



# simpleArchive as a Service

Marius Politze

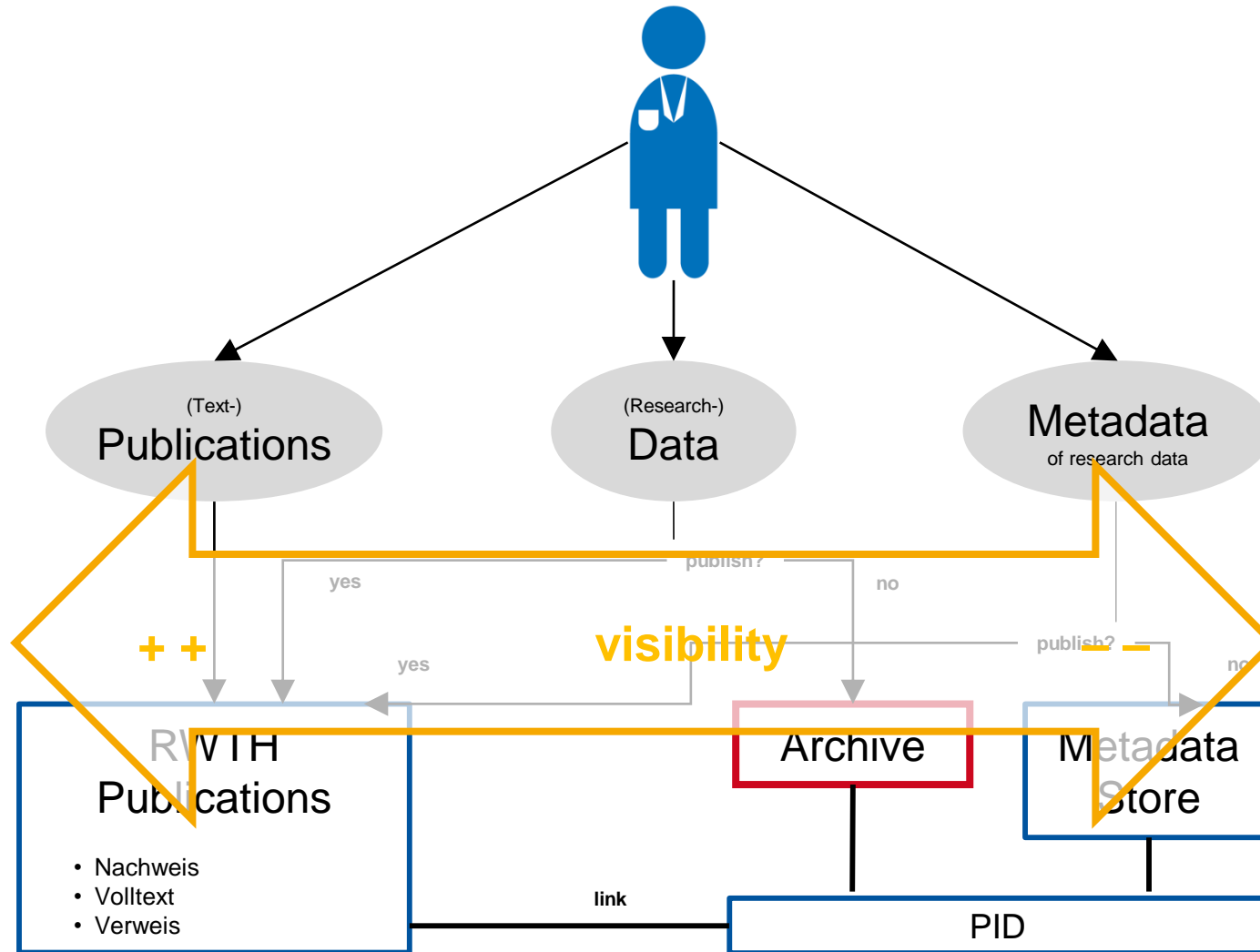
RWTH Aachen University  
IT Center

# Content

---

- Challenge: How to get researchers to archive their data?
- Our solution: make it simple
  - simpleArchive concept
  - Demo
- Scaling simpleArchive as a service
- Conclusion and future challenges

# Publications, Data, Metadata – A Research Data Infrastructure



# Archiving Until Now

- <http://www.itc.rwth-aachen.de>
- <http://www.itc.rwth-aachen.de>
- <http://www.itc.rwth-aachen.de>
- <https://doc.itc.rwth-aachen.de>

