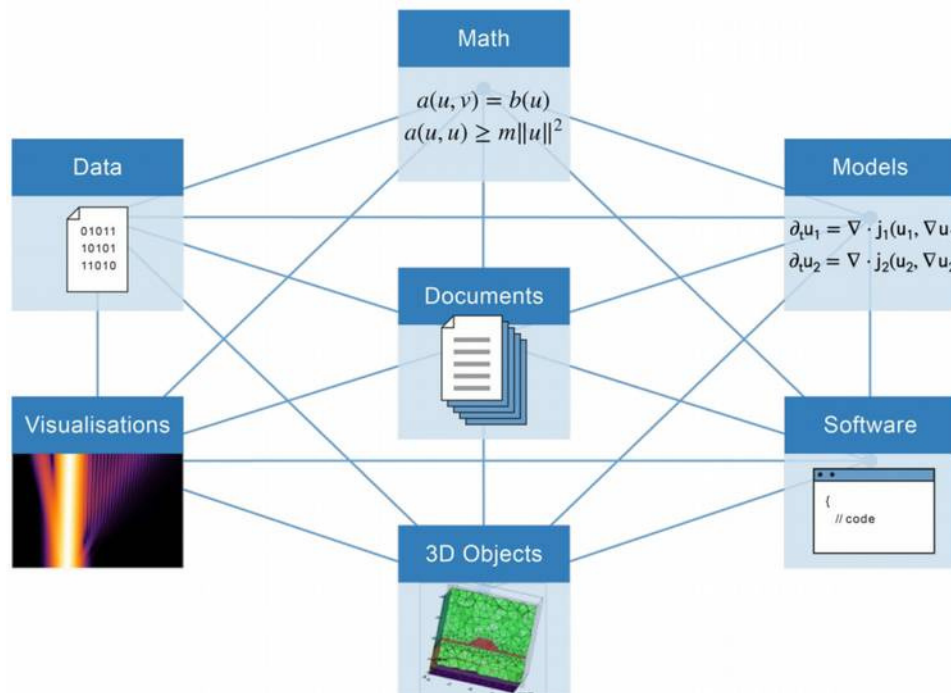


**An RDMO plugin to populate and query
the MaRDI Knowledge Graph.**

Marco Reidelbach – 29.03.2023

Mathematical Research Data Initiative



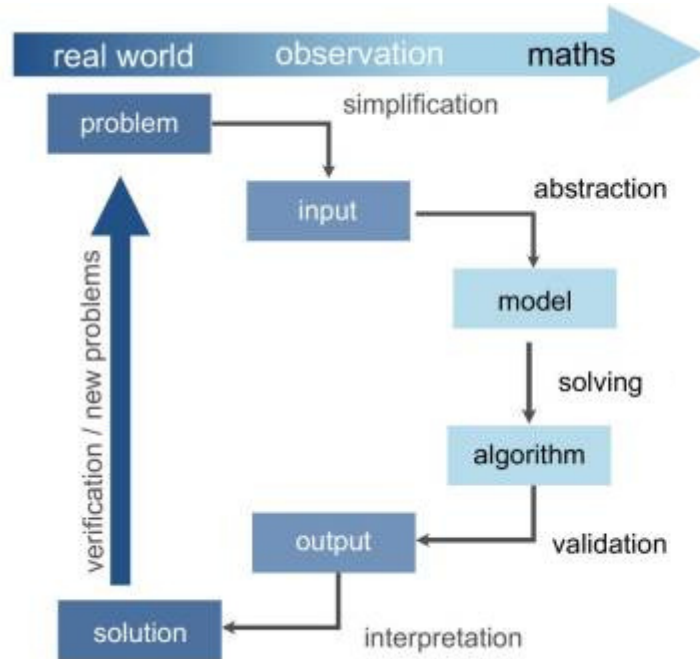
Key Aims of the NFDI:

systematically catalogize research data resulting from publicly funded research, to secure these for the future and make them findable and accessible

to develop and establish standards in research-data management and guarantee the (inter-)national linkage of research data



Model-Simulation-Optimization Workflows



Which research question was addressed?

Which mathematical model was used?

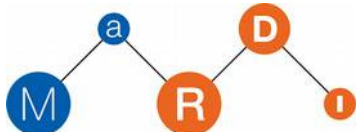
What algorithms were applied? Parameters?

Implemented in which software?

Input/Output Data Properties? Location?

Workflow reproducible? Under which conditions?

→ Read publication, supplementary,
previous publication, ...



