CLARIAH-AT

Bundesministerium
Frauen, Wissenschaft
und Forschung

Funded by
the European Union

DH INFRA

Abschlussveranstaltung
8. Juni 2026 – 10:00-12:00

Universität für
Weiterbildung
Krems

Donau Universität Krems

ÖAW Österreichische Akademie der Wissenschaften

Paris Lodron Universität Salzburg

universität innsbruck

Österreichische Nationalbibliothek

TU WIEN Technische Universität Wien

UNI GRAZ

universität wien

di:'ʌngewʌndtə

Georg Vogeler
Universität Graz – Institut für Digitale Geisteswissenschaften

Agenda

10:00	Begrüßung durch Projektleitung
10:03	Projektziele und -verlauf
10:10	Ergebnisse
11:55	Ausblick

Florian Atzenhofer-Baumgartner
Universität Graz – Institut für Digitale Geisteswissenschaften

1) Projektziele und -verlauf

DHinfra in Zahlen

9 Institutionen

3,5 Jahre (2023 - 2026)

4 Arbeitsgruppen

8 Arbeitspakete

28 Meilensteine

50+ Mitglieder

100+ Assoziierte

Projektbedingungen

Förderprogramm: “(Digitale) Forschungsinfrastrukturen”

Fördergeber: BMFWF und ERC Recovery & Resilience Facility (NextGen)

Förderziele: Beschaffung, Modernisierung, besserer Zugang

Infrastrukturprojekt – kein Forschungsprojekt

Projektziele (1/2)

DHinfra schließt Lücken …

… bei der Standard-Digitalisierung innerhalb von Bibliotheken, Archiven, Museen.

… von generischen Repositorien und Datenbanken.

… in Verfügbarkeit von Hochleistungsrechnern (‘HPC’) außerhalb der Natur-, Technik- und Lebenswissenschaften

DHinfra ermöglicht, …

… **neue** und/oder **große Datenbestände** (an Kulturerbe) zu **prozessieren** und zu **speichern**.

… diese Datenbestände mit **Maschinellem Lernen** (bzw. **KI**) zu verarbeiten.

Projektziele (2/2)

DHinfra bewerkstelligt das in 4 Bereichen:

1) Fortgeschrittene Digitalisierung

2) Fachspezifische Datenbanken und Repositorien

3) domänen-affine ML- und Recheninfrastruktur

4) Open-Source-Software

Wir eröffnen damit neue und **kompetitive Forschung**, ermöglichen **sichere Verarbeitung** von Kulturerbe-Daten, **stärken** die existierende DH-**Community** in Österreich, **fördern Nachwuchs** und **Interdisziplinarität** und sichern eine **nachhaltige Nutzung** von Infrastruktur.

Projektverlauf

Konsortialvertrag, Governance-Ansätze, Uni Wien integriert

Start Scan-Roboter-Entwicklung, IaaS-**Planung**

FOSS-AG & erste **Bedarfserhebungen** (alle Bereiche)

Anschaffung & **Test** der MSI-Systeme (XpeCAM, MISHA)

Ausbau des Datenbank-Cluster (IaaS; PLUS),

Repositorien-(Re-)**Konzeption** (UWK, PLUS)

Systematische FOSS-Bedarfserhebung, Governance-Entwurf

2023-2024:
Konsortium &
Grundlagen

2024-2025:
Anforderungen
& Vorbereitung

Projektverlauf

Scan-**Roboter** im produktionsnahen **Einsatz** (400.000+ Meldezettel, Fehlerquote < 1%)

GPU-**Cluster** (Graz, Krems) beschafft und **installiert**; Alternativlösung für Angewandte

IaaS weitgehend **abgeschlossen**; Alternativlösung statt PLUS-Ceph

FOSS-Workshops, Whitepaper, **Aufträge und Abwicklung** für zentrale DH-Software

(fast) alle geplanten Infrastruktur-Bausteine bereits im Einsatz

Governance-Kern (Betriebskonzepte, ‘Business Model’, Helpdesk-Ansatz) steht

GPU-Cluster & LLM/Embedding-APIs kurz vor Roll-out an CLARIAH-AT-Partner:innen

2025-2026:
Aufbau &
Ausbau

2026- :
Integration &
Abschluss



2) Ergebnisse

Agenda (Ergebnisse)

10:10	Digitalisierung / Data Capture
10:35	Datenbanken und Repositorien / IaaS & SaaS
10:55	Pause
11:00	Maschinelles Lernen und KI
11:35	Freie und offene Software

2.1) Digitalisierung

10:10 – 10:35

Florian Kleber
Technische Universität Wien – Institute of Visual Computing and Human-Centered Technology

Multispektral-Kameras

- Anschaffung von 2 portablen MSI Systemen
- Easy to use
- Leih-/Mietvertrag
- Schulung
- GPU HW for Processing

XpeCAM

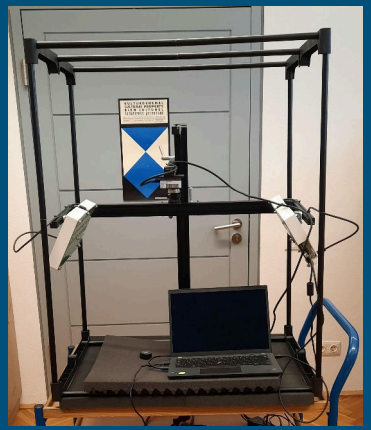
- 350-1200 nm
- 30 filter
- 6.4Mpx
- lens 35 mm

MISHA

- 365 - 940 nm
- LED based
- 20Mpx
- lens 23 mm







CVL
Computer Vision Lab

Anwendungsfälle

- Wandgemälde Burg Kreuzenstein (DA AkBild)
- Archiv der Erzabtei St. Peter, Salzburg (zb. Informationstafel aus dem 19. Jahrhundert (PAR, Akt 99))
- Archiv des Erzbistums in Zagreb, Monumenta Germaniae Historica, Handschrift auf Papier
- Nationalmuseum Montenegro, Cetinje. Pergamente



CVL Computer Vision Lab

Florian Wachter
Universität Graz – Institut für Digitale Geisteswissenschaften (u.a.)

Die Urkunden von Cetinje, Montenegro

- Fehlende Edition und Digitalisierung
- 9 Pergamente, stark beschädigt
- Multispektralaufnahmen mit MISHA von der TU Wien



MSCA PF Projekt MeSeCa,
PI Zarko Vujosevic

Günter Mühlberger, Patrick Schönegger
Universität Innsbruck – Digitalisierung und elektronische Archivierung (DEA)/ Forschungszentrum DH


FISCAL YEAR
1990-2000
CONTRACTS
FISCAL YEAR
1990-2000
CONTRACTS
INVOICES
CONTRACTS
CONTRACTS
FISCAL YEAR
1990-2000



Eingelangt Innsbruck
am 25 JAN. 1909
Städt. Meldeamt.

Meldezettel für Afterparteien, Bettgeher und Gäste.

Straße (~~Gasse, Platz~~) Nr.: 24 Stock: 2

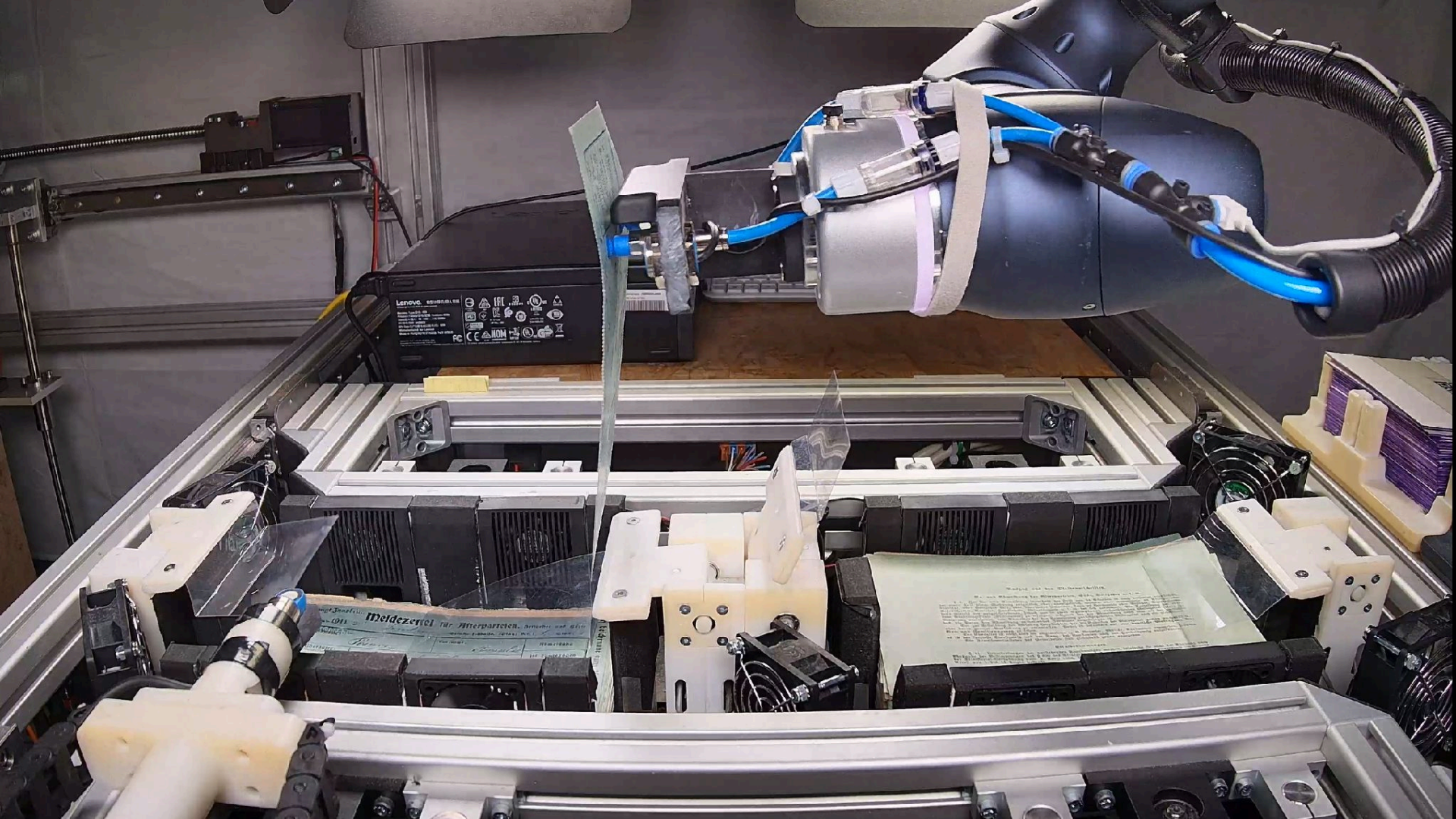
Schreibname: Spörr	**Taufname:** Ulrich	**Abmeldung**
Charakter: (Beschäftigung)	Holzarbeiter	ist ausgezogen (abgereist am) 1. II. 09
Geburtsort, Land	Wilten b. Innsbruck, Tirol	wohin? M. Thurnfeld
Heimatsort, Land	Matters	
Geburtsjahr, Religion	1883 kath.	ledig, ~~verheiratet, verwitwet, geschieden~~
Tauf- Mädchenname und Geburtsjahr der Gattin; Name u. Geburtsjahr der Kinder:		
Frühere Wohnung:	Matters Straße Nr.	

Eingelangt Innsbruck am 8 FEB. 1909 Städt. Meldeamt.

Innsbruck, 25/I 1909.

Unterschrift des Wohnungsgebers: Ulrich Spörr

Nach beschmutzt zum Amte gebracht werden.


Lenovo
Meldezettel

Projekte und Dissemination

KI-Verarbeitung historischer Dokumente

Kooperationsprojekt: FZ Digital Humanities der Uni Innsbruck mit dem Institut für Numismatik der Uni Wien

Römische Kaiserzeit
Macrinus
217–218

Denar
2,96g



477

Amtliches 10(1993)

RIC 76
c. 37

Historical Context & Identification (Top Left):

Römische Kaiserzeit: German for "Roman Imperial Period." This specifies the historical era of the coin.

