



Zusammenarbeit im Wissenschaftsnetz: Datenhaltung und Datensicherung als kooperative Dienste

Thomas Eifert

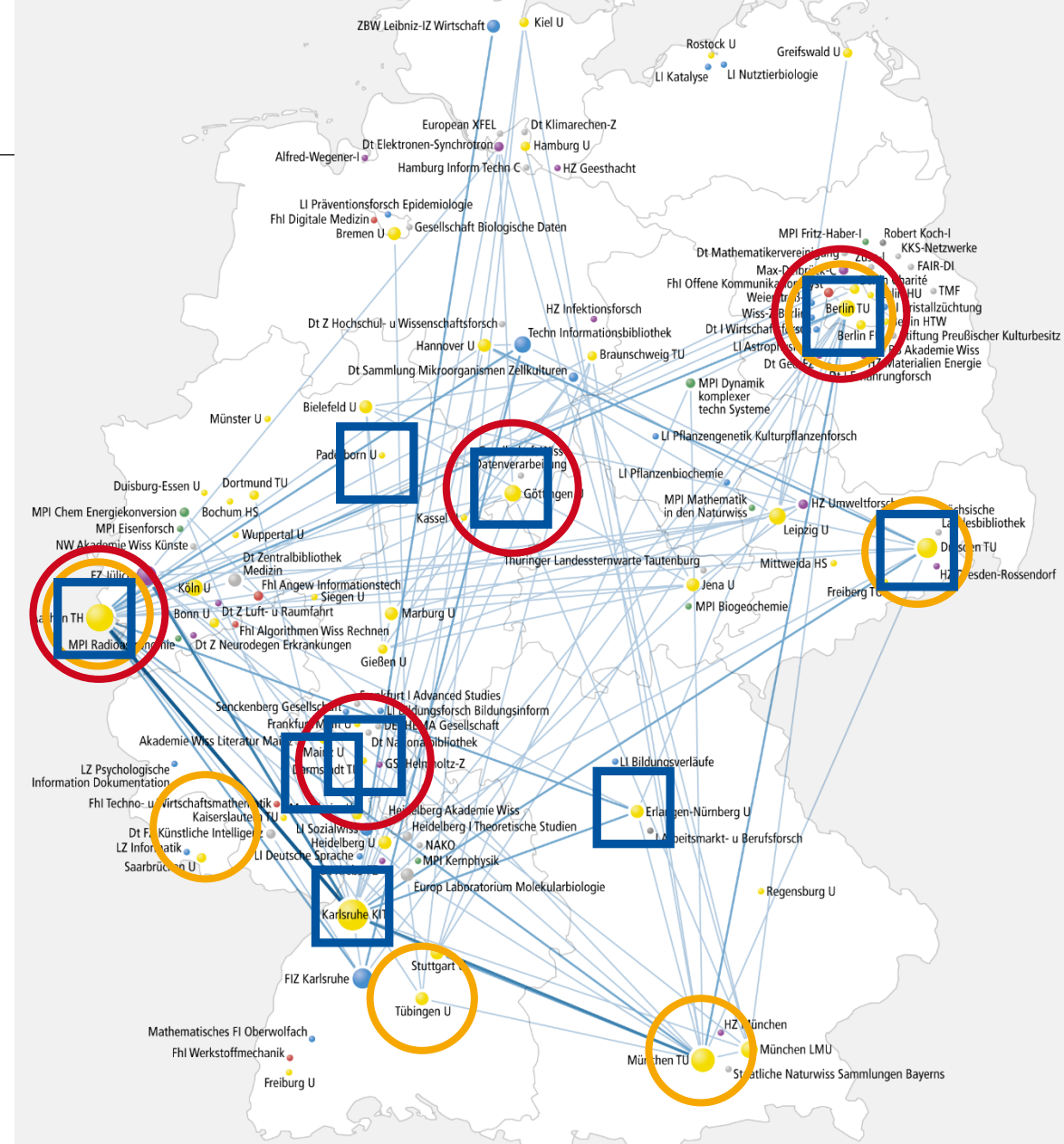
Inhalt

- Gemeinsame Datensicherung als Beispiel
- Allgemeine Betrachtungen
 - Cloud
 - Rolle RZ
 - Prozessintegration
 - IDM, AAI, OVZ
- NRW-Datendienste



Zusammenarbeit: Nationale Strukturen

- Nationale Forschungsdateninitiative
- Nationales Hochleistungsrechnen



IT-Basisdienst: Datensicherung

- Sicher gespeicherte Datenbestände als Voraussetzung für Forschung, Lehre und Verwaltung
- Durch Digitalisierung neue Nutzergruppen
- Vielfalt datenhaltender Systeme bedingt geeignete Verfahren
- Wachsende Datenmengen bedingen adäquate Sicherungs-Infrastruktur



Anforderungen an Datensicherung

- Datenwachstum
- Vielzahl der Speichersysteme
- Notwendige Dienstgüte

➔ Vergleichbare Situation an jeder Hochschule

➔ Kooperation unter den Hochschulen in NRW ➔

➔ Projekt <https://datensicherung.dh.nrw>

NRW:
Hochschullandschaft

14 Universitäten,
16 Fachhochschulen,
7 Kunst- und Musik-
Hochschulen



Überlegungen

- Sortieren der Daten nach Prozess, nicht nach Infrastruktur
 - "Geschäftsdaten"
 - Forschungsdaten
 - Labor-Daten
 - Geschäftsdaten werden von der Verwaltungs-IT verwaltet
 - Forschungsdaten sind ein wichtiger Bestandteil der wissenschaftlichen Arbeit
RDM bietet
 - Wiederverwendung von Daten
 - Wissenschaftliches Qualitätsmanagement
 - Input für spezifisches AI-Training
- verschiedene landesweite Dateninfrastrukturen
- Strategische Entscheidung für den Betrieb vor Ort





Datensicherung: Schnell gemacht?!

- Große Hardware und geeignete Software beschaffen
- Geeignete Konnektivität: DFN
- Los geht's



Daten von Uni A bei Uni B zu sichern ist gut, reicht aber nicht

Entscheidend für nachhaltige Umsetzung:

- Prozessintegration:
essenziell für Automatisierung und Skalierbarkeit
- Notwendige Änderungen auf allen Ebenen





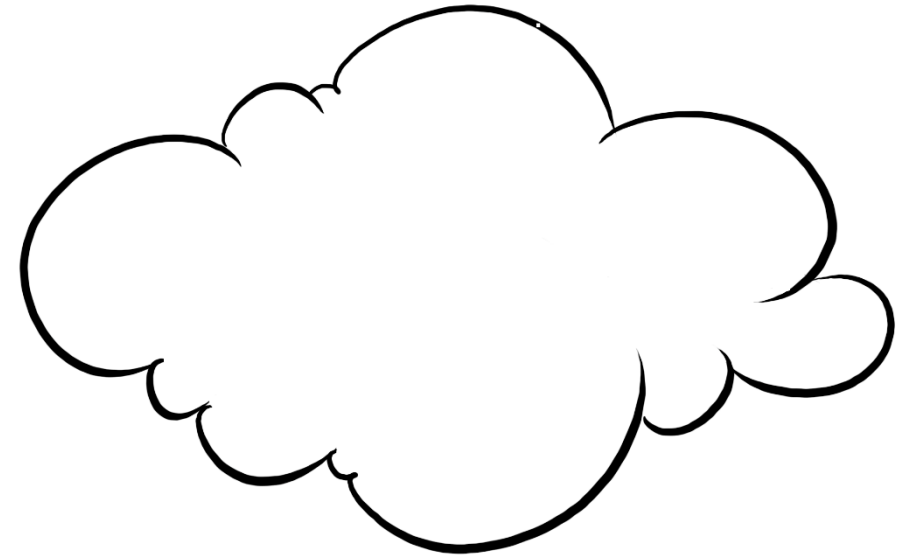
Wer loggt sich da gerade ein?

