

Das EOSC EDEN Re-Use Fitness Framework

Was, wie, für wen und warum archivieren?

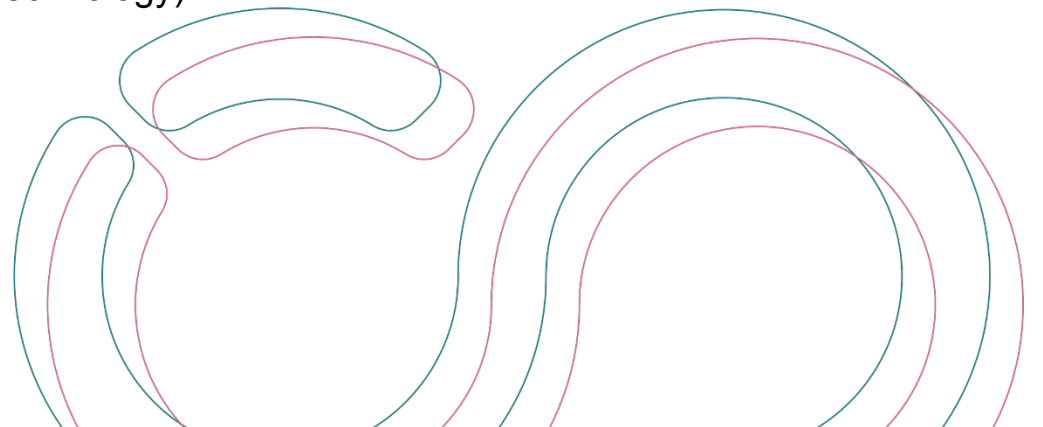
Micky Lindlar (they/them), TIB - Leibniz Information Centre for Science and Technology)
EOSC EDEN WP1 Lead



**Funded by
the European Union**

Micky Lindlar, TIB
DataWeek
01. Juni 2026, Leipzig

EOSC EDEN Grant agreement ID: 101188015



Technische Informationsbibliothek (TIB) – Leibniz-Informationszentrum Technik und Naturwissenschaften und Universitätsbibliothek

- Deutsche Zentrale Fachbibliothek für Technik und Naturwissenschaften
- Informationszentrum für die Digitalisierung von Wissenschaft und Technik
- Forschende Bibliothek
- Seit 2009 in der digitalen Langzeitarchivierung aktiv
- Seit 2017 Digitale Langzeitarchivierung als Dienstleistung
- In EOSC EDEN: Leitung des WP1 „Data and Process Framework for Long-Term Digital Preservation“



TIB Beratung zur LZA:

<https://www.tib.eu/de/lernen-arbeiten/information-und-d-beratung/tibgefragt/langzeitarchivierung-als-dienstleistung>

TIB LZA-as-a-Service:

<https://www.tib.eu/de/services/preservation-as-a-service>

TIB LZA Dokumentation:

<https://wiki.tib.eu/confluence/display/lza/Digitale+Langzeitarchivierung+an+der+TIB>

Was ist EOSC?

Vision

EOSC ist das europäische Netzwerk für FAIRe Daten & damit verbundenen Diensten für die Forschung

Leicht zu findende, zugängliche, interoperable und wiederverwendbare Forschungsdaten (FAIR)
Zuverlässige und nachhaltige Forschungsergebnisse sind innerhalb und über wissenschaftliche Disziplinen hinweg verfügbar

Heruntergebrochen in verschiedene Dinge unter dem Label EOSC

Federation

Implementiert
verbundenen
Knoten & Services

Association

256 europ.
Mitgliedsinstitutionen &
Beobachter

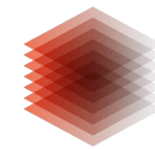
... und noch viel
mehr ...

Projekte

24 laufende, EU
geförderte Projekte

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
Partner	16 Institutionen aus 9 verschiedenen Ländern: Finnland (1), Deutschland (3), Niederlande (3), Norwegen (1), Belgien (1), UK (2), Schweiz (2), Frankreich (1), Polen (1)



Was ist „Langzeit“?

- „Langzeit ist keine Garantieerklärung über fünf oder fünfzig Jahre, sondern die verantwortliche Entwicklung von Strategien, die den beständigen, vom Informationsmarkt verursachten Wandel bewältigen können.“ – nestor Handbuch (2010)

