Jarves: The digital RDM assistant

Joint Assistant for Research in Versatile Engineering Sciences

21.03.2023

Author:
Tobias Hamann | ORCiD 0000-0002-8021-5524
RWTH Aachen University | WZL | Organizational Development | Industrial Capabilities
Campus Boulevard 30 | 52074 Aachen | GERMANY
Phone: +49 151 72921981 | t.hamann@wzl-mq.rwth-aachen.de | www.wzl.rwth-aachen.de

Tobias Hamann
WZL | RWTH Aachen University
Chair of Production Metrology and Quality Management & Institute for Information Management in Mechanical Engineering
(Prof. Robert Schmitt)

Story Time: Two researchers are managing their research data

Task: Conduct research data management for a publicly funded project

Initial situation:

- Several years of experience in research
- Would now like to manage research data, as knowledge has been lost several times after colleagues left, which was difficult or impossible to reconstruct
- Has never before managed his research data

Gerd Papernote
Research associate

Where and how should I start with RDM?

Initial situation:

- Several years of experience in research
- Would now like to manage research data, as knowledge has been lost several times after colleagues left, which was difficult or impossible to reconstruct
- Has never before managed her research data, but has heared of Jarves

Franziska Research
Research associate

https://pixabay.com/de/photos/mann-schreiben-laptop-rechner-2562325/

https://pixabay.com/de/photos/laptop-apfel-macbook-rechner-2557571/

Story Time: Two researchers are managing their research data

Task: Planning, reuse and collect data

How do I document my data management and data collection?

- Not knowing what a DMP is, Gerd reads a summary about DMPs and writes one in Word
- He has to determine his parameters from scratch, as he does not reuse any data
- He stores his data on his PC
- He additionally saves his collected data on the institute drive in a project folder so that it does not get lost

Gerd Papernote
Research associate

Now I would like to analyze my data.

How do I document my data management and data collection?

- Franziska also does not know, what a DMP is, but Jarves shows her tutorials telling her what a DMP is and what it is used for. It also sends her project data to RDMO with a click.
- Using a recommender system and based on Franziska's input, Jarves suggests repositories that contain re-usable data for her project
- Jarves suggests to store the data using Coscine

Franziska Research
Research associate

https://pixabay.com/de/photos/b%c3%bccher-forschung-bibliothek-regale-2562331/

https://unsplash.com/de/fotos/9lpSbMgYm0Q

Story Time: Two researchers are managing their research data

Task: Analyse data

How do I document my data analysis?

- Gerd evaluates his collected data by writing code for the evaluation in Python
- Gerd's code is written quickly and works, but comments in the code and a description of the required input are missing. After all, Gerd knows the input
- Gerd's code remains on his computer
- 8 months later he leaves the institute

Gerd Papernote
Research associate

Okay just a few more steps and i am finished.

How do I document my data analysis?

- Franziskas evaluates her collected data by writing code for the evaluation in Python
- Jarves points out that code should be well documented so that it can be reused and shows Franziska a training on Jupyter Notebooks
- Jarves recommends Franziska to use GitLab and links a Git tutorial
- 8 months later she leaves the institute

Franziska Research
Research associate

https://unsplash.com/de/fotos/iar-afB0QQw

https://unsplash.com/de/fotos/oZ61KFUQsus

Story Time: Two researchers are managing their research data

Charles Newguy searches for previous projects and associated data

What could Charles re-use and how?

- Publication and related published data
- There was data on the drive, but it is not sufficiently documented for Charles and the results cannot be reconstructed, amongst other things because the evaluation scripts are not available

I could hardly reuse anything because important information was missing....

Charles Newguy
Research associate

What could Charles re-use and how?

- Publication and related published data
- Raw data, evaluation script and results are available to members of the institute on Coscine and GitLab.
- Franziska's data has a higher value

I could benefit from Franziska's RDM!

https://unsplash.com/de/fotos/rWJ2RthM-gc

https://unsplash.com/de/fotos/774sCXD0dDU

What exactly is Jarves and what does it look like?

