


# RDM Platform Coscine – FAIR play integrated right from the start

Ilona Lang <sup>1</sup>Marcel Nellesen <sup>1</sup>Marius Politze <sup>1</sup>

1. IT Center, RWTH Aachen University, Aachen.

**Date Received:**

2022-02-17

**Licenses:**This article is licensed under: **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](https://git.rwth-aachen.de/coscine)**Software availability:**

Software can be found here:

[coscine.de/](https://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

17 store and manage data, however, these options usually lack in support of RDM workflows or  
18 policies.

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

## 23 1.2 Goals & Requirements

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

- 29 1. develop a data management system that replaces all existing services or
- 30 2. develop a data management system that adds a ‘FAIR’ layer to already established services.

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

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

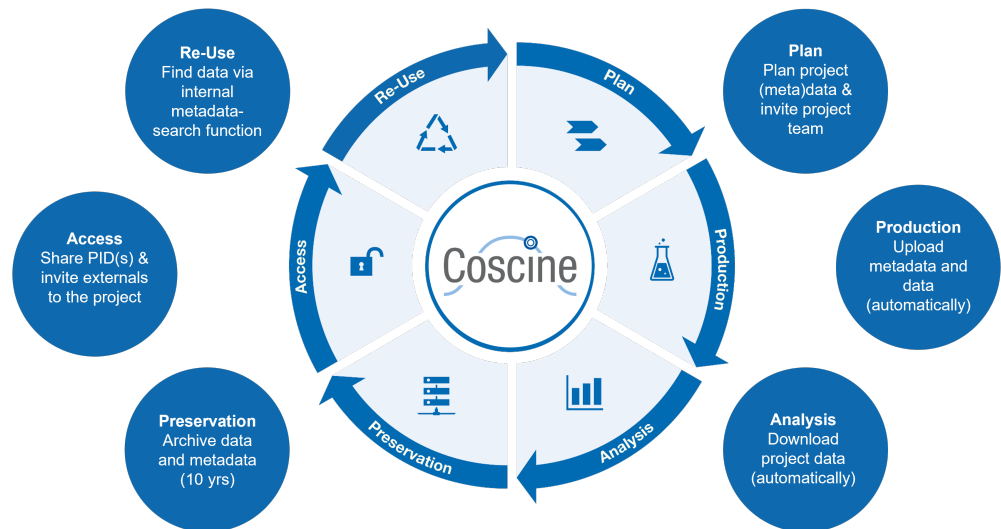
40 To create such a data management system that supports researchers during their whole data  
41 life cycle, the RDM platform Coscine was developed at the IT Center of the RWTH Aachen  
42 University (Figure 1). Since 2020, the development is further supported by two consortia of the  
43 National Research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

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

## 49 2 Core Features of Coscine

50 Coscine is a platform for the management, storage and archiving of research (meta)data gener-  
51 ated 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.

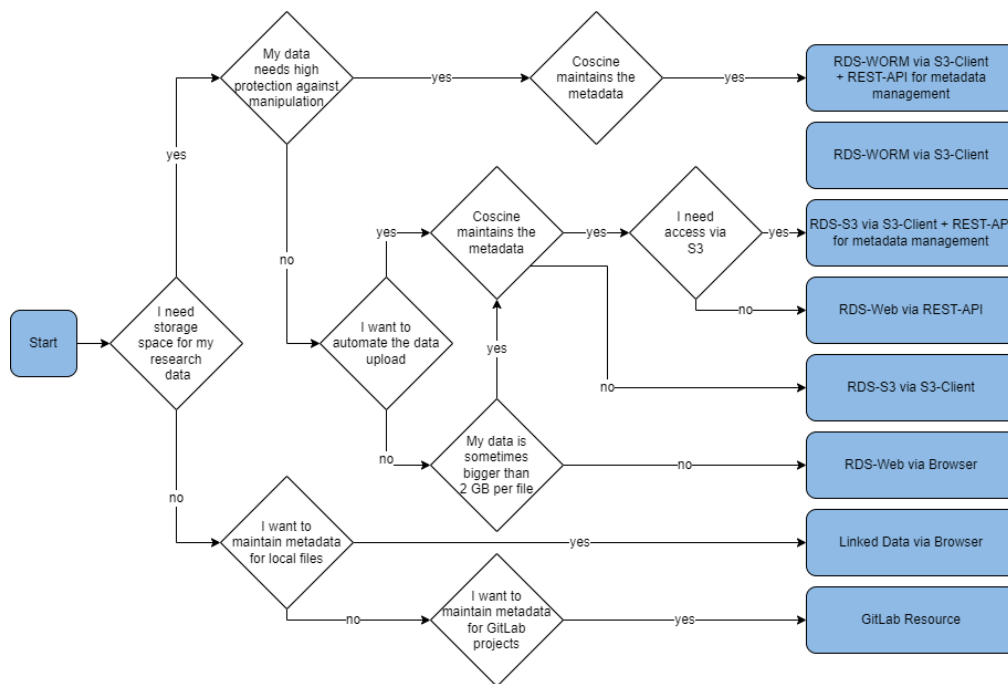
52 participants, integrating the project-related data from different resources and adding the related  
 53 metadata (Figure 5). Specifically, Coscine offers researchers the following core features:

