Analysis of current RDM applications for the interdisciplinary publication of research data

André Langer[0000−0001−7073−5377], Ellen Bilz, and Martin Gaedke[0000−0002−6729−2912]

Chemnitz University of Technology, Germany
(andre.langer, ellen.bilz, martin.gaedke)@informatik.tu-chemnitz.de

Abstract. The digital transformation of science allows researchers nowadays to expose Research Objects through many different publishing channels, so that other interested stakeholders can find and reuse it. Linked Data is an accepted mean in these meta descriptions to enhance Findability, Accessibility, Interoperability and Reusability (FAIR).
But researchers face a large variety of established publishing applications, where they have to select between general-purpose or domain-specific platforms and user interfaces of varying quality and feature set. In order to improve interoperability aspects, we want to analyze which publishing systems currently exist and to which extent they support Linked Data annotations from the very beginning.
We therefore concentrated on research data and conducted a systematic mapping of general-purpose research data management (RDM) systems currently in use, and summarize them in a tabular resource. The obtained results were then evaluated against their current support for semantic, interdisciplinary data annotation and exchange. We show, that a large set of established research data publishing solutions already exists, but that their support for Linked Data is still limited and can be improved.

Keywords: Research Data Management, Data Publishing, FAIR, Linked Data, Systematic Mapping

1 Introduction

Scientists are encouraged to publish, share and reuse research data (Open Research, stimulating Open Data and Open Access). Research data (other terms: scientific data, scholarly data) is an essential artifact of scientific work: It leads to insights, makes research reproducible and validates findings. Which also means, it is an integral part of the results of research. The number of scientific publications is growing, and consequently also its underlying data [4]. Publishing research data is a stage in common research data management life-cycle models [5], where created data is not only stored and versioned for private purposes but made available for other researchers. This affects activities such as data preparation,

annotation, provisioning and distribution. In the following, we will concentrate on digital research data that is published online as a web resource via http/s.

The digital transformation of science has offered new publication channels for research data, starting with direct data exchange in the past, traditional self-hosted webserver downloads, up to general-purpose or specialized data repositories, larger centralized or decentralized platform infrastructures with corresponding non-profit or commercial application providers. Examples are *CKAN*, *DSpace*, *EUDAT*, the *Harvard Dataverse*, *Zenodo* and derived systems. These platforms nowadays provide means to upload and publish research data sets with a general-purpose, disciplinary or institutional focus. They also have to satisfy specific needs of multiple stakeholders in different usage scenarios ([9], [7]). In order to improve interoperability, provided meta information is exposed through well-defined interfaces and well-defined protocols such as OAI-PMH to research data registries such as re3data.org or OpenAIRE in order to make them discoverable for other researchers. Additionally, many of these RDM platforms also claim to make heavy use of Linked Data principles already.

The aim of this study is to analyze the actual support of RDM systems for interdisciplinary data exchange from a user's perspective. We will describe benefits of semantic annotations for interdisciplinary data exchange and are interested in a systematic approach to get an overview over currently used scientific data publishing applications. Especially for researchers who are not technically versed, it can be a great effort to find an appropriate solution; or they will simply use a tool they are aware of in their knowledge domain or that they were told. This negative experience might also be a reason to keep scientists back from sharing their data or switch to simpler publishing approaches and tools restraining interdisciplinary research data exchange and reuse.

This research is an initial part of the PIROL PhD project [11] and has the following contributions in the context of interdisciplinary research data publishing:

1. We describe benefits of Linked Data to improve interdisciplinary and transdisciplinary research data discovery and reuse.
2. We identify the most relevant applications for research data publishing being used throughout the last ten years
3. We investigate the identified RDM publishing approaches in a comparative study on how good they support FAIR data operations in 2019 with a particular focus on Linked Data aware frontend interfaces.

The rest of the paper reflects these objectives in the following way: Section 2 introduces definitions for concepts related to interdisciplinary research data management. Section 3 describes the methodology used in this research to identify suggested research data publishing systems and describes the search strategy for reproducibility. The results of our systematic mapping are presented in section 4 and further investigated in section 5. Section 6 contrasts our work from existing studies and section 7 summarizes our results and gives an outlook to future activities.

cf. `https://www.re3data.org/metrics/software`
`https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/`

2 Conceptualization

Definitions for research data vary depending on institutions and application domains. Throughout this paper, we follow the definition, that research data "covers in principle any kind of digital artifact that is associated with scientific research" [16]. As multifaceted as research is, its artifacts can be as well: lab or measured data, audiovisual information, texts, surveys and even objects, probes and software. Activities to deal with research data and its life cycle are subsumed as *research data management*.

