

Offene Werkgenesen, Editionen und Archive. Versuch einer generischen Datenmodellierung

Bürgermeister, Martina

martina.buergermeister@onb.ac.at
Österreichische Nationalbibliothek, Österreich

Pektor, Katharina

katharina.pektor@onb.ac.at
Österreichische Nationalbibliothek, Österreich

Steindl, Christoph

christoph.steindl@onb.ac.at
Österreichische Nationalbibliothek, Österreich

Eigner, Johanna

johanna.eigner@onb.ac.at
Österreichische Nationalbibliothek, Österreich

Einleitung

Unser Beitrag zur Tagung *Open Humanities*, *Open Culture* präsentiert einen innovativen Vorschlag für die Datenmodellierung zukünftiger digitaler Werkeditionen. Das Modell ist hochgradig flexibel, indem es eine Diversität und komplexe Relationalität der Werkdaten zulässt, und zugleich stabil, wodurch es maximale Offenheit der Rezeption und Nutzbarkeit gewährleistet.

Entwickelt und erprobt wurde das Datenmodell im Zuge der Daten-Migration der Forschungsplattform *Handkeonline* (<https://handkeonline.onb.ac.at/>), bei der eine vorerst "geschlossene" Modellierung von präbibliographischen Daten der verschiedenen, genetisch interpretierten Werkmaterialien aus dem Vorlass zusammen mit bibliographischen Daten der veröffentlichten Werke des Literaturnobelpreisträgers Peter Handke in eine "offene" umgebaut werden musste – wir nennen diese Modellierung eine "offene Werkgenese".

In unserer Präsentation werden wir zuerst kurz in die Voraussetzungen der Plattform *Handkeonline* (Punkt 2) und in den aktuellen Forschungsstand der semantischen Modellierung von bibliographischen und präbibliographischen Daten (Punkt 3) einführen. Danach wird anhand von zwei Beispielen der generische Modellierungsansatz für „offene Werkgenesen“ präsentiert (Punkt 4).

Forschungsplattform *Handkeonline*

Die Forschungsplattform *Handkeonline* wurde im Zuge eines FWF-Forschungsprojekts am Literaturarchiv der Österreichischen Nationalbibliothek Wien 2011-2015 entwickelt. Sie versteht sich als virtuelles Archiv, das sämtliche auf öffentliche und private Archive in Österreich, Deutschland und der Schweiz verstreute Vorlass-Materialien (wie Notizbücher, Bleistiftmanuskripte und Typoskripte unterschiedlicher Textfassungen sowie Druckfahnen, aber auch annotierte Bücher, Fotos, Landkarten oder sogar Wanderstöcke) zum veröffentlichten Werk Peter Handkes zusammenführt. Die Materialien werden tabellarisch und in Form von Texten beschrieben und in eine entstehungschronologische Beziehung gesetzt, die in Paratexten kommentiert wird. Die werkgenetische Interpretation und Beschreibung erweitern das virtuelle Archiv zu einer Edition. Ergänzt wurde diese durch eine umfangreiche Bibliographie zur Primär- und Sekundärliteratur des über 140 Bücher sämtlicher Genres umfassenden Werks Handkes. Inhaltliche und formale Vielfalt öffnen die Plattform für ein breites Publikum. Sie wird nicht nur für die universitäre Forschung intensiv genutzt, sondern bietet Schüler*innen Unterstützung für Hausarbeiten, wird von Theatern für die Erarbeitung von Stückinszenierungen oder Begleitheften herangezogen und, wie die Aufregungen um den Nobelpreis gezeigt haben, auch von Journalist*innen konsultiert. (Der faktischen Materialdokumentation der Plattform kam etwa innerhalb der kontroversen Debatte um Handkes "Serbien-Bücher" eine wichtige Rolle zu.)

