

Ausbildungsangebote für Data Stewards an österreichischen Universitäten

Autor*innen:

Alexander Gruber, ORCID: <https://orcid.org/0000-0002-4244-3446>

Hermann Schranzhofer, ORCID: <https://orcid.org/0000-0003-0249-2726>

Sabrina Knopper, ORCID: <https://orcid.org/0000-0002-4086-4028>

Tereza Kalová, ORCID: <https://orcid.org/0000-0002-1764-7228>

Ilire Hasani-Mavriqi, ORCID: <https://orcid.org/0000-0003-0758-0805>

Institute of Interactive Systems and Data Science
Handlungsfeld Forschung – Programm Digitale TU Graz
Veränderungsprozesse und Umsetzung
TU Graz

Universitätsbibliothek Wien
Universität Wien

Kommentiert von den Mitgliedern des FAIR Data Austria Teams

DOI: <https://doi.org/10.3217/xcv1e-kc223>



Dieses Werk und dessen Inhalt steht - mit Ausnahme der Logos - unter der Creative Commons Namensnennung 4.0 International Lizenz. Um eine Kopie dieser Lizenz zu sehen, besuchen Sie bitte [Creative Commons Namensnennung 4.0 International Lizenz](https://creativecommons.org/licenses/by/4.0/)

Inhalt

Abstract	3
Deutsch.....	3
English	3
Einleitung	3
Beschreibung der drei Modelle.....	3
Kompetenzen von Data Stewards	4
Ausbildungsangebote	4
Zertifikatskurs Forschungsdatenmanagement.....	4
Masterstudium Digitales Datenmanagement	6
DataTrain: Training in Research Data Management & Data Science	8
Zertifikatskurs Data Steward	10
Ergebnisse der Gruppenarbeiten	12
Zertifikatskurs Forschungsdatenmanagement.....	12
Masterstudium Digitales Datenmanagement	12
DataTrain: Training in Research Data Management & Data Science.....	13
Zertifikatskurs Data Steward	13
Mapping – Data Steward Modelle vs. Ausbildungsangebote	14
Zusammenfassung.....	14
Annex	15
Detaillierte Ergebnisse der Gruppenarbeiten	15

Abstract

Deutsch

Mit den Erkenntnissen aus zwei vorangegangenen Workshops, welche sich mit den Modellen für und Kompetenzen von Data Stewards beschäftigt hatten, trafen sich wieder die „FAIR Data Austria“ Projektmitarbeiter*innen zu einem weiteren Workshop im Juli 2021 um Ausbildungsangebote für Data Stewards an österreichischen Universitäten zu diskutieren. Zu diesem Zweck wurden vier Ausbildungsprogramme von jeweils einer teilnehmenden Person vorgestellt und im Anschluss in einzelnen Gruppenarbeiten nach ihren Stärken und Schwächen, sowohl im Hinblick auf Eignung der gesamten Ausbildung für die Modelle als auch der einzelnen Schwerpunkte/Module innerhalb der Kurse, analysiert und bewertet. Dieser Workshop hat gezeigt, dass keiner der durchleuchteten Kurse geeignet ist, ein Modell vollständig abzudecken, wenn es darum geht, die gewünschten Kompetenzen zu vermitteln. Aus der Analyse resultieren Empfehlungen der Ausbildungsangebote für gewisse Modelle, welche nun gemeinsam mit den bisherigen Projektergebnissen in die Ausarbeitung eines Self-Assessment Tools zur Modellwahl für Forschungseinrichtungen fließen werden.

English

With the conclusions from two previous workshops, which dealt with the models for and competences of data stewards, the "FAIR Data Austria" project staff met again for another workshop in July 2021 to discuss training programmes for data stewards at Austrian universities. For this purpose, four training programmes were presented by one participant each and then analysed and evaluated in individual group work according to their strengths and weaknesses, both with regard to the suitability of the entire training for the models and the individual focal points/modules within the courses. This workshop showed that none of the courses examined is suitable to cover a model completely when it comes to imparting the desired competences. The analysis resulted in recommendations of the training courses for certain models, which will now be used together with the previous project results in the development of a self-assessment tool for model selection for research institutions.

Einleitung

Beschreibung der drei Modelle

Die drei Modelle für Data Stewards an österreichischen Universitäten wurden in einem Workshop im Herbst 2020¹ erarbeitet:

Modell 1: Anlaufstelle/Servicepoint – Ein Data Steward pro Forschungseinrichtung. Data Stewards werden als Schnittstelle für FDM-Fragen aufgefasst, nehmen Anfragen entgegen und leiten diese (z.B. an die Rechtsabteilung) weiter. Sie vernetzen und bieten allgemeine Beratung an. In dieser Rolle machen Data Stewards jedoch kein operatives Datenmanagement an den Instituten, was angesichts der Breite dieses Anforderungsprofils schwierig wäre.^{1 2}

¹ Reichman, S., Hasani-Mavriqi, I. (2021). Entwicklung eines Konzepts für Data Stewards an österreichischen Universitäten. <https://doi.org/10.25365/phaidra.243>

² Gruber, A., Schranzhofer, H., Knopper, S., Stryeck, S. und Hasani-Mavriqi, I. (2021) „Kompetenzen von Data Stewards an österreichischen Universitäten“, Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare, 74(1), S. 12–32. doi: 10.31263/voebm.v74i1.6255.

