

READ

Recognition and Enrichment of Archival Documents

D4.9 User Satisfaction

Louise Seaward, UCL

Distribution: Public

<http://read.transkribus.eu/>

READ
H2020 Project 674943

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 674943



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	30.06.2019
Actual date of delivery	05.08.2019
Date of last update	05.08.2019
Deliverable number	D4.9
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	Christophe Doin
Keywords	Dissemination, Documentation, User satisfaction, Surveys

Table of Contents

Executive Summary	4
1. User Surveys	4
1.1. Background	4
1.2. ScanTent and DocScan surveys – June and November 2018	4
1.3. Transkribus expert client survey – November 2018	5
1.4. Transkribus and HTR survey – March/April 2019	5
1.5. Focus groups	6
1.6. ScanTent and DocScan user evaluation	6
2. Contact with users.....	6
1.1. Communication.....	6
2.1. Transkribus User Conference.....	6
1.2. Transkribus on Facebook	6
1.3. Transkribus Working Groups	7
1.4. Transkribus Newsletter	7
1.5. Transkribus mailing list	7
2. Instructional Documents	7
2.1. Transkribus website	7
2.2. How to Guides.....	8
2.3. Success stories	8
2.4. Videos.....	9
3. Conclusion	9
4. Supplement	9
4.1. Transkribus Expert Client Survey 2018 – Questions and Analysis of Responses.....	9

Executive Summary

The Transkribus user network is expanding rapidly and we support our users via instructional documentation, video tutorials and various forms of online communication. User comments, questions and feature requests give us a daily insight into the user experience. We have also conducted several user surveys, addressed to specific audiences and to every registered Transkribus user. Feedback gathered through these surveys indicates that users are generally satisfied with the functionality of project tools and have been able to generate useful results with Handwritten Text Recognition (HTR) technology. Some users reported concerns relating to the usability of specific features and the long-term sustainability of the technology. Development work continues with a view to ensuring that the Transkribus expert client and web interface are user-friendly.

1. User Surveys

1.1. Background

Information gathered in surveys helps us to ensure that the Transkribus research infrastructure is meeting the needs of its main user groups: humanities researchers, archives professionals, computer scientists and members of the public.

Surveys conducted during the [tranScriptorium](#) project and the first year of READ showed that these user groups have a significant interest in the potential of HTR technology.

For a full account of our previous surveys, see [D4.7 User Satisfaction](#).

1.2. ScanTent and DocScan surveys – June and November 2018

We collected written feedback from participants in Scanathon events held in June and November 2018. For more on these events see, [D2.13 Workshops](#). Most participants agreed that the DocScan app and ScanTent device were user-friendly, appropriate for small-scale digitisation and affordable for archival institutions. Representatives from archives and libraries showed support for the idea of receiving copies of user-generated images via QR codes but pointed out that differentiating standards for image quality, formats and metadata were likely to be problematic. Some people expressed concerns that it may be difficult to supervise archival users working with a ScanTent and others registered disappointment that the tools sometimes struggled to handle documents smaller or larger than A4. These concerns should be alleviated by more sustained user engagement with the tools, which will allow people to discover the optimal conditions for digitisation. Some specific changes were suggested, such as improved on-screen instructions in DocScan and the inclusion of a rechargeable battery pack with the ScanTent. Both of these changes have been implemented. For more on the latest enhancements to DocScan and the ScanTent see, [D8.3 Open Innovation Forum](#).

1.3. Transkribus expert client survey – November 2018

The second Transkribus User Conference presented us with an opportunity to survey some of our most advanced and enthusiastic users about their experience of using the Transkribus expert client. Conference participants were given time to complete a survey as part of the event programme. We received responses from 72 people, around 70% of the users in attendance at the conference. This survey provided an insight into common desires and questions relating to Transkribus, which will be an important foundation for making Transkribus more user-friendly and building user engagement under READ-COOP.

Around half of the respondents worked with Transkribus regularly and many expressed high levels of satisfaction with the platform. There were several requests for a stronger user community and more information about best practice for HTR models. Specific changes were suggested that could enhance the usability of the Transkribus interface. A more in-depth analysis of the survey responses can be found in section 4. Supplement.

