

Events: Modellierungen und Schnittstellen

Kurzzusammenfassung

Einerseits sind Ereignisse im menschlichen Leben, und damit auch in dessen Dokumentation, allgegenwärtig. Andererseits gibt es in den Digital Humanities gerade für diesen zentralen Bereich noch deutlichen Nachholbedarf sowohl a) betreffend die Modellierung und darauf aufbauend Kodierung von Ereignissen, als auch b) betreffend die Austauschbarkeit von Daten über Ereignisse, für die es weder Standards noch – in anderen Bereichen bestehende – umfassende Normdatenquellen etwa aus dem Bibliotheksbereich (GND, VIAF, GeoNames etc.) in breit akzeptierter Form gibt.

Die EinreicherInnen haben angesichts dessen zur TEI-Konferenz 2019 eine Initiative unternommen, das Datenmodell von `tei:event` zu erweitern, ihm in Analogie zu anderen Named Entities aus dem TEI-Modul 'namesdates' ein `tei:eventName` zwecks Referenzierung im `tei:body` zur Seite zu stellen und die Möglichkeiten des `tei:listEvent`-Wrapperelements zu ergänzen. Während wir – deren Domäne die Digitale Edition ist – dieses Ziel in Abstimmung mit dem TEI-Konsortium verfolgen, streben wir außerdem einen breiteren Austauschprozess mit anderen Bereichen innerhalb der Digital Humanities an, welche sich ebenfalls mit Ereignissen beschäftigen. Das vorgeschlagene Panel versammelt ausgewiesene Expertinnen und Experten zu einer offen angelegten Diskussion zur Modellierung von Ereignissen sowie zu möglichen Einsatzszenarien einer 'eventSearch API'.

Hintergrund

Die Projektidee – in Anlehnung an `CorrespSearch` (Korrespondenzdatennetzwerk) ein Ereignisdatennetzwerk aufzubauen –, Ereignisdaten homogenisiert zu sammeln, diese zusammenzuführen und sie als historischen Background datumsbezogen zur Verfügung zu stellen, entstand aus den individuellen Interessen dreier unterschiedlicher Editionsprojekte an drei unterschiedlichen Institutionen, die von WissenschaftlerInnen verschiedener akademischer Disziplinen betrieben werden: Christiane Fritze und Christoph Steindl, Österreichische Nationalbibliothek; Infrastruktur für digitale Editionen an der Österreichischen Nationalbibliothek mit einer Tagebuchedition in TEI // Helmut W. Klug vom Zentrum für Informationsmodellierung der Uni Graz; mehrere Editionsprojekte mittelalterlicher Texte im GAMS // Stephan Kurz, Österreichische Akademie der

Wissenschaften; hybride Edition Ministerratsprotokolle der Habsburgermonarchie in TEI.

Wegen der Vielfalt der Anwendungsfälle und der edierten Materien (im einreichenden Team sind das etwa: Tagebuch, Itinerar, Kalender, Protokoll) verzichten wir auf eine normative Definition des Ereignisbegriffs abseits von „Ein Ereignis ist eine in Zeit, Frequenz und Raum verortbare Zustandsänderung eines oder mehrerer Objekte oder ihrer wechselseitigen Bezüge, die durch eine oder mehrere Quellen belegbar und mit einem oder mehreren Bezeichnungen benannt sein kann.“ In diesem Rahmen lassen sich in unseren Editionen ganz unterschiedliche Qualitäten von Ereignissen beschreiben, wie etwa: a) Frau X schreibt am 14.6.1932 am Ort Y in ihr Tagebuch, welches sie 3 Monate später an Herrn Z in A sendet; b) In dem Tagebuch berichtet sie vom Traum der letzten Nacht, in dem sie von der Geburt einer Tochter geträumt hat, c) Der Benediktinerpater F. langt am Michaelistag 1340 nach mehrtägiger Reise in Modriach an, d) Im Protokoll mit Signatur 777 steht: Die Minister A, B, D und N lehnen in der Sitzung vom 15.7.1876 das Gnadengesuch des Mörders M. ab, e) Kaiser F. lässt die Kriegserklärung übermitteln. Die Verantwortung für die Granularität, Reziprozität (nesting events) und Auswahl der für einen Datenbestand als Ereignisse erwähnenswerten Zustandsänderungen liegt bei der jeweiligen Herausgeberinnen.

