ing.grid preprints

RESEARCH ARTICLE

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

Stephan Ferenz [1]
Astrid Nieße [1]

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

Date Received:
2023-03-01

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

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

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

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

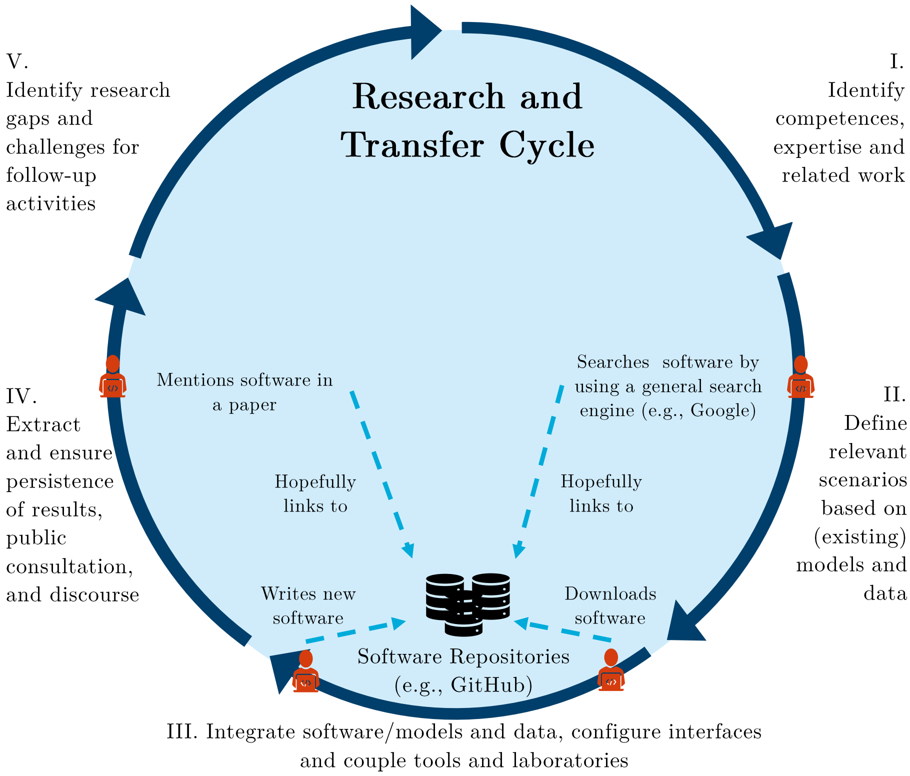

Figure 1: An exemplified research cycle of energy research with software usage

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

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

For this goal, we propose a concept for a metadata-based registry for ERS based on a good metadata scheme[2]. Our contributions are:

- We outline the state of the art in the fields of metadata schemes for research software, registries and repositories for research software, metadata schemes in the field of energy research, and ontologies in the energy domain in Section 2.
- We introduce our concept of metadata-based registry for ERS in Section 3.
- 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[1]
	Identifier	-
	Version	3.0
Accessibility	Link to repository	https://gitlab.com/mosaik/mosaik
Interoperability	Programming language	python
	Input/output formats	depends on used models
	Dependencies	-
Reusability	License	LGPL v3.0
	Link to documentation	https://mosaik.readthedocs.io/en/latest/quickstart.html

2 State of the art

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

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

2.1 Metadata for Research Software

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

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

For geosciences, Gil et al. introduced an ontology to describe research software, *OntoSoft*, with six categories: identify, understand, execute, do research, get support, and update [11]. 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

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

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

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

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

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

2.2 Registries and Repositories for Research Software

This section gives an overview of existing registries and repositories for research software. As defined in Section 1, registries only included 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, accessed 12.12.2022
9. http://models.mint.isi.edu, accessed 12.12.2022
10. http://bio.tools, accessed 12.12.2022

metadata is stored.

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

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

GitLab[14] is also a software repository based on a version control system. It includes all kind of software including research software. It collects multiple metadata for software not based on any standard but the metadata is accessible via an API.

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

Software Heritage[17] focuses on archiving public available source code from all areas. It especially allows to archive research software. For this, it uses CodeMeta as metadata standard[18].

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

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

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

The **Open Energy Platform**[20] provides a registry for frameworks and models in energy research. It includes metadata which do not follow a standard. More information on this metadata is given in Section 2.3.

