

Digital Humanities in Austria

Transkribus

A European research infrastructure for the
automated recognition and searching of historical
documents

Günter Mühlberger

University Innsbruck,

Digitisation and Digital Preservation Group

The logo consists of the word "READ" in a bold, dark blue, sans-serif font. The letters are closely spaced and have a slight shadow effect, giving it a three-dimensional appearance.

Agenda

- Overview on DH
- READ
- Handwritten Text Recognition
- Sustainability

Digital Humanities in Austria

- Austrian Centre for Digital Humanities (ACDH)
 - Common effort of the Austrian Academy of Science, Universities of Vienna and Graz as well as several other universities and colleges in Austria
 - Acts also as CLARIAH centre at the Austrian Academy of Science
- Universities of Vienna and Graz
 - Three DH professorships
 - Master studies in DH at University Graz
- National projects
 - Digital Edition Network (KONDE)
 - DI4DH (Digitisation and Information Extraction for DH)
 - GoDigital calls
 - Large number of research projects
- Highly active and lively landscape!
 - Hundreds of researchers involved

07/26/2018

ONE SCHEMA TO RULE THEM ALL

A new publication on ARCHE and the state of data repositories in Austria is available now!



ONE SCHEMA TO RULE THEM ALL.
THE INNER WORKINGS OF THE DIGITAL ARCHIVE ARCHIVE

KARL KRAUS LEGAL PAPERS

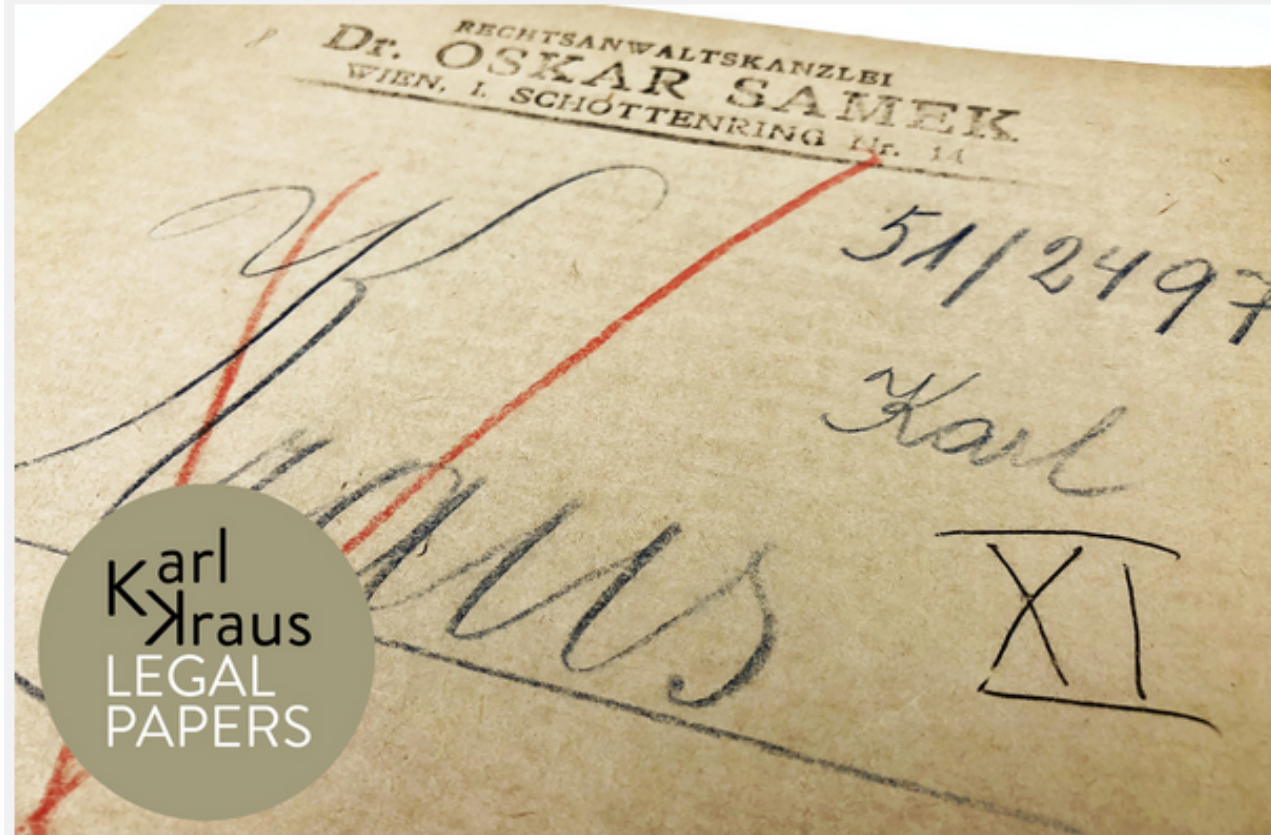


Photo by Gerhard Bauer, logo CC BY 4.0 Sandra Lehecka

In the FWF-funded project “Intertextuality in the Legal Papers of Karl Kraus. A Scholarly Digital Edition” (FWF project no. P 31138-G30), the legal papers of the Austrian satirist Karl Kraus (1874–1936) will be edited, provided digitally and contextualized with Kraus’ oeuvre as a whole. This project is carried out by the ACDH in cooperation with the [Vienna City Library](#) and the [Ludwig Boltzmann Institute for History and Society](#).

GRÜSS GOTT, SERVAS, PFIAT DI!

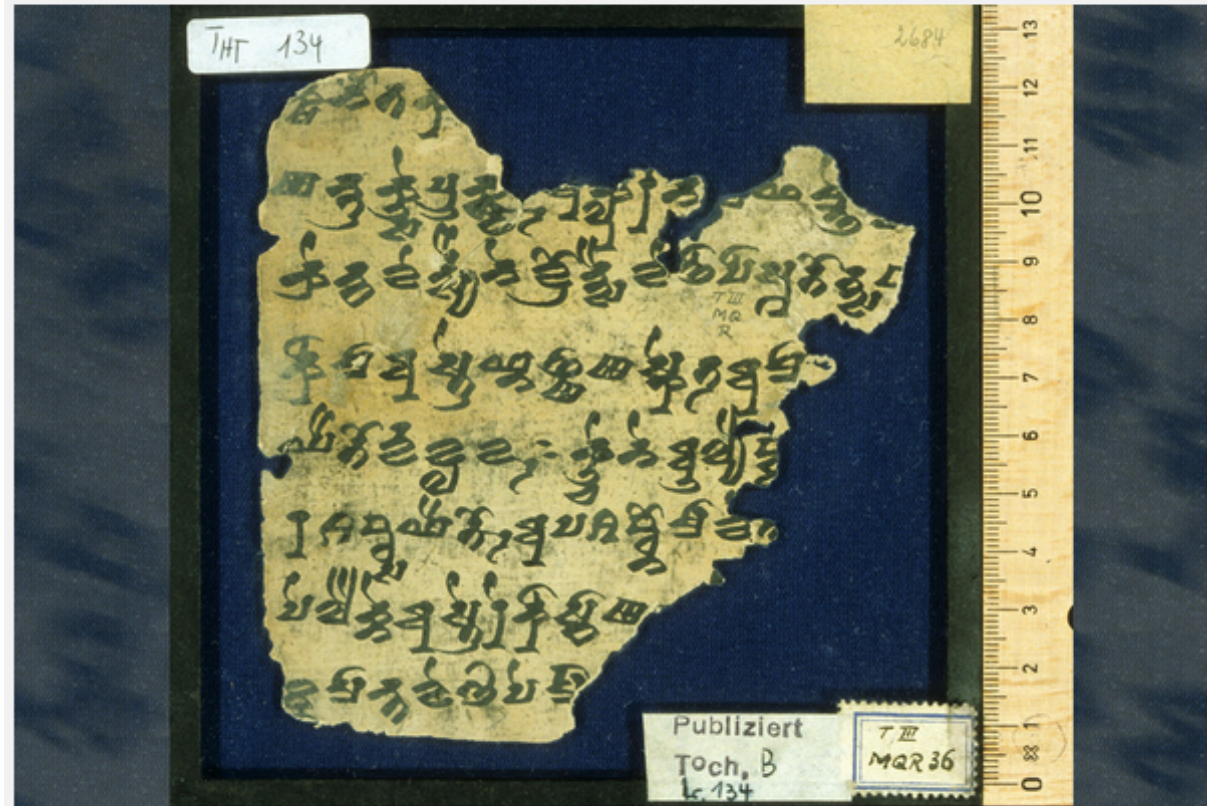
Dialekte halten sich weiterhin wacker, sagt die Sprachwissenschaftlerin Alexandra N. Lenz. Gemeinsam mit ihrem Kollegen Manfred Glauninger erforscht sie „Variation und Wandel des Deutschen in Österreich“. Warum im Alltag, in der Werbung oder der Popmusik gerne Dialekt verwendet wird, erklären die beiden ÖAW-Wissenschaftler/innen im Interview.



