

Virtuelles Treffen der nestor AG SIP-Konkretisierung am 22. Januar 2024

Teilnehmende: Tobias Steinke, Björn Steffenhagen, Jens Steidl, Kai Naumann, Franziska Schwab

Wiki-Seite

Die Wiki-Seite (<https://wiki.dnb.de/pages/viewpage.action?pageId=259632402>) nimmt Struktur an und hat bereits Inhalte, ist aber noch nicht bereit für eine weitere „Veröffentlichung“. Der Tools-Abschnitt sollte für jeden Eintrag neben einem Link zum Tool auch – wenn vorhanden – einen Link zum entsprechenden Eintrag bei COPTR bekommen. Eine weitere Spalte heißt nun „Standards“ und kann sowohl referenzierte Standards des Tools als auch durch das Tool implementierte Standards auflisten. Es wird eine Verbindung zur bzw. von der nestor-Toolbox und der dortigen Bibliographie über Zotero diskutiert. Generell kann jeder AG-Teilnehmer jederzeit an der Seite arbeiten und Inhalte ergänzen.

Berichte über aktuelle Entwicklungen

KI-Konferenz in der DNB

Im Dezember fand eine Fachtagung „KI in Bibliotheken - Neue Wege mit großen Sprachmodellen?“ in der DNB statt (Programm und Präsentationen: <https://wiki.dnb.de/pages/viewpage.action?pageId=306153594>). Obwohl dort die Bildung von Informationspaketen und Langzeitarchivierung keine Themen waren, stellt sich doch die Frage, ob große Sprachmodelle in diesen Bereichen Anwendung finden oder darauf Auswirkungen haben könnten. Trainingsdaten und die Modelle selbst könnten als SIPs betrachtet werden, wenn auch vor allem im Bereich Forschungsdaten. Standards könnten mit Hilfe von KI (großen Sprachmodellen) zukünftig weiterentwickelt werden. Die Auswahl und Bewertung, was und in welchen Objekteinheiten als SIP gebildet wird, könnte KI unterstützt erfolgen.

Bereits jetzt werden große Sprachmodelle im Retrieval von Bibliotheken eingesetzt. Dabei kann eine Anfrage in natürlicher Sprache gestellt werden und die Antwort wird dann auf Basis vorgegebener Datenquellen wie z. B. allen digitalisierten Dokumente eines Archivs gegeben. Bei den Datenquellen kann wiederum über vorhandenen Suchmaschinen eine gefilterte Auswahl für die zu berücksichtigenden Inhalte erfolgen.

DIMAG

Das Kernmodul von DIMAG wird ab diesem Jahr neugestaltet und dabei kann auch das Metadatenmodell neu angefasst werden. Bisher wurde das PREMIS-Datenmodell nicht komplett übernommen. Zukünftig könnten sich die Felder stärker den Definitionen annähern.

Allerdings hat keine Institution bzw. kein System die PREMIS-Spezifikationen komplett implementiert. Es gibt zwar bereits seit 2015 ein Dokument „Conformant Implementation of the PREMIS Data Dictionary“ (<https://www.loc.gov/standards/premis/premis-conformance-20150429.pdf>), welches auch 2021 in der nestor-Veröffentlichung „PREMIS verstehen“ (<https://www.langzeitarchivierung.de/Webs/nestor/SharedDocs/Downloads/DE/berichte/pREMIStverstehen2021.html>) benannt wurde. Bisher wird eine PREMIS-Konformität danach jedoch nicht breit bekannt gemacht von Systemen.

Bundesarchiv

Bisher werden in den digitalen Systemen keine der bekannten internationalen Standards genutzt. Zukünftig soll allerdings METS zum Einsatz kommen. Dabei sollte berücksichtigt werden, dass inzwischen METS 2 vorgestellt wurde, was Vereinfachungen gebracht und Inkonsistenzen beseitigt hat. Das weit verbreitete METS 1 bleibt aber zusätzlich bestehen.