

Open-Access-Publikationen mit OS-APS medienneutral und mit automatisiertem Corporate Design erstellen

Anforderungserhebung, Schlussfolgerungen für den Publikationsworkflow und Stand der Umsetzung

Carsten Borchert, SciFlow GmbH, Berlin

Astrid Hoffmann, Universitätsbibliothek Erlangen-Nürnberg

Zusammenfassung

Die Universitätsbibliothek Erlangen-Nürnberg, die Universitäts- und Landesbibliothek Sachsen-Anhalt und das Technologieunternehmen SciFlow in Berlin entwickeln gemeinsam die „Open Source Academic Publishing Suite (OS-APS)“, die es Verlagen ermöglichen soll, E-Books einfach und automatisiert ihren Formatvorgaben entsprechend zu erstellen. OS-APS soll einen XML-basierten Single-Source-Publishing-Workflow ermöglichen. Die Software wird Import-, Editier- und Exportfunktionen integrieren, sodass die in Verlagen üblichen Manuskriptformate (z.B. DOCX oder ODT) mit wenigen Mausclicks importiert, analysiert und in das gewünschte Format (z.B. EPUB, PDF, HTML oder XML) im Corporate Design des Verlags umgewandelt werden können. Darüber hinaus werden die Publikationsplattformen Open Journal Systems (OJS) und Open Monograph Press (OMP) sowie die Repositoriumssoftware DSpace an OS-APS angebunden. Um sich im Vorfeld ein Bild von den Anforderungen der Verlage an die Software zu machen, wurde zu Beginn des Projekts eine Anforderungsanalyse in Form einer Online-Umfrage durchgeführt. Dabei wurden Fragen zu Manuskripteingang und -bearbeitung, Publikationsformaten und Zukunftspotenzialen im Verlagswesen gestellt. Die Ergebnisse dieser Umfrage sind in die Entwicklung mit eingeflossen. Im Rahmen des vorliegenden Papers erfolgt zunächst eine kurze Einführung in die Grundlagen des Publizierens mit XML, bevor das BMBF-geförderte Projekt OS-APS inklusive des angedachten Workflows und der einzelnen Softwarekomponenten vorgestellt wird. Im Anschluss daran werden die Ergebnisse der Anforderungsanalyse sowie die daraus entstandenen Schlussfolgerungen für den Workflow ausgeführt. Abschließend wird auf die Einbeziehung der Community und die nächsten anstehenden Entwicklungsschritte eingegangen.

Summary

The University Library of Erlangen-Nuremberg, the University and State Library of Saxony-Anhalt and the technology company SciFlow in Berlin are jointly developing the “Open Source Academic Publishing Suite (OS-APS)”, which is supposed to enable publishers to create e-books easily and automatically according to their style guidelines. OS-APS will enable an XML-based single-source publishing workflow. The software will integrate import, editing and export functions so that the manuscript formats that commonly occur in publishing houses (e.g., DOCX or ODT) can be imported, analysed and converted into the desired format (e.g., EPUB, PDF, HTML or XML) in the publisher’s corporate design with just a few mouse clicks. In addition, OS-APS will connect to the publication platforms Open Journal Systems (OJS) and Open Monograph Press (OMP) as well as the repository software DSpace. In order to get an idea of the publishers’ requirements for the software in advance, a requirements analysis in the form of an online survey was conducted at the beginning of the project. Questions were asked about manuscript submission and processing, publication formats and future potential in publishing. The results of this survey have been incorporated into the development.

This paper begins with a brief introduction to the basics of publishing with XML before presenting the BMBF-funded project OS-APS, including the envisaged workflow and the individual software components. This is followed by the results of the requirements elicitation and the resulting conclusions for the workflow. Finally, the involvement of the community and the next steps regarding the development are discussed.

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5859>

Autorenidentifikation: Borchert, Carsten: GND: [1105075834](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-63864-p0011-9); ORCID: <https://orcid.org/0000-0002-3981-4517> ; Hoffmann, Astrid: ORCID: <https://orcid.org/0000-0003-4106-3561>

Schlagwörter: Medienneutrales Publizieren; automatischer Satz; Open Access; Open Source

Dieses Werk steht unter der [Lizenz Creative Commons Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/).

1. Einleitung

Die Software „Open Source Academic Publishing Suite“, kurz OS-APS, soll es kleinen und mittelgroßen akademischen Verlagen ermöglichen, ihre Arbeitsabläufe weiter zu automatisieren. Um bei Open-Access-Publikationen im akademischen Bereich bei gleichzeitig relativ geringen Ressourcen einen hohen qualitativen Standard zu erreichen, ist ein hoher Automatisierungsgrad des Publikationsworkflows unabdingbar. Das Rückgrat eines automatisierten Workflows sind medienneutrale Manuskripte, die meist auf XML basieren.¹

„XML“ steht für „eXtensible Markup Language“. Es handelt sich hierbei um eine Auszeichnungssprache, die zur Beschreibung von logischen Strukturen und Hierarchien in Texten verwendet werden kann.² Struktur und Aufbau des Inhalts von Texten (z.B. Kapitel, Unterkapitel, Abbildungen, Metadaten) werden erfasst und können zur Weiterverarbeitung genutzt werden. XML ist in einem einfachen Format in Klartext strukturiert und bietet eine plattformneutrale Möglichkeit zur gemeinsamen Nutzung und Speicherung von Daten.³ Folglich kann ein XML-Dokument von Computern mit völlig unterschiedlichen Komponenten, Schnittstellen, Konfigurationen und Betriebssystemen gemeinsam genutzt werden.⁴

Ein sogenannter „XML-first-Workflow“ bringt für Verlage die meisten Vorteile mit sich.⁵ Unter „XML-first“ versteht man diejenigen Workflows, bei denen vor Beginn der Satzarbeiten XML eingesetzt wird. Bei XML-first müssen also die Manuskriptdaten noch vor Beginn der Satzarbeiten nach XML

1 Böhm, David; Grossmann, Alexander; Reiche, Michael u.a.: Open-Access-Publikationsworkflow für akademische Bücher, Leipzig 2020.

2 Haußner, Felix: XML-basierte Anreicherung von Texten. Potentiale für Verlage, Erlangen 2014.

3 Paper, David: Web Programming for Business, New York 2015, S. 151-186.

4 Ebd.

5 Rech, David Alan: Instituting an XML-first Workflow, in: Publishing research quarterly 28 (3), 2012, S. 192-196.

konvertiert werden.⁶ Dies ist unabhängig vom Quelldatenformat, in dem die Manuskriptdaten vorliegen, immer notwendig (es sei denn, Autor*innen würden ihr Manuskript direkt in XML schreiben). Nach Abschluss der Konvertierung sollte dann eine systematische Qualitätssicherung der XML-Daten stattfinden.⁷ Dies ist insofern verständlich, als die XML-Datei letztlich als „single source“, also als einzige Quelle, dienen wird: Aus ihr werden später alle digitalen Varianten der Publikation generiert. Die auf XML basierenden, medienneutralen Manuskripte werden letztlich „[...] durch Stylesheets in die speziellen Formate der verschiedenen Ausgabekanäle überführt [...]“.⁸

Es besteht die Möglichkeit, medienneutral ausgezeichnete Manuskripte direkt von Autor*innen einzufordern, was sich in der Praxis jedoch schwierig gestaltet.⁹ Auch können eingereichte Manuskripte vom Verlag selbst oder von einem Dienstleister ausgezeichnet werden. Ein Mangel an Kompetenzen sowie an personellen und zeitlichen Ressourcen seitens der Hochschulverlage erschwert aber oftmals eine medienneutrale Umsetzung von Publikationen.¹⁰ Daher kann eine automatische XML-Extraktion aus Manuskripten für Verlage sehr hilfreich sein.

