



# Invenio as One Module within a Holistic Service Suite for Research Data Management

---

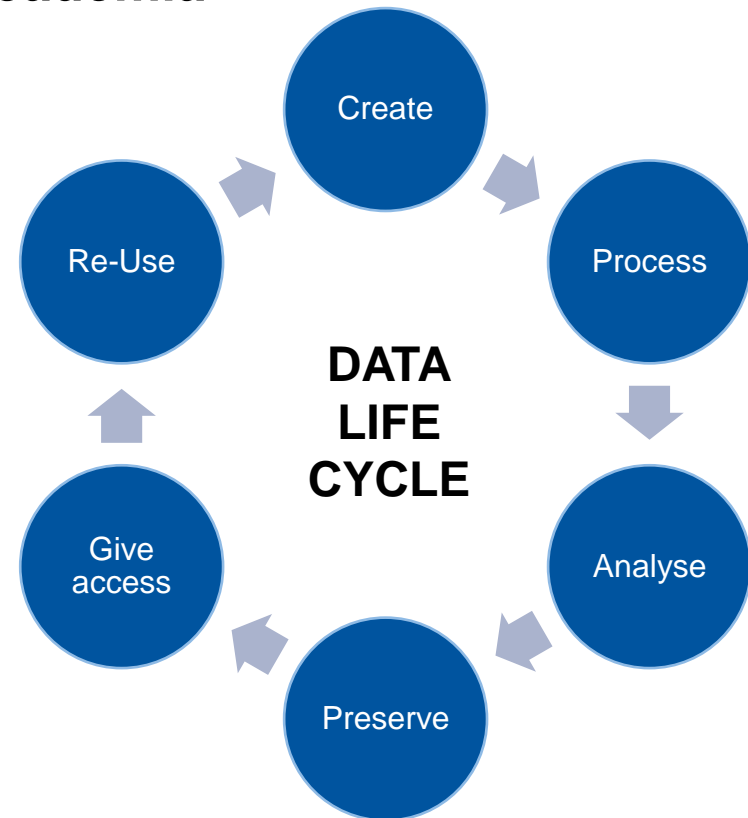
Invenio User Group Workshop 2017, Garching (Munich)  
23.03.2017

Dr. Dominik Schmitz (University Library)

# What is Research Data Management?

**A form of project management and workflow control focusing on the handling of research data in science and academia**

- ➔ Intentionally take responsibility for data
- ➔ Consider reproducibility and availability
- ➔ Organisation, documentation (metadata) and preservation



Based on: <http://www.ianus-fdz.de/it-empfehlungen/lebenszyklus>

# Research Data Management at RWTH Aachen University

---

- Project since 2014, University [guidelines](#) released in 2016
- Partners: IT Center, University library, central administration, scientific staff
- Teaching and consultancy on
  - introductory, metadata, data management plans, collaboration, publication, preservation
- ServiceDesk as the first single point of contact
- Website <http://www.rwth-aachen.de/fdm>
- Blog <https://blog.rwth-aachen.de/forschungsdaten/> (mostly in German yet)
- Technical infrastructure:

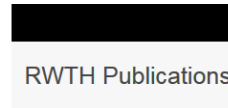
Aiming at a **modular infrastructure** that can **flexibly be integrated** into **decentral infrastructure** to support sustainable research data management at an institute level

# Data Publishing

---

## → Repositories

- Preferably a subject specific repository, see
- Generic repositories



## → Essential features

- Authority files for authors, institutes, grants
- DataCite DOI minting (delivering ORCIDS where available)
- OpenAIRE validated (EU reporting)

# Ph.D. Thesis with Published Research Data

Information | Referenzen (0) | Diskussion (0) | Dateien

**On right conjugacy closed loops and right conjugacy closed loop folders = Über rechts-konjugationsinvariante Loops und rechts-konjugationsinvariante Loop-Folder / vorgelegt von Dipl.-Math. Katharina Artig, geb. Suchantke**

Artig, Katharina<sup>\*</sup>; Hiß, Gerhard (Thesis advisor)<sup>\*</sup>; Niemeyer, Alice C. (Thesis advisor)<sup>\*</sup>

2017

Impressum Aachen 2017

Umfang 1 Online-Ressource (113 Seiten)

Dissertation, RWTH Aachen University, 2017

Veröffentlicht auf dem Publikationsserver der RWTH Aachen University

**Genehmigende Fakultät**  
Fak01

**Hauptberichter/Gutachter**  
Hiß, Gerhard; Niemeyer, Alice C.

**Tag der mündlichen Prüfung/Habilitation**  
2017-02-16

**Online**  
URN: urn:nbn:de:hbz:82-nwth-2017-019085  
DOI: 10.18154/RWTH-2017-01908

**Dokumentart**  
Dissertation / PhD Thesis

**Publikationsform**  
online

**Sprache**  
English

**Externe Identnummern**  
HBZ: HT019244747

**Interne Identnummern**  
RWTH-2017-01908  
Datensatz-ID:684327

**Beteiligte Länder**  
Germany

[...]

