

# Ist kooperativ jetzt umsonst? Die Ausweisung von Datenautorenschaft als neue Form wissenschaftlicher Reputation zur Förderung offener Forschungsdatenkulturen

**Moeller, Katrin**

katrin.moeller@geschichte.uni-halle.de

Martin-Luther-Universität Halle-Wittenberg, Deutschland

## Vorbemerkung

Die Bedeutung von Forschungsdatenmanagement ist mittlerweile umfänglich in der Wissenschaftskultur angekommen. Die Vorteile der Open Data Sciences überzeugen schnell, auch wenn die jüngst vom Rat für Informationsinfrastrukturen (RfII) pointiert formulierten Herausforderungen (RfII 2016, RfII 2017) weiterhin bestehen. Sie betreffen vor allem den Umgestaltungsprozess hinsichtlich des Aufbaus einer Landschaft von dauerhaften, stabilen Infrastrukturangeboten der Langzeitarchivierung, die durch die Anpassung und Koordination von Fördermechanismen, Personalentwicklung, Qualitätssicherung und der Entwicklung einer neuen "Forschungsdatenkultur" geprägt sind (AG Datenzentren 2017: 2-3). Den letzten Punkt möchte ich mit meinem Vortrag aufgreifen und aus dezidiert geisteswissenschaftlicher Perspektive einer datenproduzierenden und -bewahrenden Institution einen Beitrag zum Forschungsdatenmanagement formulieren, der meines Erachtens bisher kaum reflektiert wird: Es geht um die Entwicklung neuer Forschungsdatenkulturen, um Anreize und vor allem darum, welche wissenschaftliche Reputation sich mit der Nachnutzung von Forschungsdaten für den Urheber von Daten verbindet. Welche Anreize können Forschenden geboten werden, Daten tatsächlich zur möglichst flexiblen Nachnutzung frei zu geben? Diskutiert wird dazu nicht die Datenautorenschaft selbst, die über das Urheberrecht längst etabliert ist (Beer u.a. 2014: 4) und die heute bereits durch die Veröffentlichung von Forschungsdaten auf einem Forschungsdatenrepositorium über Lizenzen geregelt wird (CreativeCommons 2017, Beer u.a. 2014: bes. 24f.). Hier geht es vielmehr um eine Diskussion, welchen weiteren Weg die Datenautorenschaft nach der Veröffentlichung in einem anerkannten Repositorium nimmt und wie Datenautorenschaft innerhalb der Wissenschaft mehr Anerkennung, Sogkraft und

Reputation entfalten. Grundlegend kann dies zu einer größeren Bereitschaft von Forschenden zum Data-Sharing beitragen.

## Was beinhaltet ausgewiesene Datenautorenschaft?

Momentan werden alle Urheber eines wissenschaftlichen Forschungsergebnisses in Textpublikationen unabhängig vom inhaltlichen Beitrag als Autoren genannt. Es ist sinnvoll, hier ein Unterscheidungskriterium einzuführen. Dabei wird strikt nach den eigentlichen Textautoren und allen anderen Beiträgern eines wissenschaftlichen Forschungsergebnisses unterschieden. Textautoren sind ausschließlich diejenigen, die maßgeblich den Inhalt eines Textes verfassen, der das Forschungsergebnis und die Analyse repräsentiert. Sie sind für die Inhalte des Beitrags verantwortlich. Neben diesen Autoren für den Text werden dann alle weiteren Autoren als "Datenautor(en)" oder "Datengeber" geführt. Sie haben ebenfalls am Zustandekommen des Forschungsergebnisses wesentlichen Anteil. Diese Beteiligung kann bspw. darin bestehen, prinzipiell urheberrechtlich geschützte Daten über Lizenzen offen für eine Analyse zur Verfügung zu stellen. Wichtig ist dabei, dass die Daten des Datengebers - anders als beim Zitat - in einem wesentlichen Umfang (mehr als ein Drittel) Verwendung finden. Es ist dahingehend nicht entscheidend, ob die Daten im Datenkonvolut des Datennutzers auch einen wesentlich Bestandteil bilden oder letztlich nur einen kleinen Baustein ausmachen.

