

## **IT-Gruppe Geisteswissenschaften - DH Kompetenz- und Datenzentrum (Allgemeines)**

Die IT-Gruppe Geisteswissenschaften (ITG) versteht sich als Kompetenz- und Datenzentrum für die digitalen Geisteswissenschaften. Ihre Zuständigkeit erstreckt sich auf den gesamten Bereich der Geistes- und Kulturwissenschaften an der Ludwig-Maximilians-Universität München (LMU). Ihre Organisation beruht auf einem fakultäts- und fachübergreifenden Konzept. Zu ihren Aufgabenbereichen gehören a) IT-Infrastruktur, b) Forschung & Lehre digital und c) Forschungsdatenmanagement.

Die ITG plant, beschafft und betreibt die IT-Infrastruktur auf Server- und Clientseite, bietet Beratung und First-Level-Support an. Sie begleitet digitale Projekte in Forschung und Lehre in allen Phasen. Sie berät bei der Beantragung, erarbeitet ein Proof-of-Concept, plant, organisiert und führt die Projekte gemeinsam mit den Fachwissenschaften durch. In Zusammenarbeit mit der Universitätsbibliothek bindet die ITG digitale Projekte in ein umfassendes Forschungsdatenmanagement ein. Die ITG fühlt sich den Prinzipien von Open Access und FAIR verpflichtet. Sie verfolgt das Ziel, Forschungssoftware (einschließlich Daten, Ergebnisse, Code) mit einem Höchstmaß an Synergie, Verfügbarkeit, Transparenz, Überprüfbarkeit und Nachhaltigkeit zu versehen. Dabei ist die Softwareentwicklung grundlegend an der Idee "digital / online first" orientiert.

In allen digitalen Projekten der ITG (siehe <https://www.itg.lmu.de/projekte>) profitieren neue Projektteammitglieder von den Synergieeffekten, die sich aus der interdisziplinären Arbeitsumgebung an der ITG und dem Austausch mit Experten der Kooperationspartner der ITG ergeben.

Die ITG fördert grundsätzlich die Zusammenarbeit zwischen Geisteswissenschaften, Informatik, Statistik und Computerlinguistik sowie den neu aufkommenden data sciences. Ihr Ziel ist dabei, den interdisziplinären Dialog weiter auszubauen und damit für alle Beteiligten fruchtbare Synergieeffekte zu erzeugen.

Im "technē - Campus-Netzwerk für digitale Geistes- und Sozialwissenschaften" kooperiert die ITG mit der Friedrich-Alexander-Universität Erlangen (FAU) und der Universität Regensburg, um in den digitalen Geisteswissenschaften Studiengänge zu entwickeln und einzuführen.

Im Pilotprojekt "eHumanities - interdisziplinär: Modellvorhaben Forschungsdatenmanagement Bayern" (siehe <https://www.fdm-bayern.org>) arbeiten die ITG und ihre Partner, die Universitätsbibliotheken der FAU und der LMU sowie das ITG-Projekt VerbaAlpina (<https://www.verba-alpina.gwi.uni-muenchen.de>) und das LRZ-Projekt GeRDI (<https://www.gerdi-project.eu/>) derzeit an einem Konzept für das Forschungsdatenmanagement und dessen Umsetzung. Das Leistungsspektrum umfasst ein Organisationsmodell, ein Schulungsprogramm, Lösungen für die Erstellung von Datenmanagementplänen und die Verwendung von Metadatenmodellen (nach DataCite-Standards) sowie funktionale Erweiterungen der bibliothekarischen Recherche. Eine bedeutende Rolle in diesem Kontext spielt auch die Schaffung logischer und technischer Interoperabilität. In diesem Zusammenhang ist die ITG an der Weiterentwicklung des Konzepts der „Normdaten“ (kontrollierte Vokabulare, Ontologien und RDF-Strukturen) beteiligt.

Auf der Grundlage einer Kooperationsvereinbarung zwischen der Bayerischen Akademie der Wissenschaften, der Bayerischen Staatsbibliothek, des Deutschen Museums und der LMU ist die ITG Teil des Kompetenznetzwerks "dhmuc - Digital Humanities München" (siehe <https://dhmuc.hypotheses.org>). Dies soll in den digitalen Geisteswissenschaften den Wissenstransfer verbessern, Anwendungen digitaler Methoden verbreiten und eine Kultur der offenen Wissenschaft fördern (Open Access, Offene Lizenzen).

