

# ZENTRUM FÜR ARCHÄOLOGISCHE FELDFORSCHUNGSDATEN: DAS PROJEKT ARCHEOINF

## VERLORENES WISSEN? FORSCHUNGSDATEN SICHERN ALS ZUKUNFTSAUFGABE VON BIBLIOTHEKEN

Die langfristige Bereitstellung von Forschungsdaten, insbesondere ihre nachhaltige Sicherung und Zugänglichkeit ist in vielen, wenn nicht allen Fächern noch ein ungelöstes Problem. Auf die Erhebung der Daten erfolgen ihre Auswertung und die Publikation der Forschungsergebnisse, danach bleiben die Daten oftmals ohne Anbindung an die Forschungsergebnisse liegen. Zugang haben allenfalls noch die Wissenschaftler, die sie erhoben haben, und dies auch nur, solange die Datenformate noch gelesen werden können.

Daraus folgt im schlechtesten aber häufigen Fall, dass die wissenschaftlichen Erkenntnisse nicht von dritter Seite nachvollzogen oder überprüft werden können und die Daten selbst keinen weiteren Vorhaben zur Verfügung stehen. Die Frage ist, welche Informationsinfrastrukturen geschaffen werden müssen, um Forschungsdaten für die Wissenschaft zu archivieren und abrufbar zu machen – und wer sich dieser Aufgabe stellen kann.

Die DFG fördert seit einiger Zeit Projekte, in denen Wissenschaftler mit Bibliotheken, Archiven oder Rechenzentren kooperieren, um beispielhaft Wege zu finden, wie ein nachhaltiges Zugänglichmachen von Forschungsdaten künftig gelingen kann. Bereits in einem Positionspapier von 2006 kündigte die DFG an, die Weiterentwicklung der Strukturen zur Bereitstellung von Forschungsprimärdaten als einen Förderschwerpunkt voranzutreiben.<sup>1</sup> Allein im Mai 2011 informierte die DFG über 27 neu bewilligte Projekte, die im Förderprogramm „Informationsinfrastrukturen für Forschungsdaten“ an den Start gingen, um disziplinspezifische Datenrepositorien zu entwickeln oder bereits vorhandene Dienste zu optimieren.<sup>2</sup> Dass die Problematik nicht nur STM-Fächer betrifft, zeigt die Auswahl der geförderten Projekte: Hier sind auch Wissenschaftsdisziplinen wie die

Archäologie, die Germanistik, die Linguistik und die Sozialwissenschaften vertreten.

## DAS PROJEKTVORHABEN VON ARCHEOINF

Das Projekt „*ArcheoInf*“ wird seit dem Jahr 2008 von der DFG im Rahmen der Aktionslinie „Themenorientierte Informationsnetze“ gefördert und zählt zu den Pilotprojekten, in denen fachspezifische Bedarfe an das Management von Forschungsdaten eruiert werden sollen. In diesem Fall geht es um Feldforschungsdaten aus dem Mittelmeerraum, die im Rahmen von archäologischen Surveys gesammelt werden. Partner im *ArcheoInf*-Konsortium sind neben den Universitätsbibliotheken Bochum und Dortmund das Archäologische Institut der Georg-August-Universität Göttingen, der Lehrstuhl für Software-Technologie der Technischen Universität Dortmund sowie der Fachbereich Vermes-



**ZENTRUM FÜR ARCHÄOLOGISCHE FELDFORSCHUNGSDATEN**  
VON KAROLIN BUBKE, MAIKE LINS

sung und Geoinformatik der Hochschule Bochum. Als Projektpartner beteiligt ist unter anderem das Institut für archäologische Wissenschaften der RUB.

