

Bozner Ratsprotokolle transkribiert ✓

„Crowdsourcingprojekt“

Der Bestand: „Ratschlagbücher“

- Protokolle des Bozner Stadtrates (16.-19.Jh.), hier Ende 18. Jh.
- barocke deutsche Kurrentschrift, sauber und regelmäßig, ein Schreiber, Digraphie, viele Ziffern
- verständliches Neuhochdeutsch, typisches Amtsdeutsch, Latinismen
- **Erschließung:**
 - Vorhanden: 1470er-1840er (ca. 1-2 Bd./Jahr) + Register
 - Digital auf „BOhisto“*: 1470-1805 + Register
 - In Transkribus: 1792 – 1805 + Register (32 Bd., à ca. 650 Seiten)
- **Herausforderungen:**
 - Informationsdichte nicht durchsuchbar!
 - Typisches Layout (zwei Spalten, Überschriften, Marginalien)

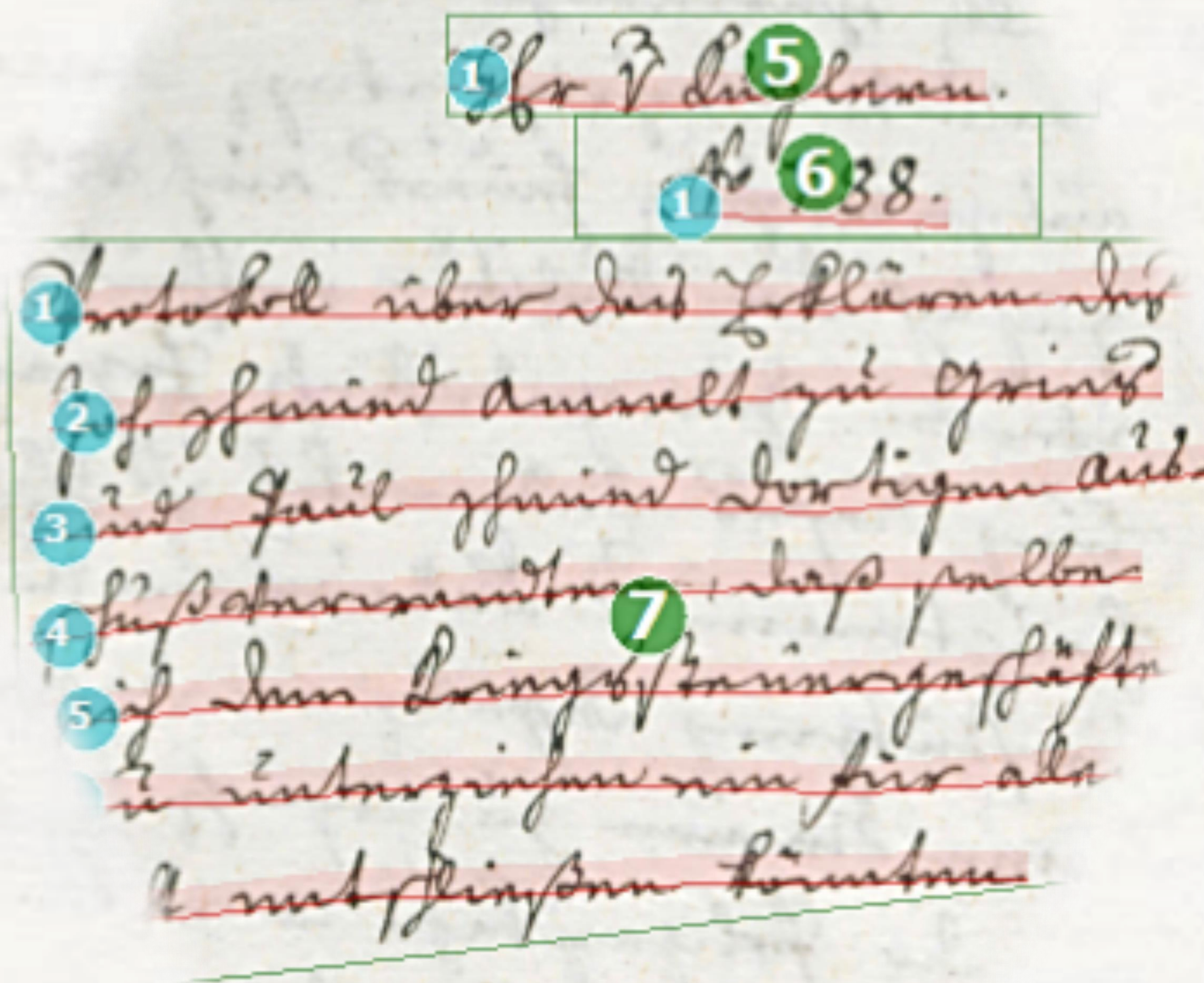
* Seite des Bozner Stadtarchivs: Bozen-Bolzano's History Online, <http://stadttarchiv-archivistorico.gemeinde.bozen.it/>



