

EOSC EDEN: Ergebnisse der Umfrage zu “Identification, Selection and Appraisal in Contemporary Digital Preservation Practice” (Auswahl)

Micky Lindlar – TIB (Leibniz Information Centre for Science and Technology)
EOSC EDEN WP1 Lead



**Funded by
the European Union**

Micky Lindlar, TIB
NFDI AG LTA Sitzung
07. Mai 2026 (Online)

EOSC EDEN Grant agreement ID: 101188015



Trigger Warnungen

„Identification, Selection & Appraisal“

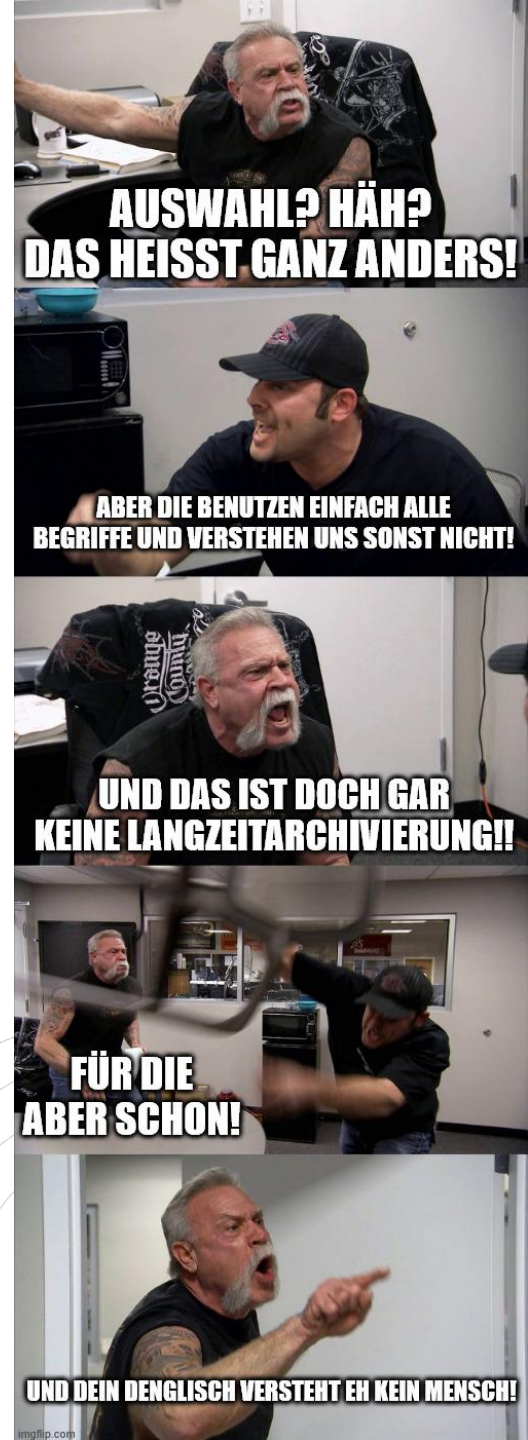
... falsche Reihenfolge, falsche Bedeutung, falscher Kontext ...

Langzeitarchivierung

... vertrauenswürdig!?, nur bitstream?, ein Repository!?, ...

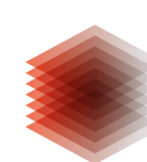
On a personal note ...

... Denglish is my language (Resultat: bunt gemischte Folien und Tonspur)



Was ist EOSC EDEN?

Projektname	EOSC EDEN
Projekttitel	Enhancing Digital preservation strategies at the European and National level
Laufzeit	1. Januar 2025 – 31 Dezember 2027
Förderer	Horizon Europe (Total: ≈8 Mio €)
Call topic	HORIZON-INFRA-2024-EOSC-01-4 Long-term access and preservation infrastructure development for EOSC, including data quality aspects
Rolle TIB	Leitung WP1, Leitung T1.2 & T1.3, Mitarbeit in allen anderen WPs
Koordinator	CSC – IT Center of Science
Partner	16 partners (incl. coordinator) from 9 different countries: Finland (1), Germany (3), the Netherlands (3), Norway (1), Belgium (1), UK (2), Switzerland (2), France (1), Poland (1)



TIB
LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY



Universität
Bremen



PREMOTECH



Swiss Institute of
Bioinformatics



Arkivum



UK Data Service



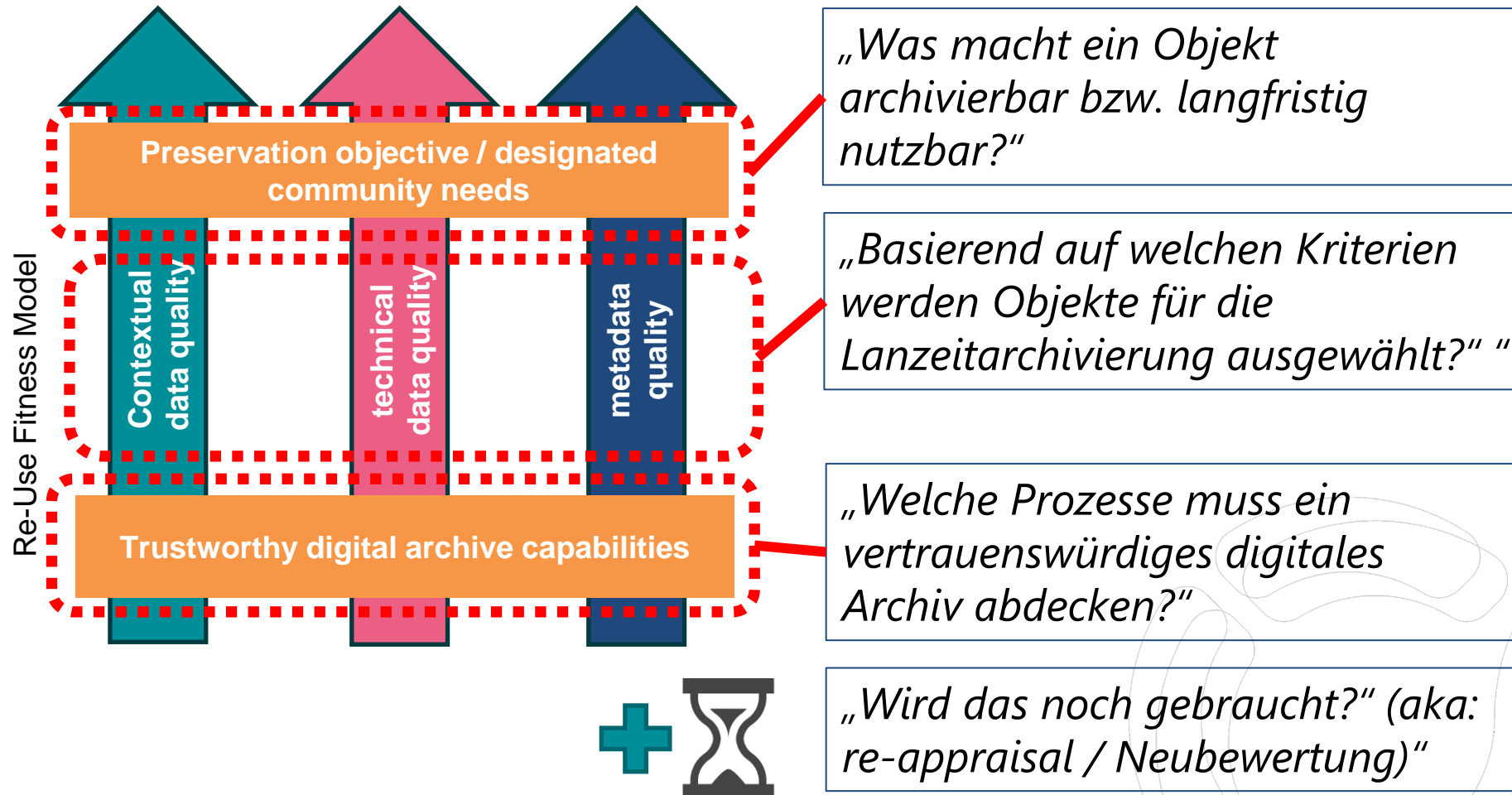
Warum EOSC EDEN?

Um sicherzustellen, dass Forschungsdaten für die jeweiligen Fachgemeinschaften verständlich & nutzbar bleiben

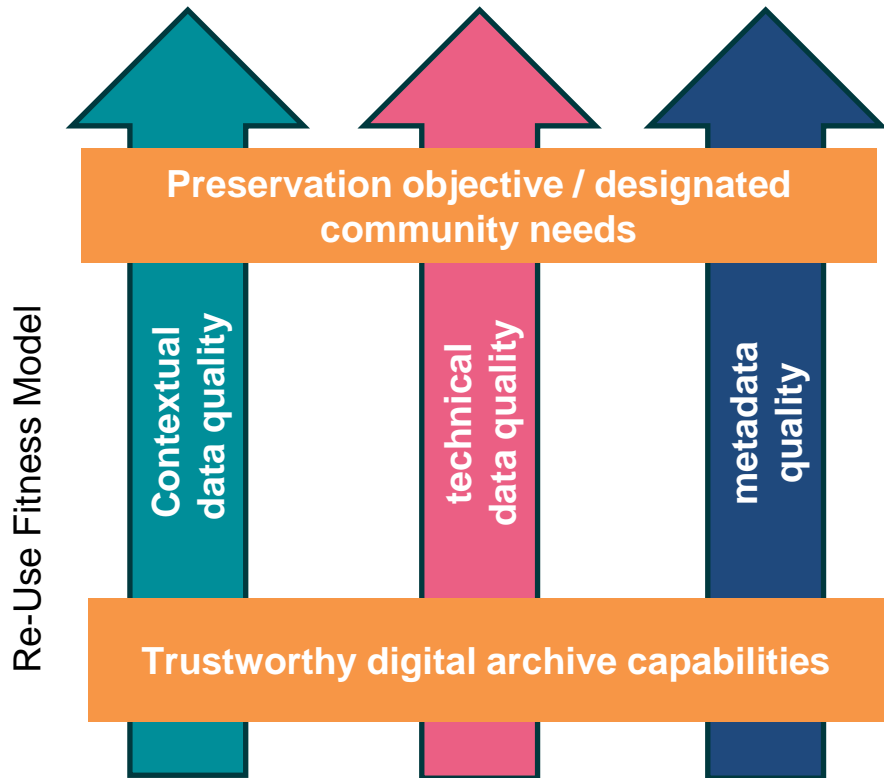
- Im Hinblick auf begrenzten Speicher und Servicekapazität stellt sich die Frage **welche Daten für die digitale Langzeitarchivierung ausgewählt werden sollen (oder können)**
- **Uneinheitliche Kenntnisstände** rund um LZA-Verfahren/Standards in den wissenschaftlichen Disziplinen führt zu Fragmentierung und Problemen der Interoperabilität
- Datenkuratierung und -archivierung sind unverzichtbare Prozesse, um die **Datenqualität insbesondere über einen längeren Zeitraum hinweg zu gewährleisten**; allerdings lassen sich die mit den langfristigen Vorteilen verbundenen Kosten oft nur schwer rechtfertigen.



