

# READ

## Recognition and Enrichment of Archival Documents

### D4.14. Integration of MoU Partners Report for Period 2

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



|                                     |   |
|-------------------------------------|---|
| <b>Project ref no.</b>              | H2020 674943  |
| <b>Project acronym</b>              | <b>READ</b>   |
| <b>Project full title</b>           | <b>Recognition and Enrichment of Archival Documents</b>                   |
| <b>Instrument</b>                   | H2020-EINFRA-2015-1   |
| <b>Thematic Priority</b>            | EINFRA-9-2015 - e-Infrastructures for virtual research environments (VRE) |
| <b>Start date / duration</b>        | 01 January 2016 / 42 Months   |
| <b>Distribution</b>                 | Public  |
| <b>Contractual date of delivery</b> | 31.12.2017  |
| <b>Actual date of delivery</b>      | 29.01.2018  |
| <b>Date of last update</b>          | 29.01.2018  |
| <b>Deliverable number</b>           | D4.14   |
| <b>Deliverable title</b>            | Integration of MoU Partners P2  |
| <b>Type</b>                         | Report  |
| <b>Status &amp; version</b>         | Final   |
| <b>Contributing WP(s)</b>           | All WPs   |
| <b>Responsible beneficiary</b>      | UIBK  |
| <b>Other contributors</b>           | All beneficiaries   |
| <b>Internal reviewers</b>           |   |
| <b>Author(s)</b>                    | Günter Mühlberger   |
| <b>EC project officer</b>           | Martin Majek  |
| <b>Keywords</b>                     | Networking activities   |

## Table of Contents

|  |    |
|--|----|
| Executive Summary .....  | 6  |
| 1. Introduction.....   | 6  |
| 2. MoU Partners from the READ proposal .....   | 6  |
| 2.1. Australian National Library (Australia).....  | 6  |
| 2.2. Gottfried Wilhelm Leibniz Bibliothek (Germany) .....  | 6  |
| 2.3. National Library of Spain – BNE (Spain).....  | 6  |
| 2.4. Digital Humanities Lab (Luxembourg).....  | 7  |
| 2.5. The Linnean Society of London (United Kingdom) .....  | 7  |
| 2.6. The Civic Archives of Bozen-Bolzano (Italy).....  | 7  |
| 2.7. The Hessian State Archive Marburg (Germany) .....   | 7  |
| 2.8. The Munch Museum (Norway) .....   | 8  |
| 2.9. The University and Research Library Erfurt/Gotha (Germany) .....  | 8  |
| 2.10. Music Instrument Museum Leipzig (Germany).....   | 8  |
| 3. MoU Partners having joined READ in 2016.....  | 8  |
| 3.1. Friedrich-Alexander-University Erlangen-Nürnberg Computer Science 5 (Pattern Recognition) .....                                   | 8  |
| 3.2. Archive of the Technical University Graz (Austria) .....  | 9  |
| 3.3. Huygens Institute for the History of the Netherlands .....  | 9  |
| 3.4. Centre for Global Heritage and Development (Leiden University, Erasmus University Rotterdam, Delft University of Technology)..... | 10 |
| 3.5. University Library “Svetozar Markovic” (Serbia) .....   | 10 |
| 3.6. Universitat Autònoma de Barcelona, Bellaterra.....  | 11 |
| 3.7. Tampere University Folkive Archives.....  | 11 |
| 3.8. Letter Edition Barlach 2020 – University of Rostock .....   | 11 |
| 3.9. Alfred Escher Foundation (Switzerland) .....  | 12 |
| 3.10. Bibliothèque nationale de France .....   | 12 |
| 3.11. University of Zurich, Historisches Seminar (Switzerland) .....   | 13 |
| 3.12. BOhisto – Bozen-Bolzano’s History Online .....   | 13 |
| 3.13. Institut für Dokumentologie und Editorik – University Cologne (Germany) .....  | 13 |
| 3.14. University Library Rostock (Germany) .....   | 14 |
| 3.15. London Metropolitan Archive (United Kingdom) .....   | 14 |
| 3.16. National Bank of Greece Cultural Foundation (Greece) .....   | 15 |
| 3.17. University Bern – Digital Edition of Joseph Dietrich’s Diary (1670-1704).....  | 15 |

|       |   |    |
|-------|---|----|
| 3.18. | Collection of charters - Christian Oertel (University Erfurt – Germany).....                          | 16 |
| 3.19. | Hochschule Hannover – University of Applied Sciences and Arts, Michael Sutor (Germany).....           | 16 |
| 3.20. | Biblioteca Nazionale Marciana (Italy) .....   | 16 |
| 3.21. | Honourable Society of the Middle Temple (United Kingdom) .....  | 17 |
| 3.22. | Archivio Storico Ricordi (Italy).....   | 17 |
| 3.23. | Society of Swedish Literature in Finland .....  | 18 |
| 3.24. | The Finnish Literature Society (Finland).....   | 18 |
| 3.25. | National Library of Norway .....  | 18 |
|       | MoU Partners having joined READ in 2017 .....   | 19 |
| 3.26. | Aarhus City Archives (Denmark).....   | 19 |
| 3.27. | Germanisches Nationalmuseum, Nürnberg (Germany) .....   | 19 |
| 3.28. | The „Der österreichische Bibelübersetzer“, University Augsburg (Germany)...                           | 20 |
| 3.29. | Western Michigan University, Kalamazoo (United States).....   | 20 |
| 3.30. | Universitätsbibliothek Tübingen (Germany).....  | 20 |
| 3.31. | Centre de Recerca d’Historia Rural, Universitat de Girona (Spain) .....                               | 21 |
| 3.32. | British Library, London (UK) .....  | 21 |
| 3.33. | Rechtsquellenstiftung des Schweizerischen Juristenvereins, Universität St. Gallen (Switzerland) ..... | 21 |
| 3.34. | Ernst-Moritz-Arndt-University of Greifswald, archive (Germany).....                                   | 22 |
| 3.35. | Department of Culture Studies and Oriental Languages, University of Oslo (Norway).....                | 22 |
| 3.36. | Universitätsbibliothek Basel (Switzerland) .....  | 22 |
| 3.37. | German Studies Department, University of Vienna (Austria).....  | 23 |
| 3.38. | Heidelberg University, Department of History (Germany) .....  | 23 |
| 3.39. | Deutsche Zentralbibliothek für Medizin, Köln (Germany).....   | 23 |
| 3.40. | University Court of the University of Aberdeen, King’s College, Aberdeen (Scotland) .....             | 24 |
| 3.41. | ALMAnaCH. Inria Paris research centre (France).....   | 24 |
| 3.42. | Fernuniversität Hagen, Institut für Geschichte und Biographie (Germany) .....                         | 24 |
| 3.43. | The international Institute of Social History (IISH) Amsterdam (Netherlands) .                        | 25 |
| 3.44. | Geopast Pty Ltd. Project “Postcard Tree” Potts Point (Australia).....                                 | 25 |
| 3.45. | University of Toronto, DEEDS Project, Toronto (Canada).....   | 25 |
| 3.46. | The oldthing R. Pröhm & M. Schrottmeyer Gbr., Berlin (Germany) .....                                  | 26 |
| 3.47. | Arizona Board of Regents, Northern Arizona University, Flagstaff (United States)                      | 26 |
| 3.48. | STELLENBOSSE HEEMKRING, Stellenbosch (South Africa).....  | 27 |

