A Primer on Open Science-driven Repository Platforms

Alessia Bardi[0000-0002-1112-1292], Paolo Manghi[0000-0001-7291-3210], Andrea Mannocci[0000-0002-5193-7851], Enrico Ottonello[0000-0003-4606-7422], and Gina Pavone[0000-0003-0087-2151]

Institute of Information Science and Technologies, National Research Council, Pisa, Italy
name.surname@isti.cnr.it

Abstract. Following Open Science mandates and FAIR principles, institutions and communities increasingly demand repositories with native support for publishing scientific literature together with research data, software, and other research products. Such repositories may be thematic or general-purpose and are deeply integrated with the scholarly communication ecosystem to ensure versioning, persistent identifiers, data curation, usage stats, and so on. Identifying the most suitable off-the-shelf repository platform is often a non-trivial task as the choice depends on the functional requirements, the programming and technical skills available, and infrastructure resources.
This work analyses four state-of-the-art Open Source repository platforms, namely Dryad, Dataverse, DSpace, and InvenioRDM, from the functional and software perspectives. This work intends to provide an overview serving as a primer for choosing repository platform solutions in different application scenarios. Most importantly, this paper highlights how these platforms reacted to some key Open Science demands, moving away from the original and old-fashioned concept of a repository serving as a static container of files and metadata.

Keywords: Repository, Repository platforms, Repository software, Dryad, Dataverse, DSpace, InvenioRDM, Open Science, FAIR, Research Data.

1 Introduction

Following Open Science mandates and FAIR principles, institutions and communities increasingly demand repositories with native support for publishing scientific literature together with research data, research software, and other scientific products. As envisioned by the COAR Next Generation Repositories Working Group [1] and other initiatives in the digital libraries' domain [2], repositories form "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". The increasing adoption of Open Science and Open Access mandates further urges this process, with many institutions, infrastructures, and research communities setting the ambitious goal of providing an "Open Science-driven" repository for their community of users. Such repositories may focus on research datasets or become the holders of all kinds of research products, including publications, datasets, and software, and aim at supporting out-of-the-box features such as FAIRness, collaboration, access control, and data curation. Their instances should be able, to some extent, to adapt to community-specific requirements and ensure interoperability with third-party scholarly communication services, be part of scientific workflows, and support an unambiguous scholarly communication ecosystem by interacting with known registries, such as ORCID, ROR.org, Crossref, DataCite, and accession numbers.

However, researchers' and policymakers' requirements often make choosing the suitable repository platform a non-trivial choice for organisations as institutional, and community demands may vary in terms of kinds of research products, metadata descriptions, and functionalities (e.g., data curation, collaboration). Typically, repository platforms may satisfy these requirements up to some extent, and then require adaptation to better fit additional specific needs [3][4]. Some solutions address general-purpose scenarios, e.g., addressing all kinds of products and enabling customisation of metadata descriptions, while others exhibit less flexibility but deliver a richer set of functionalities. Besides, local requirements, e.g., available skills and resources, may pose constraints and limitations to software choices.

In the attempt to facilitate such a choice, this work analyses four state-of-the-art Open Source repository solutions, namely Dryad, Dataverse, DSpace, and InvenioRDM, from the functional and software perspectives, addressing the selection needs of organisations willing to become repository providers – for an end-users-oriented survey see [13]. The investigation is intended as a "primer" for choosing repository platform solutions in different institutional and community application scenarios. Most importantly, the paper highlights how such known platforms reacted to Open Science demands [5], moving away from the old-fashioned concept of a repository as static containers of files and metadata.

2 Repository platforms

The four Open Source repository platforms at the centre of this analysis were picked among others due to a twofold reason. Firstly, because of their wide adoption by institutions and communities worldwide, a trend that is often followed by company uptake. Secondly, their design and functionalities evolved to address the demands of Open Science scientific workflows [6]. This paper aims to analyse such solutions to identify their specific reactions to such demands, moving away from old-fashioned repositories conceived as static containers of files and metadata.

