

LMU

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

IT-Gruppe Geisteswissenschaften
DH Kompetenz- und Datenzentrum



Geisteswissenschaftler für digitale Methoden begeistern – Das Digital Humanities Virtual Laboratory (DHVLab)

Dr. Christian Riepl, IT-Gruppe Geisteswissenschaften

Workshop "Digitale Methoden in der Forschung"
6. November 2019, GATE Garching
INDIGO/ZD.B



1. Organisatorischer Hintergrund
2. Konzept der Wissensvermittlung
3. Architektur des DHVLab
4. Komponenten des DHVLab
5. Einsatz und Erfahrungen
6. Zukunftsperspektiven



1. Organisatorischer Hintergrund
2. Konzept der Wissensvermittlung
3. Architektur des DHVLab
4. Komponenten des DHVLab
5. Einsatz und Erfahrungen
6. Zukunftsperspektiven

LMU

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

IT-Gruppe Geisteswissenschaften
DH Kompetenz- und Datenzentrum



IT-Gruppe Geisteswissenschaften DH Kompetenz- und Datenzentrum

IT-Infrastruktur

Forschungsdaten-
management



Forschung & Lehre
digital

LMU

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

IT-Gruppe Geisteswissenschaften
DH Kompetenz- und Datenzentrum



Kooperation und Vernetzung

Geisteswissenschaften

UB LMU

UB FAU

Informatik/Statistik

Dezernat VI

LRZ



dhmuc. Netzwerk DH München
(BADW, BSB, DM, LMU, ...)

téchnē – Campus-Netzwerk
Digitale Geistes- und Sozialwissenschaften
(FAU, UR, LMU, ...)

LMU

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

IT-Gruppe Geisteswissenschaften
DH Kompetenz- und Datenzentrum



Digitaler Campus Bayern

- 2016: IT for All: Digitale Datenanalyse
- 2017: BA-Nebenfach „Digital Humanities - Sprachwissenschaften“
- 2018: Forschungsdatenmanagement Bayern
- 2019: IT for All: DHVLab plus

DHVLab I

DHVLab II

DHVLab II+

Bayerisches Staatsministerium für
Wissenschaft und Kunst





1. Organisatorischer Hintergrund
2. Konzept der Wissensvermittlung
3. Architektur des DHVLab
4. Komponenten des DHVLab
5. Einsatz und Erfahrungen
6. Zukunftsperspektiven



Konzept der Wissensvermittlung

- Zielgruppe: Studierende technikferner Studiengänge
- Ausgangspunkt: Geisteswissenschaftliches Fach
 - fachnah und praxisbezogen
 - am Gegenstand historischer, kunst- und sprachwissenschaftlicher Fächer
 - fachspezifische Methodik und Kompetenzen
- Kompetenzen in Informatik/Statistik
 - Datenakquisition und Datenmodellierung
 - Datenannotation
 - Datenanalyse
 - Datenvisualisierung

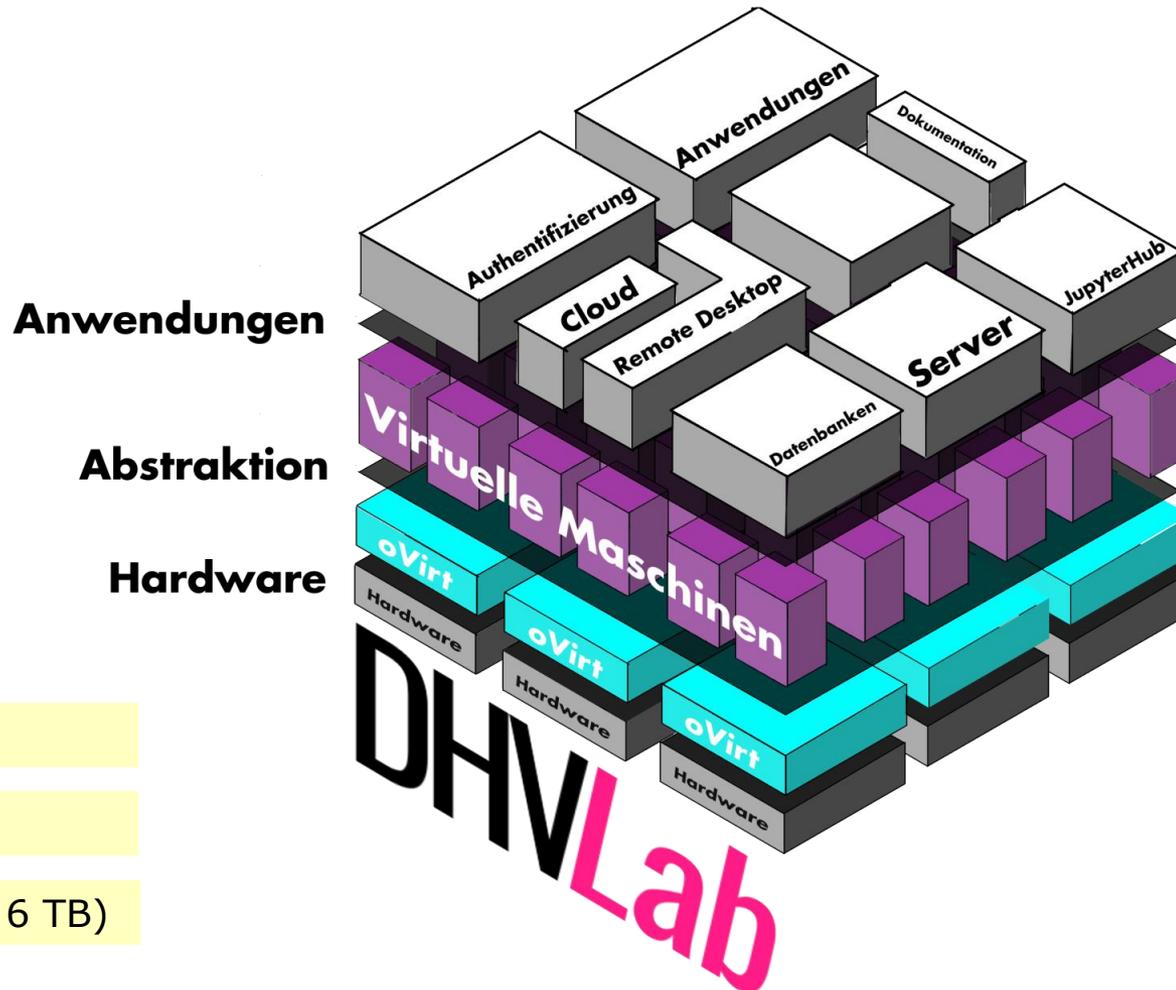


Konzept der Wissensvermittlung

- Veranstaltungen
 - interdisziplinär und kooperativ
 - Präsenzveranstaltungen basierend auf Online-Komponente
- Online-Komponente mit vier Modulen: DHVLab
 - virtuelle Rechenumgebung
 - Datenbankumgebung
 - Dokumentations-/Publikationsumgebung
 - Lehr-/Lernumgebung
 - einheitliche Software-Umgebung und niedrige Einstiegshürde



