


# Towards Improved Findability of Energy Research Software by Introducing a Metadata-based Registry

Stephan Ferez <sup>1</sup>Astrid Nieße <sup>1</sup>

1. Department of Computer Science, Carl von Ossietzky University of Oldenburg, Germany.

**Date Received:**

2023-03-01

**Licenses:**This article is licensed under: **Keywords:**

Interoperability, Digital Libraries, Energy Research, FAIR, Research Software, Metadata, Open Source Software, Software Reusability

**Data availability:**

No data used for this article

**Software availability:**

No software used for this article

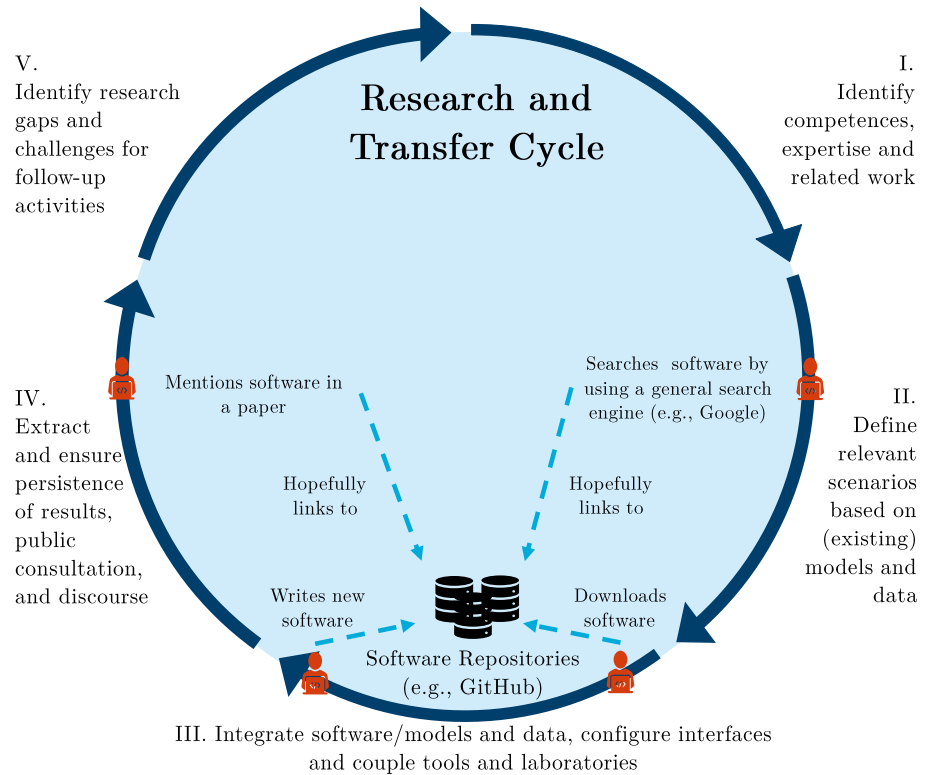
**Abstract.** Research software in the energy domain becomes increasingly important for the analysis, simulation, and optimization of energy systems and supports design decisions in the required transition of energy systems to tackle the climate crisis. To make energy research software (ERS) more findable, it should be described with metadata following the FAIR (findable, accessible, interoperable, and reusable) principles and be registered in a common registry. To this end, we motivate and present a concept for a metadata-based registry for ERS which should enable researchers to easily add new ERS as well as to find new ERS.

## 1 Motivation

1 In energy research, self-designed software is a basic tool for multiple purposes like visualization  
2 of processes and values, e.g., power quality [1], (co-)simulation of smart grids [2], or analysis of  
3 transition paths [3]. Within an exemplified research cycle, this self-designed software is often  
4 a starting point and, therefore, fundamental for producing new research results while it also  
5 presents a result of performed research (see [Figure 1](#)).  
6

7 Based on the definition of research software by Hasselbring et al. [4], we define energy research  
8 software (ERS) as software employed in the scientific discovery process to understand, analyze,  
9 improve, and/or design energy systems. With respect to the complexity of the software, ERS  
10 ranges from simple scripts, libraries, e.g., for python, and frameworks up to full software solutions.  
11 Content-wise, it can for example visualize, analyze, and/or generate (artificial) data from energy  
12 (sub)components or grids in laboratories or the real world. Alternatively, it can also represent as  
13 a model particular energy (sub)components, energy (distribution) systems, and transition paths  
14 of energy use, distribution, conversion, and/or generation to analyze the design and/or control in  
15 simulations and optimizations.

16 In the energy domain, multiple models and frameworks have been developed with partly overlap-  
17 ping and similar features. New ERS is often developed without reusing existing ones. Therefore,  
18 a lot of time is spent on (re)developing software instead of doing research slowing down the  
19 progress in research. Due to the need for interdisciplinary research in the energy domain and the  
20 growing number of simulated components, ERS will even become more complex in the next  
21 years [5]. Building these complicated ERS can be simplified by reusing existing ERS.



**Figure 1:** An exemplified research cycle of energy research with software usage

22 Different approaches to formulate FAIR criteria for research software show that metadata and  
 23 repositories for these metadata, e.g., software registries, are key elements for FAIR research  
 24 software [4], [6]–[8]. Software registries only store metadata for software without the source  
 25 code of the software. In contrast, software repositories also contain source code [9].

26 Especially the findability of ERS can be increased by describing it with useful metadata and  
 27 including it into a registry. Table 1 gives some examples for possible metadata elements with  
 28 metadata for the co-simulation framework mosaik<sup>1</sup> as example instance. Finding existing ERS  
 29 is import to enable the reuse of ERS. Therefore, good metadata and a registry are a first step for  
 30 increasing the reuse of ERS and improving the research process in energy research.

31 For this goal, we propose a concept for a metadata-based registry for ERS based on a good  
 32 metadata scheme<sup>2</sup>. Our contributions are:

- 33 • We outline the state of the art in the fields of metadata schemes for research software,  
 34 registries and repositories for research software, metadata schemes in the field of energy  
 35 research, and ontologies in the energy domain in Section 2.
- 36 • We introduce our concept of metadata-based registry for ERS in Section 3.
- 37 • We give an outlook of the further required work in Section 4.

1. <http://mosaik.offis.de/>, accessed 12.12.2022

2. Hereby, we extend our poster abstract on the same topic [10] and, therefore, reuse a graphic and text of our previous work.