Insgesamt kann ein Verlag durch einen XML-first-Workflow seine Publikationen in mehreren digitalen Formaten bereitstellen. Mehrere Formate haben wiederum den Vorteil, dass je nach Lesesituation (z.B. Lesen auf unterschiedlichen digitalen Geräten, Benutzung eines Screenreaders) das jeweils am besten passende Format gewählt werden kann. Es lässt sich daher sagen, dass die Verwendung von XML im Publikationswesen ein wichtiger Schritt ist, um die Herausforderungen der heutigen Verlagswelt zu meistern.¹¹

OS-APS verfolgt das Konzept eines XML-first-Workflows. Die Software importiert Word-Manuskripte und extrahiert automatisch die zugrundeliegenden XML-Strukturen. Nach der optionalen Optimierung über den Editor kann das Manuskript schließlich in verschiedene Formate (PDF, EPUB, HTML und XML, genauer JATS-XML) exportiert werden, wobei das Corporate Design des Verlags über Satzvorlagen, sogenannte Templates, automatisch angewendet wird. Aktuell ist die Programmierung soweit fortgeschritten, dass eine Demoversion¹² online zur Verfügung steht. Die Veröffentlichung der fertigen Software ist für Anfang 2023 geplant.

6 Ott, Tobias: Crossmediales Publizieren im Verlag, Berlin 2014.

7 Ebd.

8 Böhm, David; Grossmann, Alexander; Reiche, Michael u.a.: Open-Access-Publikationsworkflow für akademische Bücher, Leipzig 2020.

9 Ebd.

10 Ebd.

11 Rech, David Alan: Instituting an XML-first Workflow, in: Publishing research quarterly 28 (3), 2012, S. 192–196.

12 Demoversion der Open Source Academic Publishing Suite (OS-APS). Online: <<https://os-aps.de/demo/>>, Stand: 08.07.2022.

2. Kurzdarstellung des Projekts

2.1 Problem- und Zielstellung

OS-APS adressiert ein Problem, welches derzeit in vielen Publikationsworkflows auftritt: Ein Medienbruch zwischen Inhalt und Darstellung bzw. Output verhindert gute und einfache Workflows.¹³ Es gibt zum einen die Seite des Inhalts: Dazu zählen Prozesse wie das Schreiben, Überarbeiten und der Review-Prozess. Hier sind spezifische Formate wie Word (DOCX) und PDF in Verwendung. Zum anderen gibt es die Seite der Darstellung bzw. des Outputs, wo es beispielsweise auf Veröffentlichungsformate, Metadaten und die Verbreitung von Publikationen ankommt. Hier sind wiederum andere Formate von Bedeutung: Neben PDF-Dateien werden beispielsweise auch EPUB-, HTML- oder XML-Dateien benötigt. Während die Content-Seite auf Autor*innen, Lektorat und Reviewer*innen fokussiert ist, geht es auf Output-Seite hauptsächlich um die Leser*innen sowie die Produktion. Ziel von OS-APS ist es, diese Lücke durch einen medienneutralen, auf XML basierenden Workflow zu schließen.

2.2 OS-APS Kernpunkte

Bei OS-APS wird es sich um eine Software handeln, die mit möglichst wenig Aufwand professionelle Layouts in verschiedenen Formaten (PDF, EPUB, HTML, XML) erzeugen können soll. Das Programm wird Open Source verfügbar sein, wobei Nutzer*innen die Möglichkeit haben werden, sich zwischen einer „Docker On-Premise“ – oder einer Hosting-Lösung zu entscheiden. Ersteres bedeutet eine Installation, z.B. im Rechenzentrum einer Universität. Letzteres wird durch SciFlow als Dienstleistung angeboten. Neben SciFlow kommen auch noch andere potenzielle Dienstleister in Frage, da es sich um eine Open-Source-Software handelt, die jede*r nutzen und für die jede*r Dienste anbieten kann. Die beiden Lösungen unterscheiden sich dadurch, dass „Docker On-Premise“ kostenlos ist, aber selbst betreut werden muss, während beim Hosting die Betreuung an einen Dienstleister abgegeben wird, dafür aber auch Kosten anfallen. Der Code wird ebenfalls Open Source verfügbar sein.

Anbindungen an verbreitete Tools zum Publizieren sind geplant bzw. bereits in Arbeit. Open Journal Systems (OJS) und Open Monograph Press (OMP) werden als Publikationsplattformen, die beide Open Source verfügbar sind, mit eingebunden. Selbiges gilt für die Repositoriumssoftware DSpace. Die Auszeichnungssprache für Referenzen (CSL), die von Referenzmanagern wie Zotero verwendet wird, wird ebenfalls eingebunden. Pandoc als Konvertierungstool zwischen verschiedenen Formaten (u.a. MS Word zu HTML) wird mit eingebunden, da es in diesem Open-Source-Projekt Parallelen zu OS-APS gibt, weswegen OS-APS hier die Entwicklung mit unterstützen kann. Wichtig ist hierbei, dass das Ziel von OS-APS nicht darin besteht, einen „neuen Standard“ zu entwickeln – ganz im Gegenteil werden bereits bestehende Standards genutzt und eingebunden, um es Verlagen möglichst leicht zu machen, einen Workflow basierend auf OS-APS einzuführen.

¹³ Hyde, Adam: Single Source Publishing. A investigation of what Single Source Publishing is and how this 'holy grail' can be achieved. Online: <<https://coko.foundation/articles/single-source-publishing.html>>, Stand: 19.09.2022.

Das Projekt wird vom Bundesministerium für Bildung und Forschung (BMBF) gefördert¹⁴ und von drei Projektpartnern¹⁵ gemeinsam umgesetzt.

3. Vorstellung des angedachten OS-APS Workflows und der Komponenten

3.1 Der OS-APS Workflow

Ein vorliegendes Manuskript (derzeit DOCX) wird in den OS-APS Texteditor importiert. Dabei werden automatisch XML-Strukturen extrahiert. Dies bedeutet, dass beim Import des Word-Manuskriptes das DOCX-Format in ein anderes XML-Schema überführt wird. Die Überführung ist erforderlich, um semantische Auszeichnungen wie Metadaten, Strukturen und Referenzen zu ermöglichen – sei es, damit semantische Auszeichnungen, z.B. Überschriften, beim Import erkannt werden oder aber, damit sie später im Editor hinzugefügt werden können (z.B. im Falle von diversen Metadaten wie Autoreninformationen, die im Manuskript für gewöhnlich fehlen). Das XML-Format eignet sich für spätere Konvertierungen der Exportformate.

Im Texteditor kann das Manuskript anschließend bei Bedarf nachbearbeitet werden. Dies schließt auch das Erweitern der Metadaten mit ein. Schließlich wird ein Template ausgewählt und das Manuskript im gewünschten Format exportiert und gegebenenfalls auf einer Publikationsplattform (z.B. OJS, OMP, DSpace) hochgeladen. Die folgende schematische Darstellung des Workflows findet sich auch im entsprechenden BMBF-Fördermittelantrag.¹⁶

14 Förderungszeitraum: 01.02.2021–31.01.2023. Förderungsrichtlinie: Richtlinie zur Förderung von Projekten zur Beschleunigung der Transformation zu Open Access, Bundesanzeiger vom 17.06.2020. Online: <https://www.bmbf.de/bmbf/shareddocs/bekanntmachungen/de/2020/06/3044_bekanntmachung.html>, Stand: 08.07.2022.