Wie sieht die Arbeit der Archäologen vor Ort konkret aus? Im Gegensatz zu einer archäologischen Ausgrabung ist ein Survey darauf ausgerichtet, in vergleichsweise kurzer Zeit ein großes Gelände zu untersuchen und Hinweise auf frühere Besiedlungen oder andere Nutzungen zu erhalten. Die Mitarbeiter laufen dafür die Fläche ab, notieren ihre Beobachtungen zu Besonderheiten des Geländes (Bewuchs, Störung usw.) und sammeln Gegenstände auf, die Hinweise geben könnten. Dabei fällt in kurzer Zeit eine große Anzahl von Funden an, die beurteilt und dokumentiert werden müssen – Massendaten, deren ausführliche Auswertung erst später am Heimatort stattfinden kann. Die Funde selbst sind dabei meist nicht mehr greifbar, sondern verbleiben im Ursprungsland. Die Ergebnisse eines Surveys können dazu führen, dass Ausgrabungen im Anschluss gezielt durchgeführt werden können, wenn es z.B. Hinweise auf eine Nekropole oder ein bedeutendes Gebäude gibt.

Die Forschungsarbeit der Archäologen basiert also auf einer Datenbank, die parallel zur Arbeit im Feld angelegt wurde und in der eine Vielzahl von Funden mit ihren Abmessungen und Hinweisen zu den Fundumständen festgehalten sind. Diese Daten werden meist ergänzt durch Fotos und Vermessungsangaben, Karten, Litera-

turhinweise und mehr. Die Vielzahl der dokumentierten Funde kann unterschiedliche Forschungsinteressen bedienen. Meistens werden sie projektabhängig jedoch nur in Hinblick auf eine bestimmte Fragestellung betrachtet. Nach Veröffentlichung der Forschungsergebnisse verbleiben die Daten bisher in der Regel bei den Projektleitern. Selbst wenn sie für andere Vorhaben zur Verfügung gestellt würden, wäre die Interpretation der Daten nicht ohne weiteres möglich, da in der Regel keine Standards verwendet werden, wie die *ArcheoInf*-Erfahrungen der letzten Jahre gezeigt haben. Problematisch ist außerdem, dass die selbst angelegten Datenbanken meist nicht informationstechnischen Ansprüchen genügen: Fund-, Bild-, Literatur- und Geodatenbanken stehen oftmals separat nebeneinander. Die Daten sind z.T. inkonsistent und nicht in andere Systeme exportierbar. Hier setzt das Projekt *ArcheoInf* an: Die beteiligten Informatiker normalisieren die ursprünglich heterogenen Daten und bilden sie in einer Ontologie ab, um sie in einem Portal zeitgleich recherchierbar zu machen. Hierbei wird unter anderem der Standard CIDOC Conceptual Reference Model berücksichtigt, um eine spätere Austauschbarkeit der Daten mit anderen Institutionen aus dem Kulturerbe-Bereich zu gewährleisten.

Die am Projekt beteiligten Bibliotheken tragen Sorge dafür, dass der in *ArcheoInf* entwickelte Prototyp langfristig abrufbar ist. Über ein Redaktionssystem können auch nach Projektende weiterhin Daten eingespielt wer-



den, wenn sie den *ArcheoInf*-Standards entsprechen. Zu diesem Zweck wird in einem weiteren Projekt eine Datenbank entwickelt, die für Surveys und Grabungen einsetzbar ist und die auf der *ArcheoInf*-Ontologie basiert. Viele archäologische Projekte haben hieran bereits ihren Bedarf gemeldet und erklären sich im Gegenzug dazu bereit, ihre Daten für die Recherche im Portal zur Verfügung zu stellen.

## UMDENKEN FÖRDERN: OPEN DATA ALS PUBLIKATIONSFORM DER ZUKUNFT

Das Projekt *ArcheoInf* zielt also darauf ab, Forschungsdaten von unterschiedlichen Survey- und später auch Grabungsprojekten weltweit online abrufbar zu machen. Je nach Fortschritt der archäologischen Projekte kann dabei über ein Rechtemanagement gesteuert werden, wer zu welchem Zeitpunkt Zugriff auf die Daten haben darf. Ziel muss dabei insbesondere sein, die Daten nach Publikation der projektbezogenen Forschungsergebnisse offen und frei zur Verfügung zu stellen. Hier muss allerdings noch viel Überzeugungsarbeit geleistet werden: In vielen Disziplinen, darunter auch in der Archäologie, sind die Bedenken schwerwiegend und die Bereitschaft, Datasharing zu betreiben, ist vielfach nicht vorhanden. Es gibt Vorbehalte dahingehend, dass Forschungsergebnisse kontrolliert und widerlegt werden könnten, dass andere Wissenschaftler ohne großen Aufwand Nutzen ziehen könnten, ohne sich finanziell an der Datenerhebung beteiligen zu müssen usw.

