

# **Daten-Adaptation als Analysemethode für geisteswissenschaftliche Forschung. Ergebnisse einer Untersuchung zur wissenschaftlichen Kommunikation in der modernen Physik\***

Martin Fechner

*Max-Planck-Institut für Wissenschaftsgeschichte und Berlin-Brandenburgische Akademie der Wissenschaften*

## **I. EINLEITUNG**

Die aktuellen Entwicklungen auf dem Feld der digitalen Geisteswissenschaften unterstützen die Forschung durch Digitalisierung [1], Benutzung neuer Werkzeuge [2,3], Datenanalyse [4] oder Visualisierungen [5].

Die vielfältigen Möglichkeiten, die sich einerseits aus einer computergestützten Datenhaltung ergeben, werden einem durch die datenintensive naturwissenschaftliche Forschung, etwa im Bereich der Klimawissenschaften [6], vor Augen geführt. Andererseits werden im Internet die Möglichkeiten einer computerunterstützten Datenpräsentation auf professionell gestalteten Webseiten sichtbar, dort wird etwa im Bereich des Datenjournalismus [7,8] zusätzlich eine Interaktion zwischen Rezipient und Datenpräsentation möglich gemacht.

Die Geisteswissenschaften stehen vor der Frage, welchen Nutzen sie aus den vorhandenen Techniken ziehen, ohne dabei den Anspruch aufzugeben, die Forschungsrichtung zu bestimmen, und sich in ihren Forschungsfragen auf solche reduzieren zu lassen, welche die Technik scheinbar vorgibt. Statt also digitale Werkzeuge mit einer bloßen Digitalisierung oder einem automatisch erzeugten Datenkorpus zu verbinden, wird hier ein alternativer Weg vorgeschlagen, der die geisteswissenschaftliche Forschungsfrage an den Anfang stellt, davon die Auswahl und Generierung von passenden Daten abhängig macht und so eine sinnvolle Analyse ermöglicht.

## **II. PROBLEM DER FORSCHUNGSDATEN**

Alle Formen von Daten beschreiben die Forschungsobjekte aus einer speziellen Perspektive. Bei der Auswertung der Daten muss die Forschungsfrage an die Perspektive der Daten angepasst werden, um eine sinnvolle Auswertung zu ermöglichen. Übernimmt die Forschung unreflektiert die vorhandenen Techniken, so macht sie sich letztlich von den vorgegebenen, begrenzten Auswertungsmöglichkeiten abhängig.

So werden in letzter Zeit vielfach Netzwerke erforscht, wohl auch da hier scheinbar einfache Visualisierungsmöglichkeiten bestehen, die beim Auslesen eines entsprechenden Datenkorpus direkt umgesetzt werden können. Vergleicht man aber einige Netzwerkvisualisierungen [5,9], wird ersichtlich, wie komplex die Darstellung tatsächlich ist und wie individuell der Prozess der Visualisierung begriffen werden muss.

Unter dem Stichwort „Big Data“ [10] werden wiederum Möglichkeiten gesucht, große, oft automatisch erzeugte Datenmengen sinnvoll für die Forschung auszu-

werten. Bislang hat sich aber in vielen Fällen noch kein geeigneter Weg gefunden, auf dem sich dieses Ziel sicher erreichen ließe. Als schwierig hat sich zum einen die Homogenisierung heterogener Datenmengen erwiesen, viel problematischer ist allerdings die Beschränktheit der Daten selbst. Es lassen sich nicht beliebige Forschungsfragen an jedem Datenbestand klären.

## **III. ANSATZ DIESER ARBEIT**

Im folgenden soll mit der Daten-Adaptation (engl. ‘data adaptation’) eine neue Methode vorgestellt werden, die vom Autor im Rahmen des wissenschaftshistorischen Dissertationsprojektes „Kommunikation von Wissenschaft in der Neuzeit“ [11] entwickelt wurde. Mit dieser Methode wird durch ein systematisches Vorgehen eine Datensammlung generiert, die einerseits im Einklang mit der Forschungsfrage steht und andererseits durch die Heterogenität und Wiederverwendbarkeit eine Vielzahl von Analysen möglich macht. Daten-Adaptation meint hier keine technische Schnittstelle, um die Austauschbarkeit zwischen verschiedenen Datenbeständen oder technischen Geräten herzustellen, sondern eine konzeptuelle Anpassung des Datenmodells an die eigene Forschungsfrage und die damit verbundene Datengenerierung und -analyse, die einen transparenten Umgang mit Forschungsdaten vorsieht und bisher unerforschbares und damit unentdecktes offenbaren kann.

Die folgende Demonstration basiert auf zwei Fallbeispielen, die vom Autor im Rahmen des genannten Dissertationsprojektes untersucht werden.

