PRACTICE AND IMPLEMENTATION PAPER

FAIREST: A Framework for Assessing Research Repositories

Mathieu d'Aquin[1†], Fabian Kirstein[2,3], Daniela Oliveira[4], Sonja Schimmler[2,3] & Sebastian Urbanek[2,3]

[1]LORIA, Université de Lorraine, CNRS, INRIA 54506, Vandœuvre-lès-Nancy, France

[2]Fraunhofer FOKUS, 10589 Berlin, Germany

[3]Weizenbaum Institute for the Networked Society, 10623 Berlin, Germany

[4]LaSIGE, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal

Keywords: research repositories; FAIR principles; open access; research data

Citation: Mathieu, A., Fabian, K., Daniela, O., et al.: FAIREST: A Framework for Assessing Research Repositories. Data Intelligence 5(1), 202-241 (2023). doi: 10.1162/dint_a_00159

Received: March 1, 2022; Revised: May 4, 2022; Accepted: June 7, 2022

ABSTRACT

The open science movement has gained significant momentum within the last few years. This comes along with the need to store and share research artefacts, such as publications and research data. For this purpose, research repositories need to be established. A variety of solutions exist for implementing such repositories, covering diverse features, ranging from custom depositing workflows to social media-like functions.

In this article, we introduce the FAIREST principles, a framework inspired by the well-known FAIR principles, but designed to provide a set of metrics for assessing and selecting solutions for creating digital repositories for research artefacts. The goal is to support decision makers in choosing such a solution when planning for a repository, especially at an institutional level. The metrics included are therefore based on two pillars: (1) an analysis of established features and functionalities, drawn from existing dedicated, general purpose and commonly used solutions, and (2) a literature review on general requirements for digital repositories for research artefacts and related systems. We further describe an assessment of 11 widespread solutions, with the goal to provide an overview of the current landscape of research data repository solutions, identifying gaps and research challenges to be addressed.

† Corresponding author: Mathieu d'Aquin (E-mail: mathieu.daquin@loria.fr; ORCID: 0000-0001-7276-4702).

1. INTRODUCTION

The transparency of the research process and the accessibility of its outputs are core concerns for the open science movement, specifically as it seeks to ensure the reproducibility of research. The European Open Science Cloud (EOSC)[1] and, globally, funders' increasing requirements for open access to publications and research data[2] exemplify this change.

Consequently, there is a need to help institutions in their effort to provide support for researchers to deposit, store and make available research artefacts, including not only publications but also data and other resources [1]. In particular, we consider here support in the form of software solutions for research repositories, defined as systems enabling researchers to register and access such research artefacts.

There exists a multitude of solutions, software or services, that can contribute to this need for research repositories. Here, we specifically focus on solutions enabling the creation of general-purpose repositories where research artefacts (publications, data, or other) can be deposited, made accessible and reused independently from the research domain. However, in contrast with the pressing need for clear information regarding the advantages and disadvantages of those different solutions, we could not find a systematic overview that discusses their features and the aspects that would enable research performing organisations to select the most appropriate solution.

In this article, our aim is to provide such an overview, supporting decision makers who are in the process of establishing a research repository, and developers who are extending and enhancing the underlying solutions. To achieve that, we establish an assessment framework inspired by the FAIR principles [2]. In order to support a broad range of technical prerequisites, our selection is independent of the deployment mechanism. Therefore, we include on-premise/self-hosted solutions as well as online services in our selection.

The FAIR principles of findability, accessibility, interoperability and reusability for data publication and release are widely accepted as core to enabling transparent research. Therefore, these principles form a good starting point for a structured comparison of existing solutions. However, in order to create a framework comparing solutions for research repositories, a specific set of principles, inspired and redesigned from FAIR, need to be considered: First, we need to reconsider them so to be appropriate for assessing software and service solutions in relation not only to research data, as well as to include aspects of human interactions and the social context in which research artefacts are created and shared[3]. Second, they need to be adapted to the assessment of software and services making such artefacts available especially through a set of concrete metrics attached to each of the principles, enabling a structured assessment of how specific solutions (i.e. research data platforms) contribute to the practical achievement of those.

[1] https://www.eosc.eu/

[2] See for example https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm

[3] The need to include such social features in research repositories has been identified for example in [3].

We therefore adapted the FAIR principles and added three additional principles that relate to the way the solutions should enable engagement, social connections and trust, forming the FAIREST principles. Engagement encompasses usability aspects and interaction mechanisms provided. Social connections span the social context of research artefacts and social connections of researchers. Trust focuses on the reliability of the solution and of the artefacts included. In this article, we define a set of metrics to assess specific solutions with respect to the way the features they provide enable each of the FAIREST principles.

As a way of validation and illustration, but also to provide an overview of the current field, we also performed an assessment of 11 general-purpose solutions, as described below.

We conducted our research in 6 steps in order to cover a broad variety of sources, receive iterative feedback, and sharpen our findings:

1. First, we conducted a literature review to explore existing research on research repositories and related topics. We put a focus on comparative articles, solution overviews, and evaluation frameworks. The result is presented in Section 2.
2. Based on and inspired by these findings we derived a framework for evaluating and comparing features and functionalities of research repositories. The framework consists of a structured set of metrics and corresponding assessment criteria, inspired by, adapting and adding to the FAIR principles, as presented in Section 3 about the details of the FAIREST principles.
3. We applied our framework on a selection of 11 popular solutions for research repositories (see Section 4 for an overview of the solutions), covering both online and on-premise solutions. Those solutions were selected on the basis of being general (i.e. not dedicated to a specific research domain) and active (i.e. currently significantly used by various organisations). We used official documentations, online sources, concrete installations, source code and other literature as basis for the assessment of individual systems.
4. In order to validate our preliminary findings, we requested direct feedback from the developers and/or operators of each solution. Although we did not receive responses for all the solutions, this step helped refining some of the assessments and ensured that the proposed framework could be meaningfully understood and applied outside the co-author group.
5. We finalised our assessments by incorporating the responses and reviews, thus concluding with valid and approved statements. Section 4 includes a comprehensive overview of the results and presents selected highlights.
6. Finally, as presented in Section 5, we analysed our assessments to identify common gaps and recommendations for future developments and enhancements in the domain of research repositories. We conclude our work in Section 6.

2. RELATED WORK

This article focuses on research repositories, their purpose and characteristics and practical implementations. The review in this section relies on the definition of digital research repositories encompassing any repository

to store research publications (articles), research data, and other digital artefacts produced during research processes. It also spans other related systems such as academic social networks and social bookmarking systems. This definition will be narrowed down later, when it comes to assessing specific research repository solutions, to focus only on generic, domain-independent and widely used systems.

2.1 Reviewing Research Repositories

Thanks to the growing popularity of research repositories, some research exists that aims to classify and assess solutions for building research repositories.

Nicholas et al. [4] categorise repositories into institutional, subject and format repositories. Institutional repositories store research artefacts generated by an academic institution. Subject repositories specialise in a specific domain, such as computer science. Format repositories store outputs of a particular type such as electronic theses.

Amorim et al. [5] further introduce a deployment classification as an important aspect, distinguishing between installation packages (on-premise) and services (online).

Besides digital research repositories, other solutions gathered attention in the past, too. This includes preprint servers, such as arXiv[④], which aim to make publications available as early as possible. Furthermore, this also comprises academic social networks [6], including ResearchGate[⑤] and Academia.edu[⑥], whose main purpose is to connect researchers. Finally, there are also social bookmarking systems [7] such as Bibsonomy[⑦], whose aim is to share and link to scientific publications.

These systems mainly target preservation and access within the lifecycle of research artefacts. However, there are other systems targeting other phases of the research process. One of them is the Open Science Framework (OSF)[⑧], a generic tool that spans the whole research data lifecycle.

In terms of uptake, the number of research repositories has been steadily growing over the years. Due to the variety and fragmentation of existing systems, exact numbers can only be estimated by looking into topic- or system-specific repository listings [8].

OpenDOAR[⑨], a curated directory of open access repositories, reported 78 open repositories in 2005 and, as of May 2021, includes 5,663 repositories. Among these repositories, Dataverse is the most popular solution with an adoption by 39% of the repositories in OpenDOAR, followed by EPrints (11%), and WEKO (9%). Approximately 76% of these repositories host written research outputs such as articles, reports, and

④ https://arxiv.org
⑤ https://www.researchgate.net
⑥ https://www.academia.edu
⑦ https://www.bibsonomy.org
⑧ https://osf.io
⑨ https://v2.sherpa.ac.uk/opendoar

book chapters, and the remaining 24% include other research outputs such as bibliographic references, software, and more. Data is the main focus of only ≈ 2.3% of the repositories.

re3data[9] is another registry of research repositories, focusing on data depositing platforms. As of May 2021, it indexes 2,686 research data repositories, where ≈ 80% use unknown or "other" as their underlying software. The majority of the remaining repositories use Dataverse[11] (≈ 23%), followed by DSpace[12] (≈ 21%), MySQL[13] and CKAN[14]. Most of the research artefacts published in these repositories are in scientific and statistical data formats, followed by textual formats, images, and then raw data [9].

Deeper insights can be gained from several surveys and comparisons of research repositories, which target generic repositories, and which mainly focus on the FAIR principles (see Section 2.2 below on FAIR and other principles). Some representative examples will be described in the following.

Andro et al. [10] examined a mix of 10 open and proprietary software solutions, including ePrints, Invenio[15], and DSpace. The authors devised a questionnaire of 160 questions divided in 6 categories, spanning document management, metadata, engine, interoperability, user management, and Web 2.0. The main conclusion of this survey is that the criteria to choose a solution depends on the types of documents to be uploaded (contemporary or old), on the licensing of the software (open or proprietary), or on any other aspect of the 160 questions of the survey. The design of the questions follows a similar methodology to our approach. However, since the survey is close to 10 years old, an update to current needs and developments is required.

Amorim et al. [5] created a compact comparison of six established repository solutions, namely DSpace, CKAN, figshare[16], Zenodo[17], ePrints[18], and the services of EUDAT (B2DROP, B2SAFE, B2SHARE), regarding architectural and metadata characteristics. The authors derive key advantages for each solution and conclude that an extensive requirement analysis is indispensable.

Assante et al. [11] focused on analysing five scientific data repositories, namely 3TU.Datacentrum[19], CSIRO DAP[20], Dryad[21], figshare, and Zenodo. Their analysis focused on formatting, documenting, licensing, publication costs, validation, availability, discoverability and access, and citation. The authors present a thorough discussion of shortcomings and prospects for the future at the time of publishing.

[9] https://www.re3data.org
[11] https://dataverse.org
[12] http://dspace.org
[13] https://www.mysql.com
[14] https://ckan.org
[15] https://inveniosoftware.org/
[16] https://figshare.com
[17] https://zenodo.org
[18] https://www.eprints.org
[19] http://datacentrum.3tu.nl/
[20] https://data.csiro.au/
[21] https://datadryad.org/

Austin et al. [12] assessed 32 online data platforms, including CKAN, Dataverse, figshare, and Zenodo. Their survey covered the categories hardware and infrastructure, description, preservation, privacy and security, archiving, submission, accessing and sharing, collaboration, policy, administration, tabular data, and certification status.