1.4. Transkribus and HTR survey – March/April 2019

In March 2019 we launched our largest ever user survey, which was publicised to all registered Transkribus users. The aim of this survey was to evaluate the usability of Transkribus and understand how HTR is affecting access to archival documents, the nature of historical research, and methods used to study and understand historical handwriting.

The survey received over 150 responses. The survey was open for one month, from 26th March - 24th April. During this time period 1458 unique users accessed Transkribus via the expert interface. The number of logins was nearly 10.000, so around 300 logins per day (9520). Some highlights from the survey:

- 37% of the respondents regard results from the HTR as very or quite accurate and 21% still as variable (depending on the individual documents), 24% regard results as disappointing or unuseable.
- 59% of the respondents were very or quite satisfied with the communication with the Transkribus team, just 2% were not very satisfied, or not at all satisfied.
- More than 60% of the respondents used Transkribus for more than a month, 24% for more than a year, and about 10% of the respondents used it every day.

Some answers from the free text responses:

- Why will Transkribus be useful?
 - o *Because it is very difficult to know if a document will contain any useful information before deciphering it which takes a lot of time*
 - o *At the beginning I couldn't read the handwriting (German Current), it's not used any more. The letters have been "hidden away" by my Grandma. Apart from that I suppose that my Granpa's writing is not easy to read at all.*
 - o *We are a small workforce so having HTR complete even a portion of the transcription process is helpful to us.*
 - o *Because no other OCR or HTR software can assure the results obtained by Transkribus; typography is not regular but HTR works fine.*

A detailed analysis will be published in full in a forthcoming scholarly article edited by Melissa Terras.

1.5. Focus groups

Volunteers are continuing to work with Transkribus at the Civic Archives of Bozen-Bolzano and providing feedback on its usability.

In 2017, we considered setting up an academic focus group for the project that would provide advice on the future direction of Transkribus. The founding members of the READ-COOP will now fulfil this function.

1.6. ScanTent and DocScan user evaluation

During spring 2019 more than 150 ScanTents were sent out to users all over the world. They were informed that they can test the ScanTent until September 2019 under the condition that they will take part in a user survey. This user survey was drafted in June 2019 and sent out to users in July 2019 (after the end of the project). Detailed figures will be provided at the final review meeting.

However, first results are very encouraging: Based on 31 answers so far we can say that 87% of all users are very satisfied (47%) or satisfied (40%) with the Scantent. Also with the DocScan app more than 70% are satisfied or very satisfied.

2. Contact with users

1.1. Communication

At the end of the project we counted more than 25,000 registered users of Transkribus. This means that from the beginning of the project to its end the number of users raised nearly by the factor of 10. READ project staff communicate with users all over the world, sending and receiving hundreds of emails each month. User requests also come through the Transkribus Twitter account and the bug report button in the Transkribus expert client. UIBK uses a Datawiki content management system to keep a record of user enquiries.

2.1. Transkribus User Conference

The second Transkribus User Conference took place at the Technical University of Vienna on 8-9 November 2018 and was attended by around 100 users, plus partners from the READ project.

Attendees received information about the latest technological innovations, new features of the Transkribus expert client and the plans for READ-COOP. Importantly, participants were able to network with other Transkribus users, as well as developers and computer scientists from READ.

For more on the Transkribus User Conference see *D2.13 Workshops*.

1.2. Transkribus on Facebook

In May 2018 a '[Transkribus Users](#)' group was launched on Facebook by Sinai Rusinek, an academic based in Israel. The group now has over 200 members. Transkribus users in the group conduct discussions amongst themselves, with occasional interventions from READ project staff.

In February 2019, we set up an [official Facebook page](#) for the Transkribus platform, which we regularly update with news from the READ project. The Facebook page currently has over 210 followers (5th August 2019).

1.3. Transkribus Working Groups

We have also been supporting users who work on similar material through the establishment of [small working groups](#). There are now three Transkribus working groups cooperating to improve the recognition of medieval Gothic script, early modern legal documents and early modern Dutch handwriting. Users are sharing their images, transcripts and ground truth and using this data to create generic models designed to recognise particular types of writing from different collections. This kind of collaboration is laying the foundations for further user engagement in Transkribus as strong generic models may allow users to access HTR straight away without going through process of creating ground truth. Working groups are user-led but they can rely on technical support and promotional activity from READ.

