



Recognition and Enrichment of Archival Documents

D2.6. Dissemination and Awareness Plan

Günter Mühlberger (UIBK),

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	31.12.2018
Actual date of delivery	21.01.2019
Date of last update	21.01.2019
Deliverable number	D2.6
Deliverable title	Dissemination and Awareness Plan
Type	Report
Status & version	Final
Contributing WP(s)	All WPs
Responsible beneficiary	UIBK
Other contributors	All beneficiaries
Internal reviewers	Maria Kallio, Louise Seaward
Author(s)	Günter Mühlberger
EC project officer	Christophe DOIN
Keywords	Dissemination

Table of Contents

Executive Summary	4
1. Dissemination activities.....	4
1.1. Introduction	4
1.2. Some highlights.....	5
1.3. Messages and claims for Y4	9
2. Specific actions for Y3	11
2.1. Dutch conference on gathering data	11
2.2. Final Transkribus User Conference in Helsinki.....	11
2.3. HTR+	12
2.4. Training data from and for everyone.....	12
2.5. lern.transkribus.eu	13
2.6. Market ScanTent and DocScan	13
3. General dissemination and awareness channels	14
3.1. READ website	14
3.2. Transkribus website	14
3.3. Transkribus web-interface – library	15
3.4. Transkribus wiki	17
3.5. Twitter.....	17
3.6. YouTube video channel.....	18
3.7. ScriptNet	19
3.8. ZENODO	19
3.9. ScanTent/DocScan	20

Executive Summary

The final period of the project will run for 6 months – until end of June 2019. Our main objective for this period is to support the foundation of a European Cooperative Society which will run the Transkribus platform and provide services and products after the end of the READ project. For this purpose we will use and improve the several dissemination channels which we built in the past 3 years. Moreover some specific actions directed toward specific tools are planned.

1. Dissemination activities

1.1. Introduction

The READ project comprises a large number of different strands of activities which all can be summarized in one sentence:

READ revolutionizes access to archival documents

After three years of work we are not only more convinced than ever that this ambitious claim holds true, but we see that also our users share this opinion with us.

Based on a user survey carried out at the 2nd Transkribus User Conference in Vienna in November 2018 the following testimonies were provided in an anonymous user survey based on the following question:

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

- *Greatly. My most interesting job and my first funded project which I am now starting would not have been possible without it.*
- *Could do things that were not possible some years ago*
- *Expect it will affect my profession in the future dramatically.*
- *It has changed the way we planned in our research project to work on manuscripts*
- *Gave our department the ability to start a whole new project*
- *Wouldn't have started my digital edition without Transkribus*
- *Very positively, working with Transkribus has been a part-time job for me for almost two years and it has always worked very well for my research projects.*
- *Quite. it gives hope to archivist huge collections for the public.*
- *ground-breaking!*

Transkribus is the only platform for historical documents where users are enabled to transcribe, train, recognize and search handwritten documents with cutting edge technology. READ has clearly demonstrated the progress made in this respect with impressive figures concerning the ability of machine learning methods to recognize historical documents and make them searchable. Of course our work in READ strongly benefits from the progress made in general in the artificial intelligence and machine learning domain, however there is no project or platform worldwide which has put such a strong emphasis on the recognition of historical documents with such a comprehensive approach.

1.2. Some highlights

1.2.1. Workshops

As in previous years dozens of workshops and talks at dozens of conferences all around Europe were organised by the READ team. Here a selection of some of the important events:

- EuropeanaTech – Rotterdam – Netherlands (14th -16th May 2018)
- DARIAH Conference – Prague (24th April 2018)
- DigiCult Conference – Sweden (26th May 2018)
- Transkribus workshop – Austrian Academy of Science (28th May 2018)
- Mambrino Conference – Verona (5th June 2018)
- German Libraries Day – Berlin (13th June 2018)
- CLARIN Conference – Pisa (8th- 10th October 2018)
- Biblioteka Baltica conference – Rostock (4-5th October 2018)
- Tessmann conference – Bozen – Italy (26th October 2018)
- Lope de Vega 19th Conference – Madrid (7th December 2018)

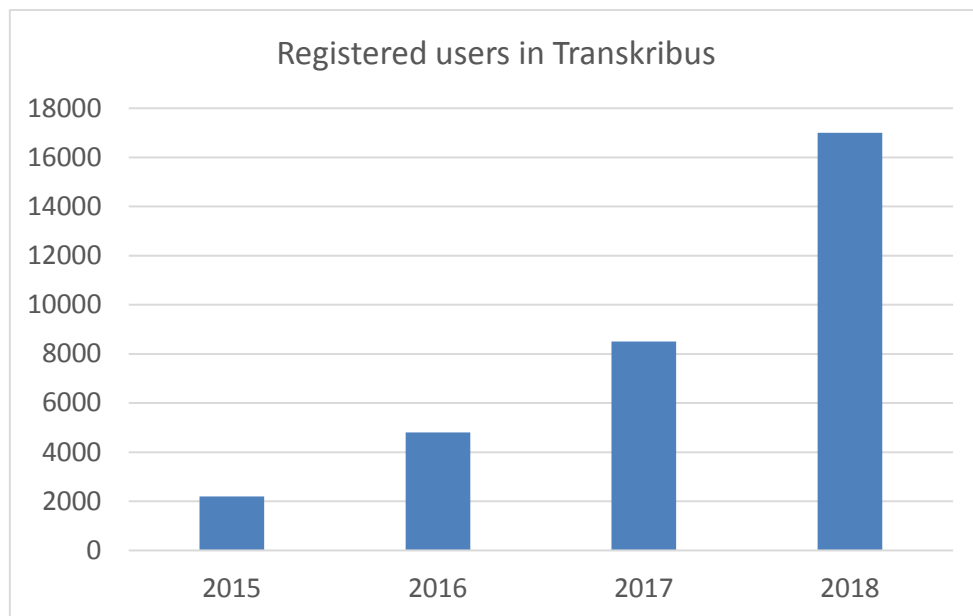
In late December 2018 we introduced the new prototypes of the ScanTent to the public at two workshops in Belgrade and Novi Sad.



Figure 1 From the Belgrade Transkribus workshop: “Democratising digitisation”

1.2.2. Registered users

In Y3 we were again able to double the number of registered users within Transkribus although we started already from a high level (8400 users).



More than 8500 users registered in Transkribus in 2018. It has to be emphasized that we are not running a crowd-sourcing project where such figures can be reached much easier, but that likely 90% of these users are interested in Transkribus and in the technology of Handwritten Text Recognition. Most of them are actually downloading the Transkribus expert client and are experimenting or working with it.

