

Kostenmodelle für digitale Archivierung – **Theoretische Modelle und Praxiserfahrung**

Arbeitskreis AUdS, Basel 28.2.2017

Dr. Tobias Wildi, t.wildi@docuteam.ch

Ausgangslage

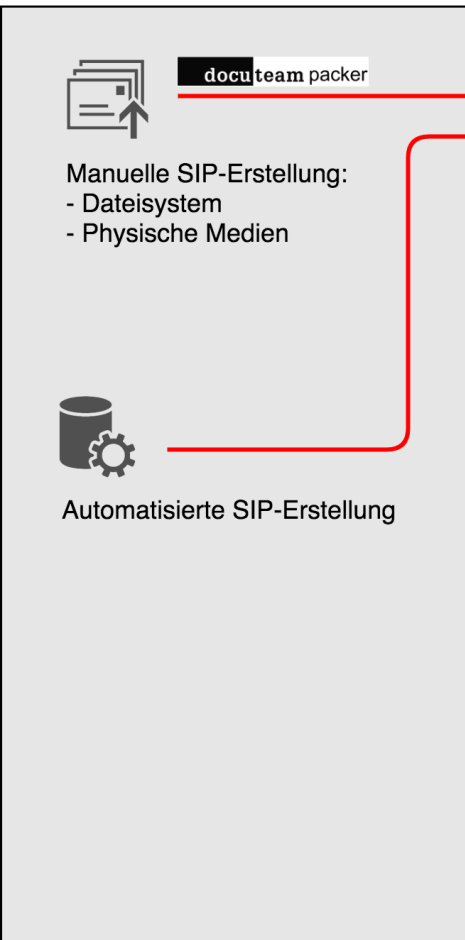
- Docuteam entwickelt seit 2006 Open Source-Software für digitale Archive mit Schwerpunkt Pre-Ingest, Ingest, Repository und Preservation Planning
- Heute rund 20 produktive Installationen mit docuteam packer, docuteam feeder und Fedora Commons
- Koppelung mit verschiedenen AIS (Scope, CMI STAR, Faust, Augias, AtoM, Ex Libris)
- Aufwand für Installation und Betrieb eines eigenen OAIS ist beträchtlich und macht nur für grössere Institutionen Sinn
- **Frage:** Wie macht man digitale Archivierung als Dienstleistungen auch kleineren Organisationen zugänglich?

Hostinglösung

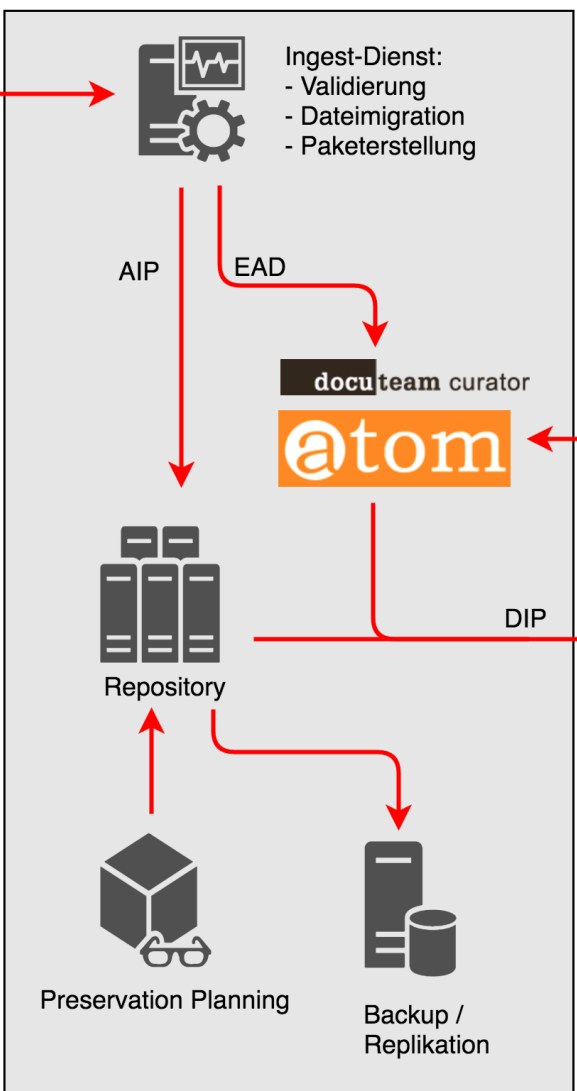
- Synergien nutzen wo möglich: Zentralisierung von Ingest-Diensten und Preservation Planning
 - Durch die Nutzung solcher Synergien wird digitale Archivierung auch für kleine Organisationen (z. B. Gemeinden) erschwinglich
 - Daten befinden sich bei jedem Prozessschritt in der Schweiz
 - Hohe Informationssicherheit: Jede Organisation verfügt über ihr eigenes, abgetrenntes Repository.
- **Was kostet das den Endkunden?**

