



DINI/nestor-WS5: Werkzeuge Breakout-Session: Metadaten in der Praxis

Moderation: Florian Krämer & Dominik Schmitz

Breakout-Session “Metadaten in der Praxis”

Rolle von Metadaten im Forschungsdatenmanagement

- Erschließen fremder Datenbestände
- Wiederauffinden von fremden oder auch nur eigenen Daten
- Sinnvolle und korrekte Nachnutzung.

Praxis? Wenige Daten konsequent durch Metadaten beschrieben.

Gründe

- hoher Aufwand
- fehlende, in den Forschungsprozess integrierte **Unterstützung durch Werkzeuge**

Session-Inhalt

Werkzeuge für das effektive und effiziente Metadatenmanagement

Drei Beiträge

- Metadatenservice zur flexiblen Integration (F. Krämer, RWTH Aachen University)
- Docuteam Packer – Vorbereitung LZA (M. El-Saad, ETH Zürich)
- Projekt MASi Metadata Management for Applied Sciences (R. Müller-Pfefferkorn, TU Dresden)



Ein Metadatenservice zur flexiblen Integration in bestehende Forschungsumgebungen

Florian Krämer & Dominik Schmitz

Forschungsdatenmanagement an der RWTH Aachen

- Projektgruppe mit Mitgliedern aus
 - Universitätsbibliothek
 - Forschungsförderung
 - IT Center

- Ziel:
Etablierung eines strukturierten und nachhaltigen Forschungsdatenmanagements an der RWTH Aachen

- Maßnahmen:
 - Beratungsangebot für Forschende
 - Weiterbildungsangebote zum Thema FDM
 - **Verbesserung der technischen Infrastruktur**

Unser Ansatz

- Angebot eines Werkzeugs zur Erzeugung und Speicherung von Metadaten, das
 - in bestehende IT-Umfelder der Forschenden integrierbar ist;
 - einfach zu benutzen ist;
 - in allen Phasen des Forschungsdaten-lifecycles genutzt werden kann;
 - interoperable mit anderen Werkzeugen ist;

Vorstellung des Werkzeugs

The image displays two screenshots of the Metadata Manager web application interface.

