Heike Neuroth, Stefan Strathmann, Achim Oßwald,
Regine Scheffel, Jens Klump, Jens Ludwig [Hrsg.]

Langzeitarchivierung von Forschungsdaten

Eine Bestandsaufnahme

Kapitel 8
Altertumswissenschaften

vwh Verlag Werner Hülsbusch
Fachverlag für Medientechnik und -wirtschaft

Universitätsverlag Göttingen

Langzeitarchivierung von Forschungsdaten

Eine Bestandsaufnahme

nestor

Heike Neuroth, Stefan Strathmann, Achim Oßwald,
Regine Scheffel, Jens Klump, Jens Ludwig [Hrsg.]

Langzeitarchivierung von Forschungsdaten

Eine Bestandsaufnahme

GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

Förderkennzeichen: 01 DL 001 B

Universitätsverlag Göttingen

vwh
Verlag Werner Hülsbusch
Fachverlag für Medientechnik und -wirtschaft

D-GRID GmbH

Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme
hg. v. Heike Neuroth, Stefan Strathmann, Achim Oßwald, Regine Scheffel, Jens Klump, Jens Ludwig
im Rahmen des Kooperationsverbundes nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland
nestor – Network of Expertise in Long-Term Storage of Digital Resources
http://www.langzeitarchivierung.de/

Kontakt: editors@langzeitarchivierung.de
c/o Niedersächsische Staats- und Universitätsbibliothek Göttingen,
Dr. Heike Neuroth, Forschung und Entwicklung, Papendiek 14, 37073 Göttingen

Die Herausgeber danken Anke Herr (Lektorat) und Sonja Neweling (Redaktion) sowie Martina Kerzel und Lajos Herpay (Gestaltung und Montage) für ihre unverzichtbare Unterstützung bei der Fertigstellung des Handbuchs.

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter http://www.d-nb.de/ abrufbar.

Die Inhalte des Buches stehen auch als Onlineversion
(http://nestor.sub.uni-goettingen.de/bestandsaufnahme/)
sowie über den Göttinger Universitätskatalog
(http://www.sub.uni-goettingen.de) zur Verfügung.
Die URN lautet: http://nbn-resolving.de/urn:nbn:de:0008-2012031401.

Die digitale Version dieses Werkes steht unter einer Creative Commons Namensnennung-Nicht-kommerziell-Weitergabe unter gleichen Bedingungen 3.0 Unported Lizenz.

Einfache Nutzungsrechte liegen beim Verlag Werner Hülsbusch, Boizenburg.
© Verlag Werner Hülsbusch, Boizenburg, 2012
http://www.vwh-verlag.de
In Kooperation mit dem Universitätsverlag Göttingen.

Markenerklärung: Die in diesem Werk wiedergegebenen Gebrauchsnamen, Handelsnamen, Warenzeichen usw. können auch ohne besondere Kennzeichnung geschützte Marken sein und als solche den gesetzlichen Bestimmungen unterliegen.

Druck und Bindung: Kunsthaus Schwanheide
Printed in Germany – Als Typoskript gedruckt –

ISBN: 978-3-86488-008-7

8 Altertumswissenschaften

Ortwin Dally, Friederike Fless, Reinhard Förtsch

8.1 Einführung in den Forschungsbereich

Spricht man von den Archäologien bzw. den Altertumswissenschaften, so meint man damit nicht ein einzelnes Fach, sondern eine Vielzahl von Disziplinen, deren Genese sich vielfach bis in die 2. Hälfte des 19. Jahrhunderts zurückverfolgen lässt und ihren Niederschlag an den Universitäten sowie an Institutionen wie dem Deutschen Archäologischen Institut (DAI) gefunden hat. Wesentlich befördert wurde diese Entwicklung durch neu einsetzende Großgrabungen in Kleinasien, Griechenland und dem Vorderen Orient, die die Ausgräber mit einer Fülle neu entdeckter Objekte konfrontierten. Die Ausdifferenzierung der universitären Landschaft ist innerhalb der Altertumswissenschaften im Verlaufe des 19. Jahrhunderts zunächst anhand von Zeugnissen vollzogen worden: die Philologie wurde zuständig für die schriftlichen Quellen, die Archäologie, ihrerseits aufgegliedert in Teildisziplinen mit unterschiedlichen zeitlichen und geografischen Arbeitsgebieten, wurde zuständig für die materiellen Hinterlassenschaften der Antike, Numismatiker spezialisierten sich auf die Untersuchung von Münzen usw. Dieser Prozess setzt sich seither unvermindert fort: Bis vor wenigen Jahren waren paläogenetische Forschungen noch nicht Teil des Fächerkanons.

Entsprechend komplex sind heutzutage archäologisch-altertumswissenschaftliche Arbeitsfelder, deren Datenproduktion ein Kompetenzzentrum intensivieren, methodisch strukturieren, wissenschaftlich integrieren und längerfristig existentiell absichern soll. Sie umfassen Ausgrabungen (Dokumentation von Architektur, Stratigrafie, Gräbern etc.), Prospektionen/Surveys (Begehungen, Testgrabungen, Keramiksammlungen), Fundbearbeitung (z.B. Keramikanalyse), Fotogrammetrie (z.B. Aufnahme von

Gebäuden), Chronometrie (unterschiedliche naturwissenschaftliche und kunstgeschichtliche Methoden zur Datierung), Klima- und Landschaftsgeschichte (Geologie, Geomorphologie, Hydrologie), anthropologische Untersuchungen (Untersuchungen von Skeletten, Ernährungsgewohnheiten, Krankheiten und der genetischen Hinweise auf Verwandtschaftsbeziehungen) sowie epigrafische, philologische, linguistische Untersuchungen (Editions- und Corpusprojekte) und die Dokumentation und Restaurierung von Gebäuden.

