



Ausblick: Auf dem Weg zur Integrationsplattform

DINI-nestor-Workshop

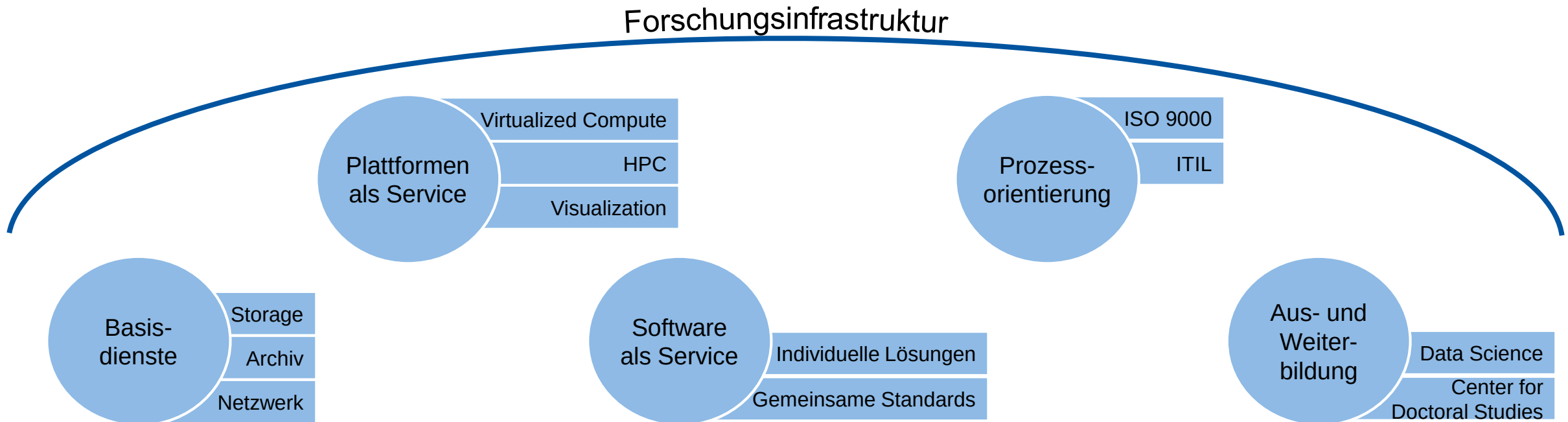
„Bedarfserhebungen - Grundlage für passgenaue Infrastrukturen?“

29./30.01.2019

Marius Politze

RWTH Aachen University

„Infrastruktur“ meint neben Hardware auch Investition in Prozesse, Dienste und „Köpfe“



(Digitale) Forschungsinfrastrukturen

„(Digitale) Forschungsinfrastrukturen“ - was ist das?

Aktuell viele Buzzwords:

Industrie 4.0, Digitalisierung...

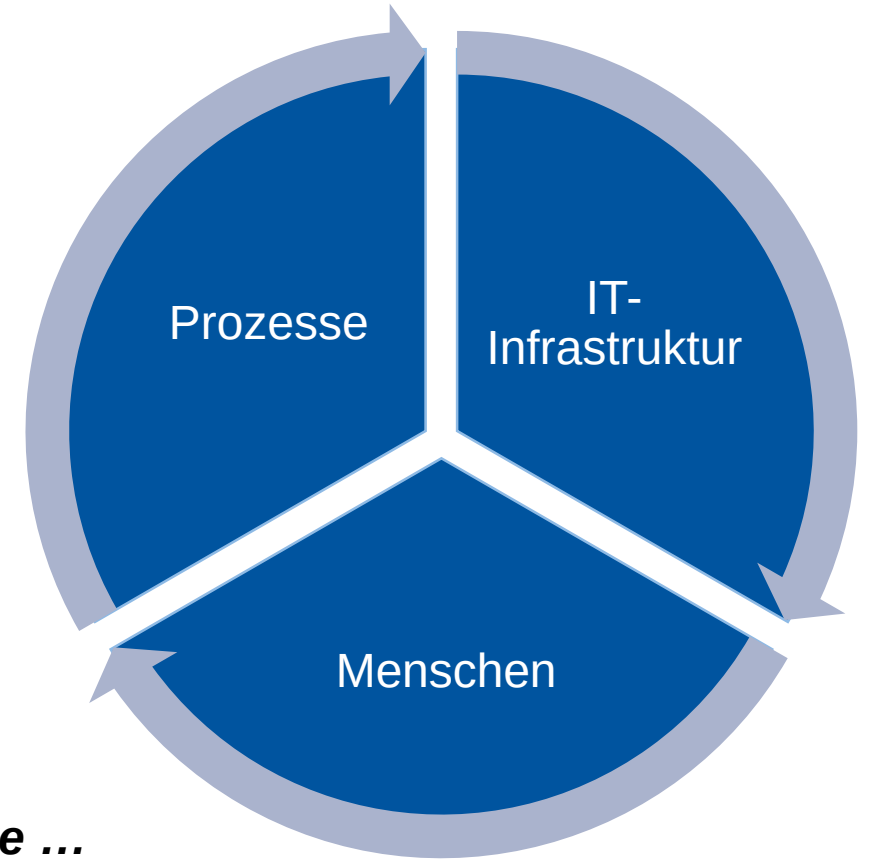
Es geht für die Hochschulen um die Frage:

