

An aerial photograph of the ETH Zurich campus, showing the main building with its large dome and surrounding structures. In the background, a large body of water (Lake Zurich) is visible, along with the city of Zurich and the surrounding hills. A large red rectangular overlay is positioned on the left side of the image, containing white text.

# Archival Visions. Neue Perspektiven auf Daten im Zeitalter von KI

Michael Gasser  
Christiane Sibille

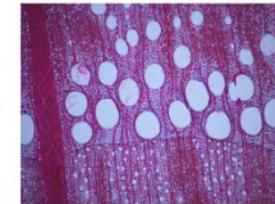
29. Archivwissenschaftliches Kolloquium |  
Archivschule Marburg  
21.05.2025

# Die Sammlungen und Archive der ETH Zürich

- [Über zwanzig spezialisierte Einrichtungen](#)
- Natur- und kulturhistorische Bestände von nationaler und internationaler Bedeutung
- Über 10 Millionen Dokumente und Objekte
- Einbettung in die ETH-Bibliothek und verschiedene Departemente der Hochschule



focusTerra



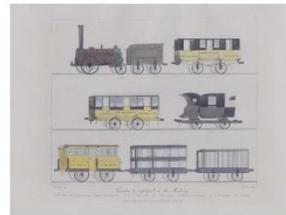
Forstwissenschaftliche Sammlung und Xylothek