Macrinus: The Roman Emperor depicted on the coin. Marcus Opellius Macrinus ruled briefly after the assassination of Caracalla.

217 – 218: The years of Macrinus's reign (217 to 218 AD), which also dates the coin.

Physical Properties (Top Right):

Denar: German for "Denarius." This was the standard silver coin denomination of the Roman Empire for centuries.

2,96g: The exact weight of this specific coin (2.96 grams). Weighing coins is crucial for authenticating them and for scholarly records.

Datenschutz? Personendaten in der Cloud?

Dissemination

Patrick Schönegger: “Low and High Tech Digitisation: From Robotic Scanning Lines to Community-Driven Capture Devices” ICARUS Convention #35 Bridging Centuries: Digital Pathways for Medieval Charters and Beyond Day 1, Wednesday, 12 November 2025

Günter Mühlberger: Robotic Scanning as Future Solution for Archives? Digitise.Transform.Inspire – Shaping The Archives Of Tomorrow, Vienna, 1-2 September 2025

Diverse Präsentationen bei der Langen Nacht der Forschung, für Archive und genealogische Gesellschaften

2.2) Datenbanken und Repositorien

10:35 – 10:55

Martin Schäler und Karoline Döring
Universität Salzburg – Fachbereich Informatik/Database Research Group

Was liefert die PLUS an Infrastruktur?



Integriert in Core Facility “Database Research Cluster” + DHinfra-Service-Layer

VM-Cluster (Proxmox + CEPH)

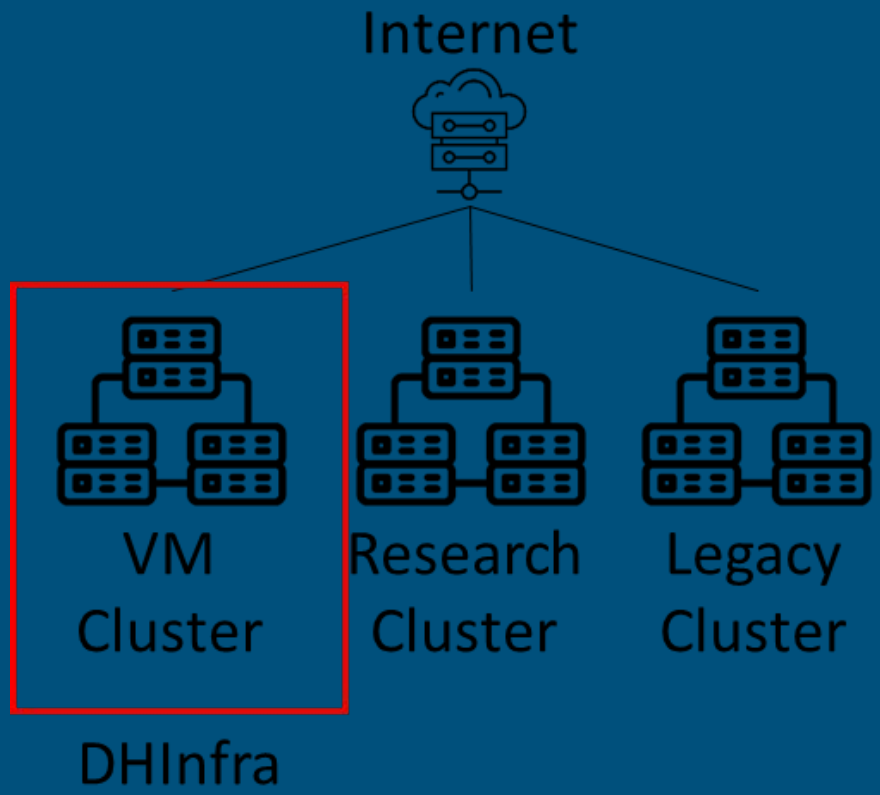
8 Server, je 32 Cores, 2 TB RAM, 200 Gbit/s Infiniband, 3GPUs,

1 Server zu 128 Kernen, 4 TB RAM

Was Projekte bekommen:

- Rechenumgebungen für Entwicklung und Test


Bundesministerium
Frauen, Wissenschaft
und Forschung
FORSCHUNGS
INFRASTRUKTUR
DATENBANK
forschungsinfrastruktur.bmfwf.gv.at


Internet
VM
Cluster
Research
Cluster
Legacy
Cluster
DHInfra

Wie DH-Projekte die PLUS-Infra nutzen (können)

IaaS / DBaaS für DH-Projekte

Test- & Dev-Umgebungen für DH-Software

Datenbank-Backends + Web-/API-Server

Integrierte Toolchain:

- VMs + Ansible-Provisionierung
- CI/Git-Runner für Forschungssoftware

Ziel:

- **reproduzierbare Setups**
- **weniger Admin-Aufwand** in Einzelprojekten
- sicherer Playground für neue Technologien

Projekte und Dissemination

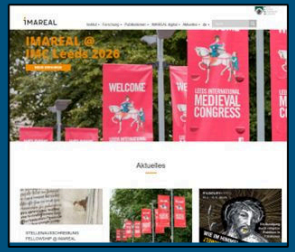
Literaturarchiv Salzburg (LAS)



KONTAKT

Lina Maria Zangerl

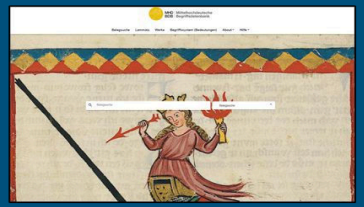
IMAREAL



KONTAKT

Isabella Nicka &
Peter Färberböck

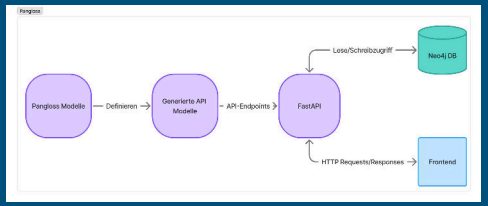
MHDBDB



KONTAKT

Katharina
Zeppezauer-Wachauer

Pangloss



Kontakt

Michael Otto

Siehe auch https://dhsalzburg.hypotheses.org/

MHDBDB als Infrastruktur-Testfall

Seit den 1970ern; heute immer noch zentral für Mediävistik

Was DHinfra hier testet:

- Betrieb auf PLUS-VM-Cluster
- offene **Graphdatenbank** statt proprietärer Lösung
- API-basiertes Backend für Nachnutzung & KI-Szenarien (cf. GPUs!)

Effekte:

- **stabilere**, **skalierbare Abfragen** + reproduzierbare Deployments
- Entkopplung: Backend/Infra vs. Usability/Frontend

Michael Otto
Universität Graz – Institut für Digitale Geisteswissenschaften

Tests auf ProxMox-VMs

Tests des VM-Clusters der Universität Salzburg anhand eines spezifischen Use-Cases

Der Cluster stellt mithilfe von ProxMox vorbereitete virtuelle Maschinen (VM) bereit

Nutzer:innen können vom eigenen Rechner aus auf die gewünschten Hardware und Software Ressourcen zugreifen

Hauptziele der Tests: Belastbarkeit und User-Freundlichkeit

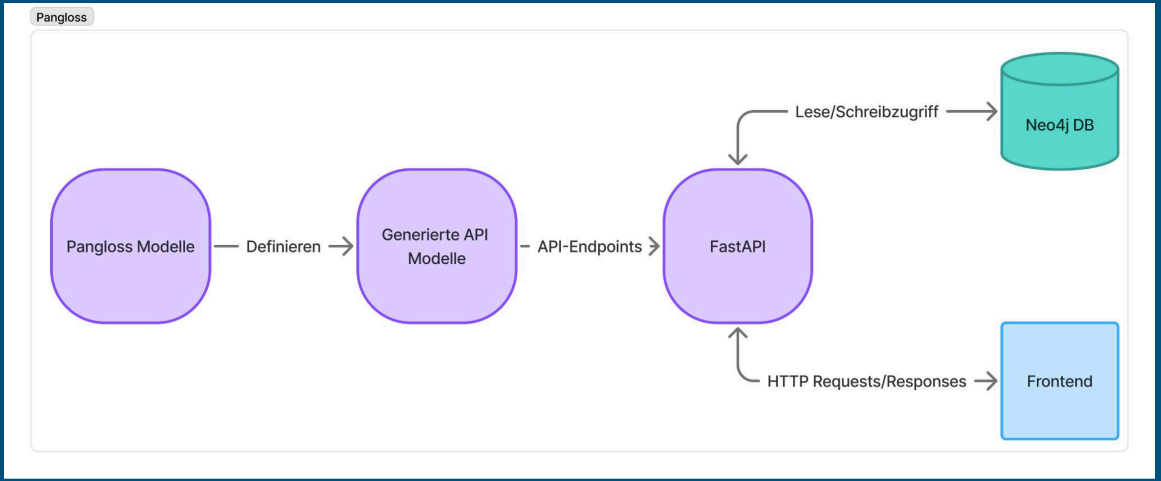
VM Spezifikationen: 64GB RAM, 8 Kerne und 32 GB Speicher; Linux Debian Betriebssystem

Tests auf ProxMox-VMs (Use-Case)

Prosopographische Datenbank “Pangloss”-Prototype von Richard Hadden als Test-Case:

- FastAPI-basierte Plattform für Datenmodellierung, -validierung, -erfassung
- Unterstützt durch Graphdatenbank Neo4j

Große, verschachtelte Modelle lassen sich einfach als verbundene Graphknoten darstellen



Tests auf ProxMox-VMs (User-Freundlichkeit)

Einfaches Deployment durch Ansible-Playbooks

Sichere und stabile Verbindung über OpenVPN Connect und SSH

Zugriff auf die gesamte VM erlaubt breite Reihe an möglichen Use-Cases

24/7 Verfügbarkeit und Access

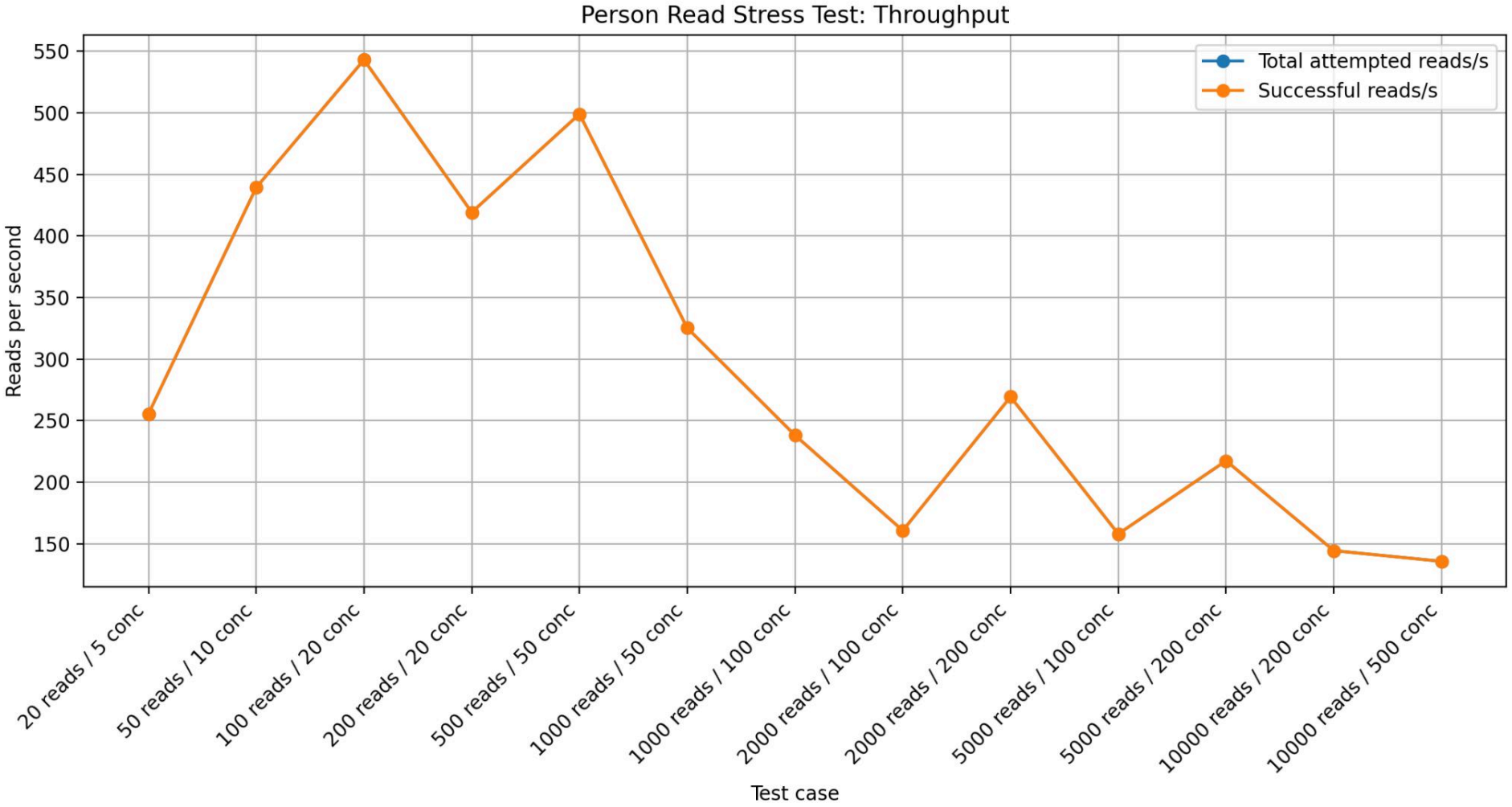
Tests auf ProxMox-VMs (Stress-Tests)

