

READ

Recognition and Enrichment of Archival Documents

D2.11 Workshops

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	31.12.2016
Actual date of delivery	28.12.2016
Date of last update	15.12.2016
Deliverable number	D2.11
Deliverable title	Workshops
Type	Other
Status & version	Public, Version 1
Contributing WP(s)	2
Responsible beneficiary	UCL
Other contributors	ABP, NAF, StAZH, UIBK
Internal reviewers	Günter Mühlberger
Author(s)	Louise Seaward
EC project officer	Martin Majek
Keywords	Conferences, Dissemination, Workshops

Table of Contents

Executive Summary	4
1. Workshops.....	4
1.1. Background	4
1.2. Preparation	4
1.3. Format.....	4
1.4. Transkribus workshops organised by READ project members	6
1.5. READ presentations at conferences	7
1.6. Feedback	9
1.7. Promotion	9
1.8. Future workshops and events.....	9
2. Conclusion	10
3. References.....	10

Executive Summary

Over the past year, we have delivered workshops and conference presentations in different European cities to audiences of archivists, scholars, students and computer scientists. During these workshops and presentations, we provide an introduction to the READ project and the Transkribus platform. We teach participants how to use Transkribus and respond to any feedback on the interface. We will continue to organise these events across the life of the project, targeting both existing and new contacts.

1. Workshops

1.1. Background

Workshops are essential to the overall objective of WP2; to disseminate the project tools and results to a variety of potential user groups. We need to communicate directly with the groups who are most likely to be interested in working with the project research infrastructure: collection holders, humanities researchers, computer scientists and members of the public. At the moment workshops focus on teaching people to use the Transkribus expert tool, which is already available for use. UIBK and UCL had previous experience of delivering presentations on Transkribus from the FP-7 funded tranScriptorium project (<http://transcriptorium.eu/>).

1.2. Preparation

Workshops are primarily led by the Dissemination Working Group, which contains representatives from ABP, NAF, StAZH, UCL and UIBK. This group cooperates on the preparation for workshops by sharing presentation techniques and resources like scripts, PowerPoint slides and handouts. The first READ workshops were led by UIBK and UCL in Austria, Germany and the United Kingdom. These workshops were used as a template to organise subsequent events.

1.3. Format

Workshops are typically provided to smaller groups comprising less than 20 participants. In advance of the workshop, participants are advised to visit the Transkribus website (<https://transkribus.eu/Transkribus/>) where they can create an account and install the expert client. They are also asked to upload a few images of historical documents to Transkribus prior to the workshop. Having the opportunity to work with their own documents helps those taking part to recognise how HTR technology is relevant to their own area of interest. If anyone does not have access to images of handwritten manuscripts, they are able to work with test documents already on the platform. Workshop participants bring their laptops along to the workshop so they can work with Transkribus straight away.

The workshops are generally divided into five sections:

1. Explanation of Handwritten Text Recognition (HTR) technology

People who are accustomed to working with digitised documents are well-acquainted with the possibilities of using Optical Character Recognition (OCR) on printed text but are less

aware of Handwritten Text Recognition (HTR) technology. In the first part of the workshop we briefly outline the difference between OCR and HTR. We also show the enormous potential of HTR technology by providing examples of the successful automatic transcription and searching of historical documents from collections which have already been processed in Transkribus. These include manuscripts from project partners UCL and StAZH, as well as the 'Konzilprotokolle' collection of administrative papers from Greifswald University written between 1775 and 1840.

2. Overview of the READ project

In the next section of the workshop, we aim to show how the READ project is making use of HTR technology. We explain the project's objective of increasing the accessibility and usability of digitised archival material through the development and application of cutting-edge technologies. The Transkribus platform is introduced, along with proof-of-concepts of other tools being built by the project, including the crowdsourcing interface, the e-learning app and the mobile scanning app.

3. Presentation on Transkribus

A full demonstration of the Transkribus platform is delivered in the third section of the workshop. The workshop leader shares their screen with the audience via a projector and gives an overview of the different elements of the interface and their functions. Participants are shown how to use Transkribus to create training data for the HTR engines. The process of creating a rich transcript (using tags and mark-up) as the basis for a digital edition is also explained.

