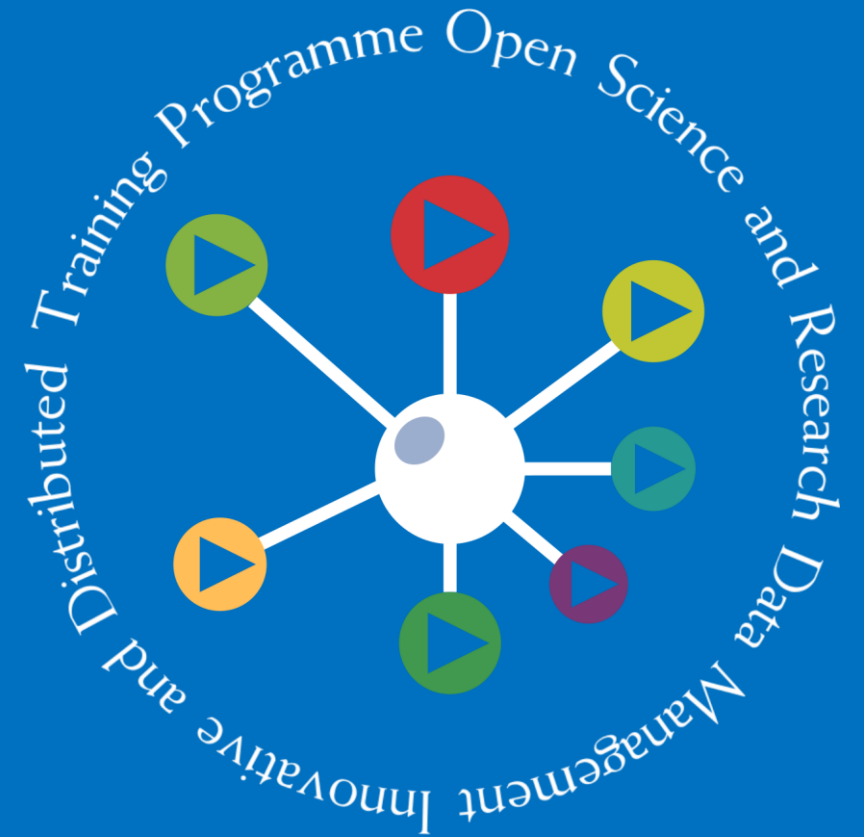


IO3: Distributed eLearning Training Platform for OS and RDM

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



Overview of IO3



🌐 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.

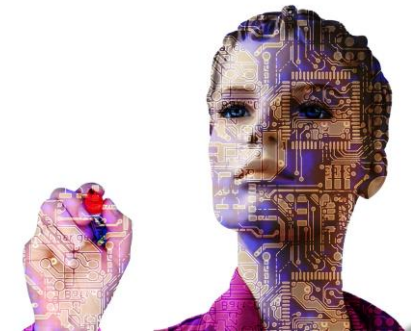
O3/A2. Implementation and first prototype evaluation



🌐 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.

We conducted a survey between partner to find out technical capabilities



- ✿ 1. Are you using an online learning platform or course management system (CMS) for teaching in your institution? I assume yes, but please let us know.
 - ✿ 1.a. If you answered Yes to Question 1 above, please tell us what platform are you using? Moodle seems to be the most frequent one.
- ✿ 2. (Only if you answered Yes in Question 1) Is your institution providing learning courses relevant to TrainRDM online through the platform in Question 1.a? Can you give us three examples of such online classes?
 - ✿ 2.a. Is there a specific IPR policy that forbids teachers using the online platform to upload their materials to an external CMS platform (Moodle being the preferred choice)? Please give a short comment on the kind of constraints in place, if any.
- ✿ 3. What kind of authorisation and authentication system is your institution using for user management?
- ✿ 4. Is there an official affiliation of your institution to EduGAIN? Please let us know the details, for example the name of the provider through which the access is provided.
- ✿ 5. Is there a current research information system (CRIS) - a database or other information system to store, manage and exchange contextual metadata for research - in use with your institution? Which one (and a link would be very appreciated if you know it)
- ✿ 6. Is there a data repository (data library or data archive like CKAN or dataverse) in use with your institution? Which one (a link would be very appreciated)
 - ✿ 6.a. If you answered yes to Question 6 above, please let us know what Data Management Policy or Open Data policy your institution uses.
- ✿ 7. For the examples of classes in Question 2, can you please describe or give example of tools and learning means teachers use for hands-on activities? This could include the use of simulators (what kind, how are they accessible - locally or remote), Matlab instances or Jupyter Notebooks / Lab containers, docker instances running specific code, etc.

