



# Projekt: Editionen mit Zukunft

Das kooperativ entwickelte Werkzeug TEI Publisher (Version 8)

Laufzeit 2021

Stand: 24. Februar 2021

## Zusammenfassung

### **Ziele**

Ziel des Projekts «Wissenschaftliche Editionen mit Zukunft. Das kooperativ entwickelte Werkzeug TEI Publisher (Version 8)» ist die Weiterentwicklung des freien und kostenlosen Tools TEI Publisher. Der TEI Publisher ist ein nachhaltiges Publikationstool für digitale wissenschaftliche Editionen, das sich konsequent an internationalen Standards orientiert und von namhaften schweizerischen Editionsprojekten verwendet wird. Der TEI Publisher leistet einen wichtigen Beitrag zur Vermittlung und Erhaltung des Kulturguts. Im Mittelpunkt der Weiterentwicklung steht die Stärkung der Nachhaltigkeit, der Zugänglichkeit und der Vernetzung der Editionen.

### **Projektteam**

Der Vorstand des Vereins e-editiones leitet das Projekt und koordiniert die Zusammenarbeit der beteiligten Editionsprojekte mit den Softwareentwickler\*innen.

### **Projektdauer**

Die Arbeiten sollen am 1.1.2021 starten und am 31.12.2021 abgeschlossen sein.

### **Status**

Mit der Zusage der Ernst Göhner Stiftung ist rund die Hälfte der Arbeitspakete finanziert bzw. durch Projektbeiträge abgedeckt. Der Vorstand wird in Rücksprache mit den beteiligten Projekten die Priorisierung vornehmen.

## Ziele und Zielgruppen

### Ziele

Das Hauptziel von *e-editiones* ist es, nachhaltige, langfristig verfügbare digitale Editionen zu ermöglichen. Mit der Version 8 soll der TEI Publisher insbesondere in folgenden Bereichen stark verbessert werden:

#### Nachhaltigkeit

- Editorische Workflows
- Langzeitarchivierung
- Wartbarkeit

Bislang konzentriert sich TEI Publisher vor allem auf die Publikation einer Edition, deren Datengrundlage bereits in erschlossener Form vorliegt. Der editorische Arbeitsprozess erfordert technische Kenntnisse, die bei den Editor\*innen nicht vorausgesetzt werden können. Er umfasst den Weg vom Faksimile über die Erstellung einer Transkription, deren semantische und analytische Anreicherung sowie die Arbeitsteilung zwischen mehreren Editor\*innen. Ziel ist, zentrale Arbeitsschritte dieses Prozesses so zu erleichtern, dass sie von Editor\*innen ohne technische Kenntnisse und in kürzerer Zeit durchgeführt werden können.

Mit dem TEI Publisher 7 (aktuelle Entwicklung) wird der technische Aufwand für den Unterhalt der Editionen auf dem Server massiv gesenkt. TEI Publisher 8 soll die inhaltliche / editorische Pflege zusätzlich verbessern. Gleichzeitig sollen Editionen ihre Daten (TEI/XML) vereinfacht langzeitarchivieren können.

#### Zugänglichkeit

- Barrierefreiheit
- Benutzerfreundlichkeit
- Vermittlung

Während TEI Publisher viele Aspekte der Richtlinien zur Barrierefreiheit erfüllt, werden einige zentrale Punkte nicht ausreichend berücksichtigt: so ist z.B. die Navigation über Tastatur bzw. Screenreader uneinheitlich und kompliziert. Mit dem TEI Publisher 8 wird die Userexperience konsistenter werden. Bisher ungelöste Einzelaspekte bei der Darstellung und Vermittlung von Daten werden gelöst.

#### Vernetzung

- Austausch zwischen Editionsprojekten
- Vernetzung von Editionen

Digitale Editionen sind heute Editionen im Internet. Für die Arbeit mit den Editionen sowie für die Forschung ist es wichtig, dass eine Vernetzung der Editionen untereinander und mit weiteren Projekten möglich ist und dass «Datensilos» vermieden werden. TEI Publisher 8 wird gezielt so ausgebaut, dass serverübergreifende Suchen in den einzelnen Editionen möglich sind. Darüber hinaus wird die gegenseitige Integration und Verlinkung von Texten bzw. Textfragmenten unterstützt.