#### 54 2.1 Integration

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

#### 61 2.2 Storage Space

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



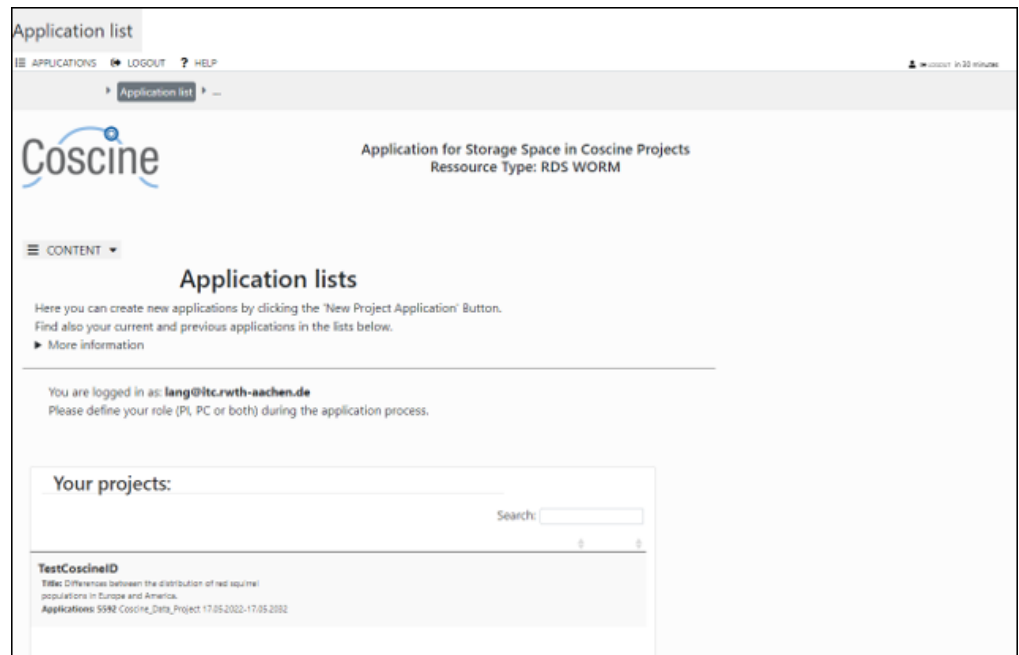
**Figure 2:** Resource Types in Coscine. To date, there are three different resource types in Coscine: RDS (subtypes: Web, Simple Storage Service (S3), write once, read many (WORM)), GitLab, and Linked Data. The decision diagram helps to select the right resource type based on different project needs.

72 prevent subsequent manipulation of the data (Figure 2).

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

83 **2.3 Collaboration**

84 Coscine allows access for all internal and external members of a research project. Users can log  
 85 in as a member of a participating organization via Shibboleth or as an external person via their  
 86 Open Researcher and Contributor ID [10] (ORCID). While the ability to request certain storage  
 87 services may be restricted, once added to a project the resource is available for all member. Basic  
 88 functionalities like project and metadata management are available to all users. Project members  
 89 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

## 90 2.4 Metadata

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

97 Individual application profiles can be created using the integrated application profile generator,  
98 developed within the DFG-funded project Applying Interoperable Metadata Standards (AIMS)  
99 [8]. This application profile generator allows researchers to create new application profiles from  
100 scratch or explore and extend already existing ones (Figure 4). New profiles can be sent as a  
101 merge request to the GitLab repository of Coscine, where they are reviewed by RDM experts to  
102 ensure a required level of technical quality and interoperability for Coscine.

