Ad fontes: e-learning platform for the handling of archival sources goes Automatic Text Recognition (ATR)

Bitterli, Laura MA
Historisches Seminar, Universität Zürich, Lehrstuhl Teuscher

1. The platform
Ad fontes is an e-learning tool that has been introducing students, researchers, and an interested lay public to working with archival sources and preparing them to go to the archives for around 20 years. The e-learning platform has been used intensively for years for blended learning.

Core values:
• Open platform with quality assurance
• High quality ensured by expert knowledge
• Research based development
• Open access
• Licensing using creative commons
• Transcription exercises for all levels
• Wide range of languages, scripts, and source types
• Self-explanatory learning modules

2. Diplomats on Ad fontes
Analog skills taught digitally
On the one hand, on our platform there are modules that introduce charters as a source type and diplomas as the associated discipline. On the other hand, many of the transcription exercises – the heart of Ad fontes – work with charters covering 700 years and several languages. We provide our images for transcription exercises ensuring readability thanks to zooming and high resolution. For this we use the IIIF standard and its annotation functionality, which enables interactivity. Transcriptions and hints about words or signs are thus available upon request (fig.1). Each exercise is embedded in a context page, a resolution page, and, if appropriate, a transcription and/or glossary.

Analog skills in a digital world
New digital tools, such as the Automatic Text Recognition (ATR) of non-standard scripts and layouts, still have high error rates (fig.2). This makes it unavoidable to clean up the data by hand. The practice of transcribing and more general of working with archival sources today takes place at the intersection of analog and digital skills. Therefore, it remains essential to acquire knowledge of the basic skills, especially in paleography, even in a digital world.

3. ATR on Ad fontes
Working with Automatic Text Recognition (ATR) currently provides a high initial hurdle on the part of interested researchers. Ad fontes is therefore developing a new learning module that will introduce students and researchers to Automatic Text Recognition (ATR), reduce this initial extra work, and at the same time teach techniques for working with the ATR output, especially in combination with conventional qualitative methods of source processing.

Conclusion
Even though Automatic Text Recognition (ATR) for non-standard scripts and layouts advances with great speed, analog skills remain imperative. Ad fontes combines teaching such analog skills with an interactive digital platform. It is currently developing a new learning module with a focus on the practical application of ATR, especially in relation to the further analog processing of the first machine read output.

Fig.1: Screenshot of a transcription exercise on Ad fontes, based on a charter of 1512. „Bestellbrief der Drei Bünde für den Hauptmann Hercules von Salis-Marschlins“

Fig.2: Screenshot of a Text Recognition Run on Transkribus, without any adjustments or corrections of layout and text recognition.