

Forschungsdaten in den Agrarwissenschaften: Management und Publikation

Birte Lindstädt¹

Abstract: Der Beitrag zeigt auf welche Schritte notwendig sind und welche Aufgaben wahrgenommen werden müssen, um Forschungsdatenmanagement umfassend zu betreiben. Dazu wird der Lebenszyklus von Forschungsdaten herangezogen, der die Grundlage eines Datenmanagementplans darstellt. Für ausgewählte Phasen und Aufgaben im Lebenszyklus wird die Situation in den Agrarwissenschaften beleuchtet, der Sachstand dargestellt und auf erforderliche Schritte hingewiesen. Einleitend erfolgt ein Überblick über die Vorteile von Open Data und Data Sharing sowie über die Rahmensetzung der Forschungsförderer im Hinblick auf das projektbezogene Forschungsdatenmanagement.

Keywords: Forschungsdatenmanagement, Open Data, Data Sharing, Datenmanagementplan

1 Einleitung

Forschungsdaten sind „Data that are used as primary sources to support technical or scientific enquiry, research, scholarship, or artistic activity, and that are used as evidence in the research process and/or are commonly accepted in the research community as necessary to validate research findings and results [Co15].“ In jeder Disziplin muss auf Basis dieser grundlegenden Definition fachspezifisch festgelegt werden, was die jeweiligen Forschungsdaten sind. In den Agrarwissenschaften sind dies u.a. Messdaten (z.B. Sediment-Analyse-Daten, Temperatur-Angaben), Geodaten (z.B. GIS-Dateien), Genomdaten (z.B. Länge extrahierter Genome von Pflanzen), Interviews (z.B. mit Experten) oder Bilder (z.B. Fotos, Satellitenaufnahmen).

Im Sinne der „guten wissenschaftlichen Praxis“ sollen Forschungsdaten langfristig gesichert und offen zugänglich gemacht werden [De15]. Bei offen zugänglichen Forschungsdaten spricht man von Open Data. Open Data sind im Internet verfügbar, liegen in einem maschinenlesbaren Format vor und sind mit einer Lizenz versehen, die jedem erlaubt, sie zu nutzen und mit anderen zu teilen [Ca15]. Shared Data sind nicht notwendigerweise Open Data, wenn sie lediglich mit einem begrenzten Personenkreis, z.B. innerhalb eines Forschungsinstituts, geteilt werden.

Die wesentlichen Vorteile von Open Data liegen in der schnellen, weltweiten Verbreitung sowie in der Nachnutzbarkeit der Daten für weitere Forschungen. In den Agrarwissenschaften kann dies einen wichtigen Beitrag zur Lösung praktischer Probleme bei der

¹ ZB MED Leibniz-Informationszentrum Lebenswissenschaften, Abt. Digitales Publizieren / Forschungsdatenmanagement, Gleueler Str. 60, 50931 Köln.

Verbesserung der weltweiten Ernährungssituation leisten, indem Entscheidungen effektiver und effizienter getroffen werden können, Innovationen befördert und für alle zugänglich werden und Transparenz zu effektiveren Organisationsstrukturen führen [Ca15].

2 Rahmensetzung der Forschungsförderer

Sowohl auf nationaler Ebene als auch europäischer Ebene werden die Forderungen nach einem offenen Zugang zu Forschungsdaten stärker, wenn er derzeit auch noch nicht verpflichtend ist. Die Deutsche Forschungsgemeinschaft hat Leitlinien zum Umgang mit Forschungsdaten formuliert, laut denen bereits in der Projektplanung Überlegungen zum Forschungsdatenmanagement erfolgen, die Daten zeitnah bereitgestellt und für mindestens 10 Jahre archiviert werden sollen. Die Kosten für das Forschungsdatenmanagement können mit einem Antrag bei der DFG eingeworben werden [De15]. Auf europäischer Ebene hat die Europäische Kommission im Rahmen des aktuellen Förderprogramms Horizon 2020 einen sog. Open Data Pilot aufgelegt. Teilnehmende Projekte werden verpflichtet einen Datenmanagementplan aufzustellen [Eu15].

3 Initiativen für Forschungsdatenmanagement und Open Data in den Agrarwissenschaften

In Deutschland wird dem Thema Forschungsdaten in den Agrarwissenschaften erst seit relativ kurzer Zeit eine Bedeutung zugemessen. Sowohl der Deutsche Bauernverband als auch die Deutsche Agrarforschungsallianz beschäftigten sich beispielsweise im Rahmen ihrer Fachtagungen im Herbst 2015 mit „Big Data“ und den daraus resultierenden Chancen und Risiken für die Landwirtschaft. International wird das Thema seit einigen Jahren diskutiert und findet Niederschlag in der Gründung verschiedener Initiativen. Ausgangspunkt für die internationale Initiative Global Open Data for Agriculture and Nutrition (GODAN) war der G8-Gipfel 2012, auf dem sich die teilnehmenden Länder darauf verständigt haben, Agrardaten untereinander und mit afrikanischen Ländern auszutauschen. GODAN „works to support global efforts to make agricultural and nutritionally relevant data available, accessible, and usable for application worldwide“ [Gl15]. Innerhalb der weltweiten Initiative Research Data Alliance (RDA), in der sich seit Frühjahr 2013 Institutionen in einer „bottom up-Bewegung“ zusammenschließen, gibt es auch eine „Agricultural Data Interest Group“ unter deutscher Beteiligung.