1. Organisatorischer Hintergrund
2. Konzept der Wissensvermittlung
3. **Architektur des DHVLab**
4. Komponenten des DHVLab
5. Einsatz und Erfahrungen
6. Zukunftsperspektiven



8 Server

4 GPUs

1 Storage (ca. 6 TB)

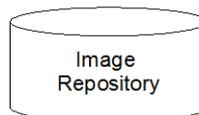


Zukünftige Entwicklung

- Von komplexer Infrastruktur hin zur Beschreibung von Diensten (Kubernetes, Docker Images)
- Bereitstellung zentraler Dienste (DNS, Authentifizierung, Git, Blog, ...)
- Einfache Integration neuer Angebote

DHVLab

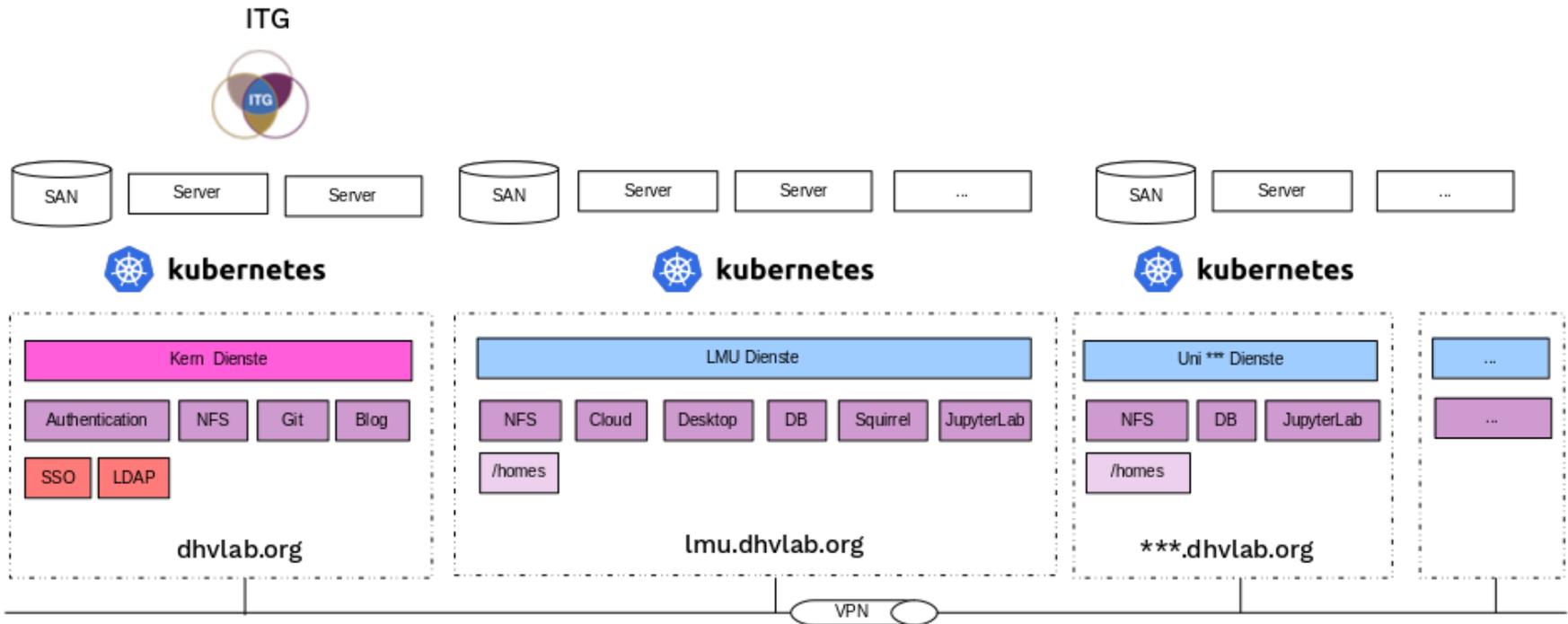
Docker Images



Helm
Charts

Kubernetes

Aufteilung in zentrale und dezentrale Dienste





1. Organisatorischer Hintergrund
2. Konzept der Wissensvermittlung
3. Architektur des DHVLab
4. **Komponenten des DHVLab**
5. Einsatz und Erfahrungen
6. Zukunftsperspektiven

LMU

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

IT-Gruppe Geisteswissenschaften
DH Kompetenz- und Datenzentrum



DHVLab

[Home](#) [Module](#) [Über](#) [Team](#) [FAQs](#) [Kontakt](#) [Vorteile](#) [Technologie](#) 

DIGITALE LEHR- UND
FORSCHUNGSINFRASTRUKTUR FÜR DIE
GEISTESWISSENSCHAFTEN

DHVLAB

Registrieren

MODULE



VIRTUELLER DESKTOP

Der Virtuelle Desktop bietet eine Reihe von Software und Tools für einen vielfältigen Einsatz in Lehrveranstaltungen und in der Forschung.



DATENREPOSITORY

Die Forschungsdaten werden im Datenrepository langfristig und sicher abgelegt. Eine Nachnutzbarkeit vorhandener Datenbestände ist möglich.



DOKUMENTATION

Die Dokumentationsplattform dient Studierenden und Lehrenden zur Vor- und Nachbereitung von Sitzungen sowie zur Publikation von Beiträgen und Analyseergebnissen.



CLOUD

Über die Cloud können Dateien zwischen der Umgebung und dem lokalen Rechner ausgetauscht werden.



SQUIRREL

Squirrel dient dem Erwerb grundwissenschaftlicher Kernkompetenzen im Bereich Transkription und Edition. Die technischen Möglichkeiten des Tools werden sukzessive erweitert.

