

Prosopographische Interoperabilität (IPIF) Stand der Entwicklungen

Vogeler, Georg

georg.vogeler@uni-graz.at
Universität Graz, Austria

Hadden, Richard

richard.hadden@oeaw.ac.at
Österreichische Akademie der Wissenschaften, Österreich

Schlögl, Matthias

matthias.schloegl@oeaw.ac.at
Österreichische Akademie der Wissenschaften, Österreich

Vasold, Gunter

gunter.vasold@uni-graz.at
Universität Graz, Austria

Prosopographie ist die Erhebung und Auswertung von individuellen Daten über historische Personen unter den Bedingungen einer eingeschränkten Quellenlage mit dem Ziel Aussagen über die Personen als Gruppe machen zu können. Auf der DHd2019 haben wir einen Vorschlag für eine geteilte RESTful API gemacht, mit der prosopographische Daten in ihren verschiedenen Inkarnationen leichter austauschbar gemacht werden sollen (Vogeler, Schlögl, Vasold 2018). Sie wird seitdem unter dem Titel “International Prosopographical Interchange Format” auf github weiterentwickelt (<https://prosopography.org/>). Seitdem haben wir an der Modellierung und der Implementierung von Prototypen gearbeitet.

Aus Sicht der **Modellierung** will IPIF nicht die Ausdrucksmächtigkeit von RDF basierten Modellen erreichen oder sich vollständig auf Upper-Level-Ontologies wie CIDOC-CRM abbilden lassen. Stattdessen zielt der Vorschlag der API darauf, technisch einfach zu implementieren zu sein. Es haben sich deshalb einige Modellierungsprobleme ergeben, zu denen wir Entscheidungen getroffen haben, die als exemplarisch für solche vereinfachten Lösungen gelten können:

Effiziente Filter: Das konzeptuelle Modell gruppiert sich um das von Jon Bradley in die Prosopographie eingeführte “Factoid” herum (Bradley & Short 2005; Pasin & Bradley 2015), das in der “Assertion” des Genealogie-Datenmodells (GedBase4All) 2010 eine Entsprechung gefunden hat. Davon sind die Entitäten *person* (Individuum), *statement* (Aussage über das Individuum) und *source* (Quellenbeleg für die Aussage) abhängig (Abb. 1)

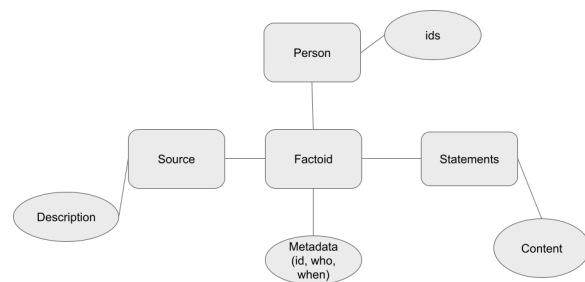


Abb. 1: Vereinfachte Darstellung des Datenmodells von IPIF

Da das Factoid aber in prosopographischer Forschung normalerweise nicht im Mittelpunkt steht, sondern Abfragen nach Personen und ihren Eigenschaften dominieren, bietet IPIF Endpunkte für genau diese Entitäten an, die leicht über ihre Beziehungen zu den anderen Entitäten gefiltert werden können: <https://example.org/ipif/v0.1/person/?name=Georg&place=Graz&from=2010&to=2015> legt für die Benutzer eine Suche nahe, die alle Personen identifiziert, über die es *statements* gibt, die ihr den Namen „Georg“ zuweisen und einen Aufenthalt in Graz zwischen 2010 und 2015. Man kann die Formulierung des Filters aber auch als Suche nach allen Personen, zu denen es je eine Aussage (*statement*) für jeden Parameter gibt, verstehen. Letztere Interpretation entspricht eher dem konzeptuellen Modell, weitet aber das Ergebnis auf schwer verständliche Art aus. IPIF verwendet deshalb die erste Interpretation als Standard, während die zweite über einen zusätzlichen Parameter `independentStatements=true` gelöst wird.

Labels für abstrakte Entitäten: Das Factoidmodell ist epistemologisch strikt und trennt das Individuum von seiner Beschreibung. So gibt es für die *person*-Identität im konzeptuellen Modell keinen menschenlesbaren Identifikator. In der Praxis ist das jedoch unrealistisch, da Listen von Personen, z.B. in *autocomplete*-Funktionen, ein menschenlesbares Label haben sollten. IPIF sieht deshalb vor, dass die jeweilige Implementation für Personenentitäten ein “*label*” erzeugt, das Informationen aus den Aussagen verwendet, die vom Standard nicht näher spezifiziert werden. Diese Label können also klassische Angaben aus Name, Lebensdaten und „Beruf“ sein, aber auch andere bei der Individualisierung helfende Daten. Sie müssen weder durchgängig konsistent mit den vorhandenen Statements sein, noch gibt die API ein Versprechen über ihre Stabilität ab.

