

Della 7. 5. anno  
questo 8. anno  
principiare nell'anno  
di 1765. C.  
li svedetti Paolo Scu-  
ri d'Anno a Parva di  
Serguiali, Merita e  
di Santi pro tempore  
e i suoi successori  
filiis sic.

Nota est f. 1. Cap. 1. in fratre  
Riccardo, fratello etiam  
Palma, fratello in qualibet for-  
me f. 1. Sic.

fidelissime am-

stato, e con promessa, ad obbligare  
e difendere in forma com-  
petente prezzo di L. 41. 8. 4. diconse-  
vata di Milano, le quali si  
alla rata decorsa come da  
perciò il pref. citt. sub la  
in ogni appurata forma  
ad il suo carone lire  
e titolo del d. prezzo, ad

de Martino  
Borghj de Com  
E sit etia quod modo  
ad tertias negotia  
zante filios et  
per idcirco de  
ad bene mun. f. 00.

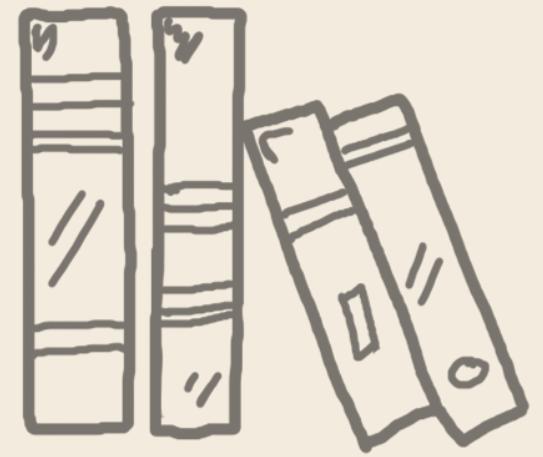
# Transkribus®

23. Februar 2023 | DigEdTnT: Workshop - Meet the Tools and Developers

Sebastian Colutto & Matthias Sorg, READ-COOP SCE



Können Sie das lesen?



16 Feb. Ao 1610.

Ehrnveste, Achtbare vnd hochgelarte,  
Großgoönstige Herrn, Nehist meiner  
alzeit willig gefliessener Dienste er-  
bietung, kan E. E. g. Ich vnvormeldet  
nit lassen, Das vorschienner Jare Marx

16 Feb. Ao 1610

Ehrnveste, Achtbare vnd hochgelarte,  
Großgoönstige Herrn, Nehist meiner  
alzeit willig gefliessener Dienste er-  
bietung, kan E. E. g. Ich vnvormeldet  
nit lassen, Das vorschienner Jare Marx

# Das Transkribus-Ökosystem:

Die digitale Verfügbarkeit ermöglichen

Bilderfassung

Bearbeitung, Datenerfassung +  
Annotation und KI-Training

KI-gestützte  
Texterkennung

Veröffentlichung



**Transkribus®**  
lite

**Transkribus®**  
eXpert

citizen&science

**Transkribus®**  
metagrapho<api>

**Transkribus®**!ai



# Was ermöglicht Transkribus?



Transkription des Textes



KI-Training



Layout-Erkennung



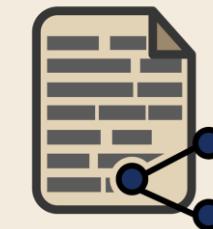
Texterkennung



Anreicherung des Textes



Durchsuchen von  
Dokumenten



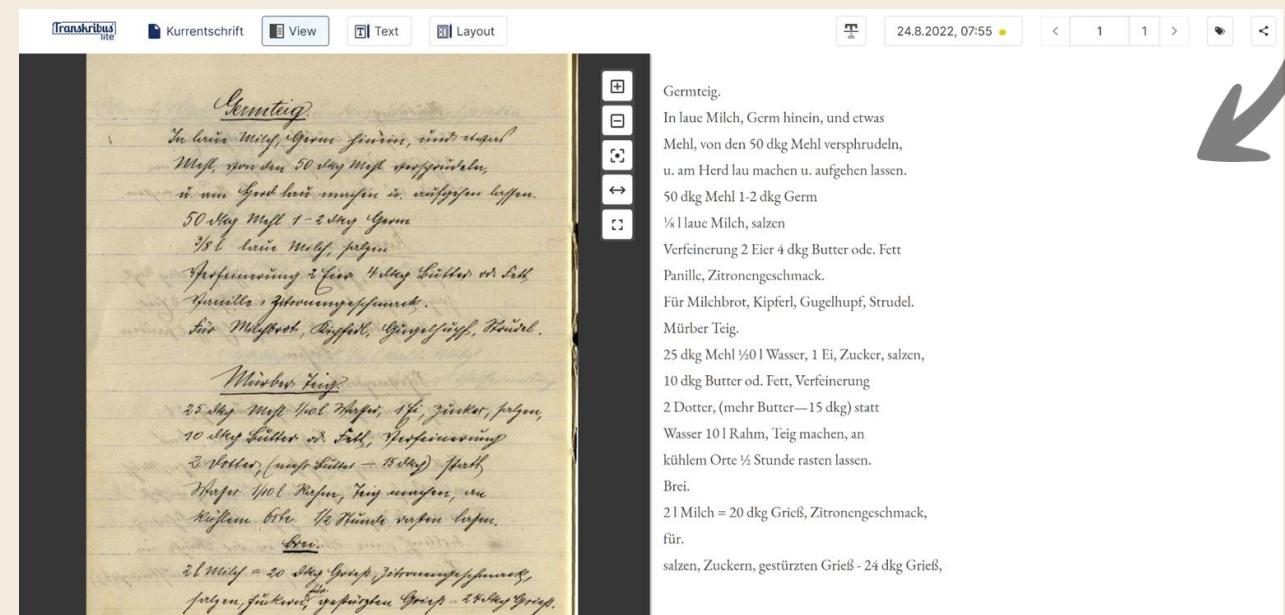
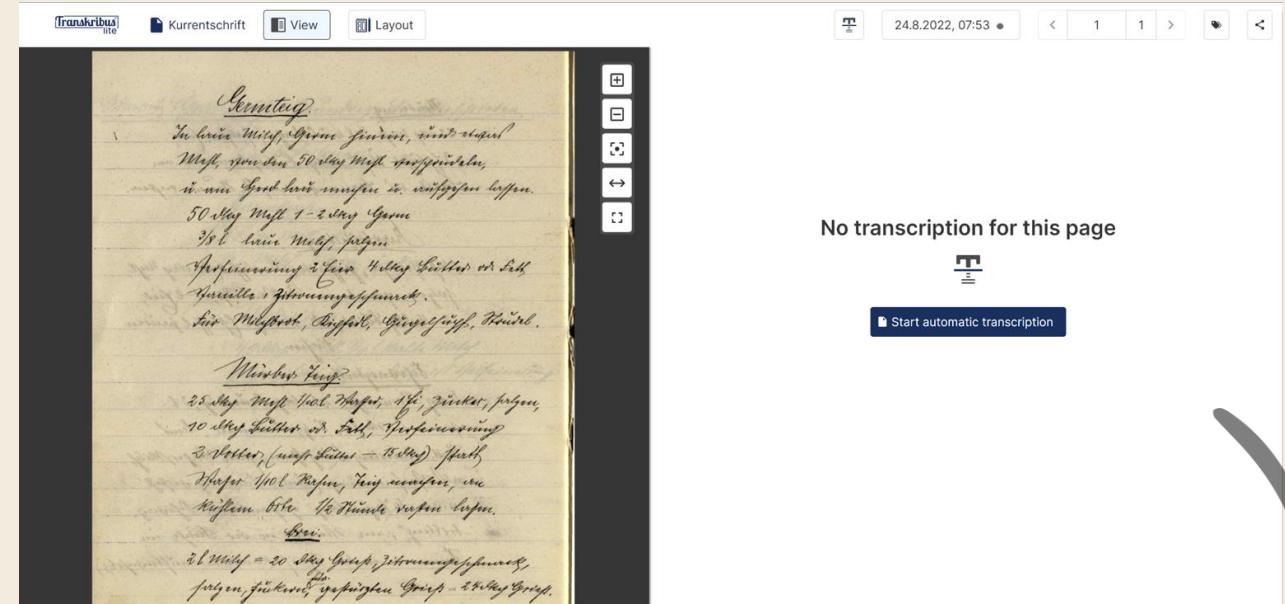
Export/Veröffentlichung



