

# READ

## Recognition and Enrichment of Archival Documents

### D3.4. European Hands

Report for Period 1 M1-M12

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>	31.12.2016
<b>Actual date of delivery</b>	29.12.2016
<b>Date of last update</b>	23.12.2016
<b>Deliverable number</b>	D3.4.
<b>Deliverable title</b>	European Hands
<b>Type</b>	Website
<b>Status &amp; version</b>	Final
<b>Contributing WP(s)</b>	WP3, WP4, WP7
<b>Responsible beneficiary</b>	UIBK
<b>Other contributors</b>	UCL
<b>Internal reviewers</b>	Frederic Kaplan (EPFL), Louise Seaward (UCL)
<b>Author(s)</b>	Günter Mühlberger
<b>EC project officer</b>	Martin Majek
<b>Keywords</b>	Marketing, Writer Identification, Crowd-sourcing

## Table of Contents

Executive Summary .....	3
1. Introduction.....	3
2. Concept outline .....	4
2.1. Core approach.....	4
2.2. Marketing background.....	5
2.3. Technical background .....	5
2.4. Technical prerequisites .....	6
2.5. Example entry .....	6
2.6. Outlook.....	8

## Executive Summary

This paper provides an outline of the marketing campaign “FamousHands”<sup>1</sup> which will be launched during 2017. The objective of this campaign is to make the Transkribus platform known among a broad public, but also archives and humanities scholars. In contrast to HTR processing of documents this campaign will put “writer identification and retrieval” in the centre and allow people to contribute directly to improve the service and technology by just providing a digital image of one page written by a famous person. The “digital fingerprint” which is created for the handwriting of a famous person can be used in various ways – mainly to find documents which are distributed among several digital archives from one person.

### 1. Introduction

The original idea of FamousHands was the following:

- organize a marketing campaign to motivate users to transcribe 30-50 pages written by a famous person
- train an HTR model of these training sets
- make the HTR model available to everyone

During the course of the project it turned out that this would be much too ambitious for a broad audience. Training data must be created carefully, otherwise the results of the machine learning process do not provide the expected high quality. It would mean that users are working with the expert tool “Transkribus” which is far too demanding for a broad public. On the other hand transcription and proof-reading are costly – e.g. 20-30 EUR can be calculated for one page.

Nevertheless already in late 2015 the READ team was approached by one of the largest commercial collectors of famous scripts worldwide: <http://brandesautographs.com/>. When talking to the CEO of the company it became obvious that there is a strong community of collectors all over the world who are fascinated by “famous persons” and their writings. These people are willing to pay enormous sums to own an autograph. Of course the “search for autographs” is therefore a lucrative business.

Based on these considerations we changed the direction of our thinking and put writer identification and retrieval, a technology which is anyway developed in the READ project and will be made available via the Transkribus platform into the centre of our considerations. This technology allows us to create a “digital fingerprint” for a writer, to compare this fingerprint with any other page and to compute the distance. Moreover users will get the chance to use a local software tool for trying out this software.

---

<sup>1</sup> The title of the campaign has been changed from „European Hands“ to „Famous Hands“ due to the fact that this title expresses much better the main objective of the campaign.

## 2. Concept outline

### 2.1. Core approach

An important cornerstones of the marketing campaign will be to tell an exciting story about the possibilities of this cutting edge technology. Such a story will consist of the following elements:

- Mozart, Leibnitz, Freud, nearly everyone in former times has left thousands of letters or manuscripts which are today spread over dozens or even hundreds of archives. Still there are many discoveries and such manuscripts achieve remarkable prices at public auctions.
- Apart from this spectacular cases scholars are highly interested in collecting documents which are spread across several archives – this is actually one of their main interests to have a complete set of documents from one (famous) person.
- The new technology enables users to exactly find the needle in the hay, respectively unknown manuscripts of famous persons among thousands or hundreds-of-thousands of digital page images.
- Crowd users can contribute to the progress in digital scholarship by providing at least one example page of the writing of a famous person. So this would be the minimum requirement for a “digital treasure hunt” and would therefore be part of great endeavour to enable access to cultural heritage in a new way.
- Archives are – on the other hand – (indirectly) encouraged to digitise large amounts of their holdings in order to enable the usage of the new technology. We will also work with archives which already have large digitised holdings to help them search their collections for writings by famous people.

Of course several examples will be prepared to showcase exactly the benefit of the technology. E.g. the Alfred Escher Dataset can be taken as a real world archive consisting of hundreds of writers and exact figures for the precision of the Writer Identification and Retrieval software can be provided. Also information on scientific competitions organised via ScriptNet will be available for users.

