RESEARCH ARTICLE

RDM Platform Coscine – FAIR play integrated right from the start

Ilona Lang [1]
Marcel Nellesen [1]
Marius Politze [1]

1. IT Center, RWTH Aachen University, Aachen.

Date Received:
2022-02-17

Keywords:
Coscine, Research Data Management Platform, FAIR Guiding Principles, FAIR Digital Object Framework, metadata

Data availability:
Data can be found here:
git.rwth-aachen.de/coscine

Software availability:
Software can be found here:
coscine.de/

Abstract. Nowadays, researchers often need to distribute their research data among a multitude of service providers with varying (if any) levels of maturity in terms of FAIR Research Data Management (RDM). To provide researchers with a single point of access to their project data and to add a 'FAIR' layer to already established services, the RDM platform Coscine was developed. Within Coscine different services (so-called resources) can be added to a project, allowing access to the associated data for all project participants. A Persistent Identifier (PID) is assigned for each resource and metadata management is integrated with flexibly definable schemas based on Resource Description Framework [4] (RDF), Web Ontology Language (OWL) and Shapes Constraint Language (SHACL). Thereby, Coscine bundles for each project the research data, metadata, interfaces and PIDs into a linked record according to the FAIR Digital Object [27] (FDO) model.

1 Introduction

For many researchers, whether from engineering sciences or other fields, an involvement with the 'FAIR Guiding Principles' [34] does not begin until the publication of an article and the sometimes-obligatory transfer of the research data to a repository. At this point, a significant amount of valuable information about the research project is often already lost. Therefore, only a fraction of the data (and metadata) collected during a research project is ever published.

1.1 A Brief Overview on RDM Platforms

But even if researchers try to follow the 'FAIR Guiding Principles' during their whole data life cycle, it is a big challenge to find a service that offers solutions for all project-related data types (e.g., managing code, collaborative work, multiple large files). Therefore, researchers typically employ a broad spectrum of IT service infrastructures for their projects that range from local to centralized, federated and external IT service providers. Central applications like Radar [17] or MASi [9] are less specific and address a wider community with more generic RDM workflows. External 'clouds' like Zenodo, Figshare or Open Science Framework (OSF) support basic RDM workflows like citation or persistent identification. By far most prominent are generic 'clouds', like the Owncloud-based tool Sciebo [32], Dropbox, Google Drive or GitLab. They are used to store and manage data, however, these options usually lack in support of RDM workflows or policies.

Taken together, the situation nowadays often leads to a fragmentation of research data among a multitude of service providers with varying (if any) levels of maturity with respect to FAIR RDM. Moreover, the amount of service providers makes it hard for researchers to keep an overview over the entirety of data related to a research project.

1.2 Goals & Requirements

Thus, a software solution is needed to get all research data under one roof while supporting the 'FAIR Guiding Principles'. Based on the focus on engineering at RWTH Aachen University and the associated high volume of research data, initial analyses and developments towards such a software solution were started at the RDM team of the IT Center in 2018. Two options were analyzed:

1. develop a data management system that replaces all existing services or
2. develop a data management system that adds a 'FAIR' layer to already established services.

The first option would require an enormous amount of human resources to cover all functions already developed by other services. A recent study shows, however, that the software development in the public sector is and will be confronted with low human resources [25]. This makes the development of a data management system that replaces all existing services an unattainable goal in the near future. The second option thus has two direct advantages:

1. the data management system does not have to cover all the functions of already established services, but can focus entirely on adding features for compliance with the 'FAIR Guiding Principles' and
2. researchers can use all their established services and still get access from one platform.

To create such a data management system that supports researchers during their whole data life cycle, the RDM platform Coscine was developed at the IT Center of the RWTH Aachen University (Figure 1). Since 2020, the development is further supported by two consortia of the National Research Data Infrastructure (NFDI): NFDI4Ing [24] and NFDI-MatWerk [6]. These consortia aim to develop RDM solutions that, at best, can be applied to other disciplines as well. For the engineering sciences, NFDI4Ing was founded to develop, disseminate, standardize and provide methods and services to make engineering research data FAIR[1].

In this paper, we show which features Coscine provides for researchers and how they support the 'FAIR Guiding Principles'– from the initial collection of data to its subsequent reuse.

2 Core Features of Coscine

Coscine is a platform for the management, storage and archiving of research (meta)data generated in the context of research projects. For each project, Coscine allows inviting all project

