



Report 3. RDMO Hackathon @

14. RDMO Community Meeting - 2025 Aug 26

Johannes Frenzel, Elena Schick, Jürgen Rohrwild



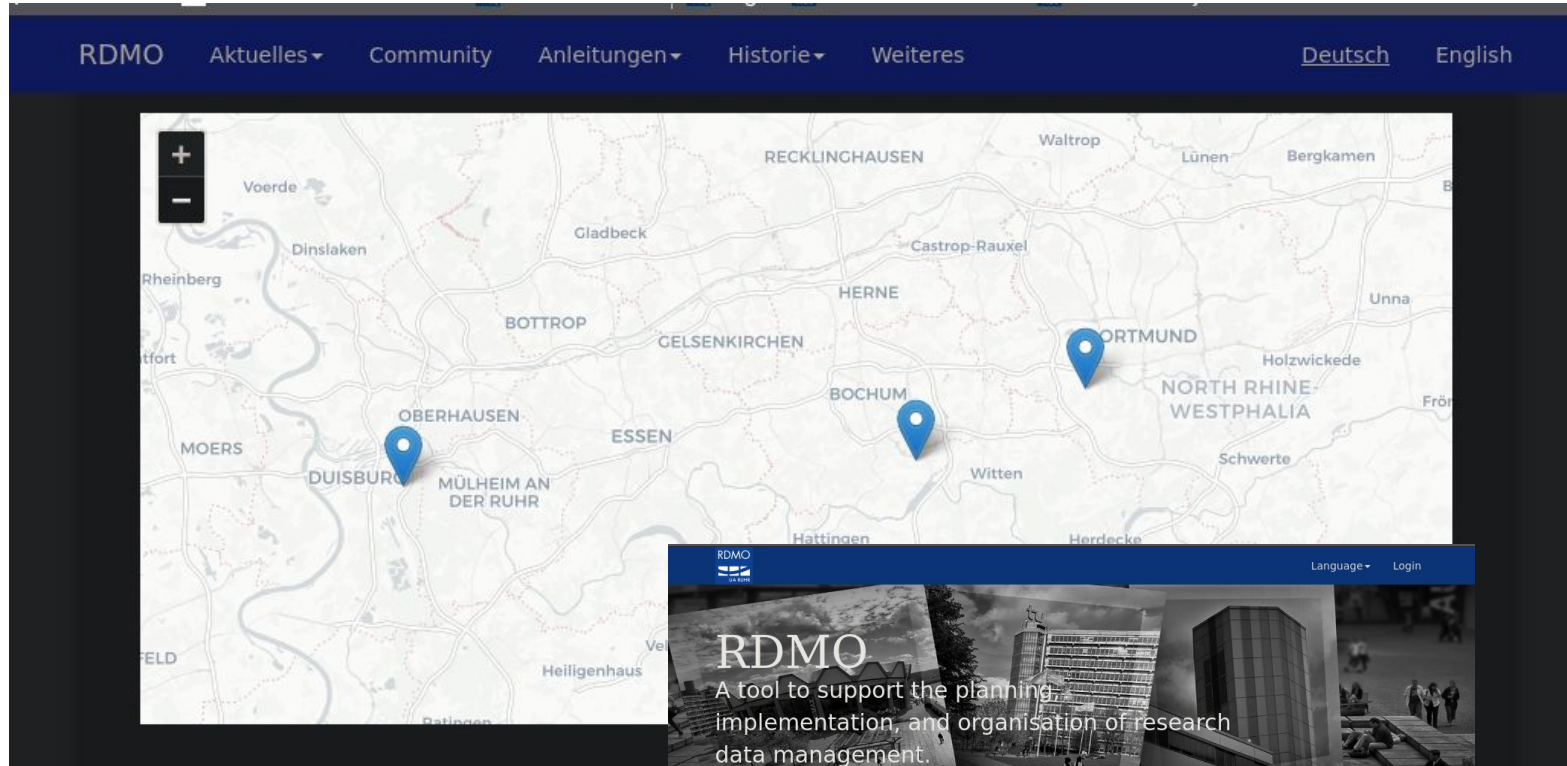
Gemeinsam besser
studieren, forschen und lehren



