

nestor«, das Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland¹, veranstaltete auch im Jahr 2017 wieder einen nestor-Praktikertag. Dieser bringt regelmäßig Fachleute aus unterschiedlichen Communities zusammen, um sich über neue Entwicklungen und praktische Ansätze in der digitalen Langzeitarchivierung zu informieren und auszutauschen.

Im Fokus stand dieses Jahr das Thema »Formaterkennung, Formatvalidierung und Tools«, welches am 28. Juni rund 75 Teilnehmerinnen und Teilnehmer in die Deutsche Zentralbibliothek für Wirtschaftswissenschaften (ZBW) nach Kiel brachte. Das Tagungsprogramm wurde maßgeblich von der nestor-AG Formaterkennung gestaltet. Die ZBW stellte ihre Räumlichkeiten zur Verfügung und organisierte die Verpflegung. Auch dieses Jahr waren verschiedene Aussteller am Praktikertag beteiligt, welche Lösungsansätze an Präsentationsständen vorstellten: scope solution ag, PDF/Association, Startext GmbH und InSoft Infotel Software GmbH.

Dem Thema Formaterkennung kommt große Bedeutung zu, weil sie einer der ersten Prozesse ist, welche Materialien in einem digitalen Langzeitarchiv durchlaufen. Durch die Formaterkennung werden Kennzahlen für die Erhaltungsplanung ermittelt und Metadaten für die bearbeiteten Objekte generiert. Sie dient auch der Risikoplanung bei der Einführung neuer Tools oder zur Abschät-

zung, ob der Bestand großflächig migriert werden sollte. Die verschiedenen Vorträge des Tages verdeutlichten die Bandbreite an Herausforderungen bei der Formaterkennung und Validierung, aber auch den Mehrwert für die vorausschauende Planung von Erhaltungsstrategien.

Nach der Begrüßung der Teilnehmerinnen und Teilnehmer durch die gastgebende ZBW und die nestor-Geschäftsstelle, beleuchtete Ange Albertini (Zürich) das Thema Dateiformate aus der Perspektive Sicherheit und Schadsoftware. Er zeigte bestehende Sicherheitslücken auf, die durch veraltete Spezifikationen und fehlertolerante Ausgabesoftware entstehen.

Durchaus kritisch betrachtete Claire Röthlisberger-Jourdan (KOST Schweiz) die Formaterkennung als Teil des Archivierungsworkflows. Da bei der Formaterkennung nur minimalste Merkmale geprüft werden, können fehlerhafte Metadaten entstehen, die erst wieder bei der Validierung gefunden und korrigiert werden können. Zudem ändern neue Formate die Erkennungsmethoden von alten Formaten, sodass jede Änderung einen neuen Formaterkennungsprozess anstoßen sollte. Da auch die Formaterkennungstools sich stetig verbessern, kam sie zu dem Fazit, dass die dauerhafte Speicherung der Format-Metadaten in Zukunft überflüssig werden könnte.

Marco Klindt (ZIB) führte auf eindruckliche Art und Weise in die Welt der Bildformate und Farbkodierung ein. Mithilfe der Bildformate lassen sich Farben als Daten abspeichern, sodass diese auf möglichst allen Ausgabegeräten korrekt angezeigt werden können. Dabei sollte auch auf die Angabe der Farbintensität geachtet werden, die bisher nur in bestimmten Farbmodellen berücksichtigt wird. Um die verschiedenen Möglichkeiten der Farbkodierung nachhaltig korrekt ausgeben zu können, ist es sinnvoll, zusätzlich ein Farbprofil abzulegen. Dadurch kann sichergestellt werden, dass die Bilddatei auf dieselbe Art und Weise wie zum Zeitpunkt der Erstellung ausgegeben wird.

Im Anschluss fand die erste Austauschrunde zwischen den Teilnehmerinnen und Teilnehmern statt. Die Gruppe diskutierte die Vorteile der Formaterkennung und tauschte sich über den Einsatz gängiger Tools und den Aufbau von Langzeitarchivierungssystemen aus.