Diese und weitere relevante Disziplinen und Forschungsmethoden sind, da sie z.B. in Feldforschungsprojekten und Texteditionen entweder Primärdaten generieren oder aber komplexe rechnergestützte Auswertungen verwenden, von jeher in besonderem Maße auf eine sorgfältige Dokumentation und visuelle Veranschaulichung ihrer Forschungsobjekte und -ergebnisse angewiesen. Waren dies noch bis vor wenigen Jahrzehnten ausschließlich Gipsabgüsse und -abklatsche, Zeichnungen, Fotografien und analoge Publikationen, geschieht dies heute überwiegend in Form digitaler Objekt- und Raumdaten, oft in lokalen geografischen Informationssystemen (GIS) vernetzt, statistischen Untersuchungen, 2-D- und 3-D-Rekonstruktionen. Diese tiefgreifenden Veränderungen werfen gerade für die Altertumswissenschaften nicht zuletzt dadurch Probleme auf, dass im Vorgang der Ausgrabung antike Befundsituationen irreversibel zerstört werden und somit die Dokumentation der Grabung zu einer einmaligen Primärquelle wird, die ihre Gültigkeit nie verliert. Aus diesem Grund sind die Altertumswissenschaften in besonderem Maße auf eine umfassende Dokumentation aller Arbeitsschritte und eine nachhaltigen Sicherung der Forschungsdaten angewiesen, was bei digitalen Daten eine besondere Brisanz besitzt.

In der aktuellen altertumswissenschaftlichen Forschung werden zudem immer komplexere Fragestellungen kultureller und gesellschaftlicher Dynamiken untersucht, so dass viele Disziplinen mit sehr unterschiedlichen Qualitäten und Quantitäten von Rohdaten und Informationen zusammenarbeiten und diese zur Beantwortung der Ausgangsfragen verknüpfen müssen. Dabei spielt auch die Vernetzung mit früher erfassten, analogen Daten eine immer größere Rolle.

Für die stark anwachsenden digitalen Datenbestände, die zunehmend den Arbeitsalltag eines jeden Altertumswissenschaftlers prägen, fehlen

jedoch fachspezifische Lösungen, die ihre analytische Durchdringung, kurz- und mittelfristige Sicherung, projektübergreifende Vernetzung und ihre Verfügbarkeit im Internet adressieren. Man kann dies etwa an der Diskrepanz zwischen Datenhaltung und -auswertung beobachten. Für die von der Retrodigitalisierung erfassten Bereiche (Bibliotheken, Archive, Stichwerke, Photographien) gibt es inzwischen einen hohen Bestand zugänglicher Daten im Netz, während die Entwicklung von Auswertungswerkzeugen, die etwa Textmining betreffen oder Semantic Web-Strategien folgen, in ihrer Verwendung in den Altertumswissenschaften noch am Anfang stehen. Daraus resultieren Defizite bei der rechnergestützten Auswertung im Bereich hermeneutischer Wissenschaftsfelder. Umgekehrt ist weltweit zu beobachten, dass rechnergestützte Analyseprozesse sich vor allem im Bereich der Quantifizierung hoch entwickelt haben, wovon in den Archäologien insbesondere der GIS-Bereich profitiert hat.

Gleichzeitig sind Basisdaten international wegweisender GIS-Analysen der deutschen Forschung für eine Nutzung in anderen Fragezusammenhängen bislang noch selten über das Internet verfügbar.

8.2 Kooperative Strukturen

Die Voraussetzungen für Kooperationen im Bereich gemeinsamer Systeme ist in institutioneller Hinsicht einerseits gut, da die Altertumswissenschaften in der Bundesrepublik Deutschland in einer weltweit einmaligen Breite institutionell verankert sind. Das Deutsche Archäologische Institut als größte außeruniversitäre Forschungseinrichtung auf dem Gebiet der Archäologie und der Altertumswissenschaften betreibt über seine Zweiganstalten im In- und Ausland Ausgrabungen und Forschungen auf allen fünf Kontinenten. Die Landesdenkmalämter haben den Auftrag, Kulturgüter in den Bundesländern zu erforschen und zu bewahren. Die Denkmalämter sind jedoch in föderalen Strukturen organisiert. An den Akademien sind Langfristprojekte u.a. im Bereich der Ägyptologie, der griechischen und lateinischen Epigrafik verankert. An den Universitäten sind nicht nur einzelne Disziplinen und Einzelprojekte verankert, darüber hinaus ist die Forschungslandschaft durch große von der DFG geförderte Forschungsverbünde gekennzeichnet.

Das zentrale Problem besteht andererseits derzeit darin, dass die institutionelle Vielfalt eine Vielfalt von Systemen und Insellösungen hervorgebracht hat. Im besten Fall ist zwischen einzelnen Institutionen ein verteiltes und plattformübergreifendes Filesharing für die Datenhaltung vorgesehen. Auf einer etwas komplexeren Ebene wird verteiltes Arbeiten mit höherer diskursiver Engführung unterstützt von webbasierten Datenbanken, die mit einem verteilten Filesharing etwa der Rohdaten verbunden sind. Dies gilt für Datenbanken wie z.B. Arachne[1] in Köln und iDAI.field.[2] Möglichkeiten der Vorhaltung und Analyse von Daten, die in einem Geoinformationssystem mit Raumdaten verbunden sind, sind in einem Web-GIS in den Altertumswissenschaften jedoch kaum entwickelt. Weitere diskursiv unterstützende Systeme wären am ehesten im Semantic Web zu vermuten, sind in der Fachcommunity aber ebenfalls nur in ersten Ansätzen zu finden.[3]

8.3 Daten und Metadaten

Der größte Problembereich im Umgang mit altertumswissenschaftlich-archäologischen Forschungsdaten besteht im noch nicht etablierten Bewusstsein für die Orientierung an Minimalstandards.[4] Hier ist in der deutschen Archäologie und Altertumswissenschaft, abgesehen von der Ur- und Frühgeschichte, ein bewusstseinsmäßiger Rückstand von zumindest einer Generation zu beobachten. In dieser Zeitspanne sind die durchaus auch vorhandenen deutschen Beiträge zu internationalen Konferenz-Plattformen wie der Computer Applications in Archaeology (CAA)[5] nicht in demselben Maße in die fachliche Praxis überführt worden wie im anglo-amerikanischen Bereich. Daher ist gegenwärtig eine Umbruchssituation zu konstatieren, die vor allem durch Heterogenität gekennzeichnet wird, die von sektoral hochentwickelten Verfahren bis hin zu gravierenden Rückständen reicht. Das Ziel muss also sein, die An-

1 Vgl. ARACHNE (2011a).
2 Vgl. Deutsches Archäologisches Institut (2011a).
3 Vgl. CLAROS (2011); Mitglieder der Special Interest Group (vgl. CIDOC (2011a)).
4 Vgl. hierzu die Publikation des DAI – Leitfaden zur Anwendung von Informationstechnik in der archäologischen Forschung (vgl. Deutsches Archäologisches Institut (2009)).
5 Vgl. CAA (2011).

wendung von datenbezogenen Minimalstandards in der Fachcommunity zu stabilisieren, um die heterogenen Primärforschungsdatenbestände auch langfristig noch analysieren, erfassen und kategorisieren zu können. Dasselbe Ziel muss auch für komplexe Datenbestände gelten, bis hin zu solchen aus 3-D-VR bzw. Grabungs- und Landschafts-GIS-Systemen.

