pSTAIX – A Process-Aware Architecture to Support Research Processes

Marius Politze, Bernd Decker, Thomas Eifert
INFORMATIK 2017 - WS18 FDM2017, 28.09.2017
Outline

• Motivation

• Problem Statement

• Modelling

• Case Study
RDM at RWTH Aachen University

• Since 2016: 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
Domain Model for Research Data

- Reseacher Working Group
- Collaborative Group (trans-regional, interdisciplinary)
- Archive
- Publishing Portals

Private Domain
- (Personal) Data Management
- Extended Data Management
- Access Rights

Group Domain

Persistent Domain

Access & Re-use

IdM / Roles / Rights / AAI

Common User-Interface
- Owncloud with RDM capabilities: metadata / DMP
- Invenio and interfaces for: Rosetta / Simplearchive

Storage / Backend Services
- ePIC / PID-Service
- Filesystem
- Archive
- Storage Technologies: TSM, ObjectStore
Problem Statement

- Existing research processes span multiple systems
- Integrated into local IT infrastructure of researchers
- Very heterogeneous (IT) system landscape
- Legacy systems often not intended for integration
Distributed Services at RWTH Aachen (pre 2014)
Consolidation in one API (since 2014)
Conceptual Model

Tier 0
Authorization and Security

Tier 1
Persistent and Temporary Storage

Tier 2
Technology Dependent Backend Interfaces

Tier 3
Standardized Access to Backend Systems

Tier 4
Process-Aware Services

Applications
Enable access to users and their identities

- Centralized Identity Management
  - Different Models: pre provisioning vs. on demand distribution

- User Identity and current Session
  - covey sessions and user information
  - between systems and steps of the process

- Protect and govern personal data
  - Enforce data minimalism
  - Protect personal data
Enable storage of processual data

- Store small amounts of data
  - Settings
  - Cache

- Reduce impact on “non-interactive” systems

- Allow different levels of storage
  - Per process (user shared)
  - Per user (process shared)
  - Per process and user (private)
**Tier 2. Technology Dependent Backend Interfaces**

**Enable access to technologies**

- Specific for backend systems
  - Legacy systems require in-depth technical knowledge
  - Modern systems often provide interfaces

- Process Independent
  - Allow re-use of backend systems
  - May allow administrative access

- Change processes base on systems
Standardize access to technologies

- Expose standardized protocols
- Enforce access in users context
- Common semantics for process entities
- Change processes orient towards software
Tier 4: Process-Aware Services

Standardize access to processes

• Integrate interfaces from systems to processes
  – Bundle mandatory steps
  – Across backend Systems

• Retain consistent semantics across processes

• Available to (external) users
  – Individualization
  – Integration
  – Automatization
End User Applications

Enabling access for end users

• Use tier 4 to deliver value services

• Allow agile software lifecycles independent from Infrastructure

• Allow individual and automated clients
Konzept: Softwarelayer

Base Applications

- Owncloud with RDM Capabilities
- Invenio
- Data Management Plans
- Metadata Tool
- simpleArchive

Infrastructure

- Virtualized Compute
- Object Store
- TSM
- Rosetta
pSTAIX - A Process-Aware Architecture to Support Research Processes
Marius Politze, Bernd Decker, Thomas Eifert
Conclusion of Case Study

- simpleArchive is available to selected researchers at RWTH Aachen since Q2 2016

- Implementation reuses existing systems and APIs
  - gigaMove
  - Backup-Portal
  - OAuth2 Service
  - REST Application Proxy

- Even a simple process need policies
  - How long is the data actually stored?
  - Who can restore the data?
  - Can archives be transferred?
  - Can archives be deleted?
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?
Conclusion

https://www.itagileshop.de/inspirieren/scrum-bierdeckel/
Thank you for your attention

Vielen Dank für Ihre Aufmerksamkeit
References


- Cantor, Scott; Kemp, John; Philpott, Rob; Maler, Eve, eds. Security Assertion Markup Language (SAML) V2.0. OASIS Standard, 2005.


References

- Kraft, Angelina; Razum, Matthias; Pothoff, Jan; Porzel, Andrea; Engel, Thomas; Lange, Frank; van den Broek, Karina; Furtado, Filipe: The RADAR Project - A Service for Research Data Archival and Publication. ISPRS International Journal of Geo-Information, 5(3):28, 2016.