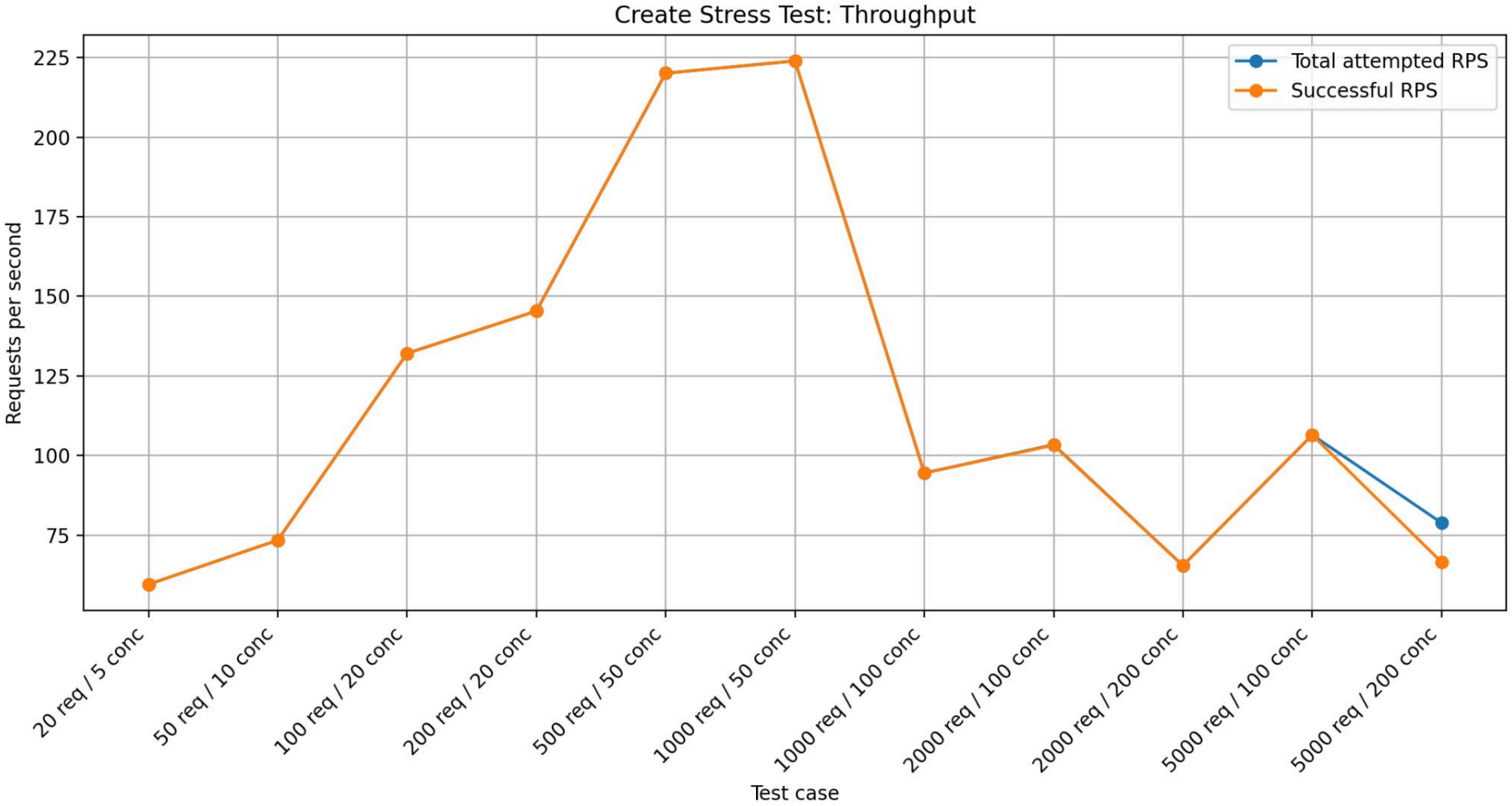
Belastbarkeit der VM auf verschiedenen Ebenen getestet:

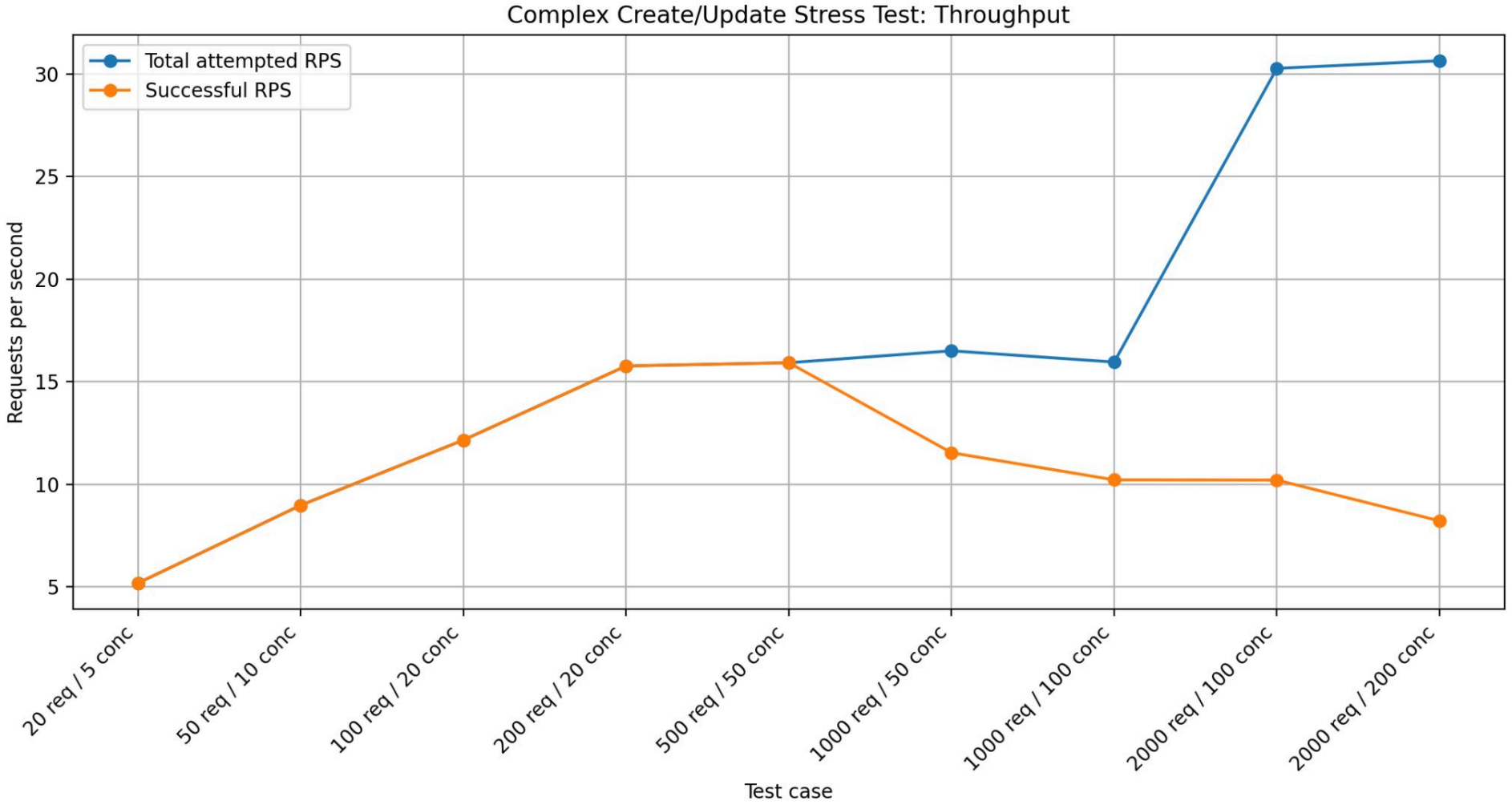
- Lesezugriff, einfacher Schreibzugriff und komplexer Schreibzugriff auf Datenbank
- Unterschiedlicher Menge an Anfragen und Gleichzeitigkeit bis Failure

Ergebnis: Für den Use-Case und dessen Menge an Anfragen definitiv geeignet

Nächste Schritte: Testen der Netzwerk-Komponenten von Pangloss und möglicher Wünsche von Nutzer:innen


Person Read Stress Test: Throughput
Total attempted reads/s
Successful reads/s
Reads per second
550
500
450
400
350
300
250
200
150
20 reads / 5 conc
50 reads / 10 conc
100 reads / 20 conc
200 reads / 20 conc
500 reads / 50 conc
1000 reads / 50 conc
1000 reads / 100 conc
2000 reads / 100 conc
2000 reads / 200 conc
5000 reads / 100 conc
5000 reads / 200 conc
10000 reads / 200 conc
10000 reads / 500 conc
Test case


Create Stress Test: Throughput
Total attempted RPS
Successful RPS
Requests per second
225
200
175
150
125
100
75
20 req / 5 conc
50 req / 10 conc
100 req / 20 conc
200 req / 20 conc
500 req / 50 conc
1000 req / 50 conc
1000 req / 100 conc
2000 req / 100 conc
2000 req / 200 conc
5000 req / 100 conc
5000 req / 200 conc
Test case


Complex Create/Update Stress Test: Throughput
Total attempted RPS
Successful RPS
Requests per second
30
25
20
15
10
5
20 req / 5 conc
50 req / 10 conc
100 req / 20 conc
200 req / 20 conc
500 req / 50 conc
1000 req / 50 conc
1000 req / 100 conc
2000 req / 100 conc
2000 req / 200 conc
Test case

Pause

10:50 – 10:55

2.3) Maschinelles Lernen (& künstliche Intelligenz)

10:55 – 11:30

Florian Atzenhofer-Baumgartner
Universität Graz – Institut für Digitale Geisteswissenschaften

Leistungen (Hardware)

State-of-the-Art Nvidia GPUs und Knotenkonfiguration

- 8xH200, 6xH200, 4xH200 + 4xRTXPro6K, 7xRTXPro6K, 4xL40s
- Knoten für unterschiedliche **Zwecke** optimiert (aber **flexibel**!)
- 4 % an vRAM des MUSICA-Clusters, aber mit mehr/besseren RAM + Cores

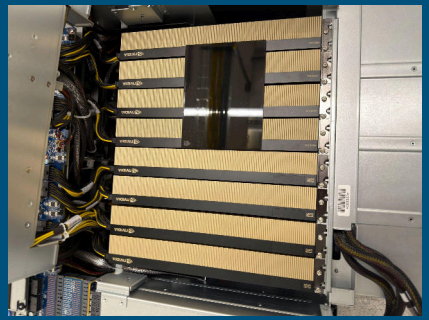
Infiniband Interconnect im Cluster

- (2x) 400Gb/s + RDMA als NFS: Alle Daten sind überall gleich(schnell) erreichbar

ACOnet Uplinks (Schnittstelle zu GÉANT) → **Single Sign-on** (SSO) + Datentransport

- Schnittstelle zu Krems: 10Gb/s, Schnittstelle von Graz: 25Gb/s (bis zu 100!)