Further surveys and comparisons or research repository solutions can be found in [12, 13, 14, 15]. Most of these overlap in the metrics and features used in their assessments. We conclude that there is a growing need for standard principles to assess research repositories against different requirements of repository managers in an organised way, which incorporates the metrics and features mentioned in the related work.

Besides these evaluations, several publications specifically analyse the usability of research repositories. Joo et al. [16] propose a method to evaluate the usability of digital libraries. Their method is based on the ISO 9241-11 standard, spanning efficiency, effectiveness, and satisfaction, extended with learnability as inspired by Nielsen's usability model [17].

Machova et al. [18] conducted a usability evaluation of governmental data portals and provided a list of best practices for improving the ability to discover, access, and reuse these online information sources.

2.2 Principles Guiding Research Repositories

As already mentioned, a widely recognised and endorsed concept regarding the publication of research data and metadata are the FAIR principles. These principles were originally introduced in 2016 by Wilkinson et al. [2] to improve the feasibility of reusing scholarly data by both machines and humans. The four high-level principles are: Findability, Accessibility, Interoperability, and Reusability, where each principle is refined into three to four sub-guidelines. For instance data and/or metadata needs to be assigned a unique and persistent identifier and a standard protocol is to be used for accessing the data. It is important to note that the FAIR principles are only universal guidelines and do not define practical implementations and technological decisions. In practice, this can cause inconsistent implementations and interoperability issues [19]. However, the principles have been widely adopted as guidance for data publishing and substantiated in various forms.

In addition, the scope of FAIR has been expanded beyond mere research data, e.g. towards research software [20, 21]. A 2017 analysis of the repositories indexed by re3data [9] recommends that, since several repositories started to operate before the publication of the FAIR principles, managers of these repositories look at these principles and update features and policies accordingly.

The FAIRsFAIR Horizon 2020 project[2] addresses the development of procedures, standards, and metrics based on the FAIR principles. Recently, the project team proposed a series of metrics to assess the extent of the FAIRness of a research artefact [22]. However, these metrics focus on adapting the principles to

[2] https://www.fairsfair.eu/

general research objects and are supposed to work in conjunction with the CoreTrustSeal principles (see below). The authors mention two use cases that align with the CoreTrustSeal stakeholders: Using these modified principles for self-assessment and for guiding development. However, they still fail to provide a framework to categorise platforms as a way to facilitate their selection by institutions.

The Research Data Alliance (RDA)[@] is a research community organization whose mission is to build the social and technical bridges to enable open sharing of data. It's working groups and publications are of high relevance for our paper. One highlight is a paper of the FAIR Data Maturity Model Working Group [23], which describes their FAIR data maturity model, spanning a broad set of indicators to assess "FAIRness" of concrete research data. These indicators form a profound consolidation of the FAIR principles, still leaving a lot of space for interpretation.

The European Open Science Cloud (EOSC)[①] is an initiative of the European Commission aiming at developing an infrastructure providing its users with services promoting open science practices. It's services, task forces and publications are also of high relevance for our paper. One highlight is a paper of the European Commission Expert Group [24], which describes the broad range of changes required to "turn FAIR data into reality". This especially includes the establishment of FAIR ecosystems, covering policies, data management plans, identifiers, standards and repositories.

Several stakeholders of the digital repository community developed the TRUST Principles [25]: Transparency, Responsibility, User Focus, Sustainability, and Technology. These principles guide the development and maintenance of digital repositories that are sustainable, reliable, and support comprehensive policies based on community practices. These principles have a strong focus on providing a reliable service for users.

In parallel with these principles, the CoreTrustSeal[@] is a community-based non-profit organization that promotes the trustworthiness of research repositories after a platform fulfils a series of requirements regarding transparency, integrity, security, and privacy of the data. These requirements are grouped in three categories: organizational infrastructure (6 requirements), digital object management (8 requirements) and technology (2 requirements). Applicants use these requirements to perform a self-assessment of their platform that must be accompanied by evidence of compliance with each requirement. The application is then peer-reviewed before being awarded a level of certification. The CoreTrustSeal certification is intended as a first step for the creation of a global framework for repository certification that increases transparency, therefore, building stakeholder confidence by demonstrating, with evidence, that a repository follows good practices. However, the uptake of the certificate is very low[@] and repository providers need to pay an administration fee to be certified.

[@] https://www.rd-alliance.org
[@] https://www.coretrustseal.org/
[@] Out of the 172 repositories listed on the CoreTrustSeal website, less than 20 have an active certification at the time of writing.

Several stakeholders of the digital repository community also came up recently with the CARE principles for indigenous data [26]: collective benefit, authority to control, responsibility, and ethics. The principles are meant to complement the FAIR principles, considering both people and purpose in their advocacy and pursuits.

In summary, research repositories have been under a lot of scrutiny in the last decade with multiple attempts to categorise and compare them. The FAIR principles provide a basis to formalise and standardise such assessments, but they need to be extended, especially with respect to engagement, the social context of research, and the trustability of solutions and artefacts. For this reason, we propose below the FAIREST principles that extend FAIR in this direction, and a set of clearly defined metrics, the FAIREST framework, to assess how much a given solution enables the realisation of each principle.

3. THE FAIREST PRINCIPLES

As described in Section 2.2, the FAIR principles aim to ensure that, from a technical point of view, research data and metadata is "optimised for reuse"[8]. In other words, those principles are seen as requirements in the way research data and its metadata must be represented and made available formally. Our aim here is to assess solutions that make available such data, the publications emanating from research, and possibly other kinds of research artefacts. For this purpose, we first extended the FAIR principles to add engagement, social connections, and trust. Engagement and social connections are inspired by functionalities subsumed as Web 2.0, Science 2.0, or Library 2.0 in the literature [27, 28], and by functionalities specifically provided by academic social networks [29] and academic bookmarking systems [7]. Trust is based on the TRUST principles mentioned in the previous section (2.2).

Those additional principles aim to address the human point of view of research repositories, including how to interact with them, how they account for the social context in which research is carried out, and the important aspect of the trustability of both the system itself and the artefacts it includes. Of course, the principles in themselves are not sufficient. Therefore, we additionally identified a set of metrics through which each solution can be assessed with respect to the way they enable each of the FAIREST principles. These measurable properties were identified, as described in the introduction of this article, through inspecting the literature and existing systems, as well as through interacting with the research repository development community. They constitute the foundation for determining comparable characteristics for individual research repository solutions. They are described, together with the principles to which they relate, in the remainder of this section.

3.1 Findability

In the objective of optimising data for reuse, the first principle borrowed from FAIR is that data first needs to be findable. Findability is seen here in a broad sense, which includes the notions of data discovery as

[8] https://www.go-fair.org/fair-principles/

well as mechanisms that can enable both human users and machines to search for data according to specific needs and criteria. Indeed, as in FAIR, the definition of this principle starts with the idea that, to be findable, research artefacts should be clearly and globally identifiable, and that an identifier should be provided with a mechanism to locate and access the artefact's description or content. Going a step further in the direction of engagement, we consider here also the search functions research repositories can include, which might apply to different aspects of research artefacts and which can enable more or less precise queries.

Table 1 shows the metrics to assess how much a research repository enables findability for the research artefacts it contains. In accordance with the requirements described above, the first metric relates to whether the repository automatically assigns a persistent identifier for all artefacts it contains (FCS1). This metric is assessed as "yes" (y) if the system automatically creates a stable, persistent way to address a research artefact it contains, and as "no" (n) if not. FCS4 adds to this, as another binary yes/no metric, the idea that those identifiers can be dereferenceable, i.e. that using this identifier as a web link will point the user to the location of the artefact in the system. In order to ensure that those identifiers are universally comprehensible and interpretable, some repositories can generate DOIs[2] (Digital Object Identifiers [30]) or other handles, as assessed by FCS2 (yes/no). Other kinds of external identifiers might also be imported by the repository to support findability, as assessed by FCS3 (yes/no).

Table 1. Findability metrics.

Type	Metric	ID	Values	Description
Content support	Persistent identifiers	FCS1	y/n	A system is assessed yes (y) if it automatically assigns a persistent identifier to a research artefact.
	Generates DOIs	FCS2	y/n	A system is assessed yes (y) if, in addition to a possible unique identifier local to the system, it also supports the generation of a DOI for a research artefact.
	External identifiers	FCS3	y/n	A system is assessed yes (y) if it allows the import of idenfiers of the research artefact, external to itself.
	Dereferenceable identifiers	FCS4	y/n	A system is assessed yes (y) if it provides a unique identifier for research artefacts that is dereferenceable to the artefact's location in the system.
Content access	Indexes and searches metadata	FCA1	y/n	A system is assessed yes (y) if it enables search over the metadata of the research artefacts.
	Indexes and searches content	FCA2	y/n	A system is assessed yes (y) if it enables search over the content of the research artefact.
	Advanced search features	FCA3	y/n	A system is assessed yes (y) if it provides advanced search features.

The metrics above are considered within the category of content support as they relate to finding artefacts directly based on their identifiers. In the content access category, we also take into consideration functions

[2] https://doi.org

that allow users to find artefacts without knowing their identifiers. The following three binary (i.e. yes/no) metrics look at different aspects of how repositories might index and enable users to search for artefacts based on criteria relating to the description (metadata) of artefacts (FCA1), to the content of artefacts (FCA2), and/or using advanced filtering and ranking mechanisms (FCA3).

3.2 Accessibility

One of the central aims of research repositories is to give access to research artefacts. Access thereby entails different actors and scopes, especially since users of the system are not only humans but also machines. For human users, access becomes easier through standardized language support. In order to enable access for machines, repositories must provide machine readable interfaces and protocols that apply domain standards. Besides accessibility for actors, repositories also need to consider the dimensions of time and availability and the degree of openness of research artefacts. Constant access to data, offered through a high availability of the service and reliable long-time preservation, are core aspects of open science, to enable new forms of research. Access control at first seems to be contradicting the open idea of research repositories. Nevertheless, many research artefacts are not meant to be open by default.

With regard to metrics for accessibility, Table 2 divides them into functionalities concerning content language and content availability. Language support (ACL1) is measured via a yes/no variable, assessing whether the functionality exists (or not). We also evaluate the availability of protocols and APIs (ACL2) for data exchange with a yes/no metric only, as we do not intend to evaluate the individual protocols. Long-term preservation (ACA1) relates to whether artefacts are processed to be available over a long period of time in the future. There are many ways to achieve long-term preservation, so this aspect is also assessed through a yes/no value, indicating whether any mechanism exists to enable this feature. Availability (ACA2) is measured by high, medium, or low, depending on the up-time of the system being higher than 99.9%, between 99.9% and 99%, or lower than 99%. This metrics is only assessed for online solutions as the availability of instances of on-premise solutions depends upon the characteristics of their deployment. The last variable, access control (ACA3), describes the possibility to restrict access to data. There are three different cases: “Closed access” means that the design of the system makes artefacts private by default or more straightforwardly. “Open” means the opposite, i.e. that the artefacts are, by default or in most cases, publicly accessible. When, in a system, the common case is for access to be enabled through a request to the provider, the value “on request” is used. The variable access control may simultaneously take on multiple values.