4. Working independently with Transkribus

In this part of the workshop, participants are able to try out the functions of Transkribus on their own laptops. They are supported by the workshop leader who explains the different elements of the platform and gives participants the opportunity to practice each function for themselves. The workshop leader circulates around the room in order to offer assistance and answer questions.

The workshop leader demonstrates the following tasks. After each demonstration, participants are given time to practice what they have learned.

- Document management – how to upload, view, save, move and export documents in standard formats (PDF, TEI, docx, PAGE XML).
- User management – how to permit specific users to view and edit documents.
- Layout analysis – how to segment documents to create training data for the HTR engines.
- Transcription – how to create a rich transcript with tags and mark-up.
- HTR – how to apply HTR models to automatically generate transcripts, how to conduct a keyword search of documents, how to assess the accuracy of computer-generated transcripts.

5. Question and Answer

The workshop closes with a Question and Answer session where participants can ask for clarification in relation to any matter they are unsure about.

These five sections explain the basic content which is covered in each workshop. Workshop leaders adjust the content and length of each element to respond to the interests of the audience and the time allotted to the workshop.

1.4. Transkribus workshops organised by READ project members

READ project members have delivered a total of 27 workshops to date, with around 500 people likely taking part. The interest has been so great that we have surpassed KPI-I11, where we set a target of organising 6 workshops in the first year of the project. Some of these workshops were organised as stand-alone events at venues belonging to project or MOU partners. Others were delivered as part of large conferences, regular seminars or alongside the presentation of other digital humanities tools and projects. These workshops have mainly been delivered to archivists and researchers interested in (digital) humanities.

- *Applied Digital Humanities: Editing, Annotating and Visualising*, University of Hamburg, Germany, 21-22 January 2016.
- *Text Recognition in Historical Documents*, University of Rostock, Germany, 9 February 2016.
- *Digital Humanities in German-Speaking Countries*, University of Leipzig, 7-12 March 2016.
- Transkribus workshop, University College London, United Kingdom, 22 March 2016.
- Transkribus workshop, The Linnean Society of London, United Kingdom, 22 April 2016.
- Online Transkribus workshop, University of Innsbruck, Austria, 24 April 2016.
- *Archives and Society Seminar*, Institute of Historical Research, United Kingdom, 17 May 2016.
- Transkribus Workshop, Austrian Academy of Sciences, Austria, 28 June 2016.
- *Philology in the Future: Nordic Seminar*, Munch Museum, Norway, 13-14 September 2016.
- Transkribus workshop, Intraanda (Göttingen), Germany, 13-14 September 2016.
- Transkribus workshop, Civic Archives of Bozen-Bolzano, 15 September 2016.
- *Automatic Text and Structure Recognition as Elementary Technologies for Digital Humanities*, Friedrich-Schiller-University Jena, Germany, 27 September 2016.
- *Demystifying Digitisation: A Hands-On Master Class in Text Digitisation*, University of Antwerp, Belgium, 29-30 September 2016.

- Transkribus workshop, The British Library, United Kingdom, 11 October 2016.
- *Googling through Archives*, National Archives, the Netherlands, 13 October 2016.
- Transkribus workshop, University of Belgrade, Serbia, 19-21 October 2016.
- Transkribus workshop, Swiss National Library, Switzerland, 20 October 2016.
- Transkribus workshop, University of Zurich, Switzerland, 20-21 October 2016.
- Transkribus workshop, The Society of Swedish Literature, Finland, 24 October 2016.
- Transkribus workshop, Somerville College, University of Oxford, United Kingdom, 25 October 2016.
- Transkribus workshop, National Archives, United Kingdom, 7 November 2016.
- *Digital Humanities and Editions*, University of Zurich, Switzerland, 10 November 2016.
- Transkribus workshop, University of Greifswald, Germany, 18-19 November 2016.
- Transkribus workshop, Digital Humanities Academy, Italy, 24-25 November 2016.
- Transkribus workshop, University of Tübingen, Germany, 2 December 2016.
- Transkribus workshop, National Archives, Switzerland, 5 December 2016.
- *Digital Humanities Austria*, Austrian Academy of Sciences, Austria, 5-7 December.