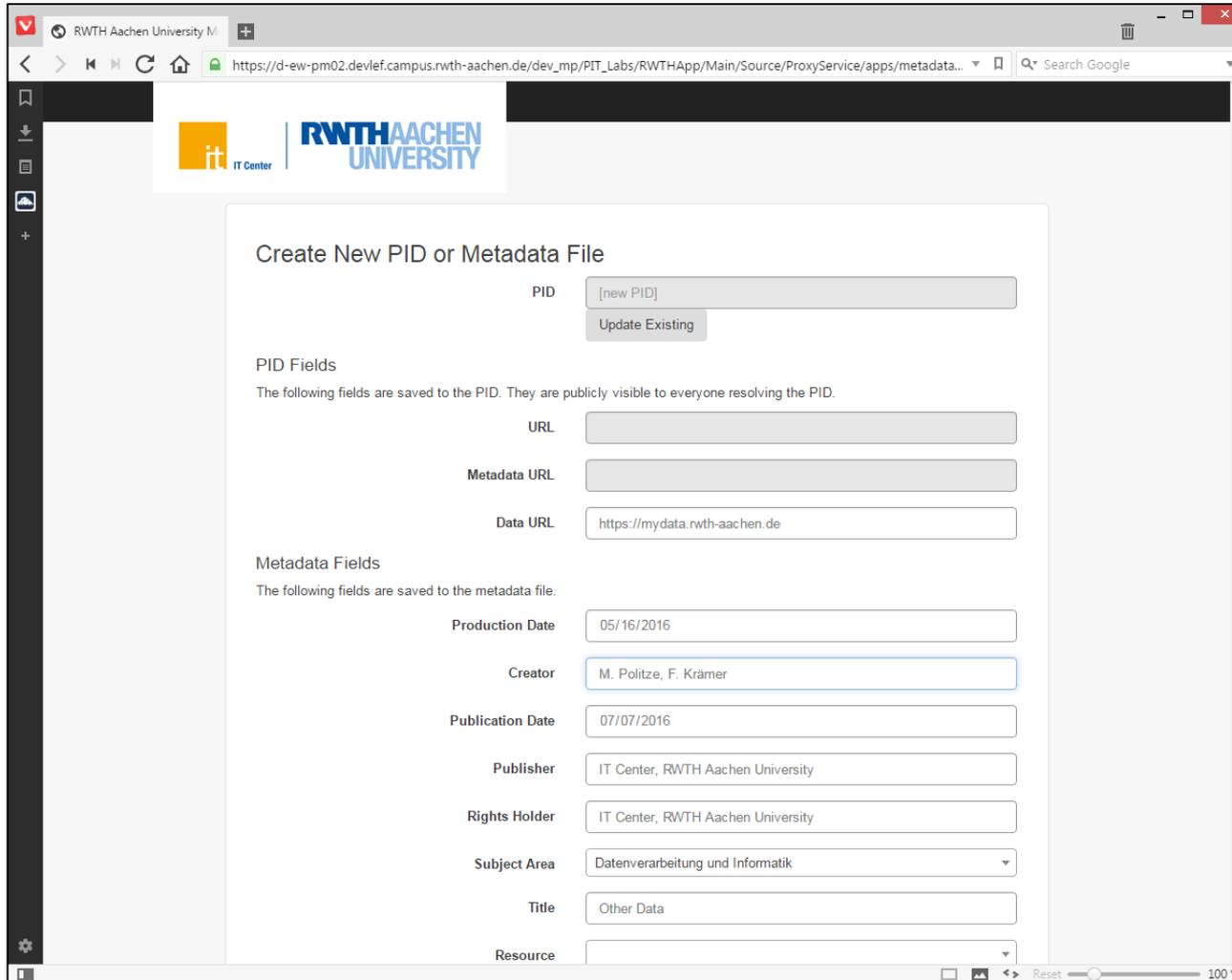
Left Screenshot: New from schema

- Header: RWTH Aachen University IT Center logo.
- Section: **New from schema**
- Input field: `radar_022000.xml`
- Section: **Created metadata**
- Output field: `http://hdl.handle.net/21.11102/dc/849726-d952-4a56-ab29-7068b24ee382`
- Footer: Metadata Manager / Version 0.1.1 / IT Center RWTH Aachen University / Datenschutzerklärung / Powered by Docker

Right Screenshot: Select Workflow

- Section: **Select Workflow**
- Option 1: **Save metadata file to integrated system** (cloud icon). Description: Select this option to create a validated metadata file automatically saved to the integrated system. The metadata is only visible to you. However it can be published. This option is recommended if you plan to publish your metadata together with the PID. **Select >**
- Option 2: **Create a local metadata file** (download icon). Description: Select this option to create a validated metadata file that can be archived on together with your data on a system of your choice. This option is recommended if you want to create metadata for datasets you do not plan to publish. **Select >**
- Form fields: Metadata URL, Data URL.
- Section: **Metadata Fields**
- Text: The following fields are saved to the metadata file.
- Fields: Production Date (mm/dd/yyyy), Creator, Publication Date (mm/dd/yyyy), Publisher, Rights Holder, Subject Area, Title, Type.
- Footer: Docker

Vorstellung des Werkzeugs: Integrierter Workflow I



RWTH Aachen University M

https://d-ew-pm02.devlef.campus.rwth-aachen.de/dev_mp/PIT_Labs/RWTHApp/Main/Source/ProxyService/apps/metadata... Search Google

IT Center | **RWTH AACHEN UNIVERSITY**

Create New PID or Metadata File

PID

PID Fields
The following fields are saved to the PID. They are publicly visible to everyone resolving the PID.

URL

Metadata URL

Data URL

Metadata Fields
The following fields are saved to the metadata file.

