

# Baustelle digitale Editions- umgebung: ediarum und Git-basierte Versionskontrolle

## Fechner, Martin

fechner@bbaw.de

Berlin-Brandenburgische Akademie der Wissenschaften,  
Deutschland

ORCID: 0000-0001-9233-0653

## Klappenbach, Lou

klappenbach@bbaw.de

Berlin-Brandenburgische Akademie der Wissenschaften,  
Deutschland

ORCID: 0000-0001-6587-3353

## Lampert, Marcus

marcus.lampert@bbaw.de

Berlin-Brandenburgische Akademie der Wissenschaften,  
Deutschland

Digitale Editionen befinden sich kontinuierlich im Wandel. Diese Eigenschaft ermöglicht sukzessive Veröffentlichungen und eine transparente Gestaltung des Forschungsprozesses sowohl innerhalb der Editionsteams als auch nach außen hin. Hierfür ist eine gründliche Dokumentation und Versionskontrolle notwendig. Die AG Digitales Publizieren betont die Bedeutung der Versionierung digitaler Texte für die Zitierbarkeit und Nachvollziehbarkeit (AG Digitales Publizieren 2021, 68). Einige digitale Editionen, wie die edition humboldt digital, setzen bereits auf regelmäßige Veröffentlichungen ihrer Forschungsdaten auf Plattformen wie Github und Zenodo, was den Nutzer:innen ermöglicht, Änderungen nachzuvollziehen und Datensätze zu vergleichen (Ette u. a. 2022). Aber nicht nur auf Ebene der Publikation, sondern auch innerhalb des Arbeitsprozesses und bei der alltäglichen Editionsarbeit sollte eine Versionierung der Forschungsdaten gewährleistet werden.

Die digitale Editions-umgebung der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW), genannt ediarum<sup>1</sup>, setzt zu diesem Zweck auf eine lückenlose Versionierung von TEI/XML-Dateien mittels Git-basierter Versionskontrolle. TELOTA, die Digital-Abteilung der BBAW, entwickelt hierfür neue Module, deren Hauptziel es ist Git-basiert zu arbeiten. Git ermöglicht es den Editor:innen einerseits, den Forschungsprozess detailliert zu verfolgen, unterschiedliche Datenstände zu vergleichen und die Historie einzelner Dateien Schritt für Schritt nachzuvollziehen (vgl. Bürgermeister 2020). Andererseits kön-

nen dann ausgewählte Datenstände kombiniert und auf einem öffentlichen Git-Branch publiziert und dokumentiert werden. Die beiden Strategien, d.h. sowohl die detaillierte Datensicherung durch kleine automatisierte Commits, als auch die Publikation von erreichten Meilensteinen innerhalb der detaillierten Versionierung, fördern die Transparenz und die gute wissenschaftliche Praxis (vgl. Deutsche Forschungsgemeinschaft 2022). Versionskontrolle verbessert die Nachhaltigkeit von Editions- und Softwareprojekten (vgl. Schrade 2017, 2–3) und sollte auch in die Datenpraktiken der Projekte Einzug erhalten. Die Integration von Versionierungssystemen wie Git in (bestehende) Publikations- und Editions-umgebungen wird aktuell innerhalb der Digital Humanities intensiv diskutiert<sup>2</sup>. ediarum reagiert darauf mit verschiedenen Entwicklungen, die auf diesem Poster vorgestellt und deren Anwendungsfälle beschrieben werden. Die neuen Lösungen unterstützen die Integration von Git in bestehende Systeme digitaler Editionen und bieten Nutzer:innen von ediarum und exist-db neue Optionen.

ediarum.WEBDAV bietet eine Lösung, die XML-Dateien während der Eingabe im Oxygen XML Editor bei jedem Speichern in ein Git-Repository auf den Server überträgt und parallel dazu in eine bestehende eXist-db synchronisiert. Für die Kommunikation zwischen dem Oxygen Editor und dem Server ist eine Webdav-Schnittstelle zuständig. Die Git-History bleibt linear und ohne Konflikte, weil nur der Server die Änderungen seriell, d.h. eine nach der anderen, committet und dies nur auf dem main-Branch erfolgt. Die WebDAV Dateisperre sorgt dafür, dass die Nutzer:innen die selbe Datei nicht gleichzeitig bearbeiten und dadurch etwa unbeabsichtigt Änderungen gegenseitig überschrieben können. Besondere Eingriffe, etwa Datenänderungen per Skript, können auf eigenen Branches durchgeführt und anschließend von den Entwickler:innen in den main-Branch überführt werden.