Question 1



1. Are you using an online learning platform or course management system (CMS) for teaching in your institution? I assume yes, but please let us know.
- UPB RO: yes, we use Moodle
 - ICI RO: yes, we use two platforms.
 - TUW AT: Moodle system (at TU Wien called TUWEL)
 - DTSL IE: -
 - UNIROMA IT: yes, we use MOODLE only for students.
 - NCI IE: a simple Google search revealed that Moodle is being used
- 1.a. If you answered Yes to Question 1 above, please tell us what platform are you using? Moodle seems to be the most frequent one.
- UPB RO: Moodle
 - ICI RO: We use Moodle and also other platform developed by us.
 - TUW AT: Moodle
 - DTSL IE: -
 - UNIROMA IT: Moodle
 - NCI IE: Moodle



Moodle is the winner!

Questions 2 and 7



- ✿ 2. (Only if you answered Yes in Question 1) Is your institution providing learning courses relevant to TrainRDM online through the platform in Question 1.a? Can you give us three examples of such online classes?
 - ✿ UPB RO: See below
 - ✿ ICI RO: No.
 - ✿ TUW AT: Yes, but we cannot provide examples because it is an internal platform.
 - ✿ DTSL IE: -
 - ✿ UNIROMA IT: No. However, a short Google search revealed several candidates (next slide)
 - ✿ NCI IE: The courses on the next slide were obtained after a short Google search +
 - ✿ Data Intensive Architectures; Scalable Systems Programming; Data Governance, Compliance and Ethics
- ✿ At UPB we teach already on Moodle classes such as:
 - ✿ Introduction to Big Data: it includes aspects to deal with open data (and students learn to use proper processing of open data)
 - ✿ Dependable Systems: it includes aspects related to making decisions starting from an exploratory data analysis; it includes aspects for working with Personal and Sensitive Data
 - ✿ Data and system security: it includes aspects related to data protection

Question 2 at UNIROMA



🌐 Data Management for Data Science (<http://datascience.i3s.uniroma1.it/it/node/5619>)

- 🌐 Large-scale data management: Distributed query evaluation, NoSQL databases, graph databases.
- 🌐 Open data management: open data, linked open data, RDF databases

🌐 Intellectual Property Competition and Data Protection Law (<http://datascience.i3s.uniroma1.it/it/node/5806>)

- 🌐 Data, contracts and rights. General principles
- 🌐 IP law: Patents and utility models, Trademarks, Copyright, IP rights in software, IP rights in databases, Trade secrets
- 🌐 Competition law: Anticompetitive agreements. General rules and specific issues (e.g. R&D agreements, licensing agreements, standard setting organizations), Sharing of commercially sensitive information, Big data and competition
- 🌐 Data protection: Data and personal data, Collection, circulation and use of personal data, Specific issues: consent, information, profiling, data transfer, binding corporate rules
- 🌐 The new EU Regulation, Cybersecurity and data protection

🌐 Statistical learning (<http://datascience.i3s.uniroma1.it/it/node/5602>) and Statistical Methods in Data Science and Laboratory (<http://datascience.i3s.uniroma1.it/it/node/5802>)

- 🌐 Statistical Methods and Applications
- 🌐 Management Engineering
- 🌐 Statistical Methods and Applications

