

Digitales Archiv NRW Köln/D

Begehung am 17.06.2013

Aufgaben	
Hauptaufgaben:	DA-NRW insgesamt: Aufbau einer nachhaltigen Infrastruktur zur langfristigen Archivierung digitaler Daten des Bundeslandes NRW HKI Köln: Entwicklung einer spartenübergreifenden Softwaresuite zur Vorbereitung und Erhaltung von Daten innerhalb einer Archiv-Infrastruktur
weitere Aufgaben:	DA-NRW insgesamt: Klärung nicht-technischen Aspekte (Betriebskonzept, Finanzierung, Rechtsfragen etc.)
nicht Teil der Aufgaben:	DA-NRW insgesamt: Entwicklung einer Forschungsumgebung HKI Köln: Entwicklung von Software-Komponenten zur fachspezifischen Datenkuratierung, Qualitätssicherung und Dokumentation von Daten
relevante Fachrichtung:	Einrichtungen des kulturellen Erbes in NRW (Bibliotheken, Archive, Museen etc.)

Allgemeine Angaben	
Rechtsform:	An der Entwicklung sind unterschiedliche Akteure beteiligt, u. a. das Institut für Historisch-Kulturwissenschaftliche Informationsverarbeitung (HKI) an der Universität zu Köln (Prof. M. Thaller), Langfristige Rechts- und Organisationsform noch nicht geklärt
Gründung/Geschichte:	Entwicklung Prototyp an der HKI: Pilotphase/Projektvorbereitung: 01.11.2010 – 30.04.2012 Ausbau: 01.05.2012 – 30.04.2014 Aktuell (12.2013): Weiterentwicklung Prototyp für zeitnahe Inbetriebnahme; insgesamt: Planungen zur Organisationsform "Digitales Archiv NRW"
Anzahl Mitarbeiter:	HKI Köln: 3 FTE Anzahl FTE (Planstellen) keine Anzahl Hilfskraftstellen (Planstellen) keine Anzahl FTE (befristet, Drittmittel) 100 % Anzahl Hilfskraftstellen (befristet, Drittmittel) unbekannt Anzahl WissMA keine Anzahl IT 3
Zertifizierung:	keine
Referenzmodell:	OAIS (SIP, AIP, DIP)

Datenbestand	
Verpflichtung Daten zu sichern?	konzipiert als Pflichtarchiv für Einrichtungen des Landes NRW (s.o.); rechtliche Grundlage noch nicht vorhanden
Findet gezielte Datenakquise statt?	nicht relevant
Anzahl an Projektarchiven:	nicht relevant

Anzahl an „grauer Literatur“:	nicht relevant
Anzahl an Doktorarbeiten:	nicht relevant
Anzahl an retrodigitalisierten Zeitschriften:	nicht relevant
Anzahl an Einträgen Nachweiskatalog:	nicht relevant
Gesamtgröße Archiv:	nicht relevant
Datensatzgröße:	unterschiedlich, da dies von den datenbesitzenden Instituten individuell festgelegt werden kann
Datenzuwachs jährlich:	nicht relevant
Altdateien vorhanden:	insgesamt bei Institutionen vorhanden, Relevanz für Entwicklung der DA-NRW Softwaresuite unbekannt

Datengenerierung	
technische Unterstützung bei Datenerhebung (z. B. durch Tools, Software etc.):	Nein
fachliche Unterstützung bei Datenerhebung:	Nein, liegt alleine bei den datenbesitzenden Institutionen

Datenupload	
Online-Upload möglich?	Ja
Auswahlkriterien für Aufnahme von Daten:	hängt von Vorgaben der Nutzer ab, die Softwaresuite selbst macht keine inhaltlichen oder technischen Vorgaben
Werden angebotene Daten abgelehnt? (Gründe)	hängt von Vorgaben der datenbesitzenden Institutionen ab
Vertragsverhältnis Datengeber - Datenarchiv:	noch nicht festgelegt
Software/Collection Registry:	nicht relevant

