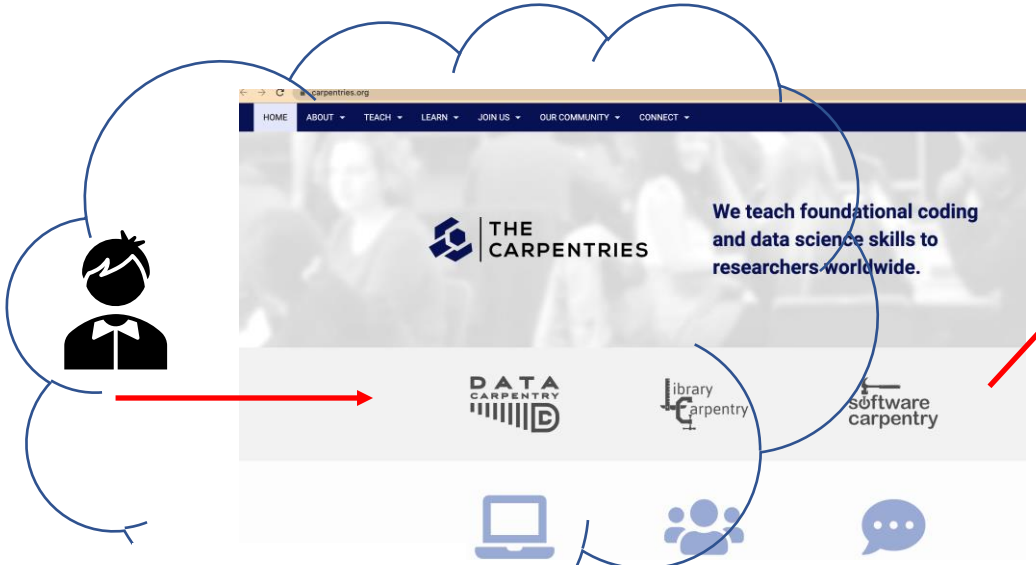
- The eduGAIN interfederation service connects identity federations around the world, simplifying access to content, services and resources for the global research and education community.
  - eduGAIN comprises over 60 participant federations connecting more than 5,000 Identity and Service Providers.
- It works with OpenID Connect and SAML 2.0
- For example, UPB is part of RoEduNet / GEANT and of eduGAIN
  - When a student from UPB tries to log into TU Wien, the education platform at TU Wien asks the eduGAIN Service Providers, who identifies the student as partaining to UPB and access the SAML 2.0 auth service at UPB for clearance / authentication

Possibilities

It solves IPR and security aspects....
Harder to implement and test and deliver on-time...



And through a carpentier (catalogue of classes / modules?)



SAPIENZA
UNIVERSITÀ DI ROMA

So what's next?

- We will develop the *centralised* Moodle platform at UPB with the first set of classes
  - To be ready for first training (common) modules → please select classes that do not suffer from IPR / authorisation problems
  - If classes are *mandatory* tought at the home institution, be prepared to manually enrol outside PhD students into the local authorisation system
- We will disseminate and collect data for a questionairre to understand the technical local training conditions for partners
  - Didn't do it earlier to not overlap with the effort in IO.1
- Develop the eduGAIN connectors with all partners supporting this
  - Already developing connectors at UPB for the integration between dataverse and SAML 2.0 (local auth system)

THANK YOU!

Follow us