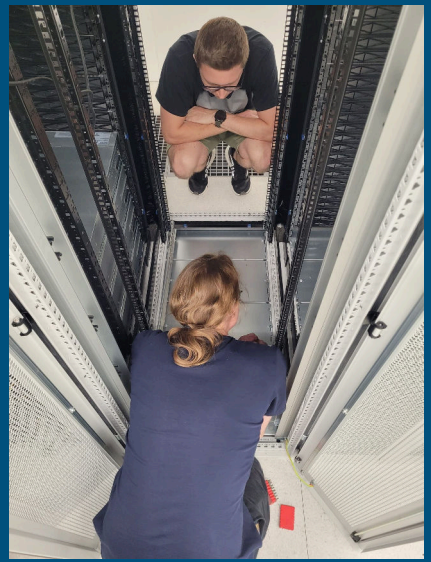


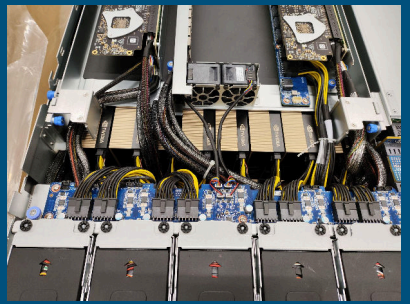












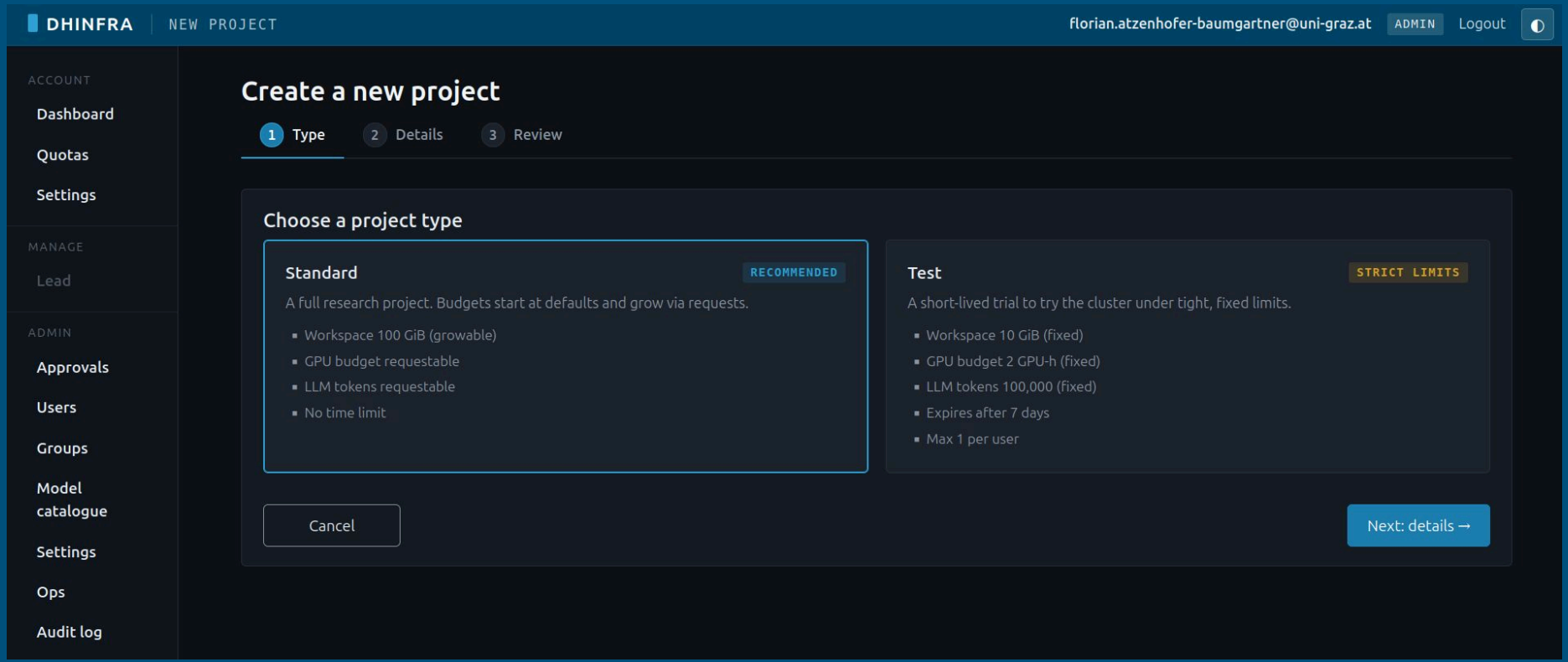
Leistungen (Software)

HPC-Elemente

- SSO, SSH
- Batch Scheduling (SLURM)
- Interaktives Arbeiten (SLURM + Jupyterhub)

Cloud-Elemente → Alleinstellungsmerkmal

- Containerisierung (Docker, Apptainer, LXCs)
- (Custom) (Embedding) APIs
- LLMs & AI (APIs und grafische Oberflächen)


DHINFRA
NEW PROJECT
florian.atzenhofer-baumgartner@uni-graz.at
ADMIN
Logout
ACCOUNT
Dashboard
Quotas
Settings
MANAGE
Lead
ADMIN
Approvals
Users
Groups
Model catalogue
Settings
Ops
Audit log
Create a new project
1 Type
2 Details
3 Review
Choose a project type
Standard
RECOMMENDED
A full research project. Budgets start at defaults and grow via requests.
Workspace 100 GiB (growable)
GPU budget requestable
LLM tokens requestable
No time limit
Test
STRICT LIMITS
A short-lived trial to try the cluster under tight, fixed limits.
Workspace 10 GiB (fixed)
GPU budget 2 GPU-h (fixed)
LLM tokens 100,000 (fixed)
Expires after 7 days
Max 1 per user
Cancel
Next: details →

Anja Grebe, Max Resch
Universität für Weiterbildung Krems – Zentrum für Kulturen und Technologien des Sammelns

Krems: GPU-Cluster

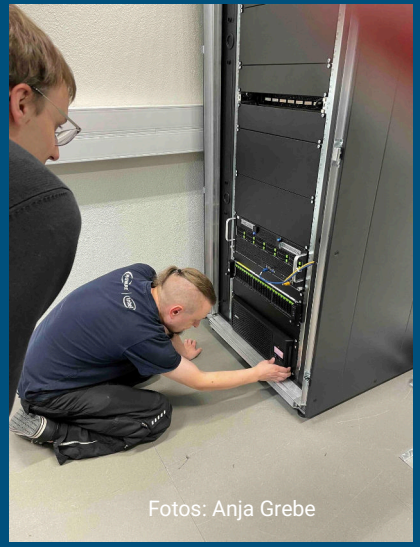
GPU-Cluster

(analog zu Cluster Graz, etwas kleiner) -

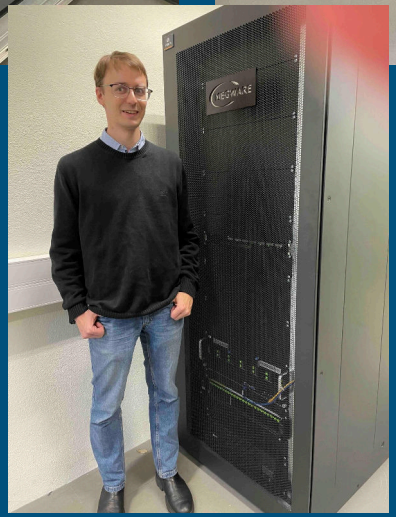
Aufbau und erste Inbetriebnahme







Fotos: Anja Grebe



Krems: CEPH-Komponente

Ceph-Storage Cluster-Komponente (konzipiert analog zu Uni Graz):

- Storage-Cluster mit insges. **4 Knoten**
- Ausstattung mit **Festplatten CMR-Typ** (Conventional magnetic recording)
- **CPU:** 1 Prozessor mit guter Single Core Leistung; mind. 16 Kerne
- **Arbeitsspeicher:** mindestens 768 GB DDR5 ECC, Verhältnis 1:1000 zu verbautem Speicher
- Boot-Speicher: 2 mind. 2 TB (NVMEe, RAID 1), DWPD-Wert 1
- **Speicher:** mind. 12 x mindestens 20 TB nutzbarer CMR Festplattenspeicher (Gesamtkapazität), SAS
- **Netzwerk:** NDR kompatibler Netzwerkadapter, 1GB RJ45 Management LAN

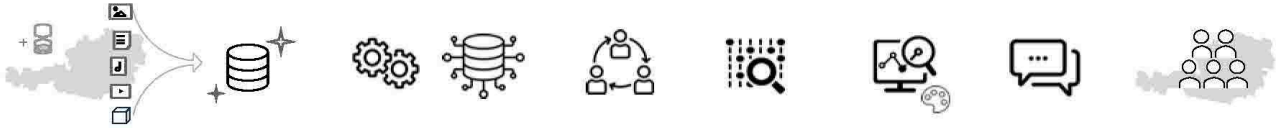
Projekte und Dissemination: Manual (in Ausarbeitung, Entwurf: Florian Windhager, UWK)



Beispiele für ML-Tools

Tool-support for DH practices & workflows

from data capture & creation to wrangling / linking, collaborating, analysis, visualization presentation (and other)



Tools & Apps

- eScriptorium (HTR)
- Whisper (speech2text)
- Tesseract (OCR)
- spacY / Stanza (NER & annotation)
- openRefine (wrangling)
- TEI publisher
- WissKI (semantic data)
- Nodegoat
- Stylo

LLMs (locally hosted) jupyter Notebooks

Projekte und Dissemination

Tools with potential benefit of being deployed on DHinfrastructure

- **eScriptorium** (capture) -- AI-assisted handwritten text recognition; GPU-accelerated, self-hostable, directly relevant for archival and manuscript collections
- **Whisper** (capture) -- Automatic speech-to-text transcription of audio and video recordings; runs locally, benefits significantly from GPU availability
- **Tesseract / OCR pipelines** (capture / wrangling) -- Optical character recognition for printed texts; runs in Jupyter, scalable across large document collections
- **spaCy / Stanza** (wrangling) -- Named entity recognition and linguistic annotation pipelines; GPU-accelerated for large corpora, identify persons, places, and dates across thousands of documents
- **OpenRefine** (wrangling) -- Data cleaning and reconciliation against authority files such as GND or Wikidata; hostable, low barrier, no coding required
- **TEI Publisher** (wrangling / presentation) -- Transforms TEI-encoded scholarly texts into web editions; requires stable hosting to remain publicly accessible
- **WissKI** (collaboration / presentation) -- Semantic data platform widely used in German-language cultural heritage institutions; requires server hosting
- **Nodegoat** (collaboration / analysis) -- Network and relational data modeling with a collaborative database layer; hostable, used across historical and biographical research projects
- **Jupyter Notebooks + locally hosted LLMs** (cross-cutting) -- interactive computing environment usable at every stage, combined with locally hosted language models for transcription support, entity suggestions, summarization, and writing assistance, keeping sensitive data off commercial servers

https://labelstud.io/ - Vorschlag von Max

Projekte und Dissemination

Use Cases – LLM-API

- z.B. **Informationsextraktion** aus mittelalterlichen Rechnungen (Aldersbacher Rechnungsbücher, FWF, PI Robert Klugseder) oder aus Kurztexten (ManMax, siehe folgende Folien)

Use Cases – Training (+ Inference)

- **Segmentierung** und **OCR** in z.B. Zeitungen (ANNO-Corpus, “JobAds”, PI Jörn Kleinert) oder der Schematismus (FWF “Unlocking the Schematismus”, PI Wolfgang Göderle; siehe folgende Folien)

Suzana Sagadin
Universität Graz – Institut für Digitale Geisteswissenschaften

Managing Maximilian (SFB 92)

Teilprojekte:

Armouring Maximilian *(Rüstung)*
(Con)Textualising Maximilian *(Lateinische Literatur)*
Creating Maximilian *(Text und Bilder)*
Depicting Maximilian *(Künstlerische Produktion)*
Digitising Maximilian *(Digitale Methoden)*
Gendering Maximilian *(Analyse der Rollen von Frauen)*
Singing Maximilian *(Musik)*
Writing Maximilian *(Korrespondenz und Kanzlei)*

Digitising Maximilian
Georg Vogeler (PI)

- *Zusammenführung der Daten aller Teilprojekte in einer (prosopographischen) Datenbank*
- *Auswertung mit digitalen Methoden*

Model - Erfassung - Extraktion

- Tiefes Datenmodell durch diverse Forschungsinteressen
- **117 Statement-Typen**
- Ursprünglich: manuelle Dateneingabe
- **LLMs** zur Unterstützung des Prozesses

Extraktions-Beispiel

Input-Text

1291

1517 November 7, Innsbruck.

Christof Geiger, Bildhauer, erhält zur Reise nach Wels von der landesfürstlichen Kammer 5 Gulden.