Production Date

Creator

Publication Date

Publisher

Rights Holder

Subject Area

Title

Resource

Reset 100 %

Vorstellung des Werkzeugs: Integrierter Workflow II

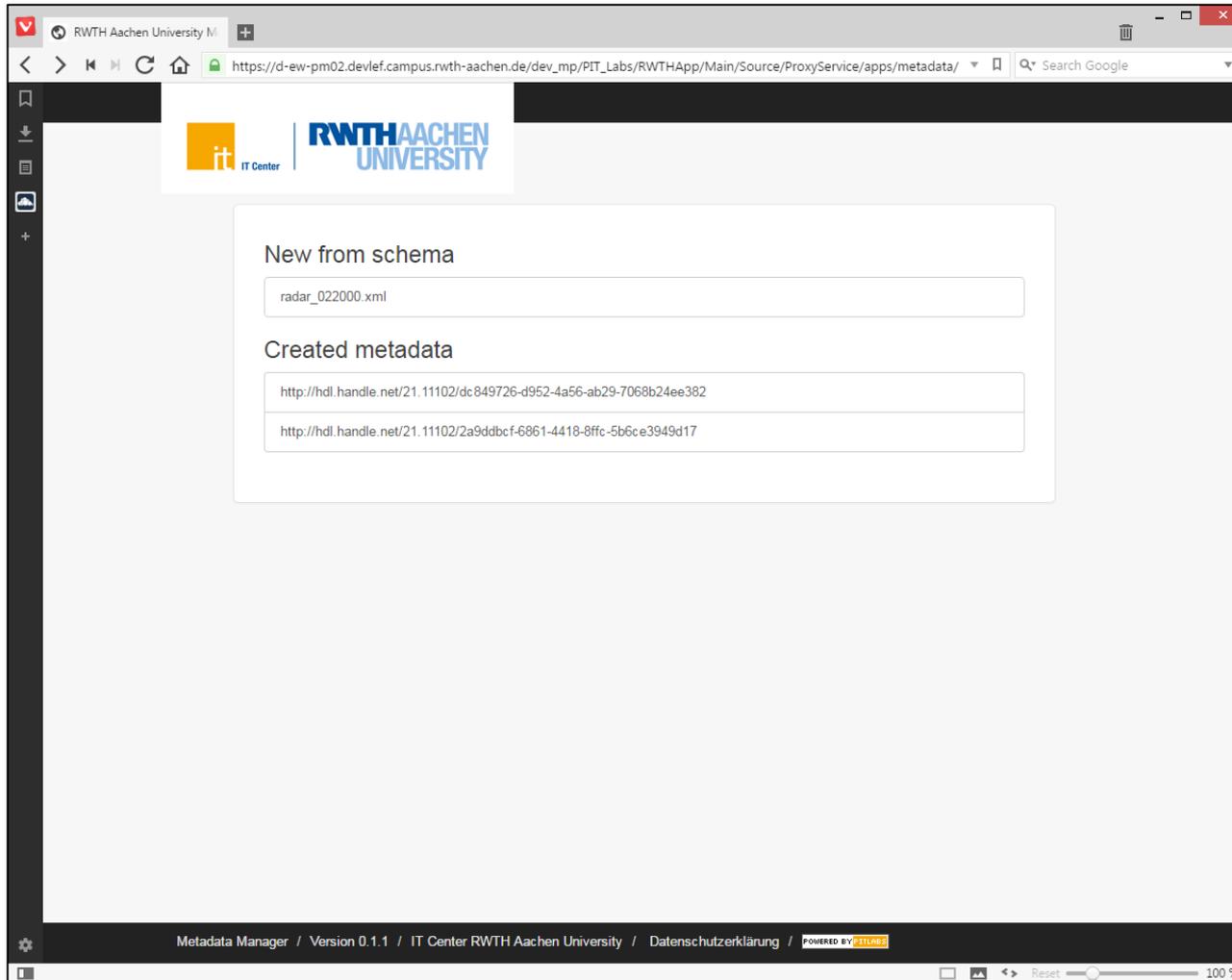
The screenshot shows a web browser window with the URL `https://d-ew-pm02.devlef.campus.rwth-aachen.de/dev_mp/PIT_Labs/RWTHApp/Main/Source/ProxyService/apps/metadata...`. The main content area is titled "Create New PID" and contains several sections:

- PID Fields:** A section with the text "The following fields are saved to the PID file." and a "Save" button.
- Metadata URL:** A text input field.
- Data URL:** A text input field containing `https://mydata.rwth-aachen.de`.
- Metadata Fields:** A section with the text "The following fields are saved to the metadata file." and several input fields:
 - Production Date:** `05/16/2016`
 - Creator:** `M. Politze, F. Krämer`
 - Publication Date:** `07/07/2016`
 - Publisher:** `IT Center, RWTH Aachen University`
 - Rights Holder:** `IT Center, RWTH Aachen University`
 - Subject Area:** `Datenverarbeitung und Informatik` (dropdown menu)
 - Title:** `Other Data`
 - Resource:** (dropdown menu)

