

Urheberrechtlich geschützte Texte nachnutzen – Der XSample-Workflow

Andresen, Melanie

melanie.andresen@ims.uni-stuttgart.de
Universität Stuttgart, Deutschland

Gaertner, Markus

markus.gaertner@ims.uni-stuttgart.de
Universität Stuttgart, Deutschland

Jacke, Janina

janina.jacke@uni-goettingen.de
Universität Stuttgart, Deutschland

Ketschik, Nora

nora.ketschik@ilw.uni-stuttgart.de
Universität Stuttgart, Deutschland

Pichler, Axel

axel.pichler@ts.uni-stuttgart.de
Universität Stuttgart, Deutschland

Eines der Hindernisse, die der freien Weitergabe von Forschungsdaten im Sinne der Open-Science-Bewegung im Wege stehen, ist das Urheberrecht (UrhG). Dieses erschwert die Einhaltung der guten wissenschaftlichen Praxis und der FAIR-Prinzipien (Wilkinson et al. 2016) insbesondere bei Forschung zu zeitgenössischen Texten. Für urheberrechtlich geschützte Texte besteht bisher nur die Möglichkeit, sog. abgeleitete Textformate (Schöch et al. 2020) für „non-consumptive access“ (Organisciak und Downie 2021) zu veröffentlichen, etwa in Form von Frequenzlisten. Geisteswissenschaftliche Forschungsprojekte sind allerdings häufig auf die Verfügbarkeit von textuellem Kontext angewiesen, der erst eine angemessene Interpretation der Daten erlaubt. Um die Nutzbarkeit von Textdaten in dieser Hinsicht zu unterstützen, wurde im Projekt XSample¹ ein Workflow entwickelt, der auf das Recht zur Weitergabe von Textauszügen aufbaut, und dabei ermöglicht, die Auszüge auf das eigene Forschungsanliegen hin zu optimieren.

Rechtslage

Im deutschen UrhG unterliegen Texte bis 70 Jahre nach dem Tod der Autor*innen dem urheberrechtlichen Schutz, der die Vervielfältigung und die öffentliche Zugänglichmachung von Texten erheblich einschränkt. Hiervon sind vor allem zeitgenössische Texte betroffen, die aus diesem Grund womöglich gar nicht erst als Ge-

genstand von Forschungsprojekten in Erwägung gezogen werden. Seit 2018 ist zumindest die Verwendung von urheberrechtlich geschützten Texten zu Zwecken des Text- und Dataminings durch § 60d UrhG legitimiert (vgl. Raue 2021). Die Nachnutzung der Daten nach Abschluss eines Projektes ist aber weiterhin nur unklar geregelt (vgl. Kleinkopf et al. 2021, Andresen et al. 2022).

Der XSample-Workflow kombiniert den §60d UrhG mit § 60c Abs. 1 Nr. 1 UrhG, der es erlaubt, bis zu 15 Prozent von Werken und auch vollständige Werke geringen Umfangs zu Zwecken der nicht-kommerziellen wissenschaftlichen Forschung zu vervielfältigen und an bestimmt abgegrenzte Personenkreise für deren eigene wissenschaftliche Forschung weiterzugeben.

Weitergabe von Auszügen im XSample-Workflow

Die Weitergabe von nur 15 Prozent eines Textes erscheint auf den ersten Blick nicht hinreichend. Um die Nützlichkeit dieser Textauszüge zu maximieren, wird im XSample-Workflow über eine Benutzeroberfläche eine gezielte Textauswahl unterstützt. So können Forschende die Textauswahl genau auf die eigenen Forschungsanliegen zuschneiden. Dafür werden auch im Korpus enthaltene Annotationen berücksichtigt, sodass bei Interesse an einem bestimmten Phänomen systematisch die mit den relevanten Kategorien annotierten Textstellen extrahiert werden können (vgl. Gärtner 2020). Die Möglichkeiten und Grenzen des Auszugskonzepts sind am Beispiel zweier Anwendungsfälle aus der Literaturwissenschaft und Linguistik erprobt worden (vgl. Andresen et al. 2022). Sie zeigen beispielsweise, dass viele Forschungsfragen davon profitieren, wenn abgeleitete Textformate, die quantitative Analysen auf dem Gesamtkorpus ermöglichen, und Auszüge, die die qualitative Interpretation erlauben, kombiniert werden.

