

University of Stuttgart

Cluster of Excellence "Data-integrated Simulation Sciences"

University Library



**Blickwechsel: Erfahrungen
bibliotheksferner
Wissenschaften bei der
Nutzung forschungsnaher
Dienste**

Sibylle Hermann | BiblioCon 2024 | 7.6.2024

Agenda

- Short introduction
- Research data management from a research perspective
- Research data management from a SimTech perspective
- Implications



I am



A MECHANICAL
ENGINEER

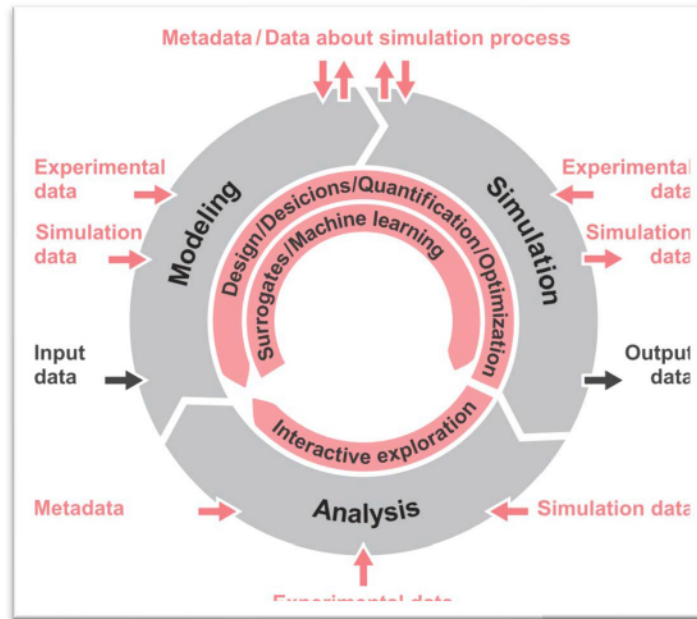


FROM INFRASTRUCTURE
(LIBRARY/SIMTECH)

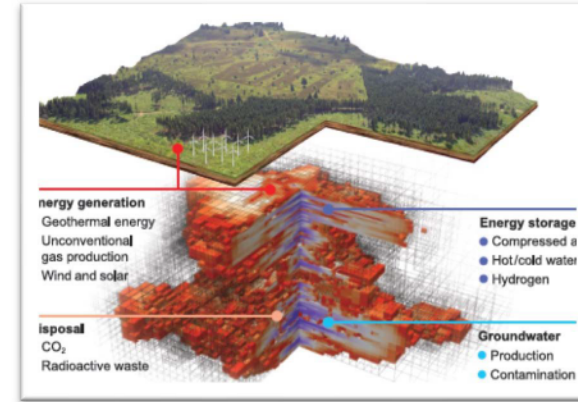


DOING RESEARCH
ABOUT RDM
(ITM/SIMTECH)

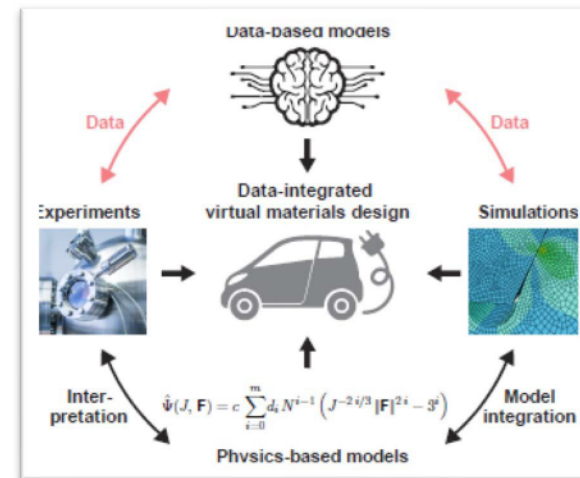
Modelling, Simulation and Analysis to Solve Visionary Examples