Neben die Autoren, Herausgeber und Übersetzer würde also eine weitere qualitative Gruppe der in einer Textpublikation genannten Autoren treten, die nun allerdings explizit keinen eigenen Beitrag am Text, wohl aber einen unmittelbaren, wesentlichen Beitrag zur Quellenbasis eines Forschungsergebnisses leisten. Dieses Kriterium ist das wesentliche Unterscheidungsmerkmal zum Textautor. In den Metadaten einer Publikation würden diese Autoren analog zu den Herausgebern eines Sammelwerks mit einem geeigneten Kürzel benannt (DA) und ggf. auch separat ausgewiesen. Für den Datenautor zählt diese Nennung aber als weitere Publikation, wenn auch mit einem abgestuften Renommee.

## Wozu brauchen wir extra ausgewiesene Datenautorenschaft?

### a) Vergleich verschiedener Methoden zur Messung wissenschaftlicher Leistung

Es ist relativ einfach zu zeigen, warum viele WissenschaftlerInnen zögerlich bleiben, Daten zu teilen. Vor allem in der Geisteswissenschaft gibt es dafür eine

Reihe von Gründen, von denen hier nur einige wenige knapp skizziert werden sollen:

- Datenproduktionen in den Geisteswissenschaften sind kosten-, zeit- und personalintensiv, für Datenbereinigung und Dokumentation müssen zusätzliche Aufwendungen gemacht werden
- Erschließungsprozesse von Medien und Quellen sind Teil des wissenschaftlichen Forschungsprozesses mit eigener Fachdisziplin (z. B. den Grund- und Hilfswissenschaften in der Geschichtswissenschaft)
- Forschungsdaten unterstehen damit - vielleicht auch im Unterschied zu sensorgestützten Datenproduktion in einigen naturwissenschaftlichen Sparten - dem Urheberrecht, da sie über eine eigenständige wissenschaftliche Leistung mit ausreichender Schöpfungshöhe verfügen. Dies gilt umso mehr, wenn Forschungsdaten nicht nur die Wiedergabe einer einzelnen Quelle in Form von Transkriptionen repräsentieren, sondern über den Erschließungsprozess mit einer Vielzahl von Annotationen und editionskritischen Anmerkungen versehen werden oder komplexe Datenstrukturen einer ganzen Serie von Quellen kombinieren.

Zunächst möchte ich kurz einen Vergleich des Umgangs mit der Datenautorenschaft in den verschiedenen Forschungsdisziplinen anstellen. Grundlegende Unterschiede liegen in der Messbarkeit der Forschungsleistung über mehr oder weniger umstrittene Methoden der Auszählung von Zitationsraten, die in den STM-Fachdisziplinen (Naturwissenschaften, Technik, Medizin) eine etablierte Forschungspraxis repräsentieren. Abgesehen davon, dass der Journal Impact Factor auch in den STM-Fächern keineswegs unumstritten und grundsätzlich davon auszugehen ist, dass Daten in Repositorien wie auch in Open Access-Veröffentlichungen durch die bevorzugten Evaluationsmetriken nicht gleichwertig erfasst werden (Herb 2010, Kap. 1.3), ist dieses Verfahren in den Geisteswissenschaften bisher grundsätzlich nicht anwendbar (Wissenschaftsrat 2006: 48ff. Jehne 2009: 59). Die wissenschaftliche Reputation durch die Zitation von Daten greift in den Geisteswissenschaften daher nicht in der gleichen Weise, wie in den STM-Fächern. Auch in den Naturwissenschaften entstehen durch die Vervielfachung der Co-Autorenschaft erhebliche neue Unschärfen bei der Leistungsbewertung von wissenschaftlichen Ergebnissen, was wiederholt zu Reglungsbedarf der Hochschulen und der DFG führte (DFG 2013: 20). Eine Trennung von Text- und Datenautorenschaft auf der Ebene der Metadatenhaltung brächte hier wesentliche Vorteile für alle Wissenschaftsdisziplinen, da sie Beiträge an Forschungsergebnissen durch klarere Definitionskriterien wieder transparenter ausweisen würde.

## b) Zitation versus Datenautorenschaft

Im zweiten Teil möchte ich durch einen Vergleich von Zitation und Datenautorenschaft die Vor- und Nachteile beider Prinzipien diskutieren und überlegen, wie man die Vorzüge beider Verwendungsweisen miteinander kombinieren könnte.