1.4. Transkribus Newsletter

We now have a monthly newsletter that reaches all registered Transkribus users.

The newsletter describes the latest features in Transkribus, project events and HTR research. Google Analytics suggests that this newsletter drives significant traffic to the READ project website, creating a broader audience for project news.

1.5. Transkribus mailing list

Feedback from user surveys, the popularity of the Transkribus User Conference and the existence of the 'Transkribus Users' Facebook group makes it clear that users benefit from sharing their experience of working with the platform.

As the user network expands, people are increasingly requesting information about documents have already been processed in Transkribus so they can build upon existing work. There is also significant potential for users to ask questions of each other and share their expertise, which would make it easier for the Transkribus team to address common queries.

UIBK set up a dedicated 'Transkribus community' mailing list, which will work similarly to a user forum. Subscribers can ask questions of those on the mailing list and access an archive of previous queries and comments. UIBK will moderate the mailing list and contribute responses where necessary. The list has been set up in June and will be made available after the end of the project as part of the READ-COOP SCE.

2. Instructional Documents

2.1. Transkribus website

Our 2018 user survey indicated that the Transkribus website is well used. However, it has become outdated over the course of the READ project. UIBK have developed and are now testing a new version of the Transkribus website which includes basic information about the platform, links for registration and installation, How to Guides and contact information. This website will become a vital hub for attracting and introducing new users to the platform.

2.2. How to Guides

Members of the Dissemination Working Group regularly check and update the existing How to Guides for the Transkribus expert client. Responses to the survey issued at the Transkribus User Conference indicated that there were some gaps in the existing guides and these have since been remedied. New guides have recently been created to ensure that users can access instructions for all aspects of the Transkribus workflow.

The following How to Guides are now available on the Transkribus wiki:

- [How To use Transkribus - in 10 steps \(or less\)](#)
- [How To Transcribe Documents with Transkribus - Introduction](#)
- [Transkribus Transcription Conventions](#)
- [How To Use Existing Transcriptions to train a Handwritten Text Recognition model](#)
- [How To Enrich Transcribed Documents with Mark-up](#)
- [How To Use the Structural Tagging Feature](#)
- [How To Process Tables in Transkribus](#)
- [How To Train a Handwritten Text Recognition Model in Transkribus](#)
- [How To use Keyword Spotting](#)
- [How To Export documents from Transkribus](#)

German translations of each of these guides are now available to help us cater to the large proportion of German-speaking Transkribus users. We also have a French translation of the 'How To use Transkribus - in 10 steps (or less)' guide. In addition, we have developed two How to Guides for the Transkribus Learn platform and for the DocScan app and ScanTent device.

2.3. Success stories

Success stories are important means of spreading the word about the project and improving comprehension of the huge potential of HTR for historical handwriting.

We write regular blog posts on the READ project website which explore the achievements of researchers and archivists and this information is disseminated further via the Transkribus newsletter and social media. As an example, a [post about the automated recognition of the writings of the French philosopher Michel Foucault](#) has garnered over 450 unique views and 14,000 impressions on Twitter since it was first posted in June 2018.

Another success story can be reported from a German researcher who trained models on Church Slavonic handwriting. Achim Rabus, who is holding the Chair of Slavic Linguistics at the University of Freiburg in Germany trained several models with rates rates going down to 3 to 5 percent. The model will be shared on the Transkribus platform and a detailed description of the training strategies is provided in the following working paper: [Recognizing handwritten text in Slavic manuscripts: A neural-network approach using Transkribus](#). The paper was seen 462 times (5th August 2019) and is therefore among the top 5% within Academia.edu.¹

¹ https://www.academia.edu/38835297/Recognizing_handwritten_text_in_Slavic_manuscripts_A_neural-network_approach_using_Transkribus_1_Achim_Rabus

2.4. Videos

The [Transkribus YouTube channel](#) offers webinars, tutorials and conference presentations. The 'How To use Transkribus - in 10 steps (or less)' video now has over 3000 views. This indicates that video tutorials play a key role in orientating new users to the Transkribus platform.

3. Conclusion

By surveying Transkribus users, we have been able to recognise how far Transkribus is affecting the research and archival domains. User feedback has helped the Transkribus team to strengthen their awareness of user needs and inspired us to open up new communication channels with users such as the newsletter and community mailing list. The latest user survey has constituted a remarkable opportunity for us to learn about a significant proportion of our user base and a forthcoming scholarly analysis of this feedback will provide important evidence about the impact of the READ project.

