On the Decentralization of IT Infrastructures for Research Data Management

Marius Politze, Thomas Eifert
Research Data Management (RDM) at RWTH Aachen University

since 2015:
Project introducing research data management (RDM),
Cooperation of University Library, IT Center and Department Research & Career

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

Challenge:
Heterogeneous (IT) environment in institutes with very diverse infrastructure
RDM / Digital Curation

What is it good for?

- Confirmability, “good Science“

- 3rd party funded work:
  - Requirement by funding agencies
  - Confirmability to Partner (own work, deliverables, priority)
  - Research results as Intellectual Property

- Later use of precious data
  - By successor (in same Team)
  - By other Teams
  - By oneself

  ➔ Deeper exploitation of once generated data
    - Impulse for scientific findings

RDM targets existing knowledge
“one time, one head“ ➔ sustainable custody
Roles and their Shares of Benefits and Burdens

<table>
<thead>
<tr>
<th>Organisation</th>
<th>Lab Scientist</th>
<th>Dept. Head / PI</th>
<th>Head of Organisation</th>
</tr>
</thead>
</table>
| Private Domain | - generate  
- annotate  
+ use data  
+ proof of priority | + annotated data  
+ use  
+ Data Exploitation  
- workload  
+ Compliance | + Compliance  
+ Intellectual Property  
+ Good Science  
+ reputation |
| Group Domain | - annotate  
+ use  
- share  
+ use colleague’s shared data | + use  
+ access control  
+ handover | |
| Persistent Domain | + store | + store  
+ reuse | |
| Access & Reuse | + reputation | + Good Science  
+ reputation | |

Method and Goal: Direct benefits for scientists
How IT fits into individual research

Generic „standard IT“ well established.

But:
• Building blocks often independent from each other
• Responsibility for combining building blocks entirely with scientist
  • Includes local components
  ➔ „glue“ invisible for organisation

Specialized (central) solutions well accepted by scientists.
Example “Research Process”
Digitally Enhanced “Research Process”

Scientific Instrument

Attached Computer

Server

Object Storage

Metadata Manager

Data

Metadata

read

connect

search

Metadata Manager
Research is cyclic
The one and only Research Life Cycle?


Sferdean, Fe; Li, Ye; York, Jeremy; Green, Jennifer (2013): Research Data Services at the University of Michigan Library. http://hdl.handle.net/2027.42/101738

Sferdean, Fe; Li, Ye; York, Jeremy; Green, Jennifer (2013): Research Data Services at the University of Michigan Library. http://hdl.handle.net/2027.42/101738
Phases ➔ Services

Re-Use
• RWTH Publications

Access
• Metadatatool
• RWTH Publications
• Gitlab

Storage / Archival
• Archiv
• SimpleArchive
• Rosetta
• ObjectStore

Planning
• RDMO
• SharePoint

Production
• Sciebo
• Fileserver
• Database
• SharePoint
• ObjectStore

Analysis
• HPC-Cluster with Software
• Virtual Infrastructure
Reality has much more cycles
Problem Statement

- Research processes span multiple systems
- Integrated into researchers’ local IT infrastructures  
  ➔ processes span separate organizational units
- Very heterogeneous (IT) system landscape
- Legacy systems often not intended for integration

Steel silos storing sunflower seed along the west side of the small West Texas town of Ralls, Texas.
Consolidation in one API (since 2014)
Integrated Research Data Management System

UI Layer
- Web UI A
- Web UI B
- Fat Client
- Custom UI A
- Custom UI B
- ...

Process Layer
- Sync & Share
- Metadata Management
- Collaboration
- Archival
- Publication
- ...

Shared Data Layer
- Local Data
- Object Store
- PID
- Metadata Store
- ...

AAI
- Identities / Roles / Groups

Trust / Segmentation
Conceptual Model
Case Study “Simple Archive”

<table>
<thead>
<tr>
<th>User</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>upload file</td>
<td>notify user</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Temporary File System</th>
</tr>
</thead>
<tbody>
<tr>
<td>save file</td>
</tr>
<tr>
<td>create PID</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ePIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>create PID</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tape Archive</th>
</tr>
</thead>
<tbody>
<tr>
<td>schedule archival</td>
</tr>
</tbody>
</table>

| |
| create temporary download |

| |
| schedule restore | restore file |
On the Decentralization of IT Infrastructures for Research Data Management
EUNIS 2019 | Marius Politze & Thomas Eifert

Architecture

Applications
Process-Aware Services
Standardized Access to Backend Systems
Technology Dependent Interfaces, Persistent and Temporary Storage

- **simpleArchive UI**
- **REST Application Proxy**
- **Proxy DB**
  - UID
  - IKZ
  - PID
  - Name

- **Backup Portal API**

- **GigaMove API**
  - Archive Client
  - File System

- **ePIC**

- **Backup Portal DB**
  - Node
  - UID
  - IKZ
  - PID

- **Tape Archive**
  - Name
  - UID
  - Link

- **GigaMove DB**

Applications
- Process-Aware Services
- Standardized Access to Backend Systems
- Technology Dependent Interfaces, Persistent and Temporary Storage

19

On the Decentralization of IT Infrastructures for Research Data Management
EUNIS 2019 | Marius Politze & Thomas Eifert

IT Center
RWTH Aachen University
Conclusion

• Lessons Learned
  – Need to break open existing silos
  – Do not be afraid of users
  – Bottom up approach from technical perspective

• Upcoming Questions
  – How to shape future IT services and service providers?
  – How to transfer technical infrastructures to business value?
Thanks for Your attention!

Are there any questions or comments?