

UNEKE –

*Vom **USB-Stick** zur **NFDI** – Entwicklung eines **Kriterien-geleiteten Entscheidungsmodells** für den Aufbau von **Forschungsdaten-Infrastrukturen***

FDMentor Workshop

29. Mai 2018

Dr. Stephanie Rehwald



UNEKE - Projektziele



Projekt UNEKE

UNIVERSITÄT
DUISBURG
ESSEN

Offen im Denken

Bela Brenger

brenger@itc.rwth-aachen.de



RWTHAACHEN
UNIVERSITY

Dr. Stephanie Rehwald

stephanie.rehwald@uni-due.de

ub

universitäts
bibliothek

UNIVERSITÄT
DUISBURG
ESSEN

Offen im Denken

Dr. Ania López

ania.lopez@uni-due.de

ub

universitäts
bibliothek

UNIVERSITÄT
DUISBURG
ESSEN

Offen im Denken

Konstantin Wilms

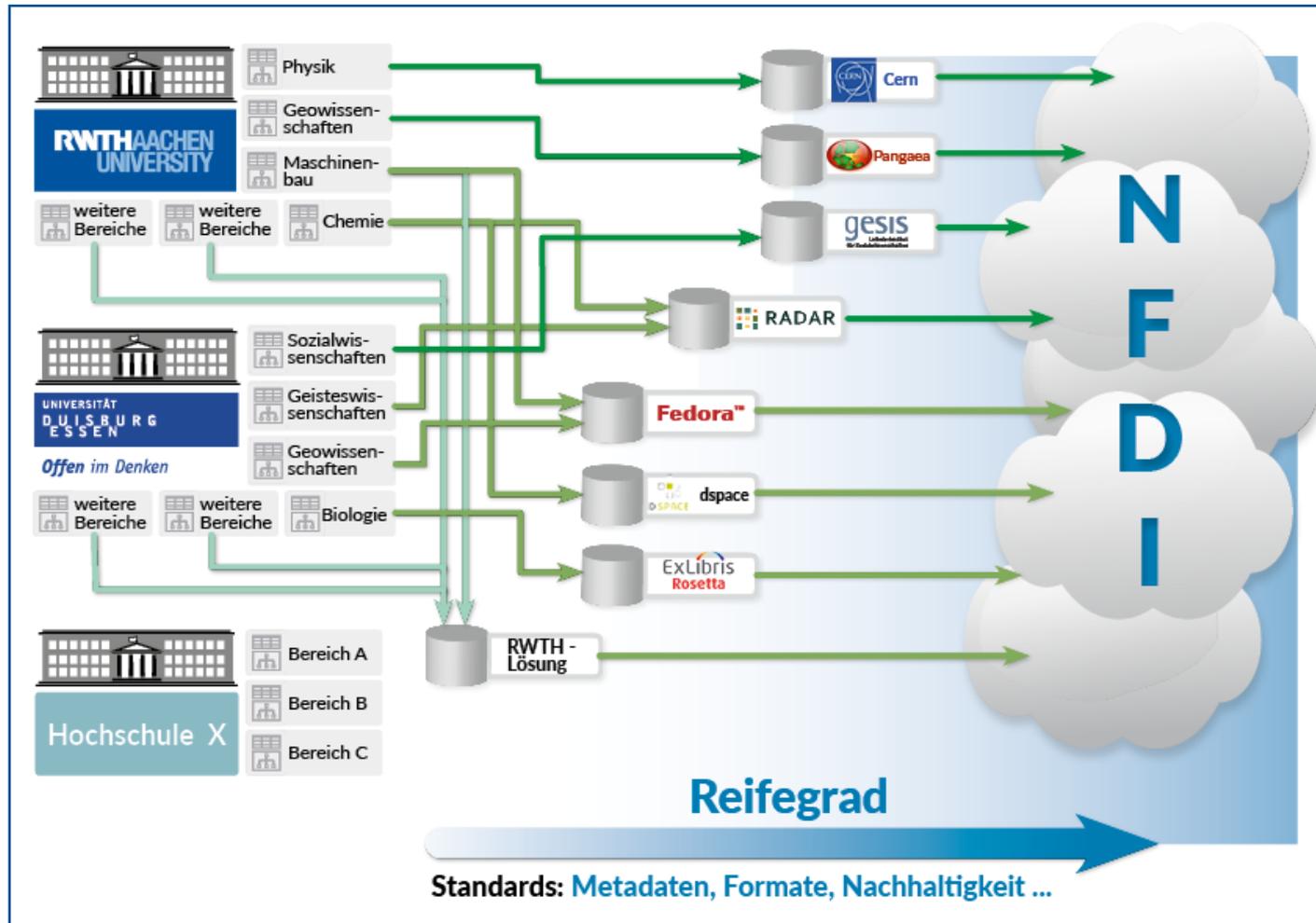
konstantin.wilms@uni-due.de



UNIVERSITÄT
DUISBURG
ESSEN

Offen im Denken

Langfristige Perspektive: Nationale Forschungsdateninfrastruktur (NFDI)



Leitfragen:

- Wie und wo werden Forschungsdaten gespeichert?
- Wie sollen Hochschulen Speicherinfrastrukturen zur Verfügung stellen? (z. B. Eigenentwicklung, Betrieb von Softwarelösungen, verschiedene Hardwaremodelle, fachliche und generische Repositorien, etc.)

→ Hochschulen brauchen Entscheidungskriterien!



Ziel des Projekts UNEKE:

Entwicklung von objektiven Kriterien für Art und Umfang des Aufbaus von Infrastrukturen. Bereitstellung eines Analysewerkzeugs und einer übertragbaren Roadmap.