4. Supplement

4.1. Transkribus Expert Client Survey 2018 – Questions and Analysis of Responses

Section 1: Introduction – Questions

How did you find out about Transkribus? Please select one option from the following list.

- **Web search**
- **At a presentation**
- **Personal recommendation**
- **Online media (e.g. blog, online article)**
- **Offline media (e.g. magazine, newspaper)**
- **Twitter**
- **Facebook**
- **Other**

How would you describe your level of expertise in Transkribus? Please select the option that you feel most closely applies to you.

- **Beginner (starting to learn about Transkribus, have installed the software)**
- **Intermediate (tested some features of the platform with my own documents)**
- **Expert (use Transkribus regularly, have trained a HTR model)**

How often do you work with Transkribus?

- **Every day**
- **Every week**

- Every month
- Every few months
- Never

What do you hope to achieve by working with Transkribus? Please select all options that apply.

- Experiment with machine learning
- Train a HTR model
- Transcribe historical documents
- Add tags and metadata to transcriptions
- Search a collection with Keyword Spotting
- Receive transcripts for my research
- Receive transcripts for scholarly editing
- Receive transcripts that can be made available to archival users
- Facilitate keyword searching for archival users
- Other

Where do you find information about Transkribus? Please select all options that apply.

- Transkribus website (<https://transkribus.eu>)
- READ project website (<https://read.transkribus.eu/>)
- Transkribus wiki
- Transkribus How to Guides
- Transkribus newsletter
- Transkribus Users Facebook group
- Videos on Transkribus YouTube channel
- Contact with Transkribus team
- Contact with other Transkribus users
- Other

How satisfied are you with the communication from the Transkribus team? Please indicate how satisfied you are on a scale from 1 to 5.

- 5 – Very satisfied
- 4 – Quite satisfied
- 3 – Middling
- 2 – Not very satisfied
- 1 – Not at all satisfied

How could we improve communication with Transkribus users?

Section 1: Introduction – Analysis of Responses

The survey garnered 72 responses, comprising a relatively even mix of beginner, intermediate and expert Transkribus users. Over half of the respondents were introduced to Transkribus by a personal recommendation, which suggests that existing users are effective ambassadors for

the platform. Nearly half of the respondents indicated that they work with Transkribus every week or month, but there were at least 13 conference attendees who had never worked with the platform.

When we asked people where they found out information about Transkribus, the most popular options were the Transkribus website, the Transkribus How to Guides and contact with the Transkribus team. The reliance on these avenues of information underlines the need to revitalise the Transkribus website and ensure that How to Guides are regularly updated. Respondents reported high levels of satisfaction in terms of communication from the Transkribus team.

Unsurprisingly, the most popular objectives for working with Transkribus were transcribing historical documents, training HTR models and Keyword Spotting. Making transcriptions available online appeared to be more important to respondents than research uses. This indicates that the conference attracted more archivists than researchers.

We received a variety of responses to the question of how we could improve communication with Transkribus users. There were many requests for some kind of user forum. Several respondents asked for additional detail in our How to Guides or expressed their desire for more information about HTR results and the future development of Transkribus. **The creation of the 'Transkribus community' mailing list should address some of these concerns.** There will be continued efforts to disseminate information to users and strengthen connections between people working on similar material under READ-COOP.

Section 2: User Experience – Questions

In the next set of questions, we will ask you to assess the usability of different features of Transkribus. Please choose a response on a scale from 1 to 5. If you have not used a particular feature of Transkribus, please leave your response blank.

- **5 – Very easy**
 - **4 – Quite easy**
 - **3 – Middling**
 - **2 – Quite difficult**
 - **1 – Very difficult**
-
- **How easy do you find it to view documents?**
 - **How easy do you find it to upload documents?**
 - **How easy do you find it to segment documents?**
 - **How easy do you find it to transcribe documents?**
 - **How easy do you find it to add tags to transcribed text?**
 - **How easy do you find it to add structural tags to a document?**
 - **How easy do you find it to export documents?**
 - **How easy do you find it to train a HTR model?**
 - **How easy do you find it to generate transcripts with a HTR model?**
 - **How easy do you find it to proof-read and correct transcriptions?**

- **How easy do you find it to manage your documents?**
- **How easy do you find it to work with Keyword Spotting?**

What do you like most about working with Transkribus?

