



Trends in der digitalen Archivierung

docuteam Community Day-Webinare, 21.10.2020
Christian Moser, Andreas Steigmeier, Tobias Wildi

Die Referenten



Christian Moser

Beratungsprojekte,
Umsetzung von digitalen
Archiven

Cloudlösung docuteam
cosmos



Andreas Steigmeier

Informationsmanagement
GEVER

Digitale Ablieferungen ans
Archiv

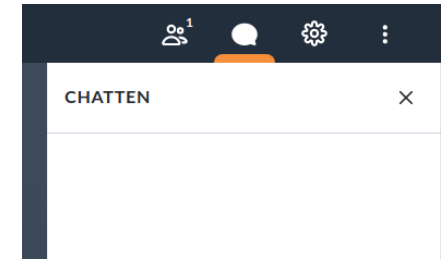


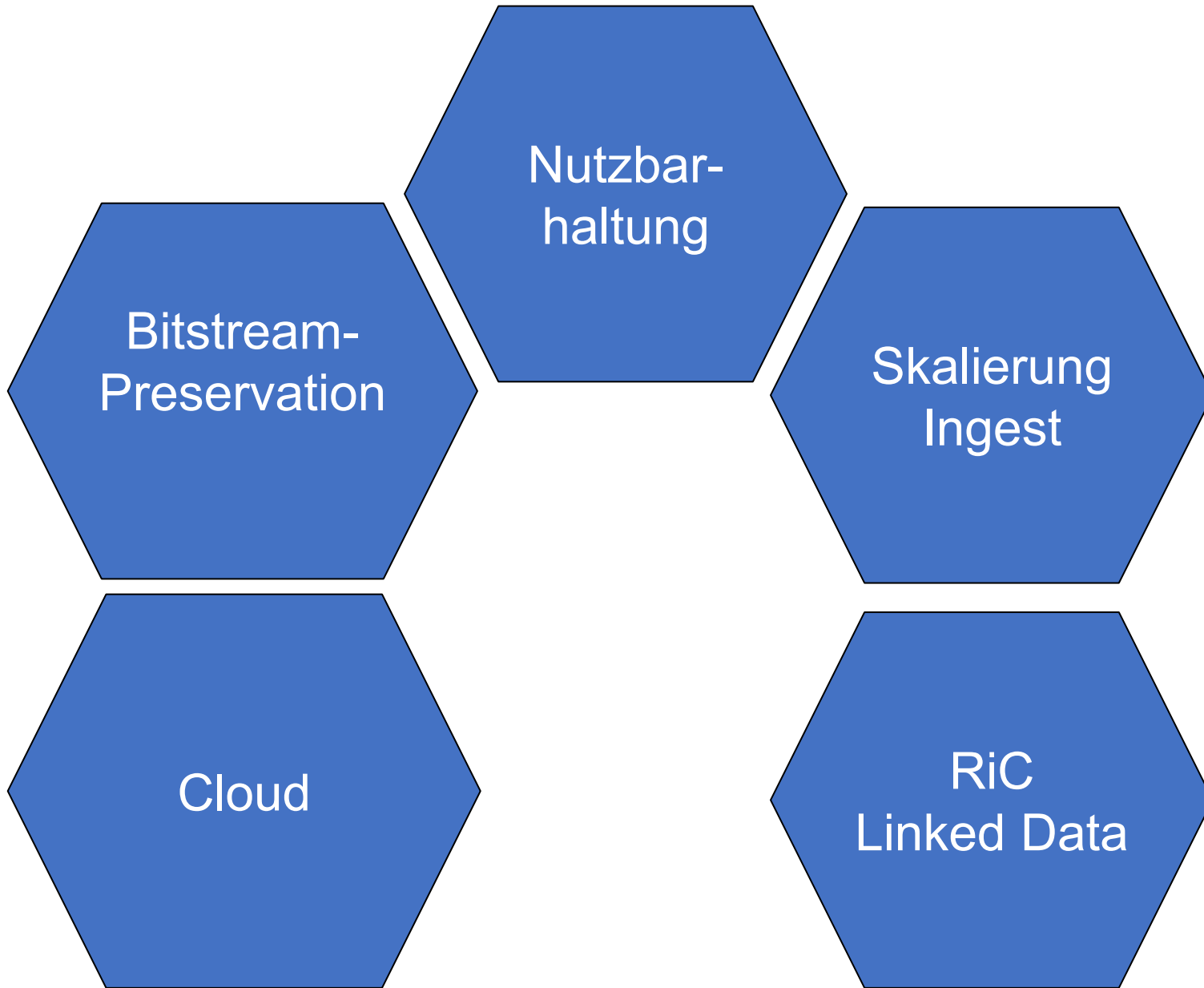
Tobias Wildi

Metadatenstandards
Infrastrukturen für digitale
Archivierung
Schutz von digitalen
Kulturgütern

Haben Sie Fragen?

- Stellen Sie uns rechts im Panel schriftlich Fragen, wir gehen während der Präsentation oder am Ende darauf ein.
- Den Link auf die Präsentation stellen wir Ihnen in der Folge zu.
- Wir stehen Ihnen später für individuelle Besprechungen bereit – per Telefon, Videokonferenz oder auch wieder persönlich





Trend 1: Archivieren in der Cloud

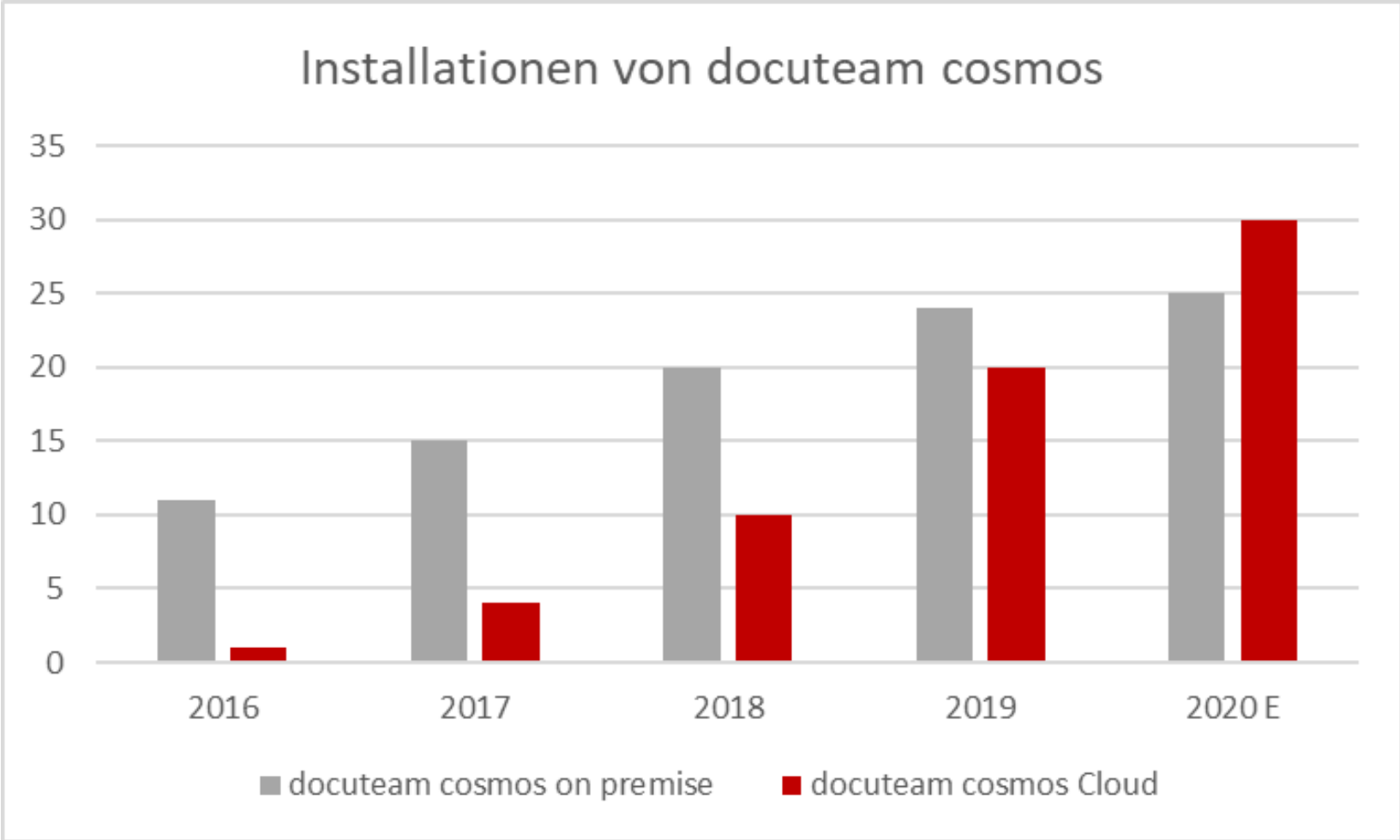
Wir beobachten, wie sich die Nachfrage für digitale Langzeitarchivierung weg von In-House-Infrastrukturen zu Cloudlösungen entwickelt. Folgende Voraussetzungen müssen aber gegeben sein:

- die Daten verbleiben in der Schweiz
- die Cloud in Schweizer Eigentümerschaft ist

Auch bestehende Kunden, die unsere Software On Premise betreiben, überlegen sich den Wechsel auf die Cloudlösung oder haben ihn 2020 durchgeführt. Dazu gehören auch öffentliche Archive.

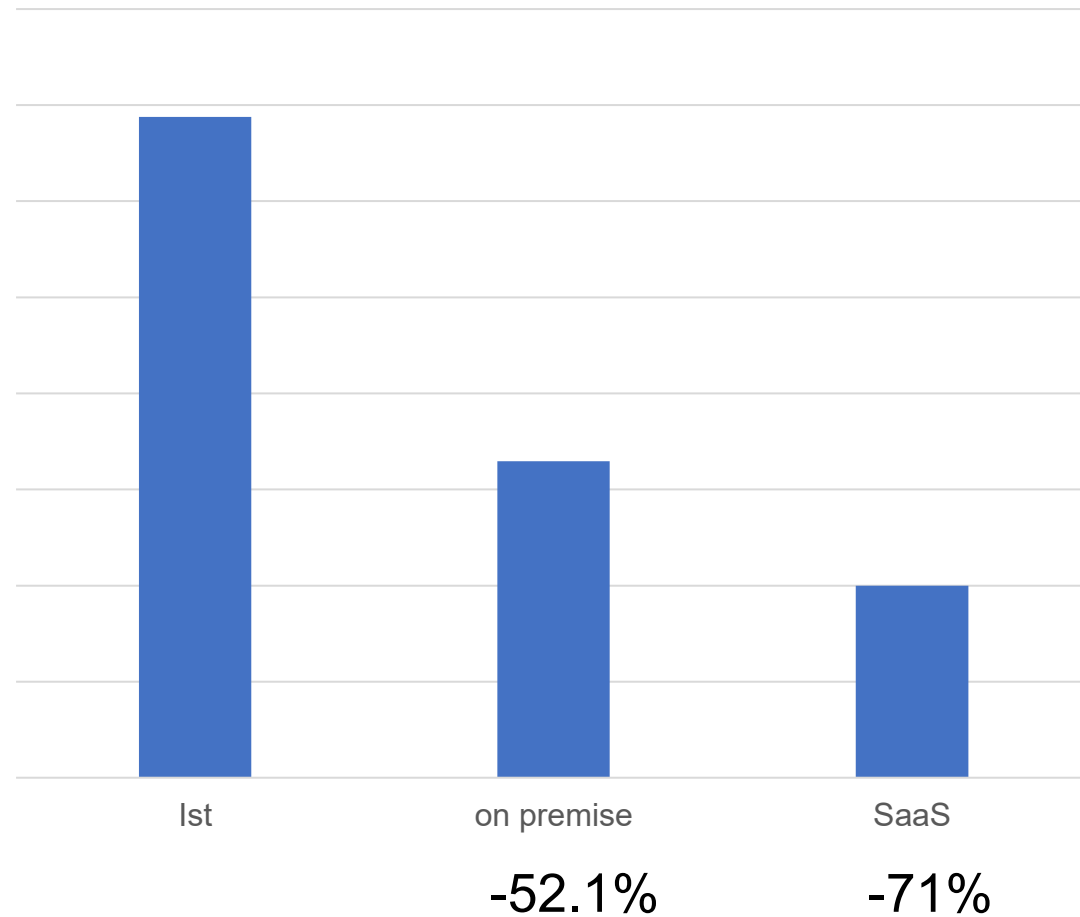
Dank Cloud können wir digitale Archive auch für kleine Institutionen als Commodity anbieten (ab CHF 4000 pro Jahr inkl. 500 GB Speicher).

Instanzen von docuteam cosmos



Senkung der Betriebskosten

Reales Beispiel
2020



docuteam cosmos basiert auf SWITCHengines



- Unsere Lösung docuteam cosmos basiert komplett auf **SWITCHengines**, der Cloudinfrastruktur von SWITCH (<https://www.switch.ch/>).
- **SWITCH** wurde 1987 als **privatrechtliche, nicht gewinnorientierte Stiftung** vom Bund und den damals acht Hochschulkantonen gegründet.
- Die Infrastruktur von SWITCHengines (<https://www.switch.ch/engines/>) läuft in zwei SWITCH-eigenen **Rechenzentren in Zürich und Lausanne**.
- Der gesamte **Technologie-Stack** ist bekannt und dokumentiert, das schafft Vertrauen.
- SWITCH bietet **Verträge mit langen Kündigungsfristen und Laufzeiten** an. Kurze Kündigungsfristen stellen bei anderen Cloudanbietern oft ein Problem für Archivanwendungen dar.
- <https://www.switch.ch/de/engines/>

Trend 2: Records in Contexts und Linked Data



Expert Group on Archival Description

Groupe d'Experts sur la Description Archivistique

- docuteam ist an der Entwicklung von Records in Contexts (RiC) im Rahmen der EGAD beteiligt
- Ablösung der bestehenden ICA-Standards durch einen neuen, auf modernen Technologien beruhenden Standard
- Weg von Freitextfeldern, hin zu Verlinkung mit (bestehenden) Ressourcen

Vereinheitlichung der ICA-Standards

Warum ein neuer Standard?

- ISAD(G)/ISAAR(CPF)/ISDF/ISDIAH sind zeitlich nacheinander entstanden und wurden nicht aufeinander abgestimmt.
- Abgesehen von ISAD sind die Standards von den Archiven und den Softwareherstellern kaum aufgenommen worden.
- Darum erfolgt ein Neustart in Form eines einzigen Standards, mit dem alle Aspekte der Kontextualisierung von Archivgut umfassend abgebildet werden können.

Warum wird nicht ein bereits verstehender Standard verwendet und weiterentwickelt? Warum das Rad neu erfinden?