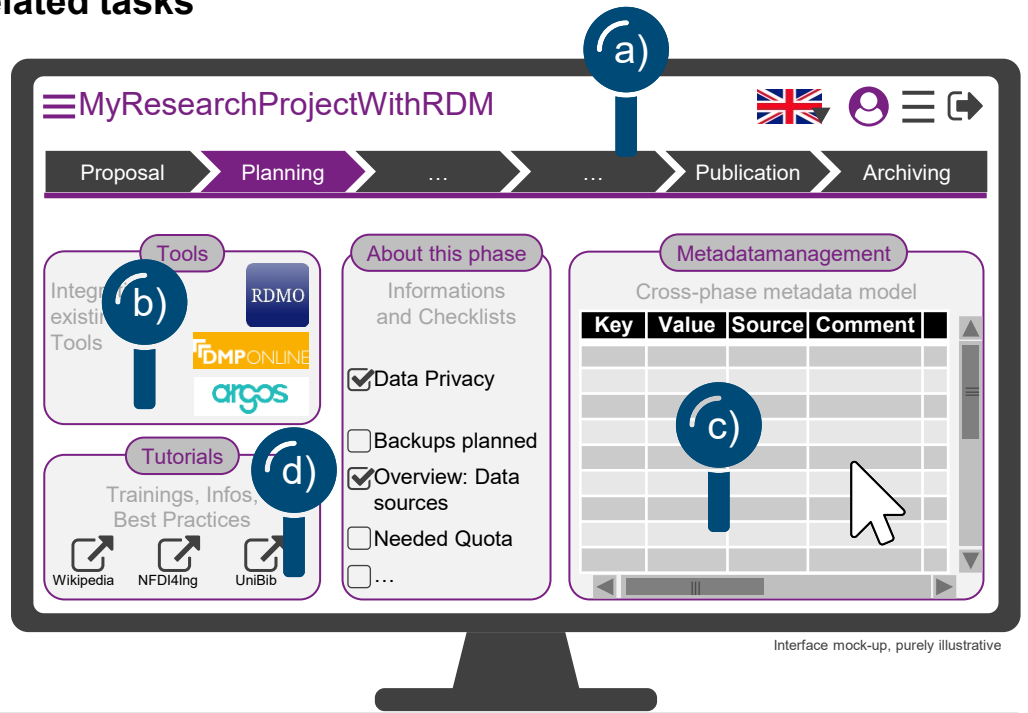
NFDi4ing

Charles Newguy
Research associate

Jarves – The idea

Based on a structured process, Jarves guides researchers through RDM, provides them with knowledge and tools for related tasks and supports them in making RDM decisions.

- Process management of (engineering) research projects in consideration of the RDM
- Support for researchers through
  - a) Support through the RDM process
  - b) Connection to existing (RDM) tools
  - c) Collection and exchange of metadata between phases of the RDM
  - d) References to best practices and training materials
- Current status : Implementation phase



Interface mock-up, purely illustrative

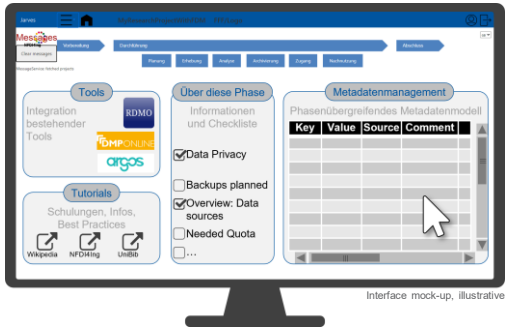
What is the core architecture of Jarves like?

NFDi4ing

Charles Newguy
Research associate

Can focus on research instead of RDM

Users



Interface mock-up, illustrative

Usage, data

Informations

Frontend: Angular

Enabling easy use of the tool by providing a graphical user interface

Request

Response

Data processing (database queries, calculations etc.)