Kuratierung von Daten	
Arbeitsschritte zur Datenarchivierung:	<ul style="list-style-type: none"> - das Paket erhält eine URN als Paketname - enthaltene Dateien werden identifiziert (soweit in PRONOM Datenbank vorhanden) - enthaltene Dateien werden, soweit konfiguriert, in LZA Formate konvertiert - Original-Dateien sowie daraus konvertierte Dateien werden durch JHOVE (Extraktion technischer Metadaten) charakterisiert - Ergebnisse aller Operationen werden in PREMIS-Datei gespeichert - das Paket wird wieder als BagIt-Tar-Archiv gepackt und enthält entsprechenden Checksum-Daten
Werkzeuge für Datenarchivierung:	<p>Eigenentwicklung ContentBroker, SIP-Builder, DigitalArchiv-Web, OfficeConverter, EAD (Encoded Archival Description)-Viewer</p> <p>verfügbare Software iRODS; FIDO (Format Identification for Digital Objects); Fedora Repository</p> <p>(kommerziell, OpenSource)</p>
Versionskontrolle:	Ja

Verhältnis manuelle/auto-matisierte Datenkuratierung:	technische Kuratierung: 100 % (voll automatisch) inhaltliche Kuratierung: 0 % (manuell erforderlich)
Aufwand Datenkuratierung insgesamt (Personen, Zeit):	nicht relevant
wissenschaftliche Bewertung von Daten:	Nein
Verwendung Repository-System:	Fedora (evtl. Wechsel weil es zu langsam ist)
Software zur Verwaltung/Ablage der Archivdateien:	Eigenentwicklung verfügbare Software (kommerziell, OpenSource) ContentBroker Bemerkung: eine Versionierung der Archivalien wird unterstützt iRODS
Software zur Verwaltung Metadaten/Rechte/Versionen:	Eigenentwicklung verfügbare Software (kommerziell, OpenSource) Rechte: SIP-Builder Metadata: keine Verwaltung bzw. Ansicht über DA-Web Suche und Metadatenansicht über Fedora

Archivierung von Daten

technischer Partner für LZA:	Projektpartner: 20 Einrichtungen aus verschiedenen Kultursparten: - Ministerium für Familie, Kinder, Jugend, Kultur und Sport des Landes NRW (Projektleitung) - 3 Universitäts- und Landesbibliotheken, Landesarchiv NRW, kommunale Archive und ihre IT-Partner, ein Museum - Hist.-Kultwiss. Inform. der Univ. Köln (Softwareentwicklung) - d-NRW (Projektmanagement) - 3 Infrastruktureinrichtungen für Bitstream Preservation (LVR Infokom, RZ der Univ. Köln, hbz)
Anzahl redundanter Datenkopien:	wird an mind. 3 Orten/Knoten archiviert
Administration Archivsystem inhouse/extern:	bei den Projektpartnern/Knotenbetreibern
garantierter Zeitraum:	nicht relevant
Erfahrungen mit technischer Migration von Archivdateien:	Ja

Bereitstellung von Daten

Arbeitsschritte zur Datenbereitstellung:	- Zusammensetzen des aktuellen Bearbeitungsstandes aus der Datei-Urversion und Deltas - Aufbau des DIP mit PREMIS und Metadaten - Übertragung des DIP an Eigentümer/Contractor - auf Langzeitspeichern des Archivs werden DIPs nicht gespeichert, sondern nach erfolgreicher Übertragung gelöscht SONDERFALL: Presentation Repository/Presentation Information Package (PIP) - eingelieferte Daten aus dem AIP werden gemäß Contract in browserfähige Formate konvertiert - die so zu jeder URN neu entstandenen Dateien werden als PIP in ein Fedora-Repository im DA-NRW System übertragen - hier werden Daten in das Fedora eigene Datenmodell integriert
Werkzeuge für Datenbereitstellung:	Eigenentwicklung verfügbare Software (kommerziell, OpenSource) prepgui DFGViewer (Browser-Webdienst zur Anzeige von Digitalisaten aus dezentralen Bibliotheksrepositorien)

