READ

Recognition and Enrichment of Archival Documents

D4.13. Integration of MoU Partners
Report for Period 1

Günter Mühlberger (UIBK)

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	31.12.2016
Actual date of delivery	18.01.2017
Date of last update	18.01.2017
Deliverable number	D4.13
Deliverable title	Integration of MoU Partners P1
Type	Report
Status & version	Final
Contributing WP(s)	All WPs
Responsible beneficiary	UIBK
Other contributors	All beneficiaries
Internal reviewers	
Author(s)	Günter Mühlberger
EC project officer	Martin Majek
Keywords	Networking activities

Table of Contents

Executive Summary

The cooperation and integration of institutions interested in the work of the READ project is one of the success stories of year 1. We were able to conclude 23 (!) Memorandum of Understanding agreements with universities, archives, libraries and memory organizations from all over Europe. Several more institutions are currently in preparation or at the signing process. Together with those institutions already having concluded an MoU in the proposal phase we can count more than 30 institutions cooperating with READ.

This exceeds indeed our expectations by far and indicates that the work done in the project is of highest interest to various organisations hosting historical documents and working with them.

1. Introduction

In 12/2015 UIBK already prepared the template of a Memorandum of Understanding and included it into the Consortium Agreement which was signed by all beneficiaries at the start of the READ project. This Consortium Agreement enabled the co-ordinator to sign on behalf of the consortium the above mentioned MoU (cf. Supplement). This turned out to be an excellent basis for negotiating with interested institutions and made the conclusion of the MoUs in most cases a much simpler task than to draft for each institution a separate contract.

In contrast to this procedure it turned out that the integration of those institutions which already had signed an MoU for the READ proposal and which are mentioned in the DOA was a much harder task. The main reason was that the transfer of 20.000 EUR to these institutions exposed a number of administrative hurdles to the participants which are partly not overcome yet. Nevertheless also in this case several institutions cooperated intensively with the READ project (partly without actually re-signing an MoU) so that they are counted here as well.

2. MoU Partners from the READ proposal

These partners were contacted at the proposal phase for the READE project. They all signed a MoU and are listed in the DOA. For their further inclusion sub-contracting money (20.000 EUR each) are foreseen but due to formal issues only three institutions were already able to benefit from these sub-contracts. The main obstacle are administrative regulations of public institutions.

In contrast to these administrative issues all initial MoU partners participated actively in the project and are providing highly valuable contributions. In the following we briefly describe the status quo and the contributions made by the individual institutions.

2.1. Australian National Library (Australia)

The library staff planned to come to the Kick-off meeting in Marburg, but due to some restructuring of the library services and cuttings it was decided to postpone the participation and to intensive cooperation in Y2 or Y3 of the project.

2.2. Gottfried Wilhelm Leibniz Bibliothek (Germany)

Members of the library took part in several meetings (Kick-off in Marburg, workshops) and are working with the Transkribus platform in order to get acquainted with the HTR technology. The main project at the library is the edition of the correspondence and papers of Johann Gottfried Leibniz, one of the largest remains of a scientist in the world. The MoU is so far not renewed, but in preparation and should be formalized in 2017.

2.3. National Library of Spain – BNE (Spain)

The cooperation with BNE is mainly carried out by the UPVLC team. Together with the Lope de Vega society which concluded an extra MoU with the READ project work is carried out in the sphere of the digital edition of Lope de Vegas papers.

2.4. Digital Humanities Lab (Luxembourg)

Due to the fact that the team and the affiliation of the DH Lab changed it was agreed to postpone the conclusion of a renewed MoU once the work plan and the collection used for the agreement can be defined in more detail. Currently considerations are made to use the budget foreseen for the several institutions from the Netherlands who have concluded an MoU (see below).

2.5. The Linnean Socienty of London (United Kingdom)

The Linnean Society is one of the success stories and a very active MoU partner in the project. The main cooperation work was done by UCL. In September 2016 an excellent public event was organised by the Linnean team in London with nearly 100 attendees from high-level institutions from the UK. Apart from the READ team several other DH projects where introduced and rounded up the programme. Reactions on the event were highly positive. Moreover Linnean prepared training material for an HTR model dealing with Linnean writing. This will form the basis for further cooperation in 2018.

2.6. The Civic Archives of Bozen-Bolzano (Italy)

The Civic Archives of Bozen/Bolzano are one of the most active MoU members. They are supported by the UIBK team. The main objective is to set up a project with volunteers for transcribing the council minutes from the late 18th century. More than 20 people are now actively taking part and are transcribing historical documents with Transkribus. Several public workshops in Bozen and Innsbruck were organized. The project is of great importance for further cooperation activities on a regional level, e.g. with universities and research organisations from South Tyrol and Trentino.