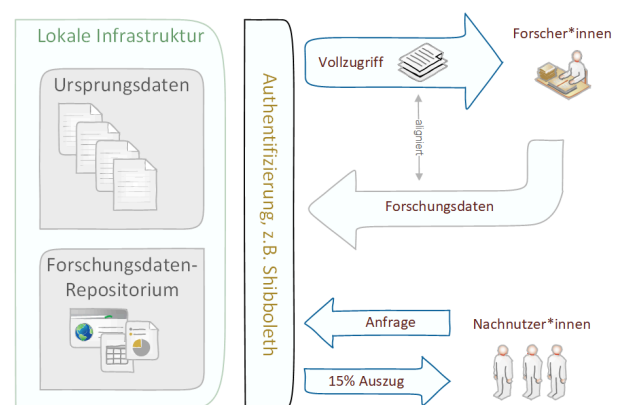


Abb. 1: Der XSample-Workflow im Überblick.

Abbildung 1 fasst das Auszugskonzept zusammen: Forscher*innen übermitteln ihre (annotierten) Forschungsdaten an Forschungsinfrastruktureinrichtungen (z. B. wissenschaftliche Bibliotheken), die diese verwalten. Nachnutzer*innen können Zugriffsanfragen an

die Infrastrukturbetreiber stellen, um Auszüge aus den Forschungsdaten zu erhalten.

Der in Abbildung 1 beschriebene Workflow wurde im Rahmen des XSample-Projekts prototypisch in Form eines Webservices implementiert.² Die urheberrechtlich geschützten Korpusdaten liegen hierbei sicher in einem Dataverse³-Repositorium. Öffentlich auffindbare Metadaten ermöglichen Nutzer*innen den Einstieg in den XSample-Workflow und leiten auf einen gesonderten Server weiter, wo die eigentliche Auszugserstellung stattfindet. Hierfür stehen verschiedene Verfahren zur Auswahl, die dabei helfen können, die Nützlichkeit der Auszugsdaten für die eigenen Bedarfe zu erhöhen. Für die anfragebasierte Auszugserzeugung setzt unsere Implementierung auf ein Java-Framework für Korpusanfragen (Gärtner 2020), wodurch gezielt nach bestimmten annotierten Phänomenen gesucht werden kann, die dann als Grundlage für die individuelle Auswahl der Auszüge dienen. Während des gesamten Prozesses haben Nutzer*innen keinen direkten Zugriff auf die geschützten Daten und erhalten vor dem Download ihres individuellen Auszugs lediglich grafische Veranschaulichungen der Auszugskomposition basierend auf dem gewählten Verfahren und/oder den Suchergebnissen (vgl. Andresen et al. 2022).

Fazit

„Open Access“ ist und bleibt die Zielvorstellung für offene und reproduzierbare Forschung auch in den digitalen Geisteswissenschaften. Das hier vorgestellte Auszugskonzept stellt demgegenüber eine „Behelfslösung“ dar, um die Nachnutzung urheberrechtlich geschützter Daten zu ermöglichen und der Tendenz entgegenzuwirken, diese Texte per se aus Forschungsprojekten auszuschließen (vgl. Gärtner et al. 2021). Die Lösung ist an der Forschungspraxis der digitalen Literatur- bzw. Geisteswissenschaften ausgerichtet, für die andere „verfremdende“ Verfahren wie abgeleitete Textformate (Schöch et al. 2020, Organisciak und Downie 2021) nur eingeschränkt nachnutzbar sind. Das Auszugskonzept ermöglicht hingegen eine größere Nähe zum Text, indem die ursprüngliche Textgestalt beibehalten wird, die für die Interpretation der Daten häufig unabdingbar ist. Darüber hinaus wird der rechtliche Rahmen durch die individuelle Auszugsgenerierung optimal ausgeschöpft und den individuellen Forschungsinteressen angepasst.