Zentrales Anliegen des *Handkeonline*-Projekts war die Präsentation und Weitergabe des erarbeiteten biographischen, literaturwissenschaftlichen und archivalischen Wissens über Handke, die Entstehung seiner Werke, und damit zusammenhängend über die Beschaffenheit und Bedeutung der einzelnen Vorlass-Materialien. Die nachhaltige Modellierung dieser Forschungsdaten, welche erst eine langfristige Verfügbarkeit sowie Ausbau- und Anschlussfähigkeit des Wissens garantiert, wurde dabei vorderhand aus zeitökonomischen und budgetären Gründen vernachlässigt. Vor allem aber fehlte noch das geeignete konzeptuelle Modell, das von Literatur-, Editions- oder Archiv- und Informationswissenschaft gemeinsam gedacht wird, und damit auch eine praktische Vorlage. Die Daten von *Handkeonline* wurden im Altsystem flach strukturiert und in einem proprietären Format vorgehalten. Die Folgen wurden bereits nach Projektende deutlich: die innovative Darstellung der Werkgenese war zwar Vorbild für andere Projekte, aber diese wurde weder in den Daten repräsentiert noch nachhaltig strukturiert. Deutlich zeigte sich das Problem in dem 2021, also nur sechs Jahre später begonnenen digitalen Editionsprojekt *Peter Handke Notizbücher* der Österreichischen Nationalbibliothek Wien und des Deutschen Literaturarchivs Marbach (<https://edition.onb.ac.at/handke-notizbuecher>): auf die wertvollen Daten über die Notizbücher konnte sogar intern nur mehr kompliziert über die dem Content-Management-System (Drupal) zugrunde liegende relationale Datenbank zugegriffen werden. Die damals getroffenen Zuordnungen der

einzelnen Datenfelder konnten wegen des geschlossenen Datenformats teilweise nicht mehr vollständig rekonstruiert werden.

Aufgrund der Bedeutung von *Handkeonline* beschloss die Österreichische Nationalbibliothek die Daten-Migration in die erst später eingerichtete *Nachhaltige Infrastruktur für digitale Editionsprojekte an der Österreichischen Nationalbibliothek* (ÖNB-DE) und nützt diese Gelegenheit zum vollständigen Re-Design der Daten und Datenstruktur. Ziel ist die Transformation sämtlicher für *Handkeonline* generierte Beschreibungsdaten zum Werk Peter Handkes nach TEI/XML und die Bereitstellung eines werkgenetischen Katalogs als *Linked Open Data*-Datensatz. Diese Umsetzung liefert einen wesentlichen Beitrag zu einer Forschungskultur, die Vernetzung und Dynamik von Forschungsprozessen ins Zentrum rückt. Folgende Herausforderungen galt es dabei zu bewältigen:

- Einen Modellierungsansatz zu finden, mit dem präbibliographische und bibliographische Daten gemeinsam repräsentiert und in Beziehung gesetzt werden können.
- Die geschlossene in eine offene Werkgenese umzuwandeln und den Datenkatalog entsprechend den FAIR-Prinzipien (Findable, Accessible, Interoperable, Reuseable) zur Verfügung zu stellen.

Modelle für präbibliographische und bibliographische Daten

Zur Erschließung und Verbreitung von bibliographischen Daten haben sich im wissenschaftlichen Bereich verschiedene bibliographische Datenbanken (*Initiative for Open Citation* (i4oc.org)) etabliert, wie z.B. *Open Citation* oder *GoogleScholar*, die sich allerdings auf die Erfassung bereits veröffentlichter Werke konzentrieren. Das gilt auch für eine Reihe weitverbreiteter Standardvokabulare für die grundlegenden Begriffe zur Dokumentation und Beschreibung bibliographischer Daten wie etwa *Dublin Core Element Set* und *Metadata Terms* (<https://www.dublincore.org/>) und *Schema.org*. Das *BIBFRAME*-Vokabular (<https://www.loc.gov/bibframe/docs/index.html>) wiederum ist ein vor allem im wissenschaftlichen Bereich gängiges (Austausch-)Format und man kann MARC- und RDA-Daten im Linked-Data-Standard RDF abbilden – dieses Vokabular ist jedoch ebenfalls ungeeignet für die Repräsentation präbibliographischer Werkmaterialien. Für das Projekt *Handkeonline* sind diese gängigen Vokabulare somit zu limitiert, da hier unveröffentlichte Werke und unterschiedlichste Werkvorstufen (Werkmaterialien) weder erfasst noch in ihrer teilweise komplexen Relationalität zu den publizierten Werken abgebildet werden können.