EOSC EDEN Re-Use-Fitness Modell (RUFit)



Umfrageergebnisse im Kontext von RUFit



- Mai – Juli 2025: Umfrage
- September 2025: M1.1 Core Preservation Processes
- Juli 2026: M1.2 Definition of Metrics for Re-Use Fitness of Information Objects

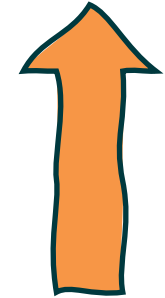
Umfrage insb. als erste Übersicht für:

- Welche Qualitätsüberprüfungen werden bei der Aufnahme von Objekten berücksichtigt?
- Welche Guidelines/Standards sind bekannt und werden genutzt?
- Wie gelangen Objekte in Archive?

Überblick

- Umfrage lief vom 2. Mai – 30. Juni 2025 im EU Survey Tool
- Gerichtet an LZA-Praktiker*innen aus verschiedenen Organisationstypen und wissenschaftlichen Disziplinen, inkl. Forschung, Kulturerbe, Industrie
- 250 valide Antworten
- Umfrageaufbau (20 top-line Fragen):
 1. Infos zur Organisation & Rolle
 2. Bekannte / genutzte Guidelines & Standards
 3. Über aktuelle Praktiken / Prozesse
 4. Disziplinspezifische Anforderungen
 5. Freiwillige Angabe von Kontaktinfos

Für weitere Infos, siehe auch:
Molloy & Lindlar „Towards Understanding
Identification, Selection & Appraisal in
Contemporary Digital Preservation Practice“
<https://phaidra.univie.ac.at/o:2198813>

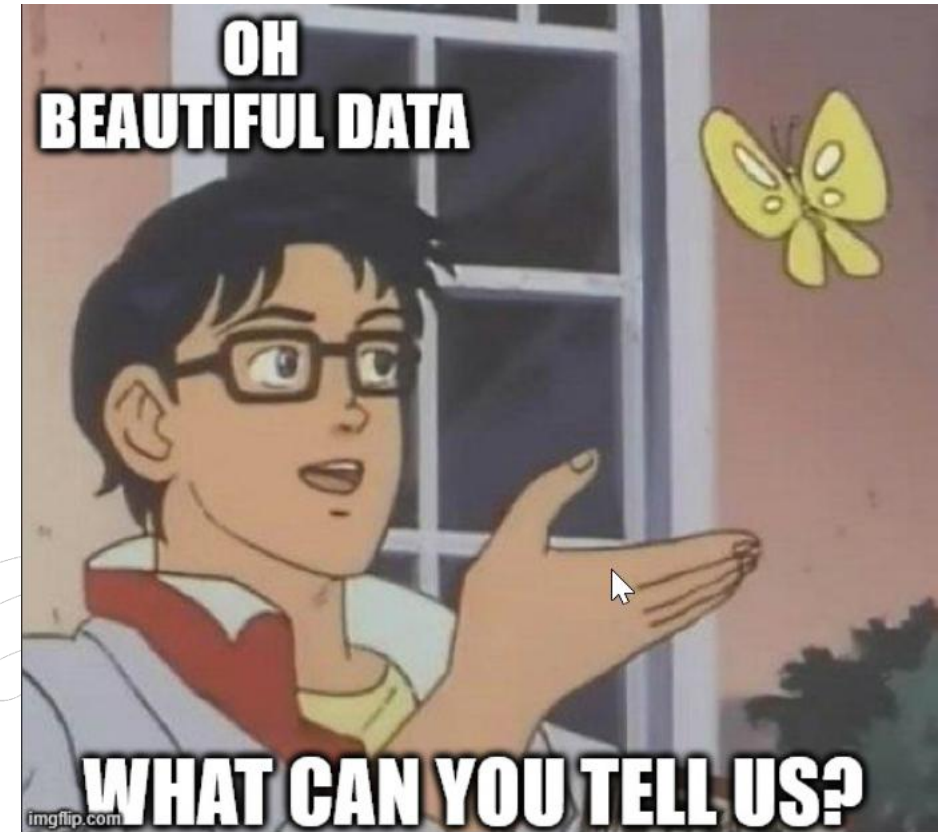


Umfrageaufbau / -design:

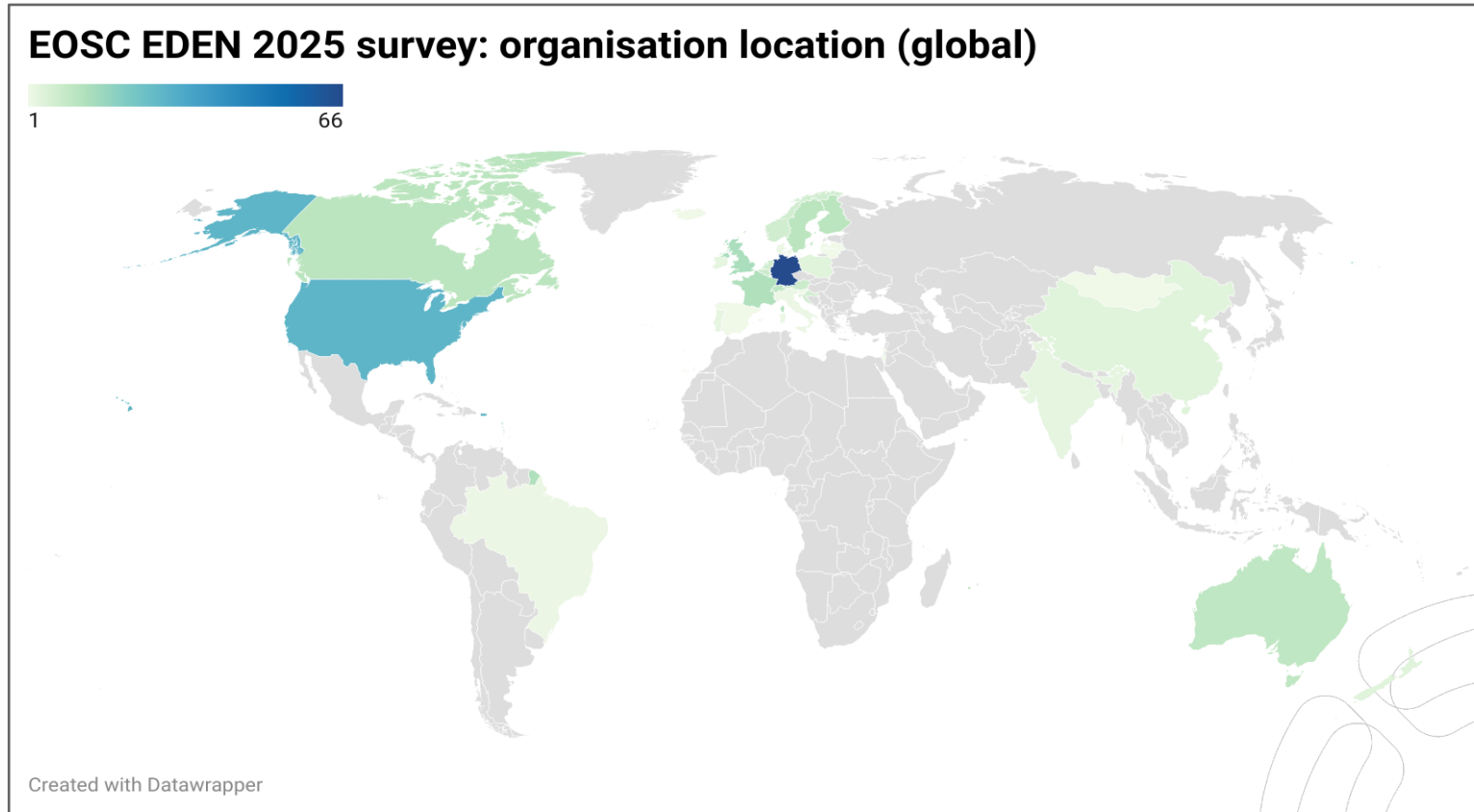
- disziplin-spezifische Sprache (inkl. technische Fachbegriffe) weitestgehend vermeiden
- mehrere Antworten pro Institutionen zulässig
- weltweite Antworten willkommen
- Keine Pflichtfragen
- „Andere“ Option für die meisten Fragen

Kurze Einblicke für heute

- Wer hat an der Umfrage teilgenommen?
- Guidelines & Standards:
Was kennen und nutzen wir?
- Wie lange ist lang ... und was passiert danach?
- Auf welchen Wegen gelangen Daten ins Archiv?
- Welche Qualitätsüberprüfungen werden im Deposit/Ingest durchgeführt?
- Einmal valide, immer valide?