The main target groups for this campaign are:

- **Public users**
  - o Our main target here is to involve people in a very simple “Citizen Science” project and to make them aware that they can contribute to the advances of Digital Humanities. Public users are often fascinated by historical handwriting as well as famous persons. The campaign gives them the chance to make a meaningful contribution to science and research (“Citizen Science”). The task is simple and can be carried out by searching the internet and copy/paste some data and documents (PDF files).
- **Humanities scholars**
  - o Many humanities scholars are still not aware of the possibilities of state-of-the-art technology in Computer Vision or Pattern Recognition. They may have heard about Handwritten Text Recognition, but may not have heard about Writer Identification and Retrieval. Therefore the campaign is a simple way to raise attention for this technology among this community.

## - Archives/libraries

- The same is true for archives and content holders in general. Their general approach towards digitisation can be summarized as “first of all lets create manual metadata – afterwards lets digitise our holdings”. With technologies such as Writer Identification it can be shown that such a “traditional” approach can be overcome with a more explorative and much simpler approach “Digitise your holdings in the order of the archive – and apply technology to create metadata – e.g. in this case the similarity of different writers and styles.” Also here the campaign shall be a tool to raise these issues in an indirect way. Instead of approaching archives directly, the campaign showcases what is possible.

## 2.2. Marketing background

The concept for our marketing campaign can be described in the following way:

- organize a marketing campaign called “FamousHands” (instead of the original title: EuropeanHands)
- encourage users to provide at least one page image of the writing of a famous person (usually downloaded from a website of an archive or similar)
- enable them to upload this page image (of course also more images are allowed as well) to the Transkribus platform via a simple web-interface
- accept page images only if a minimum of criteria are provided:
  - user needs to be registered in Transkribus
  - page image must be available in sufficient quality (200ppi minimum)
  - page image must come under a Creative Commons license
  - name of the famous person and a link to the Wikipedia (rep. WikiData) entry of this person must be provided
- display all images of all famous hands on the Transkribus website in a convenient way, so that people can browse the collection and see who is already represented in the collection with which examples
- use place of birth or similar information from WikiData to create a map of all famous persons but also a time frame with links to the Wikipedia entry
- make all images available for further research so that also tools from other groups can be used to create “digital fingerprints”
- make the software for creating the digital fingerprints available via GitHub
- explain the usage of the software which would simply allow to load large amounts of images and search for a specific handwriting

Details of the campaign need to be clarified in the first half of 2017 but from our point of view the story is suitable to be told via traditional media, such as newspapers, but also via digital media such as Facebook, Twitter, Instagram or Flickr. Local societies and online forums for history enthusiasts will also be important avenues of dissemination.

## 2.3. Technical background

As outlined in *Task 7.6 Writer Identification and Retrieval* the writing of individual persons can be used to create a “digital fingerprint” which can be compared to any other writings. This digital fingerprint can be used to identify the same or similar writings in a large set of images – independently of any HTR processing. Apart from integrating the technology into the

Transkribus platform it will also be available via the PIE (Page Image Explorer) developed by CVL. PIE enables archives and libraries to analyse large amounts of digitised documents in a convenient way. (Cf. D7.16 Writer Identification and retrieval tool)

## 2.4. Technical prerequisites

There are a number of technical prerequisites which need to be available before this campaign can be started on a broad level:

- Simple login to Transkribus
  - o This is already achieved via a Google+ registration and login button. It has to be extended to Facebook, and other large social networks.
- Simple upload of images via a web-interface
  - o This is already possible and has been developed by NAF.
- Setting up a web-interface specifically for FamousHands
  - o The general basis has been created in 2016 (Django platform)
  - o Nevertheless key features such as the database of writers, the linking to Wikipedia/Wikidata etc. needs to be developed
- Digital Fingerprint technology is available as a first prototype
  - o Integration into the Transkribus platform needs to be done
  - o A database of all fingerprints has to be developed

One of the main challenges will be to prepare the Transkribus / FamousHands website to cope with large amounts of users who will have a look to the site after having read a news article or after having seen a TV programme about the project.

## 2.5. Example entry

This is an example entry for a famous person.

- Transkribus user: [guenter.muehlberger@uibk.ac.at](mailto:guenter.muehlberger@uibk.ac.at)
- Famous person: Adalbert Stifter
- Wikipedia Link: [https://de.wikipedia.org/wiki/Adalbert\\_Stifter](https://de.wikipedia.org/wiki/Adalbert_Stifter)
- Wikidata Link: <https://www.wikidata.org/wiki/Q168542>
- Page image from the Upper Austrian Digital Library:  
<http://digi.landesbibliothek.at/viewer/image/1053/10/>