1. see https://nfdi4ing.de/about-us/

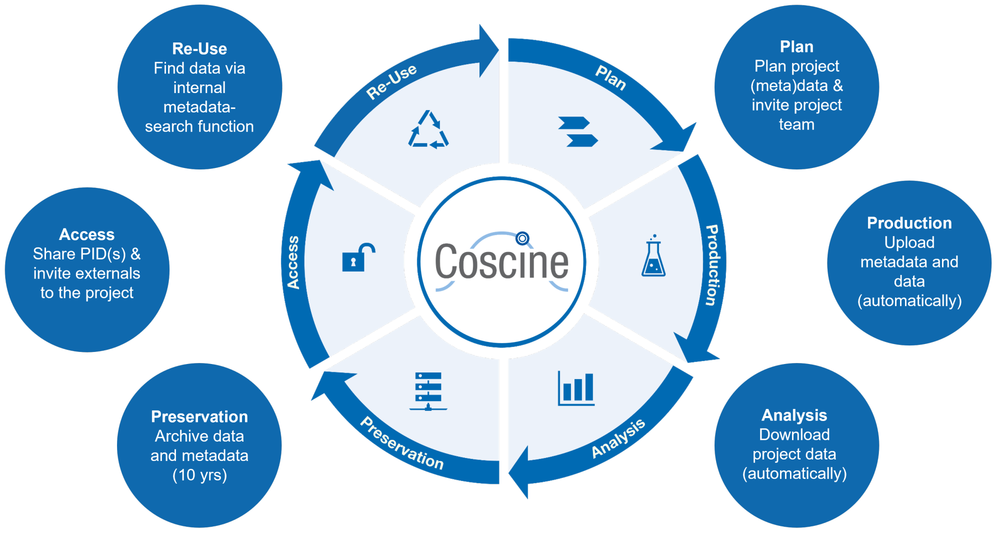

Figure 1: Using Coscine along the research data life cycle. The usage of Coscine starts at the beginning of a project, when the project-related metadata is defined and project participants are invited. During the production and analysis phase, Coscine provides access to project-related (meta)data for all project participants. Depending on the used resource type, (meta)data can be archived inside the respective resource. To access the (meta)data, Coscine assigns for each resource a PID and offers the possibility to add externals to a project. The reuse of (meta)data is supported by an internal search function.

participants, integrating the project-related data from different resources and adding the related metadata (Figure 5). Specifically, Coscine offers researchers the following core features:

2.1 Integration

By integrating various already established services, so-called resources (Figure 2), researchers can see and manage all project data in one place via the Coscine web interface or the Coscine API. Currently, resource types of the Research Data Storage [7] (RDS) (see below), Linked Data and GitLab are integrated. For the end of 2023 also cloud applications such as Sciebo and Nextcloud shall be added as resource type. Based on customer requests or market changes, additional resources can be continuously added or others replaced.

2.2 Storage Space

Coscine provides researchers of participating universities access to storage space on the RDS. The RDS is a consortial object storage system funded by the Ministry of Culture and Science of the State of North Rhine-Westphalia (MKW) and the Deutsche Forschungsgemeinschaft (DFG). When using RDS resources, a retention and archiving period of research data of ten years after the end of a research project is ensured in terms of Good Scientific Practice [5] (GSP). By default, employees of participating universities receive 100 GB of storage space per project for their research data, which they can distribute among several so called RDS-Web resources. For large amounts of data, more storage space can be requested. It is also possible to request RDS via S3 (RDS-S3) resources to interact directly with the S3 buckets or RDS-S3 with the setting WORM (RDS-WORM) resources to store research data with high protection requirements and

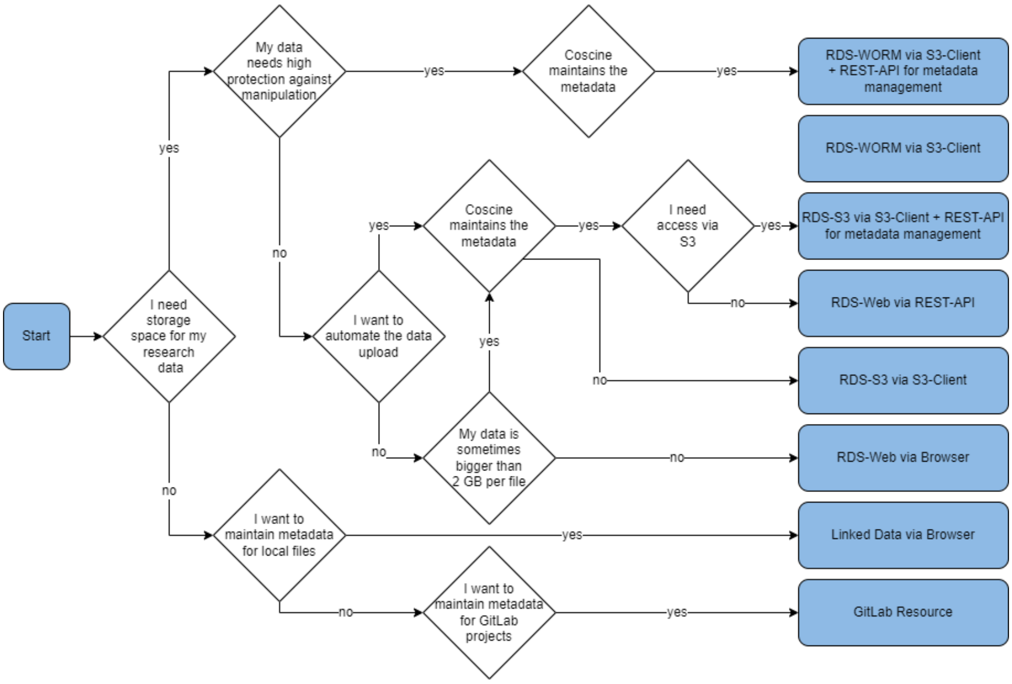

Figure 2: Resource Types in Coscine.To date, there are three different resource types in Coscine: RDS (subtypes: Web, Simple Storage Service (S3), write once, read many (WORM)), GitLab, and Linked Data. The decision diagram helps to select the right resource type based on different project needs.

prevent subsequent manipulation of the data (Figure 2).

Researchers can apply for RDS storage space using the Joint Application Review and Dispatch Service (JARDS) [13] (Figure 3). The JARDS platform allows researchers to create and manage their applications as well as RDM experts to review these applications regarding formal, technical and RDM specific feasibility. If large amounts of storage (>125 TB) are requested, a scientific review is performed to ensure the scientific value of the project. JARDS is already widely used within the high-performance computing community in Germany, so many researchers are already familiar with the platform and the procedure. Researchers thus may request storage space independently of their affiliation, however whether access is granted remains a policy of the storage provider. Especially for RDS storage, space may be provided for use cases endorsed by any NFDI consortium if they meet the formal, technical and scientific criteria mentioned above.

2.3 Collaboration

Coscine allows access for all internal and external members of a research project. Users can log in as a member of a participating organization via Shibboleth or as an external person via their Open Researcher and Contributor ID [10] (ORCID). While the ability to request certain storage services may be restricted, once added to a project the resource is available for all member. Basic functionalities like project and metadata management are available to all users. Project members can be invited to projects in a low-threshold way via their email, enabling easy collaborations.