Das Grundprinzip des Zitats ist die Belegfunktion. Ein Zitat ist nach dem Urheberrecht dann zugelassen, wenn es eigene Ideen oder Gedanken unterstützt bzw. Ideen anderer aufgreift und in den eigenen Text integriert. Ist dem Recht genüge getan, verliert nach einer Faustregel der Text auch ohne das Zitat nicht an Sinn (Schwenke 2011). Das Urheberrecht regelt im § 52a, dass in der wissenschaftlichen Forschung "kleine Teile eines Werkes, Werke geringen Umfangs sowie einzelne Beiträge aus Zeitungen oder Zeitschriften" für die wissenschaftliche Forschung genutzt werden können (Bundesministerium 2013, § 52a). Auch wenn es kein festgesetztes Limit für den Umfang eines Zitats gibt, sollte es klar begrenzt sein. Als Faustregel führt ein Zitat nicht mehr als ein Drittel eines Textes auf (Schwenke 2011). Die Verwendung großer Teile von Daten oder ganzer Datenkonvolute zur Produktion eines Forschungsergebnisses verlässt den Rahmen einer Zitation deutlich.

Der wesentliche Vorteil des Zitierens ist neben seiner festen Kanonisierung in allen Wissenschaftsdisziplinen das Prinzip der Kontaktlosigkeit. Zudem wird klargestellt, dass nur der Autor inhaltliche Verantwortung für ein Forschungsergebnis trägt. Während Autoren sich untereinander abstimmen, Inhalte diskutieren und Rechte klären müssen, kann ein Wissenschaftler durch die Zitation Ergebnisse anderer unter den genannten Voraussetzungen in seine eigene Leistung einbinden und ausweisen. Die Autorenschaft ist daher organisatorisch aufwändiger und setzt die Erreichbarkeit des Urhebers voraus, wiewohl heute mit internetbasierten Referenzsystemen wie ORCID (ORCID 2017) Voraussetzungen dafür geschaffen werden.

Letztlich verzichtet der Autor durch die Lizenzierung von Forschungsdatensätzen in Forschungsrepositorien weitgehend auf alle Nutzungs- und Verwertungsrechte zugunsten einer möglichst breiten Nachnutzung von Daten. Neben anonymen Lizenzen ist der meistgewählte Typ an Lizenzen vermutlich die Auflage zur Namensnennung, die über das Zitat erfolgt. Dies ist möglich, indem Datenrepositorien die Freistellung der Daten von allen Verwertungsrechten nach § 15 des Urheberrechtsvertrages in ihre Nutzungsverträge übernehmen. Während Wissenschaftler und Wissenschaftlerinnen mit solchen Formen der Freistellung von Daten oft keine größeren Probleme haben, möchten sie aber weitgehend nicht auf die wissenschaftliche Verwertung ihrer Daten und die damit in Verbindung stehende wissenschaftliche Reputation verzichten. Sie stellen die eigentliche Währung der Wissenschaft dar.

Zudem möchte kein Datenautor für die Verwendung seiner Daten verantwortlich gemacht werden: Neben der formalen Trennung von Daten- und Textautor, kommt auch die inhaltliche zum Tragen. Das Forschungsergebnis verantwortet hier nur der Textautor.

Daher sollten Lizenzmodelle in Forschungsrepositorien idealerweise die Vorteile der Autorenschaft (wissenschaftlicher Mehrwert) und die Vorteile des Zitierens (Kontaktlosigkeit, keine rechtliche Regelung, Klarstellung der Autorenschaft) miteinander verbinden. Möglich wäre dies, indem Datenrepositorien die Datenautorenschaft von allen Rechten freistellen (wie dies in Lizenzen der Fall ist) aber entsprechend einer Verwendung von Daten in großen Teilen die Nennung als Datenautor oder Datengeber vorschreiben. Letztlich würde sich an der bisher diskutierten Praxis der Repositorien nichts Wesentliches ändern, mit Ausnahme der veränderten Notation im Sinne eines Autors des wissenschaftlichen Ergebnisses.

Datennutzer müssten lediglich zur Meldung von Veröffentlichungen beim Repositorium verpflichtet werden. Der Datengeber kann sich auf diese Weise bei Wunsch über eigene Veröffentlichung informieren.