For this analysis, we referred to the work of the COAR Next Generation Repositories Working Group[1], for which a "next generation repository" should:

- **Support a diversity of research products** and thus manage, preserve, version, curate, and provide access to a broad range of *research products*, including published articles, pre-prints, datasets, working papers, images, and software;
- **Support a diversity of research communities and institutions**, and thus be to some extent customisable to satisfy community/local requisites [7] in terms of products, metadata, and functionalities;

[1] http://ngr.coar-repositories.org

- **Be part of an ecosystem of repositories and scholarly communication services** by interlinking via persistent identifiers their resources to relevant entities, such as author identifiers (e.g., ORCID IDs), project identifiers (e.g., FundRef), organisation identifiers (e.g., ROR.org, ISNI), other resources (e.g., data, software, literature via DOIs, ArXiv IDs);
- **Be machine-friendly and interoperable** by adopting standards [8] that enable a broader range of scientific services, such as scientific workflows, discovery, access, annotation, sharing, quality assessment, content transfer, analytics, provenance tracking, recommendations, and so on.

The following two sections point out the desiderata that can be derived from such a vision in terms of functionalities and software features. Functional features capture the ability to address Open Science resources and workflows and the proactivity expected by modern repositories in the context of scholarly communication. Software features instead frame a software project's maturity, flexibility, and modularity, pointing out the degree of customisation that the platform can meet.

2.1 Functional desiderata

After reviewing the selected repository platforms, the following relevant software functional features emerged:

Research product types. Following the approach of resource modelling recommended by the European Open Science Cloud[2] (EOSC), research products can be classified into four meta-entities: *publications,* e.g., articles, theses, reports, presentations, *research data*, e.g., tables, images, archives, *research software,* e.g., code*, and others,* i.e., all products whose nature does not match one of the other entities. The metadata descriptions of such products may differ profoundly, ranging from bibliographic descriptions to provenance and community-specific tags. Moreover, metadata may include semantic links to other products to capture the entire research lifecycle for the sake of discoverability and reproducibility.

Data (metadata & file) curation functionalities. The ability to engage scientists in data curation and validation processes is becoming prominent, as trust in research data and software is proved to be insufficient [9][10][11]. Data curation functionalities (where data means everything that is metadata or file) regard two main aspects. The first is to offer collaboration and validation tools to a group of community curators to ensure the data matches the expectation of the community at hand in terms of quality, formats, and so on. The second is to ensure the end-users, i.e., the scientists, can interact with the curators to have their enquires responded to speed up the enhancement of the data and its publishing.

Integration with entity registries. To adhere to FAIRness demands and mandates, repositories use persistent identifiers (PIDs) for scholarly communication entities. On the one hand, they provide PIDs for the products uploaded by the users. On the other hand, they enable referencing to scholarly communication entities via PIDs by connecting to the related registries; examples are ORCID for authors and ROR.org for organisations. Integration with PID systems, i.e., registries, can be supported at two different degrees. The basic integration level is one where the repository metadata

[2] https://eosc-portal.eu

includes fields dedicated to interlinking with external entity registries, managing entity identities (via PIDs, cool URLs, handles, and so on), such as DataCite, Crossref, ORCID, ROR, Commons. The approach is subject to human mistakes in the format of PID, which may be “misspelt”, or in the referencing, i.e., an existing, yet 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 PIDs are correct.

Access control. Access control provides users with different levels of restriction options and granularity regarding research product access. Users may deposit research products and fine-tune access rights (e.g., restricted, open access, embargo) for metadata and/or files and to all users or groups of users (e.g., a research community).

Table 1. Functional features.

Feature	*Description*
Research product types	Type of research products that users can deposit (e.g., publications, datasets, and/or software)
Data curation functionalities	Validation, rejection, and curation of metadata and/or files of research products
Integration with entity registries	Integration with PID systems to support and contribute to a non-ambiguous scholarly record
Access control	Users can rule access to metadata and files they deposit

2.2 Software architecture desiderata

Open Source repository platforms are deployed by organisations whose requirements may derive from local technological constraints, peculiar functionalities, or, conversely, due to lack of ICT resources, be served a ready-to-go solution. After reviewing the candidate repository platforms, the following relevant software architecture features were identified:

Software project sustainability. The maturity, traction, and licencing of a software package are key requisites for an organisation to invest in a software product.