15 Drei Partnerinstitutionen arbeiten zusammen an OS-APS: Das Technologieunternehmen SciFlow in Berlin (Entwicklung der Software), die Universitäts- und Landesbibliothek Sachsen-Anhalt in Halle (Schnittstellenanbindung) sowie die Universitätsbibliothek der Friedrich-Alexander-Universität Erlangen-Nürnberg (Anforderungserhebung und Tests).

16 Auszug aus dem von den Projektpartnern (SciFlow GmbH; Martin-Luther-Universität Halle-Wittenberg, Universitäts- und Landesbibliothek Sachsen-Anhalt; Friedrich-Alexander-Universität Erlangen-Nürnberg, Universitätsbibliothek) beim BMBF eingereichten Fördermittelantrag für das Projekt Open Source Academic Publishing Suite (OS-APS). Online: <https://os-aps.de/files/2020_08_07_BMBF_OS-APS_FAU_Antrag_kurz.pdf>, Stand: 08.07.2022.

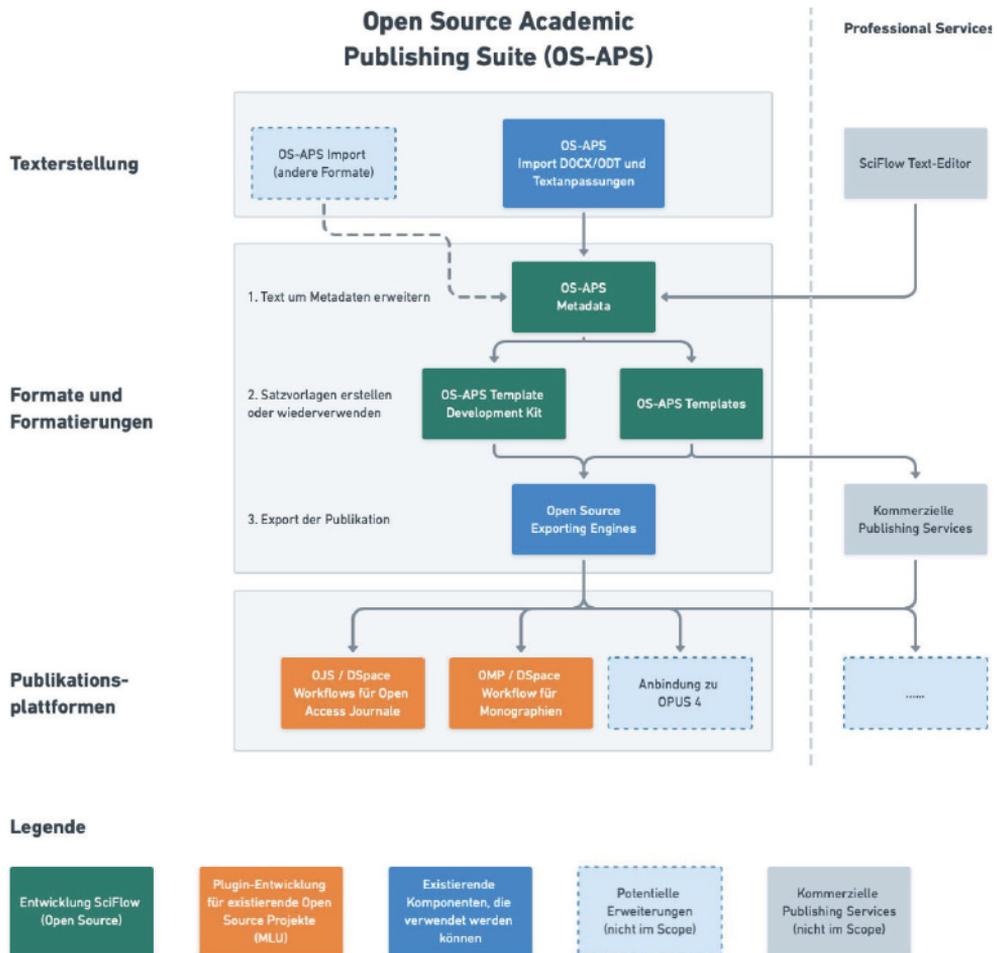


Abb. 1: Komponenten der Open Source Publishing Suite und deren Schnittstellen, entnommen aus dem BMBF-Fördermittelantrag

3.2 Vorstellung der Komponenten

OS-APS-Texteditor

Manuskripte können in den OS-APS-Texteditor importiert werden. Durch die Extrahierung von XML-Strukturen werden Elemente wie Kolumnentitel, Seitenumbrüche, Tabellen etc. erkannt. Im Editor ist die Änderung des Textes sowie der Formatierung möglich, falls Elemente nicht korrekt erkannt wurden oder schlicht von Autoren- und Verlagsseite noch Verbesserungen am Manuskript vorgenommen werden sollen. Auch können bei Bedarf die Metadaten ergänzt sowie semantische Verweise (z.B. Referenzen) hinzugefügt werden.

The screenshot shows the OS-APS text editor interface. The main window displays a document titled "Writing with OS-APS" with three sections: "Abstract", "Introduction", and "Methods". Each section contains placeholder text (Lorem ipsum). The right-hand sidebar, titled "Document outline", shows a hierarchical list of headings: "heading", "H1 Writing with OS-APS", "H1 Abstract", "H1 Introduction", "H1 Methods", "H2 A heading 2", "H2 Discussion", "H1 Ich frage mich, was das!", and "H1 References". Below the outline, there are two sections: "Add more chapters (h1)" and "Cross referencing", both with "read more" links.

Abb. 2: Ansicht im Texteditor von OS-APS

Templates

Beim Export wird das Corporate Design des jeweiligen Verlags über Templates abgebildet. Verschiedene Standardtemplates werden bereitgestellt. Weitere Templates sowie weitere Exporte können mit dem sogenannten Template Development Kit entwickelt werden. Dies ist besonders für Verlage interessant, die sehr klare Formatvorgaben haben und davon nicht abweichen wollen.

Template Development Kit

Mit Hilfe des Template Development Kits können einzelne Parameter eines Templates (z.B. eines der vorgefertigten Templates) direkt und ohne Vorkenntnisse geändert werden. Ebenso sind umfassende Veränderungen bzw. die Neuerstellung eines Templates möglich, wobei hierfür Kenntnisse in Webprogrammierung vonnöten ist. Komplett neue Exportformate können ebenfalls von Personen mit Programmierkenntnissen hinzugefügt werden. Das Template Development Kit basiert auf der Open-Source-Software Pandoc, dem bereits erwähnten Konvertierungstool zwischen verschiedenen Formaten sowie auf Eigenentwicklung seitens SciFlow.