Table 2. Accessibility metrics.

Type	Metric	ID	Values	Description
Content language	Language support	ACL1	y/n	A system is assessed as providing language support (y) if its interface is available in more than one language.
	Protocols and APIs supported	ACL2	y/n	A system is assessed as providing protocols and APIs (y) if it makes available at least one protocol and/or API for access to research artefacts.
Content availability	Long-term preservation	ACA1	y/n	A system is assessed yes (y) if it offers a convenient mechanism to perform long-term preservation beyond a simple database backup.
	Availability	ACA2	h/m/l	A system is assed high (h) if the uptime is ≥ 99.9%, it is assessed medium (m) if the uptime is between 99.9% and 99% and low (l) if it is below 99%.
	Access control	ACA3	open, closed, on-request	A system is assessed open, if it makes research artefacts public by default. It is assessed as closed if it is primary designed for closed repositories. It is assessed on-request if it offers an on-request feature.

3.3 Interoperability

Research data repositories should foster interoperability: the ability to easily share data with other systems. Therefore, it is essential that research artefacts are available in formats that are portable, open, and widely supported by other systems and platforms. To further enhance interoperability, certain data federation features are of high value. It is also important that artefacts are clearly identifiable via unique and persistent identifiers (as for findability). The associated metadata should follow commonly accepted metadata specifications, such as Dublin Core[@] or MARC 21[@]. Furthermore, data exchange should be supported by well-defined and established protocols, such as OAI-PMH[@]. The possibility to customise metadata and to (automatically) link other resources are two features that can further improve interoperability. To smoothly integrate with research processes, the repository should support manual and automatic data upload and import, as well as data download and export. In this context, the support for custom submission processes are another beneficial add-on.

Table 3 lists metrics that we used to assess interoperability. We distinguish between two types of metrics here: interoperability in content and interoperability in the interaction with content. The metrics related to interoperability of content include whether standard formats for metadata (IC1) and content (IC2) are applied, the possibility to assign persistent identifiers to the artefacts (IC3), the ability to create custom metadata schemes (IC4), and support for semantic linking between artefacts (IC5). With standards we associate any official, publicly available and established specification. Especially, this includes standards

[@] https://dublincore.org/
[@] https://www.marc21.ca/
[@] https://www.openarchives.org/

published by standardization bodies, such as IANA or W3C. These metrics are assessed through yes/no values, depending on whether the corresponding features are available in a given solution. Regarding content interaction, metadata/data upload and import (ICI1), as well as metadata/data download and export (ICI2) are important criteria. The assessment of those metrics is based on whether those functions are available, and are considered only partial if a user has to go through a web user interface to carry them out (i.e. no API function is available). Finally, the availability of a custom submission processes (ICI3) and data federation features (ICI4) are assessed via yes/no metrics.

Table 3. Interoperability metrics.

Type	Metric	ID	Values	Description
Content	Standard format for metadata	IC1	y/n	A system is assessed yes (y) if it employs a standardised and established format for providing the metadata.
	Standard formats for content	IC2	y/n	A system is assessed yes (y) if it supports the provision of data in common and standard file formats.
	Persistent identifiers	IC3	y/n	A system is assessed yes (y) if it assigns a persistent identifier (e.g. a stable URL) to the research artefacts.
	Custom metadata	IC4	y/n	A system is assessed yes (y) if it allows to customise, extend and limit the metadata schema/format.
	Linking of metadata and content	IC5	y/n	A system is assessed yes (y) if it supports the creation and publication of semantic links between different metadata and/or research artefacts.
Content interaction	Import and upload of metadata and content	ICI1	y/p/n	A system is assessed yes (y) if it supports the import and upload of metadata and data via multiple means, e.g. a web frontend or a standard API. It is assessed partially (p) if it only supports the provision via a web interface.
	Export and download of metadata and content	ICI2	y/p/n	A system is assessed yes (y) if it supports the export and download of metadata and data via multiple means, e.g. a web frontend or a standard API. It is assessed partially (p) if it only supports the download via a web interface.
	Custom submission process	ICI3	y/n	A system is assessed yes (y) it it supports the creation and maintenance of a customised submission process, including fine-grain access control and role assignment.
	Data federation	ICI4	y/n	A system is assessed yes (y) if it provides a built-in mechanism to make (part of) the metadata and data in other repositories (running the same system) available.

3.4 Reusability

Research data repositories should support and encourage the reuse of their hosted artefacts. This is especially true for artefacts other than publications, such as research data or supplementary material. Typical reuse scenarios are the distribution, aggregation, conversion, or enrichment of research data. The creation of derived and adapted work is an inherent element of many research processes. Reusability therefore needs to be supported on the technical and legal level. The application of well-defined and standardised metadata schemes, protocols, and interfaces lowers the barriers for further processing,

especially the integration into third-party tools. Furthermore, the repository solution should support and encourage the use of structured, open and machine-readable formats for research data. This facilitates the provision of data in a raw and immediate manner, attracting additional reuse scenarios. As important as the technical access is assurance about the legal conditions of reuse. Hence, a repository solution should support the provision and publication of comprehensive licence information. This includes the possibility to assign different rights and licences to different components of an artefact and a support for a wide range of established and common licences. Ideally, the users receive assistance in choosing and understanding the licence information. For instance legal wording might be translated into understandable and clear attributes, since explicit communication about reuse conditions fosters reusability.

Table 4 shows the metrics to assess the support for reusability of a repository solution. It is assessed based on three distinct metrics, which cover content depositing, access, and support aspects. First, the support for providing detailed licence information (RCD1) during the depositing process is assessed. A solution can have no support for licence information at all, hence no dedicated metadata property is provided. A research repository can offer basic support for licence information, if it facilitates the provision of licences attached to specific artefacts in any way, either as a free input field or based on a controlled vocabulary. Full support can be achieved if the solution supports a highly standardised and advanced method of providing the information, e.g. by applying a user-friendly provision mechanism and reusing existing licence vocabularies. Secondly, the same applies for accessing the licence information (RCA1). There may be no information at all, at least some basic information (e.g. a link to licence), or highly understandable and structured information is provided. Finally, the support for structured data access (RCS1) is assessed as "yes" if the actual research artefact can be accessed in a structured and machine-readable manner.

Table 4. Reusability metrics.

Type	Metric	ID	Values	Description
Content depositing	Licence support	RCD1	y/l/n	A system is assessed high (h) if it provides a highly usable and easy-to-understand mechanism to select a fitting licence for an artefact. It is assessed limited (l), if it at least provides a customizable controlled list of licences.
Content access	Licence support	RCA1	y/l/n	A system is assessed high (h) if it provides an easy-to-understand and human-readable description of the terms of use. It is assessed limited (l), if it at least provides a link to the applied licence.
Content support	Data	RCS1	y/n	A system is assessed as possessing the feature (y) if it allows users to access the content in a structured and machine-readable manner.

3.5 Engagement

Research artefacts should be provided through systems that are not only accessible, but that are also highly usable and that implement interaction mechanisms fitting and matching the workflow and culture of the research community. Specifically going beyond the FAIR principles, we consider here that research repositories are built for use also by human beings. In other words, to be successful, those repositories

should be engaging. This principle looks primarily at the usability of the repository system from the point of view of the intended user: It should be simple, visible, integrated, and practical. It should, in particular, follow modern usability standards, and minimise the effort required by both the publishers and the consumers of research artefacts in accomplishing their tasks. In addition, those repositories should aim to facilitate the work of their administrators by minimising the effort required for their maintenance.

Table 5 lists metrics used to measure how engagement is enabled by research repositories. Those metrics follow two general categories assessing, on the one hand, the usability of the system, evaluated as high, medium or low through relevant methodologies (EUA2) and through the availability of support and documentation (EUA1). On the other hand, we look through yes/no metrics, at the availability of features that are designed to support engagement. Those include in particular features targeted mostly at publishers of research artefacts through which the user is notified of a potential need for engagement (EES1) and supported through the workflow of publishing research artefacts (EES2). The availability of visualisation (EES3) and analysis (EES4) tools enabling consumers to interact with the content of the published research artefacts is also considered an important aspect in assessing how engaging a repository is.

Table 5. Engagement metrics.

Type	Metric	ID	Values	Description
Usability and ease of use	Support and documentation	EUA1	y/m/l	A system is assessed high (h) if the documentation and support are of high quality and reachability.
	Usability	EUA2	y/m/l	A system is assessed according to this metrics based on usability testing carried out.
Engagement support	Push notifications	EES1	y/n	A system is assessed as possessing the feature (y) if it includes mechanisms to notify users through mobile apps, email, or other mechanisms.
	Publication workflow support	EES2	y/n	A system is assessed as possessing the feature (y) if it provides a way to customise and manage the publication/deposit workflow.
	Visualisation tools	EES3	y/n	A system is assessed as possessing the feature (y) if it includes ways to visualise the content of publications or datasets, at least for some formats.
	Analysis tools	EES4	y/n	A system is assessed as possessing the feature (y) if it includes mechanisms to analyse the data in deposited research artefacts, including basic statistical analysis or more advanced methods.

3.6 Social connections

Research artefacts are not created in isolation. They are the result of a social process involving many different stakeholders, institutions, and collaborations. As such, it is important that they are published, made available, and consumed in a way that also takes into account this social context, not as isolated artefacts. As for engagement, this is an additional principle on top of the FAIR principles that takes into account the human dimension of research repositories and research artefacts. It considers two main categories of

features that a research repository should enable: The ability to reflect the social context in which a research artefact was created, and facilitating the creation of new social connections between researchers based on the available research artefacts.

Table 6 lists metrics used to assess how repositories enable social connections. Consistently with the definition of the principle above, metrics to assess how research repositories enable social engagement consider two main categories: Usability and ease of use, and engagement support. Related to the first dimension, this includes being able to create collections of related artefacts (SRT1), brand those collections or the whole repository (SRT3), and present research artefacts under researcher profiles that reflect their ownership and origin (SRT3).

Table 6. Social connections metrics.