Archiv Knoten anlegen

Index von

IBM Spectrum Protect Client - InstallShield Wizard

TSM Konfiguration - Archiv

Benutzung des TSM-Cli

Anmelden bei

```
1 TCPS
2 TCPP
3 COMM
```

```
dsmj -se=archive -virtualnodename(Unix)
dsm -se=archive -virtualnodename(Windows)
```

Das Argument der Option '-virtualnodename' ist der im Backupportal angezeigte Knoten Name für das Archiv, die Sie sich im Backupportal eingerichtet haben. Nach Absetzen des Befehls erscheint ein Loginfenster, in dem Sie Ihr Passwort eingeben müssen.

Hinweis: auf Linux-Systemen muss vor dem Aufruf des Clients in das TSM-Installationsverzeichnis neuwechsell werden.

Gesteuert von Atlassian Confluence 5.4.3, der Kollaborationssoftware für Teams - IT-ServiceDesk - Impressum - Disclaimer - Feedback - Anfrage/Request online

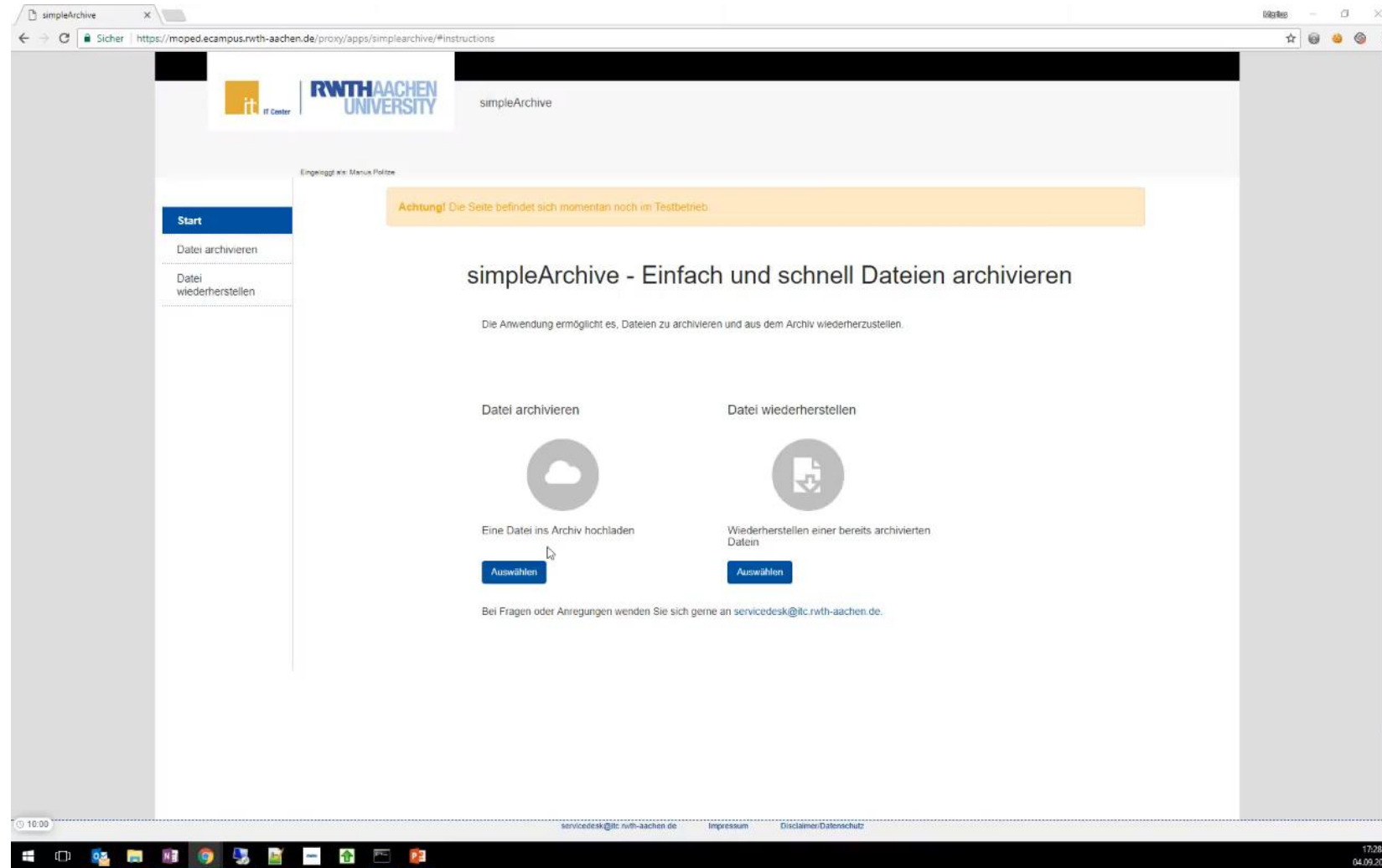
Chat Support

# Requirements

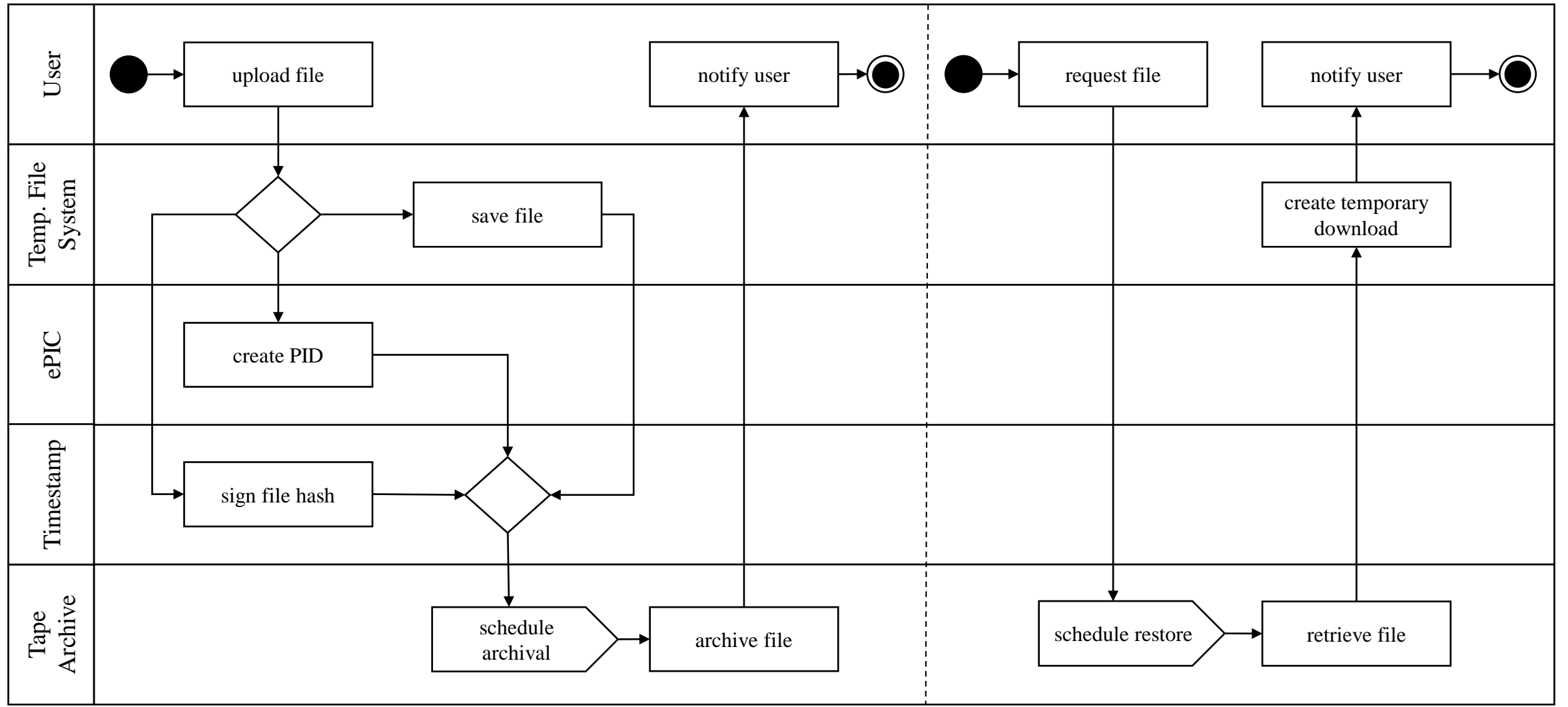
---

- Allow researchers to archive “small” files
  - Up to 2GB
  - Make it a free service so researchers will use it
  - Reduce costs by storing on tape
- Reuse existing concepts and applications
  - Allow use in federated context
  - Reduce development and maintenance costs by using available systems
- Make sharing of archived data as easy as archiving
  - Archived data is not necessarily open access
  - Let researchers restore their data
  - ... and let them share it using a simple URL
- Make archived data globally identifiable using PIDs
  - So researchers can reference it elsewhere
  - ... and can retrieve it using the PID