Modell 2: Data Steward Center/Office – Charakteristisch für dieses Modell ist eine mehr oder weniger zentrale Organisationseinheit, die unterschiedliche Kompetenzen bündelt und entsprechende Services/Beratungen anbietet; die Kompetenzen sind breiter, da auf mehrere Personen aufgeteilt.^{1 2}

Modell 3: Data Steward Netzwerk – Ein Data Steward pro Fakultät oder Field of Expertise + ein/e Koordinator/in. Das dritte Modell schließlich sieht an den Fakultäten oder Forschungsschwerpunkten angesiedelte Data Stewards vor, die über entsprechend fundiertes disziplinspezifisches Wissen/Forschungserfahrung verfügen sollten und daher besonders disziplinspezifisch beraten können und zentral koordiniert werden.^{1 2}

Kompetenzen von Data Stewards

Die Kompetenzen von Data Stewards an österreichischen Universitäten wurden in einem Workshop im April 2021² erarbeitet. Data Stewards benötigen neben ihren disziplinunabhängigen und gegebenenfalls disziplinspezifischen **Fachkompetenzen** (Ausbildung, Weiterbildung und Berufspraxis), auch weitere Kompetenzen, um ihre Aufgaben zu erfüllen. Dazu zählen **Methodenkompetenzen** (z.B. Projekt- und Konfliktmanagement, Moderationstechniken), soziale **Kompetenzen** (z.B. Kommunikations- und Motivationsfähigkeit, Hilfsbereitschaft) und **persönliche Kompetenzen** (z.B. Ergebnisorientierung, Lernbereitschaft, Anpassungsfähigkeit), die breit gefächert sind und vor allem für die kollaborative Zusammenarbeit mit Forschenden von großer Bedeutung sind.

Eine detaillierte Beschreibung der benötigten Kompetenzen von Data Stewards in den jeweiligen Modellen ist im Workshopbericht² enthalten.

Ausbildungsangebote

Im Vorfeld der Gruppenarbeit wurden vier Ausbildungsprogramme vorgestellt und analysiert.

Zertifikatskurs Forschungsdatenmanagement

Der Zertifikatskurs Forschungsdatenmanagement (FDM) an der Technischen Hochschule Köln startete erstmals im August 2021 und beinhaltet unter anderem FDM in verschiedenen Fachgebieten, Open Science, Beratung, Infrastruktur, Datenmanagement und rechtliche Aspekte.³ Der Kurs dauert ca. 10 Monate (13 Termine). Der Großteil davon (80 Stunden) ist als Online Veranstaltung vorgesehen und wird durch Selbstlernphasen (ca. 125 Stunden) ergänzt. Der gesamte Zeitaufwand wird mit 240 Stunden angegeben. Für das Zertifikat mit 8 ECTS kann eine Projektarbeit mit einem Aufwand von ungefähr 35 Stunden gemacht werden.

Die Zielgruppen sind sowohl Beschäftigte aus dem wissenschaftsnahen Infrastrukturbereich (z.B. Hochschulbibliotheken, Rechenzentren) als auch aus der aktiven Forschung (z.B. Data Stewards, Data Scientists). Vorausgesetzt werden eine abgeschlossene Berufsausbildung oder ein Studienabschluss, sowie ein Arbeitsverhältnis zu einer wissenschaftsnahen Einrichtung in Nordrhein-Westfalen. Zudem sind auch Kenntnisse im Umgang mit Forschungsdaten erwünscht. Die Teilnehmerzahl ist auf 15 Personen beschränkt was eine gute und auch individuelle Betreuung erwarten lässt. Die Kursgebühr beträgt 2.500,- Euro, es werden aber unter gewissen Kriterien Stipendien vergeben, die bis zu 80% der

³ Zertifikatskurs Forschungsdatenmanagement. https://www.th-koeln.de/weiterbildung/zertifikatskurs-forschungsdatenmanagement_82048.php (Aufruf am 24.09.2021)

Kursgebühr abdecken. Reise- und Hotelkosten für die Präsenztermine sind von den Teilnehmer*innen selbst zu tragen. Der Kurs ist nur als Gesamtpaket buchbar.

Der Inhalt ist in folgende 9 Module aufgeteilt:

Modul	Inhalt	Aufwand
Modul 1: Basismodul Grundlagen des FDM	<ul style="list-style-type: none"> • Didaktische Ansätze, Methoden und Seminaraufbau • Forschungsdaten – Lebenszyklus und Policy, DMPs, Struktur von Daten, Dokumentation, Speicherung und Backup, Langzeitarchivierung, Zugriffssicherheit, Publikation, Nachnutzung, rechtliche Aspekte 	2 virtuelle Präsenztage à 7,5 Stunden + 1 Präsenztag à 6 Stunden = 21 Stunden
Modul 2: Basismodul Open Science & rechtliche Aspekte	<ul style="list-style-type: none"> • Darstellung des Forschungskreislaufs, Rechtliche Aspekte im Forschungskreislauf, Definition Open Science, Konnex „Open Science“ und Recht im FDM 	24 Stunden E-Learning und 3 Videomeetings à 2 Stunden = 30 Stunden
Modul 3: Basismodul Forschung, FD & FDM in den Fachgebieten	<ul style="list-style-type: none"> • Forschungsprozesse, Datentypen, Tools, Oberflächen und ausgewählte Use Cases oder Best Practices einzelner Disziplinen 	18 Stunden E-Learning und 6 Videomeetings à 2 Stunden = 30 Stunden
Modul 4: Aufbaumodul Hacken & experimentieren mit Daten	<ul style="list-style-type: none"> • Umgang mit Daten, Automatisierung, Reproduzierbarkeit, Unix-Shell, Python, Git und GitHub 	16 Stunden E-Learning und 2 Videomeetings à 7 Stunden = 30 Stunden
Modul 5: Aufbaumodul (Meta-)Daten verwalten & teilen	<ul style="list-style-type: none"> • Daten strukturieren, organisieren, teilen, publizieren, finden und nachnutzen 	24 Stunden E-Learning und 2 Videomeetings zu 4 und 2 Stunden = 30 Stunden
Modul 6: Aufbaumodul Technische Infrastruktur	<ul style="list-style-type: none"> • Speichersysteme, Langzeitarchivierung, Repositorien 	24 Stunden E-Learning und 2 Videomeetings zu 2 und 4 Stunden = 30 Stunden
Modul 7: Aufbaumodul Daten- & Projektmanagement in der Forschung	<ul style="list-style-type: none"> • Planung des Forschungsdatenmanagements im Projekt von der Antragsphase bis zur Umsetzung 	24 Stunden E-Learning und 2 Videomeetings à 3 Stunden = 30 Stunden
Modul 8: Aufbaumodul FDM-Beratung & Schulung	<ul style="list-style-type: none"> • Beratungstheorie, Techniken der Gesprächsführung, Konfliktmanagement, Beratungsprozess, Beratungssettings, Bereitstellung von Informationen, Schulungen 	24 Stunden. E-Learning und 2 Videomeetings à 3 Stunden = 30 Stunden
Modul 9: Projektmodul	<ul style="list-style-type: none"> • Arbeiten an einem praktischen Projekt (z.B. Konzept für das Forschungsdatenmanagement, ein Datenmanagementplan) 	35 Stunden Projektarbeit und 1 Präsenztag à 4 Stunden = 39 Stunden