1.2.3. Transkribus User Conference

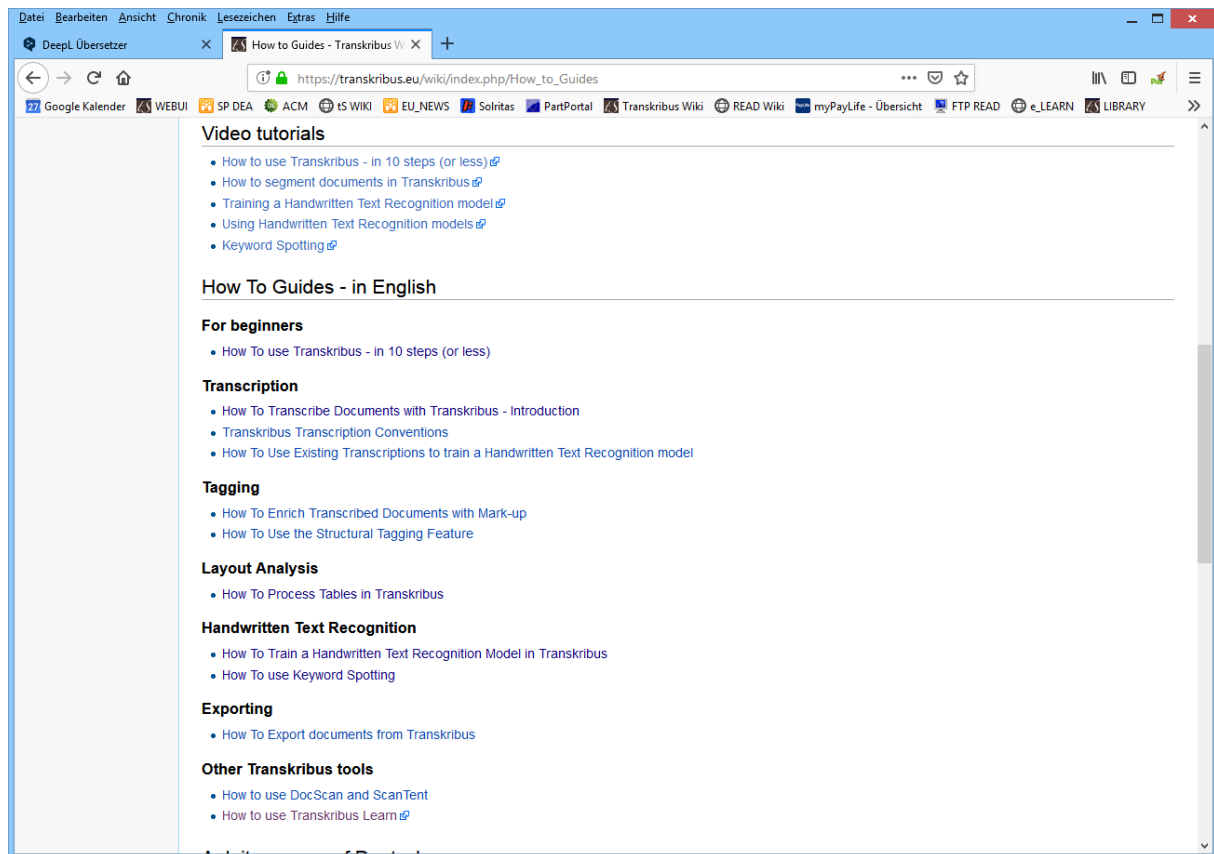
The TUC2018 became again a great success. The 120 places offered were fully booked within a few days after the first announcement.



The conference was again a mixture of project presentations (“Transkribus in practise”), technical presentations about specific issues (table recognition, keyword spotting,...) workshops for beginners and advanced users and about the planned European Cooperative. All talks are now online.

1.2.4. HowTo Guides and videos

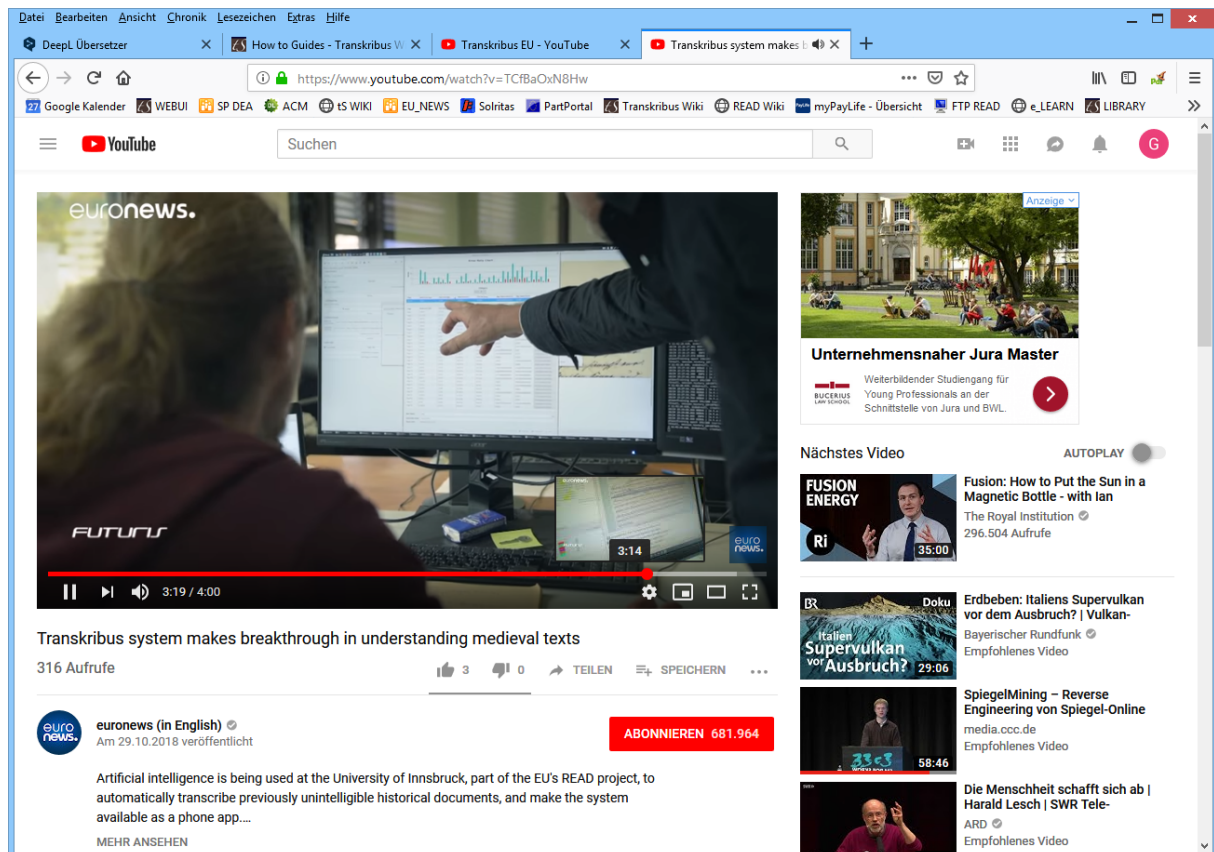
A highlight in Y3 was also the creation of several HowTo Guides providing users with detailed instructions how to use Transkribus and its tools for their purposes.



