

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

Stephan Ferez, Astrid Nieße
Department of Computer Science
Carl von Ossietzky University Oldenburg

Agenda

- Motivation
- Vision
- Related Work
- Approach
- Summary and Outlook

Motivation: Research in the Energy Domain

- Key challenges:
 - Energy transition
 - Safety-critical infrastructure
 - Interdisciplinary environment
 - Cyber-physical energy systems
- Most research is based on simulations:
 - Behavior of components (PV, batteries, ...)
 - Control strategies for energy systems
 - Grid planning
 - Transition of the energy system

› Energy Research Software (ERS) is fundamental for Energy Research

Motivation: Challenges for Energy Research Software

- Increasing complexity
- Provided code hardly findable
- Software rarely reused

→ Multiple frameworks with overlapping features

→ A lot of time spent to develop code which already exist somewhere

→ Code is often only use once

→ High development costs

› Energy Research Software needs to become more findable to be reused

Motivation: Findable Energy Research Software

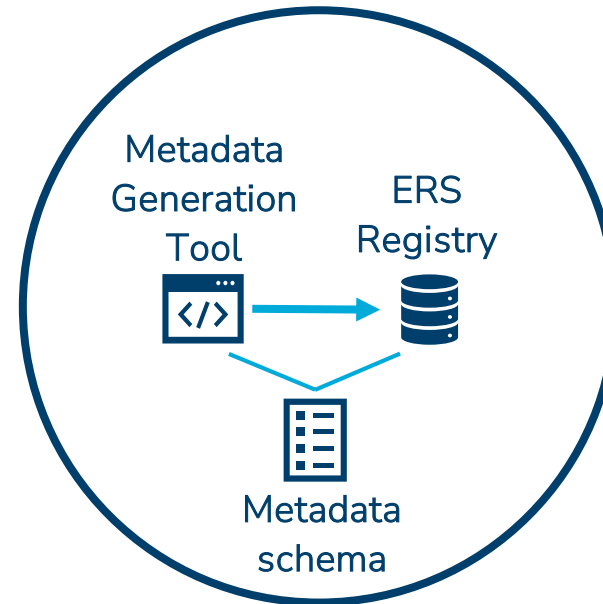
- From the FAIR criteria [1]:
 - Findability can be improved through
 - good general metadata
 - registration of a software with metadata in a registry
 - Accessibility and Interoperability can be improved through
 - good metadata as context
 - Reuse can be simplified
 - metadata including conditions for reuse as license and how to cite

➤ **Metadata and a Metadata Registry can make Energy Research Software more Findable**