In our study, we are particularly focused on research data publishing within RDM life-cycle models, which can be defined as "making a research artifact online available, discoverable, peer-reviewable, re-usable according to given rights, real-time accessible, citable, and interlinked with its research activity and associated products" [6].

It was already proven that researchers are willing to publish and share research data to other scientists [7]. The motivation for sharing data is multilayer and shaped by collaborative exchange in the same knowledge domains, institutional-specific expectations and policies and personal rewards [9]. [7] also showed, that the will to provide meta data and persistent identifiers to describe research data sets is already existing. However, only 34% of all survey participants stated that services such as Google Scholar meet their requirements for data retrieval. This implies that user's are still not satisfied with obtained search results in general-purpose research data registries when they require research data of a particular characteristic. Finding relevant information in knowledge domain-specific data catalogs might yield to better results as soon as dedicated user interfaces and an established knowledge-domain related terminologies are used [8]. But the variety of concepts and terms used in different research disciplines makes it hard to provide research data also for potential users in different knowledge domains.

One possible approach to improve this situation is to extend meta data descriptions of research data sets with additional information. Linked Data is a means to map related relevant concepts onto each other [13]. This includes the

1. Identification of relevant entities,
2. Usage of appropriate ontologies,
3. Description of associated concepts,
4. Reference to related resources

through the use of persistent identifiers, which can realize FAIR principles for data sharing [17].

We have the hypothesis, that established research data publishing platforms still use traditional user interfaces that only request basic meta information as text, such as provenance information, missing to also allow sophisticated meta descriptions and linking for interdisciplinary reuse which represents an obstacle for discovery activities by other researchers.

3 Methodology

Assuming a scenario, that a scientist of a specific domain wants to publish data of a research project that is also relevant for other researchers, we want to identify RDM publishing applications currently in use. The resulting summary can help researchers to get an overview on relevant systems and compare their feature set. To facilitate the cross-domain retrieval and reuse of published research data, we further analyze all relevant systems afterwards with a specific focus on the current support of Linked Data. In this scenario, this particularly means how existing RDM systems help a scientist to describe different characteristics of a research dataset in a structured way by reusing existing ontologies and appropriate unique concept identifiers.

A predefined set of suggested RDM platforms or a processing of data source entries in existing data catalogs was not regarded as appropriate because it could miss other relevant publishing channels being actively in use that were not considered in such a registry in advance.

The following section will provide details on the survey methodology for reproducibility concerns.

3.1 Research Question

The goal of this comparative study is to get an overview on existing RDM solutions with a special focus on publishing and sharing research data for the interdisciplinary discovery and reuse by other researchers. As so, we will concentrate on general-purpose research data publishing platforms that can be used by any researcher independent of a particular knowledge domain. The study of domain-specific solutions for describing terminologies would be interesting as well, but is out of scope of this paper.

3.2 Criteria

We ran a systematic mapping approach on scientific publications dealing with research data management and data publishing approaches. To select relevant papers, the following criteria were used.

Inclusion Criteria The main criteria for the papers selected were provided by the publishing date and the language the paper was published in:

+ only papers published between 2008 and 2018
+ only papers in English

Furthermore, the papers should focus on at least one of those criteria:

+ research data management
+ data publishing or data sharing

Exclusion Criteria Papers with a focus on other aspects in the RDM life-cycle such as data collection, data processing and data aggregation were excluded. Also papers about systems which do not exist or are not maintained anymore were excluded.

