Digitale Langzeitarchivierung am Zuse-Institut Berlin

Wolfgang Peters-Kottig

CC BY 4.0





Langzeitarchivierung wird nicht gelöst durch »den ewigen Speicher«

Langzeitarchivierung ≠ Backup

Nachnutzer der Daten sind »andere«

Ein Langzeitarchiv ist keine Software



Die Antwort auf fast alle LZA-Fragen lautet »es kommt darauf an«



Zuse-Institut Berlin (ZIB)

Ein außeruniversitäres Forschungsinstitut für anwendungs orientierte Mathematik und Informatik



Schwerpunkte sind mathematische Modellierung und Optimierung. Kooperationen mit Partnern aus Medizintechnik, Biotechnologie, Transport, Energieversorgung und Optik.

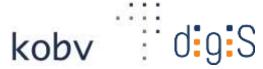


ZIB – Infrastruktur



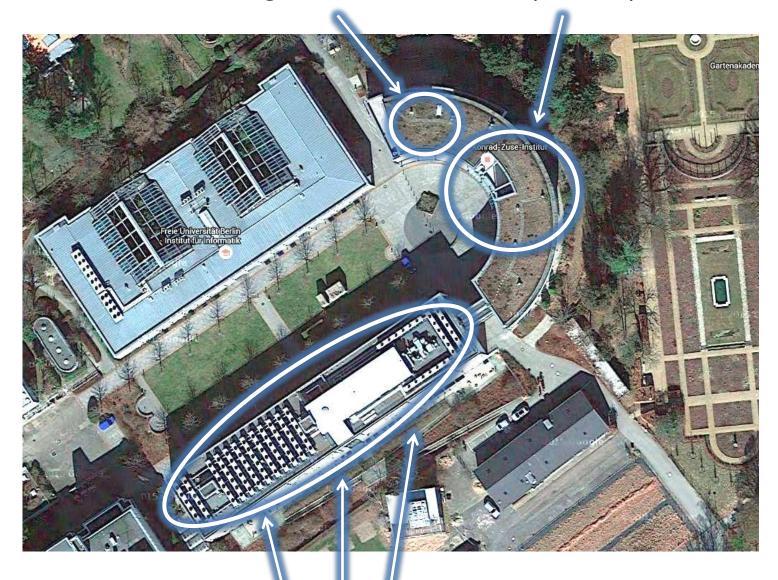
Supercomputer "Konrad"

- 1872 nodes
- 44928 cores
- 120TB RAM
- 1.4 Pflop/s



Magnetband-Archiv

Supercomputer





Kühlung

ZIB-Infrastruktur – Magnetbandarchiv



- Bewährtes Datenhaltungskonzept: Sicherung auf Magnetbandkassetten (Bandroboter)
- Am ZIB derzeit ~100 Petabytes Kapazität



Digitales Langzeitarchiv EWIG

- Service-Schicht "oberhalb" der Bitstream-Archivierung
- Entwickelt und betrieben vom Forschungs- und Kompetenzzentrum Digitalisierung Berlin (digiS) und dem Kooperativen Bibliotheks-verbund Berlin-Brandenburg (KOBV)
- Seit 2014 Gemeinsame Arbeitsgruppe "Digital Preservation" mit Mitarbeitern von KOBV und digiS



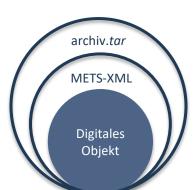


Foto: Matthias Groß, BSB München

Grundsätzliches zum Service

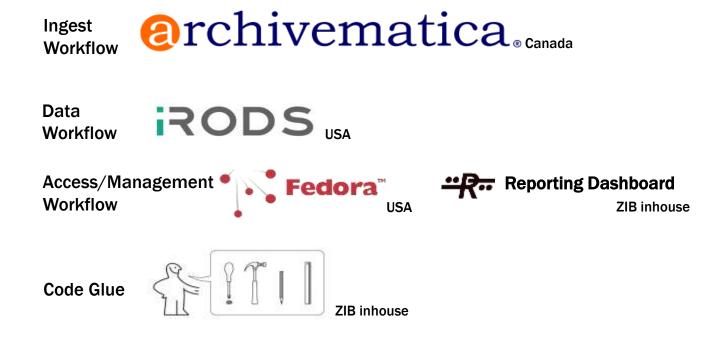
- Ziel ist Erhaltung der "dauerhaften Lesbarkeit"
- Archiviert werden "Objekte" (Dateien) und zugehörige Metadaten
- "Wir nehmen alles" und machen das Beste draus… (mit Riskikorückmeldung)
- Prinzip Zeitkapsel: Archivierung inhaltlich abgeschlossener, vollständig selbstbeschreibender Datenpakete
- Zwei Servicelevel: Active Preservation oder Passive Preservation
- EWIG ist ein "Dark Archive"
 Online-Präsentation bleibt bei den Partnern