**Linked articles:**

Dataset  
Artig, K.  
Dataset to "On right conjugacy closed loops and right conjugacy closed loop folders" : Data of the non-associative RCC loops

Dataset  
Artig, K.  
Dataset to "On right conjugacy closed loops and right conjugacy closed loop folders" : Source Code

[...]

Open Access: PDF (PDFa) (zusätzliche Dateien)

Dieses Dokument bewerten: (Bisher nicht rezensiert)

- Zum persönlichen Korb hinzufügen
- Export als BibTeX, MARC, MARCXML, DC, EndNote, NLM, RefWorks
- Request correction
- Submit fulltext

Information | Referenzen (0) | Diskussion (0) | Dateien

**Dataset to "On right conjugacy closed loops and right conjugacy closed loop folders" : Data of the non-associative RCC loops**

Artig, Katharina (Corresponding author)<sup>\*</sup>

2017

**Online**  
DOI: 10.18154/RWTH-2017-02016  
URL: http://publications.rwth-aachen.de/record/684696/files/DocumentationData.pdf  
URL: http://publications.rwth-aachen.de/record/684696/files/DocumentationData.pdf?subformat=pdfa

**Dokumentart**  
Dataset

**Sprache**  
English

**Interne Identnummern**  
RWTH-2017-02016  
Datensatz-ID:684696

[...]

Dataset  
Artig, K.  
Dataset to "On right conjugacy closed loops and right conjugacy closed loop folders" : Data of the non-associative RCC loops of order 8

Dataset  
Artig, K.  
Dataset to "On right conjugacy closed loops and right conjugacy closed loop folders" : Data of the non-associative RCC loops of order 9

Dataset  
Artig, K.  
Dataset to "On right conjugacy closed loops and right conjugacy closed loop folders" : Data of the non-associative RCC loops of order 10

Information | Referenzen (0) | Diskussion (0) | Dateien

**Dataset to "On right conjugacy closed loops and right conjugacy closed loop folders" : Source Code**

Artig, Katharina (Corresponding author)<sup>\*</sup>

2017

**Online**  
DOI: 10.18154/RWTH-2017-02035  
URL: http://publications.rwth-aachen.de/record/684715/files/QGeneratesGroup  
URL: http://publications.rwth-aachen.de/record/684715/files/QCombinesIndex  
URL: http://publications.rwth-aachen.de/record/684715/files/ComputeTT  
URL: http://publications.rwth-aachen.de/record/684715/files/ComputeGroups  
URL: http://publications.rwth-aachen.de/record/684715/files/ComputeOrbits  
URL: http://publications.rwth-aachen.de/record/684715/files/ComputeRCCLoops  
URL: http://publications.rwth-aachen.de/record/684715/files/DocumentationSourceCode.pdf  
URL: http://publications.rwth-aachen.de/record/684715/files/DocumentationSourceCode.pdf?subformat=pdfa

**Dokumentart**  
Dataset

**Sprache**  
English

**Interne Identnummern**  
RWTH-2017-02035  
Datensatz-ID:684715

# How about Data not to be Published?

- ➔ Bitstream preservation locally at RWTH Aachen university
- ➔ Free of charge to all RWTH members
- ➔ GigaMove based „SimpleArchive“
- ➔ Persistent identifier (PID via EPIC service) without disclosure of data nor metadata



**Tivoli Storage Manager**  
Client für Sichern/Archivieren

Anschlussnummer des Agenten wird überprüft ... 75%

Version 7, Release 1, Stufe 4.2

Lizenzmaterial - Eigentum der IBM. (c) Copyright by IBM Corporation und Andere vorbehalten. Ausführliche Informationen enthält die Produktlizenz. Java und alle Logos sind Marken oder eingetragene Marken der Oracle Corporation und/oder ihrer Tochtergesellschaften.



SimpleArchive

Engellogg nr: nn234344

Start

**Datei archivieren**

Datei wiederherstellen

Profildaten

Hilfe

### SimpleArchive - Einfach und schnell große Dateien archivieren

Hier haben Sie die Möglichkeit, Dateien hochzuladen bzw. zu archivieren.