- focus other than data publishing or sharing
- systems do not exist / are not maintained anymore

3.3 Search Strategy

Search Terms Formulating search terms to get relevant results was a challenge, as there is much ambiguity in the vocabulary of the subject's area. Several terms are used as quasi-synonyms, e.g., *data platform* vs. *data repository* vs. *data management system* or *research data* vs. *scientific data* vs. *scholarly data* vs. *datasets* in general. Nevertheless, the terms had to be limited due to the unmanageable amount of results in a test run. The search was conducted within the title, the abstract and the keywords (if exist), using the following terms:

- *platform*, *repository*, *data management system*
- *scientific data*, *research data*, *datasets*

As there is a focus on data sharing and publishing, the previous keywords were combined with:

- *publishing*
- *sharing*

Search Engines, Platforms and Digital Libraries The following search engines, platforms and digital libraries were used: *Google Scholar*, *ISI Web of Science*, *ACM Digital Library*, *IEEE Xplore Digital Library*, *Springer Link*, *Mendeley*, *Science Direct* , *Emerald Insight*, *SAGE Journals*. Due to the huge amount of results with descending relevance in *Google Scholar* (21,100) and *Springer Link* (21,031), only the first 1,000 results were screened by the title.

4 Identification of Research Data Publishing approaches

Conducting the search lead to 865 results after removing the duplicates. The outcome was then reassessed with reference to the criteria, which resulted in a sum of 487 papers. We reviewed all relevant publications and derived a basic systematization, before we further analyze and discuss all solutions with a focus on semantic data management.

4.1 Review of results

The publications were classified in two main categories: first, general papers including publishing and sharing research data aspects, and second, systems of different type to publish and share research data.

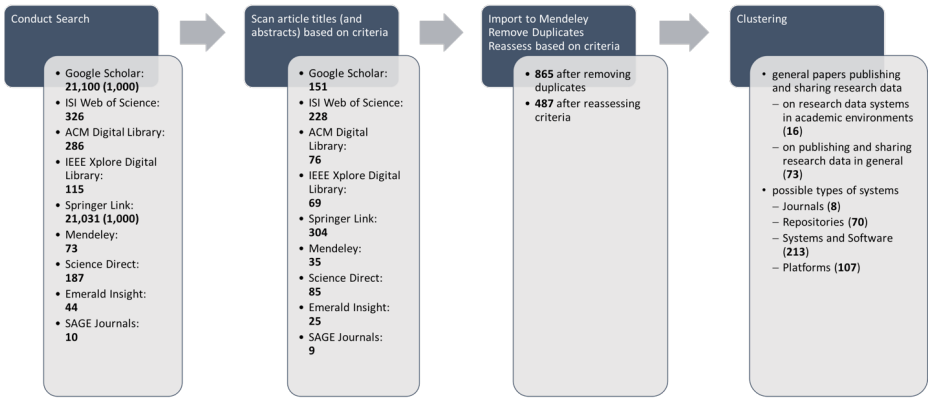

Fig. 1. Systematic Mapping Procedure

The group of general publications on publishing and sharing research allowed us to get an impression on the researcher's behavior and of systems significant in the publishing process. The second group of literature contained papers on one particular system. This included software to manage data repositories but also the research data environment of academic institutions. Even though only one system was addressed, quite often related systems were named. All mentioned unique data publishing approaches were noted.

The resulting set of available applications is depicted in table 1. The table contains software mentioned for data publishing purposes independent of the particular application focus. It is notable, that several systems are modifications or forks of another basic platform. Examples are *Dryad* which is a platform for journal data submissions built upon *DSpace*, *Zenodo* or the *EUDAT B2 services* built upon *Invenio*, or *Linkitup* built upon *Figshare*. A detailed explanation of each identified data publishing solution is omitted as we refer to the corresponding website and other surveys for details.

Interestingly, there were also data publishing solutions from related studies that did not show up in our study. Examples are *ContentDM*, *dLibra*, *Digital Commons*, *ESciDoc*, *Mendeley Data*, *Nesstar*, *RADAR*, *ResearchGate* or *Springer Nature*. They either target a different user segment, are placed under different search term strategy or are data provisioning channels not mentioned in the actual paper.

We manually removed systems that were knowledge-domain specific such as *AmeriFlux*, *Arctic Data Centre (ADC)*, *Brain-CODE*, *COINS*, *Curatr*, *ESSD*, *FID Sociology*, *GeoNode*, *GWATCH*, *IS-EPOS*, *LearnSphere*, *National Sleep Research Resource (NSRR)*, *NDA/NDAR*, *OCHEM*, *OMEGA*, *PANGAEA*, *PCORnet*, *Phenomics Ontology Driven Data (PODD)*, *reefgenomics.org*, *SchizConnect*, *SICAS Medical Image Repository*, *Sloan Digital Sky Survey*, *TypeCraft*, *Scratchpads*, *VPH-Share*, *Waveform ECG*, *WorldMap*, *xiSPEC*, *XNAT* or *XTENS*; or that were institutional-specific instances of an independent project such as *DepositOnce* as an application of *DSpace* at the *Technical University of Berlin*, *INFN OAR* as a clone of *Zenodo* in Italy, and *TIND* or *WEKO 3* as an *Invenio* instance for research data of multiple universities in the US and Japan.

