



ISTI Technical Reports

Open Science repository platforms

Paolo Manghi, ISTI-CNR, Pisa, Italy

Michele Artini, ISTI-CNR, Pisa, Italy

Sandro La Bruzzo, ISTI-CNR, Pisa, Italy

Enrico Ottonello, ISTI-CNR, Pisa, Italy

Gina Pavone, ISTI-CNR, Pisa, Italy



Open Science repository platforms

Manghi P., Artini M., La Bruzzo S., Ottonello E., Pavone G.

ISTI-TR-2022/009

Institutional and thematic repositories today play a key role in scholarly communication and more broadly in scientific workflows. Many institutions and communities have set the ambitious goal of providing an open access repository for their community of users. However, given the amount of expectations from their users, choosing the right solution is often a non-trivial choice. Some platforms may be served out-of-the-box, to be put in operation after straightforward configurations, but are in general less customizable to adhere to specific functional, non-functional, or contextual needs. Other platforms may be instead extremely customizable and flexible but require skilled personnel for their adaptation and deployment. This report performs an analysis of existing state-of-the-art Open Source repository solutions from the functional, operational, and software perspectives. As a result of the analysis, it will factor out the pros and cons of such solutions and identify typical scenarios of adoption.

Keywords: Repository, Repository platforms, Repository software, Dryad, DataVerse, DSpace, InvenioRDM.

Citation

Manghi P., Artini M., La Bruzzo S., Ottonello E., Pavone G., *Open Science repository platforms*. ISTI Technical Reports 2022/009. DOI: 10.32079/ISTI-TR-2022/009.

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"
Area della Ricerca CNR di Pisa
Via G. Moruzzi 1
56124 Pisa Italy
<http://www.isti.cnr.it>

Technical Report

Paolo Manghi, Michele Artini, Sandro La Bruzzo, Enrico Ottonello, Gina Pavone

Table of content

Table of content	1
Introduction	1
1 Repository platforms	2
1.1 Methodology	2
1.2 Analysis	5
Dryad	5
DataVerse	6
InvenioRDM	8
DSpace	10
3 Repository platform software	12
4 Discussion	14

Introduction

Institutional and thematic repositories today play a key role in scholarly communication and more broadly in scientific workflows. Due to the increasing demand of Open Science and open Access mandates, repositories are increasingly adhering to the vision of the COAR Next Generation Repositories Working Group, to “position repositories as the foundation for a distributed, globally networked infrastructure for scholarly communication, on top of which layers of value-added services will be deployed, thereby transforming the system, making it more research-centric, open to and supportive of innovation, while also collectively managed by the scholarly community”. Repositories are today the holders of research products, including publications, datasets, software, and interlinking via persistent identifiers with scholarly communication registries, such as ORCID, ROR.org, Crossref, DataCite, Fundref, CORDIS projects, OpenDOAR, FAIRSharing, re3data.org, etc..

Many institutions and communities have set the ambitious goal of providing an open access repository for their community of users. However, given the amount of expectations from their users, choosing the right solution is often a non trivial choice. Some platforms may be served out-of-the-box, to be put in operation after straightforward configurations, but are in general less customizable to adhere to specific functional, non-functional, or contextual

needs. Other platforms may be instead extremely customizable and flexible but require skilled personnel for their adaptation and deployment.

This report performs an analysis of existing state-of-the-art Open Source repository solutions from the functional, operational, and software perspectives. As a result of the analysis, it will factor out the pros and cons of such solutions and identify typical scenarios of adoption.

1 Repository platforms

This section analyses four known Open Source software repository platforms and related instances, to investigate the functional and operational solutions adopted by the community. The platforms are Dryad, DataVerse, DSpace, and InvenioRDM, all proposing different Open Source software solutions (with the exception of Figshare), meant to offer an out-of-the-box repository platform, to some extent customizable to community needs.

1.1 Methodology

Such solutions were picked among others because of their wide adoption by institutions and communities, and eventually by scientists. Their functionalities evolved over time, to adapt to the Open Science and FAIRness demands of policymakers and research communities, and address the needs of reporting of institutions and researchers (see [Next Generation Repository Working](#) Group reports). The aim of this survey is to analyze such solutions to identify their specific reactions to such demands, moving away from today's old-fashioned repository as a static container of files and metadata. The analysis will consider the features in Table 1, for open-source and non-open source software platforms and related repository instances.