Zum allgemein akzeptierten, aber bei weitem nicht realisierten State of the Art gehört, dass die verwendeten Softwarelösungen weder als Ganze proprietär sein dürfen noch proprietäre Werkzeuge bzw. Standards verwenden sollten – bis hin zu den verwendeten Dateiformaten, die nur in Formaten mit offengelegten Spezifikationen niedergelegt und mit dem Ziel informeller Verlustfreiheit über antizipierte Technologiewechsel hinweg behandelt werden sollten. So ist bei den zu verwendenden Dateiformaten für Einzeldateien die Situation relativ klar zugunsten von TIFF und XML gelöst, anders als bei Vektor- oder 3-D-Datenformaten. Auch die Abbildung von Datenbanklogiken aus beliebigen Datenbank-Management-Systemtechnologien in XML ist keinen grundsätzlichen Diskussionen mehr unterworfen. Die mittel- und langfristigen Perspektiven von Informationsverlust bei der zu Unrecht für gelöst gehaltenen Konvertierung von Dateiformaten sind Forschungsgegenstand gewesen.[6] Erfolgversprechende Lösungsansätze liegen insbesondere in der formatunabhängigen Informationsabstraktion.[7] Dagegen ist die Diskussion zur Dateninteroperabilität zwischen verschiedenen GIS-Systemen noch kaum begonnen worden und angesichts der komplexen Architekturen auch außerordentlich schwierig zu lösen.[8]

Für die Erzielung von semantischer Interoperabilität wiederum herrscht international ein offener Disput zwischen den Verfechtern komplexer oder radikal vereinfachter Strukturen, in dem sich je nach Sachgebiet unterschiedliche Lösungen gegenüberstehen.[9]

Erstaunlich wenige Überlegungen finden sich zu altertumswissenschaftlichen Datenmodellen als solchen. Zumeist existieren sie als impli-

6 Vgl. PLANETS (2011).

7 Vgl. insbesondere die Abschlussarbeiten der Kölner HKI im Rahmen von Planets: Thaller (2009).

8 Vgl. Simon et al. (2010).

9 Vgl. die Blogs von Heath (2011); Kansa (2011); Wilde (2011); Wilde (2009); vgl. auch die CAA Semantic Special Interest Group (2011).

zite Annahmen in Datenbanksystemen zu digitalen Bildern, Inschriften und Objekten bzw. in GIS-Systemen. Eine seltene Ausnahme bildet das aus dem Jahr 2001 stammende XML-Schema ArcheoML, das in einem US-amerikanischen Repository wie Open Context[10] eingesetzt wird, sich aber nicht wirklich durchgesetzt hat. Das Modell versucht, viele Fragen von unscharfem Wissen bis hin zu Interoperabilität zu adressieren und intrinsisch bereits auf der Datenebene zu lösen, die ansonsten mit hohem Aufwand durch spätere technologische Eingriffe auf der Informationssystemsebene bearbeitet werden müssen. Darunter leidet allerdings die Einfachheit der Handhabung, was gerade in einer Fachcommunity problematisch ist, die sich oft noch an der Adaptionsschwelle befindet.

Defizite bestehen auch hinsichtlich einer stärkeren Einbindung informationstechnologischer Themen in die Ausbildung, um über bisherige Ansätze hinausgehend Studenten frühzeitig mit Standards und gut zu nutzenden Werkzeugen vertraut zu machen.[11] Zum anderen bestehen ungelöste Probleme im Bereich rechtlicher Fragen. Die freie Verfügbarmachung von Daten im Open Access wird zum Teil durch urheberrechtliche Regelungen erschwert. So unterliegen beispielsweise Satellitenbilder und LIDAR-Scans sehr strengen Urheberrechtsregeln, die eine Veröffentlichung im Internet zumeist verhindern. Andere Einschränkungen ergeben sich aus Bestimmungen des Denkmalschutzes der Bundesländer und anderer Staaten, die beispielsweise die Dokumentationen von Objekten in ausländischen Museen verhindern können.

8.4 Datenhaltung und Langzeitarchivierung

Die in den Altertumswissenschaften generierten Primärdaten reichen von digitalen Fotografien, Luftbildern, Satellitenbildern, Texten, Datenbanken, Vermessungsdaten, Punktwolken, Fotogrammetriedaten, 3-D-Rekonstruktionen und -modellen, Filmen/Videos, Vektorzeichnungen bis zu groß- und kleinformatigen Scans. Diese lassen sich generell in Primär- oder Rohdaten (z.B. analoge und digitale Fotos, Luftbilder, Vermessungsdaten, Filme/Videos etc.) und Metadaten trennen (z.B. georeferenzierte

10 Vgl. The Alexandria Archive Institute (2011).
11 Vgl. Cologne Center for eHumanities (2011).

Satellitenbilder, Datenbanken, 3-D-Rekonstruktionen und Modelle, Vektorzeichnungen, groß- und kleinformatige Scans etc.).

Als ein großer Problembereich zeichnet sich ab, dass diese Daten entsprechend der Vielzahl der Institutionen und Fachdisziplinen in sehr unterschiedlichen Formen vorgehalten, gesichert und archiviert werden und eher defizitär zugänglich und nachnutzbar sind. Abgesehen von Lösungen für einzelne Projekte und Institutionen sind vorhandene grundsätzliche Standards (z.B. OAIS), Lösungen (z.B. KOPAL[12]) und LZA-Forschungsansätze (z.B. das EU-Projekt PLANETS[13] oder das Data Conservancy Project der Johns Hopkins Universities Sheridan Libraries[14]) für die Altertumswissenschaften nicht im Sinne einer lauffähigen und Fachspezifika berücksichtigenden Anwendung zusammengeführt – was unschwer möglich wäre.