Graphische Sammlung ETH Zürich



Anatomische Sammlung der Haustiere



Alte und Seltene Drucke



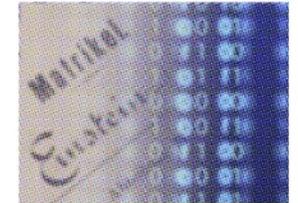
Archiv für Zeitgeschichte



gta Archiv



Vereinigte Herbarien der Universität und ETH Zürich



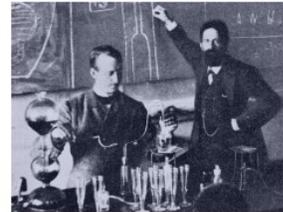
Hochschularchiv der ETH Zürich



Bestände der Baukultur ETH Zürich



Bildarchiv



Chemische und pharmakognostische Sammlung



Karten und Geoinformation



Kunstinventar der ETH Zürich



Max Frisch-Archiv an der ETH-Bibliothek



Entomologische Sammlung



Erdwissenschaftliche Sammlungen



ETH Material Hub



Sammlung wissenschaftlicher Instrumente und Lehrmittel

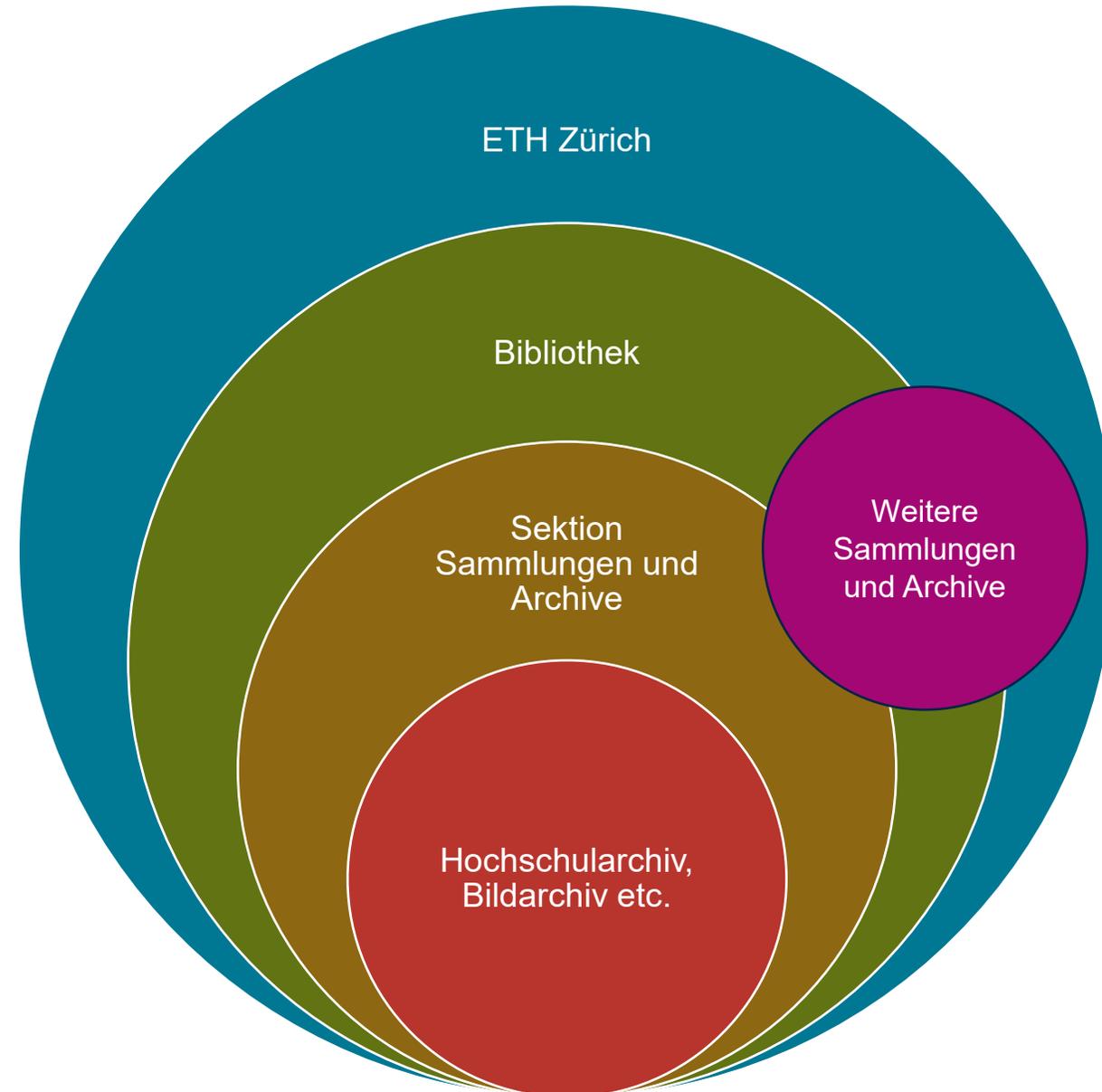


Thomas-Mann-Archiv



Zoologische Sammlung

# Organisatorische Einbettung der Sammlungen und Archive



# Alte und Seltene Drucke

Ca. 65'000 Bände

15.–19. Jhdt.





**Karten und Geoinformationen**  
Ca. 400'000 Karten (topografisch /  
thematisch)  
Digitale Karten und GIS-Daten

Hochschularchiv der ETH Zürich  
Administrative Unterlagen der Hochschule  
Privatnachlässe



ENGLIGA SVENSKA  
AKADEMIEN  
DEN 17 OKTOBER 1975  
BESLUTAT ATT MED DET  
**NOBELPRIS**  
PÅ DETTA ÅR TILLERKÄNNES DEN  
SOM GJORT DEN VIKTIGASTE  
KEMISKA UPPTÄCKTEN  
FÖR BÄTTNINGEN, MED HÄLFTEN  
TILL VARDERA, BELÖNA  
FÖR MARGUP CORNFORTH OCH VLADIMIR PRELOG,  
**VLADIMIR PRELOG**  
FÖR HANS ARBETEN RÖRANDE  
ORGANISKA MOLEKYLERS OCH  
REAKTIONERS STEREOKEMI  
STOCKHOLM DEN 10 DECEMBER 1975



