
RePlay-DH: Ein Werkzeug für Wissenschaftler, um Wissen zu erhalten und zu teilen

Sibylle Hermann and Markus Gärtner

Universität Stuttgart, Deutschland

Der im Projekt RePlay-DH entwickelte Client dient zur Unterstützung der einheitlichen Dokumentation der Arbeitsergebnisse im wissenschaftlichen Workflow. Der Client basiert auf der Versionsverwaltungssoftware git.

Forschungsdaten werden oft nicht veröffentlicht, da sie am Ende eines Projektes nur unstrukturiert vorliegen. Deshalb ist der Ansatz im Projekt RePlay-DH, von Beginn an eine lückenlose Dokumentation zu gewährleisten und somit eine mögliche Datenpublikation zu vereinfachen. Darüber hinaus wird so sichergestellt, dass die Daten auch nach einiger Zeit noch verstanden und nachvollzogen werden können. Der mit einer Prozess-Dokumentation einhergehende Aufwand neben der eigentlichen Arbeit wird von Forschenden häufig als einer der hemmenden Faktoren angegeben, wenn es um das Dokumentieren von Forschungsarbeit geht. Dieser Zusatzaufwand kann hierbei manigfaltige Formen annehmen:

die Einarbeitung in technisch anspruchsvolle Systeme, aufwändige manuelle Erstellung von Metadaten zur Dokumentation, oder selbst die komplette Umstrukturierung der eigenen Arbeitsprozesse. Für eine Akzeptanz von Seiten der Forschungscommunity ist es daher unabdingbar, dass sich angebotene Lösungen möglichst nicht-invasiv und ohne großen Mehraufwand, in etablierte Arbeitsabläufe integrieren lassen. Der von RePlay-DH entwickelte Client bietet hier einen auf die Bedürfnisse der Wissenschaftlerinnen und Wissenschaftler angepassten Kompromiss hinsichtlich einfacher Nutzbarkeit und den technischen Anforderungen an die Anwenderinnen und Anwender.

Das Werkzeug kommt ohne komplizierte Einrichtung aus und erfordert praktisch keine Umstellung von bestehenden Abläufen. Ideal ist die Verwendung für Prozesse, bei denen Daten lokal anfallen oder manipuliert werden. So verhindert die automatische Erkennung von Änderungen an lokalen Dateien im Arbeitsbereich, dass die Dokumentation potentiell wichtiger aber leicht zu übersehender Details ausbleibt.

Die Funktion des Clients, Metadaten zur Beschreibung eines Arbeitsschrittes inkrementell „on the fly“ während dessen Ausführung aufnehmen zu können, ermöglicht eine einfachere Integration in den Forschungsalltag, da der anfallende Zusatzaufwand zeitlich frei verteilt werden kann. Dazu wurde gemeinsam mit Wissenschaftlerinnen und Wissenschaftlern aus der Computer-Linguistik ein Metadatenschema entwickelt, das die Arbeitsschritte in deren Arbeitsalltag möglichst genau und mit geringem Aufwand abbildet. Es besteht die Möglichkeit, pro Workspace ein anderes auf die Bedürfnisse der Wissenschaftsdisziplin angepasstes Metadatenschema zu implementieren.

Im Client ist ein Ressourcenverwaltungsprogramm integriert, das ermöglicht, Ressourcen wie Textkorpora und Software einmalig mit Metadaten im Client zu dokumentieren, um somit immer wieder auf diese verweisen zu können. Des Weiteren besteht die Möglichkeit, Teil- und Zwischenergebnisse im institutionellen Repository zu veröffentlichen oder in einer git-Instanz zu archivieren. Der Replay-Client ist im Projektkontext auf die Bedürfnisse der Computerlinguistik zugeschnitten worden, aber auch andere Textbasierte Wissenschaftsdisziplinen können das Werkzeug verwenden. Der Replay-Client bietet dazu die Möglichkeit eigene Metadaten anzulegen. Des Weiteren können Anpassungen und Weiterentwicklungen des Replay-Clients über die öffentliche git-Instanz (<https://github.com/RePlay-DH/replay-dh-client>) vorgenommen werden.