## **Stakeholder und Zielgruppen**

- Leser\*innen
- Wissenschaftler\*innen, die digitale Editionen für ihre Forschung verwenden (Geschichtswissenschaft, Philologie, Philosophie, Theologie, Jura etc.)
- Editionsprojekte/Editor\*innen, die mit TEI arbeiten
- Gedächtnisinstitutionen, die für die Verfügbarkeit von Editionen verantwortlich sind
- Entwickler\*innen, die neue Funktionen implementieren
- Administrator\*innen, die für den technischen Betrieb von digitalen Editionen zuständig sind

## Projektbeschreibung

Mit dem aktuellen Projekt soll eine neue Version des TEI Publishers, Version 8, der ein zentrales Instrument für die beteiligten Editionen ist, entwickelt werden. Dafür hat der Vorstand von *e-editiones* die verschiedenen Desiderate aus den Editionsprojekten und der Forschungscommunity zusammengetragen, diskutiert und gewichtet. Mit dem TEI Publisher 8 kann ein wesentlicher Betrag zur Realisierung der oben beschriebenen Ziele geleistet werden.

### AP 1: Integration Transkribus

Ziele: Nachhaltigkeit, Optimierung editorischer Workflow

Beschreibung: Viele Editionsprojekte verwenden Transkribus<sup>1</sup>, um eine Rohtranskription auf Basis von Faksimiles zu erstellen. Die weitere Verarbeitung ist für technisch unerfahrene Forschende jedoch oft schwierig zu handhaben. Ziel ist deshalb, eine Schnittstelle zu Transkribus direkt in die Editionsoberfläche zu integrieren. Die in Transkribus erarbeitete Rohtranskription kann somit direkt ausgewählt und ohne Umweg für die weitere Bearbeitung in den Editionsbestand geladen werden. Zudem lässt die Qualität der von Transkribus exportierten TEI Daten zu wünschen übrig: bestimmte Informationen fehlen im Export und müssen bislang manuell nachgepflegt werden. Durch eine direktere Integration in TEI Publisher lässt sich dieses Problem lösen.

- Über die von Transkribus angebotene Serverschnittstelle (API) erhält TEI Publisher Zugriff auf alle mit Transkribus erstellten Daten und kann diese verlustfrei nach TEI transformieren.
- Innerhalb der Nutzeroberfläche von TEI Publisher wird es eine Möglichkeit geben, die in Transkribus angelegten Sammlungen zu browsen und bereits transkribierte, aber nicht weiter verarbeitete Dokumente auszuwählen.
- Benutzer\*innen können einen Export der ausgewählten Dokumente in Transkribus anstoßen.
- Das Ergebnis des Exports wird direkt in TEI Publisher geladen und kann in der von der Edition festgelegten Endansicht kontrolliert bzw. überarbeitet werden.

### AP 2: Unterstützung von Annotationen

Ziele: Nachhaltigkeit, Optimierung editorischer Workflow

Beschreibung: Ist einmal eine Rohtranskription erstellt, stehen i.d.R. weitere Bearbeitungsschritte an, die mitunter zwar aufwändig sind, theoretisch aber auch von nicht-technisch geschulten Bearbeitenden (auch Hilfskräften) durchgeführt werden können. Dazu gehört z.B. die Auszeichnung von Entitäten (wie z.B. Personen, Orten), Registerbegriffen, Datierungen oder Korrekturen. Um diese Arbeiten zu erleichtern, soll die Editionsoberfläche um entsprechende Annotationsfunktionen ergänzt werden. Direkt beim Lesen können die Bearbeiter\*innen Textsequenzen auswählen und annotieren. Ein Wechsel in die XML-Quellen entfällt damit. Voraussetzung ist allerdings ein entsprechender Workflow mit Rechtesystem, damit die Editor\*innen und die Projektleitenden die Arbeiten jederzeit kontrollieren können.

Somit werden mehrere Ziele erreicht:

- Bestimmte Arbeitsschritte können auch durch technisch nicht-geschulte Forschende vorgenommen werden.

---

<sup>1</sup> <https://transkribus.eu/Transkribus>

- Der Zeitaufwand für Korrektur und Anreicherung einer Transkription wird erheblich reduziert.
- Mehrere Forschende können ein Dokument gleichzeitig bearbeiten.
- Die breitere Öffentlichkeit kann über das Web mit eingebunden werden.
- Der Forschungsdiskurs wird an Ort und Stelle nachvollziehbar.