**Bildarchiv**

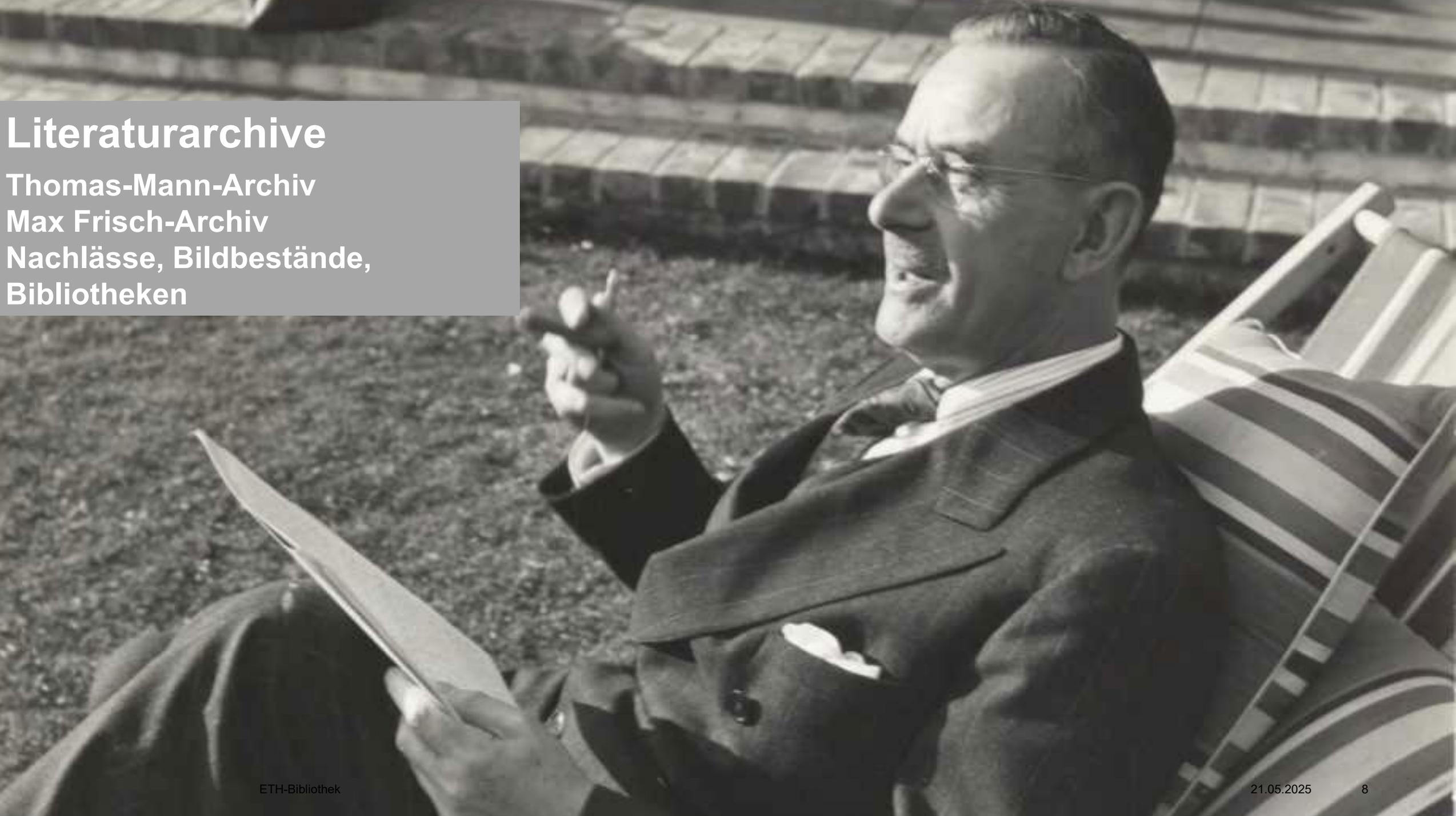
**Ca. 4 Mio. Fotografien**

# Literaturarchive

Thomas-Mann-Archiv

Max Frisch-Archiv

Nachlässe, Bildbestände,  
Bibliotheken



AD



**Graphische Sammlung ETH Zurich**  
Ca. 160'000 Werke auf Papier, 15. Jhdt. bis in die Gegenwart



**Sammlung wissenschaftlicher  
Instrumente und Lehrmittel  
Ca. 2'500 Objekte**

# Kunstinventar der ETH Zürich

Ca. 300 Kunstwerke



A 32 Hörsaal

A 32.1

# Online Plattformen (Auswahl): (Kooperative) Online-Zugänge zu digitalisierten Beständen



Alte Drucke

[e-rara](#)

~ 152'000 Titel

Public Domain Mark

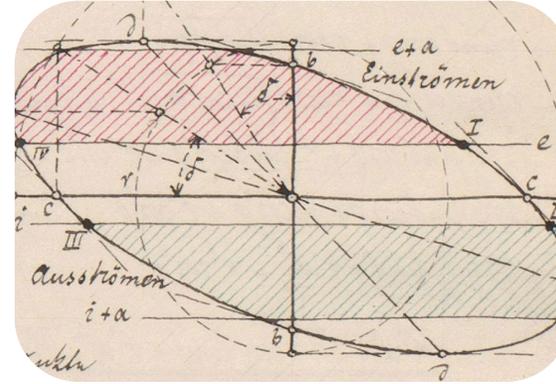


Fotografien

[E-Pics Bildarchiv](#)

~ 1 Mio. Bilder

überwiegend Public  
Domain Mark und  
CC BY-SA Lizenz

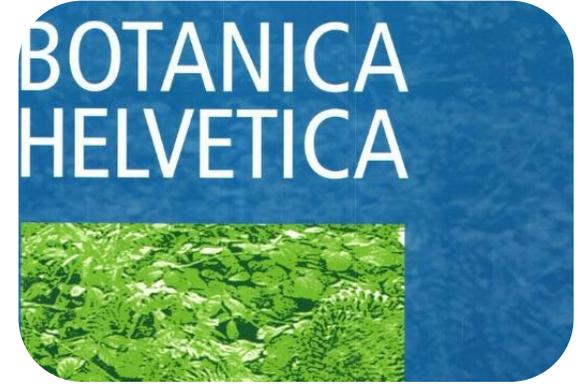


Archivalien

[e-manuscripta](#)

~ 156'000 Quellen

überwiegend Public  
Domain Mark



Zeitschriften

[E-Periodica](#)

~ 1'100 Zeitschriften

~ 10 Mio. Seiten

# Online Plattformen (Auswahl): Aufbereitete Inhalte

## Blogs / Social Media

- [ETHeritage – Highlights aus den Sammlungen und Archiven der ETH Zürich](#)
- [ETH-Bibliothek Crowdsourcing](#)
- [Blog der Graphischen Sammlung](#)
- Instagram: [eth\\_collections](#)

