Online-Datenbank zu antiken Literaturgattungen

Ursprünglicher Text

Am Ende des Projektes ist Zeit für abschließende Anpassungen im Rahmen der Veröffentlichung der Datenbank sowie Unterstützung bei der Erstellung der Printversion notwendig, damit diese aus der Datenbank exportiert werden kann.

Textbausteine aus dem Gespräch vom 5.2.24

Als digitales Ergebnis des Projekts entsteht eine virtuelle Forschungsumgebung, in der die Texte und die zwischen ihnen identifizierten Beziehungen vom Projektteam semantisch tiefenerschlossen und die Ergebnisse für Forschende publiziert und langfristig vorgehalten werden. Die virtuelle Forschungsumgebung basiert auf der DFG-geförderten open source Software WissKI[[1]](#footnote-1), die für die semantische Erschließung und Publikation kulturellen Erbes nach den FAIR Prinzipien entwickelt wurde und die am FAU CDI bereitgestellt und supported wird. WissKI unterstützt den kompletten Forschungsdatenzyklus und gewährleistet es, innerhalb der virtuellen Umgebung ein Datenmodell zu entwickeln, dieses Datenmodell mit einem benutzerfreundlichen Erfassungsinterface zu verbinden und letztlich semantische Daten zu erzeugen, die in einem Triplestore bzw. einer Graphdatenbank abgespeichert werden. Mit zahlreichen Import- und Exportschnittstellen, der semantischen Datenanreicherung und der Möglichkeit der Normdatenintegration (GND, Geonames, Getty Thesauri usw.) bildet WissKI ein ideales System für Linked Open Data, das anschlussfähig an offene Webressourcen ist. Neben der Datenmodellierung und Erfassung erfolgt über WissKI als webbasiertem System auch die Publikation der Daten. Als webbasiertes System erlaubt WissKI standort- und zeitunabhängiges sowie kollaboratives Arbeiten.

Die Anpassung von WissKI auf die projektspezifischen Anforderungen umfassen:

1. Datenmodellentwicklung

In einem ersten Schritt gilt es, ein Datenmodell zugeschnitten auf den Gegenstandsbereich und die Bedürfnisse des Projekts zu entwickeln. Dafür wird in Zusammenarbeit zwischen den Projektmitarbeitenden und dem CDI eine CIDOC CRM basierte Anwendungsontologie entwickelt. Das CIDOC CRM als ISO-Standard[[2]](#footnote-2) wird hier als sinnvoll erachtet, da es sich um Daten kulturellen Erbes handelt, die es in ihrer Komplexität mit all ihren Kontexten ereignisbasiert in einem Wissensgraphen zu modellieren gilt. Zentral sind die Texte und ihre Inhalte sowie ihre Entstehungsangaben wie Autor, Zeit und räumliche Verortung, aber auch Sekundärliteratur und wissenschaftliche Kommentare. Das Datenmodell wird in einem iterativen Verfahren entwickelt, weshalb eine Betreuung über die gesamte Laufzeit nötig ist.

→ Aufwand Datenmodellierung

* Basismodell 3 Monate E13 Stelle
* Betreuung über die restliche Laufzeit von 33 Monaten mit 20 % → 6,6 Monate
* Summe: **ca. 10 Monate E13 Stelle**

1. Programmierung eines Editors zur Annotation der Kleingattungen

Neben den Kontextinformationen der Quellentexte steht ihre Struktur im Fokus, die es präzise zu erschließen und auszuwerten gilt, um Gemeinsamkeiten und Unterschiede zwischen den verschiedenen Texten gezielt herauszuarbeiten und Rückschlüsse auf Gattungskriterien sowie mögliche Entwicklungen von Gattungsmerkmalen ziehen zu können. Dies soll in Form von Annotationen innerhalb eines Editors erfolgen. WissKI basiert auf dem Content Management System Drupal und kann eine Vielzahl bestehender Module der Drupal Community nutzen. Es gibt Texteditoren, die in WissKI integriert werden können und mit denen Fließtext erfasst werden kann. Was jedoch die Anforderungen an die spezifische Annotation und Kennzeichnung der Texte hinsichtlich ihrer strukturellen Bestandteile anbelangt, in denen sich letztlich die Kleingattungen niederschlagen, ist zusätzlicher Programmieraufwand nötig. Dabei ist auch zu beachten, dass die Annotationen von Textabschnitten auch bei Veränderungen im Text ihre Gültigkeit bewahren müssen.

→ **6 Monate E13 Stelle**

1. Webdesign- und Webentwicklungsarbeiten für die Aufbereitung und Publikation der Daten

Neben projektspezifischen Annotationsmöglichkeiten der Texte bildet die Darstellung von miteinander in Beziehung stehenden Inhalten eine weitere im Projekt zu entwickelnde Komponente. Hierbei gilt es einerseits den Nutzenden einen synoptischen Blick auf die Inhalte und damit einen unmittelbaren Vergleich zu ermöglichen. Darüber hinaus sollen die als Kleingattungen gekennzeichneten Teilinhalte über eine genau zugeschnittene Suche übergreifend recherchierbar gemacht werden. Hierfür sind zum einen Überlegungen auf Datenmodellebene nötig, zum anderen werden diese Ansichten auf der Ebene von Webdesign und Webentwicklungsarbeiten zu lösen sein.

→ Personenmonate: **mind. 6 Monate E13 Stelle**

1. Langzeitverfügbarkeit der virtuellen Forschungsumgebung und der Forschungsprimärdaten

Hosting, Maintenance und Archivierung der virtuellen Forschungsumgebung und der erschlossenen Daten erfolgen langfristig über die FAU WissKI Cloud des FAU CDI und werden über die FAU finanziert.[[3]](#footnote-3) Durch diese bestehende Infrastruktur kann die langfristige Verfügbarkeit der virtuellen Forschungsumgebung gewährleistet werden. Dadurch, dass es sich bei den entstehenden Forschungsdaten vornehmlich um Text und weniger um Mediendateien handelt, ist zu erwarten, dass das Volumen von 1 GB nicht überschritten wird und an dieser Stelle keine weiteren Kosten entstehen.

Die Forschungsdaten können darüber hinaus aus dem Triplestore in gängigen Formaten für RDF-basierte Daten wie N-Quads, Turtle oder XML-JSON exportiert und beispielsweise als Datenset separiert publiziert und zur Verfügung gestellt werden.

→ wird kostenlos vom CDI angeboten

1. Word-Export der Einträge aus der Forschungsumgebung für die Printpublikation

Zur Unterstützung bei der Erstellung der Printpublikation muss ein Modul programmiert werden, das in der Lage ist die jeweiligen Einträge in der Forschungsumgebung in ein Word-Template zu exportieren, das der geplanten Darstellung der Texte im Print möglichst nahe kommt und beim Erstellen des Manuskripts unterstützt.

Dafür gilt es zunächst bestehende Drupal-Module zu prüfen und bestehenden Code auf die Spezifika des Projekts unter Berücksichtigung der WissKI-spezifischen Darstellungsformen der semantisch angereicherten Daten weiterzuentwickeln.

* Insg. 3 Personenmonate E13 Stelle

1. Vgl. <https://wiss-ki.eu/> [↑](#footnote-ref-1)
2. Vgl. https://www.cidoc-crm.org/ [↑](#footnote-ref-2)
3. Vgl. Leistungsverzeichnis CDI WissKI <https://www.cdi.fau.de/kosten-und-leistungsverzeichnis/#collapse_0> [↑](#footnote-ref-3)