

Semantische und syntaktische Interoperabilität des RDMO-Datenmodells

Dr. Ulrike Wuttke, Olaf Michaelis

*Workshop der AG-Datenzentren zu Thema “Interoperabilität”,
Frankfurt am Main, 23.-24.01.2020*

Ulrike Wuttke
Fachhochschule Potsdam, RDMO
Twitter: [@uwuttke](https://twitter.com/uwuttke)
ORCID: <https://orcid.org/0000-0002-8217-4025>

Olaf Michaelis
Leibniz-Institut für Astrophysik Potsdam (AIP),
RDMO

RDMO

Was ist RDMO?

- RDMO...
 - ist eine webbasierte Software zur Planung, Umsetzung und Verwaltung aller Aufgaben des Forschungsdatenmanagements
 - kann als Infrastrukturangebot einer Einrichtung an deren Bedarfe, disziplinären Kontext und Corporate Identity angepasst werden
 - ist einsatzbereit in kleinen und größeren Projekten
- Quellcode ist **Open Source**, gut dokumentiert, verfügbar auf [GitHub](#)
- geschrieben in Python, nutzt Django und AngularJS
- **einfache Installation und Inbetriebnahme durch Infrastrukturdienstleister**

Hintergrund des RDMO-Projekts

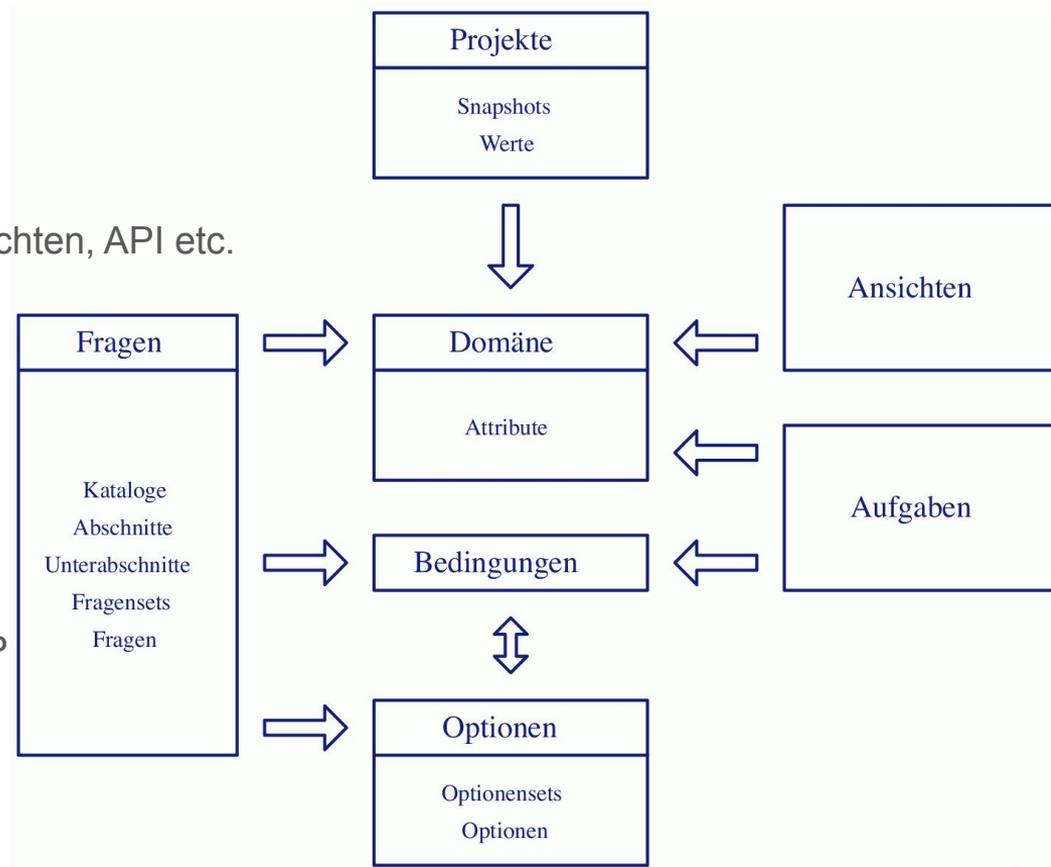
- das DFG-Projekt RDMO befindet sich in der zweiten Projektphase
- Projektpartner sind das Leibniz-Institut für Astrophysik Potsdam, die Fachhochschule Potsdam und das Karlsruher Institut für Technologie

