

Der Nachlass als Netzwerk

ZUR ENTWICKLUNG EINER NACHLASS-ONTOLOGIE AM BEISPIEL DES PROJEKTS
„STEFAN ZWEIG DIGITAL“

Christopher Pollin [1] & Lina Maria Zangerl [2]

[1] Universität Graz, Zentrum für Informationsmodellierung – Austrian Centre for Digital Humanities

[2] Universität Salzburg, Literaturarchiv Salzburg – Forschungszentrum von Universität, Land und Stadt Salzburg

christopher.pollin@uni-graz.at, linamaria.zangerl@sbg.ac.at

Keywords: Semantic Web, Nachlass, Ontologie, Stefan Zweig digital, Resource Discovery

Abstract: Within the project “Stefan Zweig digital”, a digital reconstruction of the posthumous papers of Stefan Zweig, a domain-specific ontology is modeled to represent context-sensitive cultural heritage objects flexibly and expressively according to the requirements of researchers.

Einleitung

Kulturerbeobjekte und ihre Metadaten – als geisteswissenschaftliche Daten verstanden – bringen besondere Herausforderungen mit sich: Sie sind kontextabhängig, haben bestimmte (abstrakte) Beziehungen zu anderen Objekten und weisen interne und externe Verknüpfungen zu Personen, Orten, Werken und (biographischen) Ereignissen, sowie Sekundärquellen auf. Im Hinblick auf Nachlässe ergibt sich zudem die Notwendigkeit einer Darstellung des Gesamtzusammenhangs des Materials, das sich in Unikalität, Umfang, Informalität und Individualität von klassischem Archivgut unterscheidet.¹ Ein Repository zur Langzeitarchivierung und *Semantic Web*-Technologien können dabei helfen, diesen Herausforderungen entgegenzutreten.

Im Rahmen des Projekts „Stefan Zweig digital“ (SZD),² einer digitalen Rekonstruktion des Nachlasses von Stefan Zweig, wird eine domänenspezifischen Nachlass-Ontologie modelliert. Ziel einer solchen Nachlass-Ontologie ist es, auf konzeptioneller Ebene semantische Interoperabilität, Nachvollziehbarkeit und Dokumentation zu gewährleisten, aber auch kontextsensitive Nachlassobjekte flexibel und ausdrucksstark nach den Anforderungen der Fachwissenschaftler*innen abbilden zu können.

1 Dirk WEISBROD, Langzeitarchivierung digitaler Schriftstellernachlässe. Eine Bestandsaufnahme in den Literaturarchiven Deutschlands, Österreichs und der Schweiz, in: *Young Information Scientist* 2 (2017), 1–12, hier 10, online unter: <http://eprints.rclis.org/30976/> (12.06.2019).

2 www.stefanzweig.digital.

Das Projekt Stefan Zweig digital

Im Zentrum der Forschungs- und Informationsplattform www.stefanzweig.digital, die seit Juni 2018 online frei zugänglich ist, stehen Leben und Werk des österreichischen Schriftstellers Stefan Zweig (1881–1942).³ Nicht zuletzt durch die vielfachen Stationen von Zweigs Exil ergab sich für die von ihm hinterlassenen Dokumente eine Verteilung auf zahlreiche Sammlungen in aller Welt und eine entsprechend heterogene Überlieferungslage. In der Folge standen die Originaldokumente eines der auflagenstärksten und umfassend vernetzten Autoren des 20. Jahrhunderts trotz anhaltendem Interesse an seinem Werk der wissenschaftliche Community bisher nur eingeschränkt zur Verfügung. Im Rahmen des Projekts SZD tritt das Literaturarchiv Salzburg, Forschungszentrum von Universität, Land und Stadt Salzburg, das bedeutende Originalmaterialien von und über Stefan Zweig aufbewahrt, diesen Schwierigkeiten entgegen: Erstmals werden die erhaltenen Manuskripte Zweigs sowie zugehörige Materialien systematisch online verzeichnet sowie in digitalen Faksimiles zur Verfügung gestellt. Für die Recherche auf der Plattform stehen neben den Katalogen zu Werkmanuskripten und Lebensdokumenten eine biographische Übersicht sowie ein Index der Personen und Standorte mit über 700 Einträgen zur Verfügung. Hinzu kommt ein Verzeichnis der erhaltenen Bücher aus Zweigs Bibliothek, das einen wichtigen Einblick in die von ihm wahrgenommene, gelesene und als Quelle für seine Werke genutzte Literatur bietet.⁴ In Kooperation mit dem Zentrum für Informationsmodellierung (ZIM) der Universität Graz sowie internationalen Partnern wie der Daniel A. Reed Library in Fredonia und der National Library of Israel wird der Nachlass des Schriftstellers Stefan Zweig als virtuelle Nachlassrekonstruktion der Öffentlichkeit zur Verfügung gestellt.