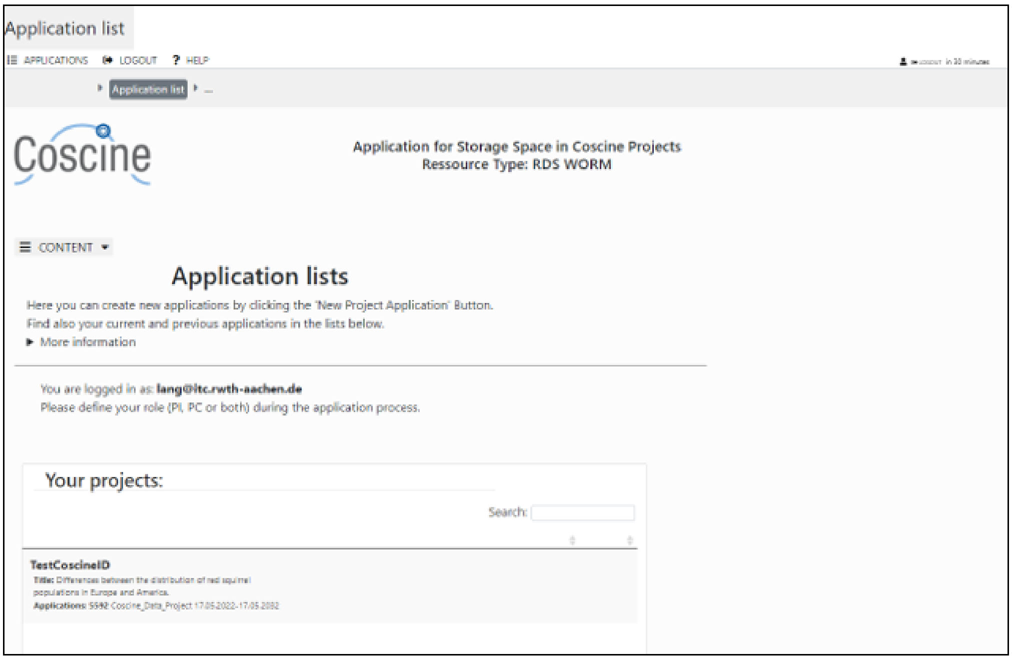

Figure 3: JARDS: Overview of ongoing and approved applications

2.4 Metadata

The use of Coscine involves three levels of metadata: at the project, resource, and data level. Adding metadata at the project and resource level is mandatory, and the necessary fields are standardized for all users and disciplines. At the data level, users can choose between different application profiles to optimally describe their research data inside a resource. All metadata are captured according to flexibly definable schemas that follow RDF, OWL, and SHACL standards. This allows a Coscine-wide search for all available metadata.

Individual application profiles can be created using the integrated application profile generator, developed within the DFG-funded project Applying Interoperable Metadata Standards (AIMS) [8]. This application profile generator allows researchers to create new application profiles from scratch or explore and extend already existing ones (Figure 4). New profiles can be sent as a merge request to the GitLab repository of Coscine, where they are reviewed by RDM experts to ensure a required level of technical quality and interoperability for Coscine.

2.5 Archiving

After completion of a research project, research data and metadata stored in resource types of RDS or Linked Data can be archived for ten years according to GSP. Thanks to the link to metadata, the assignment of a PID and the existing access for project members, Coscine facilitates the low-threshold subsequent use of the research data even during archiving.

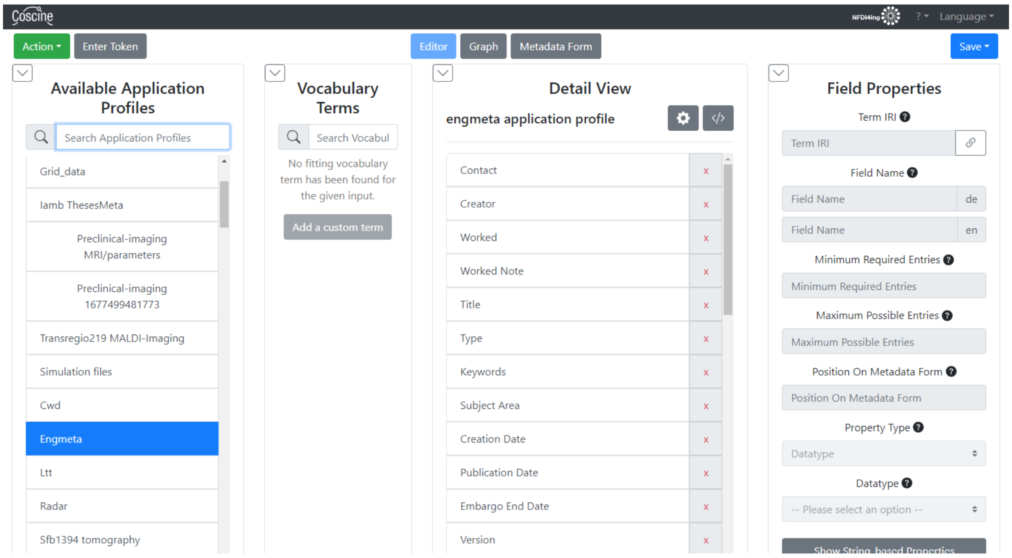

Figure 4: Screenshot of the application profile generator developed within AIMS [8].