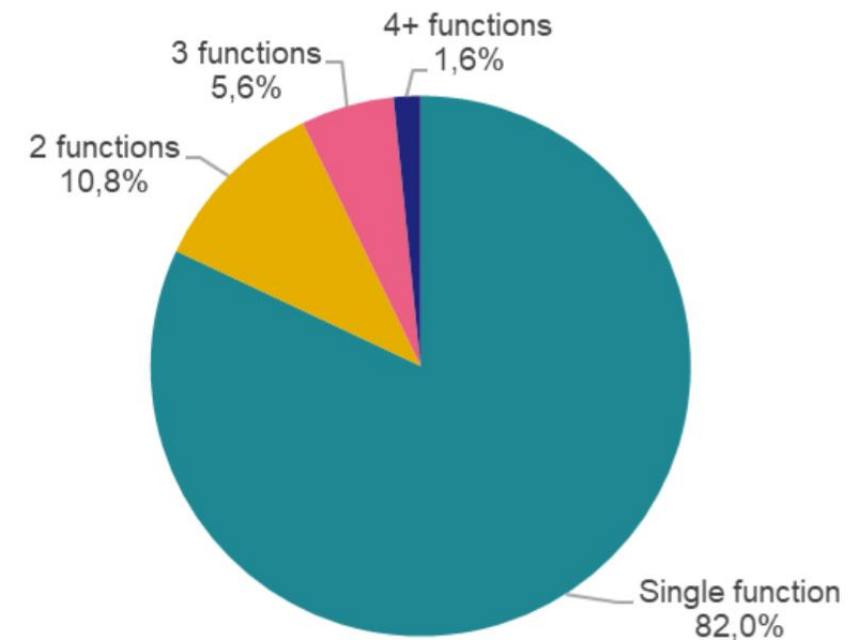
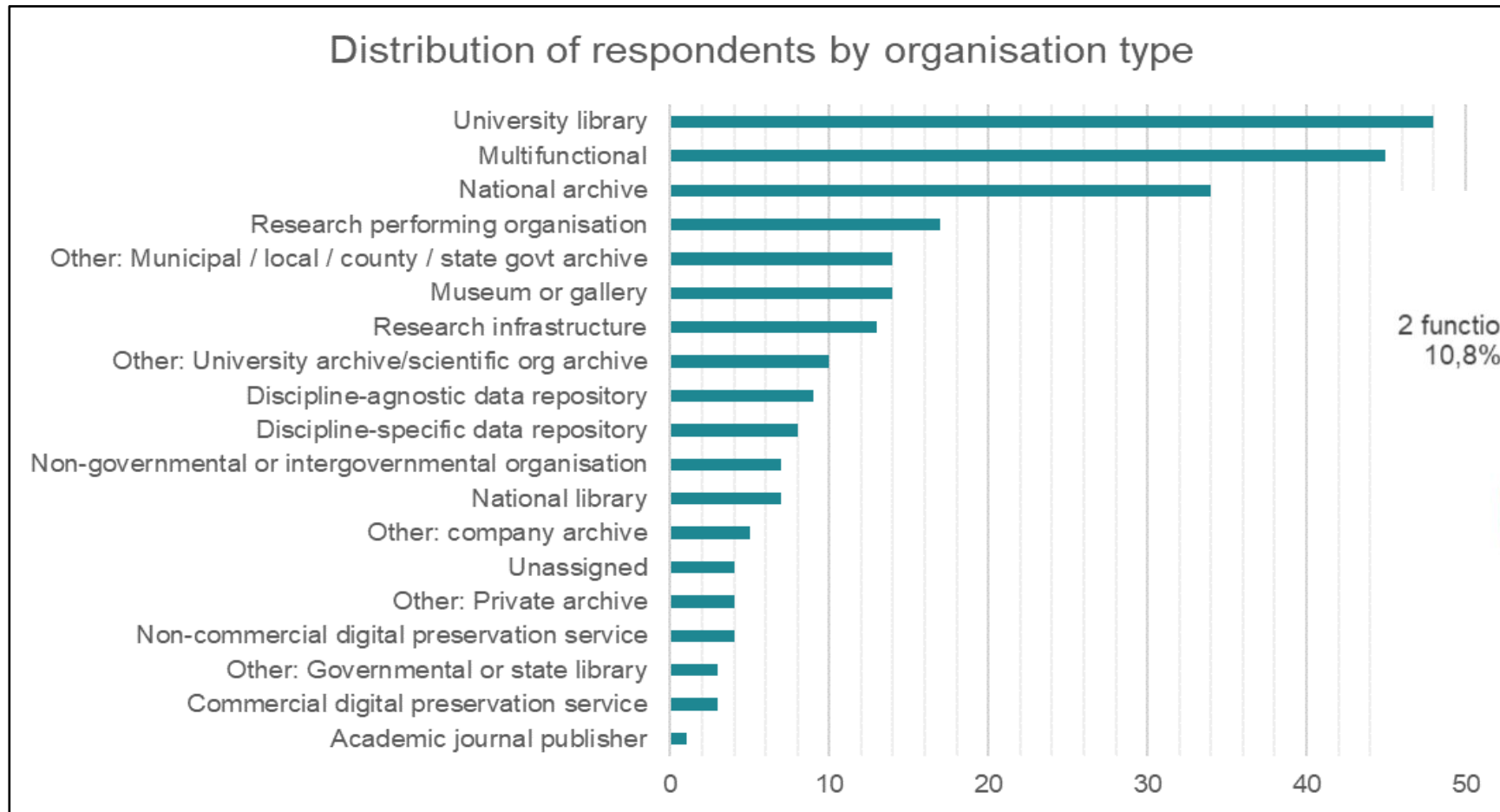
Wer hat teilgenommen – geographische Verteilung



From most to least frequent:

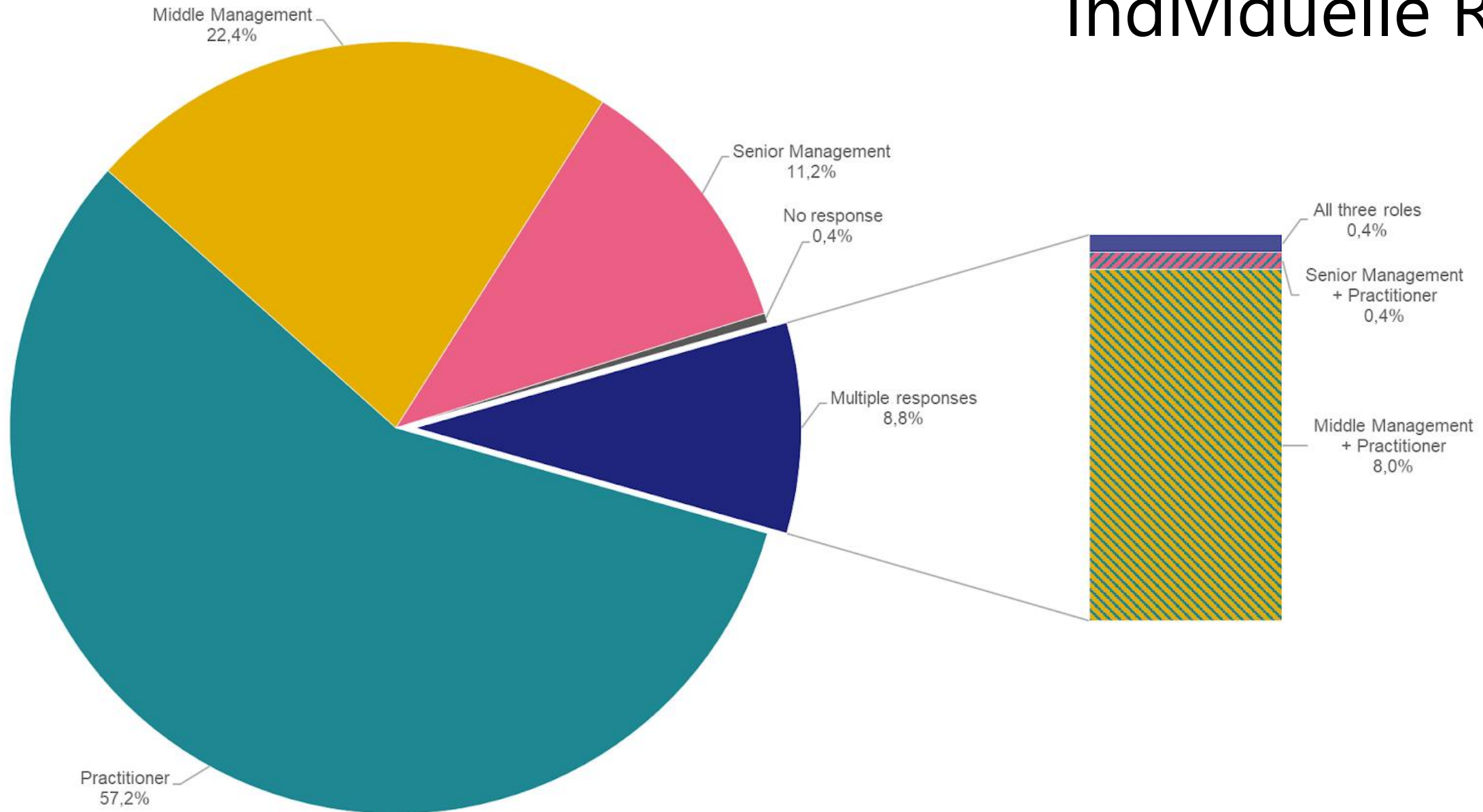
Germany 66, United States of America 29, United Kingdom 14, France 13, Canada 11, Finland 11, Sweden 11, Australia 10, Switzerland 10, Austria 8, Belgium 8, Croatia 7, Netherlands 7, Norway 7, China 4, New Zealand 4, Poland 4, India 3, Ireland 3, Portugal 3, Brazil 2, Denmark 2, Italy 2, Luxembourg 2, Iceland 1, Israel 1, Latvia 1, Lithuania 1, Mongolia 1, Slovenia 1, Spain 1, Other 1, No Answer 1.

Wer hat teilgenommen? Organisationstyp



Wer hat teilgenommen? Individuelle Rollen

Overall distribution of roles



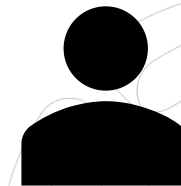
Wer: Am häufigsten vertretene Eigenschaften

Basierend auf den häufigsten Antworten auf die jeweiligen Fragen, ist die typisch teilnehmende Person:

Rolle: LZA Praktiker:in

Organisationsumfeld: in einem mittel (51-250 VZÄ) bis großen (>251 VZÄ) Archiv oder eine Universitätsbibliothek in Europa

