On the Decentralization of IT Infrastructures for Research Data Management

Marius Politze, Thomas Eifert

IT Center | RWTH AACHEN UNIVERSITY

RWTH Aachen University (2018)

~45,000 Students

~5,000 Internationals
from 117 Countries

~10,000 enrollments
in winter term 2018/19

~560 Professors

~8,500 Employees
260 Institutes

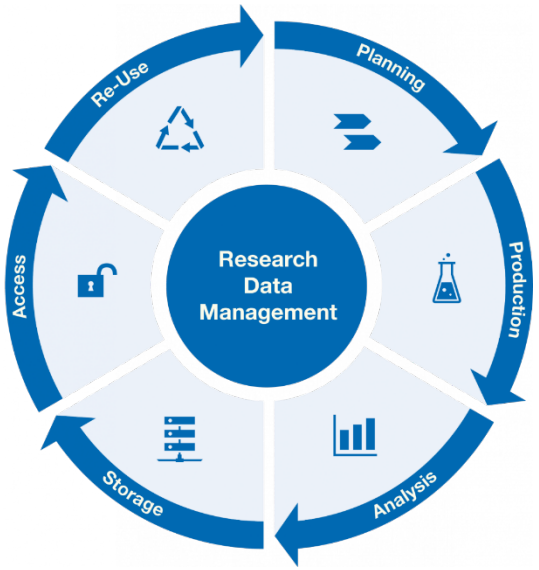
9 Faculties
157 Courses of study

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

		Organisation	
	Lab Scientist	Dept. Head / PI	Head of Organisation
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

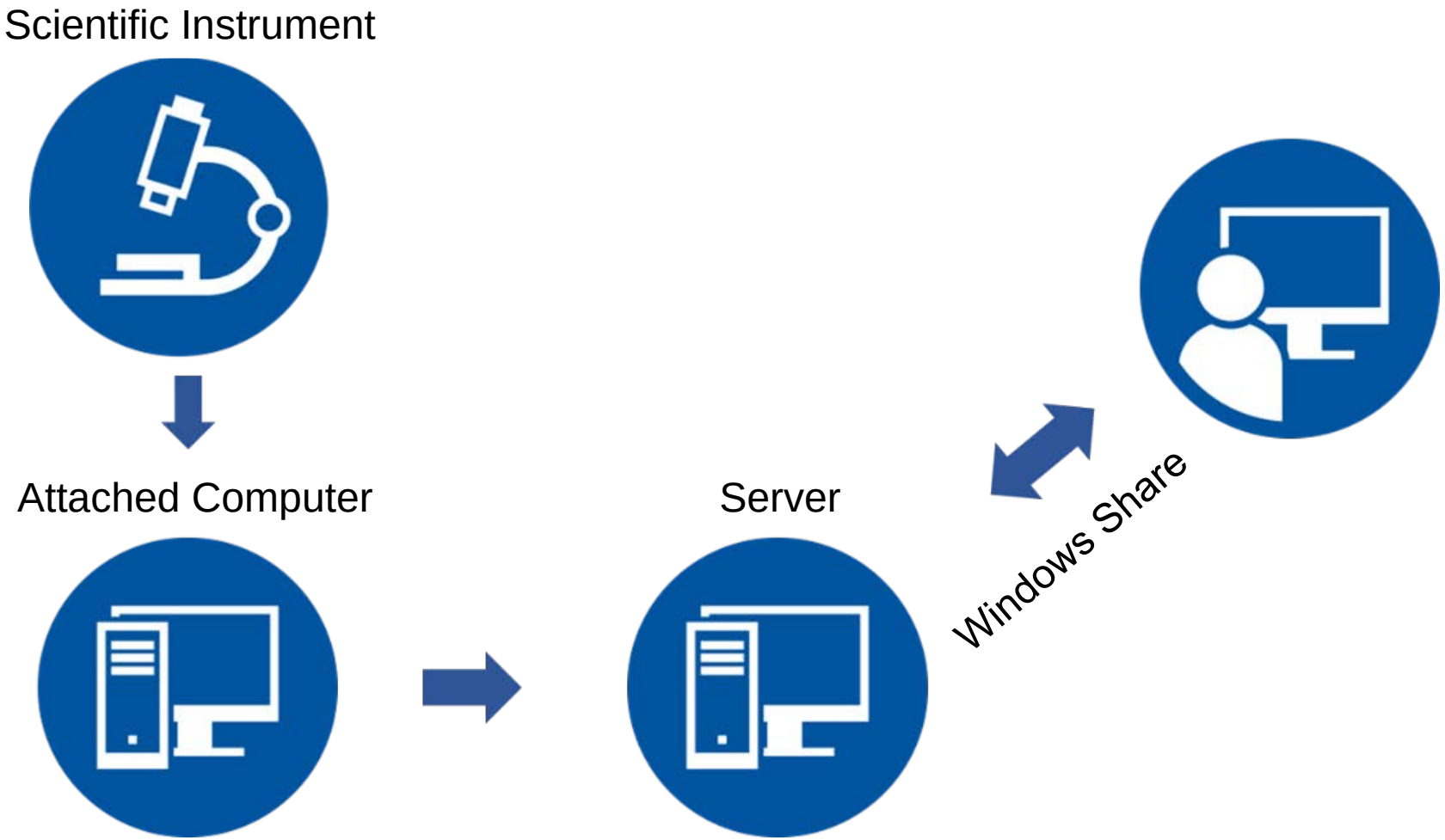
Generic „standard IT“ well established.

