**Umfrage zu Nutzungsszenarien und Kompetenzen an den verschiedenen Standorten von** [**DHInfra.at**](http://dhinfra.at)

**Thema:**

[DHInfra.at](http://dhinfra.at) baut eine **Infrastruktur für digital gestützte Forschung in den österreichischen Geisteswissenschaften** auf. Sie füllt die Lücke zwischen Standardangeboten in den Kulturerbeinstitutionen (Digitalisierung), im Forschungsdatenmanagement (kuratierte und integrierte Repositorien vs. institutionelle Repositorien), bei Softwarelösungen (fachspezifische Open Source-Produkte), und den High Performance Computing-Angeboten für die Natur-, Technik- und Lebenswissenschaften bei der Verarbeitung großer Datenmengen mit maschinellem Lernen.

**Ziel:**

In Verbindung mit [CLARIAH-AT](https://clariah.at/de/) plant [DHInfra.at](http://dhinfra.at) **die Anschaffung von Computerhardware besonders für den Bereich der künstlichen Intelligenz**, die die von den individuellen Standorten nicht bedienten Bedarfe abdecken soll. Diese Umfrage hat zum Ziel, mögliche **Nutzungsszenarien sowie vorhandene Kompetenzen an Ihrem Standort** zu erheben.

**Zielgruppe:**

Die Umfrage richtet sich an

* Mitarbeitende bei den **Kooperationspartner:innen von** [**DHInfra.at**](http://dhinfra.at), die selbst forschen oder für Forschungsinfrastruktur und forschungsunterstützende Services zuständig sind,
* Mitarbeitende an österreichischen **Forschungs- und Kulturerbeeinrichtungen**, die selbst forschen oder für Forschungsinfrastruktur und forschungsunterstützende Services zuständig sind,
* Koordinator:innen, Geschäftsführer:innen und Leiter:innen von **DH-Zentren, DH-Services, DH-Netzwerken und DH-Konsortien** an österreichischen Forschungs- und Kulturerbeeinrichtungen,
* Koordinator:innen, Geschäftsführer:innen und Leiter:innen von **disziplinären und interdisziplinären Forschungszentren und fachlichen Netzwerken,** in denen auch DH-Projekte angesiedelt sind, an österreichischen Forschungs- und Kulturerbeeinrichtungen,
* Projektleiter:innen von **einzelnen DH-Projekten und größeren DH-Verbundprojekten** an österreichischen Forschungs- und Kulturerbeeinrichtungen.

**Scope:**

[DHInfra.at](http://dhinfra.at) entwickelt in den fünf Feldern **Data Capturing und Enhanced Images** **Sensing (I.)**, **Open Source Software (II.)**, **Datenmanagement und Repositorien (III.)**, **Infrastructure as a Service (I**

**V.)** und **Machine Learning** **(V.)** eine gemeinsame Infrastruktur für digitale Forschung in den Geisteswissenschaften.

**Teilnahme:**

Bitte nehmen Sie bis zum **31. Juli** an der Umfrage teil. Senden Sie Ihre Antworten und Ihre User Story an den\*die Vertreter\*in von [DHInfra.at](http://dhinfra.at) an Ihrer Einrichtung:

* Universität Graz, Zentrum für Informationsmodellierung (Lead), Institute of Mathematics and Scientific Computing, IT-Services: Walter Scholger; walter.scholger@uni-graz.at
* TU Wien, Institute of Visual Computing & Human-Centered Technology, Computer Vision Lab: Florian Kleber; kleber@cvl.tuwien.ac.at
* Paris-Lodron-Universität Salzburg: Karoline Döring; karolinedominika.doering@plus.ac.at
* Universität Innsbruck: Günter Mühlberger guenter.muehlberger@uibk.ac.at
* Universität der Weiterbildung Krems – Department für Kunst- und Kultur­wissenschaften: Anja Grebe; anja.grebe@donau-uni.ac.at
* Die Angewandte – Universität für angewandte Kunst: Abteilung IT & Digitalisierung; itd-pk@uni-ak.ac.at
* Universität Wien: Thomas Wallnig; thomas.wallnig@univie.ac.at
* Österreichische Nationalbibliothek: labs@onb.ac.at
* Österreichische Akademie der Wissenschaften – Austrian Center for Digital Humanities and Cultural Heritage: dh@plus.ac.at

Wenn Sie keiner dieser Einrichtungen angehören, senden Sie Ihre Antworten und Ihre User Story an: dh@plus.ac.at

**User Stories**

Was ist Ihre User Story? Beschreiben Sie uns in ca. 150 Wörtern Ihren Bedarf und die Herausforderungen dabei!