Eine differenzierte bibliographische Basiskodierung, wie sie Martina Gödel und Sebastian Zimmer (2017) auf der Basis der *Functional Requirements for Bibliographic Records* (FRBR) (Coyle 2016) für die Werke und vor allem den Zettelkasten von Niklas Luhmann gemacht haben

(<https://niklas-luhmann-archiv.de/>), ist die Voraussetzung für weitere Forschungen zur Werkgenese. In dem von uns gewählten Ansatz spielt gerade dieser Aspekt deshalb von Beginn an eine Schlüsselrolle. Auch unsere Modellierung bibliographischer Daten basiert wie zuvor schon bei Stefanie Gehrke et al. (2016), Frederike Neuber (2016), Gödel und Zimmer (2017) oder Andreas Lüscho (2020) auf den Grundkonzepten von FRBR, doch wir gehen für unsere Anforderungen noch einen Schritt weiter:

Wir greifen, um die für unsere Abbildung der Werkgenesen notwendige Flexibilität (Werkdatendiversität und -referenzierbarkeit) bei gleichzeitiger Stabilität zu erreichen, die für eine Anschluss- und Ausbaufähigkeit der Daten notwendig ist, auf das im Museumsbereich gebräuchliche Format FRBR_{OO} zurück. FRBR_{OO} ist eine Harmonisierung des FRBR- *Entity-Relationship-Models* und des *Conceptual Reference Models* des *International Council of Museums CIDOC CRM* (<https://www.cidoc-crm.org/>). Letzteres wurde entwickelt, um die in der Museumsdokumentation relevanten Objekte und Konzepte (Ereignisse) zu beschreiben und institutionsübergreifend Daten austauschen zu können (Doerr, Crofts 2004, 15). FRBR_{OO} führt diese ereigniszentrierte Sichtweise nun auf Werke aller Art ein. Die Anwendung von FRBR_{OO} für die Modellierung von komplexen Werkgenesen ist, soweit wir die digitalen Editionen der letzten Jahre überblicken, neu; sie bringt viele Vorteile. FRBR_{OO} ermöglicht eine exakte Darstellung der komplexen Genese von Handkes Werken: Erstens, weil das Modell differenzierte Werkbegriffe bereitstellt, die uns erlauben, nicht nur die zahlreichen literarischen Genres, die Handke bedient (Prosa, Theater, Hörspiel, Film, Briefe, Interviews, Essays, Gedichte, Journale), sondern auch seine bislang unveröffentlichten, aber Werken entsprechenden Notizbücher zu formalisieren. Zweitens, weil FRBR_{OO} auf CIDOC-CRM basiert, ist es möglich auch individuelle, nichtbibliographische Materialien zu beschreiben. Drittens können mit FRBR_{OO} die Vorstufen eines Buches (*Manifestation Singleton*) von einer veröffentlichten Ausgabe (*Manifestation Product Type*) in den Daten klar unterschieden werden. Viertens schließlich erlaubt uns FRBR_{OO}, über das Ereignis *Expression Creation* die für eine Darstellung von Werkgenesen wichtige zeitliche und materielle Einordnung der einzelnen Fassungen.

Datenmodellierung von *Handkeonline* mit FRBR_{OO}

Für die Modellierung von *Handkeonline* gilt es die vielfältigen Werkmaterialien aus dem Vorlass mit den publizierten Werken in Zusammenhang zu bringen. Peter Handkes Notizbüchern kommt dabei für die Genese seiner Werke ein besonderer Stellenwert zu: sie sind Materialsammlung oder Skizze für mehrere publizierte Werke; sie sind erste Textfassung für die gedruckte Notizauswahl in Journalen und darüber hinaus in ihrer Gesamtheit noch unveröffentlichtes Werk, das im Zuge des digitalen Editionsprojektes *Peter Handke Notizbücher* erstmals publiziert wird. Wir gehen deshalb von zwei unterschiedlichen Werkbegriffen aus: einmal von dem von Handke autorisierten veröffentlichten Werk mit seinen verschie-

phischen Daten semantisch differenziert erfassen wollen.