**Table 1:** Examples for possible metadata elements based on the FAIR criteria with the software mosaik as example instance

Findability	Name	mosaik <sup>1</sup>
	Identifier	-
	Version	3.0
Accessibility	Link to repository	<a href="https://gitlab.com/mosaik/mosaik">https://gitlab.com/mosaik/mosaik</a>
Interoperability	Programming language	python
	Input/output formats	depends on used models
	Dependencies	-
Reusability	License	LGPL v3.0
	Link to documentation	<a href="https://mosaik.readthedocs.io/en/latest/quickstart.html">https://mosaik.readthedocs.io/en/latest/quickstart.html</a>

## 38 2 State of the art

39 Within this section, we give an overview on existing metadata schemes. First, we present current  
 40 metadata schemes for research software in general or in specific domains in [Section 2.1](#). Second,  
 41 we give an overview on existing registries and repositories for research software in different  
 42 research domains in [Section 2.2](#). Afterwards, we focus on metadata schemes in the field of  
 43 energy research and engineering in [Section 2.3](#). These are often designed for research data but  
 44 can be used as foundation for a metadata scheme for ERS.

45 Since value vocabularies are an important part of metadata schemes, the forth part ([Section 2.4](#))  
 46 outlines ontologies in the energy domain.

### 47 2.1 Metadata for Research Software

48 This section gives an overview of existing approaches for metadata schemes for research software.  
 49 While some publications directly focus on metadata, others only introduce software ontologies,  
 50 which can be used as metadata vocabulary for research software. The approaches focus on  
 51 different domains. They use the possibilities of metadata schemes in a different extent (e.g.,  
 52 mandatory elements, support for URIs, and the use of ontologies as value vocabularies) and,  
 53 therefore, follow the formal rules for metadata differently.

54 *CodeMeta*<sup>3</sup> is a community driven metadata standard for research software based on [schema.org](#)<sup>4</sup>.  
 55 Various crosswalks to other metadata schemes already exist. *CodeMeta* contains multiple  
 56 elements, some focusing on technical details like file size or supported operating systems and  
 57 others including administrative information like license. The metadata standard does not have  
 58 mandatory elements. It supports the use of Unique Resource Identifiers (URIs) for authors and  
 59 contributors as well as for licenses. The content specific metadata are limited to an application  
 60 category and keywords.

61 For geosciences, Gil et al. introduced an ontology to describe research software, *OntoSoft*,  
 62 with six categories: identify, understand, execute, do research, get support, and update [11].  
 63 They also developed an automated extraction tool for metadata and provided a registry for

3. <https://codemeta.github.io/>, accessed 12.12.2022

4. <https://schema.org/>, accessed 12.12.2022

**Table 2:** Overview of metadata scheme for research software

	Metadata scheme (s) vs. ontology (o)	Domain	Mandatory elements	Support for URIs	Use of domain ontologies as value vocabulary
<i>CodeMeta</i> <sup>3</sup>	s	General	✗	(✓)	✗
<i>OntoSoft</i> [11]	o	Geoscience	✓	✗	✗
<i>Software Description Ontology</i> [12]	o	Geoscience	✗	✓	✓
<i>Software Ontology</i> [13]	o	Bioinformatics	✗	✓	✓
<i>biotoolsXSD</i> [14]	s	Bioinformatics	✓	(✓)	✓

✓: fulfilled                      ✗: not fulfilled  
 (✓): (partly) fulfilled

64 software metadata<sup>5</sup>. Garijo et al. expanded this approach by developing the *Software Description*  
 65 *Ontology*<sup>6</sup> [12] with additional description for input and output data based on the *Scientific*  
 66 *Variables Ontology*<sup>7</sup>. Also, they aligned their approach with *CodeMeta* and enabled publishing  
 67 the metadata into an open knowledge graph including links to additional instances in the semantic  
 68 web like wikidata<sup>8</sup>. Additionally, they developed software to support researchers in the metadata  
 69 creation and to find models<sup>9</sup>.

70 The *Software Ontology (SWO)* was developed extending the bioinformatics *EDAM* ontology to  
 71 describe software in this research field [13]. The *SWO* includes licenses, programming languages,  
 72 and data formats as taxonomies. In contrast to *OntoSoft*, the use of the taxonomies improves the  
 73 usability for semantic web applications and linking. Also for bioinformatics, Ison et al. developed  
 74 the metadata scheme *biotoolsXSD* for the software registry bio.tools<sup>10</sup> [14]. The metadata scheme  
 75 is expressed as an XML scheme containing 55 elements of which 10 are mandatory. The use  
 76 of the *EDAM* ontology as value vocabulary is required for some elements like function, input,  
 77 and output. The metadata scheme also contains software specific elements like programming  
 78 language, license, and operating system for which the use of an ontology is not required.

79 Table 2 gives an overview of the diverse outlined metadata schemes and ontologies. While some  
 80 metadata schemes are less extensive, like *CodeMeta*<sup>3</sup>, others try to include value vocabularies to  
 81 improve interoperability by using semantic web technologies. Some schemes include detailed  
 82 domain knowledge based on domain ontologies like *biotoolsXSD* [14] or the *Software Description*  
 83 *Ontology* [12].

## 84 2.2 Registries and Repositories for Research Software

85 This section gives an overview of existing registries and repositories for research software. As  
 86 defined in Section 1, registries only include metadata while in repositories source code and

5. <https://www.ontosoft.org/portal/>, accessed 12.12.2022

6. <https://w3id.org/okn/o/sd>, accessed 12.12.2022

7. <https://scientificvariablesontology.org/svo/>, accessed 12.12.2022

8. [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page), accessed 12.12.2022

9. <http://models.mint.isi.edu>, accessed 12.12.2022

10. <http://bio.tools>, accessed 12.12.2022

87 metadata is stored.

88 Multiple research software repositories and registries are organized in SciCodes (Consortium of  
89 scientific software registries and repositories)<sup>11</sup> which provides an overview their properties<sup>12</sup>.  
90 In the following, we give an overview on some of these repositories and registries as well as  
91 introducing additional ones. The repositories or registries use different metadata schema as a  
92 foundation and allow different types of research artifacts in a different extent (e.g., only software  
93 or more artifacts). We also introduce, if they focus on research and/or certain domains.

94 **GitHub**<sup>13</sup> has become one of the most popular software repositories [15] for all kinds of source  
95 code including research software. It is a version control system and includes multiple metadata  
96 which do not follow any standard but are accessible via an API.

97 **GitLab**<sup>14</sup> is also a software repository based on a version control system. It includes all kind of  
98 software including research software. It collects multiple metadata for software not based on  
99 any standard but the metadata is accessible via an API.