Functionality customisation. Repository platforms are typically modular, meaning new functionalities can be easily plugged into the system, but to different degrees, with a trade-off between out-of-the-box and customisation.

Metadata model customisation. The repository can be more or less flexible with respect to the metadata model, e.g., the attributes, the vocabularies, references to external PID systems or registries in the scholarly communication infrastructure. Customisation of the metadata format measures the potential reuse across different communities and use cases, but as a counterpart, it impacts the out-of-the-box capabilities of a repository, which cannot be grounded on data model assumptions.

Custom storage infrastructure. Repository software must be configured and deployed to address potentially different scenarios, such as the cross-institutional, cross-country deployment setting, or community one. Different storage requirements may apply in terms of scalability, preservation, and availability of resources. Examples are Amazon S3 standard storage or simpler local storage solutions, typically provided by institutional data centres. The extent of customisation of the storage infrastructure is therefore relevant to making the right choice.

Integration with scientific services. Programmatic access enables third-party services to perform product depositions and/or metadata searches, exports, and downloads via APIs. The former allows for the implementation of scientific workflows capable of depositing into the repository on behalf and prior authorisation of the scientists. The latter ensures the repository can expose its content to other scholarly communication services, such as aggregators, ultimately enabling the realisation of customs UIs using the repository as a back-end (e.g., Zenodo.org).

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 concerning 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.

Usage statistics. Repository platforms are increasingly integrating with usage statistics infrastructures (e.g., IRUS-UK[3], MakeDataCount[4], OpenAIRE[5]) compliant with COUNTER Code of Practice.[6] Repositories, on their occurrence, centrally share views and downloads events of research products via the related PIDs, enabling aggregation of PID usage statistics across different repositories.

Table 2. Software features.

Feature	*Description*
Software project sustainability	Software project trust is measured by the engagement of developer communities
Functionality customisation	Modular design enabling the extension/customisation of functionalities
Metadata model customisation	Degree of research product types, metadata customisation, vocabularies, etc.
Custom storage infrastructure	Degree of customisation of storage infrastructure
Integration with scientific services	Ability to integrate with scientific services to publish products and/or provide access to products programmatically via APIs
Persistent Identifiers	Platform embeds functionality to mint PIDs for deposited research products
Usage statistics	Availability of modules to integrate with usage statistics infrastructures

2.3 Operational features

Further interesting features are the operational features, which depend on the individual installation policies and resources. The most prominent ones identified during the investigation are long-term preservation [12] and free deposition strategies:

[3] https://www.jisc.ac.uk/irus

[4] https://makedatacount.org

[5] https://www.openaire.eu

[6] https://support.datacite.org/docs/counter-code-of-practice

Long-term preservation. This is among the most important commitments for a repository service provider. It is an operational challenge that depends on the ability of the service provider to commit to long-term storage resources. For example, Zenodo's CERN and the University of Minho's Dataverse[7] storage infrastructures guarantee ten years of preservation.

Free deposition. This is always guaranteed up to a given quota per research product, which demands fees to be exceeded. Two quota typologies are adopted: by the record, where users can, for each record, deposit files for a maximum total amount (e.g., 50 GB in Zenodo.org); by the file in the record, where users can upload an arbitrary number of files but each with a given quota (e.g., 3 GB for University of Minho's Dataverse).

Such aspects will not be regarded in this study as they are instance-specific; nonetheless, they have been reported here for the sake of completeness.

3 Repository platforms analysis

3.1 Dryad

"Dryad is an open source, community-driven project that takes a unique approach to data publication and digital preservation. Dryad focuses on search, presentation, and discovery and delegates the responsibility for the data preservation function to the underlying repository with which it is integrated."[8]. Research data are uploaded with metadata representing the dataset landing pages but are formatted as an online version of a data paper that can be downloaded as an individual PDF file or as part of the complete dataset download package, incorporating all data files for all versions.