## 103 2.5 Archiving

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

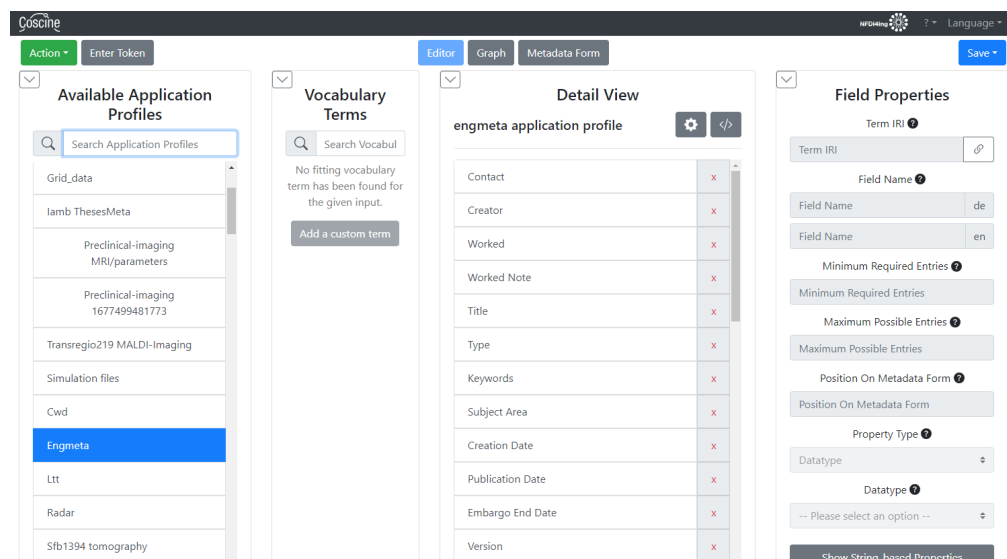


Figure 4: Screenshot of the application profile generator developed within AIMS [8].

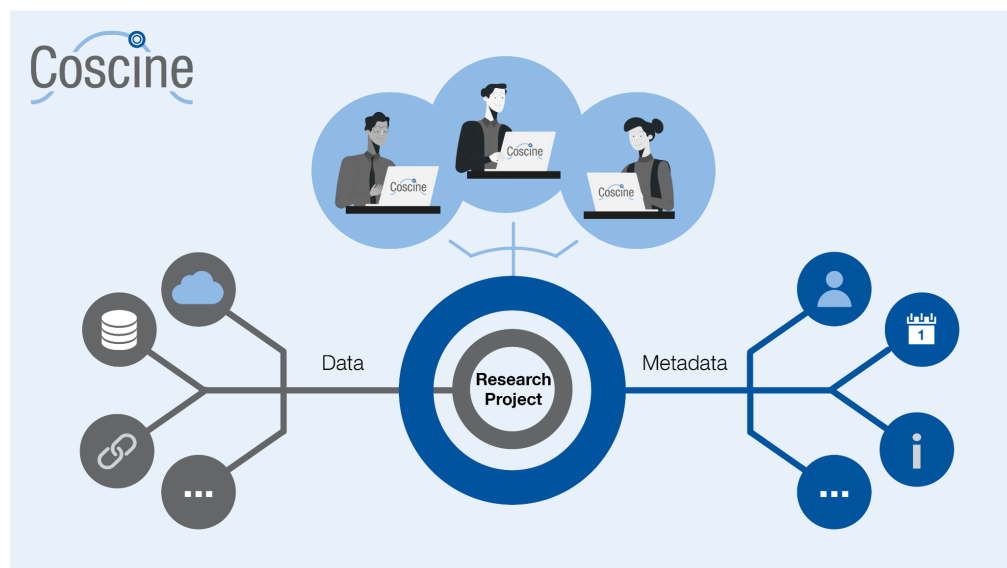


Figure 5: The project structure of Coscine. For each research project, researchers can invite all project participants (above – light blue circles), integrate the project-related data from different resources (left side – gray circles) and add the related metadata (right side – blue circles).