[Projektseite](#)



LIVE-CODE

„Jupyter Notebook“ dient dem praxisnahen Einstieg in die Programmierung. Dokumente mit Live-Code, Gleichungen und Visualisierungen können erzeugt und geteilt werden.



GITLAB

Ob für ein Projekt im Seminar oder ein individuelles Forschungsvorhaben - das Versionierungssystem Gitlab ermöglicht die kollaborative und versionierte Ablage von Dateien.



MAX

MAX - Museum Analytics ist ein modulares, dynamisches und interaktives Online-Tool zur vergleichenden Analyse musealer Datenbestände mit Hilfe von Distant Viewing.

[Projektseite](#)



ANALYSECENTER

Das explorative Werkzeug eröffnet Ihnen mit Hilfe statistischer Methoden neue Zugänge und Möglichkeiten.

Programmierkenntnisse sind nicht erforderlich.



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

IT-Gruppe Geisteswissenschaften
DH Kompetenz- und Datenzentrum



DHVLab

[Home](#)

[Module](#)

[Über](#)

[Team](#)

[FAQs](#)

[Kontakt](#)

[Vorteile](#)

[Technologie](#)



Verwaltungsfunktionen

Studierende

[Registrieren](#)

[Passwort zurücksetzen](#)

Lehrende

[Zugang](#)

[Verwaltungsoberfläche](#)



Handbücher

Grundlegende Einführungen in informationstechnologische Felder bieten die nachfolgenden Manuale, die praxisnah und den Anforderungen und Bedürfnissen geisteswissenschaftlicher Studierender entsprechend konzipiert wurden.

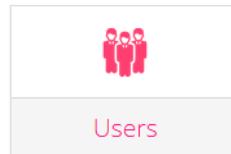
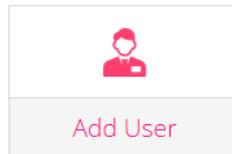
[Einführung in die Benutzung des DHVlab](#)
[Einführung in die Tabellenkalkulation \(Calc, Excel\)](#)
[Einführung in die Arbeit mit Datenbanken \(SQL\)](#)
[Einführung in die Statistik \(R\)](#)
[Einführung in die Website-Erstellung \(im Aufbau\)](#)



Anwendungsfälle

Gemeinsam mit fachwissenschaftlichen Dozierenden konzipieren wir Lehrveranstaltungen mit IT-Grundlagenausbildung. Diese Ausbildungsvorhaben stellen wir Ihnen an dieser Stelle gerne zum Selbststudium zur Verfügung.

[Anwendungsfall „Datenbankabfragen mit R und SQL“](#)
[Anwendungsfall „Cluster und Korrelationen“](#)
[Anwendungsfall „Vom gedruckten Buch zur digitalen Karte“](#)
[Anwendungsfall „Digitales Edieren in der Praxis“](#)



Management UI

- Admin (ITG)
 - legt Account mit Rolle Dozent an
- Dozent
 - legt Laboratory (Lab) an
- Studierende
 - registrieren sich mit Name, Vorname, E-Mail-Adresse und Passwort
 - melden sich an Lab an
 - Dozent schaltet frei

eine Benutzerkennung
für alle Module



VIRTUELLER DESKTOP

- mehrere Desktops und ssh
- vorinstallierte Software
 - Editoren (vi, jEdit mit xml plugin)
 - Programminterpret/Compiler (python)
 - OCR-Programme (tesseract)
 - POS-Tagger (treetagger)
 - MySQL-Client (Zugriff auf DB-Server)
 - R und R-Studio
 - QGIS
 - u.a.



The screenshot displays a Linux desktop environment with a purple and blue geometric background. In the foreground, there are two windows:

- File Manager (Dolphin):** The window title is "corpus — Dolphin". The address bar shows the path: `> courses > KorpLing_MA > 03_WiSe_1718 > Bayer > corpus`. The main pane shows a directory listing:
 - folders: `in`, `orig`, `out`, `scripts`, `temp`
 - files: `cbayer_meta.csv`, `cbayer_tokens.csv`, `convert_meta.sh`, `convert.sh`, `liste.csv`, `metakat.csv`
- Terminal (Konsole):** The window title is "out : sh — Konsole". It shows the output of a `TreeTagger` command:

```
TreeTagger DE: sdz20170908.txt.csv
  reading parameters ...
  tagging ...
1000  finished.
  reading parameters ...
  tagging ...
1000  finished.
TreeTagger DE: sdz20170909.txt.csv
  reading parameters ...
  tagging ...
5000  finished.
  reading parameters ...
  tagging ...
5000
```

The system tray at the bottom right shows the time as 4:50 PM.



DATENREPOSITORIUM

- DBMS MySQL, Neo4j (geplant)
- Benutzeroberfläche phpmyadmin
- Datenbanken
 - Projekte (zentral)
 - Labors
 - Studierende
- Nachhaltigkeit
 - Qualitätsmanagement (Lehrende)
 - dauerhafte Speicherung
 - nachnutzbar
- kollaborative Datensammlung



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

IT-Gruppe Geisteswissenschaften
DH Kompetenz- und Datenzentrum



phpMyAdmin

Server: db.dmz.dhlab.fu » Datenbank: lab_korplingma » Tabelle: cbayer_tokens

Anzeigen Struktur SQL Suche Einfügen Exportieren Importieren Rechte Operationen Trigger

Zeige Datensätze 0 - 24 (306929 insgesamt, Die Abfrage dauerte 0.0003 Sekunden.)

```
SELECT * FROM `cbayer_tokens`
```

Messen [inline bearbeiten] [Bearbeiten] [SQL erklären] [PHP-Code erzeugen] [Aktualisieren]

1 > >> | Anzahl der Datensätze: 25 | Zeilen filtern: Diese Tabelle durchsuch Nach Schlüssel sortieren: keine