Due to the important role of German users we also produced German versions of the guides. Video tutorials were also updated and made available via YouTube. The updated version of How to use Transkribus in 10 steps got more than 1500 views within 4 months.

1.2.5. EuroNews Video

In collaboration with the European Commission and EuroNews a general video was produced about Transkribus. It was displayed for several times with Euronews' Futuris channel and also uploaded to YouTube.



The video was also translated into Hungarian and accessed several hundred times.

1.2.6. Transkribus Facebook group

Sinai Rusinek from Bar-Ilan University in Israel has set up a Transkribus User group on Facebook in 2018 independently from the Transkribus team. 175 people joined the group so far.

The Transkribus team is supporting the group with posts and advices on questions raised.

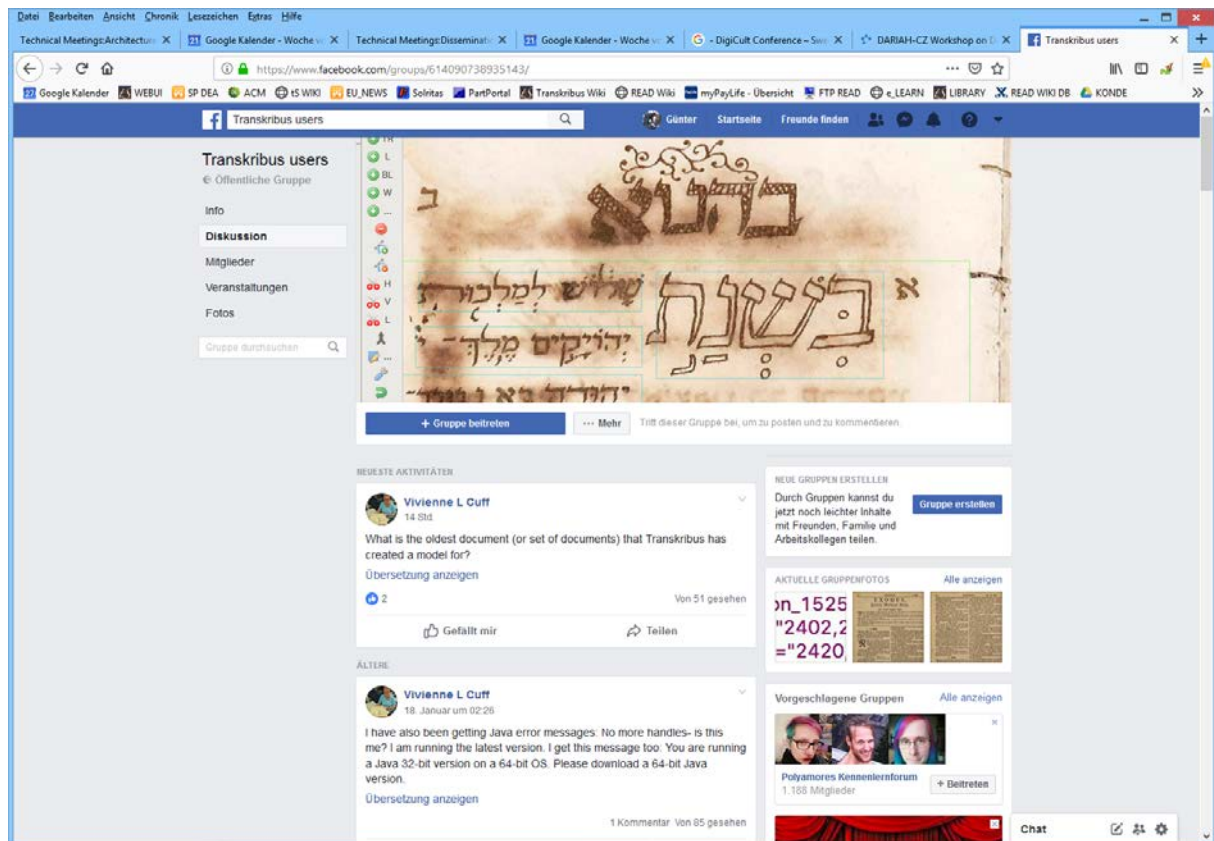


Figure 2 Transkribus User group on Facebook (external)

1.3. Messages and claims for Y4

In the last 6 months of the project the main message for all dissemination activities will be:

Due to the success of the READ project a European Cooperative Society (SCE) is set up in order to further develop and sustain the Transkribus platform

This message has a number of implications:

- READ-COOP will become a **company** – fully dedicated to provide business services for archives, libraries, researchers, the general public but also other technology providers.
- The foundation is not a stopgap solution but there is the willingness to establish Transkribus as a **key player for text recognition of historical documents** worldwide
- Well-known institutions are taking part as **member organisations** and are using Transkribus in order to improve their own service portfolio.
- READ-COOP is a **democratically organised** company – the final decisions are always taken by the General Assembly where every member has a vote.
- **Subscription and processing fees** for the Transkribus platform will be announced during 2019.

Apart from the new key message about READ-COOP we will of course follow our ten main messages as already outlined in Y2:

1. Enjoy the privileges of a **private environment**. The usage of documents in the platform does not include any making-available nor the distribution of the documents to the public therefore copyright regulations are much less important than for public

repositories or digital libraries.¹ Also benefit from the **General Data Protection** laws in the European Community which guarantee a high level of security and trust with respect to the content of the documents uploaded to the platform.

2. Work with **any kind of historical document** and upload thousands or even tens-of-thousands of scanned pages easily to your private collection in the platform. The technology is language and script agnostic and can be used for modern documents as well as for medieval ones.
3. Get independent of the digitisation efforts of archives and libraries by using your own **smartphone** and the **ScanTent** for document digitisation. Upload your files from your smartphone directly to Transkribus and share them with archives and libraries.
4. Benefit from **HTR models which were already trained** on material from other Transkribus users, e.g. run the English Writing Model on 19th and 20th C. documents and receive results on English documents without any further training.
5. **Train the HTR engine** according to your requirements – e.g. to cover a specific hand or a specific way to transcribe (i.e. abbreviations or special characters). Use **existing transcriptions** to create an HTR model in a fully automated way and also retrain the HTR models as often as new material or new technology become available.
6. **Share HTR models** with other users – without infringing copyright or personal rights.
7. Search any recognized document with **Keyword Spotting technology** – a much more powerful way to explore historical documents than with conventional full-text search.
8. **Involve users and volunteers** with a web-interface for viewing, validating and editing documents in a simplified browser environment.
9. **Encourage students and volunteers** to train themselves in reading historical handwriting with the Transkribus eLearning component.
10. **Contribute to the evolvement of research and development** in the domain of historical documents by making data available to computer scientists, namely via open scientific competitions and challenges.

