

## Kollaborativ arbeiten und annotieren – Die Forschungsinfrastruktur des Spezialforschungsbereichs Deutsch in Österreich

### Seltmann, Melanie

melanie.seltmann@univie.ac.at  
Universität Wien

### Breuer, Ludwig Maximilian

ludwig.maximilian.breuer@univie.ac.at  
Universität Wien

### Heinisch, Barbara

barbara.heinisch@univie.ac.at  
Universität Wien

Der Spezialforschungsbereich (SFB) „Deutsch in Österreich. Variation – Kontakt – Perzeption“ (FWF F 60) beschäftigt sich mit der Vielfalt sowie dem Wandel der deutschen Sprache in Österreich. Dabei behandelt er den Gebrauch und die subjektive Wahrnehmung von deutscher Sprache in Österreich und zeigt Einflüsse durch Kontaktsprachen auf. Der SFB gliedert sich in verschiedene Teilprojekte an vier verschiedenen Institutionen (Universität Wien, Universität Salzburg, Universität Graz und Österreichische Akademie der Wissenschaften), die unterschiedliche Schwerpunkte in den Fokus nehmen.

Ein Ziel des SFB ist es, die Forschungsansätze und -ergebnisse einem möglichst breiten Publikum einfach und frei zugänglich zu machen. Hierfür ist eine gut durchdachte, funktionale und vor allem flexible Forschungsplattform sowie das Zurverfügungstellen (und die Nutzung) einiger Standards und Best Practices unumgänglich. Zum einen um die Arbeit über die verschiedenen Teilprojekte und Institutionen kollaborativ und einheitlich zu gestalten, zum anderen um den Open Science-Ansatz des Projektes verwirklichen zu können und Daten sowie Forschungsmethoden und -ergebnisse bereitstellen zu können.

Hierzu ist eine breite - auch technische - Expertise nötig, da die Forschungsplattform den gesamten Forschungsprozess unterstützt und begleitet. Die Forschungsplattform dient dabei als Virtual Research Environment (VRE) (Sarwar et al. 2013: 551, Smith et al. 2011:54). Sie bietet Tools wie eine Call-Center-Maske für die Gewährspersonenakquise und -befragung, Transkriptions-, Annotations-, Kommunikations-, Arbeitsmanagement-

und Datenmanagementtools – damit das ganze Spektrum des wissenschaftlichen Arbeitsprozesses von der InformantInnenakquise über die Datenauswertung bis hin zur Publikation unterstützt werden kann. Schließlich werden alle Prozesse und Daten nicht nur nachhaltig gespeichert, sondern auch weiterentwickelt und aufbereitet, damit sie WissenschaftlerInnen sowie der interessierten Öffentlichkeit zur Verfügung gestellt werden können. Eine Herausforderung stellen die Vielfalt der theoretischen Ansätze, praktischen Methoden, Datenformate, (u.a. Textkorpora, Audio- und Videodaten) und die (Meta-)Datenaufbereitung sowie die damit verbundene Sicherung und Nutzung dar.

Die Grundlage der einzelnen Module der Forschungsplattform basiert auf Open Source-Tools, die gegebenenfalls an die Bedürfnisse des SFB angepasst werden. Dazu gehören beispielsweise Eingabemasken für eine Personendatenbank samt „Call Center-Maske“, Transkriptions- sowie Annotationsumgebungen, eine Literaturverwaltung sowie Analysetools. Die Tools werden dabei in Docker-Containern (Merkel 2014) gespeichert, u.a. um die Nachhaltigkeit der Tools zu unterstützen. Dies bietet den Vorteil, dass auch abseits des Projekts und sogar der Wissenschaft an den einzelnen Komponenten weiterentwickelt werden kann, insofern sich eine entsprechende Community bildet. Zudem bieten die Docker-Container die Möglichkeit, Tools sehr einfach für andere Projekte nutzen zu können. Die Docker-Container werden im SFB in Rancher (<https://rancher.com/>) verwaltet. Dies vereinfacht die gesamte Pipeline von der Entwicklung über das Testen bis hin zur produktiven Nutzung der Tools, da alles an einem Ort geschehen kann.

Die VRE ist zum jetzigen Zeitpunkt nur projektintern verfügbar und befindet sich in kontinuierlicher Weiterentwicklung. Eine angepasste VRE soll zum späteren Zeitpunkt auch der Öffentlichkeit zur Verfügung stehen. Andere Plattformen wie WebAnno etc. haben sich nicht als für die SFB-Bedürfnisse adaptabel erwiesen, da sie den umfassenden Aufgaben sowie der Strukturierung der Daten und Annotationen nicht gerecht werden.

Auf dem Poster soll neben der Darstellung dieser Plattform genauer auf den Bereich der Annotation eingegangen werden. Da diese auf den verschiedensten linguistischen Ebenen vorgenommen und von einer Vielzahl von WissenschaftlerInnen erarbeitet wird, muss es zum einen klare Vorgaben geben, wie annotiert werden soll, zum anderen müssen die Annotationen nach einem einheitlichen Schema gebildet werden. Dennoch soll die Flexibilität der Annotationen gewährleistet sein, um wissenschaftliches Arbeiten möglichst effizient gestalten zu können. Dabei werden die Tags zwar mit Hilfe einer m:n-Verknüpfung zwischen Attributen (Tags) und Werten (Antworten) gespeichert, sind aber hierarchisch projizierbar. Somit ist es möglich, dass die verschiedenen Teilprojekte unterschiedlich tief annotieren, aber dennoch auf die Annotationen der anderen Teilprojekte zurückgreifen können. Sie fügen sich zudem

bestmöglich in die vorhandene Forschungslandschaft ein, z.B. durch Verwendung gängiger Taggingssysteme (wie Edisyn), um auch SFB-übergreifend leicht nutz-, adaptier- und vergleichbar zu sein.

## Bibliographie

**Merkel, Dirk** (2014): „Docker: Lightweight Linux Containers for Consistent Development and Deployment“, in: *Linux J.* 239 <http://www.linuxjournal.com/content/docker-lightweight-linux-containers-consistent-development-and-deployment> [letzter Zugriff: 25. September 2017].

**Sarwar, Muhammad S. / Doherty, T. / Watt, J. / Sinnott, Richard O.** (2013): „Towards a virtual research environment for language and literature researchers“, in: *Future Generation Computer Systems* 29: 549–559 <https://doi.org/10.1016/j.future.2012.03.015>.

**Smith, Vincent / Rycroft, Simon / Brake, Irina / Scott, Ben / Baker, Ed / Livermore, Laurence / Blagoderov, Vladimir / Roberts, David** (2011): „Scratchpads 2.0: a Virtual Research Environment supporting scholarly collaboration, communication and data publication in biodiversity science“, in: *ZooKeys* 150: 53–70.