## Kuratierte Inhalte

- Places of Interest: [ETHorama](#)
- Virtuelle Ausstellungen, z. Bsp. basierend auf alten Drucken ([Exotische Reisen](#), [Ungeheure Eismassen](#))

## Interaktive Anwendungen

- Wissenschaftliche Instrumente in 3D: [ETHeritage 3D](#), [Sketchfab ETH Library Collections](#)
- Interaktive astronomische Zeichnungen: [astrorara](#)

**ETH** zürich | ETHeritage



### Sammlung 3D – Flächen messen mit dem Orthogonalplanimeter

28. Juni 2023 Roberta Flora Spano, Dorothea Regula Zimmermann

«Es gab wohl zu keiner Zeit so viele Flächeninhalte ebener Figuren zu berechnen als jetzt, wo man sich mit den ausführlichsten Landesvermessungen und den belangreichsten Strassenbauten beschäftigt.» Carl Maximilian von Bauernfeind: Die Planimeter von Ernst, Wetli und Hansen, welche den Flächeninhalt ebener Figuren durch das ... [Weiterlesen ...](#)

# Strategie Sammlungen und Archive der ETH Zürich 2025–28



Digitale Transformation vorantreiben



Dialog mit der Öffentlichkeit pflegen



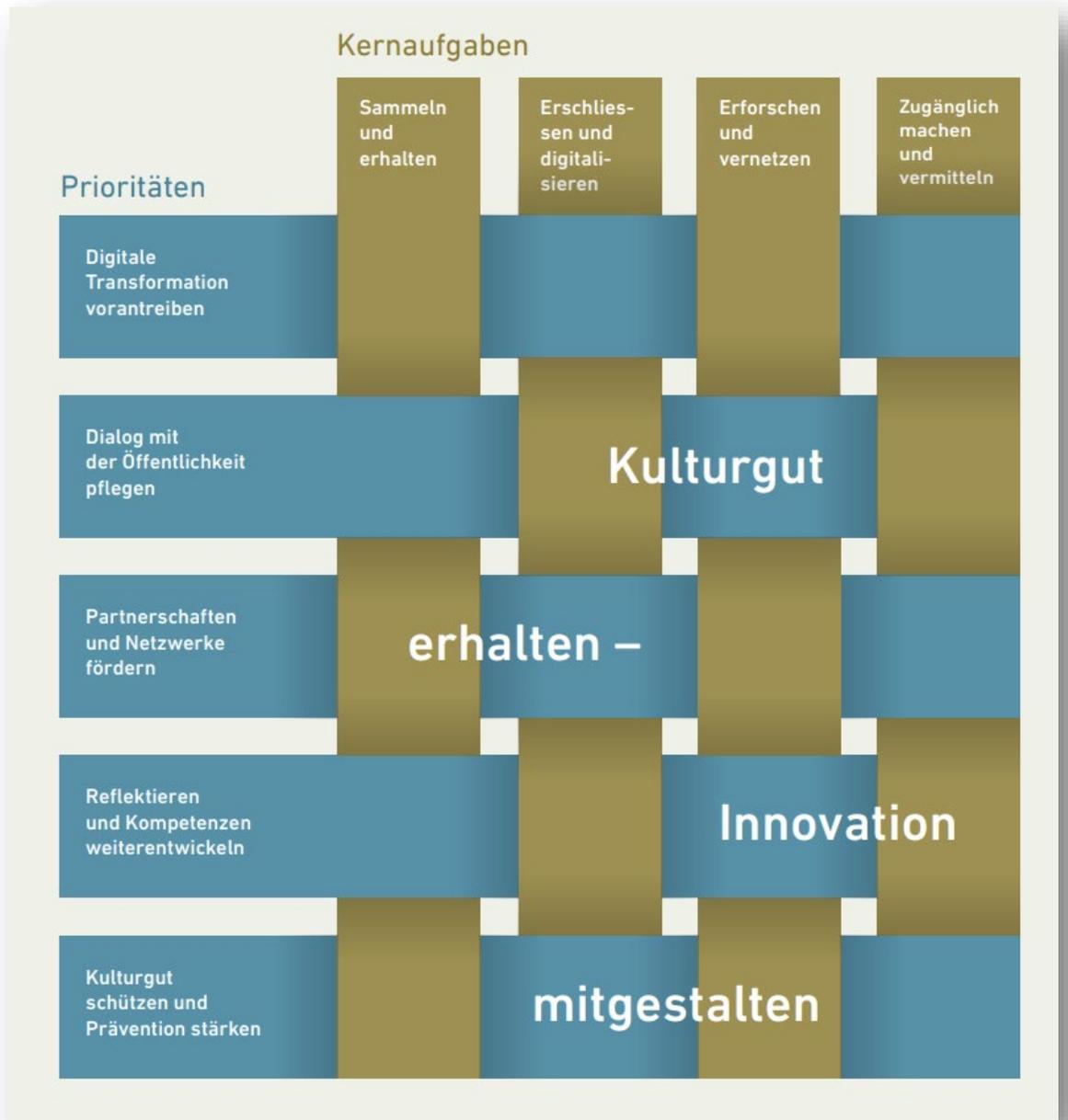
Partnerschaften und Netzwerke fördern



Reflektieren und Kompetenzen weiterentwickeln



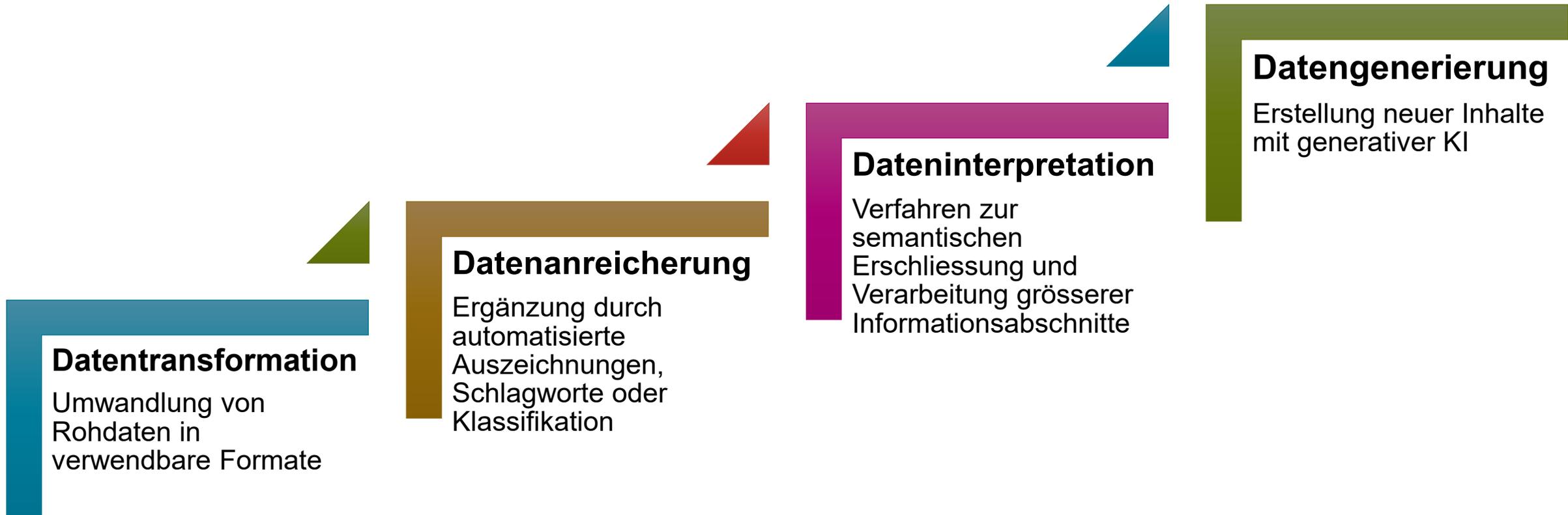
Kulturgut schützen und Prävention stärken



[DOI:10.3929/ethz-b-000733893](https://doi.org/10.3929/ethz-b-000733893)