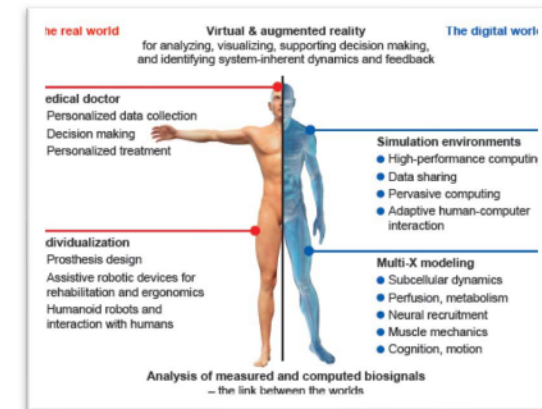
Common Methodology



Engineered Geosystems



Next-Generation Virtual Materials Design



Digital Human Model

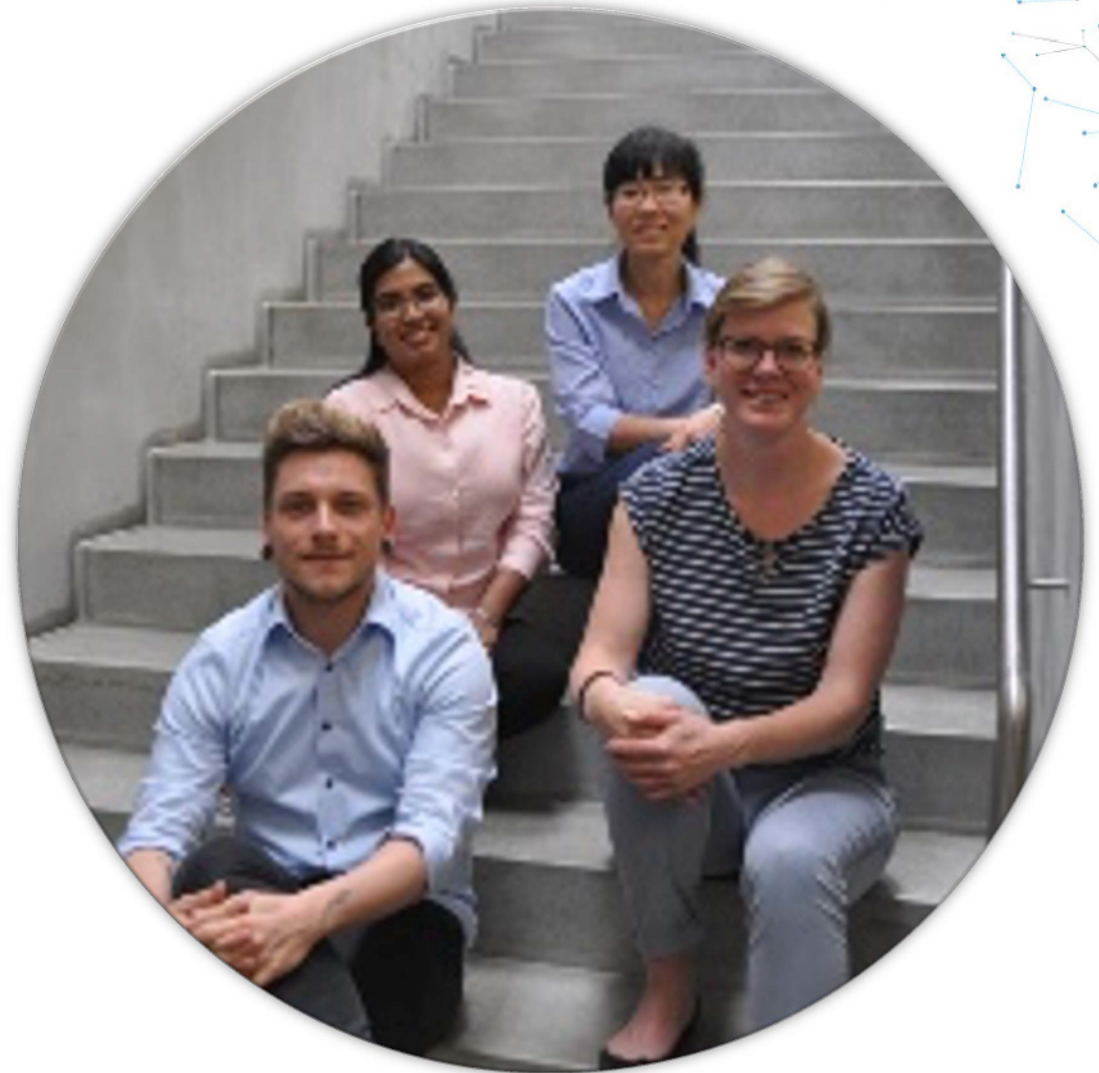
„Business Model“

- Agreement that I am delegated to the Cluster with a half-time position
- Compensation payment from the cluster to the UB for the half position
- Research Data and Software Steward
- PhD about Open Science in Simulation