Type	Metric	ID	Values	Description
Reflecting the social context	Theming / branding	SRT1	y/n	A system is assessed as possessing the feature (y) if it enables the publisher or administrator to change the aspect of pages on the system to reflect institutional affiliation.
	Creation of collections	SRT2	y/n	A system is assessed as possessing the feature (y) if it includes ways for users to create, name, and publish arbitrary sets of research artefacts.
	Individual researcher profiles	SRT3	y/n	A system is assessed as possessing the feature (y) if it provides pages for individual researchers, including at least the research artefacts they have authored/published.
Creating new social connections	Like button	SCN1	y/n	A system is assessed as possessing the feature (y) if it gives the ability to users to provide simple positive feedback (likes) on research artefacts.
	Comments	SCN2	y/n	A system is assessed as possessing the feature (y) if it enables users to comment on individual research artefacts.
	Sharing	SCN3	y/n	A system is assessed as possessing the feature (y) if it provides ways to share research artefacts with other users, on the system or other platforms (e.g. social media).
	Following	SCN4	y/n	A system is assessed as possessing the feature (y) if it enables users to follow, and receive updates from, other users, research artefacts, collections, institutions, etc.
	Discussion forums	SCN5	y/n	A system is assessed as possessing the feature (y) if it includes discussion forums for users and/or for communication with/from the administrators.
	Development community	SCN6	h/m/l/c	For a proprietary, closed system, the assessement should be closed (c). For an open system, assessement is based on the frequency of activities in the development community (daily/weekly updates: high, monthly/quaterly updates: medium, less: low).

The second dimension mostly looks at the existence of features borrowed from social networking platforms: liking (SCN1), following (SCN4), sharing (SCN3), commenting (SCN2), and forums (SCN5). While

traditional social networks might enable users to follow only other users, we consider here that being also able to follow an artefact, a collection, a topic, or an institution is desirable.

In addition, a particular metric of interest is the one related to the existence of a development community (SCN6), which is of importance to the administrator of the platform, as this community and its level of activity are indicative of whether support and extensions for the platform are likely to be available. While other metrics for this principle are assessed through yes/no values, this one is assessed based on whether the development (community) is open or closed. If it is open, we consider the level of activities in the development community according to the following guidelines: High if there are at least weekly updates, medium if updates come on a monthly basis, and low if they occur less frequently.

3.7 Trust

Trust in research repositories means that a user can rely on the provided system and information. This is enabled by demonstrating the validity, robustness, and significance of scientific artefacts. These principles are based on good scientific practice as well as the practices associated with open science [31, 32]. Another important aspect is the long-term preservation of data, i.e. that providers and consumers of research artefacts can trust that the relevant data will remain available. This aspect is therefore shared with the accessibility principle.

In addition to those, it is useful to assess how repositories include features that are explicitly designed to ensure the trustability of available artefacts, such as reviewing features, gate keeping, and authentication.

On the system level, mechanisms and techniques for long-term preservation are features that we grouped together under trust, including backup systems that prevent loss of data. The use of open source software or libraries is also part of this, since the openness of a system enables transparency, which goes hand-in-hand with trust. Adoption and the size of the user community can also be used, indirectly, as indicators of the trustability of the system.

Finally, how well personal data is protected in the research repository can play a crucial role in the trust users will have in sharing their information with it. Here, we rely on high level requirements from the European General Data Protection Regulation (GDPR[9]) to assess this aspect, such as whether the system provides 'out-of-the-box' features for personal data portability, deletion, and security.

Table 7 shows the metrics used to assess those different aspects of trust. Most metrics take simple yes/no values that indicate whether a feature is present (yes) or not (no). The metric indicator (TCS1), on the other hand, is more complex. A simple indicator represents simple usage statistics that can be quickly calculated by the application. Advanced indicators represent more sophisticated indicators, such as h-index or Altmetrics.

[9] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN

Table 7. Trust metrics.

Type	Metric	ID	Values	Description
Content	Authentication	TC1	y/n	A system is assessed as possessing the feature (y) if it includes an authentication mechanism for users.
	Gate keeping	TC2	y/n	A system is assessed as possessing the feature (y) if it provides a review functionality during submission by data stewards or other permitted organizational users.
	Review feature	TC3	y/n	A system is assessed as possessing the feature (y) if it includes a functionality for writing reviews on specific research artefacts.
Content support	Indicator	TCS1	n/s/a	A system is assessed as possessing the feature (s) if it records usage statistics and (a) if it provides advanced research indicators like h-Index or AltMetrics.
	Long-term preservation	TCS2	date	Date at which the first available artefact was deposited on the platform.
System	Open Source software and libraries	TS1	y/n	A system is assessed as possessing the feature (y) if the underlying system is open source.
	Uptake	TS2	h/m/l	A system is assessed as having high uptake (h) if it is used by a thousands of active users each month, medium uptake (m) with hundreds of active users, and low uptake (l) with lower numbers of active users.
GDPR	System backup	TG1	y/n	A system is assessed as possessing the feature (y) if it includes automated backups in a constant time interval.
	Right of information	TG2	y/n	A system is assessed as possessing the feature (y) if it provides a function to get all data about one user.
	Data deletion	TG3	y/n	A system is assessed as possessing the feature (y) if it provides a function to delete all data about one user.
	Agreement per data management process	TG4	y/n	A system is assessed as possessing the feature (y) if it allows the user to opt-in, agreeing to single personal data management processes individually.
	Portable and secure data exchange format	TG5	y/n	A system is assessed as possessing the feature (y) if it provides all personal data in an open standard format (e.g. HTML, TXT, PDF).
	Protection against data leaks	TG6	y/n	A system is assessed as possessing the feature (y) if it includes security tests for personal data.

Long-term preservation (TCS2) can also be assessed with yes/no values, but information about the time during which artefacts have been available in a given system gives a slightly richer perspective, complementing the corresponding metrics in the accessibility principle. The metric open source software and libraries (TS1) is also assessed with a binary classification, considering the underlying system and if it is mostly based on open source software. Uptake (TS2) can be hard to assess precisely and is therefore considered on a high/medium/low scale, where high means that the system is used by at least several thousands of users every month. Right of information (TG2) and data deletion (TG3) represent the availability of a function that can output or delete all personal data. For agreement per data management process (TG4), the user should have the option to decide which personal data are processed and how. The online solutions

should offer for example a control panel with checkboxes and on premise solutions need a configuration possibility for this reason. Protection against data leaks (TG6) and portable and secure data exchange format (TG5) are additional important points. All those metrics are assessed with yes/no values, where yes indicates that the repository system provides functions specifically dedicated to handling the corresponding requirement.

4. ASSESSMENT OF SOLUTIONS

In order to evaluate and assess our framework, we applied it to a variety of existing and established solutions for managing research publications and data. The main objective of our selection was to cover a variety of solutions, while restricting the sample to a manageable number of comparable and highly relevant solutions. In selecting a solution to assess, we essentially applied three criteria:

1. Eligible solutions had to be a systems/services specifically dedicated to and used for providing repositories of research artefacts.
2. They had to be a general-purpose solution and not focused on a particular domain or sector.
3. Relevant solutions had to be currently used by a significant number of research organizations and researchers.

Our selection is divided into two categories of solutions: on-premise solutions and online services. The first category includes solutions that can, or must be self-hosted on the administrator's own server. The second category covers solutions that are provided as a service by an operator. These categories are a selection criteria by themselves, but are treated neutrally in our framework.

Below, we provide a short description of the systems assessed, discussing first online services and second the systems that have to be deployed on-premise.

Academia.edu emphasises social network functionalities such as the ability to follow other researchers, as well as "metrics" that are supposedly measuring reputation and impact. In practice, it is more used as a way to publicise one's research, i.e. as a service to deposit papers to make them more accessible and visible.[6] Assessment of Academia.edu was carried out on the free version, but references to the paid version are included where relevant.

arXiv is most commonly used as a platform to publish non-peer reviewed papers. There is no limitation on the content published other than that it has to contribute to a scientific discipline. The arXiv platform is managed by moderators who check and validate the content.[4]

Bibsonomy is an online social bookmarking system created by research groups in Germany, designed to support Web 2.0 research. While the system is generic and enables bookmarking of any kind of web resources, it includes a specific section to share, comment on, and review publications.[7]

figshare is a service for the publication of generic research data, aiming to cater for the needs of institutions, publishers, and individual researchers. figshare for institutions focuses on managing research artefacts and measuring the impact of these outputs. figshare for publishers focuses on the publication of citable supplementary material. Finally, figshare for researchers relies on the public web implementation of figshare, which allows individual researchers to deposit and share their research artefacts.[8] The assessment of figshare was performed over the figshare for institutions solution as this is the most feature-complete version of the platform. However, where relevant we refer to the public version of figshare.

ResearchGate presents itself as a social network for scientists of all disciplines. Familiar social functions, such as those from Facebook or Twitter, have been adapted or newly implemented with a focus on the needs of the academic community. Users can share papers, search for collaborators, or follow specific research interests. One of the core features of ResearchGate is the RG Score, a specific indicator of the impact of a researcher.[9]

Zenodo is an open access repository that focuses on sharing data with the wider community. Datasets can be published with no restriction on the format, i.e. software, papers, measurement series, databases, and other digital artefacts can be published through this service.[7]

CKAN is a data management solution for building on-premise data repositories and is the de-facto standard for publishing public sector datasets, i.e. Open Government Data. It is maintained with support from the Open Knowledge Foundation. A vanilla installation offers basic features to publish, manage, and search for metadata with a organisation-based rights management system. It focuses on metadata, but also includes a data storage feature for binary and tabular data. A broad and vivid extension ecosystem has evolved, enabling use case-specific customisation, but those are not included in our assessments.[10]

Dataverse is an open source research data repository software. It is a web application meant to share, preserve, cite, explore, and analyse research data. A Dataverse repository corresponds to the whole installed platform, which then can host multiple virtual archives called Dataverse collections. Each Dataverse collection contains datasets, and each dataset contains descriptive metadata and data files.[11]

DSpace is an open source repository solution for digital research and educational artefacts published by an organisation or institution. Its data model reflects the structure of research organisations: Communities, collections, and items. The core data schema of an item is based on Dublin Core.[29] DSpace offers extensive support for representing the publishing workflow and its involved actors.[12]

EPrints is an open source repository solution developed at the University of Southampton to support institutions in providing open access services for their publications, with recent extensions to support educational software and research data.[13]

Invenio is an open source software framework which provides tools to implement custom institutional repositories for research data management systems. The main features are the scalability of this solution

together with its components and the long-term preservation option. The Zenodo online service has been built using the Invenio v3 framework.[8]

4.1 Overview of Assessments: FAIREST Principles

In this section, we provide a detailed assessment of each system, as structured by the seven principles of the FAIREST framework. For each principle, we provide a table showing the values for each of the metrics presented in Section 3 and highlights of interesting aspects of the way the systems, on the whole, score with respect to those metrics. This will be complemented in Section 4.2 with highlights from each of the assessed systems.

4.1.1 Findability

An overview of the assessments of each solution with respect to the metrics related to the findability principle is provided in Table 8. All solutions provide some form of persistent identifier for the research artefacts they hold (FCS1) and those identifiers are always used to create dereferenceable web links to the artefact's representation in the system (FCS4). In all but Academia.edu, external identifiers can also be imported (FCS3), which in some cases have to be standard identifiers (e.g. DOIs), and in some cases can be any identifier. A subset of the systems also have the ability to generate DOIs for the research artefacts they hold (FCS2).