# Archiving with simpleArchive



# Archive and Restore Process (simplified)

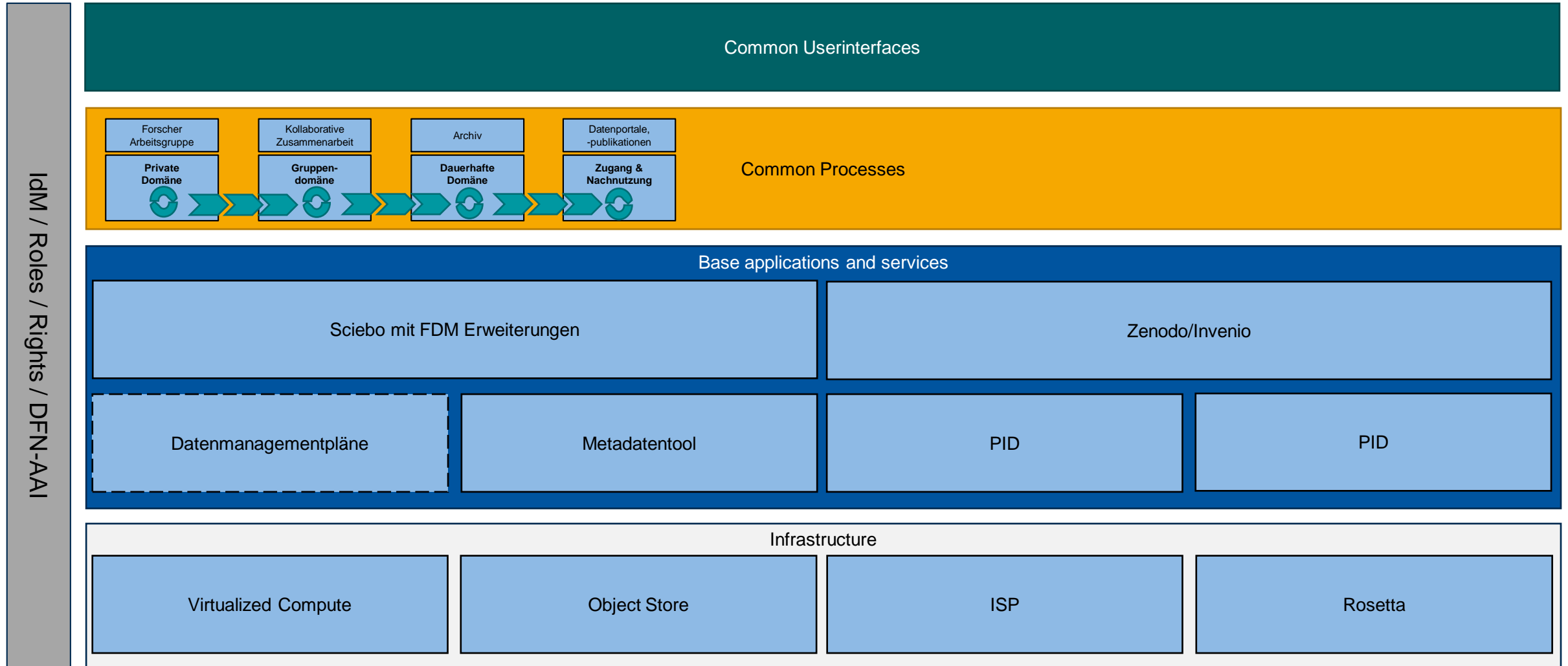


---

simpleArchive is an implementation of a process not an application!



