

# Towards a Metadata Schema for Research Software Quality

## Motivation

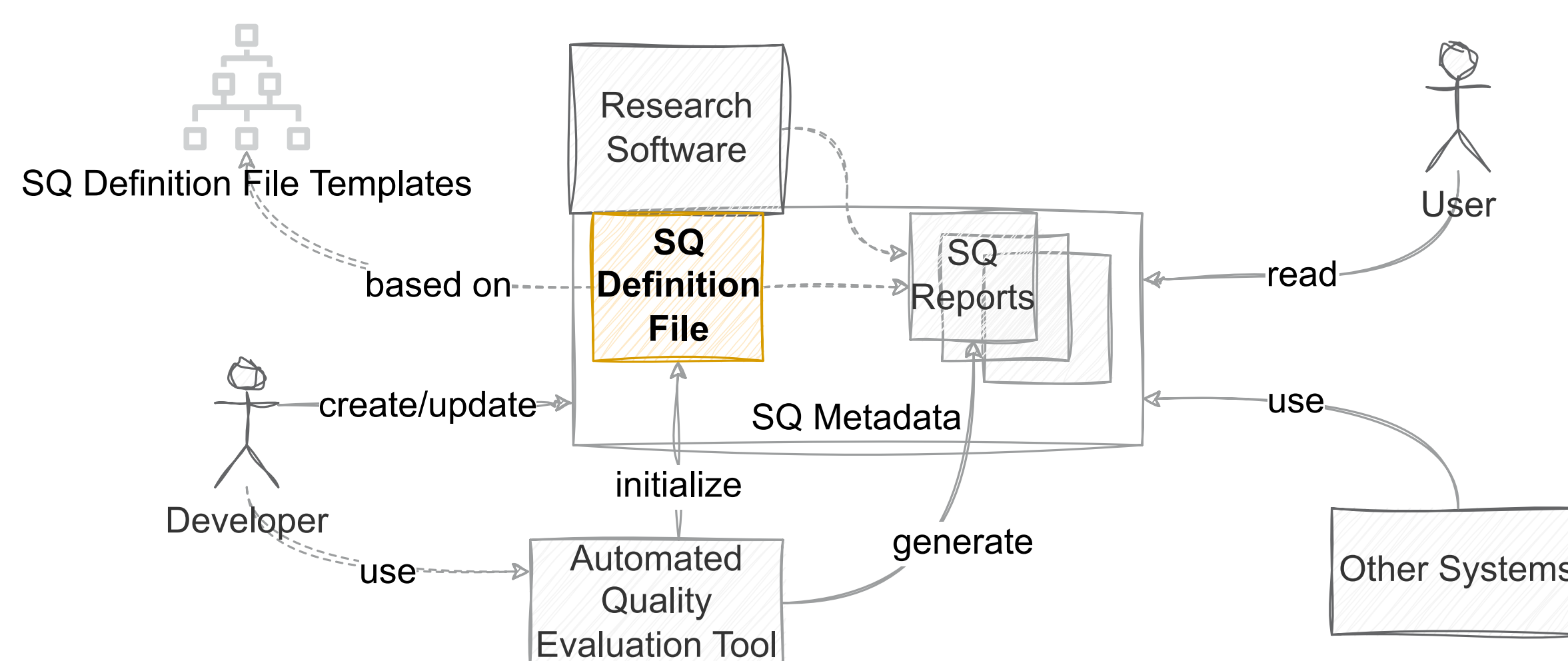
Advantages of research software (RS) quality information

- Criterion when looking for software to reuse
- Deepens our understanding of quality practices employed in RS

Challenges

- External assessment is limited
- Internal quality assurance processes are not always visible for others
- Actively developed software changes frequently

→ Enable developers to provide quality information as structured metadata



## Envisioned Concept

- Quality goals and measures are described in an *SQ Definition File* and function as a recipe
- *SQ Reports* are generated based on them and include concrete results for a specific version
- Templates and tools help developers creating and maintaining the quality metadata of their RS

FOUNDATION

IMPLEMENTATION

EVALUATION

The schema is based on the **ISO/IEC 25000 Systems and software Quality Requirements and Evaluation (SQuaRE) standards**, especially their ontology of quality concepts [3] and quality requirement specification [4].  
In previous work, we compared common quality characteristics for RS in HPC to the SQuaRE quality model and found high coverage but the occurrence of **tradeoffs between quality characteristics** [5].

To the best of our knowledge, existing metadata approaches for RS, e.g., the ones found in [2], do not cover software quality to the envisioned extent but should be used if general software metadata is needed to avoid duplication.  
However, there are **models for data quality, especially the Data Quality Vocabulary (DQV)** [1], that are partly based on SQuaRE's data quality model.