Workflow as Wiki Page on MaRDI Portal

Non-negative Matrix Factorization for Time-Resolved Raman Spectroscopy Data

PID (if applicable): doi:10.1007/s10910-020-01201-7

Contents [\[hide\]](#)

- 1 Problem Statement
 - 1.1 Object of Research and Objective
 - 1.2 Procedure
 - 1.3 Involved Disciplines
 - 1.4 Data Streams
- 2 Model
 - 2.1 Discretization
 - 2.2 Variables
 - 2.3 Parameter
- 3 Process Information
 - 3.1 Process Steps
 - 3.2 Applied Methods
 - 3.3 Software used
 - 3.4 Experimental Devices/Instruments and Computer-Hardware
 - 3.5 Input Data
 - 3.6 Output Data
- 4 Reproducibility
 - 4.1 Reproducibility of the Experiments on the original Devices/Instruments/Hardware
 - 4.2 Reproducibility of the Experiments on other Devices/Instruments/Hardware
 - 4.3 Transferability of the Experiments to
- 5 Legend

Problem Statement [\[edit\]](#) [\[edit source\]](#)

Crystallization of Paracetamol in Ethanol

Object of Research and Objective [\[edit\]](#) [\[edit source\]](#)

Determination of Intermediate States and their Kinetics along the Crystallization of Paracetamol in Ethanol using Raman Spectroscopy.

Procedure [\[edit\]](#) [\[edit source\]](#)

Data Acquisition:
Time-resolved Raman Spectroscopy to follow the crystallization of Paracetamol in acoustically-levitated Ethanol droplets. Surface, temperature (22.0 ± 1.0 °C), and relative humidity (17.5 ± 2.5 %) of the environment are controlled by Nitrogen stream.

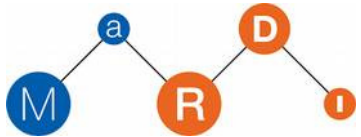
Data Extraction:
Extract the spectroscopic measurement matrix $\mathbf{M} \in \mathbb{R}_+^{n \times m}$ (n measurements of n non-negative intensities) from measurement file containing spectroscopic data and metadata.

Data Analysis:
Factorize \mathbf{M} such that $\mathbf{M} = \mathbf{W} \cdot \mathbf{H}$ ($\mathbf{W} \in \mathbb{R}_+^{n \times r}$, $\mathbf{H} \in \mathbb{R}_+^{r \times m}$ with r the rank of factorization or expected number of components) using a novel non-negative matrix factorization (NMF) approach.

Data Interpretation:
 \mathbf{W}_{rec} contains the spectra of the substances involved in the crystallization process, while \mathbf{H}_{rec} allows an inference on the kinetics. Interpretation of both matrices leads to the identification of intermediate states and the underlying kinetics.

Applied Methods [\[edit\]](#) [\[edit source\]](#)

ID	Name	Process Step	Parameter	realised / implemented by
wikidata:Q26	Singular Value Decomposition	Data Analysis - Pre-Processing	k	Matlab R2019a
wikidata:Q27	Moore-Penrose Inverse	Data Analysis - Initializing	tol	Matlab R2019a
wikidata:Q28	Perron Cluster Cluster Analysis	Data Analysis - Initializing	r	Matlab R2019a
wikidata:Q29	Nelder-Mead Algorithm	Data Analysis - Minimizing	$\alpha, \beta, \gamma, \delta, \mu$	Matlab

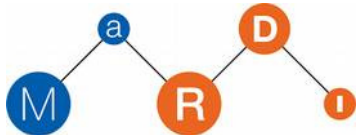


Workflow as Entity in MaRDI Knowledge Graph

Search workflows...

- ... answering specific research question ...
- ... using specific mathematical model ...
- ... using specific algorithm ...
- ... using specific input data ...

via SPARQL queries.

