Datenlebenszyklus

Bei Einhaltung der FAIR Prinzipien ergibt sich ein geschlossener Kreislauf der Datennutzung. Eine typische Möglichkeit dies darzustellen ist der Forschungsdatenlebenszyklus. Ausgehend von der Planung des Forschungsvorhabens werden alle wichtigen Schritte abgebildet, die die Daten bis zur Nachnutzung der veröffentlichten Daten durchlaufen.



Bei der **Planung** eines Forschungsvorhabens mit Hilfe eines Datenmanagmentplans werden die weiteren Phasen bereits mit gedacht und beschrieben.

Die eigentliche Studie dient oft der **Erfassung und Dokumentation** der Daten. An dieser Stelle handelt es sich zumeist um Rohdaten, also Informationen, die aufgenommen, aber noch nicht **Auswahl und Analyse** unterlaufen sind.

Dieses Dokument stammt aus der Dokuwiki-Instanz mit dem Titel FDM-Wiki von DSC und EST, die vom Rechenzentrum der Technischen Universität Clausthal bereitgestellt wird.

Kern eines durchdachten Datenmanagements ist die **Archivierung** von Daten. Das kann lokal oder in einem geeigneten Repositorium - einem Online-Datenarchiv mit beschreibenden Metadaten - geschehen. Ein frei zugängliches Repositorium (auch „Repo“) für Daten aller Art ist Zenodo.



Die Datenpublikation in nicht-kuratierten relativ unpspezifischen Repos birgt das Risiko einer weniger guten Auffindbarkeit der Daten. Geo- und Umweltwissenschaftliche Daten können beispielsweise kostenfrei in PANGAEA oder dem BONARES Repositorium archiviert werden und erhalten eine spezifische DOI (Digial Object Identifier), welch jede Datenpublikation zweifelsfrei ausweist. Wichtige Schritte aus Sicht einer Datenkuration werden vom DCC (Digital Curation Centre) hier beschrieben und dargestellt.

In einem Repo gehen **Archivierung** und **Publikation** oft einher. Um die angesprochene DOI zu erhalten müssen die Daten frei zugänglich sein. Über die Archivierung hinaus gibt es idealerweise noch eine Langzeit-Archivierung, bei der Daten und Formate an über Jahrzehnte oder gar Jahrhunderte nachnutzbar sind.[1)]

Der Datenlebenszyklus schließt sich durch eine Daten-**Nachnutzung**. Diese setzt voraus, dass die Daten auffindbar, qualitativ hochwertig und gut beschrieben sind. Die Daten werden von Nachnutzer:innen oft in Repos gefunden und können dann über ihre DOI zitiert werden. Eine gängige Lizenz dafür ist CC-BY, was bedeutet, dass Daten nachgenutzt aber die Person, welche die Daten ursprünglich erhoben hat, genannt werden muss.

Ein zusätzlicher Schritt im Datenlebenszyklus ist ein phasenübergreifendes **Management**, zum Beispiel die nachvollziehbare Dokumentation von Änderungen im Datensatz [2)]. Als wichtiger Akteur wird auch **der Mensch** selbst gesehen, der idealerweise während das Datenlebenszyklus ein verantwortungsvolles Handeln beisteuert [3)]. Dazu zählt auch die Wahrung von datenschutzrechtlichen Bestimmungen, die beispielsweise die Anonymisierung von Daten mit einschließt.

Der reibungslose Datenlebenszyklus profitiert von der Einhaltung der FAIR-Prinzipien.

1)
Markus, K., Parmaksiz, U., Svoboda, N., Schmidt, M.; Meier, K. 2021: Data for Generations - Keeping data reusable for really long periods of time. Workshop Folien, online auf https://zenodo.org/record/5786303, Zugriffsdatum 2022-06-09

2)
Stobierski, Tim 2021: 8 Steps in the Data Life Cycle, online auf https://online.hbs.edu/blog/post/data-life-cycle, Zugriffsdatum 2022-06-09

3)
Wing, Jeanette M. 2018: The Data Life Cycle, online auf https://datascience.columbia.edu/news/2018/the-data-life-cycle/, Zugriffsdatum 2022-06-09

Dieses Dokument stammt aus der Dokuwiki-Instanz mit dem Titel FDM-Wiki von DSC und EST, die vom Rechenzentrum der Technischen Universität Clausthal bereitgestellt wird.

Direkt-Link:
https://fdm-wiki.est.tu-clausthal.de/doku.php/datenlebenszyklus

Letzte Aktualisierung: **12.05.2023 06:03**

Dieses Dokument stammt aus der Dokuwiki-Instanz mit dem Titel FDM-Wiki von DSC und EST, die vom Rechenzentrum der Technischen Universität Clausthal bereitgestellt wird.