## **IV. METHODE**

In der hier vorgestellten Methode wird die Forschung durch eine Anpassung der Daten in mehreren Schritten begleitet. Zunächst werden (1) die Forschungsfragen den Forschungsgegenständen gegenübergestellt und es werden (2) durch die Einordnung in Beschreibungsmodelle die Qualitäten herausgearbeitet, die für die Beschreibung des Forschungsinteresses hilfreich sind. Mit der Entwicklung (3) eines darauf aufbauenden Datenmodells wird dann die Perspektive auf die Forschungsobjekte von der Forschungsfrage vorgegeben und (4) daran angepasste Daten können gesammelt werden. Die über die Objekte verfügbaren Informationen werden dem Datenmodell entsprechend in den Datenkorpus eingefügt. Die so bei der Quellenbetrachtung erhobenen Daten werden erst in der Forschung selbst generiert und stellen keine bloße Digitalisierung oder Kommentierung der Quellen dar. Diese

---

\* Eine detailliertere Beschreibung findet sich in der noch unveröffentlichten Dissertationsschrift des Autors [11].

Adaptation und Anpassung der Daten an die eigentliche Forschungsfrage kann einen reichhaltigen und vielschichtigen Datenkorpus erzeugen. Die anfangs formulierten Forschungsfragen lassen sich dann (5) schließlich explorativ oder (6) auch in Detailbetrachtungen erforschen.

#### A. Beispiel-Modellentwicklung

Im Rahmen der wissenschaftshistorischen Untersuchung werden Unterschiede und Gemeinsamkeiten der wissenschaftlichen Kommunikation im 19. und 20. Jahrhundert untersucht. Dabei werden für die Fallbeispiele der Entwicklung der Spektralanalyse und des Lasers die Fragen gestellt, wie aus Forschung im Labor anerkannte Entdeckungen wurden und welche Rolle die verschiedenen Medien spielten.

Aufbauend auf verschiedene Öffentlichkeitsmodelle konnte ein eigener Ansatz erarbeitet werden, in dem von Kommunikationsräumen gesprochen wird, die mit verschiedenen Eigenschaften ausgestattet sind. Ein Kommunikationsraum läßt sich hiernach etwa durch die Qualitäten Größe, Symmetrie, Organisationsgrad und thematische Ausrichtung beschreiben.

#### B. Datenmodell

Anschließend an den theoretischen Forschungsrahmen wurde ein Datenmodell entwickelt, mit welchem die zu untersuchenden Publikationsquellen beschrieben werden können. Das Datenmodell kombiniert dabei mehrere Anforderungen. Zunächst kann es die allgemeinen Angaben zu Autor, Titel, Jahr etc. als leicht verfügbare Basisbeschreibung festhalten. Daneben wurden eigene Kategorien gebildet, mit denen die Quellenobjekte dem Forschungsmodell zugeordnet werden können. Schließlich ist es mit dem Datenmodell auch möglich, Detailbeschreibungen ohne technischen Aufwand mitzunotieren. Da das Datenmodell einem Datenobjekt keine feste Anzahl von Eigenschaften zuordnet, wird mit einem XML-Schema [12] gearbeitet, das alle obigen Anforderungen erfüllt.

#### C. Datenerhebung und Quellenauswahl

Die Quellenauswahl wurde im Einklang mit dem Forschungsmodell vorgenommen und es wurden jeweils für die Fallbeispiele der Spektralanalyse und des Lasers die Buchpublikationen und wissenschaftlichen Zeitschriften der ersten zehn Jahre seit der entsprechenden Entdeckung untersucht. Bei der Datenerhebung wurde auf die bestehenden Daten aus verschiedenen Bibliothekskatalogen zurückgegriffen und eigene Informationen wurden ergänzt.

#### D. Analysevorgehen

An den erhobenen Datenbestand können durch Einsatz von XML-Techniken, wie etwa XSL-Transformationen, Reguläre Ausdrücke und Xpath, verschiedene Abfragen

gestellt werden, die die Resultate direkt oder unter Benutzung von Analyseprogrammen als Textausdrücke, Tabellen oder Graphen präsentieren. Natürlich sind nur Abfragen sinnvoll, die im Einklang mit der formulierten Forschungsperspektive stehen.

#### E. Ergebnisse

In den untersuchten Fallbeispielen werden durch die große Menge von über 1.500 untersuchten Publikationen viele Auswertungen der Daten möglich. So wird durch eine statistische Betrachtung die Unterscheidung der verschiedenen Medien Buch und Zeitschriftenartikel unterstrichen und zusätzlich wird auf eine Unterscheidung zwischen langen und kurzen Büchern hingewiesen. Die Auswertung zeigt auch eine medienabhängige Ausbreitung der Themen, indem sie Spezialisierungs- und Popularisierungsprozesse aufdeckt.

Weiterhin geben die Daten Aufschluß über geographische Verbindungen in der Verlagslandschaft der wissenschaftlichen Literatur und man kann neben einer festen Struktur auch die Veränderungen zwischen den Jahrhunderten nachweisen. Schließlich können die Daten implizite Verbindungen zwischen den Publikationen aufdecken und Hinweise auf Verhaltensweisen von Wissenschaftlern geben.