# Concept: Software Layers



# Infrastructure Since 2016

## Loadbalancer and internet connection to DFN Network

- DNS Loadbalancing
- Redundant sites in Aachen (SW23 and WW10)
- Redundant connection to DFN Network

## User Interface: `app.rwth-aachen.de`

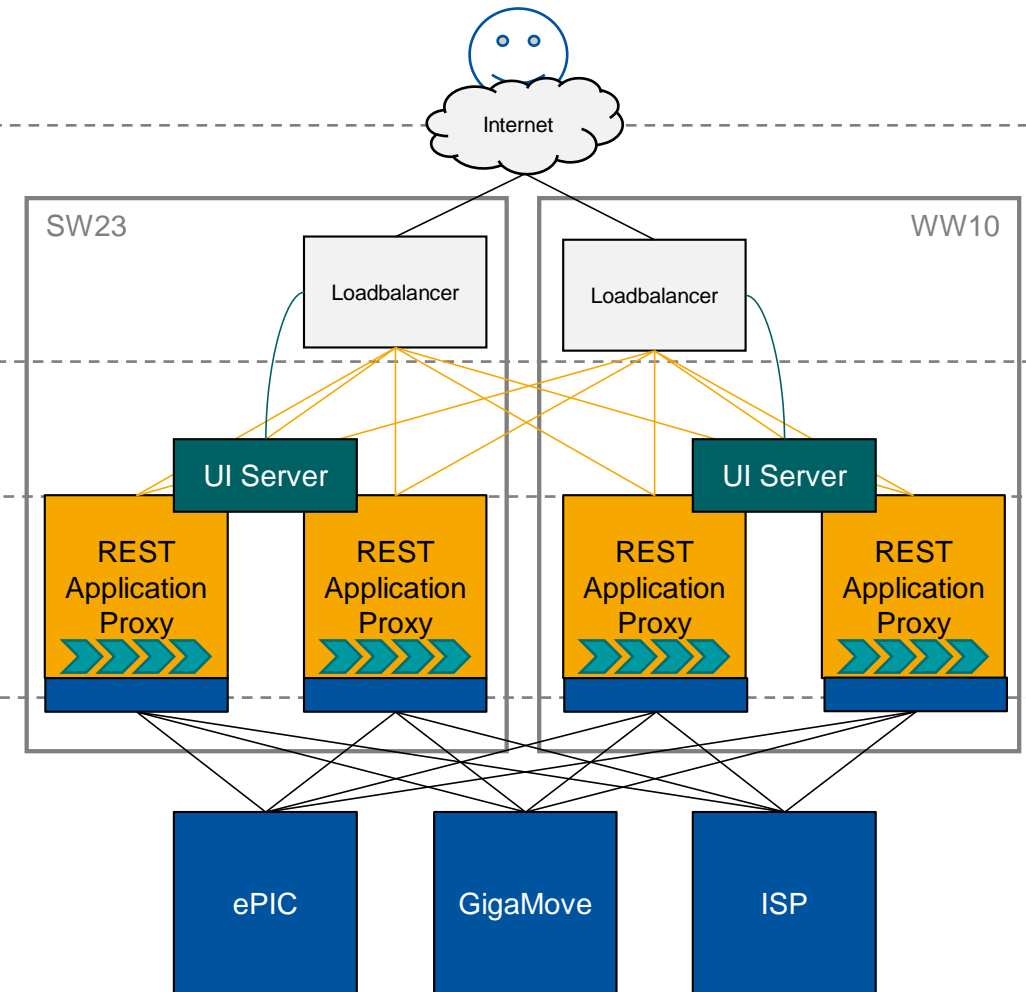
- Shared Hosts with process layer
- Accesses process layer via load balancers

## Processes: `moped.ecampus.rwth-aachen.de`

- 4 VMs at Redundant in sites Aachen (SW23 and WW10)
- Each site retains capacity to keep services available in case of site failure
- Homogeneous access to base applications and services
- Automated deployment