### 108 3 Coscine & 'FAIR Guiding Principles'

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

#### 117 3.1 Metadata Representation

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

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

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

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

146 Regarding the FAIR principles F1 and A1 [31], Coscine assigns for each resource (including  
147 data and metadata) a handle-based ePIC-PID [14, 18]. This is used to uniquely and permanently

148 identify the location of the resource and all contained files on a global level. As a result, each  
149 RDF-triple includes a PID leading to the data it describes. Within resources, fragment identifiers  
150 are used to address individual files by extending the handle URL.

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

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

### 164 3.2 Coscine & FAIR Digital Objects

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

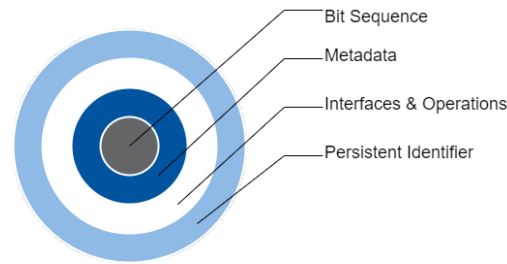
168 Although the concept of Digital Object (DO) was introduced by Robert Kahn in the early 1990s,  
169 an ecosystem of easy tools that add the FDO layers to raw data including unique identifiers and  
170 metadata is still needed [15]. This issue is most prominent in current industry grade IT solutions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

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

179 For retaining the bit sequence of the FDO Coscine relies mostly on a background storage system.  
180 In the case of the RDS the provided HTTP based S3 interface can be directly handed through to  
181 the client. For storage service that do not provide an HTTP accessible interface or in cases where  
182 access management is required, Coscine provides means for protocol translation. Coscine aims to  
183 combine approaches from two frameworks: PIDs based on Kernel Information Records (KIRs)  
184 [33] and the semantic approach of the FAIR Digital Object Framework [2] (FDOF).

185 On the one hand, the KIR work “by injecting a tiny amount of carefully selected metadata into a  
186 [PID] record” [33]. While the metadata set is typically small and rather technical key-value-pairs,  
187 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].

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

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

## 201 4 Coscine – Options for Process Automation

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

### 208 4.1 Data Upload

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

## 216 4.2 Coscine API

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

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

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

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

## 239 4.3 Taskforce ‘Coscine Technical Adaptation’

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

## 250 5 Discussion

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

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

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

## 257 5.1 Use Cases

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

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

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

## 272 5.2 Comparison of Features

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

### 278 5.2.1 Research Oriented Databases

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

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

### 286 5.2.2 Electronic Laboratory Notebooks

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

### 295 5.2.3 Knowledge Graphs

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

### 302 5.2.4 Repositories

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

## 309 5.3 Limitations

310 Coscine does not cover all steps of the data life cycle (Figure 1) completely – especially regarding  
311 the publication of research data. This is mainly due to the generic approach of Coscine, which  
312 contrasts with the recommended subject-specific publishing of data in established repositories.  
313 In addition, Coscine has been explicitly developed as an access point for so-called ‘warm’  
314 research data, thereby deliberately allowing files behind a PID to be modified during the course  
315 of the project. Coscine is continuously improved in order to promote the publication of data:  
316 Currently a contact form is established to contact advisory services (e.g. libraries). This will  
317 enable researchers to share project metadata relevant for publication with the respective advisory  
318 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/>

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

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

## 332 6 Conclusion

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

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

349 In addition, the API enables token-based authentication to automate workflows. Even for  
350 externally stored research data, Coscine allows increasing FAIRness by linking data with metadata  
351 and assigning PIDs. In this way, Coscine is a valuable contribution to the goal of NFDI4Ing:  
352 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](https://git.rwth-aachen.de/groups/coscine/-/epic_boards/539)