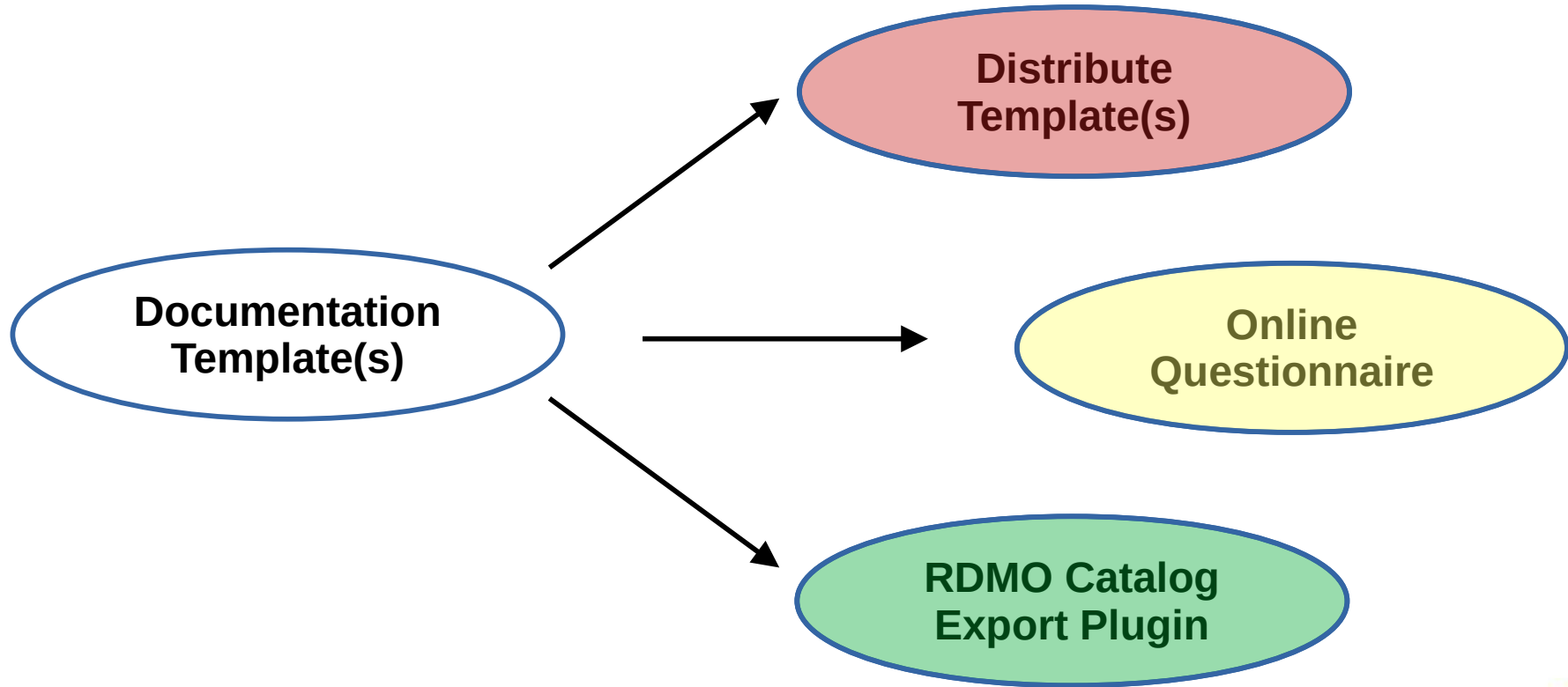


© Randy Glasbergen / glasbergen.com

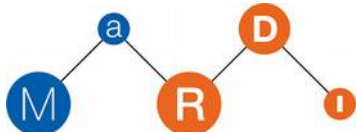
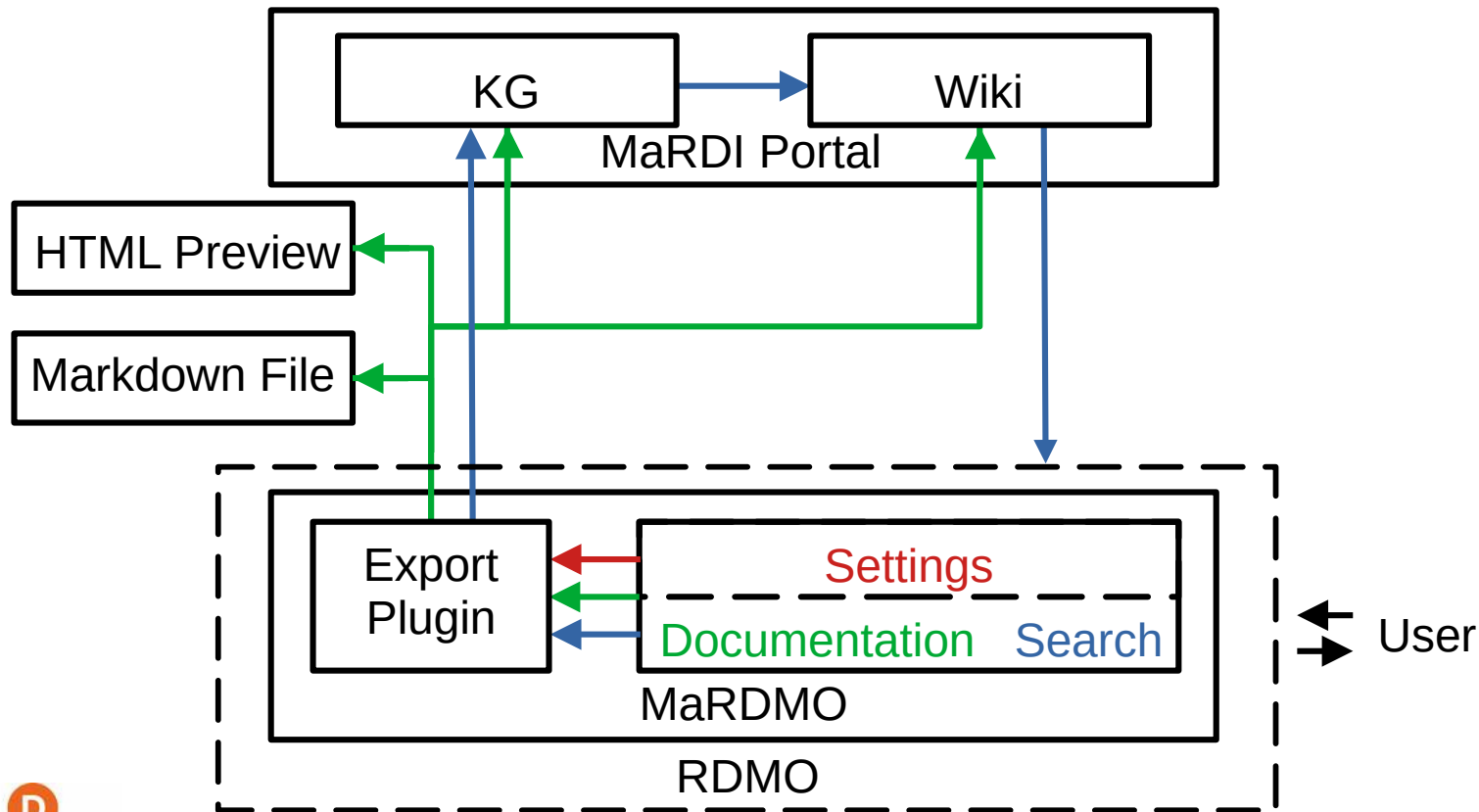


“Which is better for soul searching, Google, Yahoo or Bing?”

How to document?



MaRDMO – An Overview



MaRDMO – The Question Catalog

Sections:

- 1) MaRDI Plugin Operation Modus (1 Set, 1 Question)
- 2) Workflow Finding (1 Set, 2 Questions)
- 3) Workflow Documentation (1 Set, 2 Questions)
- 4) General Description (3 Sets, 7 Questions)
- 5) Model, Variables and Parameter (4 Sets, 14 Questions)
- 6) Process Information (7 Sets, 59 Questions)
- 7) Reproducibility (2 Sets, 9 Questions)



MaRDMO – The Export Plugin

- Simple Django App
- Workflow Documentation:
 - export:
 - download as markdown file
 - preview as html file
 - send to MaRDI Portal as mediawiki file
 - knowledge graph integration:
 - query existing databases
 - use, copy or create required entities with WikibaseIntegrator
- Workflow Finding:
 - SPARQL queries to MaRDI knowledge graph

