

Agenda/Protokoll des 11. Treffens

27.03.2019, Köln, ZB MED - Informationszentrum Lebenswissenschaften, 11:00-16:00 Uhr

Gleueler Str. 60, 50931 Köln

Sitzungsraum 4. OG (bitte am Haupteingang an der Ausleihtheke anmelden)

Anfahrt: Siehe <https://www.zbmed.de/bibliothek/medizinische-bibliothek-koeln/adresse-und-anfahrt/>

Anwesend: Andre, Heinz, Micky, Tim, Tobias, Konrad und Pia

1 Format-Policies

Präsentierende Institutionen:

- (SLUB sh. <https://www.slub-dresden.de/ueber-uns/slubarchiv/technische-standards-fuer-die-ablieferung-von-digitalen-dokumenten/>)
- Bericht von Tobias (DNB):
 - Individuelle Ingest-Policies beschreiben bei der DNB die verschiedenen Formatvorgaben für verschiedene Ablieferer.
 - Die Formatvorgaben basieren auf Pronom-Angaben.
 - Probleme werden prinzipiell direkt an Ablieferer rückgemeldet auf Basis der jeweiligen Vereinbarung (auch abhängig von dem jeweiligen Ablieferungsverfahren) und durch den Ablieferer behoben.
 - Die Prüfung erfolgt auf Basis des FITS-Tools, bei Verstoß gegen Ingest-Policy werden Tickets automatisch erstellt.
 - DNB hat Empfehlungen für Formate veröffentlicht, die sie gerne annehmen möchten (z. B. PDF/A), macht aber keine expliziten Vorgaben.
 - Bisher erfolgt nur temporäres CD-Imaging für die Bereitstellung, demnächst wird ein Workflow für die Überführung der Images nach EMil und perspektivisch für die Überführung ins Langzeitarchiv entwickelt. Formatcharakterisierung erfolgt innerhalb eines EMil-Moduls, das eine automatische Zuordnung für eine geeignete Umgebung auf Basis des Images und den Angaben aus DROID macht. Alternativ erfolgt ein Abgleich des Formats, das hauptsächlich auf einer CD zu finden ist, und dafür geeigneter Umgebungen.

Bericht von Andre (GESIS):

- Bei Aufnahme werden im einzelnen Lieferdateiformate abgesprochen.
- Bei Self-Archiving gibt es keine Vorgaben, hier wird nur Bitstream Preservation gewährleistet
- Intern allgemeine Abwendung von proprietären Formaten und Hinwendung zu Textformaten basierend auf ASCII und XML, aus denen dann bei Bedarf SPSS-Daten generiert werden können.
- Durch über Jahre hinweg erfolgte Konvertierungen von SPSS-Dateien sind Artefakte entstanden

- Versionierungsschritte aus der Mainframe-Zeit werden nicht nachgehalten, alles, was nach dem Mainframe-Wechsel kam, schon.

Bericht von Heinz (DLA)

- Nach dem Betriebsschluss beim BSZ führt das DLA SWBContent auf eigenen Servern weiter. Archivierte Netzliteratur steht Nutzenden damit unverändert zur Verfügung, aber die Harvesting-Komponente ist stillgelegt, es kommt also kein neuer Content dazu. Wie das Harvesting künftig erfolgen kann, ist derzeit noch offen.
- Beim Harvesting erfolgte keine Konvertierung (bis auf Ausnahme von Screencasts), sondern nur Archivierung in WARC
- JPGs und TIFFs aus Nachlässen bleiben unverändert
- Es gibt nur eine implizite Formatpolicy des DLA bei im Haus produzierten Daten: Bei Scans werden komprimierte TIFFs erstellt, bei Video-Mittschnitten die gesendeten Formate (MPEG und MP4(H.264), bei Audio-Aufnahmen und mittschnitten unkomprimiertes Wave, bei Texten PDF/A (mittels FreePDF) und RTF.

