

FORUM

Geodätisch.
Frei.
Beruflich.

46. Jahrgang
2020

ISSN 0342-6165

Zeitschrift des Bundes der Öffentlich bestellten Vermessungsingenieure e.V. | www.bdvi-forum.de

HEFT 4/2020

Zukunft

Neue Wege beim
Nachwuchs II

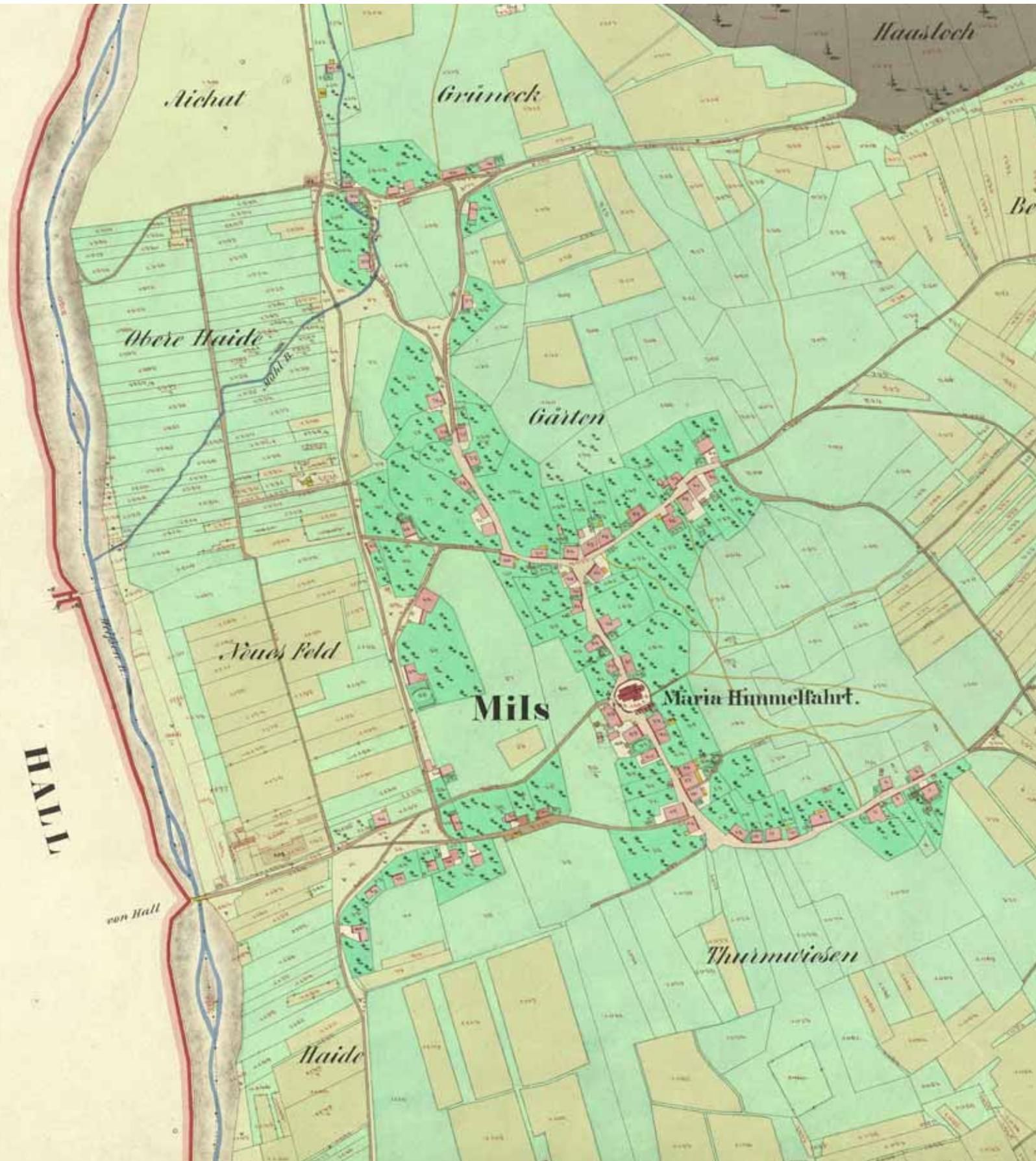
Potenzial

Transkribus
in der Praxis

Diskussion

Wettbewerb
um Anwärter







Transkribus für Katasterdokumente

Ein Werkstattbericht

GÜNTER MÜHLBERGER | INNSBRUCK

Im EU-Projekt *READ (Recognition and Enrichment of Archival Documents)* wurde die *Transkribus-Plattform* entwickelt, mit der handschriftliche Quellen automatisch texterkannt werden können. Diese Technologie war der Anlass, sich der systematischen Erschließung des Franziszeischen Katasters in Nordtirol und hier wiederum den handschriftlich verfassten Begleitdokumenten zu widmen. Auf Basis der automatisierten Handschriftenerkennung sowie einer *Digital-Science-Plattform* sollen sowohl die Katasterkarten als auch die Begleitdokumente bis Mitte 2021 für eine breite Öffentlichkeit erschlossen werden.