Research Data Management at SimTech

- Research data and software important research output
- Research data management committee with researchers and supporting staff
- Team with 4 persons, who are supporting research data management at SimTech
- Close cooperation and use of the university infrastructure
- Over 50 published research data and software sets



FoKUS

- Competence Center for Research Data Management
- FoKUS is the contact for all questions concerning research data management (FDM): from data management plans to storage space, administration, publication and archiving.
- Integrating all RDM activities at the university
- Network in the RDM community
- Participation in the NDFI4ING and MARDI
- Participation in DFG, BMBF and MWK projects
- DaRUS





Research Data Management from a Research Perspective

Why publish Research Data and Software?



Compliance with Funding
Requirements



Ethical and responsible
research



Recognition of
scientific work

Guideline 13: Providing Public Access to Research Results

- Sharing the research data, materials, and information on which the results are based
- Making the methods and software used available
- Fully explaining the work processes

Explanations:

In the interest of transparency and to enable research to be referred to and reused by others, whenever possible researchers make the research data and principal materials on which a publication is based available in recognised archives and repositories in accordance with the FAIR principles (Findable, Accessible, Interoperable, Reusable). ...

Deutsche Forschungsgemeinschaft. (2022). Guidelines for Safeguarding Good Research Practice. Code of Conduct.
<https://doi.org/10.5281/zenodo.6472827>

What to publish?

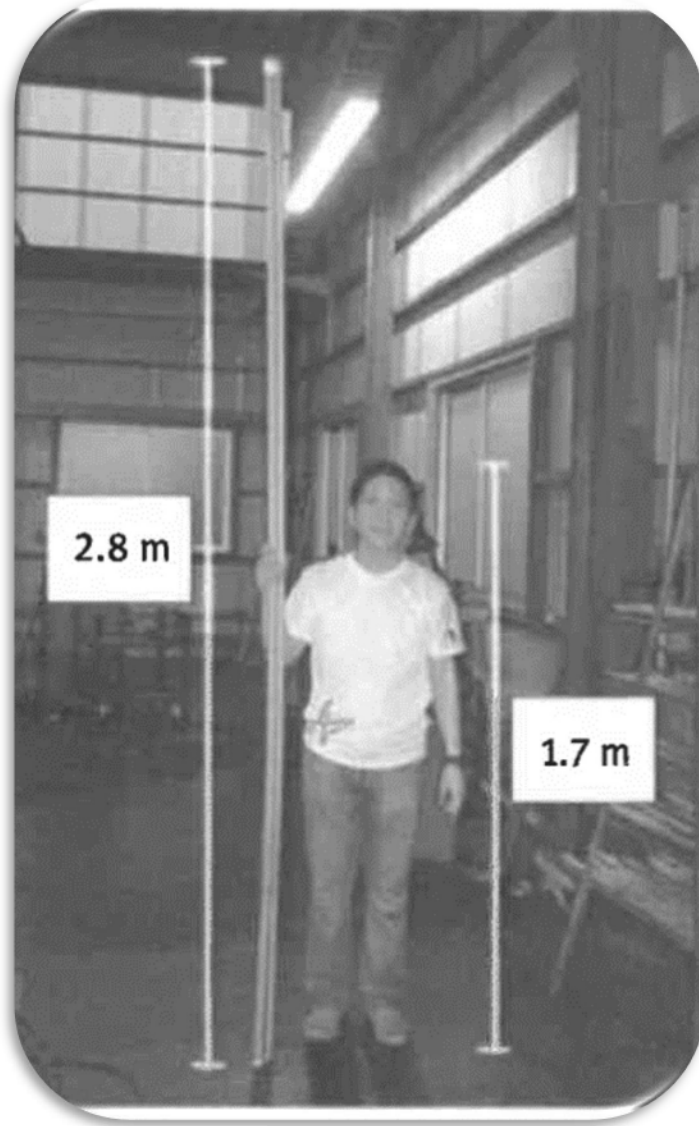
Every component of scientific research

- Experimental data
- Process and method of experiment
- Research software
- Analysing tools
- I/O data to the tools and software
- Guidelines to reuse
- ...