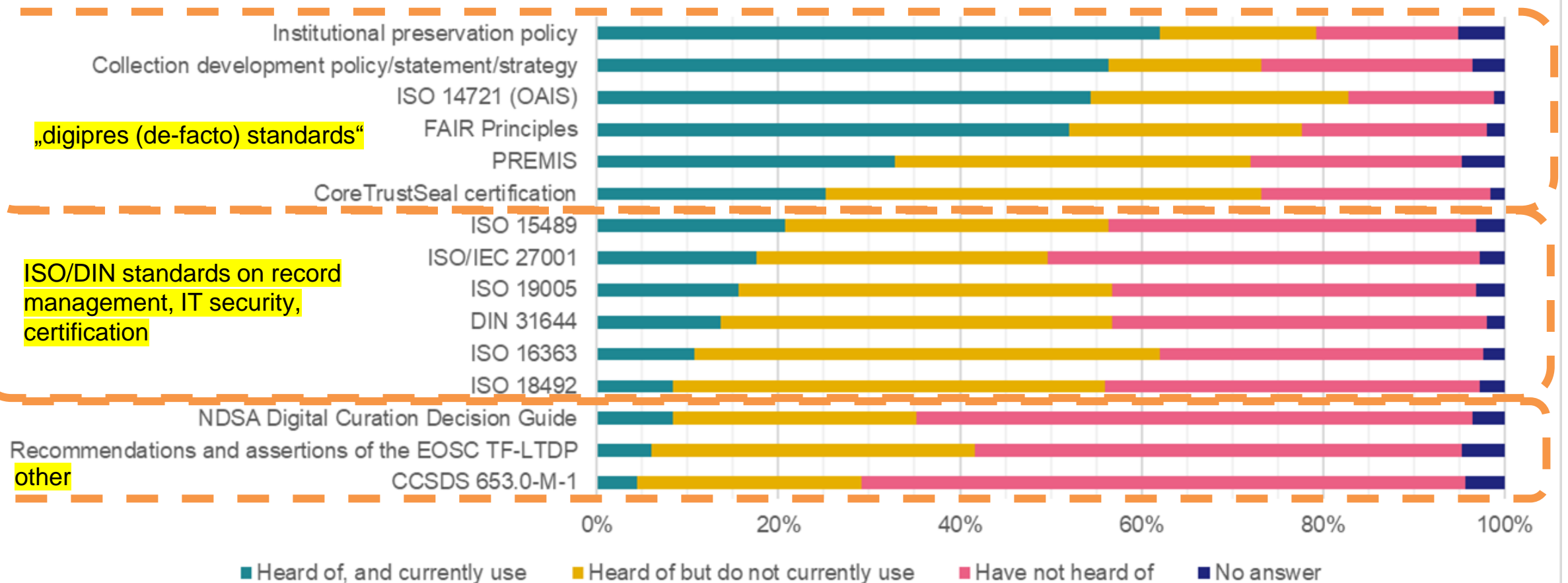
Disziplin-spezifische Ausrichtung: deren Sammlung keinen singuläre fachspezifische Ausrichtung hat, aber mehrere Disziplinen abdeckt



Durchschnittliche:r LZA-
Praktiker:in der Umfrage

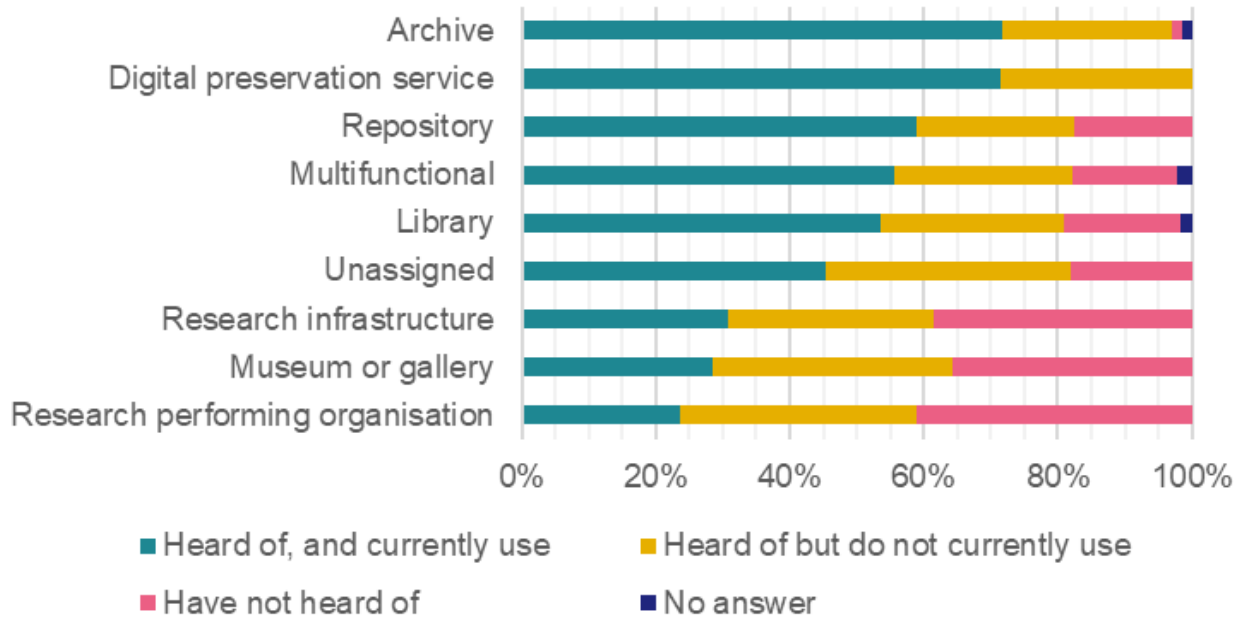
Guidelines & Standards: Was nutzen wir?

Knowledge and use of frameworks

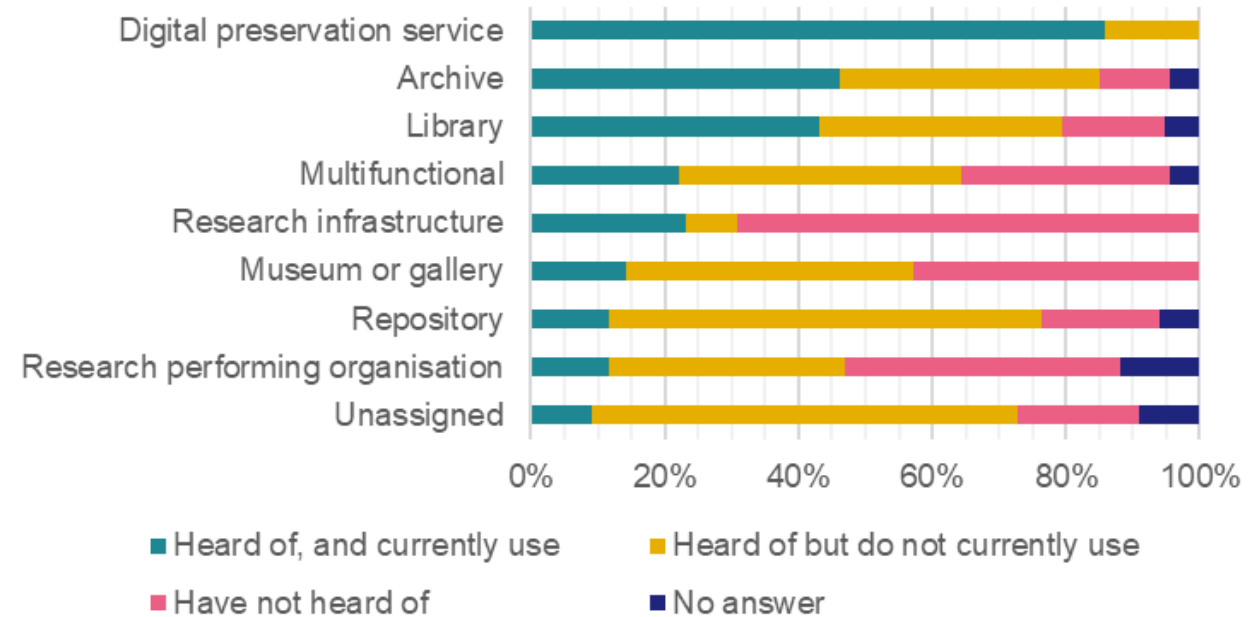


Guidelines & Standards: Was nutzen wir?

Knowledge and use of OAIS

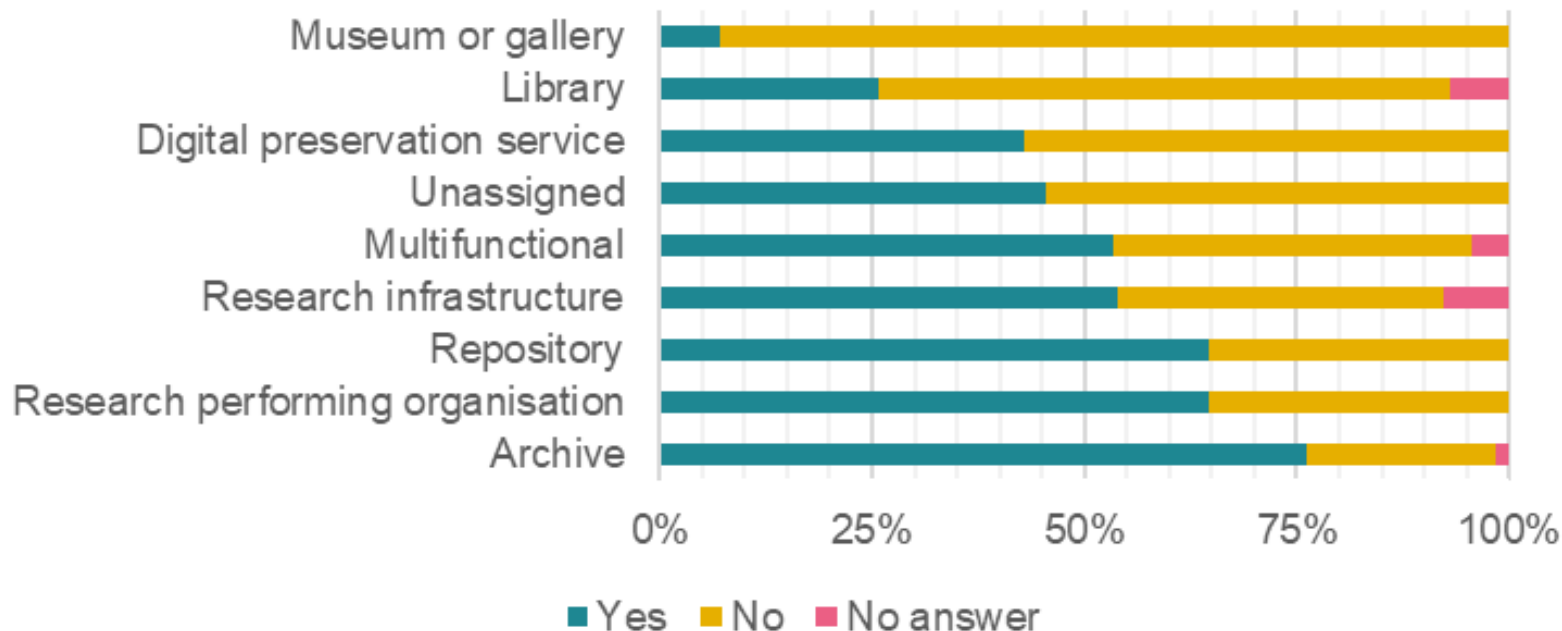


Knowledge and use of PREMIS



Wie lange ist lang?