© Shutterstock.com

Fragt man Österreicher/innen nach ihrer Muttersprache, antworten die wenigsten mit: „Deutsch“ – vielmehr spricht man Steirisch, Wienerisch, Tirolerisch oder Kärntnerisch. Und selbst innerhalb der Regionalsprachen gibt es eine bunte Vielfalt an Mundarten. Denn: Dialekte sind gelebte Sprachkultur. Lange Zeit als „hinterwäldlerisch“ verschrien und durch das überregionale, „hochdeutsche“ Fernsehen weitgehend ignoriert, erleben Dialekte als regionales Kulturgut heute ein Comeback, und zwar nicht nur im Alltag,

THE CHARACTERS THAT SHAPED THE SILK ROAD



Tocharisches Manuskript aus dem 7. Jahrhundert © Courtesy of IDP Berlin, the BBAW and the Staatsbibliothek zu Berlin

A DATABASE AND DIGITAL PALEOGRAPHY OF TARIM BRAHMI

From the 2nd century CE on, Buddhist communities and monasteries developed along the trade routes of the ancient Silk Road in and around the Tarim Basin in today's

...and many, many more

Many of these DH projects have one thing in common – they are working with historical documents and need to digitise, transcribe, annotate, search and mine the text

FileEditViewChronicleBookmarksExtrasHelp

Flüge nach...Neuer TabFahrplana...Flüge nach...TextImage...Text2vor...textImagerUW Philtag 14...http://u...hei.nz/Toggl - Ti...READ READ ...READ_D7.7_HT...

https://read.transkribus.eu

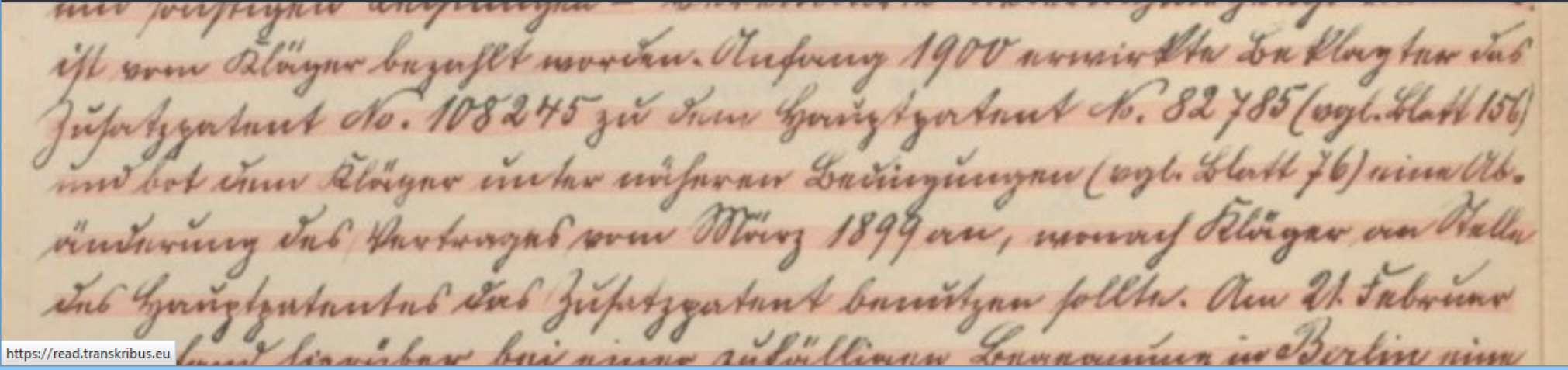
120%textimager hu

Google KalenderSP DEAACMULB PRIMOTS WIKIEU_NEWSSolritasTSXAdobe Connect LoginPARTHENOSPartPortalTranskribus WikiREAD WikimyPayLife - ÜbersichtSBB aktuell