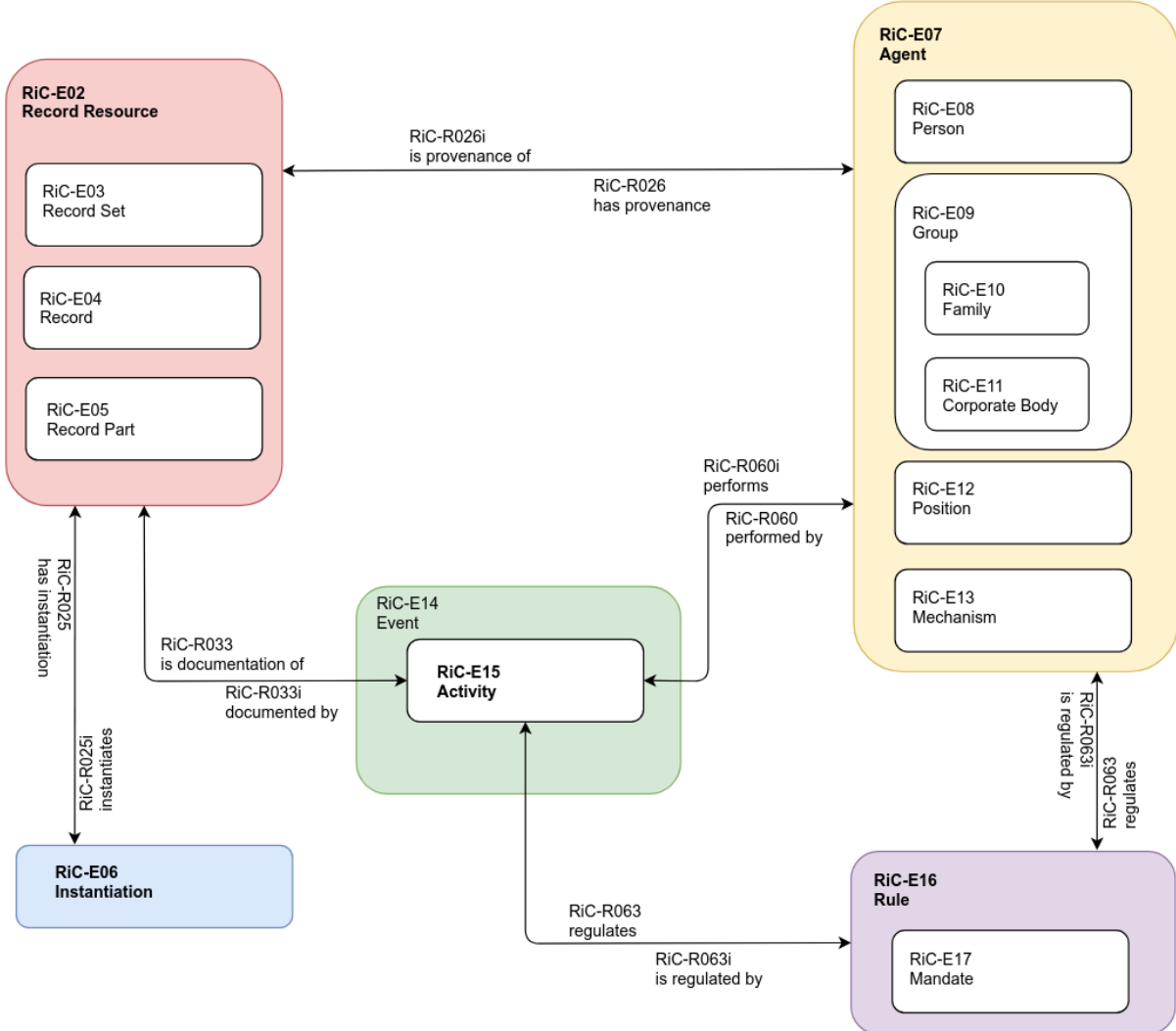
- Archive benötigen eine eigene Ontologie (Begriffssystem), um ihre spezifischen Semantiken abzubilden.
- Sie können nicht auf Modelle anderer Kultureinrichtungen zurückgreifen. CIDOC-CRM oder RDA gehen nicht.
- Zu einem späteren Zeitpunkt sollen jedoch Corsswalks zu anderen Standards und Communities definiert werden.

Records in **Contexts**

Von «multilevel» zu «multidimensional»

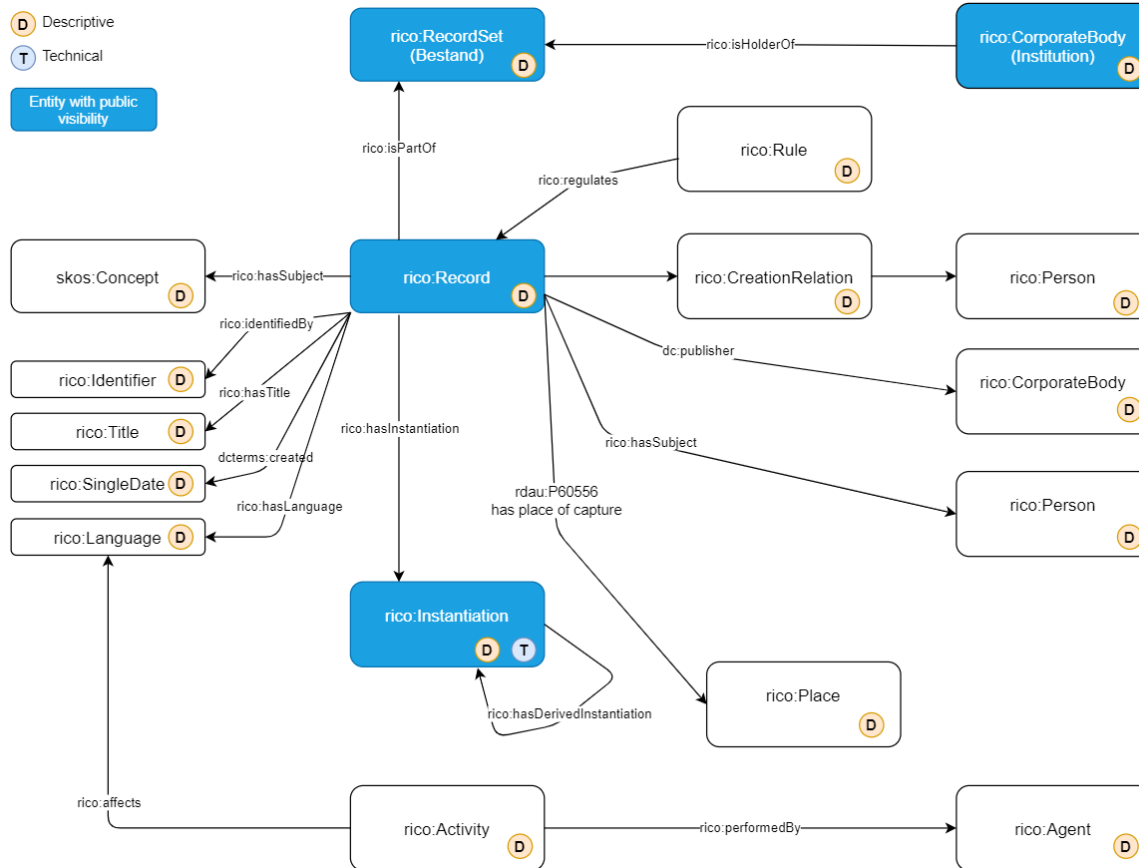
- **ISAD(G)** erlaubt heute eine "**multilevel description**" mit einem hierarchisch-klassifikatorischem Ansatz.
- **RiC** erweitert diesen Ansatz hin zu einer "**multidimensional description**", mit Menschen, Institutionen, ihren Aufgaben und Rechten, die auf mehrere Entitäten verteilt sein. Diese Entitäten können hierarchisch, horizontal oder wie auch immer miteinander verbunden sein.
- Dieser **erweiterte Provenienzbegriff** ist komplex zu modellieren („Context is complex“), weil er
 - nicht nur Personen oder Organisationen umfasst, die Registraturbildner sind
 - sondern auch andere Agents, die in den Erstellungsprozess einbezogen sind, Geschäftsaktivitäten, soziale/wirtschaftliche Kontexte.
 - und somit nicht in einer eindimensionalen Baumstruktur, sondern in einem Netzwerk dargestellt werden muss.