Raitbuch 1517, f. 261.

Strukturierter Output

① **roleoccupation**

"C. Geiger ist ein Bildhauer"

Amtsträger: C. Geiger
Amt: Bildhauer
Datum: bis 1517-11-07

② **payment**

"C. Geiger erhält von der landesfürstlichen Kammer 5 Gulden"

Zahlender: landesfürstlichen Kammer
Empfänger: Christof Geiger
Betrag: 5 Gulden
Datum: 1517-11-07
Zahlungsgrund:

└ **journey**

"C. Geiger reist nach Wels"

Reisende Person: C. Geiger
Ziel: Wels
Datum: ab: 1517-11-07

Die Pipeline

- Extraktion als Kette klassischer NLP-Aufgaben
  - Event Extraction
  - Text Classification
  - Relation Extraction
  - Template Filling
  - Entity Typing
- Lokale Ausführung mit kleinen Modellen
  - Nvidia RTX 4500
  - Kleine Modelle (z. B. Qwen3-14B_Q6, Qwen3.5-35b_Q4)

Writing Maximilian Regesten

- 128 Regesten
- im Schnitt 4x so lang wie die bisherigen Regesten
- teilweise deutlich komplexer

→ stärkeres Modell

	bisherige Regesten	Writing Maximilian
Durchschnitt	260	1091
Median	261	846
Kürzeste	103	314
Längste	410	5339

Alle Werte in Zeichen

DHinfra LLM-API

- Qwen3.5-397B
- Implementierung:
  - OpenAI-kompatibles API-Format — gleich wie bei kommerziellen Anbietern und unserer lokalen Variante.
  - Konfiguration: URL, Modellname, Authentifizierung
  - kein Umbau der Pipeline nötig

	35B (**lokal**), T=0.0	**397B thinking (cluster), T=0.0**	397B no thinking (**cluster**), T=0.0
Statements gesamt	1 395	**2 072**	2 056
davon top-level	838	**1 383**	1 236
davon nested	557	**689**	820
Laufzeit	43 h 23 min	**4 h 17 min**	23 min

David Fleischhacker
Universität Graz – Institut für Digitale Geisteswissenschaften (u.a.)

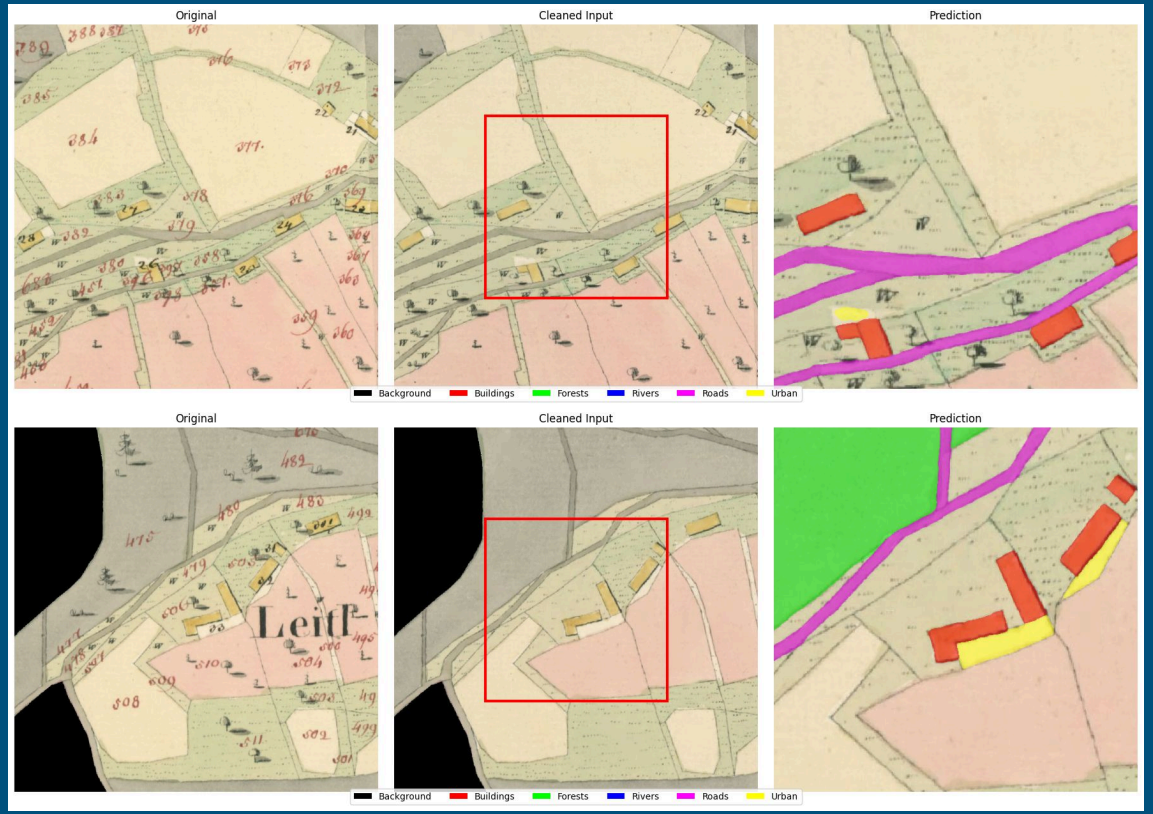
Kataster Segmentierung

(Moving in Space and Time)

Inpainting:
VLM auf ~8.000 synthetische Bilder trainiert
Entfernt Schrift und Symbole
generativer Ansatz, 20 Mrd. Parameter

Segmentierung:
Modell trainiert auf ~30.000 echten Kacheln
Erkennt 6 Kategorien
Treffsicherheit ~87% (mIoU)

Beide Modelle auf 8× NVIDIA H200 trainiert



Schematismus

(Unlocking the Schematismus)

Layout Detection:

Vortrainiert auf 200.000 fake Seiten

Fine-tuned auf ~2.000 echte Dokumente

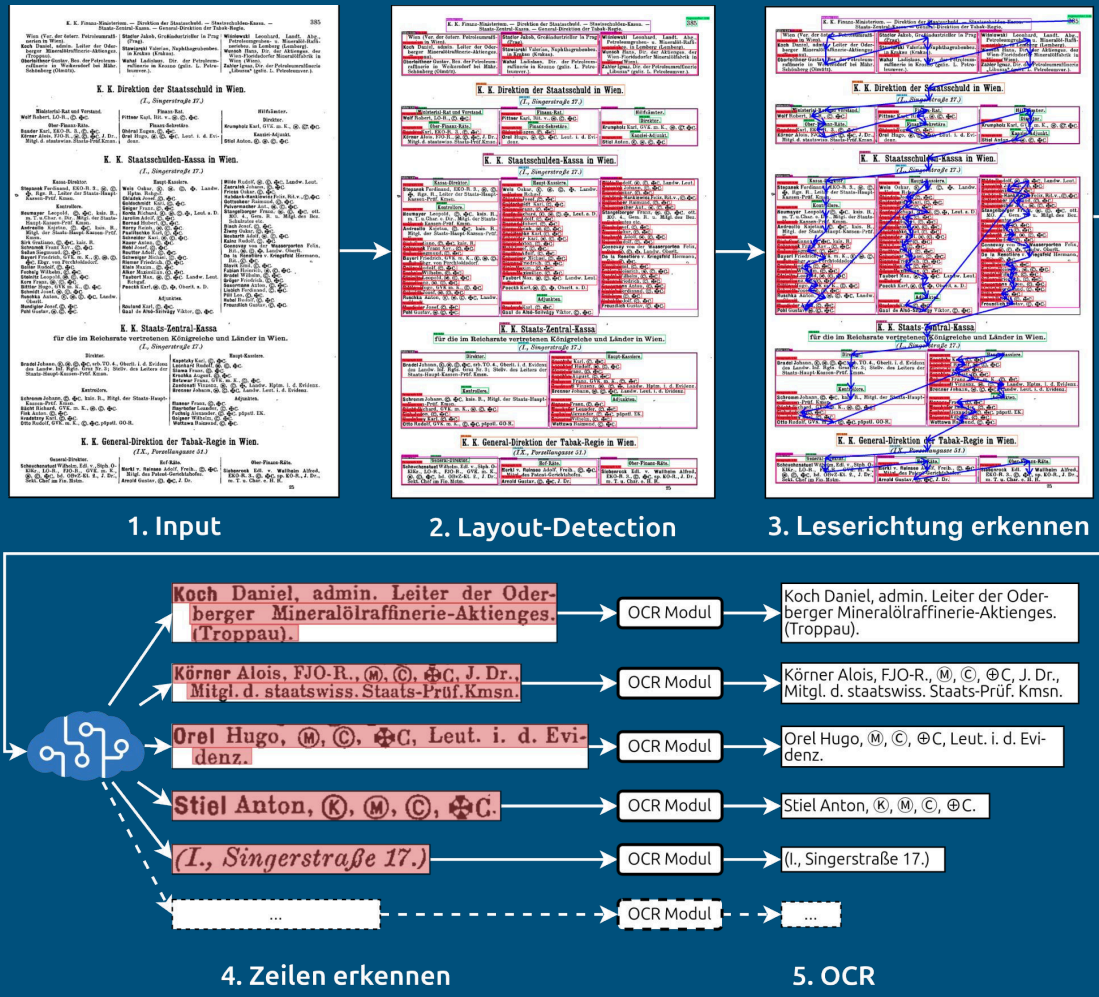
OCR:

Optimiert auf Schematismus

Erkennt Sonderzeichen

Ziel:

Wissensdatenbank



HabsburgLLM

(Teaser)

“Eine Maschine, die wie das 19. Jahrhundert denkt”

Aktueller Stand:
Sprachmodell **komplett von Null** trainiert
nur historische Londoner Texte (1800-1875)
~90GB reiner Text - kein modernes Wissen
insgesamt 100 Stunden auf 8x NVIDIA H200 trainiert



Eingabe: *“If I had to describe the idea of a flying machine, I would…”*

Ausgabe: *“...call it an apparatus of conveying souls from one planet to another.”*

Nächste Schritte:
Modell lernt Habsburger Welt - Jahrzehnt für Jahrzehnt
und dabei auch Deutsch (statt Englisch)
Daten-Stopp: kurz vor 28. Juni 1914 (Attentat Franz Ferdinand)

Raven Adam, **Florian Atzenhofer-Baumgartner**
Universität Graz – Institut für Digitale Geisteswissenschaften (u.a.)