100 **Zenodo**<sup>15</sup> is a research repository for all kind of research artifacts like data, presentations, and  
101 software [16]. Zenodo supports a general metadata standard derived from Dublin core [17]. It  
102 can be exported to DataCite and Dublin Core.<sup>16</sup>

103 **Software Heritage**<sup>17</sup> focuses on archiving public available source code from all areas. It  
104 especially allows to archive research software. For this, it uses CodeMeta as metadata standard<sup>18</sup>.

105 **OntoSoft**<sup>5</sup> is a registry for research software in the domain of GeoScience [18]. It is based on  
106 the already introduced *OntoSoft* as ontology for its metadata.

107 In the domain of BioInformatics, **bio.tools**<sup>10</sup> is a registry for research software [14]. As introduced  
108 before in Section 2.1 it uses *biotoolsXSD* as metadata scheme.

109 **CoMSES** (Net Computational Model Library)<sup>19</sup> is a repository for agent-based modeling in  
110 Social Science [19]. It includes simple metadata which not follow a standard.

111 The **Open Energy Platform**<sup>20</sup> provides a registry for frameworks and models in energy research.  
112 It includes metadata which do not follow a standard. More information on this metadata is given  
113 in Section 2.3.

114 Table 3 summarizes the relevant repositories and registries. Especially bio.tools is a good example  
115 for a well-functioning registry for research software. The Open Energy Platform provides a  
116 registry in the energy domain without using a formalized metadata scheme.

11. <https://scicodes.net/participants/>, accessed 17.07.2023

12. <https://docs.google.com/spreadsheets/d/1lWJWeEaSczu8vNH0nBr1AJ4i34YdIQPdLcgUpxoTsDE/>,  
accessed 17.07.2023

13. <https://github.com>, accessed 17.07.2023

14. <https://gitlab.com>, accessed 17.07.2023

15. <https://zenodo.org>, accessed 17.07.2023

16. <https://about.zenodo.org/policies/>, accessed 17.07.2023

17. <https://archive.softwareheritage.org/>, accessed 17.07.2023

18. <https://www.softwareheritage.org/save-and-reference-research-software/>, accessed 17.07.2023

19. <https://www.comses.net/>, accessed 17.07.2023

20. <https://openenergy-platform.org/factsheets/models/>, accessed 17.07.2023

**Table 3:** Overview of repositories and registries for research software

✓: fulfilled                      ✗: not fulfilled  
 (✓): (partly) fulfilled

	Used metadata scheme	Covered artifacts	Only for research?	Domain	Repository (repo) vs. registry (reg)
GitHub <sup>13</sup>	–	software	✗	all	repo
GitLab <sup>14</sup>	–	software	✗	all	repo
Zenodo <sup>15</sup>	DataCite/ Dublin Core	all	✓	all	repo
Software Heritage <sup>17</sup>	CodeMeta Software	software	✓	all	repo
OntoSoft Portal <sup>5</sup> [18]	Description Ontology [18]	software	✓	Geoscience	reg
BioTools <sup>10</sup> [14]	biotoolsXSD [20]	software	✓	Bioinformatics	reg
CoMSES <sup>19</sup> [19]	–	models	✓	Social Science	repo
OpenEnergyPlatform factsheets on models <sup>20</sup>	–	models and frameworks	✓	energy research	reg

### 117 2.3 Metadata in Engineering and the Energy Domain

118 In this section, metadata schemes for data and software in the engineering domain and especially  
 119 in the energy domain are introduced, starting from the broad engineering perspective. For these  
 120 metadata schemes, it is important to note, if they focus on data or software. Also, we show to  
 121 which extend they stick to formal metadata practices (e.g., formalized scheme, reuse of existing  
 122 schemes, use of value vocabularies).

123 Schembera and Iglezakis developed the metadata scheme *EngMeta* for data in computational  
 124 engineering which includes existing elements for technical and general descriptive information  
 125 from *DataCite*, *CodeMeta*, and other relevant metadata schemes. They added additional elements  
 126 for domain specific information. Controlled vocabularies and restrictions are available for  
 127 multiple elements. Schembera and Iglezakis also presented a tool for automatic metadata  
 128 extraction and included the metadata scheme in a data repository based on Dataverse<sup>21</sup>. They  
 129 validated their approach against common recommendations for good metadata schemes. [21]

130 Schwarz and Lehnhoff described a catalog of energy co-simulation components. They used a  
 131 semantic media wiki to collect information on components and the Functional Mockup Interface  
 132 (FMI) to add descriptions on the simulation interfaces. The elements of the catalog are usable as  
 133 metadata scheme but are neither formalized nor described in more detail. [22]

134 The open energy modeling initiative (openmod) includes a list of energy models in their wiki<sup>22</sup>.  
 135 For each model, administrative and descriptive metadata are listed like license, link to a code  
 136 repository, model class. The descriptive elements include detailed information on the models.

21. <https://dataverse.org/>, accessed 12.12.2022

22. [https://wiki.openmod-initiative.org/wiki/Open\\_Models](https://wiki.openmod-initiative.org/wiki/Open_Models), accessed 12.12.2022

**Table 4:** Overview of metadata scheme for energy research software

✓: fulfilled                      ✗: not fulfilled  
 (✓): (partly) fulfilled

	Focus on models (m) vs. data (d)	Formalized metadata scheme	Based on existing scheme	Use of value vocabularies
<i>EngMeta</i> [21]	d	✓	✓	✓
Catalog of energy co-simulation components [22]	m	(✓)	✗	✗
openmod <sup>22</sup>	m	✗	✗	✗
Open Energy Metadata <sup>23</sup>	d	✓	✗	(✓)
Open Energy Platform factsheets on models <sup>20</sup>	m	✗	✗	✗

137 The metadata scheme is not formalized and controlled vocabularies are neither used for the  
 138 elements nor for the values.

139 The open energy platform introduces the open energy metadata<sup>23</sup> for data. The metadata scheme  
 140 is designed for energy data and contains multiple elements. For a lot of elements, the use of  
 141 controlled vocabularies is required, e.g., for language and license. The use of an energy ontology  
 142 is not required for the description of the data but is planned as extension. The documentation of  
 143 the metadata does not link to any existing schemes used for the design of the metadata scheme.  
 144 Additionally, the open energy platform<sup>20</sup> includes information on models and frameworks. The  
 145 metadata elements are similar to the ones of the openmod wiki and also not formalized.