|       |  |    |
|-------|--|----|
| 3.49. | Ecole des hautes etudes en sciences sociales, Paris (France).....  | 27 |
| 3.50. | The Younes and Soraya Nazarian Library, University of Haifa (Israel).....  | 28 |
| 3.51. | Institute of Mathematics and Computer Science, Chisinau (Moldava) .....  | 28 |
| 3.52. | Digital Dinah Craik Project, Calgary (Canada) .....  | 28 |
| 3.53. | Institut für Germanistik, Abt. Mediävistik und Frühneuzeitforschung, Karlsruher<br>Institut für Technologie, Karlsruhe (Germany) ..... | 28 |
| 3.54. | Noord-Hollands Archief, Haarlem (Netherlands).....   | 29 |
| 3.55. | ERC Project NOCEMUS; Early Modern Science and Latin, University Innsbruck<br>(Austria) .....   | 29 |
| 4.    | Supplement .....   | 29 |

## Executive Summary

In Year 2 we were able to continue our success from Y1: More than 30 institutions from all of Europe, USA, Australia, Canada and South Africa signed a Memorandum of Understanding with the READ project.

### 1. Introduction

In preparation of the READ project 10 institutions signed an MoU. In 2016 25 institutions joined the project and in 2017 another 30 institutions joined as a MoU partner. In total we can count 65 cooperating institutions by the end of 2017.

Apart from the fact that so many exciting projects are “behind” these MoUs the feedback from all these partners was highly valuable. We hope to be able to fulfil many of the requests in further updates of the Transkribus platform.

In Y3 we will continue with this effort but we also will need to transform the existing MoU collaborations as well as new ones into the framework of a successor organisation of the READ project. READ coop offers from our point of view an excellent way for all institutions to express their interest in a tangible way.

### 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 GA.

#### 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.

#### 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.

In Y2 UPVLC worked intensively with BNE and also the Lope the Vega Society on this test project. Experiments were carried out namely with Keyword Spotting with probabilistic indexes.

#### **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.

In Y2 to cooperation was intensive and a collection of Second World War Diaries were uploaded to Transkribus. Transcriptions were done by using the Ground Truth budget and first models were trained.

#### **2.5. The Linnean Society 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 were 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 and 2019.

#### **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 18<sup>th</sup> century. In Y1 about 20 people were actively taking part and transcribed 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.

In Y2 more people joined the effort (now about 60) and several hundreds of pages were transcribed by volunteers. Moreover within a meeting with the new director of the archive we decided to organise a workshop for Italian archives in Bolzano or Trento in Y3 of the project.

#### **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.

## 2.8. The Munch Museum (Norway)

In Y1 the 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. In Y2 we had intensive exchange with the Munch Museum. First of all we carried out a first large scale test with the text2image matching tool. Moreover we trained several models. But most importantly the Museum organised in cooperation with the National Archive Norway a two day workshop where participants from Norway, Sweden, Denmark and Scotland took part. As one of the results a closer cooperation with the Bergen City Archive was created.

## 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.

In Y2 we were working closely together with the team from Gotha. Several Arabic documents were transcribed and models were trained. Moreover results were presented at the 33<sup>rd</sup> German Orientalist Conference in Jena 2017.

## 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. This work was continued in Y2.

# 3. MoU Partners having joined READ in 2016

25 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 the 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 17<sup>th</sup> and 18<sup>th</sup> centuries in the Golden Agents project (press release: <https://www.huygens.knaw.nl/wpcontentUploads/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 17<sup>th</sup> and 18<sup>th</sup> 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>).*

In Y2 the cooperation was continued. Several documents from the Golden Age project were transcribed and HTR models were trained. Moreover a meeting took place with the project

leader Charles van den Heuvel in Innsbruck to explore the further possibilities of a cooperation in Y3.

### **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.*

In Y2 parts of this collection were used for further testing the binarisation algorithms by NCSR and DUTH.

### **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.

In Y2 the collaboration was continued with the training of a model in Cyrillic with good results, as well as tests with a library catalogue in several languages. Moreover another workshop was organised by the library in December 2017 in Belgrade which was held by the READ team.

*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 ruler of modern Serbian state Milos Obrenovic and other famous public figures, scientists and artist from Serbia in 18<sup>th</sup> to 20<sup>th</sup> centuries , handwritten books from 14<sup>th</sup> till 18<sup>th</sup> 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. Universitat Autònoma de Barcelona, Bellaterra

In Y2 UPVLC cared about the collaboration and did a number of tests with the documents from Lope de Vega. Moreover team members joined the Transkribus User Conference in Vienna and expressed their interest in Transkribus. It is also considered to contribute to the software development in READ/Transkribus.