Credit: <https://wordart.com>

Deep Hole Drilling



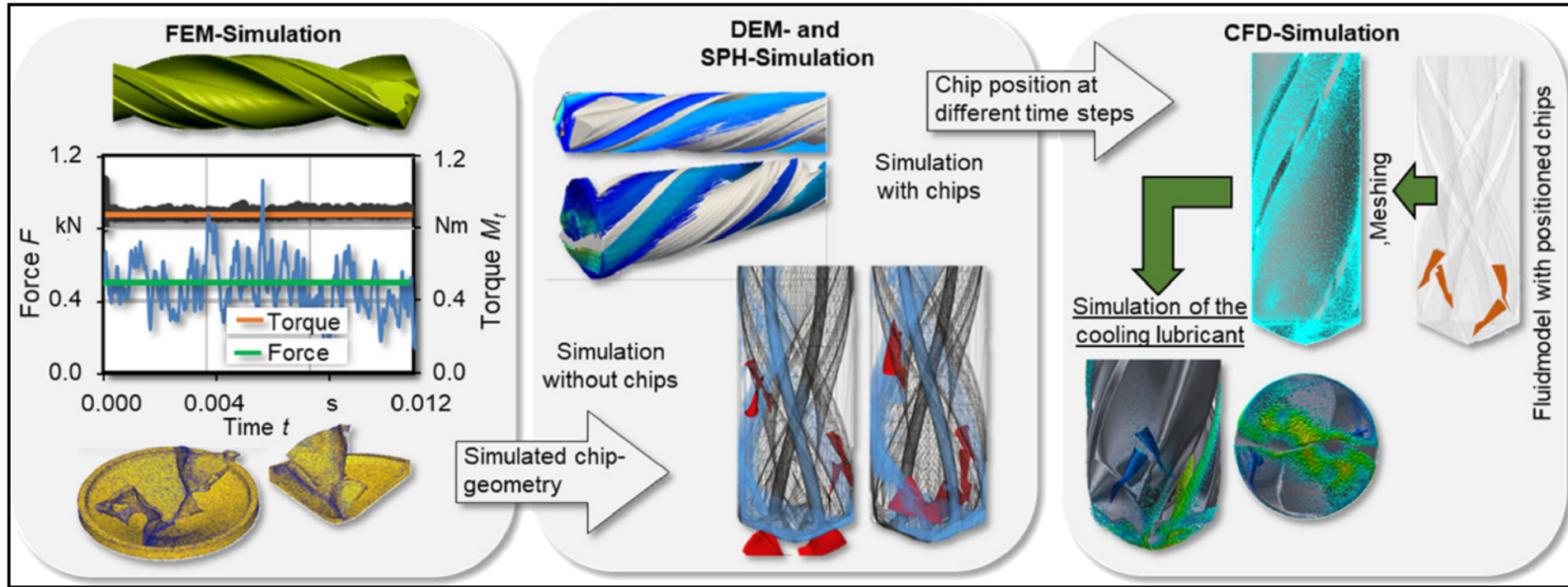
Source: WoonTnayYeao18

- Research question: how are chips removed during deep hole drilling
- Not possible to see what happens inside the drill hole in an experiment



Multi-X Simulation

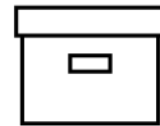
Multi-X simulation involves different spatial and temporal scales (**multiscales**), different coupled physical phenomena, fluids, materials (**multiphysics**), and different modeling approaches (**multimodel**), heterogeneous data fidelity (**multidata**) and different dimensionality and response requirements (**multicontext**).



Multi-X simulation on the example of deep hole drilling

What should be documented and published?

- Research builds on preliminary work done by other researchers who have already completed their PhDs.
- So the researcher reuses the work of his predecessors, and part of his research work is already being reused by other researchers.