Question 2 at NCI



- * Data Governance and Cybersecurity: it provides students with the knowledge, skills, and competence around security and data governance, data management, data classification, data quality, and data risk management. (<https://www.ncirl.ie/Courses/NCI-Course-Details/course/DGC>)
 - * Cybersecurity Concepts, Threats to Information & Data Processing Services
 - * Security Governance, Data Governance, Data Governance Dimensions
 - * Risk Assessment & Risk Management
 - * Legal Aspects
 - * Ethical Issues Pertaining to Data
- * Data Governance and Ethics (<http://courses.ncirl.ie/index.cfm/page/module/moduleId/67150>)
 - * Critically interpret the governance and regulatory frameworks associated with the capture, processing, and stewardship of data.
 - * Critically interpret the roles and responsibilities pertaining to data security, privacy, and data protection.
 - * Analyse and evaluate the intersection of data and ethics in socio-technical environments.
 - * Investigate and appraise the interplay of fairness, accountability, and transparency in algorithmic decision making systems.
- * Data Visualisation (<http://courses.ncirl.ie/index.cfm/page/module/moduleId/67179> hello Adriana)
 - * Analyse the theory and concepts relating to visualisation and data representation
 - * Evaluate and distinguish between visualisation techniques for specific problems to enable effective communication of data analysis
 - * Design, develop, and implement processes for data visualisation
 - * Propose a suitable visualisation design for a particular combination of data characteristics and application
- * Data Architecture (<http://courses.ncirl.ie/index.cfm/page/module/moduleId/67279>)
 - * Compare different data stores, data models, query languages, data encoding techniques and knowledge discovery techniques
 - * Summarise the constraints and trade-offs of a distributed shared-nothing architecture that is involved in the storage and retrieval of data
 - * Analyse and design a data application architecture that integrates multiple disparate data systems that are optimised for different access patterns
 - * Collaboratively implement an application data architecture
- * Scalable Data Analytics (<http://courses.ncirl.ie/index.cfm/page/module/moduleId/67297> hello Horacio)
 - * Describe and apply MapReduce and extensions for creating parallel applications on large amounts of data
 - * Describe and summarise search techniques including similarity search and search engine technologies.
 - * Distinguish between data-stream processing and specialised algorithms
 - * Develop analytical and ethical skills to employ mining and clustering algorithms on large multi-dimensional datasets

Question 2.a



- ✿ 2.a. Is there a specific IPR policy that forbids teachers using the online platform to upload their materials to an external CMS platform (Moodle being the preferred choice)? Please give a short comment on the kind of constraints in place, if any.
 - ✿ UPB RO: No IPR official policy.
 - ✿ ICI RO: No, it doesn't exist. All materials can be uploaded on the platform because the content does not contain sensitive data. For external institutions, uploading materials needs the approval of the Scientific Committee.
 - ✿ TUW AT: Upload on TUWEL is common practice. Teachers would probably not upload their materials on external platforms. Apart from this they usually do not assign open licenses to their teaching materials.
 - ✿ DTSL IE: -
 - ✿ UNIROMA IT: see below
 - ✿ NCI IE: At NCI, material owners (lecturers) typically retain ownership of materials in Moodle.
- ✿ In **Sapienza** there is a specific regulation for the use of the Moodle platform. It seems that there is not a specific IPR policy that forbids teachers using the online platform.
- ✿ Internal and external teachers can use it for 4 types of courses:
 - ✿ Frontal teaching offered by Sapienza: study courses, single courses and didactic paths provided by Sapienza internal or external teaching staff . It is for Sapienza students.
 - ✿ Training by and for internal staff: training courses by and for Sapienza staff. Use is limited to Sapienza staff.
 - ✿ Services offered for the Sapienza community: courses offered by Sapienza structures or for specific areas (libraries, laboratories, services, orientation, placement), carried out by internal staff. To Sapienza staff; to Sapienza students; to external students (e.g. high school students).
 - ✿ External services offered by Sapienza: courses provided to external organizations / entities by Sapienza staff or by external teachers.
- ✿ A course, at the author's discretion, can be:
 - ✿ 1. with free access, therefore available to anyone;
 - ✿ 2. reserved for Sapienza users only;
 - ✿ 3. reserved for a specific group of users; in this case the author of the course is responsible for the registration procedures of the users admitted to the course, according to the specific guidelines.