Tabella 1: Thematische Übersicht und Einteilung des Zertifikatskurs Forschungsdatenmanagement

Eine detaillierte Beschreibung der Inhalte findet sich im Modulhandbuch.⁴

Masterstudium Digitales Datenmanagement

Das Masterstudium Digitales Datenmanagement (DDM) ist als gemeinsamer weiterbildender Studiengang der Fachhochschule Potsdam und der Humboldt-Universität Berlin konzipiert und vermittelt Kompetenzen, die für Aufgabenstellungen im Digitalen Datenmanagement in Wissenschaft, Kultur und Wirtschaft benötigt werden. Im Rahmen dieser Weiterbildung werden Theorie und Praxis zu Forschungs- und Informationsstrukturen sowie Informationssystemen verknüpft und sowohl wissenschaftliche, ethische, wissenschaftspolitische, organisatorische, technologische als auch rechtliche Aspekte mitberücksichtigt.⁵

Da das Masterstudium Digitales Datenmanagement die drei Domänen Wissenschaft, Kultur und Wirtschaft bedient, ist die Zielgruppe breit gefächert: Interessent*innen an Datenmanagement, -prozessierung, -analyse und -bereitstellung, Forschungsreferent*innen, Mitarbeiter*innen im Forschungsservice und Wissenschaftsmanagement, Daten-Produzent*innen (z.B. Verwaltungen, Forschungseinrichtungen), Daten-Verarbeiter*innen, Mitarbeiter*innen von Serviceeinrichtungen der Informationsinfrastruktur (z.B. Bibliotheken, Museen) sowie Mitarbeiter*innen im Daten-Service-Bereich (z.B. Verlage).³

Die Zugangsvoraussetzungen umfassen einen berufsqualifizierenden Hochschulabschluss im Umfang von 180 ECTS sowie eine mindestens 12-monatige Berufserfahrung, die den Studieninhalten zuträglich ist und im Rahmen einer Vollzeitstätigkeit erworben wurde und mindestens 2 Monate im Bereich (digitaler) Informations- und Datenwissenschaft umfasst.⁶

Der Gesamtarbeitsaufwand beläuft sich auf 4 Semester mit 120 ECTS (entspricht 3000 Stunden; pro Semester sind 30 ECTS zu erbringen). Das Masterstudium setzt auf Blended Learning, bei dem E-Learning Einheiten und Präsenzveranstaltungen kombiniert werden. Pro Semester finden 4-5 Präsenztermine (2-tägig, Freitag und Samstag) statt. Die restlichen Einheiten werden über die E-Learning Plattform Moodle abgewickelt.⁷ Die erfolgreiche Absolvierung des Masterstudiums schließt mit dem akademischen Grad Master of Arts (M.A.) ab. Die Kosten belaufen sich auf 7.900 € (1975 € pro Semester) zzgl. Semestergebühren (ca. 115 €).⁸

Für jene Interessent*innen, die nicht das gesamte Studium absolvieren möchten, besteht auch die Möglichkeit einzelne Modulkurse und Module (1, 3 und 5) zu belegen, sofern dafür genügend freie Plätze im Semester zur Verfügung stehen. Die Teilnahme wird jeweils bescheinigt. Als Teilnahmevoraussetzung gilt ein einschlägiger Hochschulabschluss. Berufserfahrung ist zwar erwünscht, jedoch nicht zwingend erforderlich, um die Kurse zu besuchen. Werden alle Kurse (jeweils

⁴ Modulhandbuch Zertifikatskurs Forschungsdatenmanagement. https://www.th-koeln.de/mam/downloads/deutsch/weiterbildung/zbiw/angebote/modulhandbuch_zk_fdm_2021_22.pdf (Aufruf am 24.09.2021)

⁵ Vgl. Studieninhalte Masterstudium Digitales Datenmanagement. <https://www.ddm-master.de/studieninhalte/> (Aufruf am 24.09.2021)

⁶ Vgl. Zugangsvoraussetzungen für das Masterstudium Digitales Datenmanagement. <https://www.ddm-master.de/bewerbung/> (Aufruf am 24.09.2021)

⁷ Vgl. FAQ zum Masterstudium Digitales Datenmanagement. <https://www.ddm-master.de/faq/> (Aufruf am 24.09.2021)

⁸ Vgl. Überblick Masterstudium Digitales Datenmanagement. <https://www.ddm-master.de/> (Aufruf am 24.09.2021)

4 ECTS) eines Moduls abgeschlossen, besteht die Möglichkeit auch die Modulprüfung abzulegen. Bei erfolgreicher Absolvierung der Prüfung, wird ein Hochschulzertifikat (insgesamt 20 ECTS) ausgestellt.⁹

Das Masterstudium Digitales Datenmanagement setzt sich aus den folgenden Modulen zusammen:¹⁰