Die ITG ist ferner institutionelles Mitglied des LMU Open Science Center. Aktivitäten der ITG auf dem Gebiet „Open Science“ sind durch zahlreiche DH-Projekte ([www.itg.lmu.de/projects](http://www.itg.lmu.de/projects)) dokumentiert. Gute Beispiele dafür sind VerbaAlpina (<https://www.verba-alpina.gwi.uni-muenchen.de>: Geolinguistik, Forschungsdaten, Methodik), KiT (<http://www.kit.gwi.uni-muenchen.de>: Innovatives Publizieren im Kontext der Korpuslinguistik) oder DHVLab

(<https://dhvlab.gwi.uni-muenchen.de>: Digitale Forschungs- und Lehrinfrastruktur für die Geisteswissenschaften).

## **Umgang mit den Forschungsdaten**

Die ITG stellt eine hochmoderne IT-Infrastruktur für die permanente und nachhaltige Datenablage bereit. Eine zusätzliche Datensicherung auf die Backupserver des LRZ erfolgt mit Hilfe des Tivoli Storage Managers. Redundante Server-Cluster, die gewählte Systemarchitektur (Datenbank- und Webtechnologie in Server-Client-Architektur), Software (MySQL, WordPress, PHP), Kodierung (Unicode/UTF-8) und Datenstruktur stellen von Beginn an sicher, dass die Daten dauerhaft verfügbar sind. Die fachübergreifende Ausrichtung der ITG ermöglicht konzeptionell die interdisziplinäre Wiederverwendbarkeit der Daten.

Im Rahmen von Projekten erhobene und erzeugte Daten sowie entwickelte Software werden anderen Forschern unter einer Creative-Commons-Lizenz (CC-BY-SA) über eine datenbankgestützte Web-Oberfläche mit Such-, Analyse- und Visualisierungsfunktionen für die zukünftige Wiederverwendung zur Verfügung gestellt. Die Daten werden in plattformunabhängigen, transparenten und leicht konvertierbaren Formaten (z.B. TXT, CSV, SQL, JSON, LaTeX, XML) gespeichert.

Das Forschungsdatenmanagement der LMU, das in einem von ITG und Universitätsbibliothek getragenen DH-Datenzentrum angesiedelt ist, sorgt dafür, dass die Forschungsdaten gemäß den FAIR-Richtlinien für wissenschaftliches Datenmanagement auffindbar, zugänglich, zitierbar, interoperabel und wiederverwendbar sind. Maximale Transparenz und nachhaltige Referenzierung mit Hilfe von persistenten Identifikatoren (z.B. DOIs) wird durch die Veröffentlichung der Projektdokumentation und anderer statischer Daten auf dem Forschungsdatenrepositorium erreicht (Open Data LMU: <https://data.ub.uni-muenchen.de>; Open Access LMU: <https://epub.ub.uni-muenchen.de>). Zugleich wird damit die Sichtbarkeit des Projekts erhöht.

## **Software-Architektur und Arbeitsabläufe**

Die Konzeption und Entwicklung der Forschungssoftware erfolgt in enger Zusammenarbeit zwischen den Fachwissenschaften und der ITG.

Technisch gesehen wird die Software des Projekts mit den redundant ausgelegten Datei-, Datenbank- und Webserver-Clustern der ITG und den Backup-Services des LRZ realisiert. Zur Entwicklungsumgebung gehören das Datenbankmanagementsystem MySQL, der Webserver APACHE, eine Reihe von Programmiersprachen (z.B. PHP, AJAX und verschiedene Frameworks für die Softwareentwicklung) und das Content-Management-System WordPress mit gängigen, allgemein verfügbaren und auch selbst erstellten Plugins und zu entwickelnden projektspezifischen Erweiterungen. Der modulare Ansatz ermöglicht ein Höchstmaß an Synergieeffekten zwischen verschiedenen Projekten. Der Einsatz von Web-Technologien gewährleistet eine plattformunabhängige Verfügbarkeit über das gesamte Internet zu jeder Zeit und an jedem Ort, ferner die Perspektive einer systematischen Vernetzung des Datenmaterials mit kompatiblen Daten anderer Forschungsprojekte. Sämtliche Eigenentwicklungen sind Open-Source-Software.

Methodisch werden alle Forschungsdaten nach Möglichkeit mit dem relationalen Datenmodell zusammen mit Annotations- oder Beschreibungsstandards wie XML modelliert. Basierend auf dieser relationalen Datenbank als Backend bietet ein benutzerfreundliches Webinterface als Frontend alle erforderlichen Funktionen: Dateneingabe und -analyse, Präsentation projektspezifischer Dokumente und Artikel, Suche und Filtern nach bestimmten Kriterien sowie Visualisierung von Such- oder Analyseergebnissen. Damit kann das Webinterface sowohl als internes Arbeitsmittel als auch als Publikationsplattform genutzt werden.