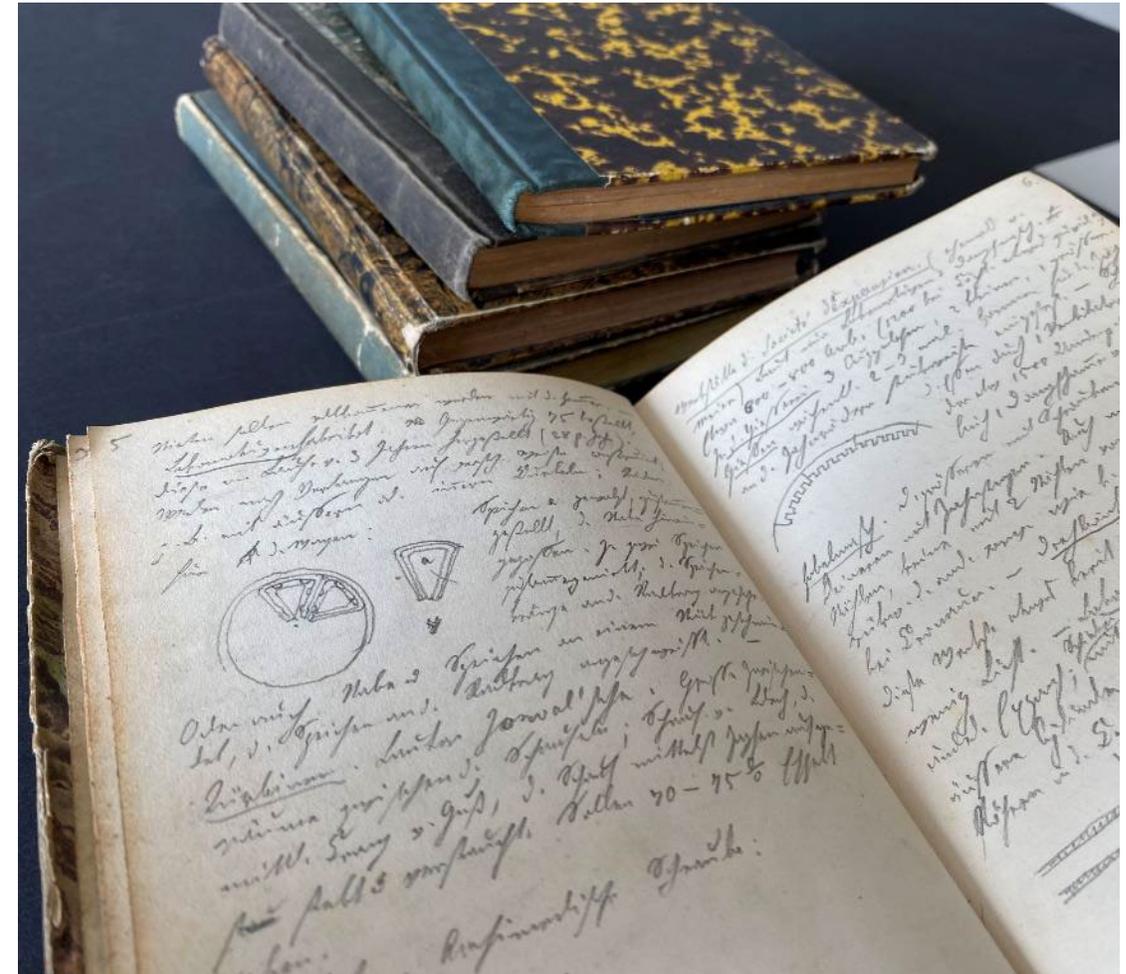
# KI Anwendungen in den Sammlungen und Archiven der ETH-Bibliothek

# Vier-Stufen-Modell zur KI-Anwendung in der Erschliessung



# Projekte innerhalb der Sammlungen und Archive

- Datentransformation
  - Max Frisch Archiv: Edition der Notizhefte von Max Frisch (**HTR mit Transkribus**)
  - ETH Hochschularchiv: Transformation grösserer handschriftlicher Quellenbestände: Reise- und Feldtagebücher von Arnold Heim, Tagebücher von Joseph Wolfgang von Deschwanden, ETH Schulratsprotokolle (**HTR mit Transkribus**)
- Datenanreicherung
  - e-rara: **OCR, Named Entity Recognition und Named Entity Linking** als Teil des Textlabs von Visual Library (kommerzieller Anbieter)
  - ETH Bildarchiv: zusätzlich zur manuellen Verschlagwortung: **automatische Verschlagwortung und Übersetzung** des gesamten digital zugänglichen Bildbestands (ca. 1 Mio Bilder)
  - E-Periodica (Kooperation innerhalb der Bibliothek): **Named Entity Recognition und Named Entity Linking** im Rahmen des Projekts „[E-Periodica – Next Level Access](#)“ (10 Millionen Seiten)



# Digital Scholarship Services

# Digital Scholarship Services

- Die Fachstelle Digital Scholarship Services entwickelt **Projekte und Services** an der Schnittstelle zwischen digitalisierten Sammlungen und digitaler Forschung.
- Im Zentrum stehen die Idee von „**Collections as Data**“ und die Nutzung von Machine Learning Technologien
- Team: 2 Fachspezialistinnen (1.6 FTE), Studentische Mitarbeitende (0.4 FTE)
- Enge Zusammenarbeit mit anderen Teams der ETH-Bibliothek
  - z.B. ETH Hochschularchiv: KIZH: KI basierter Zugang zum Hochschularchiv (9 Monate, 1 FTE)
  - Max Frisch Archiv: Edition der Notizhefte von Max Frisch
- Kooperation mit **Scientific IT-Services der ETH** (Computational & Data Science Support)
- Projektbasierte Kooperationen, z. B. „**Exploring Collections as Data**“ ([collada.ethz.ch](http://collada.ethz.ch))
- Ziel: Aufbau und Verankerung von Services zum Einsatz von KI/Data Science im Kontext der Sammlungen und Archive

# Entwicklung einer lokalen modularen und skalierbaren Pipeline zur Datentransformation und Datenanreicherung (OCR | NER | NEL etc.)