*The research group CER-Prolope (founded in 1989) has achieved international prestige in producing critical editions of Lope de Vega's theatrical plays, established on the basis of a broad variety of manuscript and printed texts of the early modern period.*

*As a Associated Partner of READ project, Prolope can offer:*

*Advice in the selection of manuscript documents related to Lope de Vega and to early modern Spanish theater contained in the collection of the Biblioteca Nacional de España, on the basis of their cultural interest, hand-writing variety and complexity. - Electronic files of the texts already edited by Prolope (about 500.000 text-lines), especially those related to BNE manuscript collection of theatrical plays. - Know-how of the material and graphic features of early modern theatrical manuscripts.*

*Assessment on the liability of the transcriptions of these texts made by Transkribus, after its confrontation with the images provided by Biblioteca Nacional de España, always in the context of READ project.*

*Advice, help and collaboration in actions oriented to the diffusion of the procedures, methods and results of READ project based on Lope de Vega's plays*

### 3.7. Tampere University Folkive 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 Folkive 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 40 000 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.*

### 3.8. Letter Edition Barlach 2020 – University of Rostock

In Y1 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 the project decided to work solely with Transkribus in order to finish their German Science Funds project in 2018. Several models were trained and tuned specifically for the MoU partner by URO.

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

### 3.9. 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 was reused in Y1 and Y2 for an open HTR competition (application was done in 12/2016) in the course of the ICDAR 2017 conference. Results from the competition are highly impressive and were published as part of the ICDAR conference proceedings.

*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 the text recognition of handwritten documents.*

### 3.10. 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). The plan to organise a workshop in Paris during Y2 of the project was postponed to Y3. Nevertheless in Y2 the collaboration was intensified with a bilateral meeting between BnF and UIBK. Furthermore the project team for a digital edition of the papers of Michele Foucault was successful with a grant application to the National Science funds and will start to work with Transkribus in Y3 of the project.

*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 65 000 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 corpora 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.

### 3.11. 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. This was actually carried out in Y2 of the project.

From the MoU:

*Collection of about 2'000 missives kept at the Burgerarchiv Thun, written in the 15th and 16<sup>th</sup> 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.*

### 3.12. BOhisto – Bozen-Bolzano’s History Online

For details see above.

*The website BOhisto – Bozen-Bolzano’s History Online (<http://stadtarchiv-archiviostorico.gemeinde.bozen.it/>) makes freely available the innumerable old Minutes of the Municipal council of Bozen-Bolzano, today’s capital of the South Tyrol region in Northern Italy. The manuscripts digitised up to now span from 1470 to 1805 and shed light on the administration, the economy and the townspeople’s life of one of the main urban centres of Tyrol, situated on the most important transalpine route between Germany and Italy. The rich database offered by the archival data is a unique playground to further explore the urban history of a central European area. Therefore, the collaboration with READ and its crowdsourcing project is of great interest to the BOhisto-project, aiming at still wider use of the whole material by the historical sciences, especially for teaching and researching purposes, and providing in the long run a better readability and an in-depth knowledge of the rich Bozen-Bolzano corpus.*

### 3.13. 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).

In Y2 a number of experiments were carried out as part of this collaboration. Mainly experiments with the text2image matching tool and with the training of several models were performed by StAZh. In addition an exchange took place concerning the TEI export of Transkribus. A number of suggestions were raised and shall be integrated in the new releases of Transkribus in Y3.

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/>

### 3.14. 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. In Y2 the digitisation was continued so that in Y3 a more intensive collaboration can be expected.

*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 15<sup>th</sup> century onwards, and the Wossidlo-Archive of ethnographic material from Northern Germany are also part of the library. UB Rostock runs a number of mass digitisation projects of various historical materials, for example 17<sup>th</sup> and 18<sup>th</sup> century prints, music sources, medieval and early modern manuscripts and archival material. Currently the digitised collections of UB Rostock contain about 10,000 items with more than 1,000,000 digital images. Furthermore UB Rostock 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 Rostock, together with CITLab Research Group at Rostock 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 Rostock is 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. Tycho with many multiscrit 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.*

### 3.15. London Metropolitan Archive (United Kingdom)

The London Metropolitan Archive is currently in the phase of testing the Transcribus GUI for producing training material for first tests with the HTR engine. In Y2 this testing was continued.

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.*

### **3.16. 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 started in Y2 of the project. Also a workshop was organised by NCSR in Athens where members of the team took part.

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 für Griechische und Lateinische Philologie Freie Universität 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 (1842-1879) 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 (pennname of N. Kalamaris), correspondences (1890-1940) of demoticists (intellectuals that advocated the use of the demotic language in literature, education and the official state affairs) et. al.*

### **3.17. 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. In Y2 an HTR model based on about 600 carefully transcribed pages was available.

Several experiments were run with the material and ground truth was created in several forms (abbreviated and expanded forms of words).

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.*

### **3.18. 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 was planned to become available in Y2 of the project. The team is still waiting on a final decision on their grant application.

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.*

### **3.19. 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 tests were continued but the result for the automated transcription did not meet the expectations of the MoU partner. 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.*

### **3.20. 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). In Y2 a separate MoU was concluded 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.*

### **3.21. Honourable Society of the Middle Temple (United Kingdom)**

This MoU was organized by the University College London. Currently the manuscripts foreseen for testing were collected. In Y3 tests shall be carried out also with a special view on the chances for keyword spotting.

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 RCO 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.*

### **3.22. 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. A number of tests were carried out in Y2 of the project and several models were trained with good results. The archive also supported the READ archive in providing test material for the next HTR competition in Y3.

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 10 000 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.