These messages will be the basis for all our dissemination and awareness activities in the last 6 months of the project. Of course they need to be detailed and adapted to specific target groups and services but we believe that they are highly important for all users who want to make their plans and to further benefit from Transkribus services.

¹ Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society

2. Specific actions for Y3

In Y3 we plan four to five specific actions towards disseminating the most important results of the READ project respectively for the Transkribus platform.

These actions are:

- Dutch conference
- Final Transkribus conference
- HTR+
- Training data from and for everyone
- learn.transkribus.eu (eLearning)
- ScanTent and DocScan

2.1. Dutch conference on gathering data

The National Archive Netherlands (NAN) will carry out a two years project with READ-COOP. The project will partly be prepared within the READ project. The main goal is to recognize 2 mill. pages from the so-called VOC collection of NAN.

As an important dissemination activity we will organise a one-day conference on 4th of April. The main goal is to gather researchers and representatives from institutions in the Netherlands to share their already available transcriptions and to make them available for the Transkribus platform. Since the Netherlands are among the most active countries in Europe with regard to digitisation and digital humanities also the number of high-potential users in Transkribus is high. We expect that 20-30 people will come to the workshop.

2.2. Final Transkribus User Conference in Helsinki

We will held our final project meeting in Helsinki from 26th to 28th June 2019. As part of our final meeting we will organise a final conference especially dedicated to the Nordic countries. Due to the high interest we received so far from Finland, Sweden, Norway, Denmark and Island, we expect that we will have 100+ participants.

The focus will be put on the following items:

- Launch of the Large Scale Demonstrator for the National Archive Finland
- Introduction to READ-coop and the business model
- Talks about key technologies and the progress made in READ for Handwritten Text Recognition, Keyword Spotting, Layout Analysis, Document Understanding, etc...
- Use cases especially from Nordic countries and stories to demonstrate how Transkribus is already being used by scholars and archives
- Updates on advances in core tools: Writer Identification, Information Extraction, Table Recognition
- Demonstrations of products: learn.transkribus, ScanTent and DocScan
- Feedback and feature requests

An important aspect of the final conference will be that users of Transkribus get to know each other but also understand that they contribute to the sustainability of the Transkribus platform if they go home and convince their archives, libraries or universities to become member in READ-coop.

Apart from the final conference we already have a number of workshops and conferences fixed for 2019. Among them are:

- Manchester – UK (24th January)
- Passau – Germany (8th February)
- ICARUS – Croatia (27th March)
- ILIDE conference – Slovakia (8th April)
- Zuerich – ETH – Switzerland (25th April)
- EOD conference – Slovenia (4th June)
- ADHO Utrecht – Belgium 2019

2.3. HTR+

In Y2 we assumed that the improvement achieved by the new HTR+ will be between 30-50% of the Character Error Rate (CER). Actually it turned out that in most cases more than 60-70% and in some cases even 80% improvement can be achieved. We have spread this message already at our Transkribus User Conference in 2018 and will of course continue with it in 2019.

As indicated last year, we already contacted more than 25 institutions and retrained some of their HTR models with the HTR+ in order to make them aware of the progress made in the project. We will continue with this information campaign and follow our standard workflow where we

- retrain a HTR model with the new engine
- contact the owner of the training data
- arrange a Skype or Hangout meeting with screen sharing
- demonstrate the new results
- take this as an opportunity to report about other achievements in the platform and
- try to intensify or re-establish the existing contact with the clear offer to join the Transkribus platform respectively READ-coop as a member

Such sessions may be organised as a single meeting or via a series of webinars with several users at once.

A rough calculation shows that about 100-150 institutions need to be contacted for this dissemination activity and about 350 new models need to be trained.

2.4. Training data from and for everyone

Also in 2019 we will follow the path which we have taken the years before and offer not only large institutions but also researchers and scholars the option to train a specific HTR model for them in exchange to transcriptions provided by archives or scholars.

Since the READ project has a dedicated budget for the generation of such training data we are able to make even more institutions and potential users aware of READ and the Transkribus platform by organising a campaign which offers a “deal” to archives and libraries, as well as humanities scholars, researchers but also family historians and genealogists.

This deal can be formulated in the following way:

- Send us all kinds of documents for which you would be interested in text recognition and information extraction.

- READ/Transkribus will create ground truth data, e.g. for Layout Analysis, Table Recognition, HTR or Document Understanding – using the remaining ground truth budget from the project
- This training data will be the basis for automated processing, but they will also be shared between the content providers and the Transkribus platform members

Since Transkribus is already well-known in countries such as Austria, Belgium, Finland, France, Germany, Ireland, Luxembourg, Netherlands, Norway, Sweden and UK we will first and foremost approach the following countries with this campaign:

- Bulgaria, Czech Republic, Croatia, Estonia, Greece, Hungary, Italy, Lithuania, Latvia, Malta, Poland, Portugal, Rumania, and Spain.

For this purpose we will contact the archives associations in these countries as well as use our address list for history departments in Europe to directly approach the two most important target groups.

2.5. lern.transkribus.eu

In Y3 the lern.transkribus.eu application was rebuilt in order to make it more attractive to users. We also proceeded with our investigations in potential users for the application and can now boast to have nearly 3000 addresses available for dissemination activities.

In the last 6 months we will (and already have) started a campaign by sending out letters with information material to the above mentioned 3000 addresses. We are now able to offer more than 40 documents in several languages and several scripts. We will strongly work with instruction videos, both for students or volunteers as well as for scholars who want to include specific documents in the Transkribus learn collection.

2.6. **Market ScanTent and DocScan**

In the last six months of the project we will also set a focus on preparing the market launch of the ScanTent. This is now planned for the time after the READ project, but will be prepared carefully within the READ project. About 300 pieces will be sent out to early adopters who contacted us in the last months. The two target groups are humanities scholars who are working professionally with archival material and have access to resources, such as project grants or students working in their team.

The other target group are archives and libraries which want to provide a cost-effective alternative for users who want to take images from their archival holdings.