System	Count	URL
GitHub	101	https://github.com/
Dryad	67	https://datadryad.org/
e!DAL	59	http://edal.ipk-gatersleben.de/
DSpace	51	https://duraspace.org/dspace/
Figshare	44	https://figshare.com/
Fedora	37	https://duraspace.org/fedora/
Eprints	35	https://www.eprints.org/
Dropbox	32	https://www.dropbox.com/
CKAN	27	https://ckan.org/
Dataverse	27	https://dataverse.org/
Zenodo	25	https://www.zenodo.org/
myExperiment	23	https://www.myexperiment.org
Globus	22	https://www.globus.org/
Virtuoso	22	https://virtuoso.openlinksw.com/
B2SHARE / EUDAT B2 Services	21	https://b2share.eudat.eu/
Drupal	20	https://www.drupal.org/
XSEDE	14	https://www.xsede.org/
Dendro	13	http://dendro.fe.up.pt/
D2R	11	http://d2rq.org/d2r-server
HUBzero	10	https://hubzero.org/
Google Drive	9	https://drive.google.com/
LabKey Server	9	https://www.labkey.org/
Greenstone	7	https://www.gdsidm.com/
SharePoint	7	https://www.sakaiproject.org/
Invenio	6	https://invenio-software.org/
eSciDoc	5	https://www.escidoc.org/
Omeka	4	https://www.omeka.org/
Open Science Framework	3	https://osf.io/
Sakai	3	https://www.sakailms.org/
SeedMe	3	https://www.seedme.org/
Archivematica	2	https://www.archivematica.org/
RODA	2	https://www.roda-community.org/
BRICS	1	https://brics.cit.nih.gov/
Clowder	1	https://clowder.ncsa.illinois.edu/
CoESRA	1	https://www.coesra.org.au
DataONE	1	https://www.dataone.org/
Galaxy platform	1	https://usegalaxy.org/
i2b2	1	https://www.i2b2.org
ISA tools suite	1	https://isa-tools.org/
Islandora	1	https://islandora.ca/
Jackrabbit	1	http://jackrabbit.apache.org/
Linkitup	1	https://github.com/Data2Semantics/linkitup
RIKEN MetaDatabase	1	http://metadb.riken.jp/
RunMyCode.org	1	http://www.runmycode.org/
SAIL Databank / UKSeRP	1	https://saildatabank.com/
SEEK	1	http://seek4science.org/
SQLShare	1	https://sqlshare.uw.edu/
Stardog	1	https://www.stardog.com/

Table 1. List of identified research data publishing solutions and their number of mentions in different publications (count) in our systematic mapping

5 Discussion

In this section, we discuss the obtained results of the systematic mapping and further investigate all identified solution with respect to their current support for Linked Data operations to improve interdisciplinary retrieval and reuse.

5.1 Discussion of study results

The landscape of solutions to support publishing and sharing research data is hallmarked by a huge diversion, especially through the variety of domain-specific specialized platforms. Researchers should become aware of the systems that exist and research should focus on lowering barriers between these systems to eliminate data silos and increase interoperability. Gaining a fast impression of a system's functionality was often difficult due to a lack of user-focused self-description and list of features. A better communication would already open them for a bigger community.

It is remarkable, that research data publishing activities often take place via *GitHub* (software links explicitly excluded). The platform is referenced more often than the *Dryad* journal platform or the underlying *DSpace* system which implies a notable current relevance for the research community.

Furthermore, the study revealed publishing channels that are traditional and reasonable, but which were not expected in advance. Examples are the data publication via a website Content Management Systems (CMS) such as *Drupal*, or the usage of commercial Cloud providers such as *Dropbox* or *Google Drive* and similar solutions. As we were looking for particular research data management software, we did not emphasize these research-external solutions, but use it as an indicator that users might stick to accustomed, easy-to-use solutions for file sharing, even if they might be commercially driven or privacy-insensitive.