### 3.23. Society of Swedish Literature in Finland

The cooperation was initiated by the National Archive Finland. UIBK already received in Y1 first images and transcripts and did some tests and trainings with these collections. Results are satisfying and SLS will continue this collaboration in Y3 of the project.

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.*

### 3.24. 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 also in this case several models were trained and applied to material from the archive.

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.*

### 3.25. National Library of Norway

In Y1 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 and the collaboration was continued in Y2. Here tests were carried out with printed books in Gothic letter. Excellent results were achieved for this type of material.

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.*

## **MoU Partners having joined READ in 2017**

30 new Organisations joined the READ project via a MoU in 2017 up to a total number of 55 current members.

### **3.26. Aarhus City Archives (Denmark)**

In Y2 the archive worked intensively with Transkribus and created some training data for the HTR processing. This shall be realized in Y3. Moreover the archive will contribute to test the web-interface for Transkribus – enabling volunteers to transcribe documents without any further need to “learn” the expert tool.

*The Aarhus City Archives is the municipal archive of Aarhus. In collaboration with digital archivists, the Aarhus City Archives have digitized and marked up a number of central collections, first of all the minutes of the City Council of Aarhus 1868-1997. The next digitization project is the transcription of the so-called parish community protocols of present-day Aarhus Municipality. The Aarhus City Archives also coordinates a digitization/transcription of city council minutes and parish community protocols on behalf of the Danish Organization of Archives.*

*Aarhus City Archives find it important to share data in open formats and is constructing an archives system that integrates digital documents and meta data. From the point of view of Aarhus City Archives a collaboration with the READ-Consortium is expected to produce a better and more agile tool for sharing digitization jobs with volunteers and fellow municipal archive institutions.*

### **3.27. Germanisches Nationalmuseum, Nürnberg (Germany)**

In Y2 a number of tests were carried out with documents from GNM. Especially documents from the 20<sup>th</sup> century created with typewriter were highly interesting and results were convincing.

*GNM is the largest museum of cultural history in the German-speaking region. It collects and preserves cultural, artistic and historical artefacts ranging in date from the dawn of Germanic cultures to the present day. Among its 23 collections the collection of historical musical instruments is one of the most important of its kind in the world. Nearly half of the objects stem from the private collector Ulrich Rück, whose acquisition correspondence is also part of the collection. More than 17,000 documents (handwritings and typescripts) from the time before, during, and after the Second World War build a unique heritage that provides detailed information on provenance, restoration, and pricing of musical instruments, Within an ongoing DFG funded research project these documents have been digitized and get edited for an online research portal. Only by a professional OCR processing a full search function Of all character based information of these documents can be provided*

### 3.28. The „Der österreichische Bibelübersetzer“, University Augsburg (Germany)

The collaboration with this group from Augsburg University was very intensive. A number of HTR models were trained and a clear benefit for the transcription team was achieved. The collaboration is planned as a long-term endeavour, actually the current Academy project is running for another 10 years.

*The aim of the project 'Der Österreichische Bibelübersetzer' is to produce a critical and commented hybrid edition of the texts by the so called Österreichischer Bibelübersetzer. This unknown author translated extensive parts of the bible into German at the beginning of the 14<sup>th</sup> century. The manuscript corpus counts ca. 100 manuscripts and fragments. Due to the planned hybrid edition, all manuscripts and fragments have to be transcribed, including the tagging of all ascertainable codicological attributes. Therefore, a collaboration of 'Der Österreichische Bibelübersetzer' with UIBK promises purposes for both sides; while I-JIBK will receive and be able to collect large amounts of test and sample data, 'Der Österreichische Bibelübersetzer' will increase the workflow-efficiency and also have the purpose of transferring the generated XML-documents directly in the used editor.*

### 3.29. Western Michigan University, Kalamazoo (United States)

This MoU is an excellent example for a Citizen Science project carried out between a university (Kalamazoo) and a volunteer, Frank Sasama, in Germany. Mr. Sasama is a very active user (retired professor of chemistry) and organized the transcription of the remains of Leopold Janussek (1827-1898). About 5000 manuscript pages are handed down. The documents are highly challenging, both from the script, the layout and the image quality. Several tests were carried out in Y2 and will be continued in Y3.

*Only one work has been published on the 900 year history of the monasteries of the Cistercian Order, *Originum Cisterciensium Tomus I* in 1877 by Leopold Janaussek. The work deals with the history of the male masteries of the Order. A proposed second volume on women's houses was not completed by Janaussek. About 7,000 images have been digitized from his research which reflect his work on Cistercian women's houses. The 7,000 sheets of notes are to be transcribed and made accessible to academics and the public in the form of a printed book and in an ongoing digital project of the Internet. Use of READ and Transcribus to test of the Handwritten Text Recognition Technology will considerably reduce the time involved to transcribe the 7,000 sheets compared to manual transcription and allow the project to focus upon the analysis and development of scholarly tools to make the information easy to use for all potential audiences in both the print and online publications. Western Michigan University has been the home of the Center of Cistercian and Monastic Studies for over 40 years and has extensive experience in the production of scholarly resources. The ambitious project of making Janaussek's "lost" work available will be immeasurably furthered by this collaboration.*

### 3.30. Universitätsbibliothek Tübingen (Germany)

In Y2 the MoU was concluded with the objective to carry out some tests on the texts below, mainly keyword spotting shall be explored.

UB/Handschriftenabteilung:

- *Martin Crusius: Tübinger Predignachschrift (in griechischer Sprache), 1563-1604, 21 Bände, (Signatur: MB 19)*

- *Martin Crusius: Tagebücher (lateinisch), 1573-1604, 9 Bände (Signatur Mh 466) (häufig benutzt) und chronologische Aufzeichnungen (lateinisch), 1585-1604, 3 Bände (Signatur: Mh 443)*

#### UB/Universitätsarchiv

- *UAT 2/1a-4/11, 47/1-38: Senatsprotokolle (87 Bände, 1524-1912, 7,5 lfm)*
- *UAT 84: Juristische Konsilien (269 Bände, 1602-1883, 30 lfm)*
- *UAT S 100: Vorlesungsnachschriften (587 Bände, 1779-1982, 7,4 lfm)*

### 3.31. Centre de Recerca d'Historia Rural, Universitat de Girona (Spain)

The CRHR has started working with the documents in Y2 and had also intensive contact with the UPVLC team. This will be continued in Y3.