Export

RDMO XML

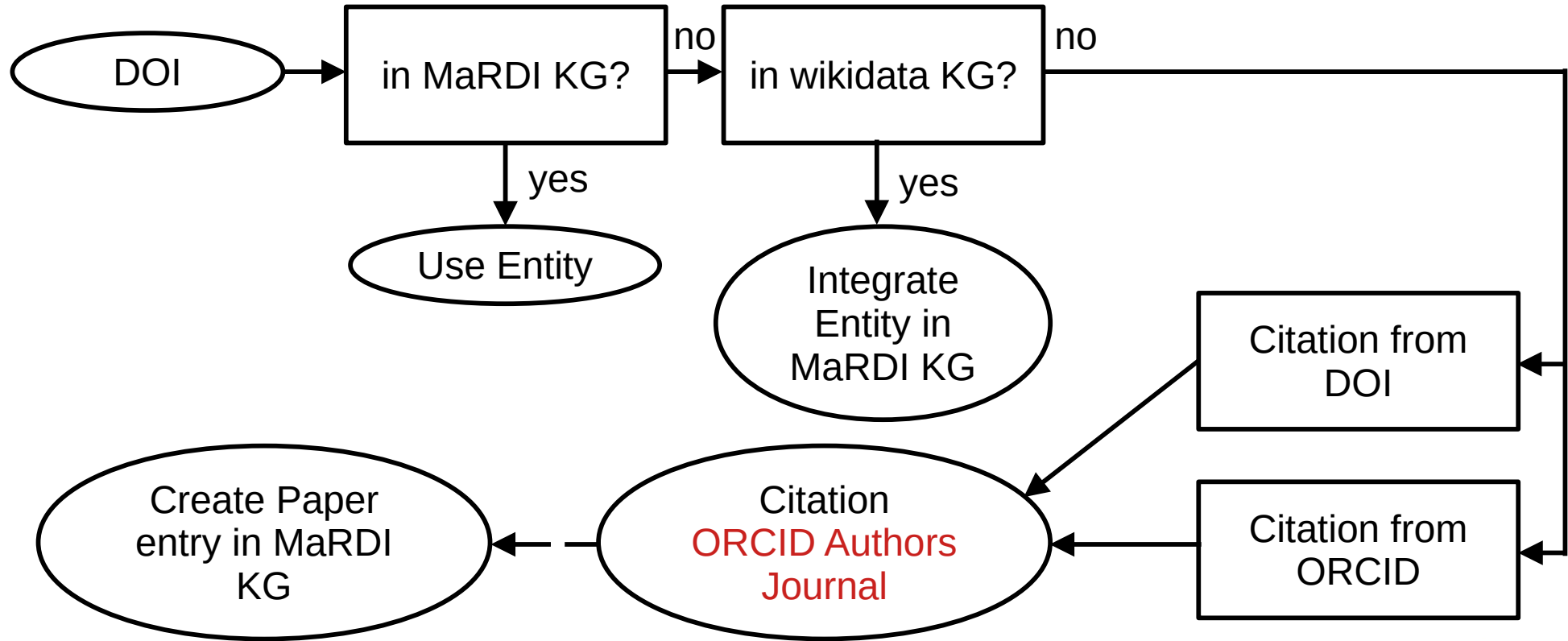
CSV comma separated

CSV semicolon separated

MaRDI Export/Query



MaRDMO – Knowledge Graph Integration



Documentation on the MaRDI Portal...

Knowledge Graph Entry

Wikipage

Non-negative Matrix Factorization for Time-Resolved Raman Spectroscopy Data (Q217)

Determination of Intermediate States and their Kinetics along the Crystallization of Paracetamol in Ethanol using Raman Spectroscopy. [edit](#)

[in more languages](#) [Compare](#)

Language	Label	Description	Also known as
English	Non-negative Matrix Factorization for Time-Resolved Raman Spectroscopy Data	Determination of Intermediate States and their Kinetics along the Crystallization of Paracetamol in Ethanol using Raman Spectroscopy.	

Statements

cites work

- Analyzing Raman spectral data without separability assumption [edit](#)
= 0 references
[+ add reference](#)
[+ add value](#)

instance of

- research workflow [edit](#)
= 0 references
[+ add reference](#)
[+ add value](#)

field of work

- Chemistry [edit](#)
= 0 references
[+ add reference](#)

Non-negative Matrix Factorization for Time-Resolved Raman Spectroscopy Data

PID (if applicable): doi:10.1007/s10910-020-01203-7

Contents [\[hide\]](#)

- 1 Problem Statement
 - 1.1 Object of Research and Objective
 - 1.2 Procedure
 - 1.3 Involved Disciplines
 - 1.4 Data Streams
- 2 Model
 - 2.1 Discretization
 - 2.2 Variables
 - 2.3 Parameter
- 3 Process Information
 - 3.1 Process Steps
 - 3.2 Applied Methods
 - 3.3 Software used
 - 3.4 Experimental Devices/Instruments and Computer Hardware
 - 3.5 Input Data
 - 3.6 Output Data
- 4 Reproducibility
 - 4.1 Reproducibility of the Experiments on the original Devices/Instruments/Hardware
 - 4.2 Reproducibility of the Experiments on other Devices/Instruments/Hardware
 - 4.3 Transferability of the Experiments to
- 5 Legend

Problem Statement [\[edit\]](#) [\[edit source\]](#)
Crystallization of Paracetamol in Ethanol

Object of Research and Objective [\[edit\]](#) [\[edit source\]](#)
Determination of Intermediate States and their Kinetics along the Crystallization of Paracetamol in Ethanol using Raman Spectroscopy.