Zur technischen Basis

Eingebettet in das „Geisteswissenschaftliche Asset Management System“ (GAMS), ein OAIS-konformes, FEDORA basiertes Repository zur Langzeitarchivierung und Publikation von Forschungsdaten in den Geisteswissenschaften, existieren Workflows zur Standardisierung, Formalisierung und semantischen Anreicherung der Nachlassobjekte und ihrer Metadaten in einem digitalen Repository.⁵ Neben einer weitgehend XML-basierte Inhaltsstrategie zur Sicherstellung des Zugriff und der Langzeitarchivierung verwendet GAMS etablierte Metadatenstandards (TEI, METS), Datenformate (XML, RDF) und Programmierschnittstellen (IIIF). Alle digitalen Objekte im System erhalten eine persistente Kennung auf Basis des „Handle“-Systems, um die Zitation zu gewährleisten. Eine weitere Standardisierung wird durch die Verwendung des „International Image Interoperability Framework“ (IIIF) zur Verwaltung und Veröffentlichung digitaler Faksimiles erreicht.

Im Rahmen des Projektes werden erfasste Daten in TEI modelliert. METS dient als Standard zur Formalisierung digitalisierter Manuskripte und ihrer strukturellen Metadaten, die über IIIF in GAMS angeboten werden. Normalisierungen von vorkommenden *Named Entities* umfassen Personen (GND), Organisationen (GND), Orte (GeoNames) und spezifische Provenienzmerkmale (SKOS). Darüber hinaus werden Verknüpfungen zu Wikidata.org hergestellt. Beim Ingest der TEI/XML-Dokumente, beispiels-

³ Als verwandte Projekte, die in den letzten Jahren in Österreich entstanden sind, wären etwa die Plattformen Handke online (<https://handkeonline.onb.ac.at/>), Karl Kraus Online (<https://www.kraus.wienbibliothek.at/>) und Ernst Jandl Online (jandl.onb.ac.at/) zu nennen.

⁴ Stephan MATTHIAS / Oliver MATUSCHEK, *Stefan Zweigs Bibliotheken*, Dresden 2018.

⁵ Vgl. Johannes STIGLER / Elisabeth STEINER, *GAMS – An Infrastructure for the long-term preservation and Publication of Research Data from the Humanities*, in: *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare* 71 (2018), S. 207–216.

weise der Werkliste⁶, in die GAMS-Infrastruktur werden bestimmte Metadaten und Objektbeziehungen als RDF/XML und somit als semantische Struktur abgebildet. Die Normalisierung von *Named Entities* ermöglicht die semantische Anreicherungen der Daten, beispielsweise über das Tool OpenRefine, durch Wikidata.org oder durch die GND. Die *Retrieval-* (Volltext) und *Discovery-*Funktionalitäten basieren auf diesen RDF-Daten. In Übereinstimmung mit den FAIR-Datenprinzipien werden die Projektdaten als *Linked Open Data* (LOD) veröffentlicht und können nachgenutzt werden.⁷

Auf dem Weg zu einer domänenspezifischen Nachlass-Ontologie

Klassische archivarische und bibliothekarische Erschließungsmaßnahmen bilden die beste Grundlage für das Auffinden und Erforschen von Nachlassdokumenten. Sie haben aber ihre Grenzen, wenn es darum geht, die komplexen Verbindungen von Einzeldokumenten und Nachlassteilen untereinander und nach außen aufzuzeigen und das (Kontext-)Wissen der Bearbeiter*innen zu den Originalen zu formalisieren. *Semantic Web*-Technologien bieten in diesem Bereich neue Möglichkeiten, solche Verbindungen, die oftmals auch für Fachwissenschaftler*innen aus der Materialautopsie allein nicht ersichtlich wären, auf Basis einer Ontologien erstmals kenntlich zu machen.