*The CRHR will provide a collection of images from the books of the Registry of Deeds of Girona, digitized by the Historical Archive of Girona. This documentary series begins in 1769 and ends in 1862. So far, only the first fourteen volumes (around 14,000 images) have been digitized. but the collection will be expanded in the coming years. Thanks to the collaboration with the READ-Consortium, the automatic transcription of this collection, which includes many of the operations carried out by notaries in the region, will address a detailed and thorough analysis of the social dynamics referring to a period of major historical changes.*

### 3.32. British Library, London (UK)

In Y2 BL explored the options for producing training data for automated transcription on their India Office records. Several models were trained and applied to the material. As a result of Y2 a blog post was published by BL in 01/2018. The collaboration will be continued in Y3.

*The British Library is the largest library in the UK and holds one of the largest collections of books and printed documents in the world. The British Library is under-going the process of digitization of its collections to increase accessibility to them. Many digitized documents are images of hand written texts from a variety of collections and the British Library wishes to collaborate to facilitate the development of automated transcribing tools which will assist in fulfilling the British Library's purposes and wider objectives. The British Library will make available certain digital material of its own choice from its collections on such terms as it may determine in its absolute discretion for the purposes of the Project, which materials shall be ones that it generally makes freely available and which are within the public domain ("BL Materials"). During the course of the collaboration the British Library will benefit from access to research tools, in being able to participate in plans for future use for the tools and technology developed and in recognition of its contribution to the Project.*

### 3.33. Rechtsquellenstiftung des Schweizerischen Juristenvereins, Universität St. Gallen (Switzerland)

StAZh is the coordinating partner for this MoU. In Y2 the selection of available and suitable transcriptions was carried out. This will form the basis for further work in Y3.

*The SSRQ is the oldest ongoing scholarly edition project of Switzerland. Since 1898 over 100 volumes have been edited spanning all geographical regions of Switzerland. Since several years, digital editions are being prepared and carried out. The institution unites highly valuable*

*punditry in the fields of standardization and production of high-class editions. The SSRQ will test the tools and algorithms provided by READ and give feedback on enhancements for possible implementations of the products for their activities.*

### **3.34. Ernst-Moritz-Arndt-University of Greifswald, archive (Germany)**

The team from the university archive Greifswald was one of the most active in the READ project. Already in Y1 first models were trained (using the HMM implementation) and this was continued in Y2. Transkribus is now professionally used for transcribing minutes of the university council (Konzilsprotokolle), and around 3000 pages were transcribed with the support of Transkribus. Due to the large amount of data also excellent models were trained. Character Error Rate is usually below 5% - enabling users to transcribe much faster than keying from the scratch. Moreover Greifswald is also one of the first test users for the ScanTent and DocScan application, as well as for the eLearning application. In Y3 the collaboration will be extended also towards keyword spotting of documents.

*The University of Greifswald was founded in 1456. It holds an extensive and valuable collection of historical scripts from the 15th to the 20th Century, which is made accessible for research and teaching through the University's Archive. Using the historic scripts as a basis, the University Archive delivers the ground truth for the training of HTR models and applies these HTR models to provide the digitised scripts as full texts on the internet.*

### **3.35. Department of Culture Studies and Oriental Languages, University of Oslo (Norway)**

In Y2 the team from NFS produced some ground and several HTR models were created. Results were excellent and NFS is now considering how to take further benefit of the models in Y3.

*Founded in 1914, the Norwegian Folklore archives (NFS) has served as a national archive of cultural-historical texts and source materials. Located at IKOS, the collection includes manuscripts and correspondence from well-known cultural figures from nineteenth-century Norway, such as Peter Christen Asbjørnsen and Jørgen Moe. Some of these documents have been digitalized as jpeg-files and are thus made available to the public on our webpage, <http://www.hf.uio.no/ikos/english/services/knownledge/norwegian-folklore>. However, in order to develop research, reach a wider audience and make our collections more visible, cooperation with Transkribus would enable us to learn more about digital tools as well as testing possibilities for crowdsourcing etc.*

### **3.36. Universitätsbibliothek Basel (Switzerland)**

In Y1 and Y2 the UB Basel supported the READ team by providing data for the Writer Identification and Writer Retrieval competition organised by CVL as part of ICDAR 2017. For Y3 the UB Basel will explore further options to integrate HTR and improved OCR as part of their internal workflow.

*UB Basel follows a comprehensive digitisation strategy resulting in a remarkable collection of digitised handwritten and printed documents from several languages and centuries. About 19'000 documents, manuscripts and archive material are already digitised so far. As part of its digitisation strategy UB Basel is also highly interested that these digital collections are made available for researchers and the public for free on the Internet. They are therefore included in the e-manuscripta.ch, e-rara.ch, e-codices and related sites. The availability of*

*handwritten documents not only as images, but also as full-text is a fascinating idea and will make the documents even more interesting to many user groups. The MOU with the READ consortium offers the chance that these collections are in the coming years used for computer science e.g. as test and training collections for scientific competitions on Handwritten Text Recognition or Automated Writer Identification. UB Basel supports therefore such usage by making the images available to the READ consortium. The READ consortium will in exchange cite the data set from UB Basel under the name "UB Basel Data Set" and promote it within the research community. READ partners are also willing to report about of the results of the competitions and tests in a way which is suitable for the UB Basel.*

### **3.37. German Studies Department, University of Vienna (Austria)**

During autumn 2017 the team for the Okopenko edition made their current transcriptions available to the READ team. Tests are currently going, we expect to come up with representative models in 2018.

