

Sammlungen als Forschungsdaten

Datennutzung als Gemeinschaftsaufgabe

Gerber, Anja

anja.gerber[at]klassik-stiftung.de
Klassik Stiftung Weimar, Deutschland
ORCID: 0000-0003-2576-1511

Gnyp, Anna

anna.gnyp[at]hu-berlin.de
Humboldt-Universität zu Berlin, Deutschland
ORCID: 0000-0003-3598-1112

von Hagel, Frank

f.v.hagel[at]smb.spk-berlin.de
Institut für Museumsforschung, Deutschland
ORCID: 0000-0003-1349-2131

Markert, Michael

michael.markert[at]fau.de
Friedrich-Alexander-Universität Erlangen-Nürnberg, Deutschland
ORCID: 0000-0002-8943-4870

Schönemann, Hendrik

hendrik.schoenemann@hu-berlin.de
Humboldt-Universität zu Berlin, Deutschland
ORCID: 0000-0002-7091-0283

Winkler, Alexander

winkler[at]zib.de
Zuse Institut Berlin, Deutschland
ORCID: 0000-0002-9145-7238

Zusammenfassung. Sammlungsdaten sind auch Forschungsdaten und als solche werden sie verstärkt wahrgenommen. Die computerbasierte Verarbeitung von Sammlungsdaten birgt jedoch gewisse Herausforderungen, da die Datenangebote die Bedarfe der Nutzung in der Forschung etwa mit Blick auf Zugänglichkeit und Interoperabilität nicht adäquat adressieren. Eine verbesserte Nutzbarkeit setzt eine enge Kooperation zwischen Sammlungen und Forschenden voraus. In diesem Workshop

sollen mittels Personaentwicklung die Anforderungen der Nutzung von Sammlungsdaten in der Forschung herausgearbeitet werden.

1 Sammlungen als Forschungsdaten

1.1 Potenziale von Sammlungsdaten für die Forschung

Objektsammlungen in Museen, Archiven, Bibliotheken und Universitäts-sammlungen besitzen große Relevanz für die Forschung und können als eigenständige Forschungsinfrastrukturen verstanden werden. Dass die der Sammlung zugehörigen Daten (Metadaten, Bilddaten u. a.) in digitaler Form vorliegen, bietet neue Möglichkeiten für Forschende, diese Infrastrukturen zu nutzen.

Auf das Potenzial dieser Daten für die Forschung hat in jüngeren Jahren die Arbeitsgruppe „Always Already Computational: Collections as Data“ hingewiesen (Padilla et al. 2018, 2023). Sie betont die Bedeutung der auf computergestützte Verarbeitung ausgelegten Verfügbarmachung in offenen Formaten, unter freien Lizenzen und mit Rücksicht auf urheberrechtliche und ethische Fragen.

Auch der Wissenschaftsrat hat in seinen Empfehlungen die Wichtigkeit der Digitalisierung für Sammlungen betont und dabei insbesondere die hierdurch beförderte „Nutzbarkeit und Nutzungsintensität“ (Wissenschaftsrat 2011; vgl. für ähnliche Statements auch Rat für Informationsinfrastrukturen 2024) der digitalisierten Sammlungsobjekte unterstrichen.

Die Potenziale einer computergestützten und datenfokussierten Forschung sind dabei vielfältig. Sie umfassen etwa quantitative Analysen, Datenverknüpfung und -visualisierung, Methoden der Computer Vision wie bildbasierte Ähnlichkeitssuchen, neue Formen der Zusammenstellung sowie viele weitere, teilweise noch unbekannte Verfahren.

1.2 Status Quo der Datenangebote

Kulturerbeeinrichtungen, Forschungseinrichtungen und Hochschulen digitalisieren seit beinahe drei Jahrzehnten ihre Bestände in größerem Maßstab und machen die Digitalisate samt Metadaten über das Internet zugänglich. Der Umfang der Daten ist beträchtlich, obgleich bisher beispielshalber lediglich die Hälfte der Museen und nur 40% der wissenschaftlichen Sammlungen Teile ihrer Bestände digital zur Verfügung stellen (Schlösser et al. 2025). Aktuell weist die Deutsche Digitale Bibliothek (<https://www.deutsche-digitale-bibliothek.de/>), die zentrale nationale Aggregationsplattform, mehr als 60 Millionen Datensätze nach, von denen über 23 Millionen Objekte auch über ein Digitalisat verfügen.

Daneben werden Sammlungsdaten aber auch über diverse Verbund- und Fachportale oder eigene Websites zur Verfügung gestellt.

Zum Flickenteppich der Datenpublikation kommt hinzu, dass die datengebenden Institutionen unterschiedliche Ziele verfolgen. Lange Zeit haben z.B. Museen und Sammlungen Objekte vorrangig publiziert, um ihre Ausstellungen und Häuser zu repräsentieren und zu bewerben. Erst mit der Etablierung der FAIR-Prinzipien (Wilkinson et al. 2016) hält allmählich auch ein Verständnis dafür Einzug, welche neuen Möglichkeiten maschinenlesbare Sammlungsdaten für die geisteswissenschaftliche Forschung bieten. Folglich ist die Datenqualität der online verfügbaren Sammlungsdaten für die forschende Nutzung oftmals noch unzureichend.

Auch der Zugriff auf maschinenlesbare Metadaten über Schnittstellen oder Datendumps ist zwar zunehmend möglich, doch immer noch mit teils beträchtlichen Hürden verbunden.