Dilbert.com DilbertCartoonist@gmail.com



3-1-10 ©2010 Scott Adams, Inc./Dist. by UFS, Inc.



Guideline 12: Documentation

- **Document information relevant to the production of a research result as clearly as possible to allow the result to be reviewed and assessed.**
- This also includes documenting individual results that do not support the research hypothesis.

Explanations:

An important basis for enabling replication is to make available the information necessary to understand the research (including the research data used or generated, the methodological, evaluation and analytical steps taken, and, if relevant, the development of the hypothesis), to ensure that citations are clear, and, as far as possible, to enable third parties to access this information. Where research software is being developed, the source code is documented.

Deutsche Forschungsgemeinschaft. (2022). Guidelines for Safeguarding Good Research Practice. Code of Conduct.
<https://doi.org/10.5281/zenodo.6472827>

Research Data Management

RDM: focus on documenting results and process with metadata

FAIR: focus on finding and legal reusability

Science: focus on decisions and ideas

Conclusion: We need to understand the whole process in order to answer the question how and what to document.

Idea: Looking deeper into the **research process** from a **philosophy of science** perspective in order to understand what has to be documented

Insights

- Research results are more than software and data
- Often focus on application instead of method
- Method and application differ depending on the discipline
- Difficult to answer what the research output is
- Research output often not so „sexy“
- Need of awareness of what to document, what to publish and why



Research Data Management from an SimTech Perspective

Multi-modal Software and Data Repository: DaRUS



1. An instance of Dataverse.
2. Organize, store, publish and share data and software.
3. Assigns DOI.
4. Subject-specific metadata.
5. Several metadata formats.
6. Cross-linked to other resources.
7. Can link web applications.
8. Wide range of licences.
9. Possibility for restricted access.

DaRUS > SFB 1313 "Interface-Driven Multi-Field Processes in Porous Media" > Benchmarks > Verification benchmarks for single-phase flow in three-dimensional fractured porous media >

Case1 - Single Fracture

Version 1.0

Gläser, Dennis, 2020, "Case1 - Single Fracture", <https://doi.org/10.18419/darus-862>, DaRUS, V1, UNF:6:103YE3NiQa3v4NNImuR2yg== [fileUNF]

Cite Dataset - Learn about Data Citation Standards.

Description This dataset contains results related to the benchmark case 1 as described in detail in the publication

I. Berre, W. Boon, B. Flemisch, A. Fumagalli, D. Gläser, E. Keilegavien, A. Scotti, I. Stefansson, and A. Tatomir. Call for participation: Verification benchmarks for single-phase flow in three-dimensional fractured porous media, arXiv e-prints, art. arXiv:1809.06926, 2018

Subject Earth and Environmental Sciences; Mathematical Sciences

Related Publication I. Berre, W. Boon, B. Flemisch, A. Fumagalli, D. Gläser, E. Keilegavien, A. Scotti, I. Stefansson, and A. Tatomir. Call for participation: Verification benchmarks for single-phase flow in three-dimensional fractured porous media, 2018 arXiv: 1809.06926

License/Data Use Agreement CC BY 4.0

Files Metadata Terms Versions

Citation Metadata

Process Metadata

Software Dumux Version: 3.1 (<https://dumux.org/>) (Timo Koch, Dennis Gläser, Kilian Ackermann, Martin Beck, Beatrix Becker, Samuel Burbulla, Holger Class, Emmert, Thomas Fetzer, Christoph Grüninger, Katharina Heck, Johannes Melanie Lipp, Farid Mohammadi, Samuel Scherrer, Martin Schneider, Gabor Stadler, Martin Utz, Felix Weinhardt, Bernd Flemisch, DuMux 3 - an open-source simulator for solving flow and transport problems in porous media with a focus on model coupling, Computers & Mathematics with Applications, 2020, doi:10.1016/j.camwa.2020.02.012)

Access Dataset

Edit Dataset

Link Dataset

Contact Owner Share

Dataset Metrics 12,965 Downloads

Access Dataset

The dataset is too large to download. Please select the files you need from the files table.

Explore Options

Explore in Browser (VIPLab)

License/Data Use Agreement

This dataset will be published under the terms specified below. Our Community Norms as well as good scientific practices expect that proper credit is given via citation.