Derzeit werden die Daten entsprechend der Vielzahl der Institutionen und Fachdisziplinen in sehr unterschiedlichen Formen gesichert, archiviert, zugänglich und nachnutzbar gemacht. Abgesehen von Lösungen für einzelne Projekte und Institutionen wie z.B. im Bereich der Denkmalpflege oder für Projekte wie Arachne bzw. iDAI.field fehlen hier grundsätzliche Lösungen.[15] In der Berlin-Brandenburgischen Akademie sind intern Überlegungen zur langfristigen Sicherung von Daten in Angriff genommen worden. In den Landesdenkmalämtern liegen vereinzelt jeweils spezifische Regelungen für den Transfer von Grabungsdaten in Archiven vor. Es fehlen aber Modelle eines routinemäßig durchgeführten Archivierungsprozesses und von Fallback-Strategien im Sinne eines Standards wie OAIS. Diese werden im Rahmen einer AG Archivierung diskutiert.[16] Im DAI sind entsprechende Regelungen erst im Aufbau be-

12 Vgl. kopal (2011).

13 Vgl. PLANETS (2011).

14 Vgl. Choudhury (2009).

15 So hat das seit 2008 geförderte Projekt ArchaeoInf (vgl. ArchaeoInf (2011)) in keiner seiner Komponenten zugängliche Daten oder Demonstratoren online im Internet.

16 Der Verband der Landesarchäologen hat hierzu 2008 ein Kolloquium ausgerichtet (vgl. Verband der Landesarchäologen (2011a)). Die AG Archivierung ist Teil der Kommission „Archäologie und Informationssysteme“ und führte eine Umfrage durch, auf Grund derer offene Fragen, Handlungs- und Klärungsbedarf ermittelt werden. Ein Bericht hierzu wird Mitte Mai 2011 auf der Homepage des Verbandes veröffentlicht (vgl. Verband der Landesarchäologen (2011b)).

griffen, und auch die Museen der Staatlichen Museen Berlin stehen erst am Anfang, entsprechende Strategien zur Langzeitarchivierung zu entwickeln.[17] Eine Zusammenarbeit mit bereits existenten Forschungsnetzen und Datenservicezentren anderer Fachdisziplinen und Institutionen in Deutschland (z.B. MPDL[18], FIZ Karlsruhe[19], RatSWD[20], GESIS-ZA Zentralarchiv für Empirische Sozialforschung[21], Arbeitsgruppe „Informationsstruktur" der Leibniz-Gemeinschaft[22]) ist bislang allenfalls in Ansätzen verwirklicht worden.

8.5 Interoperabilität

Sämtliche Überlegungen zur Strukturierung von Daten und kooperativen Prozessen haben das Ziel, durch Auffindung, Verknüpfung und Analyse von Informationen den Forschungsprozess zu fördern, qualitativ zu verbessern und bestenfalls um neuartige Diskursstrukturen zu bereichern: Archivierung, Zugänglichkeit und Nachnutzbarkeit von standardisierten Daten bilden dafür die Grundlage. "Long term preservation is an interoperability issue along the time vector."[23]

Interoperabilität von Systemen

Da gegenwärtig und auch noch mittelfristig Informationen in technisch abgegrenzten Informationssystemen gespeichert, verarbeitet und zugänglich gemacht werden, ist der Wunsch nach Interoperabilität gleichbedeutend mit dem Wunsch nach Interoperabilität zwischen Systemen.[24] Sie sollen Informationen effizient austauschen und Nutzern zur Verfügung stellen, ohne dass dafür spezielle Abstimmungen zwischen Systemen untereinander getroffen werden. Vielmehr ist die Verwendung von standardisierten Datenschnittstel-

17 Im Rahmen der im Aufbau begriffenen Deutschen Digitalen Bibliothek (vgl. Deutsche Digitale Bibliothek (2011)).
18 Vgl. MPDL (2011).
19 Vgl. FIZ Karlsruhe (2011).
20 Vgl. RatSWD (2011a).
21 Vgl. GESIS (2011d).
22 Vgl. Bibliotheksdienst (2009), S. 915f.
23 Gradmann (2008b).
24 Vgl. Gasser; Palfrey (2007).

len gefordert (etwa OAI/PMH), in die dann unterschiedliche semantische Ontologien implementiert und auf die Ausgangssysteme gemappt werden können (etwa CIDOC-CRM, Dublin Core etc.). Indem man die Daten über diese Ebene verbindet, können neue Systeme durch eine Adressierung dieser Ebene ihre Daten in den Pool einbringen, ohne die anderen teilnehmenden Systeme kennen oder mit ihnen jeweils Abstimmungen aushandeln zu müssen. Altertumswissenschaftliche Daten sind von diesen Vorgehensweisen betroffen, insofern sie entweder in EUROPAEANA[25] bzw. seinen Zuführungsprojekten wie CARARE[26] verarbeitet werden. CARARE soll die Daten aus Ausgangssystemen in Datenmodelle überführen, welche dann in EUROPAEANA verarbeitet werden können. Die größte Menge aggregierter archäologischer Daten wird in der Plattform CLAROS[27] zusammengefasst sein, die auf dem Metdadatenschema CIDOC-CRM Core beruht. Zur Kodierung und Verlinkung der Daten werden das Resource Description Format und darauf aufbauende Standards verwendet, so dass diese Systeme respektive die altertumswissenschaftlichen Daten in ihnen zugleich Teile des Semantic Web sind.

Eine Erweiterung dieses Ansatzes versucht, unter Verwendung von CIDOC-CRM die ganze Erschließungstiefe von heterogenen Datenbanksystemen abzubilden. Dies ist von den Berliner Museen, dem DAI, TOPOI und der Universität zu Köln erprobt und auf einem Workshop im DAI unter internationaler Beteiligung diskutiert worden. Dabei zeigte sich vor allem, dass trotz aller Erfolge in einzelnen Pilotprojekten diese Versuche technisch und semantisch sehr aufwändig sind.[28] Jedoch stellen sie derzeit die einzige Möglichkeit dar, um komplexere semantische Strukturen auszudrücken.