Registrierung bzw. Login für Nutzer notwendig?	Ja, für unbeschränkte Nutzung Zugriff gemäß eingeräumter Rechte
Wird die IP und der Zugriff geloggt?	unbekannt
Gibt es verschiedene Authentifizierungsstufen?	Ja
Zugriff:	public; Institution only; no publication
Wie wird ein Missbrauch von Daten verhindert/kontrolliert (z. B. wirtschaftliche Nutzung)?	SIP-Builder: Rechte-Einstellungen werden in einer Erweiterung des Premis-XML-Schemas vorgenommen
Rollenkonzept:	Archiv-Admin; Knoten-Admins; Contractor (einliefernde Institution)
Qualitätssicherung:	durch die einliefernde Einrichtung
Schutzmechanismen für bestimmte Informationen:	personenbezogene Metadaten/Dateien OAI-PMH erlaubt keine gezieltere Rechteeinschränkung raumbezogene Metadaten/Dateien OAI-PMH erlaubt keine gezieltere Rechteeinschränkung

Metadaten & Interoperabilität

Mindestanforderungen an Daten:	hängt von einliefernder Einrichtung ab
Verwendung von Standards:	PREMIS; Baglt SIP-AIP-DIP (Paketstandard)
Metadatenmodell:	Metadatenstandards: Dublin Core; MODS (Metadata Object Description Schema); XMP (RDF/XML) Metadatenmodelle: METS (Metadata Encoding and Transmission Standard); EAD (Encoded Archival Description); XMP; LIDO (Museum); EDM [Europeana Data Model] eigene Anpassungen an Metadaten? unterstützte Dateiformate werden konvertiert
vorhandene Schnittstellen:	OAI-PMH über Fedora
Sichtbarkeit der Metadaten in anderen Portalen/Aggregatoren:	Ja, z. B. Europeana, DDB
Sichtbarkeit der archivierten Dateien in anderen Portalen/Aggregatoren:	unbekannt
System für Persistente Identifikatoren:	Contractor-ID (Identifier im DA-NRW) Objektidentifier: URN (DA-NRW); Originalname (einliefernde Einrichtung); Package Name (DA-NRW); File ID (einliefernde Einrichtung); URN (einliefernde Einrichtung); HTTP-URI (DA-NRW) Anzahl PIDs insgesamt nicht relevant

Nicht-Technische Dienstleistungen

Bereitstellung Guidelines, Ratgeber etc.:	Benutzerhandbuch; Technische Dokumentation; Wiki; Buchpublikation; Videos
Antragsberatung:	Nein
Schulungen/Workshops:	auf Anfrage
Weiterentwicklung Standards:	Nein
Support Datenmanagement:	Nein

Finanzen: Einnahmen	
Jahresbudget:	davon speziell für das Datenzentrum unbekannt
Basisfinanzierung:	noch nicht festgelegt Zeitraum der vertraglich Finanzierung noch nicht festgelegt nachhaltige Finanzierung? noch nicht festgelegt
Drittmittel:	Entwicklung zu 100 % aus Drittmitteln (Ministerium NRW)
restliche Finanzierung:	unbekannt
Gebühren für das Archivieren von Daten:	nicht relevant
Gebühren für die Nutzung von Daten:	keine

Finanzen: Ausgaben		
Kostenstruktur nach Kostenträger:	Personalkosten	nicht relevant
	Planstellen	nicht relevant
	Drittmittelstellen	nicht relevant
	Hardware/Software	nicht relevant
	Reisekosten, Gebäudeinstandhaltung	nicht relevant
Kostenstruktur nach Kostenstellen:	Daten akquirieren und aufbereiten	nicht relevant
	Daten archivieren	nicht relevant
	Daten zugänglich machen	nicht relevant
	Softwareentwicklung	nicht relevant
	Kosten für PIDs	nicht relevant
Höhe Durchschnitts-Kosten pro 1 GB Daten:	nicht relevant	

Negativa	
nicht funktionierende/ optimierbare Workflows oder Prozesse:	unbekannt
nicht gelöste Probleme:	unbekannt