

ABIDA – Assessing Big Data

Vertiefungsstudie Wissenschaft – Workshopbericht (Eva Späthe)

Karlsruhe 15./16.11.2018

Datenintensive Wissenschaften – Theorie & Praxis–

Im Rahmen des ABIDA-Projekts fand vom 15.-16. November, in den Räumen des ITAS in Karlsruhe, der ABIDA Expertenworkshop unter der Überschrift „Datenintensive Wissenschaften. Theorie und Praxis“, statt.

Dabei wurden die Chancen und Herausforderungen durch die Forschung an und mit großen Datenmengen und ihrem Umgang innerhalb der Wissenschaft sowie die Auswirkungen von Big Data und Digitalisierung aus drei Perspektiven in den Blick genommen: I. Neue Formen der Logik und epistemische Opazität, II. Wissenschaft und Datenprovenienz sowie III. Herausforderungen in der Praxis.

Nachdem ABIDA in der ersten Projekthälfte durch Arbeitskreise, Fachtagungen und Bürgerkonferenzen einen umfassenden Überblick über die gesellschaftlichen Auswirkungen von und Entwicklungen durch Big Data erhalten konnte, wurde der ExpertInnenworkshop aufbauend auf den bislang erarbeiteten Ergebnissen in der zweiten Projekthälfte durchgeführt, in der es nunmehr darum geht, pointierte Positionen zu entwickeln, zu prüfen und zusätzliches Fachwissen zu gewinnen sowie Handlungsempfehlungen abzuleiten.

Der Themenkomplex ‚Datenintensive Wissenschaften‘ bietet ein breites Spektrum an zu diskutierenden Punkten: Die Datensammlung und -verwaltung, das Forschungsdatenmanagement, die Datenverarbeitung und die damit verbundene Selektion und Wertung oder auch der Kurationsaufwand um zu bearbeitbaren Daten zu kommen. Big Data bietet zweifelsfrei einen großen Rahmen an Möglichkeiten, allerdings sind Daten oft weit weniger vielseitig einsetzbar als in der Öffentlichkeit wahrgenommen.

Als ein wichtiger Aspekt wurde die Opazität als Gegenteil von Transparenz in der Wissenschaft bzw. innerhalb wissenschaftlicher Untersuchungen identifiziert. Bedingt durch die Verarbeitung großer Datenmengen und die eingesetzten Verfahren sei häufig nicht nachvollziehbar, wie genau von einem Input zu einem Output gelangt wurde. Ist dabei eine Überprüfbarkeit z.B. aufgrund der Komplexität oder Größe von Datenmengen nicht gegeben, ist eine wissenschaftliche Arbeit nicht mehr richtig bewertbar.

Oft vernachlässigt wird dabei der Aspekt des Datenmanagements. Nach den Teilnehmenden sei die wichtigste Herausforderung im Bereich Datenmanagement, das Bewusstsein für das Datenmanagement selbst. Im deutschen Wissenschaftssystem, aber auch in der Wirtschaft, sollte es viel prominenter behandelt werden, um eine größere Aufmerksamkeit sowie ein Bewusstsein zu schaffen. Ein durchdachtes Datenmanagement könne auch eine Lösung für Konflikte mit Verlagen sein. So plädierten die Teilnehmer für eine von Dritten unabhängige Lösung: Daten werden in von wissenschaftlichen Institutionen angebotenen Repositorien gespeichert, auffindbar gemacht und mit einem Rechtemanagement versehen.

Viele Probleme innerhalb der Wissenschaft, die im Zuge einer Big-Data-Debatte diskutiert werden, sind schon lange existent, werden jetzt aber verstärkt sichtbar. So ist das Verhältnis von Kausalität, Kontext und der Forschungsmotivation, welche ein hohes Maß an Reflexion über die eigene Forschung beinhalten muss, schon immer relevant, die Technisierung der Wissenschaft verstärkt diesen Aspekt aber noch.

Dazu gehört die Reflektion der wissenschaftlichen Werkzeuge, aber auch die Reflexion und Kritik der Wissenschaft selbst. Die Frage bei datenintensiven Wissenschaften ist darüber hinaus, wie Daten und die Methoden ihrer Verarbeitung reflektiert werden, hier gibt es einen normativen Klärungsbedarf.

Die Teilnehmer wiesen darauf hin, dass es in einigen (vorrangig naturwissenschaftlichen) Bereichen keine große Skepsis gegenüber den Werkzeugen gäbe. Man entferne sich in diesem Bereich von dem Ideal des Wissenschaftlers, der Wahrheiten über die Welt erforscht hin zu einem Wissenschaftler als Teil eines technologiegetriebenen Wissenschaftssystems. Big Data ermögliche es auch, bedingt durch die oft unterstellte Wertfreiheit der Verfahren, dass Wissenschaftler sich sozusagen hinter ihren Methoden verstecken. Darüber hinaus kann eine Standardisierung als

Begleiterscheinung der durch datengetriebene Wissenschaft auch zur Vernachlässigung führen, z.B. im Bereich Interkulturalität. Insgesamt müssen durch und mit datenintensive Wissenschaften viele grundsatztheoretische Diskussionen neu geführt werden, beispielsweise die nach dem Verständnis von Wahrheit oder Freiheit.

