

Sammlungen als Daten – Das Projekt „Bibliotheca Eugeniana Digital“ als Use Case aus der Österreichischen Nationalbibliothek

Eva Mayr, Annerose Tartler-Ostrizek, Florian Windhager, Simon Mayer

Zusammenfassung: Die Initiative „Sammlungen als Daten“ transformiert die Nutzung von Kulturerbe-Sammlungen, indem sie diese als maschinenlesbare Daten aufbereitet, um digitale Technologien wie Visualisierung, maschinelles Lernen und Datenanalyse zu ermöglichen. Ein Beispiel dafür ist das Projekt *Bibliotheca Eugeniana Digital*, das von der Österreichischen Nationalbibliothek gemeinsam mit der Universität für Weiterbildung Krems umgesetzt wurde. Ziel war die Rekonstruktion der Privatbibliothek von Prinz Eugen von Savoyen mittels digitaler Technologien (Maschinelles Lernen, HTR), um die Sammlung für weitere Forschung zu öffnen. Zu den Ergebnissen des Projekts zählt eine digitale Edition, ein offenes Datenset, sowie eine Visualisierung der Sammlung, die sowohl Forschenden als auch der Öffentlichkeit zugutekommt. Herausforderungen wie die Qualität der Digitalisierung, strukturierte Datenformate, Langzeitarchivierung und Zugänglichkeit, rechtliche Aspekte, sowie Nutzerfreundlichkeit und Visualisierung werden in diesem Kontext diskutiert. Dieses Projekt unterstreicht die Bedeutung des Ansatzes „Sammlungen als Daten“ für die Erforschung und Vermittlung historischen Wissens.

Schlagwörter: Collections as Data, Sammlungsvisualisierung, digitale Edition, historische Bibliothek

Gefördert durch: Österreichische Akademie der Wissenschaften, Förderprogramm „go!digital 3.0“

Collections as Data – The Project “Bibliotheca Eugeniana Digital” as Use Case at the Austrian National Library

Abstract: The “Collections as Data” initiative is transforming the use of cultural heritage collections by preparing them as machine-readable data to enable digital technologies such as visualization, machine learning and data analysis. One example of this is the project *Bibliotheca Eugeniana Digital* by the Austrian National Library and the University for

Continuing Education Krems, which is reconstructing the private library of Prince Eugene of Savoy using digital technologies (machine learning, HTR) to enable further research. The results of the project are a digital edition, an open data set and a visualization of the collection that will benefit both researchers and the public. Challenges such as the quality of digitization, structured data formats, long-term archiving and accessibility, legal aspects, as well as user-friendliness and visualization will be discussed using the project as an example. This project underlines the approach of collections as data for the research and communication of historical knowledge.

Keywords: Collections as data, collection visualisation, digital edition, historical library

Funded by: Austrian Academy of Sciences, funding program „go!digital 3.0“

DOI: <https://doi.org/10.31263/voebm.v78i1.9296>

Dieses Werk ist – exkl. einzelner Logos und Abbildungen – lizenziert unter einer Creative-Commons-Lizenz Namensnennung 4.0 International-Lizenz

1. Sammlungen als Daten in Bibliotheken

„Collections as Data“ (Sammlungen als Daten, vgl. <https://collections-asdata.github.io/>) ist eine Initiative, die die Rolle von Bibliotheken, Museen und Archiven in der digitalen Ära neu definiert. Während Sammlungen historisch als statische Aggregate behandelt wurden, die zur Lektüre, Erforschung und Bewahrung physischer Medien dienen, so eröffnet deren Digitalisierung neue Möglichkeiten. Durch den Ansatz von *Sammlungen als Daten* werden digitale Sammlungen in maschinenlesbare Formate umgewandelt, was die vertiefende Nutzung digitaler Methoden wie Datenanalyse oder maschinelles Lernen ermöglicht.

