

Für die Zukunft – EWIG

der neue Service »Digitale Langzeit-
archivierung« der KOBV-Zentrale

Wolfgang Peters-Kottig

JUL 8, 2016 @ 05:08 PM 441 VIEWS

Digital Preservation Is Society's Genome



Tom Coughlin
CONTRIBUTOR

An organism's genome is the complete set of genes or genetic material present in the organism. The genome stores all the information that the organism needs to create and maintain its organs and living functions. Likewise we might say that the genome for a civilization is the set of preserved data that defines that civilization and serves as the basis of its organization and functions.

<http://www.forbes.com/sites/tomcoughlin/2016/07/08/1975/#3c112bfe243f>

Backup ist KEINE Langzeitarchivierung

Prämisse: Daten und Dateiformate, mit denen wir gegenwärtig umgehen, werden in Zukunft nicht mehr zugänglich und verstehbar sein. Das liegt nicht nur in der Hardware (z.B. Laufwerke für 5¼"-Floppy-Disketten) als auch in der Software begründet (z.B. Textverarbeitung Wordstar 2000).

OAIS steht für Open Archival Information System und hat nichts zu tun mit OAI, was für Open Archives Initiative steht.

Langzeitarchivierung und Langzeitverfügbarkeit sind dasselbe.

Backup

Backup dient der kurzfristigen Wiederherstellung der Funktionsfähigkeit eines Systems.

Bei einem Backup können die Daten, so wie sie auf der Platte liegen, weggeschrieben werden.

Backup sichert Daten unabhängig von ihrem Zustand (auch beschädigte Dateien oder obsoleete Formate werden 1:1 gesichert).

Das Zurückspielen eines Backups in das laufende, gesicherte System erfordert minimale Metadaten.

Mit einem Backup lassen sich Daten nur solange wiederherstellen, wie das System, aus dem sie stammen lauffähig und verfügbar ist.

Langzeitarchivierung

Langzeitarchivierung ist ein fortlaufender Prozess, der sicherstellen soll, dass die archivierten Inhalte in Zukunft verfügbar und verstehbar sind.

Für die Langzeitarchivierung muss erheblich mehr organisatorischer Aufwand betrieben werden.

Langzeitarchivierung stellt hohe Anforderungen an die Qualität der zu archivierenden Daten und prüft auch die archivierten Daten laufend.

Für die Langzeitverfügbarkeit ist auf hohe Qualität und maximalen Umfang der Metadaten zu achten.

Ein Langzeitarchiv stellt sicher, dass Inhalte und Eigenschaften der archivierten Objekte in Zukunft systemunabhängig verfügbar bleiben.

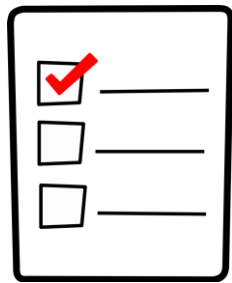


Service „Digitale Langzeitarchivierung“ der KOBV-Zentrale

- ▶ „Bereitstellung, Pflege und Weiterentwicklung eines vertrauenswürdigen, OAIS-konformen digitalen Langzeitarchivs“ → **EWIG**
- ▶ Service steht allen KOBV-Bibliotheken im Rahmen ihrer Mitgliedschaft zur Verfügung
- ▶ Technische Basis:
Speicherinfrastruktur des ZIB
(Magnetband, derzeit ~90 PB Kapazität)



Service „Digitale Langzeitarchivierung“ der KOBV-Zentrale



- ▶ Inhalte: z.B. Hochschulschriften & ähnliche Publikationen; Digitalisate von Handschriften, Büchern, Zeitschriften; AV-Werke; Forschungsdaten...
- ▶ Online-Präsentation erfolgt weiterhin durch die Bibliothek (aber direkter Zugriff für Endnutzer mitkonzipiert)
- ▶ Flexible Datenübernahme ins Langzeitarchiv: Keine grundsätzliche Beschränkung nach Datenformaten
- ▶ KOBV-Zentrale bietet Beratung zu allen Fragen der Langzeitarchivierung, auch zum Forschungsdatenmanagement