Until summer 2019 we expect that more than 20.000 users will be registered in Transkribus. These users will then be our first addressees for the market launch of the ScanTent. They will receive not only information on the ScanTent via a personal email, but also get links to videos showing how the ScanTent can be used at home for the digitisation of personal documents, but also in archives and libraries. We will also explain the idea that the ScanTent enables volunteers to easily contribute to the digitisation of archival collections via Scanathons and similar events.

3. General dissemination and awareness channels

3.1. READ website

<http://read.transkribus.eu>

In Y1 this website was set up and in Y2 and Y3 the website was maintained and updated according to the progress in the project. All technical deliverables are available as well as all new MoU partners were included.

Most importantly regular posts were published about progress in the project, about conferences or about new partnerships and short portraits of the “people behind READ”.

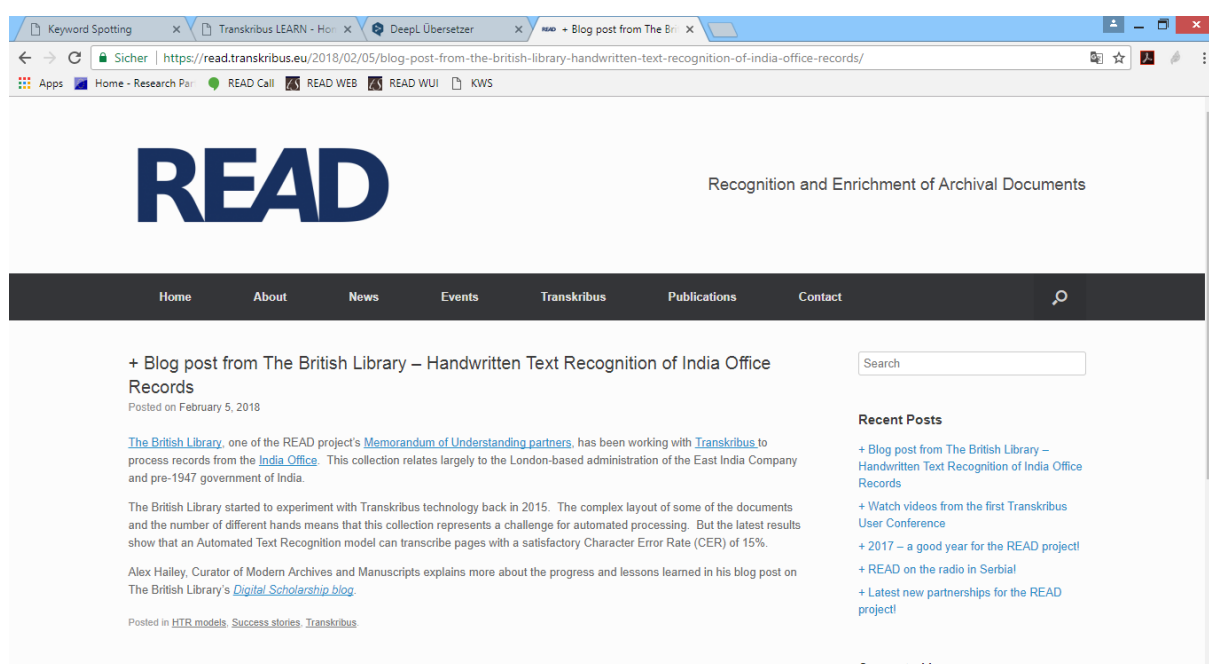


Figure 3 Blog post about a blog post by the British Library

In Y4 this website will be continued but the focus will shift to the new Transkribus website (see below).

3.2. Transkribus website

<http://transkribus.eu/>

This website was set up already before the start of READ. In Y3 we have already created a new version of the Transkribus website but not published so far. The current site is clearly outdated and needs to be adapted to the current status of the Transkribus platform. Very likely we will use a bootstrap template to construct the site in a modern and fresh way.

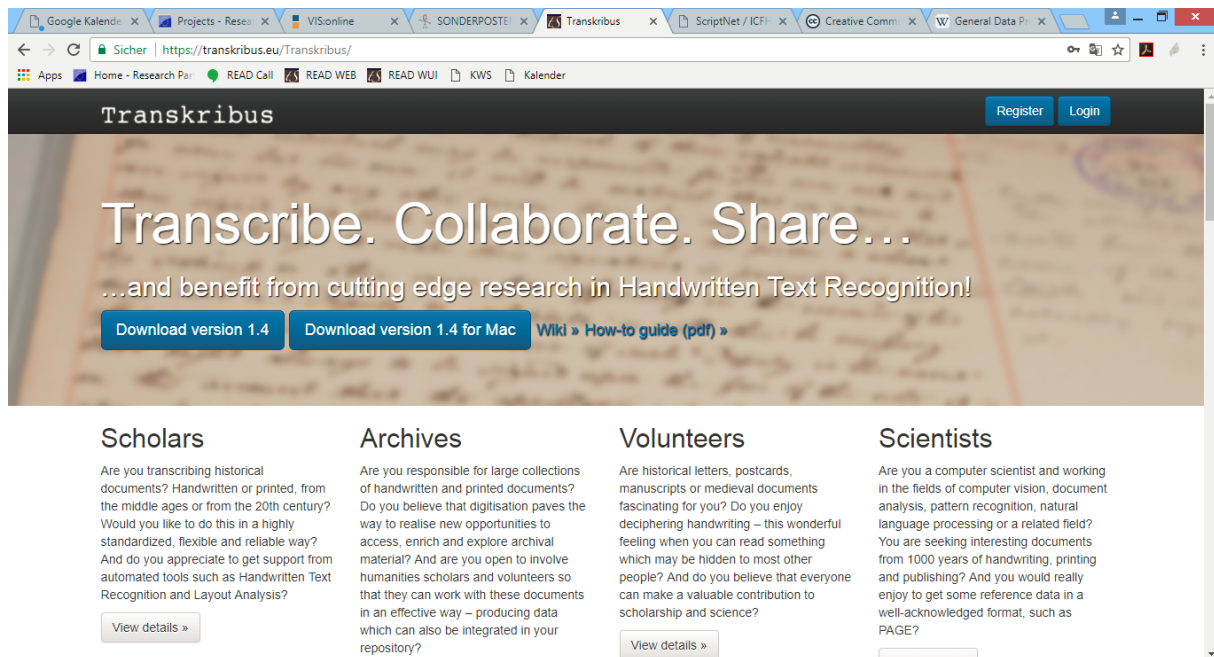


Figure 4 Transkribus website

Especially the corporate structure of the Transkribus platform shall be emphasized on the new website. Here we have to find a way to provide a good overview on the many activities going on in the READ project but also in the Transkribus platform as there are:

- Transkribus Expert client
- Transkribus document library
- ScanTent/DocScan
- learn.transkribus.eu
- ScriptNet
- FamousHands

The following new features which are dedicated to the idea of sharing and collaborating will be implemented in this website as well:

- An overview of HTR models which are available to the public. The HTR model will be described by the users who created it but also contain some example pages of the dataset so that everyone interested can easily see on which scripts the model was trained.
- An overview of research projects working with Transkribus. With the growing number of users more and more overlapping of research fields can be observed. E.g. there are several groups working on Latin documents from the middle ages, or several groups working on WW2 documents. In many cases an exchange of information or data may be of benefit for each of the groups. Of course the MoU partners will be the first addressees for this list.

