

41 www.clarin.eu/

42 Die Deutsche Nationalbibliothek (DNB), die Bayerische Staatsbibliothek, die SUB Göttingen sowie die SLUB Dresden sind bibliothekarische Mitglieder im Kompetenznetzwerk der Deutschen Digitalen Bibliothek; die DNB ist federführend an der Entwicklung von EuropeanA beteiligt; die SUB Göttingen federführend in den laufenden Projekten OpenAIRE und DARIAH.

43 Vgl. GESIS in den Sozialwissenschaften und das DKRZ für Klimadaten, a. a. O.

44 www.langzeitarchivierung.de/index.htm

45 www.wissgrid.de/workgroups/ap3.html

46 A. a. O.

47 Beispielfähig seien die Entwicklungen im Rahmen des BMBF-geförderten Projekts eAqua zur Informationsextraktion aus Quellen in den Altertumswissenschaften genannt, www.eaqua.net/

48 Das GCDH wird von den fünf geistes- und gesellschaftswissenschaftlichen Fakultäten der Universität gemeinsam mit der Staats- und Universitätsbibliothek getragen mit dem Ziel, Services, Forschung und Lehre in den IT-gestützten Geisteswissenschaften integriert zu entwickeln und anzubieten.

49 Z. B. das FIZ Karlsruhe.

50 Beginnend mit der nicht RAK-basierten Erfassung von Strukturmetadaten für Digitalisate und Beschreibungen von Internetquellen in virtuellen Fachbibliotheken ist hier auch personell die Expertise für das gemeinschaftlich mit der Deutschen Nationalbibliothek und der Max Planck Digital Library betriebene Kompetenzzentrum für Interoperable Metadaten (KIM) angesiedelt.

51 Diese Gruppe fasst die Aktivitäten des Universitätsverlags, der elektronischen Dissertationen, zu GoeScholar als institutionellem Repository und zahlreiche einschlägige Projektaktivitäten zusammen.

52 Für Mitarbeiter der Abteilung »Forschung und Entwicklung« der SUB wird diese Perspektive wohl noch in 2011 Realität werden, wenn sie in die geplanten Räumlichkeiten des Göttingen Centre for Digital Humanities (GCDH) mit einziehen werden, um dort an neuen Verbundprojekten wie DARIAH-DE oder Digital Humanities mitzuwirken.

53 www.libereurope.eu/

54 www.tilburguniversity.edu/research/institutes-and-research-groups/ticer/courses.html

55 www.wissenschaftsrat.de/index.php?id=344&L=

DER VERFASSER

Prof. Dr. Norbert Lossau ist Direktor der Niedersächsischen Staats- und Universitätsbibliothek Göttingen, Honorarprofessor Institut für Bibliotheks- und Informationswissenschaften, Humboldt Universität Berlin, 37070 Göttingen, Tel.: 0551 – 5212, Mail: lossau@sub.uni-goettingen.de

Kommentar

GERHARD LAUER

Kommentar aus Sicht eines Literaturwissenschaftlers

Virtuelle Forschungsumgebungen sind Modernisierungssphänomene. Wie alle Modernisierungen sind diese mit systemischen Irritationen etablierter Institutionen verbunden. Das gilt für die Bibliotheken, das gilt aber nicht weniger für die Wissenschaften. Und wie andere Modernisierungsprozesse auch, verläuft auch diese Modernisierung nicht linear und nicht in allen Bereichen mit derselben Geschwindigkeit. »Bibliothekare in die Forschung« – mit dieser Zuspitzung beschreibt der Beitrag von Norbert Lossau einen Modernisierungsschub, den virtuelle Forschungsumgebungen auf Bibliotheken haben und haben werden, ganz einfach weil sie die tradierten Funktionszuweisungen an die Bibliotheken verändern. Bibliotheken, darauf macht Norbert Lossau zu Recht aufmerksam, werden durch virtuelle Forschungsumgebungen zu Informationsinfrastrukturen, in denen Forschung stattfindet, genauer durch die vielfach erst Forschung möglich wird. Auch Bücher waren in dieser Hinsicht so etwas wie virtuelle Forschungsumgebungen, weil sie mit anderen Büchern in einer Systematik aufgestellt,

den Denk- und Arbeitsraum der Wissenschaften lange mitbestimmt haben. Aber die digitalen Forschungsumgebungen haben mehr als Bücher. Sie stellen neben die Bücher und Zeitschriften analytische Werkzeuge bereit, leiten neue Methoden an und modellieren die Untersuchungsobjekte. Sie eröffnen neue Formen der wissenschaftlichen Zusammenarbeit, die bisher bekannte räumliche Grenzen weitgehend überwindet. Schließlich kümmern sie sich auch um die Verbreitung und langfristige Vorhaltung der Forschungsergebnisse. Wenn Bibliotheken Häuser sind, die Bücher in unterschiedliche Räume anordnen, dann sind virtuelle Forschungsumgebungen Städte, die Häuser mit unterschiedlichen Funktionen in einer Stadt zusammenstellen.