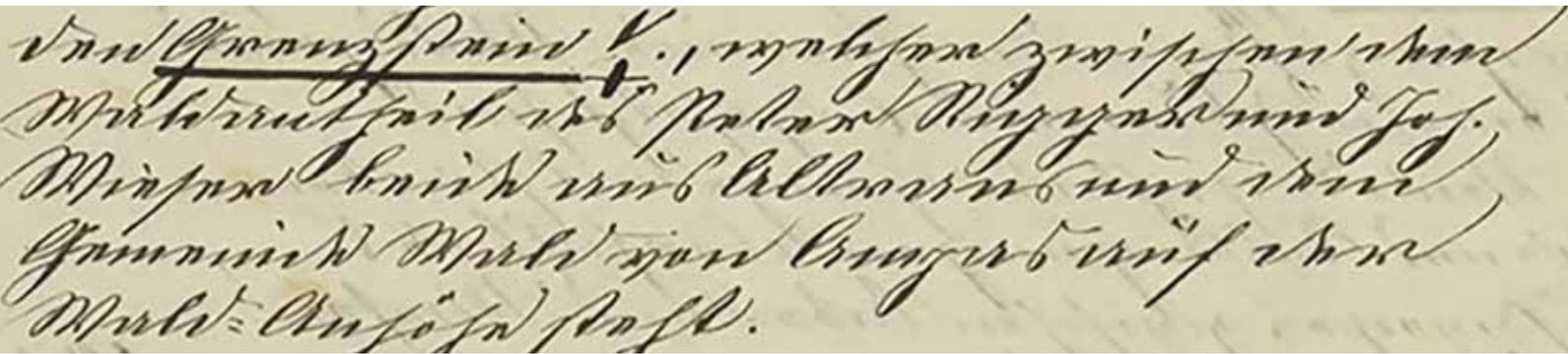


Abbildung 1 | Textausschnitt aus einer Grenzbeschreibung. Die Fehlerquote liegt ungefähr bei 5 % Character Error Rate auf Buchstabenebene.

Das – oder wie man in Österreich sagt – der Franziszeische Kataster ist der erste vollständige Liegenschaftskataster der Donaumonarchie. Er ist nach Kaiser Franz I. benannt, der die Katastervermessung durch das Grundsteuerpatent vom 23. Dezember 1817 veranlasste. Entstanden in den 1810er- bis 1870er-Jahren weist er die Grundstücke des Gebiets des Kaisertums Österreich mit dem Ziel nach, eine einheitliche Basis für die Bemessung der Grundsteuer zu schaffen. Im Kronland Tirol-Vorarlberg wurde der Franziszeische Kataster (»Urmappe«) in mehrjähriger Arbeit zwischen 1851 und 1861 im Maßstab 1:2880 aufgenommen.