5.2 Discussion of methodology

The applied systematic mapping methodology had some implications to be considered. Even though the search terms were limited, the amount of obtained search results was huge as many publications nowadays include a data publishing aspect. The applied search result pruning implies, that some systems might have been left out. The exclusion of processing and analyzing might have excluded some very interesting systems with a Linked Data focus. In contrast, the general terms *platform* or *repository* often resulted in small and very specific systems which had no huge value for the general identification of research data publishing systems. Also, the large amount of domain-specific systems and platforms did not allow to validate each solution and check its feature set. We therefore concentrated on domain-independent research data management platforms and analyzed only approaches that showed up at least 10 times or were base systems or higher ranked solutions.

Identifying the main representatives by occurrence can be an indicator to show their relevance and establishment, but not consequently their eligibility and quality. This applies especially for newer and hyped systems. Also, multiple RDM systems did not exist anymore and therefore not showed up in the mapping summary.

5.3 Discussion of FAIR and Linked Data support

To improve scientific discoverability, exchange and reuse among multiple research disciplines, we further investigated the identified research data publishing solutions for the implementation of sustainable interdisciplinary data management capabilities. In order to compare these RDM systems on assessable characteristics, we derived 15 criteria C from the FAIR principles [17] for research data management, that particularly focus on the awareness and handling of Linked Data (LD) in these data publishing solutions as shown in the following:

Findable (LD)

- C1 Is a particular research data set in a current version accessible via a unique PID?
- C2 Is the research data information through that platform indexed in data catalogs, registries and search engines?
- C3 Is a search interface available with filter possibilities for structured Linked Data?

Accessible (LD)

- C4 Can new research data be stored or referenced in an easy way?
- C5 Is the user input interface Linked Data - aware and easy to use by hiding technical terms and identifiers?
- C6 Can the research data and/or meta data be accessed directly via http(s)?
- C7 Do authentication and authorization settings for public/private/restricted access exist?

Interoperable (LD)

- C8 Is the meta data description available in an RDF serialization?
- C9 Can particular established ontologies be used to describe the research data set in a general way such as *schema.org/Dataset*, *DataCite* or *DCAT/DublinCore*?
- C10 Can domain-specific vocabularies be used to further describe the research data?
- C11 Can each concept related to the research data set be described with a corresponding resource URI?

Reusable (LD)

- C12 Can a data license be specified in a Linked Data fashion?
- C13 Can the data provenance be specified and updated in a structured way?
- C14 Are data sets set into relationship based on Linked Data and criteria such as the topic, community, used methods or similar?
- C15 Is the provided data validated or do compliance checks exist?

In order to assess every criterion, a demo instance of each approach was examined together with the provided application self-documentation. We therefore used the basic version without any enabled plugins or sophisticated configuration.

A scoring system was applied and the following symbols were used:

- \+ was assigned, if the criterion was entirely fulfilled
- o was assigned, if the criterion was partially fulfilled
- \- was assigned, if the criterion was not fulfilled
- % was assigned, if the criterion was not applicable
- ? was assigned, if it was not possible to assess the mentioned criterion
- () was assigned, if the feature is limited in the native version but might be there with plugins

Table 2 lists the results of this assessment. We thereby categorize the previously identified RDM solutions in three types: *RDM systems* that serve as a basic data management platform, *RDM applications* as a further development of these platforms, and other *tools* for data management; as a direct comparison between solutions of different application focus appeared inappropriate.

We faced a lack of Linked Data support in the identified RDM systems. They normally provide direct web access to the research data itself via a uniform, persistent resource identifier (C1, C6), manage a research upload (C4) and