Das „Santa Barbara Statement on Collections as Data – Always Already Computational“ (Padilla et al., 2019) wurde vor mittlerweile fünf Jahren veröffentlicht. In dieser Erklärung werden Kulturerbeinstitutionen dazu aufgerufen über das bloße Scannen ihrer Sammlungsobjekte hinauszugehen und diese weiter aufzubereiten, um entsprechend hochwertige Daten mit den Methoden der digitalen Geisteswissenschaften besser zugänglich zu machen – für die Forschung, Praxis und

Öffentlichkeit. Im Detail wurden zehn Prinzipien formuliert (s. Tabelle 1), die neben der Stärkung von computergestützten Methoden auch die Klärung zahlreicher anderer Fragen befürworten. Zu diesen zählen Fragen der Ethik der Digitalisierung, Zugänglichkeit und Benutzerorientierung, Desiderate von transparenter Dokumentation und Nachhaltigkeit, Interoperabilität und offener Schnittstellen, sowie ein prozessorientiertes Verständnis von Sammlungen als Daten.

1. Die Initiative Sammlungen als Daten (SaD) zielt darauf ab, die computergestützte Nutzung von digitalisierten und digitalen Sammlungen zu fördern.
2. Kurator:innen von SaD unterliegen langfristigen ethischen Verpflichtungen.
3. Kurator:innen von SaD reduzieren Barrieren für deren Nutzung.
4. SaD orientieren sich an ihren Nutzer:innen, aber nicht nur an einer speziellen Zielgruppe.
5. Geteilte Dokumentationen helfen anderen, mit SaD zu arbeiten.
6. SaD sind offen zugänglich, außer ethische bzw. rechtliche Verpflichtungen stehen dem entgegen.
7. Bei der Entwicklung von SaD wird Wert auf Interoperabilität gelegt.
8. Kuratoren von SaD arbeiten transparent, um Vertrauen und eine langfristige Verfügbarkeit zu gewährleisten.
9. SaD umfassen sowohl Daten als auch Metadaten.
10. Die Entwicklung von SaD ist ein fortlaufender Prozess, der nicht notwendigerweise mit einer finalen Version endet.

Tabelle 1: 10 Prinzipien für Sammlungen als Daten (SaD) nach Padilla et al. (2019), übersetzt durch die Autor:innen.

Trotz der Bedeutung dieser Initiative für die geisteswissenschaftliche Forschung und die Zugänglichkeit von Kulturerbesammlungen, folgt die Digitalisierung in vielen Bibliotheken noch nicht solchen hohen Standards. Die europäische GLAMlabs Community hat daher konkrete Richtlinien entwickelt, die Kulturerbeinstitutionen dabei unterstützen sollen, diesen Ansatz technisch umzusetzen (Candela et al., 2023, s. Tabelle 2). Dabei obliegt es den jeweiligen Institutionen und Projekten entsprechend ihrer Ziele und Nutzungsszenarien zu entscheiden, welche Schritte in welcher Reihenfolge implementiert werden sollen.

1. Definition einer Nutzungslizenz, die eine uneingeschränkte Wiederverwendung der SaD ermöglicht
2. Bereitstellung eines Zitiervorschlags für die SaD
3. Erstellung einer Dokumentation über die SaD
4. Verwendung einer öffentlichen Plattform, auf der die SaD offen zugänglich ist
5. Bereitstellung von Beispielen, die zeigen, wie die SaD wiederverwendet werden kann
6. Klare Strukturierung der SaD, um ihre Wiederverwendung zu erleichtern
7. Bereitstellung maschinenlesbarer Metadaten für die SaD
8. Verwendung einer Kollaborationsplattform für die gemeinsame Bearbeitung der SaD
9. Bereitstellung einer API für die SaD
10. Website zur Präsentation und Beschreibung der SaD, um seine Wiederverwendung zu fördern
11. Definition von Nutzungsbedingungen.

Tabelle 2: Workflow in 11 Schritten zur Erstellung von Sammlungen als Daten (SaD) nach Candela et al. (2023), übersetzt durch die Autor:innen.