WEITERFÜHRENDE LITERATUR ZUM FRANZISZEISCHEN KATASTER

Susanne Fuhrmann

Digitale Historische Geobasisdaten im Bundesamt für Eich- und Vermessungswesen (BEV).

Die Urmappe des Franziszeischen Kataster.

In: Vermessung & Geoinformation 1/2007, S. 24–35, www.ovg.at/de/vgi/files/fullpdf/1245/



Kurt Scharf
Kataster und Grundbuch
im Kaisertum Österreich,
Ausgangssituation und
Entwicklung bis 1866.

In: 200 Jahre Kataster –
Österreichisches Kulturgut
1817–2017, Wien 2017, S. 37–51,
www.bev.gv.at/pls/portal/url/ITEM/60D9411002373COCE050010A1F2116BD

Kurt Scharf
Der Franziszeische Grund-
steuerkataster 1817–1866.

In: Tiroler Chronist 2013/1,
Nr. 128, S. 13–17,

www.bev.gv.at/pls/portal/url/item/FB149146CB80862CE040010A1F210809

1 | EINLEITUNG

Mit Unterstützung der Tiroler Landesregierung und ihres Programms »Leuchtturmprojekte im Bereich Digitalisierung« [1] haben die Universität Innsbruck sowie die Europäische Genossenschaft READ-COOP SCE [2] in Zusammenarbeit mit dem Bundesamt für Eich- und Vermessungswesen, dem Landesarchiv Tirol, dem Bildungsservice des Landes Tirol und dem Forschungszentrum Digital Humanities der Universität Innsbruck das Projekt »Kataster Tirol – digital (KATI-digital)« [3] Anfang 2019 gestartet. KATI-digital läuft bis zum Sommer 2021 und wird dann von der READ-COOP SCE weiterbetreut und fortgeführt werden. Die Projektbetreiber hoffen, dass sich weitere Bundesländer in Österreich, aber auch Verwaltungseinheiten der Habsburger Nachfolgestaaten dem Projekt anschließen, sodass eine umfassende Erhebung der Daten erfolgen könnte.

Der inhaltliche Ausgangspunkt für KATI-digital sind die Detailvermessungsoperatte (DVO) oder Katasterbegleitdokumente, wie sie bei der Erstellung des Franziszeischen Katasters zusammen mit den gezeichneten Urmappen erstellt wurden. Diese Dokumente enthalten alle wichtigen Informationen, wie man sie von einem (modernen) Grundbuch erwartet, und bilden zum Teil heute noch eine wichtige Rechtsgrundlage. Die Zugänglichkeit der Daten sowohl für die Wissenschaft als auch für die breite Öffentlichkeit zu ermöglichen, ist das eigentliche Ziel des Projekts.

2 | QUELLENMATERIAL

Im Projekt KATI-digital werden zwei unterschiedliche Quellenmaterialien bearbeitet: erstens die bereits erwähnten Detailvermessungsoperatte (DVO). Sie machen den Hauptteil aus und enthalten in tabellarischer Form Informationen zum Kataster wie etwa die Parzellennummer sowie den Namen des Besitzers. Neben den zwei wichtigsten Dokumenten, dem Bau- und Grundparzellenprotokoll, befinden sich darunter auch Grenzbeschreibungen der jeweiligen Katastralgemeinde, alphabetische Listen von Haus- und Grundeigentümern sowie zusammenfassende Statistiktabelle über die Bodennutzungsarten. [4] Diese Do-