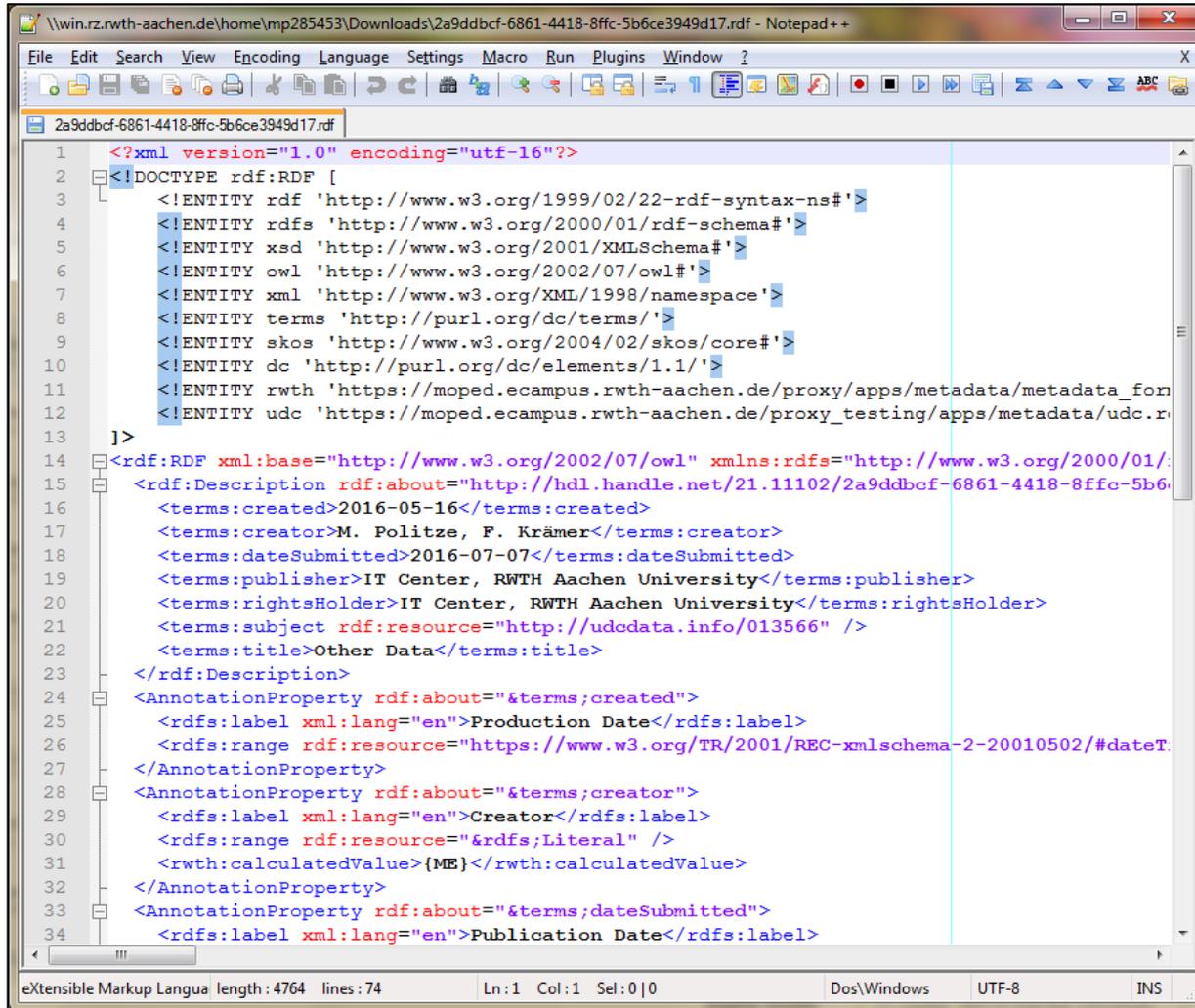
A modal dialog box is overlaid on top of the form, titled "Download '4f3f7db3-f30c-47fc-9264-1463043d8d11'". It displays file information: File Name: `4f3f7db3-f30c-47fc-9264-1463043d8d11`, File Size: Unknown, Host: `https`. There is a checkbox for "Always Save Files to Default Download Location" which is unchecked. The dialog has buttons for "Save", "Save As...", "Open", and "Cancel".