Table 1. Repository features

Feature	Description
Repository instance	
Types of research products	<i>Users can deposit publications, datasets, and/or software</i>
Storage capacity	<i>Storage limits / pay to deposit</i>
Long-term Preservation	<i>Preservation policy/ToA</i>
Metadata/file curation functionalities	<i>Validation, rejection, curation of metadata and/or deposited products</i>
EC Projects tagging	<i>Metadata includes links to EC project in Participant Portal</i>
Programmatic access	<i>Ability to integrate with scientific workflows to publish products and to access products programatically via APIs</i>
Integration with entity registries	<i>Range of integration with entity registries, with/without validation mechanisms</i>

EOSC-Core compliance	<i>Degree of EOSC-Core compliance of the instance</i>
Restricted access control	<i>Access control to metadata record and files</i>
Repository URL	<i>URL to inspected instance (if any)</i>
Repository software	
Open Source software	<i>Yes/No, if yes mention components</i>
Software modularity	<i>Modular design enabling the extension/customization</i>
Metadata model customizability	<i>Degree of customization of the metadata, standard metadata formats, vocabularies, etc.</i>
Software URL	<i>URL to software sites or documentation (if any)</i>

Types of research products Following the approach of resource modeling recommended by the EOSC, research products can be classified into four meta-entities: *publications, research data, research software, and others*, i.e. all products whose nature does not match one of the other entities. This property characterizes an instance of a repository, as typically software platforms are agnostic of the type. The classification however does not necessarily match the EOSC data model and typically requires the definition and maintenance of a mapping.

Limits to free deposition This feature depends on the specific instance of the repository, on its governance, and its sustainability strategy. Catch-all, cross-country, cross-institution, cross-discipline repositories like Zenodo, Figshare, Dryad, etc. generally offer a free storage quote per record, which require payment of fees to be exceeded. The quota policies observed are of two kinds:

- **By the record** Users are allowed to deposit a new product in the repository, uploading files for a total disk space that cannot exceed a given quota (e.g. 50GB in Zenodo.org)
- **By the file in the record** Users are allowed to deposit a new product in the repository, uploading an arbitrary number of files, each of which cannot exceed a given quota (e.g. 3GB for University of Minho's DataVerse).

The approaches have pros and cons, to be evaluated by the platform operators based on the expected average behaviour and requirement of platform users. Under the first modality, users are limited in the overall storage of one product but can upload files up to a given quota. When they need to exceed this quota, either they request and get permission for the exception, or they need to fragment their products into different (possibly semantically related) products. Under the second modality, users are not limited by the product-level quota but need to fragment their product files into smaller ones.

Another option, however not supported by the investigated solutions, is to have a quota by collection (e.g. Zenodo's communities, DataVerse's dataverses). Such an option may require a non-trivial change in the software platforms.

Long-term preservation Long-term preservation is the hardest of the commitment for the service provider. As such, it depends on the specific instance and on the ability of the service provider to commit to long-term storage.

Metadata-file curation functionalities: Data curation functionalities (where data means everything that is metadata or file) involve two main aspects. The first is to offer collaboration and validation tools to a group of community curators to make sure the data matches the expectation of the community at hand, in quality, formats, etc. The second is to make sure the end-users, the scientists, can interact with the curators to have their enquires responded to speed up the enhancement of the data and speed its publishing.

Programmatic access Programmatic access enables third-party services to perform product depositions and/or metadata searches, exports, and downloads via APIs. The former is of paramount importance today, to allow for the implementation of scientific workflows that are capable of depositing into the repository on behalf and prior authorization of the scientists. The latter is important to ensure the repository can expose its content to other scholarly communication services, such as aggregators, ultimately to enable the realization of customs UIs using the repository as a back-end (e.g. Zenodo). This practice is quite common and in some cases offered as-a-service (e.g. Figshare for institutions).

Integration with entity registries Integration with PID systems can be supported at two different degrees. The basic integration level is one where the repository metadata includes fields dedicated to the interlinking with external entity registries, managing entity identities (via PIDs, cool URIs, handles, etc.), such as DataCite, Crossref, ORCID, ROR, Commons, etc. The approach is subject to human mistakes, in the format of PID, which may be “misspelled”, or in the referencing, i.e. the wrong PID may be used. A deeper and optimal level of integration is one where the insertion of a PID is supported by direct interaction with the related registry APIs, ensuring both format and objects are correct.