Demgegenüber wäre eine Peer-to-peer-Abstimmung zwischen einzelnen Systemen ein wesentlich höherer und für größere Datenpools kaum noch

25 Vgl. Europeana (2011); Gradmann (2008a).
26 Vgl. CARARE (2011).
27 Das System soll 2011 online gehen und ist bereits funktionsfähig (vgl. CLAROS (2011)).
28 Vgl. CIDOC (2011b). Ursprünglich von Museen und anderen Institutionen als conceptual reference model entwickelt, um Kulturgüter zu verwalten, kann es genutzt werden, um heterogene Informationen aus verschiedenen Kulturerbebereichen zusammenzuführen. Es ist als internationaler ISO-Standard (ISO 21127:2006) zertifiziert. Eine Erweiterung für die archäologische Arbeitspraxis wurde von English Heritage entwickelt (vgl. University of Glamorgan (2011).

beherrschbarer Aufwand. In einzelnen Fällen inhaltlich tiefergehender Kontextualisierung behalten Peer-to-peer-Verknüpfungen von Systemen, obwohl für viele Bereiche konzeptionell überholt, ihre Bedeutung. So sind etwa ARACHNE und CENSUS als Informationssysteme beide mit der CLAROS-Plattform verbunden, können aber in ihren gegenseitigen Bezügen zwischen antiken Objekten und ihrer Rezeption wesentlich enger miteinander verbunden werden, als sich dies im CIDOC-CRM-Core abbilden ließe. Daher wird hier eine tiefergehende Verknüpfung nur dieser beiden Systeme angestrebt, die sich Dank des CIDOC-CRM aber weiterhin auf der Basis einer standardisierten Ontologie bewegt und daher auch nach außen hin weiterverwertbare Ergebnisse enthält. Ein ähnliches Beispiel wird im Rahmen eines Pilotprojektes des DAI und der BBAW umgesetzt durch die Verknüpfung eines GIS-Systems zu Felsinschriften in der Region Assuan (Oberägypten) mit philologisch-linguistischen Einträgen des Altägyptischen Wörterbuches und der archäologisch-epigrafischen Dokumentation von Felsinschriften.[29] Eine solche Verbindung eines projektbezogenen Datenbestandes mit bestehenden Informationsressourcen kann den traditionellen Bestand am Publikationssystem um eine Form bereichern, die sich nur noch in digitalen Medien wiedergeben lässt.

Wesentliches Merkmal der Interoperabilität von Systemen sollte sein, ihre Objektdaten durch ihre Schnittstellen transparent zu machen. Die Systemsicht soll also gegenüber einer möglichst reinen Objektdatensicht zurücktreten.

Moderne Repository-Systeme versuchen dasselbe Ziel nicht durch eine nachträgliche Aufrüstung zu erreichen, sondern indem sie in ihrer Architektur von vorneherein die Module Daten, Interfaces und Executable Programs in einer „separation of concerns“ trennen.[30]

Interoperabilität von Daten

Diese Sicht auf das einzelne digital repräsentierte Objekt wird im Linked-Data-Ansatz zum zentralen Konzept, wobei von einer Transparenz der Systeme ausgegangen wird bzw. von Informationsressourcen, die über-

29 Vgl. die Arbeitsstelle Altägyptisches Wörterbuch der Berlin-Brandenburgischen Akademie der Wissenschaften (2011).

30 Vgl. FedoraCommons (2011).

haupt nicht in einem komplexen Database Management System gespeichert sein müssen – dies aber nach wie vor auch sein können.[31] Sie müssen lediglich über eine URI (Uniform Resource Identifier) per HTTP direkt aufrufbar sein und ebenfalls per URI auf verknüpfte Daten hinweisen. Auch hier werden zur Kodierung und Verlinkung der Daten mit RDF und – zur Abfrage-Realisierung – SPARQL und OWL Semantic Web-Technologien verwendet. Im Bereich altertumswissenschaftlicher Daten hat sich hier das möglicherweise am schnellsten wachsende Aktivitätszentrum um das Linked-Data-Prinzip im Bereich der antiken Geografie gebildet.[32] Die Ortserwähnungen in so groß angelegten Informationsressourcen wie Google Books werden dabei vorwiegend als durch den PLEIADES-Gazetteer objektivierbare Marker für die Verknüpfung von Zitaten der Antikenrezeption seit dem Barock bis hin zur altertumswissenschaftlichen Literatur verwendet. Im zweiten Schritt werden dann die mit den Ortserwähnungen verbundenen intellektuellen Konzepte analysiert. Auf diesem Weg berühren sich dann erstmals Methoden wie Textmining mit dem Bereich der Archäoinformatik. An dieser Stelle setzen Bemühungen an, deutschsprachige archäologische Daten in diese Strukturen verstärkt einzubringen und ihr analytisches Potential zu nutzen.[33] Welche semantischen Strukturen sich in den Graphendarstellungen von Linked-Data der Fachcommunity ergeben, wird sich erst noch zeigen müssen. Bisher ist auch hier die theoretische Diskussion weitaus intensiver als die reale Umsetzung.

Interoperabilität von Datenformaten und Metadatenschemata

Gegenüber den bisher betrachteten Ansätzen verlieren die umfänglichen Thesaurus-Systeme immer mehr an Bedeutung, da sie meist sehr schwer zu handhaben sind und zudem als Solitäre konzipiert wurden, der Inter-

31 Vgl. Linked Data (2011). DBMS können ihre Objekte über eine Schnittstelle als Linked Data ausgeben.

32 Vgl. Babeu (2010), S. 56 ff. Den *state of the art* verkörpert gegenwärtig das JISC-Projekt PLEAGIOS (vgl. PLEAGIOS (2011)). Es verwendet den Linked-Open-Data-Ansatz (vgl. Linked Data (2011)) zur Aggregation mehrerer angeschlossener Datenlieferanten, die Open Annotation Collaboration Ontologie (vgl. Open Annotation Collaboration (2011)) sowie PLEIADES (vgl. PLEIADES (2011)) als Gazetteer.