The site will of course also contain usual elements such as blog posts, news, HowToPapers and links to the YouTube channel.

3.3. Transkribus web-interface – library

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

Due to technical difficulties encountered we followed a completely new concept and focused on the most important features which were to enable users to transcribe, annotate and correct the layout within a browser application.

[illegible]

In Y4 we will put a strong emphasis on this site and further develop it in strong connection also with the DocScan app.

3.4. Transkribus wiki

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

The Transkribus Wiki was set up in Y1 and also maintained in Y2. Nevertheless in Y3 we will streamline the complete support service for Transkribus users and go more in the direction of the Transkribus “How to Guides” as well as video instructions (see below).

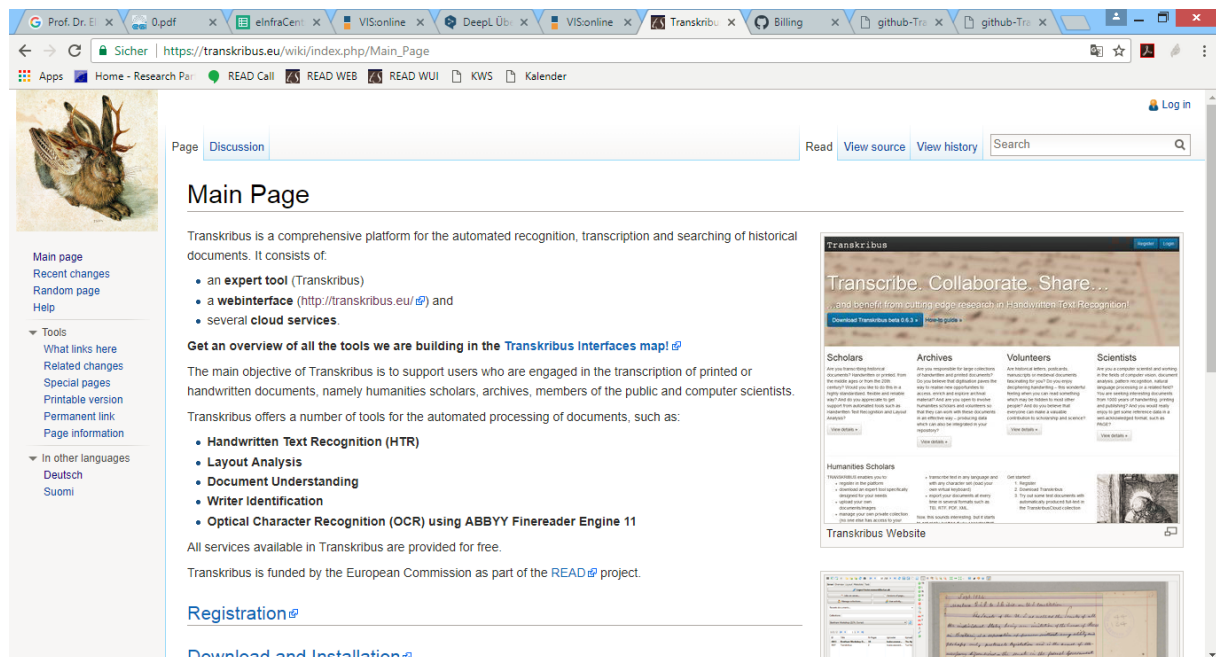


Figure 6 Transkribus Wiki

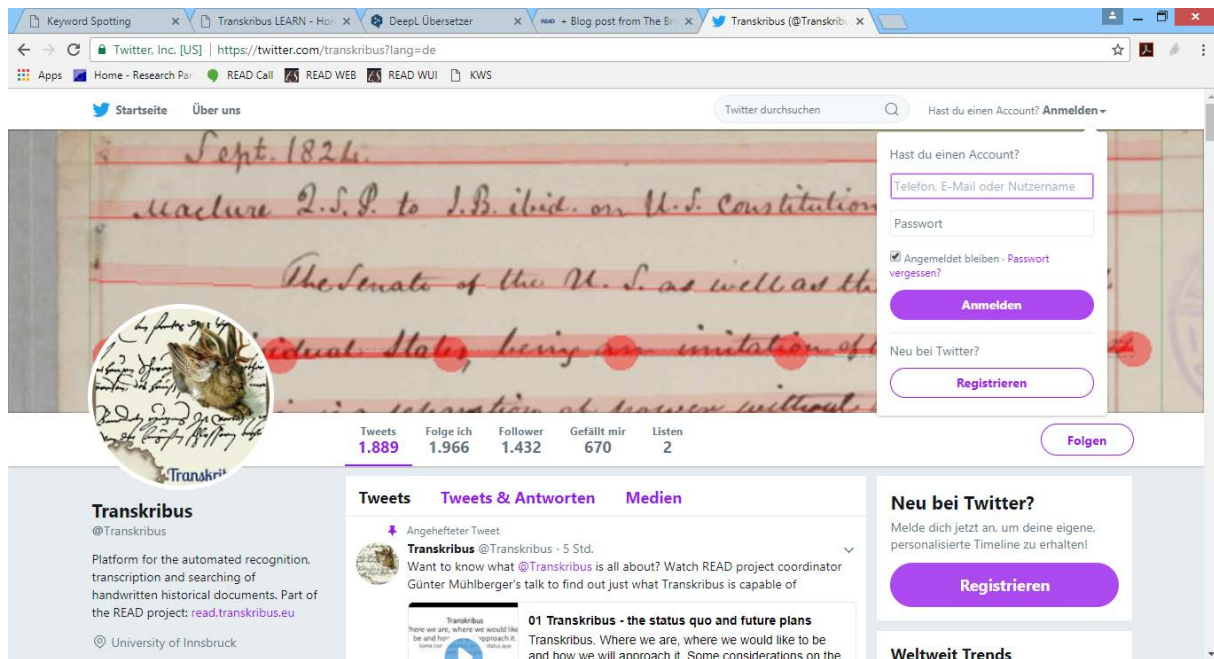
In order to reduce the maintenance effort for the Wiki we will drastically reduce the amount of content made available via this site.