Modul	Inhalt	Aufwand
Modul 1: Rahmenbedingungen des Datenmanagements	<ul style="list-style-type: none"> • Theoretische Grundlagen Datenmanagement und Data Literacy • Forschungs- und Informationsinfrastrukturen • Open Access, Open Data und Open Science • Metadaten, Standards, Interoperabilität 	20 ECTS
Modul 2: Designprojekt	<ul style="list-style-type: none"> • Agiles Projektmanagement 	10 ECTS
Modul 3: Technologien des Datenmanagements	<ul style="list-style-type: none"> • Informationstechnologische Grundlagen: Internet- und Webtechnologien • Informationstechnologische Grundlagen: Datenmanagementsysteme • Einführung in Algorithmen und Datenstrukturen • Digitale Repositorien 	20 ECTS
Modul 4: Wahlpflichtmodule 2/4	<ul style="list-style-type: none"> • Reallabor Daten • Reallabor Technologien • Reallabor Methodik • Ausgewählte Themen des Datenmanagements 	10 ECTS
Modul 5: Methoden des Datenmanagements	<ul style="list-style-type: none"> • Forschungsdatenmanagement • Datenmanagementpläne • Statistische Methoden in der Datenaufbereitung und -auswertung • Datenanalyse und Datenvisualisierung 	20 ECTS
Modul 6: Transferprojekt	<ul style="list-style-type: none"> • Agiles Projektmanagement 	10 ECTS
Modul 7: Abschlussmodul	<ul style="list-style-type: none"> • Gute wissenschaftliche Praxis • Masterarbeit und Verteidigung 	30 ECTS

Tabelle 2: Thematische Übersicht und Einteilung *Masterstudium Digitales Datenmanagement*

Eine ausführliche Beschreibung der Modulinhalte findet sich im Modulhandbuch.¹¹

⁹ Vgl. Belegung einzelner Modulkurse und Module des Masterstudiums Digitales Datenmanagement. <https://www.fh-potsdam.de/weiterbilden/weiterbildungsangebote/digitales-datenmanagement/> (Aufruf am 24.09.2021)

¹⁰ Studienverlaufsplan Digitales Datenmanagement. <https://www.ddm-master.de/wp-content/uploads/2020/12/DDM-Studienverlaufsplan.pdf> (Aufruf am 24.09.2021)

¹¹ Modulhandbuch Digitales Datenmanagement. https://www.ddm-master.de/wp-content/uploads/2019/12/DDM_MH.pdf (Aufruf am 24.09.2021)

DataTrain: Training in Research Data Management & Data Science

Das DataTrain Programm ist eine Ausbildung, die sich primär an Doktoranden der Mitgliedseinrichtungen der U Bremen Research Alliance¹² richtet und Grundlagen im Bereich Forschungsdatenmanagement und Data Science vermittelt. Die Vortragenden und Expertise stammt aus dieser Allianz und wird um Unternehmen aus der Region Bremen erweitert, die ebenfalls Interesse an Personal mit entsprechendem Wissen haben. Das Programm selbst besteht aus drei eigenständigen „Tracks“, die allerdings inhaltlich aufeinander aufbauen. Die Teilnehmer*innen können gesamte „Tracks“ durchlaufen oder auch nur an einzelnen Veranstaltungen teilnehmen. Prinzipiell können sich alle interessierten Personen für die Einheiten anmelden, aber Doktorand*innen der Allianz werden immer bevorzugt behandelt und nur bei freien Restplätzen werden diese, nach Registrierung auf einer Warteliste, an externe Personen vergeben. Einzig der „Starter Track“ (Tabelle 3) erlaubt eine freie Anmeldung da es sich um eine interaktive Überblickvorlesungsreihe handelt, mit ca. 150 möglichen Teilnehmer*innen. Aufbauend auf den „Starter Track“ gibt es dann den „Operator Track: Data Steward“ (Tabelle 4) und den „Operator Track: Data Science“ (Tabelle 5), die damit werben, die Kompetenzen eines Data Steward bzw. Data Scientist zu vermitteln bzw. mit digitalen Hands-on Workshops direkt zu erlernen. Der gesamte Data Train umfasst in seiner Dauer ca. 10 Monate mit 30 unterschiedlichen Themen und mindestens 47 einzelnen Einheiten.¹³ Dieser besteht aus 13 interaktiven Vorlesungen im „Starter Track“, mit ca. 30 Stunden Vortragszeit, und 17 Hands-on Workshops in beiden „Operator Tracks“, mit mindestens 150 Stunden. Im Folgenden wird kurz erläutert mit welchen inhaltlichen Schwerpunkten in den einzelnen „Tracks“ zu rechnen ist:

Der „Starter Track“ besteht aus vier thematischen Gruppen, die wiederum aus 13 Vorlesungen aufgebaut sind (s. Tabelle 3). Die erste Gruppe, *Digitalization / data science*, beschäftigt sich mit Methoden und Begriffen aus dem Data Science und Big Data Bereich, sowie der kritischen Auseinandersetzung mit der eigenen Forschungsarbeit und der unterschiedlichen Denkweisen der verschiedenen Disziplinen. In der zweiten Gruppe, *Data governance*, wird zuerst über die Bedeutung des Forschungsdatenmanagements und deren rechtlichen Möglichkeiten gesprochen. Danach werden die Herausforderungen und Handhabung von vertraulichen bzw. personenbezogenen Daten und qualitativen Daten erklärt. *Statistical methods for data science* bildet die dritte Gruppe und beginnt mit den wichtigsten statistischen Prinzipien, sowie der Orientierung bei der Frage nach den benötigten Daten und der geeigneten Methoden bei Forschungsfragen. Zusätzlich wird versucht eine kritische Denkweise zu fördern, wenn es darum die Bedeutung von Forschungsdaten für bestimmte Disziplinen zu ermitteln. Abschließend wird mit *Informatics* ein Überblick über die Informatik selbst und eine Auswahl an Programmiersprachen (Bedeutung, Unterschiede, etc.) gegeben. Basiskennnisse zum Thema Kryptographie und ihrer Praxistauglichkeit sowie grundlegende Sicherheits- und Datenschutztechniken runden das Angebot ab.