Table 8. Overview of assessments of solutions for findability. In this table, as well as in the six following ones, the double line separates systems operating as online services, from systems that are deployed on-premise.

	FCS1	FCS2	FCS3	FCS4	FCA1	FCA2	FCA3
Academia.edu	y	n	y	y	y	n	n
arXiv	y	n	y	y	y	n	y
Bibsonomy	y	n	y	y	y	y	y
figshare	y	y	n	y	y	n	y
ResearchGate	y	y	y	y	y	n	y
Zenodo	y	y	y	y	y	n	y
CKAN	y	n	y	y	y	n	y
Dataverse	y	y	y	y	y	n	y
DSpace	y	y	y	y	y	y	y
EPrints	y	n	y	y	y	y	n
Invenio	y	y	y	y	y	n	y

All solutions provide a search feature that is based on the metadata of research artefacts (FCA1). In most cases (all but EPrints and Academia.edu), this search feature can be considered 'advanced' as it enables the use of optional parameters and filters (FCA3). Only three solutions (EPrints, DSpace, and Bibsonomy) enable searching in the content of the research artefacts under specific conditions and configurations (FCA2).

4.1.2 Accessibility

The overview of the assessments of solutions with respect to metrics related to the accessibility principle is provided in Table 9. Here, the implementations differ greatly: While some solutions (e.g. DSpace) appear to have been developed to be highly accessible, others (e.g. Academia.edu) appear less concerned with these aspects. Interestingly, only a few solutions provide interfaces in other languages than English out-of-the-box (ACL1). Only few solutions do not provide any form of API or programmatic access protocols (ACL2). As the assessment of ACA3 makes clear, there are mostly two types of systems: The ones that are meant to support open access, and the ones with a lesser focus on this aspect. This difference, in combination with whether the solution is deployed as a cloud service or as a 'commercially supported' on-premise solution, partially explains the variable support for long-term preservation (ACA1) and lack of information about system availability (ACA2).

Table 9. Overview of assessments of solutions for accessibility. Here and in subsequent tables, — means that the information required to assess the metric is not available, while n/a means that the metric does not apply.

	ACL1	ACL2	ACA1	ACA2	ACA3
Academia.edu	n	n	-	-	open
arXiv	n	y	n	h	open
Bibsonomy	y	y	n	m	open
figshare	n	y	n	h	open
ResearchGate	n	n	n	-	open, on-request
Zenodo	n	y	y	m	open, closed, on-request
CKAN	y	y	n	n/a	open
Dataverse	y	y	n	n/a	open, closed
DSpace	y	y	y	n/a	open
EPrints	n	y	y	n/a	open, closed, on-request
Invenio	y	y	y	n/a	open, closed

4.1.3 Interoperability

The overview of the assessment of solutions with respect to metrics related to interoperability is provided in Table 10. All solutions support standard formats for the content (IC2) and persistent identifiers (IC3). Only half of the systems employ a standard and established format for the metadata (IC1), most of them being on-premise solutions. The ability to customize metadata (IC4) is also a feature provided by all on-premise solutions, while only the online solutions EPrints and figshare offer this functionality. The linking of metadata and content (IC5) is available in a few solutions with very different characteristics. Almost all solutions support the import of metadata and data (ICI1), whereas only Academia.edu offers a web interface but no API. The download of metadata and data is possible with almost all solutions via frontend and API (ICI2). Only ResearchGate and Academia.edu do not offer an API. A rare feature is the support for a custom submission process (ICI3), which is only offered by EPrints, figshare, and DSpace. Furthermore, only three

solutions allow the setup of federated repositories (ICI4): DSpace, Dataverse, and CKAN, although, the implementations are limited to harvesting mechanisms.

Table 10. Overview of assessment of solutions for interoperability.

	IC1	IC2	IC3	IC4	IC5	ICI1	ICI2	ICI3	ICI4
Academia.edu	n	y	y	n	y	p	p	n	n
arXiv	n	y	y	n	n	y	y	n	n
Bibsonomy	n	y	y	n	y	y	y	n	n
figshare	y	y	y	y	n	y	y	y	n
ResearchGate	n	y	y	n	y	y	p	n	n
Zenodo	y	y	y	n	n	y	y	n	n
CKAN	y	y	y	y	n	y	y	n	y
Dataverse	y	y	y	y	y	y	y	n	y
DSpace	y	y	y	y	n	y	y	y	y
EPrints	n	y	y	y	n	y	y	y	n
Invenio	y	y	y	y	y	y	y	n	n

4.1.4 Reusability

The overview of the assessments of solutions with respect to metrics related to reusability is provided in Table 11. The solutions tend to cover reusability in different ways with a focus on different aspects. For example, while most systems have some form of description of the licences attached to research artefacts, very few appear to focus on making the aspect of licensing prominent through clear and well documented descriptions of the available options (RCD1), and even fewer support users in understanding the impact of licences on the consumption of research artefacts (RCA1).

Table 11. Overview of assessments of solutions for reusability.

	RCD1	RCA1	RCS1
Academia.edu	n	n	n
arXiv	l	l	n
Bibsonomy	n	n	n
figshare	y	l	n
ResearchGate	l	l	y
Zenodo	y	l	n
CKAN	l	l	y
Dataverse	l	l	n
DSpace	l	l	n
EPrints	n	n	y
Invenio	l	l	n

RCS1 relates to whether the content of research artefacts is accessible in a programmatic way. As such, it relates to similar metrics to the ones mentioned in relation to the accessibility and interoperability principles. However, while APIs and access protocols might be available to access the metadata of the research artefacts, only few systems really enhance the reusability of research artefact by providing machine-readable access to content.

4.1.5 Engagement

The overview of the assessments of solutions with respect to metrics addressing engagement is provided in Table 12. As visible from this table, the aspects of usability and ease of use are, at best, hard to assess for most solutions. Indeed, while most solutions provide at least some reasonable documentation (EUA1), the large majority of systems does not provide information about usability tests or other forms of assessments of usability (EUA2).

Table 12. Overview of assessments of solutions for engagement.

	EUA1	EUA2	EES1	EES2	EES3	EES4
Academia.edu	h	-	y	n	y	n
arXiv	m	l	n	n	n	n
Bibsonomy	m	-	n	n	n	n
figshare	h	-	n	y	y	n
ResearchGate	h	m	y	n	y	n
Zenodo	h	-	n	n	y	n
CKAN	h	n	n	n	y	n
Dataverse	h	n	y	y	y	n
DSpace	h	n	n	y	y	n
EPrints	m	-	y	y	n	n
Invenio	h	-	n	n	n	n

Regarding supporting engagement, the availability of analytics tools (EES4) is absent from all the solutions. Other metrics are addressed differently by solutions. Push notifications (EES1) for example tend to be more present in online solutions that also include social connection functions (see next section), while the possibility to customise the depositing workflow (EES2) is present only in a few systems for which the ability to curate the content of the repository is particularly important. About half of the solutions allow to visualise the content of artefacts without the need to download them and use external tools (EES3).

4.1.6 Social Connections

The overview of the assessments of solution with respect to metrics addressing social connections is provided in Table 13. With the exception of arXiv and Zenodo, all assessed systems provide ways to reflect the social context in which an artefact has been created. Most on premise and few online solutions enable the customisation of system branding and themes, e.g. to clearly identify the institution (SRT1). About half

of the systems enable creating dedicated collections that can correspond to people, projects, groups, or broader organisations (SRT1). Surprisingly, only few systems create dedicated pages for researchers who are authors of deposited artefacts (SRT3). Those are the systems that tend to put more emphasis on social features.

Table 13. Overview of assessments of solutions for social connections.

	SRT1	SRT2	SRT3	SCN1	SCN2	SCN3	SCN4	SCN5	SCN6
Academia.edu	n	y	n	n	y	y	y	n	c
arXiv	n	n	n	n	n	y	n	n	c
Bibsonomy	n	y	y	n	y	y	y	n	m
figshare	y	y	y	n	n	y	y	n	c
ResearchGate	n	y	y	y	y	y	y	y	c
Zenodo	n	n	n	n	n	n	n	n	h
CKAN	y	y	n	n	n	y	y	n	h
Dataverse	y	y	n	n	n	n	n	n	h
DSpace	y	y	n	n	n	n	n	n	h
EPrints	y	n	y	n	n	n	n	n	h
Invenio	y	y	n	n	n	n	n	n	h

As mentioned above, some of the systems assessed put a strong emphasis on features that enable social connections, by providing functions such as the ability to comment on artefacts (SCN2), share artefacts with others (e.g. through social media — SCN3), or follow (artefacts, people, or collections — SCN4). ResearchGate is noticeable for providing all of those three features and also for being the only one including discussion forums (SCN5). Surprisingly, none of the systems provide explicitly a 'like button' for their artefacts (SCN1), common on social media platforms, even though ResearchGate includes a 'recommend' feature which can be seen as similar.

SCN6 is particular here as it considers social connections specifically with the development community for the system. For proprietary systems, access to the developers of the platform is naturally closed, but almost all open source systems are assessed to have highly active and reachable development communities at the time of writing.

4.1.7 Trust

The overview of the assessments of solutions with respect to metrics addressing trust is provided in Table 14. All systems provide an authentication mechanism (TC1). Interestingly, only one online and three on-premise solutions provide a gate keeping feature to implement an internal, organisational review process (TC2). Only Bibsonomy and Dataverse offer a feature, allowing users to provide reviews to research artefacts (TC3). This is surprising, since (peer) review processes are very common in the scientific domain. Regarding scientific indicators, only ResearchGate provides an advanced research indicator for their users, while other solutions only provide basic usage statistics. arXiv, Bibsonomy, and DSpace do not offer any

statistics (TCS1). Naturally, the long-term preservation metric can only be applied to online solutions and correlates with the first availability of the respective solutions. arXiv offers its services the longest, since 1991 (TCS2).

Table 14. Overview of assessments of solutions for trust.

	TC1	TC2	TC3	TCS1	TCS2	TS1	TS2	TG1	TG2	TG3	TG4	TG5	TG6
Academia.edu	y	n	n	s	2009	n	h	-	n	y	n	n	n
arXiv	y	n	n	n	1991	-	h	n	n	n	n	n	n
Bibsonomy	y	n	y	n	2005	y	-	y	n	y	n	n	n
figshare	y	y	n	s	2012	n	h	y	n	y	n	n	n
ResearchGate	y	n	n	a	2008	n	h	-	y	y	n	n	n
Zenodo	y	n	n	s	2016	y	-	y	n	n	n	n	n
CKAN	y	n	n	s	n/a	y	n/a	n	n	n	n	n	n
Dataverse	y	y	y	s	n/a	y	n/a	n	n	n	n	n	n
DSpace	y	y	n	n	n/a	y	n/a	y	n	n	n	n	n
EPrints	y	y	n	s	n/a	y	n/a	n	n	n	n	n	n
Invenio	y	n	n	s	n/a	y	n/a	y	n	n	n	n	n

Regarding the use of open source software, it is very positive, that all on-premise solutions are available as open source. In addition, the online solutions Zenodo and Bibsonomy are also based on open source software, where we could not determine it for arXiv (TS1). The number of active users is only relevant to online solutions. Here, most solutions have a high uptake with thousands of users each month (TS2). For Bibsonomy, we were not able to find usage statistics.