Wichtige Ziele der zweiten Projektphase:

- Weiterentwicklung und Verbesserung der Software
- Hinzufügen neuer Funktionalitäten in Zusammenarbeit mit den Nutzenden
- Unterstützung der Community bei der Integration von RDMO in lokale Infrastrukturen
- Entwicklung eines nachhaltigen Verstetigungsmodells

RDMO-Datenmodell

- internes Vokabular von RDMO
- Verbindung von Fragen, Antworten, Ansichten, API etc.
- gewachsen mit der Erstellung von Fragenkatalogen
- derzeit in Überarbeitung
 - neue Version zu Projektende
 - Migration bestehender Instanzen
 - systematischere Struktur
 - "compliance" mit DataCite, maDMP
 - Interoperabilität



Interoperabilität

DataCite: Was haben wir vor und warum?

- **Ziel:** höhere Interoperabilität zwischen **RDMO-Domänenmodell** und **DataCite** (Metadata Schema 4.2, 2018) > Unterstützung des **Ingest** und der **Findability** anhand eines weit verbreiteten Standards (FAIR Prinzipien)
- **Use Case:** RDMO-Domänenmodell, d. h. die **RDMO-Attribute**, sollen so angepasst werden, dass ggf. bei **Datenübertragung** in ein Repository Felder **vorausgefüllt** werden > **Reichere Metadaten** und **Vermeidung Doppeleingaben** (entsprechende API muss noch gebaut werden)
- **Fokus für Machbarkeitsstudie** auf Interoperabilität zwischen den **15 DataCite Mandatory Properties und Subproperties** (+ einige ausgewählte andere Recommended und Optional Properties) und RDMO-Domänenmodell

Mapping RDMO & DataCite: In RDMO existierende Attribute

Level of Obligation	DataCite Property	RDMO Attribute	RDMO Question
M	identifierType	dataset/pids/system	Which system of persistent identifiers shall be used?
M	creatorName	project/dataset/creator_name (note: rename to "...creator/name", modify question)	Who created the (new or reused) dataset? Please provide the full name of the creator(s) of the dataset, that is the main researcher(s) involved in producing the data, in priority order (order: family name, first name) in Latin alphabet.
M	Publisher	dataset/preservation/repository	Where will the data (including metadata, documentation and, if applicable, relevant code) be stored or archived after the end of the project?
M	PublicationYear	dataset/data_publication_date	When will the data be published (if they are)?
O	Rights	dataset/sharing/conditions	Under which terms of use or license will the dataset be published or shared?
O	Format	dataset/format	Which file formats are used?

Mapping RDMO & DataCite: Neu in RDMO zu schaffende Attribute (1)

Level of Obligation	DataCite Property	RDMO Attribute	RDMO Question
M	Identifier	project/dataset/identifier	Please provide the published data's identifier (DOI, Handle etc.).
M	Creator	project/dataset/creator	Who created the (new or reused) dataset?
M	nameType	project/dataset/creator/name_type	Please choose if the data creator is a organisation or a person from the drop down menu.
M	givenName	project/dataset/creator/given_name	What is the creator's first name?
M	familyName	project/dataset/creator/family_name	What is the creator's last name?
M	nameIdentifier	project/dataset/creator/name_identifier	What is the creators unique identifier (e.g. personal ORCID)?
M	nameIdentifierScheme	project/dataset/creator/name_identifier_scheme	What is the name of the identifier scheme (e.g. ORCID)