MIT License

CC BY-NC-ND 4.0

CC BY-NC-SA 4.0

CC BY-ND 4.0

CC BY-SA 4.0

GPL 3.0 or later

AGPL 3.0 or later

LGPL 3.0 or later

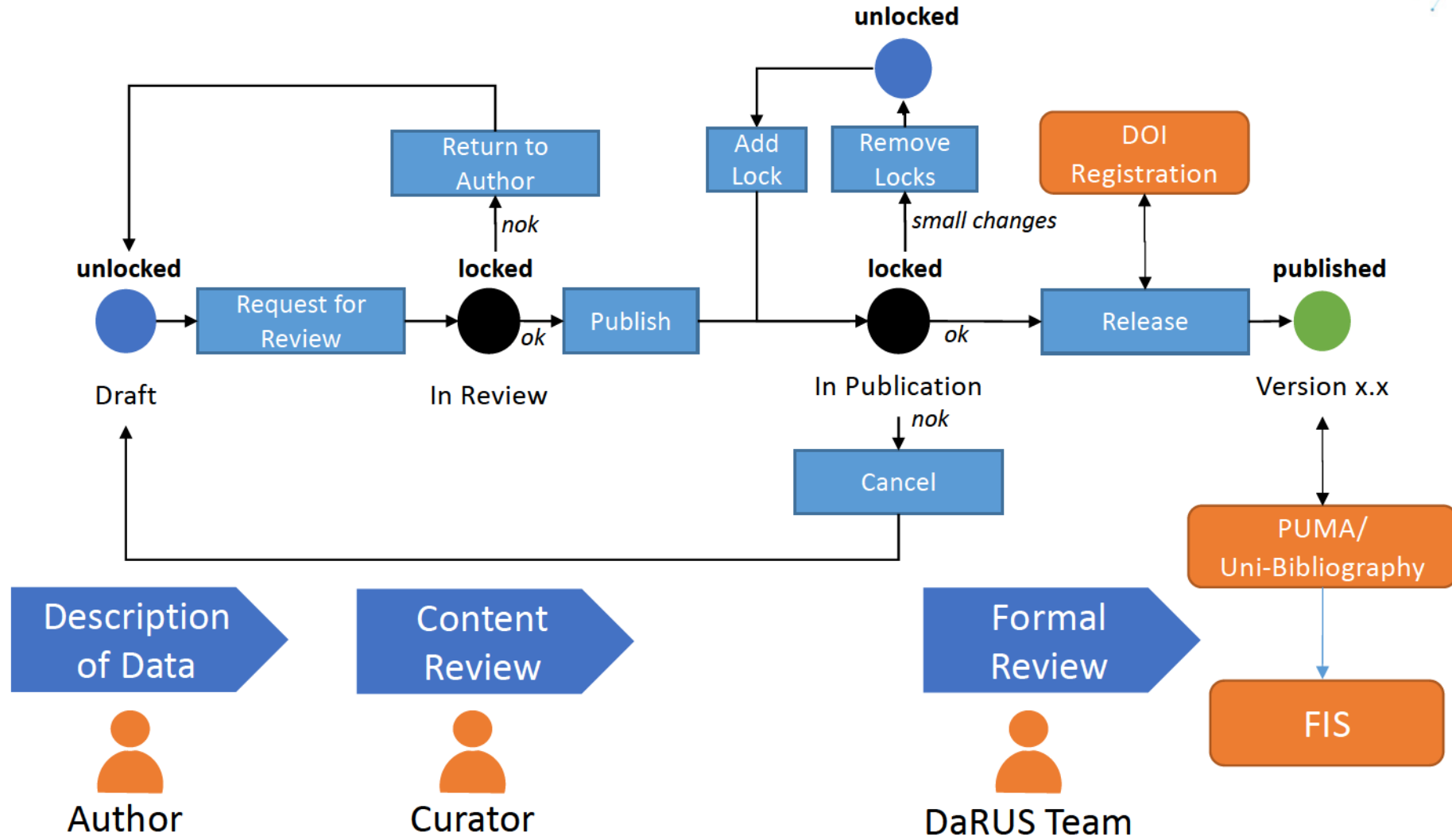
MIT License

Restricted Files + Terms of Access

Request Access

Terms of Access Information on how and if users can access restricted files in this Dataset