Fußnoten

1. <https://www.izus.uni-stuttgart.de/fokus/fdm-projekte/xsample/> (Stand: 14.12.2022).
2. Langfristig soll dieser Dienst an der Universitätsbibliothek Stuttgart angeboten werden. Die Software für den Prototypen ist unter <https://github.com/ICARUS-tooling/xsample-server> open-source verfügbar und kann somit auch von anderen Einrichtungen genutzt werden, um eine eigene Instanz des XSample-Servers zu betreiben.
3. Dataverse ist eine weit verbreitete Open-Source-Repositoriumssoftware, die über eine für XSample notwen-

dige feingranulare Rechteverwaltung verfügt und eine simple Schnittstelle zur Integration externer Services in die Benutzeroberfläche bietet.

Bibliographie

Andresen, Melanie, Markus Gärtner, Sybille Hermann, Janina Jacke, Nora Ketschik, Felicitas Kleinkopf, Jonas Kuhn und Axel Pichler. 2022. „Vorzüge von Auszügen – Urheberrechtlich geschützte Texte in den digitalen Geisteswissenschaften (nach-)nutzen.“ *Zeitschrift für digitale Geisteswissenschaften*. https://doi.org/10.17175/2022_007.

Gärtner, Markus. 2020. „The Corpus Query Middleware of Tomorrow – A Proposal for a Hybrid Corpus Query Architecture.“ In *Proceedings of the 8th Workshop on Challenges in the Management of Large Corpora*, 31–39. DOI: <https://www.aclweb.org/anthology/2020.cmlc-1.5>.

Gärtner, Markus, Felicitas Kleinkopf, Melanie Andresen, Sybille Hermann. 2021. „Corpus Reusability and Copyright – Challenges and Opportunities.“ In *Proceedings of the Workshop on Challenges in the Management of Large Corpora*, 10–19. DOI: <https://doi.org/10.14618/ids-pub-10467>.

Kleinkopf, Felicitas, Janina Jacke und Markus Gärtner. 2021. „Text- und Data-Mining – Urheberrechtliche Grenzen der Nachnutzung wissenschaftlicher Korpora bei computergestützten Verfahren und digitalen Ressourcen.“ *MMR Zeitschrift für IT-Recht und Recht der Digitalisierung* 24, H. 3: 196–200. DOI: <http://dx.doi.org/10.18419/opus-11445>.

Organisciak, Peter und J. Stephen Downie. 2021. „Research access to in-copyright texts in the humanities.“ In *Information and Knowledge Organisation in Digital Humanities*, hg. von Koraljka Golub und Ying-Hsang Liu, 157–177. Digital Research in the Arts and Humanities. London, New York: Routledge.

Raue, Benjamin. 2021. „Die Freistellung von Datenanalysen durch die neuen Text und Data Mining-Schranken.“ *Zeitschrift für Urheber- und Medienrecht* 56, H. 10: 793–802.

Schöch, Christof, Frédéric Döhl, Achim Rettinger, Evelyn Gius, Peer Trilcke, Peter Leinen, Fotis Jannidis, Maria Hinzmann und Jörg Röpke. 2020. „Abgeleitete Textformate: Text und Data Mining mit urheberrechtlich geschützten Textbeständen.“ *ZfdG* 5. DOI: https://doi.org/10.17175/2020_006.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg et al. 2016. „The FAIR Guiding Principles for scientific data management and stewardship.“ *Sci Data* 3, H. 160018. DOI: <https://doi.org/10.1038/sdata.2016.18>.