But:

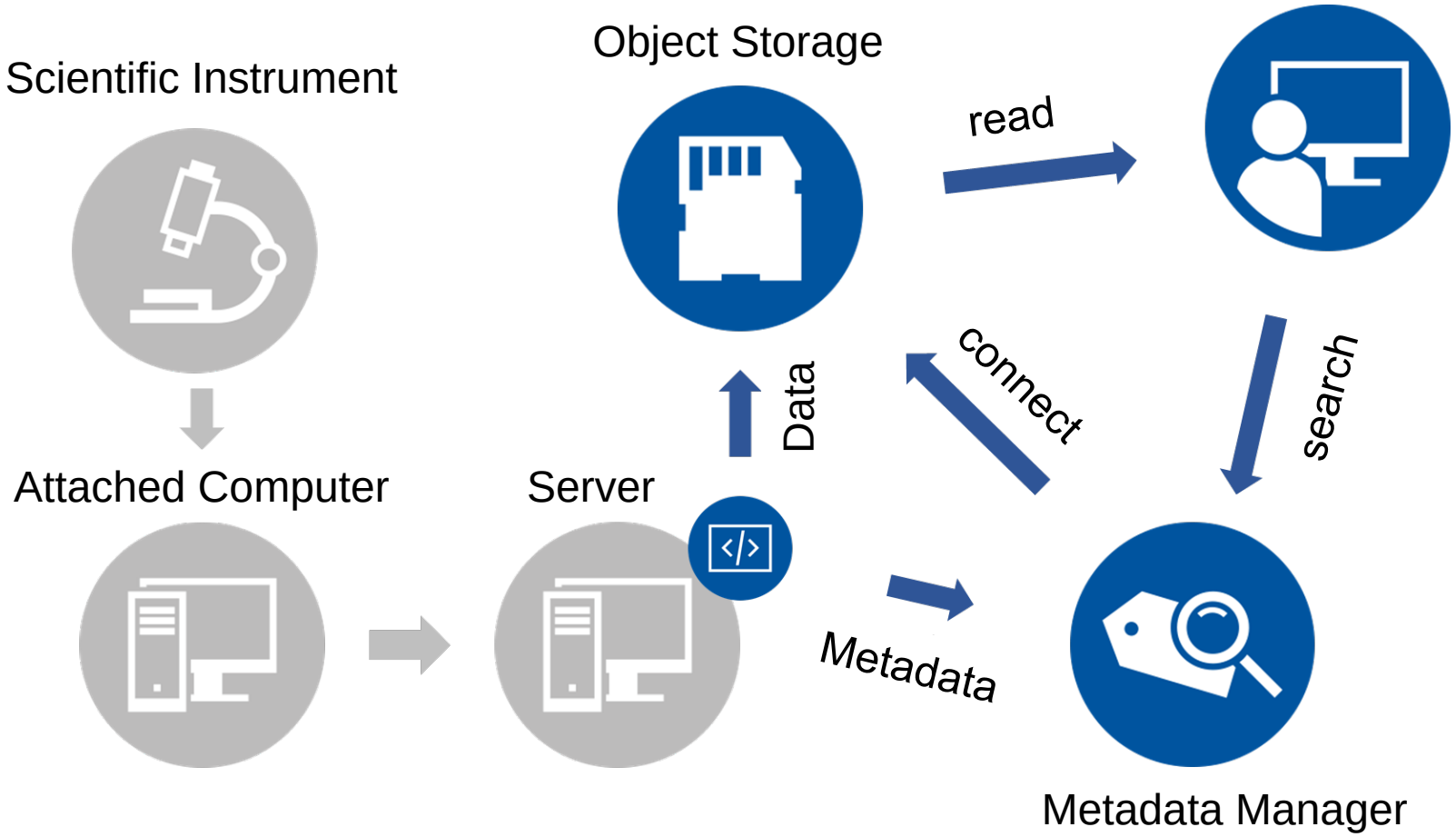
- Building blocks often independent from each other
- Responsibility for combining building blocks entirely with scientist
  - Includes local components
  - ➔ „glue“ invisible for organisation

Specialized (central) solutions well accepted by scientists.

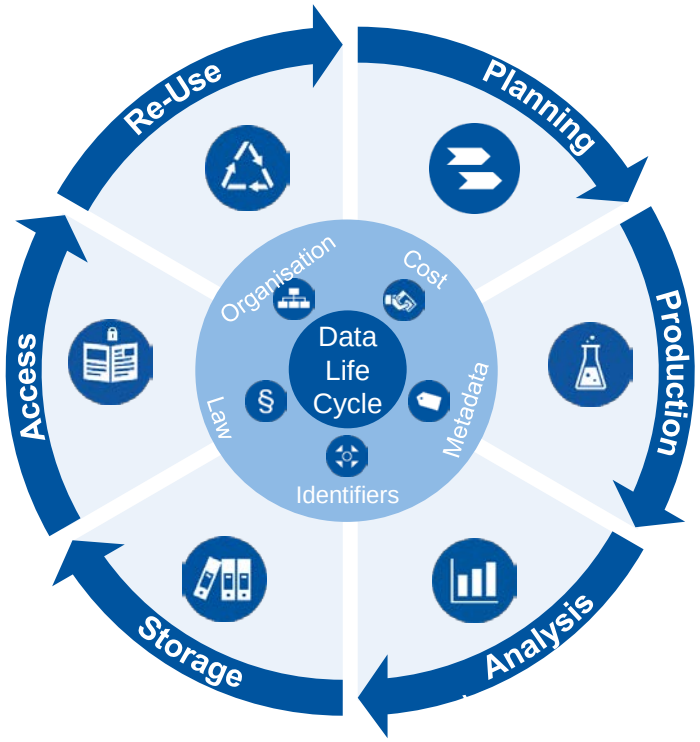
Example “Research Process”



Digitally Enhanced “Research Process”