2.7. The Hessian State Archive Marburg (Germany)

Already in beforehand of the start of the project the Hessian State Archive and READ cooperated in organising the Kick-off meeting of READ as a public event in January 2016. This was a three days event with more than 150 participants from over 20 countries, among them visitors from abroad. It is hard to imagine that a better start for the READ project had been possible without the engagement from this MoU partner. In late 2016 Marburg organized a smaller event again with READ participation. Already now there is the generous invitation to

organise either a mid-term event or a final event at the archive. A renewed MoU is in preparation.

2.8. The Munch Museum (Norway

Team members of the Munch museum were very active in promoting the READ project in the Nordic countries (DH conferences), as well as in preparing training material for the HTR engine. Since the Munch museum has some strong experience in crowd-sourcing projects it will be of high interest to compare the solutions provided by the Transkribus platform with the applications used in their former projects. This will be part of the cooperation of the coming year.

2.9. The University and Research Library Erfurt/Gotha (Germany)

The University and Research Library Erfurt is another example of an institution struggling with administrative regulations but providing highly valuable input on a voluntary basis. The library hosts one of the most important collections of oriental manuscripts from the middle-ages which were digitized with support of Qatar and are labelled as "World Heritage". This collection of 500 documents shall be made available via the READ project and serve as a basis for research on historical handwriting in Arabic and similar scripts.

2.10. Music Instrument Museum Leipzig (Germany)

The Music Instrument Museum is part of the University Leipzig and has an outstanding collection of historical instruments. The main research aim is to use Writer Identification and similar technologies to identify the signatures of historical documents and to apply them to certain workshops or craftsmen. In Y1 the preparation of the dataset was carried out.

3. MoU Partners having joined READ in 2016

23 organisations joined the READ project via a Memorandum of Understanding in 2016. In the following we provide short descriptions of their contributions and the area of cooperation.

3.1. Friedrich-Alexander-University Erlangen-Nürnberg - Computer Science 5 (Pattern Recognition)

The Pattern Recognition group (head: Andreas Maier) is very well known in Digital Humanities and related fields. Members of the team have won several competitions in Writer Identification (ICDAR 2013, 2015) and were involved in the design and setup of the Writer Identification competition which shall be launched at the ICDAR 2017. From the MoU:

FAU has access to several medieval digital databases that might be of interest to the READ consortium. Furthermore, FAU is interested in sharing software for the means of automatic writer identification and other pattern recognition methods. As a member of READ, FAU would be interested in access to software and data of other partners of the READ consortium. Moreover, FAU is interested in performing joint research and publications that might result in follow-up research projects. Joining READ, FAU will try to acquire additional third party funding independently to facilitate further joint research projects.

3.2.   Archive of the Technical University Graz (Austria)

The archive is part of an university and as such the material is highly interesting. It is not only representative for the Habsburg Empire with the well-known universities in Vienna, Prague, Budapest, Krakow, Graz, and Marburg, but all universities in Europe have very similar records of their students. The material has been delivered to Transkribus and will be subject for further testing – mainly with the table recognition tools – in Y2 and Y3 of the project. From the MoU:

One part of the historical material, which is relevant for the project, has been scanned in high resolution. These digital data are Matrikeln, Personenstandsblätter and Indices from the years 1828 to 1950 which belong to the inventory of Graz Technology University Archives. Another part of the inventory would be digitized and prepared during the project. A collaboration would be interesting and useful because the enormous amount of time needed to transcribe manually and to make the huge historical inventory of Graz Technology University Archives ready for online research.

3.3.   Huygens Institute for the History of the Netherlands

The collaboration with Huygens Institute was one of most fruitful in Y1 of the project. Not only a great deal of training material was made available to the READ project (approx. nearly 2000 pages) but also the promotion of the project among other Dutch institutions and the invitation for a talk at a national conference for Digital Archives in The Hague were highlights of the collaboration. Y2 will intensify this collaboration by providing HTR models, carrying out further tests with documents consisting of tables and information extraction in general.

Huygens ING researches texts and sources from the past with the aid of new methods and techniques. The Institute champions innovation in research methodology, as well as a better comprehension of Dutch culture and history among a broad public. The institute maintains almost 200 data sources among which source transcriptions are the largest group. For a full overview: https://www.huygens.knaw.nl/tools-en-data/?lang=en. Sources are mostly in (early modern) Dutch but also include languages like Latin, French, English, German, Spanish, Swedish etc. as a reflection of the many international (trade) relations of the Netherlands over the centuries.