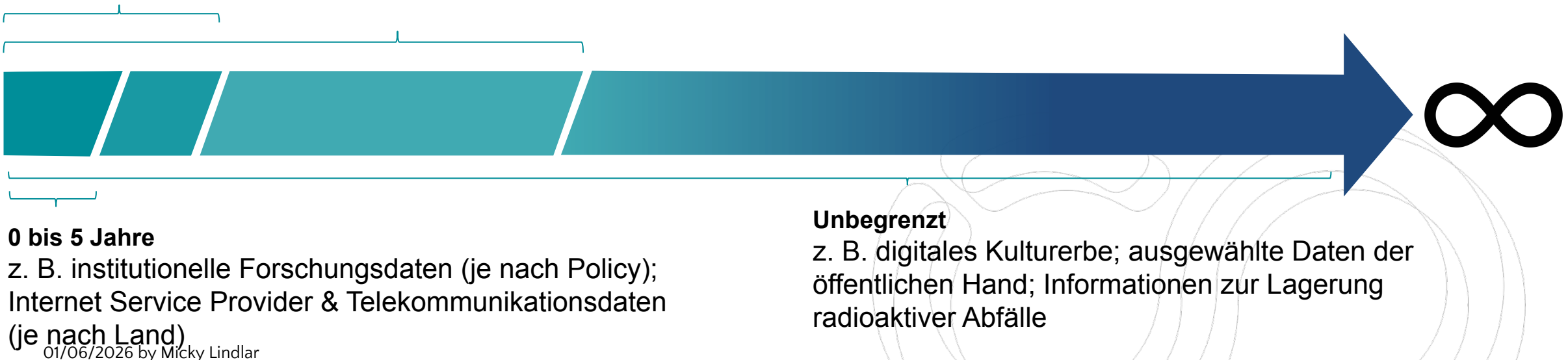
10 Jahre

z. B. Forschungsdaten
aus Drittmittelprojekten;
elektronische Rechnungen

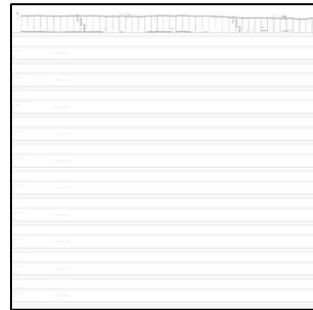
25 Jahre

z. B. Daten aus medizinischen Studien;
Fertigungs- und Konstruktionsdaten im Zusammenhang
mit der Einhaltung von Sicherheits- und Umweltnormen

Wann fängt „Langzeit“ an?



(Langzeit-)Speicherung: ein planbares Risiko?



10 Jahre

25 Jahre



Digitale Daten sind ab dem Zeitpunkt ihrer Entstehung Risiken ausgesetzt!

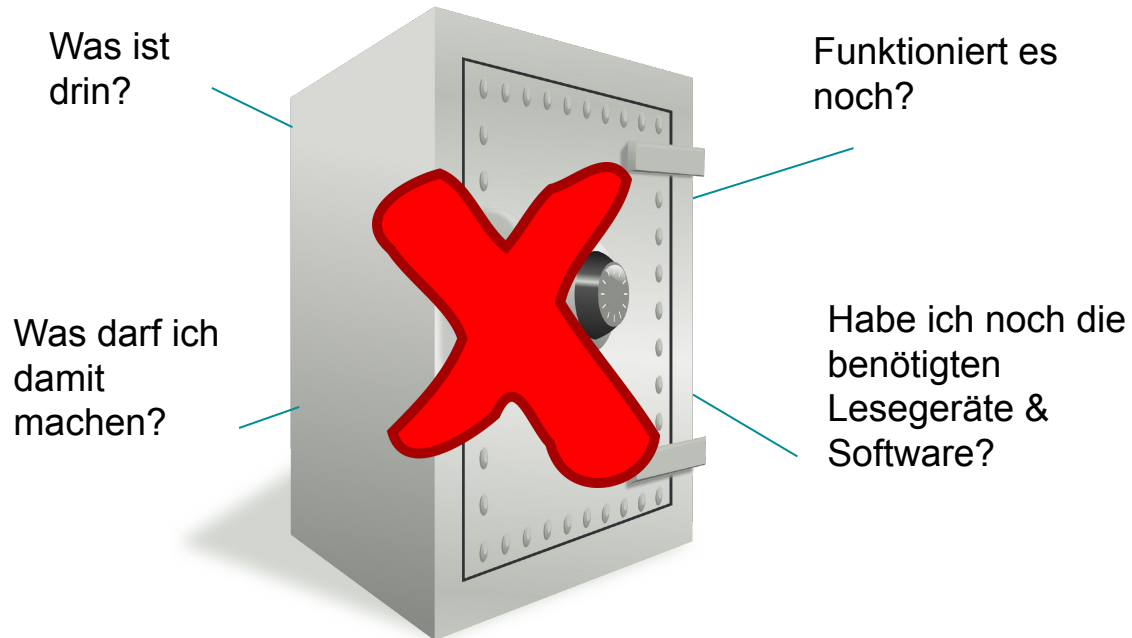


0 bis 5 Jahre

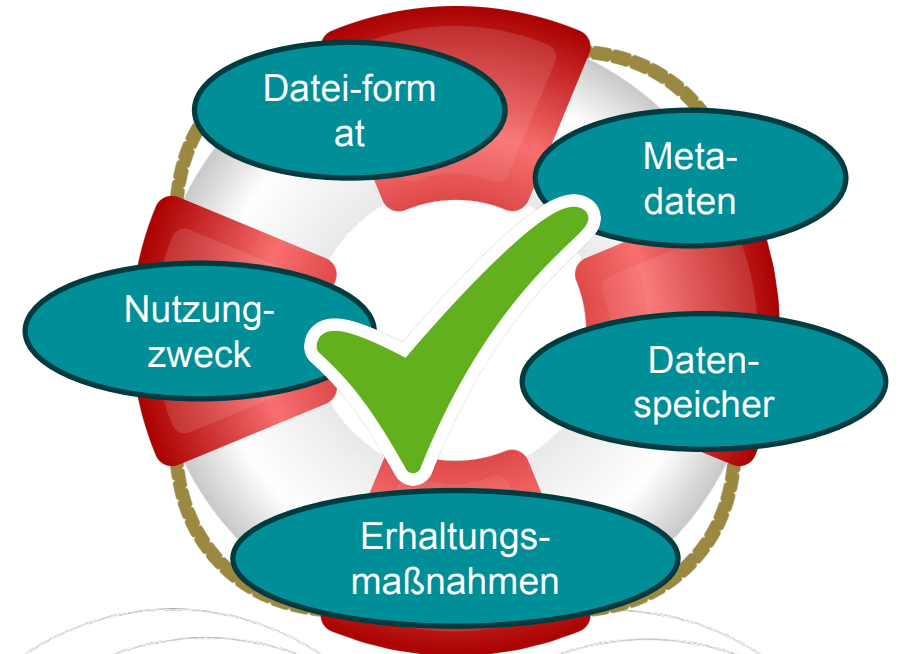
$$\frac{1}{\Gamma(s)} P_0! \, {}^n s) r^{-3} (s + \sqrt{1!} s^2 ds = \frac{1}{\Gamma(s)} P_0! \, {}^n s) r^{-3} (s! \sqrt{1!} s^2 ds \quad (6)$$



Was ist digitale Langzeitarchivierung?