## 353 7 Acknowledgements

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

## 358 8 Roles and contributions

359 **Ilona Lang:** Conceptualization, Writing – original draft

360 **Marcel Nellesen:** Conceptualization, Writing – original draft

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

## 362 References

- 363 [1] Nima Babaei et al. In: *Materials Testing* 66.2 (2024), pp. 145–153. DOI: [doi:10.1515](https://doi.org/10.1515/mt-2023-0262)  
 364 [/mt-2023-0262](https://doi.org/10.1515/mt-2023-0262). URL: <https://doi.org/10.1515/mt-2023-0262>.
- 365 [2] Luiz Bonino. *FAIR Digital Object Framework Documentation*. Oct. 27, 2022. URL:  
 366 <https://fairdigitalobjectframework.org/> (visited on 01/25/2023).
- 367 [3] Luiz Bonino, Keez Burger, and Rajaram Kaliyaperumal, eds. *FAIR Data Point*. Aug. 26,  
 368 2022. URL: <https://specs.fairdatapoint.org/> (visited on 01/25/2023).
- 369 [4] Richard Cyganiak, David Wood, and Markus Lanthaler, eds. *RDF 1.1 Concepts and*  
 370 *Abstract Syntax*. W3C, 2014. (Visited on 02/20/2020).
- 371 [5] Deutsche Forschungsgemeinschaft (DFG). *Guidelines for Safeguarding Good Research*  
 372 *Practice. Code of Conduct*. Bonn, Germany, 2019. URL: [https://www.dfg.de/down](https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp_en.pdf)  
 373 [load/pdf/foerderung/rechtliche\\_rahmenbedingungen/gute\\_wissenschaftli](https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp_en.pdf)  
 374 [che\\_praxis/kodex\\_gwp\\_en.pdf](https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp_en.pdf).
- 375 [6] Christoph Eberl et al. “Consortium Proposal NFDI-MatWerk”. In: (2021). DOI: [10.5281](https://doi.org/10.5281/ZENODO.5082837)  
 376 [/ZENODO.5082837](https://doi.org/10.5281/ZENODO.5082837).
- 377 [7] Thomas Eifert, Florian Claus, and Ania Lopez. *Research Data Storage (RDS) : Verteilte*  
 378 *Speicherinfrastruktur für Forschungsdatenmanagement : Gemeinsamer Antrag (öffentliche*  
 379 *Fassung) im DFG-Programm ”Großgeräte der Länder” : RWTH Aachen University (Kon-*  
 380 *sortialführer), Fachhochschule Aachen, Ruhr-Universität Bochum, Technische Universität*  
 381 *Dortmund, Universität Duisburg-Essen, Universität zu Köln*. de. Tech. rep. 2018. DOI:  
 382 [10.18154/RWTH-2021-04541](https://doi.org/10.18154/RWTH-2021-04541).
- 383 [8] Matthias Grönewald et al. “Mit AIMS zu einem Metadatenmanagement 4.0: FAIRe
- 384 Forschungsdaten benötigen interoperable Metadaten”. In: *E-Science-Tage 2021. Share*  
 385 *Your Research Data*. Ed. by Vincent Heuveline and Nina Bisheh. Heidelberg, Germany:  
 386 heiBOOKS, 2022. ISBN: 978-3-948083-54-0. DOI: [10.11588/heibooks.979.c13721](https://doi.org/10.11588/heibooks.979.c13721).