#### F. Probleme der Daten-Adaptation

Mögliche Probleme bei der Methode der Daten-Adaptation sind schlecht gebildete Kategorisierungen, die keine eindeutige Zuordnung zulassen, oder eine unzureichende Quellenbasis.

Ob das erste Problem auftritt und das gewählte Datenmodell nicht vollständig mit den Quellen vereinbar ist, kann schon bei der Erhebung der Daten schnell festgestellt werden. Empfehlenswert ist, das Modell an möglichst klar zu beschreibende Qualitäten der Quellen anzupassen. Bei einer möglichen und nicht aufzulösenden Uneindeutigkeit kann das Modell entweder in diesem Punkt eine offenere Beschreibung vorsehen oder es kann wie in sozialwissenschaftlichen Feldanalysen mit einem Reliabilitätsfaktor gearbeitet werden. Dem zweiten Problem kann nur durch eine umfangreiche Recherche oder einer gezielten Verengung des Untersuchungsgegenstand begegnet werden.

#### V. ZUSAMMENFASSUNG UND AUSBLICK

Die im Rahmen der Arbeit erhaltenen Ergebnisse demonstrieren, wie sich mit dem Ansatz der Daten-Adaptation neue Erkenntnisse durch Datenanalyse auch im Rahmen einer hermeneutischen und geisteswissenschaftlichen Arbeit gewinnen lassen. Durch ein mehrstufiges Verfahren, welches von fachlichen Überlegungen geleitet wird, können Forschungsfragen bearbeitet werden und implizite Strukturen offenbart werden.

Hier konnte gezeigt werden, wie sich dieses systematische Verfahren im Rahmen einer wissenschaftshistorischen Untersuchung anwenden lässt und wie sich neue Erkenntnisse und Zusammenhänge aufdecken lassen. Weitere Forschungen müssen zeigen, welche Erkenntnisse sich mit dieser Methode auch in anderen Forschungszweigen gewinnen lassen können. Eine genaue Ausformulierung der Methode folgt in der Dissertationsschrift.

## VI. LITERATUR\*

- [1] Digitalisierung und Digitale Sammlungen, Münchner Digitalisierungszentrum (MDZ) an der Bayerischen Staatsbibliothek. – <<http://www.muenchener-digitalisierungszentrum.de>>
- [2] DARIAH-DE Tools für die Geistes- und Kulturwissenschaften. – <<https://de.dariah.eu/tools>>
- [3] Werkzeuge und Dienste, CLARIN-D. – <<http://de.clarin.eu/de/sprachressourcen/werkzeuge-und-dienste.htm>>
- [4] Stéfan Sinclair und Geoffrey Rockwell, *Teaching Computer-Assisted Text Analysis: Approaches to Learning New Methodologies*, in: Digital Humanities Pedagogy: Practices, Principles and Politics, hg. v. B. D. Hirsch, Cambridge 2012. – <<http://dx.doi.org/10.11647/OBP.0024>>
- [5] Marian Dörk, Heidi Lam und Omar Benjelloun, *Accentuating Visualization Parameters to Guide Exploration*, in: CHI 2013: Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems, ACM, May 2013, S. 1755-1760. – <<http://mariandoerk.de/accentuation/>>
- [6] Klimawirkung und Vulnerabilität – Forschungsbereich II, Potsdam-Institut für Klimafolgenforschung. – <<http://www.pik-potsdam.de/forschung/klimawirkung-vulnerabilitat>>
- [7] Lorentz Mazat, Datenjournalismus, veröffentl. bei bpb.de am 26. Oktober 2011. – <<http://www.bpb.de/geellschaft/medien/opendata/64069/datenjournalismus>>
- [8] Stefan Plöching, Datenjournalismus und digitale Infografiken, veröffentl. im SZblog am 8. März 2013. – <<http://www.sueddeutsche.de/kolumne/datenjournalismus-und-digitale-infografiken-entdecken-sie-unseren-datagraph-1.1619138>>, <<http://sz.de/datagraph>>
- [9] Ulrik Brandes, Linton C. Freeman und Dorothea Wagner, *Social Networks*, in: Handbook of graph drawing and visualization, hg. v. R. Tamassia, London 2010. – <<http://nbn-resolving.de/urn:nbn:de:bsz:352-244311>>
- [10] „Big Data for the Humanities“ Workshop im Oktober 2013, während der 2013 IEEE International Conference on Big Data. – <<http://bighumanities.net/events/ieee-bigdata-2013/workshop-program/>>
- [11] Martin Fechner, Diss., Kommunikation von Wissenschaft in der Neuzeit, in Arbeit.
- [12] Tim Bray, Extensible Markup Language (XML) 1.0 (Fourth Edition) – Origin and Goals, veröffentl. vom W3C am 29. September 2006. – <<http://www.w3.org/TR/2006/REC-xml-20060816/#sec-origin-goals>>

\* Alle aufgeführten Webseiten wurden am 18.12.2013 abgerufen.