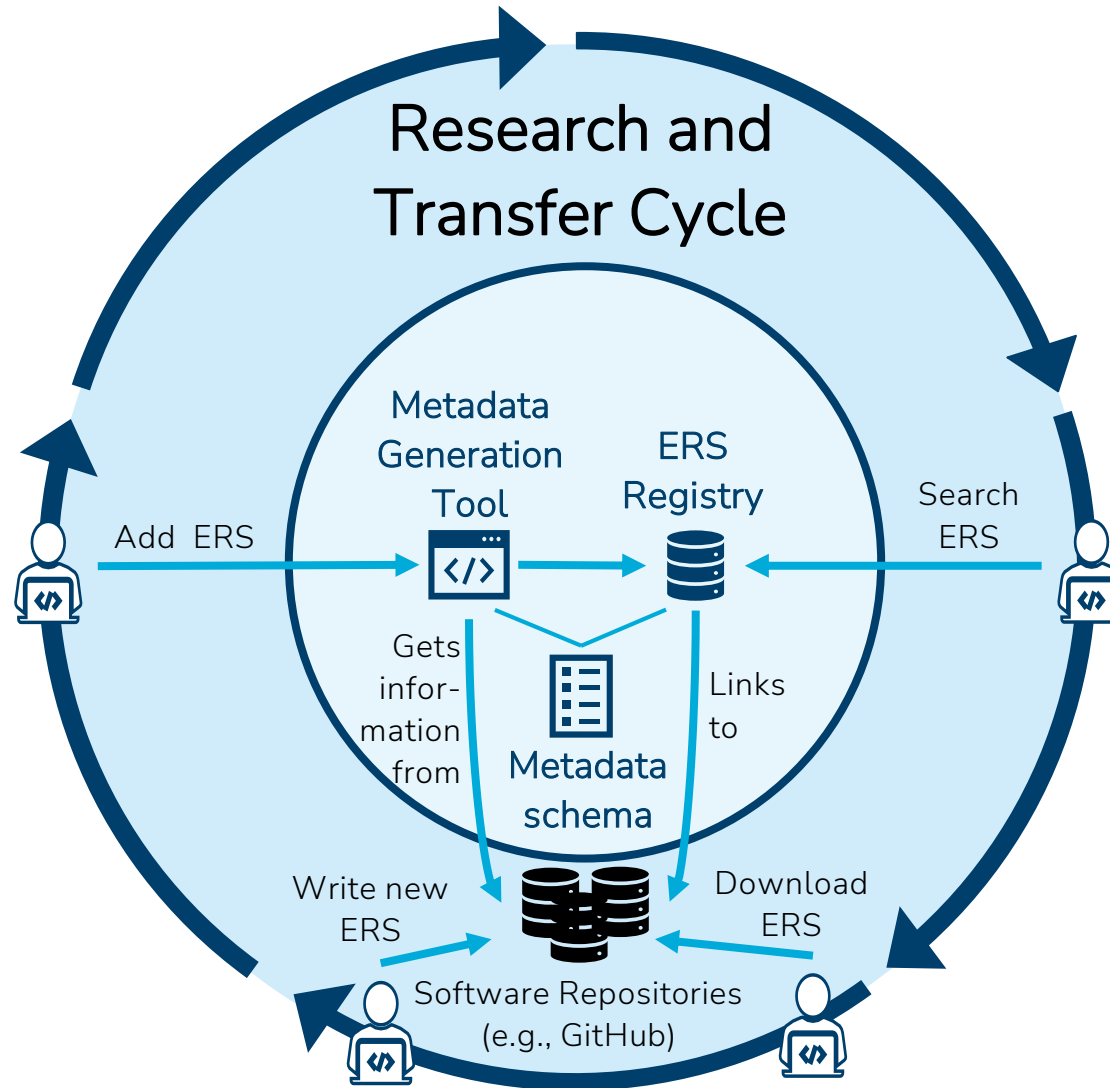
Vision

Metadata-based Registry for Energy Research Software

Metadata-based Registry for Energy Research Software



Metadata-based Registry for Energy Research Software



Requirements 1/2

- Useable for all types of energy research software
 - From long-term to short-term
- Application profile for energy research software
 - Reuses elements from existing metadata schemas
 - Improves FAIRness (Findable, Accessible, Interoperable, Reusable)
 - Follows best practice in metadata schema creation
 - Uses ontologies as value vocabularies
 - Compatible with existing metadata, e.g., PyPI



Requirements 2/2





- Registry
 - Repository for metadata
 - Includes good search functionality
 - Supports all features of the application profile
 - Gives information on link-ability of software based on their metadata
- Metadata Generation Tool
 - Usable for all researchers in the energy domain without further knowledge
 - Automatically collects metadata from GitHub, GitLab, Google Scholar
 - Supports the use of value vocabularies (existing ontologies)



Related Work



Overview

Scope					
CodeMeta	General Research Software	✗	✓	✗	~
OntoSoft	Research Software in GeoScience	✗	✓	✓	~
Bio.Tools	Research Software in LifeScience	✗	~	✓	~
EngMeta	Research Data in Engineering	~	✓	~	✓
Open Energy Platform	Research Software in Energy Research	✓	~	✓	~

Open Energy Platform

- Framework Factsheet
 - Not a formalized schema
 - Not based on existing schemas
 - Does not yet support the use of a value vocabulary
- Open Energy Ontology [10]
 - Domain Ontology for the energy domain
 - Can be used for annotating data & software