Envisioned cooperation involves - but is not limited to - large scale projects like the transcription of the resolutions and decrees of the States-General of the Dutch Republic (1576-1796); transcription of notary records at the City Archives of Amsterdam during the 17th and 18th centuries in the Golden Agents project (press release: https://www.huygens.knaw.nl/wpcontenUuploads/2016/02/NWO-large-grant-Golden-Agents 1.pdf); transcription of large numbers of shipping and transportation records from the Dutch East India Company at Batavia (modern-day Djakarta) in the 17th and 18th centuries and transcription of corpora of intellectual letters in an extension to the Circulation of Knowledge project (http://ckcc.huygens.knaw.nl and http://www. republicofletters. net).

3.4.   Centre for Global Heritage and Development (Leiden University, Erasmus University Rotterdam, Delft University of Technology)

The CGHD delivered highly interesting test material to the project which shall be subject to further tests in Y2 and Y3 of the project.

CGHD and RCE act as initiator of th[INK]back, a project initiative to improve readability of high resolution scanned manuscripts, especially those that result from mass digitization projects. CGHD and RCE regard the improvement of readability as a crucial step in the full text digitization of the content of especially old archives, kept mostly under (sub)tropical conditions and, as a result of these conditions, with substantial archival damage such as ink corrosion, fading, paper discolouration and transparency. CGHD and RCE aim to develop algorithms and online tools to contribute to the digital restoration of scanned manuscripts. Test sets and specific examples of scans of manuscripts that are already obtained from archival institutions for th[INK]back can be shared with READ for research purposes.

In return CGHD and RCE would like to learn from and make use of READ technology, open source, algorithms and benefit from the expertise of the READ consortium and partners. CGHD and RCE regard it to be mutually beneficial to all Parties if the th[INK]back results could be included as an optional plugin into the toolset as being developed by the READ project.

3.5. University Library "Svetozar Markovic" (Serbia)

The University Library has already delivered several manuscripts in Cyrillic language, as well as about 10.000 handwritten index cards from their library catalogue. Moreover the team from Belgrade produced some training data for testing the HTR engine. UIBK was strongly involved in this MoU project and took part in a national workshop with a talk about READ and a workshop about Transkribus. Y2 will bring first results and further testing.

ULSM is central library of University of Belgrade and leading academic library of Serbia. ULSM collections are specific and renowned for historic and archival materials. Important parts of these collections are made up of manuscripts and handwritten materials. Digitized collections of handwritten materials include historically important letters and other materials related to first ru ler of modern Serbian state Milos Obrenovic and other famous public figures, scientists and artist from Serbia in 18th to 20th centuries , handwritten books from 14th till 18th century and handwritten unique copies of first university textbooks in Serbia.

Collaboration with the READ-Consortium is of high interest for ULSM because of the possibility to make these unique materials searchable and because of opportunity to be involved first hand with frontrunners of development and implementation of high end technologies that merge effectively worlds of digital and humanities research.

3.6. Tampere University Folklive Archives

The collaboration with Tampere University was mainly organized by the National Archive Finland (NAF). Currently first training material is produced with Transkribus so that in Y2 tests can start. The WWII diaries are highly interesting material and may be the basis for a nation-wide crowd-sourcing project.

Tampere University Folklife Archives (KPA) (Director of the archive, PhD. Marko Tikka & project) is interested in studying how new digital technologies, e.g. OCR applications can be used to improve research possibilities and methods of large-scale collections of autobiographical material. KPA have 40 OOO letters from World War II Finland and we are interested in experimenting how this material could be searched and managed efficiently and from new perspectives. For example, how to find comparable writers and themes through

word-based search or how to study quantitatively personal writing through word frequencies. So far, this kind of collections has been studied only manually and in small scale. At the moment, we are digitalizing the letters and applying for funding.

3.7.   Letter Edition Barlach 2020 – University of Rostock

The team from the Letter Edition Barlach was strongly supported by the CITLab team from Rostock. Actually the txt2img tool from *Task 7.7. Model for semi- and unsupervised HTR Training* was developed partly with the material provided by the Barlach project. Currently more training data are produced by using already existing transcriptions. In Y2 tests will be carried out for transcribing the Barlach letters with support of HTR.

The Briefedition "Barlach 2020" is an editorial project, in which about 2,000 letters by the artist Ernst Barlach will being transcribed and commented till 2020. The fetters are provided from different institutions, like the Ernst Barlach Stiftung GOstrow or the Ernst Barlach Haus Hamburg, and are collected and processed by the workplace of the Universität Rostock.

3.8.   Alfred Escher Foundation (Switzerland)

The cooperation with the Alfred Escher Foundation is a perfect example how humanities and computer sciences can go together. The foundation has organised the complete transcription of more than 5000 letters of the Swiss politician and technical pioneer Alfred Escher. This edition shall be reused for an open HTR competition (application was done in 12/2016) in the course of the ICDAR 2017 conference. Results from this competition will not only be of interest for scientists in the field, but also for humanities scholars in Switzerland and all German speaking countries.