Questions 3 and 4



- ❁ 3. What kind of authorisation and authentication system is your institution using for user management?
- ❁ UPB RO: We use both SAML 2.0 and OpenID. A common Keycloak solution (login.upb.ro) is currently being deployed.
 - ❁ ICI RO: We use LDAP.
 - ❁ TUW AT: TU Wien authentication System (SSO)
 - ❁ DTSL IE: -
 - ❁ UNIROMA IT: Two ways for user registration on the Moodle platform:
 - ❁ Automatic for Sapienza internal users in possession of credentials (email);
 - ❁ Manual in the case of users external to Sapienza (contract teacher, tutor or student external to Sapienza) at the request of the manager who will indicate the name, surname, tax code, start and end date of the account.
 - ❁ NCI IE: Microsoft Active Directory (internal), Shibboleth for subscriptions and other services
- ❁ 4. Is there an official affiliation of your institution to EduGAIN? Please let us know the details, for example the name of the provider through which the access is provided.
- ❁ UPB RO: Yes, through RoEduNet and GEANT.
 - ❁ ICI RO: to be verified
 - ❁ TUW AT: Yes, the Austrian branch is called ACOnet
 - ❁ DTSL IE: -
 - ❁ UNIROMA IT: Sapienza is part of EduGAIN but I do not know the details.
 - ❁ NCI IE: a simple search (<https://edugate.heanet.ie/rr3/metadatalocations>) showed to be part of eduGATE.

Keycloak is an open source software product to allow single sign-on with Identity and Access Management aimed at modern applications and services.

🌐 Conclusion: eduGAIN is already in use by a lot of participants via national federations

🌐 <https://technical.edugain.org/status> (you can filter the list by country).

Quick filter

Participants (73) Voting-only members (1) Candidates (7)

Participants

- AAF(Australia)
- AAI@EduHr(Croatia)
- AAI@EduMk(North Macedonia)
- ACONet Identity Federation(Austria)
- AFIRE(Armenia)
- ARNaai(Algeria)
- ArnesAAI Slovenska izobraževalno raziskovalna federacija(Slovenia)
- Belnet Federation(Belgium)
- BIF(Bulgaria)
- Canadian Access Federation(Canada)
- CAFe(Brazil)
- CAFMoz(Mozambique) ▲
- CARS(China)
- COFRE(Chile)
- COLFIRE(Colombia)
- CSTCloudFederation(China)
- CyNet Identity Federation(Cyprus)
- DFN-AAI(Germany)
- Edugate(Ireland)
- eduID.cz(Czech Republic)
- eduID.hu(Hungary)
- eduID Luxembourg(Luxembourg)
- eduIDM.ma(Morocco, Western Sahara)
- FEBAS(Belarus)
- φEDUrus(Russia)
- FEIDE(Norway)
- FFNIX(Mexico)

Europe



World



Questions 5 and 6



5. Is there a current research information system (CRIS) - a database or other information system to store, manage and exchange contextual metadata for research - in use with your institution? Which one (and a link would be very appreciated if you know it)
- UPB RO: We made our own solution (<https://crescdi.pub.ro/>). This allows us to interconnect all databases and data never leaves the premises.
 - ICI RO: No
 - TUW AT: Yes, but it is currently being revised (data will be migrated to a DSpace platform)
 - DTSL IE: -
 - UNIROMA IT: We use IRIS for publications and their metadata
 - NCI IE: No answer, but yesterday we found it's ePrints
6. Is there a data repository (data library or data archive like CKAN or dataverse) in use with your institution? Which one (a link would be very appreciated)
- UPB RO: Currently being deployed (CKAN)
 - ICI RO: No
 - TUW AT: Currently in implementation (software InvenioRDM). In a simple search we found out also <https://researchdata.tuwien.ac.at/>
 - TU Wien Research Data is an institutional repository of TU Wien to enable storing, sharing and publishing of digital objects, in particular research data. It facilitates the funders' requirements for open access to research data and the FAIR principles by making research output findable, accessible, interoperable and re-usable. This service is developed by the TU Wien Center for Research Data Management and hosted by TU.it.
 - DTSL IE: -
 - UNIROMA IT: we are adopting DATAVERSE by Harvard
 - NCI IE: No answer, but we found out NO