Im Frontend ist zunächst mit Hilfe von WordPress-Plugins wie z.B. WP-advanced-custom-fields eine projektspezifische Eingabemaske zu programmieren. Sie stellt sicher, dass die Eingabe der Daten durch die Projektmitarbeiter möglichst einheitlich strukturiert geschehen kann. Außerdem hilft sie, mögliche Fehlerquellen bei der Eingabe zu vermeiden. Die eingegebenen Daten (u.a. Klassifizierungen, Lemmata, Glosseme, Lexikonartikel, Literatur) werden sodann strukturiert und annotiert in der darunterliegenden Datenbank abgelegt. Passend zu dieser Datenbankstruktur sind alle Ausgabefunktionen für die Webschnittstelle zu programmieren. Dazu gehören die Funktionen für die crossmediale Präsentation der Daten (sowohl online als auch in einem Printformat wie z.B. LaTeX/PDF) sowie sämtliche Funktionen für die Suche im gesamten Datenbestand einschließlich der Visualisierung der Suchergebnisse. Ferner können bei Bedarf statistische Auswertungen mit Hilfe von R und shiny in die Benutzeroberfläche integriert werden.

Durch die in WordPress implementierten Funktionen für die Verwaltung von Benutzern, Gruppen und Rechten ist es in jeder Projektphase möglich, den Zugriff auf die Daten zu regulieren. Darüber hinaus bietet WordPress social-software-Funktionen (Kommentare), die auch eine interaktive Gestaltung der Benutzeroberfläche ermöglichen. Dies kann dazu genutzt werden, den wissenschaftlichen Diskurs zu fördern. Schließlich lässt sich das Layout der Webschnittstelle nahezu beliebig anpassen und gestalten.

Zusätzlich wird die Datenbank für den internen Gebrauch zur Erfassung, Annotation, Analyse und Korrektur über weitere, generische Schnittstellen wie etwa die Webschnittstelle PhpMyAdmin (<https://pma.gwi.uni-muenchen.de:8888/>) erreichbar sein.

Erfahrungsgemäß besteht während der kompletten Projektlaufzeit ständiger Bedarf für funktionale Anpassungen und Erweiterungen, die Modifizierungen auch des Datenmodells erfordern. Solche späteren Änderungen können durchaus komplex sein, stellen aus Sicht des Datenmodells jedoch kein substantielles Problem dar und sind durch das DH-Projektteam auszuführen. Die ITG hat dieses Basiskonzept in den vergangenen Jahren im Rahmen mehrerer drittmittelgeförderter Projekte erfolgreich umgesetzt. Hervorzuheben ist in diesem Zusammenhang das DFG-Langzeitprojekt VerbaAlpina (<https://www.verbaalpina.gwi.uni-muenchen.de/>), das konzeptionell und technisch auf einer entsprechenden Architektur fußt. Im Rahmen dieses und anderer Projekte (siehe <http://www.itg.lmu.de/projekte>) wurden zahlreiche Erfahrungen gesammelt. Dabei konnten Probleme und Herausforderungen, die sich aus der konsequenten Umsetzung digitaler Konzepte und Methoden ergaben, gelöst werden. Die ITG verfügt demnach über tragfähige Konzepte und Lösungen im Hinblick auf zentrale Fragen des digitalen Projekt- und Datenmanagements wie z.B. die persistente Zitierbarkeit von digitalen Online-Ressourcen und die Sicherstellung der langfristigen Verfügbarkeit von digitalen Forschungsdaten.

Im Projektverlauf werden die erhobenen Forschungsdaten in bestimmten Intervallen versioniert, mit formalen und inhaltlichen Metadaten angereichert, je nach Forschungsinteresse auf verschiedenen Ebenen mit persistenten Identifikatoren versehen und im Forschungsdatenrepositorium Open Data LMU in verschiedenen Formaten (SQL, CSV, XML) abgelegt. Auf diese Weise wird der Datenbestand (Daten, Ergebnisse, Code) im Sinne der nachhaltigen Verfügbarkeit, Auffindbarkeit, Zitierbarkeit und Nachnutzbarkeit für die Öffentlichkeit erschlossen.

Die IT-Infrastruktur und die Dienste der ITG stehen für das beantragte Projekt für dessen volle Laufzeit unentgeltlich zur Verfügung (Eigenleistung der LMU). Die Projektdaten werden eingebettet in den Kontext aller relational strukturierter Daten, die bislang in DH-Projekten unterschiedlichster Disziplinen gesammelt wurden. Damit besteht zumindest theoretisch die Perspektive, vollständig oder selektiv Verbindungen zu Daten herzustellen, die ursprünglich im Rahmen anderer Forschungsinteressen zusammengetragen worden sind oder auch in Zukunft gesammelt werden. Auch hier ist damit zu rechnen, dass sich weitere Synergieeffekte ergeben und neue Perspektiven zu neuen Fragestellungen führen werden.