*Andreas Okopenko's unpublished diaries are part of his estate kept at the Literary Archives of the Austrian National Library. The project "Diaries from the Poet's Estate" (FWF-Project P28344) is running as a cooperation between the University of Vienna (German Department) and the Austrian National Library (Literary Archives). The annotated digital edition of Andreas Okopenko's diaries will be launched on the website of the Austrian National Library. In addition, the most significant parts of the annotated diaries are published as a print-edition. The digitization of the diaries - consisting of more than 10000 pages — was part of the ongoing project. Encoding is following the TEI standard. Working with Transkribus will accelerate the transcription and give us the opportunity to cover larger parts of Okopenko's estate than formerly expected.*

### **3.38. Heidelberg University, Department of History (Germany)**

As indicated below the test was carried out on highly interesting material (police protocols from early 20<sup>th</sup> century). Due to the fact that several hands and just 60 pages of transcription were provided, the HTR models did not reach the level of accurateness as it was expected by the MoU partner. In Y3 the options for keyword spotting will be explored which may be fully sufficient for this type of research.

*Katja Patzel-Mattern is Professor of Economic and Social History at the University of Heidelberg. Current research explores the appropriation of urban spaces in times of sociopolitical change. Funded by a Marsilius Fellowship, Prof. Patzel-Mattern analyses supervisory reports of the political police of Hamburg, written between 1892 and 1910. In order to transcribe these documents handwritten in Sütterlin and Kurrent script, Prof. Patzel-Mattern is interested in a collaboration with the READ-Consortium. Based on a 60-page transcription, the READ-Consortium may develop an HTR model which enables transcription of her collection with the Transkribus software. Other scholars may benefit from these new possibilities to transcribe handwritten German documents dating back to the late 19<sup>th</sup> and early 20<sup>th</sup> century*

### **3.39. Deutsche Zentralbibliothek für Medizin, Köln (Germany)**

Several tests were carried out with the material provided by the MoU partner. The main challenge is to recognize non-character symbols in the Medizinalkalender. Here the creation of ground truth is a real challenge since no Uni-Code characters are available for these signs. The tests will be continued in Y3 based on the new HTR implementation.

*ZB MED — Information Centre for Life Sciences is a non-profit organization with the mission to supply specialized literature and other information in the life sciences. ZB MED hosts the largest stock of life science literature in Europe and provides with LIVIVO, the most comprehensive search portal and with PUBLISSO an open access portal for the life sciences. ZB MED digitizes content in the fields of medicine, health, and nutritional, environmental and agricultural science. ZB MED's digital collections include digitized copies of its own holdings as well as content from other institutions. The majority of this content has no copyright restrictions and is made freely available by ZB MED. Currently ZB MED is seeking to cooperate with the University of Düsseldorf and the READ Consortium in order to digitize the "Reichs-Medicinal-Kalender" (RMK) from 1880 to 1943 allowing to tap into this unique source of information about the German medical and health care system. In this context READ's expertise is highly desirable for achieving high precision optical character recognition in order to transfer information contained in the RMK into a searchable database.*

### **3.40. University Court of the University of Aberdeen, King's College, Aberdeen (Scotland)**

The LACR team sent several documents with TEI transcriptions which were prepared as ground truth for further training. Several HTR models were created. Results were somehow of mixed nature, so that an improvement in the actual transcription was not achieved. Of course the quality of the models is sufficient to search with keyword spotting. Two meetings took place with LACR team members (Oslo and Innsbruck) and the project will be continued in 2018 with more data.

*The Law in the Aberdeen Council Registers project (LACR Project) is transcribing the handwritten registers of the Burgh of Aberdeen, Scotland from the periods of 1398-1511. The registers are hundreds of pages in overall length; at about halfway through, the LACR Project has transcribed some 500,000 words as well as encoding them into hisTEI, which is a TEI format suitable for historical texts. <https://aberdeenreglstersorg/>*

### **3.41. ALMANaCH. Inria Paris research centre (France)**

The MoU was part of a grant application drafted by INRIA, unfortunately rejected. In Y3 the collaboration shall be continued

*ALMANaCH is a joint Inria and EPHE research team whose pluridisciplinary research domain brings together Natural Language Processing and Computational Humanities. In the context of the ECRPER project, the team ALMANaCH wants to digitize a large corpus of French and German personal writings, from the 19th and 20th centuries, focusing on the French-German war periods.*

*The present MOIJ concerns only the documents described in annex, subject to the legal status of their use and publication. The project ECRPER intends to build a larger corpus until 2019, which will be the subject of a later MOIJ*

### **3.42. Fernuniversität Hagen, Institut für Geschichte und Biographie (Germany)**

The Distance University of Hagen will use Transkribus as a tool for their courses on digital history and paleography in Y3 of the project. In Y2 the team members were instructed how to use Transkribus. It is also planned to offer further support for using the eLearning application from Transkribus.



*Das Institut für Geschichte und Biographie der Fernuniversität Hagen führt regelmäßig Lehrveranstaltungen für Studierende des Studiengangs "Bachelor Kulturwissenschaften" durch. Im Rahmen eines geplanten Online-LehrKurses "Digital Humanities und biographische Forschung" soll als eine von sechs Übungseinheiten für Studierende die Transkriptionen von Dokumenten unterschiedlicher Quellengattungen mit Transkribus durchgeführt werden. Der Kurs entspricht einem Seminar an Präsenzuniversitäten und sollte perspektivisch in jedem Semester angeboten werden*

### **3.43. The international Institute of Social History (IISH) Amsterdam (Netherlands)**

The collaboration with IISH resulted in one of the largest training sets available in Transkribus for the Dutch language. Nearly 500.000 words were used for training of a general model with good results on historical Dutch. In Y3 the options how to use keyword spotting will be explored.