2.1 Ziele und Arbeitspakete

Werkzeuge zur
Bedarfsanalyse

Bewertung von
Repository-
Lösungen

Roadmap zur
Speicher-
infrastruktur

Workshops
Oktober/November

Fokus: Ingenieure
März - Juli

Gesamtkonzept &
Umsetzung
Juli – Jan`19

Umfragen
März- Mai +X

Fokus: Long-tail
März - Juli

Veröffentlichung &
Vernetzung
Feb. – April `19

Use Cases
März- Juli

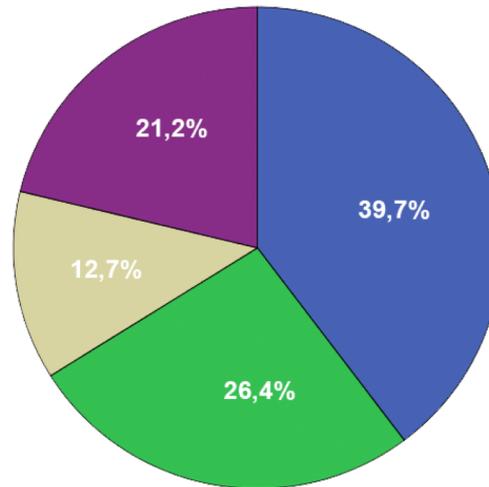
Bedarfserhebungen



Fachspezifische
Workshops

Online Umfrage

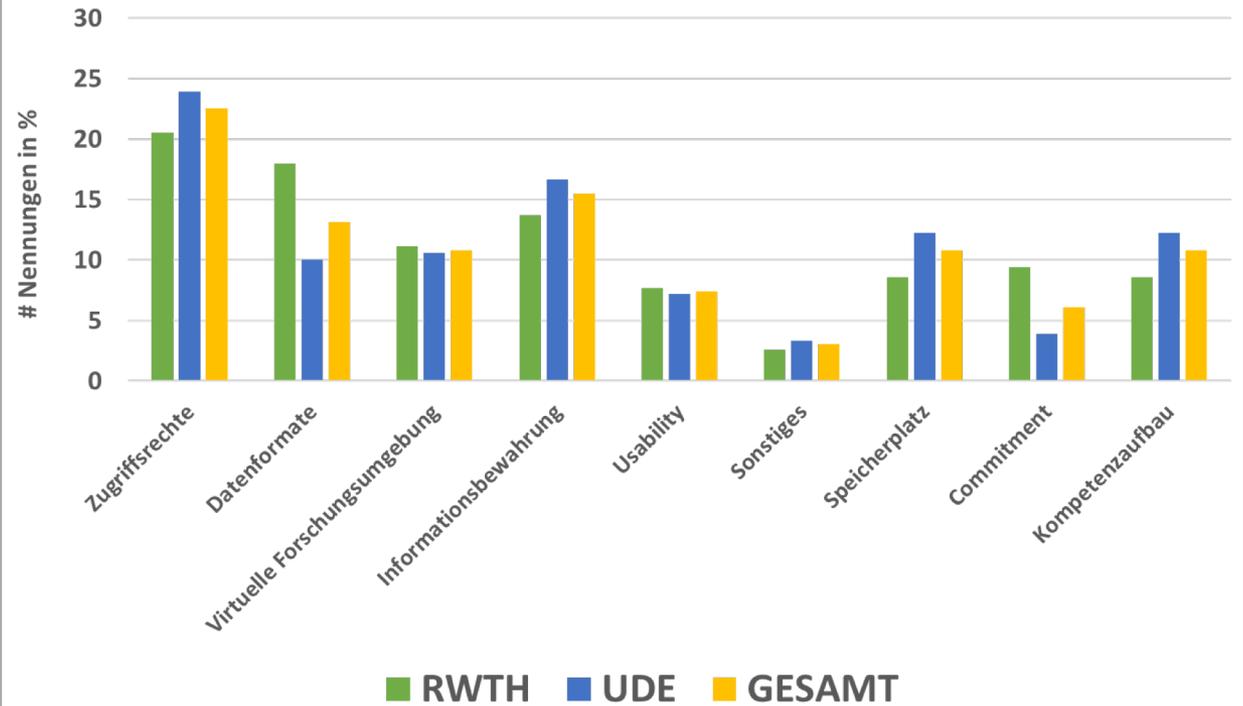
Interviews



Fachspezifische Workshops



Workshops an UDE und RWTH Bedarfe des Forschungsdatenmanagements

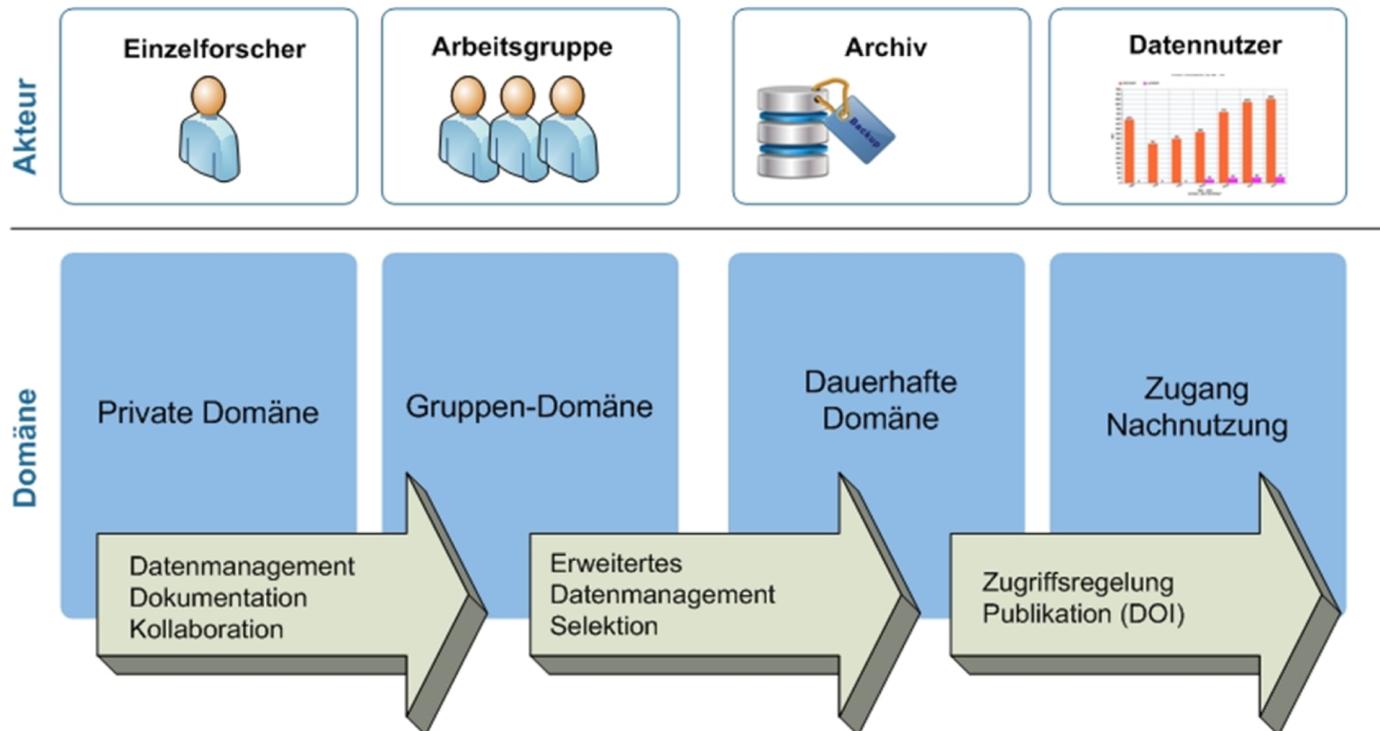


UNEKE Umfrage



Orientierung am Domänenmodell

- Aus der Perspektive der Speicherinfrastruktur
- Speicherbedarf und Verlauf während eines typischen Forschungsprojekts



Teil 1: Deskriptive Datenerhebung

- „Wie viel Speicher benötigen Sie im Durchschnitt?“
- „Welche Kriterien müsste eine Infrastruktur erfüllen um Ihren Anforderungen gerecht zu werden?“

Teil 2: Wissenschaftlicher Teil

- Abfrage evaluierter Konstrukte
- **Fokus:** Teilen von Informationen

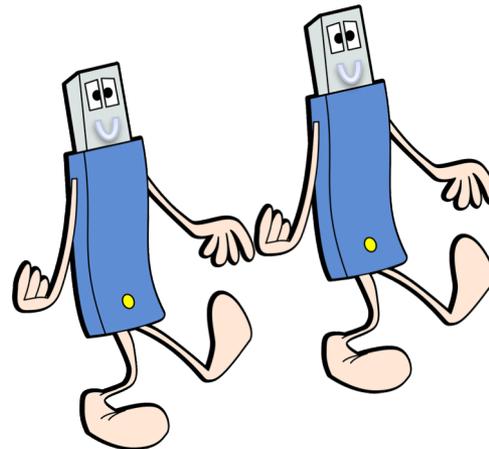
Link zur Testversion:

<https://www.soscisurvey.de/rdmHIL/?act=S432yQY3dOca2wDoCuxx1aiF>

- Fachhochschule Aachen
- RWTH Aachen
- Fachhochschule Bielefeld
- Ruhr-Universität Bochum
- Universität Bonn
- Universität Duisburg-Essen
- HHU Düsseldorf
- Universität Siegen
- BU Wuppertal

Aktuell
**9 Hochschulen
in NRW**

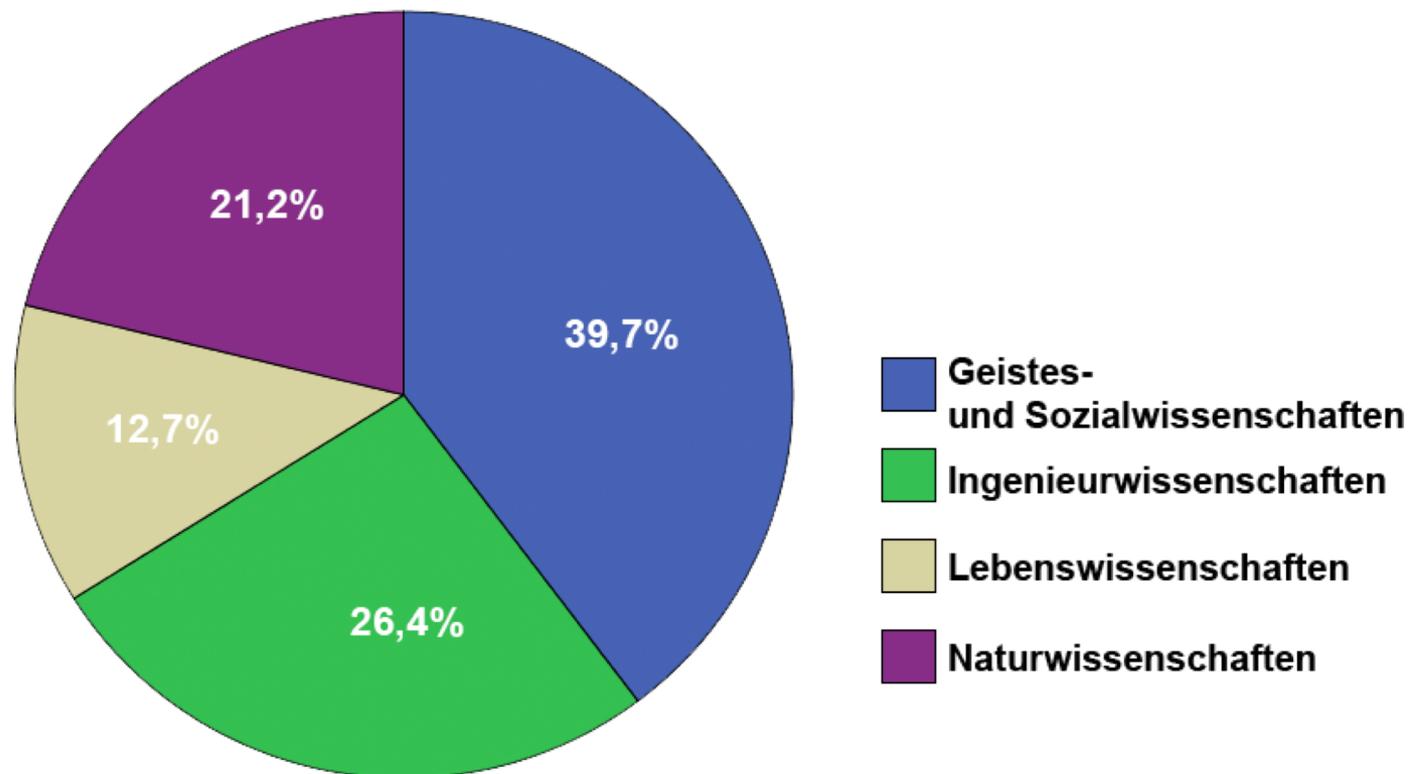
+ 4 weitere HS
deutschlandweit
ab Sommer



UNEKE Umfrage – ein erster Blick in die Daten

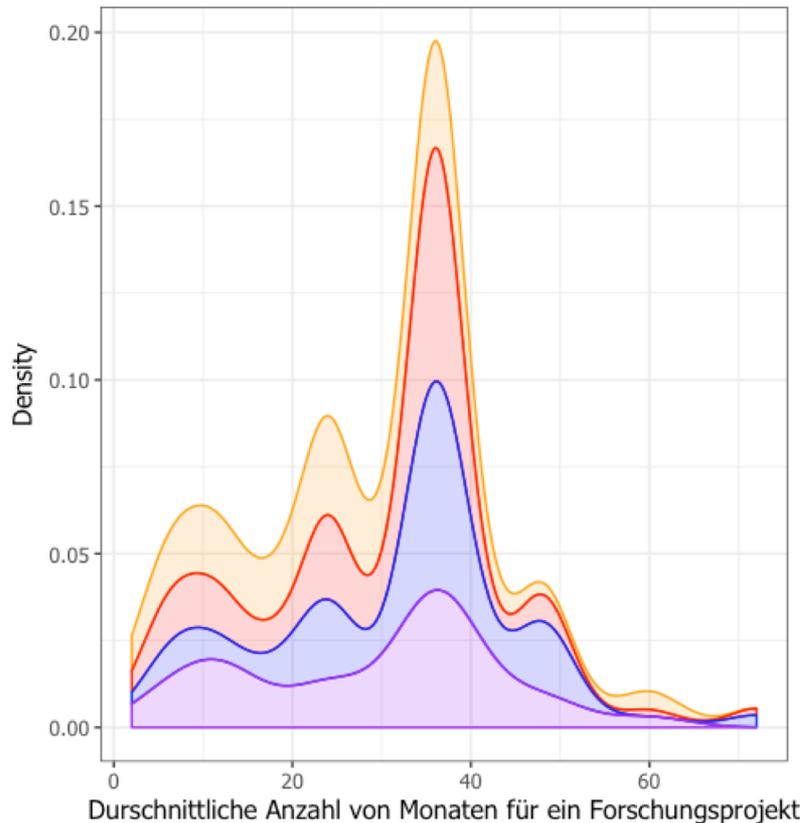


Teilnehmerzahl nach einer Woche: 292 (UDE:169; RUB:136)



Wie hoch schätzen Sie die Dauer und genrierte Datenmenge eines typischen Forschungsprojektes ein, an dem Sie arbeiten?