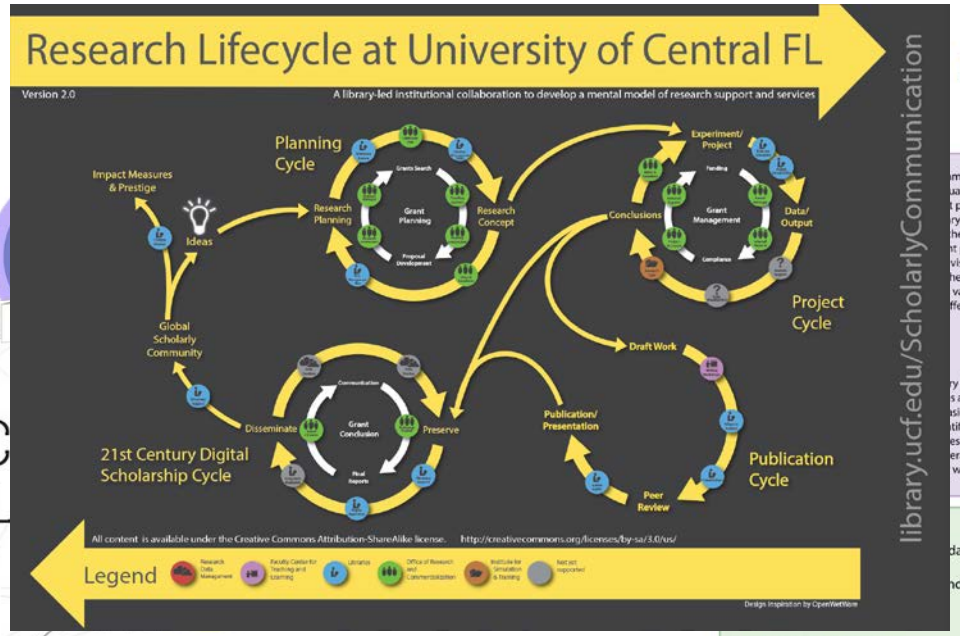


Research is cyclic

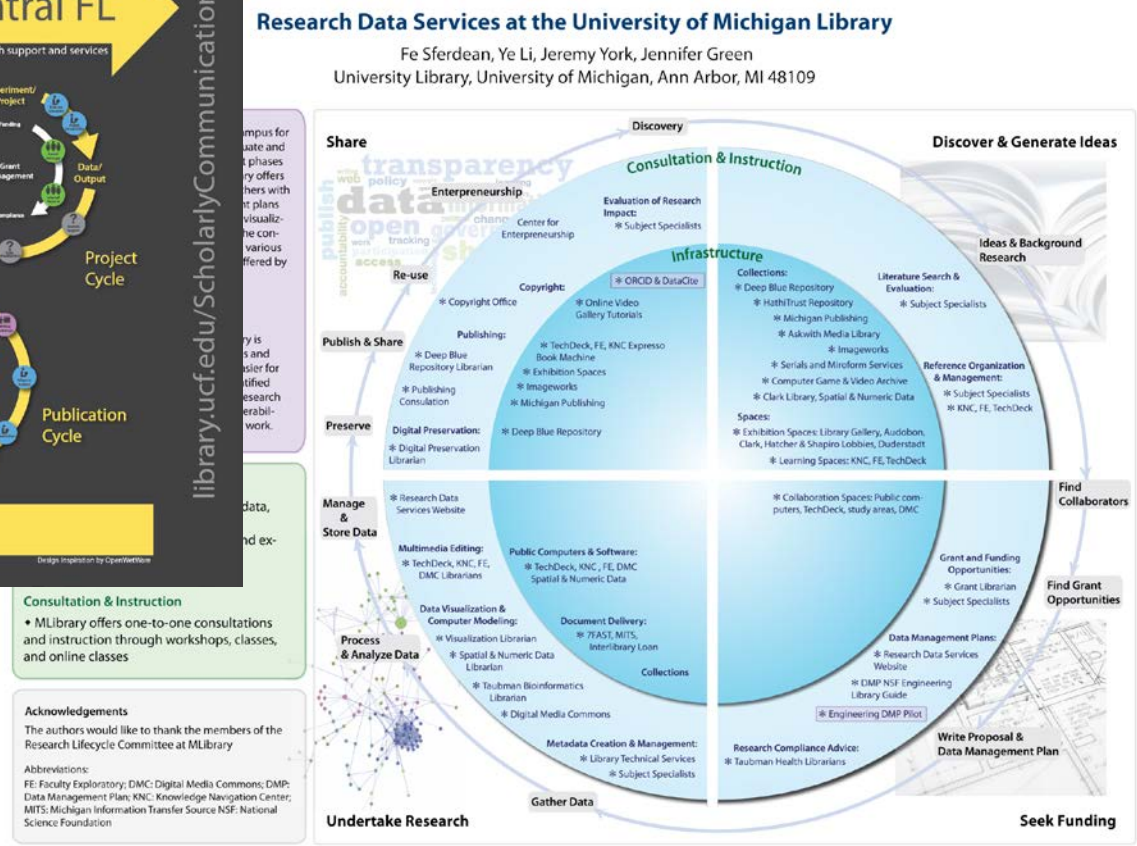


The one and only Research Life Cycle?