Nur das Speichern von Daten im „Tresor“ / (managed) Storage?



Aktivitäten entlang des Lebenszyklus?

Was kann da schon passieren? Fallbeispiele aus der Praxis

$$\int_0^1 \frac{1}{x} P_0! "s) r^3 (s + \$\# \sqrt{1! s^2} ds = \int_0^1 \frac{1}{x} P_0! "s) r^3 (s! \$\# \sqrt{1! s^2} ds$$

PDF

TIFF



Foto: Polizei Hamburg/hfr

write_record_fields(template) and template names according to attribute names (e.g. @test_data@, field_*, test_data_*, template).

4.1 Generated program with student's example

Generated program consists of several parts:

- identification part (student's ID, surname and name; in a form of comments);
- program code in C++ that includes function which generates file with test data to be entered and
- student exercises (in form of comments).

The whole student session with generation system is shown by Use-case diagram in Figure 7.

Figure 5: Example of main program code template

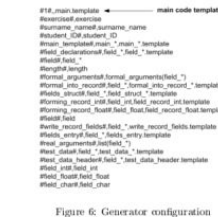


Figure 6: Generator configuration

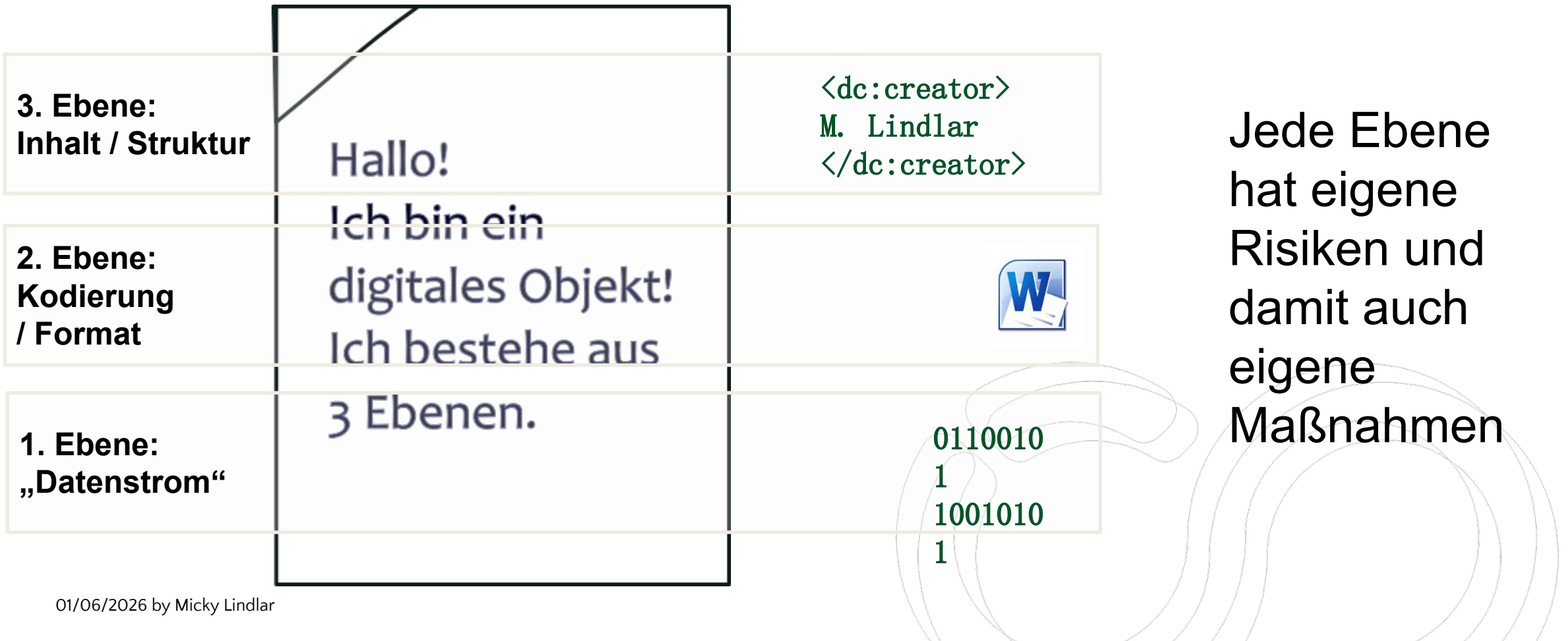
There are also some special possibilities like specifying attribute groups by asterisk (e.g. @write_record_fields@, field_*,

PDF

Figure 7: Use-case diagram of the system

Student enters his/her identification ID, surname and name in the generation system, which generates exercises together with exercises. Student compiles it on a local computer and runs it on a line compiler. The program generates test data. Student does the exercises and makes the resulting document which