Restricted access control This feature is rather important, as it supports users with different levels of restriction options and granularity. The feature depends on the software platform and is supported as functionality for its instantiations. Users may deposit research products and fine-tune access rights (restricted, open access, embargo, etc.) for metadata and/or files and to all users or a subset of users (community).

Open Source software Repository platforms may be released as Open Source and in some cases depend on other Open Source products that may be of interest in identifying the best solution out there.

Software modularity Repository platforms are typically designed to be modular, meaning that new functionalities can be easily plugged into the system. However, different degrees of modularity are possible, to the extent that for some platforms only minor changes or extensions are allowed.

Metadata model customizability Customization of the metadata format is an important feature of the repository software, which is directly related to the flexibility of adoption and therefore the scope of reuse across different use cases. On the other hand, a higher degree of customization impacts the out-of-the-box capabilities of a repository, which cannot ground

on data model assumptions. The right balance between customization and added-value functionalities is an important measure of the quality of repository platform design.

Persistent identifiers Repositories must rely on persistent identifiers, which are typically issued at the level of the record, to uniquely refer to the pair metadata-files. Software platforms may be more or less flexible with respect to the identifiers scheme to be used (e.g. handles, DOIs), by offering support to one or more specific PID Agencies (e.g. DataCite, EZID) and by enabling the integration with any PID agency. Flexibility of PID schemes is an important feature for a general-purpose redeployable software platform, which should be ready to support diverse institutional or community scenario.

1.2 Analysis

This section includes a table for each of the identified repository software solutions, describing how they satisfy the features identified in the previous section.

Dryad

Feature	Description
Repository instance	
Types of research products	Dryad is a research data repository, hence assumes the products deposited are data resulting from or useful to the process of science. For deposition of software, Dryad offers an integration with Zenodo.org (files uploaded via Dryad and published in Zenodo), developed under a partnership between the two organizations.
Storage capacity	There are a variety of paid membership plans available to institutions and publishers for depositing datasets. Pricing is based on factors such as the level of research grant funding. The limit for a data publication is 300GB, but individual files cannot exceed 10GB. Larger size submissions are possible, but an explicit request is necessary and fees are applicable.
Long-term Preservation	Content is stored indefinitely at California Digital Library's Merritt Preservation Repository , which applies a distributed storage of data clones, at different geographic locations, copied overnight.
Metadata/file curation functionalities	For the official Dryad installation, curation is performed at the system level, ensuring metadata is complete and both metadata and files comply to the platform recommendations . Self-validation tools are available (e.g. tables).
EC Projects tagging	Dryad includes integration with Funder DOIs from

	Crossref. So EC project tagging is available as far as EC projects are assigned DOI.
Programmatic access	Dryad is integrated with publishing platforms, to ensure data publishing and review of submitted data by assigning temporarily DOIs. It supports SWORD APIs and Shibboleth/OAuth2 protocols.
Integration with entity registries	The platform is integrated with ORCID, ROR, and Funder IDs from Crossref.
EOSC-Core compliance	Dryad is integrated with Make Data Count for usage data, hence usage stats are shared with the EOSC Accounting for Usage statistics. Dryad default schema is based on DataCite so its products can be onboarding into the EOSC catalogue, as long as DataCite is made fully compliant with OpenAIRE guidelines . This may vary from installation to installation as compliance is not provided out of the box.
Restricted access control	All metadata and files in Dryad are by policy under the CC0 waiver
Repository URL	http://dryad.org
Repository software	
Open Source software	The platform is Open Source and highly modular and based on the Stash software. The software is organized into three modules: Store (deposition of metadata and files), Harvest (export of metadata to third-party services and to a full-text Solr index), and Share (GeoBlacklight UIs).
Software modularity	Modular design enables the extension/customization by proper software adaptation of new export protocols and new metadata schemas.
Metadata model customizability	Metadata schema can be customized to match application needs.
Persistent Identifiers	Dryad issues DOIs from DataCite for all datasets
Software URL	https://datadryad.org/stash/our_platform https://github.com/CDL-Dryad/dryad-app