Functionality. Dryad supports research data deposition only, although software deposition is possible via integration with Zenodo.org. For the official Dryad installation, curation is performed at the instance level, ensuring metadata is complete and both metadata and files comply with the platform recommendations.[9] Automatic validation tools are available for tables. The platform is integrated with ORCID, as it requires an ORCID to log in and supports ROR.org and FundRef IDs to refer to organisations and projects/grants funding, respectively. Metadata and files in Dryad are by the policy under CC0 waiver, so no fine-grained access control is supported.

Software. Dryad's software is released on GitHub[10] under the MIT Licence and maintained by a community of 21 contributors (as of August 2022). The software is modular and based on the Stash software, organised into three modules: *Store* (deposition of metadata and files), *Harvest* (export of metadata to third-party services and a full-text Solr index), and *Share* (GeoBlacklight UIs). Its design enables the extension/customisation of metadata export protocols and personalisation of metadata schemas. The Store module supports the DataCite format but can be extended to include extra fields. It supports SWORD 2.0 to enable programmatic access from

[7] https://datarepositorium.uminho.pt

[8] https://datadryad.org

[9] https://datadryad.org/stash/faq#files

[10] https://github.com/CDL-Dryad/dryad-app

journal publishing platforms during article submission to deposit research data for peer review temporarily. The Harvest module supports OAI-PMH and ResourceSync, and can be customised to support different export protocols.

3.2 Dataverse

“The Dataverse Project is an open-source web application to share, preserve, cite, explore, and analyse research data. It facilitates making data available to others and allows you to replicate others’ work more easily”.[11] A Dataverse repository hosts multiple archives called Dataverses, each intended as a collection of datasets consisting of descriptive metadata and data files. As a design choice, Dataverse collections may also be nested. Dataverse has been conceived to automate archivists’ and librarians’ tasks and to provide services for and to distribute credit to the data creator.

Functionality. The Dataverse platform focuses on publishing research data and related supplementary material (e.g., code and docs in support of the data). Dataverses are created by super-users that can assign to users nine different roles, establishing rights for publishing (draft, public, to be validated), accessing (read-only, update, delete, access to the record, access to files), and curating (right to update and publish). By creating a Dataverse and by assigning the proper roles to users, Dataverse installations can support custom data curation workflows. The platforms support user interfaces for manually reviewing tabular data.

Dataverse repositories are integrated with ORCID and ROR.org via APIs to ensure up-to-date references to authors and institutions. Access to research data can be controlled at the level of the Dataverse collection, at the fine-grain level of metadata records and individual files. The restricted mode, i.e., download upon request to the owner, is optional.

Software. Dataverse’s software is released on GitHub[12] under Apache Licence v2.0, maintained by a community of 144 contributors (as of August 2022). Designed as a ready-to-deploy package, which can be customised in some core capabilities such as the underlying storage system and the data model. Further customisation is possible via plugins, for example, to fetch vocabulary or entity data from external information systems such as registries.

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 customised to include application-specific fields. It is possible to create new templates for the discipline metadata, to be shared with the community. The platform is also compliant with schema.org to support Google’s data search crawlers.

Dataverse user interfaces can be extended via different tools developed by third parties[13], e.g., for data previewing and data curation. Dataverse is designed to be integrated with other systems via SWORD protocol for data deposition. Examples are

[11] https://dataverse.org

[12] https://github.com/IQSS/dataverse

[13] https://guides.dataverse.org/en/latest/admin/external-tools.html#inventory-of-external-tools

OJS[14] and publisher platforms, supporting publication and data submission workflows, but also integration with existing scientific workflows, e.g., Lab notebooks in RSpace.[15] The platform also supports OAI-PMH protocol standards for metadata harvesting. Its storage layer supports object and file system preservation via S3 or Swift and is configurable at the collection and dataset level.

3.3 DSpace

"DSpace is a web application allowing researchers and scholars to publish documents and data. […] It is free and easy to install "out of the box" and completely customisable to fit the needs of any organisation."[16]. DSpace repositories are organised into *communities* (e.g., departments or institutions) that include *collections*, which are groups of *items*, i.e. data files and metadata descriptions.