Zusammenarbeit

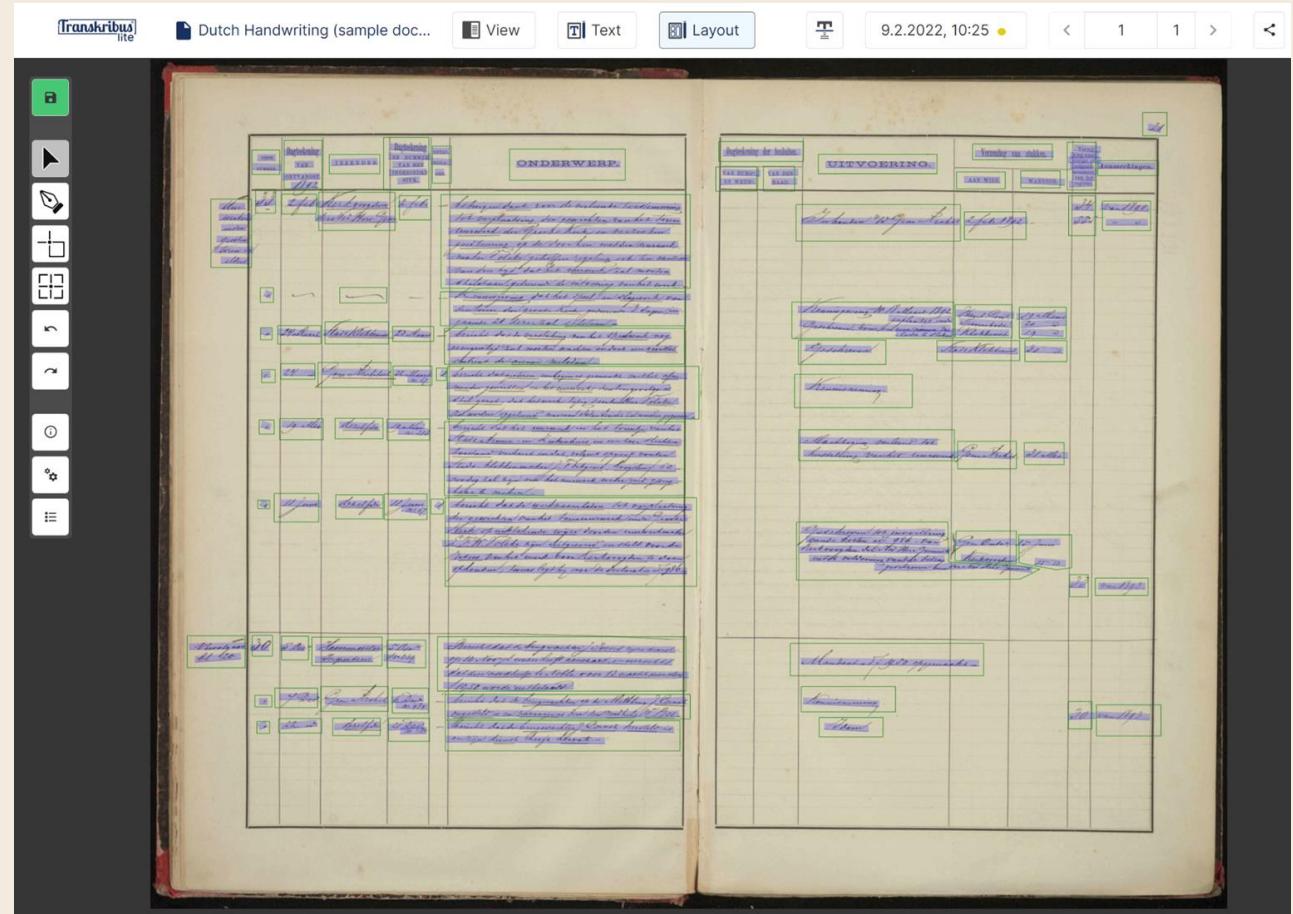
# Texterkennung

- Automatische Erkennung von Text auf Scans
- Handschriftliche und gedruckte Texte
- Möglich durch KI-Modelle



# Layout-Erkennung

- Trainierbare Layout-Erkennung
- Basis einer guten Texterkennung
- Möglichkeit der strukturierten Erkennung komplexer Dokumente



# Eigene KI-Modelle trainieren

- KI für jeden zugänglich
- Erstellung eigener Modelle möglich
- Weiterentwicklung bestehender Modelle möglich

The screenshot shows the 'Training' tab of the Readcoop AI platform. At the top, there are tabs for 'Workdesk' and 'Training', with 'Training' selected. A search bar says 'Enter search term ...' and a notification badge shows '899'. Below the tabs, a collection named 'learn@readcoop.eu Collection' is listed. On the left, a 'Text Model' is selected. The main area is titled 'Text Training' and includes a workflow diagram with four steps: 'Model Setup', 'Training data', 'Validation data', and 'Start'. Each step has a circular icon above it. Below the diagram, there are input fields for 'Model Name' (set to 'Kurrento - My Custom AI Model'), 'Description' (containing the text 'This is my own model to recognise German Kurrentschrift'), 'Language' (set to 'German'), and 'Centuries' (set to '17' and '19'). At the bottom, there are buttons for 'Transcript version' (with 'Latest Transcript' selected) and 'Ground truth only'.

# Zusammenarbeit

- Kollaboration in Teams
- User Management
- Zentrale Creditverwaltung möglich

The screenshot displays the Transkribus lite application interface. On the left, a sidebar menu includes 'Workdesk' (selected), 'Training', 'Enter search term ...', '899 users', and a user icon. The main area shows a 'Collection Overview' for 'learn@readcoop.eu Collection'. It features three document thumbnails with handwritten text. Below this is a 'Users' section with a 'Manage' button, a search bar ('Search users by mail shop@readcoop.eu'), and a 'Search' button. A list of users is shown, with 'shop@readcoop.eu' selected. At the bottom, there are sections for 'Kurrent-Rezept' and 'Test', each with a 'Select' checkbox and a '...' button.

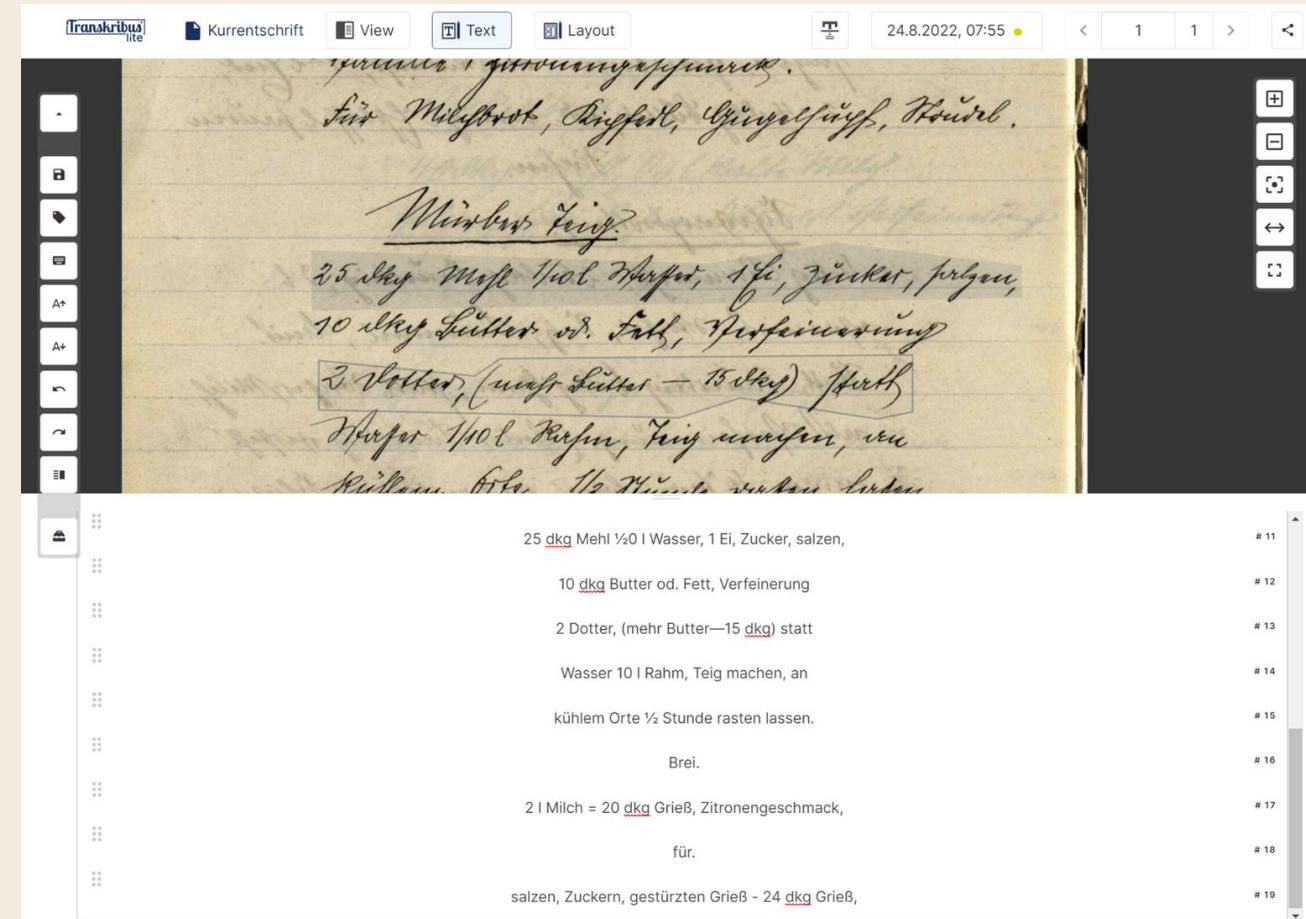
# Durchsuchen von Dokumenten

- Volltextsuche
- Smart Search, um schwierige Suchbegriffe zu finden sowie Rechtschreib- und Erkennungsfehler oder Schreibvarianten zu überbrücken

The screenshot shows the Transkribus lite interface. At the top, there is a navigation bar with 'Workdesk', 'Training', a search bar containing 'Mehl', a user icon, and a notification count of '899'. Below the navigation bar is a sidebar titled 'TRANSKRIBUS ORGANIZER' with options for 'Collections', 'Jobs', 'Tag Manager', and 'Recent Documents'. The main area is titled 'Text' and has a search input field with 'Mehl', a 'Fuzzy' checkbox, and a 'Search' button. Below the search bar are filters for 'Author', 'Uploader', 'Title', 'Collection Name', and 'Script type'. A results section shows '6 Results' for 'Kurrentschrift / Page 1'. The results list includes a snippet of text: "... Germteig. In laue Milch, Germ hinein, und etwas **Mehl**, von den 50 dkg **Mehl** versprudeln, u. am Herd ... lau machen u. aufgehen lassen. 50 dkg **Mehl** 1-2 dkg Germ ½ l laue Milch, salzen Verfeinerung 2 Eier 4 ... Teig. 25 dkg **Mehl** ½ l Wasser, 1 Ei, Zucker, salzen, 10 dkg Butter od. Fett, Verfeinerung 2 Dotter ...". Below the snippet are four horizontal rows of handwritten text fragments, each with a red box highlighting the word 'Mehl'. The right edge of the interface has a vertical 'Feedback' button.

# Transkription des Textes

- Integrierter Transkriptions-Editor
- Dokumente manuell transkribieren
- Automatisch generierte Transkriptionen bearbeiten



# Anreicherung des Textes

- Anreicherung mit Metadaten durch Markieren, Taggen und Annotieren
- Mit oder ohne Unterstützung der KI
- Aufbau einer Grundlage für eine ausfeilte Informationsextraktion

The screenshot shows a digital transcription of handwritten notes from a notebook. The notes are organized into sections with headings underlined in blue ink.