+ Optionen

ID	datei	zeitung	datum	artnr	linenr	poslinenr	sentntr	possentntr	token	wortart	lemma	chunk	chunknr
1	sdz20170623	sdz	20170623	1	1	1	1	1	Bestimmen	NN	Bestimmen	B-NC	1
2	sdz20170623	sdz	20170623	1	2	2	1	2	die	ART	die	B-NC	2
3	sdz20170623	sdz	20170623	1	3	3	1	3	Alten	NN	Alte	I-NC	2
4	sdz20170623	sdz	20170623	1	4	4	1	4	den	ART	die	B-NC	3
5	sdz20170623	sdz	20170623	1	5	5	1	5	Wahlsieger	NN	Wahlsieger	I-NC	3
6	sdz20170623	sdz	20170623	1	6	6	1	6	?	S.	?	0	4
7	sdz20170623	sdz	20170623	1	7	7	2	1	Die	ART	die	B-NC	5
8	sdz20170623	sdz	20170623	1	8	8	2	2	alteren	ADJA	alt	I-NC	5
9	sdz20170623	sdz	20170623	1	9	9	2	3	Mitbürger	NN	Mitbürger	I-NC	5
10	sdz20170623	sdz	20170623	1	10	10	2	4	werden	VAFIN	werden	B-VC	6
11	sdz20170623	sdz	20170623	1	11	11	2	5	in	APPR	in	B-PC	7
12	sdz20170623	sdz	20170623	1	12	12	2	6	diesem	PDAT	dies	I-PC	7
13	sdz20170623	sdz	20170623	1	13	13	2	7	Bundestagswahlkampf	NN	Bundestagswahlkampf	I-PC	7
14	sdz20170623	sdz	20170623	1	14	14	2	8	von	APPR	von	B-PC	8
15	sdz20170623	sdz	20170623	1	15	15	2	9	allen	PIAT	alle	I-PC	8
16	sdz20170623	sdz	20170623	1	16	16	2	10	Parteien	NN	Partei	I-PC	8
17	sdz20170623	sdz	20170623	1	17	17	2	11	intensiv	ADJD	intensiv	0	9
18	sdz20170623	sdz	20170623	1	18	18	2	12	umworben	VVPP	umwerben	B-VC	10
19	sdz20170623	sdz	20170623	1	19	19	2	13	,	S.	,	0	11
20	sdz20170623	sdz	20170623	1	20	20	2	14	zahlenmäßig	ADJD	zahlenmäßig	0	12
21	sdz20170623	sdz	20170623	1	21	21	2	15	stellen	VVFIN	stellen	B-VC	13
22	sdz20170623	sdz	20170623	1	22	22	2	16	sie	PPER	sie	B-NC	14
23	sdz20170623	sdz	20170623	1	23	23	2	17	bereits	ADV	bereits	0	15
24	sdz20170623	sdz	20170623	1	24	24	2	18	die	ART	die	B-NC	16
25	sdz20170623	sdz	20170623	1	25	25	2	19	Mehrheit	NN	Mehrheit	I-NC	16

Alle auswählen markierte:



DOKUMENTATION

- WordPress-Plattform
- veranstaltungsbegleitende Dokumentation
- Verknüpfung zu Datenbanken
- Visualisierung
- Publikationsmöglichkeit
- wiederverwendbare und nachnutzbare Module
- wachsendes Wissensrepositorium



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

IT-Gruppe Geisteswissenschaften
DH Kompetenz- und Datenzentrum



Lehre in den Digital Humanities

Ein Portal der IT-Gruppe Geisteswissenschaften der LMU

[Home](#) [Lehrveranstaltungen](#) ▾ [Vorlesungen](#) [Abschlussarbeiten](#) [Anmelden](#)

Lehrveranstaltungen

[Wintersemester 2019/20](#) | [Sommersemester 2019](#) | [Wintersemester 2018/19](#) | [Sommersemester 2018](#) |
[Wintersemester 2017/18](#) | [Sommersemester 2017](#) | [Wintersemester 2016/17](#) | [Sommersemester 2016](#) |
[Wintersemester 2015/16](#) | [DH - Summerschool 2015](#) | [Sommersemester 2015](#) |

Wintersemester 2019/20

Beschreibung, Abbildung und Visualisierung von Sprache und Text (WiSe 19/20) (WiSe 2019/20)

Einführungsvorlesung Romanische Sprachwissenschaft (WiSe 2019/20)

Thomas Krefeld, ADufter



EisenzeitDigital: Datenanalyse mit MySQL und QGIS (WiSe 19/20) (Wissenschaftliche Übung, WiSe 2019/20)

Caroline von Nicolai, Stephan Lücke





Handbücher

- Handbücher

- allgemeine Einführungen
- an Geisteswissenschaften orientiert
- DHVLab, Tabellenkalkulation, Datenbanken (SQL), Statistik (R)



Anwendungsfälle

- Anwendungsfälle

- konkrete Anwendungen zum vertieften Selbststudium
- „Datenbankabfragen mit R und SQL“
- „Vom gedruckten Buch zur digitalen Karte“
- „Digitales Edieren in der Praxis“



SQUIRREL

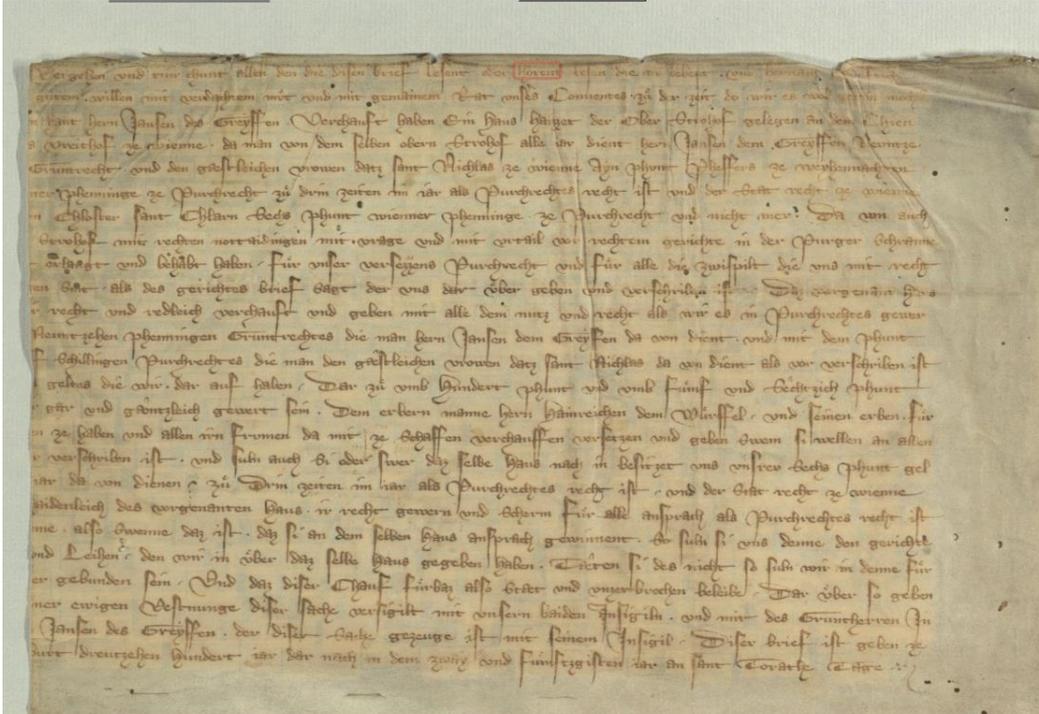


ANALYSECENTER



MAX

- Squirrel
 - Verknüpfung Bild-Text
 - Transkription und Annotation
 - digitale Edition
- Analysecenter
 - statistische Modellierung
 - Auswertung und Visualisierung
 - Daten der Kunstgeschichte
- MAX – Museum Analytics
 - Vergleichende Analyse
 - modular, dynamisch, interaktiv
 - museale Datenbestände