Table 3 summarizes the relevant repositories and registries. Especially bio.tools is a good example for a well-functioning registry for research software. The Open Energy Platform provides a 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/1lWJWeEaSczu8vNH0nBrlAJ4i34YdIQPdLcgUpxoTsDE/, 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[13]	-	software	✗	all	repo
GitLab[14]	-	software	✗	all	repo
Zenodo[15]	DataCite/ Dublin Core	all	✓	all	repo
Software Heritage[17]	CodeMeta	software	✓	all	repo
OntoSoft Portal[5][18]	Software Description Ontology [18]	software	✓	Geoscience	reg
BioTools[10][14]	biotoolsXSD [20]	software	✓	Bioinformatics	reg
CoMSES[19][19]	-	models	✓	Social Science	repo
OpenEnergyPlatform factsheets on models[20]	–	models and frameworks	✓	energy research	reg

2.3 Metadata in Engineering and the Energy Domain

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

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

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

The open energy modeling initiative (openmod) includes a list of energy models in their wiki[22]. For each model, administrative and descriptive metadata are listed like license, link to a code 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, 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
EngMeta [21]	d	✓	✓	✓
Catalog of energy co-simulation components [22]	m	(✓)	✗	✗
openmod[22]	m	✗	✗	✗
Open Energy Metadata[23]	d	✓	✗	(✓)
Open Energy Platform factsheets on models[20]	m	✗	✗	✗

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

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

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

2.4 Domain Ontologies for Energy Research

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

Cuenca et al. introduced an approach to unify different existing ontologies in the energy domain [24]. They focused especially on ontologies for energy management applications in smart grids. Their ontology is called *OEMA* (Ontology for Energy Management Applications) and is formulated in OWL2 (Web Ontology Language 2[24]). It reuses existing ontologies and adds 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

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

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

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

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

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

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

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

Overall, there exist good approaches for metadata schemes for research software in other domains like life science and geoscience, as shown in Section 2.1, which can be used as inspiration. Also, other domains already provide good examples for research software registries as described in 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-semantics/`, 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?
OEMA (Ontology for Energy Management Applications) [24]	Energy management	OWL2	English	✗
DABGEO (Domain Analysis-Based Global Energy Ontology) [25]	Energy	OWL2	English	✗
SEAS (Smart Energy-Aware Systems) [26]	Energy	Tur-tle	English	✗
OEO (open energy ontology) [27]	Energy	OWL	English	✓
EnArgus ontology [28]	Energy		German	✗
OpenADRontology [29]	Demand response	OWL	English	✗
CIM (Common Information Model) [30]	Energy	OWL	English	✓

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

3 Concept for a Metadata-based Registry for ERS

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

3.1 A Metadata Scheme for ERS

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

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

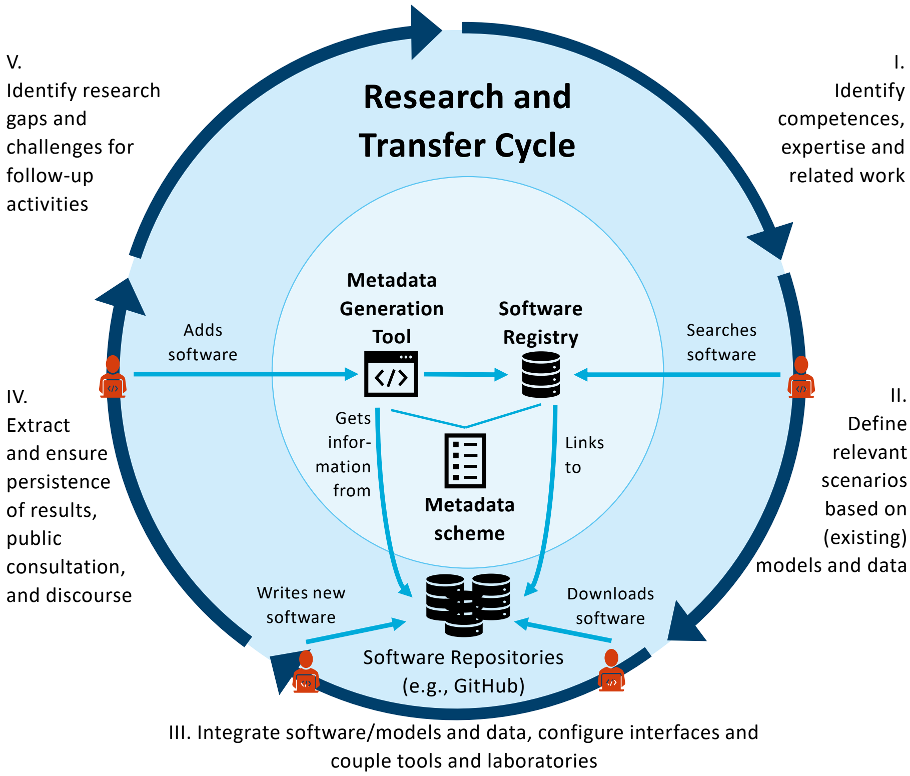

Figure 2: A metadata-based registry support the exemplified research cycle of energy research with software usage [10]