UCF Libraries (2017): Research Life Cycle at University of Central FL.
https://library.ucf.edu/about/departments/scholarly-communication/overview-research-lifecycle/



Northam, Julie (2014): Introducing the BU Research Lifecycle diagram!
http://blogs.bournemouth.ac.uk/research/2014/10/07/introducing-the-research-lifecycle/

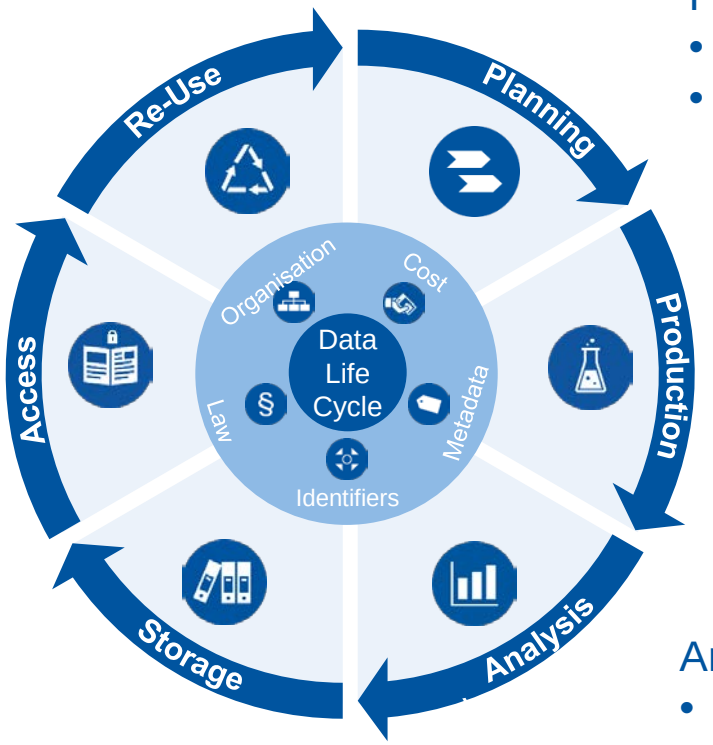


Sferdean, Fe; Li, Ye; York, Jeremy; Green, Jennifer (2013): Research Data Services at the University of Michigan Library
http://hdl.handle.net/2027.42/101738

Phases ➔ Services

Re-Use
- RWTH Publications

Planning
- RDMO
- SharePoint



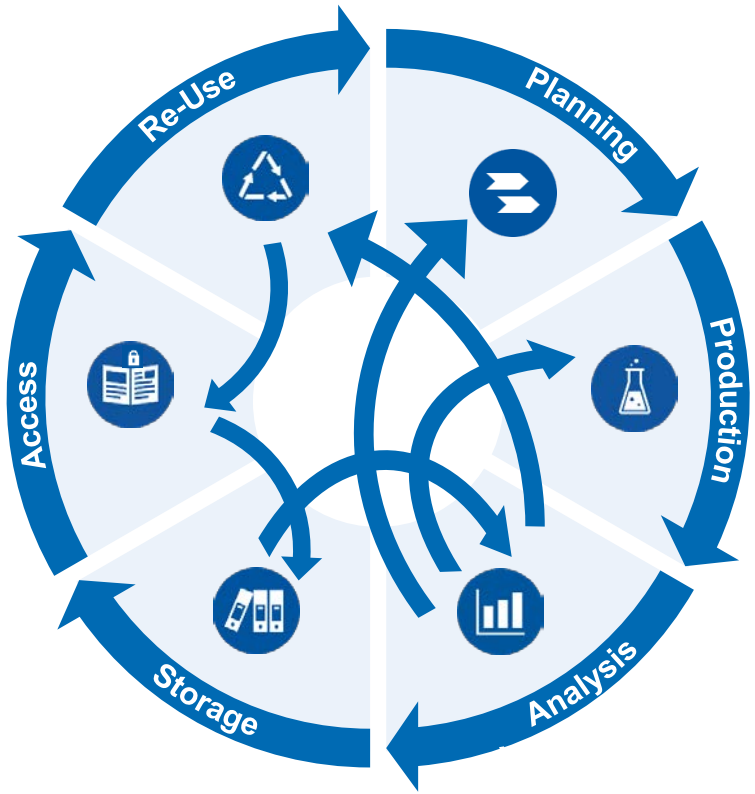
Access
- Metadatatool
- RWTH Publications
- Gitlab