¹² Universität Bremen, U Bremen Research Alliance, <https://www.uni-bremen.de/research-alliance> (Aufruf am 24.09.2021)

¹³ Zum Zeitpunkt des Workshops waren Inhalt und Umfang des Operator Track: Data Scientist nicht bekannt

Thematische Gruppe	Vorlesungen	Aufwand
Digitalization / data science	<ul style="list-style-type: none"> • Data science and big data • Philosophical reflections on data science 	virtuell: 2 Tage mit insgesamt 4 Stunden
Data governance	<ul style="list-style-type: none"> • Data and information management • Data protection licenses • Managing confidential data • Managing qualitative data 	virtuell: 4 Tage mit insgesamt 10 Stunden
Statistical methods for data science	<ul style="list-style-type: none"> • Statistical thinking • Asking the right research question in data science • About the meaningfulness of data 	virtuell: 3 Tage mit insgesamt 6 Stunden
Informatics	<ul style="list-style-type: none"> • Computer science basics for data science • Overview about programming languages • Cryptography basics • Security and privacy 	virtuell: 4 Tage mit insgesamt 8 Stunden

Tabelle 3: Thematische Übersicht und Einteilung des Data Train *Starter Track*¹²

Der “Operator Track: Data Steward” umfasst drei thematische Gruppen mit insgesamt 9 Workshops (s. Tabelle 4). Im ersten Teil, *Programming languages*, wird gelehrt wie mit Hilfe von R Data Science Methoden angewendet werden und wie Daten aus externen Quellen für statistische Analysen genutzt werden. Zusätzlich wird ein grundlegendes Verständnis für die Programmiersprachen MATLAB und Python vermittelt. Die zweite Gruppe, *Data handling*, befasst sich mit dem Versionskontrollsysteme und kollaboratives Arbeiten mittels Git/GitHub und der Erstellung eines Datenmanagementplans. Auch das Thema Reproduzierbarkeit von Forschungsergebnissen durch die richtigen Arbeitsabläufe und Tools wird behandelt. Im letzten Teil, *Data engineering*, wird der Umgang mit verschiedenen Arten von Daten (Bild-, Klima-, Gesundheits- und Quantitativdaten) geübt und die Bereitstellung von Daten mit Repositorien gezeigt. Darin enthalten ist auch das Anlegen von Datenbanken mittels Datenbankmanagementsystemen und deren Abfrage durch SQL.

Thematische Gruppe	Workshops	Aufwand
Programming languages	<ul style="list-style-type: none"> • Getting started with R • Erste Schritte mit MATLAB • Getting started with Python 	virtuell: 13 Tage mit insgesamt 92 Stunden
Data handling	<ul style="list-style-type: none"> • Git / Git Hub • Tools for FAIR data handling: DMP tool • Reproducibility in science – How and why? 	virtuell: 6 Tage mit insgesamt 16 Stunden
Data engineering	<ul style="list-style-type: none"> • Data preparation • Data provisioning • Data base skills 	virtuell: 6 Tage mit insgesamt 27 Stunden

Tabelle 4: Thematische Übersicht und Einteilung des Data Train *Operator Track: Data Steward*¹²

Der “Operator Track: Data Scientist” beschäftigt sich mit dem Thema Data Science und besteht aus drei Themenbereichen mit insgesamt 8 Workshops (s. Tabelle 5). Da zum Zeitpunkt des „FAIR Data Austria“ Workshops die Inhalte dieses „Tracks“ nicht zugänglich waren, sind nur die Bezeichnungen der einzelnen Workshops bekannt.

Thematische Gruppe	Workshops	Aufwand
Statistical methods for data science	<ul style="list-style-type: none"> • Quantitative analyses for data science • Causal learning 	-
Artificial intelligence, machine learning	<ul style="list-style-type: none"> • Machine learning algorithms • Deep learning / neural networks • Evaluation of ML/AI algorithms • Computational social sciences 	-
Visualization	<ul style="list-style-type: none"> • Visualization: Principles & critical reflections • Visual Analytics using GIS 	-

Tabelle 5: Thematische Übersicht und Einteilung des Data Train **Operator Track: Data Scientist**¹²

Zertifikatskurs Data Steward

Die Universitätsbibliothek Wien entwickelt derzeit einen neuen Zertifikatskurs Data Steward. Da sich dieser in der Konzeptionsphase befindet, können hier nur vorläufige Informationen präsentiert werden, die sich bei der Weiterentwicklung ändern können. Im vorliegenden Bericht wird sowohl der Entwicklungsstand von Juli 2021, als auch die möglichen Änderungen, die bis September 2021 vorgenommen wurden, dargestellt.

Bei dem Zertifikatskurs Data Steward handelt es sich um einen Ausbau des Zertifikatskurses Data Librarian¹⁴. Dieser wird seit 2018/2019 in Kooperation der Ausbildungsbibliotheken in Österreich (Universitätsbibliothek Wien, Universitätsbibliothek Innsbruck und Universitätsbibliothek Graz) als eine berufliche Fortbildung für Bibliothekspersonal angeboten. Der Zertifikatskurs Data Librarian vermittelt Grundlagen der Arbeit im Bereich Forschungsdatenmanagement. Da Data Stewards für ihre Tätigkeiten auch weitere spezialisierte Fähigkeiten und Kenntnisse benötigen, wird das Konzept erweitert.

Data Stewards arbeiten an der Schnittstelle zwischen Forschung und Forschungsinfrastruktur und üben somit eine wesentliche Brückenfunktion aus. Der Zertifikatskurs richtet sich daher an interessierte Personen, die über Forschungserfahrung verfügen oder bereits im Bereich forschungsunterstützender Services tätig sind als Vorbereitung auf die zukunftssträchtige Rolle von Data Stewards.

Der Zertifikatskurs ist für zwei Semester geplant und umfasst 120 Präsenz-Stunden, sowie Selbstlernphasen von 180 Stunden. Die Inhalte sind in vier obligatorische Module aufgeteilt. Das Modul 1 fokussiert sich an die Einführung in das Thema Forschungsdatenmanagement, Data Stewardship und Open Science sowie die relevanten rechtlichen und ethischen Aspekte. Die erworbenen Kenntnisse werden im Modul 2 “FAIRe Daten im Lebenszyklus” entlang des prototypischen Verlaufs eines Forschungsprojektes vertieft und um weitere Themen wie Repositorien, Metadaten für Forschungsdaten oder Persistente Identifier ergänzt. Im Modul 3 werden Grundlagen des Programmierens und von Data Science vermittelt. Im Modul 4 werden die erworbenen fachlichen Kenntnisse in typische Aufgaben von Data Stewards wie Training oder Beratung übergeführt. Der

¹⁴ <https://www.postgraduatecenter.at/weiterbildungsprogramme/kommunikation-medien/data-librarian/> (Aufruf am 24.09.2021)

Zertifikatskurs wird in einer Kombination aus Präsenz- und Online-Veranstaltungen (hybrid), berufsbegleitend angeboten und sollte jährlich durchgeführt werden. Er sollte die Teilnehmenden auf die vielfältigen Tätigkeitsbereiche und die zukunftsorientierte Rolle von Data Stewards an verschiedenen Forschungseinrichtungen vorbereiten.