Wer ist da? → Authentifizierung → Person, Identität

„Was“ ist diese Person?

- Person, die Service in Anspruch nimmt
- Person, die Service aufgrund spezifischer Berechtigung in Anspruch nimmt
 - Z.B. Services nur für Studierende, nur für Mitarbeitende, nur für Mitglieder, ..
- Einrichtung, vertreten durch Person(en)
 - Explizite Autorisierung für eine Rolle in einem organisatorischen Kontext
 - Voraussetzungen für skalierbare Umsetzung
 - Delegierte Rollenverwaltung
 - Organisationsverzeichnis, das alle Nutzer-Einrichtungen umfasst – ansonsten Fallunterscheidung pro Dienst

Perspektivwechsel



Verallgemeinerung

Wie sieht die Situation für uns (eine Hochschul-IT) aus?

- Höhere Nutzeranforderungen, die mit lokaler Infrastruktur und Personal kaum erfüllt werden können
- Notwendig sind hohe Verfügbarkeit der Dienste und einfacher Zugang zu allen Arten von Diensten für die Nutzer (ohne zentrale IT)
- Worin liegt der Beitrag der lokalen IT?

→ Vergleichbare Situation an jeder Hochschule

→ Kooperation und Sourcing



Randbedingungen: Bisher



Viele Annahmen sind möglich

Neue Realität

„Als Nutzer kann ich mich doch überall einloggen.“



Dienstleistungen von „außen“



Ja, und?

„Als Nutzer kann ich mich doch überall einloggen.“

„überall“?

Und: als was?

- Interessen der Nutzerinnen und Nutzer ↔ Interessen der Organisation

Anbieter von IT Services: wo ist „überall“?

Lokale IT

Andere Hochschulen / wiss. Einrichtungen: „Community Cloud“

Externe Anbieter, Cloud



Wie wählen Nutzer einen Dienst aus?

Erreichbarkeit

Bequemlichkeit, Kooperationsfähigkeit

- Individuelle Wahl
- Was Teammitglieder / Kooperationspartner bereits nutzen

Vorgaben

- Von Partnern oder Geldgebern vorgegebene Tools und Dienste

→ Individuelle Ziele

„Überall einloggen“ klingt so einfach

„Public“ Angebote

„Freie“ Zugänge häufig problematisch

- Datenschutz
- Professionelle Nutzung

Externe Bezahl-Zugänge problematisch

- Beauftragung,
- Abrechnung

Beide: Verbindung zur Heimathochschule

- Affiliation (wichtig für wissenschaftliche Metriken und Fördergeber-Reporting)
- Organisatorische Datenhoheit

➔ Ziele der Organisation

(nicht mehr ganz) Neue Aufgaben für IT-Service

- Unterstützung der Nutzenden: „Enabling“ ebenso wie Vorgaben
- Geeignete Rahmenverträge mit externen Anbietern
- Klärung des Rechtsrahmens
- Klärung von Randbedingungen, z.B. Schutzklassen
- Übermittlung der Zugehörigkeit

Und wer erbringt die IT Services?

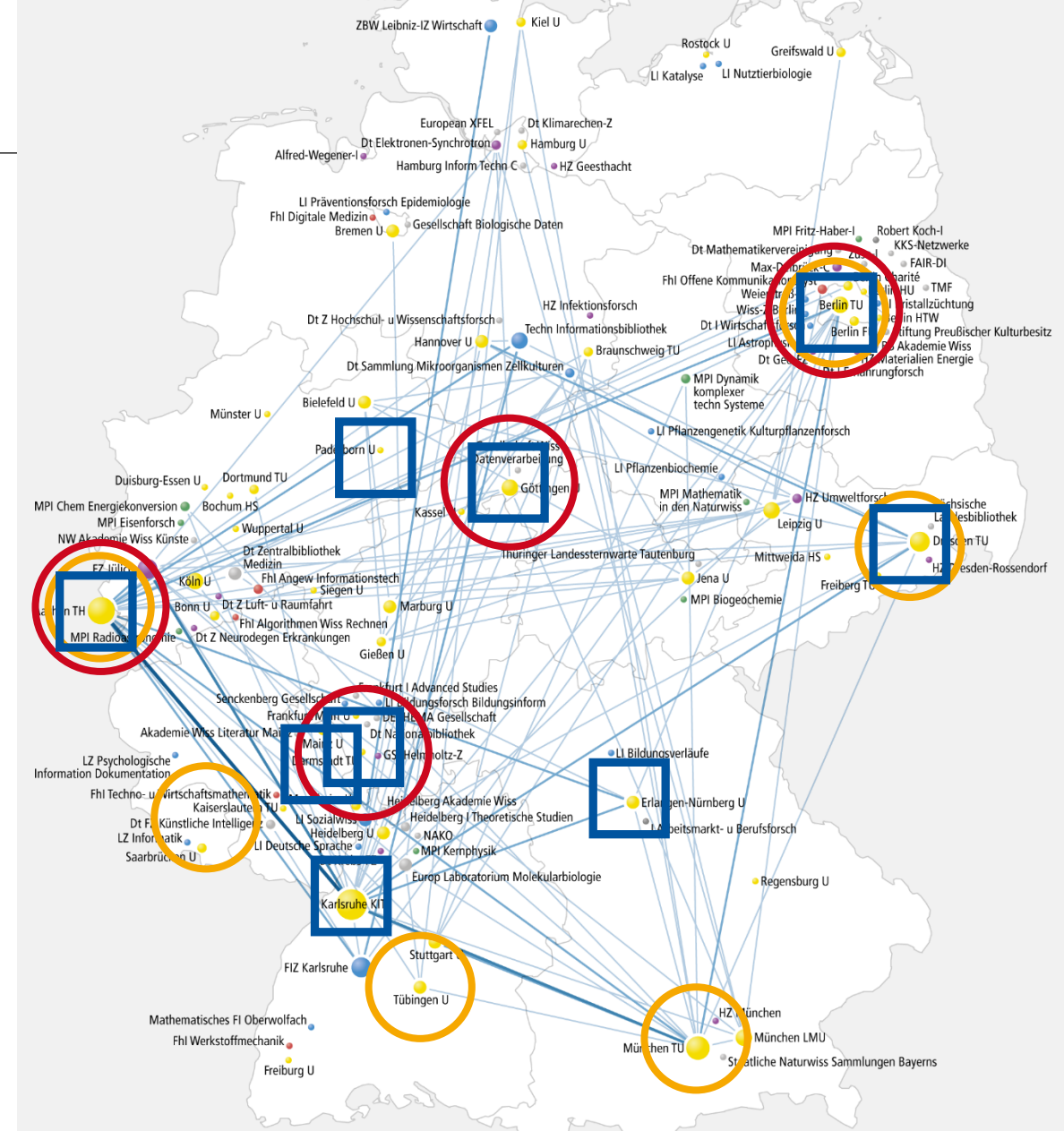
Hochschulen, Forschungszentren,
Dedizierte Organisationen