„Wissen die Hochschulen, was die Hochschulen wissen?“

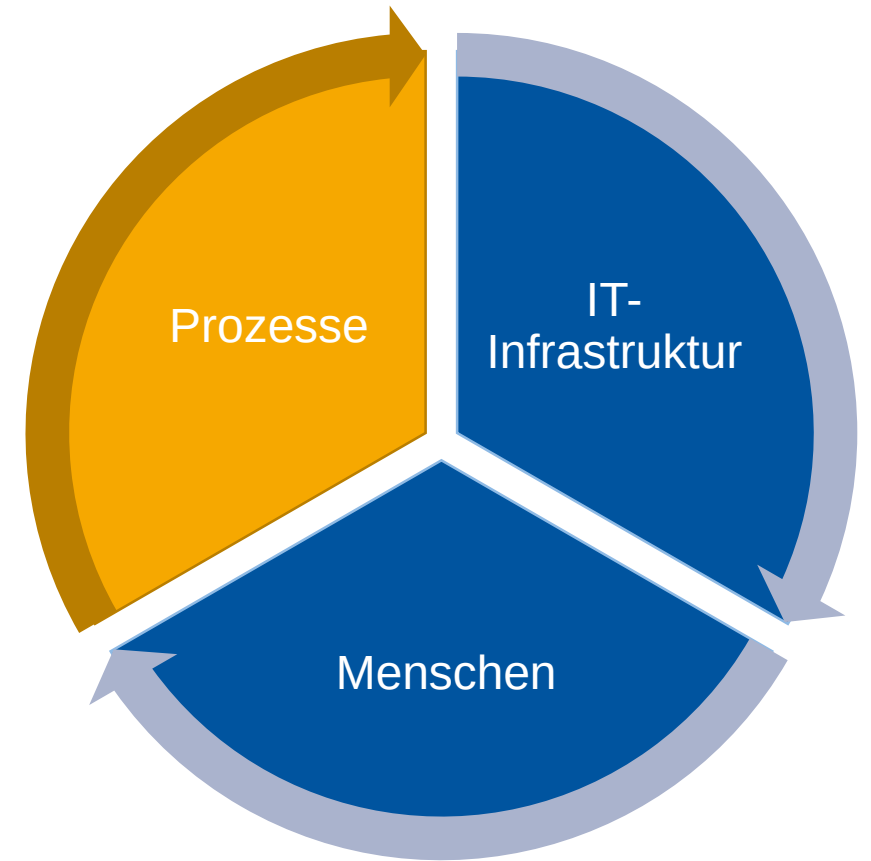
Mindestens an der RWTH Aachen gilt aktuell:

„Die RWTH weiß nicht, was die RWTH weiß!“

... zumindest nicht alles, was sie wissen sollte ...