The assessment regarding GDPR compliance is mostly negative accross all solutions. Only two on-premise solutions offer comprehensive backup mechanisms (DSpace and Invenio) and for the online solutions only Bibsonomy, figshare and Zenodo offer transparent information (TG1). Of all solutions only ResearchGate offers the functionality to retrieve all data about one user (TG2). Regarding the deletion of user-related data (right to be forgotten) the online solutions are leading, since only arXiv and Zenodo do not offer this features. Unfortunately, no on-premise solution implements this important and arguably required feature (TG3). Finally, no solution allows users to opt-in to single personal data management processes individually (TG4) or offers a features to export all personal data (TG5). In addition, we could not find any public information regarding measures to protect against data leaks for any solution (TG6).

4.2 Overview of Assessments: Solutions

Taking an orthogonal view to the previous section, we now highlight interesting aspects of the assessment of each system. As mentioned previously, those assessments have been realised by using the systems and inspecting relevant documentation as well as by requesting feedback from the developers of each platform. We received acknowledgements for our request from Academia.edu, arXiv, Zenodo, and Invenio, and detailed feedback on our framework and the assessments for Bibsonomy, figshare, ResearchGate, CKAN, and DSpace.

4.2.1 Academia.edu

As a commercial "academic social network", the system puts special emphasis on engagement and social connections of the FAIREST principles. Besides the free version evaluated in this paper, a paid version exists, providing some additional functionalities.

For findability and interoperability, the system is comparable to other online solutions, especially to ResearchGate. Some additional functionalities are available in the paid version, including advanced search functionalities for metadata and data.

Academia.edu does not fulfil any of the criteria we formulated for accessibility and reusability. Although we could not find any information about it, and hence rated it as "no", we assume that long-term preservation and availability are ensured by this commercial system.

The system scores well in the categories engagement and social connections. For instance, it provides a mechanism for email notifications. Some additional functionalities are available in the paid version, e.g. researcher profiles. It further provides some analysis tools, including the translation and summarisation of documents.

For trust, the system again does not score well. For most indicators we could not find information whether or not this functionality is available.

In summary, the system clearly focuses on engagement and social connections, and is rather limited in its functionalities when it comes to the other categories. However, in its paid version it provides some additional features, and can serve as a reliable academic social network.

4.2.2 arXiv

As one of the most considerable preprint servers in the field of science and technology, this solution offers an uncomplicated way to publish one's research. The focus lies on publications and less on other research data, but the artefacts do not receive a DOI. Instead, identifiers specific to arXiv are assigned together with a URL. Furthermore, users can also add their own DOI.

As it is provided by the US-based Cornell University, the interface is only available in English and offers no support for other languages. Overall, the range of functions is reduced to the essentials compared to other online solutions. However, the most important file formats and API protocols are supported. It is not possible to enter additional (meta) information.

An advantage is the long-term continuity of arXiv. The platform has been available, relatively unchanged, since 1991, managing a significant number of publications. As a result, however, functions related to social interactions and design improvements have been slow to appear in arXiv. A detailed view of artefacts is also not possible: Available publications can only be published, downloaded, or distributed via sharing buttons.

Even without a review feature, the service's scientific domains are managed by moderators in their structure and content (with 13,000 to 18,000 publications per month). Information about GDPR was not found on the website.

In summary, while limited in features, because of its adoption and stability, arXiv is a good solution for someone looking to ensure that artefacts are and will remain accessible through a platform that is known by researchers.

4.2.3 Bibsonomy

Built as a semantic web system, Bibsonomy tends to score high in the first three of the FAIREST principles. Indeed, the creation of unique, resolvable identifiers, the availability of APIs, and the possibility to import/export from various sources and formats are part of the design of the system. In addition, because it is meant as a social web system, Bibsonomy tends to provide many of the features considered through the social connection principle.

Since Bibsonomy focuses on the open sharing of web resources, the aspect of reusability is not well addressed, in particular with no available support for specifying the rights and licences applying to shared items. In addition, as seen from the metrics related to engagement, the system does not provide an advanced level of usability or functions one might expect from a commercial system.

On trust, while Bibsonomy is comparable to other online systems with respect to that principle, one of the few systems that provides a review function through a five star system. Additionally, the platform on which it is based is open source.

Furthermore, while we considered Bibsonomy as a shared online system, the platform on which it relies can be re-deployed to provide a repository for a specific institution, with the possibility to customise it in that case[29].

In summary, Bibsonomy can be seen as an experiment towards academic social networks. It represents a valid solution when the aspects of social interaction and the ability to share with others, beyond simple accessibility, are important.

4.2.4 figshare

figshare strongly focuses on giving users tools to enable the publication of FAIR data [33]. In terms of findability, figshare implements all assessed features, except for content search. Contrary to the public online version, figshare for institutions supports the migration of content from preexisting or legacy repositories, including handles or unique identifiers. A function to allow users to choose not to mint a new

[29] see for example https://puma.uni-kassel.de/

object identifier but use a pre-existing one was scheduled to go live in 2021. Furthermore, the ability for full-text indexing of publications was expected in 2021.

figshare accepts all file formats and offers several data viewers[83] with a well-documented API for accessing and publishing data. figshare for institutions provides tools for customizing the submission process[84] and figshare for publishers includes flexible submission workflows[85].

Despite being not strongly focused on Social Connections, figshare has some social features such as the creation of individual profiles that can be followed by other users. Generally, figshare does not provide discussion forums for users but figshare for institutions allows communication between administrators and submitting authors during the review process. Finally, figshare is a closed source software platform. However, the developers provide a detailed roadmap and a feature request forum with comments.

In summary, figshare represents a valid choice for institutions looking for a well established platform providing solid foundations towards enabling the FAIREST principles broadly.

4.2.5 ResearchGate

ResearchGate is more a social network for researchers than a repository. It offers many features to establish social connections, such as comments, sharing, following, and a type of discussion forum. Individual researchers have rich profiles and the connection to other researchers is endorsed and supported with many (automatic) recommendation functionalities.

Yet, ResearchGate offers many features to manage and publish research artefacts, e.g. publications, presentations and raw data. It also offers free DOI assignment and integrates existing DOIs. References and links are automatically extracted from publications and displayed. The metadata of all artefacts can be searched and filtered, while the content itself is not indexed. ResearchGate does not follow any established metadata standard and does not offer an API. However, metadata can be imported based on the OAI standard and citations can be exported.

The interface is only available in English. The entire submission process is highly automated, with little space for customisation. For instance licences are automatically assigned and cannot be set manually. ResearchGate is highly interactive and social as it offers push notifications, preview features for many file formats, individual profiles, comment functions, and sharing/follow features. The access control mechanism for artefacts is notable, e.g. because it allows users to only offer private content to other researchers upon request.

[83] https://drive.google.com/file/d/11N1D0e7b36SbeysmZeYc7-vP9qQyUIUx/view (Accessed in July 2021)
[84] https://support.figshare.com/support/solutions/articles/6000225218-reviewing-items (Accessed in July 2021)
[85] https://knowledge.figshare.com/publisher/workflows (Accessed in July 2021)

The RG Score offers an interesting tool to monitor the individual impact of artefacts (mostly for publications), but since its calculation is limited to the content on ResearchGate, its validity is questionable. In addition, the handling of personal data and the compliance with the GDPR is not clear in every aspect.

In summary, ResearchGate is an interesting platform to interact with other researchers and broadly disseminate artefacts. However, it is highly proprietary and promotes a strong vendor lock-in.

4.2.6 Zenodo

Zenodo is committed to enabling the publication of FAIR data[16] and scores well in the assessment of these principles, but has almost no features to enable Engagement, Social Connections or Trust.

In particular, Zenodo is assessed highly on metrics that focus on aspects already included in FAIR, including for example providing persistent, external identifiers for artefacts (findability), or using open licences (reuse). There is also a focus on the simplicity of the publication workflow and on ensuring programmatic access through a well documented and comprehensive API.

A unique feature in Zenodo is the long-term preservation policy, that guarantees availability for at least 20 years and, presumably, a migration to alternative repositories in case of a shutdown[17]. Since Zenodo is built upon the open source solution Invenio, it has an active development community and accepts contributions from community members[18].

In summary, Zenodo should be considered for managing a collaborative and public repository for publications and research data within consortia, e.g. cross-organisational research projects. The close connection to Invenio and the open development indicate transparency and durability.

4.2.7 CKAN

CKAN is not primarily intended as a research data repository solution, but it is capable to operate as one with some limitations. Its advantage is the relatively simple design and emphasis on openness. CKAN offers a powerful search index and filter features. A special data store allows to index and query tabular data very efficiently. The built-in federation feature via harvesting allows to create a network of harmonised repositories.

CKAN targets the publication of Open Government Data and has established standards in this domain. Those are, however, not always in line with the research data domain. For instance CKAN has its own metadata schema and API specification. However, it can be extended and the common DCAT format is already included. In addition, the API is very comprehensive and allows a complete interaction with the repository. The CKAN frontend is also already available in numerous languages.

[16] https://about.zenodo.org/principles/
[17] https://about.zenodo.org/policies/
[18] https://github.com/zenodo/zenodo

Notable are the integrated data preview and visualisation feature, the extensive theming module, and the well-documented extension interfaces. The latter allows users to customize CKAN individually.

CKAN does not support the generation of DOIs and has a fixed submission process, that is not suited for creating closed repositories. It also does not offer any relevant social features, such as user profiles or review functionalities. CKAN does not provide any mechanism for long-term preservation.

Despite this, CKAN can act as a foundation for creating highly customised research data repositories. Many missing features are available as community extensions, some requiring additional custom development work. The core development is active and it can be expected that CKAN will be supported and updated for the foreseeable future.

4.2.8 Dataverse

The system performs well in terms of findability, accessibility, and interoperability. It performs weaker in the reusability category, and also in the remaining categories.

In the findability category, the only feature that is missing is content search, which is only rarely available.

When looking at accessibility, Dataverse is one of just a few systems that provide different access levels, ranging from open to closed, with the option of granting access to certain user groups. Availability and long-term preservation depend on the running instance. Both aspects are left to the system administrator.

Dataverse is also very powerful in terms of interoperability. It comes with a powerful API, and provides a multitude of functionalities, including the support for different metadata standards, linking of artefacts from outside the repository, and the possibility to use it for metadata harvesting. The only feature that is missing is a customisable submission process.