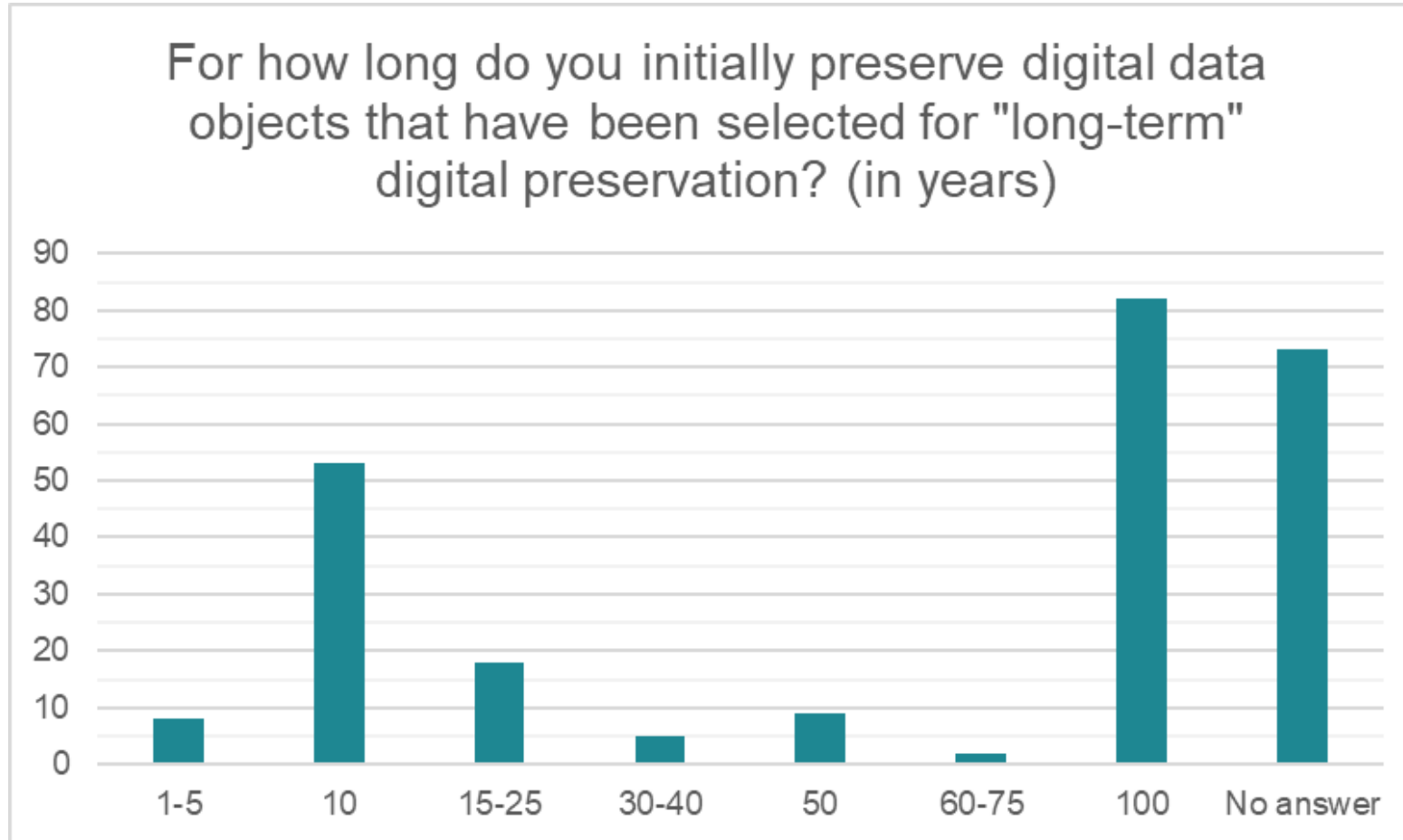
In your organisation, do you have an agreed or working definition of "long-term" in the context of digital preservation? (by "no")



Bei "ja" – Folgefrage: Wie lang ist Langzeit?

- Für immer (n=35)
- 10 Jahre (n=30)
- "Lange genug, um Obsoleszenz zu überstehen" (n=7)
- 100 Jahre (n=6)
- 50 Jahre (n=5)
- "so lange die Institution existiert" (n=4)
- ...
- "Ein Archiv ohne Nutzende ist lediglich ein trostloser Aufbewahrungsort für Informationen, auf die nie zugegriffen wird."

Wie lange werden Objekte initial archiviert?



Im Gegensatz zur Freitext-Frage nach einer Definition von „Langzeit“, konnten hier nur Zahlenwerte angegeben werden

... und was passiert am Ende von „lang“?

Once the initial "long-term" preservation period has ended, what usually happens to digital data objects?	n	Ratio (of 250)
We take no specific action.	98	39.20%
I don't know.	48	19.20%
We do something else at the end of the initial preservation period.	38	15.20%
We consult the depositor to agree on the next steps.	32	12.80%
No Answer	29	11.60%
We return the data to the depositor and delete our copy.	3	1.20%
We delete the data, without consulting the depositor.	2	0.80%

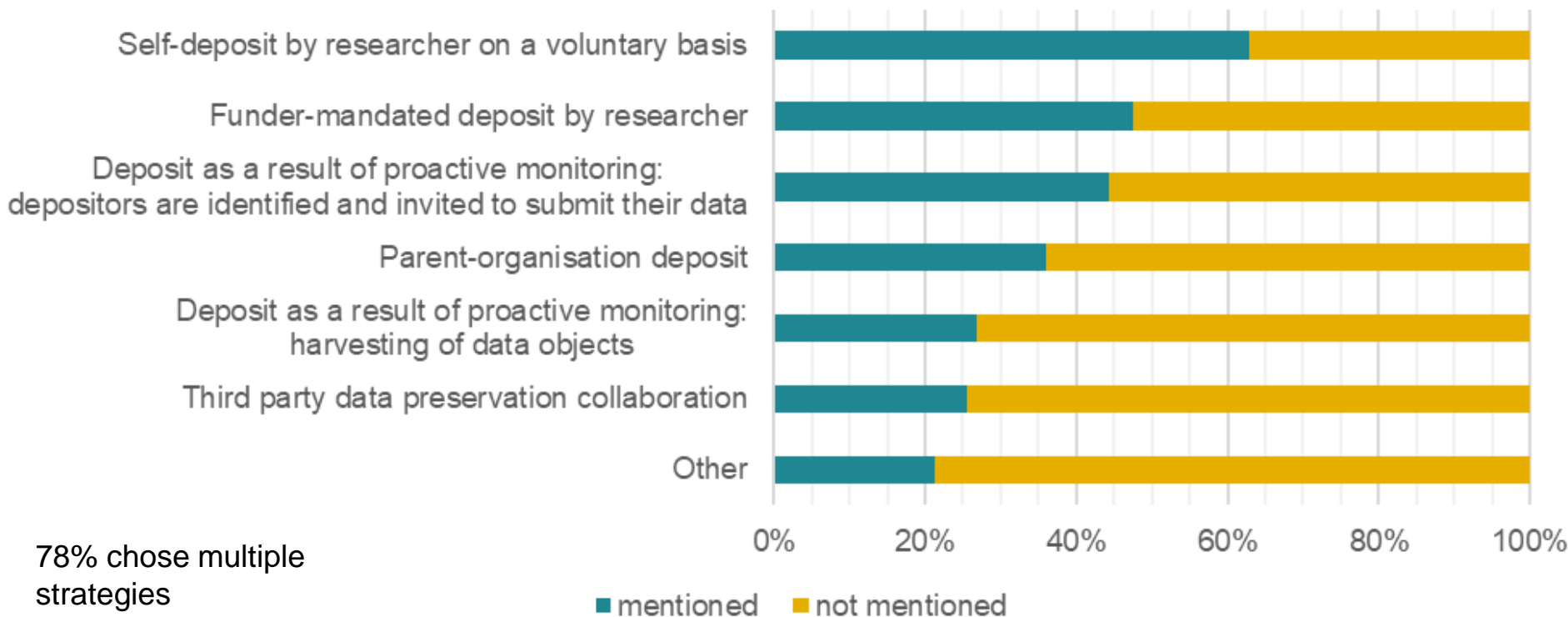
Andere (Auswahl)

- Weiterführung der Archivierung (n=20)
- Rücksprache mit abliefernder Stelle (n=8)
- Reevaluierung der Aufbewahrungsfrist (n=7)
- Wir nutzen das Konzept einer "initialen Phase" nicht (n=2)



Wie gelangen Daten ins Archiv?

Most common responses for acquisition strategies among all responses



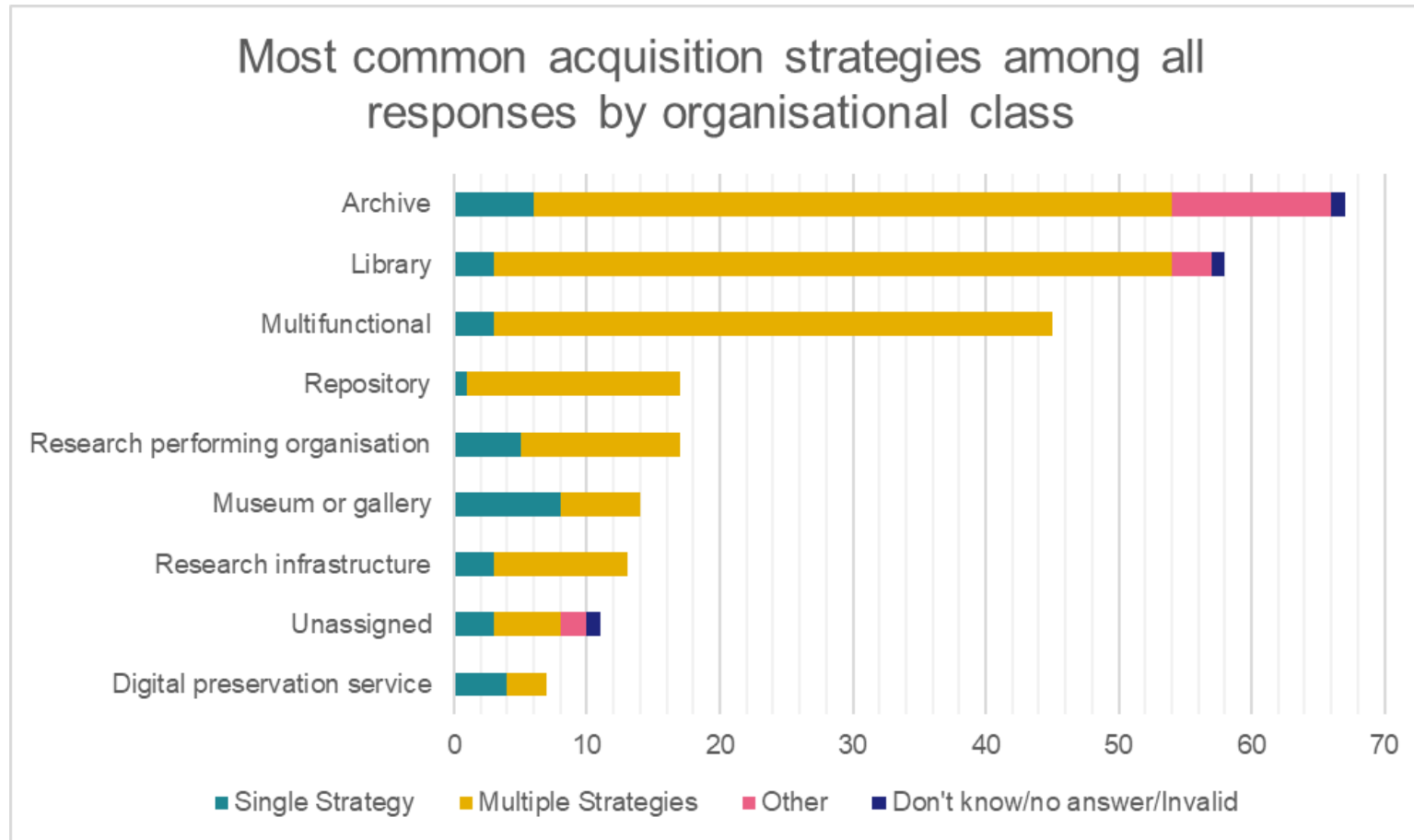
Stärkste Verbindung:

- Freiwillige Selbst-Ablieferung durch Forschende & durch Förderer verpflichtende Selbst-Ablieferung

Schwächste Verbindung:

- durch Förderer verpflichtende Selbst-Ablieferung & proaktives Monitoring / Harvesting

„Erwerbungsstrategien“ nach Organisationstyp

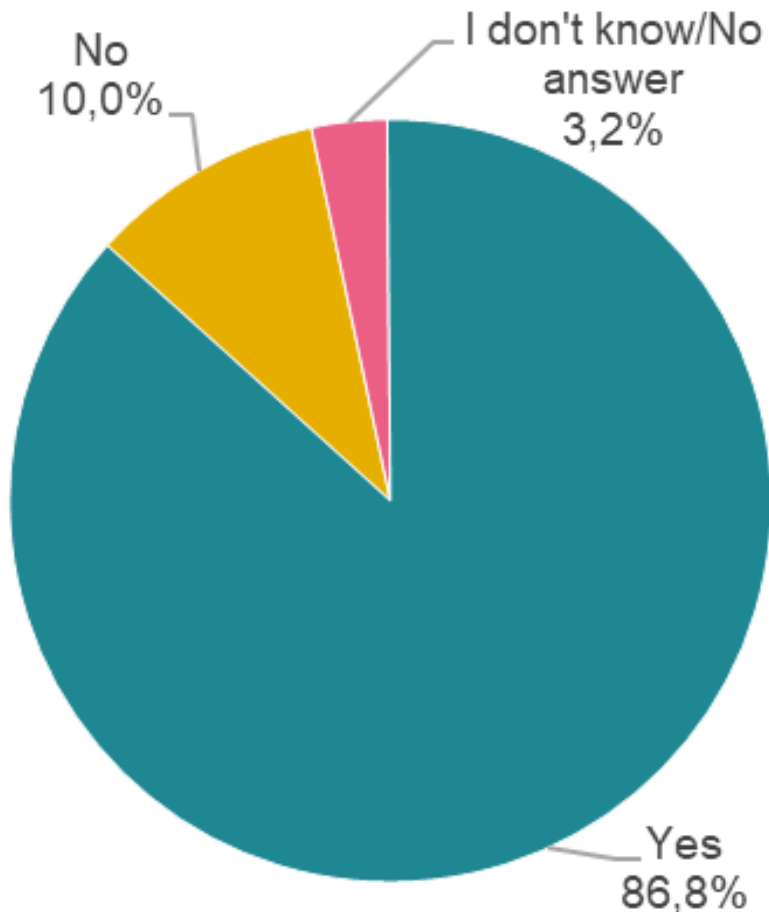


Mehrfache genannte Wege unter „Andere“ (n=19)

- Digitalisierung
- Nach-/Vorlass oder Geschenk
- Pflichtabgabe / gesetzliche Regelung / Archivgesetz
- Durch organisationsinterne Prozesse

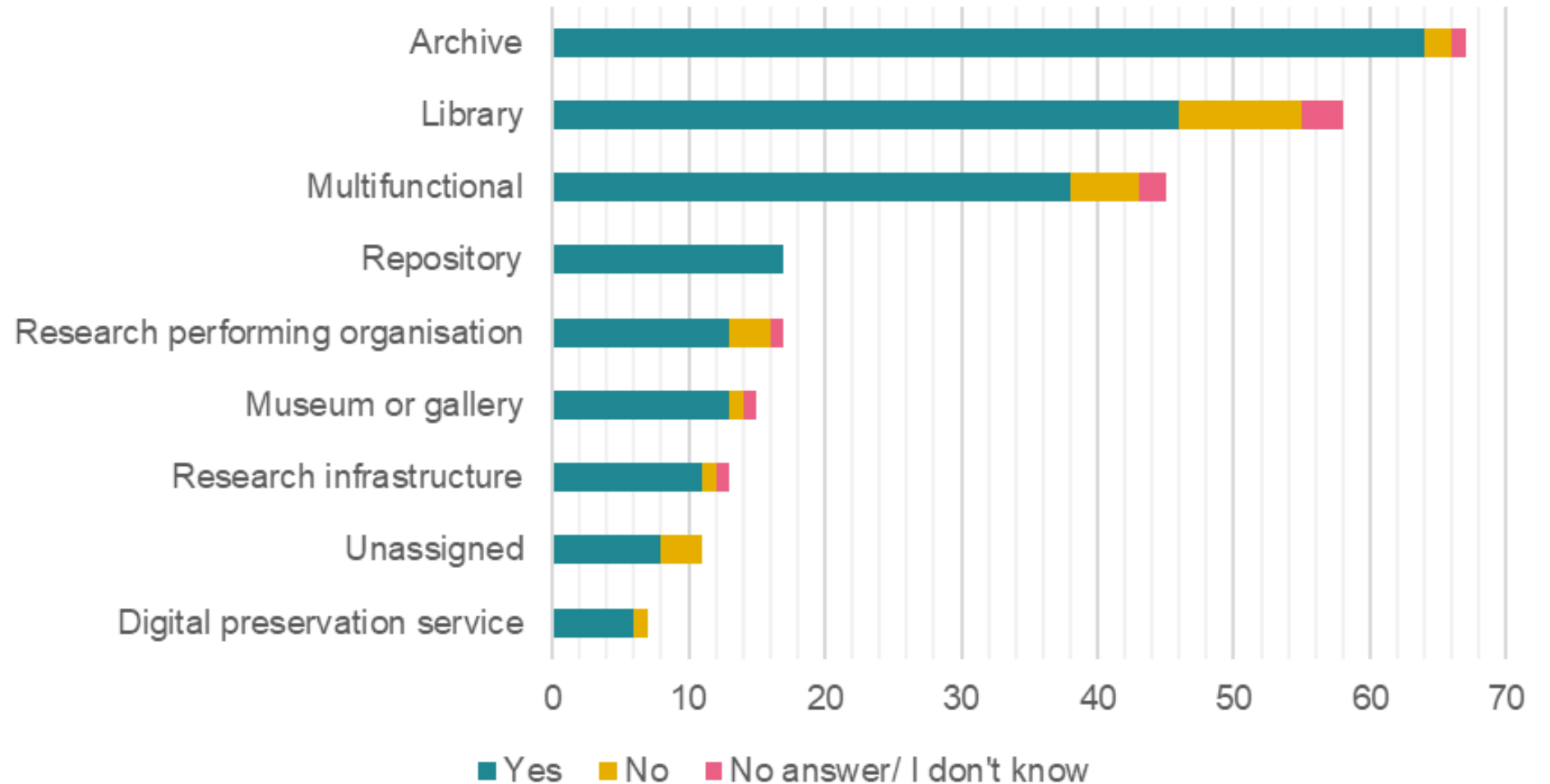
Qualitäts- überprüfungen in der Deposit/Ingest- Phase

Do you assess the quality of digital data objects
at the point of ingest for "long-term"
preservation?



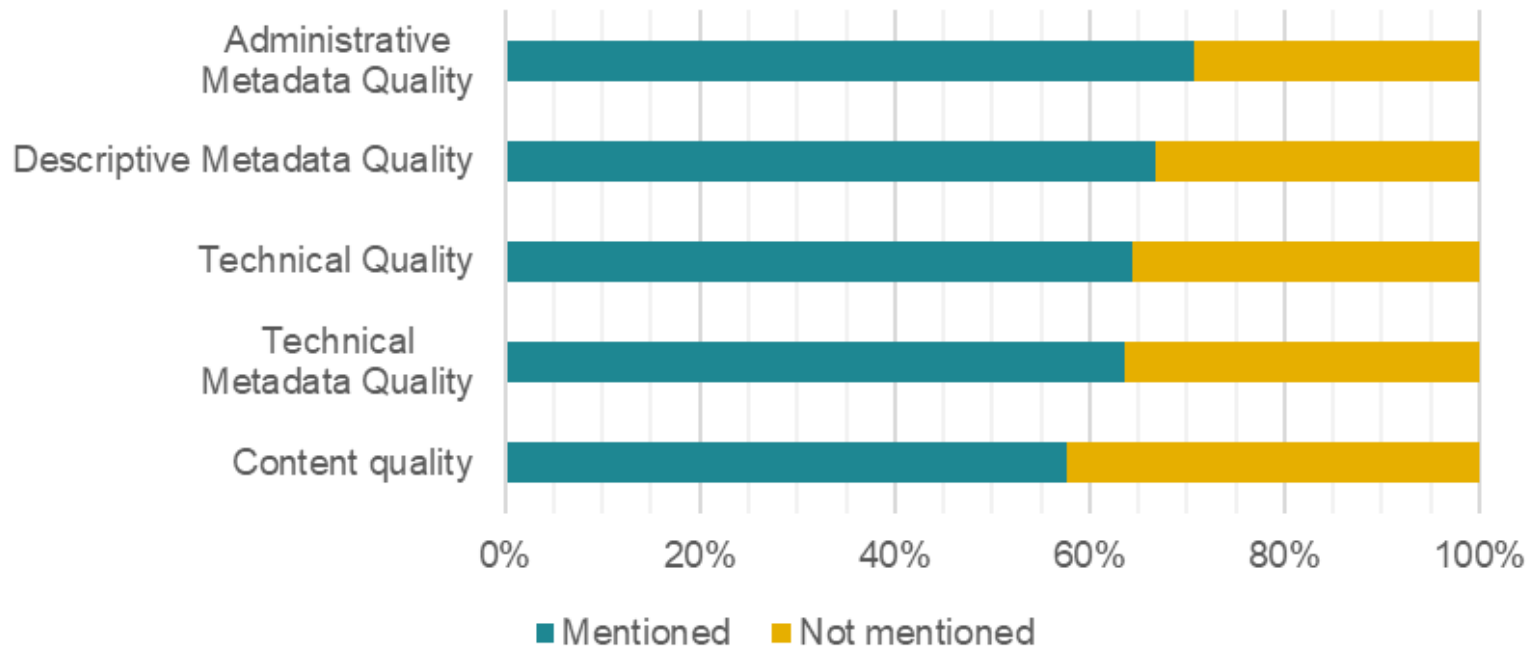
Qualitäts- überprüfungen in der Deposit/Ingest- Phase nach Organisationstyp

Do you assess the quality of digital data objects at the point of ingest for "long-term" preservation?