Aufbau docuteam cosmos 2016

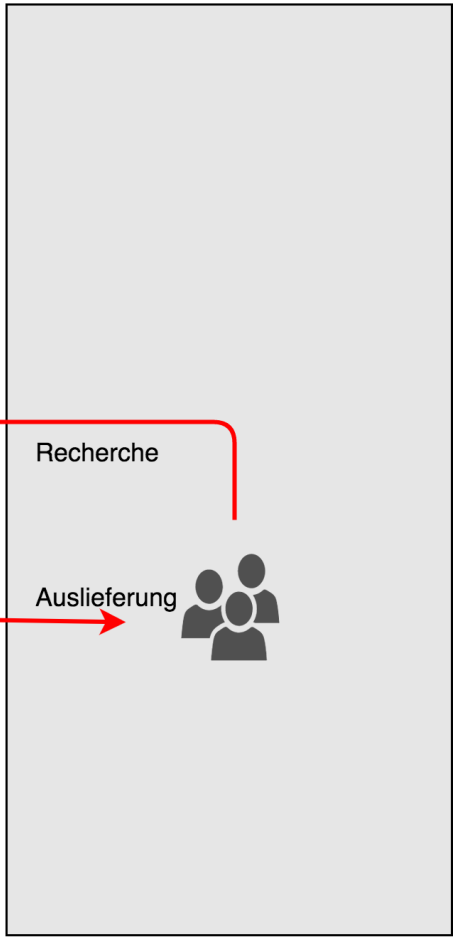
Vorbereitung



docuteam cosmos



Zugang



Kostenmodelle

Was ist ein Kostenmodell?

- Identifizierung aller Kostenfaktoren, die beim Betrieb einer Hostingplattform für digitale Archive auftreten.
- Die Quantifizierung der Kosten basiert zu Beginn auf Annahmen:
 - bezüglich der Designated Community (an wen richtet sich die Plattform? Welche Erwartungen richtet diese Designated Community an die Hostingplattform?)
 - bezüglich Einkaufspreisen von IT-Basis-Dienstleistungen (Annahmen bezüglich zukünftiger Entwicklung)
- Die Annahmen müssen periodisch überprüft und angepasst werden. Iterativer Prozess.
- Aus dem Kostenmodell lässt sich eine Vollkostenrechnung und dann auch ein Preismodell ableiten.

4C:

Collaboration to Clarify the Costs of Curation

- EU-Projekt 2013—2015
- <http://www.4cproject.eu/summary-of-cost-models/>
Zusammenfassung verschiedener Kostenmodelle
- Näher betrachtet werden im Folgenden:
 - LIFE3 Costing Model (University College London and the British Library)
 - Total Cost of Preservation (TCP) (California Digital Library)
 - DP4lib-Kostenmodell (Deutsche Nationalbibliothek)

LIFE3-Modell

- <http://www.life.ac.uk/3/>, drei Projekte 2006—2010
- Basiert auf OAIS, Workflow-orientiert

Lifecycle Stage	Creation or Purchase	Acquisition	Ingest	Bit-stream Preservation	Content Preservation	Access
Lifecycle Elements	Digitisation	Selection	Quality Assurance	Repository Admin	Preservation Watch	Access Provision
	Submission Agreement	Metadata	Storage Provision	Preservation Planning	Access Control
	IPR & Licensing	Deposit	Refreshment	Preservation Action	User Support
	Ordering & Invoicing	Holdings Update	Backup	Re-ingest	
		Obtaining	Reference Linking	Inspection	Disposal	
		Check-in				

Quelle: <http://www.4cproject.eu/summary-of-cost-models/16-community-resources/outputs-and-deliverables/105-life3-costing-model-life3/>

Reichweite des LIFE3-Modells

Wo hilft das Modell in der Praxis?