33 Vgl. European Public Policy Blog (2010).

operabilität somit schwere Hindernisse entgegensetzen. Umgekehrt muss versucht werden, die außerordentlich wertvollen Mengen an mit solchen Systemen indizierten Daten zu nutzen und nicht etwa zu verlieren; der Thesaurus des Realkataloges des DAI Rom hat im Bereich der Altertumswissenschaften hier einen durch das Ausgangssystem erheblich erheblichen Technologiewechsel zum System ZENON[34] hin überstanden. Umgekehrt sind gerade bei ZENON die großen Chancen der Anbindung an objektbezogene Informationssysteme noch ungenutzt. Und auch die inhaltliche Struktur des Thesaurus selbst weist, wie immer in solchen Fällen, erhebliche zeitgebundene Merkmale einer an der Kunstarchäologie des sog. Dritten Humanismus (20er-Jahre des 20. Jahrhunderts) orientierten Struktur auf, für die etwa eine klare Verschlagwortung politischer Einheiten der Antike keine Bedeutung besaß. Eine kontrollierte Fortschreibung derartiger Thesauri in Richtung neuer Forschungsinteressen unter Beibehaltung der Nutzbarkeit bereits verschlagworteter Altdaten sind somit erhebliche Forschungsdesiderate im Bereich einer Interoperabilität der Daten.

Content Reference-Modelle waren als Metadatenschemata u.a. auch zur Einschränkung der Komplexitäten eines gegenseitigen, kumulativen Thesaurus-Mappings gedacht, haben gegenwärtig jedoch die Rolle der „zu komplexen Systeme“ übernommen – zumindest aus Sicht programmatisch weniger aufwändiger Linked-Data-Ansätze, weil sie letzen Endes immer sehr kleine bis sehr große Ontologien, also Wissenskonzeptualisierungen sind, wohingegen die Linked-Data-„Bewegung“ die inhaltlich-semantische Strukturierung ihrer Verknüpfungsgraphen als Problem ignoriert oder als nachgeordnet betrachtet, da sie vor der Etablierung minimalster URI-Vergaben für Objekte und deren Verknüpfungen zurückzustehen habe.

Auffallendes Indiz eines Trends hin zur Vereinfachung ist, dass auch ursprünglich vergleichsweise grob gehaltene Metadatenschemata wie Dublin Core, die vor allem zur Beschreibung von Büchern und Dokumenten im Web gedacht waren, in einer leicht erweiterten Form zunehmend auch zur Beschreibung von realen Objekten verwendet werden – allen

34 Vgl. Deutsches Archäologisches Institut (2011b).

voran durch die American Agora Excavation.[35] Als Konvergenzbewegung wird das von Hause aus hochkomplexe, mit großem institutionellen Aufwand entwickelte und politisch vereinbarte CIDOC-CRM meist in einer stark reduzierten Form, als CIDOC-CRM-Dublin Core-Metadaten Verwendung finden und dabei in etwa die Komplexität des erweiterten Dublin Core-Metadatensets aufweisen.[36]

Ein Ansatz, der seit 2005 vom Verband der Landesarchäologen der Bundesrepublik Deutschland verfolgt wird, ist der Datenaustauschstandard ADeX, mit dem zentrale Informationseinheiten aus komplexen Datenbanksystemen zusammengeführt und damit gemeinsam recherchierbar und kartierbar gemacht werden.[37] Wegen der Fokussierung auf denkmalpflegerische Belange (Verortung, Kartierung, Schutzkriterien) und der Ressourcenknappheit können komplexe Zusammenhänge wie z.B. präzise chronologische Fragestellungen allerdings auf diese Art und Weise nur unzureichend oder sehr ausschnitthaft abgebildet werden. Zu nennen wären hier als Standard oder als Quasistandard aber auch MuseumDAT, CDWA Lite, LIDO, MODS und CCO[38] im Bereich der Museen und Dokumentation von Kulturgütern, FRBR[39] für bibliografische Metadaten sowie der vom Europarat empfohlene International Data Core Standard for Archeological Sites and Monuments[40].

Die Versuche, Interoperabilität durch Standardisierung zu erreichen, haben ihrerseits eine Vielzahl von Standardisierungsvorschlägen erbracht, die bereits wieder unübersichtlich und unhandhabbar zu werden drohen. Man kann daher die Kritik an diesen Entwicklungen aus der Richtung radikaler

35 Vgl. Dublin Core-Felder (2011). Zu ARACHNE in Dublin Core vgl. ARACHNE (2011b).

36 Vgl. etwa die auf http://explore.clarosnet.org/XDB/ASP/clarosHome/ angeführten Partnerinstitutionen. Die Planungen der CENSUS-Datenbank zur Implementation des CIDOC-CRM im CLAROS-Verbund sind ebenfalls weit gediehen.

37 Archäologischer Datenexport (ADeX), entwickelt von der Kommission „Archäologie und Informationssysteme“ im Verband der Landesarchäologen der Bundesrepublik Deutschland, ist ein deutschlandweiter Standard der Landesämter für den Datenaustausch archäologischer „raumbezogener“ Informationen (vgl. Verband der Landesarchäologen (2011c)).

38 Vgl. MuseumDAT (2011); The Getty Research Center (2011); LIDO (2011); MODS (2011); Cataloging Cultural objects (2011).

39 Vgl. IFLA (2011).

40 Vgl. ICOM (2011).

„lightweight"-Ansätze verstehen[41], die all diese Aktivitäten als noch immer viel zu umfangreich, präskriptiv, anfällig und nicht nachhaltig kritisieren.

Man wird dieser Kritik einzig entgegenhalten müssen, dass es beim bisherigen Stand der Linked-Data-Anwendung kaum möglich ist, Verhältnisse abzubilden, die auch nur annähernd die Komplexität herkömmlicher Forschungsdiskurse erreichen. Eine den Altertumswissenschaften erreichbare, eindeutige Lösung des Dilemmas ist derzeit nicht recht in Sicht. Repositorien werden hier oft als Lösungsmuster diskutiert, sind jedoch ihrerseits aufwändig zu betreiben, so dass im Bereich der Fachcommunity derzeit keine Implementationen bekannt sind.