Zielsetzung

Ausgehend von dem Bedürfnis, die Daten zu Ereignissen aus (den oben skizzierten, aber auch allgemein) TEI-Editionen bzw. anderen Datenquellen vergleichbar und distribuierbar zu machen, verfolgen wir mit dem Panel das übergeordnete Ziel, den wissenschaftlichen Diskurs zur Modellierung, Sammlung und Darstellung von Ereignissen nach der überaus erfolgreichen Diskussion mit der TEI-Community auch in einem breiteren DH-Kontext voranzutreiben. Eine erfolgreiche Umsetzung eines derartigen Service kann nur erreicht werden, wenn Vorschläge und Feedback eines möglichst heterogenen InteressentInnenkreises vorliegen und die Möglichkeiten anderer Dokumentationsstandards (z.B. CIDOC CRM) sowie die Interoperabilität der Ansätze in Betracht gezogen werden. Darüberhinaus müssen Fragen zur Vernetzung von editionsgetriebenen und fremddatengestützten Ereignisdaten mit weiteren Daten z.B. aus prosopographischen Forschungsunternehmungen bzw. aus dem Semantic Web (Linked Open Data) diskutiert werden.

Konkret heißt das:

- Vorstellung der Ideen zu eventSearch mit kritischer Rückmeldung
- Sammlung von Zugängen zur Ereignismodellierung aus LOD/Semantic Web

- Verbreiterung der Basis an Überlegungen durch Öffentlichkeit dieser Diskussion
- Vernetzung von Event-Interessierten (Konsortialbildung? Infrastrukturschaffung?)
- Einladung zur Mitarbeit

Methoden

Panel. Um statt der im CFP definierten 30 Minuten für die Diskussion mit dem Publikum zumindest 45 Minuten aufwenden zu können und damit notwendige Rückkoppelung und Feedback zu erhalten, sind nach einer Themeneinführung sehr kurze Einleitungsstatements (5 Minuten pro beteiligter Initiative/Person) geplant. Nach einer Diskussionsrunde innerhalb der Panel-TeilnehmerInnen zu dem Thema Modellierung von Ereignissen sowie zum Thema Schnittstelle werden Fragen und Anregungen gemeinsam mit dem Publikum diskutiert.

Stand der Diskussion zu Ereignissen im Kontext TEI-gestützter digitaler Editionen

Ereignisse sind allgegenwärtige Merkmale menschlichen Lebens zu allen Zeiten an allen Orten. Die deutschsprachige Wikipedia listet für den 27. September 65 Ereignisse in Politik und Wirtschaft ('events'), TV-Sender und Zeitungen bieten ähnliche Formate, indem sie in einer Art Panoptikum historische Ereignisse Revue passieren lassen, die EU plant ein monumentales Time Machine Projekt. Allgemein gesprochen: Es besteht immer Interesse daran, unterschiedlichste Ereignisse in Relation zueinander zu setzen. Besonders spannend wird das, wenn man dies nicht nur als eine Art Rückschau betreibt, sondern versucht, dieses Nebeneinander von Ereignissen für einen Tag, eine Periode in der Vergangenheit darzustellen. Entsprechend unausweichlich trifft man auf Erwähnung und Dokumentation von Ereignissen in historischen wie zeitgenössischen Quellen, die Grundlage geisteswissenschaftlicher Forschung sind. Aber ein Nebeneinander von Ereignissen kann auch für fiktionale Texte festgestellt werden. Historische Zeugen können Lebensdokumente (Tagebücher, Reisenotizen) genauso sein wie amtliche Dokumente (Urkunden, Protokolle, Kalender) oder Kulturüberlieferungen materieller Natur (Inschriften, Teppiche, `tei:object` u.a.). Während Ontologien wie CIDOC CRM Ereignisse als Zustandsänderungen beschriebener Objekte modellieren, gehen textbasierte Schemata wie die TEI bislang eher implizit davon aus, dass einem Überlieferungsträger auch ein (z.B. Schreib-)Ereignis ursächlich zugrunde liegt; im Text beschriebene Ereignisse sind bestenfalls im Sinne einer Referenzierung in Analogie zu anderen Klassen von Named Entities erfasst. Auch die Linked Open

Data-Welt beschreibt/referenziert Ereignisse. Den Hiatus zwischen diesen Ansätzen zu harmonisieren fällt aus unserer Sicht am ehesten konkreten Projektanwendungen zu, die bspw. im Zusammenhang mit prosopographischen Auxiliardaten für digitale Editionen stehen – deren Aufmerksamkeitsspanne endet allerdings an den Enden des konkreten Anwendungsfalles. Das vorgeschlagene Panel richtet seine Aufmerksamkeit auf diesen neuralgischen Punkt in der DH-Forschungslandschaft.

Besonderes Augenmerk der Panel-Diskussion gilt auch der Frage der Granularität von Ereignissen in den verschiedenen Ansätzen: In CIDOC CRM etwa ist prinzipiell jede distinkte kleinteilige Zustandsänderung als Ereignis konzeptionalisiert, wogegen für das Ziel eines Discovery- und Disseminationservices eher ein alltagssprachliches Verständnis von Ereignissen Nutzen verspricht (im Editionscontext mit der Lösung in TEI, dass all jenes zum Ereignis wird, das von der Editorin/dem Editor als `tei:event` ausgezeichnet ist).