Planning

- **Call for Participation via Mailing-List / Slack**
- **Program / Themes to work**
 - **Via MiroBoard**
 - **Two Tracks: Software and Content**
- **Infrastructure/ Location**
- **Rooms**
- **WiFi (Eduroam / WiFi for events (without Eduroam), projectors**
- **Power sockets and extension cables are available in all rooms**

Location in UA Ruhr



Location at UA Ruhr

Full ground floor of Sebrath Library - University Library TU Dortmund

- **One large open room:**
 - **Open coffee kitchen**
 - **Presentation and**
 - **group working areas**
- **Three meeting rooms for group sessions**
- **Duration 1.5 days**
 - **Day 1: 10am - 6pm**
 - **Day 2: 8:30am - 1pm**
- **Coffee + Pizza lunch + cakes**
- **Dinner in the centre of Dortmund**

[illegible]

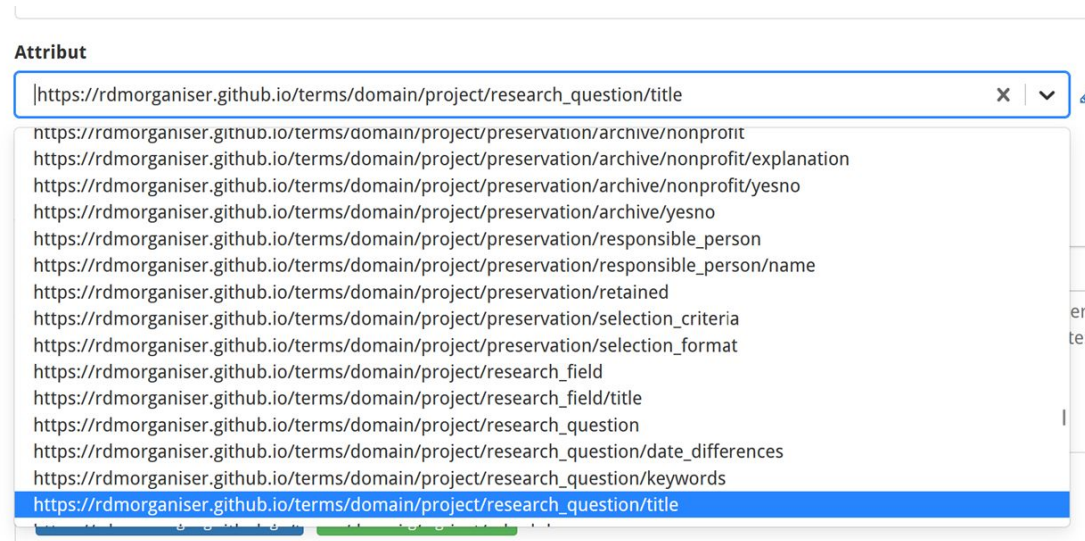
- 19 participants
- 11 Ideas for topics
- 7 Pitches
- 5 groups

[illegible]

- LLM view
- Chatbot
- LZV Katalog Design
- READ.ME view
- Domänen-Pflege
- Datenkategorien Hilfe
- ORCID- Zertifizierung

Domänen-Pflege

- Die meisten RDMO-Attribute kommen aus dem Ur-DMP im DFG-Projekt
 - keine eindeutige Struktur
 - Meist ohne Hilfe, wozu das Attribut gedacht war/ist
-
- Ca 300 Attribute
 - Für Einsteiger ist die Auswahl herausfordernd
 - Redaktionsteam braucht viel Zeit neue Attribute zu prüfen



Domänen-Pflege

Vorschlag auf Hackathon:

Neues Domänen-Schema mit neuen Attributen, die alle eine nachschlagbare „Definition“ haben.

Technisches Problem:

niemand soll gezwungen werden eigene Kataloge nachzubessern, die alte Attribute nutzen

Lösungsidee:

Neue Domäne kann importiert werden und RDMO versteht diese als „Alias“ der alten Attribute à alles geht wie bisher

Alte Attribute bekommen einen „depreciated tag“ → brauchen „tags“

Domänen-Pflege

- Umstellung = langes Projekt mit viel Arbeit
- auf Hackathon:
 - viel, viel Diskussion, was nützlich und hilfreich ist
 - Schema, dass angewand und erweitert werden kann
 - Wunsch nach „Tags“ für RDMO-Elemente:
 - Beispiel jemand sucht nach einem Attribut, um bestimmte Kosten „zu speichern“; Filter nach Tag „costs“ wäre hilfreich;
 - Auch hilfreich für Fragen, die man taggen kann (gib mir alle Fragen mit Tag „Sensitive Data“)
 - also “tags for everything”