Drei Ebenen eines digitalen Objekts

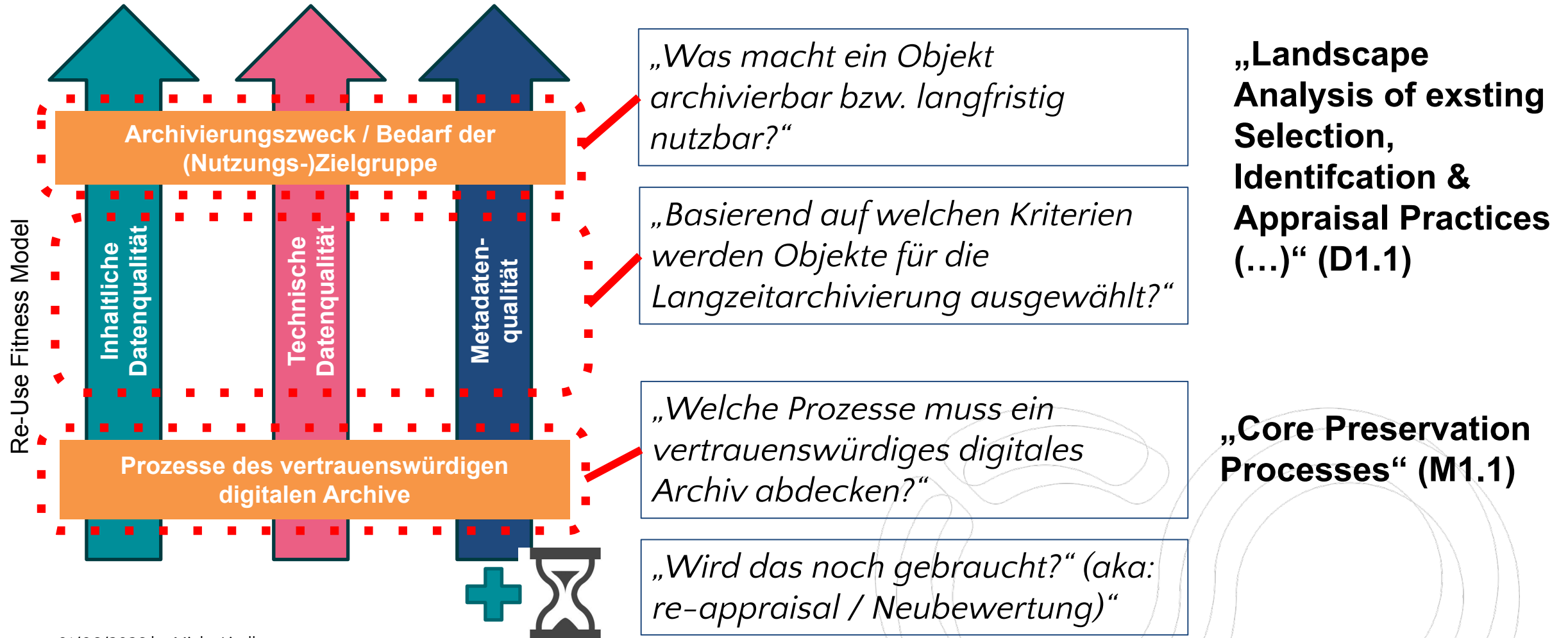


Warum EOSC EDEN?

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)



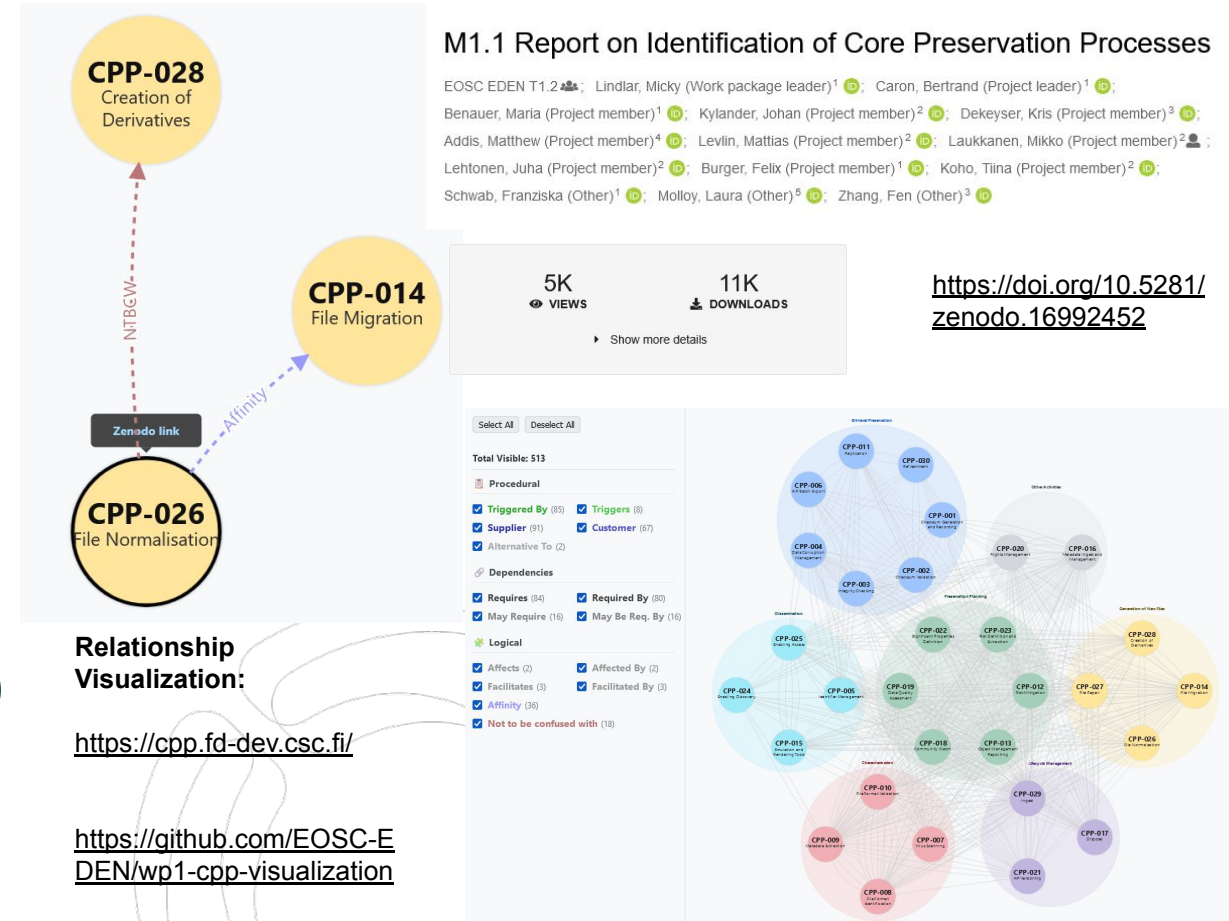
Core Preservation Processes (CPPs) – M1.1