Yvonne Tunnat (ZBW Kiel), in Vertretung für Michelle Lindlar (TIB Hannover), ging auf das Validierungstool JHOVE ein. Zum Einstieg verdeutlichte sie den Unterschied zur bloßen Formaterkennung und den Mehrwert der Validierung für den gesamten Prozess der digitalen Langzeitarchivierung. Während der Formatvalidierung werden Dateien gegen Formatspezifikation geprüft und



1 Sabrina Kistner Hidalgo begrüßt die Veranstaltungsgäste

Foto: ZBW / Sven Wied

somit die Gültigkeit der Datei festgestellt. Ein Großteil von fehlerhaften Dateien kann in diesem Prozess gefunden werden. Da Formate und Tools weiterentwickelt werden, sollten auch die Validierungswerkzeuge getestet und stets verbessert werden. Durch die Arbeitsgruppen bei nestor und der Open Preservation Foundation wurden Benchmark-Tests für JPEG² und TIFF³ durchgeführt. Zuletzt wurde JHOVE an einem Corpus synthetisch erstellter PDFs geprüft und mit anderen Validierungstools verglichen. Die Tests bestätigen JHOVE als fähiges Validierungstool und legten gleichzeitig Schwachstellen offen, die von der Community in den Arbeitsgruppen weiterhin bearbeitet werden.

Darauf folgte ein Vortrag von Dietrich von Seggern (Callas Software) zum Thema PDF-Validierung mit dem Schwerpunkt auf Schriften in PDF-Dokumenten. Ferner wurde die Entwicklung des Open-Source-Validators veraPDF im Rahmen des EU Projekts PREFORMA⁴ vorgestellt. In Folge der Entwicklung entstand unter anderem ein umfangreicher Test-Corpus von PDF/A-Dateien mit verschiedenen Konformitätsleveln. Desweiteren konnten mehrere Unklarheiten in der PDF/A-Spezifikation geklärt werden, welche demnächst in einer neuen Tech Note veröffentlicht wird. Das veraPDF-Konsortium wird die Weiterentwicklung in den kommenden zwei Jahren weiter vorantreiben.

Die Veranstaltung endete mit der zweiten Austauschrunde, die trotz fortgeschrittener Zeit rege in

Anspruch genommen wurde. Der nächste nestor-Praktikertag wird 2018 in der Deutschen Nationalbibliothek in Frankfurt am Main stattfinden. Alle Vorträge sind über die nestor-Webseite verfügbar: <https://tinyurl.com/zaqxm56>

Anmerkungen

- 1 www.langzeitarchivierung.de; Das Kompetenznetzwerk versteht sich als Anlaufstelle für Fragen rund um die digitale Langzeitarchivierung. nestor wurde im Rahmen eines BMBF-geförderten Projektes (2003-2009) begonnen und wird seit Juli 2009 von den ehemaligen Projektpartnern gemeinsam mit weiteren Einrichtungen selbstständig weitergeführt.
- 2 <http://openpreservation.org/blog/2016/11/29/jpegvalidation/>
- 3 <http://openpreservation.org/blog/2017/01/17/tiff-format-validation-easy-peasy/>
- 4 www.preforma-project.eu/

Die Verfasserinnen

Sabrina Kistner Hidalgo, Deutsche Nationalbibliothek, nestor-Geschäftsstelle, Adickesallee 1, 60322 Frankfurt am Main, Telefon 069 1525-1141, s.kistner@dnb.de

Vera Marz, Deutsche Nationalbibliothek, nestor-Geschäftsstelle, Adickesallee 1, 60322 Frankfurt am Main, Telefon 069 1525-1757, v.marz@dnb.de



2 Erfahrungsaustausch und Netzwerken in der Kaffeepause

Foto: ZBW / Sven Wied