Synergien im Konsortium: Beispiel ÖNB

Anwendungen mit hoher Schlagkraft (e.g., in https://anno.onb.ac.at/)

- Objekterkennung → Zeitungen werden klassifizierbar
- Texterkennung → Zeitungen werden durchsuchbar

Nebeneffekte

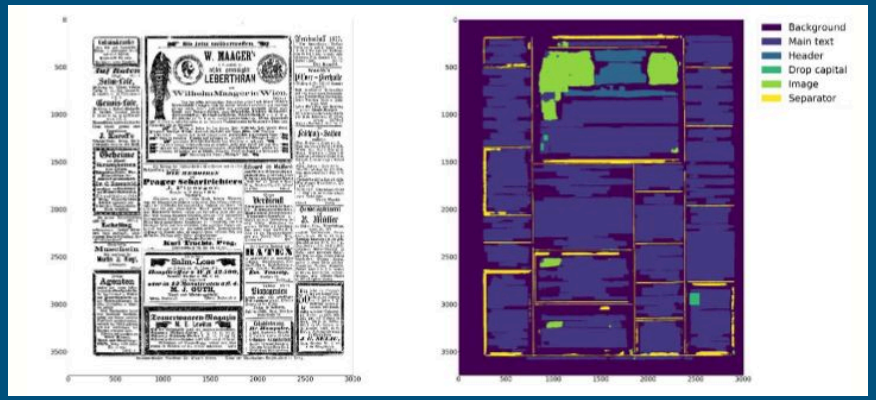
- schnellere und robustere Programme und **Schnittstellen**
- neue (institutionsübergreifende) Projekte, die zuvor unmöglich waren

Beispiel: Historische Stellenanzeigen / JobAds

DH-Beitrag: Pipeline zur Erkennung und Analyse von historischen Stellenanzeigen

Effizienzsteigerungen durch Kombination aus leistungsstärkeren GPUs, CPUs, mehr RAM und direkten Mount der Cloud inklusive Parallelisierung

https://historical-job-ads.uni-graz.at/en/about-the-project/

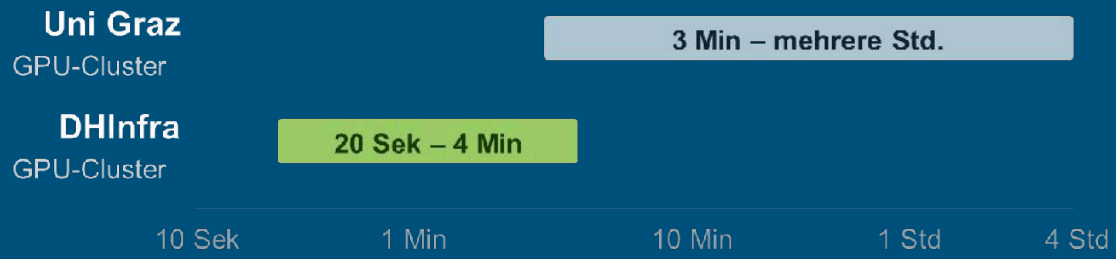




https://static.uni-graz.at/fileadmin/_files/_project_sites/_historical-job-ads/Bilder/CLARIN_job_ads.pdf

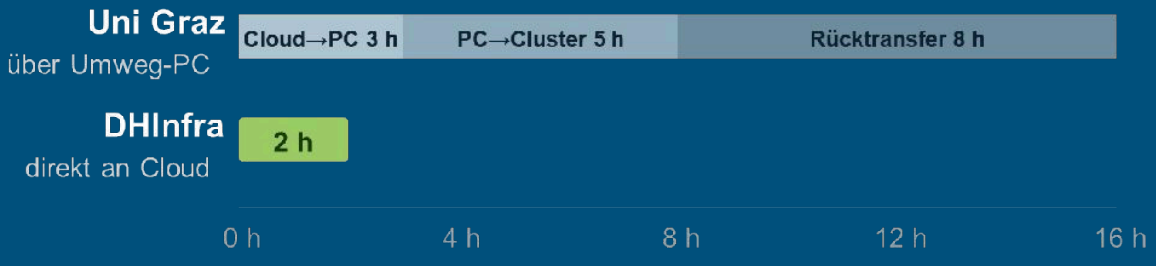
Engpass 1 — Rechenzeit pro Seite

Layoutanalyse



Engpass 2 — Datentransport pro Zeitung

Datenübertragung



2.4) Freie und offene Software

11:30 - 11:45

Florian Atzenhofer-Baumgartner
Universität Graz – Institut für Digitale Geisteswissenschaften

Prozess

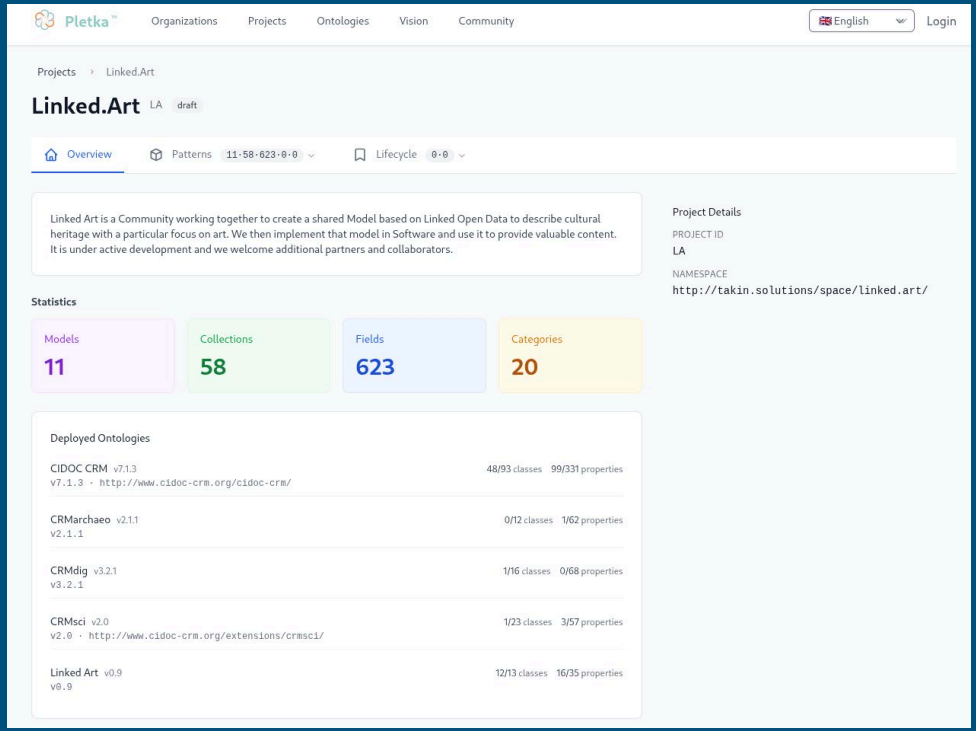
1. Bedarfserhebung (durch systematische Befragungen)
2. **Evaluierung** (mit Kriterienkatalog)
3. Bewertung (inkl. 3 **Workshops** + **Whitepaper**)
4. Ausschreibung (=vergleichende Auswahl von Devs)
5. **Entwicklung** (=Begleitung)
6. Abnahme (mit technischer und rechtlicher **Prüfung**)
7. Integration (im Konsortium und der ganzen Welt)

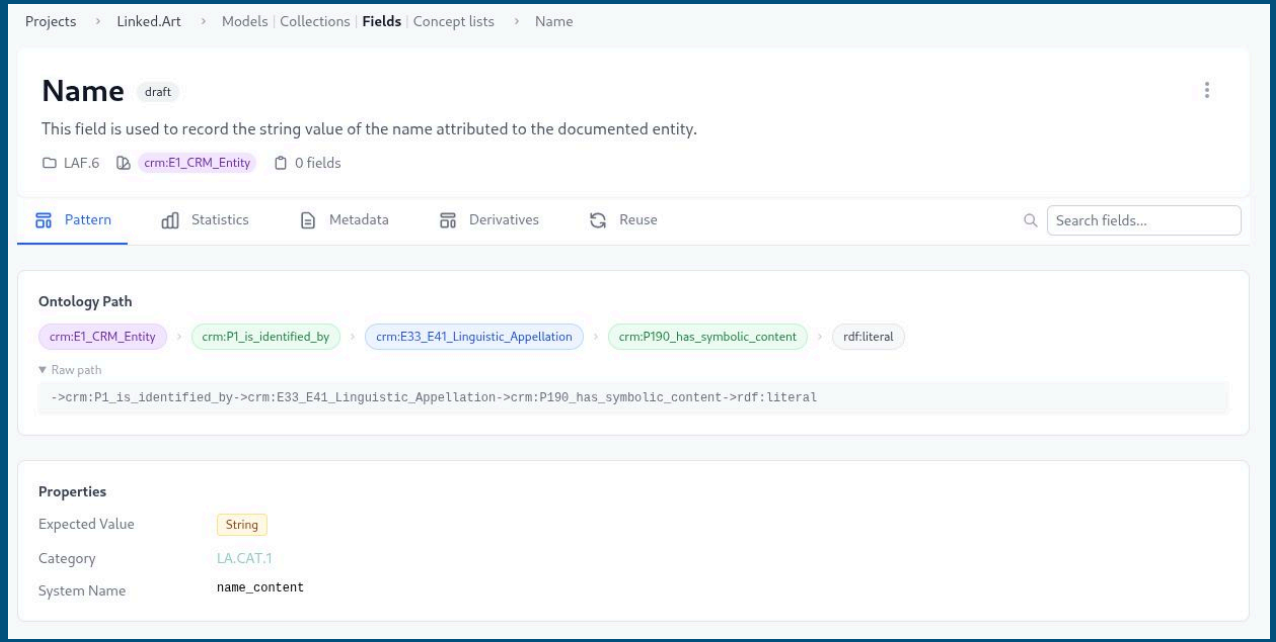


Matthias Schlögl
Österreichische Akademie der Wissenschaften – Austrian Centre for Digital Humanities

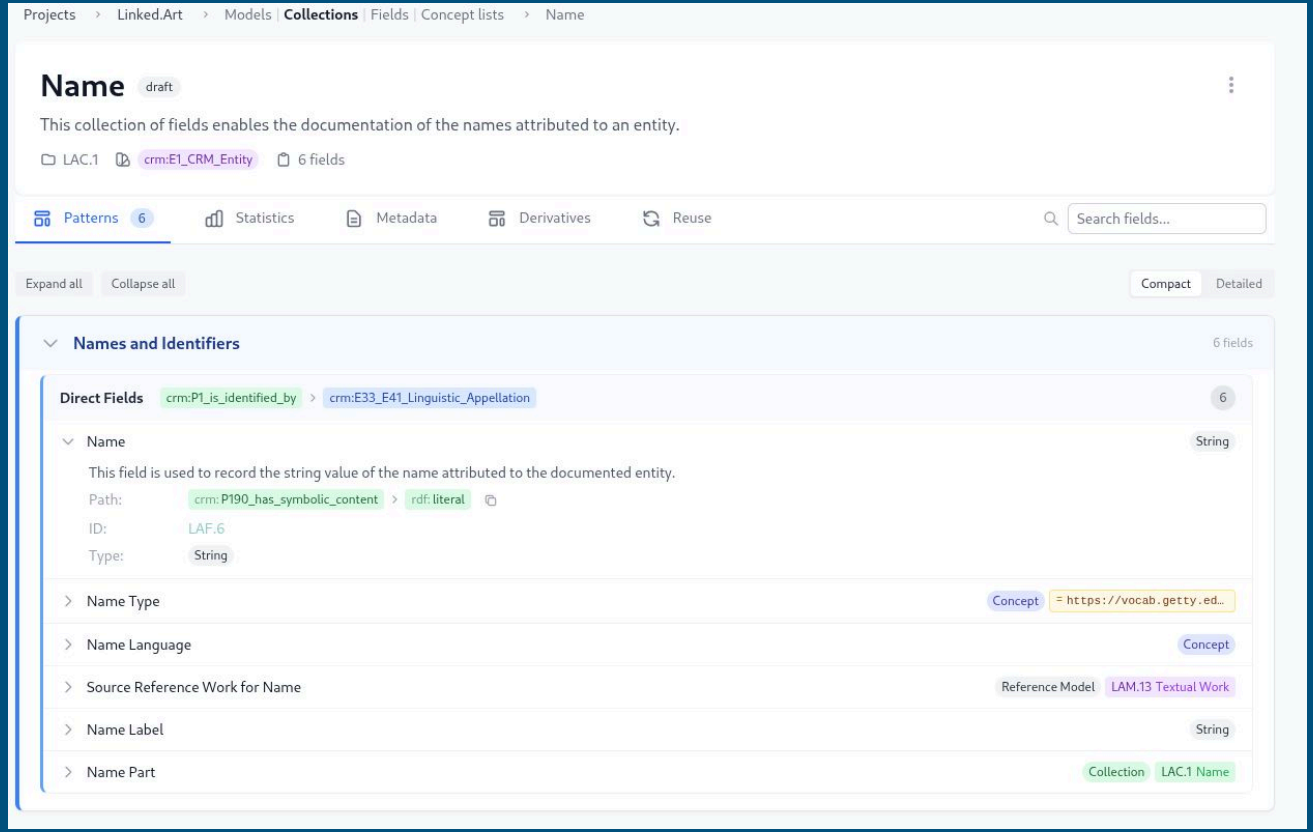
Pletka (Zellij)