The Alfred Escher-Stiftung is a charitable foundation in accordance with Article 80 et seq. of the Swiss Civil Code (https://www.alfred-escher.ch). Its head office is in Zurich and it does not pursue commercial aims. The purpose of the foundation is to promote critical historical research into Alfred Escher's life and work, and to communicate the results of this research to a wider public.

The Alfred Escher-Stiftung digitized all 5018 letters to and from Alfred Escher (1819—1882) and provides a scholarly edition (https://www.briefedition.alfred-escher.ch). The letters were collected and recorded from several private and public institutions and archives from Switzerland. A special feature of the edition can be seen in the connection between the transcription and the page image on line level. READ will organize with the provided dataset a scientific competition, which aims to support advances in in the text recognition of handwritten documents.

3.9.   Bibliothèque nationale de France

The cooperation with BNF is highly interesting for the READ project since a number of French organisations are already interested in the project (e.g. University Sorbonne). We plan to organise a workshop in Paris during Y2 of the project. From the MoU:

With more than 3 million documents, Gallica, the digital library of the BnF, offers a great documentary diversity: books, newspapers, maps, photographs... Among these, over 65 OOO

manuscripts of all times and all civilizations are available to the public through digitization. If the BnF has developed the full-text search for scanned printed with OCR, its interest is to offer similar access to the content of these manuscripts. The handwriting recognition technology developed and provided for by the READ project represents a potential solution to be applied to some corpuses as a first experiment. The challenge is to manage with this tool the various scriptures and also to the French language whereas it has been developed on other linguistic backgrounds collections.

3.10. University of Zurich, Historisches Seminar (Switzerland)

The collaboration with University of Zurich was organized by the StAZh from Zurich. Currently test material is prepared for HTR training in Y2. From the MoU:

Collection of about 2'000 missives kept at the Burgerarchiv Thun, written in the 15th and 16th century. The documents are partially transcribed by students (about 100 documents). The transcriptions (on line basis) can serve as ground-truth in order to train HTR (gothic script) and run recognition on the remaining documents. The Partner tests and evaluates Transkribus and HTR in a scholarly environment.

3.11. Institut für Dokumentologie und Editorik – University Cologne (Germany)

The cooperation with I.D.E. was very intensive and comprised not only support in evaluating the Transkribus software, but also in setting up a test project with material from the Itineria nova project from Belgium. Several thousands of transcribed pages were delivered to UIBK from IDE who are supporting partners of the Itineria Nova project. This material was used to create training data for HTR but also as input for experiments carried out in *Task 7.10 Language Models* by the University Leipzig / Allgemeine Sprachverarbeitung (ASV). From the MoU:

The Partner is going to evaluate the READ export website for HTR'ed documents regarding various aspects relevant for digital scholarly editions. The evaluation will particularly apply criteria for reviewing scholarly digital editions as listed and published on the IDE website: http://www.i-d-e.de/publikationen/weitereschriften/criteria-version-1-1/

3.12. University Library Rostock (Germany)

As indicated by the MoU the collaboration with University Library Rostock will be based on material which is currently in the process of digitisation. Y2 and Y3 shall therefore bring some concrete results.

Rostock University Library, founded in 1569, is the oldest university library in Northern Germany and today one of the region's major academic libraries. It holds printed collections of about 2.2 million volumes as well as substantial digital collections. The library's historic collections of more than 300,000 volumes include important book and manuscript collections of national and regional significance, including the libraries of the Dukes of Mecklenburg, Low German language material, early Hebrew and Yiddish prints or the personal papers of a number of important Northern German authors. Rostock University Archive, holding more than 3 running kilometers of archival material from the 15th century onwards, and the Wossidlo-Archive of ethnographic material from Northern Germany are also part of the library. UB Restock runs a number of mass digitisation projects of various historical materials, for example 17th and 18th century prints, music sources, medieval and early modern manuscripts and

archival material. Currently the digitised collections of UB Restock contain about 10,000 items with more than 1,000,000 digital images. Furthermore UB Restock is engaged in several Digital Humanities projects as well as in the development of digital library technologies, particularly MyCoRe, a software framework for digital repositories. Since 2014 UB Restock, together with CITLab Research Group at Restock University and PLANET intelligent systems GmbH, is involved in a research project funded by the German Federal Ministry of Economic Affairs on automatic text recognition in handwritten historical documents. UB Restock is in interested to collaborate with the READ-Consortium for different reasons. First of all we want to expand our successful collaborative work in the field of text recognition together with CITLab as one of the READ-Consortium members. Particularly, we intend to apply the technology to large handwritten collections currently in the process of digitisation, f. e. the correspondence of O.G. Tychsen with many multiscript and multilingual documents or the Diccionario del Espanol Medieval's data archive with medieval word forms and context information on index cards. Furthermore we aim to enhance our digital library infrastructure by connecting our repository services with the Transcribus platform to offer our users new ways of doing research with our digitised collections.