Die Unterscheidung in Text- und Datenautorenschaft ist damit vor allem eine transparente Präzisierung der einzelnen Forschungsleistungen, die rechtlich über die Nutzungsverträge der Datenrepositorien Regelung erfährt. Dieses Verfahren findet bei jenen Daten Anwendung, deren Lizenzen eine freizügige, verschneidbare Nutzung ermöglichen. Auch bei anderen Daten ist natürlich eine Trennung von Datenautorenschaft und Textproduzent möglich, hängt jedoch durch die urheberrechtlichen Beschränkungen nach wie vor an einem persönlichen Austausch der beteiligten Personen.

## Bibliographie

**AG Datenzentren des DHd** (2017): Stellungnahme der DHd AG Datenzentren und des DHd-Verbands zur Nationalen Forschungsdateninfrastruktur (NFDI), URL: [http://dig-hum.de/sites/dig-hum.de/files/DHd\\_NFDI\\_Stellungnahme\\_2017-07-31.pdf](http://dig-hum.de/sites/dig-hum.de/files/DHd_NFDI_Stellungnahme_2017-07-31.pdf) [letzter Zugriff 22. September 2017].

**Bundesministerium der Justiz und für Verbraucherschutz** (2013): Gesetz über Urheberrecht und verwandte Schutzrechte, in: *Juris*, URL: <https://www.gesetze-im-internet.de/urhg/index.html#BJNR012730965BJNE023100377> [letzter Zugriff 22. September 2017].

**Deutsche Forschungsgemeinschaft** (2013): Sicherung guter wissenschaftlicher Praxis. Denkschrift, Empfehlungen der Kommission "Selbstkontrolle in der Wissenschaft", Weinheim 2013, URL: [http://www.dfg.de/download/pdf/dfg\\_im\\_profil/reden\\_stellungnahmen/download/empfehlung\\_wiss\\_praxis\\_1310.pdf](http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf) [letzter Zugriff 22. September 2017].

**Martin Jehne** (2009): Publikationsverhalten in der Geschichtswissenschaft, in: Alexander von Humboldt Stiftung (Hg.), *Publikationsverhalten in unterschiedlichen wissenschaftlichen Disziplinen. Beiträge zur Beurteilung von Forschungsleistungen 12*, 2. erw. Auflage, S. 59-61,

URL: [https://www.humboldt-foundation.de/pls/web/docs/F13905/12\\_disk\\_papier\\_publikationsverhalten2\\_kompr.pdf](https://www.humboldt-foundation.de/pls/web/docs/F13905/12_disk_papier_publikationsverhalten2_kompr.pdf) [letzter Zugriff 22. September 2017].

**Nikolaos Beer, Kristin Herold, Wibke Kolbmann u.a.** (2014): Datenlizenzen für geisteswissenschaftliche Forschungsdaten. Rechtliche Bedingungen und Handlungsbedarf, in: *GOEDOC: Dokumenten- und Publikationsserver der Georg-August-Universität Göttingen*, Göttingen, URL: <http://webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2014-6.pdf> [letzter Zugriff 22. September 2017].

**Rat für Informationsinfrastrukturen (RfII)** (2016): Leistung aus Vielfalt, Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland, Göttingen, URL: <http://www.rfii.de/de/category/dokumente/>.

**Thomas Schwenke** (2011): Texte richtig zitieren, statt plagiiere (Anleitung mit Checkliste), in: *ILAWit. Blog zum Social Media-, Marketing, Online- und Datenschutzrecht*, Berlin, URL: <http://rechtsanwalt-schwenke.de/texte-richtig-zitieren-statt-plagiiere-anleitung-mit-checkliste/> [letzter Zugriff 22. September 2017].

**Ulrich Herb** 2010: Open Access, zitationsbasierte und nutzungsbasierte Impact Maße: Einige Befunde. Tagungsband der 12. Tagung der Deutschen ISKO, (International Society Knowledge Organization), URL: <http://scidok.sulb.uni-saarland.de/volltexte/2010/3307> [letzter Zugriff 22. September 2017].

**Wissenschaftsrat** 2006: Empfehlungen zur Entwicklung und Förderung der Geisteswissenschaften in Deutschland, Köln, URL: <http://www.wissenschaftsrat.de/download/archiv/geisteswissenschaften.pdf> [letzter Zugriff 22. September 2017].

**ORCID** (2017): Distinguish yourself in three easy steps, URL: <https://orcid.org/> [letzter Zugriff 22. September 2017].