



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Transkribus and IIF

Citation for published version:

Krull, F, Muehlberger, G & Terras, M 2019, 'Transkribus and IIF: Beneficial possibilities between image sharing and Handwritten Text Recognition frameworks', Paper presented at 2019 IIF Conference - Göttingen, Germany, Göttingen, Germany, 24/06/19 - 26/06/19.

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



[Transkribus and IIF: beneficial possibilities between image sharing and Handwritten Text Recognition frameworks](#)

Florian Krull - University of Innsbruck (Austria), [Guenter Muehlberger](#) - University of Innsbruck (Austria), **Melissa Terras** - University of Edinburgh (UK)

Presentation type: Presentation

Abstract:

The Transkribus platform enables scholars to train computer models in order to automatically recognize handwritten and printed documents. There is no restriction regarding the age, style, script or language of the documents. Codices from the Middle Ages can be processed in the same way as modern manuscripts, languages such as ancient Greek, Latin or Hebrew do not impose more challenges than modern English, German or Arabic. Transkribus is designed as an open platform in order to support the collaboration among scholars, computer scientists, archives, libraries and the public. The assumption is that the more data there is in the platform, the higher will be the benefit for all users. In February 2019 more than 20,000 users were registered in the platform, hundreds of them actively transcribing documents. More than seventy institutions, such as universities, research groups, archives, and libraries have concluded collaboration agreements indicating their interest in the platform and the technology. In January 2019 more than 220 text recognition models were successfully trained by Transkribus users, indicating their diverse interests and competences. In order to sustain the service platform several universities and institutions around Europe will set up a European Cooperative Society. This legal entity will enable the team to further develop the platform and offer services on a new scale. See Muehlberger et al (Forthcoming 2019) for an overview.

One of the new services will be to explore the options to connect the IIF technology and community with the Transkribus platform. We can think on several scenarios: First of all a simple “upload” mechanism for transferring images from a IIF server to the Transkribus platform would enable libraries and archives to enrich their holdings with textual data. The simple delivery of the ID of the manifest would start a job a text recognition job at Transkribus. In return an updated manifest could be either delivered to the hosting institution, or the manifest could also be made available via a IIF server based within the Transkribus environment. Secondly transcription interfaces – as they are increasingly developed at several institutions on basis of IIF technology – may be connected with the Transkribus API in order to not only request a recognition job, but also to use the transcribed texts as training material for machine learning processes. Other options include the making available of all images within Transkribus (daily upload of users is currently around 5000 images) via IIF, or to even move specific features, such as training or evaluation functionalities into a IIF based web-interface.

The presentation will discuss the several options and invite the audience to come up with their own ideas and suggestions as to how these two communities and frameworks can work together.

Muehlberger, G. et. al: Transforming Scholarship in the Archives Through Handwritten Text Recognition: Transkribus as a Case Study. Journal of Documentation (In press)

Topics:

- Annotation, including full-text or academic use cases,
- Using IIF material for Machine Learning and AI

Keywords:

- OCR,
- HTR,
- text recognition,
-