**Germteig:**

In einer Milzf. Germ ziehen, um sie zu  
Milzf. von den 50 dkg Mehl ausschmieden,  
u. um Spießknochen zu machen u. aufzugeben lassen.  
50 dkg Mehf 1-2 Ei 1 Germ  
3/8 l einer Milzf. salzen  
Verfeinerung 2 Eier 400g Butter vor Salz  
Zimtlauch, Zitronengeschmack.  
Für Milzbrod, Brötchen, Grießflocken, Knödel.

**Milzteig:**

25 dkg Mehf 100g Margarine, 1 Ei, Zucker, salzen,  
100g Butter vor Salz, Verfeinerung  
3 Eier, (ausst. Butter - 15 dkg) Margarine  
Margarine 100g Margarine, Teig machen, an  
Küchentisch 1/2 Stunde rasten lassen.  
**Brei:**  
1 l Milzf = 20 dkg Grieß, Zitronengeschmack,  
salzen, Zuckern, Margarine Grieß - 24 dkg Grieß,

Annotations and toolbars are visible at the top of the screen, indicating the use of digital annotation tools like Transkribus.

# Export von Dokumenten

- Verschiedenste Formate verfügbar
- z.B. XML, PDF, TEI oder Docx
- Export der Bilder möglich

The screenshot shows the Transkribus Lite interface. On the left, there's a sidebar with 'TOOLS' (Text Recognition, Layout Recognition, Upload, Download, Delete, Users-Manager), 'TRANSKRIBUS ORGANIZER' (Collections, Jobs, Tag Manager, Recent Documents), and a 'Workdesk' tab. The main area shows a 'Collection Overview' for 'learn@readcoop.eu Collection'. Three documents are listed, each with a preview image and a detailed description below it. A modal dialog box is open in the center, titled 'Download'. It contains two sections: 'Image' (with a checked checkbox for 'Images') and 'Files' (checkboxes for Page XML, Alto, PDF, TEI, Docx). Below these are sections for 'Selected documents' (checkbox for 'Kurrentschrift Pages: 1') and a large green 'Start' button. At the bottom of the interface, there are sections for 'Kurrent-Rezept' and 'Test', each with a 'Select' checkbox and three dots.

# Wer steckt hinter **Transkribus**?

## READ-COOP SCE

Eine Europäische Genossenschaft zur Aufrechterhaltung und Weiterentwicklung der Transkribus-Plattform.



2013  
transCriptorium

2016  
READ

2019  
Foundation  
of READ  
COOP SCE

today  
140+ members  
30 countries

# Wer steckt hinter Transkribus?



# Das KI-Team

30+ Mitarbeiter mit Sitz in Österreich, Deutschland, Italien und UK



\* Einige  
Gesichter  
fehlen noch :-)

# READ-COOP in Zahlen

**~ 110.000 Nutzer**

**30 Mio. Seiten texterkannt**

**50.000.000  
Hochgeladene Scans  
insgesamt**

**~ 20.000  
trainierte Modelle**

**150 +  
Länder mit aktiven  
Nutzer\*innen**

**40.000 +  
Monatliche Website-  
Besucher\*innen**

**~ 30  
read&searches online**

**1000 +  
ScanTents in Verwendung**

# Das Transkribus-Ökosystem:

Die digitale Verfügbarkeit ermöglichen

Bilderfassung

Bearbeitung, Datenerfassung +  
Annotation und KI-Training

KI-gestützte  
Texterkennung

Veröffentlichung



**Transkribus®**  
lite

**Transkribus®**  
eXpert

citizen&science

**Transkribus®**  
metagrapho<api>

**Transkribus®**!ai

read & search

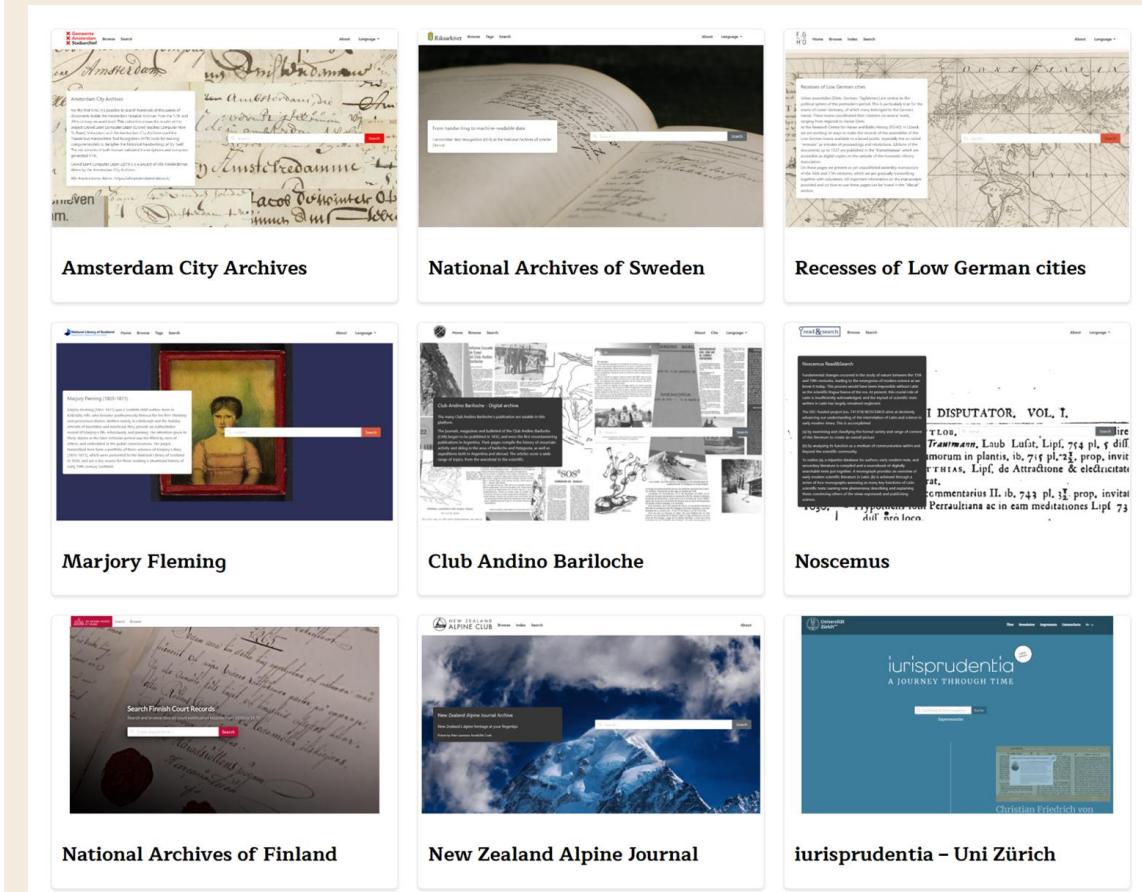
# Das ScanTent

- Hochwertiges Scannen für unterwegs
- Nutzung des eigenen Smartphones
- Eigens entwickelte DocScan App
- Leichtes, faltbares Zelt
- schnell auf- und abgebaut
- standardisierter Hintergrund
- ideale Lichtbedingungen durch integrierte LED Beleuchtung



# read&search

- Einfache und effiziente Publikationsmöglichkeit
- Ohne Programmierkenntnisse oder IT-Ressourcen nutzbar
- Teilen mit Öffentlichkeit oder privater Gruppe möglich
- Volltext- und Fuzzy-Suchfunktionen
- Smart Search integrierbar
- Tag-Suche verfügbar
- Überlagerung des Scans mit einer Text-Layer oder Gegenüberstellung möglich
- Beispiel: Amsterdamer Stadtarchiv



# metagrapho api

- Extraktion von Text- und Layout-Informationen
- Integration von KI-basierter Texterkennung in Ihren bestehenden Workflow oder Ihre Toolchain
- Erstellung völlig neuer Anwendungen zur Texterkennung
- Zugriff auf alle KI-Modelle, die Sie in Transkribus trainiert haben

```
{  
  "processId": 47725,  
  "status": "FINISHED",  
  "content": {  
    "text": "Eisenschrott",  
    "regions": [  
      {  
        "id": "region_1",  
        "coords": {  
          "points": "0,1 282,1 282,107 0,107"  
        },  
        "lines": [  
          {  
            "coords": {  
              "points": "5,106 37,109 59,92 80,91 97,105 156,108 195,76 220,69 254,82  
283,66 283,11 247,9 228,26 205,5 173,15 146,3 98,13 71,0 4,6"  
            },  
            "baseline": {  
              "points": "9,68 278,67"  
            },  
            "text": "Eisenschrott"  
          }  
        ]  
      }  
    ]  
  }  
}
```

# Transkribus On-Prem

- Ermöglicht das Hosting des Transkribus Workflows in Ihren Räumlichkeiten
- Nutzung Ihrer eigenen Infrastruktur
- Daten verlassen nie Ihre Server
- Lokale Verarbeitung der Scans
- Senden von Daten an Transkribus-Server ausschließlich zu Lizenzzwecken und Statistik (ohne Internetzugang möglich)