RiC-CM: Records in Contexts Conceptual Model



Konkrete Umsetzung: Memobase 2020

- Ziel: Ersatz der bestehenden Memobase-Datenbank
- Projekt im Auftrag von Memoriav, zusammen mit UB Basel und Outermedia
- Datenmodell: RiC und Ebucore



In der Schweiz: ENSEMEN-Projektgruppe im VSA: Neue Ziele seit Mai 2020

- ENSEMEN will keinen eigenen Standard für die Schweiz entwickeln, sondern die **Umsetzung von RiC in der Schweiz planen und unterstützen**
- Review von RiC mit direkten Rückmeldungen an die EGAD
- Kontrollierte Vokabularien für die Schweiz entwickeln
 - Record Set Types (Verzeichnungsstufen)
 - Rules (Rechte)
 - Activity Types (Kategorien von Aktivitäten, z. B. um die archivischen Prozesse zu dokumentieren)
- Experimentelle Umsetzungen von RiC in vorhandenen Systemen und Lösungen. Wie können die bestehenden Verzeichnungsmetadaten und auch Standards wie eCH-0160 und Matterhorn METS in RiC abgebildet werden?

Trend 3: Nutzbarhaltung

- Wir sprechen lieber von «Nutzbarhaltung» als von «Archivierung». Denn das Thema betrifft nicht nur die traditionellen Gedächtnisinstitutionen.
- Bidirektionale Schnittstellen zwischen Produktivsystem und Archiv
- Bedürfnis nach Access manifestiert sich jetzt. Das wird nicht durch einen AIS-Webclient abgedeckt.

Unsere Kernkompetenz als Archivarinnen und Archivare...

... ist im Grunde die digitale Nutzbarhaltung von Inhalten über grosse Zeiträume.

- Unser Referenzmodell ist OAIS.
- Eine der Funktionen in diesem Modell ist Preservation Planning.
- Wir müssen innerhalb eines OAIS in der Lage sein, Preservation Actions durchzuführen.

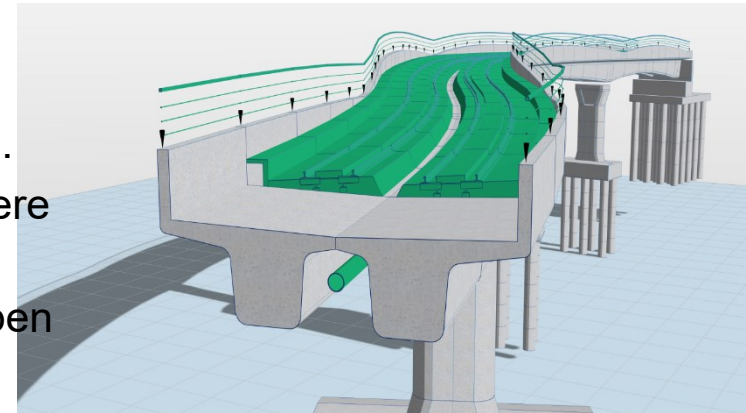
- Diese Funktionen besitzt kaum ein Produktivsystem.
- Hier haben Archive ein Alleinstellungsmerkmal.



Das Bedürfnis ist beträchtlich, wenn auch noch nicht erkannt

Viele Unterlagen können weder nach ein paar Jahren «abgeschlossen» noch nach zehn Jahren der Aufbewahrung an ein Archiv «abgeliefert» werden.

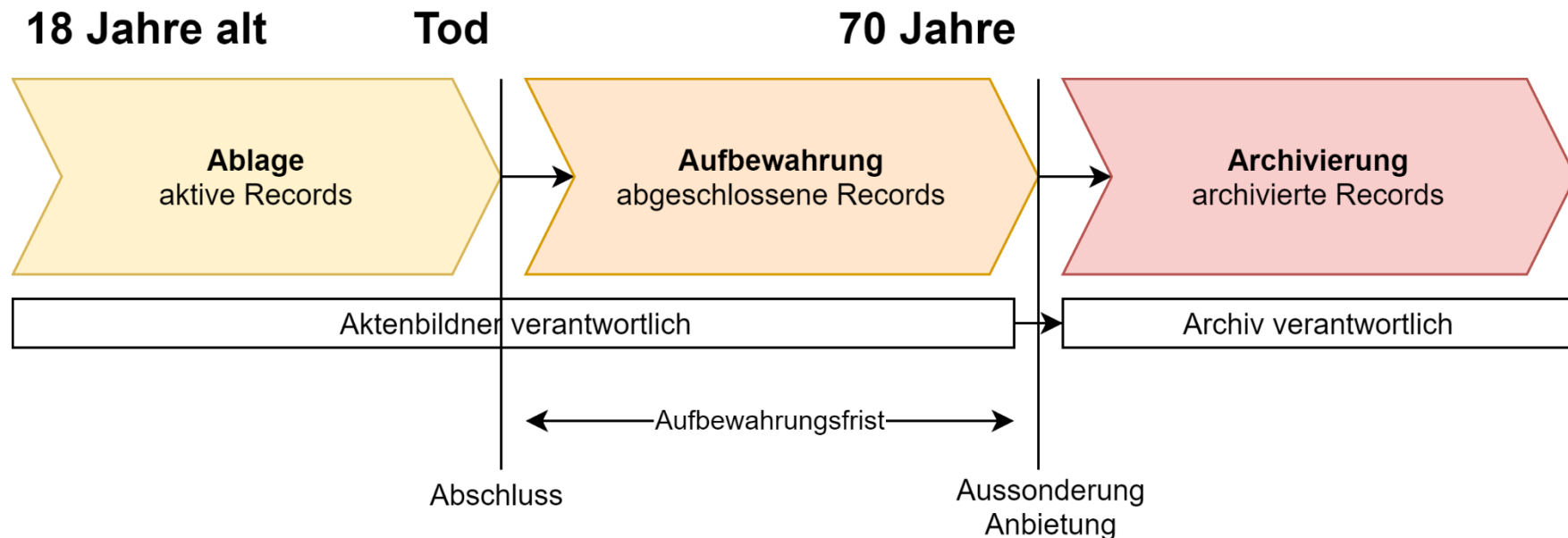
- Objektdokumentationen über Immobilien und andere Infrastrukturen
- Medizinaldossiers, die über Jahrzehnte offen bleiben
- Personaldossier langjähriger Mitarbeitender
- Fachwissen von Gedächtnisinstitutionen (Dokumentationen von Museen, Denkmalpflege, Archäologie)
- Baugenehmigungsunterlagen, die beim nächsten Vorhaben wieder verfügbar sein müssen, und zwar sofort, denn die rufen an und wollen Auskunft
- Unterlagen von Krankenkassen, Vorsorgeeinrichtungen, die ein Menschenleben lang benötigt werden