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

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

3.2 A Metadata Generation Tool

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

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

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

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

3.3 A Metadata Registry for ERS

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

4 Conclusion and Outlook

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

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

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

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

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

5 Acknowledgements

This work was funded by the Lower Saxony Ministry of Science and Culture under grant number 11-76251-13-3/19 – ZN3488 (ZLE) within the Lower Saxony “Vorab“ of the Volkswagen Foundation and was supported by the Center for Digital Innovations (ZDIN). The authors would also like to thank the German Federal Government, the German State Governments, and the Joint Science Conference (GWK) for their funding and support as part of the NFDI4Energy consortium, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 501865131.

6 Roles and contributions

Stephan Ferenz: Conceptualization, Writing – original draft

Astrid Nieße: Conceptualization, Writing – review & editing

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

References

[1] A. J. Christe, S. I. Negrashov, P. M. Johnson, D. Nakahodo, D. Badke, and D. Aghalarpour, “OPQ Version 2: An Architecture for Distributed, Real-Time, High Performance Power Data Acquisition, Analysis, and Visualization,” in *2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, Jul. 2017, pp. 1415–1419. DOI: 10.1109/CYBER.2017.8446111.

[2] C. Steinbrink, M. Blank-Babazadeh, A. El-Ama, *et al.*, “CPES Testing with mosaik: Co-Simulation Planning, Execution and Analysis,” *Applied Sciences*, vol. 9, no. 5, p. 923, Jan. 2019. DOI: 10.3390/app9050923.

[3] K. Löffler, K. Hainsch, T. Burandt, P.-Y. Oei, C. Kemfert, and C. Von Hirschhausen, “Designing a Model for the Global Energy System—GENeSYS-MOD: An Application of the Open-Source Energy Modeling System (OSeMOSYS),” *Energies*, vol. 10, no. 10, p. 1468, Oct. 2017. DOI: 10.3390/en10101468.

[4] W. Hasselbring, L. Carr, S. Hettrick, H. Packer, and T. Tiropanis, “From FAIR research data toward FAIR and open research software,” *it - Information Technology*, vol. 62, no. 1, pp. 39–47, Feb. 2020, ISSN: 1611-2776, 2196-7032. DOI: 10.1515/itit-2019-0040.

[5] C. Steinbrink, S. Lehnhoff, S. Rohjans, *et al.*, “Simulation-Based Validation of Smart Grids – Status Quo and Future Research Trends,” in *Industrial Applications of Holonic and Multi-Agent Systems*, V. Mařík, W. Wahlster, T. Strasser, and P. Kadera, Eds., ser. Lecture Notes in Computer Science, Cham, Germany: Springer International Publishing, 2017, pp. 171–185, ISBN: 978-3-319-64635-0. DOI: 10.1007/978-3-319-64635-0_13.

[6] A.-L. Lamprecht, L. Garcia, M. Kuzak, *et al.*, “Towards FAIR principles for research software,” *Data Science*, vol. 3, no. 1, pp. 37–59, Jan. 2020, ISSN: 2451-8484. DOI: 10.3233/DS-190026.

[7] D. S. Katz, M. Gruenpeter, and T. Honeyman, “Taking a fresh look at FAIR for research software,” *Patterns*, vol. 2, no. 3, Mar. 2021, ISSN: 2666-3899. DOI: 10.1016/j.patter.2021.100222.

[8] S. Druskat, O. Bertuch, G. Juckeland, O. Knodel, and T. Schlauch, *Software publications with rich metadata: State of the art, automated workflows and HERMES concept*, Preprint, Jan. 2022. DOI: 10.48550/arXiv.2201.09015. (visited on 07/07/2023).

[9] D. Garijo, H. Ménager, L. Hwang, *et al.*, “Nine best practices for research software registries and repositories,” en, *PeerJ Computer Science*, vol. 8, e1023, Aug. 2022, Publisher: PeerJ Inc., ISSN: 2376-5992. DOI: 10.7717/peerj-cs.1023. (visited on 07/06/2023).