- kollaborative Arbeit an Ontologien
- einheitliches Format und Datenmodell
- Dokumentation und Export in andere Formate (RDFS, SHACL, Researchspace …) an einem Ort
- https://beta.pletka.io/

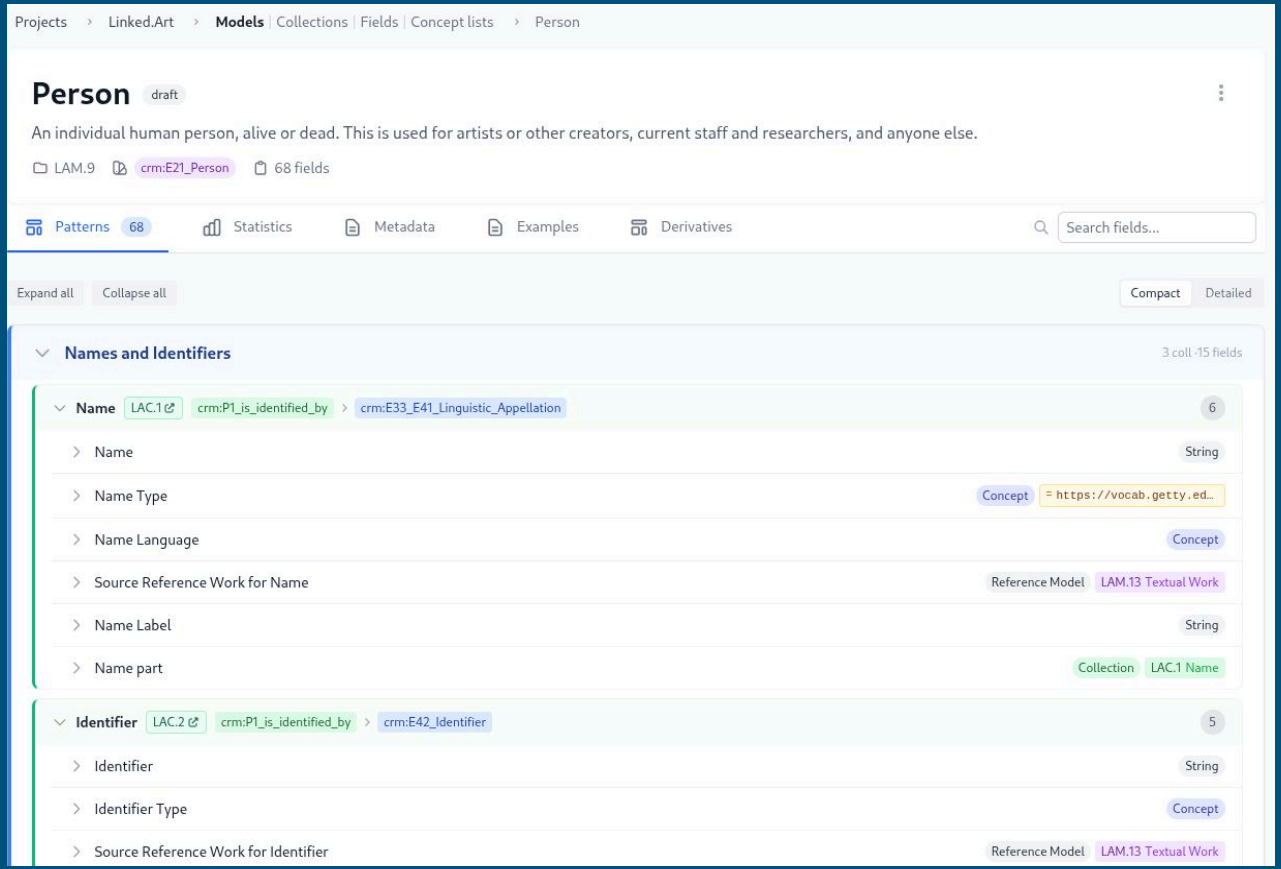




Fields, z.B. "Name": crm:E1_CRM_Entity -> crm:P1_is_identified_by -> crm:E33_E41_Linguistic_Appellation -> crm:P190_has_symbolic_content -> rdf:literal



Collections, z.B. "Name" -> Zusammenstellung von Fields die Dokumentation einer wiederkehrenden Sache (z.B. "Name") erlauben



Models, z.B. "Person" -> Zusammenstellung von Collections und/oder Fields um eine Entität zu beschreiben

Pletka als Community-Tool

- Institutionen
- Innerhalb der Institutionen: Projekte, Modelle etc.
- Veröffentlichte "pattern libraries" können wiederverwendet werden
- Versionierung
- Zusammenarbeit über Teams und Organisationen hinweg
- lokal gehostete Instanzen

DHInfra Finanzierung

- Pluggable Backend (Go + PostgreSQL statt Airtable)
- Migration bestehender Projekte
- Versionierung
- Web-basiertes Admin-Dashboard (Svelte Frontend)
- Integration Dashboard mit Backend
- Veröffentlichung unter OSI-konformer Lizenz

Florian Wachter
Universität Graz – Institut für Digitale Geisteswissenschaften (u.a.)

QLever

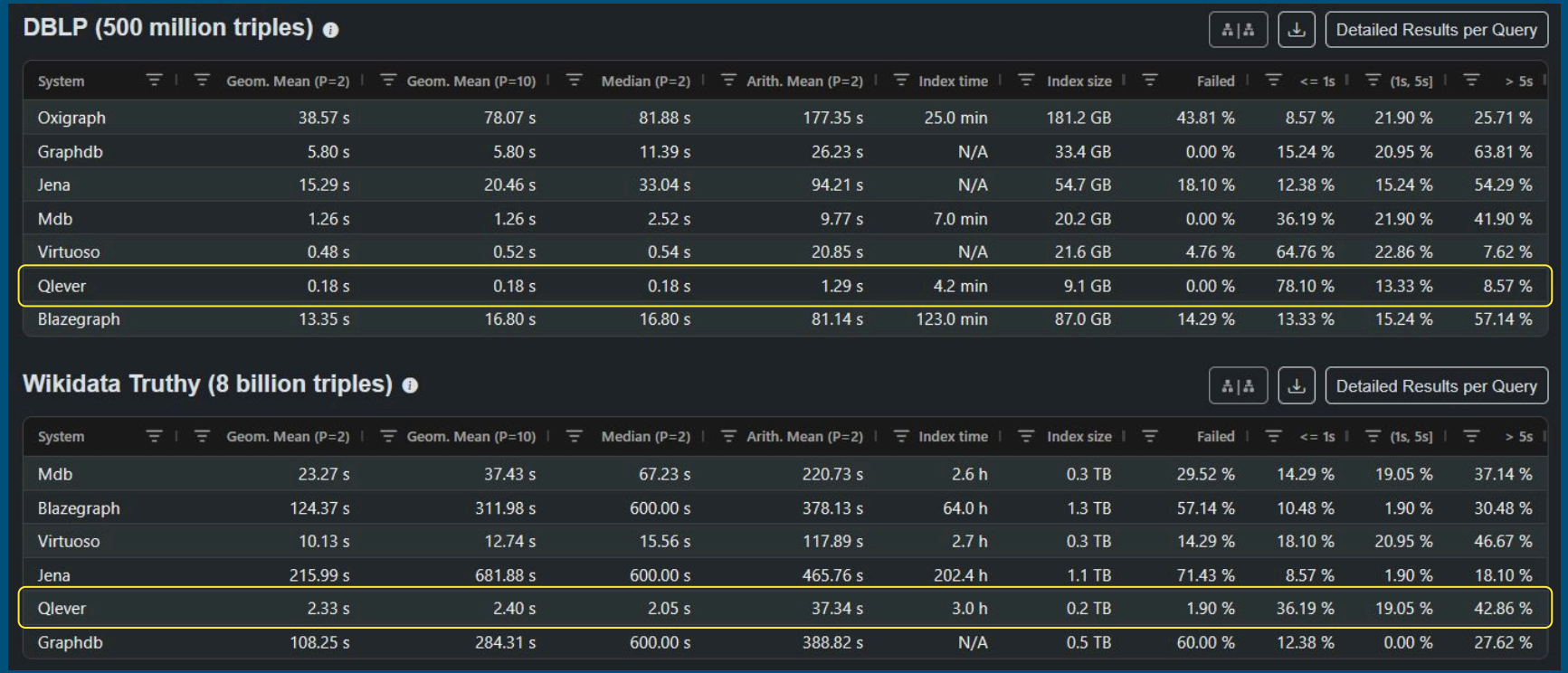
About

- Effiziente SPARQL Engine
- Über 100 Mrd. Tripel auf einem einzelnen Standardrechner
- Deutlich schnellere Antworten als bei anderen Optionen

Probleme

- Steile Einstiegshürde für neue Nutzer:innen.
- Dokumentation verstreut, lückenhaft und teils veraltet.
- Keine standardisierten, wiederverwendbaren Beispiele (Use Cases).

QLever – qlever.dev/evaluation/www/



DBLP (500 million triples)

Detailed Results per Query

System	Geom. Mean (P=2)	Geom. Mean (P=10)	Median (P=2)	Arith. Mean (P=2)	Index time	Index size	Failed	<= 1s	(1s, 5s]	> 5s
Oxigraph	38.57 s	78.07 s	81.88 s	177.35 s	25.0 min	181.2 GB	43.81 %	8.57 %	21.90 %	25.71 %
Graphdb	5.80 s	5.80 s	11.39 s	26.23 s	N/A	33.4 GB	0.00 %	15.24 %	20.95 %	63.81 %
Jena	15.29 s	20.46 s	33.04 s	94.21 s	N/A	54.7 GB	18.10 %	12.38 %	15.24 %	54.29 %
Mdb	1.26 s	1.26 s	2.52 s	9.77 s	7.0 min	20.2 GB	0.00 %	36.19 %	21.90 %	41.90 %
Virtuoso	0.48 s	0.52 s	0.54 s	20.85 s	N/A	21.6 GB	4.76 %	64.76 %	22.86 %	7.62 %
Qlever	0.18 s	0.18 s	0.18 s	1.29 s	4.2 min	9.1 GB	0.00 %	78.10 %	13.33 %	8.57 %
Blazegraph	13.35 s	16.80 s	16.80 s	81.14 s	123.0 min	87.0 GB	14.29 %	13.33 %	15.24 %	57.14 %

Wikidata Truthy (8 billion triples)

Detailed Results per Query

System	Geom. Mean (P=2)	Geom. Mean (P=10)	Median (P=2)	Arith. Mean (P=2)	Index time	Index size	Failed	<= 1s	(1s, 5s]	> 5s
Mdb	23.27 s	37.43 s	67.23 s	220.73 s	2.6 h	0.3 TB	29.52 %	14.29 %	19.05 %	37.14 %
Blazegraph	124.37 s	311.98 s	600.00 s	378.13 s	64.0 h	1.3 TB	57.14 %	10.48 %	1.90 %	30.48 %
Virtuoso	10.13 s	12.74 s	15.56 s	117.89 s	2.7 h	0.3 TB	14.29 %	18.10 %	20.95 %	46.67 %
Jena	215.99 s	681.88 s	600.00 s	465.76 s	202.4 h	1.1 TB	71.43 %	8.57 %	1.90 %	18.10 %
Qlever	2.33 s	2.40 s	2.05 s	37.34 s	3.0 h	0.2 TB	1.90 %	36.19 %	19.05 %	42.86 %
Graphdb	108.25 s	284.31 s	600.00 s	388.82 s	N/A	0.5 TB	60.00 %	12.38 %	0.00 %	27.62 %