2 Nutzung als Gemeinschaftsaufgabe

Eine effektive, zielführende und innovative Nutzung von Sammlungsdaten verlangt – so die These – eine enge Kooperation zwischen Bereitsteller:innen und Nutzer:innen von Sammlungsdaten, um die Nutzung der Datenangebote zu steigern und Hürden abzusenken. Auch ‚Intermediäre‘ (hier verstanden als Vermittlungsinstanzen für *collection data literacy*, wie z.B. *data stewards*, NFDI-Konsortien und Datenkompetenzzentren) spielen in dieser Kooperation eine wichtige Rolle.

Hierbei stellt sich die Frage, welche Rollen und Aufgaben den einzelnen Akteur:innen zukommen: Was brauchen datengebende Institutionen, um Daten in hoher Qualität, offen und standardisiert bereitzustellen? Wie können sie ihre Datenangebote und Services besser auf die Bedürfnisse der Nutzer:innen ausrichten?

Welche Aufgaben haben Forschende? Sind sie lediglich passive Datenkonsument:innen oder tragen sie auch Verantwortung, die Potenziale der Datenangebote für ihre Forschung zu prüfen oder im Sinne eines produktiven Miteinanders selbst Kompetenzen zu erwerben, um Daten nach eigenen Bedürfnissen aufzubereiten und diese an Datengeber:innen zurückzuspielen?

Wie können Intermediäre diesen Austausch befördern und neben Kompetenzvermittlung, Bedarfe, Potenziale und Herausforderungen in beide Richtungen antizipieren? Können sie auch beispielhafte Anwendungsfälle dokumentieren und standardisierte Arbeitsprozesse wie digitale Forschungsmethoden entwickeln und vermitteln?

3 Ablauf und Ziele des Workshops

Im dreistündigen Workshop sollen erste Schritte zur Bewältigung der Datennutzung als Gemeinschaftsaufgabe unternommen werden. Nach einem kurzen initialen Input tragen die Teilnehmer:innen in einem ersten Teil kooperativ einschlägige Quellen für Sammlungsdaten und ihre Merkmale, Zugriffsmöglichkeiten und die mit ihrer Nutzung verbundenen Herausforderungen zusammen. Auf dieser Grundlage soll diskutiert werden, welche Forschungsfragen sich mit den Daten bearbeiten ließen und welche Bedarfe die Datenangebote in ihrer aktuellen Form nicht abdecken.

In einem zweiten Teil werden die Teilnehmer:innen Personas entwickeln, die Forscher:innen aus der Kunstgeschichte, Geschichtswissenschaft, Archäologie oder auch Provenienzforschung darstellen mit jeweils eigenen Forschungsfragen und Nutzungsszenarien, in denen Sammlungsdaten verwendet werden. Zum Schluss sollen auf Grundlage dieser Personaentwicklung mittels einer Matrix Anforderungen an die verschiedenen Akteur:innen (Nutzer:innen, Intermediäre, datengebende Institutionen) genauer gefasst werden.

Ziel des Workshops ist es, (1) dass die Teilnehmenden sich einen kritischen und effektiven Umgang mit Datenangeboten von Sammlungen erarbeiten. (2) Mit den Personas soll ein Impuls für eine Diskussion um eine gelebte Datenkultur geliefert werden, die gerade vor dem Hintergrund der immer größeren Relevanz digitaler Methoden in den Geisteswissenschaften und des Ausbaus von Forschungsdateninfrastrukturen absehbar intensiver geführt werden wird. (3) Der erarbeitete Anforderungskatalog kann schließlich auch datengebenden Institutionen und Intermediären eine wichtige Hilfe bei der zielgruppenorientierten Weiterentwicklung ihrer Daten- bzw. Serviceangebote bieten.

Es ist geplant, die Ergebnisse des Workshops zu dokumentieren und im Nachgang in geeigneter Form unter freier Lizenz zu publizieren.

Bibliografie

- Padilla, Thomas, Laurie Allen, Hannah Frost, Sarah Potvin, Elizabeth Russey Roke und Stewart Varner. 20. Mai 2019. „*Santa Barbara Statement on Collections as Data – Always Already Computational: Collections as Data*“. <https://doi.org/10.5281/zenodo.3066209>.

- Padilla, Thomas, Hannah Scates Kettler, Stewart Varner und Yasmeen Shorish. 13. September 2023. „*Vancouver Statement on Collections as Data*“. <https://doi.org/10.5281/zenodo.8342171>.
- Rat für Informationsinfrastrukturen. 2024. *Sammlungen als multimodale Infrastrukturen. Analog und digital für die verknüpfte Nutzung erschließen*. Göttingen. <https://rfii.de/?p=11282>.
- Schlösser, Magdalene, Johannes Schäffer, Frank von Hagel und Felix Schäfer. 2025. *Überblick über das Forschungsdatenmanagement in Museen und Universitätssammlungen*. <https://doi.org/10.5281/zenodo.14926511>.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. *The FAIR Guiding Principles for Scientific Data Management and Stewardship*. *Scientific Data* 3: 160018. <https://doi.org/10.1038/sdata.2016.18>.
- Wissenschaftsrat. 2011. *Empfehlungen zu wissenschaftlichen Sammlungen als Forschungsinfrastrukturen*. <https://www.wissenschaftsrat.de/download/archiv/10464-11.html>.