Despite this, its portfolio in terms of reusability is limited. Nevertheless, the systems proves to be very flexible, e.g. leaving content support to external tools.

The system also does not perform very well in the social connections category, but is comparable to other on-premise solutions. Despite this, the system is well documented and provides several features in the engagement category.

The trust category shows a mixed picture. On the one hand, it offers a variety of features, including a review feature, which is only available in two of the reviewed systems. On the other hand, we could not find any information about GDPR compliance. Only system backups are possible, which needs to be handled by the system administrator.

In summary, Dataverse appears to be a valid solution for institutions looking for a robust approach towards providing their users with flexibility both with respect to the consumption of (i.e. access to) research artefacts, and with respect to the publication (i.e. depositing) process.

4.2.9 DSpace

Many universities, libraries, and similar organisations use DSpace, since its data model is based on the organisational structure of academic institutions. This is achieved by linking the role system with the gate keeping process. As soon as a user wants to create a publication, the responsible persons of the organisation are automatically included in the publishing process. The entire process can be highly customised. DSpace supports the ability to set an embargo so that an artefact is only made available after a particular time.

DSpace relies on Dublin Core as its core metadata standard, including options to extend and adapt it to the organisational requirements or scientific domain via configuration. However, the metadata structure of Dspace remains very flat, allowing easy editing of properties, but preventing the creation of a more complex and customised data structure. DSpace emphasizes import and export features of artefacts and their metadata, including a large variety of feature-rich API standards and protocols. Long-term preservation is also supported, allowing the creation of backups of the latest version of all artefacts. Currently, DSpace's development community is implementing GDPR-related features, allowing to view and remove person-related data.

DSpace is a proven solution to build on-premise institutional repositories for research publications and data. It is highly customisable and one of few solutions allowing to adapt the submission process to the organisational realities.

4.2.10 EPrints

Since EPrints originates from a research group specifically looking at open access research outputs and data, it tends to score well in the first three of the FAIREST principles. It provides a platform meant to enable high accessibility and interoperability, especially through the use of machine readable formats and APIs, and provides common functions to enable findability, as well as enabling content search. Regarding reusability, EPrints mostly focuses on publications, for which it supports common formats, but surprisingly does not allow to specify a licence for the publications deposited.

EPrints does not provide many of the functions and capabilities other systems might include as part of the engagement and social connection principles.

Since it is meant to provide a trusted repository for institutions, many of the metrics considered in the trust principle are positively assessed for EPrints. It is useful to mention in particular that EPrints is an open source system, and one of few that enable a customizable depositing process to support gatekeeping. As an on-premise system, however, it expects the operations related to GDPR compliance to be implemented by the local administrator rather than providing those functions already embedded in the system.

In summary, EPrints appears to provide a valuable solution for institutions wanting to focus on trust, both because of the reputation of a system originating from academia, but also because of some of the features it focuses on (e.g. the customisable depositing process).

4.2.11 Invenio

Invenio is geared towards the implementation of large-scale digital repositories and features that are not available in the base framework can be added by extending the publicly available source code. For example, Zenodo is an implementation of Invenio, showcasing a potential application of this framework.

In terms of the FAIREST principles, Invenio tends to be assessed positively for metrics related to the findability, accessibility, and interoperability principles, due to a design that favours compliance with FAIR: The use of persistent identifiers, of interoperable standards, of comprehensive APIs and others.

As a counter point to this, the aspects related to user experience tend to be assessed less positively, since Invenio was not designed to focus on those aspects. It especially makes no claim towards enabling social network-like features and therefore scores poorly in terms of engagement and social connections.

It is worth mentioning that Invenio has recently (August 2021) launched a long-term support version of InvenioRDM[39] which is a tool built on top of the Invenio framework together with Zenodo. This tool promises to enable anyone (e.g. institutions) to run a complete and customisable service similar to Zenodo.

In summary, Invenio appears to be a valid choice for institutions looking for a robust solution to enable the FAIR principles, just like Zenodo, but that is to be deployed on-premise.

5. ANALYSIS & RECOMMENDATIONS

A core objective of this article is to provide a framework, through the FAIREST principles, enabling users and, in particular, research institutions, to select a solution that best meets their requirements. Indeed, besides making a choice between deploying on-premise solutions and using online services, the framework and the individual assessments presented above can be used to make a selection, first, based on which part of the FAIREST principles is given higher priority. In a second phase, specific metrics can be considered to further refine this selection.

To make this principle-based approach to selecting a solution more concrete, Figure 1 shows a heatmap of aggregated scores for each principle and each solution. A cell represents a score for a solution in a principle, based on mapping the values of assessments to numerical values (e.g. yes is 1 and no is –1), and normalising the sum of those scores using the minimum and maximum possible scores for each principle[40].

[39] https://inveniosoftware.org/products/rdm/

[40] spreadsheet templates in the supplementary material contains the mapping of value to scores and the normalisation formula.

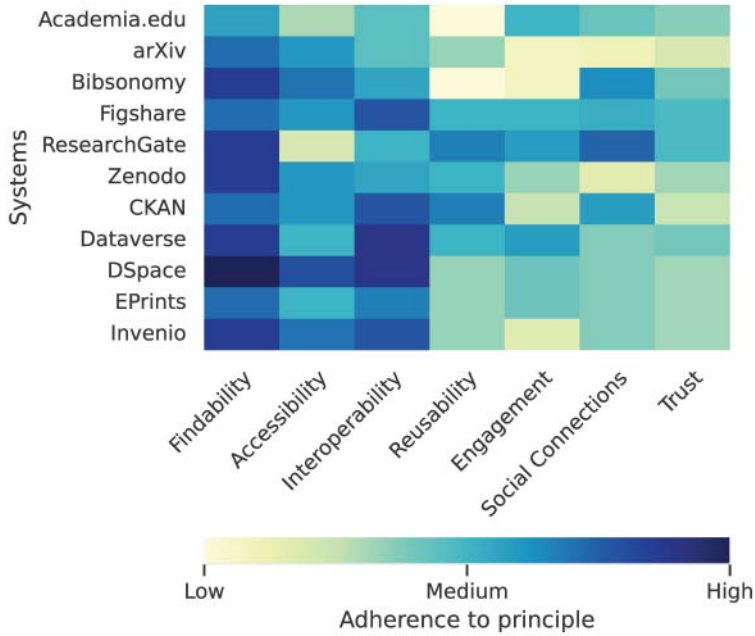

Figure 1. Level of adherence to the FAIREST principles.

As can be seen, this gives an overall picture of the degree to which each solution addresses the FAIREST principles and acts as a decision support tool for the most appropriate solution. For example, an institution that puts a strong focus on building a highly available infrastructure for their research artefact might choose DSpace since it scores comparatively high on the findability, accessibility, and interoperability principles. Another institution keen on ensuring that social connections are well represented might turn to ResearchGate instead. In a third example, an institution that is looking for a solution that is reasonably robust across all principles, might want to select figshare.

Naturally, this can only be an initial selection based on a broad overview and more specific requirements might be taken into account. Our core contribution is that the framework of metrics based on the FAIREST principles is highly applicable to a wider variety of research repository solutions. The template scoring sheets and existing assessments are openly available[6]. Any new solution assessed is then easily comparable to the ones already addressed here.

Interestingly, besides supporting the selection of a solution, the assessments in this form also give some insight into the level of development of existing solutions. Even though the metrics tend to be based on existing features in the assessed solutions, we can find areas that are less supported by these solutions. At a high level, it appears striking in Figure 1 how the first three principles of the FAIREST framework are

[6] https://doi.org/10.5281/zenodo.5282929

significantly better addressed than the rest. Trust appears especially badly covered, and, even though some solutions put some emphasis on social connections, the human element of research repositories (social connection, engagement, trust) appears not to be well developed. Below, we discuss in particular the aspects of rights management, data reuse/consumption, and usability which appear as areas where large gaps currently exist in the assessed systems.

In more detail, on rights management, due to the GDPR in the European Union, we would have expected to see research repositories including more functions related to managing requirements associated to data protection. While, even without the corresponding functions, assessed solutions can be GDPR-compliant, the assessments in the trust principle show that they tend not to include mechanisms that would facilitate enabling such compliance for the administrators and users of the platform. This lack of support for rights management is also visible in the reusability principle, where most systems score poorly in relation to enabling producers to assign a licence to their artefacts, and for users to understand the rights included in those licences for consumption of the artefacts. While this might be based on a naive view that open research implies open licences for artefacts, this could actually create a barrier, since even in open research, some artefacts require some level of legal protection.

Another aspect that could be seen as surprising is the lack of focus on the consumption of artefacts in the assessed solutions. Considering that this is an important function found in domain-specific (two times) repositories, the fact for example that none of the assessed repositories includes functions to analyse data directly on the platform, and only a few of them enable visualisations, could be seen as unexpected. That we only assessed generic (i.e. non-domain specific) solutions can possibly explain this finding, since those functions are much harder to provide for generic content. We believe that a stronger focus on the consumption side of research artefacts is essential for open research to become a global reality.

Not unrelated to the point above, the aspect of usability of the research repository solutions is something that we were not able to assess clearly. Indeed, almost none of the solutions appear to make available the results of usability tests, and it is unclear whether such tests are actually carried out. However, it is expected that, when choosing a solution, an institution or research group might put some emphasis on such aspects. It would therefore, for the development of the field as a whole, be a positive step to systematically carry out and publish usability tests.

Finally, it is worth noting that, while our definition of research repositories is broad and could include any type of artefact, the assessed solutions only focus on two types: publications and data. As shown for example in [34, 20, 35], the application of the FAIR principles can be extended to different types of research resources, including software, and some domain-specific repositories include a larger variety of artefacts. To this can be added that many of these research resources are often shared on repositories that are not specifically dedicated to research (such as Github© for example).

© https://github.com

6. CONCLUSION AND OUTLOOK

In this article, we introduced the FAIREST principles and proposed a framework, based on these principles, to assess research repositories and support institutions in selecting a solution meeting their requirements. The FAIREST principles are inspired from, adapt and add to the FAIR principles including aspects that focus more on human interaction, the social context of research, as well as the trust users can place in the repository and the artefacts they contain, than on aspects purely related to data exchange. For each of the principles, we included a set of metrics with clear guidelines on assessing them, and used those to assess 11 domain-independent and actively used solutions.

While the objective of those assessments is to provide a view of the strengths and weaknesses of each solution for the purpose of selecting one, it also provides us with an overview of the current state of development of the considered research repository solutions. It showed in particular that some aspects, such as usability, data reuse, and data rights, were not well addressed by the solutions. More focus in future development of the systems might be put on improving those aspects to better enable in particular engagement and trust.