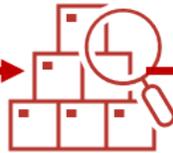
**Protokoll.**

Teilnehmer. A. Mitglieder: d. **Herren Dr. C. Chr. Bernoulli**, Universitätsbibliothek Basel; Dr. Herm. Escher, Stadtbibliothek Zürich; M. A. Aubert, Bibliothèque publique de Genève; Ferd. Aubert, Stadtbibliothek Zürich; Dr. Hans Barth, Stadtbibliothek Winterthur; Dr. E. Brugger, Stadtbibliothek Zürich; Prof. Dr. J. Dierauer, Stadtbibliothek St. Gallen; Max von Diesbach, Kantonsbibliothek Freiburg; Dr. Jak. Escher, Stadtbibliothek Zürich; H. H. Field, Concilium Bibliographicum Zürich; Prof. W. Gisi, Kantonsbibliothek Solothurn; Dr. Franz Heinemann, Bürgerbibliothek Luzern; Dr. H. Herzog, Kantonsbibliothek Aarau; Wenzel Karczewski, Musée polonais Rapperswil; Keller-Ris, Stadtbibliothek Lenzburg; P. Gabriel Meier, Stiftsbibliothek Einsiedeln; Prof. W. F. von Mülinen, Stadtbibliothek Bern; Ch. Robert, Bibliothèque de ville, Neuchâtel; Prof. Dr. E. Rudio, Bibliothek des Polytechnikums, Zürich; J. Schiesser, Landesbibliothek Glarus; Prof. Dr. Th. Vetter, Bibliothek der Museumsgesellschaft Zürich; Dr. H. Weber, Kantonsbibliothek Zürich; Dr. J. Werner, Kantonsbibliothek Zürich; Dr. Emil Wettstein, Bibliothek des Polytechnikums, Zürich; V. Wirz, Stadtbibliothek Zürich; Prof. Dr. W. von Wyss, Stadtbibliothek Zürich.

B. Gäste: Herr Dr. C. Escher, Präsident des Konventes der Stadtbibliothek Zürich und Frä. Mary E. Hawley von der John Crerar Library, Chicago.

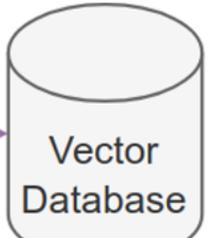
## Named Entity Tagging

Suche nach C. Chr Bernoulli in interner Datenbank



- Candidates
- Candidate 1
- Candidate 2
- ...

Abfrage der Vektoren der besten Treffer in der Vektordatenbank



Der Name (C. Chr. Bernoulli) und sein Kontext werden ebenfalls "embedded" (in Vektoren umgewandelt)