Hier sind Wissenschaftseinrichtungen und Förderinstitutionen gleichermaßen gefragt, die eine offene Bereitstellung der Daten fordern und fördern müssen. So hat bereits die „Allianz der deutschen Wissenschaftsorganisationen“ mit Mitgliedern wie der DFG, der Max-Planck-Gesellschaft, der Hochschulrektorenkonferenz oder dem Wissenschaftsrat im Juni 2010 gemeinsame „Grundsätze zum Umgang mit Forschungsdaten“ verabschiedet und verlangt darin unter anderem, Daten aus öffentlich geförderter Forschung grundsätzlich offen bereitzustellen und dabei auf Standards zurückzugreifen, so dass sie problemlos nachgenutzt werden können.<sup>3</sup>

## CHANCEN NUTZEN: OPEN BIBLIOGRAPHIC DATA UND ARCHEOINF

Bibliotheken haben ihre Position als Ansprechpartner in Sachen Open Access und Open Data eingenommen und versuchen – auch im Projekt *ArcheoInf* – die Wissenschaft aufzuklären und zu überzeugen, dass freier Zugang zu Information und Wissen nicht schadet, sondern in vielerlei Hinsicht jedem Einzelnen Nutzen bringen kann. Um mit gutem Beispiel voranzugehen, geben derzeit viele Bibliotheken ihre Online-Kataloge zur Nachnutzung frei. Dies kommt wiederum dem Projekt *ArcheoInf* zugute, da dadurch die Anreicherung der Forschungsdaten mit relevanten Literaturdaten maßgeblich erweitert werden kann.

Feldforschungsprojekte, wie die von den mit *ArcheoInf* kooperierenden Partnern durchgeführten Kampagnen, bringen in ihren Forschungsdatenbanken neben Funddaten, Geoinformationen und Bildern auch eine Vielzahl projektrelevanter Literaturhinweise mit. Diese beinhalten mitunter sogar dezidierte Vermerke einzelner Textstellen eines Titels mit Bezug zu bestimmten Fundstücken. Es findet sich hier nicht nur leicht aufzufindende Standardliteratur, sondern auch Spezialliteratur zu bestimmten, engen Themenbereichen sowie auch graue Literatur oder im Ausland erschienene Schriften, die den Weg in deutsche Bibliotheken vielfach nicht gefunden haben.

Die Tatsache, dass die von Grabungs- oder Surveymitarbeitern gesammelte (Spezial-)Literatur nicht nur für einen engen Mitarbeiterkreis auffindbar ist, sondern für jeden fachlich Interessierten nachgewiesen wird, ist der besondere Wert der Erfassung der Daten durch *ArcheoInf*. Doch auch die Verzeichnung von Fachliteratur innerhalb eines Projektes vermag es nicht, auch sämtliche verwandte Literatur zu erfassen oder gar inhaltliche Verknüpfungen zu weiteren thematisch interessanten Titeln über Tags, Schlagworte, Kategorien oder Ähnliches zu schaffen.

An dieser Stelle werden die von Bibliotheken freigegebenen Katalogdaten wirksam. Die Integration von Titeldaten, die andere große Bibliotheken bereits zur freien Nutzung verfügbar gemacht haben, in den Datenpool unseres Projektes, soll es möglich machen, die Literaturdaten aus den Partnerprojekten durch eine große Zahl weiterer Titel zu ergänzen, sie miteinander zu verknüpfen und die Suche innerhalb des Literaturrechercheportals von *ArcheoInf*, in dem diese vorgehalten werden sollen, attraktiver, ergiebiger und erfolgreicher zu gestalten. Es werden hierbei Daten nachgenutzt, die als Open Bibliographic Data verfügbar sind. Das bedeutet, dass sie für jeden ohne Einschränkung zur Nutzung und Weiterverarbeitung frei zugänglich sind.

