Recognition and Enrichment of Archival Documents

D4.8
User Satisfaction

Louise Seaward, UCL

Distribution: Public

http://read.transkribus.eu/

Project ref no.	H2020 674943
Project acronym	**READ**
Project full title	**Recognition and Enrichment of Archival Documents**
Instrument	H2020-EINFRA-2015-1
Thematic Priority	EINFRA-9-2015 - e-Infrastructures for virtual research environments (VRE)
Start date / duration	01 January 2016 / 42 Months

Distribution	Public
Contractual date of delivery	31.12.2017
Actual date of delivery	22.12.2017
Date of last update	22.12.2017
Deliverable number	D4.8
Deliverable title	User Satisfaction
Type	Report
Status & version	Public, Version 1
Contributing WP(s)	4
Responsible beneficiary	UCL
Other contributors	ABP, NAF, StAZH, UIBK
Internal reviewers	Günter Mühlberger, Joan Andreu Sanchez
Author(s)	Louise Seaward
EC project officer	Martin Majek
Keywords	Dissemination, Documentation, User satisfaction, Surveys

Table of Contents

Executive Summary

User satisfaction is of key importance in the READ project. We need to maintain and develop our relationship with the main user groups of the Transkribus research infrastructure: archivists, humanities researchers, computer scientists and members of the public. During 2017 we have continued to engage with and support our users through surveys, direct contact and instructional documentation. We also organised the first Transkribus User Conference, where users were able to receive information and training, share their experiences, provide feedback on and ask questions about our platform. User feedback is evaluated and passed on to developers at UIBK so they can improve the Transkribus user experience.

1. User Surveys

1.1. Background

Regular user surveys are necessary in order to ensure that we understand the needs and desires of Transkribus users.

We have conducted several surveys of archival users, both during 2016 and prior to the READ programme as part of the tranScriptorium project. The results of these surveys suggested that there is a broad audience interested in digitised historical material. The surveys also indicated that these users were interested in the more efficient techniques of transcription and searching that are made possible within Transkribus. The information gathered about the interests and motivations of different types of users fed into the creation of seven user personas that represent a hypothetical set of users. The READ project team refer to these personas when they are discussing the construction of tools and the writing of instructional documentation for users.

For a full account of these surveys and the user personas see *D4.7 User Satisfaction*.

1.2. Survey of Transkribus Users

We have started to collect user feedback from participants in Transkribus workshops by means of a simple survey (using Google Forms). We share this feedback with the Transkribus developers at UIBK to indicate which features of the platform users may struggle with. We have also started to collect feedback from colleagues and contacts on prototypes of new tools being built by READ, such as the ScanTent device, the Transkribus Learn e-learning app and the Transkribus My Collections web interface.

In KPI-I5 we suggested that we would undertake regular surveys to investigate user satisfaction. The size of the Transkribus user community (more than 8,300 registered users as of November 2017) means that it will be a time-consuming exercise to survey the entirety of this group and then analyse and implement their feedback. It was decided that we would organise a user conference as an intensive forum for feedback and support, in lieu of a general survey in 2017 (see section 1.3 Transkribus User Conference).

In the next reporting period, we intend to send a short survey about the usage, strengths and weaknesses of Transkribus to all users. We are investigating the possibility of including short, pop-up questions inside the Transkribus platform so that users are surveyed as they work. This feedback could be linked to an analysis of user behaviour, which is visible to the UIBK team on the Transkribus server. We also plan to work in association with a University course in Digital Humanities, where students will be able to test out and comment on Transkribus and other tools.

1.3. Transkribus User Conference

The first Transkribus User Conference took place at the Technical University of Vienna on 2-3 November 2017 and was attended by around 80 users, plus partners from the READ project. It was advertised to all registered users of Transkribus via email, as well as on the READ project website and Transkribus Twitter account. Targeted invitations were also sent to all of the project's MOU partners.

Around half of the attendees were humanities researchers and most of the other participants were drawn from the archives and libraries domain. Around 60% of the attendees were male and 40% were female.

As well as hearing presentations from READ researchers and Transkribus users, the conference included lots of opportunities for users to discuss their work with READ project partners, share experiences with other users who were interested in similar documents, suggest possible improvements to the Transkribus workflow and ask questions.

The conference included a 'Transkribus Feedback Session' where users were split into groups according to their interests and background and then asked to discuss their views of the strengths, weaknesses and potential of Transkribus. We found that users would like the READ project to provide more information about the process of training Automated Text Recognition models, make it possible to share models between similar collections and improve the processing of documents containing tables and forms. These suggestions are on the READ project agenda for 2018 and we will communicate information about their implementation to the users.

We also asked attendees at the conference to fill in an online survey about their experience of the Transkribus expert client but unfortunately there were too few responses to draw any further conclusions about user needs.

For more on the Transkribus User Conference see *D2.12 Workshops*.

1.4. Focus groups

Focus groups of interested users are an efficient way of gaining detailed feedback on Transkribus.

Established in 2016, we have a small group of volunteers at the Civic Archives of Bonze-Bolzano who are working with Transkribus to create training data in order to generate a model capable of processing the town's municipal records. These users are providing feedback on the usability of the Transkribus platform.

We have also taken the first steps to set up an academic focus group for the project commenting on Transkribus and the overall direction of the READ project. We made contact with several interested researchers with a view to organising the first meeting of such a

group at the Transkribus User Conference. Conflicting schedules of the proposed participants meant that a meeting was not possible, but we will make a second attempt to organise a meeting of this group during the next reporting period.

2. Contact with users

1.1. Communication

There are more than 8,000 registered users of Transkribus (as of November 2017). 2103 of these can be defined as active users, who have saved their documents at least 5 times in the platform.