Technisch erfordert dies, dass alle Annotationen zunächst separat vom Haupttext gespeichert werden, bis ein Bearbeitender diese jeweils überprüft und freigegeben hat. Alle Annotationen beziehen sich daher immer auf einen Base-Text, der bis zum Ende des Bearbeitungsschritts eingefroren bleibt. In einem zweiten Schritt werden die Annotationsvorschläge evaluiert und entweder angenommen oder verworfen. Im dritten Schritt wird der jeweils erreichte Stand auf Knopfdruck veröffentlicht und konstituiert damit einen neuen Base-Text, der wiederum erneut einer weiteren Iteration unterzogen werden kann.

Die Speicherung der Annotationen geschieht in einem Format kompatibel zum W3C Web Annotations Standard. Damit sind die Annotationen auch mit anderen Systemen austauschbar. Zudem liegen gewisse Funktionskomponenten bereits in anderen Werkzeugen vor und können nachgenutzt werden. Eine kompatible Implementierung des Web Annotation Standards wurde für ein Projekt schon früher erstellt, so dass wir uns vorrangig auf die Integration in die Nutzeroberfläche konzentrieren können.

Wegen der Komplexität des Vorhabens schlagen wir ein mehrstufiges Vorgehen vor: Ziel der Basisversion ist, zunächst die Erfassung der einfacheren Elemente des TEI wie Personen- oder Ortsnamen, Registerbegriffe, Abkürzungen, Datumsangaben oder Korrekturen. Diese Liste deckt bereits einen grossen Teil der in einem Projekt üblicherweise anfallenden Auszeichnungsschritte ab, ist aber insofern einfacher zu handhaben, als keine strukturellen Änderungen anfallen. Damit bleibt das Vorhaben beherrschbar und wir können eine stabile Funktionalität garantieren. In einer späteren Ausbaustufe können diese Basiselemente um komplexere, z.B. textkritische Anmerkungen erweitert werden.

### **AP 3: Versionierung/Langzeitarchivierung**

Ziele: Nachhaltigkeit, Optimierung editorischer Workflow, Langzeitarchivierung

Beschreibung: Eine einfachere Versionierung der Textdaten einer Edition mit Git<sup>2</sup> ist das Ziel dieses Arbeitspakets. Der gängige Weg über die Kommandozeile stellt für Editor\*innen oftmals eine grosse Hürde dar oder kann gar nicht geleistet werden. Eine direkte Integration des entsprechenden Workflows in die Editions Oberfläche ermöglicht eine unkomplizierte und damit auch häufigere und sichere Versionierung. Gleichzeitig sollen publizierte Texte automatisiert in Zenodo<sup>3</sup> langzeitarchiviert werden.

### **AP 4: Persistente und sprechende URLs**

Ziele: Nachhaltigkeit, Benutzerfreundlichkeit, Vernetzung

---

<sup>2</sup> <https://git-scm.com> Git ist ein anspruchsvolles Programm, um Versionen zu verwalten. Es ist der de facto Standard in der Softwareentwicklung und kann sehr gut auch für die Verwaltung der Rohdaten der Editionen verwendet werden.

<sup>3</sup> <https://zenodo.org> Zenodo ist ein Langzeitarchiv für Forschungsdaten, das von der Europäischen Union finanziert und vom CERN betrieben wird. Standardisierte XML-Daten wie das verwendete TEI sind ideal für die langfristige Archivierung. Als einfache Textdaten sind sie jederzeit leicht lesbar, durch die Dokumentation und die Standardisierung sind die Daten auch inhaltlich langfristig interpretierbar.

Beschreibung: Sprechende und persistente URLs für Dokumente werden zur Verfügung stehen, um die Referenzierbarkeit der Texte zu verbessern und die langfristige Zitierbarkeit zu garantieren. Technisch muss hierzu die Beziehung zwischen einer URL und der entsprechenden physischen Ressource auf dem Server weiter entkoppelt werden. Diese Beziehung sollte insoweit frei konfigurierbar sein, als die je besondere Struktur einer Edition in sprechenden URLs abgebildet werden muss. Gleichzeitig wird jedoch ein standardisierter Rahmen vorgegeben, der editionsübergreifend stets gleichbleibt, um den Nutzer\*innen die Orientierung zu erleichtern.

Zudem sollte es innerhalb von TEI Publisher eine zentrale Registratur für URL-Pfade, deren Auflösung und die Verarbeitung von URL-Parameter geben. Nur so kann garantiert werden, dass von Nutzer\*Innen gemerkte URLs auch immer auf die beabsichtigte Seite verweisen.