*The International Institute of Social History (IISH) is one of the world's leading research institute and archives in the field of social history. Established in 1935, it has collections from all over the world, comprising archives, library and audio-visual material with the thematic emphasis on social and emancipatory movements. The Institute holds over 4,000 archives, more than 1,000,000 printed volumes, and an equivalent number of audio-visual items including posters. The Global Labour History research program examines how work and labour relations developed globally in the last five centuries, and comprises among others projects on slavery, strikes, labour unions and industrialization.*

*The transcriptions made are part of the research project between local debts and global markets. Explaining slavery in South and Southwest Asia 1600-1800 funded by a Veni research grant awarded to Dr. Mathias van Rossum as part of the Innovational Research Incentives Scheme of the Netherlands Organisation for Scientific Research (NWO) (<https://socialhistory.org/en/projects/projects/between-local-debts-and-global-markets-explaining-slavery-south-and-southeast-asia-1600>). The scans of archival material are property of and provided by the Nationaal Archief in the Hague ([www.gahetna.nl](http://www.gahetna.nl))*

### **3.44. Geopast Pty Ltd. Project "Postcard Tree" Potts Point (Australia)**

In Y2 we got a great dataset from Geopast in order to carry out some tests. With the new layout analysis exactly this kind of documents can be processed with high precision. In Y3 we will implement all features of the new LA (mainly rotation scheme) and proceed with the tests.

*The PostcardTree project consists of a high quality digital archive of handwritten and postally used postcards from the past. In collaborating with the READ-Consortium, it is hoped we can contribute to further advancement of HTR.*

### **3.45. University of Toronto, DEEDS Project, Toronto (Canada)**

In order to support a grant application of the DEEDs team we carried out several tests and were supported by the Department of Computerscience of the University of Toronto. The grant application will be decided in 2018.

*DEEDS, which stands for Documents of Early England Data Set, was originally established to analyze the content of medieval English Latin charters, especially those issued between the time of the Norman Conquest in 1066 to the end of the reign of King Edward I in 1307. The chronological range represents the period when 95% of such documents were issued without*

*dates, a custom introduced to England by the Normans and one which was shared with Normandy until that territory was absorbed into the French realm in 1204. It has been a major objective of the Project since the beginning of the current century, in collaboration with the Department of Statistics at the University of Toronto, to develop algorithms leading to the attribution of dates to undated property-transfer documents. This objective has been achieved and can be actively engaged using 'Dater' on the DEEDS website: <http://deeds.library.utoronto.ca>. In 2017, the DEEDS database consisted of nearly 51,000 charters, of which 29K from England, Scotland and Wales, and the rest from French- and German-speaking lands. 20K from England are dated to within a year of issue, sourced from 406 published cartularies or charter collections (approximately 25% of extant cartularies from the region). Additional Latin charter material, including those from the Anglo-Saxon period, is added as means permit.*

*Through collaboration with the READ-Consortium, DEEDS would make its extensive database of transcribed charters readily available to the international research community, while at the same time sharing in, and contributing to, the technical advancement of the Transkribus software. The next major objectives are to obtain digital images of the sources (both original charters where they exist and cartulary copies) from archival repositories, to place them beside their transcribed counterparts on the DEEDS website, and to use them to train Transkribus to read and transcribe a wide range of paleographical styles.*

### **3.46. The oldthing R. Pröhm & M. Schrottmeyer Gbr., Berlin (Germany)**

Similar to Geopast also in this case a test project with post cards will be carried out once the layout analysis is implemented with all features in Transkribus. This will be done in Y3.

*oldthing is an online marketplace that currently offers over 5 million historical postcards. The cards are available in digital form as image files (jpgs). Each card is stored with front (picture) and back (hand written card greetings). Understanding and decrypting the contents of the back pages and making them searchable would be extremely exciting. The challenge is to identify millions of different handwritings from 1880 to the present day. We would like to participate in this exciting project.*

### **3.47. Arizona Board of Regents, Northern Arizona University, Flagstaff (United States)**

A first set of documents was delivered by the Peirce team and first HTR models were trained. This was used as a starting point for another grant application which will be decided in 2018.

*The published writings of the great American polymath Charles S. Peirce run to approximately 12,000 pages, or 24 volumes. His known unpublished manuscripts run to approximately 100,000 pages, held mostly at Harvard and in the National Archives. It would take approximately 100 additional volumes to bring everything into print. Several selective print editions have appeared, as well as a 32-reel microfilm. The main effort currently ongoing is a thirty-volume critical edition being created by the Peirce Edition Project and published by the Indiana University Press. Starting in the 1970s, the edition has so far produced seven volumes. Given the current rate of progress, it will take approximately 100 more years to complete the series. Digital images of the available manuscripts recently have been published online for the first time by the Scalable Peirce Interpretation Network (SPIN) project. Peirce worked for more than thirty years as a philosopher, logician, and scientist. In his published work, he made first-rate contributions to a great variety of disciplines, including gravitational research, astronomy,*

*mathematical economics, cartography, metrology, chemistry, topology, cognitive psychology, and scientific methodology. The majority of Peirce's intellectual output in philosophy, logic, and semiotics has not been transcribed, and there are good reasons to believe that making the massive trove of surviving Peirce manuscripts more readily available in a form that is more readily discoverable and searchable online would be of enormous value to researchers working in a great variety of areas.*

### **3.48. STELLENBOSSE HEEMKRING, Stellenbosch (South Africa)**

The team from Stellenbosch created a good portion of training data and several HTR models were processed. Results for HTR are satisfying. Nevertheless the next step will be to cope with the table structure of the documents described below. This shall be tackled in Y3 of the project.

