# **Tools & Experts**

## **ediarum: Arndt, Nadine**

### ***Tool***

Die Benutzung digitaler Mittel für die Erstellung und Publikation von wissenschaftlichen Editionen nimmt im Arbeitsprozess einen großen Stellenwert ein, der nicht mehr wegzudenken ist. Neben der fachwissenschaftlichen Forschung stehen Editor\*innen heutzutage daher vor der Herausforderung, sich auch mit digitalen Werkzeugen und Methoden auseinanderzusetzen. Aufgrund der Vielzahl an Tools und Methoden ist dies aber meistens nicht allein zu bewältigen.

Die digitale Arbeits- und Publikationsumgebung ediarum wird seit 2012 von TELOTA, der Digital Humanities Initiative an der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW), (weiter)entwickelt und bereits in mehr als 20 Projekten zur Erstellung digitaler Editionen eingesetzt. Die aus mehreren Modulen bestehende Softwarelösung unterstützt Wissenschaftler:innen mittels einer grafischen Arbeitsoberfläche dabei, Transkriptionen von Manuskripten, Dokumenten und Drucken in TEI-konformem XML anzufertigen, diese mit einem Text- und Sachapparat zu versehen, ergänzende Personen-, Orts- und weitere Register zu erstellen, die Ressourcen untereinander zu verknüpfen und anschließend als digitale Edition im Web zu veröffentlichen.

ediarum setzt auf bereits existierende Software von Dritten und ergänzt diese um eigene Komponenten. Die Entwicklungsarbeit fokussiert sich so auf spezifische Bedürfnisse und Anforderungen der editionswissenschaftlichen Community.

### ***Expert***

Nadine Arndt ist wissenschaftliche Mitarbeiterin bei TELOTA an der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) und für das Kooperationsprojekt „Redaktions-/Onlinesystem für Online-Editionen des Bundesarchivs“ zuständig. Sie ist Teil der ediarum-Koordination und in dieser Funktion an der Weiterentwicklung beteiligt. Zudem koordiniert sie das ediarum.MEETUP. Ihre Forschungsschwerpunkte sind digitale Editionen, Datenmodellierung und die stetige Verbesserung der Usability von eingesetzten Werkzeugen.

Tool details: <https://www.ediarum.org/>

## **Transkribus: Matthias Sorg & Sebastian Colutto**

### ***Tool***

Transkribus ist eine Software, die automatische Handschriftenerkennung durch künstliche Intelligenz möglich macht. Einzigartig ist dabei die Trainierbarkeit der KI-Modelle durch die NutzerInnen selbst, die zahlreichen Möglichkeiten, die erkannten Texte mit verschiedenen Tools anzureichern und die Einbindung von Transkribus in ein rundes Ökosystem. Mit diesem kann der gesamte Digitalisierungsprozess vom Scannen über die Texterkennung bis hin zur nahtlos eingebundenen Online-Veröffentlichung abgebildet werden. All das geschieht im Sinne der Kooperation, da alle Produkte und Lösungen dieses Ökosystems von der europäischen Genossenschaft READ-COOP SCE zur Verfügung gestellt werden - alle Einnahmen fließen somit in die Weiterentwicklung der Transkribus-Plattform zurück, da keine Gewinne ausgeschüttet werden dürfen. Matthias Sorg und Sebastian Colutto zeigen in ihrer Präsentation die Möglichkeiten des Transkribus-Ökosystems und sprechen über die folgenden Technologien und Lösungen:

* Nutzung künstlicher Intelligenz in Texterkennung und Layoutanalyse
* Verbesserte Suche durch Smart Search
* Ermöglichung von Citizen-Science-Projekten durch spezielle Webanwendungen (KATI)
* Technische Einbindung von Transkribus mithilfe der metagrapho-API

### ***Expert (Matthias Sorg)***

Matthias Sorg ist seit November 2021 bei der READ-COOP im Sales und Projektmanagement beschäftigt. Dabei verantwortet er den gesamten Prozess vom ersten Kontakt mit dem Kunden bis zum Abschluss der aus dem Sales-Prozess resultierenden Projekte zusammen mit dem Sales- und Projektmanagementteam. Neben jahrelanger Erfahrung im Sales-Bereich hat er einen akademischen Hintergrund in Betriebswirtschaftslehre (BA) und strategischem Management (MSc).

### ***Expert (Sebastian Colutto)***

Sebastian Colutto ist seit der Gründung der READ-COOP vor 3 Jahren dort als Softwareentwickler tätig. In den vorangegangenen zwei EU-Projekten Transcriptorium und READ war er am Aufbau der Transkribus-Plattform führend beteiligt. Derzeit beschäftigt er sich vor allem mit der Entwicklung und Integration neuer Tools aus dem Machine-Learning-Bereich in die Plattform. Sebastian Colutto hat Informatik an der Universität Innsbruck studiert. Dieses Studium hat er im September 2012 mit seiner Doktorarbeit auf dem Gebiet Image Processing & Numerical Optimization abgeschlossen.

Tool details: <https://readcoop.eu/de/transkribus/>

## **FairCopy: Nick Laiacona**

### ***Tool***

FairCopy is a specialized word processor for the study and transcription of primary sources in literature and history. We are using FairCopy in the Native Bound Unbound (NBU) project to structure and annotate transcriptions and translations in a variety of formats. For each primary source, the transcription and images flow in from FromThePage. The translation is then imported as plain text, exported from MS Word. These are assembled into a single TEI Document in FairCopy.

FairCopy allows us to markup texts in a deeper way than is possible in a public crowdsourcing environment. Once in FairCopy, the editor can markup the structure of the document to prepare it for publishing on the web. The interface is customized for the NBU project so that editors just have the elements and attributes they need for the job. Users can also markup up names of people and places, as well as identify events. Fortunately, TEI is quite flexible in this regard and FairCopy makes the process easier and more approachable to our non-technical collaborators.