READ

Recognition and Enrichment of Archival Documents

HomeNewsMissionNetworkResearchServicesPublicationsContact



https://read.transkribus.eu

Desktopc60817

Links

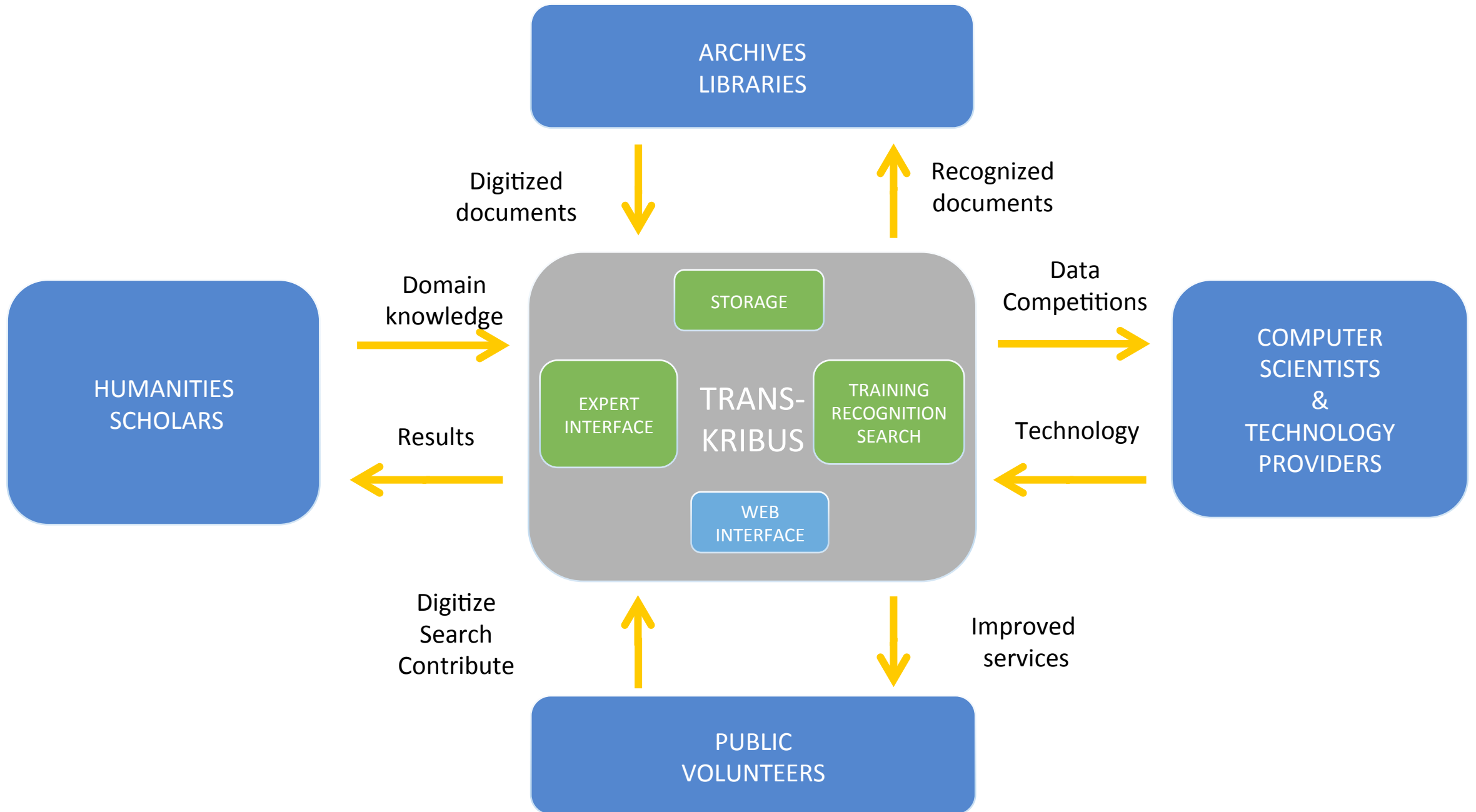
DEU06:59

Consortium

- **University Innsbruck (Coordinator / Austria)**
- **Technical University Valencia (Spain)**
- **University College London (UK)**
- **National Center for Scientific Research “Demokritos” (Greece)**
- **University of London Computer Centre (UK)**
- **Technical University Vienna – Computer Vision Lab (Austria)**
- **University Rostock (Germany)**
- **University Leipzig (Germany)**
- **Naver Labs (France)**
- **Technical University Lausanne (Schwitzerland)**
- **National Archive Finland (Finland)**
- **Statearchive Zurich (Switzerland)**
- **Passau Diözesanarchiv (Germany)**
- **Democritus University Thrace (Greece)**

READ - Transkribus

- Facts and figures
 - Horizon 2020 eInfrastructure Project
 - 8,2 mill. EUR funding
 - Duration: 1.1.2016 – 30.6.2019
 - 14 partners, coordinated by the University of Innsbruck
- Research infrastructure
 - Digitisation, transcription, recognition, search in historical documents
 - Cover the complete workflow so that users need no other tools to come from images to fully functional text
- Culture of collaboration
 - Foster cooperation between all stakeholders in the field: archives/libraries, humanities scholars, computer scientists and public users (volunteers)
 - Open Access
 - Share research and data but enable users to work in a private environment



Memorandum of Understanding Partners

- **Australia**
- [Geopast](#)
- [National Library of Australia](#)
- **Austria**
- [University Vienna, German Studies Department](#)
- [Technical University Graz – University Archive](#)
- **Canada**
- [DEEDS Project – University of Toronto](#)
- **Denmark**
- [Aarhus City Archives](#)
- **Finland**
- [Finnish Literature Society](#)
- [The Society of Swedish Literature in Finland](#)
- [Tampere University Folklife Archives](#)