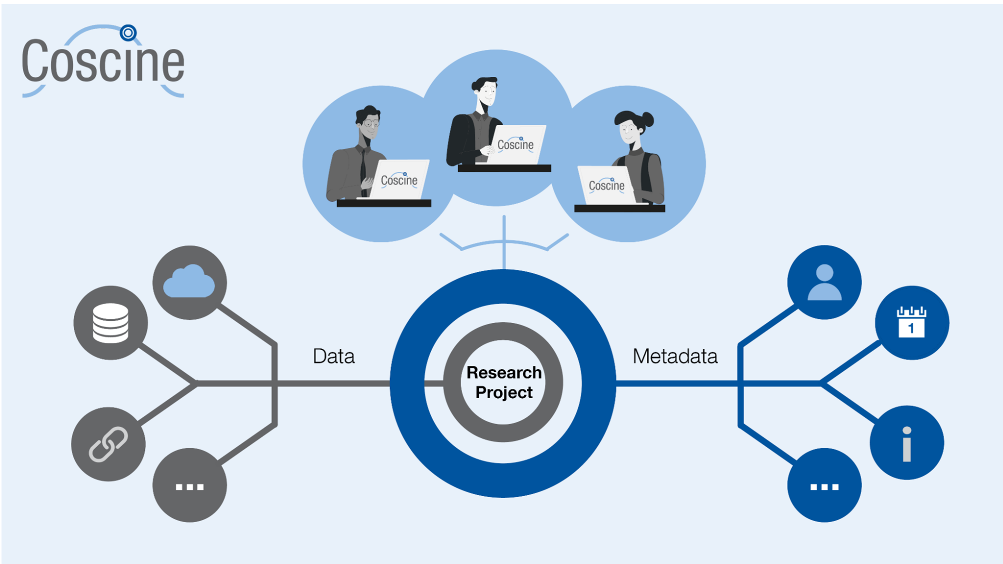

Figure 5: The project structure of Coscine. For each research project, researchers can invite all project participants (above – light blue circles), integrate the project-related data from different resources (left side – gray circles) and add the related metadata (right side – blue circles).

3 Coscine & 'FAIR Guiding Principles'

To enable the accessibility of research data in line with the 'FAIR Guiding Principles' across institutional borders, Coscine can be accessed either through participating universities or at a low-threshold level via ORCID. After registration, researchers can create a research project and invite all project-related participants. The project creator is automatically the project owner and can choose between three different roles for the other participants (owner, member, or guest). In line with A1.2 of the 'FAIR Guiding Principles' [31] the mandatory registration of project participants ensures the authentication of all data owners and contributors for each dataset, while the role management enables the definition of user-specific rights.

3.1 Metadata Representation

For research projects, metadata is collected at three levels and automatically linked to the research data. The first level of metadata relates to the research project (including name, description, Principal Investigators (PIs), discipline). The second level of metadata describes the resources, which are assigned to the research project (including resource name, discipline, keywords, metadata visibility, license). The third level of metadata is realized via application profiles that describe the uploaded or linked research data. For this step the researchers must select for each resource an application profile from various predefined profiles, e.g., for engineering research data the established EngMeta profile can be used. If a suitable application profile has not yet been added to Coscine, the AIMS Application Profile Generator [8] can be used to create a profile with individual and discipline-specific metadata. When using the storage resource type RDS-Web, file upload is only possible after entering the associated metadata in the application profile. In this way, Coscine makes metadata entry a direct part of the researcher's workflow, supporting the FAIR principles.

The World Wide Web Consortium (W3C) standards RDF [4] and SHACL [16] are used for the technical representation and validation of all metadata stored in Coscine. This largely complies with the FAIR principles regarding findability, interoperability, and reusability of metadata [31]. By using the AIMS Application Profile Generator [8] researchers without knowledge regarding RDF and SHACL can still create an application profile that suits their needs while being FAIR regarding the technical representation and validation.

Following the recommendations of the FAIR principle F4, the (meta)data are indexed in Coscine in a searchable resource via ElasticSearch. To also publish the semantically-rich and machine-actionable metadata, we work on implementing FAIR Data Point [3] (FDP) as a standardized interface [26]. Moreover, a connection to the NFDI4Ing metadata hub is currently realized via ”FAIR Digital Object” interfaces.

To support researchers' processes as much as possible and to align with A1 [31], Coscine provides open, free and universally implementable protocols to access data based on the resource type, either via a browser, using a REST API or directly via an S3 interface. This allows for high performance transfer of even large amounts of research data.

Regarding the FAIR principles F1 and A1 [31], Coscine assigns for each resource (including data and metadata) a handle-based ePIC-PID [14, 18]. This is used to uniquely and permanently identify the location of the resource and all contained files on a global level. As a result, each RDF-triple includes a PID leading to the data it describes. Within resources, fragment identifiers are used to address individual files by extending the handle URL.

Even though the technical standards used by Coscine to represent metadata are featuring a set of complex technologies, they are mostly hidden for the average user of the web user interface or the REST API respectively. Hence, a researcher in a lab or even a data scientist storing, annotating and accessing data can make use of the underlying standards without going into technical details. This is slightly different for data stewards, who are often required to configure projects or create application profiles. The creation process is partly supported by the AIMS Application Profile generator, however advanced use cases will likely require some knowledge of RDF to (re-)use or define vocabulary terms or thesauri. Most advanced users could create complex queries using SPARQL Protocol And RDF Query Language [29] (SPARQL) for the metadata stored in SHACL validated graphs [20]. In turn, this requires in depth knowledge of the used technologies and terminologies.

The layers in Coscine (metadata, interfaces & operations and persistent identifiers) that increase the FAIRness of the research data can be best described with the framework of FDOs.

3.2 Coscine & FAIR Digital Objects

The FAIR principles are about making data findable, accessible, interoperable and reusable both for humans and machines. To reach these aims, RDM software requires a framework to store and disseminate digital objects in a robust and informative way.