Prozesse



Aktivitäten im Forschungsdatenmanagement

seit 2015:

Projekt zur Einführung eines Forschungsdatenmanagements (FDM),
enge Zusammenarbeit UB, IT Center, Dez. 4.0 Forschungsförderung

2016: umfangreiches Weiterbildungsprogramm zum FDM

2016: abgestuftes Beratungskonzept

2017: simpleArchive & Metadatentool:

Archivierung und Beschreibung von Forschungsdaten

2018: Git & ObjectStore:

Speicherstrukturen für das Management von Forschungsdaten

2016: Nutzung externer Dienste, z.B. GWDG

2017: Intensive Kooperation mit dem FZ Jülich

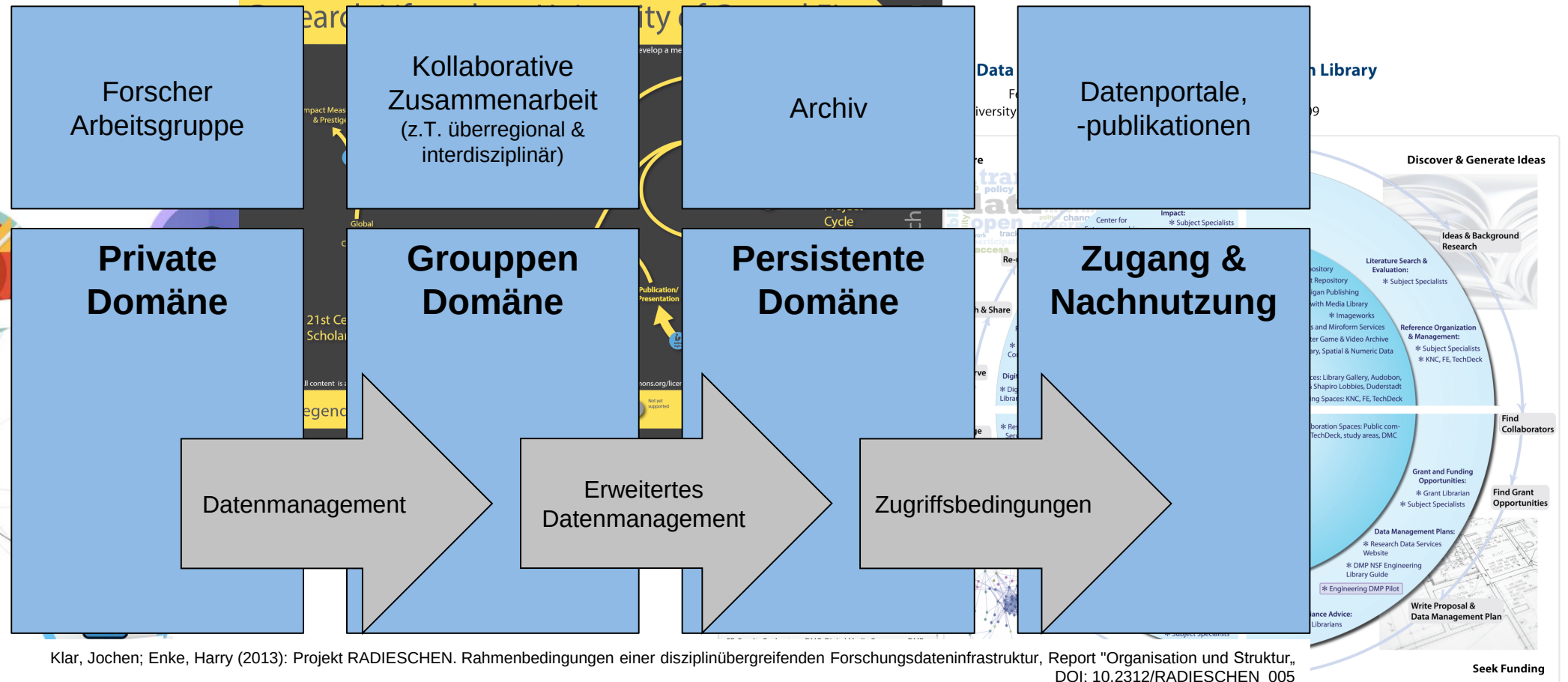
2017: Intensive Kooperation mit der TU Darmstadt

2018: Dienstleistungen im Rahmen von NFDI4Ing



Der eine Forschungs(-daten) Lebenszyklus?

