

Woche 03 Anwendung: Forschungsdaten

Skript

Erarbeitet von

Ann-Kathrin Selker und Ania López

Lernziele	1
Interview	1
Quellen.....	4
Weiterführendes Material	5
Disclaimer.....	5

Lernziele

- Forschungsdaten definieren können
- Die FAIR-Prinzipien erklären können

Interview

Ann-Kathrin: In diesem Video geht es heute um Forschungsdaten und das Forschungsdatenmanagementkompetenzzentrum der HHU. Ich sitze hier mit Frau Dr. López, wenn Sie sich bitte kurz vorstellen können.

Dr. López: Ja gerne, vielen Dank. Mein Name ist Ania López, und ich koordiniere das FDM Kompetenzzentrum der Heinrich-Heine-Universität seit 2021.

Ann-Kathrin: Sie arbeiten also mit Forschungsdaten. Was genau versteht man darunter denn?

Dr. López: Mit Forschungsdaten sind die (digitalen) Daten gemeint, die während des Forschungsprozesses entstehen, oder dessen Ergebnis sind. Das können z.B. empirische Messdaten sein, Labordaten, Transkriptionen von Texten, Daten aus Korpora, Umfragedaten, Softwarecode, usw. Eine eindeutige Definition gibt es nicht, da Forschungsdaten je nach Fachdisziplin sehr unterschiedlich sind. Durch die hohe Heterogenität fallen Forschungsdaten auch in sehr unterschiedlichen technischen Formaten an. Das können Textdaten (strukturiert und nicht strukturiert), Bilddaten, Audiodaten sein, gerne auch in proprietären Formaten (z.B. Daten, die mit einem bestimmten Mikroskop erstellt werden).

Ann-Kathrin: Warum sind Forschungsdaten denn so wichtig und warum beschäftigt man sich mit dem Management von Forschungsdaten?

Dr. López: Forschungsdaten sind wichtig, um Forschungsergebnisse nachvollziehbar und reproduzierbar zu machen. Die Nachvollziehbarkeit von Forschungsergebnissen ist ein immanenter Bestandteil der guten wissenschaftlichen Praxis. Wenn Forschungsergebnisse auf der Auswertung von Forschungsdaten basiert, sollten diese Daten entsprechend dokumentiert sein. Bei der Reproduzierbarkeit geht es auch darum, manchmal recht aufwändige und teure Datenerhebungen nicht mehrmals machen zu müssen. So gibt es auch Forschungsdaten (z.B. Klimadaten), die man kein zweites Mal erheben kann. Um dies zu gewährleisten, versucht man, Forschungsdaten nach den sogenannten FAIR-Prinzipien zu handhaben.

Ann-Kathrin: Sie haben jetzt gerade die FAIR-Prinzipien erwähnt, was genau versteht man denn darunter?

Dr. López: Das Akronym FAIR steht für

F- findable

A – accessible

I – interoperable

R – re-usable

Einblendung FAIR-Prinzipien (Quelle [1])

Dabei handelt es sich wirklich um Prinzipien. In der Praxis bedeutet die „Übersetzung“ dieser Prinzipien, je nach Forschungsdatentyp, unter Umständen auch was sehr Unterschiedliches. Ich versuche mal, das an einem Beispiel festzumachen.

Wenn wir uns jetzt einfach F- findable angucken: Um Daten auffindbar zu machen, müssen Forschungsdaten mit sog. Metadaten versehen werden. Metadaten sind dann beschreibende Daten dieser Daten. Man kennt das im privaten Kontext, also z.B. wenn man Fotos oder Musikdateien hat. Diese Dateien haben oft Metadaten. Das kann das Entstehungsdatum sein oder der Entstehungsort, sowie der Autor/Künstler, ggf. auch Daten

des aufnehmenden Gerätes, also der Kamera enthalten, etc. Mit Hilfe dieser Metadaten, kann man die zugrunde liegenden Fotos oder Musikdateien in entsprechenden Programmen durchsuchbar machen, sortieren, werten etc. Man kann manchmal auch selbst Metadaten ergänzen (z.B. Tags). Die Art und die Tiefe der Metadaten helfen dabei, die ursprünglichen Daten „findable“ zu machen. Dabei ist aber die Art der Metadaten, die man z.B. für ein Mikroskopbild benötigt, andere, als die Metadaten, die man z.B. für Umfragedaten nutzt. Man spricht hier von unterschiedlichen und disziplinspezifischen Metadatenschemata.

Ann-Kathrin: Das hört sich ja sehr kompliziert an. Wie kann denn dabei das Forschungsdatenmanagementkompetenzzentrum der HHU helfen?

Dr. López: Das FDM Kompetenzzentrum unterstützt die Forschenden der HHU in allen Belangen rund um das Forschungsdatenmanagement. Wir bieten in diesem Kontext Beratungen, Schulungen und Tools zum Forschungsdatenmanagement an.