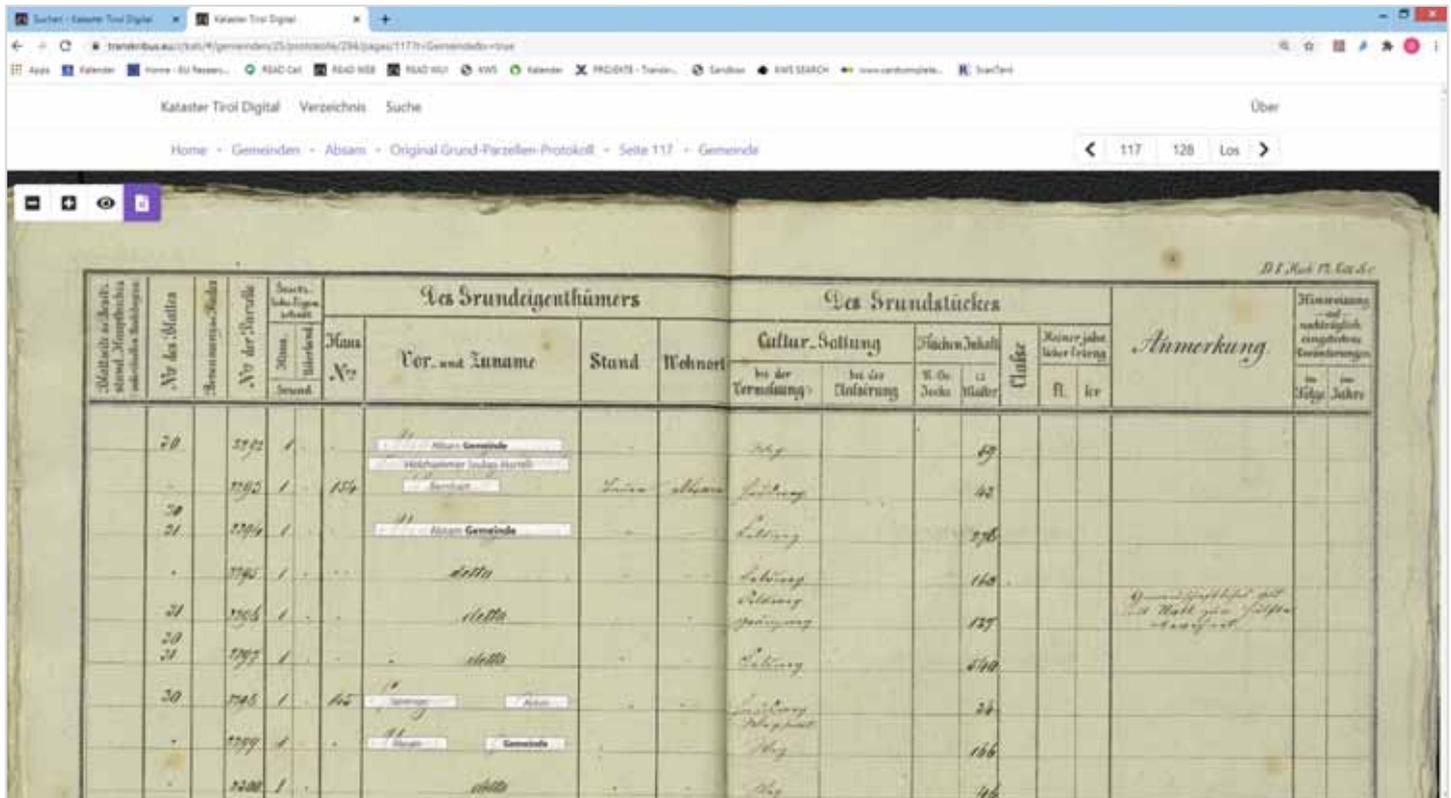


Abbildung 2 | Anzeige des Textes als Überblendung – zu- und abschaltbar

kumente liegen vollständig für alle 280 historischen Katastralgemeinden in (Nord-)Tirol vor und wurden vor der eigentlichen Digitalisierung durchgesehen und verzeichnet. Bei der Digitalisierung selbst wurden alle Dokumente, die zu einer bestimmten Gemeinde vorhanden waren, gescannt, also auch alle Kopien, Nachträge und Zusätze. Insgesamt wurden über 90.000 digitale Seitenbilder erstellt, der allergrößte Teil davon doppelseitig, sodass von rund 180.000 Einzelseiten für (Nord-)Tirol ausgegangen werden kann.

