

Bericht zur Tagung

Herausforderung Datenqualität – Forschung im digitalen Wandel, Hannover, 27.–28.02.2020

Es mag zunächst fast widersinnig erscheinen, dass eine Institution wie der Rfll zu einer [Konferenz zum Thema Datenqualität](#) einlädt, denn schließlich sollte es eine Selbstverständlichkeit sein, dass Wissenschaft, die den Namen verdient, auf Grundlage hochwertiger Daten stattfindet. Zur Perspektivierung lohnt es sich deshalb, den Untertitel „Forschung im digitalen Wandel“ zu bedenken und festzustellen, dass die Wissenschaft eine durch Digitalität bedingte Transformation durchlebt, deren Mühlen langsam mahlen. Ihre Tragweite im Alltagsgeschäft wird allzu leicht aus den Augen verloren: Wir sind live dabei und stecken mittendrin. Wenn wir das so akzeptieren, können wir auch die Bemühungen um die NFDI als ein Symptom des Transformationsprozesses auffassen, und dass in diesen Tagen die Vergabeentscheidungen der ersten Runde bekanntgegeben werden, können wir als Indiz dafür werten, dass wir dem Anfang dieses Prozesses deutlich näher sind als dem Ende.

Wenn am Ende dieses Wandels aber eine neue Dimension der Datenverfügbarkeit und daraus folgend bisher ungeahnte Möglichkeiten für die Forschung liegen, dann scheint es allemal angemessen, sich innerhalb zweier halber Tage mit dem titelgebenden Thema Datenqualität auseinanderzusetzen. Tatsächlich ist das Thema so umfassend und für alle Wissenschaftsbereiche relevant, dass es auch im Rahmen einer ganzwöchigen Veranstaltung mit mehreren Parallelsessions kaum abschließend behandelt hätte werden können. Es liegt deshalb wohl in der Natur der Sache, dass sich ihm im Rahmen der Plenarvorträge und einer Podiumsdiskussion nur exemplarisch anhand von Beispielen genähert werden konnte. Die präsentierten Disziplinen (Werkstoffforschung, Agrarwissenschaft, Klimaforschung, Ingenieurwissenschaften, Physik, Sozialwissenschaften, Lebenswissenschaften) lassen sich dabei durchaus als „datenintensiv“ beschreiben. Angesichts dieser thematischen Auswahl erscheint es als positiv, dass von Vortragenden – unter diesen die ehemalige Vorsitzende des Wissenschaftsrats Martina Brockmeier – betont wurde, dass eine Auseinandersetzung mit Daten und Datenqualität nicht dazu führen dürfe, dass es eine Unterscheidung in bessere und schlechtere Wissenschaftsbereiche gibt, die sich entlang der Datenfrage entwickelt.

Die Zusammenstellung der Themen kann gleichzeitig auch als Erinnerung verstanden werden, dass es bei den Gedanken um Datenqualität nicht nur um rein wissenschaftsinterne Überlegungen gehen kann, sondern dass in vielen Wissenschaftsbereichen auch eine wirtschaftliche Komponente mitgedacht werden muss. So war bereits im ersten Panel der Fraunhofer-Vorstand für Technologiemarketing, Ralf B. Wehrspohn, als Redner zu Wort gekommen. Noch vor diesem hatte aber auch der parlamentarische Staatssekretär im BMBF, Michael Meister, den möglichen

außerwissenschaftlichen Wert qualitativ hochwertiger Daten ausgedrückt, indem er die Hoffnung äußerte, dass „Data Made in Europe“ ein Gütesiegel im internationalen Wettbewerb werde.

Bedenkt man also, dass die Auseinandersetzung um Daten und Dateninfrastrukturen nicht in einem abgeschotteten Raum stattfindet, sondern dass dieser Debatte stets auch eine politische oder volkswirtschaftliche Dimension anhaftet, erscheint auch der Zeitpunkt der Veranstaltung gut gewählt. Fast zeitlich mit der Konferenz startete die Online-Konsultation zur Datenstrategie der Bundesregierung, die bis zum 3. April 2020 lief.

Vor dem Hintergrund eines Szenarios, in dem Datenqualität auch aus einer außerwissenschaftlichen Perspektive betrachtet wird, erscheint der im Kontext der Tagung als Herausforderung formulierte Punkt, dass in der Auseinandersetzung mit Datenqualität Themen- und Methodenvielfalt zu erhalten sind, fast als Warnung vor einer drohenden Fehlentwicklung. An anderer Stelle wurden Mängel der gegenwärtigen Wissenschaftsförderungspraxis klar benannt, etwa dahingehend, dass die bisherige Praxis der Projektförderung keine Zeit bzw. Mittel für eine langfristige Kuratierung der Daten vorsieht und die Möglichkeit, Projektfinanzierung speziell zu diesem Zweck zu verlängern, als Lösungsansatz aufgezeigt. Auch der Hinweis, dass der Aufbau und Betrieb von Qualitätssicherung mit extremen Kosten verbunden sei, und dass das niemand alleine schaffe, ist so treffend wie trivial, illustriert aber dennoch, dass die NFDI einen bestehenden Bedarf anspricht. Kaum strittig dürfte sein, dass die Herausbildung entsprechender kollaborativer Strukturen das genaue Gegenteil von trivial ist, doch die präsentierten Beispiele zeigen auch, dass es Disziplinen gibt, die den Herausforderungen mit Erfolg begegnen.

