

Virtuelles Treffen der nestor AG SIP-Konkretisierung am 19. April 2023

Teilnehmer*innen: Tobias Steinke, Jens Steidl, Franziska Schwab, Björn Steffenhagen

Die Veröffentlichung der Dokumente zu konkreten SIP-Spezifikationen ist nach wie vor nicht erfolgt, da der Beitrag vom BSZ nicht wie zugesagt von Herr Wolf gekommen ist. Es wird daher entschieden, alle vorliegenden Beiträge zu veröffentlichen. Mögliche Aktualisierungen (etwa die Überprüfung von Links) sollten möglichst bald (innerhalb der kommenden Woche) erfolgen.

Die Struktur der öffentlich zugreifbaren Wiki-Seite der AG bei <https://wiki.dnb.de/pages/viewpage.action?pageId=259632402> wird als sinnvoll angesehen. Die Übersicht am Anfang für alle Partner müsste noch vervollständigt werden. Idealerweise kann dann dort auf die an anderer Stelle bei nestor publizierten Beiträge verlinkt werden. Für eine bessere Nutzbarkeit, Suche und Übersicht wird verabredet, die Listen in den einzelnen Abschnitten als Tabellen anzulegen, ergänzt mit geeigneten Spalten zur Sortierung. Eine Überführung aller Abschnitten in eine große Tabelle wird als zu unübersichtlich abgelehnt. Im Abschnitt "Tools" soll auch ein Verweis auf die nestor-Toolbox erfolgen.

Berichte über aktuelle Entwicklungen

Base4NFDI-Antrag zur LZA

Im großen NFDI-Rahmen (viele Projekte zum Aufbau einer nationalen Forschungsdateninfrastruktur) gibt es einen Unterbereich für übergreifenden Dienste (Base4NFDI) und darin wiederum wird ein Projektantrag zur Langzeitarchivierung vorbereitet. Gegenstand ist die Konzeption und spätere Implementierung einer Art Broker-Dienst, welcher über einheitliche Verfahren und Schnittstellen vorhandene LZA-Dienste und -Systeme zur Archivierung von Forschungsdaten vermittelt. Dabei geht es auch um Standards und somit potenziell ebenfalls um SIP-Formate. Hier könnte die Handreichung unserer nestor AG herangezogen werden.

Erfahrungen mit Problemen bei Ingest-Workflows

Von der TIB wird von Problemen mit Formeln (mathematisch, chemisch, etc.) in Metadaten bei der Übernahme aus der OAI-Schnittstelle berichtet. Bei E-Journal-Artikel haben Verlage offensichtlich Probleme mit einer geeigneten Kodierung solcher Formeln, die auf deren Webseiten in verschiedenen Formen integriert sein können (TeX, XML, als Grafiken), weshalb über die OAI-Schnittstelle dann Metadaten unvollständig und fehlerhaft sind. Das macht die Bedeutung von Qualitätssicherungsmaßnahmen in den verschiedenen Phasen des (Pre-)Ingest deutlich. Ggf. könnte die AG zukünftig eine weitere Handreichung erstellen, die eher prozessorientiert den Ingest-Vorgang betrachtet. Dafür müsste allerdings der nestor/DIN-Ingest-Leitfaden berücksichtigt werden.

Grundsätzlich sind Erfahrungsberichte von Problemen oder gar Fehlschlägen bei der Langzeitarchivierung sinnvoll und lehrreich. Ein Bereich, bei dem es schon mehrfach Fehlschläge gab, ist der Aufbau einer umfangreichen internationalen File Format Registry für

die digitale Langzeitarchivierung. Der ambitionierteste Ansatz war wohl die UDFR (<https://web.archive.org/web/20220120034056/http://udfr.org/project/UDFR-final-report.pdf>). Zudem gab es einen gescheiterten EU-Projektantrag und ein daraus hervorgegangenes Vorhaben mehrere Nationalbibliotheken, welches an rechtlichen Problemen scheiterte. Letztlich hat sich nur PRONOM durchgesetzt, was allerdings viele Schwächen hat. Wikidata wird vielfach als mögliche Ergänzung bzw. Basis für eine eigene File Format Registry gesehen