Backend: Python Django with REST API

Communication via Django

e.g., RDMO, Terminology Service, …

Request

Response

Request

Response

Storage of data provided by users

Database: PostgreSQL/SQLite

Database: Global data	Database: Project-related data	Database: User data and authentication

Other Services

RDMO

Coscine

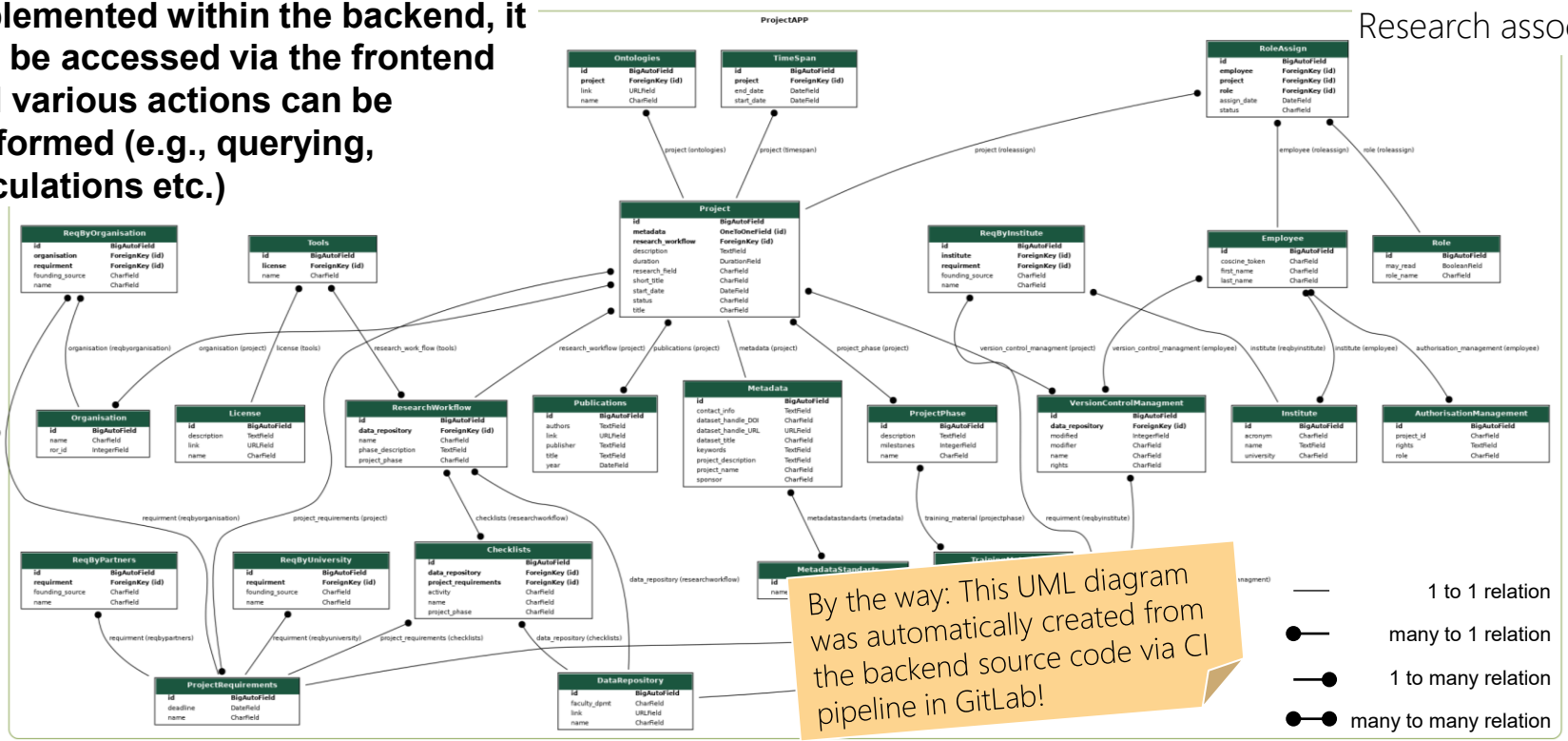
GitLab

What does the data model look like?

NFDi4ing

Charles Newguy
Research associate

The current data model (WIP) is implemented within the backend, it can be accessed via the frontend and various actions can be performed (e.g., querying, calculations etc.)