## Base applications and services

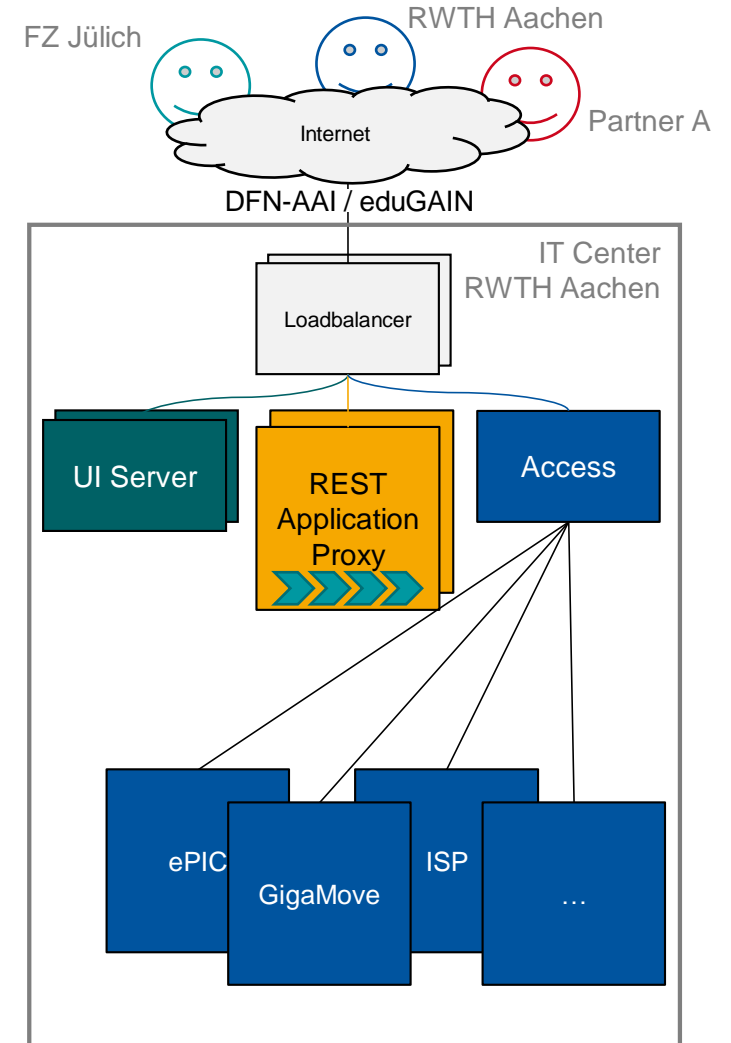
- Base on specific OLAs with the service providers
- Partially redundant, cold standby or disaster recovery
- Failures in these systems impact only dependent processes



# Scaling Out: Vision 2018

## Providing FDM Processes and Infrastructure as a service

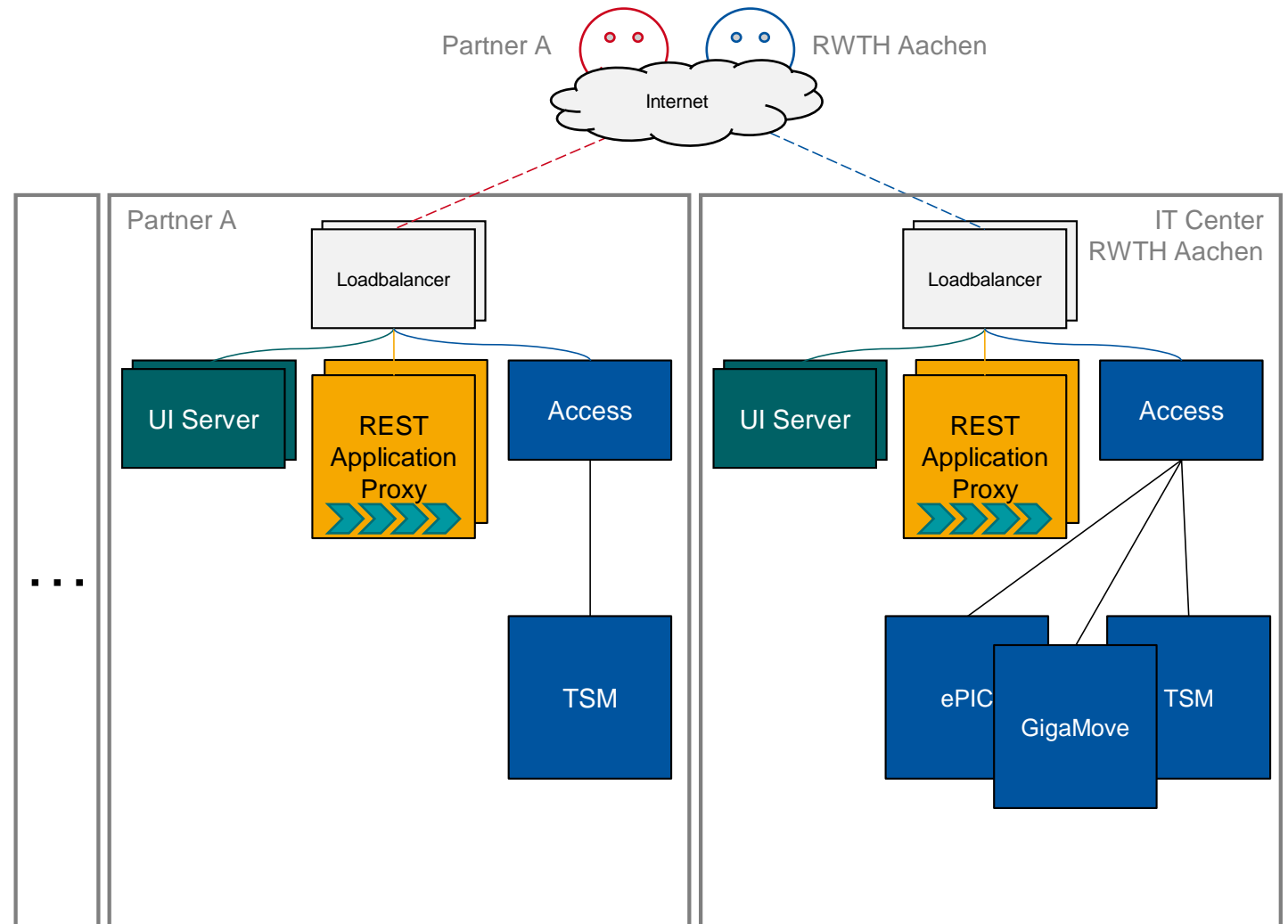
- Pro
  - Simple for customers and providers
  - Only single instance reduces maintenance costs
  - Reuses already available federated infrastructures like DFN-AAI
- Con
  - Failure in the instance impacts all customers
  - Does not scale for data or compute intensive services
  - Researchers and service providers often want to keep services local



