

E-Science-Tage 2019: Data to Knowledge

27. bis 29. März 2019 in Heidelberg

Details zum Beitrag

147

Art des Beitrags / Konferenztrack: Einreichung eines Posters

OpenDACHS: Ein Citation Repository zur nachhaltigen Archivierung zitierter Online-Quellen

Matthias Arnold¹, Hanno Lecher², Sebastian Vogt²

Organisation(en): 1: Heidelberg Research Architecture - Heidelberg Centre for Transcultural Studies, Universität Heidelberg; 2: Zentrum für Ostasienwissenschaften, Universität Heidelberg

eingereicht von: Matthias Arnold

(Universität Heidelberg, DE), ID: 1085

Themen: Fachspezifische oder disziplinübergreifende Herausforderungen, Institutionelle Lösungsansätze, Open Data, Open Science, Open Access, Datenformate und Standards, Datenarchivierung, Reproduzierbarkeit und Nachnutzung

Zusammenfassung

Das Internet ist seit dem Ende des vorigen Jahrhunderts eine zunehmend wichtige Plattform verschiedenster Publikationen sowie ganz allgemein des sozialen Diskurses. Die hier veröffentlichten Inhalte sind jedoch extrem flüchtig, da sie jederzeit inhaltlich verändert werden können oder ganz aus dem Netz verschwinden. Gleichzeitig sind sie jedoch für unterschiedliche Forschungsfragen von großer Bedeutung, und das Zitieren von Online-Quellen ist mittlerweile Alltag in der akademischen Forschung. Die Flüchtigkeit dieser Quellen und damit deren Nachprüfbarkeit ist zwar durch verschiedene Untersuchungen gut dokumentiert und belegt, die daraus resultierenden Folgen für die Forschungspraxis wurden aber bislang weitgehend ignoriert. Zwei Beispiele verdeutlichen die Problematik:

1. In Artikeln der Ausgabe 1.2010 der vom Exzellenzcluster Asia and Europe veröffentlichten Zeitschrift „*Journal of Transcultural Studies*“ wurden insgesamt 6 Online-Ressourcen zitiert, von denen heute nur noch 3 funktional sind. Die anderen 3 zitierten Quellen sind aus dem Netz genommen und nicht mehr nachprüfbar.
2. Für die Monographie *A Continuous Revolution* von Barbara Mittler (Cambridge, 2012) wurde im Webarchiv DACHS (*Digital Archive for Chinese Studies*) des Instituts für Sinologie ein Citation Repository angelegt. Obwohl von den zitierten 76 Internetquellen heute über 50% nicht mehr erreichbar sind, können via DACHS alle zitierten Quellen in ihrer ursprünglichen Form eingesehen und nachgeprüft werden.

In diesem Poster stellen wir eine Erweiterung des DACHS-Projekts vor: OpenDACHS. Das bisher auf sinologische Inhalte konzipierte DACHS wird derzeit in einer Kooperation zwischen dem Institut für Sinologie am Zentrum für Ostasienwissenschaften (ZO) und der Heidelberg Research Architecture (HRA) des Heidelberg Centre for Transcultural Studies (HCTS) zu *OpenDACHS* umgestaltet. Es wird als Service aufgebaut, funktional erweitert und zunächst für das neue Centre for Asian and Transcultural Studies (CATS) geöffnet.

Außerdem werden Arbeitsabläufe etabliert, die auch Fragen der Katalogisierung und der Verwaltung von Speicherplatz einbeziehen.

DACHS nutzt das Datenformat WARC, das als ISO Standard 28500:2017 veröffentlicht ist. Als Webcrawler kommt die Open-Source Software Heritrix zum Einsatz, die vom Internet Archive geschrieben wurde und heutzutage vielfach im Einsatz ist. Archivdateien können über Tools wie den Betrachter der *Wayback Machine* angesehen werden.

Zwar bieten frei zugängliche Services wie *WebCite* oder das Internet Archive an, einzelne Webseiten zu archivieren. Der Funktionsumfang dieser Services ist allerdings stark eingeschränkt. Open DACHS bietet dagegen die Möglichkeit, eine ganze Website oder mehrere Teile davon zu archivieren sowie regelmäßig wiederkehrende Archivierungen einzurichten, die Pfadtiefe frei zu definieren, sowie Seiten des "Deep Web" oder dynamische Inhalte zu berücksichtigen.