Schon lange etabliert im Bereich HPC

NHR

NFDI-Infrastruktur-Provider

DFN



IT-Anbieter für die eigene Hochschule

- Bereitstellung von Dienstleistungen für Hochschulangehörige

Cloud-Nutzung

IT-Anbieter für die Gemeinschaft

- Dienstleistungen für andere Hochschulen

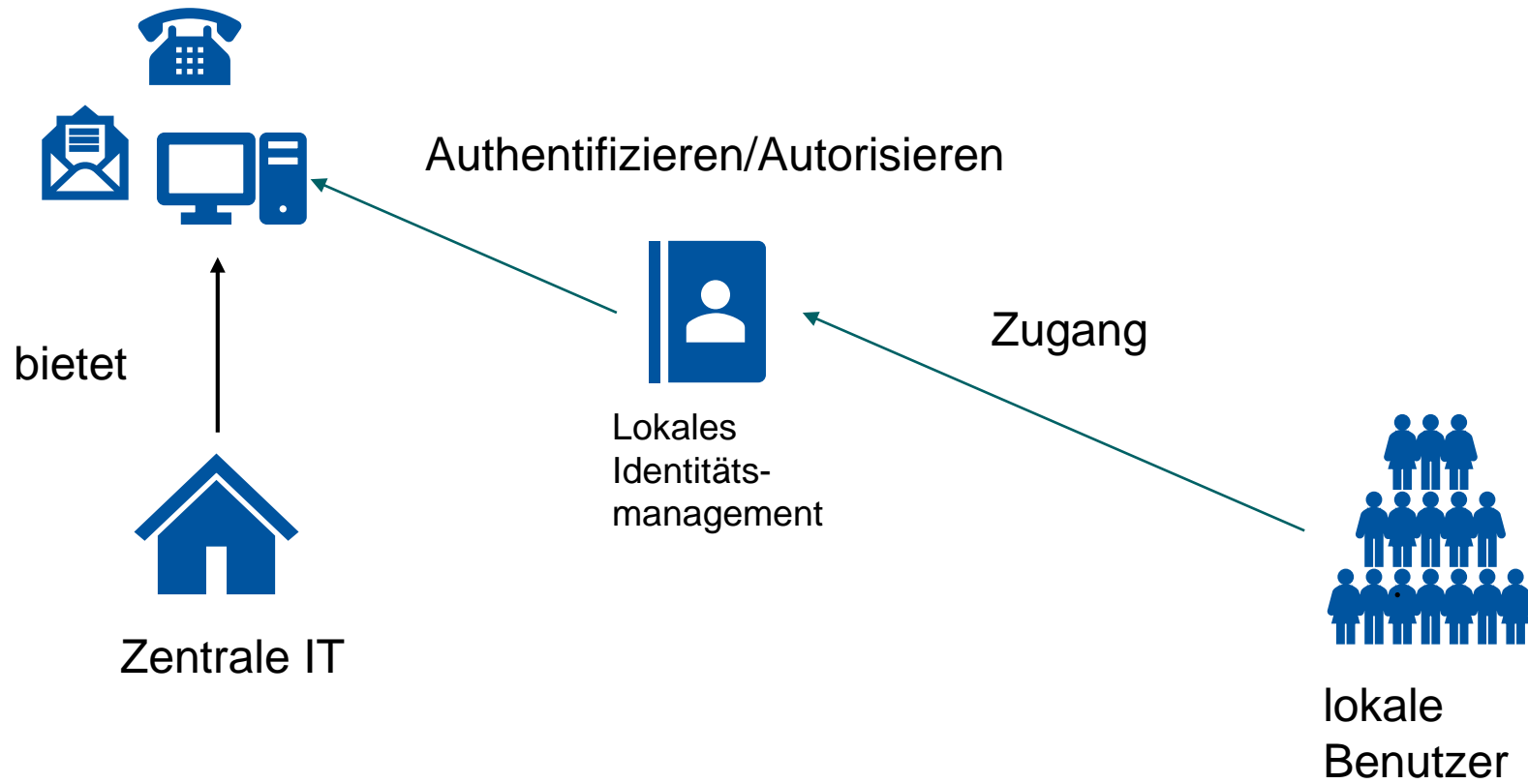
Cloud-Nutzung

Service-Broker für Community- und Public-Cloud-Dienste

- Integration von Cloud-Diensten in universitäre Prozesse
- Der Dienst wird direkt von den Nutzern in Anspruch genommen