Dennoch scheitert die Umsetzung solcher Projekte oft an fehlenden Ressourcen und fehlender Infrastruktur. Bei der Schaffung von Infrastrukturen, die die quantitative und qualitative Erforschung von großen Datenmengen unterstützen, können Bibliotheken einen wichtigen Beitrag leisten. So folgt auch die Österreichische Nationalbibliothek (ÖNB) mit den ÖNB Labs diesem Ansatz, bereitet sukzessive einzelne Sammlungen aus ihren Beständen als Datensätze digital auf, und macht sie auf ihrer Plattform offen zugänglich (Kaiser, 2023). Ein Projekt im Rahmen dieser Initiative ist Bibliotheca Eugeniana Digital, welches wir in diesem Beitrag vorstellen wollen, um danach zu reflektieren, welche konkreten Schritte unternommen wurden, um die Prinzipien der Initiative Sammlungen als Daten umzusetzen.

2. Bibliotheca Eugeniana Digital

Bibliotheken gelten als traditionelle Bewahrer von verschriftlichtem Wissen, doch dessen Verfügbarkeit variiert je nach Aufbewahrungs- und Katalogisierungsmedium. Insbesondere durch den verstärkten Einsatz digitaler Technologien in den letzten Jahrzehnten ist diese Diskrepanz noch offensichtlicher geworden. So wurde an der ÖNB

beispielsweise der historische Zettelkatalog, als Vorgänger des Online-Katalogs, direkt in diesen überführt, während weitere, vor allem handschriftliche Quellen aufgrund des benötigten langwierigen Aufarbeitungsprozesses oftmals keine Beachtung fanden.

Dies ist auch der Fall für den handschriftlichen Katalog der *Bibliotheca Eugenia* (Unbekannt, 1700), der Privatbibliothek des Prinzen Eugen von Savoyen (1663–1736), die 1738 für die Hofbibliothek Wien (heute ÖNB) angekauft wurde. Die etwa 15.000 gedruckte Bücher und 2.400 Manuskripte umfassende Bibliothek stellt heute ein bedeutendes Zeugnis des wissenschaftlichen und künstlerischen Geistes des 17. und 18. Jahrhunderts dar und wurde als solche von der UNESCO als Teil des „Gedächtnisses von Österreich“ anerkannt. Trotz ihrer historischen Bedeutung blieb der Umfang, die Zusammensetzung und der aktuelle Status dieser Sammlung innerhalb der ÖNB bisher weitgehend unerforscht. Beschreibungen der Bibliothek beziehen sich in der Regel auf eine Reihe von gut dokumentierten Objekten, wie die berühmte Tabula Peutingeriana, aber Informationen über die Zusammensetzung der Sammlung in ihrer Gesamtheit sind begrenzt, da ihre schiere Größe und fehlende Dokumentation eine vollständige Rekonstruktion bis dato verhinderten.

Bibliotheca Eugenia Digital ist ein im Rahmen des „go!digital 3.0“-Calls der Österreichischen Akademie der Wissenschaften gefördertes Projekt (Projektlaufzeit 11/2022–10/2024), das sich erstmals der Aufarbeitung dieser Sammlung und ihres historischen Kontexts mit Methoden der Digital Humanities und Data Science annimmt. Großflächige Digitalisierung, algorithmisch unterstützte Workflows und die Anwendung von künstlicher Intelligenz bieten hierbei neue Möglichkeiten, die Büchersammlung digital zu rekonstruieren und im Sinne der Initiative Sammlungen als Daten zugänglich zu machen. Die digitale Zusammenführung des vormals nur physisch vorliegenden handschriftlichen Katalogs und einer Machine Learning-basierten Klassifizierung der Supralibros-Einbände Prinz Eugens bietet hierbei den Ausgangspunkt für eine umfassende Erforschung und Verfügbarmachung der Sammlung.