3.13. London Metropolitan Archive (United Kingdom)

The London Metropolitan Archive is currently in the phase of testing the Transkribus GUI for producing training material for first tests with the HTR engine. From the MoU:

London Metropolitan Archives (LMA) is the regional local authority archive for the Greater London area. It holds over 100km of archives which date from 1067 to the present day. The collections cover almost every aspect of life in London, including government, law, crime, medicine and education. LMA holds a wide range of digitised collections. We believe that the digitised minutes of the organisations that have governed London since the medieval period will be of particular interest in the context of the READ project as much of the early material is handwritten. Within these volumes we expect to find a fascinating narrative which will be crucial to our understanding of the development of London and the government of its people. LMA could potentially collaborate in a number of areas including: Contributing scanned documents; producing training data; providing feedback to the user interfaces; providing additional data (e.g. lists of person names, etc.); user testing.

3.14. National Bank of Greece Cultural Foundation (Greece)

This collaboration was initiated by the National Centre for Scientific Research – Demokritos in Athens. It will be of high interest to experiment with Greek handwriting. The actual preparation of the test documents will start in Y2 of the project. From the MoU:

The archives collection of the Hellenic Literary and Historical Archive-Cultural Foundation of the National Bank of Greece (ELIA/MIET) comprises more than 1300 private archives (records of individuals, families, associations, institutions, companies etc.). Among them are included more than two hundred literary archives from the 19th century to the present day. A feature of great importance, among most of the literary archives, is the extant correspondence since it allows the reader to trace and examine the movement and intertwining of ideas and artistic/aesthetic currents, the construction and evolution of the author's image and identity, the interactions between literary communities, the language of private and public address and countless other subjects open to research. The collaboration with the READ Consortium will greatly facilitate research both in house and online.

An initial selection from the archives of ELIA/MIET could comprise: The correspondence between the author Stratis Tsirkas and the literary critic M.M. Papaioannou (c. 130 letters of Papaioannou 1955-1975 and c. 150 letters of Tsirkas 1955-1963) which has been digitized and transcribed while the texts have been thoroughly annotated by professors Miltiadis Pechlivanos (Institut fUr Griechische und Lateinische Philologie Freie Universitat Berlin) and Ioannis Papatheodorou (University of Ioannina, Greece). Tsirkas' long-time correspondence with Lily Tsafitsa (1931-1962) has also been digitized.

Correspondences of high importance that could be also digitized include the following: the whole body of correspondence of the Stratis Tsirkas archive (approximately 2.500 letters of the period 1928-1979), the correspondence (1 842-1879) of the poet Aristotle Valaoritis, the correspondence (1861-1918) of the author Angelos Vlachos with Greek and European intellectuals, the correspondence (1886-1919) of the author, critic and translator Kostantinos Chantzopoulos, the correspondence of Stefanos Pargas, an editor with critical influence on the greek intellectuals of Alexandria in Egypt, mainly through his magazine Grammata (1911-1921), the correspondence (mainly from the 1930's) of the poet and art critic N. Kalas (penname of N. Kalamaris), correspondences (1 890-1940) of demoticists (intellectuals that advocated the use of the demotic language in literature, education and the official state affairs) et. al.

3.15. University Bern – Digital Edition of Joseph Dietrich's Diary (1670-1704)

This project is of great interest to the READ project since it deals with several diaries written by one hand. Images and available transcriptions were already delivered and matched with the image. Early in 2017 an HTR model based on about 600 carefully transcribed pages will be available and create the basis for further testing. From the MoU:

The project aims to transcribe all known diaries (by now 18) that were written by Pater Joseph Dietrich from the monastery of Einsiedeln between 1670 and 1704. The primary goal is to analyze its content regarding the author's weather observations to learn more about the locale climate conditions during this time span. It furthermore wants to make information about the daily life within and outside the monastery public available. The project has access to all known diaries, which are already scanned. We seek collaboration for the transcribing process, since it might help to save resources and to improve Transkribus' underlying algorithms.

3.16. Collection of charters - Christian Oertel (University Erfurt – Germany)

The project by Christian Oertel provides the READ team with the option to deal with medieval charters, a document type which is interesting but still experiments for this type of scripts are missing. The further collaboration will depend strongly on the successful evaluation of a German Science Funds project which shall become available in 2017. From the MoU:

The collection which can be provided by Dr. Oertel consists of approximately 1000 digitized charters of the German and Bohemian king Wenceslaus IV (1361-1419). These charters and a number of transcriptions sufficient to "train" the software Transkribus will be transferred to the READ project. Once the software has learned to read the handwriting it will be used to exploit the rest of the collection and to digitally search the handwritten charters.