146 Table 4 summarizes the most relevant metadata schemes in the engineering and energy domain.  
 147 While *EngMeta* [21] presents a good broad metadata scheme in the engineering domain following  
 148 best practices, a formalized metadata scheme for ERS is still missing. However, the open energy  
 149 platform<sup>20</sup>, the openmod wiki<sup>22</sup>, and the work of Schwarz and Lehnhoff [22] present good  
 150 starting points for developing a formalized metadata scheme for ERS.

#### 151 2.4 Domain Ontologies for Energy Research

152 Domain ontologies are necessary to improve the interoperability of metadata by being used  
 153 as value vocabularies. Wierling et al. [23] give a broad overview of ontologies in the energy  
 154 domain. In the following, we only give a short overview and refer to Wierling et al. for further  
 155 information. For the ontologies, it is important to note their scope which is important for their  
 156 usability as value vocabulary. Also, it is relevant if they are still maintained or not.

157 Cuenca et al. introduced an approach to unify different existing ontologies in the energy do-  
 158 main [24]. They focused especially on ontologies for energy management applications in smart  
 159 grids. Their ontology is called *OEMA* (Ontology for Energy Management Applications) and  
 160 is formulated in OWL2 (Web Ontology Language 2<sup>24</sup>). It reuses existing ontologies and adds  
 161 additional concepts. It has eight parts: infrastructure, energy and equipment, geographical,

23. <https://github.com/OpenEnergyPlatform/oemetadata>, accessed 12.12.2022

24. <https://w3.org/TR/owl2-overview/>, accessed 12.12.2022

162 external factors, person and organization, energy saving, smart grid stakeholders, and units. The  
163 authors extended their work by introducing *DABGEO* (Domain Analysis-Based Global Energy  
164 Ontology) [25]. To increase the reusability compared to *OEMA*, this ontology is constructed in a  
165 way that makes it easy to extend with application specific vocabulary. Using OWL2 *DABGEO*  
166 includes five parts: energy equipment, infrastructure, energy performance, energy external  
167 factors, and smart grid stakeholders.

168 Lefrançois presented the *SEAS* (Smart Energy-Aware Systems) ontology. It is a modular ontology  
169 published in Turtle<sup>25</sup>. *SEAS* contains the following modules: DeviceOntology, ForecastingOntol-  
170 ogy, OptimizationOntology, TradingOntology, and SmartMeterOntology. In the design process,  
171 he tried to follow the current best practices in ontology engineering. [26]

172 Booshehri et al. introduced the *OEO* (open energy ontology) as an ontology for energy systems  
173 analysis. The *OEO* is developed using OWL (Web Ontology Language<sup>26</sup>) and consists of  
174 oeo-modal, oeo-social, oeo-physical, and oeo-shared. The *OEO* includes concepts from other  
175 ontologies as the Financial Industry Business Ontology and the Unit Ontology. It can be used to  
176 annotate scenarios, factsheets, and data used in energy systems analysis. The ontology will be  
177 further developed on Github making it possible for everyone to contribute. [27]

178 Oppermann et al. introduced an ontology for EnArgus<sup>27</sup>, the funding information system for  
179 energy research in Germany. The ontology aims to map the whole domain of energy research  
180 and is used to improve the search in a project database. The ontology is not yet publicly available  
181 but the authors see its publication as a future work. Also, the authors admitted that they followed  
182 less strict ontology engineering rules than the developers of the *OEO*. [28]

183 Fernández-Izquierdo et al. gave a good overview on ontologies in the energy domain and intro-  
184 duced an ontology for demand response to improve interoperability between demand response  
185 stakeholders. The ontology described in OWL is based on OpenADR<sup>28</sup>, an open exchange model  
186 and global smart grid standard. The ontology is available on Github<sup>29</sup>. [29]

187 *CIM* (Common Information Model) is a group of IEC standards and a domain ontology which  
188 can be used to describe multiple aspects in energy systems. The standard is mainly applied in  
189 the electric utilities sector and has a high industry focus. The description is given in the Unified  
190 Modeling Language (UML) and a conversion to OWL is possible by using the CIMTool<sup>30</sup>. [30]

191 There exist different ontologies for the energy (research) domain which are broadly summarized  
192 in Table 5. It seems promising that one or a combination of multiple ontologies, for example of  
193 the *OEO* [27] and *CIM* [30], can be used as value vocabulary for describing ERS.

194 Overall, there exist good approaches for metadata schemes for research software in other domains  
195 like life science and geoscience, as shown in Section 2.1, which can be used as inspiration. Also,  
196 other domains already provide good examples for research software registries as described in  
197 Section 3.3. The existing approaches for ERS often lack formalization, are not based on existing

25. <https://w3.org/TR/turtle/>, accessed 12.12.2022

26. <https://w3.org/TR/owl-semantic/>, accessed 12.12.2022

27. <https://enargus.de/>, accessed 12.12.2022

28. <https://openadr.org/>, accessed 12.12.2022

29. <https://github.com/albaizq/OpenADRontology>, accessed 21.07.2021

30. <https://wiki.cimtool.org/>, accessed 12.12.2022



**Table 5:** Overview of ontologies in the energy domain

✓: fulfilled      ✗: not fulfilled

	Focus	Ontology language	Language	Maintained?
<i>OEMA</i> (Ontology for Energy Management Applications) [24]	Energy management	OWL2	English	✗
<i>DABGEO</i> (Domain Analysis-Based Global Energy Ontology) [25]	Energy	OWL2	English	✗
<i>SEAS</i> (Smart Energy-Aware Systems) [26]	Energy	Turtle	English	✗
<i>OEO</i> (open energy ontology) [27]	Energy	OWL	English	✓
EnArgus ontology [28]	Energy		German	✗
OpenADRontology [29]	Demand response	OWL	English	✗
<i>CIM</i> (Common Information Model) [30]	Energy	OWL	English	✓

198 approaches and/or do not use value vocabularies like shown in Section 2.3. The last part of the  
 199 related work showed that there exist multiple ontologies which can be used as value vocabularies  
 200 for a metadata scheme for ERS.

