



Towards a distributed research data management system

Marius Politze & Florian Krämer

Contents

- Introduction
 - Research Data Management at RWTH Aachen University
 - What are Metadata and why do I need them?
 - Basic idea of our approach
- Walkthrough Metadata Tool
 - Metadataschemas
 - Storage location
 - Private Workflow
 - Integrated Workflow
 - PID handling
- Technical implementation
 - Workflow design
 - Architecture
 - Extensibility – towards a distributed system
 - Metadata and Metadataschema requirements
 - RDF, OWL, and XML
 - Future Work

Research Data Management at RWTH Aachen University

- Project group with members from the
 - University Library
 - Department Research and Career
 - IT Center
- Goal:
Establishing a structured and sustainable Research Data Management at RWTH Aachen University
- Measures:
 - support structures for researchers
 - training in RDM topics
 - improving the technical infrastructure

What are Metadata and why do I need them?

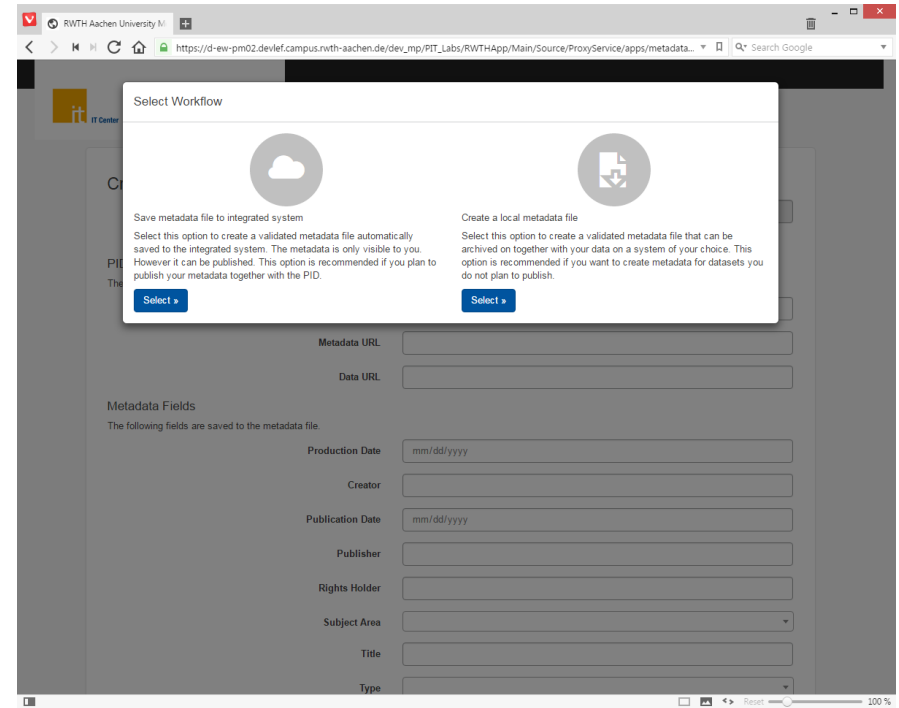
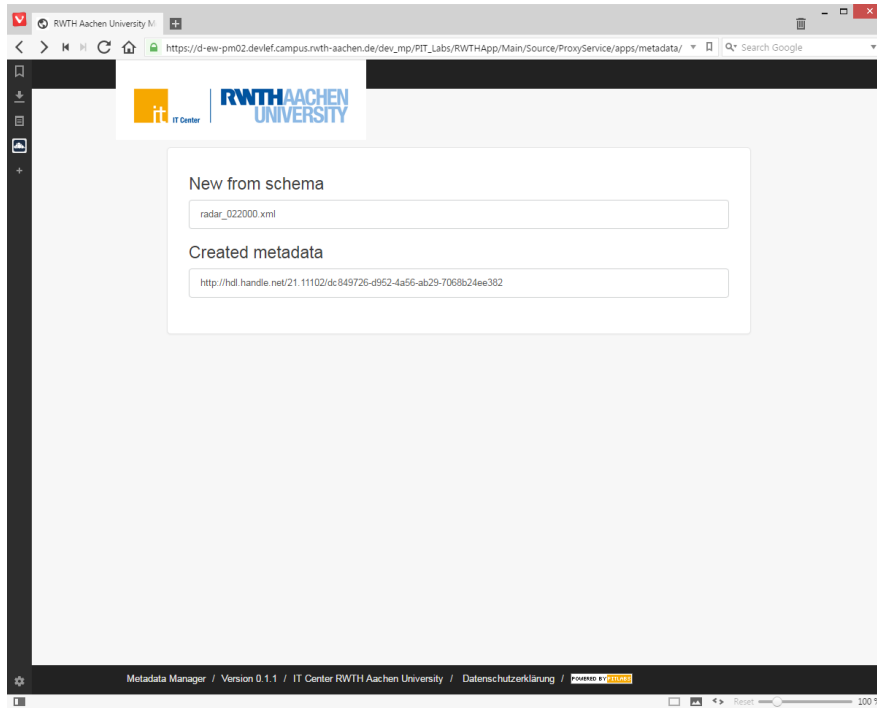
- Metadata are data describing data
- Metadata helps me to find an re-use data
- Metadata needs to be created in a systematic and structured way