Type	Name	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	Total
RDM Systems	CKAN	+	+	o	+	-	+	-	(-)	-	-	-	o	+	o	-	**(-)**
	e!DAL	+	+	-	+	o	+	-	+	+	?	o	+	+	?	+	**o**
	ePrints	+	+	o	+	-	+	o	-	-	-	-	+	+	o	-	**-**
	Dataverse	+	+	o	+	+	+	o	+	+	-	o	+	+	o	-	**+**
	DSpace	+	+	o	o	(o)	+	+	(-)	+	(-)	(-)	o	+	o	-	**(o)**
	HUBzero	o	o	o	o	-	+	-	-	-	-	-	-	%	o	-	**-**
	Invenio	+	+	o	+	-	+	-	-	o	-	o	-	o	-	+	**-**
RDM Applications	Dryad	+	+	o	+	-	+	-	-	+	-	-	-	-	-	-	**-**
	EUDAT B2	+	+	+	+	+	+	-	-	+	-	o	+	+	+	-	**+**
	Figshare	+	+	o	+	o	+	o	-	+	-	o	o	+	+	-	**o**
	Globus	o	-	o	o	-	o	+	-	+	-	-	-	o	-	-	**-**
	myExperiment	+	o	o	+	+	+	+	+	+	-	+	+	+	+	-	**+**
	XSEDE	o	-	o	o	-	+	-	-	-	-	-	-	-	-	-	**-**
	Zenodo	+	+	o	+	o	+	+	+	+	o	+	+	+	+	o	**+**
RDM Tools	D2R	+	%	+	%	-	o	%	%	-	-	o	%	%	%	%	**o**
	Dendro	+	%	+	o	+	+	o	+	+	o	+	o	+	+	+	**+**
	Fedora	+	%	%	+	%	+	+	+	%	%	+	%	+	%	%	**+**
	Virtuoso	+	%	+	+	-	+	+	+	+	+	+	%	%	+	-	**+**

Table 2. Systematized mapping results with assessment of Linked Data support

request in an input form additional meta information that is mapped to a *DCAT*, *Dublin Core* or *schema.org* vocabulary. However, most of the meta information is still based on literal string input (C5). This means, that the strength of Linked Data to set other concepts into a comprehensive, distinct relationship is not used at all. This situation could be improved for user input interfaces in the frontend by providing recommended and more sophisticated input operations [12]. It was also not possible to provide a meta description file with predefined values in an automated fashion.

The generated meta information can only be reviewed on every RDM platform and the provided export format is only in a few cases available in an RDF serialization (C8). Provided metadata was mostly limited to basic information and provenance data (C9). Other relevant, domain-specific vocabularies are seldom taken into consideration or offered in the user interface (C10). If such a structured research data description is possible, it was limited to a rudimentary provision of property URIs and object values to identify particular topics.

Managing versions and provenance aspects was a key feature, where the majority of the tested RDM solutions succeeded (C13). However, this provenance information was seldom part of the meta description export (e.g., by relying on the *PROV-O ontology*) which could be of high relevance for reusing research data and existing meta information. Automatic data quality assessment checks were not run by any of the tested systems and only concentrated on a valid meta data input (C15). Instead, platform providers include a manual review process with a confirmation step before publishing a submitted resource.

Outstanding solutions with Linked Data awareness that should be highlighted and considered are currently the *Harvard Dataverse*, the *EUDAT B2 services*, *Zenodo* and *DSpace* with TripleStore and RDF extensions and/or a *Fedora* sublayer. Toolwise, *Dendro* has to be pointed out as a tool to specify potential relevant meta data.

6 Related Work

Research data management systems were already compared in several studies.

Kökörčený and Bodnárová [10] focused on digital library systems and gave an overview on appropriate existing systems and their information architecture for libraries, but without dealing with data management in particular. The provision of data repository content to the Semantic Web was discussed by Becker et al. [3], focusing particularly on technical aspects to reduce data silos and implementing open protocols to extend a traditional RDM application with Linked Data capabilities on the example of DSpace.

Poole [15] provided a critical literature review and discussed the current state in digital curation also with respect to human factors. The current necessary transition of data platforms was then described by Poline et al. [14], who encouraged to see data publishing as the next, higher step in comparison to data sharing. Assante et al. [2] concentrated on 5 selected scientific data repositories and their capability to use them for Open Science and data publishing activities. In another study, Amorim et al. [1] compared 6 out of 15 preselected RDM solutions from an institutional perspective and gave a comprehensive description and overview of the feature set for each system.

However, all of these surveys assumed a pre-defined basic set of relevant systems to evaluate. We reassessed the establishment of RDM solutions in 2019 by focusing on an analysis of existing publications and a systematic mapping. Furthermore, we particularly highlight the role and support of Linked Data in these data management systems.

7 Conclusion

In this paper, we investigated current research data management platforms and their capabilities to request and expose Linked Data descriptions of research data sets for interdisciplinary discovery and reuse. We therefore conducted a comparative study where we identified 18 web-based general-purpose RDM platforms that are used in 2018 to publish research data.