Die derzeit umfangreichste **Implementierung** ist für das prosopographische Framework APIS in Entwicklung. APIS ist ein Entity-store, in der die aus dem Österreichischen Biographischen Lexikon (ÖBL) extrahierten Entitäten und ihre Beziehungen gespeichert werden (Schlögl & Leitowicz 2017). Das Framework ist damit ein generell für prosopographische Daten geeignetes Werkzeug und wird inzwischen in verschiedenen Projekten am ACDH-CH verwendet (APIShub: <https://apis-hub.acdh-dev.oeaw.ac.at/>). APIS bietet eine eigene RESTful API und Serialisierungen der Daten in TEI und CIDOC-CRM (<https://apis.acdh.oeaw.ac.at/apis/api2/>). Zusätzlich haben wir eine IPIF Schnittstelle für die APIS-Daten entwickelt. Die technische Lösung ist eine Solr basierte Bibliothek, die einen für APIS spezifischen Export aus APIS verwendet. Der Suchindex repräsentiert die Endpunkte als Dokumente und definiert für die von IPIF definierten Filter spezifische Indexe. Die Lösung übersetzt also nicht das konzeptuelle

elle Modell von IPIF direkt, sondern verwendete ein auf Performanz getrimmte Variante.

Einen ähnlichen Ansatz verfolgt der in Entwicklung befindliche IPIF-Python-Client, der Traversierungen zusammenfasst, um komplexere Abfragen zu vereinfachen, in denen die API in mehreren Schritten abgefragt werden müsste. Damit können Benutzungsfälle abgebildet werden, die auf Joins zwischen den drei Hauptentitäten beruhen, also z.B. Ergebnisse eines Filters auf *statements* an den *sources*-Endpunkt weitergeben.

Aus diesen Fällen lassen sich folgende Ansprüche an pragmatische RESTful APIs generalisieren, die wir mit dem Poster zur Diskussion stellen wollen:

1. Bevorzuge Filter, die einen kleinen Datenausschnitt erzeugen, denn es ist einfacher, mehrere API-Aufrufe clientseitig zu kombinieren als ein zu großes serverseitiges Ergebnis clientseitig einzuschränken.
2. Keine URI ohne Label: abstrakte Konstrukte sollten immer eine menschenlesbare, semantisch besetzte Alternative besitzen, deren mangelnde semantische Präzision (es können fehlerhafte, veraltete oder umstrittene Angaben sein) hingenommen werden muss.
3. Erlaube Implementationen, die die Definitionen der API performant umsetzen, auch wenn sie nicht explizit das von der API verwendete konzeptuelle Datenmodell realisieren.

Bibliographie

Bradley, John / Short, Harold (2005): „Texts into Databases: The Evolving Field of New-Style Prosopography“, in: *Literary and Linguistic Computing* 20/Suppl 3–24. doi:10.1093/lilc/fqi022.

Pasin, Michele / Bradley, John (2015): „Factoid-Based Prosopography and Computer Ontologies: Towards an Integrated Approach“, in: *Literary and Linguistic Computing* 30/1 86–97. doi:10.1093/lilc/fqt037.

Schlögl, Matthias / Lejtovicz, Katalin (2018): „A Prosopographical Information System (APIS)“, in: Antske Fokkens, ter Braake, Serge, Sluijter, Ronald, Arthur, Paul, and Wandl-Vogt, Eveline (eds.): *BD-2017. Biographical Data in a Digital World 2017. Proceedings of the Second Conference on Biographical Data in a Digital World 2017. Linz, Austria, November 6-7, 2017*, CEUR Workshop Series 2119, [Budapest: CEUR] 53-58 <<http://ceur-ws.org/Vol-2119/paper9.pdf>> [letzter Zugriff 10.7.2021]

Vogeler, Georg / Vasold, Gunter / Schlögl, Matthias (2019). „Von IIF zu IPIF? Ein Vorschlag für den Datenaustausch über Personen“. Patrick Sahle (Hrsg.). *DHd 2019. Digital Humanities: multimedial & multimodal. Konferenzabstracts. Universitäten zu Mainz und Frankfurt 25. bis 29. März 2019*. Frankfurt am Main. doi:10.5281/zenodo.2600812.

Zeditz, Jasper (2010): “Gedbas4all -- New Data Model for Genealogy“. In: *GenWiki* <<http://wiki-en.genealogy.net/Gedbas4all/Article>> [letzter Zugriff 10.7.2021]