Publication Workflow



Checklists and more information: <https://www.izus.uni-stuttgart.de/en/fokus/darus/publication/>

Build Infrastructure to facilitate Publication Workflow

SimTech Software Website:
<https://simtech-software-website.vercel.app>

Flemisch, B., Hermann, S., Herschel, M. *et al.* Research Data Management in Simulation Science: Infrastructure, Tools, and Applications. *Datenbank Spektrum* (2024).
<https://doi.org/10.1007/s13222-024-00475-4>

The screenshot displays the SimTech Software Website interface. On the left, a navigation menu lists 'All topics' including Data Analysis, Data Models, Database Tools, Machine Learning, Publication Tools, Simulation Software, Visualisation, and Workflow Manager. The main content area features a grid of software tool cards. The 'Data Models' section is highlighted with a count of 4 items. The cards shown are: 'EnzymeML' by Range, Jan; Pleiss, Jürgen (tags: biocatalysis, enzymology, research data model); 'Porous Media Flow Model' by Range, Jan; Fuhrmann, Tobias (tags: continuous flow dynamics, data management, piv measurements, research data model); 'Software-driven Research Data Management' by Range, Jan (tags: python, data management, data models, markdown); 'ZnH5MD' by Zills, Fabian (tags: python, h5md, molecular dynamics); and 'sdRDM Database' by Range, Jan; Häußler, Max (tags: python, data management, graphql, object relational model). A 'Firefox' browser notification is visible in the bottom left corner.

Dataset Reviews: Current State

- Carried out via **email** communication from reviewer to user

Workflow

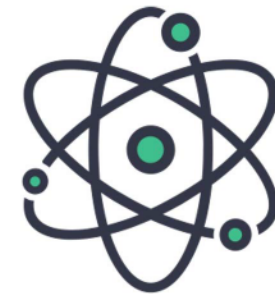
1. User submits dataset for review
2. Admin review contents and checks files
3. Manually written report
4. User implements changes



Dataset Reviews: Requirements to a Software Solution

Distribute reviews

- Invite experts to review datasets via URL
- Per field guidelines for any new reviewer