Die Verantwortung für den nötigen Wandel auf Geldgeber und Förderprogramme abzuschieben wäre jedoch der Tragweite der bestehenden Probleme nicht gerecht geworden, und so richtete sich der Blick auf den notwendigen Wandel auch nach innen. Es wurde gefordert, dass Daten- und Softwareprodukte als wissenschaftliche Leistungen anerkannt werden müssen, dass Datenqualität in den Strategien der Wissenschaftseinrichtungen verankert werden sollte, dass ein disziplinübergreifender Qualitätsdiskurs stattfinden soll, und dass Wissenschaftler*innen die Weiterentwicklung von Qualitätsstandards als eigene Aufgabe begreifen sollen – kurz gefasst: ein umfassender Wandel in der Wissenschaftskultur. Zugleich wurde angemahnt, dass obwohl eine allgemeine Ausrichtung des Forschungsdatendiskurses auf die FAIR-Prinzipien positiv ist, deren Einhaltung alleine nicht mit hoher Datenqualität einhergeht. Mit der Forderung des Kulturwandels hin zur Anerkennung von Datenleistungen erging auch eine Mahnung an die Wissenschaft: Die datenpflegerischen Fähigkeiten, die auf dem Weg zur Professur nicht anerkannt werden, würden bei anderen Arbeitgebern wie Google hochgeschätzt, und die Abwanderung datenqualifizierten Personals stellt eine Konsequenz der gegenwärtigen Verhältnisse dar.

Grundsätzlich herrschte disziplinübergreifend Einigkeit darin, dass sich Datenqualität nicht absolut greifen lasse. Das aber von einigen Seiten aufgebrachte Konzept der Datenqualität als „fitness for use“ wurde von anderen stark zurückgewiesen – zwar existierten Unterschiede in der Eignung zur Nachnutzung, doch ließe sich das aus der Wirtschaft stammende Fitness-Konzept

nicht ohne weiteres auf die Wissenschaft übertragen. Bei einigen Vorträgen wirkte es, als erschöpfe sich Datenqualität im Wesentlichen in Dokumentation, und dass sie sich damit weitgehend mit guter wissenschaftlicher Praxis gleichsetzen ließe. An anderer Stelle wurden jedoch auch weiterführende Gedanken angebracht, die auf grundlegende Probleme hinweisen, wie Daten in einigen Disziplinen überhaupt entstehen. So sind zum Beispiel die durch Satelliten erhobenen Messdaten der Klimaforschung nicht vergleichbar mit dem Ablesen eines Thermometers, sondern das Produkt umfangreicher Berechnungen und als solche aufs Engste mit dem dahinterliegenden Code verknüpft.

Die Überlegungen der meisten Vorträge und Diskussionen bezogen sich stark auf die Forschung. Nur in einem Vortrag wurde die Frage von Datenqualität in Verbindung mit den Lehrcurricula gebracht, daneben wurde jedoch auch auf die Notwendigkeit der Grundlagenarbeit im primären und sekundären Bildungsbereich hingewiesen, und ein weiterer Teilnehmer betonte am Beispiel der Forschung zur Künstlichen Intelligenz die Notwendigkeit von fortwährender Weiterbildung, da die Entwicklungen in diesem Bereich so schnell vor sich gehen, dass Absolvent*innen wenige Jahre nach der Promotion mit dem bis dahin erworbenen Bildungsstand schon nicht mehr arbeitsfähig wären.

Da sich die anwesenden Fachwissenschaftler*innen dem Thema oft mit Blick durch ihre eigene disziplinspezifische Brille näherten, waren die abschließenden Äußerungen von Petra Gehring erfreulich allgemeingültig: Es gibt keine Patentlösungen, man kann Qualitätskontrolle bei Datenpublikation nicht analog zur Forschungspublikation denken, das Prinzip Kontrolle ist keine Lösung – „wenn Wissenschaft nur gut ist, wo sie kontrolliert wird, dann wäre sie nicht gut“ – und Qualität muss immer im Bezug zu Methoden stehen.

Die NFDI wurde nur verhältnismäßig selten direkt angesprochen. Auch hier fielen jedoch einige gewichtige, teils auch widersprüchliche, Aussagen von verschiedenen Anwesenden, so etwa, dass die NFDI mit den beschränkten ausgeschriebenen Mitteln für eine Beutegemeinschaft schlecht geeignet ist, was aber nicht bedeutet, dass die Konsortien in Zukunft keine wichtige Rolle zu spielen haben werden, schon alleine durch die Aufmerksamkeit die diese für die Thematik erzeugten. Von anderer Seite wurde betont, dass es von großer Bedeutung sei, dass die NFDI auch bei den kleineren bzw. nicht aktiv teilnehmenden Institutionen ankomme – die lokalen Strukturen würden bei allen Institutionen ihre Bedeutung behalten und nicht durch die NFDI ersetzt. Ebenfalls angemahnt wurde ein Zusammenwachsen auf Augenhöhe zwischen Wissenschaft und Infrastruktur. Und während von der einen Seite der Projektcharakter der NFDI hervorgehoben wurde, argumentierten andere, dass die NFDI gerade über diesen hinausgehen müsse, und dass die Konsortien im Zusammenspiel von Konkurrenz und Kooperation agil handeln sollen. Besonders prägnant erschienen in der Diskussion um die NFDI die Aussagen, dass es auch Irrwege geben werde und – noch viel weiter greifend – dass wir gegenwärtig überhaupt erst erleben, was Digitalität eigentlich bedeutet.

Verfasst von Matthias Fingerhuth