Der fünfbändige zeitgenössische Katalog wurde hierzu zunächst eingescannt und der thematische Teil (Bände 1–3) anschließend mithilfe der Handschrifterkennungs-Plattform Transkribus transkribiert. Um die im handschriftlichen Katalog befindlichen Informationen mit dem

modernen (Online-)Bibliothekskatalog der ÖNB zu verbinden, wurden die Einträge im Anschluss über eine Ähnlichkeitssuche semiautomatisch dem entsprechenden Katalogeintrag zugeordnet. Während dieser Prozess mit Blick auf die große Datenlage ein gutes Ergebnis liefert, ist er doch mit einiger Unsicherheit verbunden. So ist es möglich, dass ein Eintrag mit einem Buchtitel zu 100% übereinstimmt, es sich jedoch um ein Exemplar aus einem anderen Bestand handelt. Diese Unsicherheit kann nur dann vollständig geklärt werden, wenn weitere Provenienzmarker berücksichtigt werden können.

Eine weitere wichtige Quelle für die digitale Rekonstruktion sind die Supralibros-Einbände der historischen Sammlung. Diese verschiedenfarbigen Ledereinbände, die mit dem Wappen von Prinz Eugen versehen sind, dienen als zusätzlicher Indikator für die Provenienz der Bücher. Mithilfe von maschinellem Lernen wurden die Einbände der gesamten digitalisierten historischen Buchsammlung der ÖNB analysiert, um sowohl das entsprechende Wappen zu identifizieren, als auch die taxonomisch relevante Farbe der Einbände. Diese Methode diente als weiterer Indikator für die datenbasierte Zuordnung der Bücher zur Bibliotheca Eugeniana.

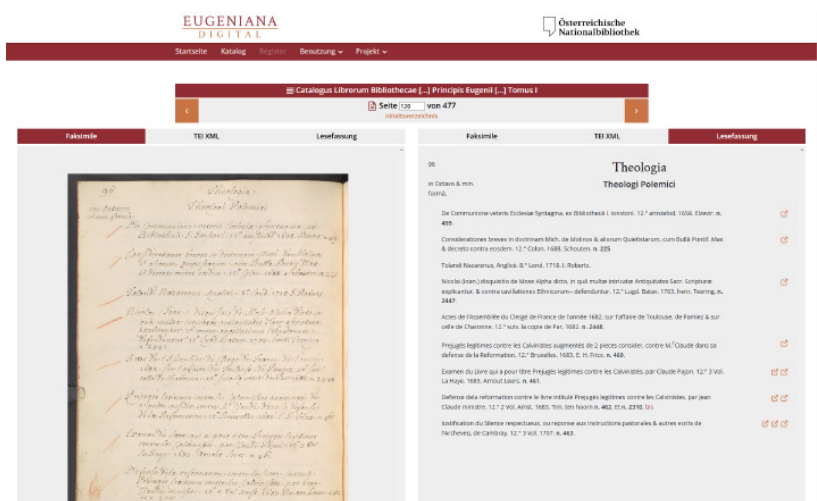


Abbildung 1: Digitale Edition des historischen handschriftlichen Katalogs der Bibliotheca Eugeniana. Ein Faksimile wird dabei einer Lesefassung mit Verknüpfungen zu den Einträgen im modernen Katalog gegenübergestellt.

Die im Projekt erzeugte Digitale Edition (siehe <https://edition.onb.ac.at/context:eugeniana>) kombiniert Informationen aus dem handschriftlichen Katalog, den Supralibros-Analysen und dem modernen Katalog der ÖNB und kann Forschenden eine direkte Verbindung zwischen historischem Eintrag und Originalquelle bieten (Abbildung 1).