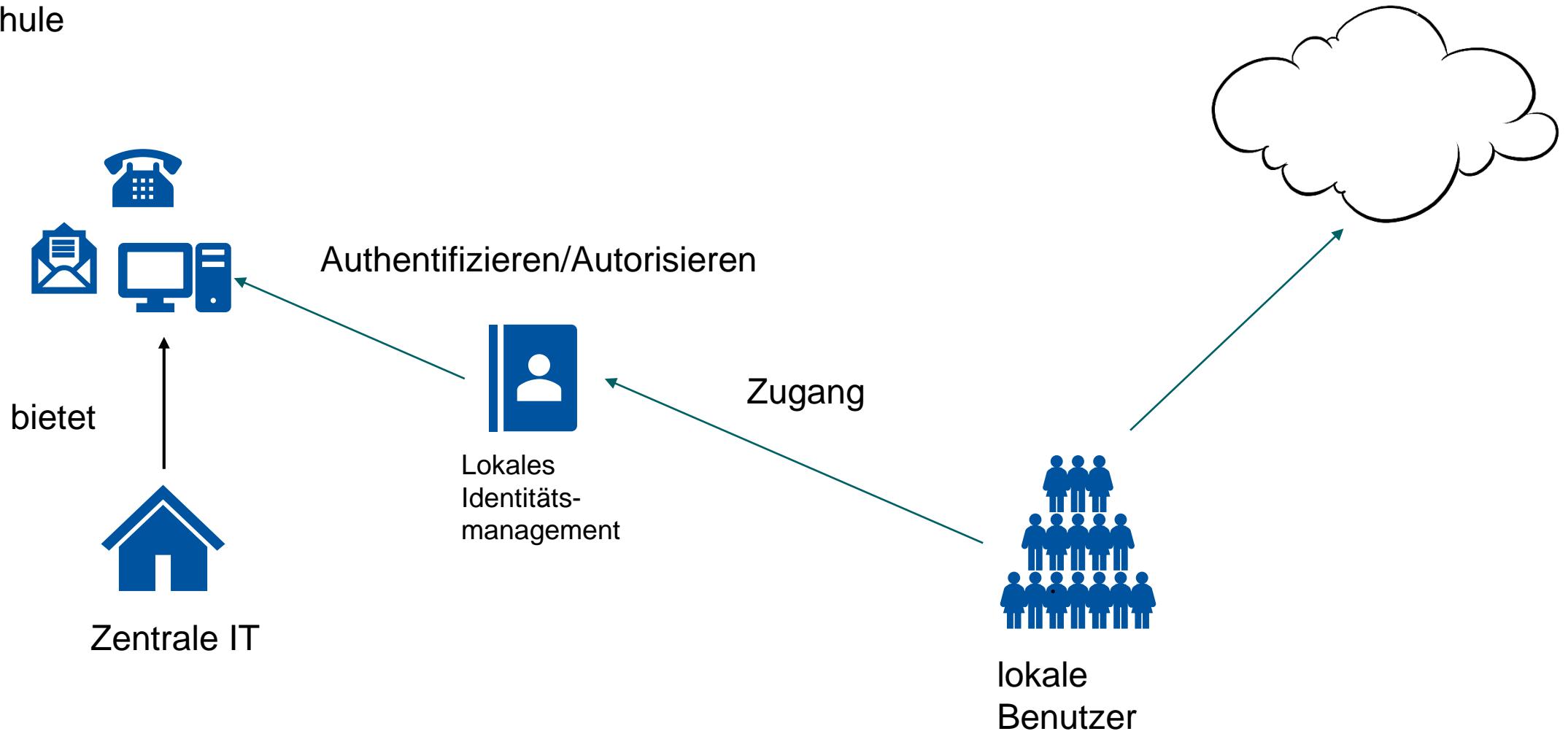
Perspektive Anbieter: Lokale Anbieter

Hochschule



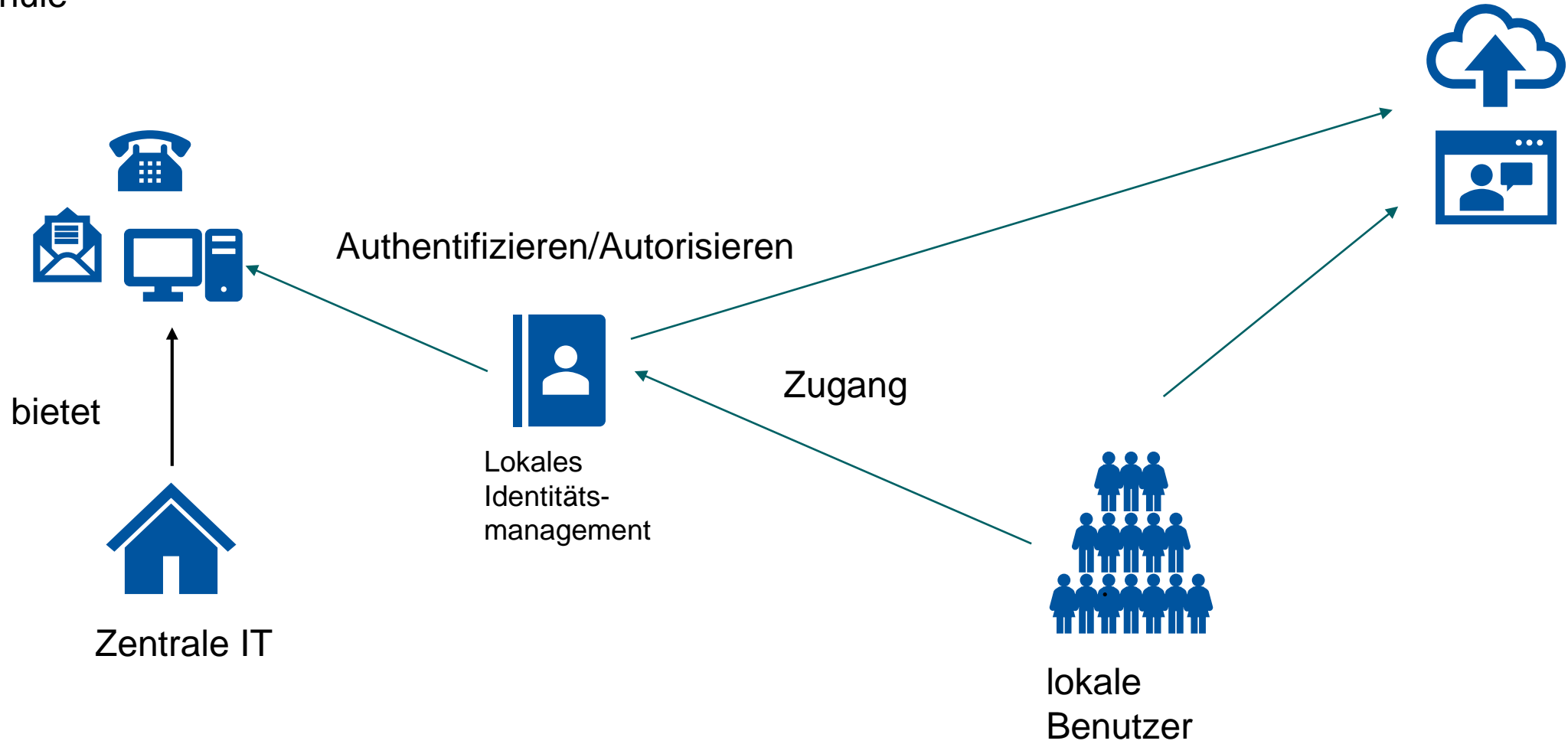
Perspektive Anbieter: Lokale Services und „wilde“ Cloud-Nutzung