## AUSWAHL UND NACHNUTZUNG FREIER TITELDATEN DURCH ARCHEOINF

Allem voran steht hierbei zunächst die Auswahl der relevanten zu nutzenden Daten anhand ihrer Herkunftsbibliotheken. Im Falle von Universitätsbibliotheken – und um solche handelt es sich bei den liefernden Institutionen zumeist – ist das entscheidende Kriterium dabei zunächst das Vorhandensein eines archäologischen Fachbereiches an der betreffenden Universität – nur so ist ein nennens- und nutzenswerter Bestand an archäologisch relevanter Fachliteratur zu erwarten.

Auf der Open-Data-Plattform *CKAN – Comprehensive Knowledge Archive Network* – findet der Interessierte gleichsam ein Katalogsystem von Open Knowledge Paketen und Projekten. Die Plattform ist eine Einrichtung der Open Knowledge Foundation, die seit einigen Jahren den Aufbau eines Pools von frei verfügbaren Informationen aller Art stark vorantreibt.<sup>4</sup> Die Website bietet eine umfangreiche Liste von Datenpaketen und hält diese zum Download vor. Eine Gruppen-Zuordnung der verfügbaren Daten ermöglicht es, leicht die Pakete aufzufinden, die für die eigene Verwendung von Interesse sein können.

Aus der Liste der mit *Bibliographic Data* überschriebenen Gruppe<sup>5</sup> boten sich als geeignete Datenquellen für *ArcheoInf* die Daten der British National Library, der

Universitäts- und Stadtbibliothek Köln, der Universitätsbibliothek Tübingen sowie der Library of Congress an.

Institutionen, die ihre bibliothekarischen Daten zur freien Nutzung bereitstellen, tun dies zwar unter Verwendung verbreiteter bibliothekarischer Standards, dennoch erfordert die Einbindung fremder Daten in das System von *ArcheoInf* in den meisten Fällen eine ganze Reihe von vorbereitenden Schritten der Anpassung.

Die Architektur der verschiedenen Bestandteile des Literaturrechercheportals von *ArcheoInf* wurde bereits mit Rücksicht auf anerkannte und zukunftsfähige Standards für die Speicherung bibliographischer Daten gestaltet. Die offen erreichbaren Daten stellen uns vor der möglichen Integration in ein vorhandenes System jedoch vor neue Herausforderungen: Die Zahl der bibliothekarischen Standards ist groß und ebenso ist es die Menge der vorhandenen Datenformate, in denen die offenen Daten zur Nutzung bereitstehen. Vor der Einbindung der Daten steht also die Untersuchung der verwendeten Formate und die Vertiefung in die Struktur der tatsächlich vorhandenen Titeldaten: in Katalogisierung, Sacherschließung und Klassifikationen. Nicht immer liegen die Katalogdaten in Formaten vor, die es erlauben, sie ohne weitere Zwischenschritte in unseren Datenpool aufzunehmen.

Ist dies der Fall, gilt es, die Konvertierung der Daten aus dem verfügbaren Format in ein mit den im Projekt *ArcheoInf* eingesetzten Formaten kompatibles Format zu bringen.

Die Entwicklung des *ArcheoInf*-Literaturrechercheportals sieht es vor, Literaturdaten in dem von der Library of Congress entwickelten *Metadata Object Description Schema MODS* – ein XML-Format für bibliografische Metadaten – einzuspielen; auch dies ist ein hinlänglich anerkannter Standard.

Ist die gewünschte Datei mit Titeldaten in einem nichtkompatiblen Format, so wird im Rahmen einer Konkordanz eine Liste von Entsprechungen der in dem Ur-

sprungsformat verwendeten Kategorien, Elemente oder Auszeichnungssprachen zu MODS erstellt. Diese wird wiederum als Grundlage benutzt, um zur Definition von Umwandlungsregeln ein XSLT-Stylesheet zu erstellen, auf dessen Basis anschließend eine Konvertierung erfolgen kann.

Ist dies geschehen, so gilt es, aus der Menge der jeweils verfügbaren Katalogdaten die für eine Einbindung in unser Projekt wesentlichen Titeldaten herauszufiltern. Es sollen hier schließlich nur Daten einfließen, die für eine archäologisch basierte Suche im Portal lohnend sein können. Dies erfolgt über die verwendeten Sacherschließungselemente – Klassifikationen und Schlagwörter. Sinnvoll ist zumeist eine Selektion über – falls vorhanden – Klassifikationen, da die Ergebnismenge höher, gleichzeitig aber die Zahl der für unseren Zweck untauglichen Titel zu vernachlässigen ist.