„Alter Bozner Ratssaal“ (Foto: Gemeinde Bozen)



Infotreffen im Stadtarchiv, Sept.. 2016 (Foto: Denicolò)



Das Projekt: 2016 – 2018

Forschungsplattform Transkribus* (<https://transkribus.eu/Transkribus/>)

Doktorandin à 6 Stunden/Woche

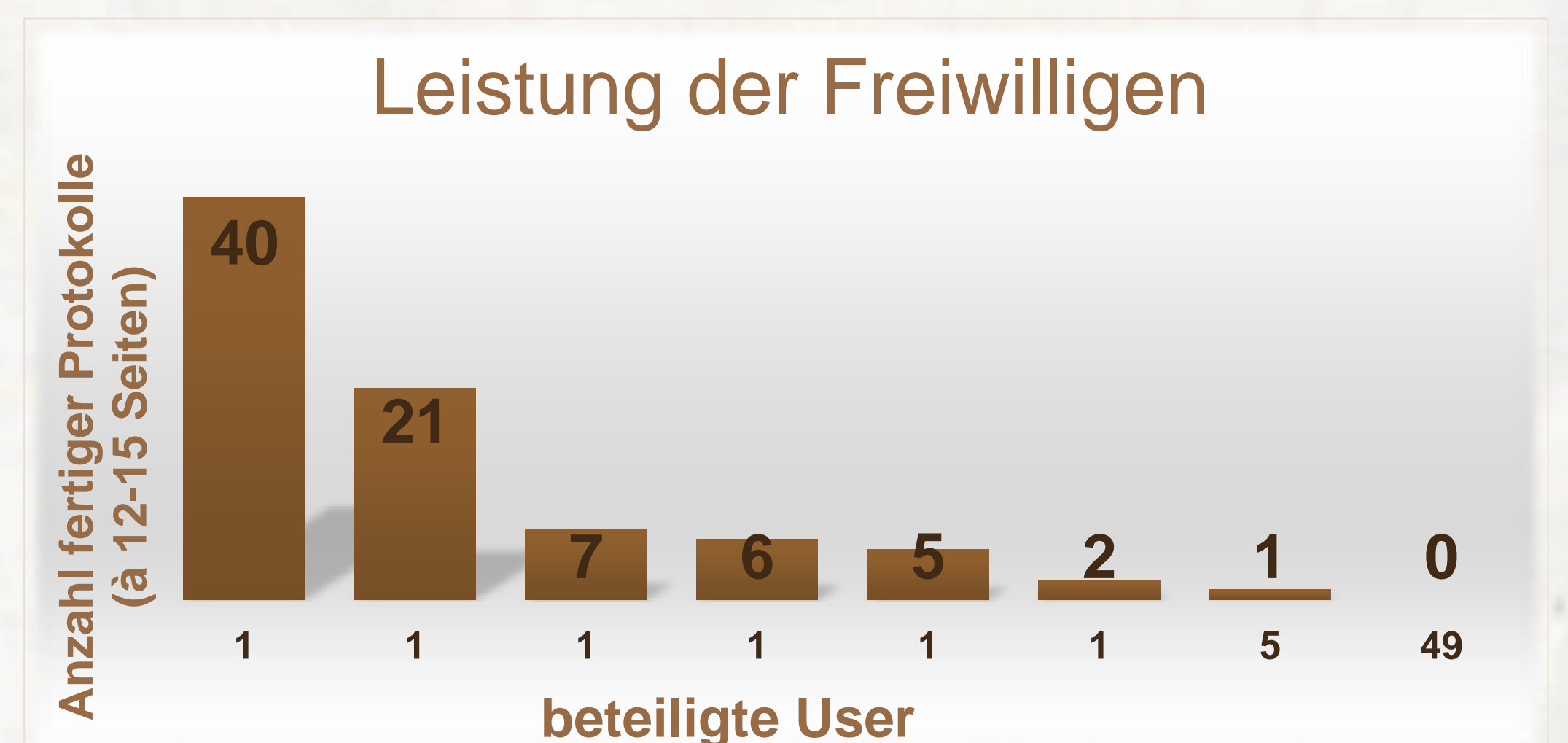
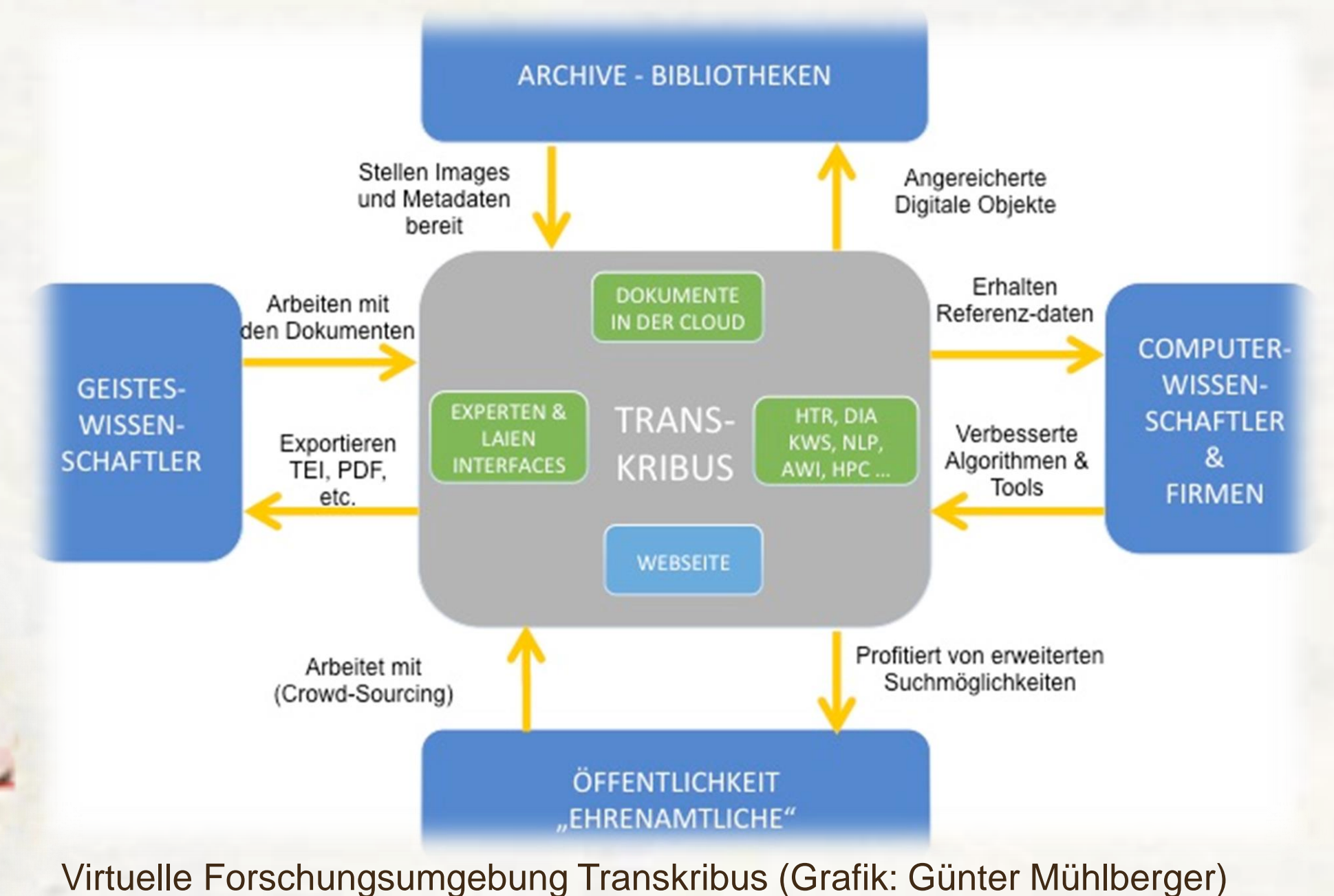
DEA + READ: „Recognition and Enrichment of Archival Documents“