DataVerse

Feature	Description
Repository instance	

Types of research products	The DataVerse platform is focused on research data. The platform is highly configurable to include custom vocabularies.
Storage capacity	The instances we have analyzed, namely the University of Minho and Harvard, do not impose any costs on the deposition of datasets but limit users to deposit files smaller than 3GB and 5GB respectively. Each record can however include an arbitrary number of files. The software allows storage on a local university cluster or, more generally, via S3.
Long-term Preservation	Preservation policy/ToA for the University of Minho is 10 years.
Metadata/file curation functionalities	DataVerse repositories are organized into collections called DataVerses to which super-users can assign users with 9 different roles, establishing rights of publishing (draft, public, to be validated), access to the files (read-only, update, delete, access to the record, access to files), and curation (right to update and publish). By creating a DataVerse and by assigning the proper roles to users, DataVerse installations can support a data curation process.
EC Projects tagging	Some DataVerse instances include links to external project directories (Consortio Madroño), so an extension to include EC Participant Portal APIs or OpenAIRE APIs to project is doable.
Programmatic access	DataVerse is designed to be integrated with other systems, e.g. OJS and publisher platforms. Scientific workflow examples exist (lab notebooks in RSpace). Workflow can create a draft deposition for review.
Integration with entity registries	The platform is designed to be integrated with external systems and exploit API to insert values in research data fields via UIs. Integration with ORCID APIs is available, but others can be built.
EOSC-Core compliance	The platform is OpenAIRE compliant by default. Once the platform is registered (onboarded) as an EOSC Service, its research data can be integrated into the EOSC catalogue.
Restricted access control	Access to research data can be controlled at the level of the DataVerse collection, at the fine-grain level of metadata record and individual file.
Repository URL	https://datarepositorium.uminho.pt/ https://dataverse.harvard.edu
Repository software	

Open Source software	The software is Open Source and developed as a monolithic block, which can be customized in some of the core capabilities such as the underlying storage system and the data model.
Software modularity	The software is designed as a ready-to-go platform, where customization is possible via plug-ins, to fetch data from external systems
Metadata model customizability	DataVerse software supports three levels of metadata: metadata for citation (standard DDI), metadata for journal info (linking to external publications), and disciplinary metadata (provided with six default templates). The three come with a DataVerse schema which can be further customized to include application-specific fields for citation, journal, and discipline. It is possible to create new templates for the discipline metadata, to be shared with the community.
Persistent identifiers	The Dataverse Software currently supports creating identifiers using DOI and Handle.
Software URL	https://dataverse.org https://www.iq.harvard.edu/roadmap-dataverse-project

InvenioRDM

Feature	Description
Repository instance	
Types of research products	Users can deposit publications, datasets, and/or software.
Storage capacity	For the demo installation of Invenio RDM, storage capacity is limited to 10GB per file and based on CERN's data center.
Long-term Preservation	As for the data center of CERN, preservation is guaranteed for 10 years
Metadata/file curation functionalities	InvenioRDM includes communities to model collections of research products. Research products, i.e. depositions, can be managed by multiple users (shared submissions). Communities support curation/management workflows, where different users with different roles (curator, manager, reader, owner) are involved to

	<p>ensure smooth, tracked, deposition workflows. Deposition of metadata and files is structured as a “pull request” in software repositories, in which submitter and curators (who can modify the metadata) are engaged in a discussion via internal ticketing systems.</p> <p>Workflows can be customized, to include specific steps of approval, at the community level: assigning roles of submitters subject to validation and curators notified of new submissions and in charge of the evaluation. Multiple curators can interact with the same submitter for the same submission. Also, requests for extra storage may be sent and handled by community managers.</p>
EC Projects tagging	Metadata includes links to EC projects in Participant Portal or other funders (via OpenAIRE APIs), but more in general UIs can be adapted to access third-party APIs to fetch values.
Programmatic access	<p>Authorized services can deposit files. EOSC AAI is being integrated to enable programmatic deposition of research products under the delegation of EOSC users.</p> <p>The index offers indexing synchronization functions, which make sure that a new deposition is indexed in InvenioRDM full-text index but also in external indexes (provided authorization to access). This ensures synchronization with external systems.</p>
Integration with entity registries	Users can specify ORCID and ROR PIDs for creators and related affiliations via the UIs. The selection of PIDs is enabled by ORCID and ROR APIs, otherwise, textual values can be typed in by users.
EOSC-Core compliance	The platform is by default OpenAIRE compliant, thus ready to onboard research products into the EOSC Catalogue. It will provide in 2022 also full integration with EOSC AAI to support authorization delegation, hence the capacity of services to deposit in InvenioRDM on behalf of users.
Restricted access control	Access can be controlled at the level of the community (apply restrictions to community members and to non-community members) or at the level of the record, at the granularity of the metadata and the files. The embargo function ensures that a