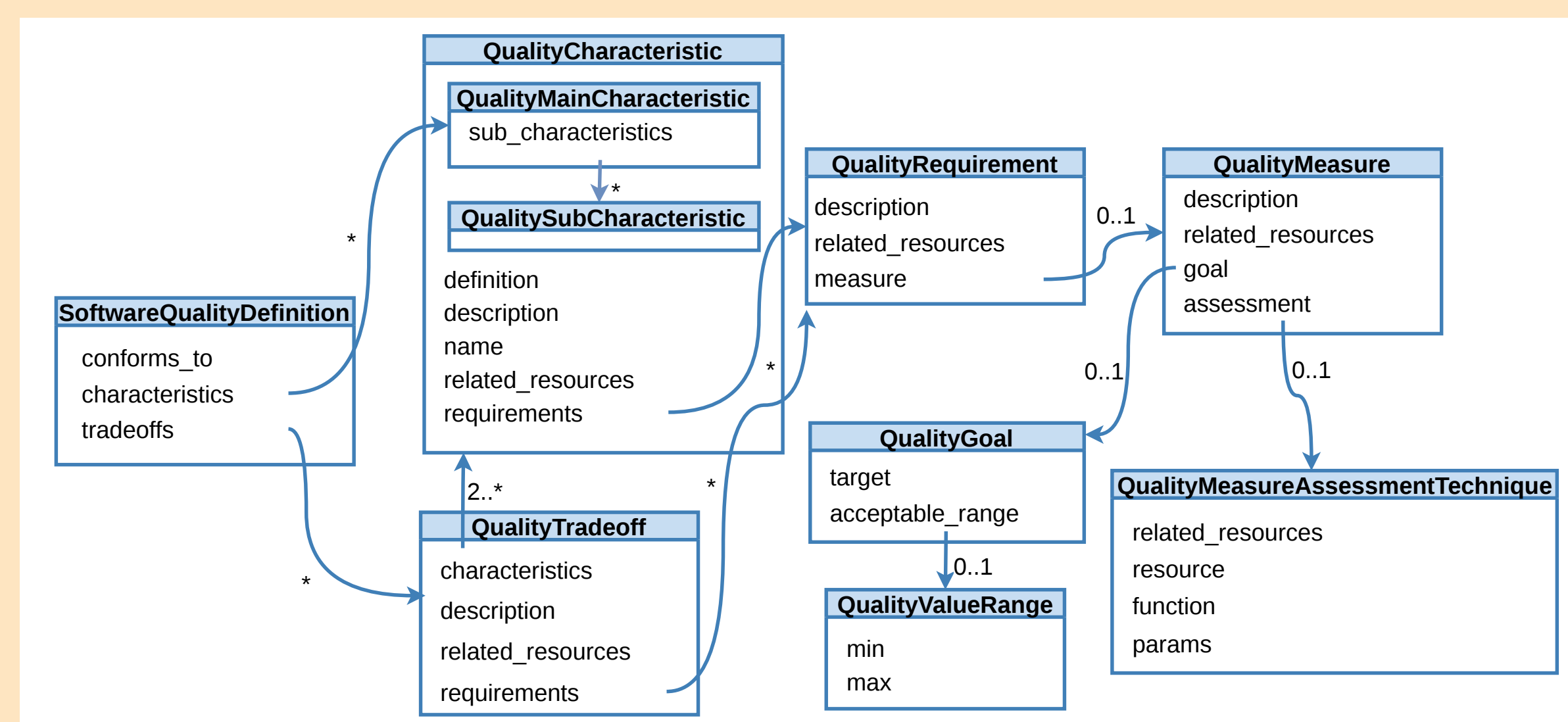
We derived the following **requirements from the envisioned concept**:

- Structured format, readable for humans and machines
- Technical details to allow report generation
- Option to add custom requirements and measures, as well as linking external resources