Zusätzlich zu der Digitalen Edition, die sich hauptsächlich an Wissenschaftler:innen richtet, wurde eine Visualisierung der historischen Büchersammlung entwickelt, um Nutzer:innen neue Möglichkeiten zur digitalen Erkundung und Analyse der Sammlung zu bieten (siehe <https://labs.onb.ac.at/bed/eugeniana/>, sowie das Tutorial zum Einstieg). Sie begünstigt neben der visuellen Analyse auch die freie Erkundung der Bibliothek und ist auch für Nutzer:innen ohne Vorkenntnisse zugänglich. Das Interface ermöglicht die Exploration der Sammlung anhand verschiedener Dimensionen wie geografischer oder historischer Herkunft, Sprache und thematischer Klassifikation der Bücher (siehe Abbildung 2). So bietet es neben Überblicksvisualisierungen zu ausgewählten analytischen Fragen für Fachleute auch eine ansprechende Aufbereitung der Sammlung für die breite Öffentlichkeit.

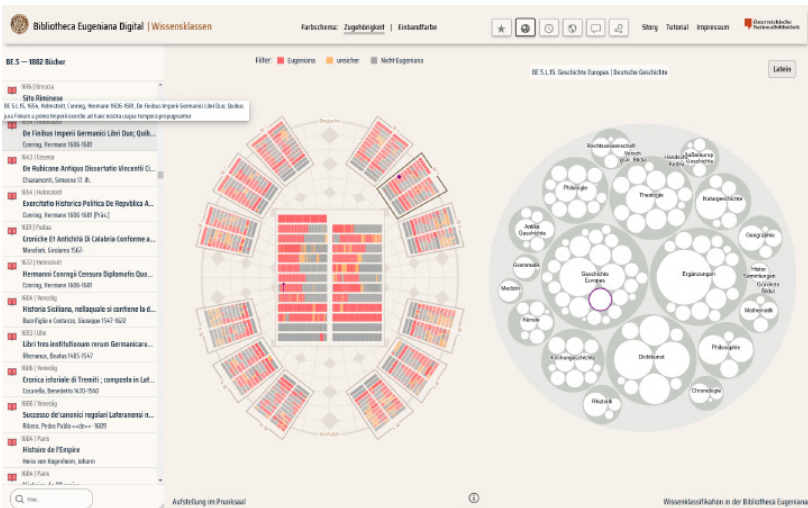


Abbildung 2: Visualisierung der Bibliotheca Eugeniana. Nutzer:innen können die Sammlung anhand ihrer Aufstellung im Prunksaal, sowie anhand der historischen Wissensklassifikation erkunden. Über die Buchliste können die Volltexte aufgerufen werden.

3. Bibliotheca Eugeniana Digital im Licht von Sammlungen als Daten

Die Digitalisierung einer historischen Buchsammlung im Rahmen von Sammlungen als Daten erfordert nicht nur die Umwandlung physischer Bücher in digitale Dateien, sondern eine umfassende Umstrukturierung der Daten, um sie für maschinelle Analyse nutzbar zu machen. Die Qualität der Digitalisierung, standardisierte Datenformate, nachhaltige Speicherung und Interoperabilität sind zentrale Anforderungen der Initiative Sammlungen als Daten (Padilla et al., 2019; Candela et al., 2023). Gleichzeitig müssen rechtliche und ethische Aspekte berücksichtigt und den Nutzer:innen benutzungsfreundliche Werkzeuge zur Analyse und Visualisierung bereitgestellt werden (Padilla et al., 2019). Welche und wie viele von diesen Anforderungen eine Sammlung als Daten erfüllt, obliegt den jeweiligen Projektanforderungen – dem Projektziel, dem Kontext, den Inhalten, den Benutzer:innen und der intendierten Nutzung der Daten (Candela et al., 2023). Im Projekt *Bibliotheca Eugeniana Digital* wurden folgende Schritte gesetzt:

1. Qualität der Digitalisierung: Die Digitalisierung des historischen Katalogs muss in hoher Qualität erfolgen, um sicherzustellen, dass alle Details originalgetreu reproduziert werden. Für spätere Analysen sind fehlerfreie Daten essenziell. Daher wurden die handgeschriebenen Texte zunächst automatisiert mit Transkribus digitalisiert, jedoch in einem zweiten Schritt durch Projektmitarbeiter:innen manuell überprüft. Dies stellte sich speziell für unregelmäßige Seiten, die spätere Ergänzungen durch einen zweiten Schreiber beinhalteten als essenziell heraus. Aus dem digitalisierten Text wurden die Metadaten über Wissensklasse, Buchformat und Titeleintrag extrahiert, um eine spätere Datenanalyse und Visualisierung zu ermöglichen. Aus dem modernen Bibliothekskatalog wurden für jeden einzelnen Eintrag die Metadaten zu Autor, Jahr, Erscheinungsort und Sprache übernommen, die bereits größtenteils in standardisierter Form vorliegen und damit maschinell einfacher zu verarbeiten sind.
2. Strukturierte Datenformate: Der Text der digitalen Edition wurde in einem TEI (Text Encoding Initiative) XML-Format gespeichert, das die semantische Analyse unterstützt. Durch die auf textuelle Quellen ausgerichteten, klar definierten Strukturen von TEI XML können ver-

schiedene Elemente wie Wissensklassen, Buchformate und Katalogeinträge gekennzeichnet werden, was eine gezielte Analyse erleichtert. Daneben wurden die Katalogeinträge mit den entsprechenden Signaturen der Bücher im modernen Bibliothekskatalog verknüpft.

3. Langzeitarchivierung und Zugänglichkeit: Speicherlösungen wie XML, JSON und RDF gelten als stabile, nachhaltige Formate, die langfristig zugänglich und kompatibel mit künftigen Technologien bleiben. Die zugrundeliegenden Daten des Projekts *Bibliotheca Eugeniana Digital* sind über die projekteigene GitLab-Instanz offen zugänglich, sodass Forschende leicht darauf zugreifen und sie für weiterführende Analysen verwenden können. Zudem sichert die Integration der Projektergebnisse in die ÖNB-Infrastruktur eine weit über den Projektzeitraum hinaus reichende Bereitstellung und Wartung von Technologien und Schnittstellen, die für die unmittelbare Einsichtnahme in die Daten, ohne zusätzlichen Aufwand wie das eigene Herunterladen und Visualisieren, erforderlich sind.
4. Rechtliche Aspekte: Als historisches Werk ist der historische Katalog der *Bibliotheca Eugeniana* urheberrechtsfrei. Auch die digitale Edition und die im Projekt entstandenen Daten können unter der Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) Lizenz wiederverwendet werden.
5. Nutzerfreundlichkeit und Visualisierung: Um den Umfang der *Bibliotheca Eugeniana* für Forschende und die interessierte Öffentlichkeit greifbarer zu machen, wurde zusätzlich zur digitalen Edition in einem nutzungszentrierten Designprozess eine Sammlungsvisualisierung entwickelt. Innerhalb dieser Visualisierung ist auch eine Suchfunktion implementiert, die das Auffinden einzelner Werke in der Sammlung oder von Werken zu bestimmten Themen erleichtert.

Mit diesen Schwerpunkten setzte das Projekt *Bibliotheca Eugeniana Digital* wesentliche Prinzipien der Initiative *Sammlungen als Daten* (Padilla et al., 2019) um. Innerhalb des Projekts wurden bereits einige weiterführende digitale Analyseverfahren (z.B. visuelle Suche, Texterkennung, semantische Ähnlichkeitssuche, Visualisierung) genutzt, um die Daten weiter anzureichern und die Buchsammlung Prinz Eugens besser zu verstehen. Durch die Bereitstellung der Sammlung als Daten steht sie aber auch interessierten Forscher:innen für weiterführende

Analysen und der Öffentlichkeit zur Exploration dieses UNESCO Kulturerbes zur Verfügung.