	record is made public at the expiring date, without users to perform any manual action.
Repository URL	https://inveniordm.web.cern.ch/
Repository software	
Open Source software	InvenioRDM is a software that enables to deploy a repository platform out of the box, customized to match specific application needs (storage and data model). The software is developed as a customization of the Invenio Framework v3.0 gluing known Open Source tools such as Elastic Search, OpenSearch, Postgres, and based on JSON and DataCite format.
Software modularity	Being based on the Invenio Framework, the software is modular. Storage can be of any kind (e.g. S3), as well as indexing and database systems. Adaptation to other indexing/db solutions requires development actions.
Metadata model customizability	The data model implements the DataCite guidelines but can be customized with extra fields. Interaction with vocabularies can be implemented by integrating external APIs to vocabulary systems or PID registries (e.g. ORCID, ROR).
Persistent identifiers	InvenioRDM can register DOIs with DataCite for all records, and allows you to write plugins for other identifier schemes.
Software URL	https://github.com/inveniosoftware/invenio-app-rdm

DSpace

Feature	Description
Repository instance	
Types of research products	All types of research products (customizable).
Storage capacity	Details about storage capacity were not available at the visited sites, likely due to the fact they are operating institutional repositories, part of a family of Library Services sharing a common institutional data center.
Long-term Preservation	Access to institutional repositories was not possible, so details about limits to individual uploads or at the record level are not available.
Metadata/file curation	DSpace enables the definition of "Tasks" via a

functionalities	customizable curation system (https://wiki.lyrasis.org/display/DSDOC6x/Curation+System#CurationSystem-Tasks). Tasks enable checks and controls over metadata and files.
EC Projects tagging	Integrated with OpenAIRE APIs to fetch EC Project IDs.
Programmatic access	The platform offers APIs for the deposition of files and metadata.
Integration with entity registries	The platform is integrated with ORCID.
EOSC-Core compliance	The platform is by default OpenAIRE compliant, thus ready to onboard research products into the EOSC Catalogue.
Restricted access control	DSpace allows you to control read/write permissions site-wise, per community, per collection, per item and per file. You may also delegate administrative permissions per community or per collection.
Repository URL	https://uhcl-ir.tdl.org/
Repository software	
Open Source software	DSpace is an out-of-the-box repository platform.
Software modularity	The couple UI/index (metadata store) is decoupled from the file storage. Files in DSpace can be stored either using a local filesystem (default) or a cloud-based solution, such as Amazon S3. DSpace comes with a suite of tools (batch ingest, batch export, batch metadata editing, etc.) and plugins for translating content into DSpace objects.
Metadata model customizability	By default, DSpace uses a Qualified Dublin Core (QDC) based metadata schema. Institutions can extend that base schema or add custom QDC-like schemas. DSpace can import or export metadata from other major metadata schemas such as MARC or MODS.
Persistent identifiers	DSpace support handle system by default, but also integration with DOI DataCite and EZID identifiers (ARK, DOIs).
Software URL	https://duraspace.org/dspace/download/ https://github.com/DSpace/DSpace/releases/ https://github.com/DSpace/dspace-angular/releases/

2 Software analysis

Open Source repository platforms are maintained by a community of developers, and can be deployed by institutions and/or research organizations. In particular, the optimal software solution should at least meet the following criteria:

Sustainable Open Source project Building new software from scratch is not an option, due to the cost and the time required to build trust around a community of developers. The software should be mature, adaptable to new requirements, and count on a lively community of developers.

Modular with respect to functionalities Local requirements may not be supported out-of-the-box by the software. Examples are updates required to ensure compliance to the OpenAIRE guidelines (aka EOSC Interoperability Framework), or the integration of tools for measuring data FAIRness and data curation.

Flexible with respect to data model Software could be more or less adaptable to store any kind of research product, customizable in the data model (e.g. metadata structure), in the vocabularies of given metadata fields, and in terms of integrating with PID systems or registries in the scholarly communication infrastructure.

Decoupling the web application from the storage infrastructure Since the software has to serve potentially different scenarios, such as the cross-institutional, cross country deployment setting, or the institutional one, the application layer of the software should allow for the integration with different storage systems, as made available by the repository provider. Examples are Amazon S3 standard storage or simpler local storage solutions, typically provided by institutional data centers.

Data Curation The software should provide out-of-the-box data curation tools, to support communities at defining collections and implement data validation workflows respecting community criteria and policy for data quality.

Programmatic access Enabling deposition and access to records to third-party services is a key feature to support the implementation of scientific workflows driven by services, delegated by humans to perform discovery, analysis, and publishing of research data or software.