Projektdauer



Typisches Projekt dauert 3 Jahre und produziert 20 GB Daten

Generierte Datenmenge

Min: 10 MB

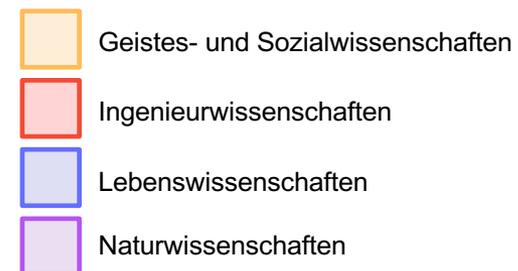
Max: 250 TB

Mean: 1664 GB*

Median: 20 GB

(von 10 GB (GE) bis 40 GB (ING))

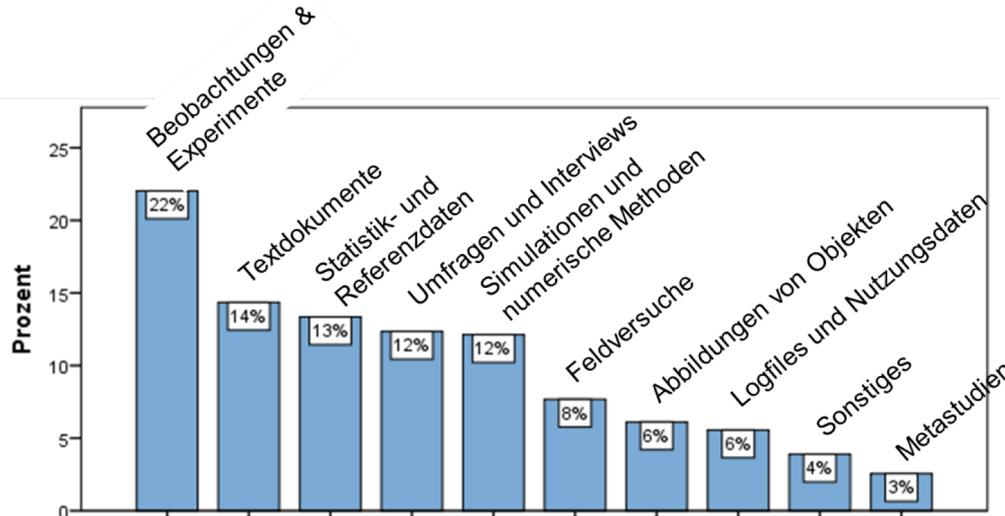
*verzerrt durch das Maximum



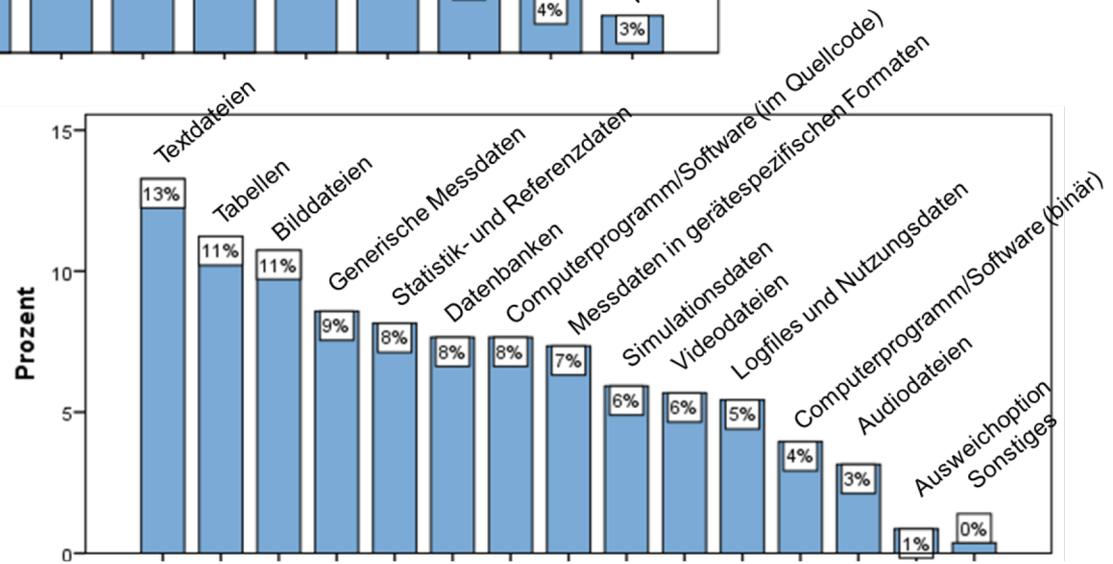
Datenquellen und Datenformate:

Auf welche Art und Weise bzw. aus welchen **Quellen** gewinnen Sie Ihre Forschungsdaten? Welche Datenformate fallen dabei an?