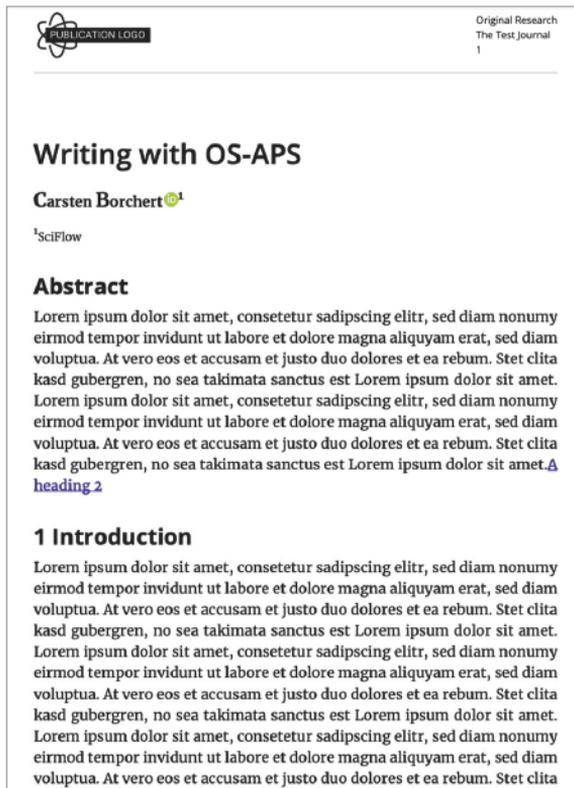


Abb. 3: Unter Verwendung eines Templates aus OS-APS exportiertes Testdokument



Abb. 4: Änderung einzelner Parameter auf der linken Seite; Erstellung eines neuen Templates auf der rechten Seite

Anbindung von OJS, OMP und DSpace

Im Rahmen des vorgesehenen Workflows erfolgt ein Export von OJS- und OMP-Daten nach DSpace mit anschließender Rückgabe von DOI-Informationen. Generell sind OJS und OMP im Rahmen des Projekts als Darstellungsplattformen für Leser*innen und DSpace zur Langzeitarchivierung angedacht. Für Details zur technischen Umsetzung sei an dieser Stelle auf einen 2022 erschienenen Artikel in der Zeitschrift ABI-Technik verwiesen.¹⁷

4. Anforderungserhebung durch eine Umfrage

4.1 Kontextualisierung des Fragebogens

Um die genauen Anforderungen zu erheben, wurde eine Umfrage durchgeführt. Bei den teilnehmenden Verlagen handelte es sich weitgehend um Hochschulverlage, wobei auch einige kommerzielle Verlage vertreten waren. Die Teilnehmer*innen waren meist in der Position der Verlagsleitung. Fast alle wissenschaftlichen Disziplinen waren repräsentiert, jedoch ließ sich eine deutliche Tendenz hin zum Bereich Geistes-, Kultur- und Sozialwissenschaften erkennen. Der Link zur Online-Umfrage wurde per E-Mail bzw. über Mailverteiler an die verschiedenen Zielgruppen¹⁸ gesendet.

Die Umfrage konnte für zwei verschiedene Publikationstypen ausgefüllt werden: Monographien oder Periodika (Journals). Diese Auswahl musste zu Beginn der Umfrage getroffen werden. Möglich und auch ausdrücklich gewünscht war es allerdings, die Umfrage mehrfach auszufüllen. So konnten Teilnehmer*innen, deren Institutionen beide Publikationstypen verlegen, auch für beide Typen den Workflow darstellen. Zudem konnten auf diese Weise auch unterschiedliche Workflows innerhalb desselben Publikationstyps (z.B. bei verschiedenen Journals) dargestellt werden. Die überwiegende Mehrheit (78,8%) der Teilnehmenden beantwortete die Umfrage für Monographien. Für Periodika (Journals) füllten 21,2% der Teilnehmer*innen den Fragebogen aus.

Größtenteils handelte es sich um Multiple-Choice-Fragen. Sofern logisch sinnvoll, war pro Frage die Auswahl mehrerer Antworten möglich. Auch konnten Fragen übersprungen werden, falls sie beispielsweise als unzutreffend auf die jeweilige Institution empfunden oder nicht verstanden wurden. In etwas geringerem Umfang enthielt die Umfrage auch offene Fragen, für deren Beantwortung ein Freitextfeld zur Verfügung stand.

Insgesamt wurde die Umfrage 52-mal eingereicht. Somit beträgt N (die Gesamtzahl der Grundgesamtheit) 52, während die Anzahl der Antworten pro Frage n (Teilmenge aus der Grundgesamtheit)

17 Putnings, Markus; Borchert, Carsten; Cozatl, Roberto: Ein Einblick in das BMBF-Projekt Open Source Academic Publishing Suite (OS-APS), in: ABI-Technik 42 (3), 2022, S. 166–173. Online: <<https://doi.org/10.1515/abitech-2022-0030>>

18 AG Universitätsverlage, AEUP, Enable!-Community, Peergroup Produktion der IG Digital, GeSIG, Library Publishing Coalition, Association of University Presses, OASPA, ACUP / APUC – Association of Canadian University Presses, Association of American University Presses, The Association of Japanese University Presses, Kooperationspartner wie OA-STRUKTKOMM, DEval, CeDiS der FU Berlin, Foren wie das Open Access Books Network sowie das deutsche PKP Community Forum.

je nach Frage variiert, da nicht jede*r Teilnehmende alle Fragen beantwortet hat – entweder weil Fragen übersprungen wurden, oder weil sie aus logischen Gründen nicht erschienen (z.B. keine weiterführenden Fragen zu Dienstleistern, wenn im Vorfeld angegeben wurde, dass Dienstleister nicht in den Publikationsprozess einbezogen werden). Bei 52 Einreichungen der Umfrage wurden folglich manche Fragen tatsächlich 52-mal beantwortet, andere aber z.B. nur 40-mal, was sich wiederum auf die Aussagekraft der Ergebnisse auswirkt und daher nicht unerwähnt bleiben soll.

4.2 Manuskripteingang

Da eine Priorisierung von Importformaten seitens der Entwickler*innen notwendig ist, wurde unter anderem abgefragt, welche Formate für die Einreichung von Manuskripten akzeptiert (n=52) werden sowie im Anschluss, welches Format seitens der Verlage bevorzugt wird (n=50).

Als Einreichungsformat für Manuskripte werden Microsoft Word bzw. DOCX (39,4%), PDF (23,9%), LibreOffice bzw. ODT (20,2%) und LaTeX (15,6%) akzeptiert. Zudem wurde auch Markdown, eine vereinfachte Auszeichnungssprache, die die Formatierung von Texten im Internet ermöglicht, unter der Antwortoption „Andere“ (0,9%) genannt.

Die deutliche Mehrheit (68%) der Verlage bevorzugt das Format DOCX. Mit 24% werden PDF-Dateien als Einreichungsformate deutlich weniger stark präferiert. Des Weiteren gaben 6% an, kein Format zu präferieren, und 2% bevorzugten LaTeX. ODT stellt für 0% der Teilnehmer*innen ein bevorzugtes Einreichungsformat dar. Demensprechend wird OS-APS beim Import zunächst DOCX unterstützen. Eine Unterstützung von ODT wäre insofern sinnvoll, als es sich bei LibreOffice um ein Open Source verfügbares Schreibprogramm handelt (im Gegensatz zu Microsoft Word). Priorität hat mit Blick auf die Umfrageergebnisse jedoch DOCX aufgrund seiner Relevanz in der Praxis.