Packern Sie mehrere Dateien bitte in inhaltlich sinnvolle Einheiten zusammen (z.B. als ZIP, RAR). Alle Felder mit \* sind Pflichtfelder.

**Datei archivieren**

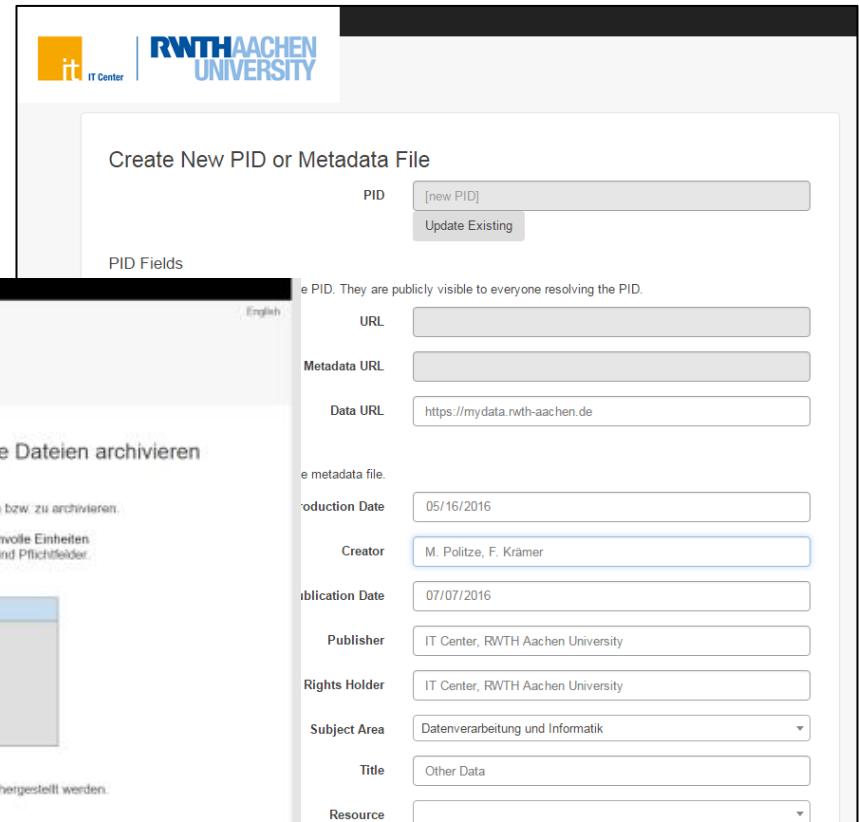
maximale Dateigröße: 2 GB

Datei auswählen \*

Durchsuchen... Keine Datei ausgewählt.

Datei archivieren

Die bisher archivierten Dateien können hier wiederhergestellt werden.



IT Center | RWTH AACHEN UNIVERSITY

### Create New PID or Metadata File

PID

PID Fields

...e PID. They are publicly visible to everyone resolving the PID.

URL

Metadata URL

Data URL

...e metadata file.

Production Date

Creator

Publication Date

Publisher

Rights Holder

Subject Area

Title

Resource

# Publication with Preserved, Non-Published Research Data

The image shows two side-by-side screenshots of the RWTH Aachen University Publications website. The left screenshot shows a standard article page for 'Line segments in homogeneous scalar turbulence'. The right screenshot shows the same article but with a 'Data Reference(s)' section highlighted by a paperclip icon, indicating the presence of non-published research data.

**Left Screenshot (Standard Article Page):**

- ub Universitätsbibliothek | RWTH AACHEN UNIVERSITY | RWTH Publications
- Suchen | Publizieren & Eintragen | Personalisieren | Hilfe
- Information | Referenzen (1) | Diskussion (0) | Dateien
- Line segments in homogeneous scalar turbulence**
- Gauding, Michael (Corresponding author) ; Göbbert, Jens Henrik ; Hasse, Christian ; P...
- In**  
Physics of fluids 27(9), Seiten/Artikel-Nr.:095102
- 2015
- Impressum [S.I.] : American Institute of Physics
- Umfang 21 S.
- ISSN 1089-7666
- Online**  
DOI: 10.1063/1.4929450
- Einrichtungen**  
1. Lehrstuhl und Institut für Technische Verbrennung (411410)
- Thematische Einordnung (Klassifikation)**  
DDC: 530