**Umfrage**

**I.**               **Data Capturing und Enhanced Images Sensing**

Die TU Wien hat mit XpeCam und MISHA zwei mobile Systeme zur multispektralen Photographie angeschafft, die an nicht-mobile Forschungsobjekte wie Kirchen, Museen, Grabungen o. ä. verlegt und einfach gehandhabt werden können. Die Universität Innsbruck wird voraussichtlich ab Anfang 2025 über ein robotergesteuertes System (Scan-Roboter) verfügen, das auf die Digitalisierung von losen Einzelblättern bis A4 optimiert ist. Es eignet sich besonders für das sorgfältige und (hoffentlich) weitgehend autonome Scannen von speziellem Papier (z. B. Meldezettel, Karteikarten von Museen, etc.).

* Gibt es Sammlungen an Ihrer Institution oder an Institutionen, mit denen Ihre Institution kooperiert (Museen, Archive…), die für das automatisierte Scannen loser Zettel in Betracht kommen?
* Wenn ja, welchen Umfang hätten diese Sammlungen?
* Können diese nach Innsbruck gebracht werden, oder muss der Roboter vor Ort eingesetzt werden?
* Welche rechtlichen Voraussetzungen und Rahmenbedingungen bestehen schon für die Digitalisierung dieser Sammlungen? Welche Vereinbarungen müssten neu getroffen werden?
* Gibt es Personen in Ihrer Institution, die bereits Kenntnisse und Kompetenzen für die Bedienung solcher oder ähnlicher Systeme besitzen?
* Welche Haftungsfragen stellen sich dabei?

**II.**             **Open Source Software**

[DHInfra.at](http://dhinfra.at) entwickelt Open Source-Lösungen für sechs Softwarepakete: Blazegraph, FEDORA Commons, Semantic MediaWiki, WissKI, Tools für NLP und Korpusanalyse und Annotations-Bibliothek

* Welche Software ist bereits an Ihrer Institution für DH-Anwendungen in Gebrauch?
* Welche konkreten Aufgaben erledigen Sie damit?
* Welche Aufgaben können damit bislang nicht oder nur unzureichend erledigt werden?
* Welche technischen Kenntnisse und Kompetenzen erfordert die Anwendung dieser Software?
* Gibt es Personen in Ihrer Institution, die diese Kenntnisse und Kompetenzen bereits besitzen?
* Welche konkreten Aufgaben wollen Sie in Zukunft computergestützt erledigen?
* Welche Softwarelösungen haben Sie dafür bereits erprobt und welche eignen sich nicht, um Ihre Aufgaben zu erledigen?
* Welche Probleme traten konkret auf?
* Handelte es sich dabei um kostenpflichtige oder Open Source Software?
* Soll kostenpflichtige Software durch eine Open Source-Lösung ersetzt werden? Wenn ja, welche Software soll ersetzt werden?
* Welche personellen und / oder finanziellen Ressourcen gibt es bei Ihnen für die langfristige Weiterentwicklung der ausgewählten Open Source Software?

**III.**            **Datenmanagement und Repositorien**

Für die gemeinsame Nutzung baut [DHInfra.at](http://dhinfra.at) die vorhandene Hardware-Infrastruktur im Bereich der integrierten Infrastruktur und integrierten Datenablage aus und verschränkt sie mit der von den Partnern bereits im Aufbau befindlichen generischen Infrastruktur an der Universität Graz und der ÖAW Wien.

* Benötigen Sie (nur) einen Ablage- und Speicherort, z. B. für die Archivierung von Digitalisaten, Publikationen und Forschungsdaten oder benötigen Sie (auch) eine Umgebung für rechen- und datenintensive DH-Anwendungen und Analysen? (siehe auch unten IV.)
* Haben Sie einen konkreten Bedarf für die langfristige Speicherung von Digitalisaten und/oder Datensammlungen?
* Über welchen Zeithorizont soll das Repositorium verfügbar sein?
* Um welche Datenmengen handelt es sich?
* Um welche Datentypen handelt es sich? (Metadaten? Dokumentation? Lizenzbedingungen?)
* Um welche Datenformate handelt es sich?
* Welche Art von Zugriff auf die Daten wünschen Sie sich? (von einfacher Auflistung der Metadaten bis zu komplexen projektspezifischen Webinterfaces für Suche und Anzeige; APIs zur Anbindung von Frontends; gefilterter Datensatz über eindeutige Identifikatoren zitierbar)
* Was wäre der Grund, dass Sie diese Daten nicht lokal z. B. an Ihrer Universität speichern können oder wollen?
* Welche technischen Anforderungen muss ein Repositorium erfüllen, das an Ihrer Institution aufgebaut wird? (Einbindung von Normdaten, Schnittstellen für Metadatenharvesting usw.)
* Welches Berechtigungskonzept ist für ein an Ihrer Institution aufzubauendes oder gehostetes System möglich? Welche Zugriffsrechte sind ausgeschlossen? (z. B. vollständig öffentlich und indiziert vs. nur registrierte Nutzende vs. privat)
* Welche technischen Kenntnisse und Kompetenzen erfordern die Einrichtung und die Betreuung eines Repositoriums?
* Gibt es Personen in Ihrer Institution, die diese Kenntnisse und Kompetenzen bereits besitzen?
* Welche personellen und / oder finanziellen Ressourcen gibt es in Ihrer Institution für die langfristige technische Betreuung eines Repositoriums?
* Welche personellen Ressourcen gibt es in Ihrer Institution für das Datenmanagement?
* Werden die Daten an Ihrer Einrichtung kuratiert? Wenn ja:
	+ Wer übernimmt in Ihrer Einrichtung die Pflege und Wartung der Daten?
	+ Benötigen Sie eine Versionierung des Datenbestandes?
	+ Sollen die einzelnen Versionen über dauerhafte Identifikatoren zitierbar sein?