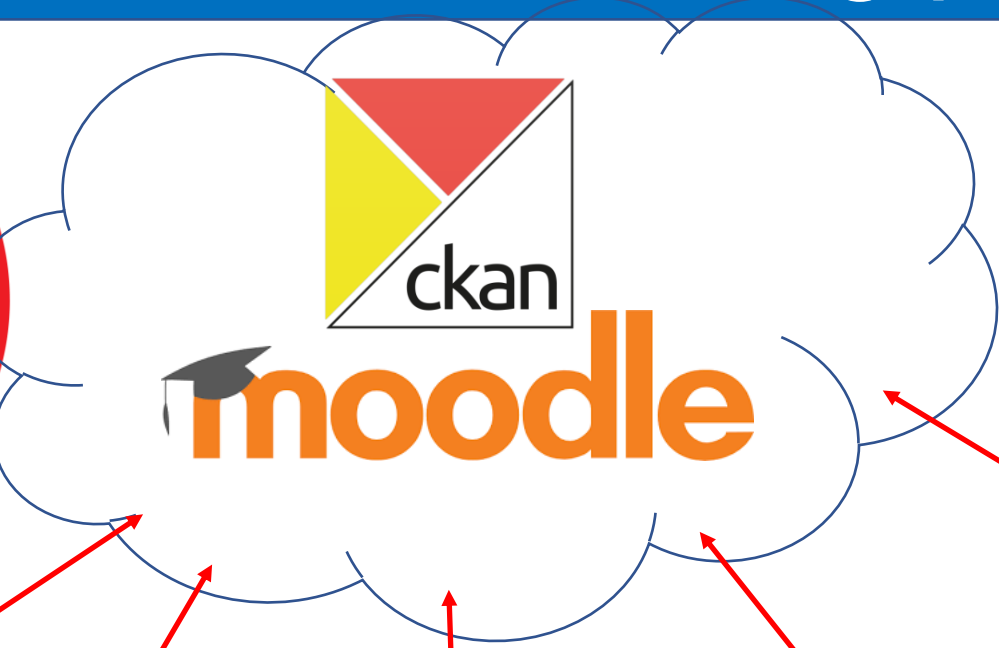
Questions 6 and 7



- 6.a. If you answered yes to Question 6 above, please let us know what Data Management Policy or Open Data policy your institution uses.
- UPB RO: No
 - ICI RO: -
 - TUW AT: RDM policy [1]
 - DTSL IE: -
 - UNIROMA IT: we are working on it
 - NCI IE: An institutional Data Management Policy will be soon released aligned with the Horizon Europe guidelines
7. For the examples of classes in Question 2, can you please describe or give example of tools and learning means teachers use for hands-on activities? This could include the use of simulators (what kind, how are they accesible - locally or remote), Matlab instances or Jupyter Notebooks / Lab containers, docker instances running specific code, etc.
- UPB RO: No official institutional guidelines for teachers is set
 - ICI RO: -
 - TUW AT: free course material can be found in the project FAIR Data Austria under <https://phaidra.univie.ac.at/o:1168881> (TU Wien is a partner there, but it is not TU Wien specific)
 - DTSL IE: -
 - UNIROMA IT: -
 - NCI IE: AWS EC2 Ubuntu instances, OpenStack Ubuntu instances, Hadoop, Spark

[1] <https://www.tuwien.at/index.php?eID=dms&s=4&path=Directives%20and%20Regulations%20of%20the%20Rectorate/Policy%20for%20Research%20Data%20Management.pdf>

So remember last time the proposal Phase 1 – centralised training platform



Easy to have – everyone is using Moodle
Can work for a common curricula like the one for Reproducibility Science
Not working when already existing materials need to be moved somewhere else (IPR / GDPR?)