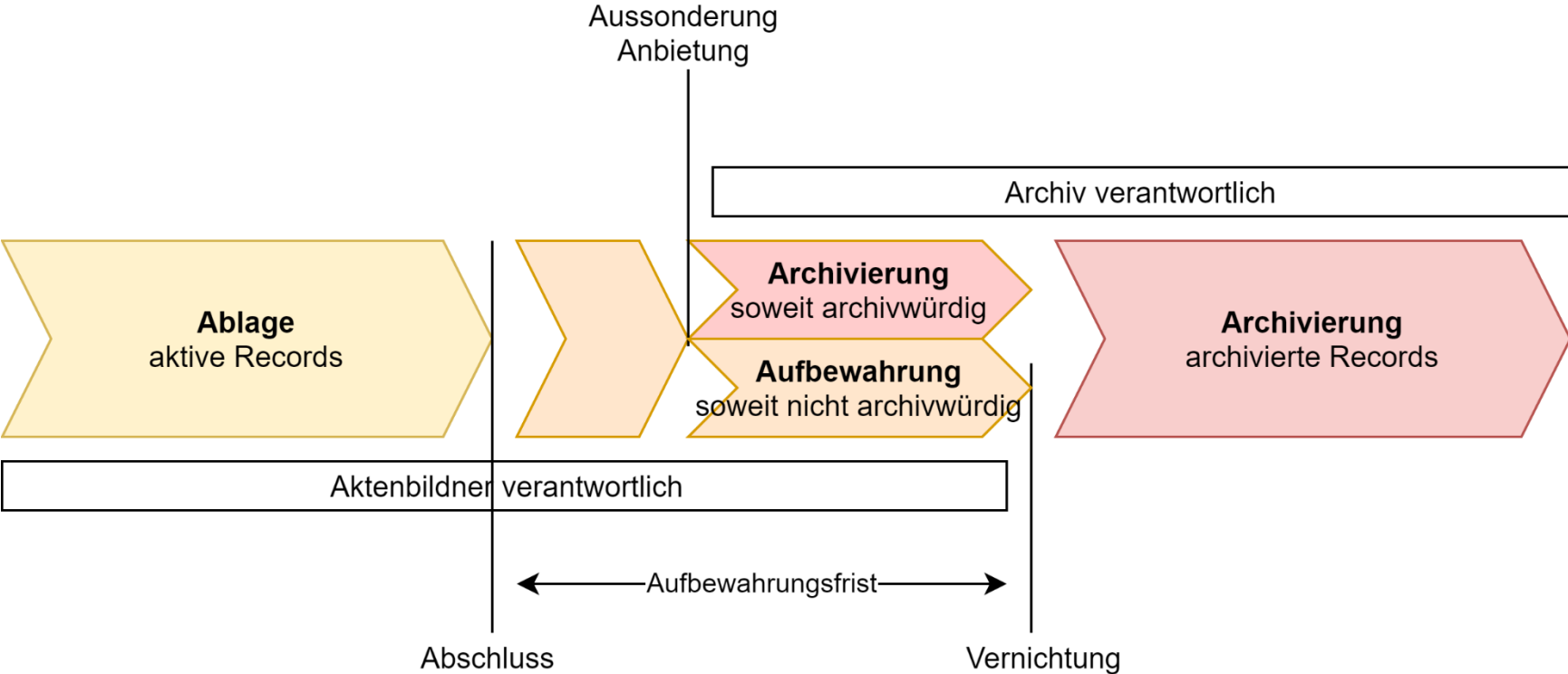
Lange Fristen in jeder Phase möglich

Es spielt keine Rolle, ob die Unterlagen lange aktiv bleiben oder ob sie eine überlange Aufbewahrungsfrist haben. Manchmal kommt sogar beides vor.

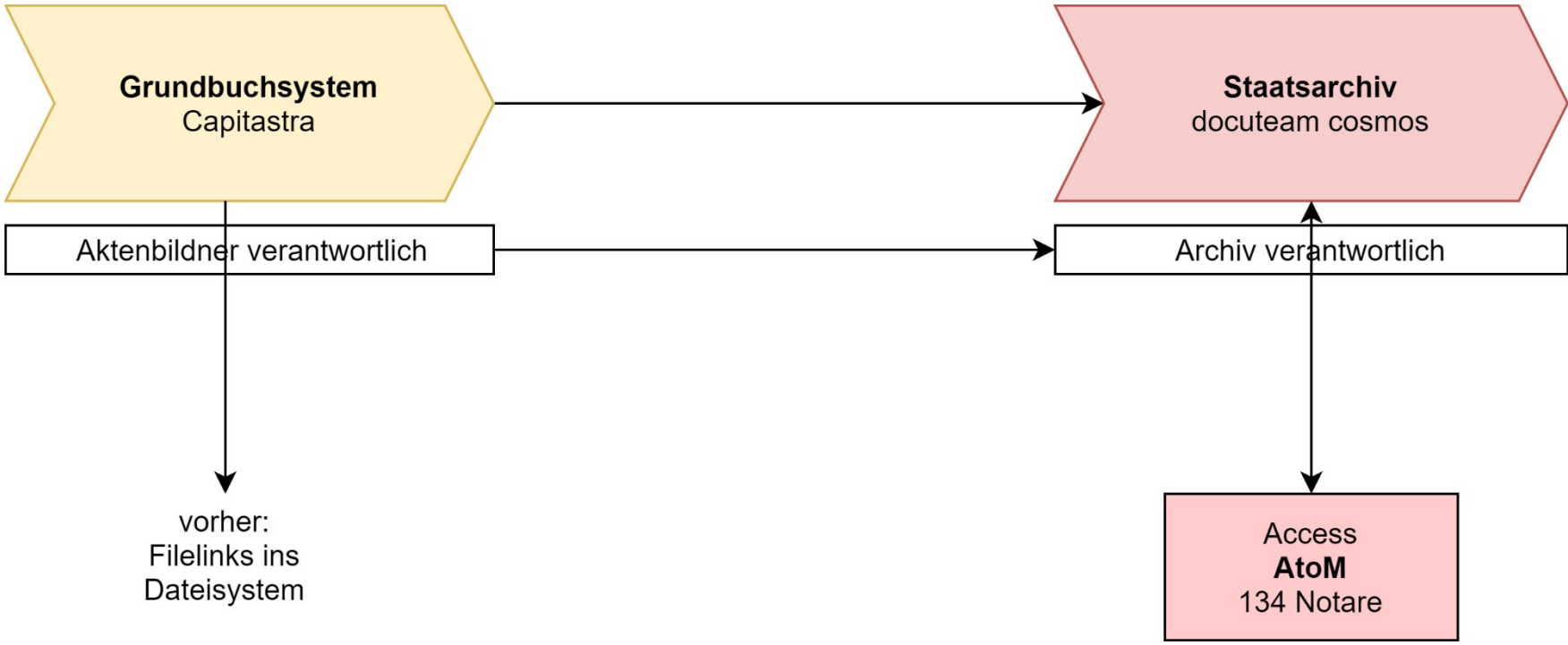
Klientenakten einer Institution, die Schwerbehinderte betreut:



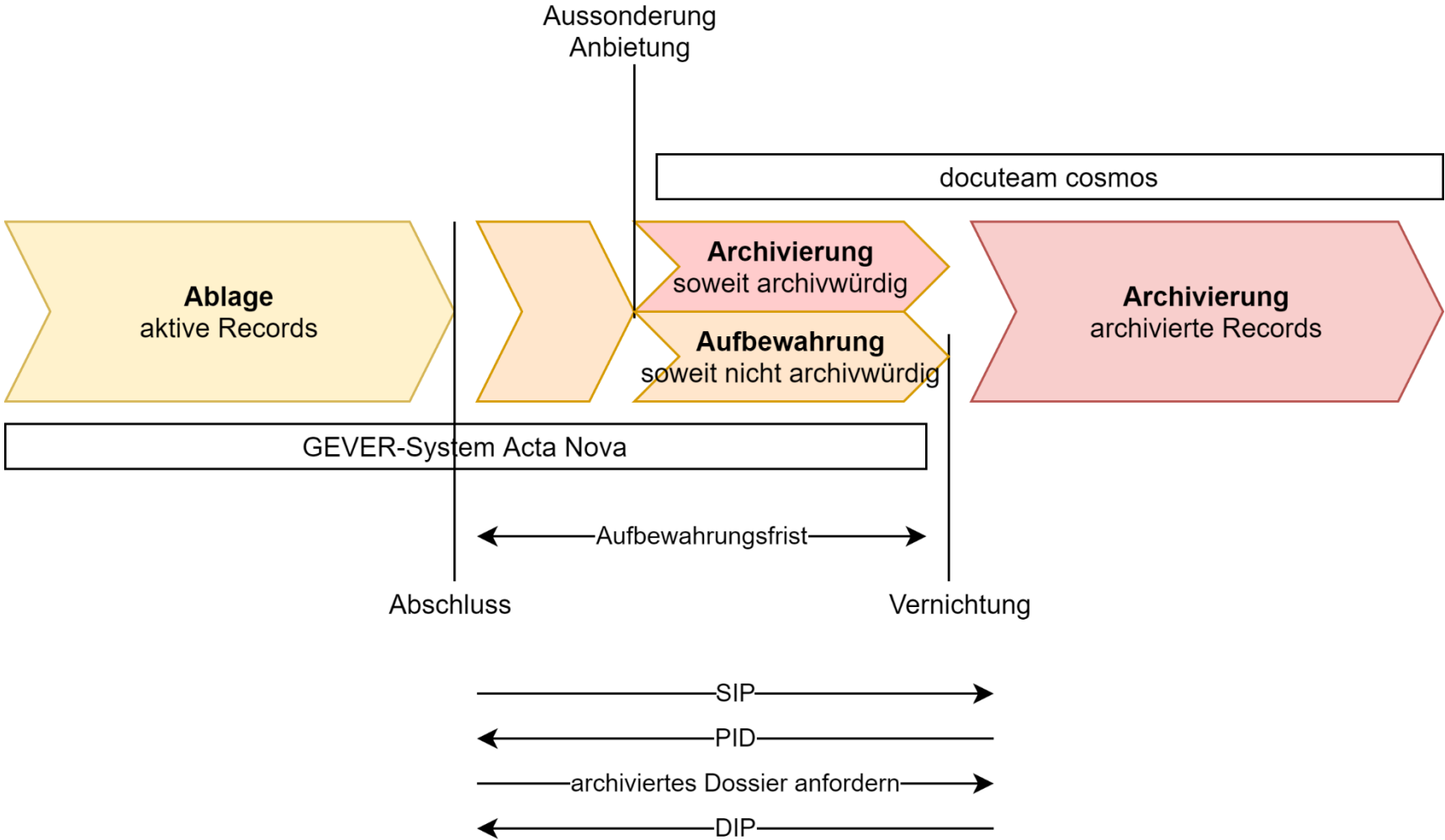
Alternative Lebensphasen-Modelle: Triagemodell (eCH-0164)



Beispiel 1: Grundbuchamt eines Kantons



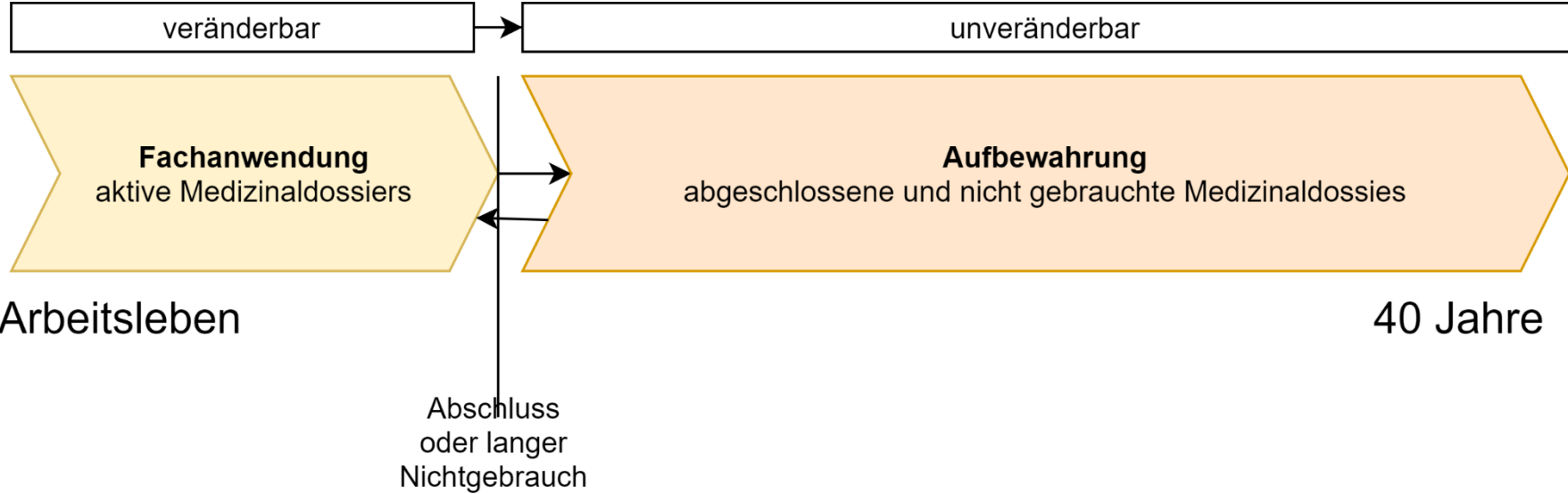
Beispiel 2: GEVER-System einer Stadtverwaltung



Beispiel 2: Konsequenzen

- Das Datenmodell muss es erlauben, einen PID für jedes Datenobjekt zu generieren. Dokumente müssen einzeln bestellt werden können.
- Im digitalen Archiv werden pro Dokument zwei, später mehr Repräsentationen vorrätig gehalten.
- Es ist möglich, im digitalen Archiv zu löschen, zum Beispiel nach einer Aufbewahrungsfrist von 50 Jahren.

Beispiel 3: Medizinaldossiers einer grossen betriebsärztlichen Unternehmung



Weitere Anbindungen

- Projektunterlagen eines Industriekonzerns (> 10'000 Dateien pro Projekt) werden an docuteam cosmos abgeliefert, dort geschieht der Ingest, die AIP werden wiederum im Open Text Content Server gespeichert.
- Ein Museumskatalogsystem, das auch die Verwaltung und Veröffentlichung von Dateien (z.B. Fotos der Objekte) ermöglicht, liefert im Hintergrund an docuteam cosmos ab, wo die Dateien gesichert werden. Im Katalogsystem verbleiben für die Nutzung und Publikation bloss Nutzungskopien.

Vorteile eines alternativen Lebensphasen-Modells 1/2

- Man ist gewohnt, im Dateisystem alles verfügbar zu haben, was jemals digital produziert wurde. → Verzicht fällt schwer.
Die gezeigten Modelle ermöglichen es, Unterlagen der digitalen Nutzbarhaltung zu übergeben, ohne darauf verzichten zu müssen.
→ Schlüssel zur Ablieferungsbereitschaft für digitale Unterlagen
- Dateikonvertierung kann an einem zentralen Punkt erfolgen: im Ingest.
 - Die Mitarbeitenden im Tagesgeschäft werden nicht damit belastet.
 - Konvertierung ist eine archivische Fachaufgabe, die die Konservierungsplanung präjudiziert und die wir selber im Griff haben wollen.
 - Dateikonversion im Ingest bedeutet, mehr und sicherere Konvertoren im Einsatz zu haben als in verteilten Produktivsystemen und die richtigen technischen Metadaten erzeugen zu können (Premis).