We chose to focus our exploration of existing solutions on domain-independent, actively developed and used systems. However, the framework itself as a set of metrics and a scoring mechanism using those metrics is straightforwardly applicable to other research repositories. It would, in particular, be interesting to compare the results obtained here with assessments of domain-specific repositories, to see if similar conclusions can be drawn and what the main differences are. Even if our framework was designed to enable such use, additional elements to assess, in the form of new metrics, might also be uncovered in this way.

Finally, when designing the FAIREST framework, our assumed target user was someone in a position to decide on the solution to use or deploy within a university, a research institute or another similar organisation. The end-users of the solution, such as researchers, were considered implicitly included since they would be taken into consideration by the person making the decision. However, researchers looking to decide on which solution to use to make available their research artefacts might not focus on the same criteria as what we have established here, and our framework might need to be adapted to support this scenario.

Another category of users of the FAIREST framework are the developers of research repository solutions. While those might find that some additional metrics are required to cover new features they are developing[©], we hope and expect that developers of new and existing solutions will use the FAIREST framework as a guide to evolve their platform and address some of the gaps identified.

[©] For example, ResearchGate having developed a mobile application for their platform, they might consider mobile availability a relevant metric.

ACKNOWLEDGEMENTS

We thank Hannes Wünsche for many fruitful discussions, and for contributing to the development of the framework. We also thank Leonard Mack for critically reviewing the paper.

This research was supported by the Fundação para a Ciência e a Tecnologia through the LASIGE Research Unit, UIDB/00408/2020 and UIDP/00408/2020.

This work was supported by the Federal Ministry of Education and Research of Germany (BMBF) under grant no. 16DII128 ("Deutsches Internet-Institut").

AUTHOR CONTRIBUTION STATEMENT

Mathieu d'Aquin, Fabian Kirstein, Sonja Schimmler and Sebastian Urbanek proposed the research problem. All authors contributed parts of performing the research, of the design of the research framework, and of collecting and analysing the data. All authors wrote sections of the manuscript and contributed to its revision.

SUPPLEMENTARY MATERIAL

Assessment tables for all 11 systems analysed as well as the template table for analysing new systems are available at https://doi.org/10.5281/zenodo.5282929.

REFERENCES

[1] Nicholas, D., Rowlands, I., Watkinson, A., et al.: Have digital repositories come of age? the views of library directors. Webology (2013)

[2] Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., et al.: The FAIR guiding principles for scientific data management and stewardship. Scientific Data 3 (2016)

[3] Borrego, A.: Institutional repositories versus ResearchGate: The depositing habits of spanish researchers: Institutional repositories versus ResearchGate. Learned Publishing 30(3), 185–192 (2017)

[4] Nicholas, D., Rowlands, I., Watkinson, A., et al.: Digital repositories ten years on: what do scientific researchers think of them and how do they use them?. Learned Publishing (2012)

[5] Amorim, R.C., Castro, J.A., Rocha da Silva, J., et al.: A comparison of research data management platforms: architecture, flexible metadata and interoperability. Universal Access in the Information Society 16(4), 851–862 (2017)

[6] Manca, S.: ResearchGate and Academia.edu as networked socio-technical systems for scholarly communication: a literature review. Research in Learning Technology 26(0), (2018)

[7] Benz, D., Hotho, A., Jäschke, R., et al.: The social bookmark and publication management system bibsonomy: A platform for evaluating and demonstrating Web 2.0 research. The VLDB Journal 19(6), 849–875 (2010)

[8] Arlitsch, K., Grant, C.: Why so many repositories? examining the limitations and possibilities of the institutional repositories landscape. Journal of Library Administration 58(3), 264–281 (2018)

[9] Kindling, M., Pampel, H., van de Sandt, S., et al.: The landscape of research data repositories in 2015: A re3data analysis. D-Lib Magazine 23(3), (2017)
[10] Andro, M., Asselin, E., Maisonneuve, M.: Digital libraries: Comparison of 10 software. Library Collections, Acquisitions, and Technical Services 36(3), 79–83 (2012)
[11] Assante, M., Candela, L., Castelli, D., et al.: Are scientific data repositories coping with research data publishing?. Data Science Journal 15(0), 6 (2016)
[12] Austin, C.C., Brown, S., Fong, N., et al.: Research Data Repositories: Review of Current Features, Gap Analysis, and Recommendations for Minimum Requirements. IASSIST Quarterly 39(4), 24(2016)
[13] Kim, S.: Functional requirements for research data repositories. International Journal of Knowledge Content Development & Technology 8(1), 25–36
[14] Marcial, L.H., Hemminger, B.M.: Scientific data repositories on the web: An initial survey. Journal of the American Society for Information Science and Technology 61(10), 2029–2048 (2010)
[15] Murphy, D., Gautier, J.: A comparative review of various data repositories (2017). Available at: https://dataverse.org/blog/comparative-review-various-data-repositories
[16] Joo, S., Yeon Lee, J.: Measuring the usability of academic digital libraries: Instrument development and validation. The Electronic Library 29(4), 523–537 (2011)
[17] Nielsen, J., Landauer, T.K.: A mathematical model of the finding of usability problems. In: Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems, pp. 206–213 (1993)
[18] Máchová, R., Hub, M., Lněnička, M.: Usability evaluation of open data portals: Evaluating data discoverability, accessibility, and reusability from a stakeholders' perspective. Aslib Journal of Information Management 70(3), (2018)
[19] Jacobsen, A., de Miranda Azevedo, R., Juty, N., et al.: FAIR Principles: Interpretations and Implementation Considerations. Data Intelligence 2(1–2), 10–29 (2020)
[20] Hasselbring, W., Carr, L., Hettrick, S., et al.: From FAIR research data toward FAIR and open research software. it - Information Technology 62(1), 39–47
[21] Lamprecht, A.-L., Garcia, L., Kuzak, M., et al.: Towards FAIR principles for research software. Data Science Journal 3(1), 1–23 (2019)
[22] Devaraju, A., Mokrane, M., Cepinskas, L., et al.: From conceptualization to implementation: FAIR assessment of research data objects. Data Science Journal 20(1), 4 (2021)
[23] Research Data Alliance FAIR Data Maturity Model Working Group: FAIR Data Maturity Model: Specification and guidelines. Report from the Research Data Alliance (2020)
[24] European Commission, Directorate-General for Research and Innovation: Turning fair into reality: final report and action plan from the european commission expert group on fair data. Publications Office (2018)
[25] Lin, D., Crabtree, J., Dillo, I., et al.: The TRUST principles for digital repositories. Scientific Data 7(1), 144 (2020)
[26] Carroll, S.R., Herczog, E., Hudson, M., et al.: Operationalizing the CARE and FAIR Principles for Indigenous data futures. Scientific Data 8(1), 108 (2021)
[27] Assante, M., Candela, L., Castelli, D., et al.: Science 2.0 Repositories: Time for a Change in Scholarly Communication. D-Lib Magazine 21(1/2), (2015)
[28] Williams, M.L.: The adoption of Web 2.0 technologies in academic libraries: A comparative exploration. Journal of Librarianship and Information Science 52(1), 137–149 (2020)
[29] Ovadia, S.: ResearchGate and Academia.edu: Academic Social Networks. Behavioral & Social Sciences Librarian 33(3), 165–169 (2014)
[30] Paskin, N.: Digital object identifier (doi®) system. Encyclopedia of library and information sciences 3, 1586–1592 (2010)

[31] Chan, L., Cuplinskas, D., Eisen, M., et al.: Budapest open access initiative (2002). Available at: https://www.budapestopenaccessinitiative.org/read

[32] Fecher, B., Friesike, S.: Open Science: One Term, Five Schools of Thought. In: Bartling, S., Friesike, S. (eds.) Opening Science, pp. 17–47. Springer International Publishing, Cham (2014)

[33] Splawa-Neyman, P.: Figshare and the fair data principles (2018). Available at: https://figshare.com/articles/journal_contribution/Figshare_and_the_FAIR_data_principles/7476428/1

[34] Goble, C., Cohen-Boulakia, S., Soiland-Reyes, S., et al.: FAIR computational workflows. Data Intelligence 2(1), 108–121 (2020)

[35] de Sousa, N.T., Hasselbring, W., Weber, T., et al.: Designing a generic research data infrastructure architecture with continuous software engineering. In: Software Engineering Workshops 2018 (vol. 2066 of CEUR Workshop Proceedings), pp. 85–88. CEUR-WS.org (2018)

AUTHOR BIOGRAPHY

Daniela Oliveira is the metadata management lead of the Semantics team in Data42 @ Novartis. She was previously an Invited Assistant Professor in the Department of Informatics of the Faculty of Sciences, University of Lisbon, Portugal and a researcher at LASIGE in the research line of Health and Biomedical Informatics and Data & Systems Intelligence. Her research interests broadly focus on data integration to enable the publication of meaningful knowledge on the Web, with a special interest in the publication of health data to help find and access currently available knowledge that can enable the generation new knowledge in the domain.
ORCID: 0000-0003-0559-8737

Fabian Kirstein is a researcher and software developer at the Fraunhofer Institute for Open Communication Systems and a PhD candidate at TU Berlin. His work focuses on the area of Linked Open Data, Open Science, service-oriented architectures, scalable distributed systems, and decentralised data management. In those domains he participated in several national and international research and industry projects. In his Ph.D. project, he focuses on provenance and traceability in Open Data ecosystems. Since 2015 he is part of the core development team of data.europa.eu and contributes to the central data management component, the overall system architecture and the Linked Data integration.
ORCID: 0000-0002-9064-2546

Mathieu d'Aquin is professor of computer science and head of the K Team at the LORIA laboratory in Nancy, France. He was previously director of the Data Science Institute in NUI Galway Ireland, and head of the Data Science Team at the Knowledge Media Institute of the Open University, UK. His research interests include knowledge intensive systems and understanding the way in which computational representations of knowledge can contribute to artificial intelligence. He has worked on those topics in various application domains, including medicine, education, smart cities and the digital humanities.
ORCID: 0000-0001-7276-4702

Sonja Schimmler is research group lead at the Fraunhofer Institute for Open communication Systems and at the Weizenbaum Institute. She is also an associated researcher at Technical University of Berlin. In her research, she focuses on the digitalisation and opening up of science and puts a special emphasis on research data infrastructures. Her research interests range from semantic web and linked data over data science and artificial intelligence to software engineering and human-centered computing. She holds a Ph.D. in Computer Science from the University of the Federal Armed Forces Munich. She studied Computer Science at the Technical University of Munich and at the Georgia Institute of Technology (USA).
ORCID: 0000-0002-8786-7250

Sebastian Urbanek is a researcher at the Berlin University of Applied Sciences in the Department of Construction Engineering and Geoinformatics. As a PhD candidate of the Institute for Geography at the Otto-Friedrich-University Bamberg, he is developing a visualization of lost places for use by immersive technologies. In his scientific work, he focuses on data management of geodata, related standards, and their visualisation. He worked at the Fraunhofer Institute for Open Communication Systems in the scope of Open Data before and supported the development of the European Data Portal.
ORCID: 0000-0003-0925-1781