Modul	Inhalt	Aufwand
Modul 1: Grundlagen des Forschungsdatenmanagements und Open Science	<ul style="list-style-type: none"> • Grundlagen des Forschungsdatenmanagements und Datenarchivierung, Rolle von Data Stewards • Open Science, Open Data, Open Access • Gute wissenschaftliche Praxis und Ethik • Rechtliche Aspekte, insbesondere Datenschutz und Urheberrecht • Finanzierungsmodelle, Kostenschätzung und Ressourcenaufwand 	2 ECTS, 30 Unterrichtseinheiten (UE)
Modul 2: FAIRe Forschungsdaten im Lebenszyklus	<ul style="list-style-type: none"> • Forschungsvorhaben planen (z. B. Daten- und Projektmanagement (in der Forschung); Datenmanagementpläne und DMP-Tools) • Daten erheben (z. B. Forschungsdaten und FDM in den Fachgebieten; Beispiele von Tools für Datenmanagement) • Daten aufbereiten und analysieren (z. B. Datenaufbereitung und -bereinigung, Statistische Auswertung) • Daten teilen und publizieren (z. B. Metadaten(schemata) für Forschungsdaten, Persistente Identifier, Ontologien) • Daten archivieren (z. B. Speicherung und Langzeitarchivierung, Datensicherheit, Repositorienmanagement) • Daten nachnutzen (z. B. rechtliche Rahmenbedingungen, Nutzungsbedingungen von Repositorien, Bibliometrie) 	5 ECTS, 45 UE
Modul 3: IT und Data Science Grundlagen	<ul style="list-style-type: none"> • IT-Grundlagen - Datenbanksysteme, Übersicht von Programmiersprachen • Grundlagen: Git und GitHub, SQL, Machine Learning, Python, XML • Hacken und experimentieren mit Daten 	3 ECTS, 30 UE
Modul 4: Forschungsdatenmanagement-Support	<ul style="list-style-type: none"> • Beratung und Service • Wissenstransfer und Didaktik • Bedarfserhebung und -analyse, Requirements Engineering 	2 ECTS, 15 UE

Tabelle 6: Thematische Übersicht und Einteilung Zertifikatskurs Data Steward (Stand Juli 2021)

Ergebnisse der Gruppenarbeiten

Zertifikatskurs Forschungsdatenmanagement

Der Zertifikatskurs Forschungsdatenmanagement (FDM) an der TH Köln fordert doch einen hohen Arbeitsaufwand und wird hier daher eher nicht als Zusatzausbildung gesehen, sondern rechtfertigt sich wohl nur dann, wenn die erworbenen Kenntnisse auch tatsächlich in der täglichen Arbeit eingesetzt werden können.

Der Kurs hat einen hohen Praxisbezug und ist für Personen gedacht, die täglich mit der Thematik zu tun haben. Besonders hervorzuheben ist das Modul 3, in dem unterschiedliche Fachrichtungen in Hinblick auf das Forschungsdatenmanagement beleuchtet werden. Die Voraussetzungen für die Teilnahme sind sehr allgemein formuliert, was es möglicherweise schwierig macht die Inhalte auf Personen mit unterschiedlichen Vorkenntnissen anzupassen.

Das Modul 7 wird von der Gruppe für keines der Modelle als geeignet angesehen. Alle anderen Module bieten für die Modelle 2 und 3 wertvolle Inhalte an. Für das Modell 1 werden nur die Module 1, 2, 5 und 9 empfohlen. Modul 9 ist für alle Modelle besonders wichtig, da erst bei einem eigenen Projekt klar wird was man aus dem Kurs wirklich mitgenommen hat und wo die Herausforderungen und Schwierigkeiten liegen.

Zusammenfassend wird festgehalten, dass die Kursinhalte als durchwegs positiv wahrgenommen werden und dabei Modul 3 besonders hervorgehoben wird. Ein großer Nachteil ist, dass nur Personen aus Nordrhein-Westfalen an diesem Kurs teilnehmen können.

Masterstudium Digitales Datenmanagement

Das Masterstudium Digitales Datenmanagement (DDM) vermittelt in vier Semestern umfangreiches Wissen über die Anforderungen an Datenmanagement in Forschung, Wirtschaft und Kultur (Abschluss: Master of Arts). Es ist als Blended Learning Studium konzipiert, das einen hohen Anteil an E-Learning mit wenigen Präsenzterminen pro Semester kombiniert. Für Teilnehmer*innen, die großen Wert auf direkte Interaktion in Form von Präsenzlehre legen, ist dieses Studium wenig geeignet.

Personen, die sich zu spezifischen Themen des Datenmanagements weiterbilden möchten, haben auch die Möglichkeit einzelne Kurse eines Moduls oder auch gesamte Module zu absolvieren, inkl. Modulabschlussprüfung zum Erhalt eines Hochschulzertifikats. Leider ist es nicht möglich das Reallabor, welches konkrete Praxisbeispiele behandelt, einzeln zu belegen (sondern nur im Rahmen des Studiums). Dies hätte sich als Mehrwert für Teilnehmer*innen erwiesen, die bereits einen theoretischen Hintergrund im Datenmanagement aufweisen und sich nun genauer über den Transfer in die Praxis informieren möchten. Nichtsdestotrotz stellt die Belegung von einzelnen Kursen und Modulen eine sinnvolle Ergänzung zu anderen Weiterbildungsangeboten von Data Stewards dar.