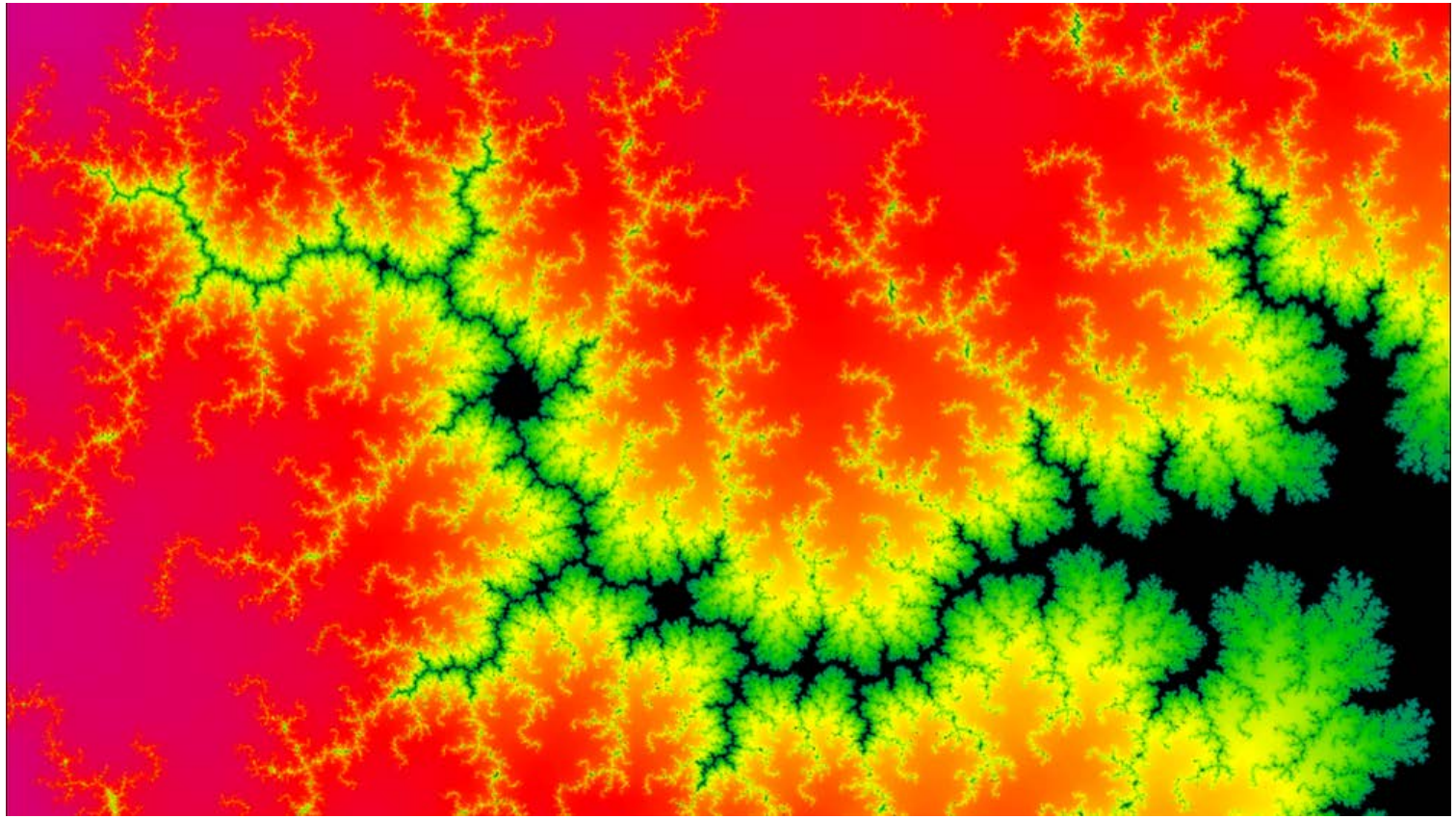
Production
- Sciebo
- Fileserver
- Data bases
- SharePoint
- ObjectStore