QLever

Dokumentation

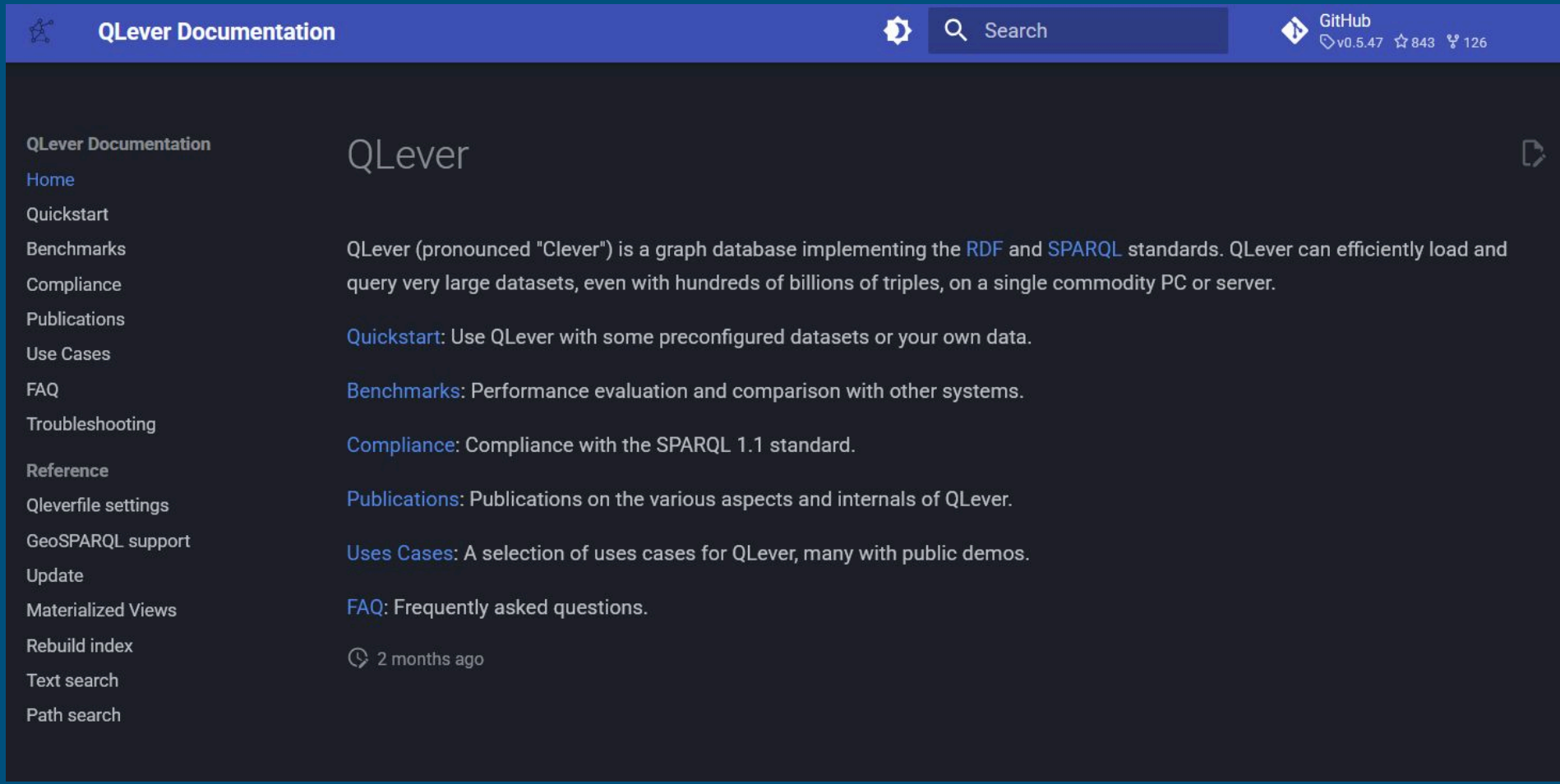
- Komplett neu: **https://docs.qlever.dev/** (seit Ende 2025), laufend erweitert
- Quickstart, “Qleverfile”-Referenz, Troubleshooting

Use-Cases

- Vereinheitlichte Qleverfiles (24 Commits seit Dez. 2025)
- Öffentliche Bereitstellung per Pull-Request → qlever setup-config
- Übersicht: **https://docs.qlever.dev/use-cases/** mit Größe, Ressourcen, Live-Demos

Deployment

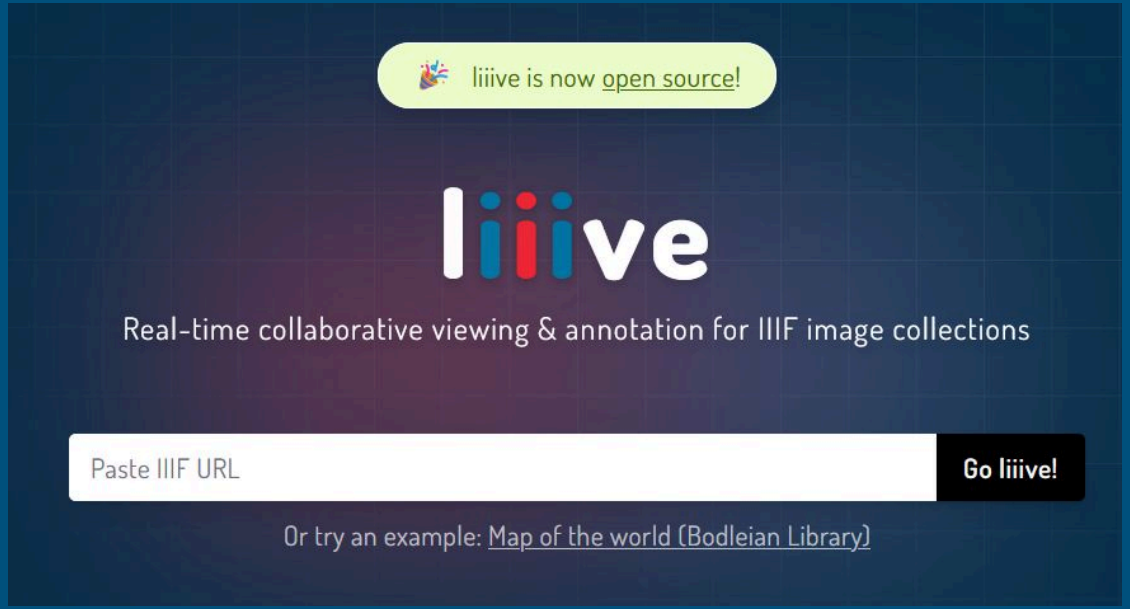
- Offizielle Pakete für Debian/Ubuntu/macOS (seit 21.01.2026)
- Docker-Images: QLever UI & Petrimaps; qlever ui nativ und per Docker
- Reproduzierbar und vor jedem Release getestet
- Lizenz: Apache v2



docs.qlever.dev

Lukas Waldhofer
Universität Graz – Institut für Digitale Geisteswissenschaften

https://liiive.now



https://github.com/rsimon/liiive

liiive.now – Features

Funktionen

- **Echtzeit-Zusammenarbeit**
- Gemeinsame Navigation
- Cursor-Tracking live
- Kommentare / Threads
- **IIIF API Support**
- Annotation Export
- Anonyme Sessions via Link

Tech-Stack

- Web-App (React + Astro)
- WebSockets (Realtime)
- Supabase Datenbank
- IIIF Image + Presentation API
- Self-hostable
- MIT Lizenz

Hosted Version

- keine Installation
- sofort nutzbar
- ideal für Lehre / Demo

Self-Hosting

- eigener Server / VPS
- Uni / Institution / Cloud
- volle Datenkontrolle

Architektur

- Client (Browser)
- Realtime Sync (WebSockets)
- IIIF Integration
- Session-basiert (Link Sharing)

liiive.now – Use Cases

Digital Humanities

- Manuskriptanalyse
- Quellenannotation
- Forschungsdiskussionen

Lehre

- Seminare
- Gruppenarbeit
- interaktive Quellenarbeit

Kultureinrichtungen

- Museumsarbeit
- Kuratorische Zusammenarbeit
- Dokumentation

Forschung

- Remote Collaboration
- gemeinsame Annotation
- synchrone Diskussion

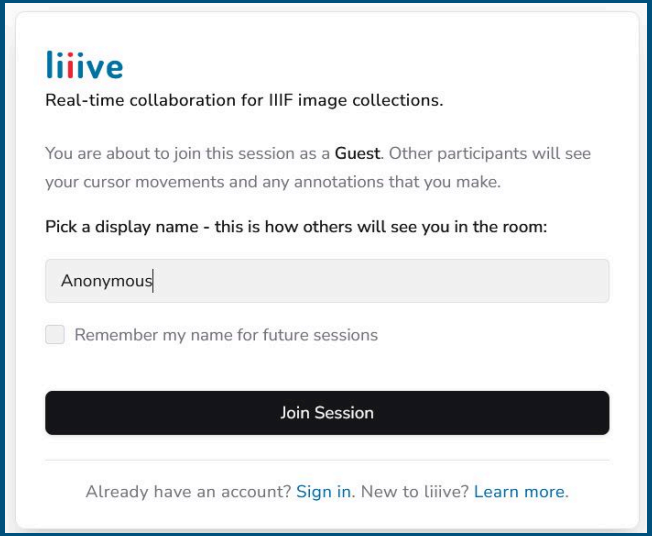
Vorteile

- Echtzeit
- Open Source
- IIIF Standard
- einfache Zusammenarbeit via Link
- kein Login nötig
- geeignet für verteilte Teams

1) IIIF-Link eingeben



2) Raum öffnen




Share
3:58
AN
3) Annotieren, kommentieren etc.
4) Teilen, downloaden etc.
1 / 1

3) Ausblick

Walter Scholger, Florian Atzenhofer-Baumgartner
Universität Graz – Institut für Digitale Geisteswissenschaften

Ausblick

Juni 2026	dhinfra.at – Relaunch
	Ende der (internen) Testphase
Juli 2026	CLARIAH-AT Projektförderung (6 Monate)
Q4/2026	clariah.at – Relaunch
(2027- ?)	Nationale Verstetigung via CLARIAH-AT und Service Desk(s)
(2027- ?)	Internationale Verstetigung via HORIZON

Danksagungen

ASC/VSC sowie ACA/EuroCC für Expertise

ASHPC24, ASHPC26, DHd26: Community Feedback & Stipendien

IT Services aller Institutionen, v.a., Universität Graz und Krems

Uni Graz, Institut für Mathematik und Wissenschaftliches Rechnen für Input

TU Graz, Institute of Machine Learning and Neural Computation für Input

Universität Wien zur (finanziellen) Förderung des FOSS-Whitepapers

BMFWF und EU-RRF via “(Digitale) Forschungsinfrastrukturen”

CLARIAH-AT

Bundesministerium
Frauen, Wissenschaft
und Forschung

Funded by
the European Union

DH INFRA

Universität für
Weiterbildung
Krems

DONAU UNIVERSITÄT KREMS

ÖAW ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN

Danke!

PARIS
LODRON
UNIVERSITÄT
SALZBURG

universität
innsbruck

Österreichische
Nationalbibliothek

TU WIEN TECHNISCHE UNIVERSITÄT WIEN

UNI GRAZ

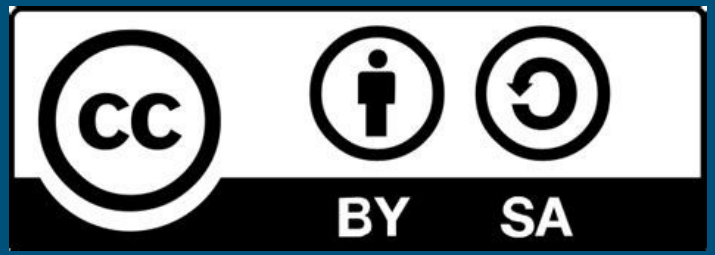
universität
wien

di:'ʌngewʌndtə

Q&A

Fragen? Gerne JETZT, ansonsten dhinfra@uni-graz.at

Die Inhalte dieser Folien sind unter CC-BY-SA nachnutzbar, außer anders gekennzeichnet.



Zitate zu wissenschaftlichem Gebrauch unterliegen den Lizenzen der jeweils angegebenen Quellen.