Domänen-Pflege

	A	B	C	D	E	F	G	H	I	J
1	Attribute (old name)	1st Level	2nd Level	3rd Level	Further Levels	End	Attribute (new name)	Label	Status	Comment
2	project/acronym	project/	organi...	coordination/	acronym		project/organizational/coordination/acronym		first mapping done	
3	project/additional_rdm_policy/requirements	project/	legal...	rdm_policies/	requirements		project/legal_aspects/rdm_policies/requirements	Ethics & Legal	first mapping done	
4	project/additional_rdm_policy/yesno	project/	legal...	rdm_policies/		yesno	project/legal_aspects/rdm_policies/yesno	yes/no Ethics & Le	first mapping done	
5	project/coordination/name	project/	organi...	coordination/	name		project/organizational/coordination/name	Person	first mapping done	
6	project/costs/creation/non_personnel	project/	organi...	coordination/	creation/non_personnel/	costs	project/organizational/coordination/creation/non_personnel/costs	Infrastructure Costs	first mapping done	
7	project/costs/creation/personnel	project/	organi...	coordination/	creation/personnel/	costs	project/organizational/coordination/creation/personnel/costs	Person Costs	first mapping done	
8	project/costs/ipr/ipr_differences	project/	legal...	ipr/	differences/	costs	project/legal_aspects/ipr/differences/costs	Ethics & Legal Cos	first mapping done	
9	project/costs/ipr/non_personnel	project/	legal...	ipr/	non_personnel/	costs	project/legal_aspects/ipr/non_personnel/costs	Infrastructure Ethic	first mapping done	
10	project/costs/ipr/personnel	project/	legal...	ipr/	personnel/	costs	project/legal_aspects/ipr/personnel/costs	Ethics & Legal Per	first mapping done	
11	project/costs/metadata/differences		content/	metadata/	differences/	costs	content/metadata/differences/costs	Costs	first mapping done	project or dataset?
12	project/costs/metadata/non_personnel		content/	metadata/	non_personnel/	costs	content/metadata/non_personnel/costs	Infrastructure Costs	first mapping done	project or dataset?
13	project/costs/metadata/personnel		content/	metadata/	personnel/	costs	content/metadata/personnel/costs	Person Costs	first mapping done	project or dataset?
14	project/costs/other_research_output/non_personnel	project/	content/	output?/	other/non_personnel/	costs	project/content/output?/other/non_personnel/costs	Infrastructure Costs	first mapping done	new 3rd Level?
15	project/costs/other_research_output/personnel	project/	content/	output?/	other/personnel/	costs	project/content/output?/other/personnel/costs	Person Costs	first mapping done	new 3rd Level?
16	project/costs/pid/non_personnel	project/	techni...	pid/	non_personnel/	costs	project/technical/pid/non_personnel/costs	Infrastructure Costs	first mapping done	
17	project/costs/pid/personnel	project/	techni...	pid/	personnel/	costs	project/technical/pid/personnel/costs	Person Costs	first mapping done	
18	project/costs/pid/pid_costs_differences	project/	techni...	pid/	differences/	costs	project/technical/pid/differences/costs	Costs	first mapping done	
19	project/costs/preservation/costs_differences		preser...	storage/	differences/	costs	preservation_sharing/storage/differences/costs	Costs	first mapping done	project or dataset?
20	project/costs/preservation/cover_how	project/	preser...	storage/	coverage/	costs	project/preservation_sharing/coverage/costs	Costs	first mapping done	Path indicates a connection to
21	project/costs/preservation/non_personnel		preser...	storage/	non_personnel/	costs	preservation_sharing/storage/non_personnel/costs	Infrastructure Costs	first mapping done	project or dataset?
22	project/costs/preservation/personnel		preser...	storage/	personnel/	costs	preservation_sharing/storage/personnel/costs	Person Costs	first mapping done	project or dataset?
23	project/costs/sensitive_data/anonymization/non_personnel	dataset/	legal...	sensitive_data/	anonymization/non_personnel/	costs	dataset/legal_aspects/sensitive_data/anonymization/non_personnel/costs	Infrastructure Costs	first mapping done	
24	project/costs/sensitive_data/anonymization/personnel	dataset/	legal...	sensitive_data/	anonymization/personnel/	costs	dataset/legal_aspects/sensitive_data/anonymization/personnel/costs	Person Costs	first mapping done	
25	project/costs/sensitive_data/costs_differences	dataset/	legal...	sensitive_data/	anonymization/differences/	costs	dataset/legal_aspects/sensitive_data/anonymization/differences/costs	Costs	first mapping done	