- 387 [9] Richard Grunzke et al. “The MASi repository service. Comprehensive, metadata-driven  
388 and multi-community research data management”. In: *Future Generation Computer*  
389 *Systems* 94 (2019). PII: S0167739X17305344, pp. 879–894. ISSN: 0167-739X. DOI:  
390 [10.1016/j.future.2017.12.023](https://doi.org/10.1016/j.future.2017.12.023).
- 391 [10] Laurel L. Haak et al. “ORCID: a system to uniquely identify researchers”. In: *Learned*  
392 *Publishing* 25.4 (2012), pp. 259–264. ISSN: 0953-1513. DOI: [10.1087/20120404](https://doi.org/10.1087/20120404).
- 393 [11] Benedikt Heinrichs, Marius Politze, and M. Amin Yazdi. “Evaluation of Architectures for  
394 FAIR Data Management in a Research Data Management Use Case”. In: *Proceedings*  
395 *of the 11th International Conference on Data Science, Technology and Applications*  
396 *(DATA 2022) / editors: Alfredo Cuzzocrea, Oleg Gusikhin, Wil van der Aalst and Slimane*  
397 *Hammoudi ; [sponsored by the Institute for Systems and Technologies of Information,*  
398 *Control and Communication (INSTICC)]. 11. International Conference on Data Science,*  
399 *Technology and Applications, Lisbon (Portugal), 11 Jul 2022 - 13 Jul 2022. Setúbal:*  
400 *SCITEPRESS - Science and Technology Publications, July 11, 2022. DOI: [10.5220/00](https://doi.org/10.5220/0011302700003269)*  
401 *[11302700003269](https://doi.org/10.5220/0011302700003269).*
- 402 [12] Thomas C. A. Hitch et al. *The Human Intestinal Bacterial Collection Website*. 2023. URL:  
403 <https://hibc.otc.coscine.dev/> (visited on 06/22/2023).
- 404 [13] Florian Janetzko. “JARDS Ein Softwarewerkzeug zur Handhabung von Ressourcenver-  
405 gabeprozessen”. In: *ZKI-AK Supercomputing Herbsttagung* (Berlin, Germany). Sept. 26,  
406 2019. URL: <https://juser.fz-juelich.de/record/868324>.
- 407 [14] Tibor Kálmán, Daniel Kurzawe, and Ulrich Schwarzmann. “European Persistent Identifier  
408 Consortium - PIDs für die Wissenschaft”. In: *Langzeitarchivierung von Forschungsdaten.*  
409 *Standards und disziplinspezifische Lösungen*. Ed. by Reinhard Altenhöner and Claudia  
410 Oellers. Berlin, Germany: Scivero Verl., 2012, pp. 151–164. ISBN: 978-3-944417-00-4.
- 411 [15] Christine Kirkpatrick. “Is FAIR FAIR? An Overview of FAIR Digital Objects”. In: *Inter-*  
412 *national Data Week 2022*. June 23, 2022. URL: <https://fairdo.org/library/>.
- 413 [16] Holger Knublauch and Dimitris Kontokostas, eds. *Shapes Constraint Language (SHACL).*  
414 *W3C*, 2017. URL: <https://www.w3.org/TR/shacl/> (visited on 06/10/2018).
- 415 [17] Angelina Kraft et al. “The RADAR Project – A Service for Research Data Archival and  
416 Publication”. In: *ISPRS International Journal of Geo-Information* 5.3 (2016), p. 28. ISSN:  
417 2220-9964. DOI: [10.3390/ijgi5030028](https://doi.org/10.3390/ijgi5030028).
- 418 [18] Florian Krämer, Marius Politze, and Dominik Schmitz. *Empowering the Usage of Persis-*  
419 *tent Identifiers (PID) in Local Research Processes by Providing a Service and Integration*  
420 *Infrastructure*. In collab. with RD Alliance. Garching, Germany, 2016.
- 421 [19] Fadi Maali and John Erickson, eds. *Data Catalog Vocabulary (DCAT)*. W3C, 2014. URL:  
422 <http://www.w3.org/TR/vocab-dcat/> (visited on 06/10/2018).
- 423 [20] Marius Politze, Sarah Bensberg, and Matthias S. Müller. “Managing Discipline-Specific  
424 Metadata Within an Integrated Research Data Management System”. In: *Proceedings of*  
425 *the 21st International Conference on Enterprise Information Systems*. 21st International  
426 *Conference on Enterprise Information Systems (Heraklion, Crete, Greece)*. Ed. by Joaquim  
427 *Filipe et al. SCITEPRESS - Science and Technology Publications, 2019, pp. 253–260.*  
428 *ISBN: 978-989-758-372-8. DOI: [10.5220/0007725002530260](https://doi.org/10.5220/0007725002530260).*