Basic idea of our approach

- Providing a tool to create and store metadata that
 - integrates into existing environments;
 - is easy to use;
 - can be used in all phases of the research process;
 - inter-operates with other tools;

Walkthrough Metadata Tool (I)

Metadataschemas / Storage location



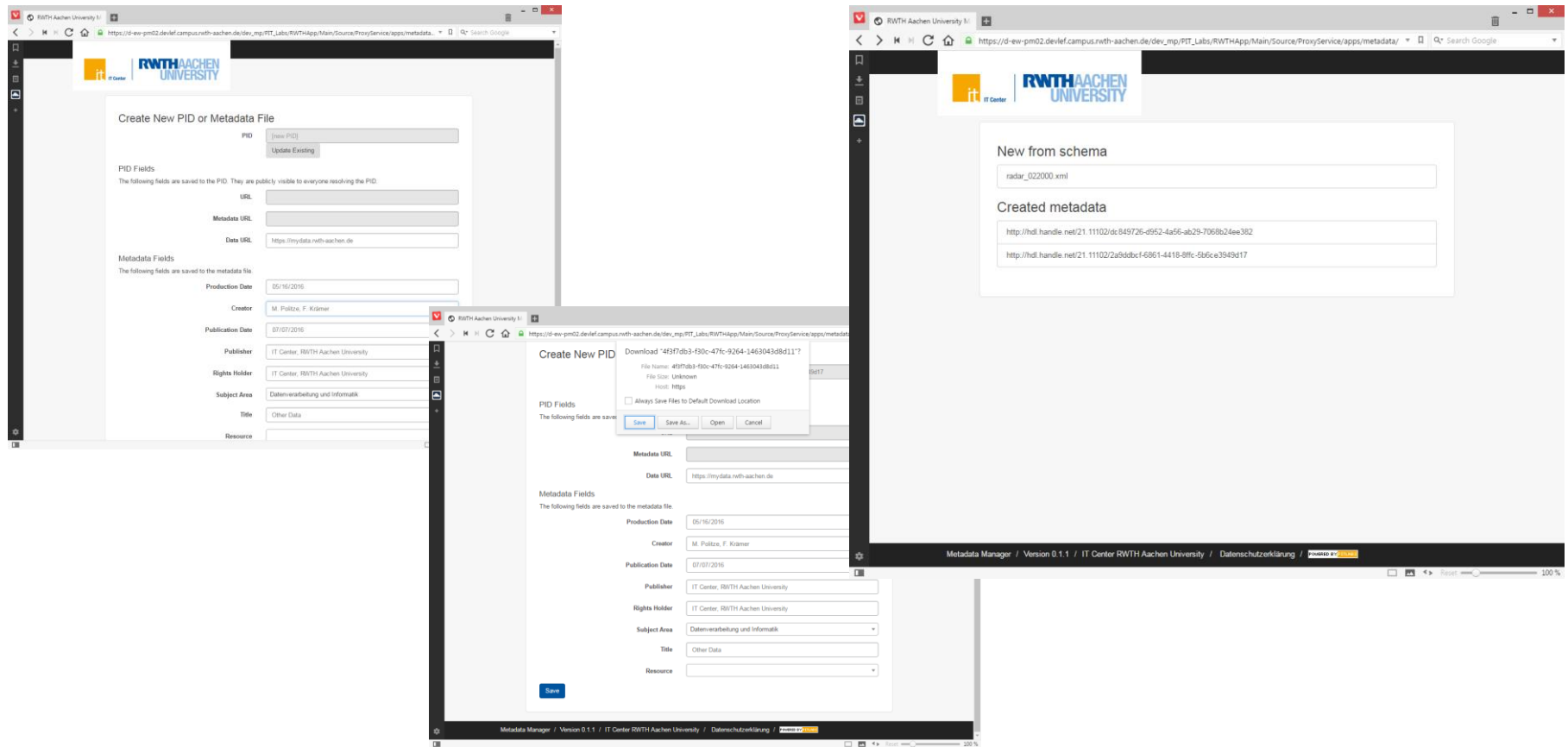
Walkthrough Metadata Tool (III)

Private Workflow

The screenshots illustrate the workflow for creating a new PID or metadata file. The interface includes sections for PID Fields (URL, Metadata URL, Data URL) and Metadata Fields (Production Date, Creator, Publication Date, Publisher, Rights Holder, Subject Area, Title, Type). A 'Download' dialog box is shown, indicating the file name and size. The 'Save' button is visible at the bottom of the form.