Kontext

vergehen und tun chunt allen den ,die disen
brief lesent oder hörent lesen ,die in lebent und
hernach chu+ontfich |

Publicatio

gutem willen mit verdachtem mut und mit
gemainem rat unse+rs Conuentes ,zu der zeit do
wir es wol getu+en mochte(n) |

[...]n hant hern Jansen des Greyffen verchauft

Person

haben ein haus haizzet der
ober Strohof gelegen an dem Chien |

Ort

A Protokoll Kontext Eschatokoll
B
C
D



DHVLab
+

- Übersicht
- Deskriptive Analysen
- Ähnlichkeitsanalysen <
- » 1-Alle-Vergleich
- » Alle-Alle-Vergleich
- Kombinationsanalyse
- Direkte Bildadressierung <
- Information

Auswahl

Künstler

Kunstwerk

Künstler

Marc, Franz

Lokale Gewichtung

Binäre Projektion

Suchwortdichte

Normalisierte Suchwortdichte

Globale Gewichtung

Keine Gewichtung

Inverse Dokumenthäufigkeit

Entropie

Ähnlichkeitsmaß

Kosinus-Ähnlichkeit

Jaccard-Ähnlichkeit

Mindestzahl gemeinsamer Schlagwörter

1 500

1 51 101 151 201 251 301 351 401 451 500

Springe zu Seite

Berechnen

Marc, Franz

1880 - 1916, 303 Kunstwerke

Tags: 61997 (gesamt), 4439 (verschieden)

Links: Artigo Getty

Gauguin, Paul

1848 - 1903, 186 Kunstwerke

Tags: 44487 (gesamt), 3816 (verschieden)

Links: Artigo Getty

Vor- und Nachname	Geburt	Werke	G.Tags	Ähnl.
Gauguin, Paul	1848	186	1833	44,54 %
Cézanne, Paul	1839	262	1860	43,21 %
Boccioni, Umberto	1882	84	1421	42,64 %
Corinth, Lovis	1858	201	1869	42,08 %
Trübner, Wilhelm	1851	133	1721	41,39 %
Vallotton, Félix	1865	164	1849	41,37 %
Kirchner, Ernst Ludwig	1880	78	1211	40,84 %
Gogh, Vincent van	1853	166	1706	40,77 %
Dillis, Georg von	1759	178	1607	40,70 %
Max, Gabriel	1840	167	1514	40,20 %

Zurück 1 2 3 4 5 ... 435 Nächste



- Dashboard
- Preprocessing
- Visualisierung

Ein Online-Tool zur vergleichenden Analyse musealer Datenbestände

● Alte Pinakothek ● Neue Pinakothek ● Pinakothek der Moderne



Museum laden

Museum	Bestand	
Joconde	590962	+
National Library of Denmark	421563	+
Riksantikvarieämbetet	204157	+
Rijksmuseum	189827	+
Rijksdienst voor het Cultureel Erfgoed	152616	+
Zeeuwse Bibliotheek	105411	+
National Library of France	103952	+
Domkirkeodden	75969	+
Malmö museer	74261	+
Gool en Vecht Historisch	64304	+

Filter

Museum importieren

Name des Museums

Datei auswählen



CLOUD



LIVE-CODE



GITLAB



Schnattern

- Datenaustausch mit virtuellem Desktop
- jupyter note-book: Programmierung
- Gitlab: versionierte Datenablage
- mattermost: Kommunikation (wieder geplant)



1. Organisatorischer Hintergrund
2. Konzept der Wissensvermittlung
3. Architektur des DHVLab
4. Komponenten des DHVLab
5. Einsatz und Erfahrungen
6. Zukunftsperspektiven



Einsatz in der akademischen Lehre

- seit Sommersemester 2016 (DH-Studienangebot an der LMU)
- Kunstgeschichte, Geschichte, Sprachwissenschaften, Archäologie
- Summerschools (dhmuc und téchnē)

Einsatz in der Forschung

- Forschungsprojekte: individuelle Labore
- Qualifizierungsarbeiten (Master, Promotion)

Demo-Lab

- Interesse? Testen Sie das DHVLab, indem Sie sich am Demo-Lab registrieren.