What do you dislike most about working with Transkribus?

Section 2: User Experience – Analysis of Responses

The questions about the ease of working with the different features of Transkribus indicated that most users find it manageable to undertake the workflow of ground truth creation (viewing documents, uploading images, segmentation and transcription). More advanced functionalities relating to tagging and HTR training appear to be more challenging. The survey also indicated that fewer respondents had engaged with the later stages of the HTR process. This evidence suggests that users may benefit from more detailed How to Guides and videos about advanced features.

When people were asked what they liked most about working with Transkribus, they conveyed their enthusiasm for HTR and its transformative potential. Several people commented that the platform was user-friendly and there was praise directed towards elements including Layout Analysis tools, export options and Keyword Spotting.

Complaints about Transkribus related to recurring bugs, the complexity of the interface and the difficulty of organising documents and collections. Requests for improved information about existing HTR models and best practices were repeated.

Section 3: Improvements – Questions

In the next set of questions, we will ask you to suggest ways in which we could improve the different features of Transkribus. If you have no improvements to suggest, please leave your response blank.

- **How could we improve document management?**
- **How could we improve segmentation tools?**
- **How could we improve options for transcribing and tagging text?**
- **How could we improve HTR training?**
- **How could we improve Keyword Spotting?**
- **How could we improve export options?**

Are there any features that you think are missing from Transkribus? If so, please provide a short explanation.

Section 3: Improvements – Analysis of Responses

This section of the survey provided lots of useful feedback on potential enhancements for the different features of Transkribus.

In terms of document management, many users asked about the possibility of sharing HTR models between collections. This is an existing functionality in Transkribus but one that needs to be publicised more widely. Users also suggested that the 'Server' tab should be streamlined to provide a better overview of document metadata. Respondents wanted to be able to rename and move their documents with greater ease.

When it came to segmentation tools, some people stated that tools for manual segmentation and correction were not user-friendly. Respondents asked for improved table recognition, which is an ongoing area of research.

The survey also asked for opinions on the process of transcription and tagging. Several respondents asked about tag shortcuts, another existing Transkribus feature that needs to be communicated more clearly to users. There were some complaints about the usability of the tagging interface and suggestions of extra functionalities like the possibility of working with non-Unicode fonts.

Requests for improved documentation and dissemination were repeated in response to the question about HTR training. People wanted to have more information about the technical side of the training, such as the meaning of 'epochs' and the optimal size of the training set. Many respondents pointed out that they would find it useful to access existing models that might be applicable to their script. At the time of writing, models can only be shared between specific collections that individual users have access to. Respondents also looked forward to future developments in the technology relating to the recognition of abbreviations and deleted text.

The Keyword Spotting feature in Transkribus is still in prototype form and so many of the comments offered by users will be integrated in due course. Users wanted to be able to correct the results of Keyword Spotting, access the data behind each spotted word and work with the results in an online interface.

Users suggested a few improvements to the export process. The most frequent requests related to optimising the existing export options to ensure that they generated useable files in specific use cases. Users also asked for new export options, such as a HTR transcript where potentially incorrect words are highlighted.

Finally, we asked users to tell us if particular features were missing from the platform. Some of the proposals included more documentation on the API, a public library of models, a clearer workflow for HTR correction, direct links to other systems such as IIF and Recogito and the implementation of new features like Named Entity Recognition.

Section 4: Final Questions – Questions

How has working with Transkribus impacted your professional and/or scholarly activity?

Do you have any concerns about working with Transkribus and HTR technology?

Do you have any questions that you would like to ask the Transkribus team?

Finally, is there anything else you would like to tell us about your experience of Transkribus?

Section 4: Final Questions – Analysis of Responses

This section gave Transkribus users the opportunity to offer more generalised feedback.

Many people reported that working with Transkribus had made a big impact on their work, helping them to improve their skills, analyse documents in new ways and make collections more accessible. There was also much acclaim for the platform and the team behind it.

When asked to share their concerns about Transkribus and HTR, people again referenced the need for more information about HTR results. Other issues arising included server bottlenecks, sustainability and disappointment about the closed source elements of the platform.