Über eine Anbindung der vorgefundenen Sacherschließungselemente an die in *Archeolnf* erstellten Thesauri, die eine Vielzahl von archäologischen Begriffen bein-

halten und einen Suchzugang auch über die hier eingegebenen Konzepte erlauben, bietet sich dem Benutzer schließlich die Möglichkeit, ganz gezielt die Suchmaschine anzufragen. Er erhält neben den Literaturdaten der *Archeolnf*-Partnerprojekte in der Ergebnismenge nun auch relevante Titel aus Fremdkatalogen und in Verbindung mit den weiteren Forschungsdaten der Projekte eine umfangreiche Informationsbasis.

Ganz im Sinne der Open-Data-Initiative erfahren die freigegebenen Daten der Bibliotheken in ihrer Weiterverwendung und -verarbeitung durch *Archeolnf* weiteren Nutzen, kommen durch ihre Anbindung an die Daten des *Archeolnf*-Systems einer noch größeren Zahl von Menschen in neuen Zusammenhängen zu Gute und erhalten so einen neuen Wert – und bieten einem Projekt wie *Archeolnf* ganz neue Perspektiven.

Dr. Karolin Bubke ist Fachreferentin für Archäologie, Ur- und Frühgeschichte, Geschichte, Kunstgeschichte, Klassische Philologie, Anglistik sowie Projektkoordinatorin für das DFG-Projekt *Archeolnf*.

Maike Lins ist Mitarbeiterin im DFG-Projekt *Archeolnf*.

## ENDNOTES

<sup>1</sup> Vgl. Wissenschaftliche Literaturversorgungs- und Informationssysteme: Schwerpunkte der Förderung bis 2015, Bonn 2007, S. 7–9.

<sup>2</sup> DFG, Aufbau von Informationsinfrastrukturen für Forschungsdaten: DFG bewilligt 27 Projekte, 2011. [http://www.dfg.de/foerderung/info\\_wissenschaft/archiv/2011/info\\_wissenschaft\\_11\\_18/index.html](http://www.dfg.de/foerderung/info_wissenschaft/archiv/2011/info_wissenschaft_11_18/index.html) (abgerufen am 13.10.2011).

<sup>3</sup> Grundsätze zum Umgang mit Forschungsdaten, Berlin 2010.

<sup>4</sup> OPEN KNOWLEDGE FOUNDATION DEUTSCHLAND E.V., Mission Statement. Open Knowledge Foundation Deutschland, 2011. <http://okfn.de/organisation/mission/> (abgerufen am 03.01.2012).

<sup>5</sup> COMPREHENSIVE KNOWLEDGE ARCHIVE NETWORK, Bibliographic data. the data hub, 2011. <http://thedatahub.org/group/bibliographic> (abgerufen am 03.01.2012). Derzeit enthält diese Liste 51 Einträge.

## LITERATURVERZEICHNIS

Grundsätze zum Umgang mit Forschungsdaten, Berlin 2010 [= Working Paper Series des Rates für Sozial- und Wirtschaftsdaten, Bd. 156].

COMPREHENSIVE KNOWLEDGE ARCHIVE NETWORK, Bibliographic data. the data hub, 2011. <http://thedatahub.org/group/bibliographic> (abgerufen am 03.01.2012).

Wissenschaftliche Literaturversorgungs- und Informationssysteme: Schwerpunkte der Förderung bis 2015, Bonn 2007.

DFG, Aufbau von Informationsinfrastrukturen für Forschungsdaten: DFG bewilligt 27 Projekte, 2011. [http://www.dfg.de/foerderung/info\\_wissenschaft/archiv/2011/info\\_wissenschaft\\_11\\_18/index.html](http://www.dfg.de/foerderung/info_wissenschaft/archiv/2011/info_wissenschaft_11_18/index.html) (abgerufen am 13.10.2011).

OPEN KNOWLEDGE FOUNDATION DEUTSCHLAND E.V., Mission Statement. Open Knowledge Foundation Deutschland, 2011. <http://okfn.de/organisation/mission/> (abgerufen am 03.01.2012).