3.17. Hochschule Hannover – University of Applied Sciences and Arts, Michael Sutor (Germany)

The collaboration with the Hochschule Hannover comprises an interesting diary from the 1960ies and 1970ies. A first small test set of about 50 pages was already delivered to the READ team and transcribed. First tests confirm that the material is challenging since the script is hard to read and misses distinctive characters. In Y2 we will continue with tests and increase the number of training material. From the MoU:

The partnership is going to help to facilitate and support the transcription of a little more than a dozen handwritten records of former time well-known German filmmaker Heinz Sielmann. Those records are of historical value.

3.18. Biblioteca Nazionale Marciana (Italy)

The manuscript 379 is subject to a research project of the Swiss National Foundation project, HumaReC, on the great trialling New Testament manuscript (Arabic, Greek, Latin, 12th century, Marciana Gr. Z. 11 (379)) led by Claire Clivaz, Head of Digital Enhanced Learning, Swiss Institute of Bioinformatics (Lausanne, CH). A separate MoU is in preparation with this team. The MoU is an excellent example how scholarship, archives/libraries and technology providers (READ) need to collaborate in order to create the best research results. From the MoU:

A large number of manuscripts of the BNM is digitized, in particular with regard to the music and geographical manuscripts. Important are the funds of the Greek manuscripts, Latin and Italian. For the purposes of this project might be interesting to apply to the manuscript Gr. Z. 11 (=379) thirteenth, with the biblical text trilingual (Greek, Latin, Arabic) with notes for liturgical use. It belongs to the library of Cardinal Bessarion, donated in 1468. The library offers free digitization and information on the manuscript, and free use of the images in the project. The images will not be published or inserted in websites without the express agreement of the Library. The images will not be transferred to others, either free of charge or sold. The library will participate in the project results.

3.19. Honourable Society of the Middle Temple (United Kingdom)

This MoU was organized by the University College London. Currently the manuscripts foreseen for testing are collected and will be made available for further testing in early 2017. From the MoU:

The Middle Temple Is one of the four Inns of Court in England. The Middle Temple Archive holds records of the Inn dating back to 1501. Manuscript records of Middle Temple's governing body or Parliament have been digitized and made available via the Inn's website. These records are of significant interest to scholars In a range of disciplines. They are not indexed or machine-searchable. It Is hoped that collaboration with RCAO will result In a section of these digitized manuscript records benefitting from Handwritten Text Recognition and that the resulting searchable transcriptions will greatly improve their accessibility and usefulness.

3.20. Archivio Storico Ricordi (Italy)

Though the MoU with the Archivio Storico Ricordi was set into force not earlier than December 2016 UIBK already received a first batch of digitised letters together with correct transcripts. This has been prepared in Transkribus to be used as training data for HTR recognition. The material is challenging, since the writing comes from carbon copies. Tests in 2017 will reveal

how the standard models for HTR recognition are able to cope with such material. From the MoU:

The Archivio Storico Ricordi (ASR) holds the historical records of Italy's most important music publisher: some 8,000 manuscript scores more than 10,000 librettos, approximately 15,000 letters addressed to Casa Rtcordi by authors and librettists. performers and composers, some 10 OOO stage and costume designs, and over 6,000 photographs posters, drawings, and prints Founded in 1808, for almost two hundred years Ricardi played a central and unique mediating role within Italian musical life This role is documented by an immense trove of letterbooks, a repository of some 600,000 outgoing letters that enables a reconstruction of the firm's business correspondence, shedding light not only on music, but also on the historical social and economic context, as well as on aspects of artistic and everyday life. The ASR wants to gradually make this heritage freely accessible to scholars and general public through its website.

Since this material is mainly manuscript, until now this intention was deemed almost unthinkable due to the daunting need of manually transcribing the letters. Even with resort to appropriate projects of crowdsourcing the dimension of the task looks discouraging for a single organization. The READ project opens new perspectives to the feasibility of such an undertaking.

The ASR started a pilot project digitizing and transcribing 280 letters, totalling some 700 pages, basically written by one same writer, from 1888 until 1909. These letters will be made available to the READ-Consortium in order to allow the necessary experiments and evaluations of the tools and services developed by the project.