Storage / Archival
- Archiv
- SimpleArchive
- Rosetta
- ObjectStore

Analysis
- HPC-Cluster with Software
- Virtual Infrastructure

Reality has much more cycles



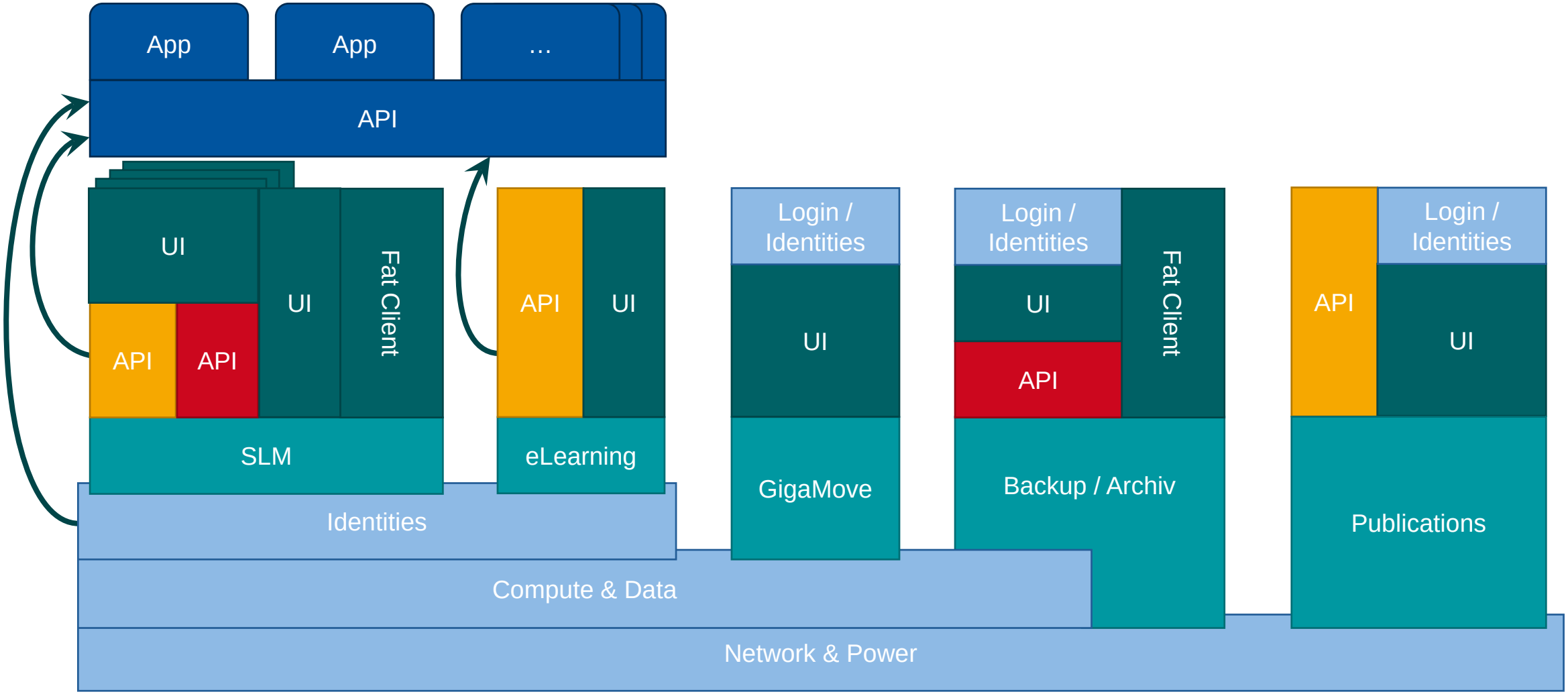


Problem Statement

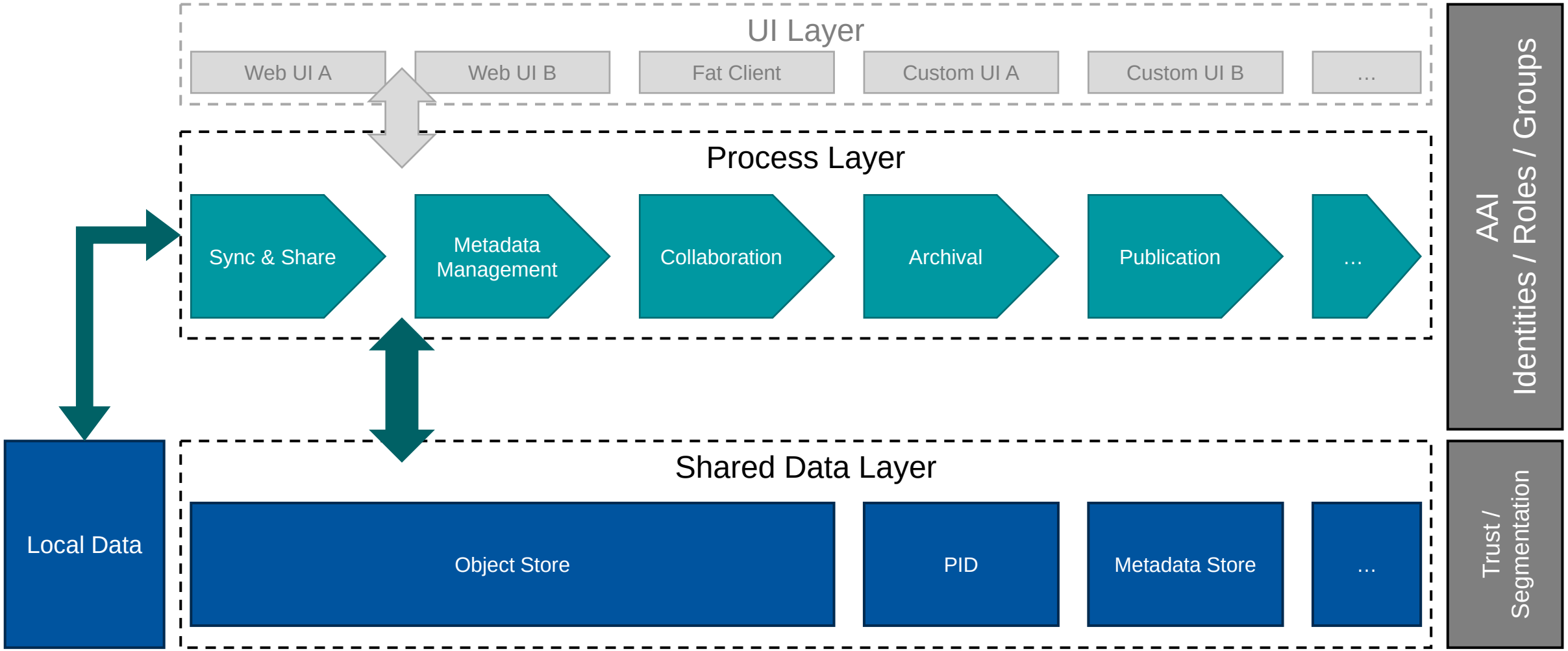
- Research processes span multiple systems
- Integrated into researchers' local IT infrastructures
  ➔ processes span separate organizational units
- Very heterogeneous (IT) system landscape
- Legacy systems often not intended for integration

Steel silos storing sunflower seed along the west side of the small West Texas town of Ralls, Texas.
By Wikipedia User leaflet. https://commons.wikimedia.org/wiki/File:Ralls_Texas_Grain_Silos_2010.jpg