Zur Formatierung der Manuskripte gibt es in der Regel Formatierungsrichtlinien für die Autor*innen (n=50). Bei den meisten der befragten Verlage gibt es textuell beschriebene Formatierungsvorgaben (68%). Hingegen gaben 22% an, dass keine derartigen Vorgaben bei ihnen existieren. International standardisierte Styles als Formatvorgaben, beispielsweise von der Modern Language Association (MLA), spielen mit 8% bei den Verlagen eine untergeordnete Rolle. Unter „andere“ (2%) wurde im entsprechenden Freitextfeld die Angabe gemacht, dass sowohl textuell beschriebene Richtlinien als auch Word- und LaTeX-Templates gibt. Eine Folgefrage speziell zu Templates wurde in der Online-Umfrage gleich im Anschluss an die Frage nach Formatierungsvorgaben gestellt; es war jedoch immer nur eine Frage auf einmal sichtbar.

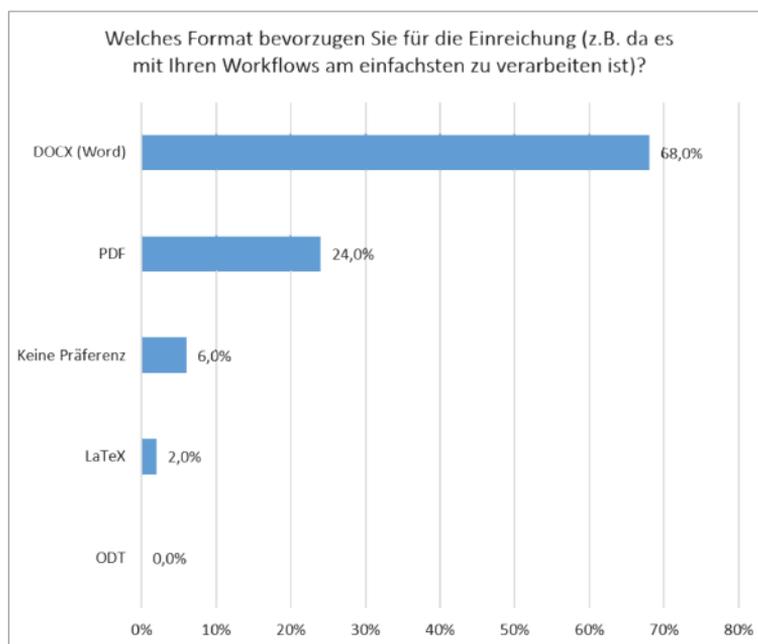


Abb. 5: Ergebnisse zur Frage „Welches Format bevorzugen Sie für die Einreichung [...]?“ (n=50)

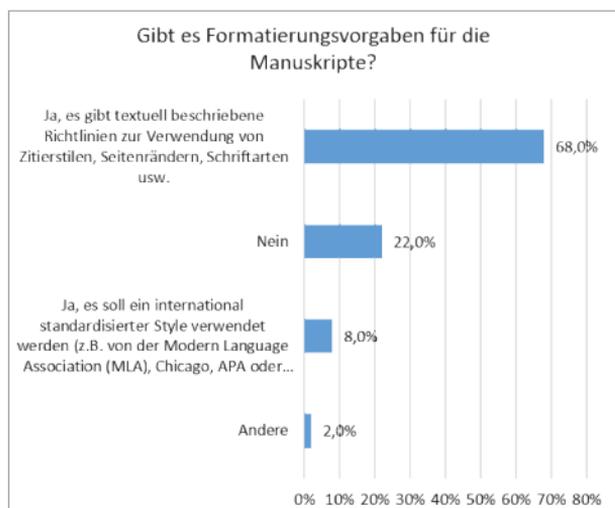


Abb. 6: Ergebnisse zur Frage „Gibt es Formatierungsvorgaben für die Manuskripte?“ (n=50)

Die Mehrheit der befragten Verlage stellt jedoch keine konkreten Formatvorlagen (sogenannte Templates), die die Autor*innen selbst nutzen können, bereit. Auf die Frage hin gaben 56,9% der Teilnehmer*innen an, dass ihr Verlag kein eigenes Template bereitstellt. Für 23,5% der Verlage ist die Verwendung einer verlagseigenen Formatvorlage für die Autor*innen optional oder wird empfohlen; bei 17,6% ist die Verwendung Pflicht. Unter „andere“ (2%) wurde zudem eine vorformatierte Word-Datei als Vorlage genannt und der Punkt weiter ausgeführt, allerdings ohne Angabe, ob die Verwendung für die Autor*innen verpflichtend oder optional ist. In jedem Fall handelt es sich hier jedoch um die Verwendung eines Word-Templates.

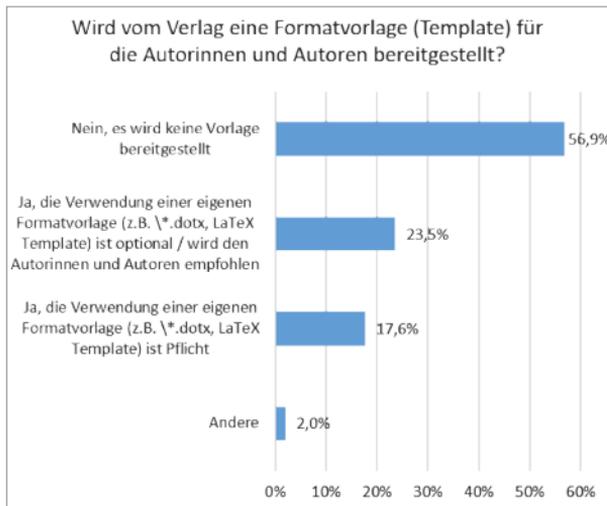


Abb. 7: Ergebnisse zur Frage „Wird vom Verlag eine Formatvorlage (Template) für die Autorinnen und Autoren bereitgestellt?“ (n=51)

Einzureichende Bilder werden in den Formaten JPEG, TIFF und PNG akzeptiert (n=52). Sie sollten in der Regel eine Auflösung von 300 dpi haben (n=48).

Im Rahmen der Manuskripteinreichung werden die (Meta-)Daten meist per E-Mail gesendet (n=52) und anschließend in einem Formular oder in einem separaten Content-Management-System gehalten (n=49). An Autorendaten werden besonders häufig die E-Mail-Adresse und die Instituts- bzw. Firmenzugehörigkeit erhoben (n=51), während bei Manuskriptdaten insbesondere die Schlagwörter sowie die Gesamtseitenzahl manuell erhoben werden (n=51).

Auch diverse Sonderzeichen, etwa für Formeln, sowie nicht-lateinische Schriften kommen insbesondere in wissenschaftlichen Publikationen abhängig vom Fachbereich häufig vor. Daher wurde abgefragt, ob die teilnehmenden Verlage auch Inhalte in nicht-lateinischer Schrift publizieren (n=49), wobei verschiedene nicht-lateinische Schriften auswählbar waren. Während 20,4% angaben, ausschließlich in lateinischer Schrift zu publizieren, benötigen 19,4% der befragten Verlage zumindest Einzelbuchstaben aus anderen Schriften, z.B. griechische Buchstaben für Formeln. Griechische

Schrift wurde von 17,2% ausgewählt, während jeweils 12,9% der Befragten auch in ostasiatischen Schriften sowie arabischer Schrift publizieren. Etwas seltener wird kyrillische Schrift (11,8%) von den Verlagen benötigt.