Jarves – Availability and further steps

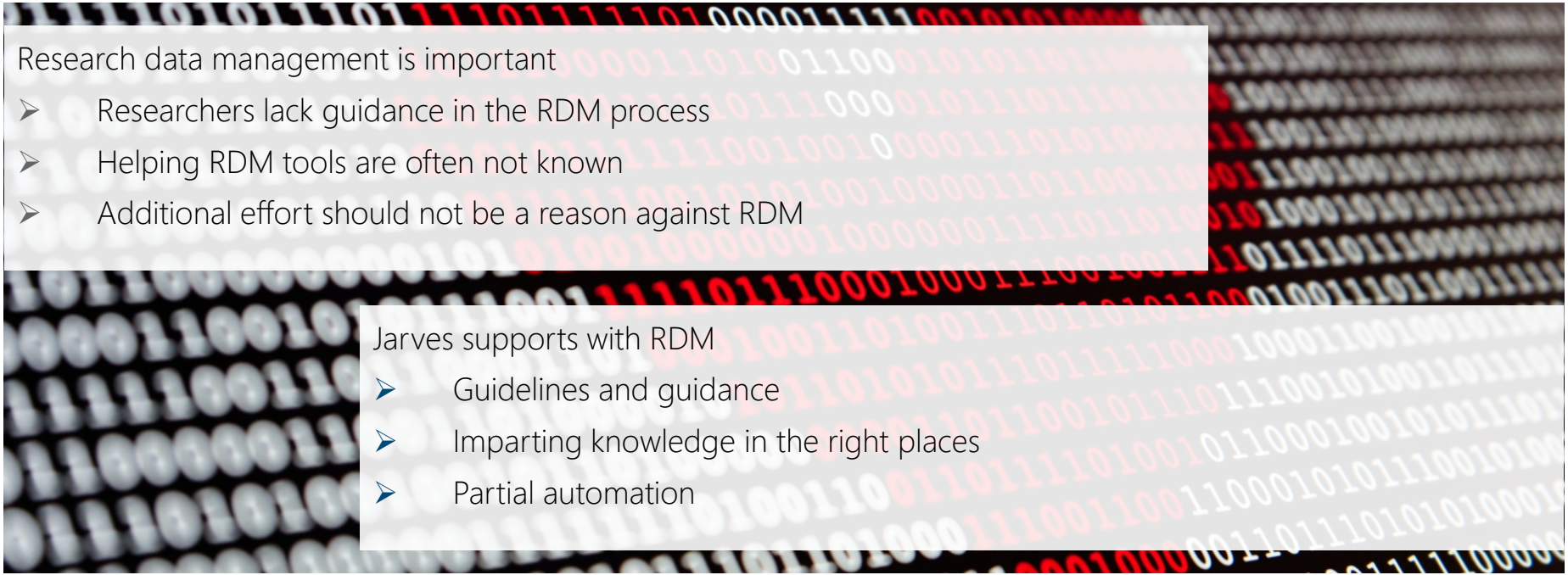
Charles Newguy
Research associate

Charles himself wants to use Jarves to simplify his RDM.

- Creation of the data model ✓
  → End of March 2023
- Define all workflow phases and associated tools
  → End of April 2023
- Implement API to RDMO and Coscine
  → End of August 2023
- Completion of the MVP
  → End of September 2023
- Afterwards: pilot phase and gradual expansion
  →Continuous

Take aways

What can you learn from this story?



Research data management is important

- Researchers lack guidance in the RDM process
- Helping RDM tools are often not known
- Additional effort should not be a reason against RDM

Jarves supports with RDM

- Guidelines and guidance
- Imparting knowledge in the right places
- Partial automation

https://unsplash.com/de/fotos/KgLtFCgfC28

Thank you for your attention!

You can find this lecture at: https://doi.org/10.5281/zenodo.7715684

Tobias Hamann
Contact:

- WZL of the RWTH Aachen University
- +49 151 72 92 19 81
- t.hamann@wzl-mq.rwth-aachen.de

License notice

Author: Tobias Hamann
WZL | RWTH Aachen University
Organizational Development | Industrial Capabilities

Campus Boulevard 30
52074 Aachen | GERMANY
Phone: +49 151 72921981
t.hamann@wzl.rwth-aachen.de
www.wzl.rwth-aachen.de
ORCiD 0000-0002-8021-5524