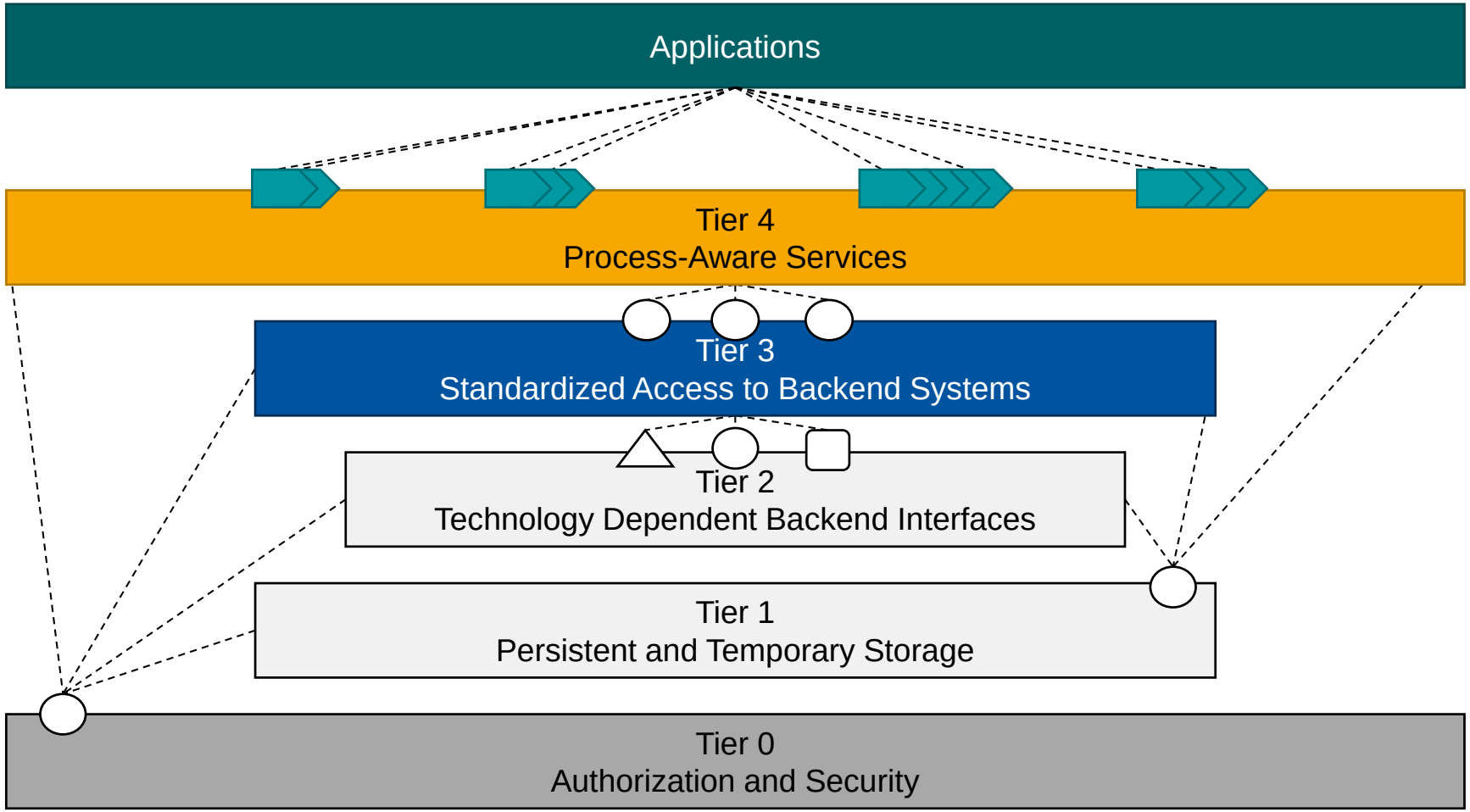
Consolidation in one API (since 2014)



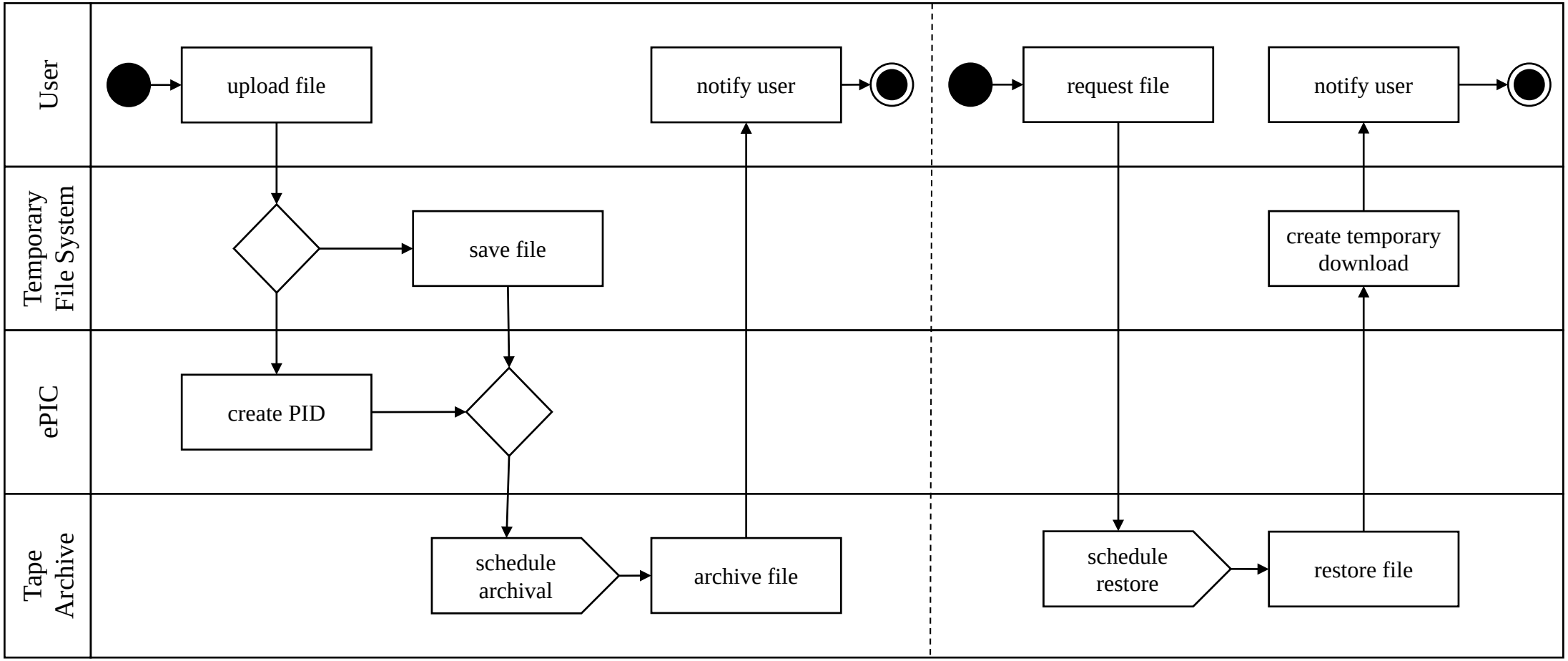
Integrated Reseach Data Management System