Hochschule



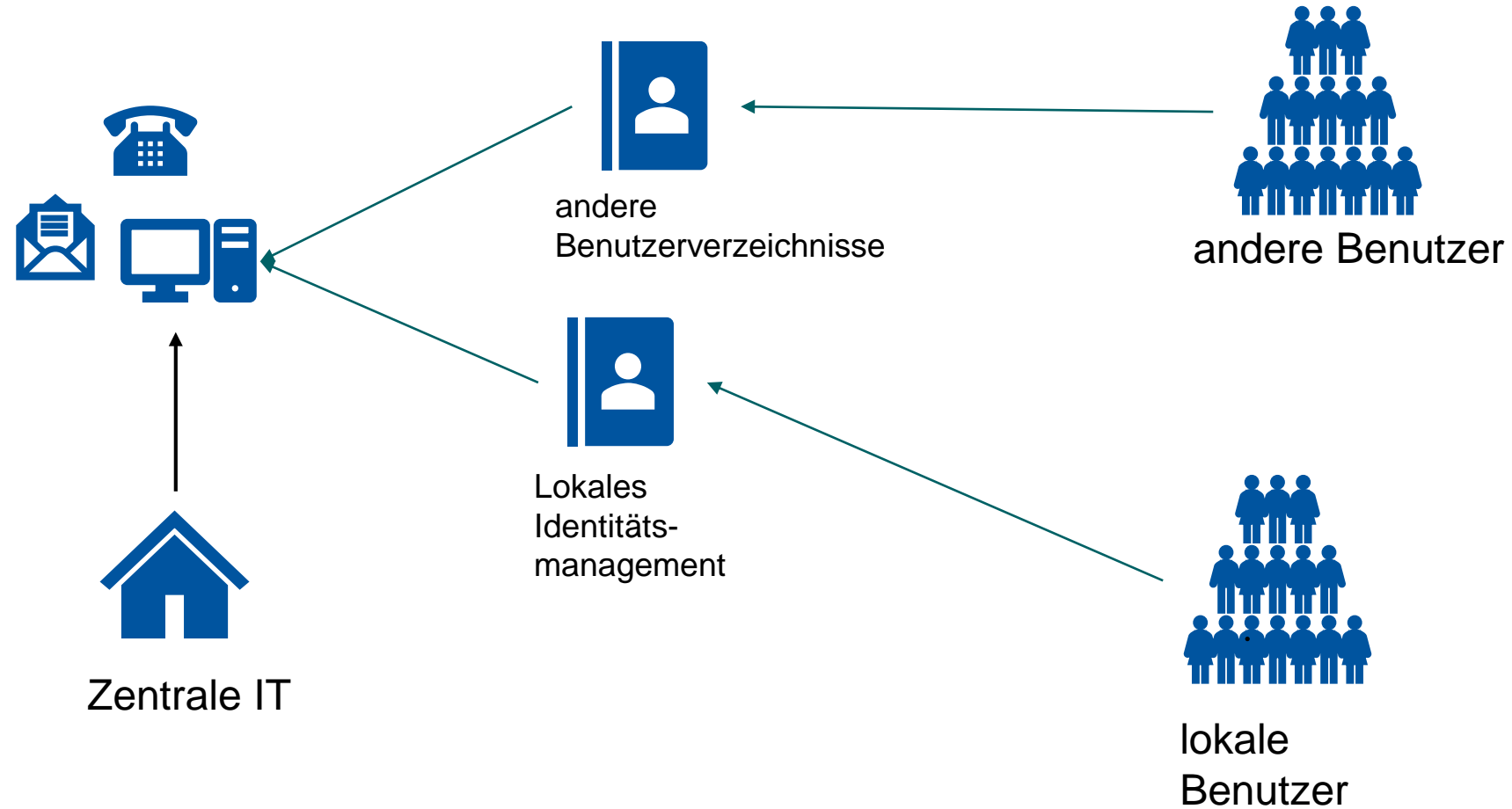
Perspektive Anbieter: Externe Services

Hochschule



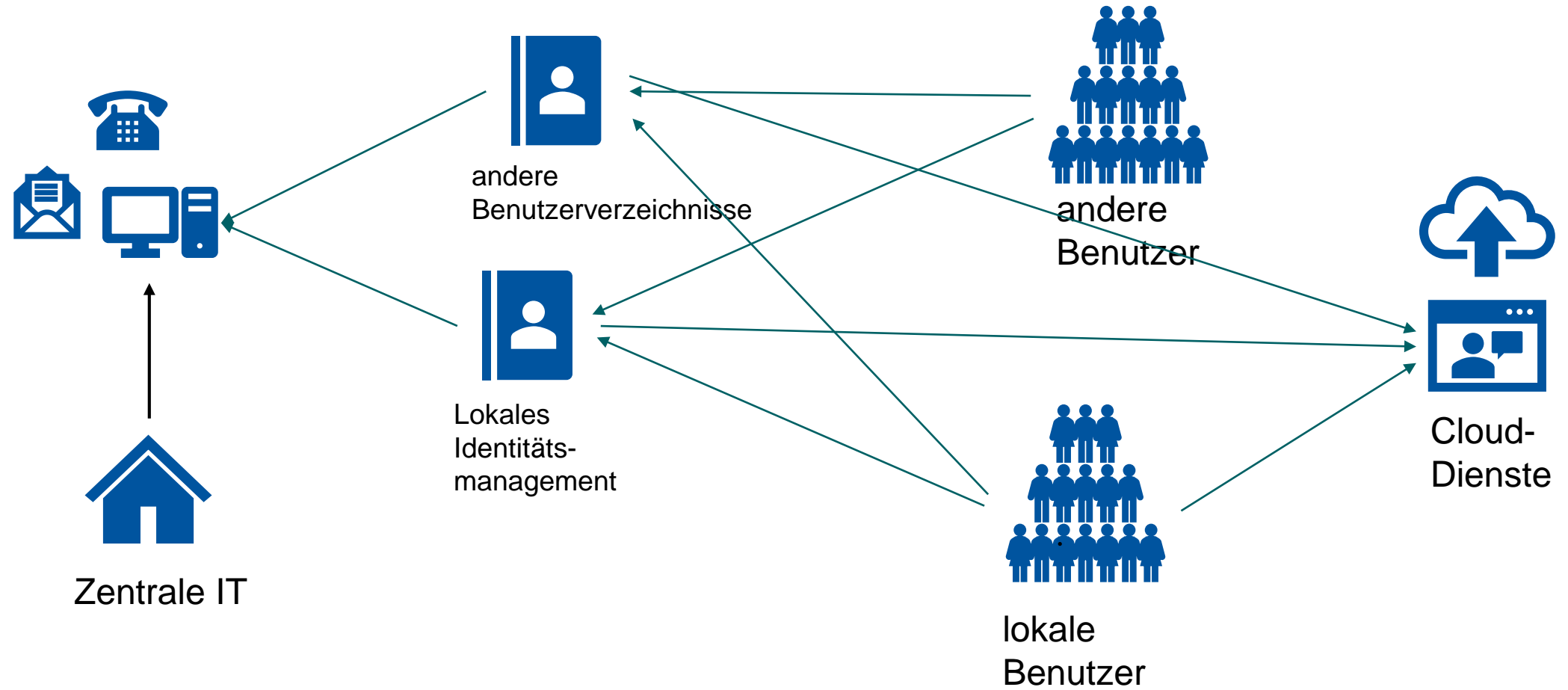
Perspektive Anbieter: Nutzer kommen von externen Organisationen

Hochschule



Perspektive Anbieter: Lokale und externe Nutzer, Lokale und externe Services

Hochschule



Merkmale der Rollen

Kriterien	IT-Anbieter für die eigene Hochschule	IT-Anbieter für die Gemeinschaft	Service-Broker für Community- und Public-Cloud-Dienste
Unterstützung	Hohe Beteiligung	Mittlere Beteiligung (1. Ebene "Heimatinstitution", 2. + 3. Ebene IT-Zentrum)	Geringe Beteiligung (hauptsächlich 1. Ebene)
Möglicher Grad der Individualisierung	hoch	mittel	Niedrig
Betrieb und Instandhaltung	Hohe Beteiligung	Hohe Beteiligung	Geringe Beteiligung (Fokus auf Integration in universitäre Prozesse)
Dienstleistungsentwicklung und Entscheidungsfindung	von der Hochschule kontrolliert	Kontrolliert durch die Gemeinschaft	Kontrolliert durch den Anbieter
Integration in andere Hochschuldienste	hoch	mittel	Niedrig (Normen+ Schnittstellen)

Schlussfolgerung

- Wir (IT) werden zu einem Multi-Sourcing-Ansatz gedrängt
- Rollen helfen, die Veränderungen zu beschreiben
- Die Änderungen gelten für verschiedene Bereiche: IT, Servicebereitstellung und -nutzung, Support, IAM, Entscheidungsfindung, ...
- Änderungen:
 - Konzentration des lokalen Fachwissens auf den lokalen Bedarf und auf die Vermittlung
 - Auslagerung derjenigen Teile der IT, die nicht lokal angepasst werden müssen