Eine solche zentrale Registratur ist zudem Voraussetzung für eine verbesserte Barrierefreiheit.

#### **AP 5: Barrierefreiheit und Usability optimieren**

Ziele: Barrierefreiheit, Benutzerfreundlichkeit

Beschreibung: Eine standardisierte Navigation zwischen den zentralen Funktionen einer Edition wird realisiert: z.B. Vor-/Zurückblättern in einem Dokument, Zurück zur Suche / Einstiegsseite, Grundfunktionen für die Recherche. Die Richtlinien zur Barrierefreiheit geben vor, dass eine sinnvolle Navigation durch die Seite allein über Tastatureingaben möglich sein sollte.

Die dynamische Anpassung des Layouts für Benutzer\*innen mit Seh- oder Farbschwäche soll umgesetzt werden. Dies kann zum einen automatisch geschehen – abhängig von z.B. der gewählten Schriftgröße, zum anderen durch auswählbare Layout-/Darstellungsalternativen, die dem Benutzenden angeboten werden können (grosse Schrift, hoher Kontrast, verschiedene Farbschemata etc.). Funktionalitäten wie z.B. das Ein- und Ausschalten von farbigen Tags hilft allen, die Texte ohne Ablenkung durch Farben lesen möchten.

Zudem benötigen Screenreader für Blinde eine sorgfältige Auszeichnung aller relevanten Elemente mit semantischen Hinweisen nach den Accessible Rich Internet Applications (ARIA)<sup>4</sup> Richtlinien. Alle Komponenten von TEI Publisher sind entsprechend zu überarbeiten oder zu ergänzen.

#### **AP 6: Personalisierte Sammlungen**

Ziele: Benutzerfreundlichkeit

Beschreibung: Forschenden wird die Möglichkeit geboten, ausgewählte Texte oder Bilder nach eigenen Kriterien zusammenzustellen und in eine personalisierte Sammlung zu speichern. Auf diese Weise lassen sich einfach z.B. Reader für Lehrveranstaltungen erstellen. Hierfür ist der personalisierte Zugang Voraussetzung, der in AP 2 umgesetzt wird.

#### **AP 7: Progressive Web App**

Ziele: Nachhaltigkeit, Benutzerfreundlichkeit

---

<sup>4</sup> <https://www.w3.org/TR/wai-aria/>

Beschreibung: Mit der Umsetzung des TEI Publishers als «Progressive Web App» können die personalisierten Sammlungen und Texte auch offline genutzt werden.

Eine Webanwendung kann dem Browser mitteilen, welche Bestandteile vorab geladen und für eine Offline-Benutzung zur Verfügung gestellt werden sollen. Dies betrifft zunächst die für die grundlegende Funktionalität erforderlichen Programmteile, kann sich aber auch auf die Textdaten selbst erstrecken.

Dafür muss die Webanwendung verschiedene Standards erfüllen, die jedoch auch für AP 4 und 5 benötigt werden. Sind diese beiden Arbeitspakete einmal implementiert, lässt sich auch eine Offline-Nutzbarkeit relativ leicht erreichen.

#### **AP 8: Zeitleiste**

Ziele: Vermittlung, Benutzerfreundlichkeit

Beschreibung: Dieses Arbeitspaket ermöglicht einem Editionsprojekt die Visualisierung der Texte nach Zeiträumen mit Hilfe einer Zeitleiste zum intuitiven Suchen und Darstellen von Inhalten. Eine Zeitleiste veranschaulicht auf einen Blick, wo Texte zeitlich verortet sind und welche Zeiträume wie stark vertreten sind. Zudem erleichtert sie eine Suche nach Datierungen. Hierzu soll eine weitere Komponente erstellt werden, die von Editionen bei Bedarf in die Suche bzw. das Browsing von Dokumenten integriert werden kann.

#### **AP 9: Darstellung mathematischer Formeln**

Ziele: Vermittlung, Benutzerfreundlichkeit

Beschreibung: Es werden (auch komplexe) mathematische Formeln im TEI Publisher in den Texten darstellbar. Für die Anzeige mathematischer Formeln im Browser gibt es gängige Funktionsbibliotheken. Diese einzubinden erfordert jedoch Programmierkenntnisse, die von den Editionen nicht erwartet werden können. Es soll daher auf Basis vorhandener Code-Bibliotheken eine generische, einfach zu verwendende Komponente erstellt werden, welche die Ausgabe mathematischer Formeln ohne zusätzliche Programmierung ermöglicht.