4. Diskussion und Ausblick

Der digitale Wandel in Bibliotheken wirft wichtige Fragen bezüglich des nicht digital verfügbaren Wissens und dessen Zugänglichkeit auf. Fehlende Verfügbarkeiten verzerren den Zugang zu historischen Sammlungen und beschränken die Möglichkeiten der umfassenden Erforschung und Vermittlung. Gleichzeitig bieten digitale Methoden neue Wege, Informationen zu verifizieren und zu überprüfen, was zur Sicherung und Validierung historischer Quellen beiträgt. Einen wichtigen Beitrag hierzu liefert die Initiative Sammlungen als Daten. In diesem Artikel wurde diskutiert, wie im Projekt *Bibliotheca Eugenia Digital* die Empfehlungen dieser Initiative umgesetzt wurden, um diese historische Sammlung in der Form von Daten zugänglich zu machen, die von hoher Qualität, wohlstrukturiert und interoperabel, nachhaltig zugänglich, benutzerfreundlich und offen sind.

Die so entstandenen Daten konnten bereits im Rahmen des Projekts genutzt werden, um die tradierte Aussage, dass die Farbe des Einbandes mit bestimmten Wissensklassen in Zusammenhang steht, als nur teilweise korrekt zu identifizieren (Tartler, Mayr, Windhager & Mayer, 2025). Doch viele Fragen über die historische Buchsammlung Prinz Eugens sind noch ungeklärt. Die Initiative Sammlungen als Daten und Projekte in ihrem Sinne laden die wissenschaftliche Gemeinschaft dazu ein, die neu verfügbaren Daten für diese und weiterführenden Forschungsfragen zu nutzen.

Dr.ⁱⁿ Eva Mayr
ORCID iD: <https://orcid.org/0000-0001-8402-5990>
Universität für Weiterbildung Krems
eva.mayr@donau-uni.ac.at

Annerose Tartler-Ostrizek, BA MA
ORCID iD: <https://orcid.org/0009-0009-4141-4270>
Österreichische Nationalbibliothek
atartler@naturalsciences.be

Mag. Dr. Florian Windhager
ORCID iD: <https://orcid.org/0000-0002-5170-2243>
Universität für Weiterbildung Krems
florian.windhager@donau-uni.ac.at

Dr. Simon Mayer
ORCID iD: <https://orcid.org/0009-0008-3322-5130>
Österreichische Nationalbibliothek
simon.mayer@onb.ac.at

Literatur

Candela, G., Gabriëls, N., Chambers, S., Dobрева, M., Ames, S., Ferriter, M., Fitzgerald, N., Harbo, V., Hofmann, K., Holownia, O., Irollo, A., Mahey, M., Manchester, E., Pham, T.-A., Potter, A. & Van Keer, E. (2023), „A checklist to publish collections as data in GLAM institutions“, *Global Knowledge, Memory and Communication*, preprint. <https://doi.org/10.1108/GKMC-06-2023-0195>. **Kaiser**, M. (2023). Digitale Sammlungen als offene Daten für die Forschung: Strategische Zielsetzungen der Österreichischen Nationalbibliothek. *Bibliothek Forschung und Praxis*, 47(2), 200-212. <https://doi.org/10.1515/bfp-2023-0021>. **Padilla**, T., Allen, L., Frost, H., Potvin, S., Russey Roke, E. & Varner, S. (2019). *Santa Barbara Statement on Collections as Data — Always Already Computational: Collections as Data* (Version 2). Zenodo. <https://doi.org/10.5281/zenodo.3066209>. **Tartler**, A., Mayr, E., Windhager, F. & Mayer, S. (2025). Digitale Erschließung historischer Bibliotheken: Erkenntnisse und Perspektiven aus dem Projekt Bibliotheca Eugeniana Digital. *Bibliothek Forschung und Praxis*, <https://doi.org/10.1515/bfp-2024-0074>. **Unbekannt** (1700). *Catalogus librorum bibliothecae principis Eugenii de Sabaudia*. Vol. I-III. *Signatur ÖNB Cod. 14376–14387*. <http://data.onb.ac.at/rec/AC13948858>, <http://data.onb.ac.at/rec/AC13956022>, <http://data.onb.ac.at/rec/AC13956023>.