3.21. Society of Swedish Literature in Finland

The cooperation was initiated by the National Archive Finland. UIBK already received first images and transcripts which are currently prepared for further tests. From the MoU:

SLS is a scholarly society for the preservation and dissemination of knowledge about Finland-Swedish culture. SLS publishes scholarly literature, maintains archives and libraries, funds research, awards literary and scholarly prizes, and gives grants for scholars. SLS was established in 1885 in memory of the Finnish national poet Johan Ludvig Runeberg. SLS has worked for many years with archival material and transcriptions in critical editions. The READ-Consortium offers tools to provide researchers better search possibilities and tools to work with handwritten archival material. SLS offers as pilot texts material from the Topelius collection and the Edelfelt collection, as well as other collections if needed.

3.22. The Finnish Literature Society (Finland)

Also in this case the National Archive of Finland made the connection and is in close contact with the Finnish Literature Society. Members of the society are already working with the Transkribus platform to create training data. In 2017 this will form the basis for further tests with the HTR engine. From the MoU:

The SKS is a learned society, founded in 1831, which stores, preserves and studies Finnish culture. It is Finland's largest scholarly and non-fiction publisher in the field of the humanities.

The collection comprises approximately 260 letters and drafts of letters from Elias Lennrot (1802-1884), the creator of the Kalevala, medical doctor and professor of Finnish, to Frans Johan Rabbe (1801-1879). The letters and drafts are written in Swedish and date from the

1830s to the early 1850s. The majority is owned and preserved by the Archives of the Finnish Literature Society. A small number by the Finnish Historical Society, deposited to the National Archives of Finland.

3.23. National Library of Norway

The National Library of Norway took part in a workshop organised by the National Archives Finland as part of the DH Conference in Oslo (September 2016). Test data of already existing transcriptions were delivered to UIBK and a first model was trained for the writings of Christiane Koren (27 July 1764 – 28 January 1815). Results were encouraging so that cooperation will be extended in Y2 of the project. From the MoU:

The National Library of Norway has an extensive digitization programme. We are digitizing all Norwegian books, and also private archives, letters and manuscripts in the National Library's collection. Everything without copyright restrictions is made available on our website. We are also transcribing and editing letters and manuscripts, and a collaboration with the READ Consortium would therefore be of interest to us.

4. Supplement

Attached is the template for the MoU as it is part of the Consortium Agreement. The coordinator is entitled by the READ beneficiaries to sign the MoU on behalf of the consortium as long as this template is used.

Memorandum of Understanding

made by and between

the Consortium relating to the Action
entitled "READ (Recognition and Enrichment of Archival Documents)" or "Project",
which is being funded by the European Union under its Horizon 2020 Programme, Project No. 674943, the "READ-Consortium"
represented by its Coordinator, the University of Innsbruck,
Innrain 52, 6020 Innsbruck, Austria, in short "UIBK", Institute of German Studies, Digitization and Digital Preservation Group,
represented by the authorized representative for the Project Mag. Dr. Günter Mühlberger,

and

the [Name of Partner Organization, Address, represented by NAME],
in short [Partner's short name]
the "Partner"
hereinafter, jointly or individually, referred to as "Parties" or "Party"

WHEREAS:

The Parties, having considerable experience in the field concerned, wish to specify binding commitments among themselves and, therefore, set out the terms and conditions for collaborating relationships in regards to READ in this Memorandum of Understanding (in short "MoU").

NOW, THEREFORE, IT IS HEREBY AGREED AS FOLLOWS:

Purpose
The objectives of this MoU are the following:
- The READ-Consortium wants to showcase that the envisaged platform for the automated indexing and recognition of large amounts of archival documents meets the needs of memory organization (archives, libraries, collections), academic and higher education institutions (universities, research organizations) as well of the broad public (volunteers, crowd-users) in Europe and abroad.
- Consequently, the READ-Consortium is highly interested in establishing relationships with such institutions in order to understand the specific needs and requirements arising from the various user groups and their specific needs.
- Moreover, the READ-Consortium is interested in receiving and collecting large amounts of test and sample data (digitized historical documents or parts of it) in order to carry out experiments and to evaluate the tools and services developed in READ.
- Furthermore, READ is designed as an open platform that will allow institutions to take part, to receive support in utilizing the platform for their own purposes or for

implementing own projects and to benefit in general from the developed technology and services as well as the network set up within READ.

The READ-Consortium welcomes [Partner's short name] as a Partner of READ. Partnership shall involve:
- cooperating in the implementation of READ,
- make available to the READ-Consortium and the Project the digitized documents or parts of them from the archival documents described below
- receiving confidential information, including, but not limited to, information produced and/or acquired by the Consortium members either as part of the Project ("Results") or before the Project ("Background").

Description of the Partner and its collection as well as the potential fields of collaboration

(1) For publication on the READ/Transkribus website: Describe in some sentences your (digitized) collections and why a collaboration with the READ-Consortium could be of interest for you/the Partner.
(2) Confidential notes: Describe further regulations etc. This will not be published on the website]

Funding
For the sake of clarity it is agreed among the Parties that this MoU is not a commitment for any funding provided by READ to the Partner specified above and no rights for funding shall be derived out of it. READ and the Partner carry all expenses arising from any activities in connection with this MoU by themselves.