**Transkribus<sup>®</sup>**  
On-Prem

# Laufende Projekte



# Das Landesarchiv Berlin

## Einwohnermeldekartei Berlin

~3,6 Millionen Karteikarten

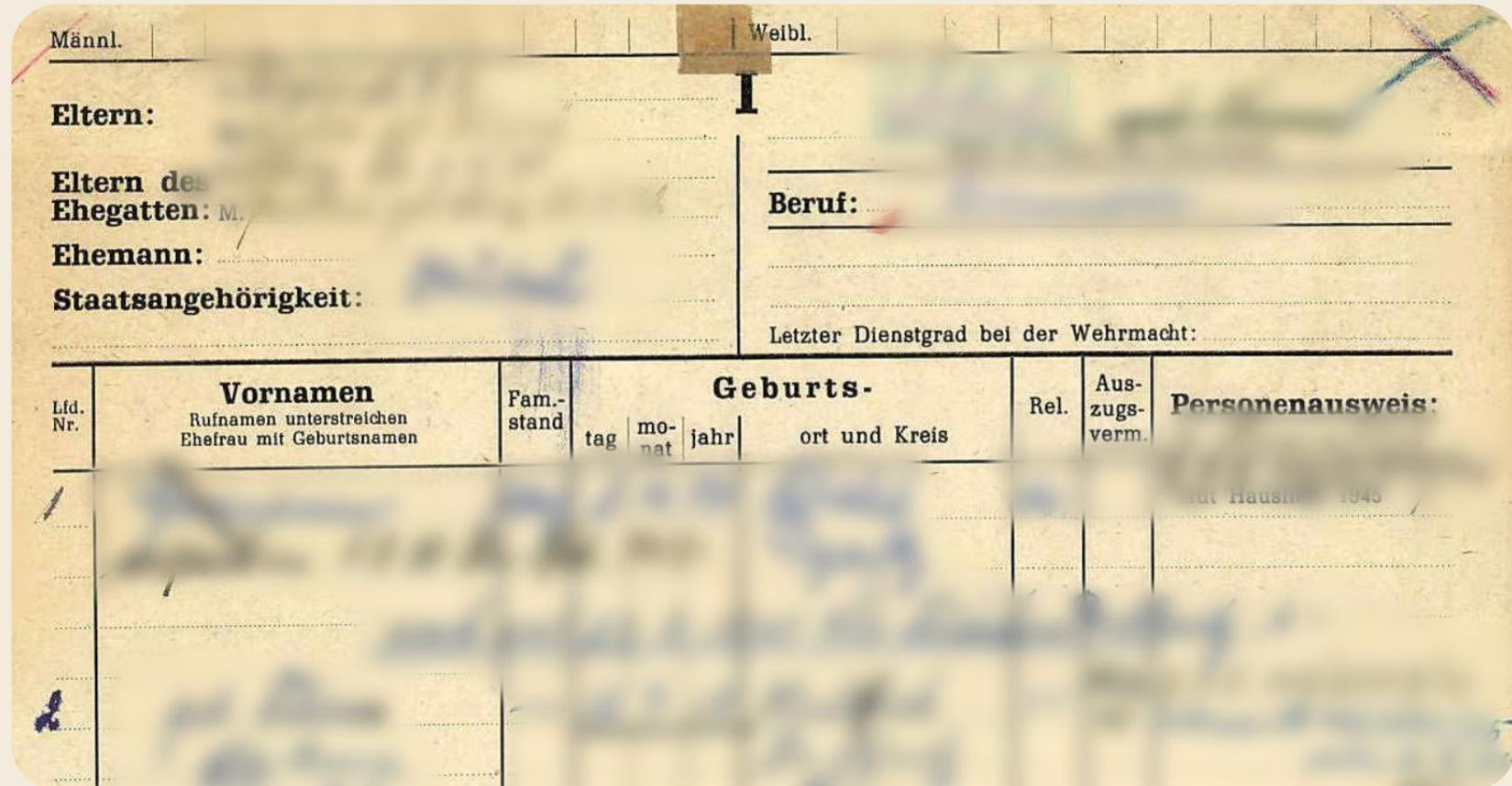
**Zeitraum: mehr als 50 Jahre** (erste Hälfte des 20. Jahrhunderts)

**Handschriftliche Texterkennung**

Layoutanalyse für **Formulare**

**Web-Interface** für rasche Korrekturen

Veröffentlichung der Ergebnisse über **read&search**



Landesarchiv Berlin



# zeit.punkt – Zeitungsportal

## Zeitungsportal (Nordrhein-Westfalen)

15 Millionen Zeitungsblätter

Universitäts- und Landesbibliothek  
Bonn

Verwendung des **NewsEye-Modells**  
(EU-Projekt koordiniert von der  
**Universität La Rochelle**)

**Texterkennung**

**Automatische experimentelle  
Artikel-Separierung (AAS)**

The screenshot shows a historical newspaper page from the "Wattenscheider Zeitung". The header includes the issue number (Nr. 153), date (Samstag, den 5. Juli 1913), year (45. Jahrgang), and a red mark (640). The main title is "Wattenscheider Zeitung" in large, bold, black letters. Below it is the subtitle "(Neueste Nachrichten.)" and the text "in Stadt und Amt Wattenscheid, Gevelsberg, Wetter, Hückelhoven, Höntrop, Eschendorf, Dinslaken, Überruhr u. a.". The page features several columns of text and three small illustrations at the bottom right depicting scenes from World War I.



# Das Staatsarchiv Amsterdam - VeleHanden

“Crowd bringt dem Computer das Lesen bei”

Citizen-Science-Plattform **VeleHanden**

Geplant: 4.000 Seiten  
Aktuell: **42.000 Seiten**  
von Freiwilligen mithilfe der Transkribus-Plattform transkribiert

**300.000 Seiten** über **read&search** zugänglich gemacht

Interessante Geschichte:  
Ein **neues Rembrandt-Gemälde** wurde entdeckt

The screenshot shows a digital interface for reading historical documents. At the top, there are navigation links for 'Gemeente Amsterdam Stadsarchief' and buttons for 'Browse' and 'Search'. On the right, there are links for 'About' and 'Language'. The main area displays several pages of old, cursive handwriting. A central box contains text about the Amsterdam City Archives and the project 'Crowd Leert Computer Lezen'. It mentions that for the first time, it's possible to search hundreds of thousands of documents from the 17th and 18th century on word level. The text also describes the use of Transkribus Handwritten Text Recognition (HTR) tools for training computer models to decipher historical handwriting. Below this text, there is a link to 'Alle Amsterdamse Akten'. At the bottom of the page, there are logos for 'READ co-op' and 'VeleHanden.nl', along with links for 'Imprint' and 'Privacy'.

**X Gemeente  
X Amsterdam  
X Stadsarchief**

  
**VeleHanden.nl**

# Bibliotheca Hertziana/Max-Planck Institut für Kunstgeschichte

Humanities Connect

read&search-Website

Texterkennung und Strukturanalyse

Änderung der **TEI Export-Funktion + Tagging, doppelte Erkennung von Abkürzungen** (abgekürzt und erweitert)

Aron ein bruder mo ysi vnd höhster bischoff	XVIII	Alkeyro die königlich statt in Egypto wer die ge parot habe	XXII
Abel der ander sone ade vnd der erst mertrer ein ansang d kirche	IX	ltar des opfers wie d gestalt sey	XXXIII
Abgötterey wen vn wo sie vrsprung gehabt habe.	XXIX	Alter der werlt das vierdt anfang	XLVI
Abrahams geschlecht ist anzaigt	XXI XXII	Alter der werlt fünftes anfang	LXIII
Abrahā hat seinen sone ysaac geopffert	XXII	Alexander der gross kaiser regirt tc.	LXXV
Abdon vnd Sennen hohberümbter man zwen	CXX	Alexandria grosse statt in Egypto	LXXVII
Abbas siculus ein mechtiger doctor	CCXLII	Alexandra em weib alexandri des iudische Königs	LXXXVIII
Agter thüngen mit mist hat Saturnus erdacht.	XXXV	Alter der werlt das sechsten anfang	XCV
Achacius ein ritter Cristi	CXII	Alexander der erst babbt ein römer	CXI
Achaia die gegent vnd was darin sey	CCLXXIII	Alexander römischer kaiser	CXVII
Accursius von Florenz der rechten hohgelerter man	CCXI	Alexander ritter aus der schar Thebeorum	CXXIII
<b>A</b> dam wo vnd wie er erschaffen sey	VI	Alexius heiliger beichtiger ein Römer	CXXXV
Adam vnd Eva sind aus dem Paradyes gestossen	VII	Alaricus ein König gothorum	CXXXV
Adam vnd Eva wielang sie gelebt vnd wieviel kinder sie gehabt haben	IX	Alexander ein berümbter arzt	CXXXVI
		Alboinus König Longobardorum was der Wel- sche land gezwang thün habe	CXLVII
		Albinus bischoff zu andegauo	CLXIII
		Alcuinus lehrmaister Caroli des grossen	CLXVI
		Aldegisius ein sone Desiderij ist zu den gezeitten er- nyder gelegen	CLXVII
		Albania die gegent vnd ir gelegenheit	CCCLXXXIII