Arten von Qualitätsüberprüfungen in Deposit/Ingest-Phase

Most frequently mentioned quality indicators that are assessed at point of ingest for "long-term" preservation



Qualität Administrativer Metadaten

z.B. Rechte, Lizenzen, Zugriffsmodalitäten

Qualität Deskriptiver Metadaten

z.B. Creator, Description, Subject, Forschungsthema

Technische Qualität

z.B. korrekte Darstellung (Rendering), Erfüllung von Formatspezifikationen

Qualität technischer Metadaten

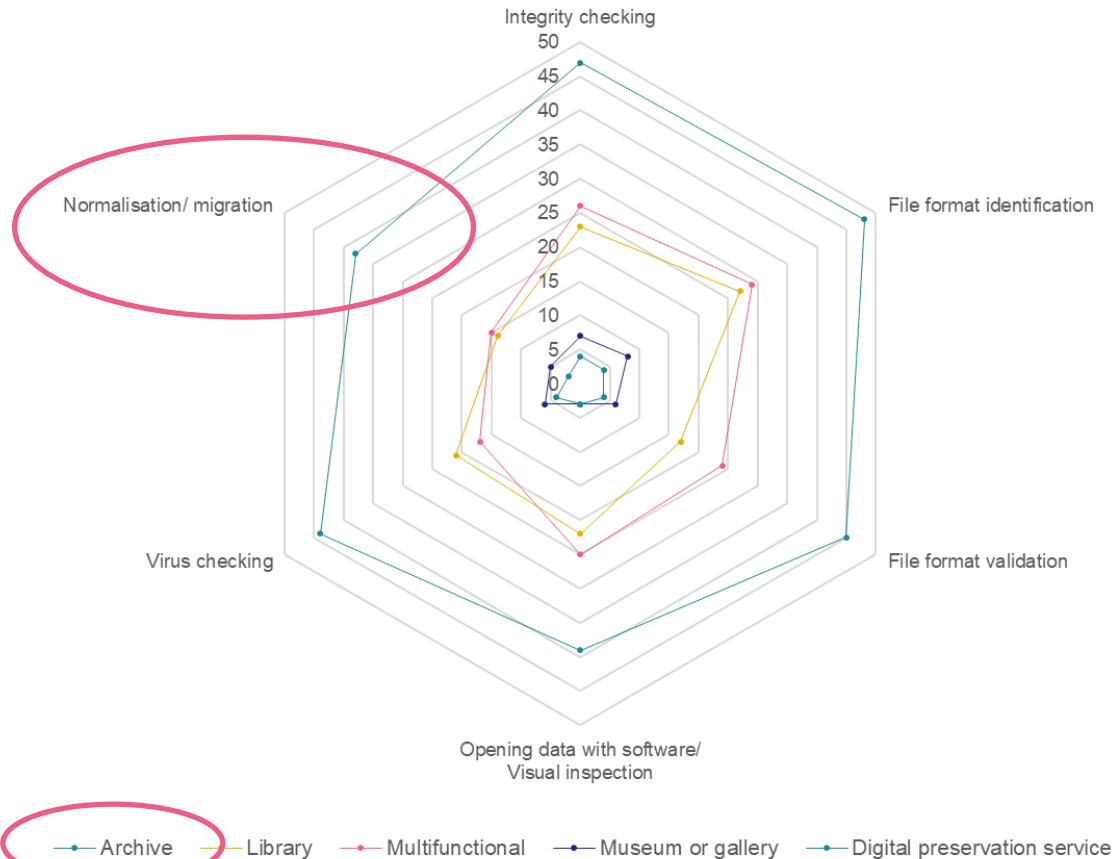
z.B. Format, Auflösung, MD5 Checksumme

Inhaltliche Qualität

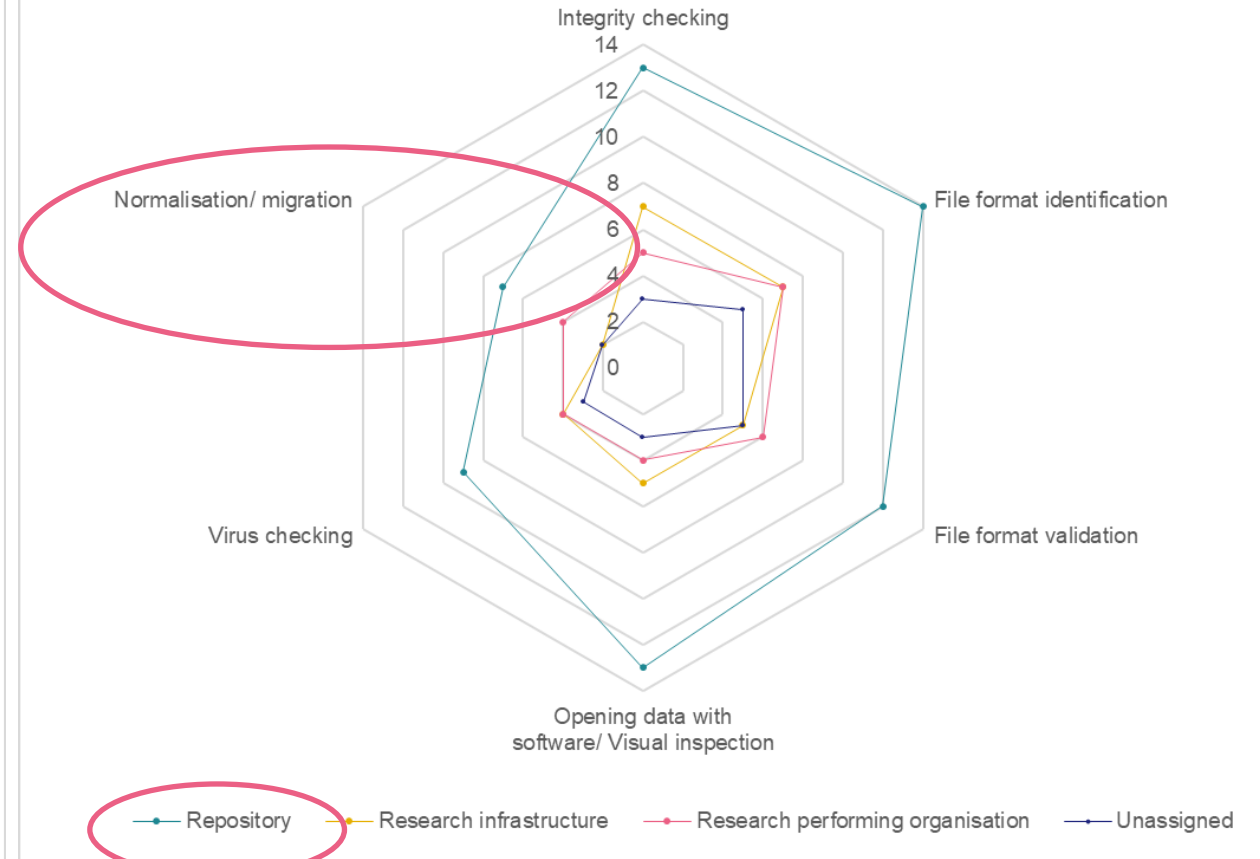
z.B. Authentizität, Einzigartigkeit, Verlässlichkeit, Vollständigkeit

Arten technischer Qualitätsprüfungen

Which assessments do you undertake to assess the technical quality of the data? (I)

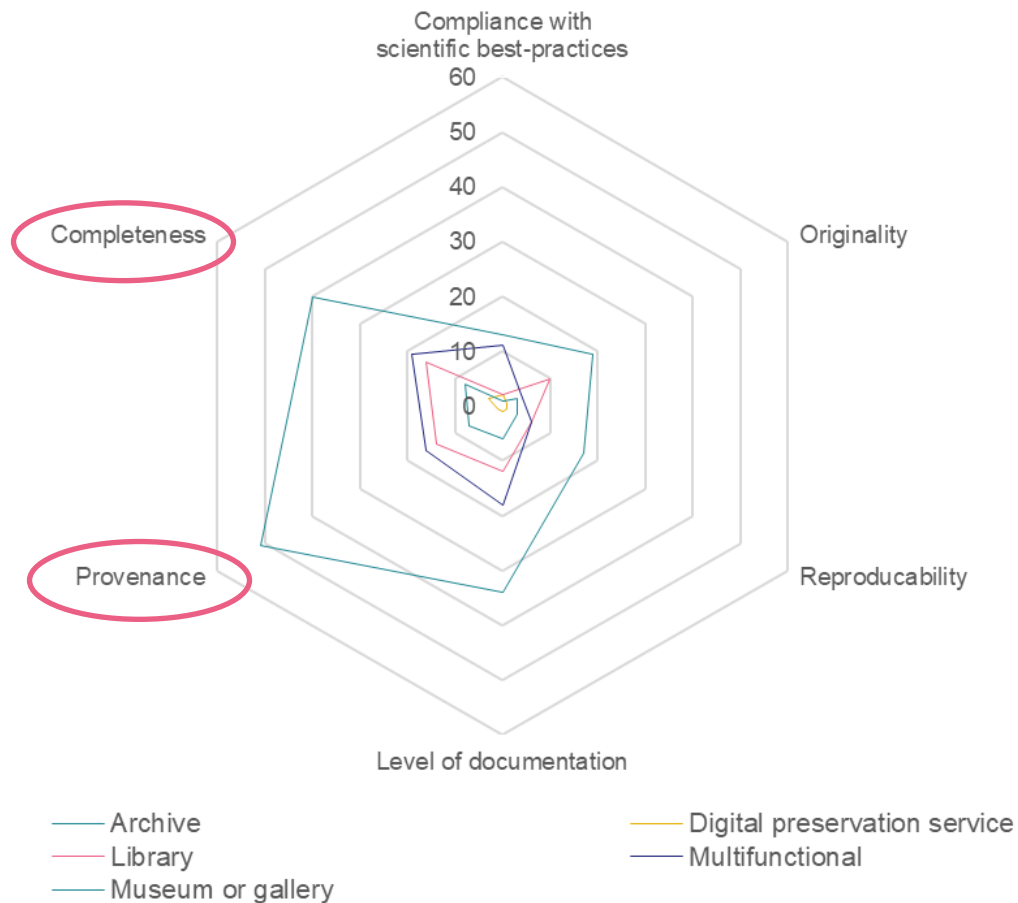


Which assessments do you undertake to assess the technical quality of the data? (II)