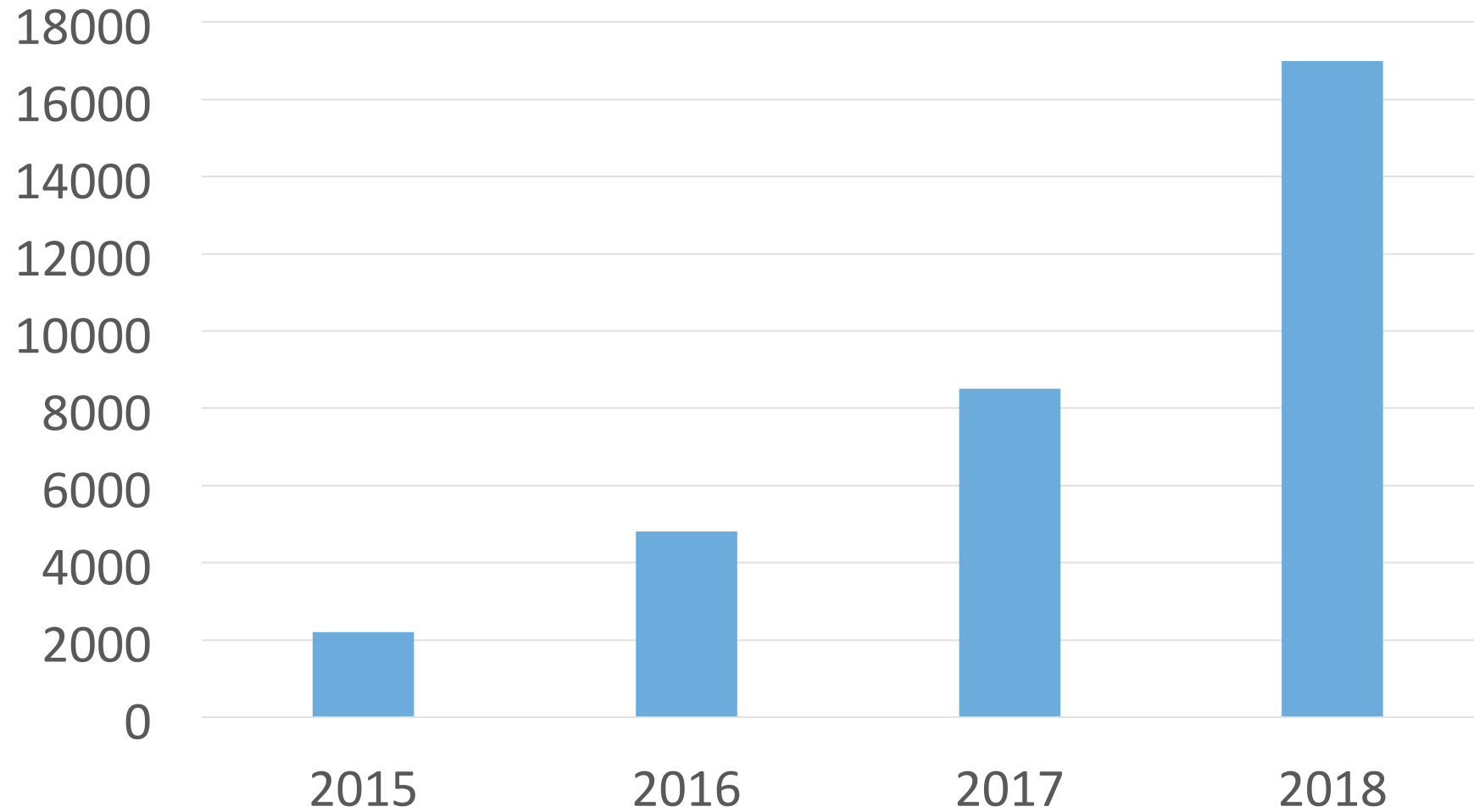
- **France**
- [ALMAAnaCH – Inria Paris Research Centre](#)
- [National Library of France](#)
- **Germany**
- [Deutsche Zentralbibliothek für Medizin](#)
- [Ernst-Moritz-Arndt-University of Greifswald](#)
- [Fernuniversität Hagen – Department for History and Biography](#)
- [Friedrich-Alexander-Universität Erlangen-Nürnberg – Pattern Recognition Lab](#)
- [Germanisches Nationalmuseum](#)
- [Gottfried Wilhelm Leibniz Bibliothek](#)
- [Heidelberg University, Department of History](#)
- [The Hessian State Archives](#)
- [Institute for Documentology and Scholarly Editing](#)
- [Museum for Musical Instruments of Leipzig University](#)
- [PLANET](#)