At the bottom of the browser window, there is a footer: "Metadata Manager / Version 0.1.1 / IT Center RWTH Aachen University / Datenschutzerklärung / POWERED BY PITLINE".

Vorstellung des Werkzeugs: Integrierter Workflow III



Vorstellung des Werkzeugs: Integrierter Workflow IV



```
1 <?xml version="1.0" encoding="utf-16"?>
2 <!DOCTYPE rdf:RDF [
3   <!ENTITY rdf 'http://www.w3.org/1999/02/22-rdf-syntax-ns#'>
4   <!ENTITY rdfs 'http://www.w3.org/2000/01/rdf-schema#'>
5   <!ENTITY xsd 'http://www.w3.org/2001/XMLSchema#'>
6   <!ENTITY owl 'http://www.w3.org/2002/07/owl#'>
7   <!ENTITY xml 'http://www.w3.org/XML/1998/namespace'>
8   <!ENTITY terms 'http://purl.org/dc/terms/'>
9   <!ENTITY skos 'http://www.w3.org/2004/02/skos/core#'>
10  <!ENTITY dc 'http://purl.org/dc/elements/1.1/'>
11  <!ENTITY rwth 'https://moped.ecampus.rwth-aachen.de/proxy/apps/metadata/metadata_for'
12  <!ENTITY udc 'https://moped.ecampus.rwth-aachen.de/proxy_testing/apps/metadata/udc.r
13 ]>
14 <rdf:RDF xml:base="http://www.w3.org/2002/07/owl" xmlns:rdfs="http://www.w3.org/2000/01/
15 <rdf:Description rdf:about="http://hdl.handle.net/21.11102/2a9ddbfc-6861-4418-8ffc-5b6
16   <terms:created>2016-05-16</terms:created>
17   <terms:creator>M. Politze, F. Krämer</terms:creator>
18   <terms:dateSubmitted>2016-07-07</terms:dateSubmitted>
19   <terms:publisher>IT Center, RWTH Aachen University</terms:publisher>
20   <terms:rightsHolder>IT Center, RWTH Aachen University</terms:rightsHolder>
21   <terms:subject rdf:resource="http://udcdata.info/013566" />
22   <terms:title>Other Data</terms:title>
23 </rdf:Description>
24 <AnnotationProperty rdf:about="&terms;created">
25   <rdfs:label xml:lang="en">Production Date</rdfs:label>
26   <rdfs:range rdf:resource="https://www.w3.org/TR/2001/REC-xmlschema-2-20010502/#dateT
27 </AnnotationProperty>
28 <AnnotationProperty rdf:about="&terms;creator">
29   <rdfs:label xml:lang="en">Creator</rdfs:label>
30   <rdfs:range rdf:resource="&rdfs;Literal" />
31   <rwth:calculatedValue>{ME}</rwth:calculatedValue>
32 </AnnotationProperty>
33 <AnnotationProperty rdf:about="&terms;dateSubmitted">
34   <rdfs:label xml:lang="en">Publication Date</rdfs:label>
```

eXtensible Markup Language length: 4764 lines: 74 Ln:1 Col:1 Sel:0|0 Dos\Windows UTF-8 INS

Vorstellung des Werkzeugs: Privater Workflow I

The screenshot shows a web browser window with the URL `https://d-ew-pm02.devlef.campus.rwth-aachen.de/dev_mp/PIT_Labs/RWTHApp/Main/Source/ProxyService/apps/metadata...`. The page title is "Create New PID or Metadata File".

PID

PID Fields
The following fields are saved to the PID. They are publicly visible to everyone resolving the PID.

URL

Metadata URL

Data URL

Metadata Fields
The following fields are saved to the metadata file.