Erst durch die hier besprochene Öffnung der Daten entsteht die Affordanz zur Repräsentation in vollem Umfang. Die Effekte sind vielfältig: Die Migration der isolierten Daten und deren Re-Modellierung nach anerkannten Standards machen das Material langzeitarchivfähig und schaffen die Grundlage für LOD. Durch den Einsatz von FRBR_{OO} wird die Interoperabilität der vernetzten Informationen sichergestellt. Insgesamt werden durch die Öffnung neue Wege der Analyse geschaffen. Abfragen zu werkhistorisch relevanten Aspekten werden möglich: Wie die differenzierte Abfrage der Werkkomponenten oder zu aufführungsspezifischem Material oder sogar zu parallelen Entwicklungsprozessen; gleichzeitig wird auch eine Visualisierung von Stammbäumen zur Entstehung der Werke möglich. Wir erwarten mit Spannung, wie sich unsere Anwendung der formalen Semantik von FRBR_{OO} in Zukunft auf die Reflexion über Werk und Werkgenesen auswirken wird.

2020: *Spielräume: Digital Humanities zwischen Modellierung und Interpretation*. <https://zenodo.org/record/4621704#.YuedhRzP2Uk> (zugegriffen: 01.08.2022).

Neuber, Frederike. 2016. „Stefan George Digital“ In *DHd 2016: Modellierung - Vernetzung - Visualisierung: Die Digital Humanities als fächerübergreifendes Forschungsparadigma*. <https://zenodo.org/record/4645060#.YuefAhzP2Uk> (zugegriffen: 01.08.2022).

Bibliographie

Arbeitsstelle für Standardisierung. 2006. *Funktionale Anforderungen an Bibliografische Datensätze. Abschlussbericht der IFLA Study Group on the Functional Requirements for Bibliographic Records*. Leipzig/Frankfurt a.M./Berlin: DNB. <https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frbr/frbr-deutsch.pdf> (zugegriffen: 01.08.2022).

Bekiari, Chryssoula et al. 2015. *FRBR object-oriented definition and mapping from FRBR ER, FRAD and FRSD* (version 2.4). https://www.cidoc-crm.org/frbroo/sites/default/files/FRBRoo_V2.4.pdf (zugegriffen: 01.08.2022).

Coyle, Karen. 2015. *FRBR Before and After. A Look at Our Bibliographic Models*. Chicago: Ala Editions 2016. ISBN 978-0-8389-1364-2 (PDF). <http://192.248.73.38/bitstream/handle/94ous-I/926/RDA%20FRBR%20book%20for%20reading%20%20978-0-8389-1364-2.pdf> (zugegriffen: 01.08.2022).

Doerr, Martin und Patrick LeBoeuf. 2006. *FRBR object-oriented definition and mapping to FRBR ER (version 0.8.1). International Working Group on FRBR and CIDOC CRM Harmonisation*. https://archive.ifla.org/VII/s13/wg-frbr/FRBR_oo_V.0.8.1c.pdf (zugegriffen: 01.08.2022).

Gehrke, Stefanie, Pauline Charbonnier und Frunzeanu Eduard. 2016. „Biblissima - Semantic Web Application für Handschriften, Inkunabeln und historische Sammlungen - Zwischenbericht“ In *DHd 2016: Modellierung - Vernetzung - Visualisierung: Die Digital Humanities als fächerübergreifendes Forschungsparadigma*. <https://zenodo.org/record/4645060#.YuefAhzP2Uk> (zugegriffen: 01.08.2022).

Goedel, Martina und Sebastian Zimmer. 2017. „Niklas Luhmanns Werk- und Lesekosmos - DH in der bibliographischen Dimension“ In *DHd 2017: Digitale Nachhaltigkeit*. <https://dh-abstracts.library.cmu.edu/works/10641> (zugegriffen: 01.08.2022).

Lüschow, Andreas. 2020. „Automatische Extraktion und semantische Modellierung der Einträge einer Bibliographie französischsprachiger Romane“ In *DHd*