Zusammenfassend gibt dieses Studium einen guten Überblick über alle relevanten Aspekte des Datenmanagements. Sehr positiv zu bewerten ist das Reallabor, das durch eine hohe Praxisorientierung die Herausforderungen des Alltags aufzeigt und Lösungen dazu erarbeitet. Auch die Zahl der Teilnehmer*innen ist passend, um die Themen in kleinen Gruppen auszuarbeiten.

DataTrain: Training in Research Data Management & Data Science

Das Data Train Programm der Universität Bremen vermittelt in drei speziellen Tracks Grundlagenwissen zu den Themen Datenmanagement, Data Science und Awareness Building. Die Absolvierung aller Einheiten dauert etwa ein Jahr und kann auch von organisationsfremden Personen, d.h. von Personen außerhalb der Uni Bremen Research Alliance, besucht werden.

Da der Starter Track nur aus Frontalvorträgen besteht, ist hier eine einfache Anmeldung ausreichend. Die beiden anderen Tracks haben ein begrenztes Kontingent und werden nach Organisationszugehörigkeit und Zeitpunkt der Anmeldung vergeben. Die Referentinnen und Referenten sind Experten auf ihrem jeweiligen Gebiet, was sich bereits in der Beschreibung der einzelnen Module zeigt, die frei kombiniert werden können.

Da das Thema Förderung und Didaktik komplett ausgeklammert wird, fehlen diese Inhalte bei Modell 1. Generell sind die Inhalte des Data-Train-Programms besser für die Modelle 2 und 3 geeignet, und mit dem Data Scientist Track werden auch einige der in Modell 3 geforderten Kompetenzen trainiert.

Zusammenfassend lässt sich sagen, dass das Data Train Programm eine kostengünstige Alternative zu anderen Trainingsprogrammen ist, die sich auf die Verwaltung, Erstellung und Verarbeitung von Daten konzentriert und auch die Aspekte der IT und IT-Sicherheit anspricht. Die didaktischen und moderierenden Fähigkeiten, die für die Sensibilisierung für gute RDM-Praxis erforderlich sind, sind schwer zu finden.

Zertifikatskurs Data Steward

Der Zertifikatskurs vermittelt in hybrider Form in vier obligatorischen Modulen die Grundlagen des Forschungsdatenmanagements inklusive technischer, rechtlicher und ethischer Aspekte sowie die Entwicklung und Durchführung von unterstützenden FDM-Services. Die Kursinhalte wurden überwiegend positiv betrachtet. Alle vier Module vermitteln relevante Inhalte für die Modelle 2 und 3. Obwohl die Inhalte im Modul 3 und 4 zum Teil das Aufgabenspektrum im Modell 1 sprengen, so sollte die Teilnahme trotzdem in Betracht gezogen werden, um sich auf mögliche Änderungen des Modells an der eigenen Einrichtung vorzubereiten. Es wird empfohlen, die Module auch einzeln als Weiterbildungen anzubieten sowie den Wissenstransfer in der Form einer Projektarbeit zu sichern.

Als positiv bewertete die Gruppe die offene Zielgruppe des Kurses. Ein Kritikpunkt zum Masterabschluss als eine viel zu hohe Hürde für die Teilnahme (Stand Juli 2021) wurde durch ein Stufenmodell, dass relevante Berufserfahrung vermehrt berücksichtigt, ersetzt (Stand September 2021). Die Entwicklung des Angebots anhand bestehender ähnlicher Aus- und Fortbildungsangeboten wurde ebenso gut angenommen.

Als Schwäche des Kurses wurde die unzureichende Förderung der Sozialkompetenz sowie Management Skills (relevant besonders für die Koordinator*innen im Modell 3) genannt. Der angemessene Arbeitsaufwand sowie die Konzeption des Kurses als eine hybride berufsbegleitende Fortbildung, wurden als besonders positiv angesehen. In Kombination mit den relevanten Fach- und Methodenkompetenzen sollten Data Stewards durch den Zertifikatskurs auf ihre Rolle im allgemeinen Forschungsdatenmanagement gut vorbereitet werden.

Mapping – Data Steward Modelle vs. Ausbildungsangebote

Ziel dieses Workshops ist es unter anderen klar darzustellen ob ein Ausbildungsangebot die gewünschten Kenntnisse für Data Stewards in den einzelnen Modellen ausreichend vermittelt. Als ausreichend wird, im Rahmen des Workshops, eine Abdeckung von mindestens 2/3 der Fachkompetenzen eines Modelles angesehen. In Tabelle 7 ist die Eignung eines Angebotes für ein Modell mit einem „x“ markiert. Bei drei von vier Angeboten ist zusätzlich noch „eingeschränkt“ hinzugefügt worden, da keines der bestehenden Angebote die disziplinspezifischen Herausforderungen abdecken kann, mit welchen Data Stewards im Netzwerk Modell konfrontiert sind. Diese Anforderungen können nur durch spezielle Aus- und Weiterbildungsmöglichkeiten in den entsprechenden Bereichen erworben werden und sind daher als Ergänzungen zu den behandelten Angeboten zu sehen.

Angebot / Modelle	Anlaufstelle	Data Steward Center	Data Steward Netzwerk
Zertifikatskurs Forschungsdatenmanagement	x	x	x (eingeschränkt)
Masterstudium Digitales Datenmanagement	x	x	
Data Train		x	x (eingeschränkt)
Zertifikatskurs Data Steward	x	x	x (eingeschränkt)