**IV.**           **Infrastructure as a Service**

Um projektspezifischen Speicher- und Rechenbedarf abdecken zu können, schafft [DHInfra.at](http://dhinfra.at) einen Cluster mit 5 Serverknoten an. Der Cluster bietet eine Database as a Service Infrastruktur für a) relationale Datenbanksysteme, b) Key-Value Stores und c) Graph-Datenbanken sowie eine Infrastructure as a Service-Umgebung in Form von virtuellen Maschinen für rechenintensive DH-Anwendungen und Analysen.

* Benötigen Sie (nur) einen Ablage- und Speicherort, z. B. für die Archivierung von Publikationen und Forschungsdaten (siehe auch oben III.)  oder benötigen Sie (auch) eine Umgebung für rechenintensive DH-Anwendungen und Analysen?
* Welche Datenbanken werden in Ihrer Institution betrieben?
* Wer soll wie auf die Daten zugreifen können?
* Welche Datenbanktechnologie wird eingesetzt?
* Welches Datenmodell wird verwendet (z. B. CSV, relationale Daten, Graphdaten, Bilddaten etc.). Bitte alle Daten beachten, also auch Metadaten etc.
* Welches konkrete System wird für die Speicherung der Daten eingesetzt?
* Welche Rechnerkapazitäten sind dafür notwendig?
* Welcher Technologie-Stack ist für den Aufbau Ihres (Teil-)Services notwendig?
* Welche Sicherheitsanforderungen sind für Sie relevant (auch gegenüber der Partei, die den Service anbietet)?
* Welche Softwarearchitektur hat die bestehende Lösung? Bezogen auf diese Lösung:
	+ Existiert eine Unterteilung in Anwendungslogik und Persistenzschicht z.B. als Client-Server Infrastruktur?
	+ Wie greift die Software auf die Daten zu? Z. B. WebAPI, Datenbanktreiber (wie JDBC), als Teil eines Frameworks.
	+ Wie ist die Anwendungslogik implementiert?
* Welche technischen Kenntnisse und Kompetenzen erfordern die Einrichtung und der technische Betrieb dieser Datenbanken?
* Gibt es Personen in Ihrer Institution, die diese Kenntnisse und Kompetenzen bereits besitzen?
* Welche personellen und / oder finanziellen Ressourcen gibt es bei Ihnen für den langfristigen Betrieb der diesen Datenbanken zu Grunde liegenden technischen Systeme?

**V.**             **Machine Learning**

Machine Learning wird immer wichtiger in den DH: NLP, HTR, Computer Vision für Kunstgeschichte, Large Language Models, Speech2Text oder die Erkennung von Bewegtbildern sind bereits bekannte Anwendungsszenarien. Events zu AI verbreiten sich und Schulungen im Prompt Engineering sind gefragt. Publikationen und Konferenzbeiträge unter Einsatz von Verfahren, die auf Machine Learning aufbauen, sind nur schwer zu übersehen. [DHInfra.at](http://dhinfra.at) möchte den österreichischen DH die nötige Hardware zur Verfügung stellen, um die Chancen und Grenzen des erkenntnisgeleiteten Einsatzes von AI in der geisteswissenschaftlichen Forschung auszuloten.

* Welche Datenmengen sollen verarbeitet werden (in Anzahl Objekten, in GB/TB)?
* Welche Rechtsfragen sind mit dem Prozess verbunden?
* Wofür möchten Sie maschinenlernende Verfahren einsetzen?
* Welche konkreten Aufgaben wollen Sie damit erledigen?
* Welche Kompetenzen besitzen Sie im Team diese Aufgaben technisch zu lösen (Programmierung, Datenhandling)?
* Gibt es schon konkrete Technologien / Software Stacks, die Sie für diesen Zweck einsetzen wollen?
* Sind diese Aufgaben eher experimenteller Natur oder für den Produktivbetrieb vorgesehen?