Production Date

Creator

Publication Date

Publisher

Rights Holder

Subject Area

Title

Type

Vorstellung des Werkzeugs: Privater Workflow II

RWTH Aachen University M

https://d-ew-pm02.devlef.campus.rwth-aachen.de/dev_mp/PIT_Labs/RWTHApp/Main/Source/ProxyService/apps/metadata... Search Google

IT Center | **RWTHAACHEN UNIVERSITY**

Create New PID or Metadata File

PID

PID Fields
The following fields are saved to the PID. They are publicly visible to everyone resolving the PID.

URL

Metadata URL

Data URL

Metadata Fields
The following fields are saved to the metadata file.

Production Date

Creator

Publication Date

Publisher

Rights Holder

Subject Area

Title

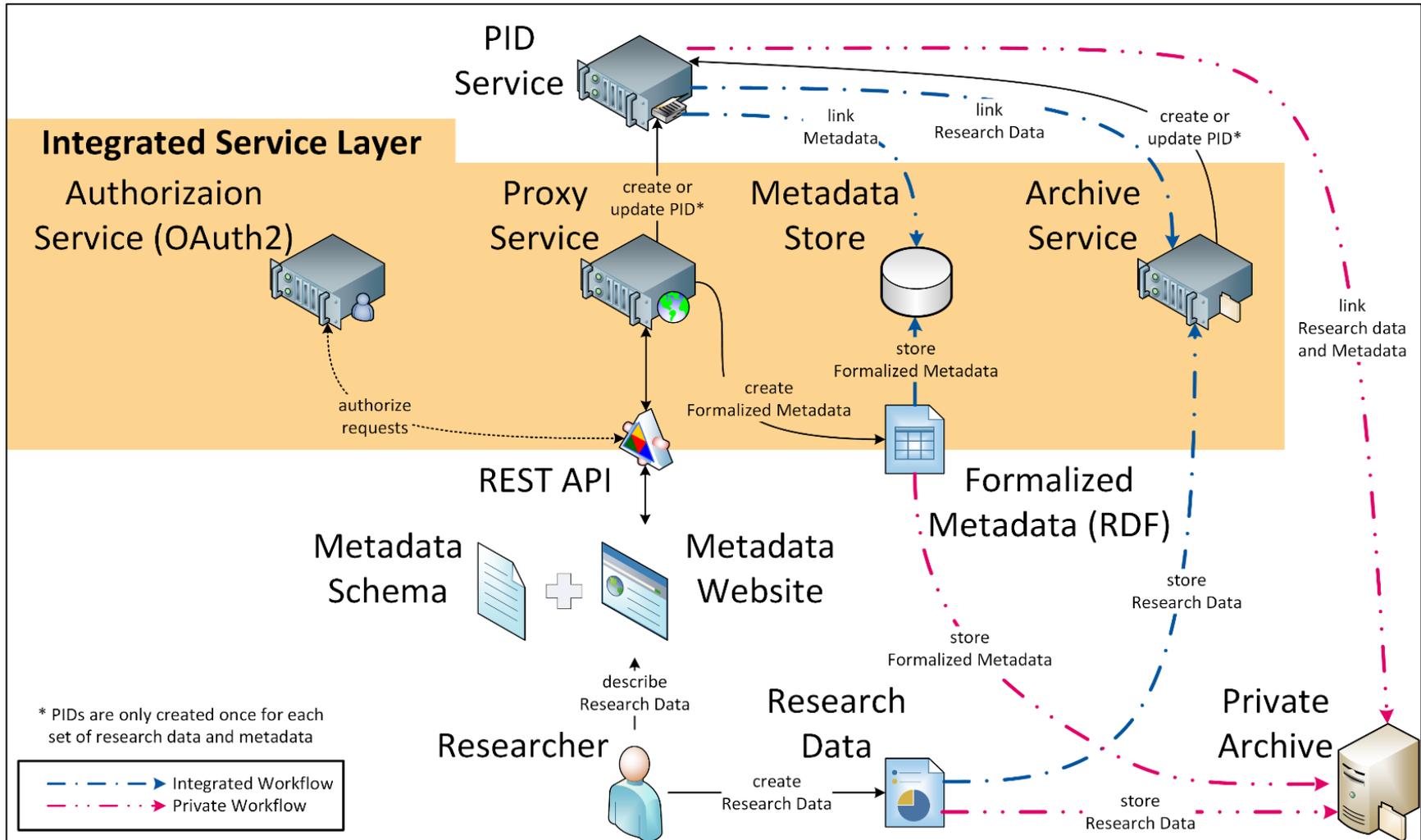
Type

Reset 100 %

Vorstellung des Werkzeugs: PID-Integration

The image displays two screenshots of a web application interface for PID integration. The top screenshot shows a modal dialog titled "Enter PID OTA" with the instruction "Please enter PID OTA to edit an existing PID." and a text input field containing a long alphanumeric string: "RrhNigtODUzOS1MDAxNjlmZTAwMmYlLCJleHhAiom51bGwslmlhdCl6MTQ2NTM5NDcyNH0.wRH90XS1_Q7aADrdkij_TN8YamBgnZ66tvlT2fpw". An "OK" button is visible at the bottom right of the dialog. The bottom screenshot shows the main "Create New PID or Metadata File" form. It includes fields for "PID" (containing "21.11102/8e5e0e05-5425-4a68-8539-b00169fe002f"), "OTA" (containing "eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJpc3MiOiJodHRwczovL3NoeXJlcG9pbmQ"), "URL", "Metadata URL", "Data URL" (containing "http://mydata.rwth-aachen.de"), "Production Date" (containing "05/24/2016"), "Creator" (containing "M. Poltze, F. Krämer"), "Publication Date" (containing "06/09/2016"), "Publisher" (containing "IT Center, RWTH Aachen University"), and "Rights Holder" (containing "IT Center RWTH Aachen University"). The form also includes an "Update Existing" button and descriptive text for "PID Fields" and "Metadata Fields".

