

Technology meets Scholarship, or how Handwritten Text Recognition will Revolutionize Access to Archival Collections. With a special focus on biographical data in archives

Convention in Marburg, 19 – 21 January 2016

Dear Mr. Hedwig,
dear Günter, dear colleagues and dear friends,

I remember quite well a day in October 2014 when I got an email from a German email-list, because its content was of great interest to archives and archivists: The sender was looking for partners to provide content – this meant digitised archival material. I did not hesitate to contact the sender of the email, Günter Mühlberger; at this time, his possible project called READ – Recognition and Enrichment of Archival Material – had not been chosen to be granted, yet. We started talking about possible cooperations in case of its granting – and discovered something very useful: READ would be a network of highly specialized computer scientists, some of them world leading in their special fields. The aim of the network would be to improve an existing software in order to recognize Handwritten Texts! An absolutely fascinating goal! Therefore the future project leader was looking for archival partners, and that's why he wrote the aforementioned email. On the other hand the Hessian State Archives in Marburg, this is what I told Günter Mühlberger on the telephone, is a member of the International Center for Archival Research (ICARUS), a network of more than 160 archives, museums and universities from all over Europe, from Canada and the United States with the common goal to bring Europe's written heritage to its people – by digitising it! Within ICARUS, a group of 17 partners from 11 countries had been chosen in August 2014 to be funded by the European Union under the project name “co:op – community as opportunity”, which will be presented by Kerstin Muff later on, and this project was to be coordinated in Marburg. To sum up in a very short way: On one hand, a network of computer scientists and private enterprises looking for content and historical and archival know-how, and on the other hand another one composed of archivists and scholars with huge amounts of content in search of ways to manage these quantities and to bring them to the people. To Günter Mühlberger and me this seemed to be a perfect situation, and we very easily came to the conclusion that the Hessian State Archives in Marburg, one of three State Archives of the State of Hesse, should become part of READ in case it would be granted by the European Union. We decided to organise a huge convention together, and this is, obviously, the reason we gladly welcome you here in Marburg.

Our convention is, at the same moment, the Kick-Off-Meeting for the READ partners, and therefore the READ members will not be with us on the following two mornings, but will conduct their Kick-Off-Meeting in the rooms of the Archives School, which is right next to our building.

On behalf of READ and the organising committee of our convention, I thank its director Dr. Irmgard Christa Becker for the hospitality in the near neighbourhood!

As I said, this convention is one of the projects READ and the Hessian State Archives, READ and co:op will conduct together; let me point out another one, which is to teach the software not only to transcribe handwritten text, but also to recognize a certain handwriting as such. For this goal we provided some thousands of scans with the handwriting of the world famous Grimm brothers. Dear colleagues, imagine a software able to transcribe a handwritten text and furthermore able to find its writer in other documents! From our point of view, the keepers of the treasures, there is absolutely no doubt about the fact that this would be a tremendous innovation for everybody who is concerned with handwritten documents, and this Archive is proud to contribute to this really revolutionary project!

The huge interest in this project and its results is shown by the fact that on 17 December, more than four weeks before the start of the convention, we had to close the registration due to security and logistic reasons. But anyway, we welcome more than 160 people from 17 European and 4 non-European countries, while my list contains more than 211 people that would have liked to follow us! The attention it attracts is also demonstrated by the fact that the participants of our convention come from quite a lot of different kinds of archives, libraries, from different types of museums, academies, from quite all kinds of memory institutions.

When you take a look at the programme you may see what we intended to do: We tried to make two worlds meet. Firstly: What is Handwritten Text Recognition? Let's ask the computer scientists from the READ project! The first day will be dedicated to this topic, and since the different presentations belong to one main topic – Handwritten Text Recognition and its presentation to humanity scholars – we planned one Question and Answer Forum at the end of the panel – so, please keep your questions in mind. In the morning session of the second day, the humanity scholars will talk about the topic: Which archival items contain biographical data in an aggregated way, and in which form do these data occur? We invited leading experts in the field of 7 different types of archival items. Furthermore, we want to present to you, and first of all to our foreign guests, archival items you will – due to very bad historical circumstances – find only in Germany. In the afternoon, we swap the roles in comparison to the first day: Now it's the humanity scholars' turn to explain to our READ partners the most important issues of the respective archival items they have dealt with in the morning session. We hope that the computer scientists will have a lot of questions, and this is why there is a little time left for discussion after every short presentation that can be based upon a questionnaire created by the computer scientists and sent to the humanity scholars. The third day's general topic is easier to explain: If we had a perfectly working Handwritten Text Recognition Programme (day one), and if we knew exactly about aggregated biographical data (day two), what could we do with these data? Seven

experts will present their projects, which are based on the work with or the processing of biographical data. The convention, to sum it up, tries to describe the whole process of gaining new knowledge, starting with new methods provided by the computer scientists.

Transkribus, that is the name of the service platform which will integrate the many tools which will be developed and made available in the READ project. Depending on the nature of the handwriting, the scanning quality and the language resources, it is already possible today to recognize handwritten documents nearly as effectively as printed texts. Needless to say that one will still discover a lot of errors, but for searching, data mining and many other applications this opens up complete new horizons. So in a certain sense, let's think big, starting with the Kick-Off Meeting of the READ project and the convention here in Marburg! Before we do so, just let me add some annoying but necessary administrative issues: Organising this convention we tried to welcome as many guests as possible – possible in terms of security and logistics. This is why some issues are inevitable.

Firstly, let me emphasize that all of you are warmly invited to the public reception from 5 pm to 6:30 pm this evening with some finger food and a glass of wine or sparkling wine. But we hope for your understanding that the dinner after the public reception is reserved for the members of both projects, the colleagues of the Hessian State Archives and, of course, for our speakers. Since this is a convention dedicated to make two worlds meet, the guided tours are primarily for the ones who have never had an insight into archival storage facilities, for example our READ partners; for tomorrow we have prepared five guided tours, for 20 persons and 20 Minutes each, four in English and one in German. Please register in the appropriate circulating forms. The guided tours will take place tomorrow, you will find the exact dates with other useful information regarding the internet connection and the Twitter-hashtag at the registration office and next to the door to this hall, too. The tours are one part of tomorrow's afternoon programme: Besides the tours you are invited to have a coffee or to go to one of our conference rooms, which is signposted, to visit the Archive's little exposition of those types of items our colleagues talked about in the morning session. As you see, the convention is being filmed, and we plan to upload the speakers' contributions on the ICARUS' Channel within Youtube. If you do not agree, don't hesitate to let me know.

Dear guests, dear speakers, you have seen our Archive's director, you have seen me and you will see the coordinator of READ, Günter Mühlberger, and within the next days a lot of other colleagues. But let me express my sincere thanks to all those you will not see consciously although we all know that a convention of these dimensions would never be possible without the support of Thomas Brozat, Daniel Buhl, Sabine Dietzsch-Uhde, Christina Dölle, Nicole Enke, Sabine Fees, Mareike Hoff, Helmut Klingelhöfer, Harald Kremp, Marina Laube, Kerstin Muff from ICARUS, Dr. Dirk Petter, Robert Reiter, Kurt Sachau, Matthias Seim, Klaus Schleiter and Dr. Wolfhard Vahl and last but not least the whole storage facilities team of our Archive!