

Virtuelles Treffen der nestor AG SIP-Konkretisierung am 16. Dezember 2025

Teilnehmende: Tobias Steinke, Franziska Schwab, Alina Efken, Svenja Zilian, Kai Naumann

Wiki-Seite

Der Eintrag zu Matterhorn METS wurde überarbeitet. Dieses SIP-Format wird im kommerziellen System docuteam cosmos genutzt, was u. a. von der Schweizerischen Nationalbibliothek eingesetzt wird. Als weitere Einträge könnten eingesetzte Tools bei der TIB, sowie Formate des Systems Preservica ergänzt werden. Der unvollständige Eintrag von Yvonne Tunnat in der Spezifikationentabelle sollte noch einen Namen und einen Link erhalten.

Berichte über aktuelle Entwicklungen

iPRES 2025

Die iPRES fand dieses Jahr in Neuseeland statt und konnte auch virtuell live oder in Aufzeichnungen verfolgt werden. Hier ein kurzer Bericht zu einigen Präsentationen dort mit Bezug zum Themenfeld der AG:

In der Session „In the Lab: Tools & Technologies“ stellt zuerst Andy Jackson von der DPC eine Studie Quantität von Dateiformaten in Format Registries vor, wobei die Anzahl der Dateierendungen gezählt wurden. Das Spektrum reichte dabei von 7628 in Wikidata bis 1678 in PRONOM. Er schätzt, dass in der Praxis ca. 12.000 Extensionen genutzt werden, also viele, die noch nicht in diesen Datenbanken erfasst sind. Preservica stellte ihre „Rapid Assessment Methodology for Evaluating File Format Obsolence Risk“ vor. Dabei werden acht Kriterien nacheinander durchgegangen und sobald dabei als Ergebnis Risiko bzw. kein Risiko erkannt wird, ist das automatisch die Gesamteinschätzung. Zuerst wird überprüft, ob der Hersteller des Formats das Ende des Supports angekündigt hat (Risiko). Dann ob aktuelle Systeme es öffnen können (kein Risiko), ob es von einem Standards-Gremium oder von einer Open Source Organisation gepflegt wird (jeweils kein Risiko), die damit assoziierte Software nicht mehr verfügbar ist (Risiko), die Nachfolgesoftware das Format nicht mehr unterstützt (Risiko), die Webpräsenz zum Format nicht mehr existiert (Risiko) und schließlich ob es niemals eine Webpräsenz dazu gab (Risiko). Die FH Potsdam stellte das unter museum-digital.org verfügbare digitale Archivsystem auf Basis von Archivemata vor. Auch der Dienstleister CSC aus Finnland zeigte seinen National Digital Preservation Service mit SIPs im eigenen METS Profile inklusive PREMIS. Ein medizinisches Institut aus Peking präsentierte ihre Archivierungslösung für Prompts für große Sprachmodelle mit eigenem AIP inkl. Informationen zum Modell und Parametern.

In einer weiteren Session „Formats in Focus: Strategies“ stellte das DPC eine Unterscheidung von Hazards und Risks bei der Formatidentifikation vor. Während Risks in die Zukunft verweise, seien Hazards Probleme, die bereits jetzt relevant sind. Letzteres wird wiederum in zwei Kategorien betrachtet: beschädigter, falsch geformter oder gefährlicher Inhalt, sowie unbekannter, falsch erkannter oder fehlender Inhalt. Zudem seien unentdeckte Abhängigkeiten zwischen Dateien ein großes Problem.

Unter dem vielversprechenden Titel „Disruptive Intelligence: Rethinking Archives with AI and Blockchain“ stellten zwei Vorträge die KI-gestützte Gesichtserkennung für Bildarchive, sowie die Erhaltung von Authentizität mit Blockchain vor.

Bei der Session „Decoding Formats & Meaning“ ging jemand aus Melbourne auf die Problematik von ZIP-Dateien als Black Boxes in Langzeitarchiven am Beispiel von Preservica ein. Dabei stelle sich die grundsätzliche Frage, ob ZIP-Dateien entpackt werden und zusammen mit den Inhalten archiviert werden sollen und was in den

Metadaten dazu stehen soll. Preservica arbeitet an einer eigenen Lösung für die Problematik. Der Umgang mit ZIP-Dateien wird auch in der AG besprochen. Die TIB entpackt ZIPs vor der Archivierung. Probleme könnte dabei durch Referenzierungen aus HTML-Dateien auf die ZIP-Datei entstehen, die dann nicht mehr funktionieren. Das Bundesarchiv hat bei Akten keine ZIP-Dateien, da dies bei E-Akte Bund kein erlaubtes Dateiformat ist. Allerdings wäre das bei der E-Mail-Archivierung für Anhänge potenziell relevant. Bei einem weiteren Vortrag in der Session wurde vom Einsatz von ARKs als Persistent Identifiern in französischen Archiven und Museen berichtet.

EOSC EDEN

Am 1. Dezember gab es ein Webinar zur Vorstellung der im EU-Projekt EOSC EDEN definierten Core Preservation Processes (<https://zenodo.org/records/16992452>). Eine Aufzeichnung davon ist verfügbar: <https://www.youtube.com/watch?v=OVBQmRYoNCg>