Rollen gelten pro Service

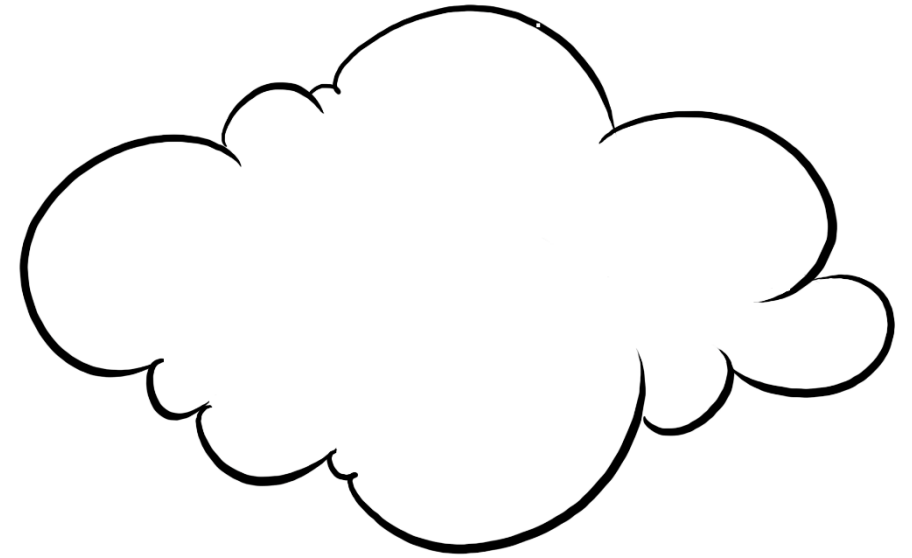
- IT-Organisation kann gleichzeitig mehrere Rollen haben, lokaler und Community-Anbieter und Broker
- Sinnvollerweise nicht in einem Service:
 - Prozesse, Randbedingungen, Annahmen zu unterschiedlich
 - Service müsste unterschiedliche Prozesse bedienen
 - Lokale Nutzerinnen und Nutzer sind gleichzeitig Teil der Community
- (unsere) Lösung:
 - Bei Community-Services ist die Heimathochschule Teil der Community
 - Gleiche Prozesse, Zugänge, .. für Interne und Externe

Ebenfalls notwendig für Skalierbarkeit: Organisationsverzeichnis

- Services für Einrichtungen (meist Basis-IT)
 - Einrichtung: Kontext für Rolle, um Service wahrzunehmen (Netzwerk-Verantwortliche, IT-Besteller, Backup-Admin, ...)
→ Verzeichnis der Einrichtungen
 - Einrichtungen unterschiedlicher Art
 - „normale“ Institute und Fachbereiche
 - An-Institute, „nahestehende“ Institute, Leibniz, Fraunhofer, ..
 - Selbstorganisation: Fachschaften, Gremien
 - Studentische Vereinigungen
 - Institute und Fachbereiche anderer Hochschulen
- Organisationsverzeichnis muss die ganze Vielfalt abbilden
- oder der einzelne Service muss das leisten
 - oder es gibt unterschiedliche Prozesse

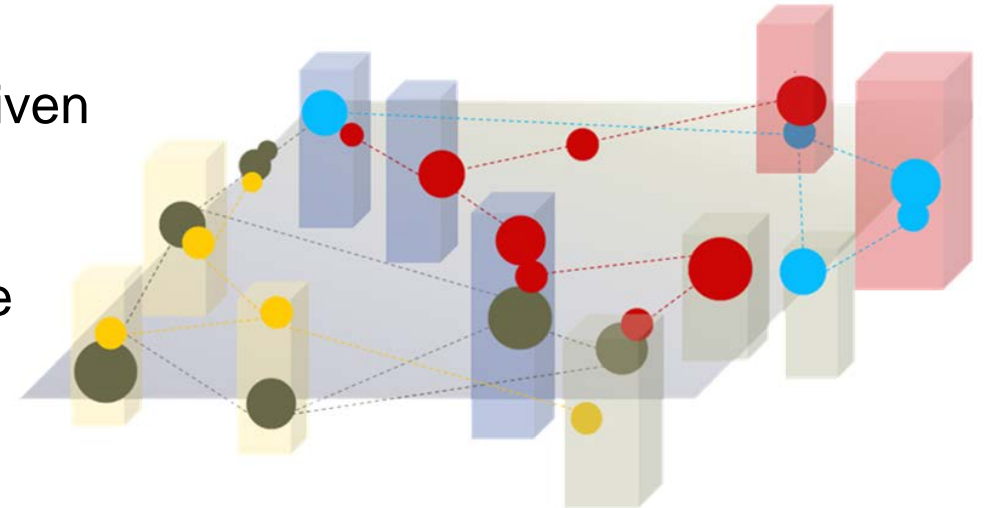
s. DFN Mitteilungen 106

Perspektivwechsel zurück



Zurück zu Datendiensten

- Daten sind eine Voraussetzung für Forschung, Lehre und Verwaltung
- Neue Nutzergruppen und Spezialitäten in der "IT-intensiven Arena" durch Digitalisierung
- Wachsende Datenmengen erfordern eine angemessene Infrastruktur
 - ➔ Zu viel für viele Hochschulen
 - ➔ Zu ineffizient für den Betrieb pro Hochschule
 - ➔ Viele Gründe für den Aufbau einer hochschulübergreifenden Servicelandschaft



Feststellungen

- Kein Kostenvorteil bei Cloud-Nutzung
- Expertise bei den Hochschulen würde trotzdem benötigt

➔ Zusammenarbeit zur kooperativen Selbstversorgung

- Bündeln von Kompetenzen
- Entlastung
- Skalierbarkeit

- Lösungsansatz:
 - Wenige Service Provider
 - (Deutlich mehr) Experten für Sicherungsszenarien





Wo wir gerade dabei sind ..

- Sowieso eine massive Änderung des Service → **Wie** sollte er werden?
 - Umfrage¹ unter Hochschulen, deren IT and Forschungsabteilungen wichtigste Forderungen:
 - Dezentraler Selfservice für Client-bezogene Aufgaben ((de-)-registration, backup, restore, ..)
 - Unterschiedliche Ebenen (Hochschul-IT ↔ Fachbereichs-/Einrichtungs-IT)
 - Lokaler Support
- Änderung für einige Hochschul-ITs

1) EUNIS 2021, <https://doi.org/10.29007/t9bc>

Datensicherung: Besonderheit an Hochschulen

- Zentrale IT für Verwaltung und Campus-Management
- Im wissenschaftlichen Bereich sehr dezentrale IT „by design“:
 - Team-spezifische und individuelle Software
 - Laborsysteme mit Ankopplung an Experimente und Prüfstände u.v.a.m.

Beispiel RWTH:

- ~250 dezentrale IT-Teams
- ~90% des Datenvolumens sind dezentral und „unstrukturiert“
- ~80% der Daten existieren in nur einer Version

➔ Sicherung dieser Datenbestände über eine Backup-Software auf zentrale Infrastruktur



Selfservice

- Nur Betreiber von datenhaltenden Systemen kennt Bedarf
 - Neue Clients anlegen
 - Sicherungspolicy
 - Backup / Restore
 - Clients abmelden
- ➔ Nur Realisierung als Selfservice sorgt für bedarfsgerechten Service **und** Skalierbarkeit

- Zuordnung:
 - Clients ⇔ Organisationseinheit
 - Admin (Rolle!) ⇔ Organisationseinheit

 - Clients ⇔ Admin
 - ➔
 - Schwierigkeiten bei Vertretung / Übergabe
 - Keine Möglichkeit bei „shared Admins“



Wer loggt sich da gerade ein? (hatten wir schon)

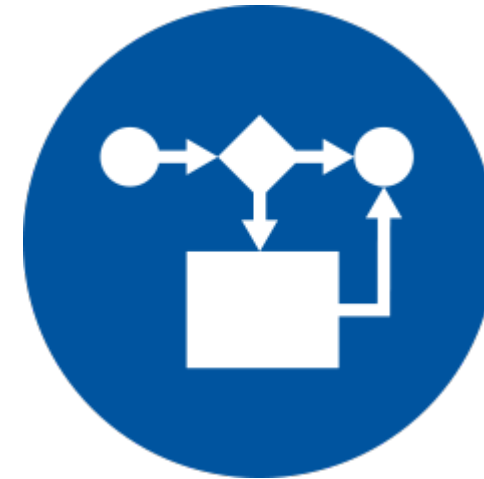
Wer ist da? → Authentifizierung → Person, Identität

„Was“ ist diese Person?