BIBLIOTHECA HERTZIANA  
MAX-PLANCK-INSTITUT FÜR KUNSTGESCHICHTE

# KaTi: Universität Innsbruck / Land Tirol

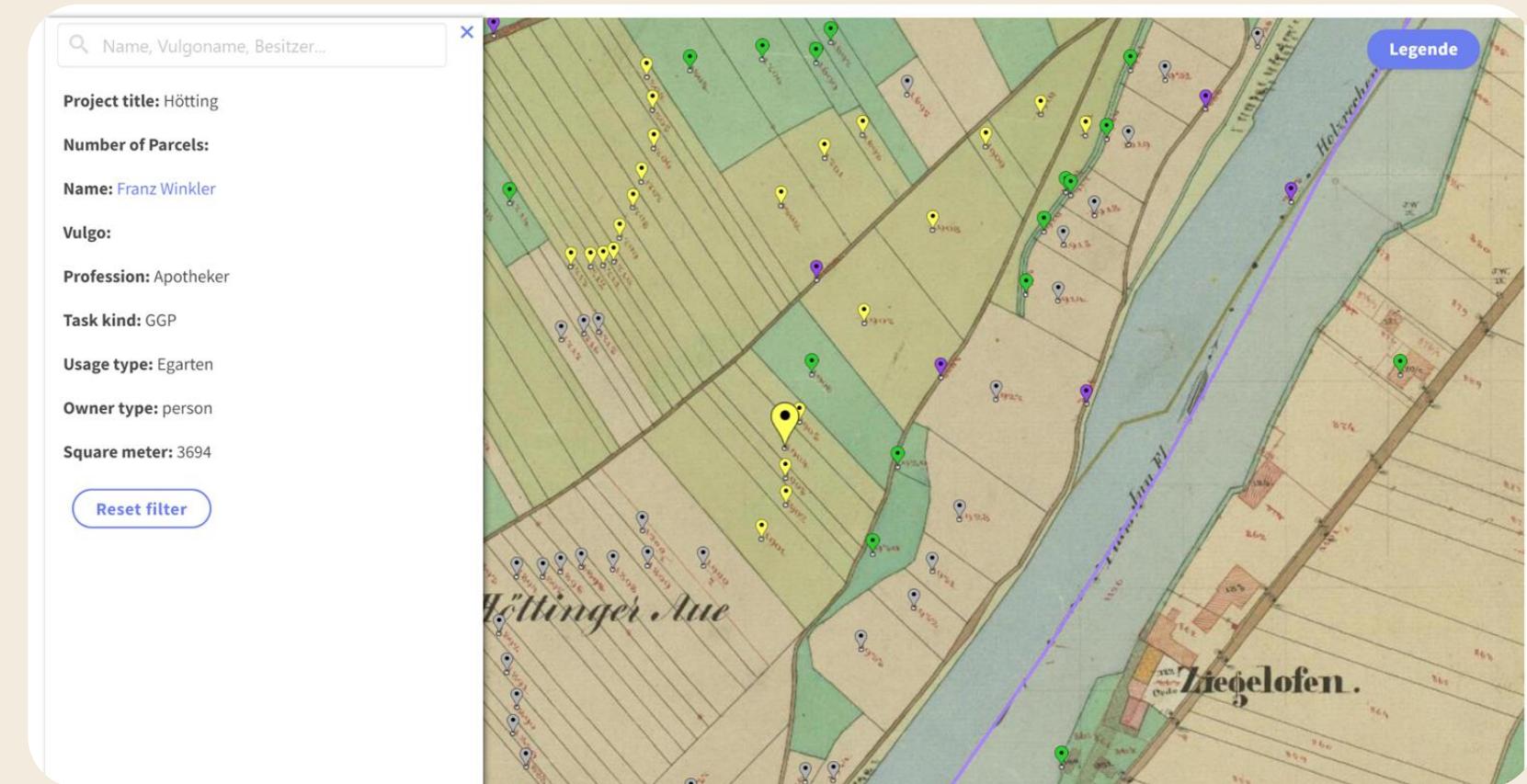
## Citizen-Science / Crowdsourcing

Erster österreichischer **Kataster**  
**in Tirol - 1860-1865**

Citizen Scientists arbeiten mit  
**Grundbüchern (Tabellen) und**  
**Karten**

**mehr als 500.000 Einträge**  
bereits in der Datenbank (=ca.  
30%)

Ziel – Dashboard, das der  
Öffentlichkeit die **Suche und**  
**Visualisierung aller Daten**  
ermöglicht



# EnrichEuropeana+

**Unterstützung der Crowd-Arbeit für das europäische Kulturerbe**

**Weiterentwicklung** der Transcribathon-Crowd-Plattform

**Transkriptions-Editor**

Zugang zu HTR über die **metagrapho API**



# Österreichische Akademie der Wissenschaften + READ-COOP

## Optical Music Recognition

Entwicklung von Musikerkennung  
für mittelalterliche Dokumente

wird Open Source gestellt

Zusammenarbeit mit der  
**Universität von Alicante** (Jorge  
Calvo)

Transkribus Expert Client v1.19.3 (16\_05\_2022\_12:28), Loaded doc: A-VOR253-GT-MUSIC-ANNOTATION, ID: 17392, Page 1, file: A-VOR253-001r.jpg [Image Meta Info: (Resolution:300.0, w\*h: 3247 \* 4407) ] [ curr... ]

Server Overview Layout Metadata Tools

Logout guenter

Document... Find

Document Manager User Manager

Versions Jobs

Recent Documents... User activity

Collections: OEW Music Manuscripts (5104, Owner) Col-ID

Documents Model Data

1-7 / 7 Doc-ID

ID	Title	Pa...	Uploader	Uploaded	Col...
42...	Sample-2 (Bilder-2)	106	guenter	Mon Mar...	(O)
42...	Sample-1 (bilder-1)	394	guenter	Mon Mar...	(O)
42...	TEST_CITlab_OEAW Musi...	4	guenter	Wed Mar...	(O)
24...	A-VOR253_duplicated	492	hubert	Wed Aug...	(O)
23...	TEST_CITlab_OEAW Musi...	6	guenter	Wed Jul ...	(O)
17...	<b>A-VOR253-GT-MUSIC-...</b>	<b>492</b>	<b>guenter</b>	<b>Mon Ma...</b>	<b>(O)</b>
14...	A-VOR253_sample_page	1	guenter	Mon Mar...	(O)

10C Filter

1-1 (c4)-dcde--ge-f--fed--c-ef-g--ga--kj-hg-fe--d-cd-e  
2-1 pulis Moyses et helyas cum domino loquentes  
3-1 (c4)-eg-agfddcdee-jkjhg-fed-efg-fe  
4-1 apparuerunt  
5-1 (c4)-e-d-fg--h-h--g-ef  
6-1 Quoniam Christi evan  
7-1 (c4) e-e--g-e-f  
8-1 (c4)-e-d-fg--h-h--efg-fe-e--g--ghj-kjhg

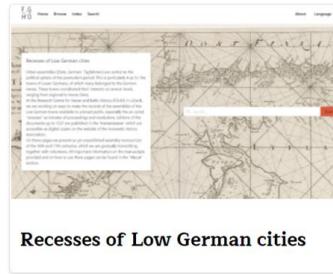
# Viele, viele weitere Projekte



Amsterdam City Archives



National Archives of Sweden



Recesses of Low German cities



Czech and Slovak Fracture  
(Schwabacher)



Grenzregelungsausschuss



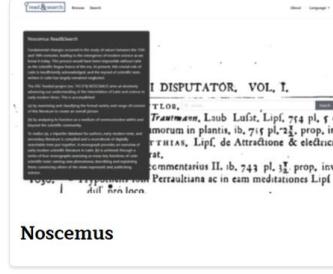
Mining Hub



Marjory Fleming



Club Andino Bariloche



Noscemus



Mountain Hub



Regesta Pomeraniae Monastica



Deutsche Ratsprotokolle 17. bis  
19. Jh. (Bautzen)



National Archives of Finland



New Zealand Alpine Journal



iurisprudentia - Uni Zürich



Tiroler Grundbuchblätter



Historischer Roman



Transkriberte handskrifter

Della 7. 1. 1765  
questo 8. anno  
principiare nell'anno  
di 1765. C.  
li svedetti Paolo Scu-  
ri d'uno a Parva di  
ve Svervali, il quale  
di sani pro tempore  
e iorni. affliggitur  
**Pecunia est** ~~pro~~ ~~cop.~~ ~~l'att~~ in fidei-  
scienz ~~per~~ ~~l'att~~ ~~re~~ ~~re~~  
Palma ~~per~~ ~~l'att~~ ~~re~~ ~~re~~  
mei ~~l'att~~ ~~re~~ ~~re~~  
fiduciamme an-

stato, e con promessa, ad obbligare  
a difendere in forma com-  
petente prezzo di L. 41. 8. 4. diconse-  
vata di Milano, le quali si  
alla rata decorsa come da  
perciò il pref. citt. sub la  
in ogni appurata forma  
ad il suo carone lire  
e titolo del d. prezzo, ad a-

de Martino  
Borghj de Com  
E sit etia quod modo  
ad tentias negotiis  
zanki filij et  
per idcirco de  
ad bene mun. f. 00.

The Next Generation Of

Transkribus®



The screenshot shows a digital transcription of a handwritten document. The top half displays two pages of cursive German handwriting. The left page contains a single paragraph, while the right page begins with a sentence and continues with a second sentence starting with 'des'. Below the text area is a search bar with fields for 'comment', 'type', and 'number', and a 'SEARCH' button. To the right of the search bar are several small icons for navigating the document. The bottom half of the interface features a large, bold title 'Moving Transkribus to the web'.

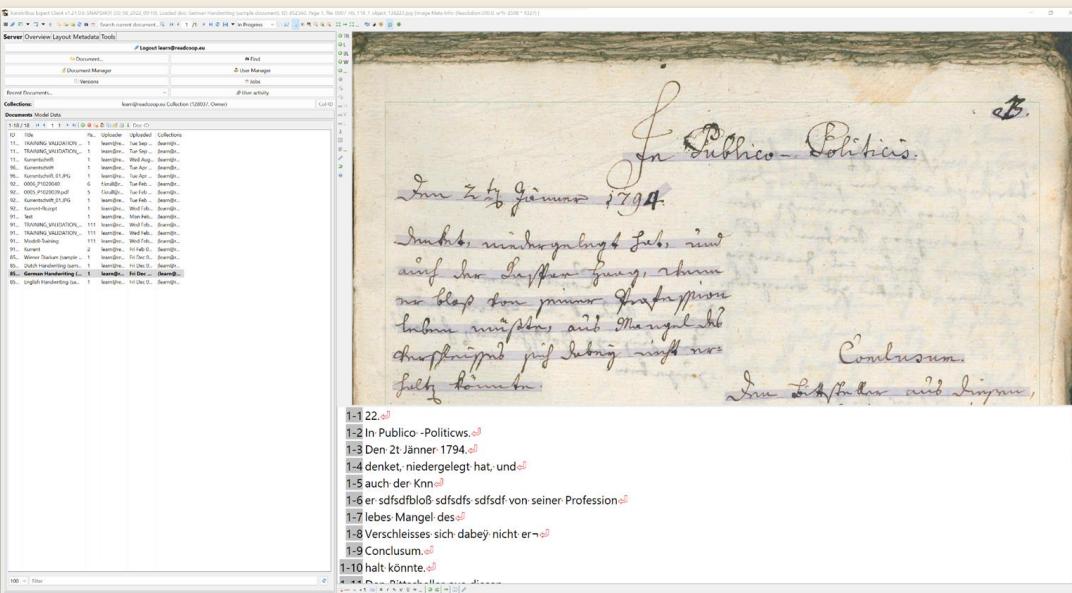
mit noch manchmal zu kommen können), aber  
der Vierfüßerindien mich sie ger nicht. Von den  
steigstrichen im Greger Bett bis zu dem  
aufrechten Verstecken auf dem Herkunftsplatz  
ist es so weit, dass es mir die Phantasie Kugge  
überwindet, aber das wird ja erst die allgemein-  
heitlich darüber hinweg, etwas die Vorstellung, zu  
verringeren, dass ich z. B. in Venedig in Gesellschaft  
mit Tagessie (ich kann mir allein eben) da ver-  
trete. + und es ist Bla. Taus. Über ihm werden

aber Vorfel nach meine Erfahrung bei ihrem letzten  
Proces kommt sie nicht wieder kommen, das, wie ich  
ja weiß ich hätte Ihnen gern schon geschildert, wenn  
mir nicht das Brief-Schreiben allmählich so schwer  
würde wie das Reden und wenn nicht sogar das  
Brief-Schreiben schwieriger machen würde,  
denn einen Brief hätte ich für Sie schon fertig, &  
ist aber immer, alte Dinge werden nicht anders.