3.5. Twitter

<https://twitter.com/transkribus/>

In the digital humanities domain the use of Twitter is – in contrast to the computer science domain – rather popular and a good means to reach this target group.

UCL, StAZh and CVL were very active on the Transkribus user account. An impressive number can be reported (as of 5th February): 1.889 tweets, 1432 follower and 670 likes. This will be continued in Y3.



3.6. YouTube video channel

<https://www.youtube.com/channel/UC-txVgM31rDTGIBnH-zpPiA>

Already in Y1 all talks from the Kick-off Meeting in Marburg were published as videos on YouTube. Later on a specific YouTube channel was created for Transkribus. Main content are HowTo videos for beginners. This aspect shall be emphasized in Y3: All HowToPapers will be accompanied by short screencast videos.

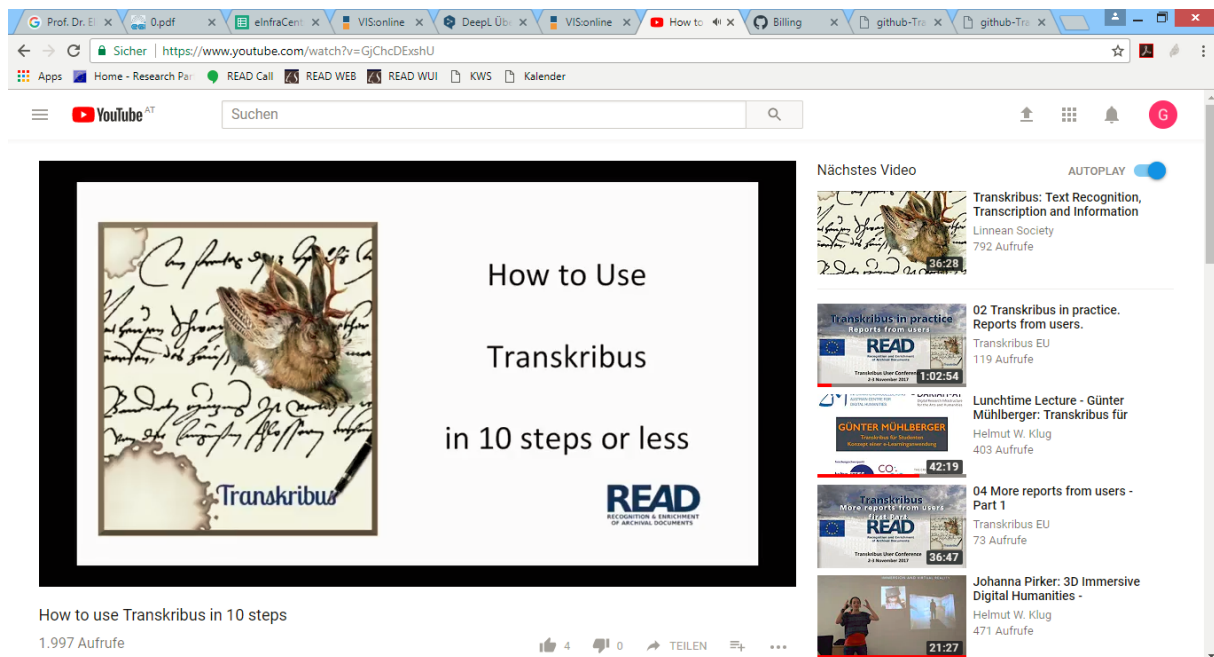


Figure 7 YouTube channel for Transkribus

The 10 steps guide was viewed nearly 2000 times at the YouTube site of Transkribus.

Due to some great support from volunteers we were also able to publish all talks (more than 10h) from the Transkribus User Conference.

3.7. ScriptNet

<https://scriptnet.iit.demokritos.gr/>

The ScriptNet website is dedicated to organise scientific competitions in the READ project (and above). It was developed in Y1 and further improved in Y2.

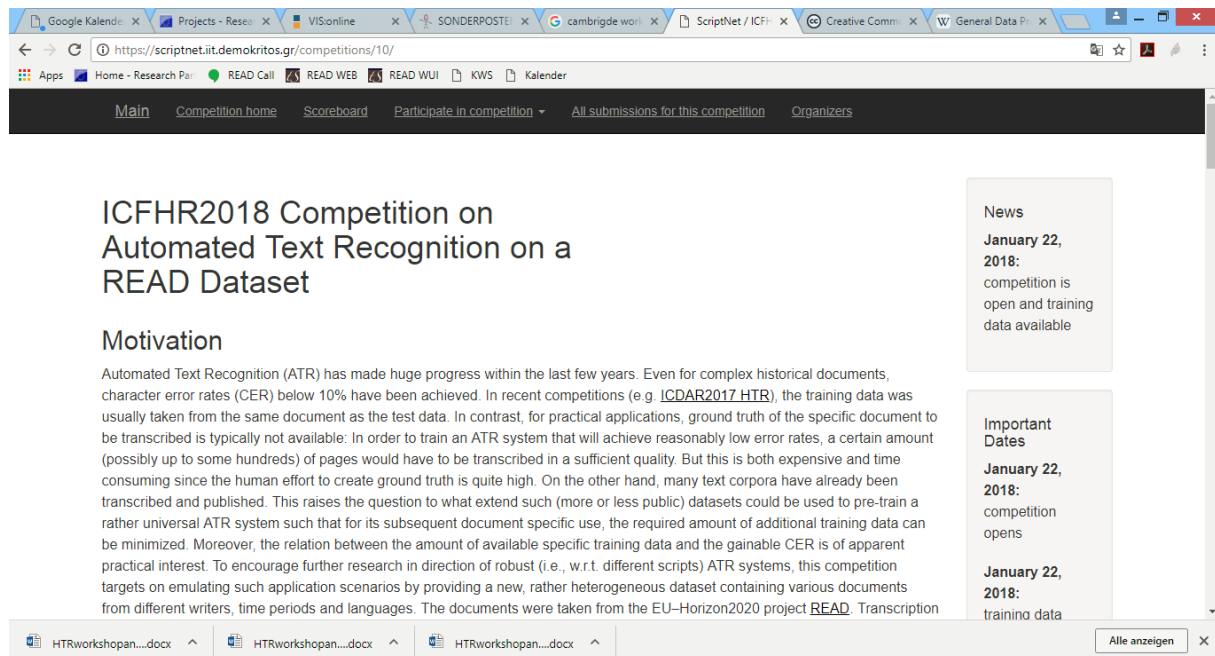


Figure 8 ScriptNet site with ICFHR 2018 HTR competition site

Five competitions were organised in 2017, and for 2018 already an HTR competition was accepted by the International Conference on Frontiers in Handwriting Recognition (ICFHR).