- [University and Research Library Erfurt/Gotha](#)
- [Hochschule Hanover – University of Applied Sciences and Arts](#)
- [Österreichischer Bibelübersetzer – Bavarian Academy of Science at the University of Augsburg](#)
- [University Library Tübingen](#)
- [University of Rostock – Edited Letters – Barlach 2020](#)
- [University of Rostock Library](#)
- **Greece**
- [National Bank of Greece Cultural Foundation](#)
- **Italy**
- [Archivo Storico Ricordi](#)
- [BOhisto – The Civic Archives of Bozen-Bolzano](#)
- [Marciana National Library](#)
- **Luxembourg**
- [Centre virtuel de la connaissance sur l'Europe – Digital Humanities Lab](#)

- **Netherlands**
- [Center for Global Heritage and Development \(Leiden University, Erasmus University Rotterdam, Delft University of Technology\)](#)
- [Cultural Heritage Agency](#)
- [Huygens Institute for the History of the Netherlands](#)
- [International Institut of Social History](#)
- **Norway**
- [Department of Cultural Studies and Oriental Languages, University of Oslo](#)
- [The Munch Museum](#)
- [National Library of Norway](#)
- **Serbia**
- [University Library Belgrade](#)
- **Spain**
- [National Library of Spain](#)
- [PROLOPE](#)
- [University of Girona – Center de Recerca d’Historia Rural](#)

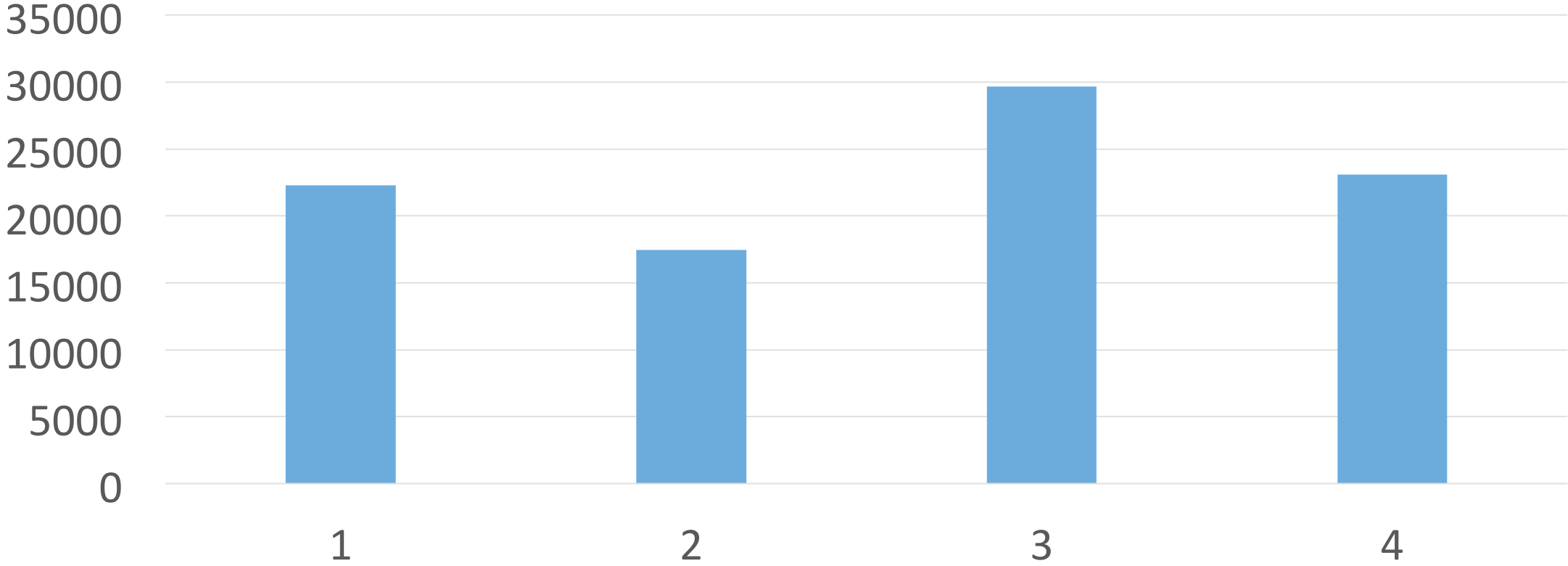
- **Switzerland**
- [Alfred Escher Foundation](#)
- [Rechtsquellenstiftung des Schweizerischen Juristenvereins](#)
- [University Library Basel](#)
- [University of Bern, Digital Edition of Joseph Dietrich's Diary](#)
- [University of Zurich, History Department](#)
- **United Kingdom**
- [The British Library](#)
- [The Honourable Society of the Middle Temple – Library and Archive](#)
- [The Linnean Society](#)
- [London Metropolitan Archives](#)
- [The University of Aberdeen](#)
- **United States**
- [Northern Arizona University](#)
- [Western Michigan University – Kalamazoo](#)