Private und Integrierter Workflow



Metadata Schema in RDF, OWL und XML

```
...
<!ENTITY rdf 'http://www.w3.org/1999/02/22-rdf-syntax-ns#'>
<!ENTITY rdfs 'http://www.w3.org/2000/01/rdf-schema#'>
<!ENTITY terms 'http://purl.org/dc/terms/'>
<rdf:RDF>
...
  <AnnotationProperty rdf:about="&terms;creator">
    <rdfs:label xml:lang="en">Creator</rdfs:label>
    <rdfs:range rdf:resource="&rdfs;Literal" />
  </AnnotationProperty>
  <AnnotationProperty rdf:about="&terms;dateSubmitted">
    <rdfs:label xml:lang="en">Publication Date</rdfs:label>
    <rdfs:range
      rdf:resource="https://www.w3.org/TR/2001/REC-xmlschema-2-20010502/#dateTime" />
  </AnnotationProperty>
  <ObjectProperty rdf:about="&terms;subject">
    <rdfs:label xml:lang="en">Subject Area</rdfs:label>
    <rdfs:range rdf:resource="http://udcdata.info/078887" />
  </ObjectProperty>
  <AnnotationProperty rdf:about="&terms;title">
    <rdfs:label xml:lang="en">Title</rdfs:label>
  </AnnotationProperty>
...
</rdf:RDF>
```

Beschreibung eines Datensatzes in RDF, OWL und XML

```
...
<rdf:RDF>
  <rdf:Description
    rdf:about="http://hdl.handle.net/21.11102/df8f04ac-d698-483e-bb24-cb135112737b">
    <terms:created>2016-05-24</terms:created>
    <terms:creator>M. Politze, F. Krämer</terms:creator>
    <terms:dateSubmitted>2016-06-09</terms:dateSubmitted>
    <terms:publisher>IT Center, RWTH Aachen University</terms:publisher>
    <terms:rightsHolder>IT Center, RWTH Aachen University</terms:rightsHolder>
    <terms:subject rdf:resource="http://udcdata.info/013566" />
    <terms:title>Some Data</terms:title>
  </rdf:Description>
  ...
</rdf:RDF>
```

Geplante Erweiterungen

- Nutzung des Systems als Interface für die PID-Registrierung
- Bereitstellen von Metadaten für Archiv- und Publikationsdomäne
- Durchsuchbarkeit der gespeicherten Metadaten
- Beispielskripte zur automatisierten Übertragung existierender Metdatensätze
- **Akzeptanz schaffen!**

**Vielen Dank
für Ihre Aufmerksamkeit**