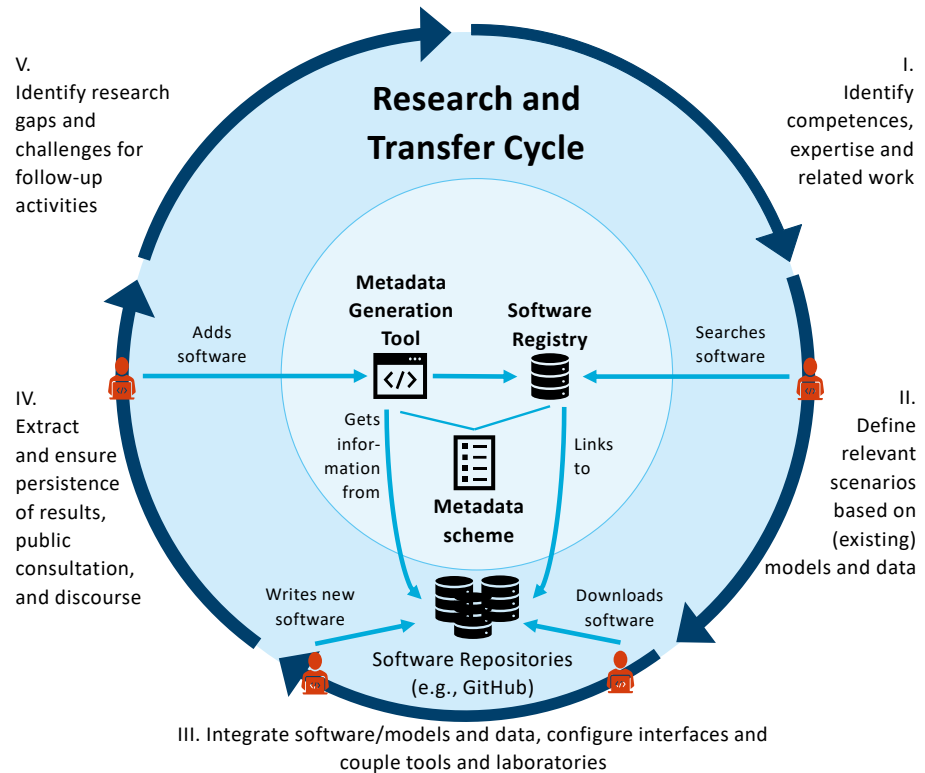
### 201 3 Concept for a Metadata-based Registry for ERS

202 We introduce our concept for a Metadata-based Registry for ERS in Figure 2. As a first step in  
 203 the research cycle, researchers should be able to use the software registry in our concept to find  
 204 relevant ERS for their research problem. All listed software should link to the corresponding  
 205 software repositories (e.g., GitHub). We describe the registry in more detail in Section 3.3. After  
 206 downloading the software, researchers will use it, extend it and/or write additional code for  
 207 their research. The new or extended software should be published on any existing software  
 208 repository. Then, researchers should be able to add the software to the software registry by using  
 209 the metadata generation tool. It should extract as many information as possible from the software  
 210 repository and, therefore, helps the researchers with creating metadata. We further describe the  
 211 concept of the metadata generation tool in Section 3.2. Both the metadata generation tool and  
 212 the registry are based on a common metadata scheme for ERS. Therefore, we first give more  
 213 details on that in Section 3.1.

#### 214 3.1 A Metadata Scheme for ERS

215 A metadata scheme for ERS should be usable for all different types of ERS to increase their  
 216 findability. It is the foundation for the other two artifacts. A metadata scheme comprises elements  
 217 describing the categories for the metadata, guidelines for creating according metadata, a syntax  
 218 and constraints for the metadata, e.g., ontologies as value vocabularies [31].

219 The metadata scheme should be developed as an application profile. An application profile  
 220 combines existing metadata elements from different ontologies and schemes into a new metadata  
 221 schema for specific use cases [32] here to describe ERS. Existing methods like the approach



**Figure 2:** A metadata-based registry support the exemplified research cycle of energy research with software usage [10]

222 of Curado Malta and Baptista [33] should be used for the development of the metadata scheme  
 223 which should be based on a domain model containing desired elements with short descriptions,  
 224 examples, and their relations. These elements and their description should only be formulated  
 225 in English to be usable by the international research community. The development of this  
 226 metadata scheme for ERS should be based on an environmental scan [34] to incorporate existing  
 227 approaches, especially the ones in semantic-web compatible formats, like the general work on  
 228 metadata for research software, e.g., *CodeMeta*<sup>3</sup>, and the domain-specific developments like the  
 229 metadata schemes of the Open Energy Platform<sup>20</sup>. For each element of the metadata scheme,  
 230 it should be specified if it is mandatory or optional and if constraints exist, e.g., if the use of  
 231 a controlled value vocabulary is required or not. Especially for the domain specific elements,  
 232 value vocabularies should be used. It should be carefully analyzed if one or multiple ontologies,  
 233 like *OEO* [27], can be integrated as value vocabulary. By using the AIMS platform [32] existing  
 234 and news terms should be combined to the metadata scheme.

235 The metadata scheme should be used by the registry and the metadata generation tool, as described  
 236 in the following sections.

### 237 3.2 A Metadata Generation Tool

238 A metadata generation tool should support all researchers to create high-quality metadata for  
 239 their ERS. It should lower the entrance barrier for creating metadata for all researchers in the  
 240 energy domain without the need for a deeper understanding of the underlying technologies.

241 Therefore, the tool consist of two main functionalities: First, the tool should extract as many  
242 metadata as possible from software repositories and other sources. Second, the tool should allow  
243 researchers to check, curate, and add the metadata for their ERS.

244 **Metadata Extraction** If the ERS is already in a software repository like GitLab or GitHub,  
245 the tool should start by extracting as many metadata as possible from the repository. Using  
246 the API of GitLab or GitHub at least some general metadata as author, name, codeRepository,  
247 programmingLanguage, dateCreated, dateModified, and license can be extracted. Mao et al.  
248 developed the framework SoMEF which extracts additional metadata out of readme files [35], e.g.,  
249 citation and installation information. Druskat et al. developed HERMES<sup>31</sup> which automatically  
250 publishes research software with metadata but also automatically harvests research software  
251 metadata from multiple sources, e.g., structured metadata in different formats (CodeMeta, citation  
252 file format and others) and unstructured files like readme, configuration, and other files. [8]  
253 Betty's Research Engine also extracts software metadata from different sources and should  
254 also be considered for the Metadata Generation Tool [36]. The extracted metadata should be  
255 combined and mapped to the metadata scheme for ERS as described in Section 3.1.

256 **Metadata Curation and Creation** The envisioned tool should allow the users to check and  
257 correct all automatically extracted metadata. Also, the users should add additional metadata  
258 which can not be extracted automatically. The tool should support and encourage the use of  
259 controlled value vocabularies as defined in the specification of the metadata scheme and should  
260 allow to integrate additional ontologies to include many links to other semantic web sources.  
261 The tool should guide the users through the metadata creation process and should be simple to  
262 use.