Alternativen zu CIDOC CRM und/oder TEI, stehen zur Verfügung und sollten ebenfalls diskutiert werden:

- The Simple Event Model Ontology, <https://semanticweb.cs.vu.nl/2009/11/sem/>
- Normdaten zu historischen Einzelereignissen z.B. aus der GND-Ontologie, die die bibliothekarischen Schlagwortketten in super/sub-Relationen modelliert
- CMIF (nicht erst in der Version 2, siehe Dumont et al. 2019) als ausmodellierter Spezialfall: `correspActions` als Subklasse von Ereignissen
- ...

Ereignisse im Namensraum TEI:

Werden Quellen und Texte unabhängig von ihrem Fiktionalitätsgrad digital ediert, hat sich zwar die Kodierung nach den Richtlinien der TEI inklusive der Erfassung von Named Entities durchgesetzt, jedoch fehlt eine gängige Praxis für die semantisch eindeutige Erfassung und Auszeichnung von Ereignissen. In den TEI-Guidelines werden zum Zeitpunkt der Einreichung Ereignisse (`tei:event`) geradezu nachlässig behandelt und es wird ihnen keine Eigenständigkeit zugesprochen: Das `tei:event`-Element ist derzeit nur recht eingeschränkt verfügbar und kann ausschließlich als Kindelement anderer Elemente aus dem Names-Dates-Modul verwendet werden: Im Allgemeinen sind Ereignisse anderen Konzepten (Date, Person, Organisation usw.) untergeordnet. Der Eindruck, dass 'Events' nicht besonders im Fokus der TEI-Community stünden, täuscht allerdings, denn seit 2010 gibt es die unterschiedlichsten Diskussionen im Rahmen der TEI-Mailingliste und im TEI-C GitHub. Dabei geht es ganz allgemein um die Auszeichnung von Ereignissen, aber es werden auch vielversprechende Themen wie die Einführung eines `<eventName>`-Elements besprochen und natürlich, dass es notwendig wäre, das Event-Element zu stärken und es z.B. den vergleichbaren

Elementen aus dem Names-Dates-Modul anzugleichen. Mit der TEI P5 Version 3.6.0 wurden `tei:ptr` und `tei:idno` als Kindelemente von `tei:event` erlaubt. Dieses Event untermauert den Bedarf an einer tiefergehenden und viele AkteurInnen einschließenden Diskussion zum Ereignis. Hinzu kommt, dass zum Ereignis gehörige grundlegende Informationen wie ein zeitlicher Hinweis, eine Lokalisierung oder die Angabe zu beteiligten Personen nicht ohne Weiteres möglich ist. Im entsprechenden Vorschlag zur Änderung/Erweiterung der TEI Guidelines wird die Kodierung von Ereignissen im Rahmen der TEI angepasst und werden dem TEI-Konsortium umfassend ausgearbeitete Änderungsvorschläge für eine universelle Auszeichnung von Events vorgelegt. Dabei wird eine möglichst flache Auszeichnung angestrebt, um die Datenmigration für Ereignisdaten so niedrigschwellig wie möglich zu halten. Das vorgeschlagene Modell wird auf bestehende Editionsprojekte der InitiatorInnen angewendet, um so ein exemplarisches Korpus an `eventSearch`-Daten bereitzustellen.

Elemente einer Verschaltung von Ereignissen

Ein Vorschlag, den die InitiatorInnen des Panels verfolgen, ist die Erstellung einer Programmierschnittstelle (API) basierend auf dem erarbeiteten und mit dem TEI-Konsortium abgestimmten und tw. noch abzustimmenden Datenmodell für Ereignisse sowie einer webbasierten grafischen NutzerInnenschnittstelle für in digitalen Editionen kodierte Ereignissen nach dem Vorbild von `correspSearch`. Events aus externen Quellen werden ebenso eingebunden und ermöglichen eine Betrachtung einzelner Ereignisse im historischen Kontext. So bietet eine kalendarische Übersicht raschen Überblick über singuläre Ereignisse, die möglicherweise von mehreren Zeugen und Zeugnissen betrachtet worden sind oder die koinzidieren. Die Übersicht lädt so zur Interpretation und Kontextualisierung ein. Im Gegensatz zu der bloßen Auflistung von Ereignissen zu einem bestimmten Zeitpunkt, wie sie leicht etwa über Wikipedia (historische Jahrestage) einzusehen ist, erfahren die NutzerInnen hier über die Verlinkung in die jeweiligen Quellen individuelle Wahrnehmung und Bewertung: Aus welcher Quelle/Edition stammen die Informationen, wessen Sicht auf ein bestimmtes Ereignis wird wiedergegeben, wer stellt das Ereignis/den chronologischen Kontext des Ereignisses wie dar?