Kernprozesse der digitalen Langzeitarchivierung =
Operative Prozesse, die jedes Langzeitarchiv ausführen sollte (wahlweise selbst oder durch einen externen Service)

M1.1 Report on Identification of Core Preservation Processes

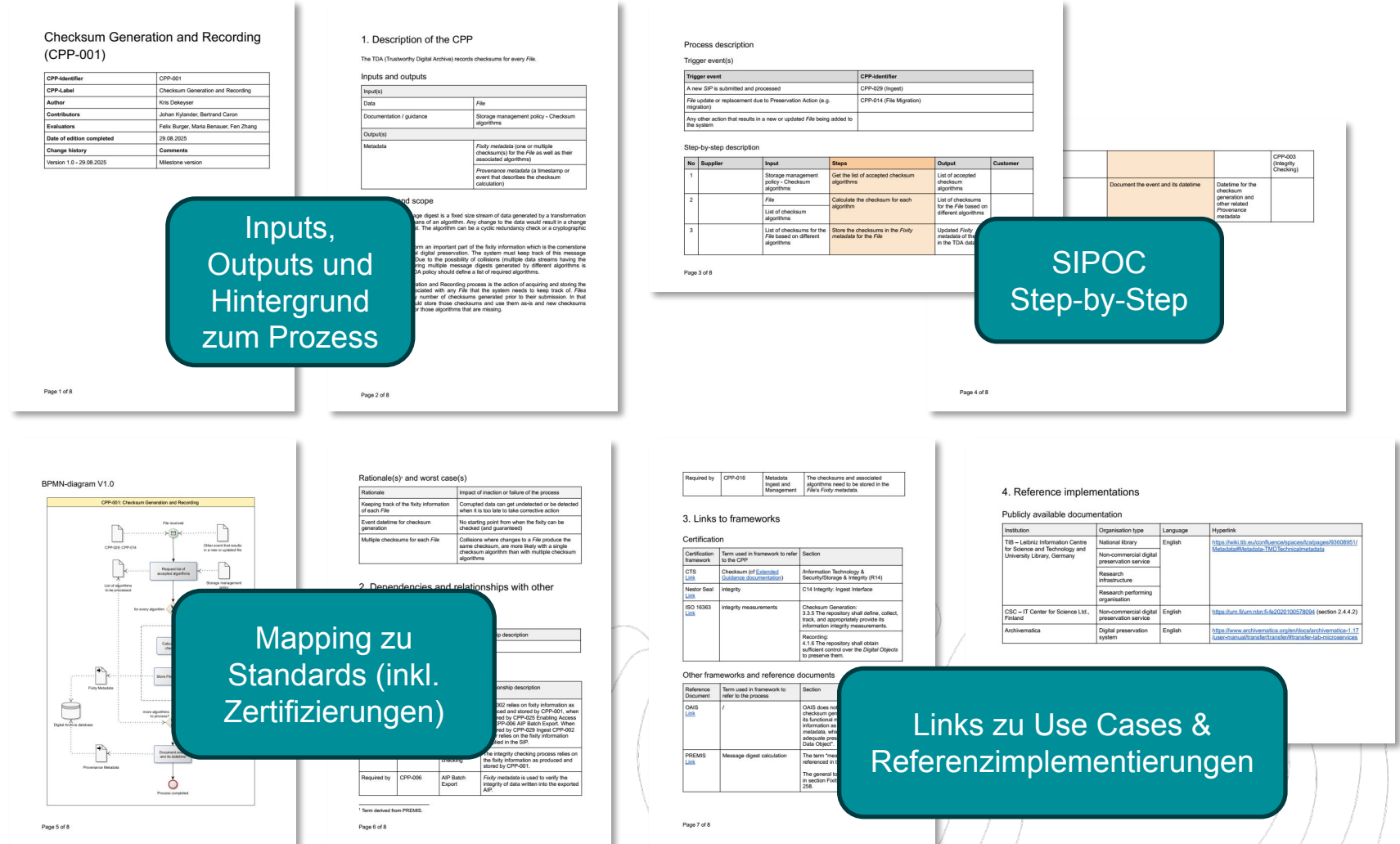
- 30 CPP Beschreibungen
- Glossar
- Design & Guidance Report
- Visualisierungstool