Bericht von Konrad (LATH)

- Haben Übernahmeformate definiert, aber es gibt noch keine strikte Policy
- System ist momentan noch nicht im Produktivbetrieb
- Konvertierung erfolgt momentan nur in PDF

Bericht von Tim (BArch)

- BArch hat momentan keine Formatpolicy
- Aber: Im Zusammenhang mit dem Beschaffungsprozess für neue Magazinierungssoftware macht BArch sich Gedanken über die Formate, die es im LZA haben möchte
- Interne und externe Digitalisierung: LZA-Format ist TIFF, Migration der Daten in JPEGs für den Viewer bei Invenio. Bei hochauflösten Daten (z.B. Karten) ist trägerlose Übertragung momentan noch schwierig.
- Bei Filmen: mxf- dcp- Repräsentationen auf Speicher, Nutzungsformate: MP4 und avi
- Bei Audio wave, als Nutzungsformat aber MP3 angedacht
- Für einen groben Überblick bei der Inventur alter Speichersysteme nutzt Tim [TreeSize](#), um Dateistrukturen auf alten Speichern auszulesen. Bisher erfolgt noch keine Migration.
- BagIt wird für SIPs genutzt, TAR für AIPs
- Neue Magazinierungssoftware soll voraussichtlich nicht-deskriptive Metadaten in die SIPs schreiben.
- Im Zusammenhang mit Anschaffung eines neuen Bandspeichersystems ist zukünftig auch eine georedundante Speicherung im BArch vorgesehen

Bericht von Pia (ZB MED)

- ZB MED hat im Haus bis auf bei den Dissertationen, die als PDF abgegeben werden sollen, keine Format-Policy. Die Abteilung "Forschungsdatenmanagement" arbeitet gerade an allgemeinen Empfehlungen.

- Um Open-Access-Veröffentlichungen voranzutreiben und niemanden von der Publikation über die ZB-MED-Plattformen abzuhalten, macht ZB MED keine Vorgaben für Textpublikationen oder Forschungsdaten, sondern nimmt prinzipiell alle Formate an.
- Im LZA-System sind momentan hauptsächlich PDFs mit ein paar vereinzelt Bilddateien. Migrationen sind bisher nicht erfolgt

Bericht von Micky (TIB)

- Regelung für Dissertationen: Empfohlen wird PDF als Ablieferungsformat, aber keine strenge Vorgabe; Überprüfung abgelieferter Dissertationen erfolgt mittels JHOVE, keine Dateiformatvorgaben für Anhänge von Dissertationen. Nicht-wohlgeformte / nicht-valide Objekte werden durch das Team LZA untersucht und zur Behebung zurück an den Promovenden gegeben. Hier erfolgt Beratung bzgl. regelkonformer Erstellung bis hin zur Erstellung einer regelkonformen Datei durch die TIB und anschließende Überprüfung und Freigabe durch den Ablieferer.
- In-House Formatpolicy für Digitalisate 2D: TIFF mit eingebetteten ICC-Profilen inkl. Vorage bzgl. Tags (neben Formatvalidierung somit Policyvalidierung)
- In-House Formatpolicy für AV-Digitalisate (aktuell im DELFT Projekt): Film- und Video in Archivformat MKV mit FFV1v3 und WAVE / PCM Linear, Accessformat mit h264 Codec.
- Mindestanforderung im Archiv ist, dass Datei identifizierbar ist
- Rosetta erlaubt regelbasierte Behandlung von Validierungs- und Identifizierungsergebnissen, z.B. Mapping auf bestimmten PUID bei multiple-format hits und homogenen, zuvor geprüften Datenstand
- Ansonsten keine Formatvorgaben für Ablieferungen, aktuell 98 verschiedene Dateiformate (nach PUID) im LZA-System der TIB, darunter aber auch verschiedene PDF-Versionen

2 Updates der Arbeitspakete

R.O.S.T.

Siehe auch 3.

Passend zum Beginn des iPRES-Review-Prozesses ab 18.3.2019 soll die ROST-Umfrage erneut aktiviert werden (Preview bis dahin:

<https://survey.gesis.org/nestor/index.php/survey/index/sid/811173/newtest/Y/lang/de>). Sie wurde etwas überarbeitet, um auch unvollständige Angaben besser auswerten zu können. Die Umfrage soll auf der iPRES beworben werden (ist bisher allerdings auf die deutschsprachige Community begrenzt) und läuft daher mindestens bis Ende 2019.

Zu diskutieren: Wie und wo und durch wen wird die 2. Runde der Umfrage beworben?

- Es soll bei der Bewerbung darauf hingewiesen werden, dass TeilnehmerInnen, die bereits an der Umfrage teilgenommen haben, nur noch einmal teilnehmen sollen, wenn es substantielle Änderungen bei ihnen gibt.
- Heinz entwirft einen kurzen Ankündigungstext auf Basis der ersten Runde.