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

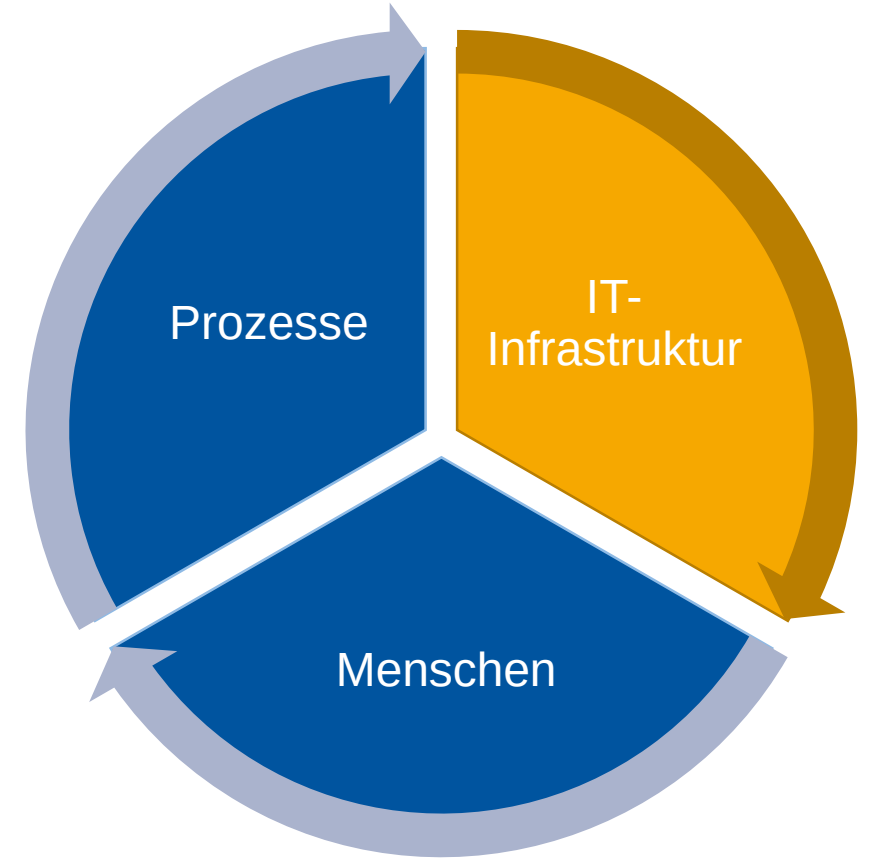


Klar, Jochen; Enke, Harry (2013): Projekt RADIESCHEN. Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur, Report "Organisation und Struktur," DOI: 10.2312/RADIESCHEN_005

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>

IT-Infrastruktur



Systemlandschaft an der RWTH Aachen (unvollständig)