Einblendung FDM-Website (Quelle [2])

Beratung kann ganz unterschiedlich sein, also z.B. bei der Beantragung eines Projektes beraten wir, oder aber auch bei konkreten Fragen zum Umgang mit Forschungsdaten selbst. Schulungen bieten wir zu ganz allgemeinen Themen rund um das Thema Forschungsdatenmanagement für verschiedene Zielgruppen. Bei den Tools gibt es natürlich neben den zentralen Speicherdiensten auch beispielsweise Software für den Einsatz im Labor (sog. Elektronische Laborbücher).

Ann-Kathrin: Was für Tipps und Tricks haben Sie denn für Nachwuchswissenschaftler, die das allererste Mal mit Daten arbeiten? Worauf muss geachtet werden?

Dr. López: Ich denke das wichtigste ist, dass es eben nicht „das eine“ Forschungsdatenmanagement als Methode oder Schema F gibt; dennoch gibt es meistens immer irgendwo Optimierungsmöglichkeiten in der eigenen Handhabung der Daten. Wer kennt das halt nicht? Man arbeitet so vor sich hin, und der lästige Dokumentationskram fällt so ein bisschen hinten über. Unserer Erfahrung nach ist der beste Ratschlag so früh wie möglich anzufangen, sich zu überlegen, wie man mit den Forschungsdaten umgehen wird. D.h. vor allem frühzeitig die Forschungsdaten mit geeigneten Metadaten zu versehen. Wenn man nämlich ansonsten erstmal seine Forschungsfrage beantwortet hat, ist das nachträgliche Bearbeiten der Forschungsdaten unglaublich zeitaufwendig. Darüber hinaus sollte man natürlich auch immer das Thema Backup im Hinterkopf behalten. Also niemals Forschungsdaten nur an einer Stelle speichern. Wenn die Festplatte kaputt geht oder der USB-Stick verloren geht, sind die Forschungsdaten einfach weg. Lieber von vorne herein auf zentrale und gesicherte Speicherorte gehen.

Ann-Kathrin: Welche rechtlichen Hintergründe gibt es denn zu beachten?

Dr. López: Es gibt jede Menge rechtliche Fragestellungen rund um das Thema Forschungsdaten. Zum einen sind da natürlich die Belange des Datenschutzrechtes. Wenn Forschungsdaten personenbezogene Daten enthalten (das kann bei Patientendaten oder

Daten aus Interviews sein), unterliegen diese Daten natürlich besonderen Schutzpflichten, sie müssen also datenschutzkonform gehandhabt werden. Methoden der Anonymisierung und Pseudonymisierung spielen hier eine wichtige Rolle, aber auch die grundsätzliche Fragestellung, auf welchen Servern und ob sie überhaupt gespeichert werden können. Abseits vom Datenschutzrecht ist aber auch das Thema Nutzungs- und Lizenzrechte relevant, z. B. wenn Daten aus sozialen Netzwerken (also Twitter, Facebook, Instagram, ...) ausgewertet werden sollen, ist dies rechtlich nicht so ohne weiteres möglich. Selbst wenn die Daten technisch einfach abzugreifen sind, muss geklärt werden, ob das eben lizenzrechtlich möglich ist. Ebenso verhält es sich mit Daten aus Zeitungen/Zeitschriften, die online verfügbar sind. Darüber hinaus ist auch die Weiterverwendung von selbst erzeugten Daten rechtlich immer an Einzelfallbetrachtungen gekoppelt, also einfach nicht trivial.

Ann-Kathrin: Vielen Dank für das informative Gespräch. Welche weiterführenden Materialien können Sie denn noch für die Studierenden empfehlen?

Dr. López: Für alle diese Themen gibt es mittlerweile recht gute Informationsquellen. Ich persönlich kann die Seite forschungsdaten.info sehr empfehlen, aber natürlich auch unser Team vom FDM Kompetenzzentrum, also kommen Sie einfach gerne auf uns zu.

Einblendung Team des FDM Kompetenzzentrums der HHU (Quelle [3])

Ann-Kathrin: Vielen Dank!

Dr. López: Danke auch.

Quellen

- Quelle [1] Grafik: Paulina Halina Sieminska / Bearbeitung: Dr. Ilona Lang / CC BY-SA 4.0
<https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/faire-daten/>
- Quelle [2] <https://fdm.hhu.de>, angerufen am 15.02.23
- Quelle [3] FDM Kompetenzzentrum HHU

Weiterführendes Material

<https://www.fdm.hhu.de/>

<https://forschungsdaten.info/>

<https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/faire-daten/>

<https://www.forschung.hhu.de/wissenschaftliche-integritaet/grundsaeetze-guter-wissenschaftlicher-praxis>

Disclaimer

Transkript zu dem Video „Woche 03 Anwendung: Forschungsdaten“, Ann-Kathrin Selker und Ania López.

Dieses Transkript wurde im Rahmen des Projekts ai4all des Heine Center for Artificial Intelligence and Data Science (HeiCAD) an der Heinrich-Heine-Universität Düsseldorf unter der Creative Commons Lizenz [CC-BY](https://creativecommons.org/licenses/by/4.0/) 4.0 veröffentlicht. Ausgenommen von der Lizenz sind die verwendeten Logos, alle in den Quellen ausgewiesenen Fremdmaterialien sowie alle als Quellen gekennzeichneten Elemente.