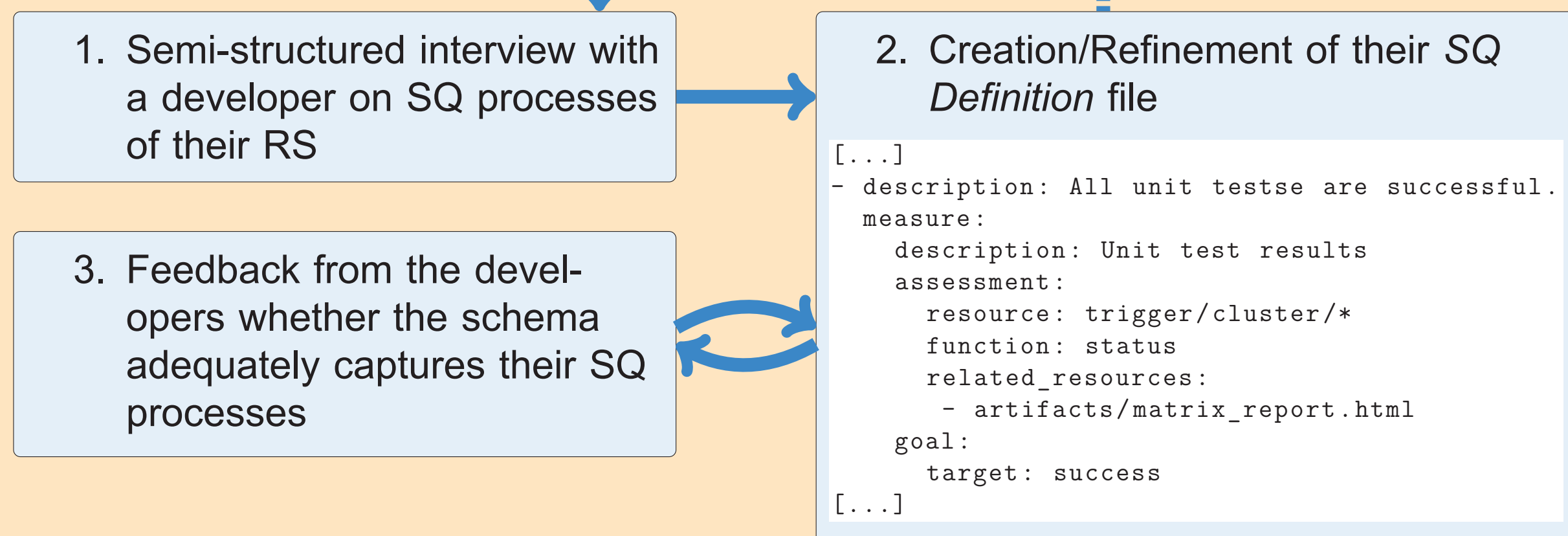
## Software Quality Definition Schema

- Implemented using the Linked Data Modeling Language (LinkML) [6]
- *SQ Definition* schema is separated into a base schema and an extension that predefines the SQuaRE characteristics
- Schema is written in YAML but artifacts into different formats are created
- Tooling includes documentation and validation

Online documentation:



## Iterative Process



So far conducted with two projects:

- MUST (MPI runtime error detection tool)
- RWTH fork of JARDS (resource allocation management service)

## Preliminary Results & Discussion

- ✓ Created files are **overall understandable and accurately represent the project's quality processes**
- Finding suitable quality characteristics led to discussions as practices often fit multiple characteristics
  - Using standards for requirements and measures could facilitate creation and comparison
- The term *quality* was perceived as potentially off-putting by some
  - Documentation and schema terms should emphasize processes rather than judging quality
- Further areas for refinement were identified:
  - Wording can be confusing regarding whether it is meant to describe the current state or the goals
  - It is unclear whether functional and technical capabilities of the software should also be included
  - A formal way to describe at which process stages assessments occur is missing

## Next Steps

- Address remaining evaluation feedback
- Extend the framework
  - Extend the schema with *SQ Reports*
  - Implement tools for automation to facilitate adoption
- Conduct broader evaluation
  - Repeat the process with additional HPC RS projects
  - Evaluate usability when RSEs create the metadata independently

## References

- [1] Riccardo Albertoni and Antoine Isaac. Introducing the Data Quality Vocabulary (DQV). In: *Semantic Web 12.1* (Nov. 2020), pp. 81–97. DOI: 10.3233/sw-200382.
- [2] Leyla Jael Castro et al. Results on a Survey on Research Software Metadata in the NFDI Consortia. In: *Zenodo* (2024). DOI: 10.5281/ZENODO.12704414.
- [3] International Organization for Standardization. *Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model overview and usage*. 2024.
- [4] International Organization for Standardization. *Systems and software engineering — Systems and software quality requirements and evaluation (SQuaRE) — Quality requirements framework*.
- [5] Camilla Lüttgens and Marius Politze. Quality Characteristics for Software in HPC Environments: A Systematic Literature Review. In: *Electronic Communications of the EASST 85* (2025). DOI: 10.14279/ECEASST.V85.2690.
- [6] Sierra AT Moxon et al. The Linked Data Modeling Language (LinkML): A General-Purpose Data Modeling Framework Grounded in Machine-Readable Semantics. In: *ICBO 3073* (2021), pp. 148–151.