Registered users in Transkribus

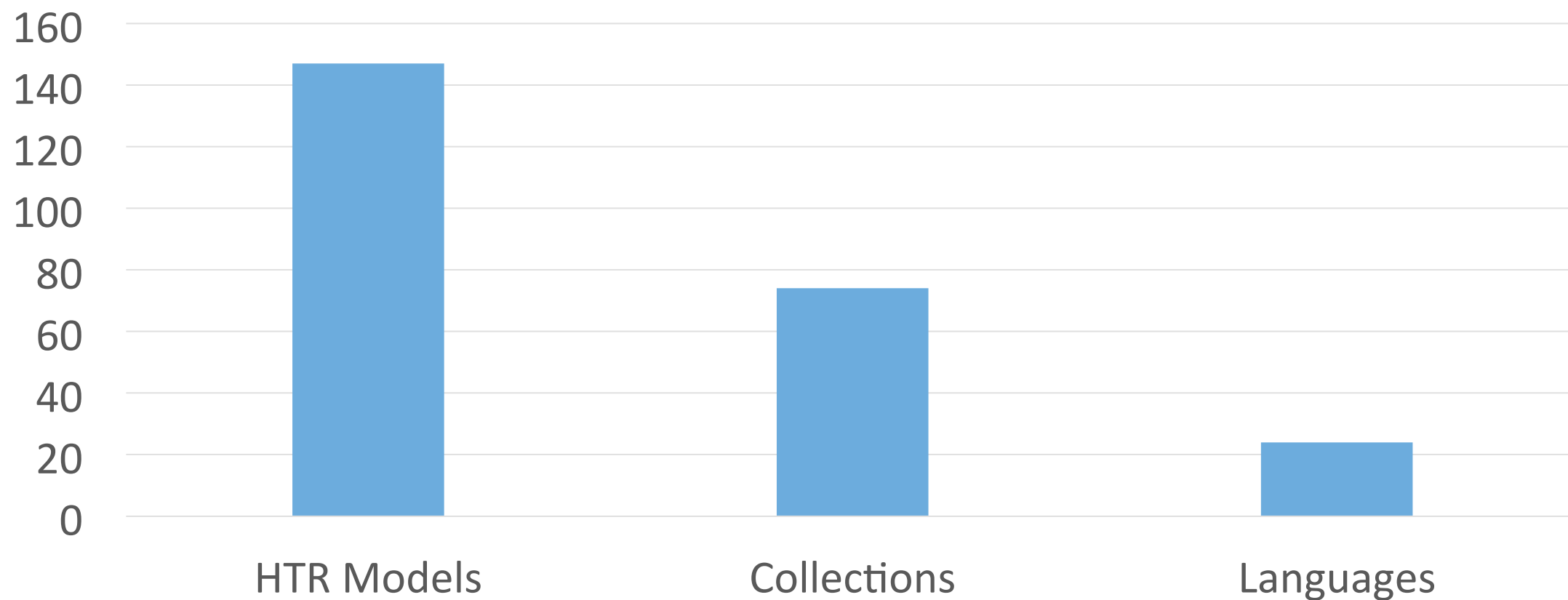


Activities from September 2018

Uploaded page images in Sept. 2018 / per week
/ 92.448 altogether



Handwritten Text Recognition (HTR) models trained by users in September 2018



Value generated by users in Sept. 2018

- Software changes – data remain!
- Users transcribed **6525 pages** for their training = ground truth
- 1 page = 1 hour work = 6525 hours = **3,8 person years**
- E.g. one hour calculated with 20 EUR sums up to **130.500 EUR**

Handwritten Text Recognition

Council Protocols from Bozen/Italy

- Collaboration with City Archive Bozen “Bohisto”
- Volunteers are transcribing documents from late 18th century
- Supervision by University of Innsbruck – Barbara Denicolo
- Model trained with 562 Seiten, 44.521 words
- Results (until now): +10% CER
- Results with new Handwritten Text Recognition Engine: 4,73% CER

In Publico

Am 31. Xbr 1792.

Ausstellung der Gesundheits-
zeugnisse für die in die Stadt
einkommenden Waaren sind
auf den vorigen Gemeinderath zurück-
geführt worden, wozu
laut den pünktigen Bericht vom
19ten v. M. Hoffnung vorhanden

283. 283

In Publico - Politico - Publico - Politico

Den 31. Xbr 1792.