FairCopy also integrates with ArchivEngine, which provides version control and team collaboration. Using ArchivEngine, the whole team can browse the repository of texts. Team members can check out texts to edit and check them in when they are done. They can also change the project schema, which is then reflected to all other users. ArchivEngine has both a public and an authenticated REST API. The public API is used for publishing texts via IIIF. The private API is used by FairCopy and the NBU database. Once the texts have been checked into ArchivEngine, the NBU database can check the documents out and scan for people, place, and event names. These are ingested into a database and reconciled against existing records, merging any duplicate records with new information. This database of people, places, and events serves as the backend to the public facing website.

### ***Expert***

Nick Laiacona is a partner at Performant Software Solutions LLC. Performant serves clients in the Digital Humanities throughout North America and Europe. Laiacona has developed tools for critical digital editions including: Juxta, Digital Mappa, TextLab, and now FairCopy. He has also helped produce a number of critical editions, including “Secrets of Craft and Nature in Renaissance France” and the “Melville Electronic Library.”

Tool details: <https://faircopyeditor.com/>

## **IIIF: Glen Robson**

### ***Tool***

IIIF is a community-driven technology led by world-leading research, national and state libraries, archives, museums, companies and image repositories committed to providing access to high quality digital resources. It is a framework for publishing images and metadata in a way that can make the content interoperable and can be particularly relevant for digital editions. The framework supports annotations and encourages the reuse of content in various tools including transcription and crowdsourcing systems. This presentation will introduce the concepts of IIIF and showcase examples of how different institutions are making use of it.

### ***Expert***

Glen Robson works for the IIIF Consortium as the IIIF Technical Coordinator giving training and assisting the community to implement IIIF. Before this he spent 13 years working at the National Library of Wales (NLW), latterly as the Head of Systems. Glen started working with the IIIF standard in 2013 and implemented the IIIF standard within the library on its Newspaper, Photograph, Archive, Map and Crowdsourcing Systems.

Tool details: <https://iiif.io/>

## **FromThePage: Ben Brumfield**

### ***Tool***

The web-based transcription system FromThePage is known for large, public crowdsourcing projects at the National Archives of Sweden or the UK National Archives, but it was originally built to create digital scholarly editions. An easy-to-use interface lets users focus on encoding text, while import/export features built on IIIF and TEI provide data integration with publication systems.

This presentation will cover the following topics:

* Collaborative manuscript transcription and OCR text correction
* Case studies on scholarly editions using FromThePage
* Workflows for public crowdsourcing within digital scholarly editions
* Annotation and indexing features
* Integration workflows with Transkribus, Omeka-S, and FairCopy
* Beyond transcription: structured data indexing for forms, index cards, and ledgers

If time allows, a hands-on editathon will be conducted using 20th-century German documents from the USA.

### Expert

Ben Brumfield is a software engineer in Austin, Texas, USA. He founded the collaborative digital edition platform FromThePage in 2005 and has been researching crowdsourced manuscript transcription since 2007. FromThePage has been used by the Frederick Douglass Papers, the Civil War and Reconstruction Governors of Mississippi Digital Edition, and the Wilford Woodruff Papers to transcribe, index, and annotate historic documents.

Tool details: <https://fromthepage.com/>

## **Abendvortrag/Keynote*:* Ulrike Henny-Krahmer**

### ***Tools für digitale Editionen als Helfer und Hürden: Wie nehmen wir die Werkzeuge in die Hand?***

Für die Erstellung digitaler Editionen sind Tools, die einzelne Aufgaben im Arbeitsprozess übernehmen oder unterstützen, an vielen Stellen unverzichtbar, sei es zur Digitalisierung der zu edierenden Gegenstände, zur Transkription und Codierung von Inhalten und Form oder zur weiteren Erschließung und Publikation der digitalen Edition. Bestehende Software ist hier essentiell, um die Arbeit an einer digitalen Edition überhaupt erst zu ermöglichen und vor allem effizient zu machen. Tools sind aber nicht nur Helfer, sie können auch Hürden darstellen. Zum Beispiel ist eine Einarbeitung in die Handhabung einzelner Werkzeuge erforderlich und für den Editionsprozess als Ganzes müssen oftmals verschiedene Tools kombiniert und aufeinander abgestimmt werden. Auch können Tools die Handlungsfreiheit zum Teil begrenzen bzw. lenken sie Arbeitsprozesse in eine bestimmte Richtung, indem sie Optionen für Ein- und Ausgabeformate und durchführbare Operationen vorgeben. Schließlich müssen Tools auch längerfristig gewartet werden können und verfügbar bleiben, wenn sie wiederholt oder dauerhaft eingesetzt werden sollen. Wie können Editor\*innen und Entwickler\*innen von Editionstools den ermöglichenden Charakter der Werkzeuge stärken? Der Vortrag nimmt hierfür Aspekte nachhaltiger Softwareentwicklung in den Blick, aber auch die Bedeutung von Data und Coding Literacy sowie der Einbindung von Editions-Tools in den wissenschaftlichen Diskurs.

### ***Expert***

Ulrike Henny-Krahmer ist Akademie-Juniorprofessorin für Digital Humanities an der Universität Rostock, wo sie mit dem Vorhaben Uwe Johnson-Werkausgabe zusammenarbeitet. Sie ist eine der Managing Editors der Rezensionszeitschrift RIDE, die vom Institut für Dokumentologie und Editorik (IDE) herausgegeben wird. Ihre Forschungsschwerpunkte sind digitale Editionen, Textanalyse und Evaluation und Nachhaltigkeit digitaler Forschung.