Members of the READ project spend a lot of time communicating with new and existing users of Transkribus. Most user enquiries come to the Transkribus email account; users can also contact members of the Dissemination Working Group directly, send a message to the Transkribus Twitter account or request new features and report bugs using the appropriate button in the Transkribus platform.

At the time of writing, the Transkribus email account has received more than 1200 emails from users all over the world. In addition to this, hundreds of emails relating to READ and Transkribus have been sent to the personal accounts of the project team.

UIBK has set up a Datawiki content management system, where we will be able to keep better track of user enquiries and ongoing projects.

1.2. Mailing list

In reference to KPI-I15, we have continued to operate without a regular mail-out for Transkribus users. We occasionally email all users about important news like the Transkribus User Conference. Most of the time we disseminate news through the READ project website and the Transkribus Twitter account, instead of contacting all users.

The success of the Transkribus User Conference demonstrated that users benefited from sharing their experiences with each other and that there is an appetite for a more active user community. We plan to set up a Transkribus reddit page where more experienced users can communicate. The Dissemination Working Group will pre-populate the reddit page with the answers to frequently asked questions, and then promote and moderate the page once it is made publicly available in 2018.

2. Instructional Documents

2.1. Transkribus website

An improved version of the Transkribus website is currently under development by UIBK (see wireframes produced by NAF in D8.8 Layout Analysis and Crowdsourcing). The new website will include a streamlined landing page, updated explanations of Transkribus and READ project technology, images of documents currently being processed in the platform, user guidelines and information about successful text recognition models that have already been

generated. This reworked website will be linked to the Transkribus My Collections web interface, which is also in development.

2.2.   Transkribus wiki and How to Guides

The Transkribus wiki remains a central repository for instructional documentation to help users navigate Transkribus. It includes help on installing the platform, background information on Automated Text Recognition, How to Guides and videos, along with a section of frequently asked questions.

Members of the Dissemination Working Group ensure that the information on the wiki remains up-to-date.

We are also working on new How to Guides. The first, on Text2Image matching, will help users prepare their files so that training data for text recognition can be generated automatically. This process should be popular with users as it allows them to work with their existing transcriptions to train a model relatively easily. The second guide, on tables and forms, will explain how users can apply Transkribus segmentation tools to documents set out in a tabular layout (see D8.11 Keyword Spotting in Registry Books). This information is particularly important for users because automatic segmentation tools cannot yet process tables and forms with complete accuracy.

2.3.   Success stories

We have started to publicise success stories from users who have worked with our Automated Text Recognition technology and generated impressive results. This information is important as it allows users to understand that text recognition can work on a range of different document types and reach high levels of accuracy.

We post regularly about good examples of Automated Text Recognition that have been trained by our MOU partners or other collaborators on the 'News' page of the READ project website. There is also a section on the Transkribus wiki which provides images of material that has been automatically processed in Transkribus.

We have produced a detailed factsheet based on an Automated Text Recognition model created for the Archivio Storico Ricordi, an Italian archive of musical material. This factsheet explains the amount of training data submitted by the archive and the average accuracy of the computer-generated transcriptions. It also includes an image of a sample document and its automated transcription so that potential users can get a sense of how much the engine has recognised correctly.

We are now preparing similar factsheets based on other models that have been trained with good results. We also plan to create a page on the new version of the Transkribus website providing details of the different collections that we are working with and the accuracy rates of the strongest models that we have generated. There are currently 3 publicly available models and we hope to open up more examples so that users can experiment with models that might work on their collections. This information should be encouraging for potential users who may be unsure whether their documents can be processed. It should also help users with similar collections to collaborate on the preparation of training data, benefit from data already in our system and share the resultant models.

## 2.4.	Webinars and videos

ABP created our first instructional video which explains ['How to Use Transkribus - in 10 steps (or less)'](#). UIBK delivered a live webinar on Transkribus on 14 December 2017 for members of IMPACT and the Open Preservation Foundation, two organisations focused on digital research training and sustainability. The webinar was similar in format to a Transkribus workshop and included a demonstration of the platform, an overview of the workings of Automated Text Recognition and discussion of some of the new tools being built by READ. During 2018, we would like to provide regular live webinars open to all, where participants can sign up and learn about Automated Text Recognition and the features of Transkribus,

We are also planning to develop new videos focused on specific aspects of the Transkribus workflow such as document upload, management and export, image segmentation, tagging and Text2Image matching. Users who get in touch with frequently-asked questions could be directed to these videos, meaning that we would need to spend less time writing explanations by email.

We are also investigating the possibility of developing a Transkribus Massive Open Online Course (MOOC), which would include videos on READ project tools, the technical research of READ project partners and the workflow and philosophy of humanities projects who are working with our technology. This would help us promote the project to new audiences and also provide guidance and background information to help users work with the platform.

# 3.	Conclusion

The automated processing of large collections of documents could be transformative for the main user groups of Transkribus: archivists, humanities researchers, computer scientists and members of the public. At the moment, the most active Transkribus users come from the first two of these groups. The Transkribus User Conference gave us an opportunity to offer direct support to, and gain feedback about the needs and desires of our most enthusiastic users. We have continued to communicate through emails, documentation and videos to ensure that users are able to get the most out of the Transkribus platform. We have plans for more user surveys in 2018, which can then be fed into UIBK's development work on Transkribus. It will also be necessary to gauge users' response as new tools like the Transkribus My Collections web interface and the Transkribus Learn e-learning app become ready for general use. We also want to work on developing the Transkribus user community, with initiatives like communication on reddit, a webpage showcasing the work of the user network and the second Transkribus User Conference.