sz uurichtlich aus kommen sönnen, als  
nit  
des Caupthinkerug mich sie ger nicht. Von den

# Moving Transkribus to the web

# So we went from



From this

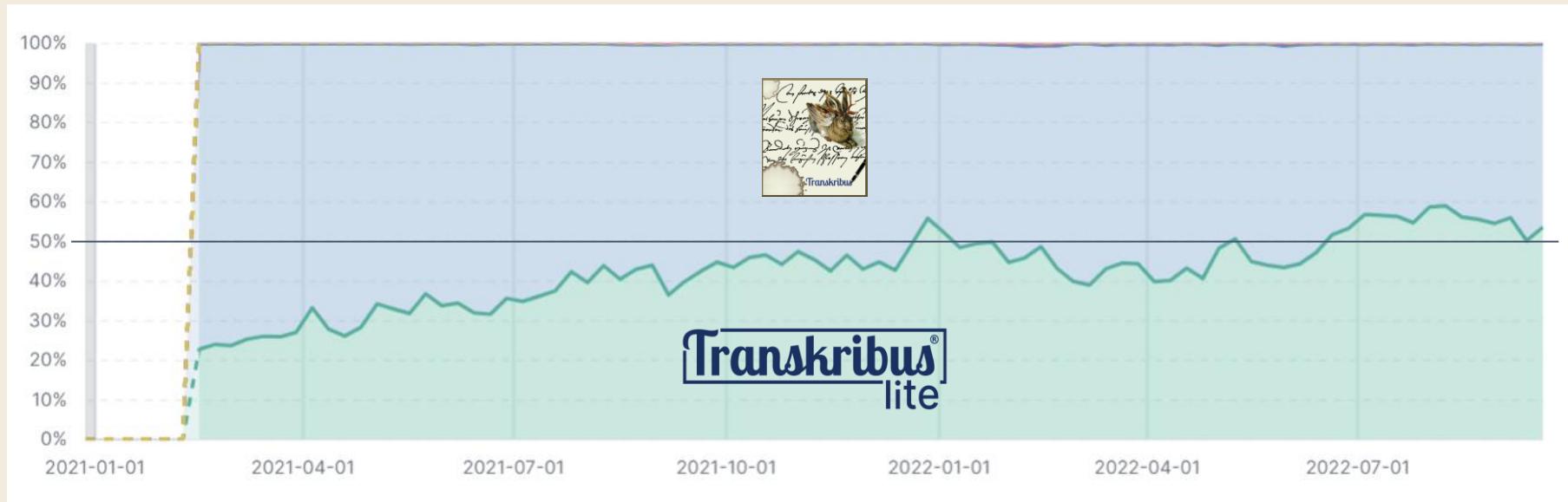
In Publico - Politici.  
Den 2t Jänner 1794.  
Dankt, niederlegt & lebt, und  
auf den Leibn zu gieß, schmeck  
so biss der Junius. In Publico  
leben wirkt, und Mangelsch  
durchgängig ist lebend nicht nur  
fertig kommt.

Conclusum.  
Den 2t Jänner 1794.

1-1 22.  
1-2 In Publico - Politici.  
1-3 Den 2t Jänner 1794.  
1-4 denkt, niedergelegt hat, und  
1-5 auch der Knn.  
1-6 er sdfsdflob sdfsdfs sdfsfv von seiner Profession.  
1-7 lebes Mangel des.  
1-8 Verschleiss sich dabē nicht er.  
1-9 Conclusum.  
1-10 halt könnte.

To this

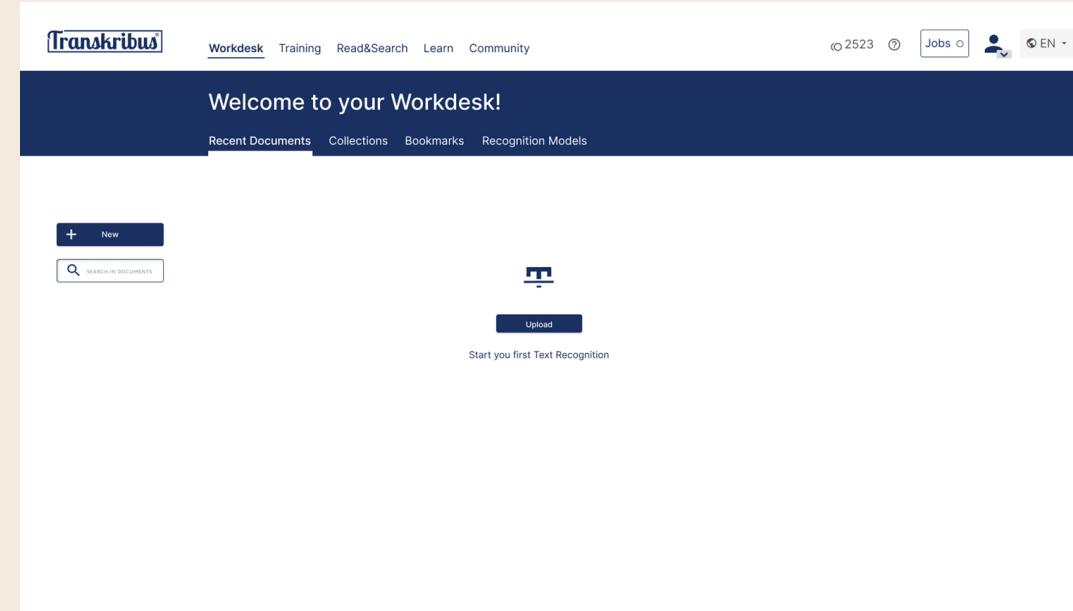
# Increasing importance of our web application



More than 50% of Transkribus  
Users use **Transkribus** Lite

It's just

# Transkribus®



[beta.transkribus.eu](https://beta.transkribus.eu)

Making Transkribus simple *and* powerful

# Transkribus Learn

← Back to collection

Cooking book (Sütterlin), 1926

Page 1

1 / 5

Text Text Enter word Text

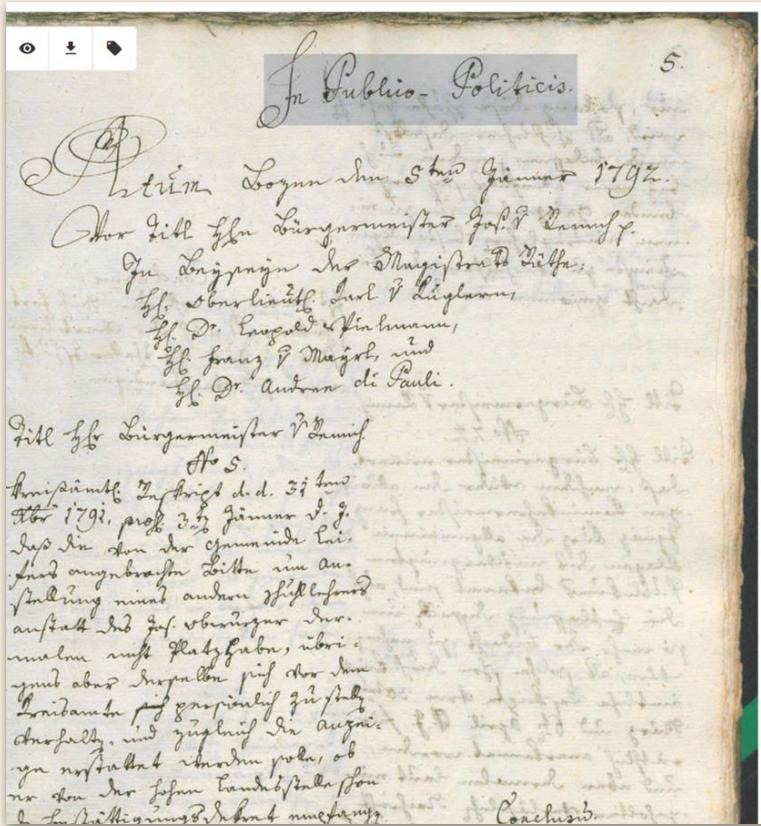
Check word

Score: 0/5

Learn how to read old  
handwriting with Transkribus

- Set and edit tasks
- Choose words or select randomly
- Statistics for tasks
- Learn groups

# read&search CMS



5  
In Publico- Politicis.  
74  
Actum Bozen den 5 ten Jänner 1792  
Vor Titl Herrn Bürgermeister Jos. v. Remich et.  
In Beyseyn der Magistrats Räthe:  
Herr Oberlieut. Karl v. Kuglern,  
Herr Dr. Leopold Spielmann,  
Herr Franz v. Maýrl, und  
Herr Dr. Andree di Pauli.  
Titl Herr Bürgermeister V Remich.  
No 5.  
Kreisämt. Reskript d. d. 31 ten  
Xbr 1791, praes. 3 ten Jänner d. J.  
daß die von der Gemeinde Lei= fers angebrachte Bitte um An= stellung eines andern Schuhlehrers anstatt des Jos. Obwurzer der malen nicht Platzhabe, übri= gens aber derselbe sich vor dem Kreisamte sich persönlich zustellen.  
verhalten, und zugleich die Anzei= gen erstattet werden solle, ob er von der hohen Landesstelleschon Conclusu.  
das Bestätigungsdekret empfangen  
66  
Ist dem Anwalt zu Leifers

Publish your material online with  
read&search

- Lightweight CMS System
- Publish your Transkribus Collections
- Powerful search features

Transkribus®

# metagrapho<api>

Processing API for text recognition

# Processing API

- First use case: EnrichEuropeana+
  - Integration of Transkribus text recognition in the Transcribathon platform
- Simplifies HTR processing for single images
  - “Image in, layout+text out”
  - No persistent storage of image data
  - No document/collection management
- Technical documentation according to OpenApi Specification v3 (Swagger UI: <https://transkribus.eu/processing>)
- Many text recognition models available
  - Your private models trained in Transkribus
  - All public models <https://readcoop.eu/transkribus/public-models/>
- Already featured in various applications
  - <https://transkribus.ai>
  - CCS docWizz
  - Goobi
  - Europeana Transcribathon



The Dean stated, that this meeting was called in consequence of a letter he had received from several members, which he desired might be read. It was accordingly read by the clerk and is as follows:

"Edinburg, 25 Nov. 1804  
Sir, - We request that you will call a meeting of the Members of the Faculty, for the purpose of considering a bill lately brought into the House of Lords, entitled, 'An Act touching the Administration of Justice in Scotland, and touching Appeals to the House of Lords.'



