

# Voraussetzungen für die Nutzung von Forschungsdaten

René Schneider, Haute école de gestion, Genf (HES//SO)

*Forschungsdaten und ihre Nachnutzung werden die akademische Landschaft in den kommenden Jahren umgreifend verändern. Dies jedoch nur, sofern an den richtigen Stellen die richtigen Voraussetzungen geschaffen werden. Dabei haben zwei Elemente sowohl eine Hebel- als auch eine Scharnierfunktion: Metadaten und persistente Identifikatoren.*

Der Begriff der «Nutzung» von Forschungsdaten lässt auf den ersten Blick vermuten, dass es sich allein um die Nutzung aktueller, häufig spricht man auch von aktiven Forschungsdaten handelt, also all jener Daten, die die Forscher in ihrer täglichen wissenschaftlichen Arbeit durch Messung, Beobachtung, Modellierung, Anreicherung, Ableitung oder ganz einfach Digitalisierung erzeugen. Diese Daten werden dann, wenn man von einem Idealbild ausgeht, mit anderen interessierten Forschern oder der Allgemeinheit, die in der Regel die finanziellen Mittel zur Verfügung stellt, geteilt oder ausgetauscht; ganz im Sinne des Open Access.

Parallel dazu gibt es eine – je nach Perspektive – zumindest gleichberechtigte Betrachtungsweise, die eher auf die «Nachnutzung» der Forschungsdaten ausgerichtet ist und vom Gedanken der Langzeitarchivierung geleitet wird.

## Nutzung und Nachnutzung

Geht es in beiden Fällen um die Bereitstellung von Forschungsdaten, unterscheiden sich diese zwei Perspektiven jedoch um einen ganz wichtigen Aspekt, jenen der Zeit nämlich. Der Einfachheit halber könnte man sagen, dass es im einen Fall um ein «So schnell wie möglich», im anderen Fall um ein «So lange wie möglich» geht; einmal darum, die Forschungsdaten so schnell wie möglich zur Verfügung zu stellen und zu teilen, ein anderes Mal darum, die Forschungsdaten so zu bearbeiten, dass sie möglichst lange aufbewahrbar und wiederverwendbar sind. Im einen Fall sprechen wir von einem Zeitpunkt, der möglichst schnell nach Ablauf des Forschungsprojekts beginnt, wenn nicht sogar schon – je nach Laufzeit des

Projekts – während des Projekts. Im anderen Fall um Zeiträume, die mindestens Jahrzehnte umfassen.

Trotz dieser Gegensätzlichkeit gibt es zugleich ein oder mehrere verbindende Elemente, unabhängig davon, zu welchem Zeitpunkt Forschungsdaten zur Verfügung gestellt oder benötigt werden. Dieser Zusammenhang wird häufig als das FAIRness-Prinzip bezeichnet. Dies besagt, dass Forschungsdaten auffindbar (Findable), zugänglich (Accessible), interoperabel (Interoperable) und wiederverwendbar (Re-usable) sein müssen. Die Daten müssen also ab dem Zeitpunkt der Publikation oder der Übergabe an ein Langzeitarchiv dauerhaft identifizierbar, referenzierbar und zitierbar sein ... und alles Weitere folgt darauf.

## Metadaten

Konkret liegen dafür zwei Instrumente bereit: die Metadaten sowie die – weniger geläufigen – sogenannten persistenten Identifikatoren (PID). Der Bereich der Metadaten, d.h. der Daten, die die eigentlichen Forschungsdaten beschreiben, lässt sich grob in technische und deskriptive Metadaten unterteilen, wobei technische Metadaten häufig automatisch erstellt werden. Problematischer sind die deskriptiven Metadaten, die Auskunft darüber geben sollen, was sich hinter den eigentlichen Daten verbirgt. Häufig können die Forscher sie ansatzweise selbst erstellen, sie bedürfen in der Regel aber der Nachbearbeitung seitens eines Datenkurators.

Diesem Datenkurator, dessen Berufsbild sich erst allmählich eigenständig zu manifestieren beginnt, obliegt es, die für eine (Nach-)Nutzung relevanten Daten so aufzubereiten, dass sie entweder möglichst schnell in entsprechenden Repositorien als FAIRe Daten zur Verfügung stehen oder dass sie so transformiert werden, dass sie auch noch während oder nach zehn oder mehr Jahren trotz einer bis dahin veränderten Hard- und Softwarelandschaft in diese nahtlos bzw. nach entsprechenden Migrationen eingefügt bzw. wieder ausgewertet werden können.

### Persistente Identifikatoren

Ähnlich komplex verhält es sich mit den persistenten Identifikatoren, die – wie eine ISBN für Bücher – eine eindeutige (oder mathematisch korrekt, eine eineindeutige) Zuordnung zwischen einem Objekt und einer Kennziffer erlauben. Diese PID werden an Daten, Personen, Organisationen u.v.m. vergeben. Was die Daten der Geisteswissenschaften betrifft, sind dies vor allen Dingen die Digitalisate und hier teils sehr feingranulare Bestandteile davon sowie die darauf verweisenden Ketten von Referenzen und Zitationen. Eine besondere Rolle bei den Personen spielen dabei neben den Forschern auch die reellen bzw. in der Realität sehr wirkmächtigen virtuellen Personen, von denen in den einzelnen Werken die Rede ist. Hierfür stehen erste Lösungen zur Verfügung, etwa DOIs, ARK/N2T, ORCID, ISNI usw. Die wissenschaftliche Landschaft aber ist noch weit entfernt von einer Vernetzung aller zu vergebenden PID. Zudem werden Mechanismen benötigt, die in der Lage sind, die Vielfalt der Metadaten und der Identifikatoren dynamisch zu verwalten, so dass es gelingen kann, den Forschern neben dem «Mehr an Arbeit», das das Forschungsdatenmanagement von ihnen verlangt, auch Anreize zu bieten, bspw. über die Berechnung eines h-Indexes für die Zitation ihrer Daten oder andere geeignete Altmetrics, etwa um die Ansichten oder Downloads bereitgestellter Datensätze nachzuweisen.

### Handlungsbedarf

Es zeigt sich auch, dass hierbei insbesondere Institutionen, die sich selbst über den Begriff der Perennität definieren oder deren Daseinszweck mit der Perennität der zu verwaltenden Objekte verbunden ist, also genauer Bibliotheken und Archive, bei der Vergabe und Verwaltung der PID eine entscheidende Rolle spielen können, sofern sie es denn möchten oder aber dazu in der Lage sind. Aber selbst darüber hinaus bedarf es in der Schweiz und hier insb. im Bereich der Geisteswissenschaften noch einer Reihe gemeinsamer Anstrengungen seitens aller mit der Forschung befassten Partner, damit Forschungsdaten schnellstmöglich und dauerhaft geteilt und nachgenutzt werden können.

---

### Zum Autor

#### René Schneider



Prof. ord. Dr. phil. René Schneider, M.A., ist Professor für Informationswissenschaft an der Haute école de gestion – HES//SO in Genf und Leiter des Masterstudiengangs Information Science. Seine Hauptinteressen gelten dem Forschungsdatenmanagement und der Nützlichkeit von Information.