Die CPPs sind kein Archiv, sie sind eher „Anleitungen“ für ein Langzeitarchiv



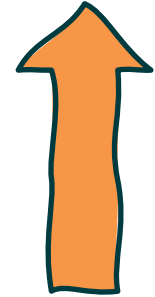
Core Preservation Processes (CPPs) – M1.1

- CPP-001 Checksum Generation and Recording
- CPP-002 Checksum Validation
- CPP-003 Integrity Checking
- CPP-004 Data Corruption Management
- CPP-005 Identifier Management
- CPP-006 AIP Batch Export
- CPP-007 Virus Scanning
- CPP-008 File Format Identification
- CPP-009 Metadata Extraction
- CPP-010 File Format Validation
- CPP-011 Replication
- CPP-012 Risk Mitigation
- CPP-013 Object Management Reporting
- CPP-014 File Migration
- CPP-015 Emulation and Rendering Tools
- CPP-016 Metadata Ingest and Management
- CPP-017 Disposal
- CPP-018 Community Watch
- CPP-019 Data Quality Assessment
- CPP-020 Rights Management
- CPP-021 AIP Versioning
- CPP-022 Significant Properties Definition
- CPP-023 Risk Definition and Extraction
- CPP-024 Enabling Discovery
- CPP-025 Enabling Access
- CPP-026 File Normalisation
- CPP-027 File Repair
- CPP-028 Creation of Derivatives
- CPP-029 Ingest
- CPP-030 Refreshment



Umfrage Landschaftsanalyse

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



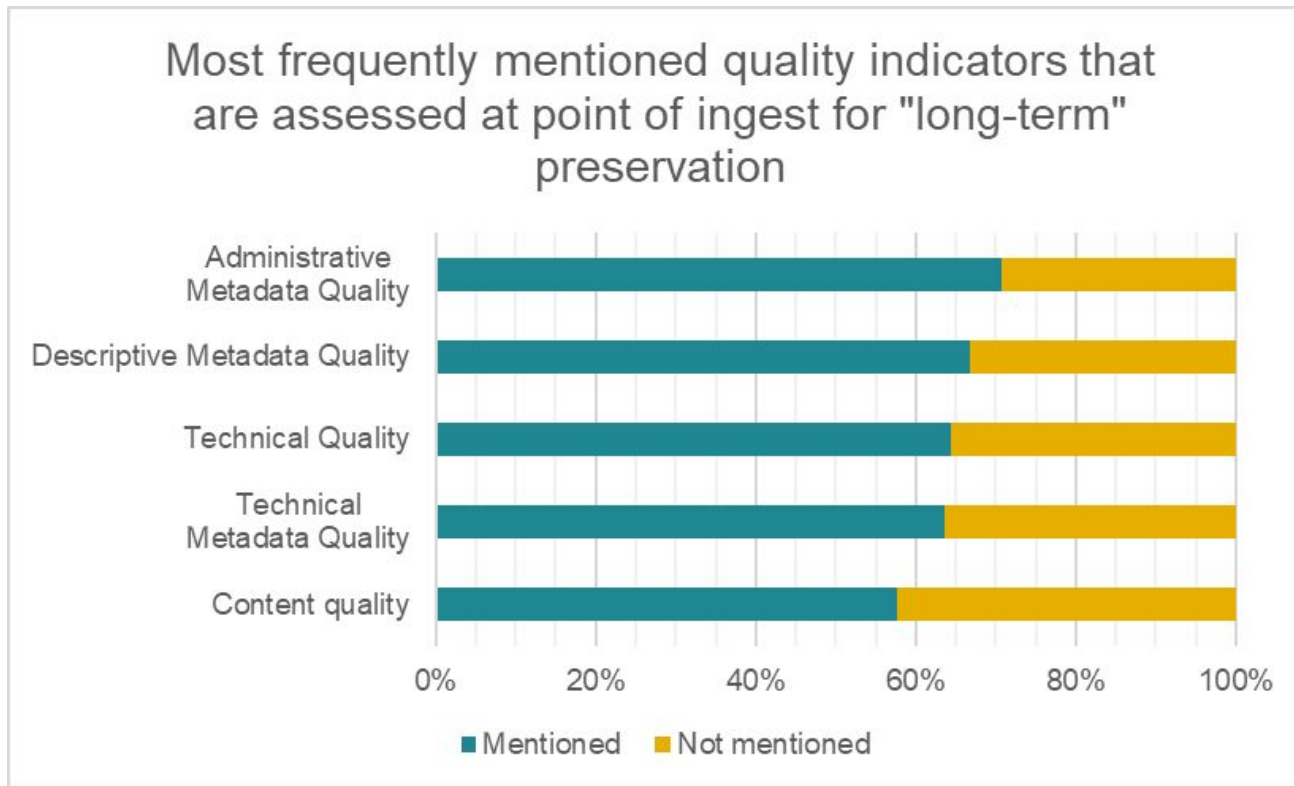
- Umfrage lief vom 2. Mai – 30. Juni 2025 im EU Survey Tool
- Gerichtet an Langzeitarchivierungs-Praktiker*innen aus verschiedenen Organisationstypen und wissenschaftlichen Disziplinen, inkl. Forschung, Kulturerbe, Industrie
- 250 valide Antworten
- Umfragenaufbau (20 top-line Fragen):
 1. Infos zur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