The Dean stated, that this meeting was called in consequence of a letter he had received from several members which he desired might be read. It was accordingly read by the clerk and is as follows.

"Edinburg, 25 Nov. 180

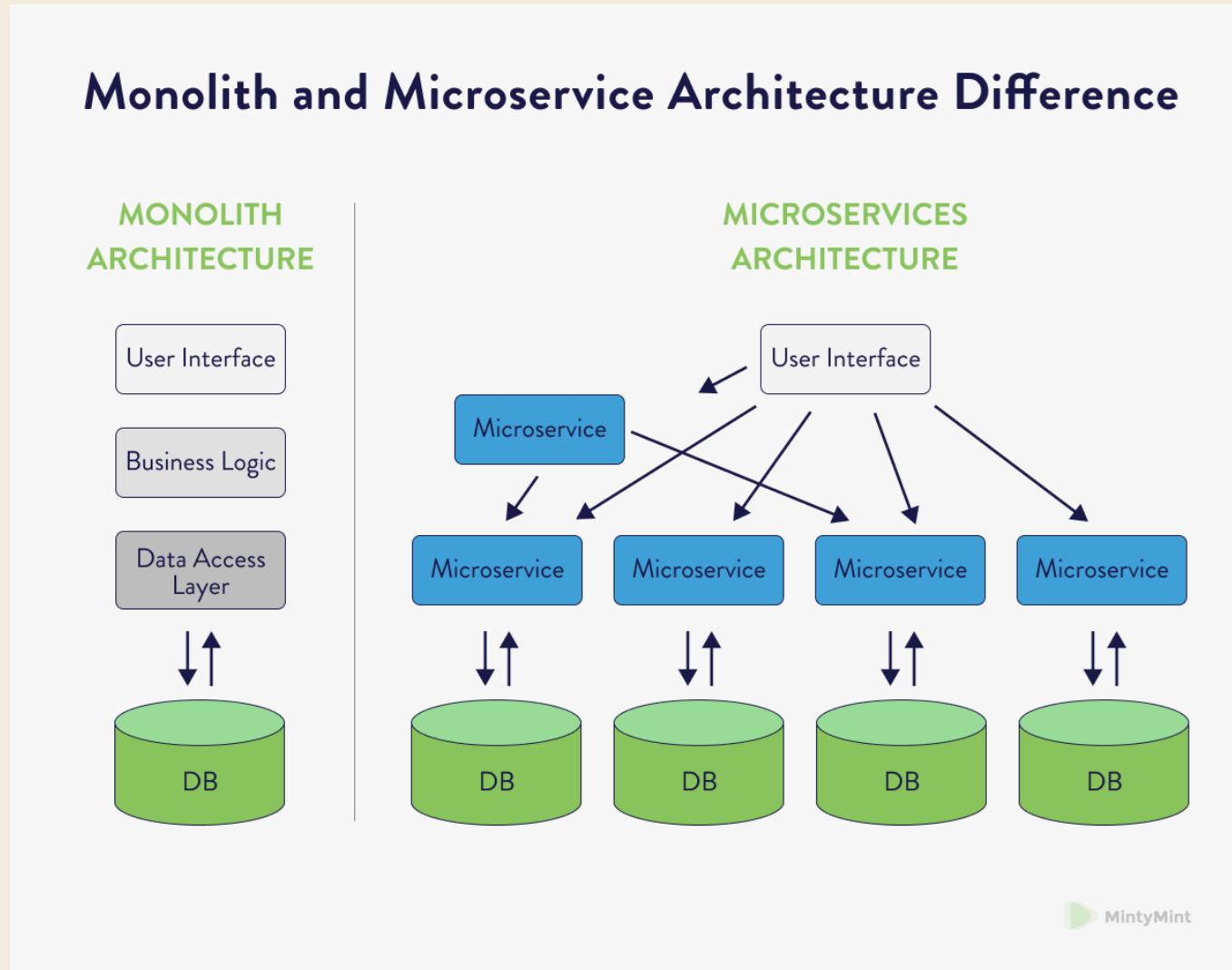
11

Sir, - We request that you will call a meeting of the Members of the Faculty, for the purpose of considering a bill lately brought into the House of Lords, entitled, An Act touching the Administration of Justice in Scotland, and touching Appeals to the House of Lords.

# Transkribus®

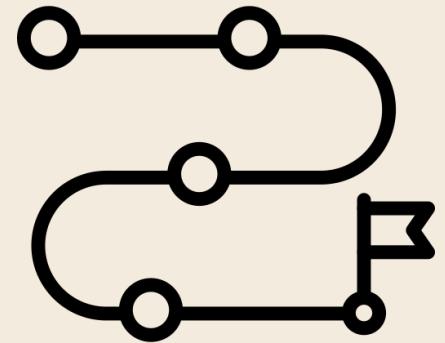
## On-Prem

# Current work on architecture: Microservices

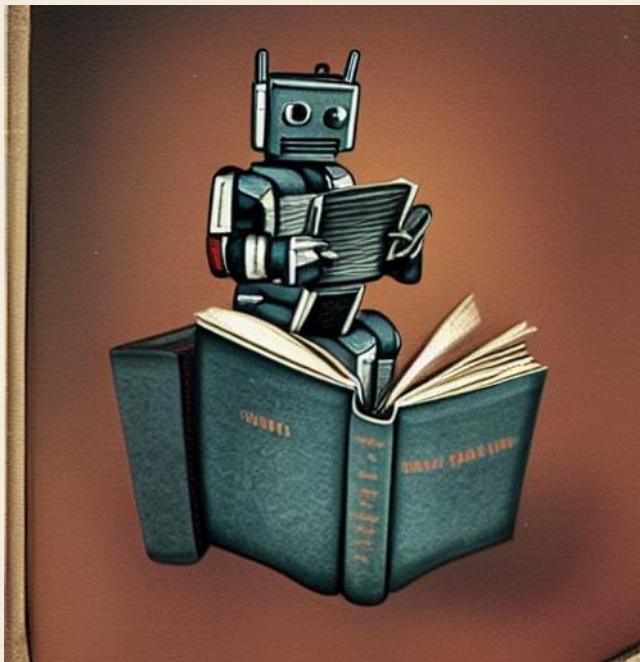


# Rough Roadmap

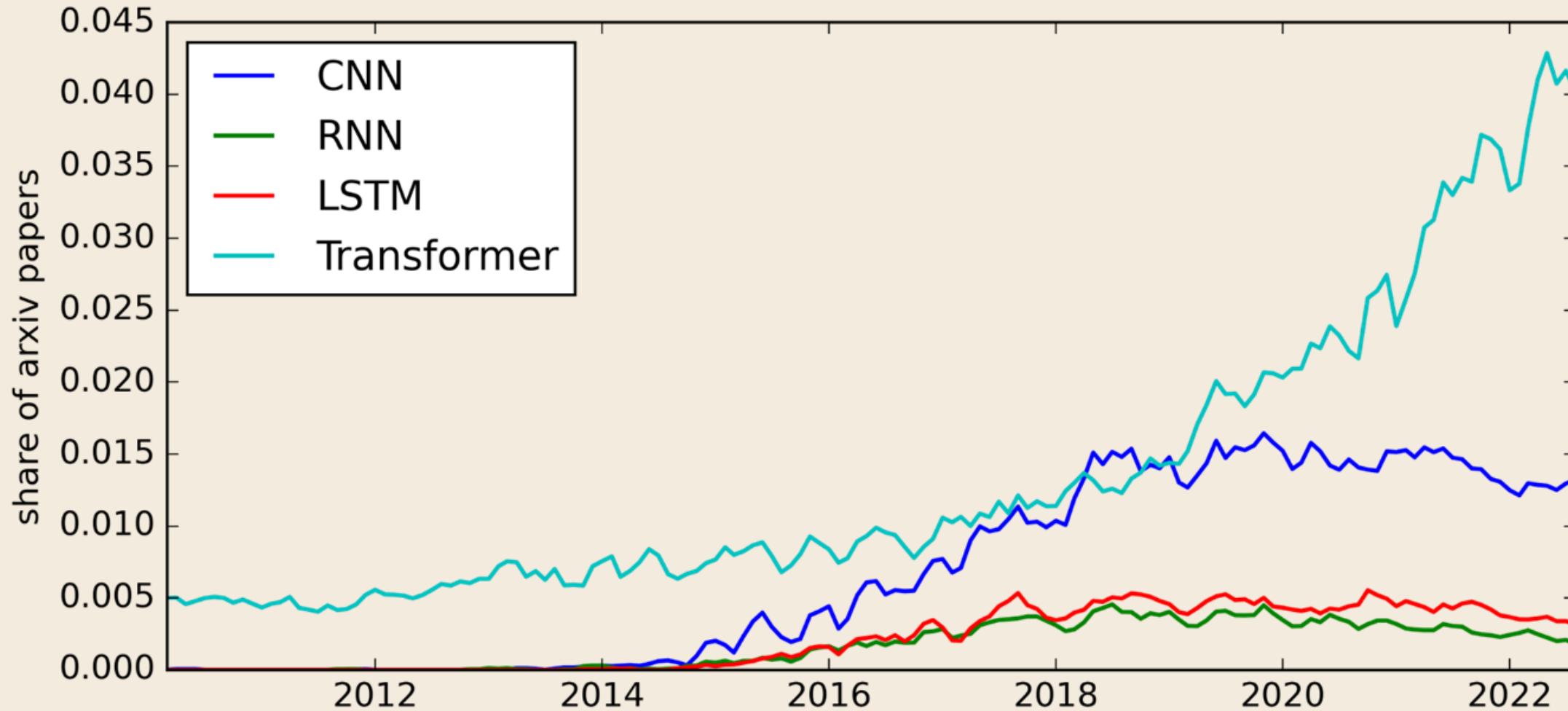
- First phase (completed)
  - Client implementation as production-grade collection of scripts + local java GUI
- Second phase (ongoing, Q4 2022/Q1 2023)
  - Docker images for core services
  - Install in Kubernetes environment
- Third phase (sometime during 2023):
  - Integration of additional GUIs: Transkribus (web version) and r&s