#### **AP 10: Distributed Text Services (DTS)**

Ziele: Vernetzung von Editionen

Beschreibung: Die Umsetzung des Distributed Text Services (DTS)<sup>5</sup> Standards erlaubt es, Textsammlungen, Texte und Textfragmente miteinander zu verlinken, auszutauschen und weiterzuverwenden (z.B. um ein Zitat in einem Popup-Fenster einzublenden).<sup>6</sup> Ziel ist auch eine echte, verteilte Suche über eigene und Fremdbestände sowie die Anzeige externer Dokumente, ohne diese in den lokalen Bestand aufnehmen zu müssen. Dies wird durch das DTS Protokoll derzeit noch nicht abgedeckt und sollte in enger Abstimmung mit der DTS Community entwickelt werden.

---

<sup>5</sup> <https://distributed-text-services.github.io/specifications/>

<sup>6</sup> Die Umsetzung des DTS-Standards stellt sicher, dass die Editionen auch FAIR sind. Die FAIR Daten Prinzipien haben zum Ziel, dass Datensätze auffindbar ("findable"), zugänglich ("accessible"), kompatibel ("interoperable") und wiederverwendbar ("re-usable") sind. Das sind zugleich die Voraussetzungen für nachhaltig nachnutzbare Forschungsdaten.

Der TEI Publisher implementiert bereits einen Teil der DTS Spezifikation und umfasst derzeit die folgenden Features:

- Browsing in einer Sammlung mit Anzeige der wichtigsten Metadaten
- Auswahl einzelner Dokumente
- Import eines Dokuments in den lokalen TEI Publisher und Anzeige gemäss den vom Server ebenfalls übermittelten Formatvorlagen (ODD-Files)

Zudem stellt jede mit dem TEI Publisher erzeugte Edition ihrerseits eine DTS Schnittstelle zur Verfügung, die lediglich aktiviert werden muss. Damit kann prinzipiell jede Edition auf Basis von TEI Publisher mit jeder anderen Edition kommunizieren und Daten austauschen.

Somit ermöglicht die DTS-Unterstützung in TEI Publisher derzeit vor allem eine Exploration anderer Datenbestände und die Integration ganzer Dokumente in den eigenen Bestand, um diese weiter zu verwenden.

Die bestehenden Features sind wie folgt zu erweitern, um eine breitere Nutzbarkeit zu gewährleisten:

- Unterstützung für die Anzeige von Textfragmenten statt ganzer Dokumente
- Direkte Verlinkung aus einem Editionsbestand auf Fragmente eines anderen Bestandes innerhalb der Leseansichten
- Anzeige von Fragmenten eines fremden Bestandes, z.B. in Fussnoten oder Popups
- Eine echte, verteilte Suche über Eigen- und Fremdbestände sowie die Anzeige externer Dokumente, ohne diese in den lokalen Bestand aufnehmen zu müssen. Dies wird durch das DTS Protokoll derzeit noch nicht abgedeckt und wird in enger Abstimmung mit der DTS Community entwickelt werden.

## Projektteam und beteiligte Projekte

### **Projektleitung und Koordination (Vorstand von e-editiones)**

- Dr. Andreas Kränzle, Karl Barth-Archiv, Basel
- Dipl. Soz. Wolfgang Meier, eXist Solutions GmbH, Lenzkirch/Berlin
- Dr. Roberta Padlina, Koordinatorin NIE-INE, Universität Basel
- Dr. Pascale Sutter, Rechtsquellenstiftung des Schweizerischen Juristenvereins, Universität St. Gallen
- Lic. phil. Arman Weidenmann, Stadtarchiv der Ortsbürgergemeinde der Stadt St. Gallen
- Dr. Joseph Wicentowski, Historian, Office of the Historian, U.S. Department of State

### **Beteiligte Projekte/Institutionen**

- Karl Barth-Stiftung / Karl Barth-Gesamtausgabe / Karl Barth-Archiv
- Rechtsquellenstiftung des Schweizerischen Juristenvereins
- Basler Edition der Bernoulli-Briefwechsel
- Office of the Historian, U.S. Department of State

### **TEI Publisher (Kern-Entwickler)**

- eXist Solutions GmbH, Lenzkirch/Berlin