Umfragenaufbau / -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

Arten von Qualitätsüberprüfungen in Deposit/Ingest-Phase

86,8% gab an, die Qualität der Daten zum Zeitpunkt der Aufnahme für die langfristige Archivierung zu überprüfen



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

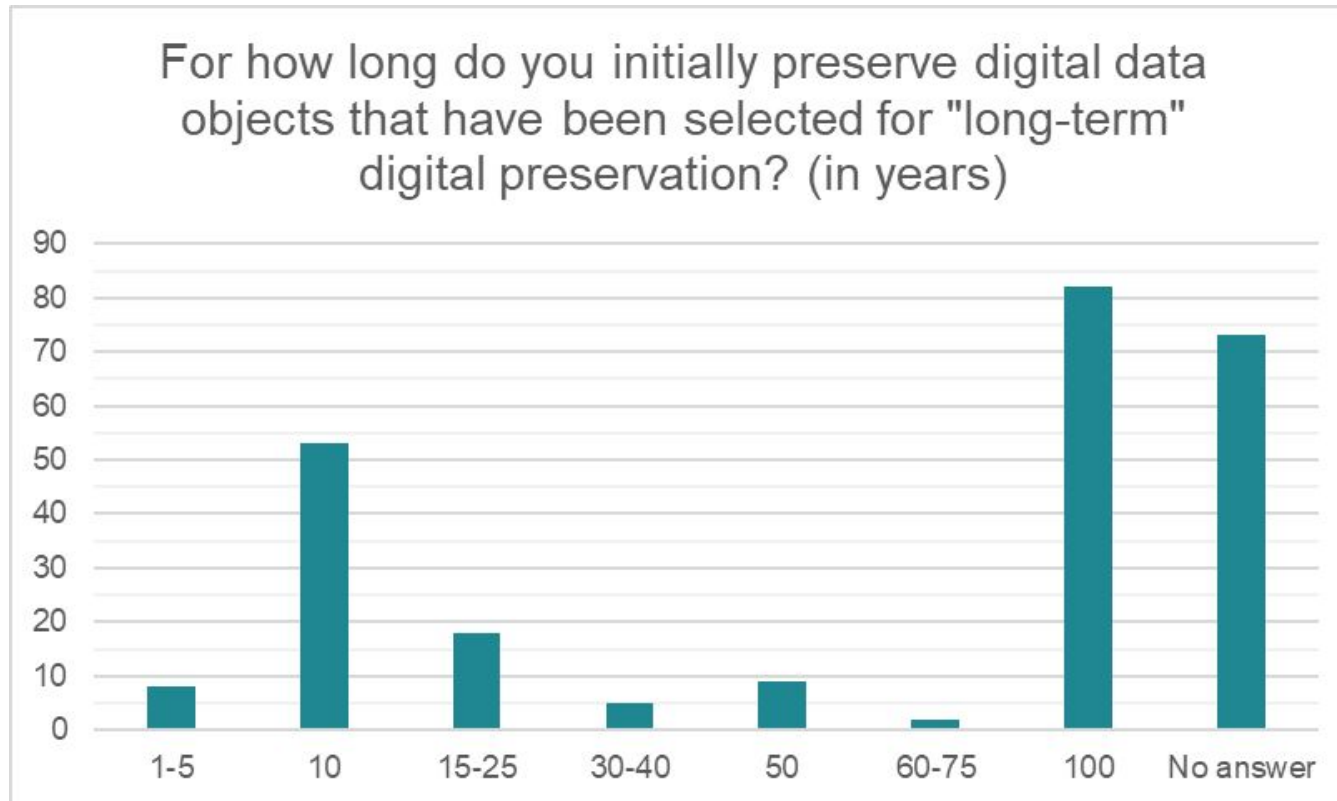
Qualität technischer Metadaten

z.B. Format, Auflösung, MD5 Checksumme

Inhaltliche Qualität

z.B. Authentizität, Einzigartigkeit, Verlässlichkeit, Vollständigkeit

Wie lange werden Objekte initial archiviert?



Langzeit liegt also
irgendwo zwischen 10
und 100 Jahren 😊



... 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%

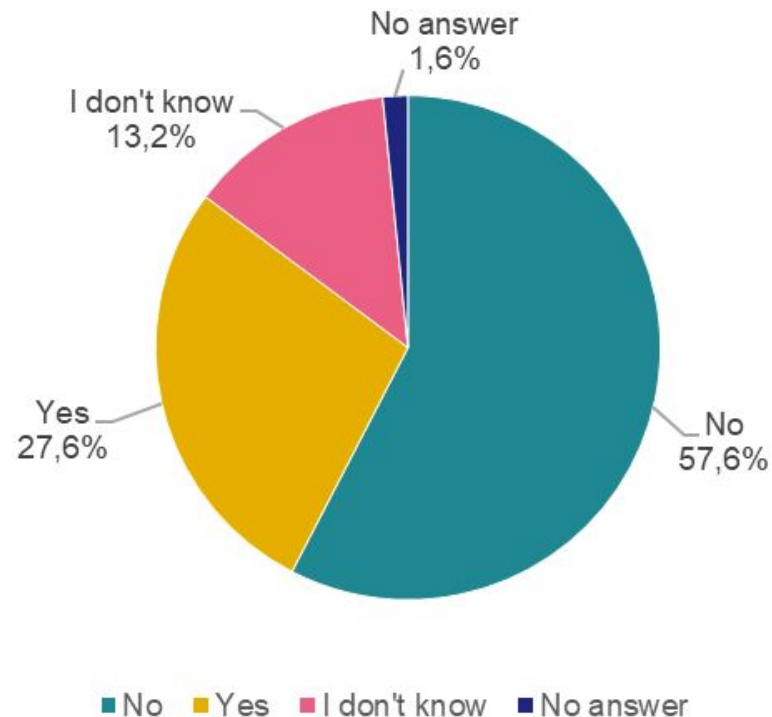
Andere (Auswahl)

- 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)



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)
- ...