Walkthrough Metadata Tool (II)

Integrated Workflow



Walkthrough Metadata Tool (IV)

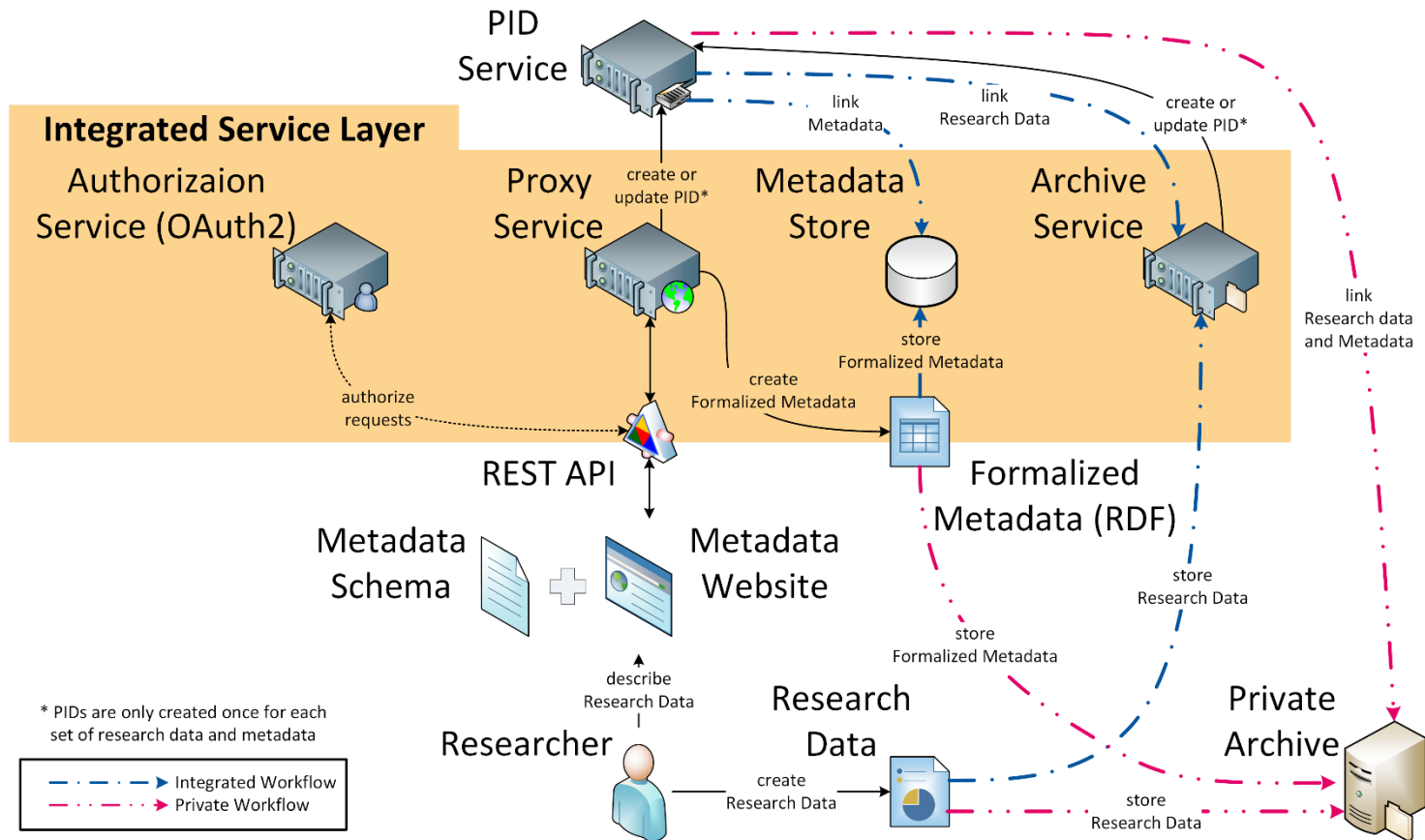
PID handling

The screenshot displays the RWTH Aachen University Metadata Tool interface. The main form is titled "Create New PID or Metadata File" and contains the following fields:

- PID:** 21.11102/8e5e0e05-5425-4a68-8539-b001696e002f
- OTA:** eyJ0eXAoARJYUQlCjhsCmCjOUHlZj11NjU5 eyJpc3MjOLtoeFRwczovL3NoeXJic09pbnQ
- PID Fields:** URL, Metadata URL, Data URL (http://mydata.rwth-aachen.de)
- Metadata Fields:** Production Date (05/24/2016), Creator (M. Politze, F. Krämer), Publication Date (06/09/2016), Publisher (IT Center, RWTH Aachen University), Rights Holder (IT Center RWTH Aachen University)

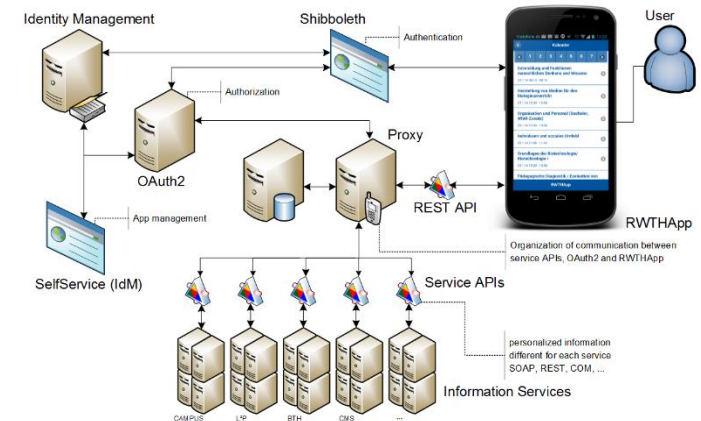
A modal dialog titled "Enter PID OTA" is open, prompting the user to "Please enter PID OTA to edit an existing PID." The input field contains the OTA value: RjNjgtODUzOS1MDAANjZTAWMmYlCjhsHAOM5tsGwslmhD06MTQ2NTM5NDcyNH0_wRH50XS_1_Q7aDdckll_TNYamBgnZ66-vIT2pnl. An "OK" button is visible at the bottom right of the dialog.

Private and Integrated Workflow



Architecture

- REST Webservices
 - Automation of metadata creation early in the research process
 - Use (part of) the workflows to support individual processes at the institutes
- User Interface
 - Easy to use with basic functionality
 - To get started without programming knowledge
- Integrated into Infrastructure at RWTH Aachen
 - OAuth2 subsystem for authorization
 - Caching for faster response times
 - Redundancy to maximize availability



Extensibility I

- PID One Time Access Tokens (OTA)
 - Used to hand over control of PID between systems
 - Based on JSON Web Token
- Web Services using OAuth
 - Each operation can be called by external applications
 - Authorizations can be passed and revoked at any time
- Workflows can be combined
 - Private and integrated workflow can be combined
 - Allows maximum flexibility to fit existing research processes
- Data can be moved from private to integrated
 - for collaboration private
 - for integrated for long term storage / archive

Extensibility II

- Many metadata schemas are available as RDF+OWL
 - Domain specific as well as independent
 - Can be combined with other dialects such as RDF+SKOS can be
- However they have to be adopted or extended
 - Extensions are easy as multiple ontologies can be linked
 - Ontologies can be reduced
- Ontologies can describe properties of the metadata schema itself
 - Default and calculated values
 - Localized Descriptions and Labels
 - Domain and Ranges

Metadata and Metadata Schema Requirements

- Metadata and metadata schemas in machine readable format
 - Descriptions of metadata fields
 - Multi Language (German, English)
- Format should be consistent, flexible and self explanatory
 - For domain specific and domain independent metadata schemas
 - Readable in 10-15 Years from now
- Availability of already existing schemas
 - Reuse and adhere existing standards
 - Combine and extend when necessary

RDF and OWL

- RDF (Resource Description Framework)
 - W3C Standard model for data interchange in the Semantic Web
 - RDF documents form a labelled graph
 - Node in the graph are denoted by URIs



- OWL (Web Ontology Language)
 - W3C Semantic Web language to represent knowledge graphs
 - Based on RDF
 - OWL documents lift graphs to ontologies by adding semantics
 - Properties of relations can be defined
- → Metadata Schema and Metadata form a Linked data graph

A Metadata Schema in RDF, OWL, and 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>
```


Description of a Dataset in RDF, OWL, and 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>
```

Future Work

- Enhance system to function as interface for PID registration
- Provide metadata for archive and publication domain
- Implement browsing of stored metadata (&data)
- Provide sample scripts that automatically transfer existing to be adopted by researchers
- Create acceptance!

Thank you for your attention

Vielen Dank für Ihre Aufmerksamkeit