Umreißen wir etwa folgendes Beispiel: Im Zuge seiner Beschäftigung mit der Französischen Revolution schrieb Stefan Zweig als eines von vier umfangreichen Werken 1932 die Biographie *Marie Antoinette – Bildnis eines mittleren Charakters*. Unter den von ihm hinterlassenen Papieren fand sich bisher keine komplette Fassung des Manuskripts von *Marie Antoinette*, das als Buch in der Erstausgabe beinahe 650 Druckseiten umfasst. Allerdings sind aus mehreren Arbeitsstufen und der Korrekturphase Skizzen, einzelne Kapitel oder Fragmente davon erhalten geblieben. In Zweigs Autographensammlung, die ihm wichtige Anregungen für seine literarischen Arbeiten bot, ist außerdem eine beachtliche Zahl von Handschriften aus dem Umfeld Marie Antoinettes und der Zeit der Französischen Revolution nachweisbar. *Marie Antoinette* wurde eines der meistverkauften Bücher des Jahres und noch im Erscheinungsjahr der Erstausgabe in zahlreiche Sprachen übersetzt. Wie zu zahlreichen seiner Werke haben sich auch zu *Marie Antoinette* Verträge mit ausländischen Verlagen erhalten. Zur besseren Übersicht über die Vermarktung seines weltweit vertriebenen Werkes wurde ab August 1932 im Büro Zweigs das so genannte Hauptbuch angelegt, in dem auf vorgedruckten Doppelseiten die verkauften Rechte für jedes Werk und die dafür in Verträgen vereinbarten Konditionen mit Verlagen, Übersetzern und Filmstudios festgehalten wurden. Auch eine Doppelseite zu *Marie Antoinette* ist darin zu finden. Unter den Bänden in Zweigs Bibliothek sind keine Quellenwerke für *Marie Antoinette* mehr vorhanden, allerdings liegen zahlreiche Belegexemplare von Übersetzungen des Buches als sogenannte „Hausexemplare“ vor.

Abbildung 1: Verbindungen rund um Zweigs Werk Marie Antoinette

Das Wissen um die hier beschriebenen Verbindungen kann nach der ausführlichen Erschließung jedes erhaltenen Objekts aus archivarischer und geisteswissenschaftlicher Perspektive durch die Modellierung semantischer Strukturen und Zurverfügungstellung als LOD menschen- und maschinenlesbar

6 TEI/XML der Werkliste, http://stefanzweig.digital/o:szd.werke/TEI_SOURCE.

7 RDF/XML Serialisierung der Werkliste, <http://stefanzweig.digital/o:szd.werke/RDF>.

zugänglich gemacht werden.⁸ Die Formalisierung auf Basis einer domänenspezifischen Nachlass-Ontologie kann den hierfür notwendigen formalen und kontextuellen Rahmen schaffen.

Das formale Modell der Ontologie ist dabei auf gängige Standards wie RNAB⁹, die GND-Ontologie¹⁰, RiC¹¹ und CIDOC-CRM¹², sowie FRBROO¹³ referenzierbar. Im Kontext des zur Verzeichnung von Archivgut entwickelten neuen Standards *Records in Contexts* (RiC) werden Überlegungen zur semantischen Interoperabilität in der Archivwelt diskutiert und umgesetzt. Ziel ist es, unter Berücksichtigung der Beschreibung durch etablierte Standards¹⁴ in der Archivwelt, die hierarchischen Organisationsstrukturen in Archiven aufzubrechen und Archivalien in unterschiedlichen Kontexten abzubilden. Diese Anstrengung werden durch RiC-O als Ontologie formalisiert,¹⁵ die auch für Nachlässe nutzbar ist.¹⁶ CIDOC-CRM ist bereits ein etabliertes semantisches Framework im *Semantic Web*, das als ‚gemeinsame Sprache‘ dazu dienen soll, Kulturerbeobjekte und ihre Dokumentation zu beschreiben, um deren konzeptionelle Interoperabilität sicherzustellen. Einzelnen Spezifikationen und Erweiterungen von CIDOC-CRM, die für bestimmte Anwendungsbereiche entwickelt wurden, fördern die semantische Modellierung unterschiedlichster Ressourcen. Eine für SZD relevante Spezifikation ist FRBROO. Dabei handelt es sich um eine Ontologie mit dem Ziel, den Austausch zwischen biographischen und musealen Informationen zu erleichtern.

Solcherart kann eine Nachlass-Ontologie nicht nur als Wissensbasis begriffen werden, sondern auch als konzeptionelle Grundlage für die Nutzung und Zurverfügungstellung von *Linked Open Data* sowie darauf fußender erweiterter Funktionalitäten zum „digitalen Entdecken“, dem *Resource Discovery*.¹⁷ Diese Funktionalitäten folgen dem Paradigma der *Exploratory Search*,¹⁸ die Joemon Jose mit dem Satz „*I don't know what I'm looking for, I'll know it when I find it*“¹⁹ auf den Punkt bringt. Die Formulierung von ‚klassischen‘ Suchanfragen, kann nur dann geschehen, wenn ein gewisses Vorwissen über einen Bestand vorherrscht. SZD versucht beispielsweise über biographische und thematische Einstiegspunkte solches Wissen zu vermitteln.