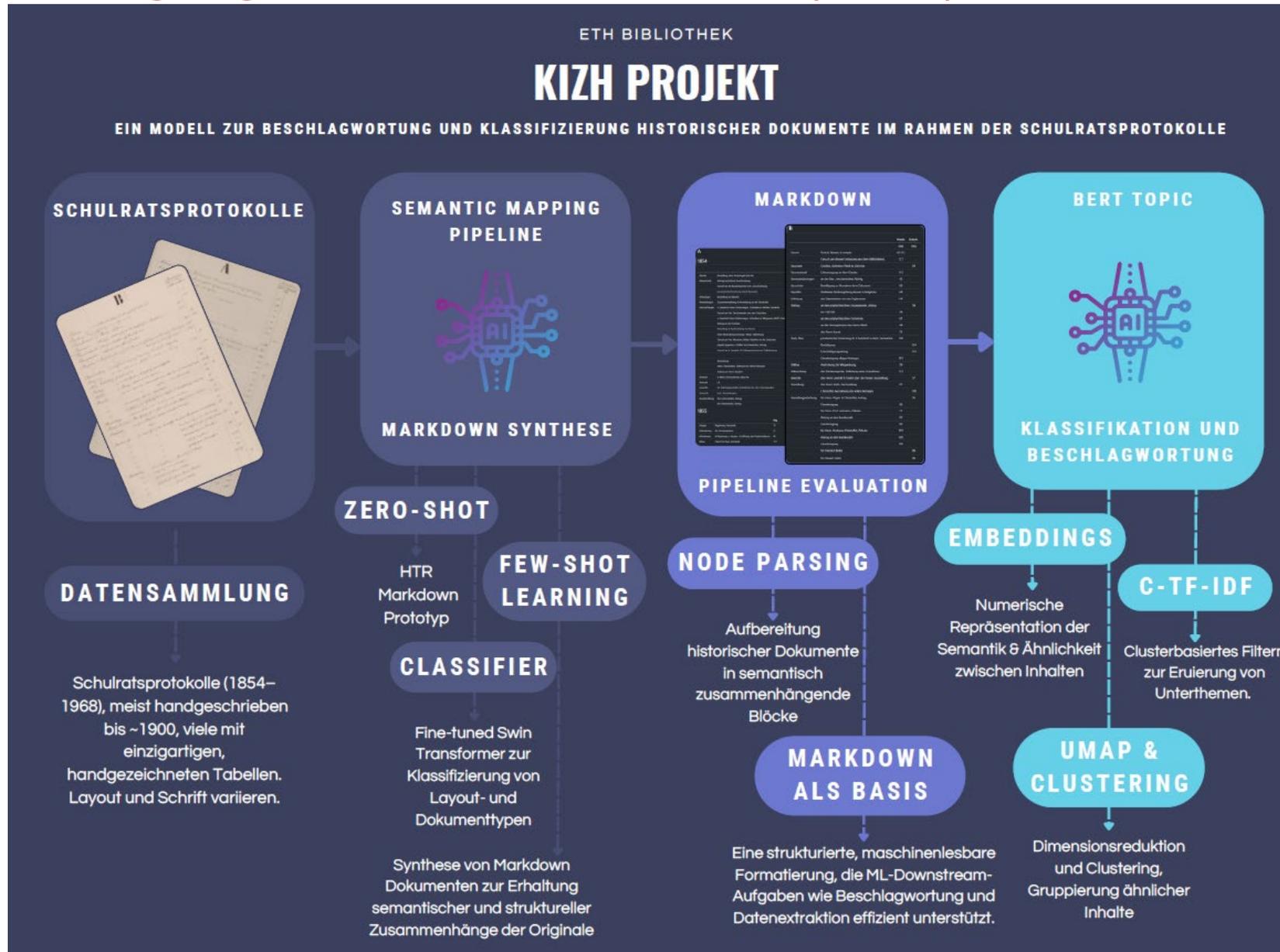


Verhältnis der beiden Vektoren wird berechnet

Verknüpfung des Namens mit dem besten Treffer



# Dateninterpretation: KI basierter Zugang zum Hochschularchiv (KIZH)



# Weiterbildungen: Introduction to Python

## Exploring Collections as Data

[Home](#) [About the project](#) [Focus Collections](#) [Activities and Network](#)

## Activities and Network

### Introduction to Python



As part of the CollaDa project, [ETH Scientific IT-Services](#) developed a six-day course providing a hands-on introduction to Python programming and textual data analysis. The first round of the course run from March to May 2023. 15 participants (students, PhDs, Postdocs and people working in the GLAM sector) participated in the course. They evaluated the course as very good and helpful.

[The next round of the six-day event will start in May 2025.](#)

# Lessons Learned und Perspektiven

# Herausforderungen und Möglichkeiten

- Rahmenbedingungen für den Einsatz von KI-Technologien
  - Ressourcen
  - Umsetzungskompetenzen („Machen“) + Methodische Kompetenzen („Reflektieren“)
  - Stabilität der Services | Dynamik der Entwicklungen
- Grundlegende Veränderungen
  - Integration von maschineller „Interpretation“ in bestehende Prozesse und Standards
  - Strukturierte Datenaufbereitung | natürlichsprachige Interaktion
- Entwicklungsmöglichkeiten
  - Neue Formen der Kooperation (Geschichtswissenschaft/GLAM/Computer Science)
  - Neue Nutzendengruppen (mit Interesse an unterschiedlichen Formen der Datenerschliessung)

# Strategische Janusköpfigkeit / Preparedness

- Blick zurück/nach Innen
  - Daten und Prozess „Readiness“
    - „Ground Truth“ für Training und Evaluation von Modellen
    - Maschinell nachnutzbaren Datensets, auch für neue Nutzergruppen (z. B. Data Science)
  - Massnahmen zur Interpretierbarkeit und Erklärbarkeit (Paradaten, Reproduzierbarkeit/deterministische Verfahren, Dokumentation)
- Blick nach vorne/nach aussen
  - KI Nutzung im Archivwesen im Kontext
  - Beobachtung technologischer Entwicklungen
  - Einbringen von archivischen Kompetenzen in die Entwicklung von Datensets und KI-Modellen (z. B. Kuratierung, Bewertung und Dokumentation von Datensets)



Robert van Audenaerde, Carlo Maratti, Kopie nach: Janus am Tor mit Allegorien auf Frühling und Winter  
ETHZ Graphische Sammlung: [https://doi.org/10.16903/ethz-grs-D\\_042297](https://doi.org/10.16903/ethz-grs-D_042297)

Michael Gasser  
Leitung Sammlungen und Archive  
[michael.gasser@library.ethz.ch](mailto:michael.gasser@library.ethz.ch)  
+41 44 632 21 82

Dr Christiane Sibille  
Verantwortliche Digital Scholarship Services  
[christiane.sibille@library.ethz.ch](mailto:christiane.sibille@library.ethz.ch)  
+41 44 633 89 54

ETH-Bibliothek  
Rämistrasse 101  
8092 Zürich

[www.library.ethz.ch](http://www.library.ethz.ch)