4 Der Lebenszyklus von Forschungsdaten als Grundlage des Forschungsdatenmanagements

Der Lebenszyklus von Forschungsdaten besteht aus den Schritten Planung/Erstellung,

Auswahl, Ingest/Übernahme, Speicherung/Infrastruktur, Erhaltungsmaßnahmen und Zugriff/Nutzung [En13]. Er wird ergänzt durch übergreifende Aufgaben, die unabhängig von der Planung und Durchführung der einzelnen Schritte bewältigt werden müssen. Als übergreifende Aufgaben des Forschungsdatenmanagements gelten (Gesamt-) Organisation, Kostenberechnung und -träger, rechtliche Rahmenbedingungen, Beschreibung/Erstellung von Metadaten und Vergabe von persistenten Identifikatoren [En13]. An ausgewählten Aufgaben des Forschungsdatenmanagements soll im Folgenden die Situation in den Agrarwissenschaften dargestellt werden.

4.1 Speicherung/Infrastruktur

Die Speicherung von Forschungsdaten erfolgt derzeit zum größten Teil auf externen Datenträgern wie CDs, USB-Sticks, Festplatten etc. der einzelnen Forschenden bzw. der Institute. Ziel sollte es sein, Forschungsdatenrepositorien zu nutzen, die die „sichere“ Aufbewahrung von mindestens 10 Jahren und einen offenen Zugang im Sinne von Open Data gewährleisten. Dies können generische Datenrepositorien, wie z.B. zenodo.org, sein. Fachspezifische Repositorien bieten den Vorteil einer für die Agrarwissenschaften konzipierten Beschreibungsmöglichkeit von Daten und einer stärkeren Verankerung in der Fachcommunity. Das Angebot fachspezifischer Datenrepositorien in den Agrarwissenschaften ist derzeit noch relativ begrenzt. Die Food and Agriculture Organisation of the United Nations (FAO) bietet mit FAOSTAT, CountrySTAT und dem Global Information and Early Warning System on food and agriculture (GIEWS) entsprechende Repositorien. Von deutschen Forschungsinstituten werden u.a. Mansfeld's World Database of Agriculture and Horticultural Crops (Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK)) und openLandscapes (Leibniz-Zentrum für Agrarlandschaftsforschung (ZALF e. V.)) betrieben. Die Vernetzung der existierenden Repositorien im Sinne einer Verknüpfung der Daten sowie der Aufbau neuer Forschungsdatenrepositorien für die Speicherung weiterer bzw. aller Forschungsdatenarten in den Agrarwissenschaften sind wünschenswert.

4.2 Metadaten

Beschreibende Metadaten für Forschungsdaten stellen einen Schlüssel für die Erklärung und damit Nachnutzbarkeit von Daten dar. Bei der Speicherung von Forschungsdaten in den oben beschriebenen Repositorien kommt den Metadaten eine hohe Bedeutung zu. Eine Einheitlichkeit und damit die Möglichkeit zur Verknüpfung von Daten werden gewährleistet, indem Standards bei der Datenbeschreibung verwendet werden. Wichtige Standards in den Agrarwissenschaften sind der Thesaurus AGROVOC (der FAO) oder die Tools zur Generierung von Metadaten agINFRA (von Agricultural Information Management Standards AIMS) oder AgriMetaMaker (von FAO).

4.3 Persistente Identifikatoren

Persistente Identifikatoren wie der Digital Object Identifier (DOI), der Uniform Resource Name (URN) oder handle werden für digitale Objekte vergeben, um sie bei einer Publikation dauerhaft auffindbar und zitierbar zu machen. Dies ist vor dem Hintergrund, dass auch die Publikation von zitierfähigen Forschungsdaten den Forschenden einen Vorteil bringt, eine wichtige Aufgabe im Forschungsdatenmanagement. Für die Vergabe von DOIs für Forschungsdaten in den Agrarwissenschaften in Deutschland ist ZB MED zuständig.

5 Ausblick

Die Agrarforschung schenkt dem Forschungsdatenmanagement derzeit verstärkt Aufmerksamkeit. Eine kooperative Lösung der anstehenden Aufgaben ist erforderlich, da keine Institution sie allein lösen kann. Im Mittelpunkt stehen Aufklärungs- und Beratungsarbeit gepaart mit der Schaffung geeigneter Informationsinfrastrukturen.

Literaturverzeichnis

- [Ca15] Carolan, Liz et al.: How can we improve agriculture, food and nutrition with open data? 2015.
- [Co15] Consortia Advancing Standards in Research Administration Information (CASRAI), http://dictionary.casrai.org/Research_data, Stand: 12.11.2015.
- [De15] Deutsche Forschungsgemeinschaft: Leitlinien zum Umgang mit Forschungsdaten, 2015.
- [En13] Enke, H. et al.: Leitfaden zum Forschungsdaten-Management. Handreichungen aus dem WissGrid-Projekt, Verlag Werner Hülsbusch, Glückstadt, 2013.
- [Eu15] European Commission, Directorate-General for Research & Innovation: Guidelines on Data Management in Horizon 2020, 2015.
- [Gl15] Global Open Data for Agriculture and Nutrition, www.godan.info, Stand 12.11.15.