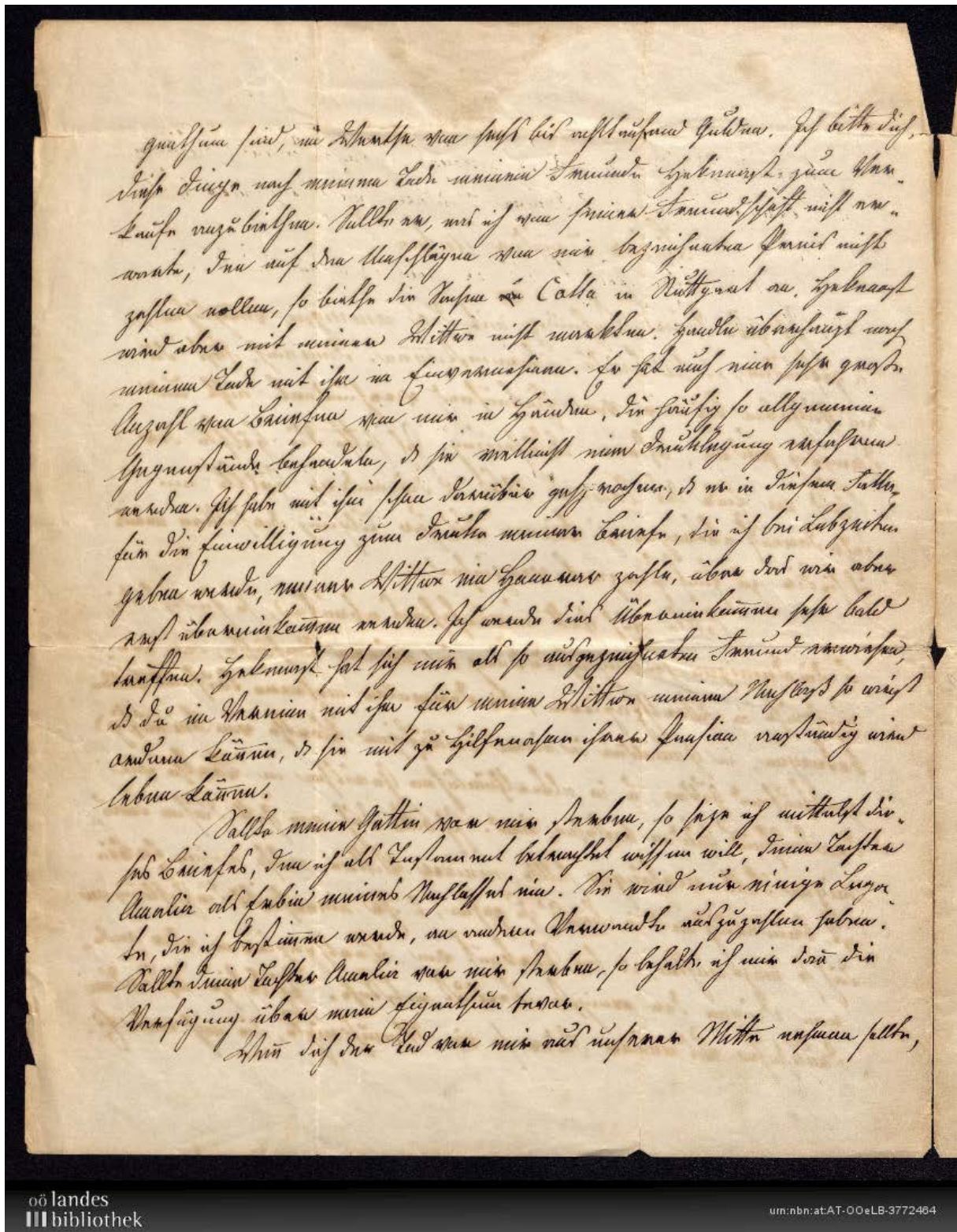


Figure 1 Page Image from Adalbert Stifter (1805-1868)

Data which can be easily derived from the Wikipedia/Wikidata link (just examples)

- Place of birth: Horní Planá
- Coordinate location: 48°46'3"N, 14°1'57"E
- Image (related to Horní Planá):
- [https://commons.wikimedia.org/wiki/File:Horn%C3%AD\\_Plan%C3%A1.JPG](https://commons.wikimedia.org/wiki/File:Horn%C3%AD_Plan%C3%A1.JPG)
- Date of birth: 23. October 1805



- Date of death: 28. January 1868
- etc.

## 2.6. Outlook

In 2017 we will draft a detailed work plan for implementing the FamousHands campaign. This will comprise two main strands:

- First the marketing side of the campaign: Work on this can be started early in 2017 by the Dissemination Working Group headed by UCL. Work here will include investigations on the best ways to launch the “story”, preparing press releases, instruction papers and setting up e.g. a Facebook site, etc. Moreover example entries for famous persons can be created straight away to provide test data for the technical development.
- Secondly the technical implementation itself. Though there is a common understanding of the technical features the implementation itself will need to be done carefully having in mind, that thousands of users all over the world will be interested to take part in this campaign. Technical work in this area will very likely not start before autumn 2017.