Framework Factsheet: Mosaik

Overview

Last updated ?	2021-05-08
Open Source ?	✓
License ?	GNU Lesser General Public License v3.0



General Information

Mathematical Description

Openness

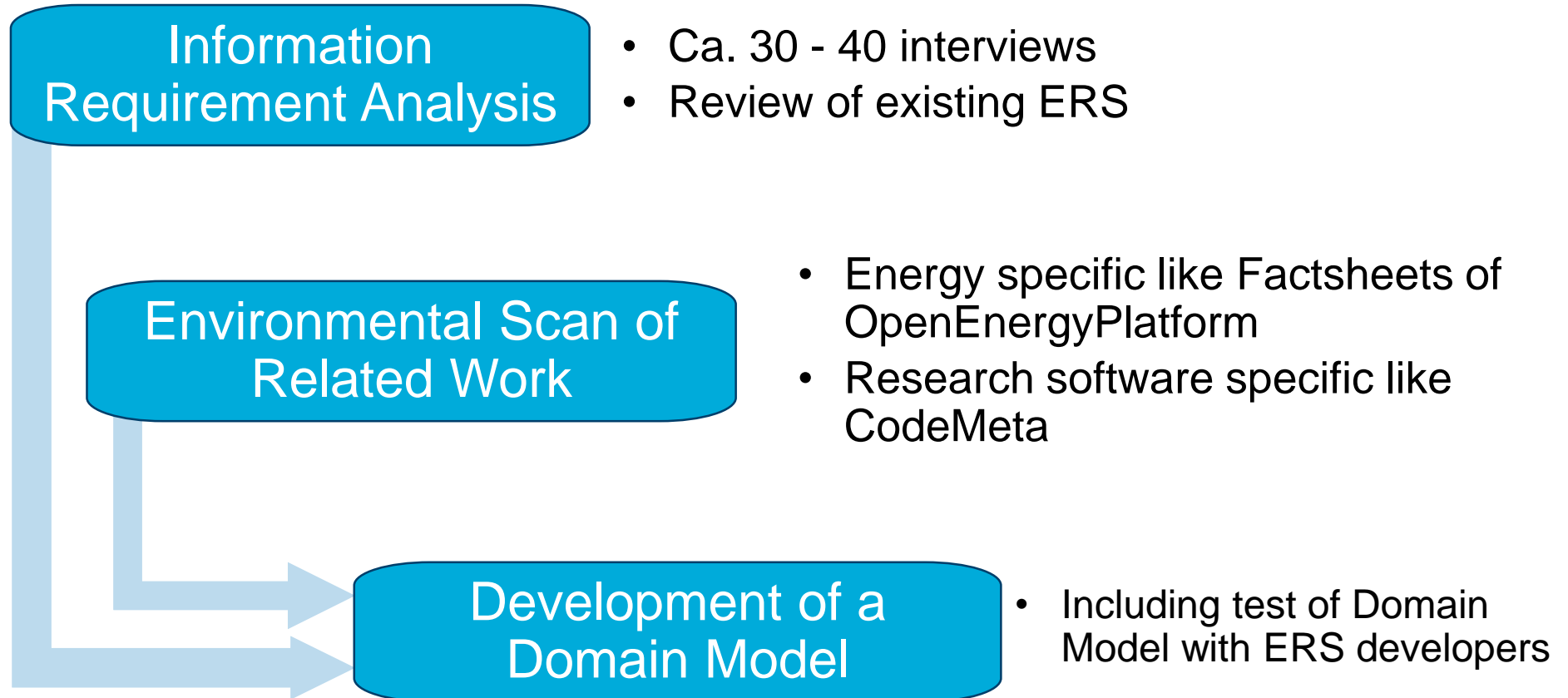
Model Building

References

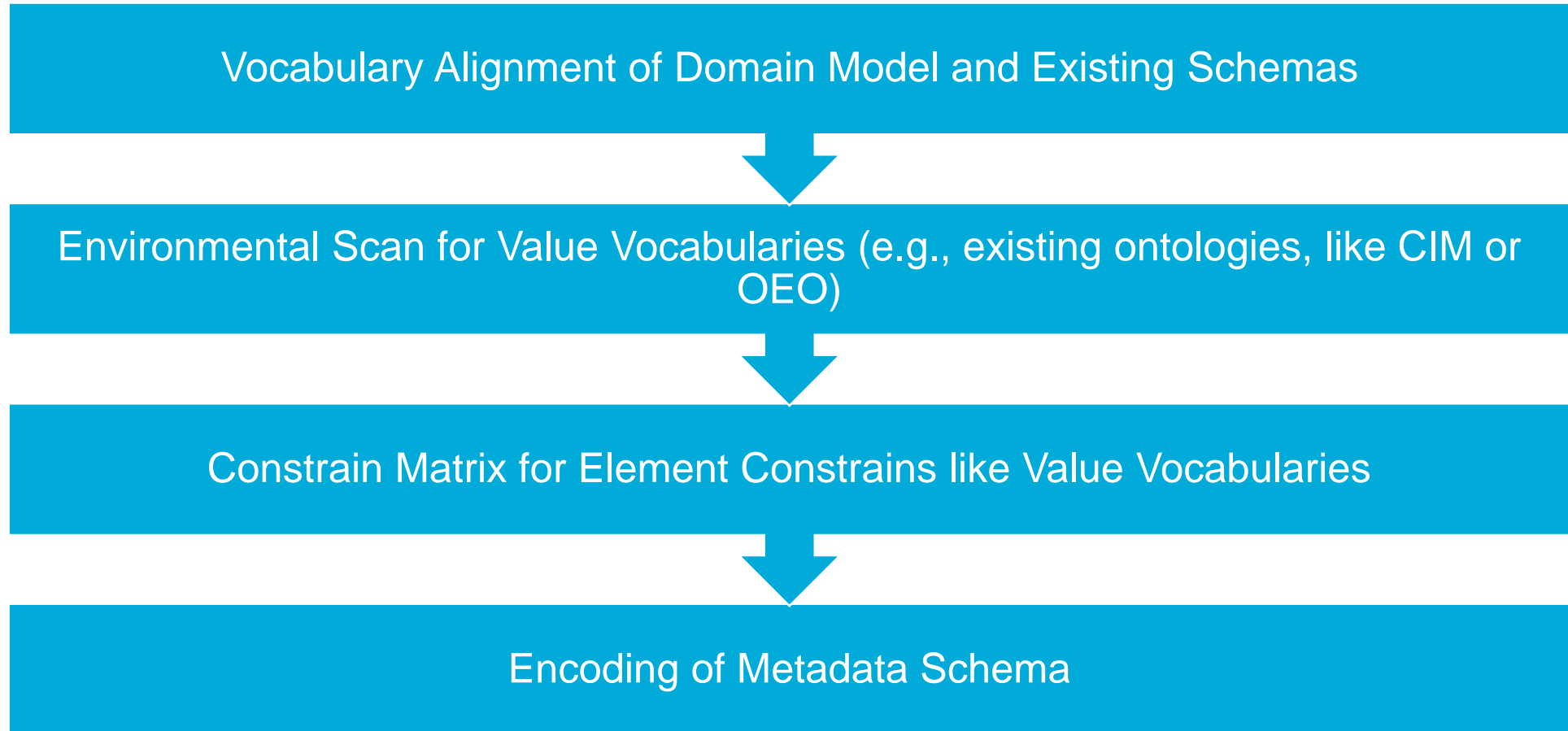


Own Approach

Approach Step 1: Development of Domain Model (like proposed by [11]) Based on a Requirement Analysis



Approach Step 2: Formalization of a Metadata Schema for ERS



Approach Step 3: Development of a Registry for ERS



- Repository for metadata
- Use of the metadata schema
- Based on existing framework like ckan (similar to TIBs Leibniz Data Manager)
- Easy usable artifact overview pages like on bio.tools
- Development of search technology based on semantic web technologies