Trotz der Herausforderungen, die in der Wartung, Verknüpfung und Abfrage des *Semantic Web* liegen, erlauben eine domänenspezifische Ontologie und die Zurverfügungstellung als LOD die Daten und ihre Kontexte – im Fall von SZD also die Objekte des Nachlasses und das Wissen über sie – so zu

-
- 8 Eine wesentliche Herausforderung im Bereich der Forschungsdaten ist es, Datensilos zu vermeiden. Mit dem *Semantic Web* wird ein Technologie-Stack angeboten, um die von Tim Berners-Lee definierten *Linked Open Data* (LOD)-Paradigmen umsetzen zu können. Vgl. Krzysztof JANOWICZ u.a., Five stars of linked data vocabulary use, in: *Semantic Web* 5/3 (2014), 173–176.
 - 9 Ressourcen-Erschließung mit Normdaten in Archiven und Bibliotheken, <https://nbn-resolving.org/urn:nbn:de:101-2019051405> (10.06.2019).
 - 10 GND Ontology, <https://d-nb.info/standards/elementset/gnd> (12.02.2019).
 - 11 Records in Contexts - Conceptual Model, <https://www.ica.org/en/egad-ric-conceptual-model> (01.06.2019).
 - 12 Definition of the CIDOC Conceptual Reference Model, <http://www.cidoc-crm.org/Version/version-6.2.5>, Version. 6.2.5, Stand Februar 2019.
 - 13 Object-oriented Functional Requirements for Bibliographic Records, <http://www.cidoc-crm.org/frbroo/home-0>, Version 3.0, Stand Oktober 2017.
 - 14 ISAD(G), ISAAR(CPF), ISDF and ISDIAH.
 - 15 Llanes-Padrón DUNIA / Juan-Antonio PASTOR-SANCHEZ, Records in contexts: the road of archives to semantic interoperability. *Program* 51/4 (2017), 387–405.
 - 16 Philipp MESSNER, Records in Contexts – vom Baum zum Netz, in: *Arbido* 3 (2017), online unter <https://arbido.ch/de/ausgaben-artikel/2017/metadaten-datenqualit%C3%A4t/records-in-contexts-vom-baum-zum-netz> (12.06.2019).
 - 17 Beispiele für solche konkreten Anwendungsszenarien, die es erlauben, formalisierte semantische Strukturen nutzen zu können, finden sich etwa in den Projekten *ResearchSpace* des British Museum (<https://www.researchspace.org>) oder *Burghardt Source* (<http://burckhardtsource.org>).
 - 18 Ryen WHITE / Resa A. ROTH, Exploratory search: beyond the query-response paradigm, San Rafael, Calif. 2009, 14.
 - 19 Joemon M. JOSE, Personalisation Techniques in Information Retrieval, in: *DigiCULT* 6 (2004), 22.

formalisieren, dass sämtliche Struktur den Daten immanent ist. Das heißt die Daten ‚wissen‘ so viel über sich selbst, dass sie durch andere automatisiert weiterverarbeitet werden können. So kann der materialreiche und inhaltlich vielfältige Nachlass Stefan Zweigs in einen wesentlich weiter gefassten Zusammenhang des kulturellen und intellektuellen Lebens seiner Zeit gestellt werden.

Literaturverzeichnis

- Llanes-Padrón DUNIA / Juan-Antonio PASTOR-SANCHEZ, Records in contexts: the road of archives to semantic interoperability. Program 51/4 (2017).
- Krzysztof JANOWICZ u.a., Five stars of linked data vocabulary use, in: Semantic Web 5/3 (2014).
- Joemon M. JOSE, Personalisation Techniques in Information Retrieval, in: DigiCULT 6 (2004).
- Stephan MATTHIAS / Oliver MATUSCHEK, Stefan Zweigs Bibliotheken, Dresden 2018.
- Philipp MESSNER, Records in Contexts – vom Baum zum Netz, in: Arbido 3 (2017), online unter: <https://arbidoc.ch/de/ausgaben-artikel/2017/metadaten-datenqualit%C3%A4t/records-in-contexts-vom-baum-zum-netz> (12.06.2019).
- Johannes STIGLER / Elisabeth STEINER, GAMS – An Infrastructure for the long-term preservation and Publication of Research Data from the Humanities, in: Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare 71 (2018), S. 207–216.
- Dirk WEISBROD, Langzeitarchivierung digitaler Schriftstellernachlässe. Eine Bestandsaufnahme in den Literaturarchiven Deutschlands, Österreichs und der Schweiz, in: Young Information Scientist 2 (2017), online unter: <http://eprints.rclis.org/30976/> (12.06.2019).
- Ryen WHITE / Resa A. ROTH, Exploratory search: beyond the query-response paradigm, San Rafael, Calif. 2009.