Die zweite Gruppe an Quellen umfasst die eigentlichen Mappen, also die jeweiligen Kartenblätter des Franziszeischen Katasters. Für Tirol in den heutigen Grenzen liegen mehr als 3.600 Mappen vor. Um sie in einem Geographischen Informationssystem (GIS) verwenden zu können, müssen die einzelnen Blätter zu einem Gesamtbild zusammengefügt und georeferenziert werden. Dies wurde von der Tiroler Landesregierung im Rahmen des Projekts »Historische Karten Tirols« [5] für einige Testgemeinden (darunter Innsbruck und Umgebung) durchgeführt.

Für unser Projekt entscheidend ist jedoch, dass die Mappen über eine Programmschnittstelle des ArcGIS-Servers [6] der Tiroler Landesregierung zugänglich sind und auch die historischen Katastralgemeindengrenzen hinterlegt sind. So können die Mappen unmittelbar vom Transkribus-System aus technisch angesprochen und mit den im Projekt entwickelten Programmen auch visualisiert und bearbeitet werden.

3 | KATI-DIGITAL

Bei der Konzeption des Projekts sind wir davon ausgegangen, dass sich das Projekt vor allem an drei Zielgruppen richten wird: (1) Wissenschaftler aus den Bereichen Geografie und Wirtschaftsgeschichte, (2) Vermessungstechniker und Rechtsanwälte (bis heute muss in manchen Streitfällen auf die Grenzbeschreibungen bzw. die Urmappen zurückgegriffen werden) und (3) Familienhistoriker bzw. die breite Öffentlichkeit, die sich gerne über die Besitzverhältnisse ihrer Vorfahren bzw. ihres Wohnortes um 1860 informieren möchten. Diese Gruppe wird im Projekt durch das Tiroler Bildungsforum bzw. die Tiroler Ortschronisten vertreten.

Der Projektplan sieht nun die Entwicklung von drei Anwendungen vor, die zwar miteinander verlinkt sind, aber ihren je eigenen Zweck erfüllen sollen. Eine zentrale Rolle spielen dabei die automatisierte Handschriften- und Layoutanalyse sowie die Einbindung von Freiwilligen für die Korrektur der Ergebnisse.

3.1 | Suchinterface

Ein einfaches Suchinterface ermöglicht die Suche im Volltext für alle Dokumente. Besonderes Augenmerk wird dabei auf die Bau- und Grundparzellenprotokolle gelegt und hier wiederum auf die Vor-, Familien- und Vulgonamen. Familienforscher und