Vielfalt:
Quellen
von
Daten

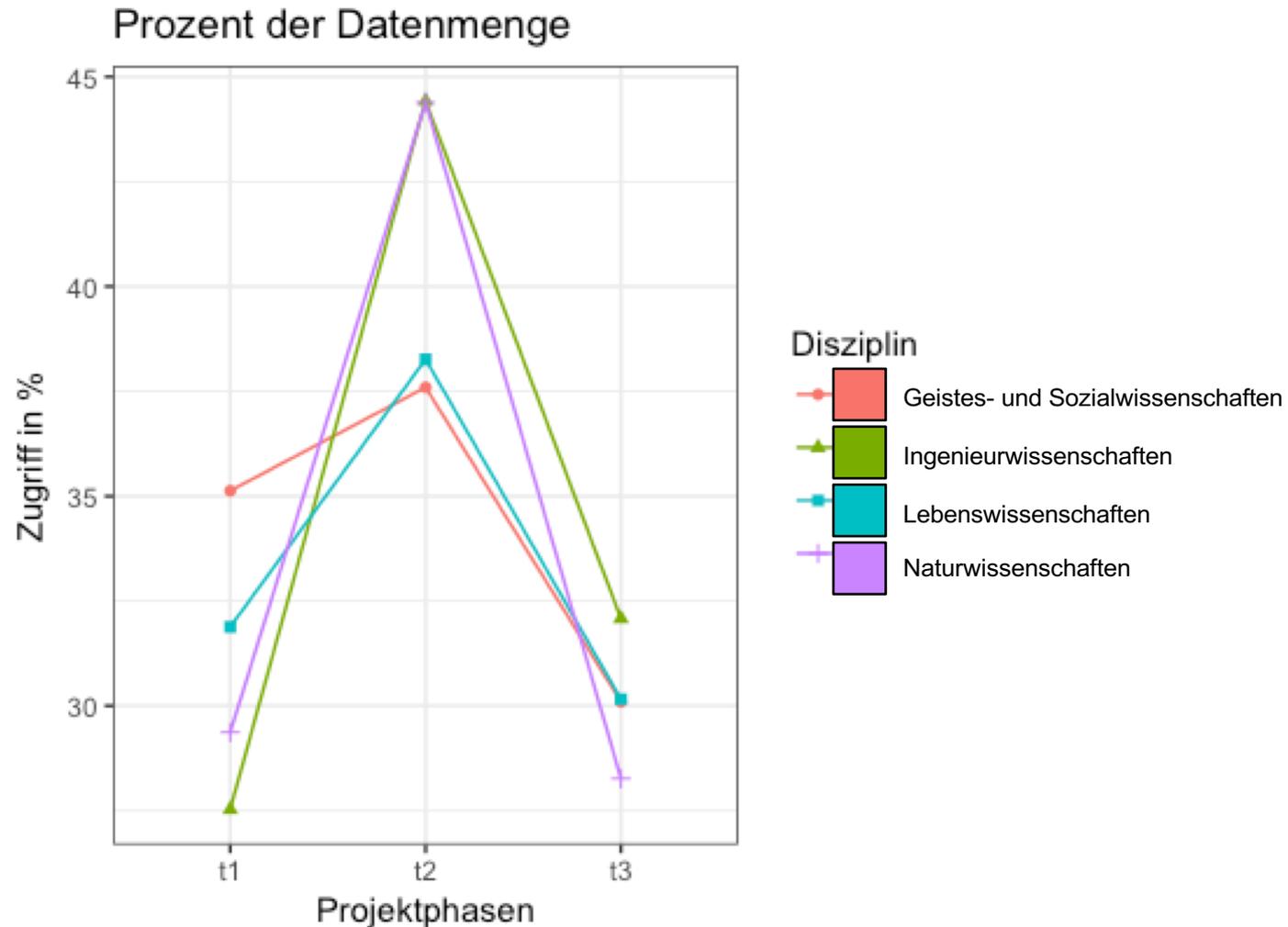


Diverse
Formate
von
Daten



..der Anfangsphase (T1)
..der Zwischenphase (T2)
..der Endphase (T3)

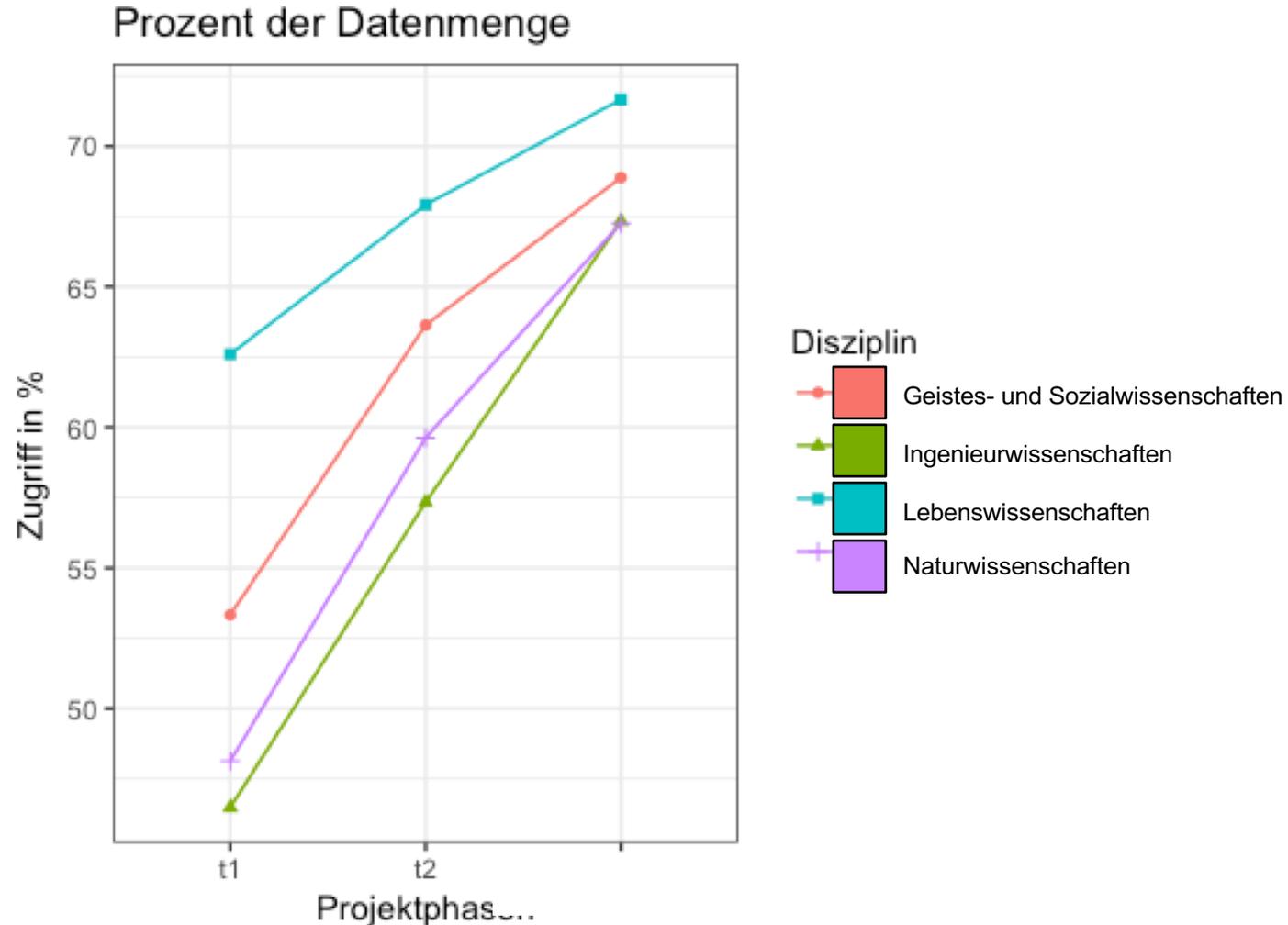
Daten fallen
weniger zu
Beginn und
Ende des
Projekts an



Wieviel Prozent der Gesamtdaten müssen (schätzungsweise) im Sinne der guten wissenschaftlichen Praxis 10 Jahre gespeichert werden?

..der Anfangsphase (T1)
..der Zwischenphase (T2)
..der Endphase (T3)

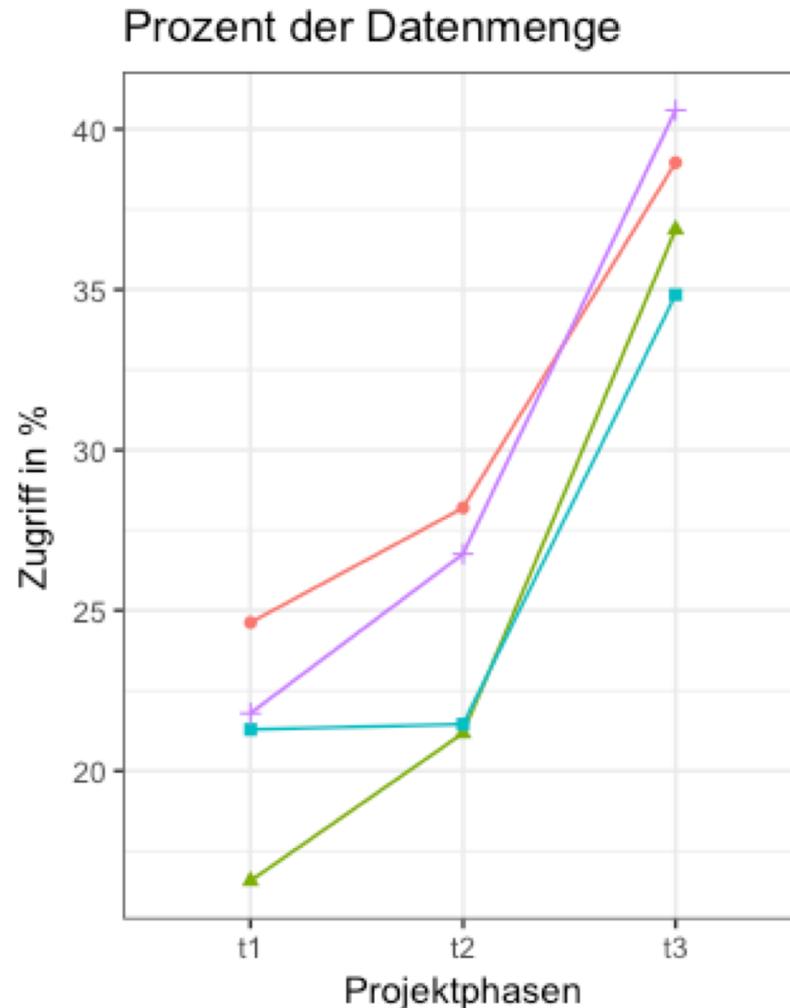
2/3 bis 3/4
der Daten in
Langzeit-
verfügbarkeit
(10 Jahre)