Ausstellung der Gesundheits-

zeugnisse - zeugnisse für die in die Vene-

tianische gehenden Waaren wird

auf die vorige Gewonheit werde

zurückgeführt werden, wozu

laut diesseitigen Bericht vom

19ten v. M. Hoffnung vorhanden

seyn solle.

Conclusum.

a-Da der angezogene Bericht

nur auf eine vom H: Joseph v

Gumer dem H: Bürgermeister

dieses Magistrats ertheilte

spezielle Auskunft - Auskunft erstattet

worden, als ist von demselben

die Äußerung abzufordern, - Äußerung abzufordern, wie

weit die in dieser Sache nach

Venedig - Venedig unterhaltene Korre-

spondenz bereits gekommen,

auch ob, - auch ob, und bis welche Zeit die

Wiederherstellung des aus

Keyword spotting

Mitterlehner - Moiveshekner

The screenshot displays a software interface for handwriting analysis. On the left, a sidebar contains a list of names with associated page numbers. The main area shows a preview of a handwritten document with blue wavy lines indicating detected text regions. The list of names is as follows:

Page	Name
1	
12	
26	ca...
5	sted
388	ed
101	14-...
322	
522	lic...
79	
79	
236	
603	

The main area shows a preview of a handwritten document with blue wavy lines indicating detected text regions. The text in the image is:

Mitterlehner Maurts
in Thon ing and for Ros

The list of names is as follows:

- 16 hetaes
- 17 1 Jäerer
- 18 Moiveshekner Maurts ded siusbesit
- 19 if

European historical document collections can be recognized, automatically transcribed and made searchable in a previously unknown way – if scholars, archives/libraries, computer scientists and the public are collaborating.

Sustainability

Spin off

- How to transform a project into a legal entity?
- Which governance model, which business model?
- How to sustain the public mission?
- How to be flexible and strong enough for the market?
- How to enable collaboration?



GROWTH

Internal Market, Industry, Entrepreneurship and SMEs

EUSurvey

[User survey questionnaire](#)

European Commission > Growth > Sectors > Social economy in the EU > Cooperatives > The European Cooperative Society (SCE)



Search



Single Market and Standards

Industry

Entrepreneurship and SMEs

Access to finance for SMEs

Sectors

Social economy

Social enterprises ☐

Associations and foundations

Mutual societies

Cooperatives ☒

European Cooperative Society

Social economy - links

News

Events

Tools and Databases ☐

Contracts and grants ☐

Public consultations

Publications

The European Cooperative Society (SCE)

The European Cooperative Society (SCE) is an optional legal form of a cooperative. It aims to facilitate cooperatives' cross-border and trans-national activities. The members of an SCE cannot all be based in one country. The SCE is required to unite residents from more than one EU country.

The Statute for a European Cooperative Society (SCE)

The regulation of the [Statute for a European Cooperative Society](#) (2003) aims to facilitate [cooperatives'](#) cross-border and trans-national activities. The statute also provides a legal instrument for other companies wishing to group together to access markets, achieve economies of scale, or undertake research and development activities.

The Statute also enables 5 or more European citizens from more than one EU country to create a European Cooperative Society. This is the first and only form of European company that can be established from the beginning and with limited liability.

The Statute for an SCE parallels [the Statute for a European Company](#) (SE).

What the SCE is

- a legal entity that allows its members to carry out common activities, while preserving their independence
- its principal object is to satisfy its members' needs and not the return of capital investment
- members benefit proportionally to their profit and not to their capital contribution.

Main characteristics of the SCE

The European Cooperative Society (SCE)

- *is a legal entity that allows its members to carry out common activities, while preserving their independence;*
- *its principal object is to satisfy its members' needs and not the return of capital investment;*
- *members benefit proportionally to their profit and not to their capital contribution.*

European Cooperative: READ-COOP

- Enables us
 - To collaborate with all kinds of institutions – archives, libraries, universities, research organisations, associations, and even private persons
 - To include new members (which are also owners) with minimum effort
 - To carry out business with these members (but also with non-members)
 - To gain profit from services
 - To use profit to improve services
 - To sustain the Transkribus research infrastructure and its public mission
- Status
 - Currently negotiations are ongoing with potential partners
 - We hope to set up SCE during first half of 2019

READ - EOSC

- What we offer
 - Highly standardized data
 - Implemented FAIR principles
 - All services as web-services via REST API
- What we would like to see from EOSC
 - Collaboration – maybe as member of READ-COOP
 - Storage and computing facilities for processing large amounts of historical documents

Thank you a lot for your attention

More information

<https://www.oeaw.ac.at/acdh/>

<https://read.transkribus.eu/>

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 674943.