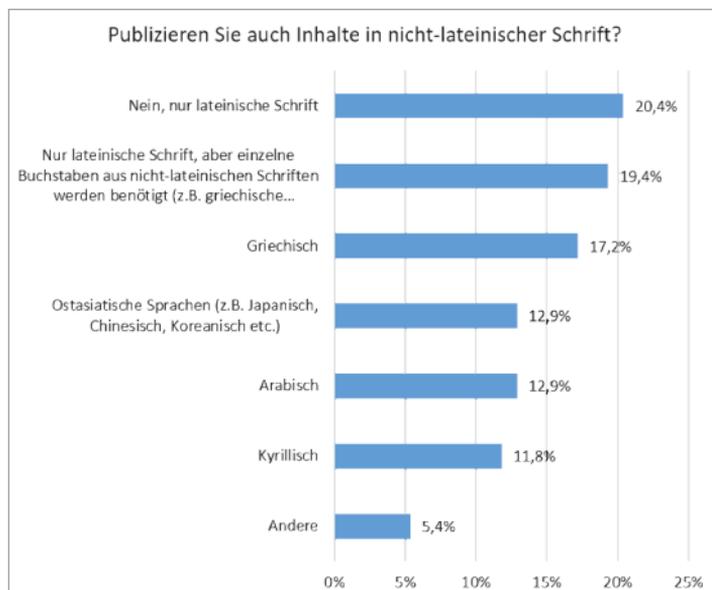


Abb. 8: Ergebnisse zur Frage „Publizieren Sie auch Inhalte in nicht-lateinischer Schrift?“ (n=49)

Unter der Option „andere“ (5,4%) erschien ein Freitextfeld, in das weitere Schriften eingetragen werden konnten. Hier wurden diverse andere Schriften, z.B. Hebräisch, Sanskrit, Keilschrift und Gotisch, genannt. Diese werden laut Angaben der Befragten häufig in Verbindung mit lateinischem Text publiziert. Zudem wird, je spezieller der Zeichensatz, teilweise mehr Eigenleistung seitens der Autor*innen erwartet.

4.3 Manuskriptbearbeitung

Die Manuskripte werden weitgehend von den Verlagen druckreif gemacht (n=52). Auf die Frage hin, ob im Workflow des jeweiligen Verlags bereits Maßnahmen zur Verbesserung der Barrierefreiheit umgesetzt sind (n=51), antworteten 54,9%, dass Maßnahmen aktuell zwar nicht umgesetzt, jedoch für die Zukunft geplant sind. Des Weiteren gaben 29,4% der Befragten an, dass aktuell zwar bereits Maßnahmen umgesetzt sind, diese jedoch in Zukunft noch erweitert werden müssen. Somit besteht bei 84,3% der befragten Verlage Handlungsbedarf in Sachen Barrierefreiheit. Jeweils 7,8% gaben an, dass entweder aktuell umgesetzte Maßnahmen zur Barrierefreiheit ausreichend sind oder dass es keine Maßnahmen dahingehend gibt und derzeit keine Notwendigkeit dazu besteht.

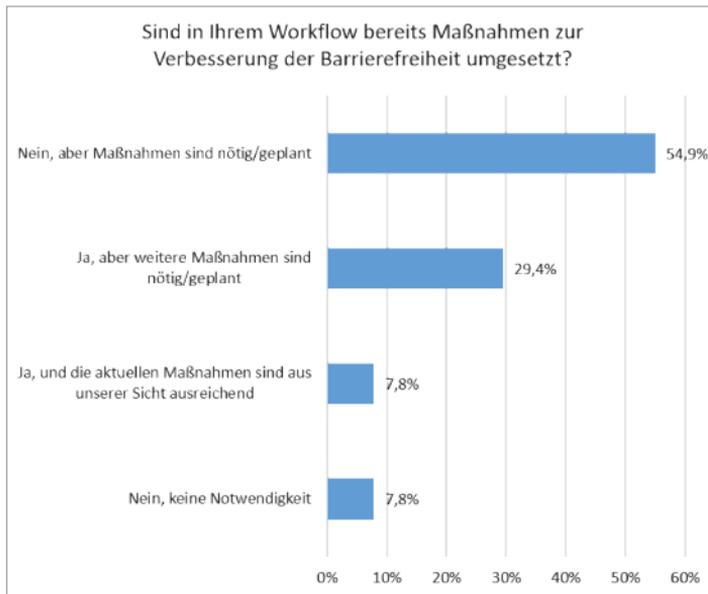


Abb. 9: Ergebnisse zur Frage „Sind in Ihrem Workflow bereits Maßnahmen zur Barrierefreiheit umgesetzt?“ (n=51)

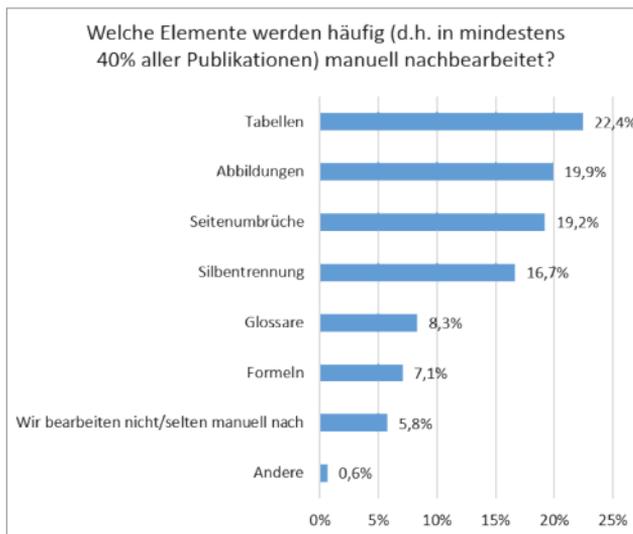


Abb. 10: Ergebnisse zur Frage „Welche Elemente werden häufig [...] manuell nachbearbeitet?“ (n=50)

Sofern bereits Arbeitsschritte zur Verbesserung der Barrierefreiheit stattfinden, sollten diese beschrieben werden. Besonders die Hinterlegung von Alternativtexten sowie die Nachbearbeitung von Tabellen wurden oft erwähnt (n=12).

Häufig in Publikationen verwendete Elemente sind Fuß- und Endnoten, Zitationen und Literaturverzeichnisse, einfache Tabellen sowie absatztrennende Abbildungen¹⁹ (n=51). Damit die verschiedenen Elemente gut lesbar und korrekt formatiert angezeigt werden können, ist häufig eine manuelle Nachbearbeitung nötig: Besonders oft müssen Tabellen (22,4%), Abbildungen (19,9%), Seitenumbrüche (19,2%) und die Silbentrennung (16,7%) manuell nachbearbeitet werden (n=50).

Erheblich weniger häufig werden Glossare (8,3%), Formeln (7,1%) sowie andere Elemente (0,6%) manuell bearbeitet. Nur 5,8 Prozent gaben an, nicht oder selten manuell nachzubearbeiten.

4.4 Publikationsformate

Aktuell werden von den befragten Verlagen vor allem verschiedene Arten von PDF-Dateien als Outputformate erzeugt (n=51). Dies sind PDF/X für den Druck (23,7%), PDF/A für E-Book (21,8%) sowie PDF allgemein (21,8%). Auf diese drei Formate folgen in absteigender Reihenfolge die Formate EPUB (12,2%), HTML (9,6%) und MOBI (6,4%). Nur sehr selten werden mit 2,6% JATS(-XML) sowie mit 1,9% BITS als Outputformate generiert. Keine Rolle spielen AZW/AZW3/KF8, MECA und andere Formate mit jeweils 0%.