Vorteile eines alternativen Lebensphasen-Modells 2/2

- Nutzbarhaltung ist auch bei langen aktiven Phasen und langen Aufbewahrungsfristen jederzeit gewährleistet. Im digitalen Archiv werden Preservation Actions durchgeführt.
- Bewertung, wo nötig, ist 2 Jahre nach Aktenabschluss einfacher als 10 Jahre danach
- Entlastet die Produktivsysteme bzw. deren Datenbanken → Performance

Konsequenzen

- Die Aufgabe des Archivs verändert sich: **Es wird zum Dienstleister für Nutzbarhaltung**. Hierfür braucht es möglicherweise mehr Ressourcen.
→ kriegt das Archiv etwas vom IT-Budget?
- Es muss geregelt werden, wer für die Unterlagen verantwortlich ist.
 - Gehen mit der Ablieferung eines SIP die Unterlagen in die Verantwortung des Archivs über?
 - Und dieses wiederum erteilt der abliefernden Stelle eine generelle Bewilligung, auf die selbst abgelieferten Unterlagen zuzugreifen?
 - Darf die abliefernde Stelle die aus dem Archiv zurückgeholten Unterlagen auch verändern und neu abliefern (als neue Repräsentation)?

Wie gewährleisten wir das?

- docuteam cosmos verfügt über eine Schnittstelle («docuteam bridge»), die offen dokumentiert ist.
- docuteam bridge besteht aus REST APIs, die normalerweise eine JSON-Antwort zurückgeben.
- Der Schnittstelle ist es mehr oder weniger egal, in welchem Format die SIP daherkommen:
METS, BagIt (Dublin Core), XDOMEA (Deutschland), EDIDOC (Österreich), SEDA (Frankreich), eCH-0160 (Schweiz)
- Der Hersteller des Produktivsystems kann aufgrund der offenen Dokumentation direkt damit beginnen, seine Schnittstelle zu programmieren. In der Regel reichen wenige konzeptionelle Besprechungen für die Realisierung eines Anbindungsprojekts. → Eröffnet die Möglichkeit, mit relativ geringem Aufwand mehrere Produktivsysteme anzuschliessen.

Pause 5 Min.

Trend 4: Skalierung Ingest

- Ablieferungsstau GEVER
- Ingest muss in der Lage sein, auf industriellem Niveau Daten zu übernehmen

Es herrscht Ablieferungsstau!

Ganz allgemein gilt: automatisierte Ablieferungen aus GEVER-, DMS- oder e-Akte-Systemen, ja auch aus Fachanwendungen haben noch kaum stattgefunden.

Gründe?

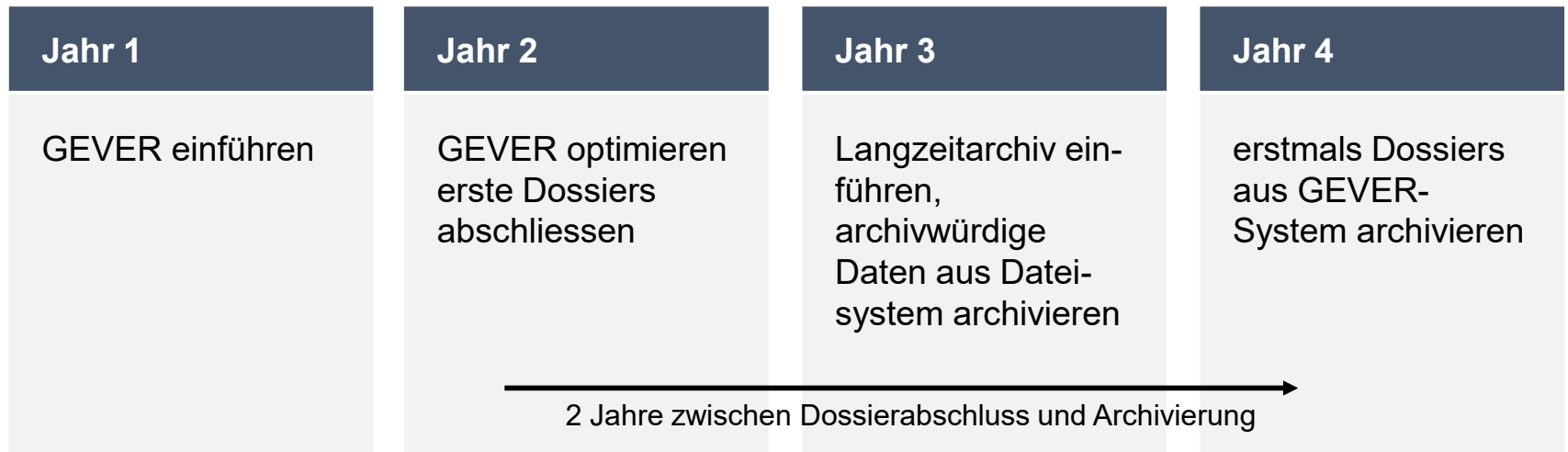
- fehlendes Verständnis
- werden hinausgeschoben
- es getraut sich keiner

Konsequenzen

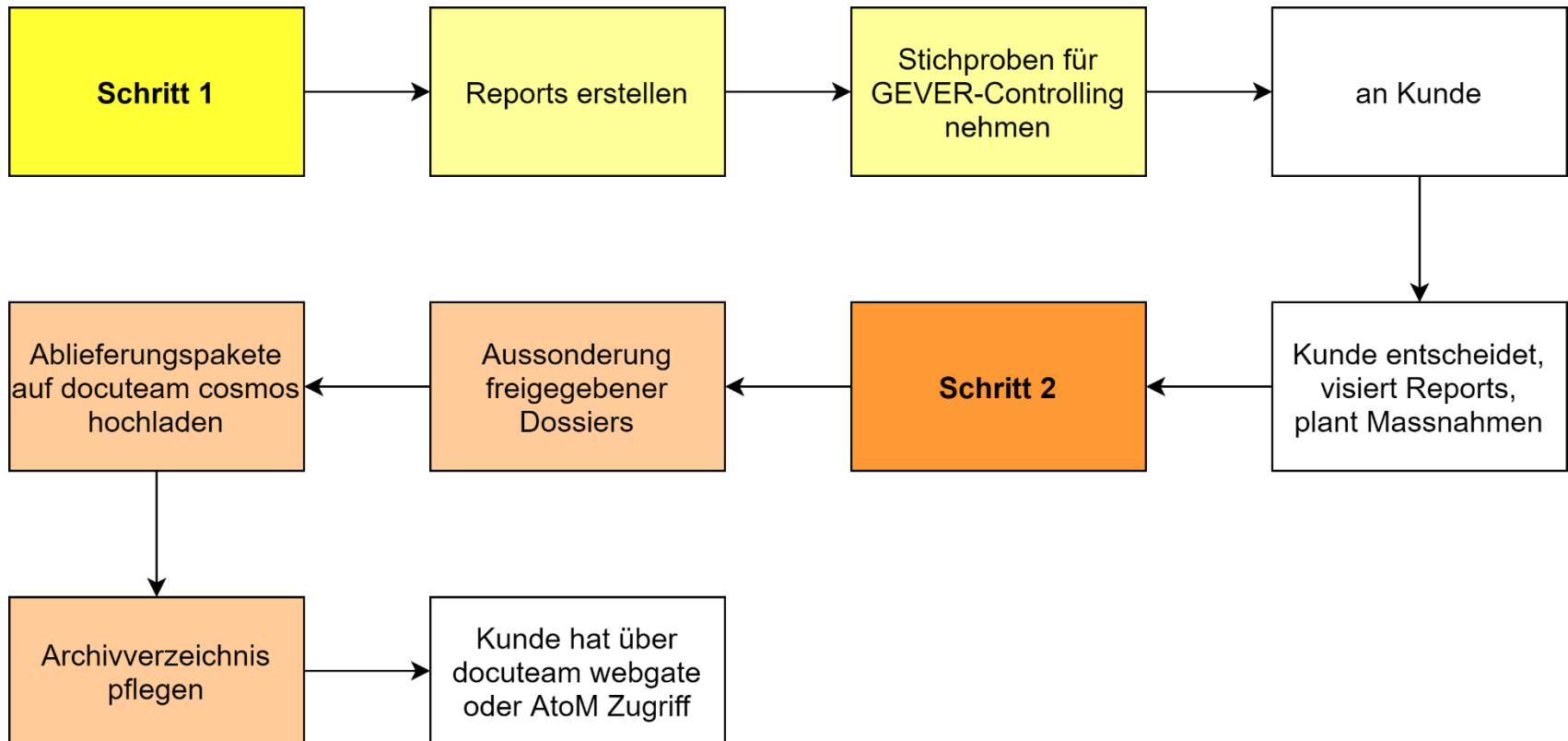
- Aufgabe wird riesig
- Dateiformate nicht mehr lesbar
- systematische Fehler werden nicht entdeckt
- Datenbanken werden zu gross