### 263 3.3 A Metadata Registry for ERS

264 A registry for ERS should help researchers to find the right software based on multiple search  
265 criteria, e.g., researchers wanting to do a grid simulation can look for a library or framework  
266 compatible with the models in a certain programming language they already use. For the  
267 development of such a registry, existing approaches from other domains like bio.tools<sup>10</sup> should  
268 be considered. The registry should be based on the new developed metadata scheme for ERS. It  
269 remains an open questions, how the registry should handle deleted repositories and if and how  
270 published and archived version of research software should be included.

## 271 4 Conclusion and Outlook

272 There already exist a lot of ERS and even more ERS will be needed in the future. ERS is often  
273 developed without reusing existing software. One reason for that is that the existing software is  
274 not simple to find. The overall findability of ERS can be improved when ERS is registered in  
275 a registry with relevant metadata. Therefore, we introduced the concept of a **metadata-based**  
276 **registry for ERS** to enable this process.

31. <https://project.software-metadata.pub/>, accessed 10.07.2023

277 We presented the state of the art in four parts. First, we described general approaches on metadata  
278 for research software as well as approaches focusing on research software in other domains.  
279 Second, we gave an overview on registries and repositories for research software. Third, we  
280 outlined metadata approaches in the energy domain where most current approaches focus on  
281 research data. The approaches already focusing on research software are not formalized and,  
282 therefore, do not meet our requirements. Forth, we gave an overview on existing ontologies in  
283 the energy domain which can be serve as value vocabularies for a metadata scheme.

284 Based on the presented state of the art, we defined our concept consisting of three main artifacts.  
285 The metadata scheme for ERS presents the foundation and should be developed following best  
286 practices in metadata scheme development. It should include mandatory elements and value  
287 vocabularies. The metadata generation tool should assist researchers in creating metadata by  
288 extracting as many metadata as possible from different sources, e.g., GitLab or GitHub. The  
289 registry should be the place where ERS can be found. It should store metadata of ERS and should  
290 use advanced search technologies.

291 As future work, we want to further develop, improve our concept and implement a prototype.  
292 Therefore, we will first perform a detailed requirement analysis with multiple interviews with  
293 energy researchers. Based on these requirements a domain model of the metadata scheme  
294 will be developed. Afterwards, the different parts of the concept will be implemented within  
295 NFDI4Energy<sup>32</sup> by reusing as many existing approaches (e.g., terms from existing metadata  
296 schemes and ontologies, as well as source code of existing registries and metadata extraction) as  
297 possible to achieve a high interoperability. The developed artifacts should later be tested.

## 298 5 Acknowledgements

299 This work was funded by the Lower Saxony Ministry of Science and Culture under grant number  
300 11-76251-13-3/19 – ZN3488 (ZLE) within the Lower Saxony “Vorab“ of the Volkswagen  
301 Foundation and was supported by the Center for Digital Innovations (ZDIN).

302 The authors would like to thank the German Federal Government, the German State Governments,  
303 and the Joint Science Conference (GWK) for their funding and support as part of the NFDI4Energy  
304 consortium. The work was partially funded by the German Research Foundation (DFG) –  
305 501865131 within the German National Research Data Infrastructure (NFDI, [www.nfdi.de](http://www.nfdi.de)).

## 306 6 Roles and contributions

307 **Stephan Ferez:** Conceptualization, Writing – original draft

308 **Astrid Nieße:** Conceptualization, Writing – review & editing

32. <https://nfdi4energy.uol.de/>, accessed 10.07.2023

309 **References**

- 310 [1] A. J. Christe, S. I. Negrashov, P. M. Johnson, D. Nakahodo, D. Badke, and D. Aghalar-  
311 pour, “OPQ Version 2: An Architecture for Distributed, Real-Time, High Performance  
312 Power Data Acquisition, Analysis, and Visualization,” in *2017 IEEE 7th Annual Interna-*  
313 *tional Conference on CYBER Technology in Automation, Control, and Intelligent Systems*  
314 *(CYBER)*, Jul. 2017, pp. 1415–1419. DOI: [10.1109/CYBER.2017.8446111](https://doi.org/10.1109/CYBER.2017.8446111).
- 315 [2] C. Steinbrink, M. Blank-Babazadeh, A. El-Ama, *et al.*, “CPES Testing with mosaik: Co-  
316 Simulation Planning, Execution and Analysis,” *Applied Sciences*, vol. 9, no. 5, p. 923, Jan.  
317 2019. DOI: [10.3390/app9050923](https://doi.org/10.3390/app9050923).
- 318 [3] K. Löffler, K. Hainsch, T. Burandt, P.-Y. Oei, C. Kemfert, and C. Von Hirschhausen,  
319 “Designing a Model for the Global Energy System—GENeSYS-MOD: An Application  
320 of the Open-Source Energy Modeling System (OSeMOSYS),” *Energies*, vol. 10, no. 10,  
321 p. 1468, Oct. 2017. DOI: [10.3390/en10101468](https://doi.org/10.3390/en10101468).
- 322 [4] W. Hasselbring, L. Carr, S. Hettrick, H. Packer, and T. Tiropanis, “From FAIR research  
323 data toward FAIR and open research software,” *it - Information Technology*, vol. 62, no. 1,  
324 pp. 39–47, Feb. 2020, ISSN: 1611-2776, 2196-7032. DOI: [10.1515/itit-2019-0040](https://doi.org/10.1515/itit-2019-0040).
- 325 [5] C. Steinbrink, S. Lehnhoff, S. Rohjans, *et al.*, “Simulation-Based Validation of Smart  
326 Grids – Status Quo and Future Research Trends,” in *Industrial Applications of Holonic and*  
327 *Multi-Agent Systems*, V. Mařík, W. Wahlster, T. Strasser, and P. Kadera, Eds., ser. Lecture  
328 Notes in Computer Science, Cham, Germany: Springer International Publishing, 2017,  
329 pp. 171–185, ISBN: 978-3-319-64635-0. DOI: [10.1007/978-3-319-64635-0\\_13](https://doi.org/10.1007/978-3-319-64635-0_13).
- 330 [6] A.-L. Lamprecht, L. Garcia, M. Kuzak, *et al.*, “Towards FAIR principles for research  
331 software,” *Data Science*, vol. 3, no. 1, pp. 37–59, Jan. 2020, ISSN: 2451-8484. DOI:  
332 [10.3233/DS-190026](https://doi.org/10.3233/DS-190026).
- 333 [7] D. S. Katz, M. Gruenpeter, and T. Honeyman, “Taking a fresh look at FAIR for research  
334 software,” *Patterns*, vol. 2, no. 3, Mar. 2021, ISSN: 2666-3899. DOI: [10.1016/j.pattee](https://doi.org/10.1016/j.pattee)  
335 [r.2021.100222](https://doi.org/10.1016/j.pattee.2021.100222).
- 336 [8] S. Druskat, O. Bertuch, G. Juckeland, O. Knodel, and T. Schlauch, *Software publications*  
337 *with rich metadata: State of the art, automated workflows and HERMES concept*, Preprint,  
338 Jan. 2022. DOI: [10.48550/arXiv.2201.09015](https://doi.org/10.48550/arXiv.2201.09015). (visited on 07/07/2023).
- 339 [9] D. Garijo, H. Ménager, L. Hwang, *et al.*, “Nine best practices for research software reg-  
340 istries and repositories,” en, *PeerJ Computer Science*, vol. 8, e1023, Aug. 2022, Publisher:  
341 PeerJ Inc., ISSN: 2376-5992. DOI: [10.7717/peerj-cs.1023](https://doi.org/10.7717/peerj-cs.1023). (visited on 07/06/2023).
- 342 [10] S. Ferez, “Towards More Findable Energy Research Software by Introducing a Metadata-  
343 based Registry,” in *Abstracts of the 11th DACH+ Conference on Energy Informatics*, Anke  
344 Weidlich, Gunther Gust and Mirko Schäfer, Ed., Springer, 2022, ISBN: 2520-8942. DOI:  
345 [10.1186/s42162-022-00215-6](https://doi.org/10.1186/s42162-022-00215-6).
- 346 [11] Y. Gil, V. Ratnakar, and D. Garijo, “OntoSoft: Capturing Scientific Software Metadata,”  
347 in *Proceedings of the 8th International Conference on Knowledge Capture*, ser. K-CAP  
348 2015, New York, NY, USA: Association for Computing Machinery, Oct. 2015, pp. 1–4,  
349 ISBN: 978-1-4503-3849-3. DOI: [10.1145/2815833.2816955](https://doi.org/10.1145/2815833.2816955).