Although the concept of Digital Object (DO) was introduced by Robert Kahn in the early 1990s, an ecosystem of easy tools that add the FDO layers to raw data including unique identifiers and metadata is still needed [15]. This issue is most prominent in current industry grade IT solutions on the market, as used for the RDS. While these usually provide high scalability at reasonable costs, their focus is clearly on (mostly) standardized storage of and access to binary information rather than (global) identification or (fine granular) description of the data itself.

Using the notion of the FDO as shown Figure 6, Coscine adds on to the bit sequences in a storage system with required elements as successive layers: metadata, interfaces & operations and finally a persistent identifier. All the elements of the FDO form a logical unit that can be distributed and fully interpreted in solitude. While FDO supplies a generic architecture, different frameworks exist for their representations [11].

For retaining the bit sequence of the FDO Coscine relies mostly on a background storage system. In the case of the RDS the provided HTTP based S3 interface can be directly handed through to the client. For storage service that do not provide an HTTP accessible interface or in cases where access management is required, Coscine provides means for protocol translation. Coscine aims to combine approaches from two frameworks: PIDs based on Kernel Information Records (KIRs) [33] and the semantic approach of the FAIR Digital Object Framework [2] (FDOF).

On the one hand, the KIR work “by injecting a tiny amount of carefully selected metadata into a [PID] record” [33]. While the metadata set is typically small and rather technical key-value-pairs, directly adding it into the PID provides basic information about the described FDO without the

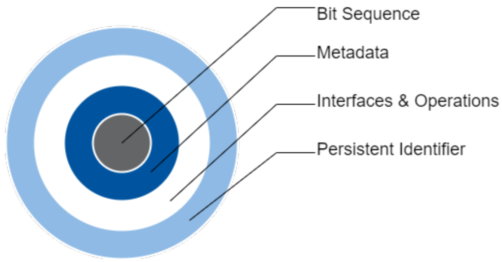

Figure 6: A layered model of an FDO with the elements needed to make the data FAIR: bit sequence, metadata, interfaces & operations and the persistent identifier [27].

need of querying additional metadata indexes. The FDOF, on the other hand, provides a set of conventions that suggest "predictable resolution behaviour" [3] for accessing bit sequences and binding rich and discipline specific semantic metadata in the form of linked documents. An FDOs implemented with the combination of both frameworks thus is machine and human actionable, technically and semantically meaningful, and widely technologically independent.

The KIR is used by Coscine to store information about the (file) type of the DO and how it can be accessed. Additionally, Coscine provides links that can be followed to access the bit stream and the semantic metadata documents. The semantic representations can be retrieved from using interfaces compliant to the FDP specification that builds upon Linked Data Plattform [30] (LDP) and extends Data Catalog Vocabulary [19] (DCAT) with a metadata service. While LDP and DCAT allow discovery of data along the hierarchies defined by projects, resources and files, FDP defines the access to the rich semantic metadata and the respective application profiles for the different levels of the aforementioned hierarchy.

4 Coscine – Options for Process Automation

Many approaches to RDM consider an ideal scenario where researchers start from scratch with a new project. However, this is often not the case, since research projects have a very long lifetime and sometimes a correct management of the data and the corresponding metadata was not originally considered. In addition, research projects are generating increasing amounts of data, which requires flexible automation of data handling processes. Thus, supporting this type of projects in Coscine is important as it allows easier adaption of the platform on a larger scale.

4.1 Data Upload

Depending on the requirements of the researchers, different resource types and ways for interactions (e.g., web UI, REST API, S3 protocol) are available in Coscine, of which RDS-S3 in particular is suitable for handling large amounts of (already existing) research data (Figure 2). The RDS-S3 resource type allows an easy interaction with the underlying storage system. Research data can be directly uploaded to the S3 bucket through a variety of programs, e.g., rclone or minio. Moreover, for each RDS-S3 resource there are two access keys available with different permissions (writing and reading), thereby also allowing easy reuse of the data.

4.2 Coscine API

After resource creation and before uploading the research data, the associated metadata must be entered into the application profile through a form on the website, which supports the use of suitable metadata default values and editing a batch of files at once. This approach of metadata management is especially feasible for smaller data sets, but for working with large amounts of research data, we recommend using the Coscine API[2].

The API allows the use of all functions that are available on the website through scripts. To secure the access, a token is required, which can be created on the website. A token belongs personally to a unique user and allows the use of all functions that the user could access through the website. During creation, each token is assigned a time frame, in which it is valid. The maximum time frame is one year, thereby a regular revision of the access rights is ensured. Every token can be revoked at anytime, in case a token is no longer required or if it has been compromised.

The token can be used to interact with the API, which comes with an extensive documentation of all endpoints, parameters, and return values [22]. Swagger, an open-source tool set for API development, interaction and documentation [28], is used to allow the exploration and execution of example queries through a website. An option exists to create commands for every query that can be used to a create a custom script to upload the metadata. Through the detailed documentation and the possibility to copy snippets with working queries it is possible for users without a background in computer science using the API and automate parts of their workflow.

Existing research project have often already research data available that can be extracted from the environment or from some files that are stored along with the research data. With the tools described above, it is also possible to write a script that allows adding the locally available metadata to the files that are uploaded to Coscine.

4.3 Taskforce 'Coscine Technical Adaptation'

To support researchers with the technical adaptation of the RDM platform Coscine, a group of developers and data stewards has been established – the Coscine Technical Adaptation Group (CTA). The CTA is in direct contact with research groups from different disciplines. Its aim is at firstly understanding the researchers' workflows in order to provide scripts, programs, tools, and best practices for the interaction with the platform [23]. The provided material is publicly available under an open-source-license and researchers are encouraged to get involved with the development. Of course not every workflow can be generalized, however frequent exchange with the researchers allows a better understanding of the requirements and challenges for the adaptation of Coscine and improves the quality of RDM in the different research groups (e.g. automation of metadata collection).