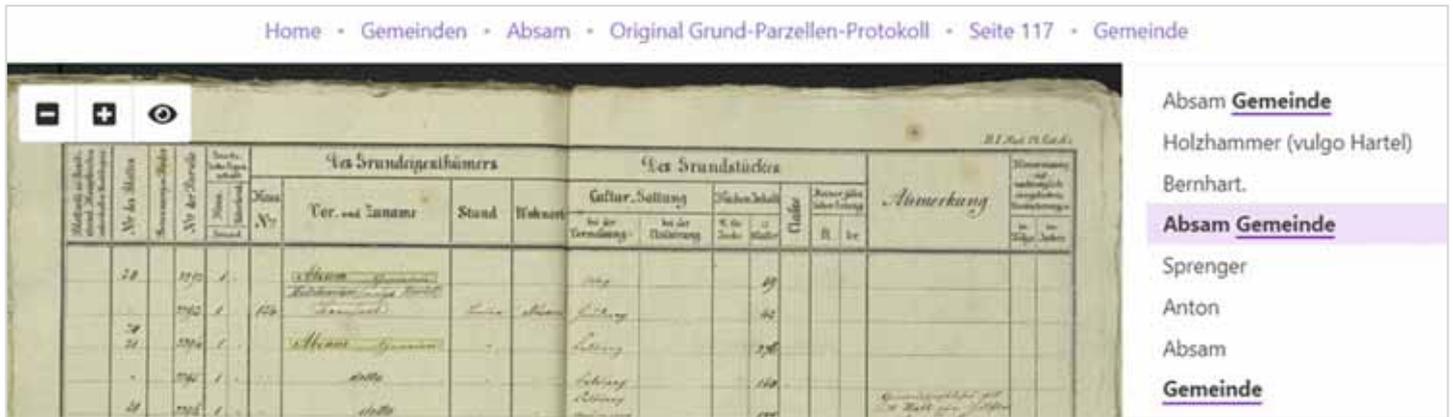


Abbildung 3 | Anzeige des erkannten und gesuchten Textes neben dem Bild der Seite

die Öffentlichkeit können so rasch eine Recherche mit Bezug zu ihren Familien- oder Hausnamen starten. Das Suchinterface ermöglicht aber auch die Suche in allen anderen Dokumenten (z. B. den Grenzbeschreibungen). Ebenso wird für jede Gemeinde ein Verzeichnis aller verfügbaren Dokumente erstellt und jedes Dokument kann darüber hinaus durchblättert und ein PDF exportiert werden. Zur besseren Lesbarkeit gibt es den Volltext der Dokumente als Überblendung am Bild (d. h., der automatisch erkannte Text wird direkt über das Bild der Seite geschrieben) oder als Zusatzinformation am Rand. So können auch Personen, die Kurrent nicht lesen können, anhand der automatisierten Transkription den Text entziffern.

Im Suchinterface wird der Text für die Benutzer sowohl als Überblendung im Bild als auch in einem eigenen Textfeld sichtbar sein.

3.2 | Citizen-Science-Anwendung

Wesentlich umfangreicher ist die Citizen-Science-Anwendung, die sich an die breite Öffentlichkeit sowie Familienforscher wendet und diese Zielgruppe gewinnen möchte, um die in den Mappen und den Bau- und Grundparzellenprotokollen enthaltenen Daten in einer Datenbank zu erfassen.

Das Interface weist zwei Methoden aus, die von den Freiwilligen selbst gewählt werden können: einmal die rein manuelle Erfassung (»Abtippen«) der Einträge und einmal eine semiautomatisierte Übernahme. Der Verlauf des Projekts wird zeigen, welche Methode von den Freiwilligen bevorzugt wird.

Besonderes Augenmerk wird in beiden Fällen darauf gelegt, die Benutzerfreundlichkeit des Interface möglichst zu erhöhen, sodass die Arbeit mit den Dokumenten im besten Fall sogar »Freude« bereiten kann und der Benutzer nicht durch unlogische oder verwirrende Schritte in seinem Arbeitseifer behindert wird.

Wichtig scheint uns auch, dass die Resultate der Arbeit für die Bearbeiter möglichst unmittelbar sichtbar werden – etwa in der Form von Fortschrittsstatistiken oder Visualisierungen der eingegebenen Daten.

Konkret besteht das Citizen-Science-Interface aus zwei getrennten Teilen: einem Editor für die Erfassung der Parzellennummern in den Mappen und einem Editor für die Grund- und Bauparzellenprotokolle.

Eine Verknüpfung zwischen den Dokumenten und den dazugehörigen Mappen wird automatisiert durch eine Verknüpfung in der Datenbank ermöglicht.

Erkannter Text	Korrektur Text
den Grenzstein., welcher zwischen dem Waldanteil des Peter Riggés und Joh. Wieser beich aus Aldrans und dem Gemeinde Wald von Anras auf der Wald- Antohe steht.	den Grenzstein V+ , welcher zwischen dem Wald theil des Peter Riggés und Joh. Wieser beide aus Aldrans und dem Gemeinde Wald von Ampas auf der Wald- Anhöhe steht.

Als Grundlage für das Suchinterface dienen ausschließlich völlig automatisiert gewonnene Daten. Zum Einsatz gekommen ist hier das Transkribus-System. Das speziell für die Kataster trainierte Modell auf der Basis von ca. 41.000 Wörtern weist eine sehr gute Erkennungsqualität von 1,3 % Character Error Rate auf dem Validierungstest auf.