Arbeit wird in der Content-Gruppe (AG Attribute) nach Hackathon-Schema fortgeführt

Abstimmung-Sprint – „Data Sets“

- Problem: RDMO-Kataloge haben oft die Option, Fragen für Datensets wiederholt zu beantworten
- In der Praxis:
 - Nutzende legen 10+ Datensets, an da diese aus ihrer Fachexperten-Sicht sehr unterschiedlich sind
 - Es gibt viele Rückfragen

AG Redaktionsprozesse hat einen Vorschlag für eine Verbesserung erarbeitet

Hackathon soll Feedback von viel RDMO-Managern einholen

Abstimmung-Sprint – „Data Sets“

Ergebnisse:

- keinesfalls den Namen „Datensatz“ anpassen (im Englischen → „Data Set“)
- Texte nochmal angepasst

The information about the data generated or used in the project is structured according to 'data sets'. The definition of what constitutes a data set is an important conceptual decision that needs to be made and carefully considered for each project.

For example, a data set may consist of a number of files of different types (numerical, graphical, textual...) grouped together, all containing the results of a series of tests on a defined test object.

Please fill in the form for each tab. The same tabs may be used later on other pages. You can add a new tab using the green button. Once created, you can edit or delete tabs using the buttons in the top right corner.

Data Set I

+ Data set



Abstimmung-Sprint – „Data Sets“

Ergebnisse:

- keinesfalls den Namen „Datensatz“ anpassen (im Englischen → „Data Set“)
- Texte nochmal angepasst

Data Sets

The following questions deal with the content, format, and origin of the research data generated or reused in the project and describe their management. The best way to address these questions is by clustering the data in logical units. For each of those “data sets” the same questions will be asked. Data sets may represent rather broader categories with similar constraints/properties regarding data management.

Assign a meaningful but short name to each “data set” with reference to the research question, and use it to label the corresponding tab in RDMO.

► Some tips for choosing data sets

Please fill in the form for each tab. The same tabs may be used later on other pages. You can add a new tab using the green button. Once created, you can edit or delete tabs using the buttons in the top right corner.

Data Set I

+ Data set



Abstimmung-Sprint – „Data Sets“

Ergebnisse:

- keinesfalls den Namen „Datensatz“ anpassen (im Englischen → „Data Set“)
- Texte nochmal angepasst

The following questions deal with the content, format, and origin of the research data generated or reused in the project and describe their management. The best way to address these questions is by clustering the data in logical units. For each of those “data sets” the same questions will be asked. Data sets may represent rather broader categories with similar constraints/properties regarding data management.

Assign a meaningful but short name to each “data set” with reference to the research question, and use it to label the corresponding tab in RDMO.

▼ Some tips for choosing data sets

You can proceed like that: Gain an overview of the data being generated or used within your research activity. Evaluate each data category with respect to the following criteria:

- origin/generation: generated (observation/measurement, experiment, survey, simulation, web-crawl,...) or reused;
- storage (space) requirements: e.g.: <1GB, ~10TB
- documentation requirements: e.g. extensive description of survey sampling and survey implementation or detailed references to meta-data and re-used data retrieved from a research data repository
- data privacy aspects: personal data, sensitive non-personal, public data
- relevance / reuse potential: high (reference data), middle (scientific main study), low (preliminary study)
- data publication
- long term preservation

Examples of two project with data organised in categories is to be found [on the GitLab page](#).

RDMO2ReadMe

RDMO erfragt viele Infos über Daten

Idee:

- nutze Infos, um ersten Aufschlag für ReadMe (für Zenodo, Dryad, ...) zu erstellen
 - spart Forschenden Arbeit
 - hilfreich für Datenstewards, um für RDMO zu werben
- auf Hackathon zunächst via “View” dann um Datensätze einzeln auswählen zu können via RDMO Plugin

Thank You