# Scaling Out: Vision 2018

## Scaling by adding new sites

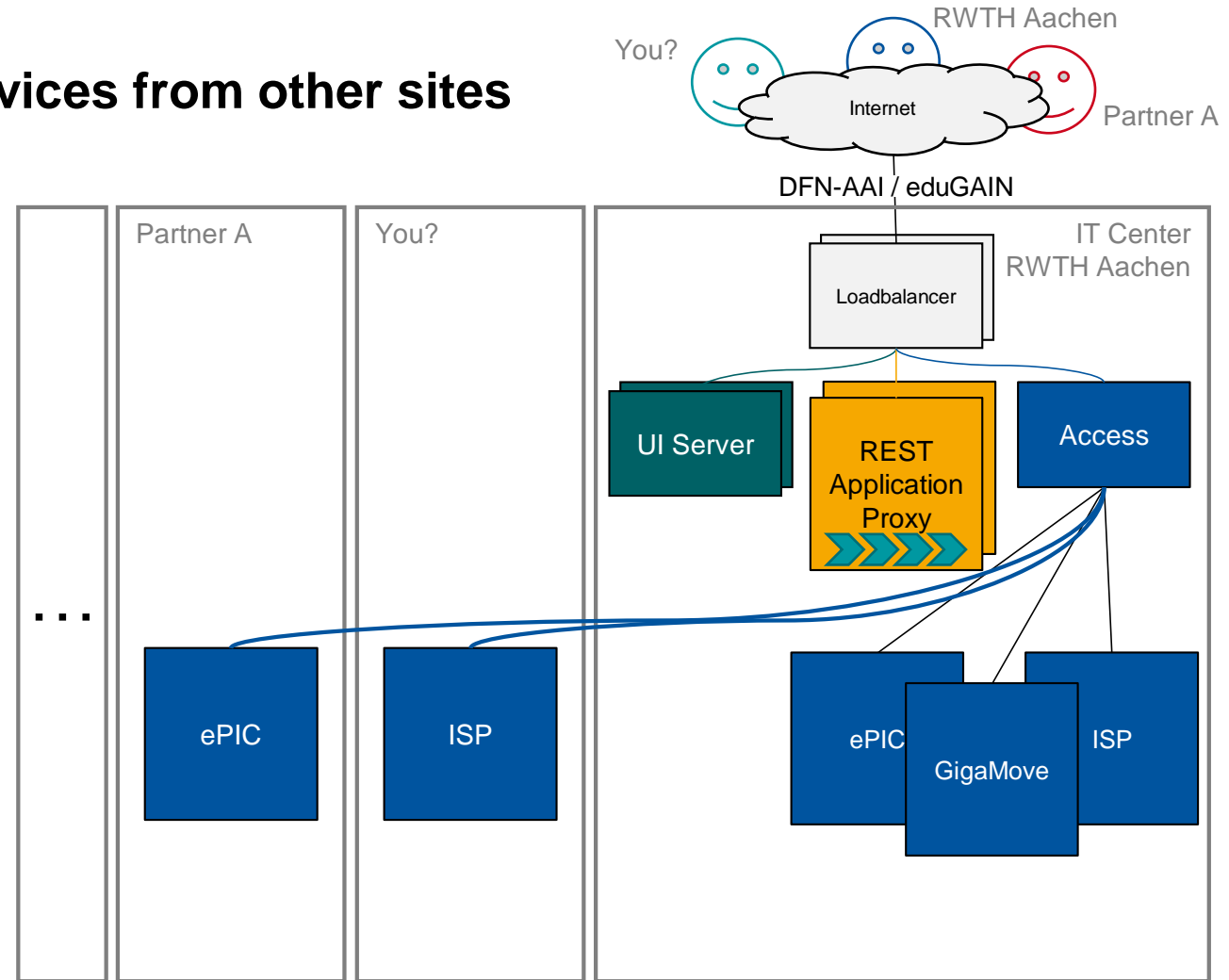
- Pro
  - Mirroring infrastructure components increases redundancy
  - Local services remain for local users and
  - Services can be used cross-site
- Con
  - Maintaining multiple infrastructures becomes expensive
  - Instead of core scientific processes sites may degenerate to support only local services