Functionality. DSpace items can be set to model any kind of research product. For data curation, DSpace enables the definition of "Tasks" as plugins via a customisable curation framework.[17] Tasks enable checks and controls over metadata and files upon their deposition. The platform is integrated with ORCID for login and reference to authors. DSpace allows controlling read/write permissions at the instance level, per community, collection, item, and file. Administrative permissions per community or per collection can be delegated to users.

Software. DSpace is a mature project released on GitHub[18] under a BSD 3-Clause License, maintained by a community of 166 contributors (as of August 2022), including companies doing business out of its custom extensions and installations.

The component pair 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 (e.g., batch ingest, batch export, batch metadata editing) and plugins for translating content into DSpace objects. 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. DSpace supports the Handle system by default but also integrates with DOI DataCite and EZID identifiers (ARK, DOIs). The platform offers custom APIs for the deposition of files and metadata and OAI-PMH for harvesting metadata.

3.4 InvenioRDM

"InvenioRDM is a turn-key research data management repository platform based on Invenio Framework and Zenodo".[19] Its instantiations offer deposition and access

[14] https://openjournalsystems.com
[15] https://www.researchspace.com
[16] https://dspace.lyrasis.org
[17] https://wiki.lyrasis.org/display/DSDOC6x/Curation+System#CurationSystem-Tasks
[18] https://github.com/DSpace/DSpace
[19] https://inveniosoftware.org/products/rdm

functionalities to a set of communities (i.e., collections) of research products, which in turn can be of any kind.

Functionality. InvenioRDM includes communities to model collections of research products. Deposited research products can be managed by multiple users (i.e., shared submissions). Communities support curation/management workflows, where different users with different roles (i.e., curator, manager, reader, owner) are involved to ensure smooth, tracked deposition workflows. Deposition of metadata and files is structured as a "pull request" in software repositories, in which the submitter and curators (who can modify the metadata) are engaged in a discussion via an internal ticketing system. Workflows can be customised 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.

Users can specify ORCID and ROR IDs for creators and related affiliations via the UIs. The selection of PIDs is enabled by ORCID and ROR APIs for validation, otherwise, textual values can be typed in by users.

Access can be controlled at the level of the community (restrictions to community and 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 performing any manual action.

Software. InvenioRDM is a rather young project, active since June 2019, released on GitHub[20] under an MIT license, and maintained by a community of 32 contributors (as of August 2022). The software is developed as a specialisation of the Invenio Framework v3.0 glueing known Open Source tools such as Elasticsearch, OpenSearch, and Postgres, and based on JSON and DataCite format. The software comes with a ready-to-deploy configuration to deliver a repository instance similar to Zenodo.org.

The metadata data model implements the DataCite guidelines but can be customised with extra fields to match community requirements. Interaction with vocabularies can be implemented by integrating external APIs into vocabulary systems or PID registries (e.g., ORCID, ROR).

Due to the flexibility of the Invenio Framework, the software is highly modular: storage can be of any kind (e.g., S3, file systems), and using different indexing and database systems is possible via programming efforts. The software supports OAI-PMH protocols and offers custom APIs for file and metadata deposition. The index offers indexing synchronisation functions, which mirror a new deposition in the InvenioRDM full-text index on external indexes.

4 Discussion

The four platforms offer ready-to-go mature or experience-based software solutions and can also benefit from companies for the configuration and installation of custom solutions. However, differences exist; Table 3 and Table 4 offer a high-level

[20] https://github.com/inveniosoftware/invenio-app-rdm

comparison of the functionality and software features we have identified for the four platforms. Summarising:

Research Products and metadata customisation beyond research data are addressed by InvenioRDM and DSpace. Dryad and Dataverse are instead specialised in research data, offering dedicated and rich data management functionalities. In general, all platforms offer a degree of customisability of the metadata descriptions. Dataverse, however, is designed to be extremely flexible in this respect, supporting a set of community profiles and a fully flexible metadata framework.