- Transkriptionen dienen als Trainingsmaterial für HTR
- Transkribus als Crowdsourcing-Tool weiterentwickeln
- Transkribus Expertinterface verbessern

Stadtarchiv Bozen:

- Erschließung und Aufwertung der Digitalisate
- Neues, partizipatives, digitales Angebot

* vgl. Kahle/Colutto/Hackl/Mühlberger, transkribus – a Plattform for Transkription, recognition and Retrieval of Document Images (2017)



Die „Crowd“

- **ca. 60 Beteiligte**
 - (davon abgemeldet – auf Nachfrage! 11)
 - 38 Studierende, 17 Berufstätige mittleren Alters, 4 SeniorInnen
 - Technikbegeisterte, HobbyhistorikerInnen, FamilienforscherInnen OrtschronistInnen
 - Ca. 25 aus Nordtirol (Innsbruck, Umgebung), ca. 20 aus Südtirol, ca. 5 aus Ludwigshafen D, Ravenna I, Vorarlberg, Wien, Graz)
- **Rekrutierung**
 - Universität, Facebook, Tageszeitung, Institutionen, Mundwerbung
- **Tätigkeit**
 - Segmentieren - Transkribieren – Korrigieren - Taggen
- **Angebote**
 - Infotreffen, „Offenes Arbeiten“, Übersetzung eng. Begriffe, HowTo-Papers, Checklisten etc., Google- und fb-Gruppe, eLearning
- **Unterstützung**
 - Ground Truth erstellen, Anwerbung von Freiwilligen, Organisation der Freiwilligen und ihrer Leistungen, Kommunikation und Korrespondenz, Transkriptionsrichtlinien aktualisieren, Korrektur und Feedback an die Freiwilligen, Normalisieren der Tags etc.
- **Herausforderungen**
 - Große Entfernungen, unterschiedliche Zielgruppen, Bedürfnisse und Ressourcen, fehlende Englischkenntnisse, wenig bis gar keine Zeit

Ergebnisse

„Crowd“

- Transkribiert: ca. 1500 Seiten, ca. 180.000 Wörter
- Tags: 3948 Personen, 950 Orte, 1954 Daten, 4439 Abkürzungen, 2566 andere Tags (Organisationen, Einfügungen, Unklarheiten...)

READ

- Neuronale Netze zur automatischen Handschriftenerkennung (HTR)
- Trainingsmodell (GT ca. 44.500 Wörter): 9 von 10 Buchstaben richtig
- Dictionary mit 10.000 Wörtern
- Feedback zur Benutzerfreundlichkeit von Transkribus

Stadtarchiv

- Übergabe der angereicherte Daten als XML bzw. ALTO
- Weiterverarbeitung/Präsentation der Ergebnisse geplant

Fazit

- „Der/die ideale Freiwillige“ ist eine Seltenheit
- Anwerbung und Kommunikation zeitaufwendig
- Einstiegshürden durch Expertclient (Vereinfachung durch Webinterface)
- Fehlende öffentliche Sichtbarkeit dämpft Motivation
- Bahn frei für großangelegte automatische Erkennung und Erschließung
- Mehr Motivationsmaßnahmen und Anreize
- Effektives HTR-Modell bringt Zeitersparnis und Lesehilfe für EinsteigerInnen
- Quelle für Episoden aus dem (Verwaltungs-)alltag einer Kleinstadt