Interoperabilität von Text- und Objektdaten

Ein klassisches Problem der Geisteswissenschaften ist das Spannungsfeld von textueller und visueller Kommunikation und darin konkreter der Text-Bild-Bezug. In informationstechnologischer Hinsicht gehört dies Problem in den Bereich der Aggregation von Informationen über grundlegende Quellengattungsgrenzen hinweg. Seine besondere Ausprägung erhält das Problem durch die Tatsache, dass die Altertumswissenschaften sowohl objektbasierte als auch textbasierte Untersuchungen und Disziplinen mit einschließen. Ist im Textmining die Entity Extraction grundsätzlich gelöst und auch in den textbasierten Altertumswissenschaften bereits angewandt worden, so fehlen bisher Ansätze, mit denen Referenzen zu Objekten aus archäologischen Texten extrahiert und mit Objektdatensätzen aus entsprechenden Webdatenbanken automatisiert verbunden werden. D.h. vor der herkömmlichen Verbindung unterschiedlicher Ojekte durch Äquivalenzerkennung (record linkage, object identification, entity resolution, reference reconciliation etc.) müsste eine relevante Menge von Objektreferenzen aus Forschungstexten bereitgestellt und identifizierbar gemacht werden. Als nicht minder komplex stellt sich die Verbindung von unterschiedlichen Texten dar. Hier spielen Probleme wie Multilingualität, Begriffsgleichheit bei unterschiedlichem Sinngehalt („Teekesselchen") und automatisierte Textanalysewerkzeuge (Textmining, Zitationsstrukturen etc.) eine entscheidende Rolle.[42]

---

41 Vgl. die Blogs von Heath (2011); Kansa (2011); Wilde (2011), Wilde (2009); vgl. auch die CAA Semantic Special Interest Group (2011).

42 Vgl. eAqua (2011); Büchler; Heyer (2008/9); Büchler et al. (2010); Romanello (2010).

8.6 Pre-processing und Analyse

Altertumswissenschaftliche und archäologische Daten sind Sekundärquellen für die Forschung und haben als solche in vielen Fällen auch konservatorischen Charakter. Ihr Forschungsbezug bedeutet jedoch nicht, dass für alle Datenarten auch formalisierte Analyseverfahren existieren bzw. existierten. So konnten etwa Bilddaten nicht auf dieselbe Weise analysiert werden wie Texte. Der sehr große Bereich visueller Dokumentation bzw. Information war bisher jedoch weitgehend analytischen Verfahren entzogen. Mit der nach jahrzehntelangen Bemühungen nun in den Bereich einsetzbarer Algorithmen gelangenden Pattern Recognition, die u.a. auf die großen im Internet residierenden Sammlungen des „Photo Tourism" zielt[43], stehen hier gravierende Änderungen bevor. In Verbindung mit den immer dichter werdenden Geolokationsdaten dieser digitalen Sammlungen sind nicht nur Abgleiche mit älteren Aufnahmen und die Übertragung der Geodaten auch auf diese möglich, sondern auch zunehmend automatisiert generierte 3-D-Rekonstruktionen.

Zudem entzogen sich hermeneutische Fragestellungen lange Zeit der Formalisierbarkeit, bis die Konzipierung des Semantic Web hier Abhilfe versprach.

Viele Datenpublikationen hatten daher die Qualität von Aufbereitungen, die allenfalls als pre-processing gelten konnten und deren Wert daher oftmals angezweifelt wurde. Die allgemeine Tendenz, Digitalisierung als Datenproduktion zu perhorreszieren und nicht mehr zu fördern, ist ein klassisches Symptom dieser Haltung. Es kann jedoch kein Zweifel sein, dass Datenanalysen ohne Daten nicht existieren könnten; und dass Datenanalysen von Daten, die eigens für diese Analysen produziert werden müssen, von vornherein in der Gefahr sind, Zirkelschlüssen zu unterliegen. Es zeigt sich zunehmend, dass auch die bisher auf der Ebene des pre-processing residierenden Datensammlungen im Begriff sind, in die Reichweite analytischer Verfahren zu gelangen.

Auf mathematisch und statistisch analysierbare Fragestellungen beziehen sich seit jeher die herausragenden Analyseverfahren der Archäoinfor-

43 Vgl. Hayes (2009); Hartley; Zisserman (2004); die entsprechenden Anwendungen in CLAROS (2011).

matik.[44] Insbesondere in Verbindung mit Geografischen Informationssystemen haben diese Verfahren Eingang in große Teile der archäologischen Praxis gefunden, werden aber auch als Standalone-Methoden angewandt wie etwa Netzwerkanalysen.[45] Diese haben sogar einen sehr hohen Verallgemeinerungsgrad, d.h. sie können auf die unterschiedlichsten Felder angewandt werden, von Handelsprozessen bis hin zu Metadatenstrukturen.[46]

Mit sprachlicher Informationsverarbeitung, speziell dem Textmining, entstand ein zweiter Zweig analysestarker Verfahren aus dem frühesten Gebiet der Digital Humanities, d.h. der Editorik. In eAQUA hat das altertumswissenschaftliche Textmining einen exemplarischen Impuls bekommen, der sich hoffentlich fortsetzen wird.[47]

8.7 Interne Organisation: Perspektiven und Visionen

Die im Rahmen der Altertumswissenschaften produzierten digitalen Daten umfassen wie geschildert nahezu alle denkbaren Formen der Heterogenität. Aufgrund dieser Sachlage und der starken Ausdifferenzierung der Fächerlandschaft fehlt bislang eine zentrale Struktur, die ein Qualitätsmanagement für die Altertumswissenschaften insgesamt übernimmt, eine Adresse im internationalen Kontext darstellt, die Sicherung von Daten vornimmt, hierzu Empfehlungen, die auch zu Richtlinien der DFG werden könnten, ausspricht, juristische Fragen klärt sowie Informationsbedarf zu technischen Entwicklungen und gemeinsamen Codes abdeckt. Ein solches *Kompetenzzentrum*, wie es im Frühjahr 2011 von der DFG in einer vorbereitenden Stufe bewilligt wurde, ist im Bereich der Archäologie und Altertumswissenschaften ein dringliches Desiderat, da es nicht nur eine koordinierende Funktion für thematisch miteinander verbundene Fachgruppen oder übergreifende Forschungsfelder wahrnehmen könnte, sondern auch zu politischen sowie auch zu technischen Aspekten

44 Zum methodischen Spektrum der internationalen Konferenz CAA (Computer Applications and Quantitative Methods in Archaeology) seit 1973 vgl. allgemein CAA Proceedings (2011).