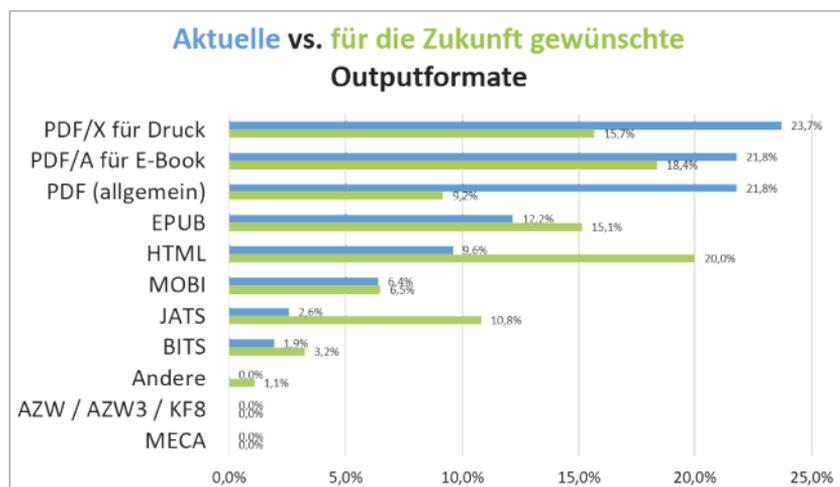


Abb. 11: Ergebnisse zu den Fragen „Welche Publikationsformate erzeugen Sie aktuell?“ (n=51, blau) und „Welche Publikationsformate möchten Sie in Zukunft erzeugen?“ (n=52, hellgrün) gegenübergestellt

Auch für die Zukunft gewünschte Output-Formate wurden abgefragt (n=52). Zukünftig möchten die teilnehmenden Verlage wesentlich stärker HTML als Format für die Veröffentlichung nutzen (20%).

¹⁹ Absatztrennende Abbildungen sind Abbildungen, bei denen der Text immer oberhalb und unterhalb der Abbildung platziert ist und sich automatisch verschiebt, sofern die Abbildung verschoben wird (im Gegensatz zu anderen Darstellungsarten, z.B. „textumfließend“, „hinter dem Text“ etc.).

Dies stellt eine deutliche Steigerung im Vergleich zu den aktuell bei 9,6% der Teilnehmenden erzeugten HTML-Dateien als Output-Formate dar.

Verschiedene PDF-Dateien bleiben relevant, jedoch ist für alle drei Antwortoptionen ein Rückgang zu verzeichnen. Auffällig ist auch, dass das PDF/A für E-Books mit 18,4% zukünftig häufiger genutzt werden soll als das PDF/X für Druck mit 15,7%. PDF allgemein soll in Zukunft deutlich weniger (9,2%) genutzt werden.

Die Relevanz des Formats EPUB steigt leicht von aktuell erzeugten 12,2% zu für die Zukunft gewünschten 15,1%. Eine deutlich dramatischere Steigerung, ähnlich der des Formats HTML, erfährt JATS(-XML) von 2,6% auf 10,8%. Leichte Steigerungen sind auch bei MOBI (6,4% auf 6,5%) und BITS (1,9% auf 3,2%) ersichtlich.

Zur Erzeugung und Bearbeitung von Publikationsformaten werden in den meisten Fällen InDesign (31,5%), Adobe Acrobat Pro (26,8%) oder Microsoft Word (20,5%) genutzt (n=49). Die meisten Verlage verwenden keine semantischen Auszeichnungen (45,2%; n=49), also Textelemente, die dem Inhalt des Textes (z.B. Schmidt, 2013) eine besondere Bedeutung geben (z.B. die eindeutige Referenz). Sofern doch semantische Auszeichnungen verwendet werden, handelt es sich mehrheitlich um Referenzen (33,9%) und nur selten um Glossareinträge o.ä.. Von den Verlagen, die keine semantischen Auszeichnungen verwenden, plant etwas mehr als die Hälfte semantische Auszeichnungen für die Zukunft (51,9%; n=27). Das am häufigsten standardmäßig unterstützte Layout ist A4 (n=48).

Mehr als drei Viertel der befragten Verlage publiziert hybrid (n=51). Unterschiede in den Workflows für Print- und Onlineveröffentlichungen gibt es meist nicht (n=45). Sofern jedoch Unterschiede in den Workflows für Print- und Onlineveröffentlichungen bestehen, konnten diese in einem Freitextfeld beschrieben werden (n=14).

Falls ein Dienstleister in den Publikationsprozess eingebunden ist, benötigt dieser meist ein Word-Dokument, unabhängig von dessen Qualität (z.B. bezogen auf die semantische Auszeichnung) (n=14). Daraufhin verarbeitet der Dienstleister das eingereichte Dokument in nahezu der Hälfte der Fälle über einen auf InDesign basierenden Workflow (n=14).

4.5 Zukunftspotenziale

Besonders viel Potenzial zur Optimierung des Publikationsworkflows sehen die Teilnehmer*innen in der weiteren Automatisierung des Publikationsworkflows und der darin begründeten Zeitersparnis (n=41). Auch weitere Standardisierung sowie Workflows mit XML, ggf. auch XML-first wurden auf die offene Frage hin häufig als Potenziale genannt. Die Erweiterung des Verlagsgeschäfts betreffend liegen Zukunftspotenziale laut den Umfrageteilnehmer*innen besonders in der Erschließung neuer Formate und damit einhergehend neuer Verbreitungswege (33%; n=47). Bei der Frage nach IT-Lösungen, die bei der Umsetzung der beschriebenen Potenziale helfen könnten, setzte die Mehrheit der Teilnehmer*innen auf neue Software-Konzepte (n=33).

5. Schlussfolgerungen für den Workflow

5.1 Verwendung des DOCX als Importformat

Da das Format DOCX (Microsoft Word) von den Teilnehmenden sowohl am häufigsten als Einreichungsformat akzeptiert als auch als solches bevorzugt wurde, wird es für OS-APS in jedem Fall einen DOCX-Import geben. Ebenfalls geplant ist der Import von ODT-Dateien (LibreOffice), da es sich bei LibreOffice um eine Microsoft Word ähnliche Open-Source-Software handelt, deren Format ODT als Einreichungsformat nach DOCX und PDF am dritthäufigsten akzeptiert wird. Bevorzugt wird ODT jedoch von keinem der befragten Verlage.

5.2 Formatierungsvorgaben und Formatvorlagen (Templates) seitens der Verlage

Die befragten Verlage gaben mehrheitlich an, zwar textuelle Formatierungsrichtlinien für ihre Manuskripte zu haben, jedoch in der Regel kein entsprechendes Template für die Autor*innen, d.h. in diesem Fall eine Formatvorlage z.B. in Word oder LaTeX, bereitzustellen. Entsprechend dieser Ergebnisse ist es für das Projekt OS-APS wichtig, dass es für die Verlage nicht erforderlich ist, Formatvorlagen für den Import zu entwickeln, um OS-APS nutzen zu können.

5.3 Zulassung von nicht-lateinischen Schriften

Die deutliche Nachfrage nach diversen nicht-lateinischen Schriften zeigt an, dass dieses Feature von Anfang an unterstützt werden sollte. Da Aufgrund der Vielzahl an möglichen nicht-lateinischen Schriften eine Priorisierung seitens der Entwickler*innen nötig werden wird, bieten die Ergebnisse zudem erste Anhaltspunkte, welche Schriften am meisten benötigt werden und daher am ehesten unterstützt werden sollten.