Wie viel Prozent der Gesamtdaten, sollten aus Ihrer Sicht mindestens 25 Jahre, oder darüber hinaus archiviert werden?

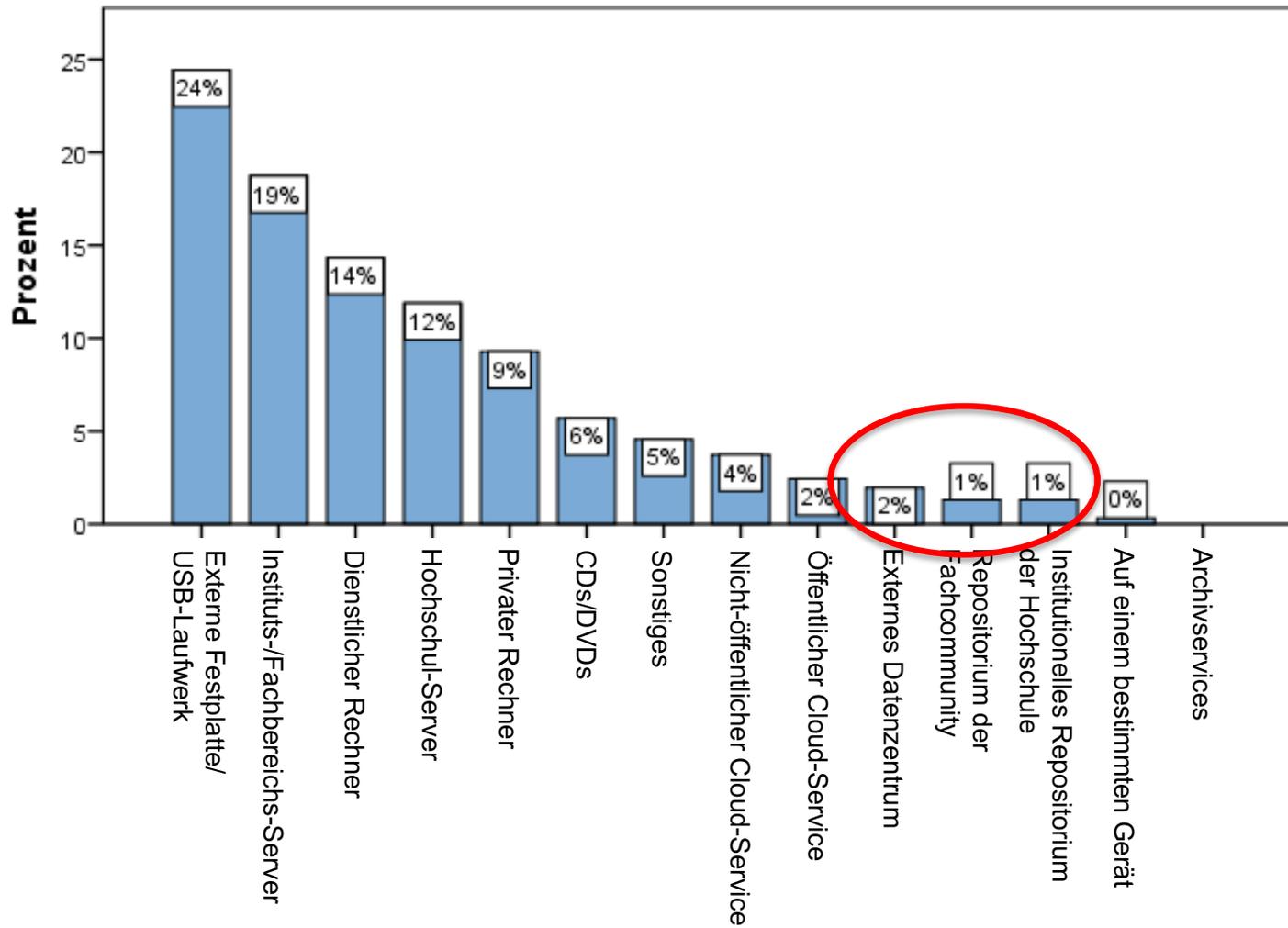
..der Anfangsphase (T1)
..der Zwischenphase (T2)
..der Endphase (T3)

> 1/3 der
Daten in
Archivierung
(25 Jahre)



Archivort:

Wo/Wie archivieren Sie die Forschungsdaten (für mindestens 25 Jahre), die im Rahmen eines typischen Forschungsprojekts anfallen, an dem Sie arbeiten?



> 1/3 der Daten in Archivierung (25 Jahre)

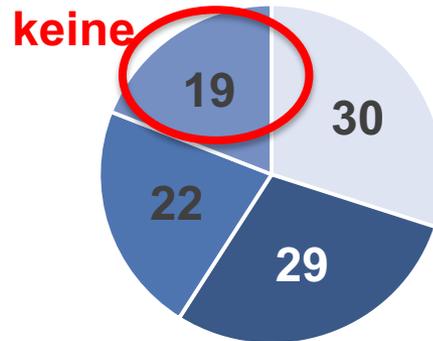
Metadaten:

Welche Metadaten erfassen Sie zur Wiederauffindbarkeit und Zuordnung Ihrer Daten?

Metadatenschemata:

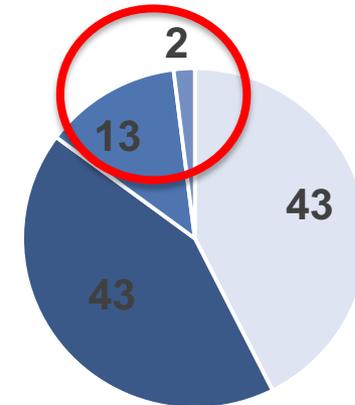
Nutzen Sie zur Erfassung der Metadaten ein existierendes Schema bzw. Standard?