- Person, die Service in Anspruch nimmt
- Person, die Service aufgrund spezifischer Berechtigung in Anspruch nimmt
 - Z.B. Services nur für Studierende, nur für Mitarbeitende, nur für Mitglieder, ..
- Einrichtung, vertreten durch Person(en)
 - Explizite Autorisierung für eine Rolle in einem organisatorischen Kontext
 - Voraussetzungen für skalierbare Umsetzung
 - Delegierte Rollenverwaltung
 - Organisationsverzeichnis, das alle Nutzer-Einrichtungen umfasst – ansonsten Fallunterscheidung pro Dienst

Zuordnung

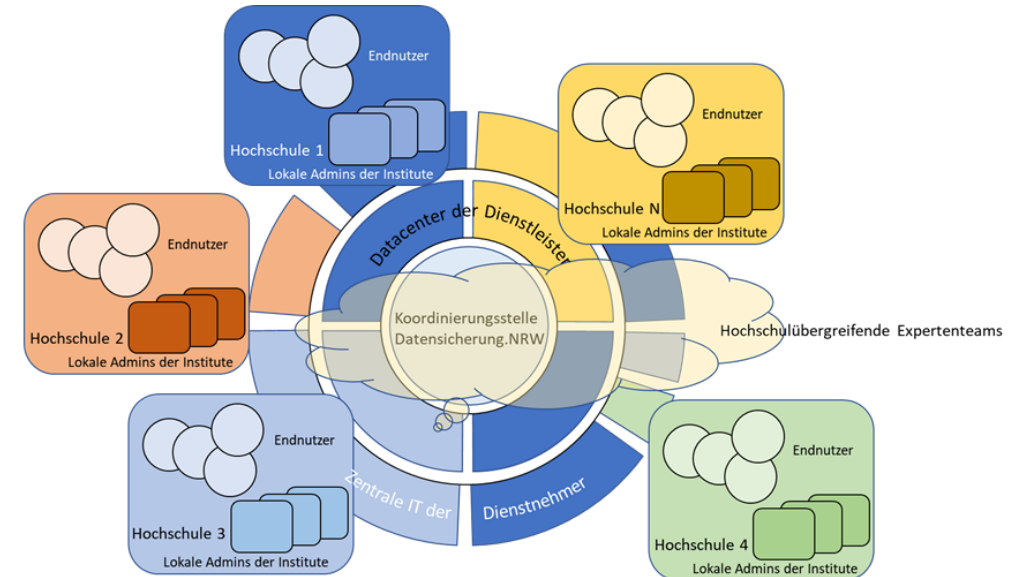
- Admin im Kontext einer Organisationseinheit einer Hochschule
 - Verantwortlich für dezentrale IT
- Verantwortung liegt bei Leitung der Einheit
 - Leitung ist für eigene Forschung und Lehre verantwortlich
 - Benennt lokale IT-Verantwortliche
 - Zuweisung der Rolle wiederum als Selfservice
- Delegation an Einheit



Realisierung

→ Rollenkonzept zur Abbildung der Zuständigkeiten

- Basierend auf Identitäten
 - Gepflegt im lokalen IdM der einzelnen Hochschulen
 - Ermöglicht Authentifizierung über Hochschulgrenzen
- Gemeinsame Definition von
 - Strukturierten Kontext-Informationen
 - Rollenbezeichnung
- Transport vom Verzeichnisdienst der Heimat-Hochschule zum zentral bereitgestellten Web-Portal



Technische Umsetzung

- Zwei Dienstleister: Universität Duisburg-Essen und RWTH Aachen
 - Objektspeicher mit lokaler Redundanz
 - Commvault-Software
 - BSI: Sicherungsdaten jedem(!) möglichen Zugriff entziehen
 - Dedizierte Speicherplattform bei UzK mit Retention lock
 - Leistungsfähige Verbindungen: DFN: UDE, UzK, RWTH: 100 Gbit/s
- ➔ 2. Kopie nur für ausgewählte Datenbestände
- ➔ Alle Hochschulen müssen betriebskritische Prozesse und darunterliegende IT identifizieren



Also ..

Projekt Datensicherung.nrw

- Bedeutet große Umstellung
 - Betrieb: neue Software, neue Infrastruktur
 - Betriebsmodell: lokal → hochschulübergreifend
 - Alle Ebenen betroffen
- Bedeutet Chance
 - Versorgung aller NRW-Hochschulen mit leistungsfähiger Datensicherung
 - Änderung mit Pilot-Charakter

Forschungsdaten-Infrastruktur

- Schlüsselbestandteil einer landesweiten Strategie + landesweite Verbindung
 - Kombination von Software, Fähigkeiten, Ausbildung, Beratung, ...
 - Verknüpfung mit anderen wissenschaftlichen Ressourcen