- 350 [12] D. Garijo, M. Osorio, D. Khider, V. Ratnakar, and Y. Gil, “OKG-Soft: An Open Knowledge  
351 Graph with Machine Readable Scientific Software Metadata,” in *2019 15th International  
352 Conference on eScience (eScience)*, Sep. 2019, pp. 349–358. DOI: [10.1109/eScience  
353 .2019.00046](https://doi.org/10.1109/eScience.2019.00046).
- 354 [13] J. Malone, A. Brown, A. L. Lister, *et al.*, “The Software Ontology (SWO): A resource for  
355 reproducibility in biomedical data analysis, curation and digital preservation,” *Journal of  
356 Biomedical Semantics*, vol. 5, no. 1, p. 25, Jun. 2014, ISSN: 2041-1480. DOI: [10.1186  
357 /2041-1480-5-25](https://doi.org/10.1186/2041-1480-5-25).
- 358 [14] J. Ison, H. Ienasescu, P. Chmura, *et al.*, “The bio.tools registry of software tools and data  
359 resources for the life sciences,” *Genome Biology*, vol. 20, no. 1, p. 164, Aug. 2019, ISSN:  
360 1474-760X. DOI: [10.1186/s13059-019-1772-6](https://doi.org/10.1186/s13059-019-1772-6).
- 361 [15] V. Cosentino, J. Luis, and J. Cabot, “Findings from GitHub: Methods, datasets and limita-  
362 tions,” in *Proceedings of the 13th International Conference on Mining Software Reposito-  
363 raries*, ser. MSR ’16, New York, NY, USA: Association for Computing Machinery, May  
364 2016, pp. 137–141, ISBN: 978-1-4503-4186-8. DOI: [10.1145/2901739.2901776](https://doi.org/10.1145/2901739.2901776).
- 365 [16] I. Peters, P. Kraker, E. Lex, C. Gumpenberger, and J. I. Gorraiz, “Zenodo in the Spotlight  
366 of Traditional and New Metrics,” *Frontiers in Research Metrics and Analytics*, vol. 2,  
367 2017, ISSN: 2504-0537. [Online]. Available: [https://www.frontiersin.org/artic  
368 le/10.3389/frma.2017.00013](https://www.frontiersin.org/article/10.3389/frma.2017.00013).
- 369 [17] A. Bucciero, E. Demetrescu, B. Fanini, A. Chirivì, and F. Taurino, “An approach to extend  
370 the metadata schema of Zenodo for Cultural Heritage datasets,” en, *Research Ideas and  
371 Outcomes*, vol. 9, e93859, Mar. 2023, Publisher: Pensoft Publishers, ISSN: 2367-7163.  
372 DOI: [10.3897/rio.9.e93859](https://doi.org/10.3897/rio.9.e93859). (visited on 07/17/2023).
- 373 [18] Y. Gil, D. Garijo, S. Mishra, and V. Ratnakar, “OntoSoft: A distributed semantic registry for  
374 scientific software,” in *2016 IEEE 12th International Conference on e-Science (e-Science)*,  
375 Oct. 2016, pp. 331–336. DOI: [10.1109/eScience.2016.7870916](https://doi.org/10.1109/eScience.2016.7870916).
- 376 [19] M. A. Janssen, L. N. Alessa, M. Barton, S. Bergin, and A. Lee, “Towards a Community  
377 Framework for Agent-Based Modelling,” en, *Journal of Artificial Societies and Social  
378 Simulation*, vol. 11, no. 2, p. 6, 2008, ISSN: 1460-7425. [Online]. Available: [https://w  
379 ww.jasss.org/11/2/6/6.pdf](https://www.jasss.org/11/2/6/6.pdf).
- 380 [20] J. Ison, H. Ienasescu, E. Rydza, *et al.*, “BiotoolsSchema: A formalized schema for bioinfor-  
381 matics software description,” English, *GigaScience*, vol. 10, no. 1, 2021, ISSN: 2047-217X.  
382 DOI: [10.1093/gigascience/giaa157](https://doi.org/10.1093/gigascience/giaa157).
- 383 [21] B. Schembera and D. Iglezakis, “EngMeta – Metadata for Computational Engineering,”  
384 *International Journal of Metadata, Semantics and Ontologies*, vol. 14, no. 1, p. 26, 2020,  
385 ISSN: 1744-2621, 1744-263X. DOI: [10.1504/IJMSO.2020.107792](https://doi.org/10.1504/IJMSO.2020.107792).
- 386 [22] J. Schwarz and S. Lehnhoff, “Ontological Integration of Semantics and Domain Knowl-  
387 edge in Energy Scenario Co-simulation,” in *Proceedings of the 11th International Joint  
388 Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Man-  
389 agement*, Vienna, Austria: SCITEPRESS - Science and Technology Publications, 2019,  
390 pp. 127–136, ISBN: 978-989-758-382-7. DOI: [10.5220/0008069801270136](https://doi.org/10.5220/0008069801270136).