Verwendete Metadaten



- Inhaltliche Metadaten
- Administrative Metadaten
- Technische Metadaten
- keine Metadaten

Metadatenstandards



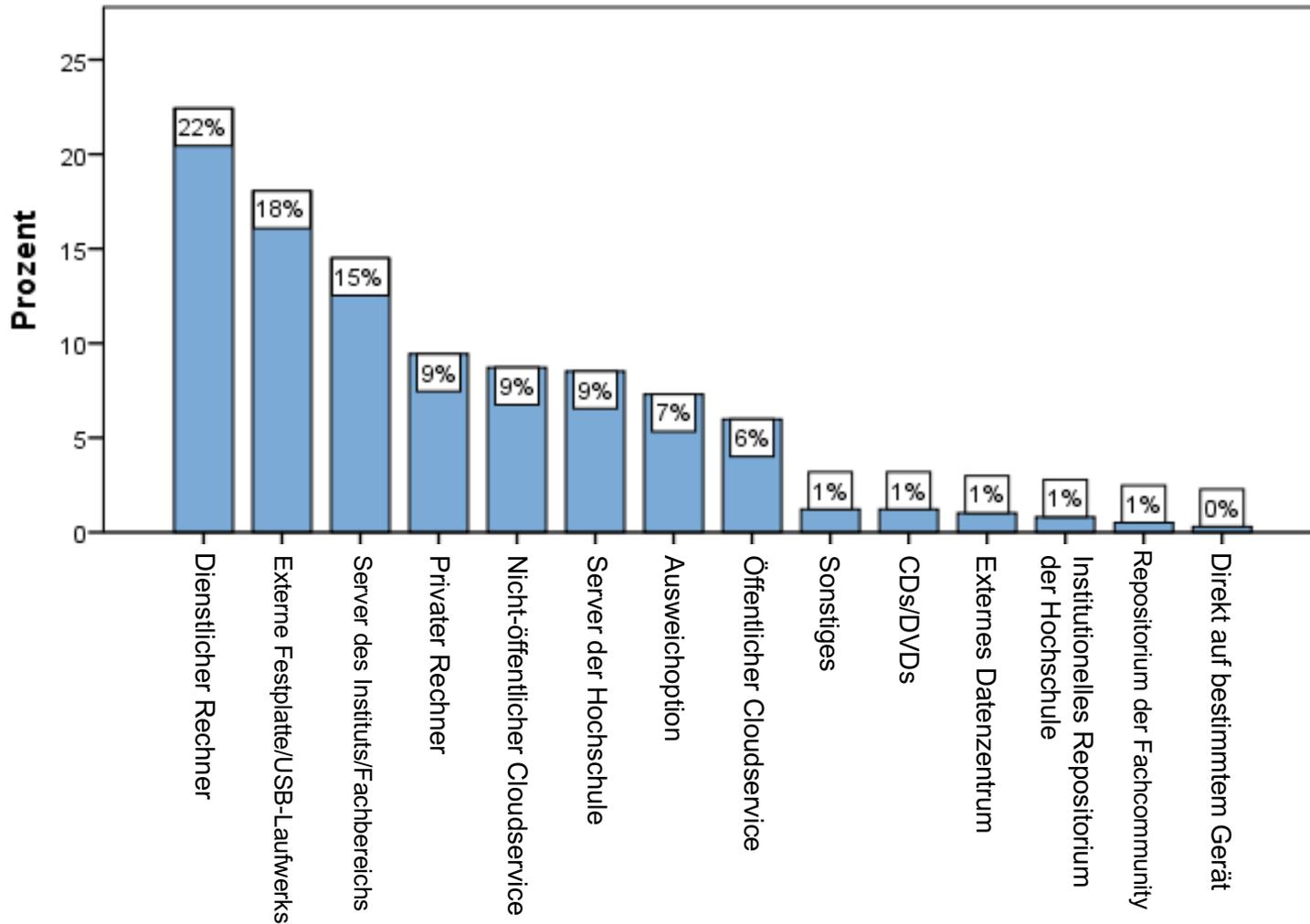
- Kein Schema / kein Standard
- Eigenes Schema
- Schema des Instituts/Fachbereichs
- Allgemeine / fachspezifische Metadatenstandards

80 % der
Forscher
erfassen
Metadaten

15 % der
Forscher
verwenden
Metadaten in
Absprache

Speicherort:

Wo speichern Sie die Forschungsdaten, die im Rahmen eines typischen Forschungsprojekts anfallen, an dem Sie arbeiten?

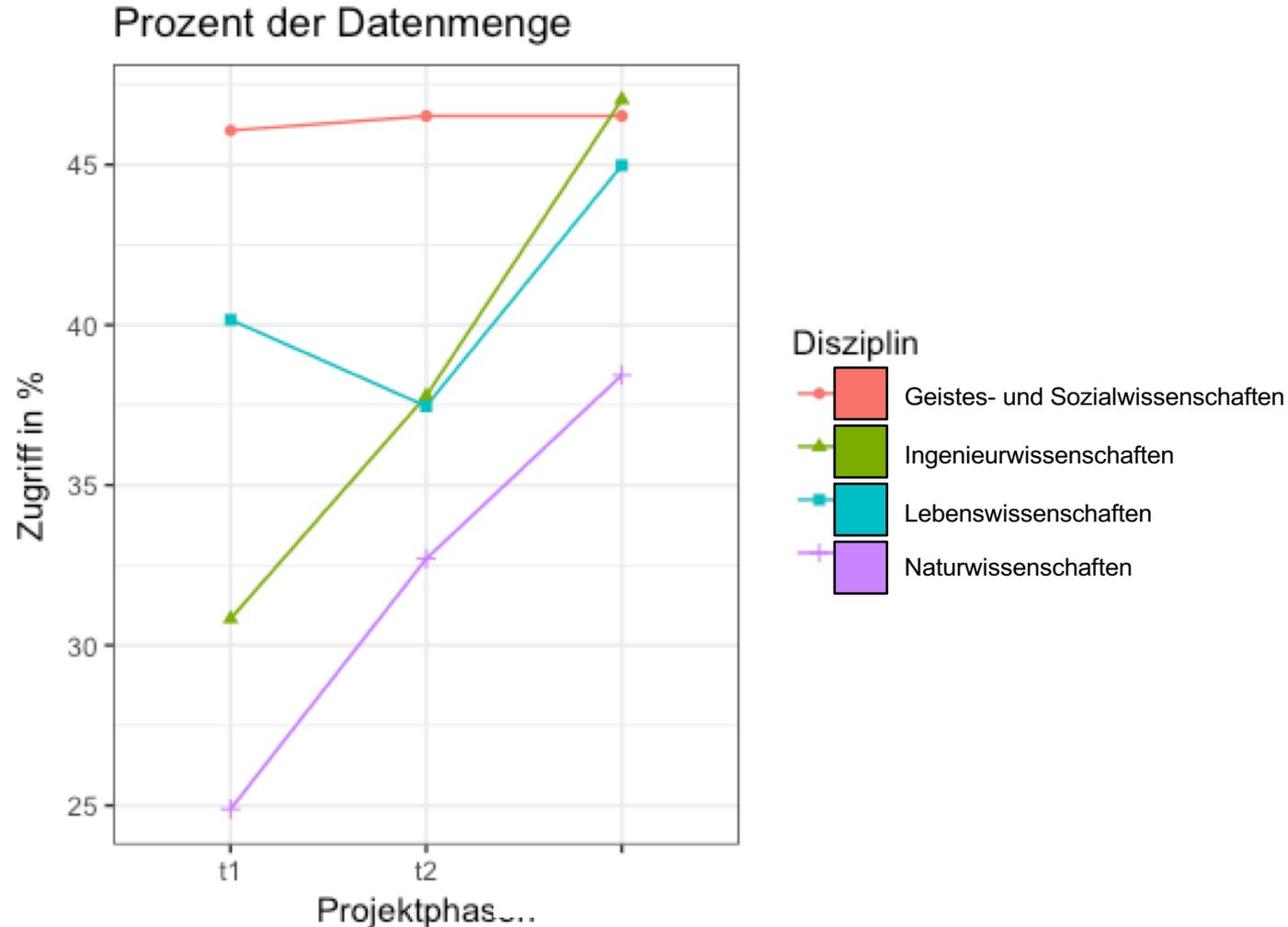


Auf ~40% der Daten Zugriff von mehreren Personen

Auf wieviel Prozent der Gesamtdaten greifen (schätzungsweise) mehrere Personen zu, um diese zu nutzen und zu bearbeiten?