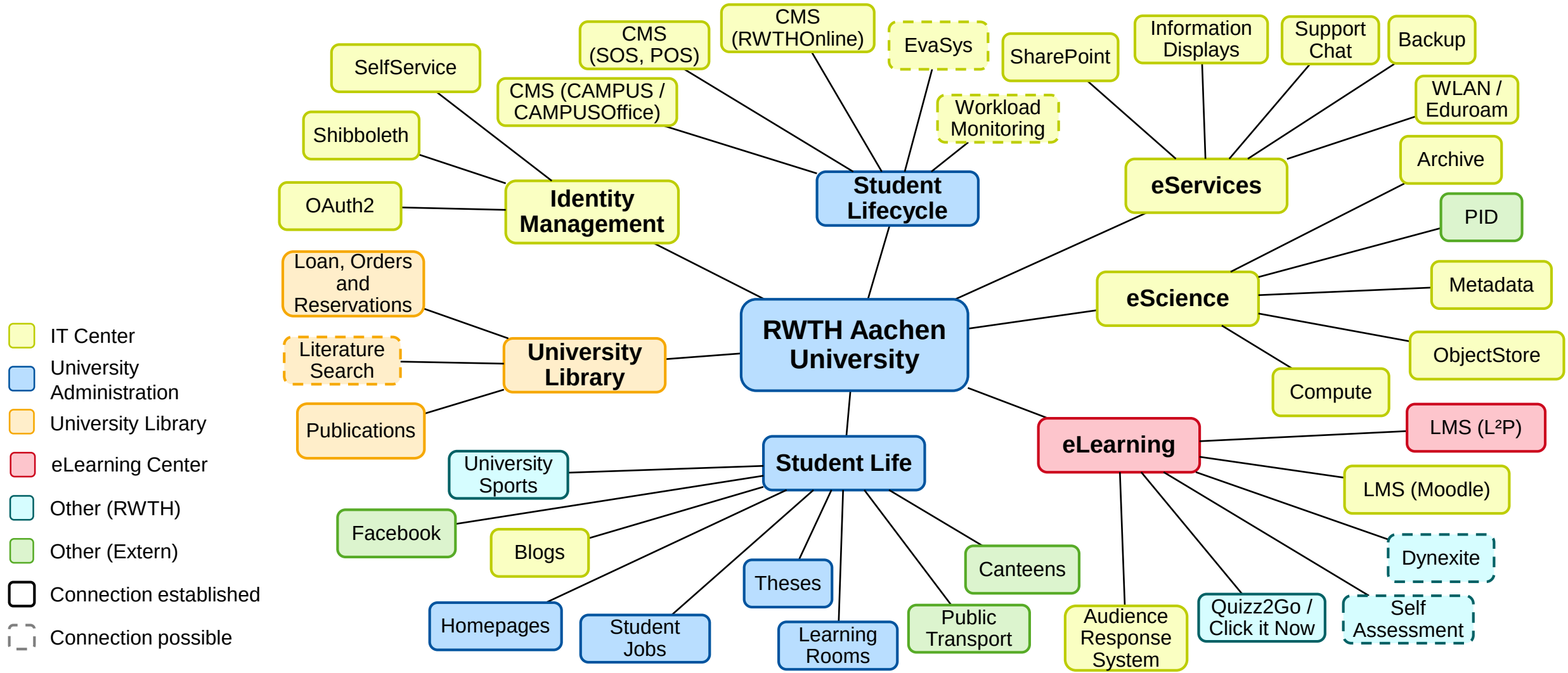