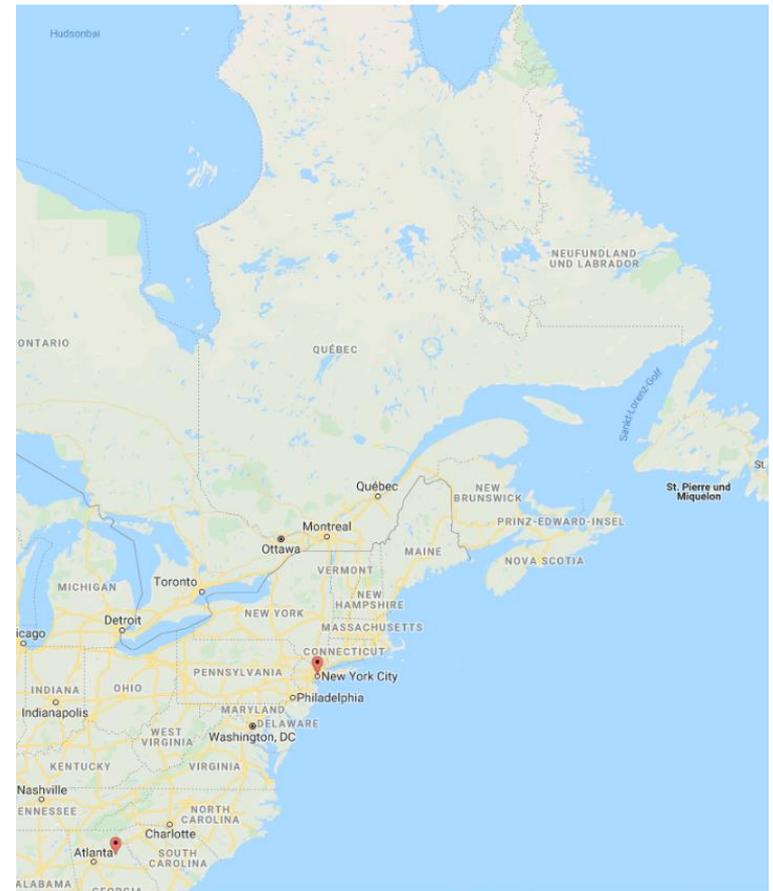
LMU

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

IT-Gruppe Geisteswissenschaften
DH Kompetenz- und Datenzentrum



DHVLab



872 User

65 Labs

46 Universitäten



Erfahrungen

- positives Feedback
 - einfache Labor- und Benutzerverwaltung
 - niedrige Einstiegshürden und Zeitersparnis durch einheitliche Umgebung und vorinstallierte Software
 - zeit- und ortsunabhängige Nutzung
 - kollaboratives Arbeiten im Team
 - Skalierbarkeit der Systemressourcen (VMs am LRZ)
 - Erweiterbarkeit der Module (Squirrel, GitLab; demnächst Neo4j)
 - Nachnutzung und Langzeitverfügbarkeit
 - Forschungsdatenmanagement (FAIR)



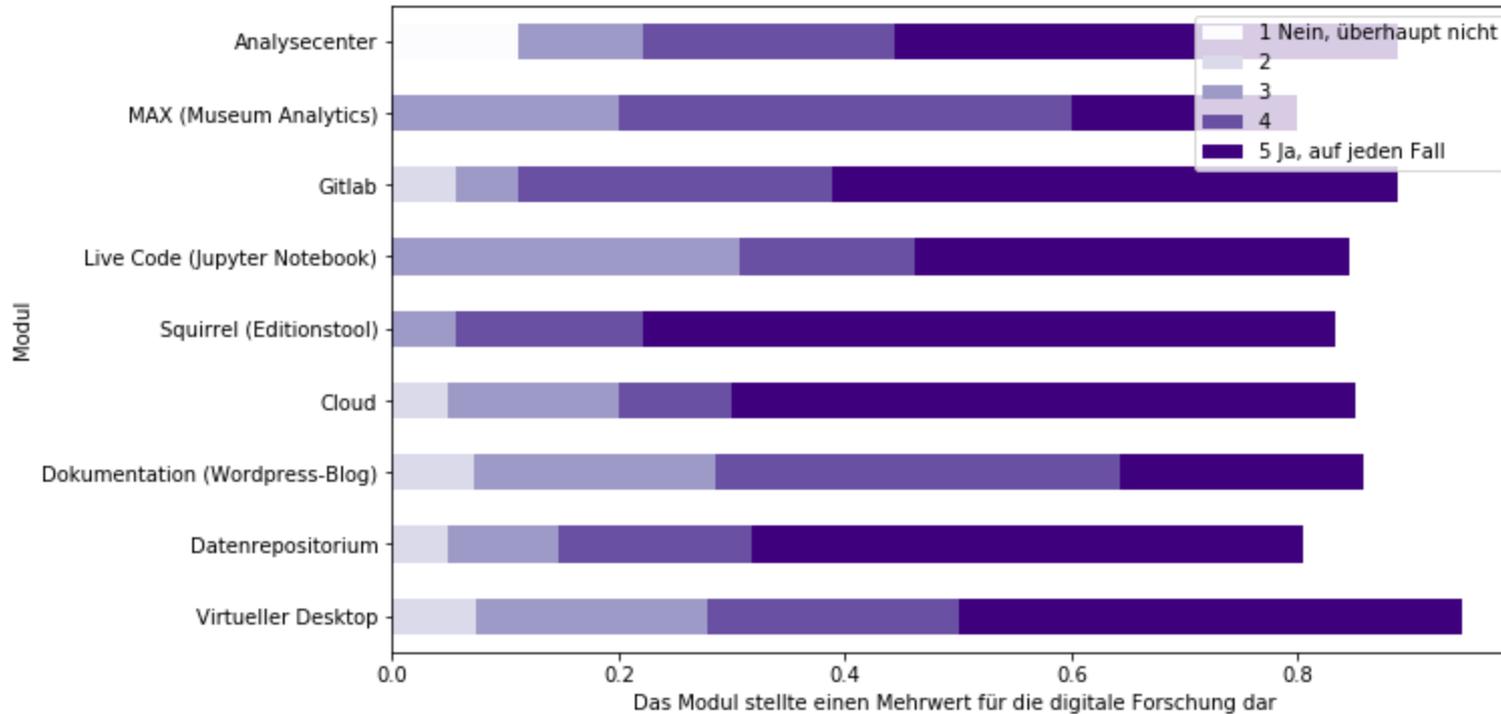
Erfahrungen

- Problembereiche
 - leistungsfähige und störungsfreie Netzanbindung
 - lizenzpflichtige Software
 - hauptspeicherintensive Software (z.B. Oxygen)
 - Anbindung an bestehende Authentifizierungsdienste (Shibboleth)
- Umfrage Herbst 2019
 - 587 User angeschrieben, 111 User teilgenommen
 - Rücklaufquote: 19%