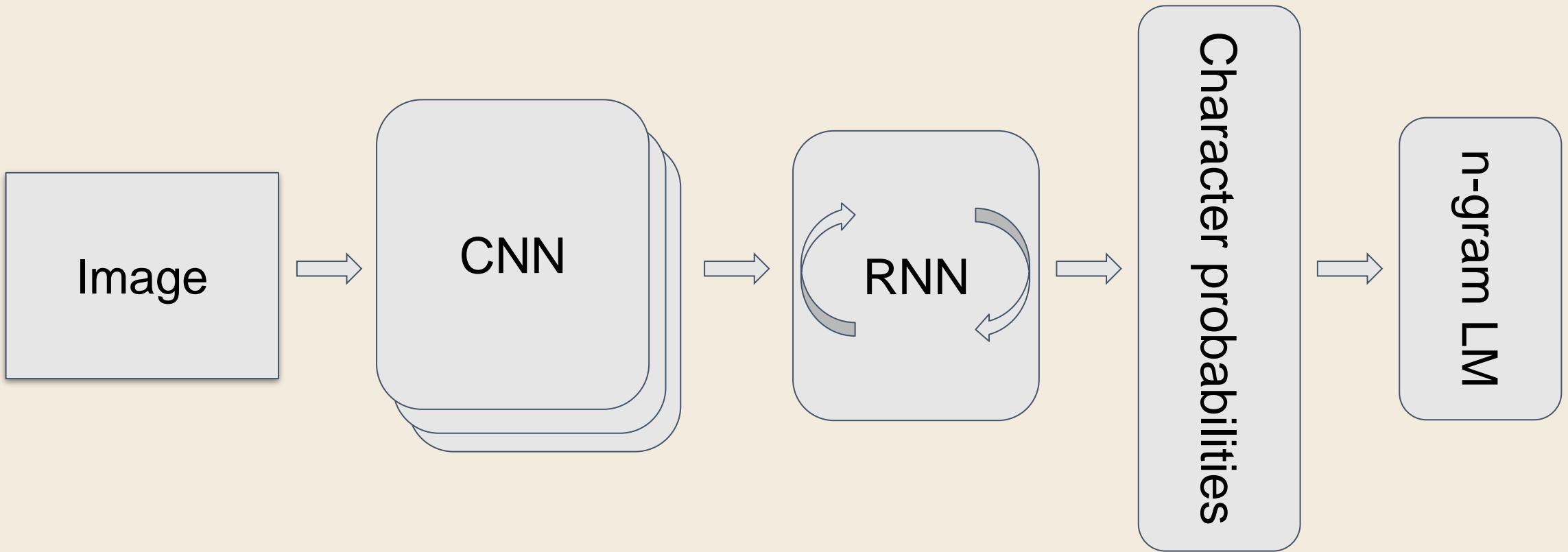
# Transformer Models



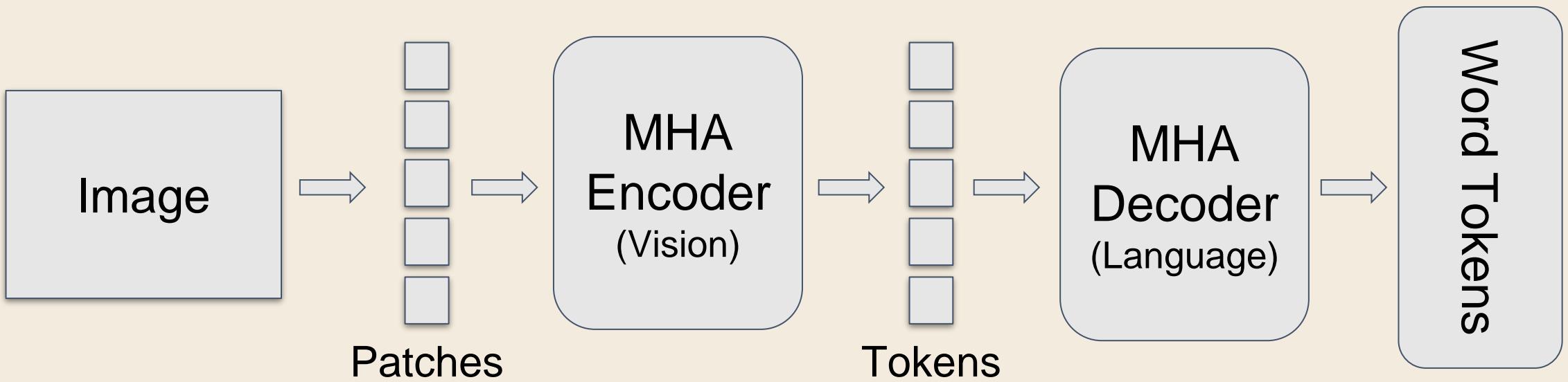
# State of research



# Classic HTR architecture (e.g. PyLaia)



# Transformer HTR



# First trHTR model now available in Transkribus!

- One big general-purpose model
- Trained on gigantic dataset
- 6+ Languages
  - English, French, German, Finnish, Swedish, Dutch
- Ideal for trying out recognition or creating GT
- Rollout has started

# Trainable Layout Analysis & Tables & Forms Recognition

ORDEN NUMMER.	Dagtekening VAN ONTVANGST. Nog 2	INZENDER.	Dagtekening EN NUMMER VAN HET INGEZONDEN STUK.	GETAL BIJLA- GEN.	ONDERWERP.	
Mer. werken indien geestes- toevoer elkaar	33	2 febr. Kerkvoordien drecht. Hore Goeie	2 febr.	-	Betrekking dient voor de verlengde toelating tot verplaatsing der gewichten van der toren toren en verwerk der Grote Kerk en vertrekken postkantoor op de door hem medezen aangevraagde staats Polleke/ geschaffen regeling ook ten aanzien van den tijl dat het verwerk dat moet schilderen, gedurende de uitvoering van het werk.	-

Dagtekening der beschuit.		UITVOERING.	Verzending van stukken.		Vervijf- zing naai- verigo of volgends nummers van het register.	Anmerkingen.
VAN BURG. EN WETH.	VAN DEN RAAD.		AAN WIEN.	WANNEER.		

# Trainable LA + Tables & Forms Recognition

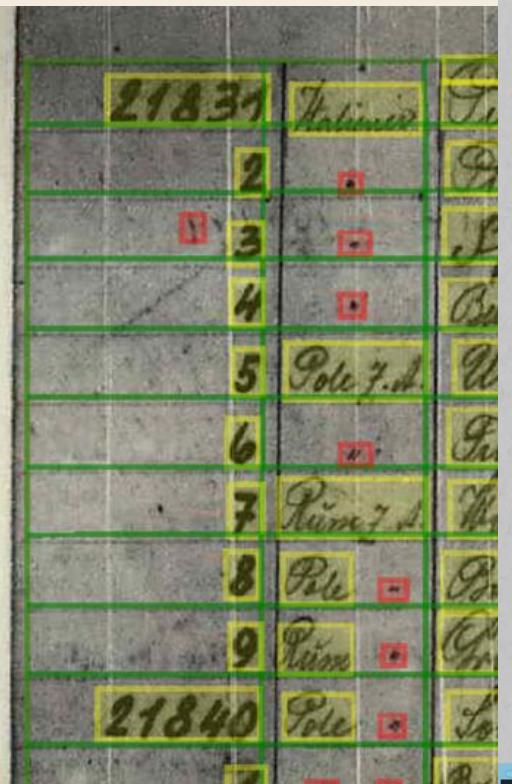
- New Layout Analysis (LA) coming to Transkribus
  - Not general purpose, but trained on certain collections / documents
- Based on the open-source Detectron2 framework
- Allows for different kinds of segmentation
  - Semantic
  - Instance



**Detectron2**

# Form segmentation

- Trainable LA based on existing PageXML's
- Neighboring instances can be segmented





NUMMER welke ieder STUK Ontvangt.	DATUM der Registratie.	NUMMER der ontvangene STUKKEN.	AANWIJZING van het OBJEKT van IEDER STUK.	AUTORITEIT of Particulier die de ontvangene Stukken heeft geschreven of aan wie men schrijft.	ZAKELIJKE inhoud van IEDER STUK.	Aan wie de zaak is Gedemandeerd.	DISPOSITIE.	Bemerkingen.
						INHOUD.	INHOUD.	
16	1891	100	100	100	100	100	100	100
17	1891	101	101	101	101	101	101	101
18	1891	102	102	102	102	102	102	102
19	1891	103	103	103	103	103	103	103
20	1891	104	104	104	104	104	104	104

# READ co · op

Unlock our written past.