- Micky kontaktiert Verantwortliche von [MediaPedia bei der NLA](#) zur möglichen Nachnutzung von Holotypes zur Auswertung, um auf MediaPedia-Datenbank verlinken zu können.
- Andre regt an, eine Reminder-Strategie zu entwickeln, um Teilnehmende, die die Umfrage unterbrochen haben, an Beendigung der Umfrage zu erinnern. Heinz schlägt vor, nicht die Teilnehmenden konkret anzumailen, sondern allgemein an die Umfrage zu erinnern, z.B. auch beim Praktikertag.
- Alle AG-Mitglieder sprechen mögliche Interessenten aus ihrem Umfeld gezielt an. Insbesondere auch solche, die sich in der ersten Runde interessiert gezeigt, aber nicht teilgenommen haben, oder auch erst später durch die Ergebnisse davon erfahren haben.
- Micky kontaktiert nestor-Geschäftsstelle zur Bewerbung im Newsletter und außerdem sollen verschiedene Mailinglisten genutzt werden.

3 iPRES-Beitrag der AG zu R.O.S.T.

Andre, Micky, Anne Krause und Heinz haben mehrere Telkos durchgeführt und ein Proposal für ein E-Poster für die iPRES 2019 erarbeitet und eingereicht, das die R.O.S.T.-Umfrage und deren (Zwischen-)Ergebnisse vorstellt. Über die Annahme wird bis zum 26.4.2019 entschieden. Im positiven Fall wird auch Andre zur iPRES fahren können. Micky, Pia und Heinz sind bereits unabhängig davon angemeldet (edit: Wurde leider abgelehnt).

Anne Krause hat ein faires Angebot erstellt, um das E-Poster und die neuen Ergebnisse der Umfrage zu gestalten (vgl. ihre Arbeit für die Ergebnisse von 2018). Die Finanzierung durch nestor ist gewährleistet, jedoch müssen aus formalen Gründen zwei weitere Vergleichsangebote eingeholt werden.

4 Praktikertag 2019 in Göttingen am 5. Juni 2019

<https://www.langzeitarchivierung.de/Subsites/nestor/DE/Veranstaltungen/TermineNestor/praktikertag2019.html> führt unsere AG bisher nicht auf.

Rückmeldung aus nestor-Koordinierungsgruppe (26.3.): Wenn die AG beim Praktikertag vortragen möchte, besteht die Möglichkeit, dies in einem ca. 20 Minuten-Slot unter „Aktuelle Aktivitäten“ zu machen. (ab 15:30 Uhr) Rückmeldung dazu bitte an die nestor-Geschäftsstelle bis Ende der Woche.

Heinz möchte gerne die Zwischenergebnisse der ROST-Umfrage auf dem Praktikertag präsentieren, ggf. kommt Andre auch mit. Pia gibt die Info an die nestor-Geschäftsstelle weiter.

[Nachrichtlich 6.6.2019: Hier ist die Präsentation vom Praktikertag: [NPT-2019-Gö-AG-Formaterkennung-Kramski_v04.pdf](#).]

5 Bericht von der 23. AUdS Prag

- Informationen zum Vortrag des LATH beim AUdS-Camp: Vortrag wurde von Christine und Daniel Wittmann ausgearbeitet und gehalten; der vollständige Vortrag macht in unser Runde eher keinen Sinn. Konrad kann die Kernpunkte vortragen, so dass wir uns dazu austauschen können.

- Dieser und andere Vorträge werden in Kürze auf der [Seite des Staatsarchivs St. Gallen](#) veröffentlicht.

6 Verschiedenes

Welchen Status hat eigentlich der Webrelaunch, dessen Preview (https://test04.gsb.itzbund.de/Webs/nestor/DE/Arbeitsgruppen/AG_Formaterkennung/ag_for_materkennung_node.html) wir im Dezember 2018 angesehen und teilweise kommentiert haben?

--> Beim Treffen der nestor-Koordinationsgruppe am 26.3. wurde der Relaunch für Mitte Mai angekündigt und gesagt, dass es noch eine finale Runde zur Korrektur geben soll. Pia leitet Zugangsdaten an den AG-Verteiler weiter, sobald diese vorliegen.

Tim hat auf die Veranstaltung "[E-ARK - EArchiving Building Block - Auch für uns in Deutschland?](#)" am 25.4. in Berlin hingewiesen.

Der DFN-Verein hat seit ca. 11/2018 einen schönen, voraussetzungsarmen, neuen Dienst für Video- und Telefonkonferenzen: <https://www.conf.dfn.de/>. Da sicher viele hier auch Mitglied im DFN-Verein sind, dürfte das für diese AG eine interessante Infrastruktur darstellen. Der Zugang für Gäste erfordert keinerlei Account, nur eine URL. Heinz richtet der AG einen Raum ein.

7 Nächstes Treffen

5.11.2019 in Kiel