Entgeltmodell „Digitale Langzeitarchivierung“



- ▶ Im Rahmen der KOBV-Mitgliedschaft kostenloses Basis-Speichervolumen für
 - Universitäts- und Hochschulbibliotheken
 - beide Landesbibliotheken
 - beide Verbünde öffentlicher Bibliotheken
- ▶ Archivierung für die Dauer der KOBV-Mitgliedschaft
- ▶ Größe der Basis-Speichermenge ist gestaffelt nach „Gruppenzugehörigkeit“ im KOBV

Gruppe	Institution	kostenlose Basis-Speichermenge (TB)
große Berliner Univ. + ZLB	FU	25
	HU	25
	TU	25
	ZLB (inkl. VÖBB)	25
Brandenburgische Univ.	EUV	10
	UP	10
	BTU	10
Berliner HS/Univ. + SLB	UdK	5
	Charité	5
	Beuth	5
	HWR	5
	HTW	5
	SLB	5
Sonstige	Alice Salomon HS	1
	HS Schauspiel	1
	HS Musik	1
	KunstHS	1
	THW	1
	THB	1
	FHE	1
	Filmuni	1
	FHP	1
	VÖB	1

Entgeltmodell „Digitale Langzeitarchivierung“



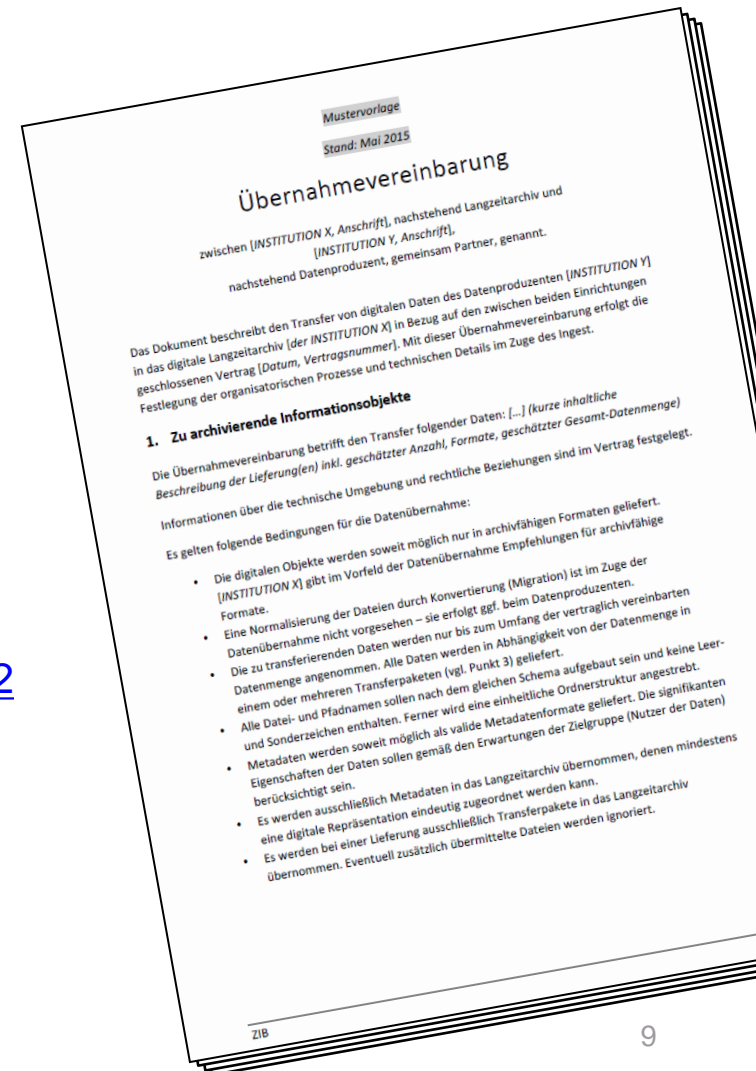
- ▶ Entgelt für zusätzlichen Speicherbedarf über die Basismenge hinaus: 500 € pro TB pro Jahr
- ▶ Alle weiteren KOBV-Mitglieder:
Leistung im Rahmen von bilateralen Kooperationsvereinbarungen (Zusatzleistungen), es gilt ebenfalls ein Entgelt von 500 € pro TB und Jahr
- ▶ Eine Abrechnung erfolgt jährlich (bzw. einmalig bei Kooperationsvereinbarung)
- ▶ Kleinste Einheit ist 1 TB

Wie kommen die Daten zu uns?