[10] S. Ferenz, “Towards More Findable Energy Research Software by Introducing a Metadata-based Registry,” in *Abstracts of the 11th DACH+ Conference on Energy Informatics*, Anke Weidlich, Gunther Gust and Mirko Schäfer, Ed., Springer, 2022, ISBN: 2520-8942. DOI: 10.1186/s42162-022-00215-6.

[11] Y. Gil, V. Ratnakar, and D. Garijo, “OntoSoft: Capturing Scientific Software Metadata,” in *Proceedings of the 8th International Conference on Knowledge Capture*, ser. K-CAP 2015, New York, NY, USA: Association for Computing Machinery, Oct. 2015, pp. 1–4, ISBN: 978-1-4503-3849-3. DOI: 10.1145/2815833.2816955.

[12] D. Garijo, M. Osorio, D. Khider, V. Ratnakar, and Y. Gil, "OKG-Soft: An Open Knowledge Graph with Machine Readable Scientific Software Metadata," in *2019 15th International Conference on eScience (eScience)*, Sep. 2019, pp. 349–358. DOI: 10.1109/eScience.2019.00046.

[13] J. Malone, A. Brown, A. L. Lister, *et al.*, "The Software Ontology (SWO): A resource for reproducibility in biomedical data analysis, curation and digital preservation," *Journal of Biomedical Semantics*, vol. 5, no. 1, p. 25, Jun. 2014, ISSN: 2041-1480. DOI: 10.1186/2041-1480-5-25.

[14] J. Ison, H. Ienasescu, P. Chmura, *et al.*, "The bio.tools registry of software tools and data resources for the life sciences," *Genome Biology*, vol. 20, no. 1, p. 164, Aug. 2019, ISSN: 1474-760X. DOI: 10.1186/s13059-019-1772-6.

[15] V. Cosentino, J. Luis, and J. Cabot, "Findings from GitHub: Methods, datasets and limitations," in *Proceedings of the 13th International Conference on Mining Software Repositories*, ser. MSR '16, New York, NY, USA: Association for Computing Machinery, May 2016, pp. 137–141, ISBN: 978-1-4503-4186-8. DOI: 10.1145/2901739.2901776.

[16] I. Peters, P. Kraker, E. Lex, C. Gumpenberger, and J. I. Gorraiz, "Zenodo in the Spotlight of Traditional and New Metrics," *Frontiers in Research Metrics and Analytics*, vol. 2, 2017, ISSN: 2504-0537. [Online]. Available: https://www.frontiersin.org/article/10.3389/frma.2017.00013.

[17] A. Bucciero, E. Demetrescu, B. Fanini, A. Chirivì, and F. Taurino, "An approach to extend the metadata schema of Zenodo for Cultural Heritage datasets," en, *Research Ideas and Outcomes*, vol. 9, e93859, Mar. 2023, Publisher: Pensoft Publishers, ISSN: 2367-7163. DOI: 10.3897/rio.9.e93859. (visited on 07/17/2023).

[18] Y. Gil, D. Garijo, S. Mishra, and V. Ratnakar, "OntoSoft: A distributed semantic registry for scientific software," in *2016 IEEE 12th International Conference on e-Science (e-Science)*, Oct. 2016, pp. 331–336. DOI: 10.1109/eScience.2016.7870916.

[19] M. A. Janssen, L. N. Alessa, M. Barton, S. Bergin, and A. Lee, "Towards a Community Framework for Agent-Based Modelling," en, *Journal of Artificial Societies and Social Simulation*, vol. 11, no. 2, p. 6, 2008, ISSN: 1460-7425. [Online]. Available: https://www.jasss.org/11/2/6/6.pdf.

[20] J. Ison, H. Ienasescu, E. Rydza, *et al.*, "BiotoolsSchema: A formalized schema for bioinformatics software description," English, *GigaScience*, vol. 10, no. 1, 2021, ISSN: 2047-217X. DOI: 10.1093/gigascience/giaa157.

[21] B. Schembera and D. Iglezakis, "EngMeta – Metadata for Computational Engineering," *International Journal of Metadata, Semantics and Ontologies*, vol. 14, no. 1, p. 26, 2020, ISSN: 1744-2621, 1744-263X. DOI: 10.1504/IJMSO.2020.107792.

[22] J. Schwarz and S. Lehnhoff, "Ontological Integration of Semantics and Domain Knowledge in Energy Scenario Co-simulation," in *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, Vienna, Austria: SCITEPRESS - Science and Technology Publications, 2019, pp. 127–136, ISBN: 978-989-758-382-7. DOI: 10.5220/0008069801270136.