Weitere Themen aus der Umfrage (Auswahl)

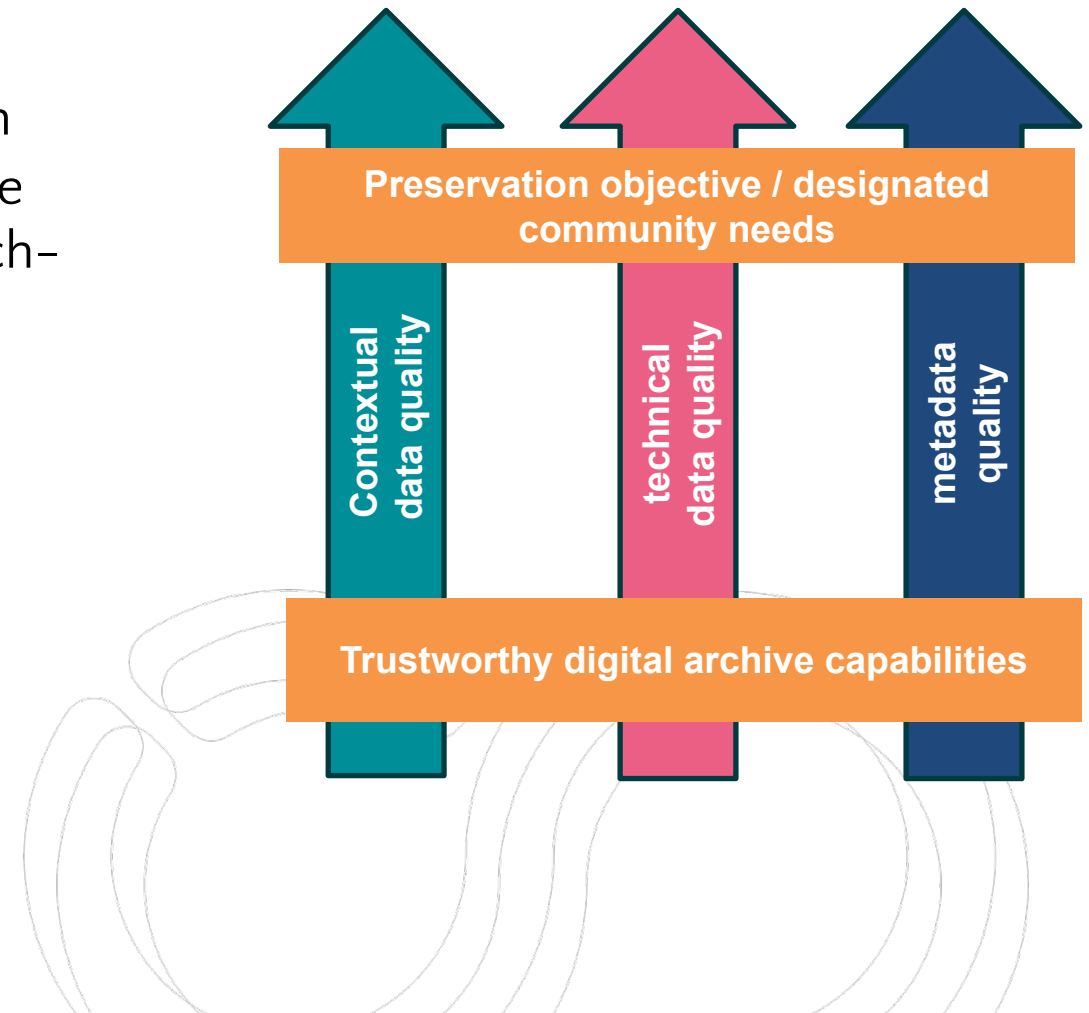
- Welche Standards- und Guidelines werden für die Auswahl von digitalen Objekten für die Langzeitarchivierung genutzt?
- Welche deskriptiven Metadaten-Standards werden genutzt?
- Existiert eine Liste bevorzugter Dateiformate?
- Hat Ihre Designated Community Bedarfe, denen aktuell nicht begegnet werden kann?

Für den gesamten Report (in englischer Sprache), siehe *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 im Projekt weiter?

- Zur Fortschreibung der Landschaftsanalyse werden aktuell Interviews durchgeführt (Ziel: tiefer gehende Informationen darüber, wie Qualitätsanalysen durchgeführt werden)
- Parallel läuft die Definition von Indikatoren für die Qualitätssäulen (*M1.2 Definition of Metrics for Re-Use Fitness of Information Objects*)
- T1.3 *Re-Appraisal* start im Juli 2026 und baut auf Ergebnissen auf; Ziel: Entwurf eines Prozesses für die Neu-Bewertung von Daten im Archiv





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**

01/06/2026 by Micky Lindlar, TIB

EOSC EDEN Grant agreement ID: 101188015