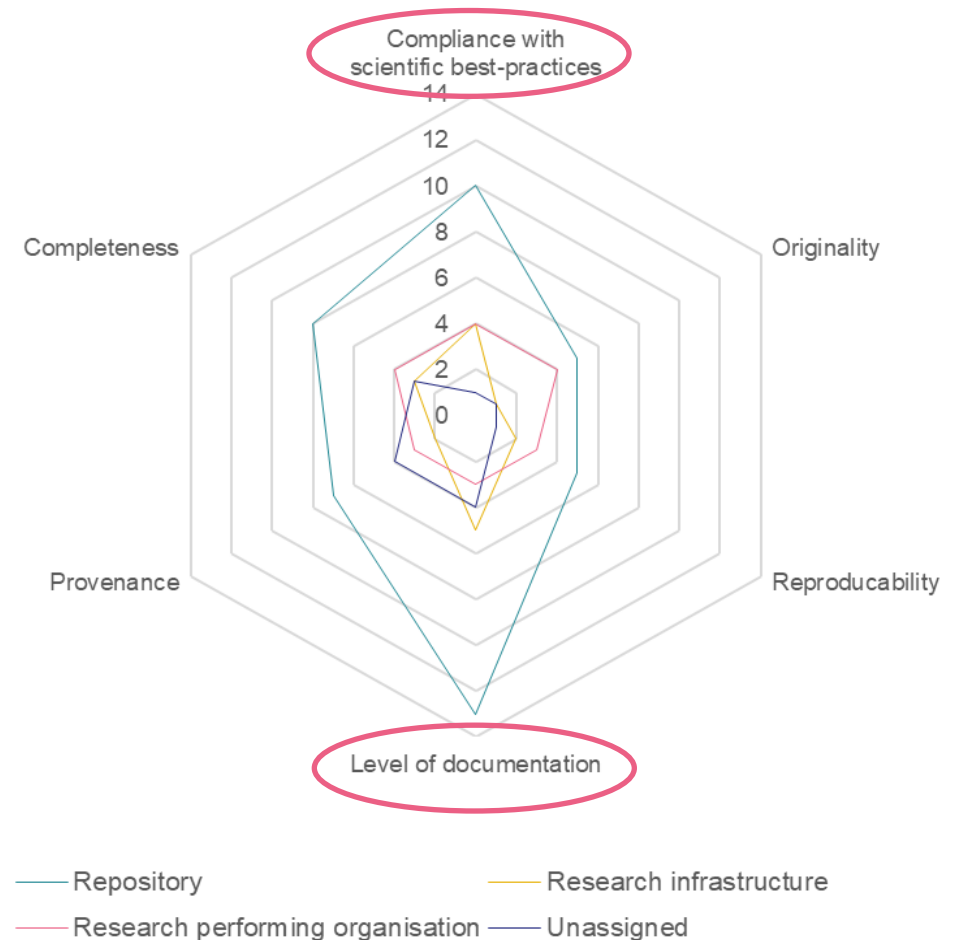


Arten inhaltlicher Qualitätsüberprüfungen

Which quality assessments do you undertake to assess the content/information quality of the data?

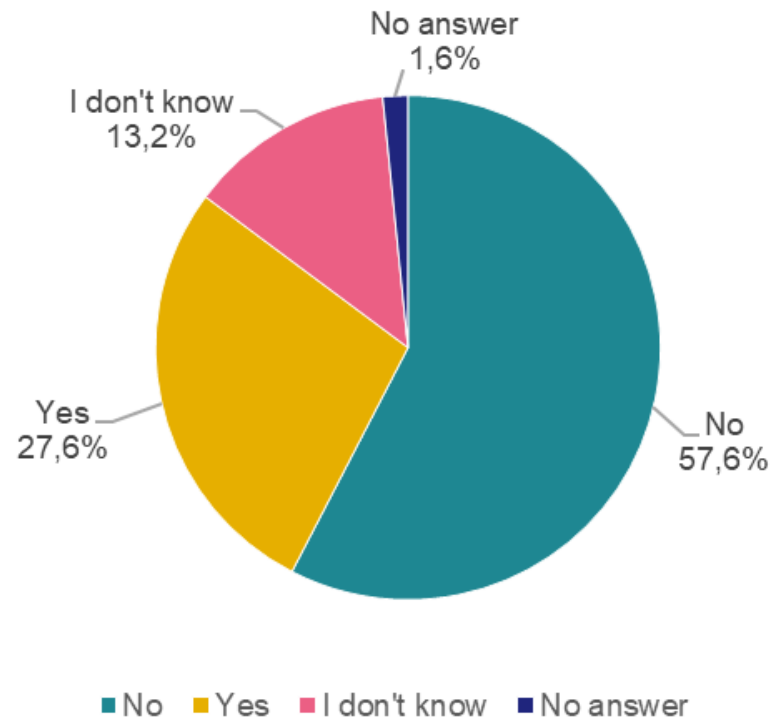


Which quality assessments do you undertake to assess the content/information quality of the data?



Einmal überprüft, immer valide

Do you regularly re-assess the quality of your preserved digital data objects?



Falls „ja“ – Folgefrage bzgl. der Frequenz von Neubewertungen:

- Jährlich (n=15)
- Verschieden (n=14)
- Bei Bedarf (n=7)
- Monatlich (n=5)
- Alle 5-9 Jahre (n=4)
- Alle 10-19 Jahre (n=3)
- Regelmäßig / täglich (n=3)
- Quartalsbezogen (n=3)

• ...

Und es gibt noch so viel mehr

Weitere Themen der Auswertung beinhalten

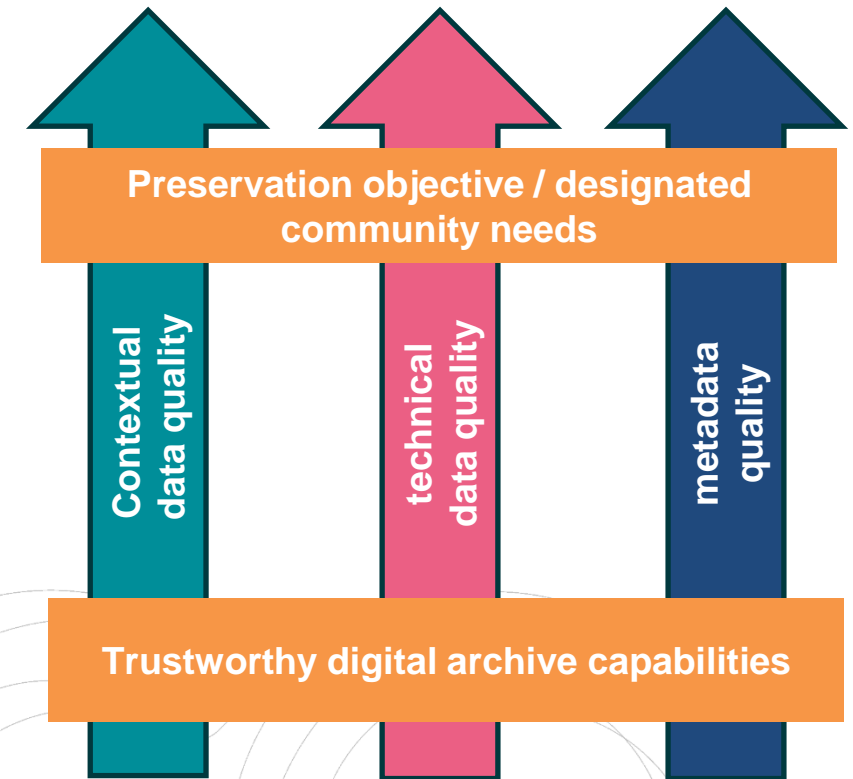
- Welche deskriptiven Metadaten-Standards werden genutzt?
- Existiert eine Liste bevorzugter Dateiformate?
- Bitte nennen Sie die am meisten vertretenen Dateiformate in Ihrem Archiv
- Hat Ihre Designated Community Bedarfe, denen aktuell nicht begegnet werden kann

And the countless possibilities in combining the data to all the questions!

Full report (D1.1 Overview of Identification, Selection and Appraisal Practices for Long-Term Preservation of Data): <https://doi.org/10.5281/zenodo.17984752>

Und wie geht es weiter?

- Aktuell laufen Vorbereitungen für Folgeinterviews
- Angaben zu Qualitätsprüfungen und Collection Policies wurden in laufenden Arbeiten zu *M1.2 Definition of Metrics for Re-Use Fitness of Information Objects* berücksichtigt
- *M1.2 Definition of Metrics for Re-Use Fitness of Information Objects* → Juli 2026
- T1.3 Re-Appraisal start im Juli 2026 und baut auf Ergebnissen auf
- WP3: Umfrage insb. unter Forschenden und "Repository Personnel" zu Qualitätsmetriken





eden-fidelis.eu



linkedin.com/company/eosc-eden



[@eosc-eden.bsky.social](https://eosc-eden.bsky.social)



[@EOSC-EDEN](https://www.youtube.com/@EOSC-EDEN)

Thank You



micky.lindlar@tib.eu



digipres.club/@mickylindlar



**Funded by
the European Union**

08/05/2026 by Micky Lindlar, TIB

EOSC EDEN Grant agreement ID: 101188015