Für die Erkennung der restlichen Dokumente wird ebenfalls ein spezielles Modell verwendet werden, das derzeit gerade auf der Basis von knapp 2.000 Seiten der Grenzbeschreibungen erstellt wird. Wir gehen von einer Fehlerquote von ungefähr 5 % Character Error Rate aus.



Abbildung 4 | Beispiel für Durchstreichungen (131) und Teilungen (196/1)

Das Geo-Interface

In den Mappen sind die Bauparzellen mit schwarzer Schrift, die Grundparzellen in roter Schrift erfasst. Zudem wurden Korrekturen in den Mappen durchgeführt, d. h., manche Grundstücksteilungen und Zusammenlegungen wurden ebenfalls eingezeichnet. So finden sich also durchgestrichene Parzellennummern, aber auch Parzellennummern, die eine zusätzliche Nummer bekommen haben.

Im vorliegenden Projekt werden diese Einzelheiten in einem Datenbankeintrag erfasst. Der Workflow für den Benutzer ist einfach: Nach dem Drücken auf »Erstellen« erscheint ein Fadenkreuz mit Punkt.

Damit wird nun eine Parzellenummer mit zwei Klicks unterstrichen (in obiger Abbildung durch dünne rote und schwarze Striche gekennzeichnet), ein Datenbankfeld öffnet sich und die Nummer kann in das Datenbankfeld eingetragen werden.

Auch die Art (Bauparzelle vs. Grundparzelle) oder ob es sich um eine unleserliche Parzellenummer handelt, kann hinterlegt werden. Nach dem Speichern kann der Benutzer ohne Verzö-

gerung weiterarbeiten, d. h., es erscheint gleich wieder das Fadenkreuz und die Vorauswahl der Felder bleibt bestehen, sodass z. B. eine Reihe von Grundparzellen und dann wieder eine Reihe von Bauparzellen in einem Durchgang erledigt werden können.

Zusätzlich findet gleichzeitig eine Abfrage in der Datenbank statt, ob die eingegebene Parzellenummer in den Bau- und Grundprotokollen bereits erfasst wurde. Falls dies – wie im obigen Beispiel – der Fall ist, wird der Name des Besitzers gleich eingeblendet.

Als Korrektur- und Prüfhilfe können die eingegebenen Nummern am Bild ebenfalls angezeigt werden, sodass fehlerhafte Eingaben rasch erkenntlich sind. Sobald eine Gemeinde komplett bearbeitet ist, kann der Administrator die Gemeinde als »Fertig« kennzeichnen.

Die Eingabe der Parzellennummern benötigt wenig Zeit: In einer Stunde können mehr als 200 Nummern eingegeben werden.

Auch größere Gemeinden wie Wilten, die über 179 Bauparzellen und 1.179 Grundparzellen verfügen, können somit in wenigen Stunden abgearbeitet werden.



Abbildung 1 | Textausschnitt aus einer Grenzbeschreibung. Die Fehlerquote liegt ungefähr bei 5 % Character Error Rate auf Buchstabenebene.

Das Interface zur Erfassung der Bau- und Grundprotokolle

Dieses Interface besitzt eine ähnliche Grundstruktur wie das Geo-Interface. Hier setzt der Benutzer allerdings nur einen einzelnen Punkt (vereinbarungsgemäß neben der Parzellenummer). Auch dann geht automatisch ein Fenster auf, in dem die für diese Tabelle wesentlichen Felder ausgefüllt werden können: Parzellenummer, Hausnummer, Vorname, Vulgoname, Familienname, Wohnort, Nutzungsart des Bodens, Joch, Klafter und Anmerkungen.