- Deckt die grundlegenden Aufgaben im digitalen Archiv ab (haben wir an alles gedacht?)

Wo hilft das Modell in der Praxis nicht weiter?

- Kosten für Netzwerk, Rechenzentrum, IT-Plattform müssen auf die einzelnen Aufgaben verteilt werden
- Zur Zeit, als das Modell entwickelt wurde, waren noch wenig Praxiserfahrungen verfügbar.

Total Cost of Preservation (TCP)-Modell

- California Digital Library (CDL), Total Cost of Preservation (TCP), Whitepaper, 2013.
- Das Modell umfasst 10 Kategorien basierend auf OAIS, die zusammen alle Aktivitäten abdecken, die in einem digitalen Archiv anfallen.
- Jede Kategorie repräsentiert einen Kostenfaktor
- Basierend auf den Kostenfaktoren wurden zwei Preismodelle entwickelt
 - Pay-as-you-go
 - Paid-up

The 10 categories are:

- ▶ content owners
- ▶ submission streams
- ▶ preservation system (ingest, data management, access)
- ▶ servers
- ▶ storage
- ▶ consumers
- ▶ preservation planning
- ▶ interventions (e.g. migrations)
- ▶ administration
- ▶ management

TCP Preismodell

- **Pay-as-you-go price model:** periodische (jährliche) Verrechnung für Organisationen mit gesicherten, verlässlichen Budgets.
- **Paid-up price model:** Einmalige Verrechnung für eine bestimmte Anzahl Jahre. Oft die einzige Möglichkeit für Organisationen, Projekte, Initiative, die nur während einer bestimmten Zeit bestehen (z. B. Forschungsprojekte mit Drittmitteln).

DP4lib-Kostenmodell (Dt. Nationalbibliothek)

- Beschreibung und Begründung des Kostenmodells und Excel für Berechnungen
- Orientiert sich an den Aufgabenblöcken des OAIS, bildet diese mehrstufig ab
- Nach unserer Erfahrung ist dies ein taugliches, praxisnahes Modell

	Ingest	Curation	Access
Hardware			
Software			
Personal			
Räumlichkeiten			
Externe Services			