Die direkte Bearbeitung von TEI/XML-Dateien in einem Git-Repository erhöht die Datensicherheit und vermeidet Datenverluste, die beispielsweise bei einem Absturz einer XML-Datenbank oder dem Entfernen semantisch relevanter Leerzeichen auftreten können. Diese Integration schafft Vertrauen bei den Editor:innen der digitalen Editionen. Eine Git-basierte Architektur bietet zudem Unabhängigkeit von spezifischen Technologien und Datenbanken und erleichtert die Migration in andere Systeme. Da einfache TEI/XML-Dateien von verschiedenen Programmiersprachen und Web-Frameworks weiterverarbeitet werden können, ist eine flexible und vielseitige Nutzung gewährleistet.

ediarum.BACKEND stellt die in Git-versionierten Daten performant über eine REST-API zur Verfügung und bietet somit eine Alternative zu gängigen XML-Datenbanken. Über eine Konfigurationsdatei (einfaches XML) wird mittels XPath definiert, welche Felder in bestimmten TEI/XML-Dateien ausgelesen werden sollen. Diese kann von den Nutzer:innen selbst verwaltet werden. Per Skript werden die ausgewählten Informationen in eine SQL-Datenbank überführt und können so sehr performant über die Rest-API von ediarum.BACKEND abgefragt werden. An-

schließlich können sie von einem Webfrontend und der oXygen-Restschnittstelle, die bei der Dateneingabe über ediarum.BASE.edit benutzt wird, weiterverarbeitet werden.

Das Poster visualisiert die Integration von Git in den Prozess der Erstellung und Bearbeitung von TEI/XML-Daten digitaler Editionen bei TELOTA und stellt die neuesten ediarum-Erweiterungen vor. Dies soll helfen Editor:innen und Entwickler:innen auf die Notwendigkeit der Versionskontrolle aufmerksam zu machen und mit ihnen über mögliche Lösungswege zu diskutieren.

## Fußnoten

1. ediarum bietet eine auf exist-db und oXygen XML-Editor basierende Infrastruktur für die Erstellung und Publikation digitaler Editionen. Das ediarum-Universum bietet bereits viele Funktionen und wird ständig erweitert, um neuen Anforderungen aus der Forschung und der fortschreitenden Datafication gerecht zu werden.
2. Über diese Thematik wurde beispielsweise beim letzten ediarum.MEETUP am 10. Juni 2024 gesprochen, siehe [https://www.ediarum.org/meetups.html#meetup\\_06\\_2024](https://www.ediarum.org/meetups.html#meetup_06_2024).

## Bibliographie

**AG Digitales Publizieren.** 2021. „Digitales Publizieren in den Geisteswissenschaften: Begriffe, Standards, Empfehlungen“. 10.17175/WP\_2021\_001\_V2.

**Bürgermeister, Martina.** 2020. „Extending Versioning in Collaborative Research“. In *Versioning Cultural Objects Digital Approaches*, Schriften des Instituts für Dokumentologie und Editorik, hg. Roman Bleier und Sean M. Winslow, 171–90. Norderstedt: BoD.

**Deutsche Forschungsgemeinschaft.** 2022. „Guidelines for Safeguarding Good Research Practice. Code of Conduct“. 10.5281/zenodo.6472827.

**Ette, Ottmar, Stefan Dumont, Annika Geiser, Carmen Götz, Tobias Kraft, Ulrike Leitner, Ulrich Päßler, Florian Schnee, und Christian Thomas.** 2022. „TEI-XML-Datenset der Tagebücher, Briefe, Dokumente, Forschungsbeiträge, Chronologieeinträge und Register der edition humboldt digital“. 10.5281/zenodo.10457932.

**Schrade, Torsten.** 2017. „Nachhaltige Softwareentwicklung in den Digital Humanities. Konzepte und Methoden.“ Vortrag bei der DHd 2017. <https://digicademy.github.io/2017-dhd-sustainable-software/> (zugegriffen: 23. Juli 2024).