Ziel: Umfassende Infrastruktur für Forschung und Lehre

Anforderungen/Bausteine

Transparenter Workflow der Ressourcenanwendung

Hochschulübergreifende AAI

Software-Entwicklung

HPC

AI/ML

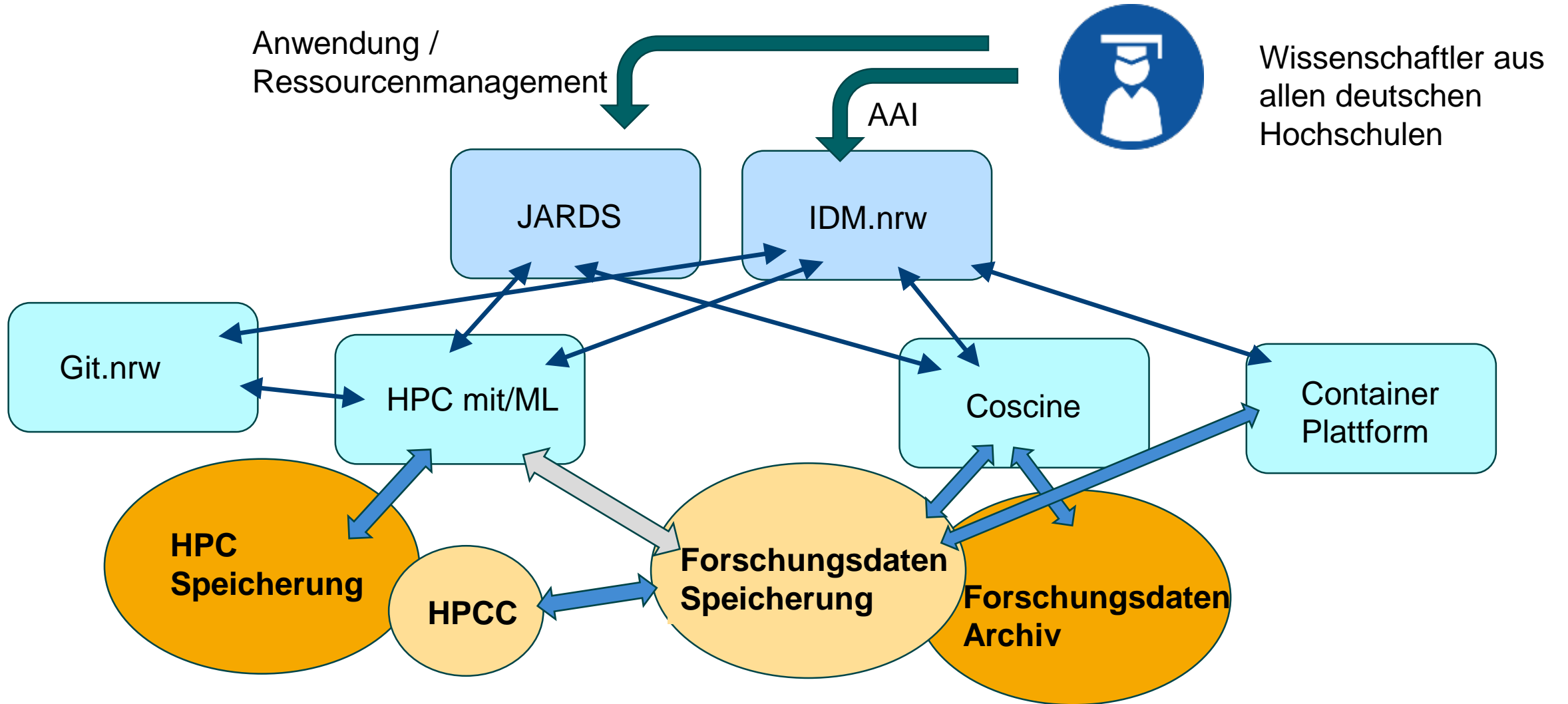
Cloud-ähnliches "Low Performance"-Rechnen

Verwaltung von Forschungsdaten, kompletter Lebenszyklus

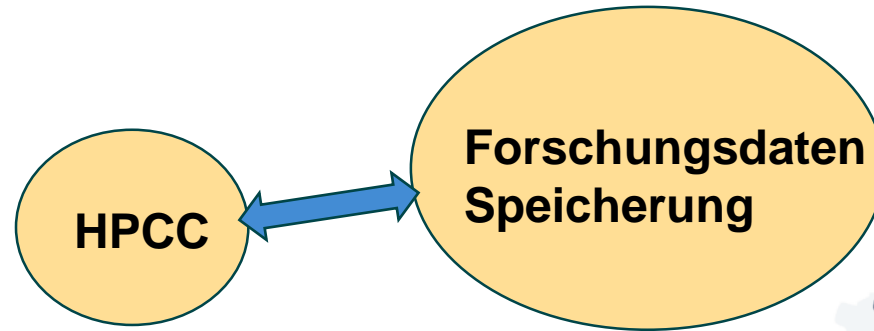
- Erzeugung (Experimente, Simulation, ..)
- Analyse der Daten
- Veröffentlichung von Daten
- Langfristige Verfügbarkeit/Zugänglichkeit



Umfassende Infrastruktur für Forschung und Lehre



Speicherung von Forschungsdaten



Objektspeicher, verteilt auf 4 Hochschulen

Große Entfernungen (Aachen ⇔ Paderborn: >200 km)

Einzelner (redundanter) Servicepunkt mit Coscine-Software



Schlussfolgerung

Eigene Daten sind das Lebenselixier für alle Bereiche einer Hochschule

Geeignete Infrastruktur muss sein

- Zuverlässig
- Sicher
- Skalierbar
- Schnell

"Infrastruktur" ist das gesamte System, das die Kernprozesse unterstützt:
Hardware, Software, Menschen

Unvermeidlich: Betrieb „on premise“ (bei einer Hochschule, nicht unbedingt die eigene)

Nicht effizient als lokale Lösung → kooperative Dienste

Landesweite Datendienste in NRW (Auswahl)

- Datensicherung
 - Bedient Geschäftsdaten, Labor-IT, ...
 - kooperativer Dienst: 2 Universitäten (Duisburg-Essen + RWTH Aachen) für (fast) alle anderen Hochschulen
 - Schutz für alle universitären Systeme, insbesondere für dezentralisierte (meist wissenschaftliche) IT
 - 2. Kopie an einer anderen Universität (Köln) für geschäftskritische Daten

- Forschungsdaten-Infrastruktur
 - Geografisch verteilte Objektspeicherung
 - Groß genug
 - Schnell genug
 - Kombiniert mit eigenem Software-Stack Coscine

<https://datensicherung.dh.nrw/>

<https://about.coscine.de/en/>

➔ Beispiele für produktiv umgesetzte hochschulübergreifende Services

Rückschluss auf kooperative IT

Annahme: Es ist sinnlos, mit kommerziellen Angeboten zu konkurrieren

Es sei denn, es liegt ein erheblicher Mehrwert vor

- Muss die IT konkurrieren?
- Oder gibt es Alleinstellungsmerkmale?
- Oder kann der Wert der verfügbaren Dienstleistung erhöht werden?

Multi-Cloud-Umgebungen

Annahme: Wir leben bereits in einer Multi-Cloud-Umgebung

Nicht alles und jeder (Cloud-)Dienst muss verwaltet werden, aber

- Wir müssen Prozesse (neu) definieren, die für jeden Dienst gelten
 - wie Abrechnungs- und Buchführungsvorschriften
 - Unterstützung
- Wir brauchen Sicherheitsrichtlinien, z.B. Unterstützung bei der Datenklassifizierung

das Bewusstsein der Nutzer stärken – es geht nicht „transparent“



Änderungen in der zentralen IT

Stärkung der technischen Kernkomponenten

- Zentrale AAI einschließlich Gruppenmanagement
- Sicherheit ("Zero Trust", Firewalling in der Cloud)
- Zusätzlicher Nutzen für lokale Dienstleistungen

Die Komponenten müssen Standards folgen, um den Anforderungen von "außerhalb der Hochschule" zu entsprechen.

Bedarf der Hochschule an Beratung

- Welcher Dienst für ihr Szenario zu verwenden ist
- wie sie Dienste aus verschiedenen Quellen in ihren Arbeitsablauf integrieren können

Unterstützungsprozesse anpassen

- Derzeitiger Zustand nicht skalierbar für die Anzahl der angestrebten Dienste
- Mehr Interaktion/Integration mit externen Anbietern
- Unterscheidung zwischen "lokal bereitgestellten Diensten" und "verwalteten Cloud-Diensten".

Was bringt die Zukunft?

Mehr organisatorische als technische Änderungen

nicht alle lokalen Dienste müssen lokal gehalten werden

Welche Fähigkeiten brauchen wir?

- Kernkomponenten: hochspezialisiertes technisches Wissen
- Beratung: breites IT-Wissen
- konzeptionelle Arbeit, Festlegung von Organisationsstrukturen, Weichenstellung

Fragen?
Anmerkungen?