*The Dutch Vereenigde Oostindische Compagnie, which dominated global trade in the seventeenth and eighteenth centuries, was renowned for its meticulous record keeping. This was also the case at the outpost which was founded at the Cape in 1652 and during its subsequent expansion.*

*Amongst these are: Minutes Of the "Collegie van Landdrost en Heemraden van de Colonies van Stellenbosch en Drakenstein"*

*This "Collegie" was the joint governing body of the first two inland settlements, Stellenbosch, founded in 1679, and Drakenstein, founded in 1687, after the establishment in 1652 of the Cape of Good Hope by the Vereenigde Oostindische Compagnie (VOC) of the Netherlands at the Cape.*

*The minutes cover the period 1691 until 1827 when the British changed the system of government. They document intimate aspects of the colonial expansion under Dutch and British rule including, until it was subdivided, an area up to about 370 km north to 650 km east, up to the Great Fish River.*

*Although the minutes have been consulted by historians, many of the details will remain hidden until made more accessible and computer searchable through transcription.*

### **3.49. Ecole des hautes etudes en sciences sociales, Paris (France)**

A first set of documents was delivered to READ in late 2017 and will be used for training HTR models. However, the images are very challenging (bitonal) due to the capturing quality (microfilm, PDF). We expect therefore that a large amount of data will be needed to produce good results.

*Eugène Wilhelm's (1866-1951) Personal Diary is an exceptional, hitherto unpublished document. This 7,964 pages long diary was written over a period of 66 years. It recounts the unusual destiny of a Franch-German bisexual jurist giving contextual elements on his professional career, his daily life, his intellectual trajectory as well as the overall European political situation between 1885 and 1951. Thanks to the READ-Consortium it should possible to propose the very first transcription of this exceptional and unpublished document in order to realize, in a second step, a critical edition of it.*

### **3.50. The Younes and Soraya Nazarian Library, University of Haifa (Israel)**

The collaboration with the University of Haifa was one of the most intensive ones in Y2. Apart from a workshop held in Haifa in June 2017, also several university courses were working with Transkribus and explored its feasibility for Arabic and Hebrew documents. Also Haifa was among the first institutions which are now using the training interface from Transkribus which will be opened up to everyone during Y3 of the project.

*The Younes and Soraya Nazarian Library at the University of Haifa is one of the largest academic libraries in Israel staffed by a team of motivated professionals offering services supported by advanced technology and information systems.*

*The library initiates and focuses on the development, digitization and preservation of special collections and archives, including manuscripts, documents, pictures and print material in Hebrew, Arabic, Turkish, German and other languages. The collaboration will center on improving OCR and HTR for some of these collections, such as the mostly typewritten IFA-ASAY, the largest collection of Jewish folktales in Israel and the world over, Historical Hebrew and Arabic periodica, Ottoman print material, and historical documents in Hebrew, German and Arabic from the Aba Khoushi Archive, the Spinozeum Archive, and the Druze archive.*

### **3.51. Institute of Mathematics and Computer Science, Chisinau (Moldava)**

The institute is exploring the options for further improving HTR and OCR for Cyrillic documents in Y3 of the project.

*IMSC works in OCR of both printed and handwritten old Romanian texts presented in the Cyrillic script (up to 1862). We use freely accessible sources of scanned texts like Internet collections, manuscript publications with modern transcription, ad hoc scans, etc. Collaboration with the READ-Consortium will be of interest for IMCS as it will give us the access to the newest technologies of optical character recognition and will permit to make available for scientific research and wide audience the old Romanian handwritten texts.*

### **3.52. Digital Dinah Craik Project, Calgary (Canada)**

A first set of documents was sent in late 2017 and in 2018 several tests will be carried out.

*Digital Dinah Craik is a digital humanities project based at the University of Calgary.. Our research team is working to transcribe, encode, and annotate more than 1000 manuscript letters according to the standards of the TEI Text Encoding Initiative in XML. Transcription is the most difficult and time-intensive part of this process, and we are excited to Together, these encoded letters will form the first digital edition of the correspondence of the bestselling Victorian novelist, Dinah Mulock Craik. We plan to analyze these encoded letters to uncover the personal and professional networks that could help a Victorian woman writer maintain a forty year career. Our ultimate goal is to illuminate the question of how to make it as a nineteenth-century woman writer.*

### **3.53. Institut für Germanistik, Abt. Mediävistik und Frühneuzeitforschung, Karlsruher Institut für Technologie, Karlsruhe (Germany)**

Also in this case the MoU was concluded late in the year and further collaboration will be done in Y3 of the project.

*Our digitalized collections consist of fourteen manuscripts of Rudolf von Ems's 'Barlaam und Josaphat' dating from the 13th to the 15th century, defying traditional views of stemmatology and philology. In order to grant open access to this complex matter, we are planning to use up-to-date digital technology.*

### **3.54. Noord-Hollands Archief, Haarlem (Netherlands)**

Though this MoU was concluded in late 2017 already in Y1 members of the the archive took actively part in the READ project and delivered highly interesting material. The collaboration was intensified in 2017 with tests comprising the upcoming table recognition features in Transkribus. The archive made an official announcement of the MoU on their website in early 2018.

*Noord-Hollands Archief in Haarlem, The Netherlands, has a special interest in large scale serial archives. These archives comprise kilometers of documents in our repositories, are information-rich, but can rarely be searched online. Our aim is to reveal these 'unseen' documents to the public, by implying HTR on the so-called indexes made at the time (nineteenth and twentieth century documents), These documents are complex because they are table structured, but are highly uniform qua layout and content.*

*Noord-Hollands Archief holds the archives of the province of Noord-Holland, decentralized national governmental institutions, and twelve municipalities in the area of Kennemerland and Amstel- en Meerlanden. In sum, they contain more than 50 kilometers of documents.*

### **3.55. ERC Project NOSCEMUS; Early Modern Science and Latin, University Innsbruck (Austria)**

NOSCEMUS is one of the most prestigious ERC projects of the University of Innsbruck and will run until 2023. Transkribus will be used as a research repository to store, process, enrich and deliver about 200 modern Latin texts. In Y1 READ provided some support to the ERC grant application, in Y2 several tests were carried out and in Y3 actual work in the project will start.

*NOSCEMUS plans to publish on its homepage (<https://www.uibk.ac.at/projects/noscemus/>) 200 early modern texts of natural science in a machine readable format. To this end, the expertise of the READ Consortium and especially the tool "Transkribus" are an invaluable help.*

## **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)