Non-disclosure agreement
As the READ-Consortium members have pre-existing obligations with respect to the confidentiality of Results, Background and confidential information, the Partner shall keep confidential, as indicated below, any Results, Background or other Confidential Information that may be disclosed to the Partner by any member of the READ-Consortium. In addition, Confidential Information may be disclosed to the Partner by other partners who are not members of the READ-Consortium. In this agreement, any information disclosed to the Partner in whatever form or mode of transmission, relating to Results and/or Background and/or any information disclosed by any READ-Consortium member or other READ partner which has been identified as confidential at the time of disclosure, shall be collectively referred to as "Confidential Information" and the party owning or holding rights to such Confidential Information, who shall be entitled to enforce the obligations contained herein, shall be referred to as the "Discloser". To avoid doubt, the READ-Consortium has approved the use of this MoU.

By signing below, the Partner agrees to the following:
– to take all reasonable steps to ensure that all Confidential Information disclosed to the Partner in regards to READ remains confidential during the duration of READ and for a period of five (5) years after the end date of the Project;
– not to become involved in any commercial, manufacturing, scientific, literary or any other exploitation of the Confidential Information, whether alone or in conjunction with another party (by licence or otherwise), or use Confidential Information otherwise than for undertaking the Partner's duties as a Partner of READ and for the implementation of the Project without the written consent of the Discloser;
– not to disclose the Confidential Information either directly or indirectly to any third party without the written consent of the Discloser.

- to return to the Discloser upon request all Confidential Information which has been supplied to or acquired by the Recipient including all copies thereof and to delete all information stored in a machine readable form.
- not to disclose and to keep strictly confidential the Confidential Information received, except to its employees, representatives or agents who need to have access to the Confidential Information for the purpose of carrying out the collaborating relationship in connection with READ. The Partner will inform them about the confidential quality of the information provided and shall ensure that their agreement is obtained to keep it confidential on the same terms as set forth in this MoU before any Confidential Information is disclosed to them. Hence, the Partner shall be responsible for ensuring that the obligations of confidentiality and non-use contained herein will be strictly adhered to and shall assume full liability for the acts or omissions made by its personnel, representatives or agents.

In addition, the Partner agrees that the above obligations of confidentiality and non-use shall not apply in the following circumstances:
- is or becomes publicly available through no action or inaction of the Recipient;
- can be demonstrated by the Recipient to have been already in its possession prior to the disclosure of the Discloser;
- is rightfully obtained by the Recipient from a third party without a breach of such third party's obligations of confidentiality;
- is independently developed by the Recipient without use, access or knowledge, either directly or indirectly, of the Discloser's Confidential Information, as shown by documents and other competent evidence in the Recipient's possession;
- is required to be disclosed by the Recipient to comply with applicable laws or governmental regulations, provided that the Recipient provides prior written notice of such disclosure to the Discloser; or
- is approved in writing by the Discloser for the Recipient to release.

Duration
This MoU shall become effective upon signature ("Effective Date") by the duly authorized representative(s) from the Partner and the Coordinator on behalf of the READ-Consortium. It shall continue in full force and effect until the end of READ (presumably at the end of 2018) or – in the case that the Grant Agreement is prematurely terminated – this MoU shall automatically terminate with exception of any non-disclosure provisions.

Furthermore, a Party shall be entitled to terminate this MoU with exception of any non-disclosure provisions by giving at least a three months written notice in advance.

Miscellaneous
This Agreement shall be governed and construed in accordance with Belgian law and the Belgian courts shall have exclusive jurisdiction over it.

Any ancillary agreements, amendments or additions hereto shall be made in writing and signed by a duly authorized representative of each contracting Party.

In consideration of the invitation to be a Partner to READ, the Partner accepts the terms and conditions set out within this MoU.

Contact details

University of Innsbruck – Coordinator of the READ-Consortium
Günter Mühlberger
Digitisation and Digital Preservation group
Innrain 52, 6020 Innsbruck
Austria / Europe
tel: ++43-512-507-87258454
e-mail: guenter.muehlberger@uibk.ac.at

[Name Partner]
[address]
[tel:]
[e-mail:]

ACCEPTED AND AGREED BY

On behalf of the READ-Consortium:

Name of the representative(s) in capital letters:

GÜNTER MÜHLBERGER

Innsbruck, _____   _____
(place and date)   (signature)

On behalf of the Partner:

Name of the representative(s) in capital letters:

_____

[place], _____   _____
(place and date)   (signature)