Customisability of functionalities and storage is equally addressed. InvenioRDM and DSpace seem to be the solutions that meet, at best, a scenario where customisability of functionalities is a strong requirement. InvenioRDM's software has been designed after the lesson learned in realising Zenodo.org using the Invenio Framework and meeting the requirements of Zenodo users and Invenio repository providers. Similarly, DSpace 7 has built on the core platform and experience reached up to the release of DSpace 6 to bring a "single, modern user interface and REST API and integrates current technological standards and best practices". As such, the platforms offer a good balance of out-of-the-box functionalities and flexibility of customisation and software extension. Dataverse also shows a high degree of customisability and a rich set of functional modules shared by the community. The four platforms allow the integration of different kinds of storage layers by modularly decoupling them from the user interfaces.

Data Curation is addressed in InvenioRDM by enabling interaction via UIs between data curators and end-users, integrating validation modules, and going beyond the validate-reject workflows. Similarly, but disregarding end-users -curators interactions, DSpace offers a "Task" framework that developers can use to implement research data validation checks. Dryad and Dataverse offer manual validation and rejection procedures at the repository instance and at the Dataverse level, respectively.

Integration with entity registries is well covered by Dryad, followed by InvenioRDM, and then DSpace and Dataverse, which only support ORCID. All platforms enable the integration of entity registries of reference via custom plugins.

Persistent identifier minting is thoroughly addressed by DSpace, supporting the open Handle system and optionally DOIs from DataCite and EZID identifiers. Dataverse follows with DOIs and Handles, while Dryad and InvenioRDM support DOI from DataCite. Plugins to other PID services are, in general, allowed.

Programmatic access is supported by all platforms for both deposition and harvesting. SWORD is provided by Dryad, DSpace, and Dataverse, while InvenioRDM implements a proprietary API. All platforms implement OAI-PMH, while DSpace offers ResourceSynch out-of-the-box.

Usage statistics are supported by all platforms via Make Data Count and COUNTER Code of Practice implementation.

Table 3. Functional features comparison.

Feature	*Dryad*	*Dataverse*	*DSpace*	*InvenioRDM*
Research product types	Research Data	Research Data	All research products	All research products
Data curation functionalities	Manual rejection or	Manual rejection or	customisable data "curation tasks" as	customisable data curation

	approval at repository instance level	approval at "dataverse" level	validation controls over metadata and files upon deposition.	workflows, interaction between submitters and collection managers/curators
Integration with entity registries	ORCID, ROR, FundRef	ORCID	ORCID	ORCID, ROR.org
Access control	Records and files are under CC0 waiver by default	At the granularity of "dataverses", record, and files in the record	At the granularity of "sites", "community", collection, item and files per item	At the granularity of "communities", record, metadata and files of the record; embargo functionality

Table 4. Software feature comparison.

Feature	*Dryad*	*Dataverse*	*DSpace*	*InvenioRDM*
Software project sustainability	21 contributors under MIT Licence	144 contributors, under Apache Licence v2.0	166 contributors, under BSD 3-Clause *Licence*	32 contributors, under MIT *Licence*
Functionality customisation	Customisation of metadata export protocols	Integration with entity registries, customisation metadata exports and UI functions.	Extendible software, a large pool of extensions is available	Extendible software. Building on the Invenio integration framework.
Metadata model customisation	Metadata is DataCite but can be customised.	Sets community metadata schemas which can be customised; supports schema.org	Metadata is Qualified Dublin Core and can be customised	Metadata is DataCite but can be customised.
Custom storage infrastructure	Decouples storage from indexing and web portals.	Decouples web portals from storage.	Decouples web portals from storage.	Decouples web portals from storage and indexing
Integration with scientific services	SWORD and OAI-PMH	SWORD and OAI-PMH	Deposition API, OAI-PMH, ResourceSynch	Deposition API, OAI-PMH, index synchronisation
Persistent identifiers	DOIs via DataCite	DOIs and Handles	Handle system, optional DOI from DataCite and EZID identifiers (ARK, DOIs)	DOIs via DataCite
Usage statistics	Make Data Count	Make Data Count	COUNTER Code of Practice	COUNTER Code of Practice