1. Beratungsgespräch(e): Was? Warum? Wieviel? Wie lange? Wer? Wann?...
2. Einigung auf technisch-organisatorische Parameter (Inhalte, Ordnerstruktur, Zugriffsrechte, Metadatenformate):
Übernahmevereinbarung

→ Mustervorlage des ZIB,
Stand Mai 2016

<http://nbn-resolving.de/urn:nbn:de:0297-zib-59452>



Wie kommen die Daten zu uns?

3. Erstellung von Transferpaketen

```
transfer 1/  
-- object_001  
|-- AIG10_DTS_log.pdf  
|-- AIG10_DTS_log_r.png  
|-- AIG10_P_and_S-wave.pdf  
|-- DataCite.xml  
|-- DIF.xml  
|-- escidoc.xml  
-- object_002  
|-- DIF.xml  
|-- DataCite.xml  
|-- data.csv  
-- object_003  
|-- submissionDocumentation  
|   |-- setup.png  
|   |-- handoverDocu.txt  
|-- DIF.xml  
|-- DataCite.xml  
|-- data.csv  
-- submission-manifest.txt
```



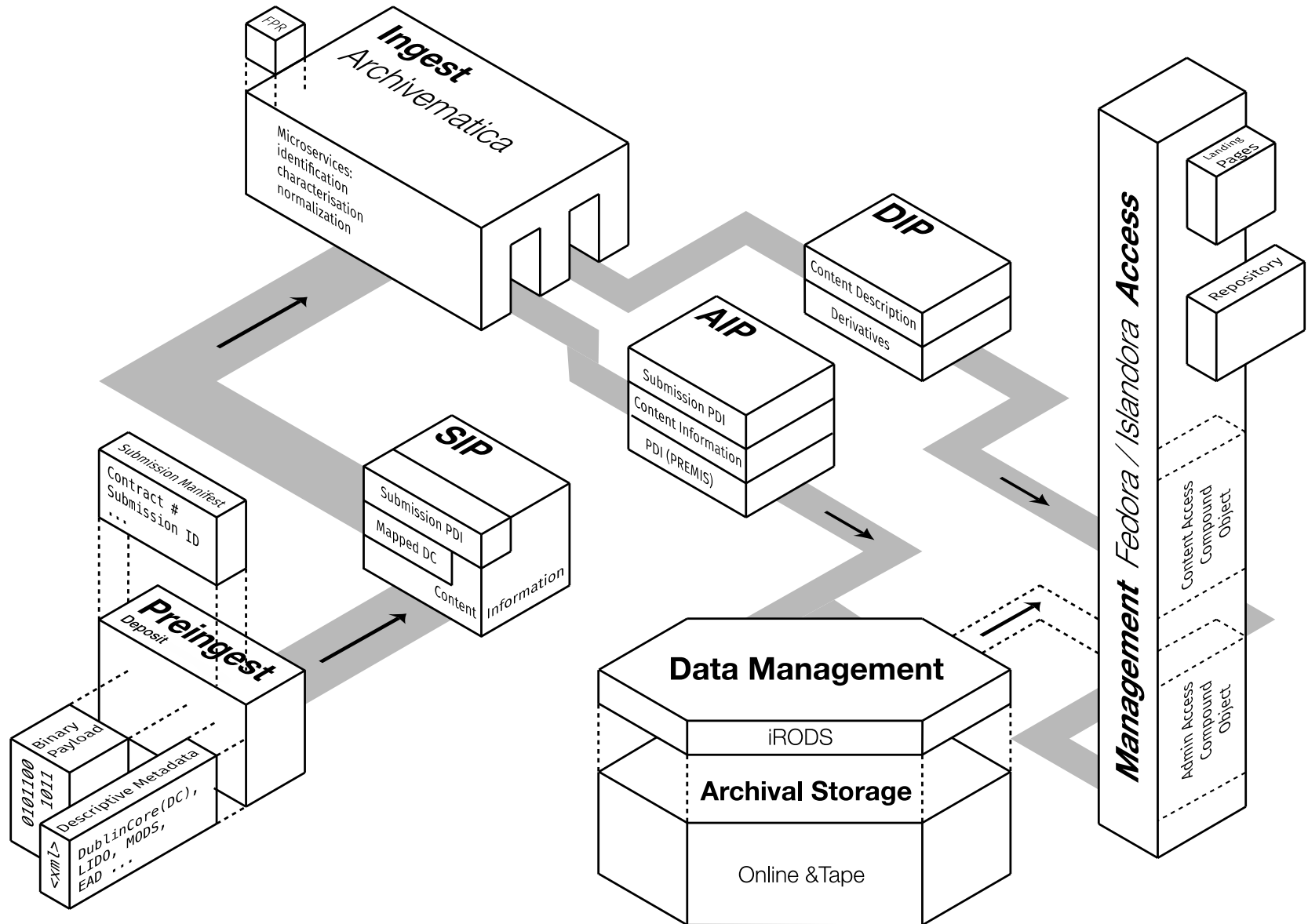
4. Datenübertragung ins ZIB

- online (z.B. via Berliner Wissenschaftsnetz BRAIN)
- externe Festplatte...

5. Ingest in EWIG

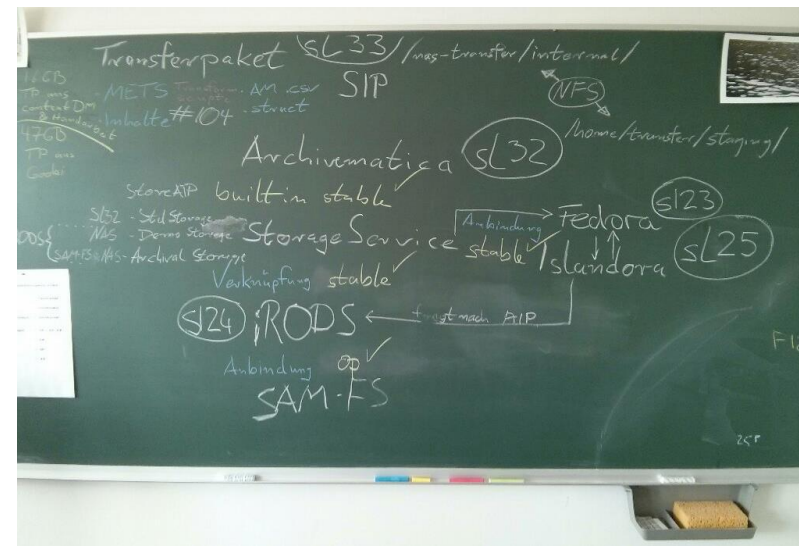
Ab hier Verantwortungsübernahme
durch die KOBV-Zentrale

Workflow der Datenverarbeitung in EWIG



Aufgaben in 2016

- ▶ EWIG-Weiterentwicklung, z.B.:
 - erhöhte Automatisierung, u.a. der Datenannahme:
→ „Goobi-Projekt“ mit Daten der HU
 - Reporting/Statistik-Tool
 - Dokumentation
 - Empfehlungen für bevorzugte Archivformate
- ▶ Objektexport aus digitalen Editionen der Geisteswissenschaften (Projekt Humanities Data Centre)



Fragen Sie uns, wir vereinbaren gerne ein Beratungsgespräch!

Wolfgang Peters-Kottig
peters-kottig@zib.de

Team Digital Preservation
(KOBV + digiS)
dp@zib.de