Die von den PanelinitiatorInnen angedachte `EventSearch`-Infrastruktur ermöglicht es, die aufgezeichneten Ereignisse einzelner digitaler Editionen sichtbar und leichter auffindbar zu machen. Eine Rückverlinkung von der Webschnittstelle/API zu digitalen Editionen gewährleistet, dass Ereignisbezüge einfach in der originalen Fassung angezeigt und analysiert werden können. Eine niederschwellige Möglichkeit zu

partizipieren wird angeboten, sodass Editionsprojekte unabhängig von Größe und Budget nicht daran gehindert werden, ihre Ergebnisse zu teilen.

Bestätigte TeilnehmerInnen

- eine Person aus dem Kreis der InitiatorInnen:
Wir stellen die Ergebnisse der Event-Extraction aus unseren eigenen Editionsdaten bis dato vor:
 - 5 Editionsprojekte, zu denen wir in TEI modellierte `listEvents` bereits vorliegen haben (CF/ÖNB: Okopenko-Tagebuch; SK/ÖAW-IHB: Ministerratsprotokolle Habsburgermonarchie, Mächtkongresse, HWK/ZIM-ACDH: Itinerar Santonino, Wirtschaftskalender, Heiligenkalender)
 - Mockups von Timelines, Kalendern, Widgets als Vorentwurf
- Matthias Schlögl, Österreichische Akademie der Wissenschaften; Projekt Austrian Prosopographical Information System (APIS):
 - Mapping von APIS-Daten zu CIDOC CRM und TEI
- Sascha Grabsch, Berlin-Brandenburgische Akademie der Wissenschaften, `correspSearch`
 - Ereignisse im Rahmen von Briefeditionen; Correspondence Metadata Interchange Format (CMIF) als Modell für Überschneidungen bei der Entwicklung von übergreifenden Infrastrukturen zur Aggregation, Speicherung und Dissemination von (`[correspAction]event`-)Daten,

Bibliographie

Anderson, C. / Eide, Ø. / Orłowska, A. / Pindl, K. / Tomasek, K. / Vogeler, G. (2016): Modeling semantically Enhanced Digital Edition of Accounts (MEDEA) for Discovery and Comparison on the Semantic Web, 2016. <https://hcommons.org/deposits/item/hc:24347>

CIDOC Conceptual Reference Model Version 6.2, Mai 2015, <http://www.cidoc-crm.org/Version/version-6.2>

Dumont, Stefan (2016): `correspSearch` – Connecting Scholarly Editions of Letters, in: Journal of the Text Encoding Initiative, Issue 10, 2016, <http://journals.openedition.org/jtei/1742>.

Dumont, Stefan et al. (2019): Correspondence Metadata Interchange Format (CMIF). In: Encoding Correspondence. A Manual for Encoding Letters and Postcards in TEI-XML and DTABf. Edited by Stefan Dumont, Susanne Haaf and Sabine Seifert. Berlin 2019. <https://encoding-correspondence.bbaw.de/v1/CMIF.html>

Grishman, Ralph (2015): Information Extraction, in: The Oxford Handbook of Computational Linguistics, hrsg. v. Ruslan Mitkov, 2015 <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199573691.001.0001/oxfordhb-9780199573691-e-009>.

Piotrowski, Michael (2012): Natural Language Processing for Historical Texts, in: Synthesis Lectures on Human Language Technologies, Bd. 5, San Rafael 2012, S. 1–157.

Mersch, Dieter (2002): Was sich zeigt. Materialität, Präsenz, Ereignis, München, Fink

Johnson, Uwe (1970ff): Jahrestage. Frankfurt: Suhrkamp.

Kritsotaki, Athina / Doerr, Martin (2006): Documenting Events in Metadata. In: The 7th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST 2006. <http://www.cidoc-crm.org/sites/default/files/Documenting%20Events%20in%20Metadata.pdf>

Van Hage et al. (2009): The Simple Event Model Ontology. <http://semanticweb.cs.vu.nl/2009/11/sem/>

Vogeler, Georg (2019): The ‘assertive edition’, in: International Journal of Digital Humanities, 1,2, 2019, S. 309–322. <http://dx.doi.org/10.17613/M6JS9H76P>

Buckland, M. / Ramos, M.R. (2010): Events as a structuring device in biographical mark-up and metadata. Bulletin of the American Society for Information Science and Technology 36, 26–29. <https://doi.org/10.1002/bult.2010.1720360209>

Fokkens, A. / ter Braake, S. / Ockeloën, N. / Vossen, P. / Legêne S. / Schreiber G. / de Boer, V., (2018): BiographyNet: Extracting Relations Between People and Events. arXiv preprint .

Peroni, S. / Tomasi, F. / Vitali, F., (2013): Reflecting on the Europeana Data Model. https://doi.org/10.1007/978-3-642-35834-0_23