

► Mapping the Unknown

Datenintegrationstool für Geisteswissenschaften

Einreichung im Rahmen der Ausschreibung: HRSM-Projekt Digitalisierung und Informationsaufbereitung für die Digital Humanities (DI4DH)

Univ.-Prof. Dr. Erich Kistler

PD Dr. Christian Heitz

Dr. Martin Auer

Dr. Birgit Öhlinger

► Institut für Archäologien

Leopold-Franzens-Universität Innsbruck

Langer Weg 11

6020 Innsbruck

Dr. Gerald Hiebel

► Institut für Grundlagen der Technischen Wissenschaften, AB Vermessung und Geoinformation

Leopold-Franzens-Universität Innsbruck

Technikerstraße 13

6020 Innsbruck

Dr. Eva Zangerle

► Institut für Informatik

Leopold-Franzens-Universität Innsbruck

Technikerstraße 21a

6020 Innsbruck

1. Projektziel

Ziel ist es, mittels einer archäologischen test case study einen workflow zum infrastrukturellen Aufbau von geisteswissenschaftlichen semantisch organisierten Datenbanken zu entwickeln, die Datenbestände für spezifische Fragestellungen aus unterschiedlichen Datenquellen integrieren. Semantisch bedeutet in diesem Fall, dass die Daten nach Standards des semantic web aufbereitet werden. Zudem wird das CIDOC CRM, ein ISO zertifiziertes ontologisches Datenmodell zur Kulturgutdokumentation, zur Datenrepräsentation verwendet. Dieses Modell wurde für die Datenintegration geschaffen und wird international eingesetzt, was es ermöglicht, die geschaffenen Datenbestände mit den Daten anderer Institutionen zu verlinken und damit den Datenpool für die Forschungsfragen zusätzlich auszudehnen. Ausgehend von archäologischen Fragestellungen wird ein Datenintegrationstool aufgebaut und auf seine Anwendbarkeit geprüft, das grundsätzlich für alle Geisteswissenschaften anwendbar sein wird. Das Projekt leistet somit nicht nur die unverzichtbare Basisarbeit für eine ausbaufähige komparatistische Archäologie-Datenbank, sondern legt zudem den Grundstein für den weiteren Einsatz im Rahmen der Digital Humanities.

Den Ausgangspunkt hierfür bilden drei langjährige Grabungsprojekte, die am Institut für Archäologien der Universität Innsbruck angesiedelt sind. Bislang werden die über die Grabung generierten Datenmengen (Befunde, Funde) jeweils unterschiedlich verwaltet (ArchaeoData, Filemaker, Excel, Access u.a.) und stellen isolierte Einheiten dar. Erfüllen sie zwar den jeweiligen Zweck der Einzelprojekte, fehlt dennoch die Möglichkeit, die Informationen auf einer übergeordneten Ebene zusammenzuführen und für innovative vergleichende orts-, zeit-, und kulturunabhängige Fragestellungen nutzbar zu machen um den antiken Menschen als Akteur ins Zentrum zu rücken. Durch einen Meta-Vergleich von ausgewählten Fundkontexten werden bislang unbekannte Konsummuster erkennbar, die ohne die Möglichkeit der Verarbeitung und Visualisierung großer Datenmengen unsichtbar bleiben. Erst durch die komparatistische Zusammenführung dieser Daten eröffnen sich völlig neue Perspektiven auf das archäologische Material und letztlich das Fach.

Das Projekt erarbeitet einen workflow für eine semantisch integrierte archäologische Datenbank anhand eines Beispieldatensatzes. Herangezogen werden hierfür die beiden süditalischen Siedlungen am Monte Iato (Sizilien) bzw. in Ascoli Satriano (Apulien) sowie die römische Stadt Aguntum (Osttirol). Alle drei Projekte verfügen über eine lange Forschungsgeschichte und entsprechend umfangreiches Fundmaterial. Die case study wird sich auf jeweils drei klar definierte architektonische sowie materielle Ensembles konzentrieren, die als Sample für die Entwicklung eines entsprechenden workflows dienen.

2. Workflow

Um zu den genannten Zielen zu gelangen bedarf es eines dreistufigen Verfahrens. Nach der Auswahl eines repräsentativen archäologischen Samplesets wird zu allererst ein übergeordneter Thesaurus für das enthaltene Fundmaterial erstellt um die Daten vergleichbar zu machen. Diese Arbeit wird das Team der AntragstellerInnen übernehmen. In einem zweiten Schritt werden die Daten koordiniert durch Dr. Gerald Hiebel mittels eines Werkvertrages aufbereitet. Über ein sogenanntes semantisches mapping werden die Daten in ein einheitliches Datenmodell mit dem zuvor entwickelten übergeordneten Thesaurus konvertiert und in einem Triple Store (einer semantischen Datenbank) gespeichert. Ziel des dritten Schrittes ist es, die im Triple Store gespeicherten semantisch vernetzten Daten für ForscherInnen und Öffentlichkeit in einem Webinterface zur Verfügung zu stellen, was es schließlich jedem Nutzer und jeder Nutzerin ermöglichen soll, eigene Abfragen zu erstellen. Für einen Proof of concept wird eine gezielte Auswahl an Abfragen möglich

sein. Zur Umsetzung des dritten Schrittes wird ein weiterer Werkvertrag vergeben, in dessen Rahmen in Zusammenarbeit mit Dr. Eva Zangerle zuerst tools evaluiert werden, die ein entsprechendes Webinterface zu Triple Stores zur Verfügung stellen um anschließend das bestgeeignete tool für die Umsetzung zu adaptieren und zu implementieren.

3. Nachhaltigkeit und Ausblick

Mit der Erstellung eines workflows zur semantischen Datenintegration wird einerseits die Basis für die Nutzung durch andere geisteswissenschaftliche Forschungen geschaffen, indem lediglich der Thesaurus der jeweiligen Forschungsfrage angepasst werden muss. Andererseits können zukünftig nicht nur sämtliche archäologische Forschungsprojekte der Universität Innsbruck, sondern auch externer Partner in eine entsprechende Datenbank integriert und semantisch organisiert werden. Die Evaluierung der Möglichkeiten zur Datenpräsentation und benutzerfreundlichen Abfrage ist notwendig um die Auswahl der modellierten Daten falls nötig anzupassen. Zudem wird mit der online Präsentation der Daten die Forderung der scientific community und Fördergeber nach der öffentlichen Zugänglichkeit der Forschungsdaten erfüllt. Diese web-basierte Struktur erlaubt darüber hinaus den Wissenstransfer in die breite Öffentlichkeit, der ebenfalls eine immer größere Bedeutung gewinnt. Mit der Umsetzung der hier geplanten Datenstruktur wird jedoch nicht ein reines Repositoryum erstellt, sondern ein Datenintegrationstool geschaffen, das es den NutzerInnen ermöglicht eigene Forschungsfragen an das Material zu stellen.