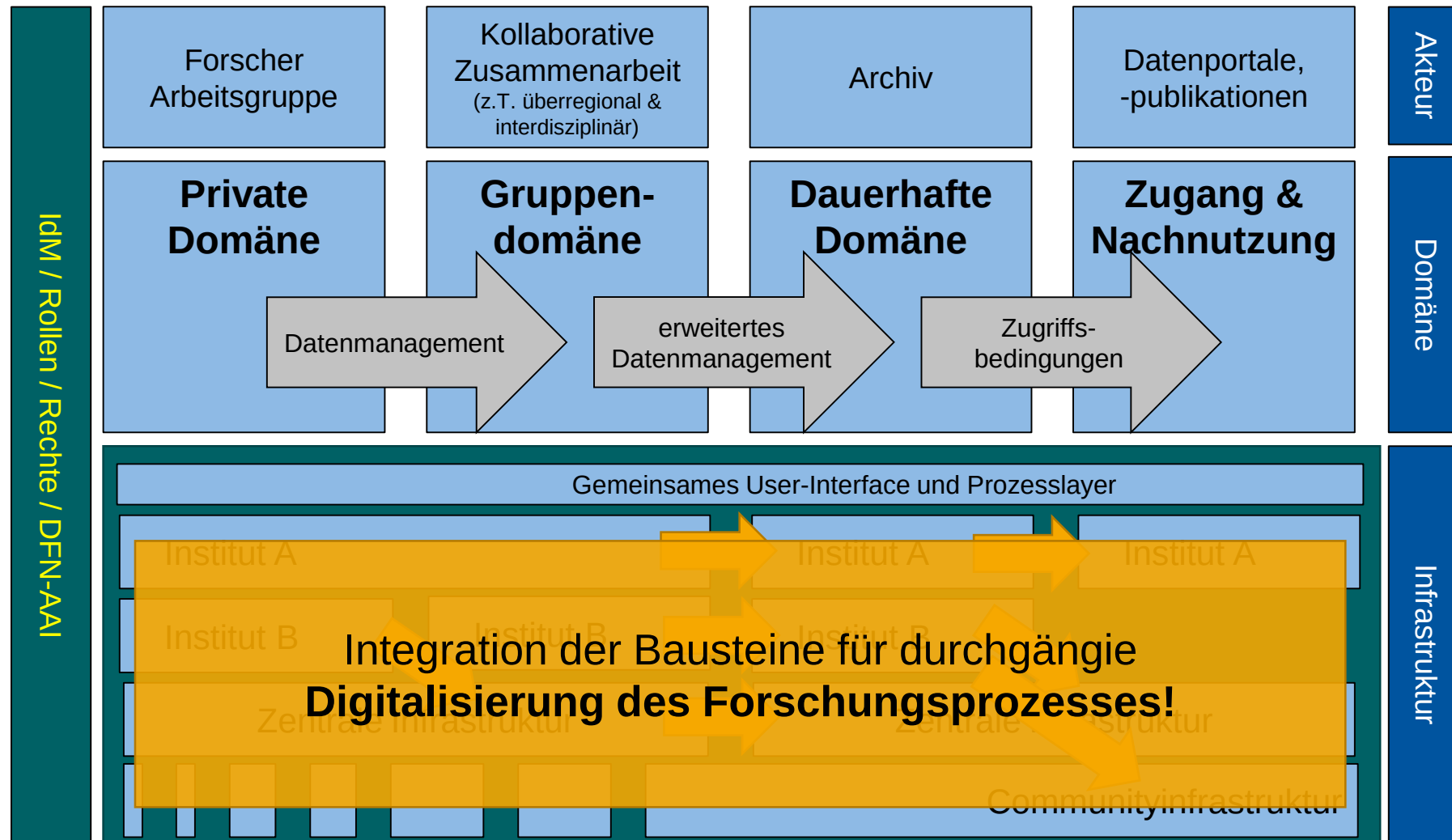