- 429 [21] Jan R uth, Torsten Zimmermann, and Oliver Hohlfeld. “Hidden Treasures — Recycling  
430 Large-Scale Internet Measurements to Study the Internet’s Control Plane”. In: *Passive  
431 and Active Measurement*. Springer International Publishing, 2019, pp. 51–67. ISBN:  
432 978-3-030-15986-3. DOI: [10.1007/978-3-030-15986-3\\_4](https://doi.org/10.1007/978-3-030-15986-3_4).
- 433 [22] IT Center of the RWTH Aachen University. *Coscine API Documentation*. Feb. 14, 2023.  
434 URL: <https://docs.coscine.de/de/advanced/api/>.
- 435 [23] IT Center of the RWTH Aachen University. *Coscine Technical Adaption Project*. Feb. 14,  
436 2023. URL: <https://coscine.pages.rwth-aachen.de/community-features/coscine-technical-adaption/>.
- 438 [24] Robert H. Schmitt et al. *NFDI4Ing - the National Research Data Infrastructure for  
439 Engineering Sciences*. 2020. DOI: [10.5281/zenodo.4015201](https://doi.org/10.5281/zenodo.4015201).
- 440 [25] Frederik Schulze Sp untrup, Frederik Braun, and Neslihan Ana S onmez. *Action, bitte! Wie  
441 der  ffentliche Sektor den Mangel an digitalen Fachkr ften meistern kann*. Jan. 1, 2023.  
442 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>.
- 445 [26] Luiz Olavo Bonino da Silva Santos et al. “FAIR Data Point: A FAIR-Oriented Approach  
446 for Metadata Publication”. In: *Data Intelligence* (Feb. 2023), pp. 1–21. ISSN: 2641-435X.  
447 DOI: [10.1162/dint\\_a\\_00160](https://doi.org/10.1162/dint_a_00160). eprint: [https://direct.mit.edu/dint/article-pdf/doi/10.1162/dint\\_a\\_00160/2070149/dint\\_a\\_00160.pdf](https://direct.mit.edu/dint/article-pdf/doi/10.1162/dint_a_00160/2070149/dint_a_00160.pdf). URL:  
448 [https://doi.org/10.1162/dint\\_a\\_00160](https://doi.org/10.1162/dint_a_00160). URL:  
449 [https://doi.org/10.1162/dint\\_a\\_00160](https://doi.org/10.1162/dint_a_00160).
- 450 [27] Koenraad De Smedt, Dimitris Koureas, and Peter Wittenburg. “FAIR Digital Objects for  
451 Science: From Data Pieces to Actionable Knowledge Units”. In: *Publications 8.2* (Apr.  
452 2020), p. 21. DOI: [10.3390/publications8020021](https://doi.org/10.3390/publications8020021). URL: <https://doi.org/10.3390/publications8020021>.
- 454 [28] SmartBear Software. *Swagger*. Feb. 16, 2023. URL: <https://swagger.io/>.
- 455 [29] *SPARQL 1.1 Overview*. W3C, 2013. URL: <https://www.w3.org/TR/sparql11-overview/>  
456 (visited on 06/10/2018).
- 457 [30] Steve Speicher, John Arwe, and Ashok Malhotra, eds. *Linked Data Platform 1.0*. W3C,  
458 Feb. 26, 2015. URL: <https://www.w3.org/TR/ldp/> (visited on 01/25/2023).
- 459 [31] GO FAIR International Support and Coordination Office (GFISCO). *FAIR Principles*.  
460 Feb. 10, 2023. URL: <https://www.go-fair.org/fair-principles/>.
- 461 [32] Raimund Vogl et al. ““sciebo — theCampuscloud” for NRW”. In: *Proceedings of the 21st  
462 EUNIS Congress*. Ed. by Michael Turpie. Dundee, Scotland, 2015.
- 463 [33] Tobias Weigel et al. “RDA Recommendation on PID Kernel Information”. en. In: (2018).  
464 DOI: [10.15497/RDA00031](https://doi.org/10.15497/RDA00031). URL: <https://www.rd-alliance.org/group/pid-kernel-information-wg/outcomes/recommendation-pid-kernel-information>.
- 466 [34] Mark D. Wilkinson et al. “The FAIR Guiding Principles for scientific data management  
467 and stewardship”. eng. In: *Scientific data 3* (2016). Journal Article. DOI: [10.1038/sdata  
468 a.2016.18](https://doi.org/10.1038/sdata.2016.18). eprint: [26978244](https://arxiv.org/abs/26978244).