Digital
Technology
Skills Ltd



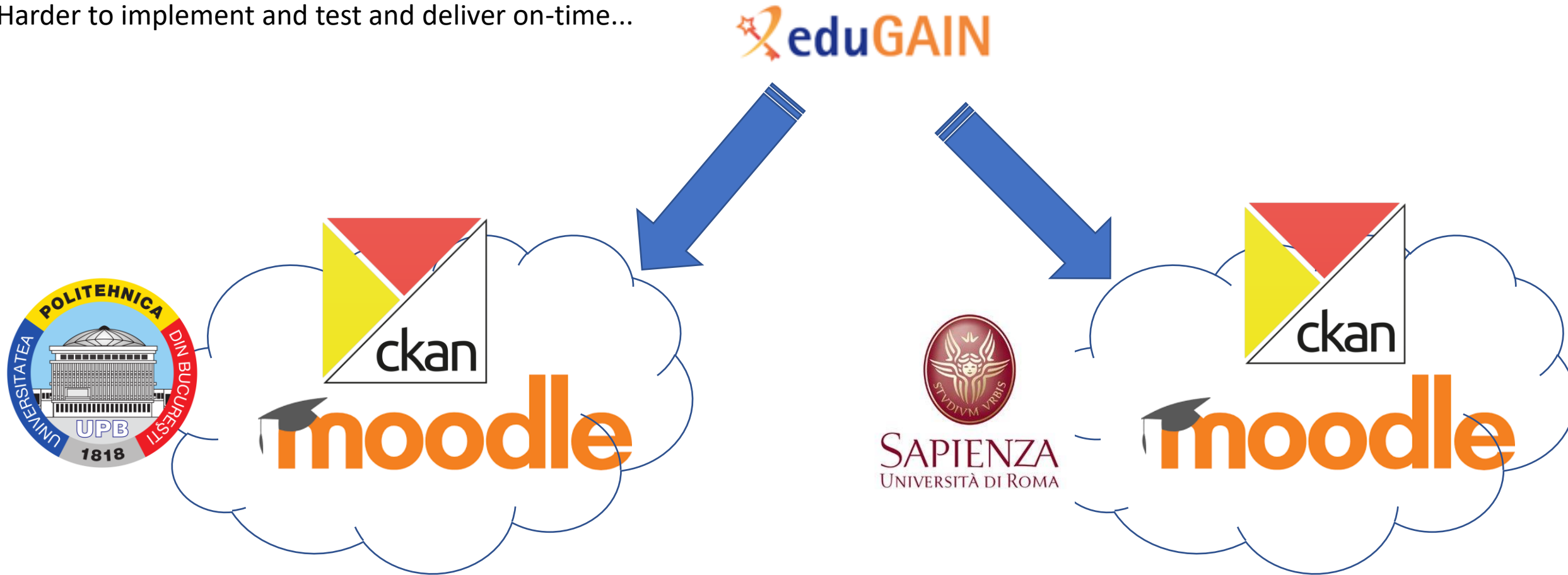
SAPIENZA
UNIVERSITÀ DI ROMA



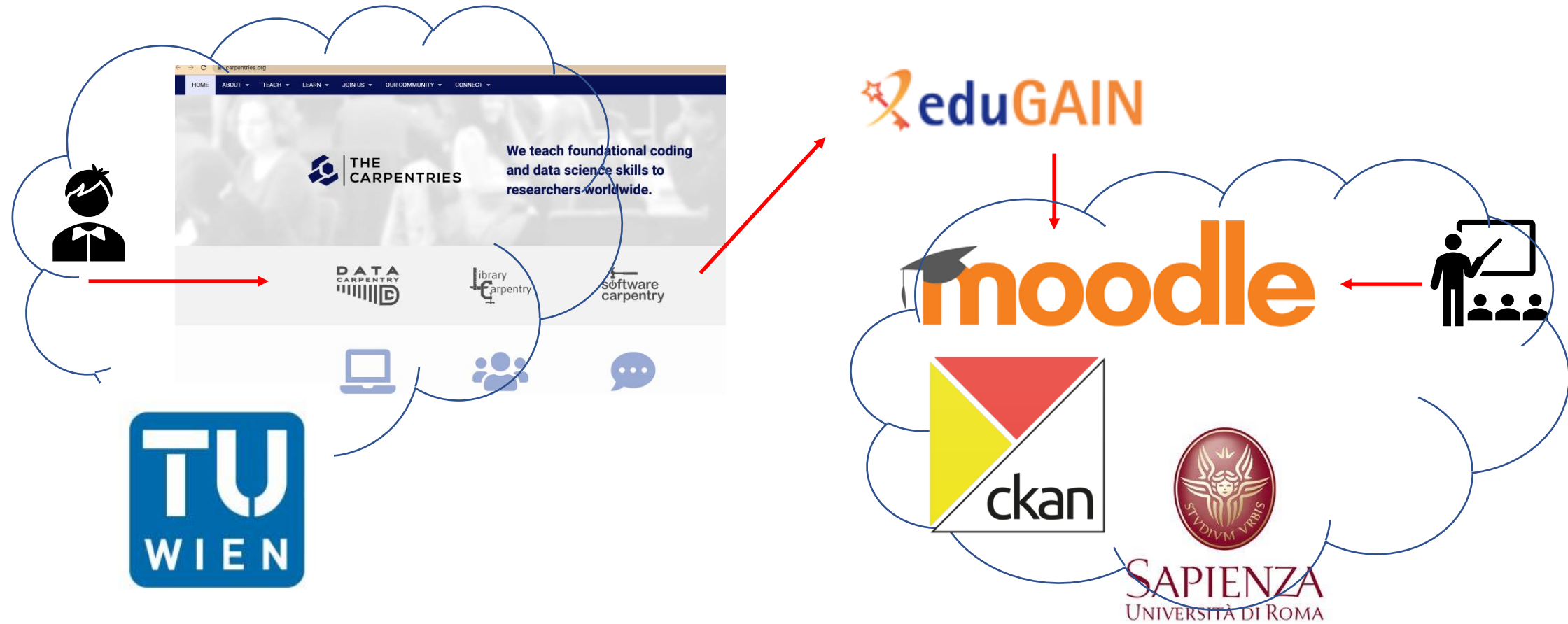
Phase 2: In terms of access, eduGAIN



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



Phase 3: carpentier (catalogue of classes / modules?)



Steps



- ❁ We started developing 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* taught 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 questionnaire 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)


Phase 1: Moodle set-up



moodle.dfilip.link

TrainRDM

You are not logged in. (Log in)



Available courses

Visualising Data

This course aims to teach the following topics:

- Introduction to data visualization
- Data visualization techniques
- Data visualization tools
- Issues in data visualization

Data Cleaning with OpenRefine

This course aims to teach the following topics:

- Use essential OpenRefine functionality
- Advanced OpenRefine functions
- Reference resources for further learning

Working with Personal and Sensitive Data

<https://moodle.dfilip.link/>

Visualising Data

Data Cleaning with OpenRefine

Working with Personal and Sensitive Data

Archiving Data

Sharing Data

Working with Data

Data Management Planning

Understanding Research Data

Inside each course has modules



Working with Personal and Sensitive Data



[Dashboard](#) / [Courses](#) / [WPSD](#)

Turn editing on

 Announcements

How to collect, share, store, and protect the sensitive data

Understand the principles of good research data management

GDPR and data protection regulations, and what these mean for research and research data

Relevant services and resources available to researchers at the university

About EduGAIN



- 🌐 We started with the eduGAIN documentation [1] [2]
- 🌐 One can add own identity provider by SAML if the elements from [2] are still specified in the metadata exposed by SAML and be signed with the specifications in that document
- 🌐 Finally, the IDP should appear at [3]
- 🌐 If we expose a service provider in eduGAIN, we have to talk to the member federation we will belong to [4] in order to be exposed to the other federations (UPB has to talk with RoEduNetID)

[1] <https://technical.edugain.org/documents>

[2] <https://technical.edugain.org/doc/eduGAIN-saml-profile.pdf>

[3] <https://mds.edugain.org/edugain-v1.xml>

[4] <https://wiki.geant.org/display/eduGAIN/How+to+Join+eduGAIN+as+Service+Provider>

And about data repository....