**Right Screenshot (Article Page with Data Reference):**

- ub Universitätsbibliothek | RWTH AACHEN UNIVERSITY | RWTH Publications
- Suchen | Publizieren & Eintragen | Personalisieren | Hilfe
- Information | Referenzen (1) | Diskussion (0) | Dateien
- Line segments in homogeneous scalar turbulence**
- Data Reference(s):**
  - EPIC: 21.11102/4d9960cf-15d6-11e6-8eb3-e41f1343fb94

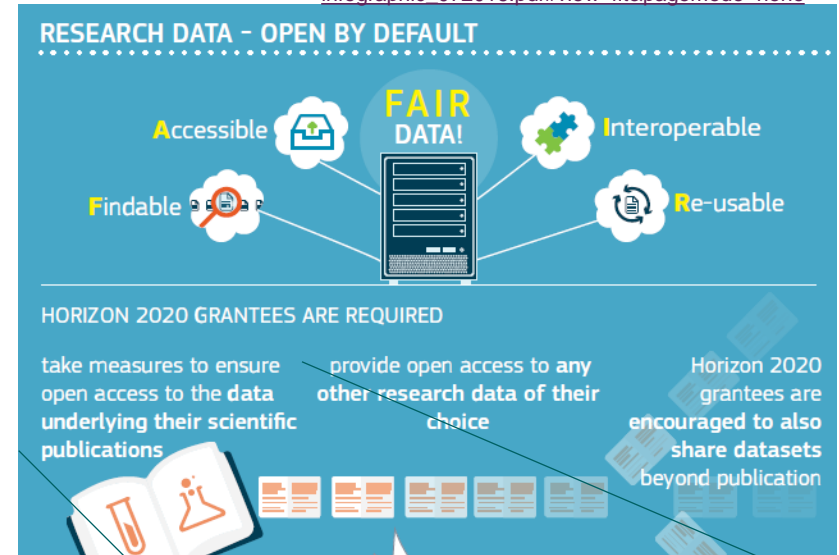
**PID Information Section:**

- it IT Center | RWTH AACHEN UNIVERSITY
- PID Information**
- This website gives information on all ePIC persistent identifiers (PID) registered by RWTH Aachen University.
- What is a PID?
- ✓ Your PID has been successfully resolved!
- In most cases the referenced data is not publicly available. Via the contact form you can request access to it.
- Contact**
- Your Name:

# Linking Publications and Research Data

- ➔ Linking Publications and Published Research Data via MARC 787  
“Linked Records”
- ➔ Linking Publications and Non-Published Research Data via MARC 786  
Tab “References”
- ➔ Update currently manually via  
“Request correction”
- ➔ Automated updates via EPIC metadata  
planned for

[http://ec.europa.eu/research/press/2016/pdf/opendata-infographic\\_072016.pdf#view=fit&pagemode=none](http://ec.europa.eu/research/press/2016/pdf/opendata-infographic_072016.pdf#view=fit&pagemode=none)



take measures to ensure open access to the data underlying their scientific publications



## Open Issues

---

- ➔ Data exists much earlier than being connected to a publication
  - API for publishing or linking if necessary metadata already available (no need for manual updates)
- ➔ Offer integrated workflow to publish, archive or link data to a publication
- ➔ Flexible “delivery” infrastructure, e.g.
  - I want to grant Tibor from CERN the rights to access my data for two weeks
- ➔ Long-term preservation
  - Currently best efforts for best practise
  - Federal State of Northrhine Westfalia: ExLibris Rosetta - piloting with use case above (154 TB)
- ➔ Collecting research data from elsewhere similarly to publications  
to document scientific performance
- ➔ Interaction with data management plan tools
- ➔ ...



# Thank you!

# What is Research Data?

---

documents (text, word)

spectra

simulations

laboratory notebooks

test responses

„Research data might include measurement data, laboratory values, audiovisual information, texts, survey data, objects from collections, or samples that were created, developed or evaluated during scientific work. Methodical forms of testing such as questionnaires, software and simulations may also produce important results for scientific research and should therefore also be categorised as research data.“ ([DFG, 2015](#))

artifacts

software

slides

models

High diversity reflecting the wide range of different scientific disciplines, research interests, and research methods

algorithms

questionnaires

transcripts

# Challenges in Research Data Management

---

Recommendations from RfII-Position paper „Leistung aus Vielfalt“, 2016

- Adjusting funding mechanisms
- Efficiency and coordination based on a (distributed) national infrastructure
- Responsible data culture
- Monitoring and quality assurance
- Comprehensive human resources development
- Stronger international ties
- Actively steering the transition