# Scaling Out: Vision 2018

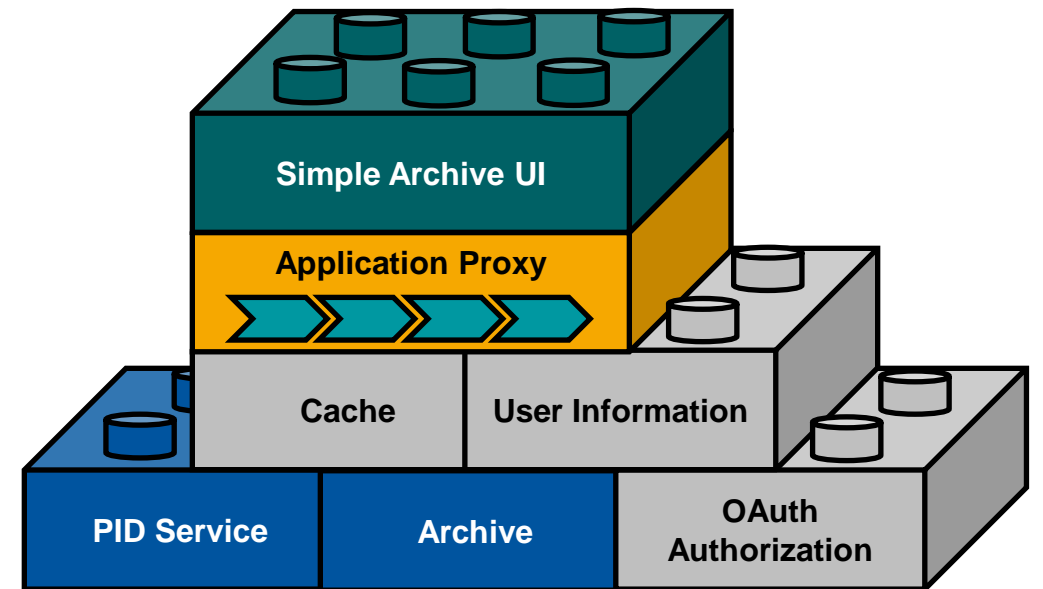
## Scaling by adding base applications and services from other sites

- Pro
  - Compute and data capacity provided locally
  - Easy cross-site reuse of services
  - Using available federative infrastructures
  - Standardized processes allow interoperability
- Con
  - Failure in process layer impacts all users
  - OLAs required to control users and processes



## Conclusion & Future Challenges

- simpleArchive is available to all researchers at RWTH Aachen since Q3 2016
- Implementation of process reuses existing systems and APIs
- Focusing on the process rather the technology reduces vendor-lockin
- Process needs to be backed by local policies
  - How long is the data actually stored?
  - Who can restore the data?
  - Can archives be transferred?
  - Can archives be deleted?
- Combine scaling methods to build a process oriented cloud-like ecosystem



**Thank you for your attention**

**Vielen Dank für Ihre Aufmerksamkeit**