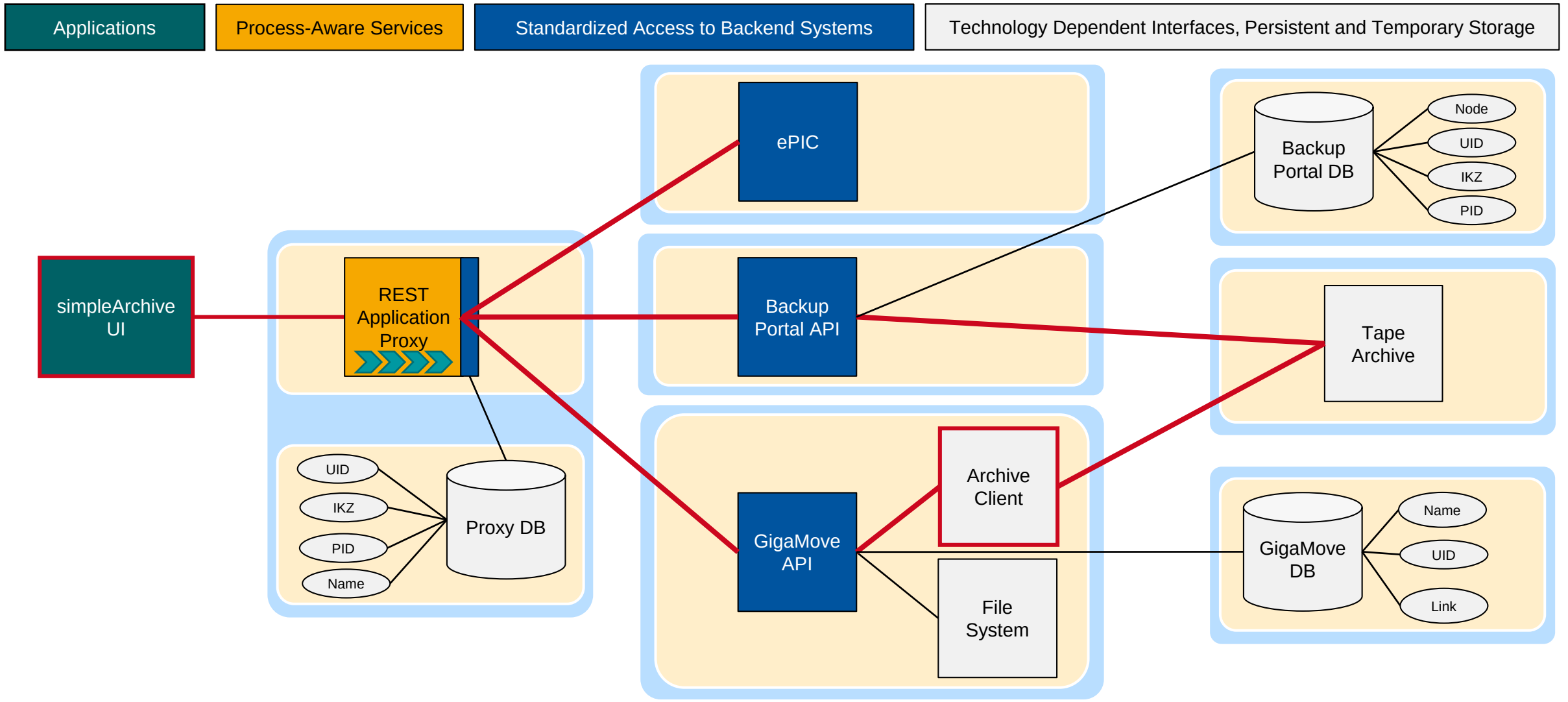
Conceptual Model



Case Study “Simple Archive”



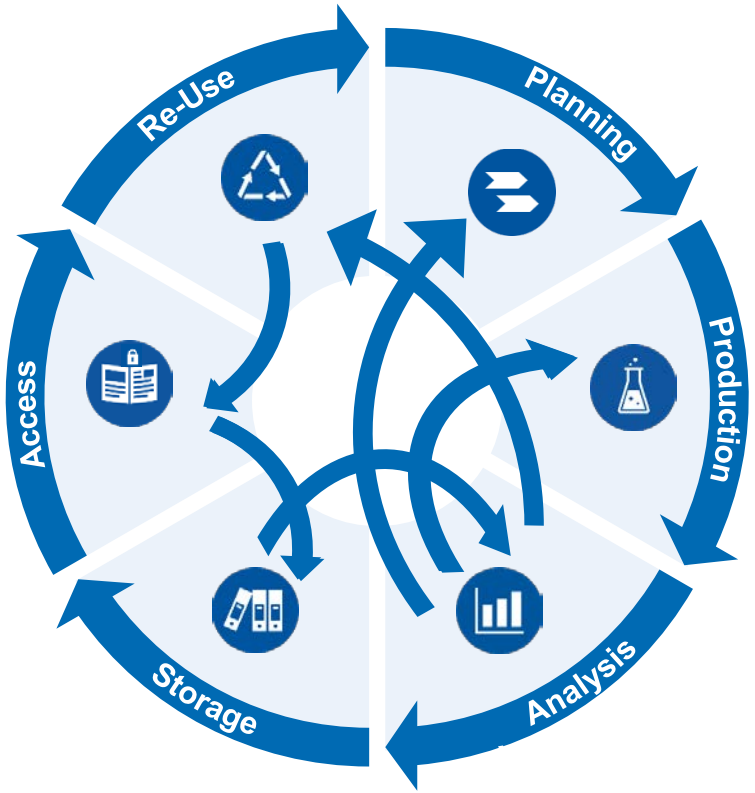
Architecture



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?