Umfrage Herbst 2019

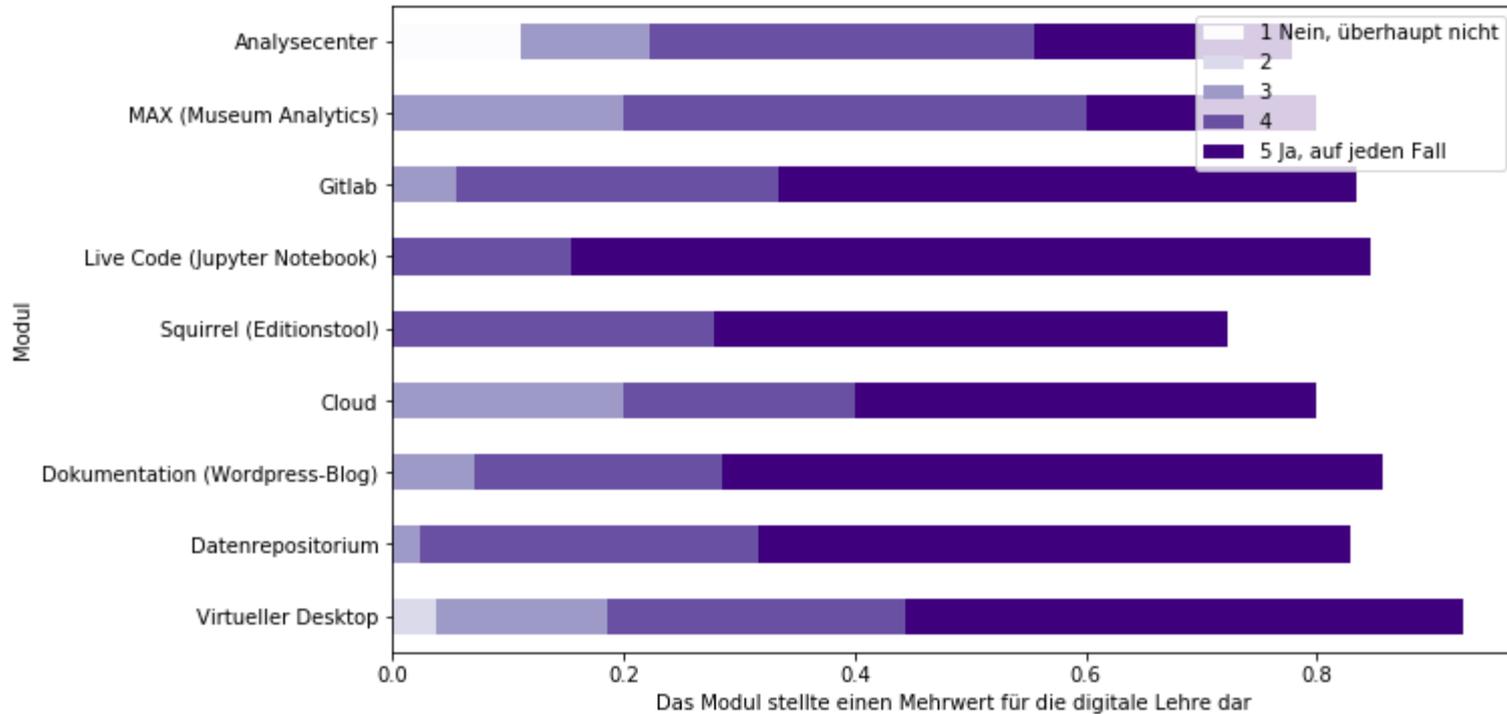
Mehrwert für digitale Forschung





Umfrage Herbst 2019

Mehrwert für digitale Lehre

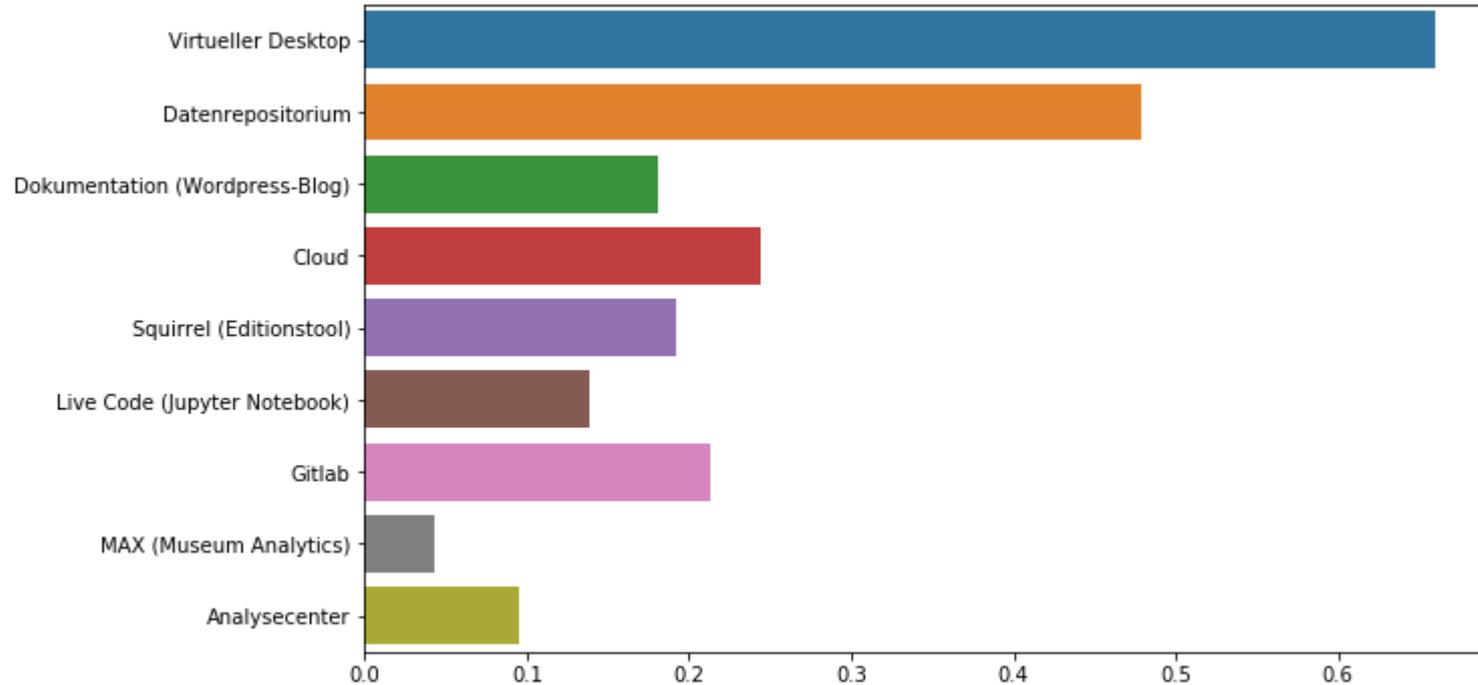


Das Modul stellte einen Mehrwert für die digitale Lehre dar



Umfrage Herbst 2019

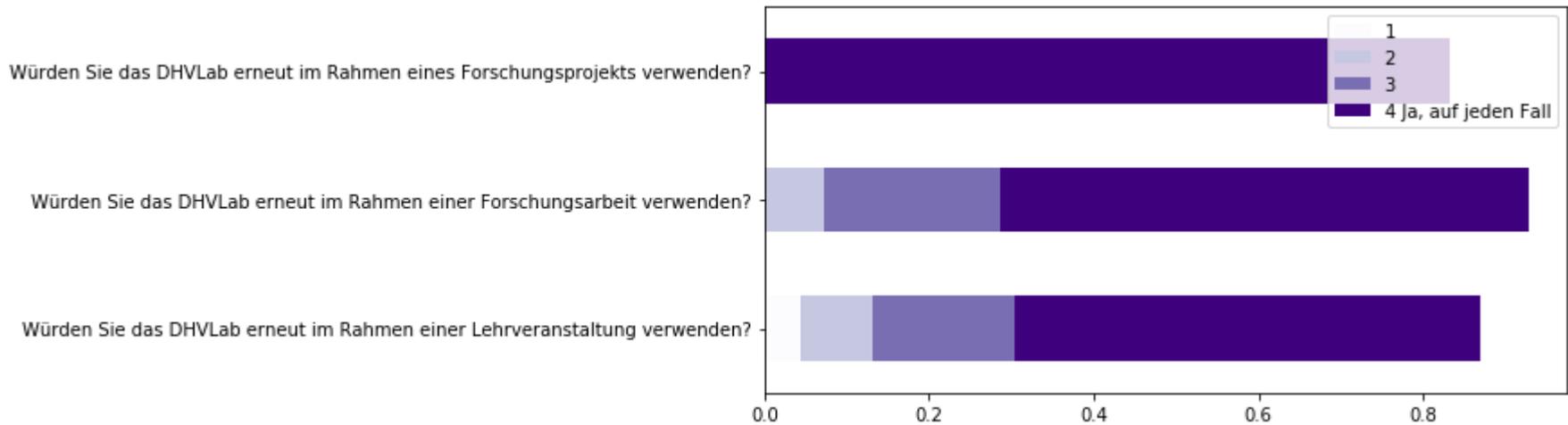
Verwendungshäufigkeit





Umfrage Herbst 2019

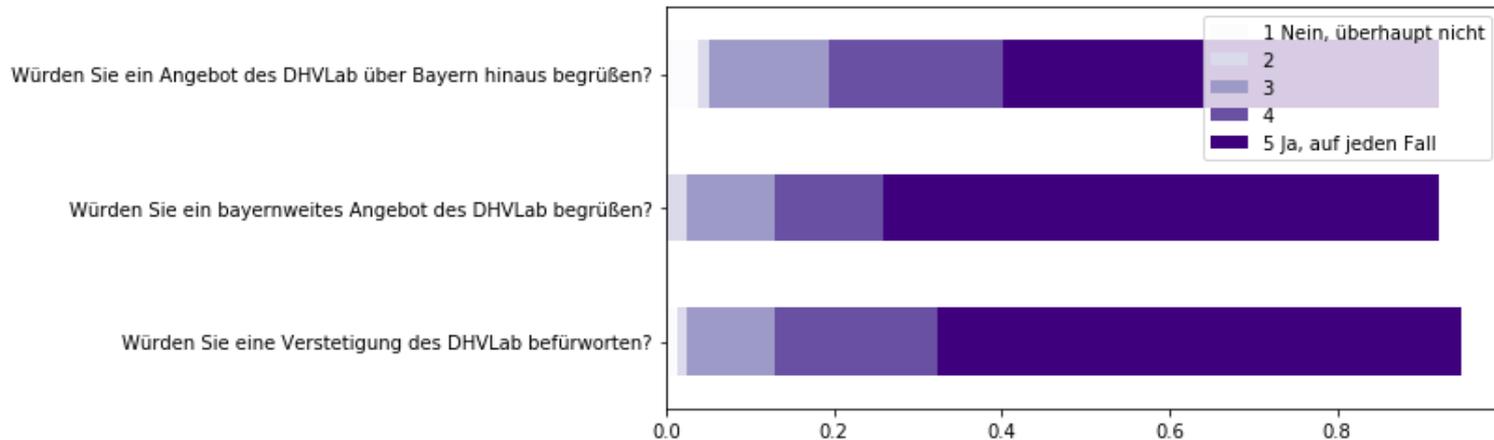
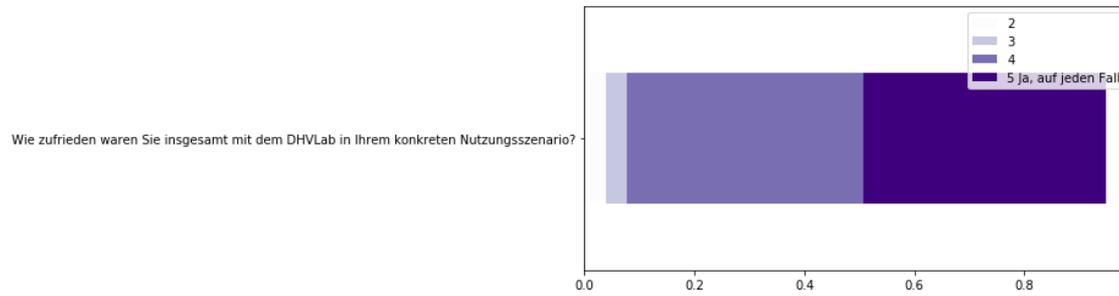
Wiederverwendung





Umfrage Herbst 2019

Zufriedenheit und Verstetigung





1. Organisatorischer Hintergrund
2. Konzept der Wissensvermittlung
3. Architektur des DHVLab
4. Komponenten des DHVLab
5. Einsatz und Erfahrungen
6. Zukunftsperspektiven



Zukunftsperspektiven

- Vom „DHVLab II“ ...
 - Modifizierung der Benutzerverwaltung (1 User – n Labs)
 - Integration des Systems „dh-lehre“ im Modul „Dokumentation“
 - Datenschutz (Zustimmung) und Lizenzmodell (CC-BY-SA 4.0)
- ... zum „DHVLab II plus“ (April 2019 bis März 2020)
 - Neue, verteilte und an Diensten orientierte Systemarchitektur
 - Vorschlag für Geschäftsmodelle für den dauerhaften Einsatz an anderen (bayerischen) Universitäten – und darüber hinaus
 - PROFiL: Schulungsangebot DHVLab auch für bayerische Universitäten
 - Einsatz auch in anderen Kontexten denkbar, z.B. an Schulen