..der Anfangsphase (T1)
..der Zwischenphase (T2)
..der Endphase (T3)

Umfang des
Zugangs zu
Daten
fachspezifisch



Forschungsfrage:

Which determinants influence researchers' decision to share research data with others?



Universität Duisburg-Essen

Abteilung für Informatik und Angewandte Kognitionswissenschaft

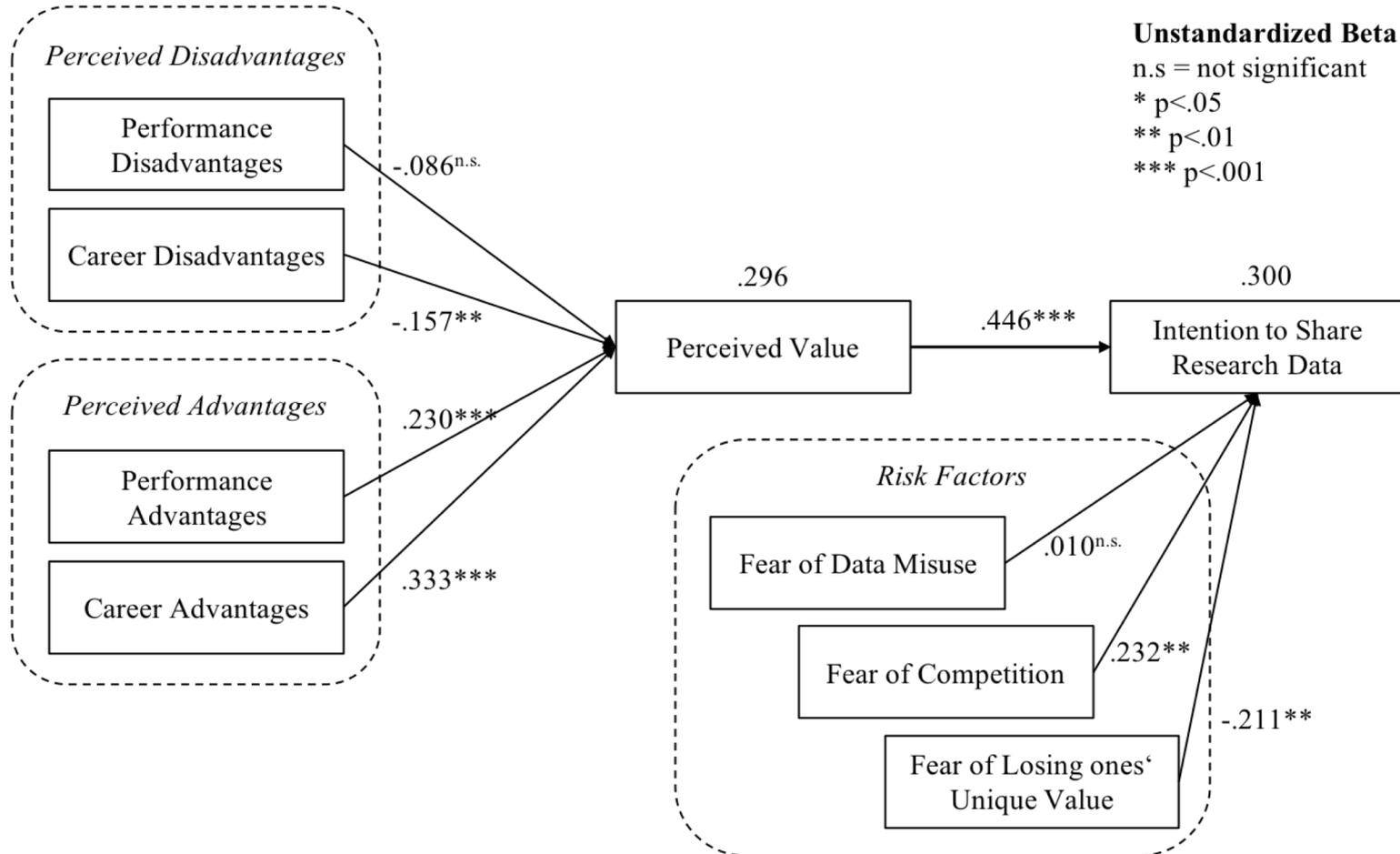
Professionelle Kommunikation in elektronischen Medien/Social Media

(Prof. Dr. Stefan Stieglitz)

Ansprechpartner: **Konstantin Wilms, M.Sc.**

www.uni.due.proco.de

„Intention to share research data“



- Standardisierte Projektlaufzeiten
- 2/3 bis 3/4 der Daten sollen langzeitverfügbar (10 Jahre sein)
- >1/3 der Daten sollen archiviert werden
- nur 4% der Forscher nutzen Repositorien & Datenzentren
- 80% der Forscher verwenden Metadaten
- Nur 15% verwenden dafür gemeinsame Schemata / Standards
- 50% der Daten liegen auf dem dienstlichen/privaten PC + Festplatten
- Auf 40% der Daten müssen mehrere Personen zugreifen



- Fachspezifische Analyse
- Beratung zu Metadaten, Information über Repositorien, Transparente Benutzung

Umfrage nachnutzbar für alle Hochschulen!

- Anpassung einzelner Fragen möglich
- Ergänzung von Fragen zu weiteren Themen, wie z. B. Beratungsbedarf
- Umfrage auf Deutsch und Englisch
- Informationen und Vorlagen zu Datenschutz und Durchführung stehen zur Verfügung

Link zur Testversion:

<https://www.soscisurvey.de/rdmHIL/?act=S432yQY3dOca2wDoCuxx1aiF>

Open science

- alle Ergebnisse und Materialien werden öffentlich auf Zenodo zur Nachnutzung zur Verfügung gestellt.

...WO stecken DEINE
Forschungsdaten
in 10 Jahren?!



...DEIN Backup?



...gute
wissenschaftliche
Praxis???



Vielen Dank für Ihre
Aufmerksamkeit 😊
www.uneke.de



Ordentliches
Forschungsdatenmanagement?!

...Forschungsdatenmanagement?



WO
liegen
DEINE
Daten...

...in
10 Jahren?



...DATEN FÜR DIE EWIGKEIT