The target group for this site are computer scientists but also archivists and libraries who want to get an impression how their documents are processed as part of a scientific competition. This cross domain collaboration was also emphasized by respective blog posts on the READ website. In Y3 this aspect shall become even more important.

3.8. ZENODO

<https://zenodo.org/communities/scriptnet/>

In strong connection with ScriptNet and the concept of Open Research Data the datasets created and used in READ are – as far as owners agree – made available publicly via the ZENODO repository. This process was initiated in Y1 and became “routine” in Y2. Also a specific ScripNet – READ community was created.

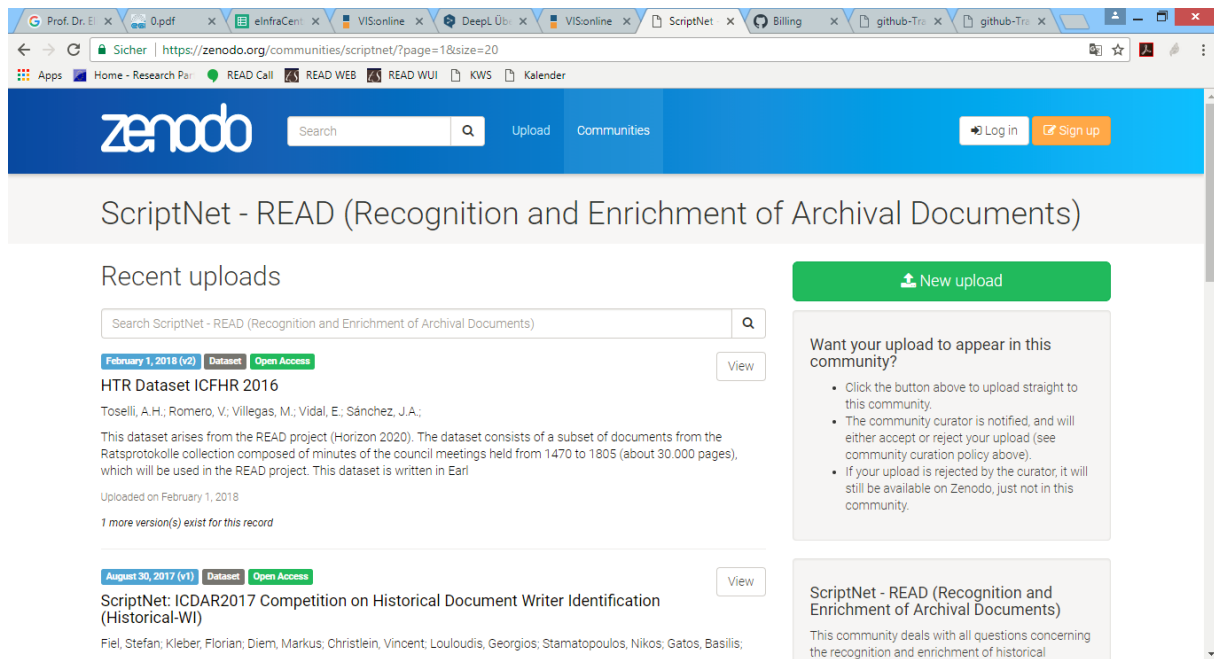


Figure 9 ScriptNet - READ datasets at ZENODO

In Y3 we will continue with this work but are also considering to make it even more convenient for Transkribus users to include their data as datasets in ZENODO. From a technical point of view it would be a doable task to use the ZENODO API directly for this purpose and to include a “ZENODO Export” directly in the Transkribus platform. The complete package including metadata could be directly sent to ZENODO and be made available via ScriptNet-READ. The implementation of such a service will depend on the availability of resources.

3.9. ScanTent/DocScan

<https://scantent.cvl.tuwien.ac.at/>

In Y2 CVL and UIBK set up a dedicated website for the ScanTent and DocScan. We did again a soft launch in order to avoid requests and raise expectations too high, but the site serves as a good starting point for those users who became already aware of this development or are part of the testing of the software and the device.

The site contains also two instructional videos how to set up the ScanTent.

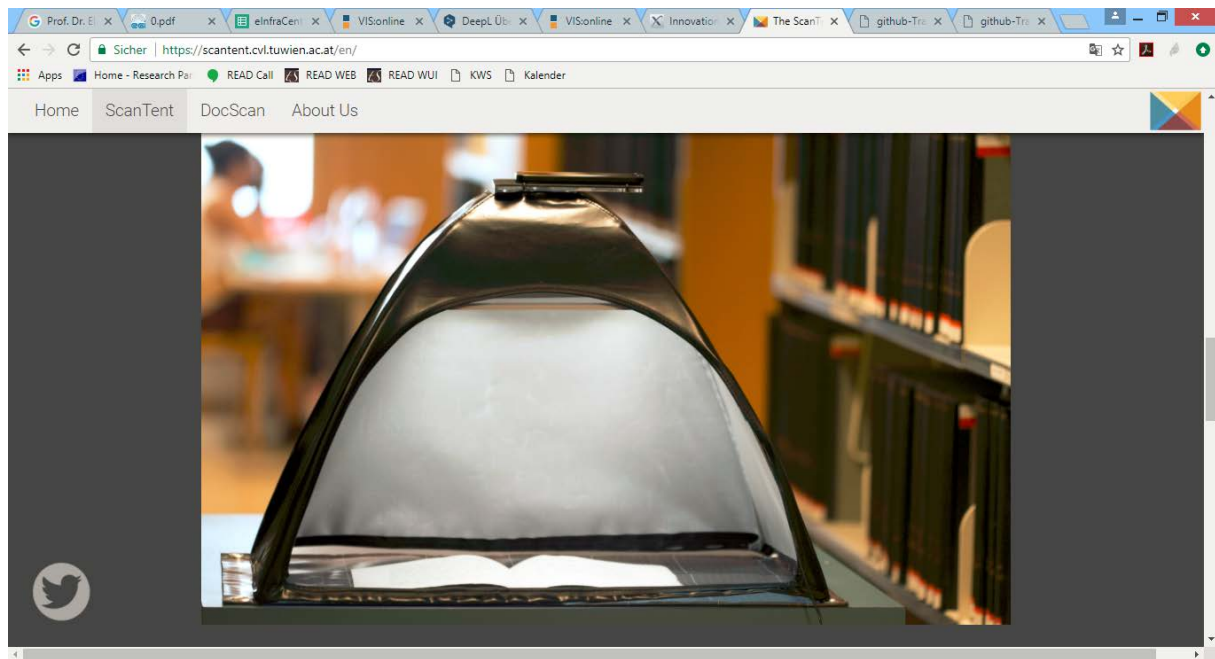


Figure 10 ScanTent website

The site will be adapted according to the progress made in Y3.