Zeitbombe durch Triage-System vermeiden



Jährlicher Prozess statt Grossprojekt, am schnellsten per Remote-Zugriff



Beispiele Skalierung

Beispiele aus diesem Jahr

- Grundbuchamt Kanton Wallis
 - 3,7 Mio Dokumente, seit 1912
 - Ingest von 5000 Archivpaketen pro Tag (1,8 Mio/Jahr)
- Universitätsbibliothek Basel: e-codices
 - 120 TB mit Formatmigration
- Landesarchiv Niederösterreich
 - Initialingest von 3 Mio EDIDOC-Paketen
 - danach jährlich > 1.5 Mio

Performanter Ingest erforderlich

- Die Ingest-Werkzeuge müssen solche Massenablieferungen bewältigen können.
 - Andere OAIS-Systeme haben hierfür
 - keine genügenden Prozesse und Routinen,
 - erfordern manuelle Eingriffe mit vielen Klicks pro SIP.
 - Wieder andere produzieren nichts ausser Wartungskosten, weil sie nicht genutzt werden.
- Belasten Sie Ihren Ingest, der muss arbeiten, auch übers Wochenende!

Wir können bereits jetzt hohe Durchsatzzahlen erreichen.

Wir sind aber dabei, die Werkzeuge mit Blick auf die Skalierung weiterzuentwickeln.

→ Fordern Sie uns heraus! Stellen Sie Anforderungen an unsere Werkzeuge!

Trend 5: Bitstream Preservation robuster machen

- Eine verlässliche und ausfallsichere Datenstrom-Sicherung ist die Voraussetzung für die langfristige Nutzbarhaltung digitaler Objekte.
- Wir setzen uns für ein nationales Sicherheitsnetz ein zur Sicherung digitaler Kulturgüter

- In Betrieb seit 1979, Erweiterungsbau 2012
- Fast 80'000 Mikrofilme eingelagert
- Stilles Archiv, keine eigene Nutzung
- Ziel: Redundanz zu schaffen
- Mikrofilme nach definierten Standards hergestellt
- Baulich gesicherte Rückfallebene (Kaverne)



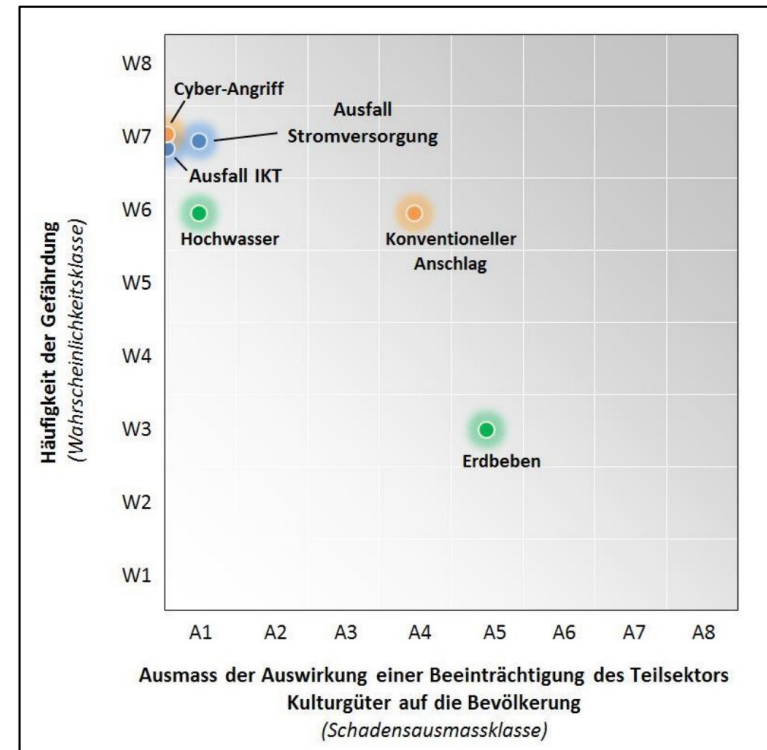
→ **Wie sieht das Äquivalent des Mikrofilmarchivs für digitale Kulturgüter aus?**

Bitstream Preservation robuster machen

Bedrohungen für digitale Objekte sind Cyber-Angriffe, Ausfälle der IKT, Ausfall der Stromversorgung, Konventionelle physische Angriffe, Feuer/Hochwasser/Erdbeben*

Anforderung an eine national koordinierte Infrastruktur

- Denzentral und verteilt, sowohl Technik wie auch Organisation
- Vollautomatisierung vs. Teilautomatisierung
- Datenintegrität muss überprüfbar sein
- Einsatz von nachhaltigen und ressourcenschonenden Technologien



*Quelle: Risiko- und Verwundbarkeitsanalyse des kritischen Teilssektors Kulturgüter, 2017.
<https://www.babs.admin.ch/de/aufgabenbabs/ski.html>

Studie für den nationalen Kulturgüterschutz

Informationsmanagement und
Archivdienstleistungen

Docuteam GmbH
Im Langacker 16
CH-5405 Baden-Dättwil

Dr. Tobias Wildi
t.wildi@docuteam.ch
Telefon +41 56 470 03 37

docuteam

Ein Bergungsort für digitale Kulturgüter

Studie im Auftrag des Fachbereichs Kulturgüterschutz KGS,
Bundesamt für Bevölkerungsschutz BABS

Tobias Wildi, 2. März 2020

Unsere Thesen 2020

- Digitale Archivierung wird sich noch verstärkt in die Cloud verlagern
- Linked Data und RiC werden zu Kerntechnologien für die Archive
- Ausgewählte Datenbestände müssen für die Aktenbildner über sehr lange Zeiträume ohne Zeitverzug nutzbar bleiben
- Im Ingest werden automatisierte Ablieferungen zum Normalfall. Der Ingest muss zunehmend grössere Datenvolumen bewältigen können
- Grundlage für die Nutzbarhaltung über lange Zeit ist eine wasserdichte Bitstream-Preservation-Strategie

Fazit: Welche Rolle übernehmen Archive in einer digitalisierten Welt?

Archive sind nicht mehr einfach Endlager, sondern Dienstleister für die Nutzbarhaltung von fixierten (unveränderbaren), aber aktiv genutzten Informationen.

Les archives ne sont plus de simples dépôts, mais des fournisseurs de services pour l'utilisation d'informations fixes (immuables) mais activement utilisées.

Pause 5 Min.

Anschliessend Beantwortung der Fragen aus dem Chat

Vielen Dank für Ihre Aufmerksamkeit.

- Den Link auf die Präsentation stellen wir Ihnen in der Folge zu.
- Wir stehen Ihnen später für individuelle Besprechungen bereit – per Telefon, Videokonferenz oder auch wieder persönlich

Christian Moser

c.moser@docuteam.ch

+41 56 511 11 79

Andreas Steigmeier

a.steigmeier@docuteam.ch

+41 56 470 03 03

Tobias Wildi

t.wildi@docuteam.ch

+41 56 470 03 38