Ein Auto-Vervollständigen bei der Eingabe von Familiennamen, aber auch die Kopie kompletter Einträge mit einem Klick soll die Bearbeitung möglichst erleichtern. Auf einer Tabellenseite finden sich üblicherweise 20 Einträge. Durchschnittlich müssen dafür ca. 20 Minuten Eingabezeit gerechnet werden. Für die Gemeinde Wilten sind somit rund 30 bis 35 Stunden Arbeitszeit zu veranschlagen. Um den Benutzern einen Überblick zu geben, wie weit die Arbeit schon fortgeschritten ist, werden Fortschrittsanzeigen auf Seitenbasis gezeigt.

Neben diesem rein manuellen Interface wird auch die semi-automatisierte Erfassung erprobt. Hier werden mit Transkribus die wichtigsten Spalten der Tabelle erkannt. Der Benutzer kann dann die Werte für einen Eintrag durch einen einfachen Klick auf das jeweilige Feld »zusammentragen«.

Wir gehen davon aus, dass diese Vorgehensweise zu einer deutlichen Reduktion der notwendigen Arbeitszeit führen wird. Allerdings ist das Interface noch in Entwicklung und eine endgültige Beurteilung kann daher noch nicht abgegeben werden.

3.3 | Das Visualisierungsinterface

Für Gemeinden, die von Citizen Scientists vollständig erfasst wurden, liegen nicht nur außerordentlich interessante statistische Daten vor, diese sind zudem mit den Mappen verknüpft. Informationen aus den Dokumenten können somit innerhalb einer Mappenansicht genutzt werden.



Abbildung 1 | Textausschnitt aus einer Grenzbeschreibung. Die Fehlerquote liegt ungefähr bei 5 % Character Error Rate auf Buchstabenebene.

Für die Visualisierung der Statistikdaten bieten sich Tabellen und Diagramme an. So können etwa die größten Grundbesitzer einer Gemeinde gelistet werden oder aber jene Besitzer mit den meisten Obstgärten oder Wäldern. Die Unterscheidung in Privatperson und Organisation (Gemeinde, k. k. Ärar, Kirche, ...) erlaubt einen Blick auf die Besitzverhältnisse.

Die gewonnenen Daten sollen zudem als CSV-Dateien (Comma Separated Values) zum Download angeboten werden, um weitere wissenschaftliche Bearbeitungen zu ermöglichen.

Für die Familienforscher hingegen wird die Visualisierung unter Einbeziehung der Mappen von besonderem Interesse sein. So können bei jeder Parzellenummer die Daten aus den Grund- und Bauparzellenprotokollen angezeigt werden, also der Name des Besitzers, sein Beruf und Wohnort.

Durch Auswahl (oder Suche) des Familiennamens können dazu auch noch alle Parzellen eines Eigentümers in den Mappen einblendet werden.

Aber auch Nutzungsarten, wie etwa der »Egarten«, können anhand der Parzellenummern zwar nicht mit ihren umschließenden Polygonen, aber doch mit ihrer Parzellenummer und d. h. mit ihrem ungefähren Standort einblendet werden.

Anmerkungen

- [1] Vgl. <https://www.tirol.gv.at/arbeit-wirtschaft/wirtschaftsfoerderung/technologiefoerderungsprogramm/digitalisierungsfoerderung/leuchtturmprojekte/>
- [2] Transkribus wird von der READ-COOP SCE gehostet und weiterentwickelt. <http://readcoop.eu/>
- [3] Das Projekt wird zusammen mit Kurt Scharr, Gerhard Siegl und Berthold Ulreich durchgeführt.
- [4] Für jede Katastralgemeinde wurde ein gedrucktes Verzeichnis ausgefüllt, in das die vorgesehenen Dokumente eingetragen wurden. Darüber hinaus wurden insbesondere von den Grenzbeschreibungen oftmals mehrere Kopien angefertigt.
- [5] Historische Karten Tirols. <https://hik.tirol.gv.at/>
- [6] ArcGIS. <https://enterprise.arcgis.com/>



Mag. Dr. Günter Mühlberger
Leopold-Franzens-Universität
Innsbruck, Forschungszentrum
»Digital Humanities«
guenter.muehlberger@uibk.ac.at