- 391 [23] A. Wierling, V. J. Schwanitz, S. Altinci, *et al.*, “FAIR Metadata Standards for Low Carbon  
392 Energy Research—A Review of Practices and How to Advance,” *Energies*, vol. 14, no. 20,  
393 p. 6692, Jan. 2021. DOI: [10.3390/en14206692](https://doi.org/10.3390/en14206692).
- 394 [24] J. Cuenca, F. Larrinaga, and E. Curry, “A Unified Semantic Ontology for Energy Manage-  
395 ment Applications,” in *2nd International Workshop on Ontology Modularity, Contextuality,  
396 and Evolution*, Vienna, Austria, 2017. [Online]. Available: [http://www.edwardcurry  
397 .org/publications/WOMoCoE\\_17.pdf](http://www.edwardcurry.org/publications/WOMoCoE_17.pdf).
- 398 [25] J. Cuenca, F. Larrinaga, and E. Curry, “DABGEO: A reusable and usable global energy  
399 ontology for the energy domain,” *Journal of Web Semantics*, vol. 61-62, p. 100 550, Mar.  
400 2020, ISSN: 1570-8268. DOI: [10.1016/j.websem.2020.100550](https://doi.org/10.1016/j.websem.2020.100550).
- 401 [26] M. Lefrançois, “Planned ETSI SAREF Extensions based on the W3C&OGC SOSA/SSN-  
402 compatible SEAS Ontology Paaerns,” Amsterdam, Netherlands, Tech. Rep., Sep. 2017.  
403 [Online]. Available: <https://hal-emse.ccsd.cnrs.fr/emse-01638275>.
- 404 [27] M. Booshehri, L. Emele, S. Flügel, *et al.*, “Introducing the Open Energy Ontology: En-  
405 hancing data interpretation and interfacing in energy systems analysis,” *Energy and AI*,  
406 vol. 5, p. 100 074, Sep. 2021, ISSN: 2666-5468. DOI: [10.1016/j.egyai.2021.100074](https://doi.org/10.1016/j.egyai.2021.100074).
- 407 [28] L. Oppermann, S. Hirzel, A. Güldner, *et al.*, “Finding and analysing energy research  
408 funding data: The EnArgus system,” *Energy and AI*, vol. 5, p. 100 070, Sep. 2021, ISSN:  
409 2666-5468. DOI: [10.1016/j.egyai.2021.100070](https://doi.org/10.1016/j.egyai.2021.100070).
- 410 [29] A. Fernández-Izquierdo, A. Cimmino, C. Patsonakis, *et al.*, “OpenADR Ontology: Semantic  
411 Enrichment of Demand Response Strategies in Smart Grids,” in *2020 International  
412 Conference on Smart Energy Systems and Technologies (SEST)*, Sep. 2020, pp. 1–6. DOI:  
413 [10.1109/SEST48500.2020.9203093](https://doi.org/10.1109/SEST48500.2020.9203093).
- 414 [30] M. Uslar, M. Specht, S. Rohjans, J. Trefke, and J. M. Vasquez Gonzalez, *The Common  
415 Information Model CIM (Power Systems)*. Berlin, Germany and Heidelberg, Germany:  
416 Springer Berlin Heidelberg, 2012, vol. 66, ISBN: 978-3-642-25215-0. DOI: [10.1007/97  
417 8-3-642-25215-0](https://doi.org/10.1007/978-3-642-25215-0).
- 418 [31] M. L. Zeng and J. Qin, *Metadata*, Third edition. London, Great Britain: Facet Publishing,  
419 2022, ISBN: 978-1-78330-588-9.
- 420 [32] M. Grönewald, P. Mund, M. Bodenbrenner, *et al.*, “Mit AIMS zu einem Metadatenman-  
421 agement 4.0: FAIRe Forschungsdaten benötigen interoperable Metadaten,” de, 2022,  
422 Publisher: heiBOOKS. DOI: [10.11588/HEIBOOKS.979.C13721](https://doi.org/10.11588/HEIBOOKS.979.C13721).
- 423 [33] M. Curado Malta and A. A. Baptista, “A Method for the Development of Dublin Core  
424 Application Profiles (Me4DCAP V0.2): Detailed Description,” in *Proceedings of the 2013  
425 International Conference on Dublin Core and Metadata Applications*, Lisbon, Portugal:  
426 Dublin Core Metadata Initiative, 2013, pp. 90–103. [Online]. Available: [https://dcpap  
427 ers.dublincore.org/pubs/article/viewFile/3674/1897.pdf](https://dcpapers.dublincore.org/pubs/article/viewFile/3674/1897.pdf).
- 428 [34] M. Curado Malta and A. A. Baptista, “The Development process of a Metadata Application  
429 Profile for the Social and Solidarity Economy,” in *Developing Metadata Application  
430 Profiles*, M. C. Malta, A. A. Baptista, and P. Walk, Eds., Hershey, PA, USA: IGI Global,  
431 2017, pp. 98–117, ISBN: 978-1-5225-2221-8. DOI: [10.4018/978-1-5225-2221-8.ch  
432 005](https://doi.org/10.4018/978-1-5225-2221-8.ch005).

- 433 [35] A. Mao, D. Garijo, and S. Fakhraei, “SoMEF: A Framework for Capturing Scientific  
434 Software Metadata from its Documentation,” in *2019 IEEE International Conference on*  
435 *Big Data (Big Data)*, Dec. 2019, pp. 3032–3037. DOI: [10.1109/BigData47090.2019](https://doi.org/10.1109/BigData47090.2019.9006447)  
436 [.9006447](https://doi.org/10.1109/BigData47090.2019.9006447).
- 437 [36] V. Seibert, A. Rausch, and S. Wittek, *Betty’s (Re)Search Engine: A client-based search*  
438 *engine for research software stored in repositories*. en, Publisher: ing.grid Preprint Repos-  
439 itory, Mar. 2023. [Online]. Available: [https://preprints.inggrid.org/repositor](https://preprints.inggrid.org/repository/view/16/)  
440 [y/view/16/](https://preprints.inggrid.org/repository/view/16/) (visited on 07/17/2023).