Mapping RDMO & DataCite: Neu in RDMO zu schaffende Attribute (2)

Level of Obligation	DataCite Property	RDMO Attribute	RDMO Question
O	schemeURI	project/dataset/creator/scheme_uri	What is the URI of the name identifier schema?
M	affiliation	project/dataset/creator/affiliation	What is the organizational or institutional affiliation of the creator?
R	Contributor	project/dataset/contributor	Did others ... contribute to the dataset (e. g. collecting, managing, distributing, or otherwise contributing to the development of the resource)?
R	contributorType	project/dataset/contributor/type	Please choose the type of the contribution.
M	Title	project/dataset/title	Under which name or title will the dataset be published?
M	ResourceType	/project/dataset/resource_type	Please provide a single term of some detail that characterises the dataset (e.g. Census Data or Conference Abstract).
M	resourceTypeGeneral	/project/dataset/resource_type_general	Please provide a general publication type that characterises the dataset (e.g. text, audiovisual, software).

Semantisch und Syntaktisch

- Semantisch, der Bedeutung nach
 - Publisher -> dataset/preservation/repository
 - Unterschied zwischen “stored” oder “archived” und “to publish”
 - RDMO fragt nach ersterem “after the end of the project”
 - DataCite benutzt diese Property für die Quellenangabe “citation”
 - creatorName -> project/dataset/creator/name
 - in RDMO kann “creator_name” Person oder Institution sein, Freitext
 - DataCite verlangt Namen einer Person und gibt gleichzeitig Format vor “Familienname, Vorname”
- Syntaktisch, die Struktur betreffend
 - betrifft alle Attribute, die in RDMO “fehlen”
 - sie müssen erstellt werden, damit sie formal auf DataCite-Attribute gemappt werden können
 - die neu erdachten RDMO-Fragen sollten inhaltlich zu DataCite passen

DataCite: Was haben wir gemacht & Status Quo

- Status Quo: Von den 15 DataCite Mandatory Properties und Subproperties konnten **4 M Properties (+2 O) 1:1 auf RDMO-Attribute** gemappt werden
- Grund: RDMO konzeptuell vorrangig für die Planung des Datenmanagements gedacht und nicht explizit für die Erfassung von Metadaten für die Publikation
- Anspruch von RDMO, den **gesamten Prozess des Datenmanagements** abzubilden, spricht für **Erweiterung**
- Deshalb haben wir **Vorschläge für neue RDMO-Attribute** (und Fragen) erarbeitet, die wir momentan intern diskutieren
- **Herausforderungen:** u. a.
 - im Gegensatz zu RDMO macht DataCite teilweise recht restriktive Vorgaben bzgl. der Eingabe (Allowed Values, Occurrence), z. B. nur DOIs (aber siehe auch DataCite Best Practice)
 - eigentlich alle Eingaben auf Englisch erforderlich

Ausblick

- **Nächste Schritte:** interne Konsolidierung und Vorstellung an RDMO-Anwender-Community mit grundsätzlicher Frage, ob wir gleich weitere Properties (und welche) einbeziehen sollten, weil jede Umstellung des RDMO-Domänenmodells einen schwerwiegenden Eingriff bedeutet, der sich auf alle Instanzen auswirkt
- **Herausforderung:** Anwender*innen können RDMO-Fragenkataloge beliebig anpassen, wir könnten eigentlich letztendlich nur Empfehlungen aussprechen, dass bestimmte Attribute bzw. Fragen zum “Grundstock” gehören müssen, um Interoperabilität z. B. mit DataCite zu gewährleisten

Kontakt

Ulrike Wuttke

E-Mail: wuttke@fh-potsdam.de

ORCID: <https://orcid.org/0000-0002-8217-4025>

Olaf Michaelis

E-Mail: omichaelis@aip.de

- [RDMO Webseite](#)
- [RDMO Testinstanz](#)