5 Discussion

Coscine offers a technical environment to follow the 'FAIR Guiding Principles', however, the platform does not replace the need for subject-specific RDM knowledge – e.g., provided by

2. see https://docs.coscine.de/de/advanced/api/

data stewards employed in research projects. For example, the level of richness in metadata (reusability) is determined by the selection and completion of the application profile by the researchers. Furthermore, the link to domain-specific vocabularies and ontologies during the creation of application profiles depends on the expertise of the creating researchers.

5.1 Use Cases

As Coscine is a general service offering, most researchers are able to integrate Coscine into their day-to-day work without further assistance of the core team. Nevertheless, we can present some illustrative projects using the platform that happened to come to our knowledge due to the feedback of the respective data stewards.

Jan Rüth et. al. presented their formerly evolving dataset of incoming ICMP internet traffic [21]. Over a timespan of about a year, several gigabytes of daily log files were collected and made accessible to the public. Daily metadata to the files could include the current version of the application used. The daily datasets are stored in an S3 resource and are linked from the projects' website.

Thomas Hitch et. al. create a continuously growing collection of bacterial strains, isolated from the human gut [12]. Again data is stored in an S3 resource and annotated with various metadata fields describing the cultivation, isolation, and genome assembly which were previously stored in an SQLite database. Data can be accessed by a specially created web application that allows filtering for different aspects of metadata using the REST APIs of Coscine.

5.2 Comparison of Features

Looking back at the previously mentioned RDM platforms in subsection 1.1 there are some key differences that led to the realization of Coscine. However, it is to be noted that this aims to give a broad overview and not a rigorous review nor an exhaustive comparison. Alternative producs are compared in three rough categories: Research Oriented Databases, Electronic Laboratory Notebookss (ELNs), Knowledge Graphs, and Repositories.

5.2.1 Research Oriented Databases

This category includes discipline oriented databases and workflow management systems. Examples for these are FurthrMind[3], or idCarl[1] explicitly target specific disciplines. These systems usually support the direct recording of data and metadata within the database. As the contents are based on the often discipline specific schemas or formats, they usually do not provide means to support usage by multiple disciplines at the same time or require heavy customizations to do so. If at all, these applications only consider discipline specific standards rather than overarching standards e.g. those provided by the W3C.

3. see https://www.furthrmind.com/

5.2.2 Electronic Laboratory Notebooks

Almost like the Databases presented in the previous section ELNs are typically motivated by individual scientific disciplines e.g. eLabFTW [4] (chemistry), chemmotion [5] (chemistry), or labfolder[6] (life sciences). Nevertheless, they can often be applied across disciplines. They typically focus on recording individual steps along laboratory experiments. Tabular data can often be embedded or attached with other binaries. Most ELNs feature group and project structures and some have agreed on a common standard to exchange information[7]. ELNs, however, often lack ability to validate metadata according to semantic profiles and do not support storage of very large files.

5.2.3 Knowledge Graphs

Knowlegde Graphs offer a form of databases that allow interlinking of datasets within a database. Some like CaosDB[8] use proprietary formats, others like Semantic Media Wiki [9] are based on standard technologies like RDF. While the data model is also used for metadata stored in Coscine, Knowledge Graphs usually do not validate their contents based on application profiles and only link binaries from external sources. Furthermore, knowledge graphs typically only offer instance level access permissions and do not have a more fine-grained access management.

5.2.4 Repositories

The last regarded category are traditional data repositories like Radar[10], or Dataverse[11]. These applications usually offer instance wide metadata schemas or are even limited to bibliographic metadata. Further more repositories likely focus on finalized data sets and do not consider changing data after it was uploaded. This is surely required for publication but is a use case explicitly covered by Coscine allowing collaboration in early stages of research before data is ready for publishing.

5.3 Limitations

Coscine does not cover all steps of the data life cycle (Figure 1) completely – especially regarding the publication of research data. This is mainly due to the generic approach of Coscine, which contrasts with the recommended subject-specific publishing of data in established repositories. In addition, Coscine has been explicitly developed as an access point for so-called ‘warm’ research data, thereby deliberately allowing files behind a PID to be modified during the course of the project. Coscine is continuously improved in order to promote the publication of data: Currently a contact form is established to contact advisory services (e.g. libraries). This will enable researchers to share project metadata relevant for publication with the respective advisory centers.

4. see https://www.elabftw.net/
5. see https://chemotion.net/
6. see https://labfolder.com/
7. https://github.com/TheELNConsortium
8. see https://caosdb.org/
9. see https://www.semantic-mediawiki.org
10. see https://www.radar-service.eu/radar/de/home
11. see https://dataverse.org/

Moreover, the core development team of Coscine can not provide access to very specific service providers for single communities due to limited resources. However, since Coscine is being developed as an open-source platform, the addition of other community-specific resource types could also be realized by external development teams [12]. For contributions from external developers, the core development team monitors pull requests, has set up a publicly available issue tracker for discussions[13] and makes the strategic decision processes publicly available for discussion [14].

While the source code of Coscine is available to everyone under an open-source license, the application is built as a service offering. Much like OSF, there is currently little to no support by the maintainers for local installations. This is mostly due to dependencies and access requirements to administrative interfaces of PID services and storage providers that could require significant adaptations when transferring to a local installation. However, further development will likely go into the direction of a more federated service based on the FDO concept.

6 Conclusion

Coscine is a strong partner for researchers in their daily RDM: Thanks to the access to storage space, interfaces for automation as well as extensive collaboration possibilities, Coscine enables compliance with the 'FAIR Guiding Principles'. This spans from the very first storage of data by bundling raw data, metadata, interfaces and PIDs to a linked record according to the FDO concept. Coscine ensures that these data objects are also independently findable and accessible via the API. The API allows researchers to easily enter their data and metadata into the system and facilitates subsequent use of the same.