Solche Städte gibt es bereits, etwa in der Astro- oder Teilchenphysik. Hier sind unter der D-Grid-Initiative im AstroGrid oder teilweise schon in Fachrepositorien wie arXiv.org die Beobachtungsdaten, Publikationen und im Grid inzwischen auch die Tools für die Forschung zusammengeführt.¹ Der Umbau ist in die



Foto: privat

Gerhard Lauer

virtuelle Forschungsumgebungen als Städte des Wissens

sen Fächern wie der Physik kein Zufall, sondern hängt mit einem Umstand zusammen, der gar nicht genug hervorgehoben werden kann: der Entstehung von datenintensiver Forschung.² Sie ist ein Kind des digitalen Zeitalters. Die gewaltigen Datenmengen, die hier die Forschung ausmachen, waren bis vor kurzem noch kaum vorstellbar, noch weniger, wie sie als Infrastruktur überhaupt zu organisieren wären. Der Large Hadron Collider in Genf produziert im Jahr selbst dann, wenn der größte Teil seiner Daten gar nicht archiviert wird, immer noch derzeit zehn Petabyte Daten, das sind 20 Millionen CDs übereinandergeschichtet, ein Datenberg höher als der Mont Blanc. DNA-Sequenzierautomaten, wie sie heute in den Lebenswissenschaften zum Einsatz kommen, analysieren Daten im Umfang von einem Terabyte pro Minute. Datenhäuser wie die GenBank führen solche Datenströme zusammen.³ Facebook erzeugt ca. 30 Milliarden Postings im Monat, die als Quelle für sozialwissenschaftliche Daten prinzipiell von Interesse sind.⁴ Und es bleibt nicht bei diesen Fächern stehen. In Disziplinen wie der Archäologie, deren Arbeitsalltag sich noch gestern nicht wesentlich von dem ihrer Kollegen vor hundert Jahren unterschieden hat, können jetzt ganze Grabungsgelände in 3D-Scannern aufgenommen und für die weitere Forschung zur Verfügung gestellt werden. In den Textwissenschaften entsteht mit TextGrid gerade die erste Grid-basierte virtuelle Forschungsumgebung in den Geisteswissenschaften, die korpusorientierte Forschung in einem ganz neuen Umfang möglich machen wird.⁵ In der Linguistik werden Forschungsergebnisse erarbeitet, die auf der Untersuchung von mehr als tausend Jahren Sprachdaten beruhen,⁶ und Google Books stellt mit Culturomics Tools für die Untersuchung von mehr als fünf Millionen Büchern zur Verfügung.⁷ Manche wie Tony Hey sprechen vom vierten Paradigma der datenintensiven Forschung, dem die Zukunft gehöre,⁸ andere wie der Chefredakteur des Magazins »Wired« gar vom Petabyte-Zeitalter, das Daten anstelle von Theorien setze.⁹ Das mag übertrieben sein, aber mindestens einzelne Fächer haben bereits die Schwelle zu den virtuellen Forschungsumgebungen überschritten, die datenintensive Forschung möglich macht. Wenn der Sloan Digital Sky Survey demnächst seine 1998 begonnene Durchmusterung des Himmels abgeschlossen haben wird und das auf diesen Beobachtungsdaten aufruhende Virtual Observatory an den Start geht, dann werden nicht Teleskope auf den Himmel gerichtet, sondern Tools auf schier unvorstellbar große Men-

gen strukturierter Daten.¹⁰ Damit wird Forschung erst in einer virtuellen Forschungsumgebung möglich.

Der Beitrag von Norbert Lossau macht darauf aufmerksam, dass dieser Umbau der Forschung durch virtuelle Forschungsumgebungen mehr als nur eine Erweiterung der Bibliotheken ist. Wenn Bibliotheken bis in die rechtlichen Dimensionen solche datenintensive Forschung ermöglichen sollen, das heißt dann auch Reputationsverteilung über Data-Sharing organisieren müssen,¹¹ werden sie selbst zu einem Teil der Forschung, weil an ihrer Strukturierung der Datenmengen und -typen wesentlich hängt, welche Forschung überhaupt möglich sein wird. Zu Recht macht er daher auf die besondere Rolle von Metadatenschemata aufmerksam, deren Entwicklung und Standardisierung Wissenschaftler und Bibliothekare zusammenführt, oder sollte ich nicht genauer sagen: Informationswissenschaftlerinnen und -wissenschaftler. Denn mit den klassischen Aufgaben hat und wird das immer weniger zu tun haben. Informationswissenschaftliche Forschungsabteilungen, die datenintensive Forschungsinfrastrukturen entwickeln und pflegen können, werden an vielen Bibliotheken zum Alltag gehören. Umgekehrt müssen nicht nur Bibliothekare in die Forschung, sondern Forscherinnen und Forscher in die Informations-Infrastrukturen. Der Rollenwechsel findet auf beiden Seiten statt. Das gehört zu dieser digitalen Modernisierung, an deren Anfang wir gerade erst stehen.

¹ www.d-grid.org.

² Vgl. das Sonderheft von »Science« (2011), www.sciencemag.org/site/special/data/

³ www.ncbi.nlm.nih.gov/genbank/

⁴ Z. B. D. Lazer, A. Pentland et al. (2009), Life in the network: the coming age of computational social science, *Science* 323 (5915), 721–723.

⁵ www.textgrid.org

⁶ E. Lieberman et al. (2007), Quantifying the evolutionary dynamics of language, *Nature* 449, 713–716.

⁷ www.culturomics.org

⁸ T. Hey et al. (2010), The Fourth paradigm. Data-Intensive Scientific Discovery, auch online: <http://research.microsoft.com/en-us/collaboration/fourthparadigm/>

⁹ C. Anderson (2008), The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, *Wired*, auch online: www.wired.com/science/discoveries/magazine/16-07/pb_theory

¹⁰ www.sdss.org

¹¹ B. Nelson (2009), Empty Archives, *Nature*, www.nature.com/news/2009/090909/pdf/46160a.pdf

DER VERFASSER

Prof. Dr. Gerhard Lauer ist Ordinarius für Deutsche Philologie an der Universität Göttingen, Seminar für Deutsche Philologie, 37073 Göttingen, Tel.: 0551 – 39 7526, Mail: gerhard.lauer@phil.uni-goettingen.de