[23] A. Wierling, V. J. Schwanitz, S. Altinci, *et al.*, "FAIR Metadata Standards for Low Carbon Energy Research—A Review of Practices and How to Advance," *Energies*, vol. 14, no. 20, p. 6692, Jan. 2021. DOI: 10.3390/en14206692.

[24] J. Cuenca, F. Larrinaga, and E. Curry, "A Unified Semantic Ontology for Energy Management Applications," in *2nd International Workshop on Ontology Modularity, Contextuality, and Evolution*, Vienna, Austria, 2017. [Online]. Available: http://www.edwardcurry.org/publications/WOMoCoE_17.pdf.

[25] J. Cuenca, F. Larrinaga, and E. Curry, "DABGEO: A reusable and usable global energy ontology for the energy domain," *Journal of Web Semantics*, vol. 61-62, p. 100 550, Mar. 2020, ISSN: 1570-8268. DOI: 10.1016/j.websem.2020.100550.

[26] M. Lefrançois, "Planned ETSI SAREF Extensions based on the W3C&OGC SOSA/SSN-compatible SEAS Ontology Paaerns," Amsterdam, Netherlands, Tech. Rep., Sep. 2017. [Online]. Available: https://hal-emse.ccsd.cnrs.fr/emse-01638275.

[27] M. Booshehri, L. Emele, S. Flügel, *et al.*, "Introducing the Open Energy Ontology: Enhancing data interpretation and interfacing in energy systems analysis," *Energy and AI*, vol. 5, p. 100 074, Sep. 2021, ISSN: 2666-5468. DOI: 10.1016/j.egyai.2021.100074.

[28] L. Oppermann, S. Hirzel, A. Güldner, *et al.*, "Finding and analysing energy research funding data: The EnArgus system," *Energy and AI*, vol. 5, p. 100 070, Sep. 2021, ISSN: 2666-5468. DOI: 10.1016/j.egyai.2021.100070.

[29] A. Fernández-Izquierdo, A. Cimmino, C. Patsonakis, *et al.*, "OpenADR Ontology: Semantic Enrichment of Demand Response Strategies in Smart Grids," in *2020 International Conference on Smart Energy Systems and Technologies (SEST)*, Sep. 2020, pp. 1–6. DOI: 10.1109/SEST48500.2020.9203093.

[30] M. Uslar, M. Specht, S. Rohjans, J. Trefke, and J. M. Vasquez Gonzalez, *The Common Information Model CIM* (Power Systems). Berlin, Germany and Heidelberg, Germany: Springer Berlin Heidelberg, 2012, vol. 66, ISBN: 978-3-642-25215-0. DOI: 10.1007/978-3-642-25215-0.

[31] M. L. Zeng and J. Qin, *Metadata*, Third edition. London, Great Britain: Facet Publishing, 2022, ISBN: 978-1-78330-588-9.

[32] M. Grönewald, P. Mund, M. Bodenbrenner, *et al.*, "Mit AIMS zu einem Metadatenmanagement 4.0: FAIRe Forschungsdaten benötigen interoperable Metadaten," de, 2022, Publisher: heiBOOKS. DOI: 10.11588/HEIBOOKS.979.C13721.

[33] M. Curado Malta and A. A. Baptista, "A Method for the Development of Dublin Core Application Profiles (Me4DCAP V0.2): Detailed Description," in *Proceedings of the 2013 International Conference on Dublin Core and Metadata Applications*, Lisbon, Portugal: Dublin Core Metadata Initiative, 2013, pp. 90–103. [Online]. Available: https://dcpapers.dublincore.org/pubs/article/viewFile/3674/1897.pdf.

[34] M. Curado Malta and A. A. Baptista, "The Development process of a Metadata Application Profile for the Social and Solidarity Economy," in *Developing Metadata Application Profiles*, M. C. Malta, A. A. Baptista, and P. Walk, Eds., Hershey, PA, USA: IGI Global, 2017, pp. 98–117, ISBN: 978-1-5225-2221-8. DOI: 10.4018/978-1-5225-2221-8.ch005.

[35] A. Mao, D. Garijo, and S. Fakhraei, “SoMEF: A Framework for Capturing Scientific Software Metadata from its Documentation,” in *2019 IEEE International Conference on Big Data (Big Data)*, Dec. 2019, pp. 3032–3037. DOI: 10.1109/BigData47090.2019.9006447.

[36] V. Seibert, A. Rausch, and S. Wittek, *Betty’s (Re)Search Engine: A client-based search engine for research software stored in repositories.* en, Publisher: ing.grid Preprint Repository, Mar. 2023. [Online]. Available: https://preprints.inggrid.org/repository/view/16/ (visited on 07/17/2023).