1.5. READ presentations at conferences

In addition to active demonstrations in workshops, we have also disseminated the outcomes of the READ project through formal conference presentations.

The technical partners in the project have presented their research at various conferences during 2016. This section focuses on presentations relating to the Transkribus platform. Other presentations are compiled in D2.1 General Dissemination Report.

READ project members have helped to organise four large events across the course of the past year.

- *Technology meets Scholarship, or how Handwritten Text Recognition will Revolutionize Access to Archival Collections*, Hessian State Archives, Marburg, 19-21 January 2016.

This event was co-organised by the EU-funded co:op project (<http://coop-project.eu/>) which is promoting cooperation between archives and the general public. All of the READ project partners presented a summary of their work. A selection of German archivists then gave an overview of their collections and the READ technical partners produced live demonstrators to indicate how computer systems could respond to some of the issues relating to the legibility and layout of their documents. More than 150 people were in attendance. The attendees were mostly drawn from the archives domain but humanities scholars and computer scientists were also present. Slides and videos of the presentations were made available online, in March and June respectively

(<https://coop.hypotheses.org/category/coop-read-convention-marburg>). The most popular video has been Roger Labahn (URO) on 'Handwritten Text Recognition: Key Concepts' and this has been viewed 335 times (as of 15 December 2016).

- *DocEng Symposium*, Technical University Vienna, 13-16 September 2016.

At this conference, academic and industrial researchers came together to hear papers and participate in workshops. The project coordinator Günter Mühlberger delivered the keynote address about how the Transkribus platform allows Document Engineering to contribute to digital humanities.

- *Automatic Handwritten Text Recognition in Historical Documents*, Passau Diocesan Archives, Passau, 20 September 2016.

ABP is part of a strong network of German archivists and this event allowed us to spread the message about HTR technology among them. This event was a German-language symposium which coincided with a READ project meeting in Passau. An audience of archivists and researchers from Bavaria heard three presentations from READ project partners and an additional talk from the director of the Bavarian state archives. Slides from these presentations were uploaded to the READ website

(<http://read.transkribus.eu/2016/09/29/read-partners-in-passau-update-on-our-latest-meeting/>).

- 'What should be in your Digital Toolbox?', The Linnean Society, London, 10 October 2016.

This was a one-day digital humanities conference organised by UCL and The Linnean Society (MOU partner in READ). There were over 70 attendees including archivists, curators, librarians, researchers and students from the humanities and digital humanities. Three READ project partners presented their research in the morning and the rest of the day was dedicated to showcasing other digital humanities projects from around the United Kingdom. A survey of the participants indicated that many had decided to attend in the hope of learning more about HTR and Transkribus. Slides and videos of these presentations are now online on Vimeo (<https://vimeo.com/album/4245164>).

In addition to these four conferences, READ project partners have also presented to archivists, researchers and students at several other events.

- *6th Library Congress*, Leipzig, Germany, 14-17 March 2016.
- *Transcription collaborative et édition numérique de manuscrits: enjeux, outils et perspectives*, National Archives, Paris, France, 16 March 2016.
- *Annual Meeting of European Botanical and Horticultural Libraries Group*, Edinburgh, United Kingdom, 6-8 April 2016.
- *International Society for Utilitarian Studies Conference*, Catholic University of Lille, France, 6-8 July 2016.
- *Archives and Records Association Conference*, London, United Kingdom, 1-2 September 2016.

- *Texterfassung historischer Dokumente*, Berlin-Brandenburg Academy of Sciences, Germany, 6-7 September 2016.
- *Digital Editions and Mark-Up Language*, Bavarian Academy of Sciences and Humanities, Germany, 21-22 November 2016.
- *Genealogy PLUS: Partnership, Open Access and Machine-Aided Indexing*, Hessian State Archives, Germany, 24-25 November 2016.
- *Persistent Identifier Services for the Humanities*, The British Library, United Kingdom, 9 December 2016.