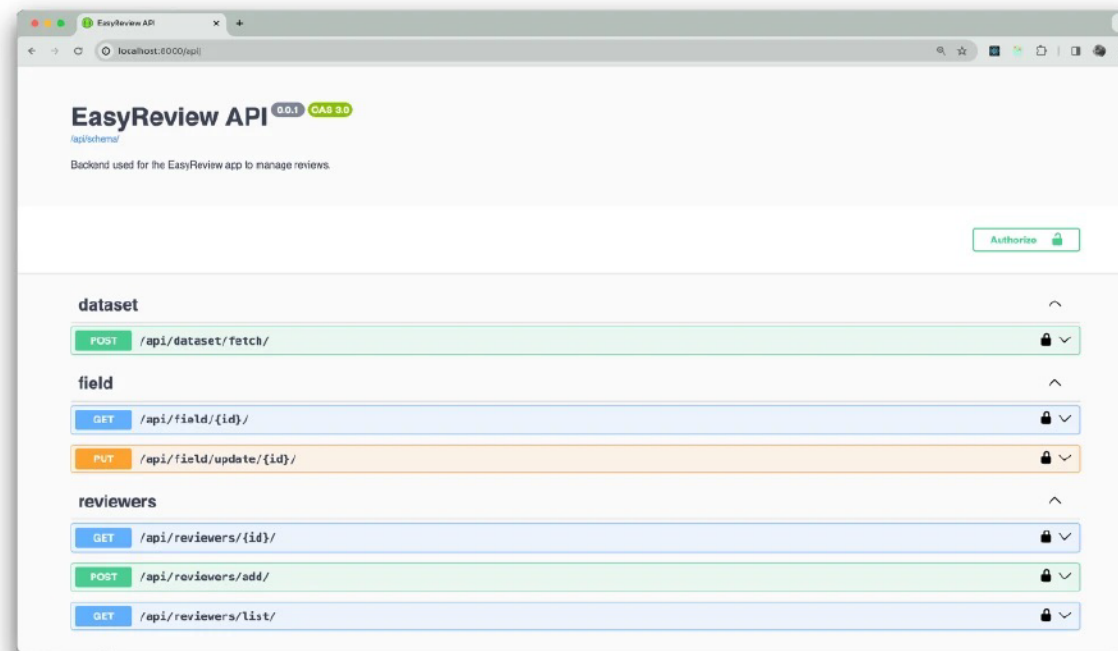
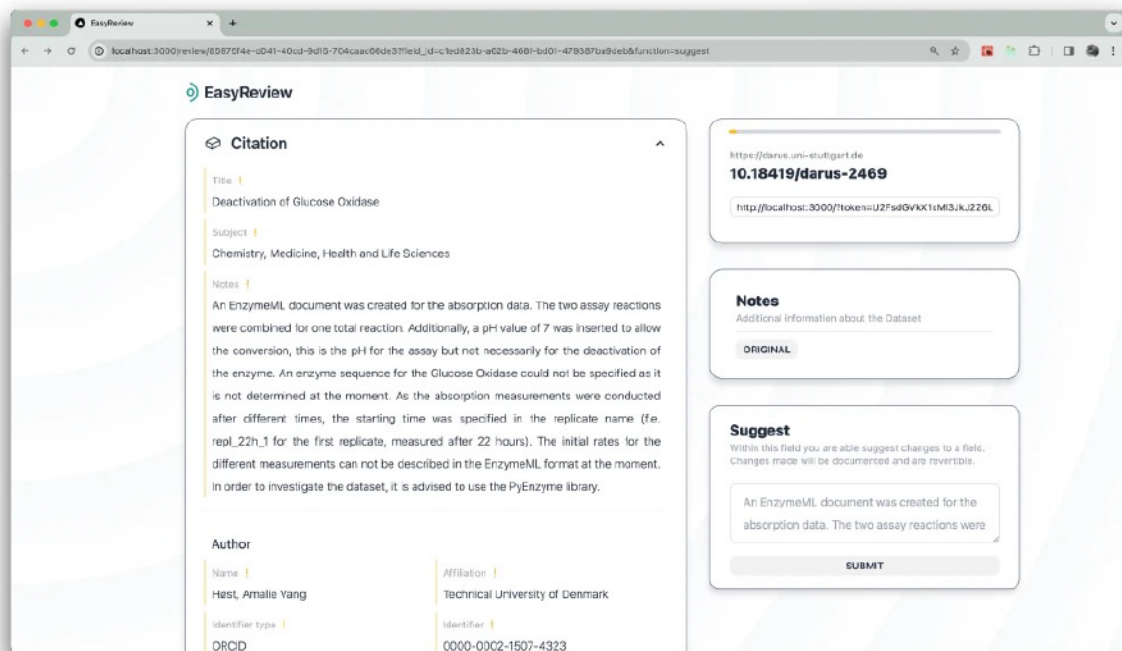


EasyReview

Collect multiple reviews

- Provide a way to present multiple review to a user

Easy Review: Prototype



Graphical user interface of EasyReview (*left*), which displays the metadata block in the *left column*. Visual indicators display the state of a field. In the *right column*, review statistics, field history and a text input for suggestions are shown. OpenAPI Swagger documentation of EasyReview's backend REST API (*right*) to automate dataset reviews

Implications

- The early integration of the infrastructure into the development process facilitates the identification of further developments by the scientific community.
- Enable the scientific community to further develop the existing infrastructure.
- The path also works in reverse, with a colleague starting at SimTech and then working half time on a library project.

Acknowledgements

- **Jan Range**
University of Stuttgart (SimTech)
- **Fangfang Wang**
University of Stuttgart (SimTech)
- **Sarbany Roy**
University of Stuttgart
(SimTech/FoKUS)
- **Dorothea Iglezakis**
University of Stuttgart (FoKUS)
- **Andreas Baumann**
University of Stuttgart (ITM)
- **Jörg Fehr**
University of Stuttgart (ITM)

Sibylle Hermann
University of Stuttgart
Cluster of Excellence “Data-integrated Simulation Sciences”
And University Library
email: Sibylle.hermann@simtech.uni-stuttgart.de

Questions?