45 Zu Netzwerkanalysen vgl. exemplarisch Brughmans (2010).

46 Vgl. European Public Policy Blog (2010).

47 Vgl. eAqua (2011).

einer digitalen Infrastruktur in den Altertums- und darüber hinaus in den Geisteswissenschaften Stellung beziehen könnte. Wesentliche Ziele wären nicht nur die langfristige Sicherung von Daten in einem digitalen Archiv – hier liegt eines der drängendsten Probleme für viele Institutionen aus dem Bereich der Archäologien und der Altertumswissenschaften, sondern auch die Garantie von deren langfristiger Interoperabilität, die (möglichst) freie Verfügbarkeit von Daten und Verfahren *(distributed resources),* die Veröffentlichung erfolgreicher Anwendungen *(best practices)* sowie die Vernetzung von Arbeitsgruppen und Ressourcen durch einfache Austauschmechanismen.

Um diese Ziele zu erreichen hat die DFG eine Arbeitsgruppe eingerichtet.[48] Sie hat es sich zur Aufgabe gemacht, im Rahmen eines mehrstufigen Projekts die Einrichtung eines solchen Kompetenzzentrums zur verfolgen. In der nun bewilligten ersten Projektphase soll es zunächst um eine Evaluierung vorhandener Lösungsansätze gehen.[49] Hierzu zäh-

48 Mitglieder sind: Prof. Dr. Friederike Fless, FU Berlin, Sprecherin des Exzellenzclusters TOPOI von 2007–2011, als Vertreterin großer, von der DFG geförderter Forschungsverbünde und universitärer Einzelprojekte; Prof. Dr. Johannes Müller, CAU Kiel, Sprecher der Graduiertenschule „Human Development in Landscapes“ (Exzellenzinitiative), als Vertreter großer, von der DFG geförderter Forschungsverbünde und der Nachwuchsförderung; Prof. Dr. Ernst Pernicka, Curt-Engelhorn-Zentrum Archäometrie in Mannheim, als Vertreter universitärer Projekte und An-Institute für den Bereich der Naturwissenschaften; Dr. Heike Neuroth, Leiterin der Abteilung Forschung und Entwicklung an der SUB Göttingen und Konsortialführerin des BMBF Projektes „TextGrid – Vernetzte Forschungsumgebung in den eHumanities“ sowie langjährige Expertin der Themenbereiche digitale Langzeitarchivierung und Forschungsdaten; Prof. Dr. Ulrich Lang, Leiter des Rechenzentrums der Universität zu Köln, als Vertreter universitärer Projekte zur Datenzusammenführung, -vorhaltung und -langzeitspeicherung; Prof. Dr. Stephan Seidlmayer, Berlin-Brandenburgische Akademie der Wissenschaften in Berlin, als Vertreter der Akademien mit ihren Langfristprojekten u.a. im Bereich der Ägyptologie, der griechischen und lateinischen Epigrafik; Prof. Dr. Andreas Scholl, Direktor der Antikensammlung in den Staatlichen Museen Berlin, als Vertreter der Museen; Prof. Dr. Jürgen Kunow, Leiter des Rheinischen Amtes für Bodendenkmalpflege in Bonn und Vorsitzender des Verbandes der Landesarchäologen in der Bundesrepublik Deutschland als Vertreter der Landesdenkmalämter; Prof. Dr. Ortwin Dally, Generalsekretär des Deutschen Archäologischen Instituts in Berlin als größter außeruniversitärer Forschungseinrichtung auf dem Gebiet der Altertumswissenschaften mit Zweiganstalten im In- und Ausland sowie Feldprojekten auf allen fünf Kontinenten. Zur Vorbereitung haben unterschiedliche Veranstaltungen stattgefunden: DFG/ANH-Tagung der American School at Athens und des DAI im November 2009 in Athen; DFG-Rundgespräch der DFG-Arbeitsgruppe im Januar 2010 unter Leitung von Stephan Seidlmayer.

49 Zur Vorbereitung haben unterschiedliche Veranstaltungen stattgefunden: DFG/ANH-Tagung der American School at Athens und des DAI im November 2009 in Athen; DFG-Rundgespräch der DFG-Arbeitsgruppe im Januar 2010 unter Leitung von Stephan Seidlmayer.

len etwa ArcheoInf[50] als Frontend für GIS-Anwendungen, eAqua[51] mit seinen Tools für Text-Mining, eSciDoc[52] als Fedora basiertes Repositorium der MPDL, Propylaeum[53] als virtuelle Fachbibliothek der Altertumswissenschaften in Deutschland und Arachne als größte deutschsprachige Online-Datenbank zu antiken Objekten. Auch Software-Lösungen der Landesdenkmalämter sowie für Kartierungen die Geodatenportale von GDI-DE[54] mit den entsprechenden Länderlösungen sind hier zu nennen wie auch einige international betriebene Angebote, von denen vor allem der Archaeological Data Service[55], das e-depot voor de Nederlandse archeologie[56], OpenContext[57], The Stoa Consortium[58], Pleiades Project[59], Perseus Digital Library[60] und die Alexandria Digital Library[61] in den Altertumswissenschaften eine hohe Bedeutung besitzen und besonders fortschrittliche Ansätze vertreten. In einer zweiten Projektphase sollte dann vor dem Hintergrund einer dann spezifizierten Bedarfsanalyse ein Geschäftsmodell für das Kompetenzzentrum umgesetzt werden.

50 Vgl. ArcheoInf (2011).
51 Vgl. eAqua (2011).
52 Vgl. eSciDoc (2011).
53 Vgl. Propylaeum (2011).
54 Vgl. GDI-DE (2011).
55 Vgl. ADS (2011).
56 Vgl. EDNA (2011).
57 Vgl. OpenContext (2011).
58 Vgl. STOA (2011).
59 Vgl. PLEIADES (2011).
60 Vgl. Perseus (2011).
61 Vgl. Alexandria Digital Library (2011).