5.4 Priorisierung der Darstellungsoptimierung der Elemente, die häufig manuell nachbearbeitet werden

Im Rahmen der Online-Umfrage wurden auch diejenigen Elemente in Publikationen ermittelt, die besonders häufig manuell nachbearbeitet werden müssen. Hier muss es das Ziel sein, den dadurch entstehenden, oftmals enormen Mehraufwand für die Verlage zu verringern. Damit liegt die höchste Priorität in der Darstellungsoptimierung für OS-APS den Umfrageergebnissen zufolge bei Tabellen, Abbildungen, Seitenumbrüchen und Silbentrennung.

5.5 Aktuell erzeugte und für die Zukunft gewünschte Output-Formate

Insgesamt ist festzustellen, dass die Verlage vor allem die Output-Formate HTML und JATS-XML für die Zukunft zu nutzen planen. Daher werden diese beiden Formate von OS-APS unterstützt werden. Ebenfalls angeboten wird das Format PDF, da es aktuell sehr häufig von Verlagen generiert wird und auch in der nahen Zukunft seine Relevanz trotz eines leichten Rückgangs behält. Zuletzt wird

als viertes mögliches Exportformat EPUB angeboten. EPUB ist aktuell das zweithäufigste Output-Format nach diversen PDF-Formaten und wird den Umfrageergebnissen zufolge in Zukunft leicht an Relevanz dazugewinnen.

5.6 Besondere Berücksichtigungen zu Anforderungen der Barrierefreiheit

Für die deutliche Mehrheit (84,3%) der befragten Verlage besteht aus ihrer eigenen Sicht Handlungsbedarf bei der Verbesserung der Barrierefreiheit ihrer Publikationen. Diese Ergebnisse bedeuten für OS-APS, dass Maßnahmen zur Verbesserung der Barrierefreiheit im Publikationswesen ein zentrales Thema darstellen. Hier kann über ein Folgeprojekt zur Entwicklung barrierefreier Exporte nachgedacht werden. Generell werden die Dokumente durch die Nutzung der Software OS-APS formaler, d.h. der Inhalt unterliegt einer festen Struktur und einzelne Elemente, beispielsweise Überschriften oder Referenzen, sind als solche ausgezeichnet. Dadurch können Dokumente in andere Formate umgewandelt werden. Wird zum Beispiel HTML erzeugt, ist das Dokument sowohl am Desktop als auch am Mobilgerät lesbar. Sollte JATS-XML erzeugt werden, kann es durch Screenreader, z.B. eLife Lens²⁰, interpretiert und angezeigt werden.

Dementsprechend kann die Lösung für bessere Barrierefreiheit auch im Umstieg auf alternative Formate liegen. HTML beispielsweise ist für die Barrierefreiheit insofern besser geeignet, als der Text in einem strukturierten Format vorliegt und sich an verschiedene Lesesituationen besser anpassen kann (Darstellung hell/dunkel, Erhöhung der Schriftgröße, Darstellung mobil/Desktop). Da das PDF jedoch hohen Ansprüchen an Ästhetik und Lesekomfort seitens der Leser*innen besser gerecht werden kann als andere Formate, wäre es auch vorstellbar, einen Standard für Barrierefreiheit²¹ auszuwählen und diesem Standard entsprechende PDF-Exporte zu entwickeln.

6. Einbeziehung der Community und Ausblick

OS-APS wird unter Einbeziehung der Community weiterentwickelt. Hier sind insbesondere die beiden Beiräte des Projekts zu nennen, die sich aus kleineren Verlagen einschließlich Hochschulverlagen, wissenschaftlichen Bibliotheken, drittmittelgeförderten Projekten und weiteren Akteur*innen im Bereich digitales Publizieren zusammensetzen. Der Wissenschaftliche Beirat hat eine beratende Funktion bei der strategischen Planung und Ausgestaltung des Projekts inne, während der Anwendungsbeirat zu Projektfragen und -anliegen sowie für praktische Tests kontaktiert werden kann. Jeden ersten Mittwoch im Monat findet zudem ein öffentlicher „Demo-Day“²² statt, bei dem Feedback von den Teilnehmenden eingeholt wird.

20 eLife Lens, Online: <<https://lens.elifesciences.org/>>, Stand: 08.09.2022.

21 Beispielsweise PAC-Konformität basierend auf dem PDF Accessibility Checker (PAC 3), Online: <<https://www.access-for-all.ch/ch/pdf-accessibility-checker-pac.html>>, Stand: 07.07.2022.

22 Einwahldaten zu den „Demo-Days“ der Open Source Academic Publishing Suite (OS-APS). Online: <<https://os-aps.de/demo/>>, Stand: 08.07.2022.

Auch befindet sich OS-APS im Austausch mit anderen, ähnlich ausgerichteten Projekten.²³ Der Hauptunterschied der Software OS-APS im Vergleich zu anderen Projekten ist, dass sie als schlankes Tool für den Browser entwickelt wurde, was ohne hohe initiale Aufwände genutzt werden kann. Bisher waren große Projekte und Individualsoftwarelösungen erforderlich. Auch die Einbeziehung bereits in Verlagen etablierter Standards, z.B. der Anschluss an OJS und OMP, trägt hierzu bei.

Die Entwicklung fokussiert sich aktuell (Stand August 2022) auf das Erstellen von Monographie-Templates für den Projektpartner UB Erlangen-Nürnberg. Der an die Bibliothek angegliederte Hochschulverlag FAU University Press verfügt über umfangreiche Formatvorgaben für verschiedene Schriftenreihen, die im Rahmen der Entwicklung von OS-APS noch innerhalb der Förderphase als praktisches Beispiel umgesetzt werden sollen.

Langfristig soll OS-APS besonders Hochschulverlagen sowie weiteren non-profit Verlagen (z.B. publizierenden Fachgesellschaften) eine Professionalisierung im technischen Bereich erlauben, ohne dass dafür notwendigerweise technisches Wissen aufgebaut werden muss.

Literaturverzeichnis

- Böhm, David; Grossmann, Alexander; Reiche, Michael u.a.: Open-Access-Publikationsworkflow für akademische Bücher, Leipzig 2020.
- Haußner, Felix: XML-basierte Anreicherung von Texten. Potentiale für Verlage, Erlangen 2014.
- Hyde, Adam: Single Source Publishing. A investigation of what Single Source Publishing is and how this 'holy grail' can be achieved. Online: <<https://coko.foundation/articles/single-source-publishing.html>>, Stand: 19.09.2022.
- Ott, Tobias: Crossmediales Publizieren im Verlag, Berlin 2014.
- Paper, David: Web Programming for Business, New York 2015, S. 151–186.
- Putnings, Markus; Borchert, Carsten; Cozatl, Roberto: Ein Einblick in das BMBF-Projekt Open Source Academic Publishing Suite (OS-APS), in: ABI-Technik 42 (3), 2022, S. 166–173. Online: <<https://doi.org/10.1515/abitech-2022-0030>>
- Rech, David Alan: Instituting an XML-first Workflow, in: Publishing research quarterly 28 (3), 2012, S. 192–196.

23 Beispielsweise OA-STRUKTKOMM (<<https://open-access.network/vernetzen/open-access-projekte/oa-strukt-komm>>, Stand: 27.09.2022) oder TU9 Monos (<<https://open-access.network/vernetzen/open-access-projekte/tu9-monos>>, Stand: 27.09.2022).