Kostenmodell docuteam

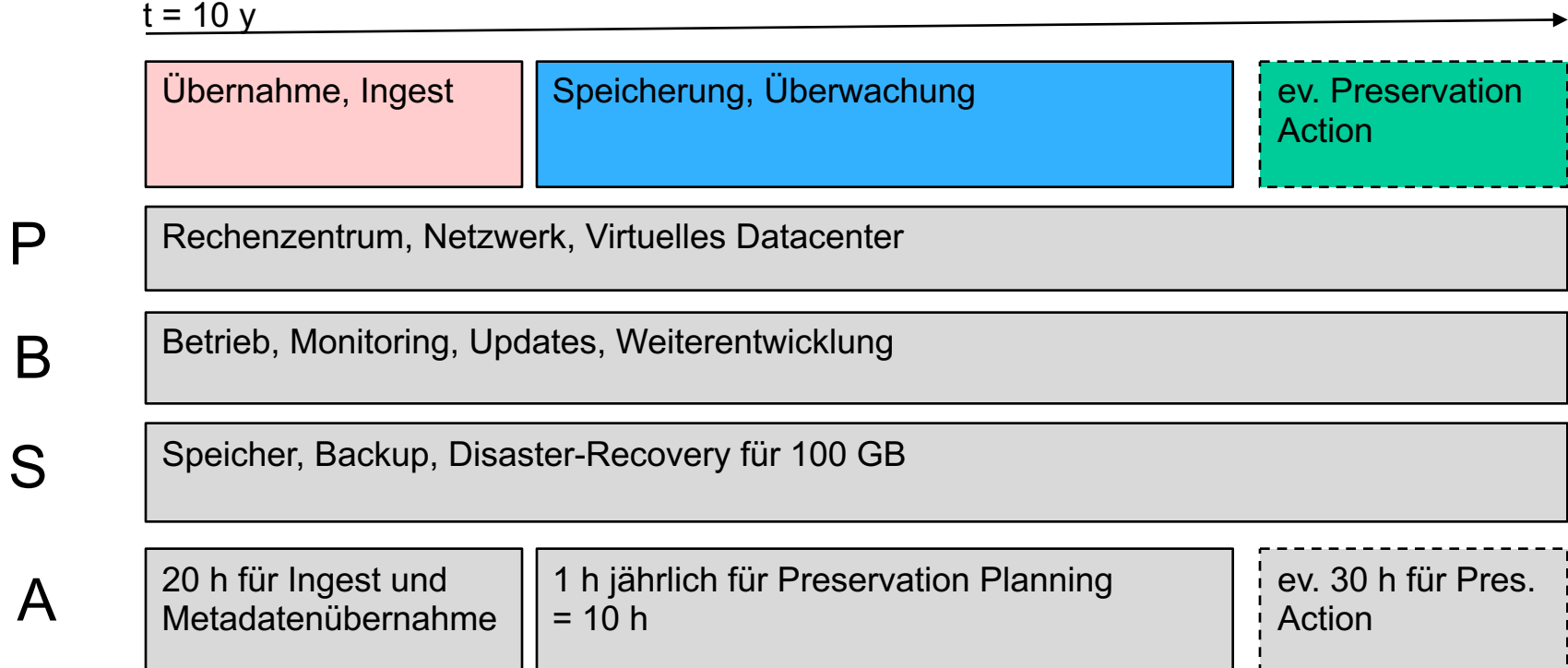
docuteam-Kostenmodell: Kategorien

Plattform	<ul style="list-style-type: none">– Rechenzentrumsdienste (Sicherheit, Strom, Kälte)– Netzwerk, Internetanbindung, Firewall– Virtuelles Datacenter bis und mit Hypervisor
Betrieb	<ul style="list-style-type: none">– Betriebssysteme, Serverdienste– Betrieb, Wartung Applikationen (von docuteam und Dritten)– Überwachung Server und Speichersysteme– Beitrag an Weiterentwicklung der Software
Speicher	<ul style="list-style-type: none">– Repository– Backup (Georedundanz in zweitem RZ)– Disaster Recovery (Tape oder Kopie zum Kunden)
Archivarbeit	<ul style="list-style-type: none">– Übernahme und Aufbereitung von Daten und Metadaten– Preservation Planning (fachliche Überwachung)– Preservation Actions (bestandserhaltende Massnahmen)– Unterstützung bei Benutzung / Access

Kostenfaktoren auf Beispiel umgelegt

- Übernahme und Speicherung von 2000 PDF/A-Dateien à 50 MB (~100 GB) während 10 Jahren. Die beschreibenden Metadaten werden in einer Excel-Datei angeliefert.

t = 10 y



Preismodell

Der Preis für docuteam cosmos setzt sich aus drei Bündeln von Kostenfaktoren zusammen:

Software as a Service	Basispreis pro Jahr, orientiert sich an der Grösse des Kunden
Speicher	Abrechnung pro GB / TB (quartalsweise)
Archivdienstleistung	Abrechnung Arbeitsstunden nach Aufwand (Archivdienstleistung)

Fazit

- Hilfreich sind systematisierte (vollständige) Kostenmodelle,
 - die **aufgabenorientiert** das funktionale OAIS-Modells abbilden
 - und andererseits **mehrstufige** IT-Architekturen berücksichtigen
- **Rechenzentrums- und Netzwerkdienste** müssen als beträchtliche Kostenfaktoren berücksichtigt werden, hier greifen die bestehenden Modelle zu kurz.
- **Kostentreiber** sind die Speicherkosten (= steigende Fixkosten).
- **Personalkosten** treten punktuell und planbar auf (Preservation Actions) und können entsprechend über einmalige Projekte finanziert werden.