Abbildung im Forschungsdaten Lebenszyklus



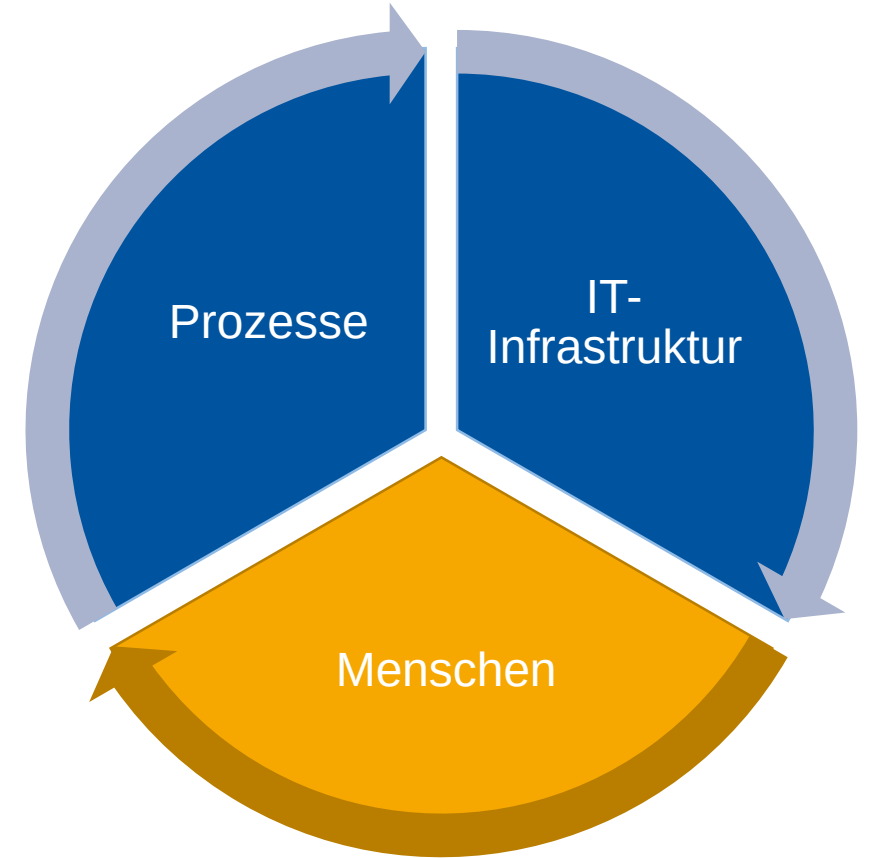
Existierende Infrastrukturen

- Aktuelle Prozesse nutzen eine Vielzahl von Systemen
- Integration in den lokale Abläufe der Forschenden
- Sehr heterogene IT-Systemlandschaft
- Teilweise lange Lebenszeit von Altsystemen
- Manueller Datentransfer zwischen Systemen



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

Menschen

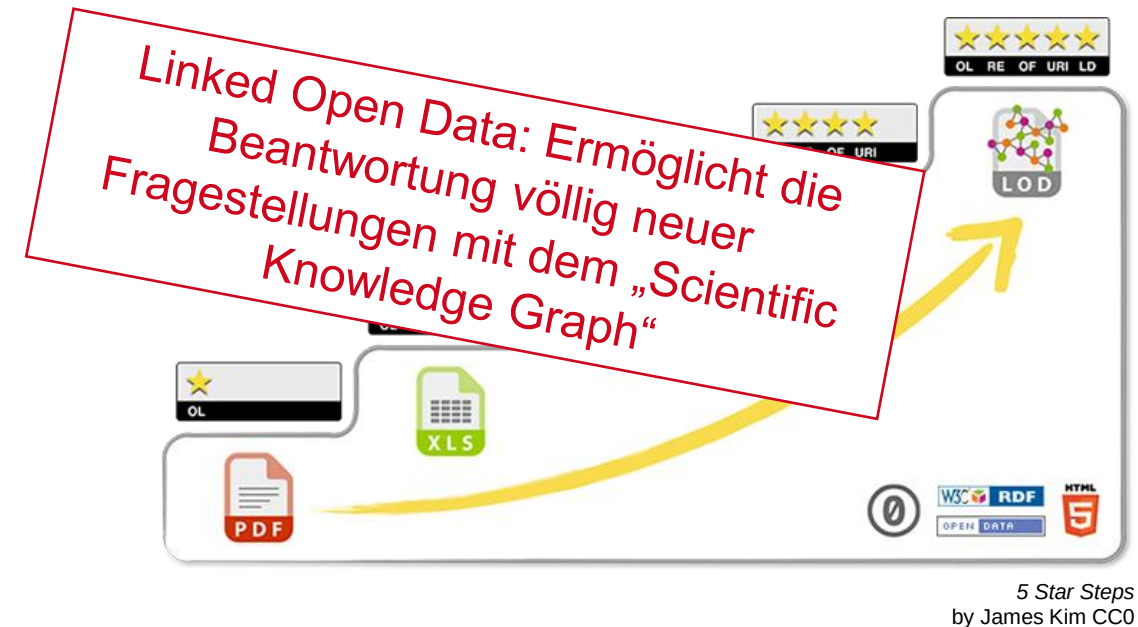


Binäre Einordnung von Services: „pain relievers“ vs. „gain creators“

- F. Piller, Chair for innovation management, RWTH Aachen

- Pain Relievers
 - Notwendig (aber nicht „geliebt“)
 - Nur sichtbar, wenn defekt
- Gain Creators
 - Adressieren den Forschenden in seinen Tätigkeiten
 - Forschung
 - Lehre
 - Projektmanagement
 - Werden als Wettbewerbsvorteil gesehen und geschätzt

Infrastrukturen unterstützen „FAIR guiding principles“ und „Linked Open Data“



Und nun?