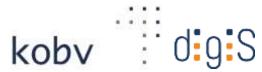




Grundsätzliches zum Service II

► Einsatz von Free & Open-Source-Software als lose verbundene Komponenten, einzeln austauschbar





Grundsätzliches zum Service III

Zielgruppe sind Kulturerbe- und Forschungseinrichtungen Derzeit Mandat der Länder Berlin & Brandenburg für digiS- Partner und KOBV-Mitgliedsbibliotheken

KOBV-Mitglieder: kostenloses Basis-Speichervolumen für Hochschulund Landesbibliotheken

 digiS: kostenlos für alle Projektpartner im Landesförderprogramm Digitalisierung

Externe: Nur beiForschungskooperationen

Der Weg ins Archiv

- Schritt 1: Vertrag
 - Rollen Langzeitarchiv / Datengeber
 - Technische Realisierung (Einlagerung, Speicherung, Bereitstellung),
 - Ansprechpartner, Nutzungsrechte, Haftung
- ► Schritt 2: Übernahmevereinbarung

Konkrete Vorgaben zu:

- Ordnerstruktur
- Transferwege (Festplatte, Internet)
- Beschreibung eines Transferprozesses
- "Packzettel": notwendige administrative Metadaten einer Datenlieferung





Übernahmevereinbarung

 - swischen (MSTITUTION II; Anschoff), nuchsiehend Lengsellarshiv und (MSTITUTION Y; Anschoff);
 - eachstehend Detemprodusent, geneinsam Partinet, genunnel.

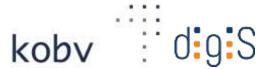
Das Dickoment beschreibt den Fransfer von eigstelen Lienen des Datemprobweisten (MS77/UTON F) in das dystale Langsretunine (der MS77/UTON E) in Bedig auf des parisches beiten Errichtungen geschlichsenen Vertrag (Distam, Vortragsnammer). Mit dieser übernahmenenhamme gefolgt die Festbeurung dies odsprektionschen Prossone und technischen Gestalt im Zust der treuet.

I. Zu archivierende informationsobjekte

Die Übernahmeseinlinkarung betrifft den Transker folgender Deten: [...] (kurze inhet/liche Beschreibung der Lieferungfen) mid, geschätzter Arsant, Farmete, geschätzter Gesenn-Datenmenge)

Informationen über die tachnache Umgebung und rechtliche Bezehungen sind im Vertrag festgelegt

- Die digitalen Objekte werden soweit m\u00f6glich nur in andr\u00f6rligen Formaten gellerlert. (NETIT/TOW-I) gilbt im Verfeld der Debar@berneitene Erngfelhangen f\u00f6r anth\u00f4flige Formate.
- Eine Normalisierung der Delejen durch Kenverfenung (Migration) ist im Juge der
- Die zu brinderteinniten Daten werden nur bis zum Unstang der wertragfich erreinkanten Datenmerge angenommen. Alle Daten werden in Abhängigheit von der Datenmerge in einem dater nebenen Transferpakenen hijd. Prutet 19. geleintert.
- Alle Datel- und Pfadramen zollen nach dem gleichen Schemu aufgebaut zein und keine Liegund Sonderzeichen erdhalten. Ferner wird eine einheitliche Ordnerchrahtur absetziebt.
- Mediadaten werden sowert religiich als salide Metaslatenformate geliefert. Die signifikanten Bigerschaften der Daten sollen gemäß den Erwertungen der Zielgruppe (Notzer der Daten)
- Sz werden auszchließlich Metadaten in das Langsedanthis übernommen, zenen mindesten eine zigitale Reprisentation eindeutig zugeordnet werden kann.
- Es werden bei einer Lieferung ausschließlich Transfetpaliete in das Langseitsrichte übernommen. Ewintwell zusätzlich übermittelbe Dateien werden ignoriert.



216

"Packzettel": Metadaten über eine Lieferung

SubmissionManifestVersion: 1.3

SubmittingOrganization: Datengeber

Contact:

Kontakt Lieferinstitution, alternativ Urheber

ContactEmail:

TransferCurator:

Bearbeiter beim Datenlieferant

TransferCuratorEmail:

SubmissionName:

Benennung der Lieferung, zur Identifizierung

SubmissionDescription: beschreibt Inhalt der Lieferung

AccessRights:

Angaben zu Zugriffsrechten

License: Nutzungsbedingungen als Lizenzangabe

RightsDescription: Beschreibung von Rechteinhabern und Beschränkungen

DataSourceSystem: Software aus der exportiert wurde

MetadataFile:

Verweis und Format; deskriptive

MetadataFileFormat: Metadaten



Lieferung als Transferpaket

Packzettel submission-manifest.txt

Master-Dateien z.B. TIFF, WAV beschreibende Metadaten z.B. LIDO, MODS