References

1. Rodrigues, Eloy, Bollini, Andrea, Cabezas, Alberto, Castelli, Donatella, Carr, Les, Chan, Leslie, Humphrey, Chuck, Johnson, Rick, Knoth, Petr, Manghi, Paolo, Matizirofa, Lazarus, Perakakis, Pandelis, Schirrwagen, Jochen, Selematsela, Daisy, Shearer, Kathleen, Walk, Paul, Wilcox, David, & Yamaji, Kazu. (2017). Next Generation Repositories: Behaviours and Technical Recommendations of the COAR Next Generation Repositories Working Group. Zenodo. https://doi.org/10.5281/zenodo.1215014
2. Dempsey, L. (2016). Library collections in the life of the user: two directions. LIBER Quarterly: The Journal of the Association of European Research Libraries, 26(4), 338–359. https://doi.org/10.18352/lq.10170
3. Claire Austin, Susan Brown, Nancy Fong, Chuck Humphrey, Amber Leahey, & Peter Webster. (2015). Research Data Repositories: Review of current features, gap analysis, and recommendations for minimum requirements (Version 0). https://doi.org/10.29173/iq904
4. Assante, M., Candela, L., Castelli, D. and Tani, A., 2016. Are Scientific Data Repositories Coping with Research Data Publishing?. Data Science Journal, 15, p.6. DOI: http://doi.org/10.5334/dsj-2016-006
5. Burgelman Jean-Claude, Pascu Corina, Szkuta Katarzyna, Von Schomberg Rene, Karalopoulos Athanasios, Repanas Konstantinos, Schouppe Michel, *Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century*, Frontiers in Big Data, (2), 2019, doi:10.3389/fdata.2019.00043
6. RepOSGate: Open Science Gateways for Institutional Repositories, Michele Artini, Leonardo Candela, Paolo Manghi & Silvia Giannini
7. Jwa, A. S., & Poldrack, R. A. (2022). The spectrum of data sharing policies in neuroimaging data repositories. Human Brain Mapping, 43 (8), 2707–2721. https://doi.org/10.1002/hbm.25803
8. Forero, D.A., Curioso, W.H. & Patrinos, G.P. The importance of adherence to international standards for depositing open data in public repositories. BMC Res Notes 14, 405 (2021). https://doi.org/10.1186/s13104-021-05817-z
9. Siaw-Teng Liaw, Jason Guan Nan Guo, Sameera Ansari, Jitendra Jonnagaddala, Myron Anthony Godinho, Alder Jose Borelli, Jr, Simon de Lusignan, Daniel Capurro, Harshana Liyanage, Navreet Bhattal, Vicki Bennett, Jaclyn Chan, Michael G Kahn, Quality assessment of real-world data repositories across the data life cycle: A literature review, Journal of the American Medical Informatics Association, Volume 28, Issue 7, July 2021, Pages 1591–1599, https://doi.org/10.1093/jamia/ocaa340
10. Löffler F, Wesp V, König-Ries B, Klan F (2021) Dataset search in biodiversity research: Do metadata in data repositories reflect scholarly information needs? PLoS ONE 16(3): e0246099. https://doi.org/10.1371/journal.pone.0246099
11. Bashir, Saimah, Sumeer Gul, Shazia Bashir, Nahida Tun Nisa, and Shabir Ahmad Ganaie. Evolution of Institutional Repositories: Managing Institutional Research Output to Remove the Gap of Academic Elitism. Journal of Librarianship and Information Science 54, no. 3 (September 2022): 518–31. https://doi.org/10.1177/09610006211009592.
12. Barrueco, J.M. and Termens, M. (2022), Digital preservation in institutional repositories: a systematic literature review, Digital Library Perspectives, Vol. 38 No. 2, pp. 161-174. https://doi.org/10.1108/DLP-02-2021-0011
13. Boch, M. *et al.* (2022). A Systematic Review of Data Management Platforms. In: Rocha, A., Adeli, H., Dzemyda, G., Moreira, F. (eds) Information Systems and Technologies. WorldCIST 2022. Lecture Notes in Networks and Systems, vol 469. Springer, Cham. https://doi.org/10.1007/978-3-031-04819-7_2