Die Teilnehmer wiesen darauf hin, dass immer berücksichtigt werden müsse, dass Datenerstellung nie wertneutral sei und oftmals aus einer praktischen Handhabung eine Methodologie entwickelt würde. Es gäbe zu wenige Theorieansätze, die prädiktive Modelle erstellen, das mache Modelle fehleranfällig. Eine antitheoretische Grundhaltung in der Wissenschaft führe zu einer erheblichen Erweiterung der Datenmengen. Zu beachten ist dabei, dass Big Data, nicht mit Smart Data gleichgesetzt werden kann. Außerdem bedeute Datenintensive Wissenschaft auch, dass Daten, oft bereits auf der Ebene der Sensoren, selektiert werden müssen, und die meisten Daten nicht verwertet werden. Von großer Relevanz ist die Frage wie sich zuverlässigen Voraussagen durch richtig gewählte Modellierungen oder Experimentalsysteme erzeugen lassen, wenn die Gewährleistung eines intellektuellen Verstehens nicht gegeben ist. So können Muster in Daten beschrieben, gleichzeitig aber die kausalen Beziehungen unverstanden bleiben. Als Grundbedingungen der Nichtwissengesellschaft beschrieben die Teilnehmer einen Umgang mit Nichtwissen und nicht den Versuch vorhandene Wissenslücken zu schließen.

Dabei gebe es spezifisches und unspezifisches Nichtwissen, bei dem der Suchraum nicht abgesteckt sei. So entstehen Grenzbereiche der (Natur)Wissenschaften, die sich durch eine starke Technisierung auszeichnen und in gleicher Weise auf klassische Methodik oder Theoriebildung verzichten. Es bleibt fraglich, wie sich diese Entwicklung in Bezug auf Transparenz und Methodenbildung weiterentwickelt. Auf den Aspekt der Smart Data beziehend ist der Punkt der Komplexitätsreduktion für und durch den menschlichen Verstand elementar.

Auch dieser ist aber nicht erst mit Big Data aufgekommen, er ist lediglich sichtbarer geworden. Das essentielle, nämlich das Verstehen, ist nach wie vor auf menschliche Beteiligung angewiesen. Maschinen aber sind bezüglich der Verarbeitungsgeschwindigkeit und der bewältigbaren Datenmengen unerreichbar und damit auch nicht nachvollziehbar, woraus sich unüberwindbare Differenzen ergeben. Zu spezifizieren wäre, welche Art des Wissens in der sog. Black Box verborgen bleibt (was passiert, wie passiert es etc.), sodass die Opazität entsteht.

Im wissenschaftlichen Umgang mit sozialen Medien besteht darüber hinaus das Problem der Bereitstellung der Daten die sich im Unternehmensbesitz befinden. Die Teilnehmenden tauschten sich darüber aus, wie diesem Datenmonopol zu begegnen ist. So seien zum einen politischer Druck, zum anderen aber auch eine Solidarisierung bzw. Verbündung der Wissenschaft denkbar. Da die politische Lage derzeit aber wenig aussichtsreich sei, wäre der Dialoge mit den Plattformbetreibern denkbar, der aber in ebenfalls wenig gewinnbringend scheine.

Zu diskutieren bleibt, ob das Anspruchsdenken an einer Partizipation der Daten gerechtfertigt ist, in anderen Bereichen ist die Akzeptanz eines Geschäftsgeheimnisses größer, was möglicherweise in den technischen Möglichkeiten begründet ist. Laut den Teilnehmern ist die Frage nach der Öffentlichkeit elementar, je nach Definition von digitaler Öffentlichkeit ist der Anspruch an den Daten begründet, offen bleibt dagegen, inwieweit persönliche Datenrechte und wirtschaftliche Interessen mit allgemeinen Forschungsinteressen korrelieren.

In Bezug auf neue Medien, aber auch andere Wissenschaftssektoren ergibt sich das Problem einer qualitativ schlechten Datenlage, die zwei Konsequenzen zulässt: Keine Forschung zum Thema oder eine Verbesserung bzw. eine Qualitätssicherung der Datensätze.

Die Teilnehmenden kamen überein, dass insgesamt eine breite forschungsethische Debatte empfehlenswert ist, die derzeit noch zu wenig stattfindet.

Dazu kommt, dass sich die datenrechtliche Lage lokal teilweise stark unterscheidet, sodass der Diskurs in den jeweiligen Forschungsgemeinschaften unterschiedlich ist. In Europa wird der Diskurs in besonders kritischer Weise geführt, was die unternehmerische Freigabe von Daten mitbestimmt.