All identified web applications were then evaluated with respect to their current support for dealing with Linked Data to follow FAIR principles for data sharing. We came to the conclusion, that the support for Linked Data differs among all examined systems and that there is room for improvement, especially from a UI input point of view. Nevertheless, a basic support is already given in systems such as the *Dataverse*, *EUDAT B2 Services*, *Zenodo* or *DSpace*. In particular, this is limited to discovery meta information with links to author's ORCIDs and basic DublinCore properties in most of the cases. Although dedicated ontologies with unique concept identifiers already exist in multiple knowledge domains that can be used to describe relevant aspects of a data set, they are hardly used due to a lack of appropriate user interfaces and terminology services.

A closer look, especially on the huge variety in knowledge domain-specific systems, could be a benefit for interdisciplinary research and for the exchange of experiences and best practices

References

1. Amorim, R.C., Castro, J.A., Rocha da Silva, J., Ribeiro, C.: A comparison of research data management platforms: architecture, flexible metadata and interoperability. Universal Access in the Information Society 16(4), 851–862 (2017)
2. Assante, M., Candela, L., Castelli, D., Tani, A.: Are Scientific Data Repositories Coping with Research Data Publishing? Data Science Journal 15, 1–24 (2016)
3. Becker, P.N., Erstgutachter, D., Paschke, A., Schulte, C.: Repositorien und das Semantic Web – Repositorieninhalte als Linked Data bereitstellen (2014)
4. Bornmann, L., Mutz, R.: Growth rates of modern science: A bibliometric analysis. CoRR abs/1402.4578 (2014)
5. Charalabidis, Y., Zuiderwijk, A., Alexopoulos, C., et al.: The Multiple Life Cycles of Open Data Creation and Use. pp. 11–31. Springer, Cham (2018)
6. Dallmeier-Tiessen, S., Khodiyar, V., Murphy, F., et al.: Connecting Data Publication to the Research Workflow: A Preliminary Analysis. International Journal of Digital Curation 12(1), 88–105 (2017)
7. Jianhui, L., Chao, W., Lili, Z., Chengzan, L., Lianglin, H.: Survey and analysis of scientific data publishing. China Scientific Data 1(1) (2016)
8. Karam, N., Müller-Birn, C., Gleisberg, M., et al.: A Terminology Service Supporting Semantic Annotation, Integration, Discovery and Analysis of Interdisciplinary Research Data. Datenbank-Spektrum 16(3), 195–205 (2016)
9. Kim, Y., Nah, S.: Internet researchers' data sharing behaviors: An integration of data reuse experience, attitudinal beliefs, social norms, and resource factors. Online Information Review 42(1), 124–142 (feb 2018)
10. Kökörčený, M., Bodnárová, A.: Comparison of digital libraries systems. Advances in Data Networks, Communications, Computers pp. 97–100 (2010)
11. Langer, A.: PIROL : Cross-domain Research Data Publishing with Linked Data technologies. In: La Rosa, M., Plebani, P., Reichert, M. (eds.) Proceedings of the Doctoral Consortium Papers Presented at the 31st CAiSE 2019. pp. 43–51. CEUR, Rome (2019)
12. Langer, A., Göpfert, C., Gaedke, M.: URI-aware user input interfaces for the unobtrusive reference to Linked Data. IADIS International Journal on Computer Science and Information Systems 13(2) (2018)
13. Parsons, M.A., Godøy, Ø., Ledrew, E., et al.: A conceptual framework for managing very diverse data for complex, interdisciplinary science. Journal of Information Science 37(6), 555–569 (2011)
14. Poline, J.b., Chodacki, J., Gulick, A.E.V.: Data sharing to data publishing [version 2] pp. 1–14 (2019)
15. Poole, A.H.: How has your science data grown? Digital curation and the human factor: a critical literature review. Archival Science 15(2), 101–139 (2015)
16. Sousa, R.B., Cugler, D.C., Malaverri, J.E.G., Medeiros, C.B.: A provenance-based approach to manage long term preservation of scientific data. In: Proceedings - International Conference on Data Engineering. pp. 126–133. IEEE (2014)
17. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., et al.: The fair guiding principles for scientific data management and stewardship. Scientific data 3 (2016)