Approach Step 4: Building a Metadata Generation Tool for ERS



- Tool supporting researchers in generating metadata
- Support for the use of value vocabulary
- Automatically extracts metadata from GitHub, GitLab and Readme files

Summary and Outlook

- What is needed to set up a metadata-based registry which can improve the Findability of Energy Research Software?
- Approach

• Domain Model based on a Requirement Analysis



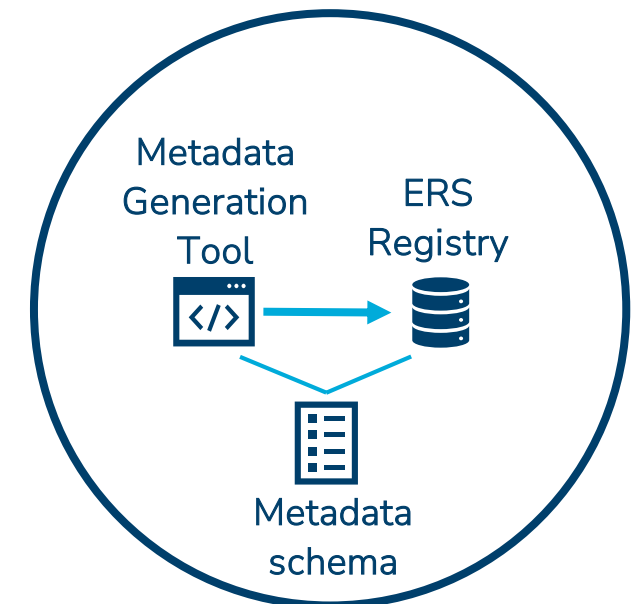
• Metadata Schema for ERS



• Registry for ERS



• Metadata Generation Tool for ERS



Appendix: Literature 1/2

[1] M. D. Wilkinson *et al.*, “The FAIR Guiding Principles for scientific data management and stewardship,” *Sci Data*, vol. 3, no. 1, Art. no. 1, Mar. 2016, doi: 10.1038/sdata.2016.18

[2] A.-L. Lamprecht *et al.*, “Towards FAIR principles for research software,” *Data Science*, vol. 3, no. 1, pp. 37–59, Jan. 2020, doi: 10.3233/DS-190026.

[3] 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, doi: 10.1515/itit-2019-0040.

[4] D. S. Katz, M. Gruenpeter, and T. Honeyman, “Taking a fresh look at FAIR for research software,” *PATTER*, vol. 2, no. 3, Mar. 2021, doi: 10.1016/j.patter.2021.100222.

[5] M. B. Jones *et al.*, “CodeMeta.” Github, 2017. Accessed: Oct. 20, 2021. [Online]. Available: <http://ssi1.eprints-hosting.org/id/eprint/2/>

[6] Y. Gil, V. Ratnakar, and D. Garijo, “OntoSoft: Capturing Scientific Software Metadata,” in *Proceedings of the 8th International Conference on Knowledge Capture*, New York, NY, USA, Oct. 2015, pp. 1–4. doi: 10.1145/2815833.2816955.

[7] 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.

Appendix: Literature 2/2

- [8] J. Ison 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, doi: 10.1186/s13059-019-1772-6.
- [9] B. Schembera and D. Iglezakis, “EngMeta -- Metadata for Computational Engineering,” *IJMSEO*, vol. 14, no. 1, p. 26, 2020, doi: 10.1504/IJMSEO.2020.107792.
- [10] M. Booshehri et al., “Introducing the Open Energy Ontology: Enhancing data interpretation and interfacing in energy systems analysis,” *Energy and AI*, vol. 5, p. 100074, Sep. 2021, doi: 10.1016/j.egyai.2021.100074.
- [11] M. Curado Malta and A. A. Baptista, “A Method for the Development of Dublin Core Application Profiles (Me4DCAP V0.2): Detailed Description,”
- [12] 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, 2019, pp. 127–136. doi: 10.5220/0008069801270136.