~~<https://dataverse.org/>~~

<https://ckan.org/>

The screenshot shows the Dataverse website interface. At the top, there is a navigation bar with 'Dataverse' logo, 'Search', 'User Guide', 'Support', 'Sign Up', and 'Log In'. Below the navigation bar, the main content area displays search results for 'Test pub'. The results include a list of datasets with columns for 'Publication Year', 'Author Name', 'Subject', 'Keyword Term', and 'Deposit Date'. A blue arrow points from this interface towards the CKAN interface on the right.

The screenshot shows the CKAN website interface. At the top, there is a navigation bar with 'ckan' logo, 'Seturi de date', 'Organizații', 'Grupuri', 'Despre', and a search bar. Below the navigation bar, the main content area displays search results for 'Test2' and 'TestDataset'. The results include a list of datasets with columns for 'Organizații', 'Grupuri', 'Cuvinte cheie', 'Formate', and 'Licențe'. A blue arrow points from the Dataverse interface on the left towards this interface.

Advantages CKAN



- ❁ Open Source software widely used for open data publications like e.g. European Data Portal, <https://www.data.org.uk/> or <https://data.gov.ro/>
- ❁ User friendly web interface for all activities associated with data publication and subscription
 - ❁ It is highly customizable in both terms of Look&Feel and functionalities.
- ❁ Capable of advanced data management
 - ❁ All datasets are organized and described with metadata, which allows it to be easily discoverable, with the use of search phrases and customizable filters (e.g.: tags, categories, data formats)
 - ❁ It is possible to publish one dataset in different data formats, not only as downloadable files but also as links to web service, web API or links to external WWW resources
- ❁ Datasets can be stored in CKAN, along with version history and dataset statistics, which allows to monitor the interest in datasets
- ❁ It provides functionalities for collaboration, community participation and providing feedback, such as comments, ratings and sharing
- ❁ It provides very rich RESTful JSON API, which allows other applications to discover and access the datasets
- ❁ It can be integrated easily with Semantic Web technologies such as RDF data model and SPARQL.

First phase - institutional authorization



odule.php/core/loginuserpass.php?AuthState=_15f022bfde5a5101a1129d7dd4774caa41211e2123%3Ahttps%3A%2F%2Fidp.pub.ro%2Fsaml2%2Fidp%2FSSOService.



UNIVERSITATEA POLITEHNICA DIN BUCURESTI



Single Sign-On



Username

ciprian.dobre

Password

.....

[Forgot password](#)

Login

Institutional data repository over ckan



The screenshot shows the CKAN website interface for an institutional data repository. The page is titled "Seturi de date" (Data Sets) and features a navigation menu with options like "Seturi de date", "Organizații", "Grupuri", and "Despre". A search bar is located in the top right corner.

The main content area is divided into two columns. The left column contains a sidebar with the following sections:

- Organizații**: Universitatea Polit... (2)
- Grupuri**: MobyLab (1), UPB (1)
- Cuvinte cheie**: There are no Cuvinte cheie that match this search
- Formate**: XLSX (2), CSV (1)
- Licențe**: There are no Licențe that match this search

The right column displays the search results for "2 seturi de date găsite" (2 data sets found). The search bar contains "Search datasets...". The results are ordered by "Relevanță" (Relevance).

The first result is "Test2", marked as "PRIVAT" (Private). It has an "XLSX" format and a description: "Acest set de date nu are descriere" (This data set has no description).

The second result is "TestDataset", also marked as "PRIVAT". It has "XLSX" and "CSV" formats and a description: "Acest set de date nu are descriere".

At the bottom of the page, there is a footer with the text "Despre CKAN" and "API CKAN" on the left, and "Alimentat de" followed by the CKAN logo on the right.

What's next



- ✿ We aim to finish with eduGAIN connectors (Moodle and ckan)

 - ✿ Institutional channels sometimes take time

- ✿ We need to finalise the construction of classes

 - ✿ Train of trainers: May 2022 → proposal to host on the centralised Moodle for now the materials (quite soon)

 - ✿ For Train of trainees: let's collect all course materials and IPR constraints

- ✿ Carpenter of course catalogues

 - ✿ Some will be taught using the local Moodle instances (eduGAIN + user registration)

 - ✿ Some will possibly be formed and may be taught over the Carpenter solution

THANK YOU!



Follow us