While the creation or adaptation of some kind RDM platform was inevitable, choosing to implement a new open-source service offering based on existing W3C standards was a bold step. It would likely not have been successful if the accompanying projects had not started at the same time. On the other hand, was the clear need for an implementation that picks up the semantic web technologies and makes them available to a broad user community. Apart from the implementation and operation work for the platform, sufficient work power needs to be available for data stewards, community management, and engagement in the further development of the standards taking place in various working groups in the NFDI, Research Data Alliance (RDA), W3C and several small independent working groups.

In addition, the API enables token-based authentication to automate workflows. Even for externally stored research data, Coscine allows increasing FAIRness by linking data with metadata and assigning PIDs. In this way, Coscine is a valuable contribution to the goal of NFDI4Ing: foster proper RDM in engineering sciences that implements the 'FAIR Guiding Principles'.

12. see https://git.rwth-aachen.de/coscine
13. see https://git.rwth-aachen.de/coscine/collaboration/issues/-/issues
14. see https://git.rwth-aachen.de/groups/coscine/-/epic_boards/539

7 Acknowledgements

The work was partially supported with resources granted by NFDI4Ing, funded by Deutsche Forschungsgemeinschaft (DFG) under project number 442146713 and NFDI-MatWerk, funded by Deutsche Forschungsgemeinschaft (DFG) under project number 460247524 and AIMS funded by Deutsche Forschungsgemeinschaft (DFG) under project number 432233186.

8 Roles and contributions

Ilona Lang: Conceptualization, Writing – original draft

Marcel Nellesen: Conceptualization, Writing – original draft

Marius Politze: Conceptualization,Writing – original draft, Supervision, Project administration

References

[1] Nima Babaei et al. In: *Materials Testing* 66.2 (2024), pp. 145–153. DOI: doi:10.1515/mt-2023-0262. URL: https://doi.org/10.1515/mt-2023-0262.

[2] Luiz Bonino. *FAIR Digital Object Framework Documentation*. Oct. 27, 2022. URL: https://fairdigitalobjectframework.org/ (visited on 01/25/2023).

[3] Luiz Bonino, Keez Burger, and Rajaram Kaliyaperumal, eds. *FAIR Data Point*. Aug. 26, 2022. URL: https://specs.fairdatapoint.org/ (visited on 01/25/2023).

[4] Richard Cyganiak, David Wood, and Markus Lanthaler, eds. *RDF 1.1 Concepts and Abstract Syntax*. W3C, 2014. (Visited on 02/20/2020).

[5] Deutsche Forschungsgemeinschaft (DFG). *Guidelines for Safeguarding Good Research Practice. Code of Conduct*. Bonn, Germany, 2019. URL: https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp_en.pdf.

[6] Christoph Eberl et al. "Consortium Proposal NFDI-MatWerk". In: (2021). DOI: 10.5281/ZENODO.5082837.

[7] Thomas Eifert, Florian Claus, and Ania Lopez. *Research Data Storage (RDS) : Verteilte Speicherinfrastruktur für Forschungsdatenmanagement : Gemeinsamer Antrag (öffentliche Fassung) im DFG-Programm "Großgeräte der Länder" : RWTH Aachen University (Konsortialführer), Fachhochschule Aachen, Ruhr-Universität Bochum, Technische Universität Dortmund, Universität Duisburg-Essen, Universität zu Köln*. de. Tech. rep. 2018. DOI: 10.18154/RWTH-2021-04541.

[8] Matthias Grönewald et al. "Mit AIMS zu einem Metadatenmanagement 4.0: FAIRe Forschungsdaten benötigen interoperable Metadaten". In: *E-Science-Tage 2021. Share Your Research Data*. Ed. by Vincent Heuveline and Nina Bisheh. Heidelberg, Germany: heiBOOKS, 2022. ISBN: 978-3-948083-54-0. DOI: 10.11588/heibooks.979.c13721.

[9] Richard Grunzke et al. “The MASi repository service. Comprehensive, metadata-driven and multi-community research data management”. In: *Future Generation Computer Systems* 94 (2019). PII: S0167739X17305344, pp. 879–894. ISSN: 0167-739X. DOI: 10.1016/j.future.2017.12.023.

[10] Laurel L. Haak et al. “ORCID: a system to uniquely identify researchers”. In: *Learned Publishing* 25.4 (2012), pp. 259–264. ISSN: 0953-1513. DOI: 10.1087/20120404.

[11] Benedikt Heinrichs, Marius Politze, and M. Amin Yazdi. “Evaluation of Architectures for FAIR Data Management in a Research Data Management Use Case”. In: *Proceedings of the 11th International Conference on Data Science, Technology and Applications (DATA 2022) / editors: Alfredo Cuzzocrea, Oleg Gusikhin, Wil van der Aalst and Slimane Hammoudi ; [sponsored by the Institute for Systems and Technologies of Information, Control and Communication (INSTICC)]*. 11. International Conference on Data Science, Technology and Applications, Lisbon (Portugal), 11 Jul 2022 - 13 Jul 2022. Setúbal: SCITEPRESS - Science and Technology Publications, July 11, 2022. DOI: 10.5220/0011302700003269.

[12] Thomas C. A. Hitch et al. *The Human Intestinal Bacterial Collection Website*. 2023. URL: https://hibc.otc.coscine.dev/ (visited on 06/22/2023).

[13] Florian Janetzko. “JARDS Ein Softwarewerkzeug zur Handhabung von Ressourcenvergabeprozessen”. In: *ZKI-AK Supercomputing Herbsttagung* (Berlin, Germany). Sept. 26, 2019. URL: https://juser.fz-juelich.de/record/868324.