The table below describes the four Open Source solutions we have identified in more detail, investigating the extent to which they meet the requirements above.

	InvenioRDM	Dryad	DataVerse	DSpace
Sustainable Open Source project	InvenioRDM has a community of 31 contributors	Dryad has a community of 21 contributors on GitHub .	DataVerse has a community of 134 contributors on GitHub .	DSpace has a community of 158 contributors on GitHub .

	<p>on GitHub. It's the youngest project, active since Jun 2019, building on Invenio Framework.</p> <p>License: MIT License</p>	<p>License: MIT License</p>	<p>License: Apache License, v2.0</p>	<p>including companies who make a living out of the realization of extensions and modules.</p> <p>License: BSD 3-Clause License</p>
Modular with respect to functionalities	<p>Ready to go platform. Modular software, where storage of files, metadata database, indexing, web portals, are modularly related via Invenio Framework, but can be replaced with technology of choice.</p>	<p>Ready to go platform. The software is modular as it separates store, from indexing, and web portals. It allows the customizaton of export protocols.</p>	<p>Ready to go platform, customizable via modules to interact with external sources or export data.</p>	<p>Ready to go platform. The software is modular as it separates store, from indexing/web portals. Modules are made available, by companies and contributors.</p>
Flexible with respect to data model	<p>Metadata is DataCite centric but customizable. All research products are supported.</p>	<p>The software is research data centric and based on DataCite metadata schemas.</p>	<p>The software is research data centric, metadata schemas are offered for several communities, but can be customized.</p>	<p>The software is Dublin Core centric, but can be customized to include community fields. All research products are supported.</p>
Decoupling the web application from the storage infrastructure	<p>The software decouples storage modules and indexing modules from the application layer.</p>	<p>The software decouples storage from indexing and web portals.</p>	<p>The software decouples application from storage.</p>	<p>The software decouples application from storage.</p>
Data Curation	<p>The software includes advanced data curation workflows,</p>	<p>Data curation is performed at the platform level, by platform</p>	<p>The software support data curation in terms of validation or</p>	<p>The software support customizable data "curation tasks" as</p>

	included as a major innovation of the state-of-the-art in the field, where interaction between submitters and curators is possible.	administrators/curators, which ensure the data meets minimal criteria of eligibility.	rejection of deposition of a controlled collection (dataverse)	controls over metadata and files.
Persistent identifiers	Support to DOIs via DataCite, plugins to other identifier schemes are possible	Support to DOIs via DataCite	Support to DOIs and Handle	Support to handle system by default, optional integration of DOI from DataCite and EZID identifiers (ARK, DOIs)
Programmatic access	The software supports deposition APIs under the control of security context, which will include EOSC AAI Federation.	The platform allows deposition of datasets from publishing platforms, to support the deposition of temporary datasets during submission and reviews of research papers.	The platform allows deposition of datasets from publishing platforms, to support the deposition of temporary datasets during submission and reviews of research papers.	The platform offers APIs for the deposition of research products.

3 Discussion

Among the four Open Source solutions investigated in this report, InvenioRDM and DSpace seem the solutions that meet at best a scenario where customizability is a strong requirement. InvenioRDM's software has been designed after the lesson learned in realizing Zenodo.org using the Invenio Framework and meeting the requirements of Zenodo users. As such it is a good balance of out-of-the-box functionalities and flexibility of customization and software extension. Moreover, it is designed as a catch-all repository, targeting all kinds of products, with special attention to the data curation functionalities. Similarly, DSpace 7 has built on the core platform and experience reached up to the release of DSpace 6, to bring a "a single, modern user interface and REST API and integrates current technological standards and best practices". Comparing the two, InvenioRDM addresses data curation

enabling an interaction via UIs between data curators and end-users, so going beyond metadata and file checks, while DSpace 7 shows a stronger users base, with more than 2000 installations, and pool of contributors.

DataVerse follows immediately after, due to its high degree of flexibility in the customization and its large developer community, which delivers state-of-the-art software modules, adding functionalities to any deployment. It is research data-oriented, which leaves any different usage a customization yet to be experimented.

Dryad and DataVerse are the best choice when the scenario requires a data repository to be set up. The two platforms offer ready-to-go solutions which require minimal effort of deployment and customization while offering a rich set of functionalities, for data management, access, and control.



This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101017452 (project *OpenAIRE-Nexus Scholarly Communication Services for EOSC users*).