Procedure [\[edit\]](#) [\[edit source\]](#)



Search a Workflow on the MaRDI Portal...

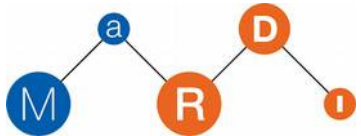
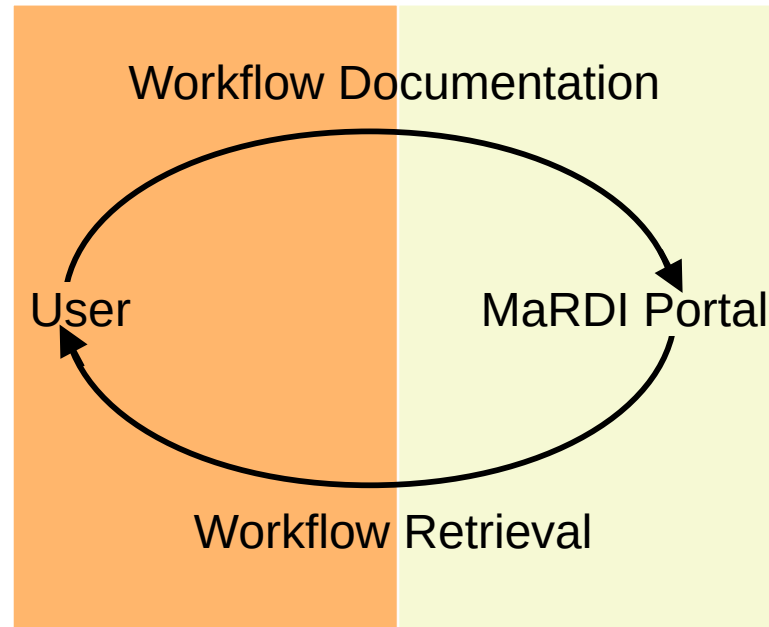
The screenshot displays the MaRDI Portal interface. At the top, a dark blue header contains the text 'example.com', 'Management - Admin', and 'Language - marco -'. The main content area is divided into several sections:

- Non-negative Matrix Factorization for Time-Resolved Raman Spectroscopy Data**: The title of the project.
- Description**: A table with one row: 'No description available.'.
- Catalog**: A table with one row: 'MaRDI Workflow Documentation'.
- Members**: A section with a description: 'Here you can see who can access the project and invite additional members. You can use the user roles to manage which rights the benefits have. Unless you are the last owner, you can leave the project with the button next to your name.' Below this is a table with columns 'User', 'E-Mail', and 'Role'. It contains one row: 'marco', 'reidelbach@zib.de', 'Owner'.
- Snapshots**: A section with a description: 'Snapshots allow you to save all responses at a given point in time and preserve a certain stage of the project. Later the snapshot can be used to create views, and the project can also be reset to a previous snapshot if needed.' Below this is a table with one row: 'No snapshots found.'.
- Options**: A list of actions: 'Answer questions', 'View answers', 'Update project information', 'Update project catalog', 'Update parent project', 'Update project tasks', 'Update project views', 'Delete project', 'Add member', 'Create snapshot', 'Back to projects overview'.
- Export**: A list of export options: 'RDMO XML', 'CSV comma separated', 'CSV semicolon separated', 'MaRDI Export/Query'.
- Import values**: A section with a description: 'Import from file'. Below this is a table with one row: 'Select file'.

The footer contains links: 'About RDMO', 'Contact', and 'Imprint and data protection'. Below these links is a paragraph: 'The Research Data Management Organisation supports the systematic planning, organisation and implementation of the data management throughout the course of a'.



MaRDMO Knowledge Loop



Summary

- Workflows should be documented explicitly
- MaRDMO provides a question catalog and export plugin to...
 - ... document workflows locally.
 - ... publish workflows as wikipedia on the MaRDI portal.
 - ... integrate workflows into the MaRDI knowledge graph.
 - ... query the MaRDI knowledge graph to find specific workflows.
- Future Tasks:

Quality Control?
Authorship?



Thank you for your attention!

Check out the MaRDI Portal:

portal.mardi4nfdi.de

Check out MaRDMO:

github.com/MarcoReidelbach/MaRDMO