[14] Tibor Kálmán, Daniel Kurzawe, and Ulrich Schwardmann. “European Persistent Identifier Consortium - PIDs für die Wissenschaft”. In: *Langzeitarchivierung von Forschungsdaten. Standards und disziplinspezifische Lösungen*. Ed. by Reinhard Altenhöner and Claudia Oellers. Berlin, Germany: Scivero Verl., 2012, pp. 151–164. ISBN: 978-3-944417-00-4.

[15] Christine Kirkpatrick. “Is FAIR FAIR? An Overview of FAIR Digital Objects”. In: *International Data Week 2022*. June 23, 2022. URL: https://fairdo.org/library/.

[16] Holger Knublauch and Dimitris Kontokostas, eds. *Shapes Constraint Language (SHACL)*. W3C, 2017. URL: https://www.w3.org/TR/shacl/ (visited on 06/10/2018).

[17] Angelina Kraft et al. “The RADAR Project – A Service for Research Data Archival and Publication”. In: *ISPRS International Journal of Geo-Information* 5.3 (2016), p. 28. ISSN: 2220-9964. DOI: 10.3390/ijgi5030028.

[18] Florian Krämer, Marius Politze, and Dominik Schmitz. *Empowering the Usage of Persistent Identifiers (PID) in Local Research Processes by Providing a Service and Integration Infrastructure*. In collab. with RD Alliance. Garching, Germany, 2016.

[19] Fadi Maali and John Erickson, eds. *Data Catalog Vocabulary (DCAT)*. W3C, 2014. URL: http://www.w3.org/TR/vocab-dcat/ (visited on 06/10/2018).

[20] Marius Politze, Sarah Bensberg, and Matthias S. Müller. “Managing Discipline-Specific Metadata Within an Integrated Research Data Management System”. In: *Proceedings of the 21st International Conference on Enterprise Information Systems*. 21st International Conference on Enterprise Information Systems (Heraklion, Crete, Greece). Ed. by Joaquim Filipe et al. SCITEPRESS - Science and Technology Publications, 2019, pp. 253–260. ISBN: 978-989-758-372-8. DOI: 10.5220/0007725002530260.

[21] Jan Rüth, Torsten Zimmermann, and Oliver Hohlfeld. “Hidden Treasures — Recycling Large-Scale Internet Measurements to Study the Internet’s Control Plane”. In: *Passive and Active Measurement*. Springer International Publishing, 2019, pp. 51–67. ISBN: 978-3-030-15986-3. DOI: 10.1007/978-3-030-15986-3_4.

[22] IT Center of the RWTH Aachen University. *Coscine API Documentation*. Feb. 14, 2023. URL: https://docs.coscine.de/de/advanced/api/.

[23] IT Center of the RWTH Aachen University. *Coscine Technical Adaption Project*. Feb. 14, 2023. URL: https://coscine.pages.rwth-aachen.de/community-features/coscine-technical-adaption/.

[24] Robert H. Schmitt et al. *NFDI4Ing - the National Research Data Infrastructure for Engineering Sciences*. 2020. DOI: 10.5281/zenodo.4015201.

[25] Frederik Schulze Spüntrup, Frederik Braun, and Neslihan Ana Sönmez. *Action, bitte! Wie der öffentliche Sektor den Mangel an digitalen Fachkräften meistern kann*. Jan. 1, 2023. URL: https://www.mckinsey.de/~/media/mckinsey/locations/europe%20and%20middle%20east/deutschland/publikationen/2023-01-25%20it%20talent%20im%20public%20sector/action%20bittemckinsey.pdf.

[26] Luiz Olavo Bonino da Silva Santos et al. “FAIR Data Point: A FAIR-Oriented Approach for Metadata Publication”. In: *Data Intelligence* (Feb. 2023), pp. 1–21. ISSN: 2641-435X. DOI: 10.1162/dint_a_00160. eprint: https://direct.mit.edu/dint/article-pdf/doi/10.1162/dint_a_00160/2070149/dint_a_00160.pdf. URL: https://doi.org/10.1162/dint%5C_a%5C_00160.

[27] Koenraad De Smedt, Dimitris Koureas, and Peter Wittenburg. “FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units”. In: *Publications* 8.2 (Apr. 2020), p. 21. DOI: 10.3390/publications8020021. URL: https://doi.org/10.3390/publications8020021.

[28] SmartBear Software. *Swagger*. Feb. 16, 2023. URL: https://swagger.io/.

[29] *SPARQL 1.1 Overview*. W3C, 2013. URL: https://www.w3.org/TR/sparql11-overview/ (visited on 06/10/2018).

[30] Steve Speicher, John Arwe, and Ashok Malhotra, eds. *Linked Data Platform 1.0*. W3C, Feb. 26, 2015. URL: https://www.w3.org/TR/ldp/ (visited on 01/25/2023).

[31] GO FAIR International Support and Coordination Office (GFISCO). *FAIR Principles*. Feb. 10, 2023. URL: https://www.go-fair.org/fair-principles/.

[32] Raimund Vogl et al. “”sciebo — theCampuscloud” for NRW”. In: *Proceedings of the 21st EUNIS Congress*. Ed. by Michael Turpie. Dundee, Scotland, 2015.

[33] Tobias Weigel et al. “RDA Recommendation on PID Kernel Information”. en. In: (2018). DOI: 10.15497/RDA00031. URL: https://www.rd-alliance.org/group/pid-kernel-information-wg/outcomes/recommendation-pid-kernel-information.

[34] Mark D. Wilkinson et al. “The FAIR Guiding Principles for scientific data management and stewardship”. eng. In: *Scientific data* 3 (2016). Journal Article. DOI: 10.1038/sdata.2016.18. eprint: 26978244.