1.6. Feedback

The Dissemination Working Group discuss their own impressions of the workshops, explaining successes and elements that might be improved. At the moment, we collect informal feedback from workshops and events in the project handbook. The reaction of attendees and questions raised are fed back to UIBK and the Dissemination Working Group in this way. We intend to begin gauging the experience of those participating in workshops more formally in the next year by using surveys.

1.7. Promotion

We have compiled a list of scholarly mailing lists and events webpages in the project handbook that we use to advertise upcoming conferences, workshops and presentations. We also use the READ website (<http://read.transkribus.eu/news/>) and the Transkribus Twitter account (<https://twitter.com/Transkribus>) (508 followers as of 15 December 2016) to publicise events.

We also take steps to disseminate the events beyond the people who were able to be present on the day. We used the Transkribus Twitter account to live-tweet our conferences in Passau and London. Engagement with the latter event led to the conference hashtag trending on Twitter for the London area. Where possible, we arrange for slides and videos of conference presentations to be made available online and showcase these using the Transkribus YouTube channel (<https://www.youtube.com/channel/UC-txVgM31rDTGIBnH-zpPiA>). We also summarise past events with blogs on the project website.

1.8. Future workshops and events

We plan to continue delivering a significant number of workshops and presentations in the next phase of the project. The dissemination work we have conducted this year has extended our network of contacts and given us more possible venues for future events. We need to target also countries such as France, Spain and the Eastern European nations. We also need to start looking further afield, particularly to the United States of America where several large international conferences are held for archivists, humanities researchers and digital humanities scholars. As we have begun to formulate a business plan to sustain the Transkribus research infrastructure beyond the life of the READ project, we will be able to offer business services at book fairs and archives days.

It will be necessary to refine the structure and content of our workshops and presentations to integrate the new tools being built by READ including the e-learning app and the crowdsourcing platform. We will also need to consider the best way of presenting READ and HTR technology to the general public in preparation for launching the FamousHands initiative. The surveys conducted as part of D4.7 indicate that non-academic users may be held back by a limited knowledge of technology.

2. Conclusion

We have organised a large number of events over the past year, where we teach interested parties to use Transkribus and present the possible benefits of using HTR technology. We will continue to work with our existing contacts to arrange workshops and presentations, as well as making efforts to reach new audiences.

We will also concentrate on honing the delivering of workshops over the coming period. This will involve recording and reflecting upon more detailed feedback from participants and thinking about the most effective methods of presenting new tools and communicating with new audiences of genealogists and volunteers.

3. References

[1] Horizon 2020, 2014, 'Communicating EU Research and innovation guidance for project participants', Version 1.0:

http://ec.europa.eu/research/participants/data/ref/h2020/other/gm/h2020-guide-comm_en.pdf

[2] T. Causer, S. Arango, R. McNicholl, G. Mühlberger, P. Kahle and S. Colutto, 2015, 'tranScriptorium. D6.3.2: Evaluation of Public DIA, HTR & KWS Platforms', Public report, tranScriptorium: <http://transcriptorium.eu/pdfs/deliverables/tranScriptorium-D6.3.2-31December2015.pdf>

[3] J. Martin, S. Arango, R. Davis, G. Mühlberger, P. Kahle, S. Colutto, T. Causer and K. Grint, 2014, 'tranScriptorium. D6.2.2: Evaluation of Public DIA, HTR & KWS Platforms', Public report, tranScriptorium: <http://transcriptorium.eu/pdfs/deliverables/tranScriptorium-D6.2.2-February2014.pdf>

[4] A. Mollet, D. Moran and P. Dunleavy, 2011, 'Using Twitter in university research, teaching and impact activities. A guide for academics and researchers', Report, LSE Public Policy Group: http://blogs.lse.ac.uk/impactofsocialsciences/files/2011/11/Published-Twitter_Guide_Sept_2011.pdf