Aufbau einer IT-System übergreifende FDM-Infrastruktur: Integrationssystem

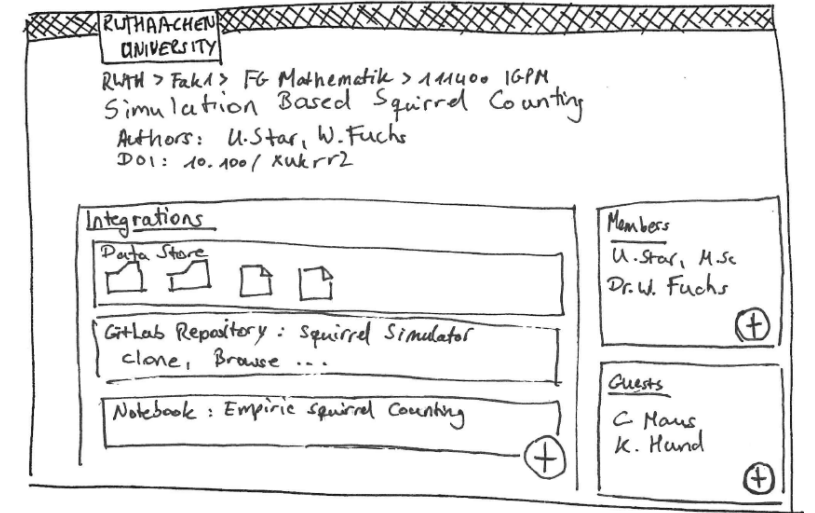
- Sammlung von Anforderungen an das System in den Kategorien
 - Technologie
 - Speicherung
 - Organisation
 - Berechtigungen
 - Integration
- Priorisierung
 - Insgesamt 33 Anforderungen
 - Nicht trennscharf, Teilweise widersprüchlich
- Umfrage unter IT-Verantwortlichen und Forschern
 - Verteilung der Prioritäten 1 = sehr wichtig, 2, 3, ..., 10, 0 = unwichtig
 - Jede Priorität (außer 0) darf nur 1x vergeben werden
- Vergabe in zwei Kategorien
 - Short-Term: Features für die nächsten 12 – 24 Monate
 - Long-Term: Features für den langfristigen Aufbau der FDM-Infrastruktur

Featurematrix

Erfolgsfaktoren (sowohl Short-Term als auch Long-Term)	
S1	Daten in beliebiger Größe können abgelegt werden. Das System kann sowohl mit sehr großen, als auch mit vielen kleinen Dateien umgehen.
I5	Eine Anbindung an FD-ObjectStore ermöglicht mit dort abgelegten Dateien zu interagieren.
T3	Das System nutzt state-of-the-art Technologien um die Anforderungen bestmöglich zu erfüllen.
O5	Die übergreifende Suchfunktion sucht in erfassten Metadaten und in Volltexten von gängigen Dateiformaten.
B2	Gruppen von Nutzenden können flexibel gebildet werden.

Ausblick: Ein FDM-Integrationssystem

- Integrationssystem: onboarding für FDM Dienste mit Anbindung an
 - Object Store, Archiv
 - GitLab, sciebo
 - HPC Cluster, ...
- Management Web Interface mit Single Sign-On
 - RWTH, DFN, Géant: SAML/Shibboleth
 - Offen für Social Sign-On für Gäste
- Abbildung der Organisationsstrukturen und -Rollen z.B.
 - Administratoren
 - Forschungsmanager
- Übergreifende Suche, Metadaten und Prozesse
 - Flexible Metadatenschemata, automatische Extraktion
 - FDM Basisprozesse: Versionierung, Archivierung, Veröffentlichung, Kollaboration
 - “Scriptable” Interfaces
- Open-Source Strategie
- Pilotkunden ab Q4 2019



- Prozesse
 - Individuell und spezialisiert
 - Wissenschaftliche Praxis steht im Vordergrund
- Infrastruktur
 - Integrationsfähigkeit und Interoperabilität
 - Nicht nur an IT denken: “Prozesse, Dienste und Köpfe”
- Menschen
 - Short-Term-Gain für Forscher notwendig
 - Kein One-Size-Fits-All



<http://www.agilemanifesto.org/>