Tabelle 7: Eignung der Ausbildungsangebote für Data Steward Modelle

Zusammenfassung

In diesem Workshop-Bericht ist die Beurteilung von vier potenziellen Ausbildungsmöglichkeiten für Data Stewards an österreichischen Universitäten festgehalten. Dabei wurden die Kurse zuerst in ihrem Aufbau und Inhalt vorgestellt und dann die Ergebnisse der Analyse dieser durch Gruppenarbeiten festgehalten. Die Bewertungen der Ausbildungen haben gezeigt, dass keiner der Kurse tatsächlich ein einzelnes oder mehrere Modelle zur Gänze abdecken kann, wenn es darum geht, die gewünschten Kompetenzen zu vermitteln. Gewisse Kurse sind für einzelne Modelle besser anzuwenden, aber es sind immer begleitende Maßnahmen oder Schulungen notwendig, um angehenden Data Stewards vertiefendes Wissen und spezielle Werkzeuge mitgeben zu können. Auch hat keine der Ausbildungsangebote die Möglichkeit geboten, die sehr wichtigen sozialen und persönlichen Kompetenzen eines Data Stewards² zu erweitern oder zu verbessern. Management Skills, die besonders für die Koordinator*innen von Data Stewardship Programmen von Relevanz sind, werden in den diskutierten Ausbildungsangeboten ebenso nicht vermittelt. Es wird des Weiteren als zentral angesehen, eine eigene Projektarbeit (ca. 35 Stunden) zu absolvieren, um sich mit gewissen Aspekten des Forschungsdatenmanagements viel intensiver beschäftigen zu können und Praxisbezug zu erlangen. Mit der zunehmenden Professionalisierung und Etablierung von Data Stewards — besonders bei dem Einsatz des dezentralen Modells 3 — ist in Zukunft mit einem wachsenden Bedarf an disziplinspezifischen Aus- und Weiterbildungsangeboten zu rechnen.

Annex

Teilnehmer*innen

Breakout-Session Zertifikatskurs Forschungsdatenmanagement

Hermann Schranzhofer (Moderator), Thomas Haselwanter, Theresa Macher

Breakout-Session Masterstudium Digitales Datenmanagement

Sabrina Knopper (Moderatorin), Sarah Stryeck, Christiane Stork, Eva Maria Asamer

Breakout-Session Data Train

Alexander Gruber (Moderator), Ilire Hasani-Mavriqi, Susanne Blumesberger, Michael Kranewitter

Breakout-Session Zertifikatskurs Data Steward

Tereza Kalová (Moderatorin), Heike Thöricht, Vitali Bodnar, Claire Jean-Quartier

Detaillierte Ergebnisse der Gruppenarbeiten

Zertifikatskurs Forschungsdatenmanagement

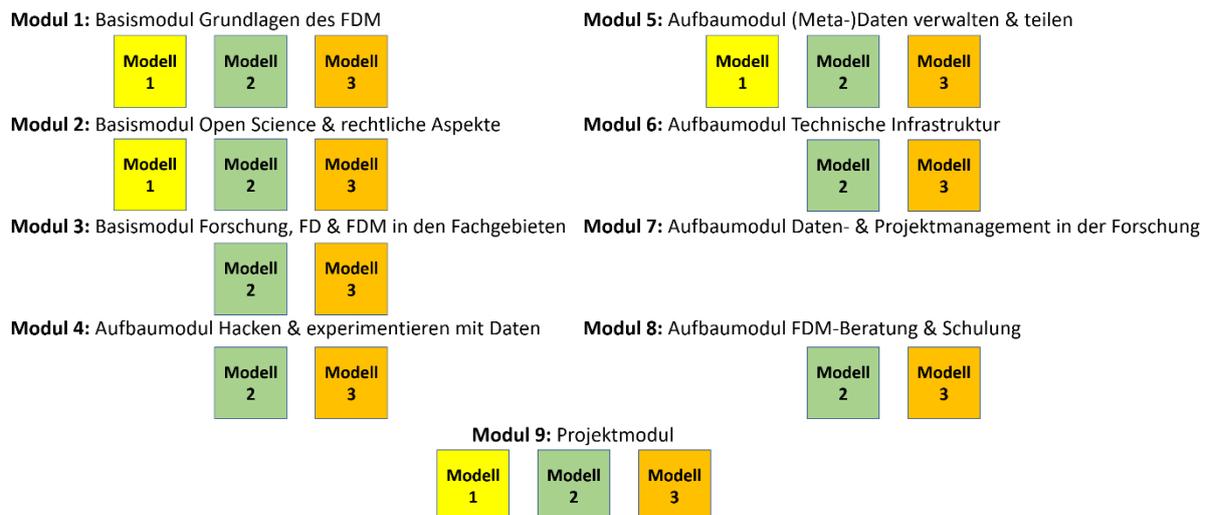


Abbildung 1: Zertifikatskurs Forschungsdatenmanagement mit den beurteilten Modulen in Hinblick auf die jeweiligen Data Steward Modelle

Der Zertifikatskurs Forschungsdatenmanagement (FDM) an der TH Köln fordert doch einen hohen Arbeitsaufwand und wird hier daher eher nicht als Zusatzausbildung gesehen, sondern rechtfertigt sich wohl nur dann, wenn die erworbenen Kenntnisse auch tatsächlich in der täglichen Arbeit eingesetzt werden können.

Positiv ist zu erwähnen, dass der Kurs einen hohen Praxisbezug hat. Trotzdem scheint der Kurs weniger für Forschende geeignet zu sein, sondern doch eher für Personen, die täglich mit der Thematik zu tun haben (z.B. Data Stewards). Datenschutzgrundlagen sind in dem Kurs kaum vorhanden, aber dafür könnten einzelne Teile des Kurses auch in der Lehre eingesetzt werden. Das Thema Projektmanagement wird kaum behandelt, allerdings gibt es im Vergleich zu den anderen Kursen ein

Alleinstellungsmerkmal, nämlich Modul 3. Hier werden unterschiedliche Fachrichtungen im Hinblick auf das Forschungsdatenmanagement beleuchtet. Die Voraussetzungen für die Kursteilnahme sind sehr allgemein und breit formuliert was dazu führen könnte, dass ein Teil der Teilnehmer*innen kaum über spezifisches Vorwissen verfügt, während ein anderer Teil möglicherweise bereits detaillierte Vorkenntnisse hat. Dies könnte eventuell dazu führen, dass der Kurs für die einen zu anspruchsvoll ist und sich andere aber unterfordert fühlen. Durch diese Inhomogenität könnte es möglicherweise schwierig sein die Inhalte auf eine spezifische Gruppe mit/ohne Kenntnisse anzupassen.

Module 1 und 2 bilden die Grundkenntnisse ab und sollten in allen Modellen vorkommen. Modul 3 ist sehr interessant, da es hier ermöglicht wird die Forschenden in ihren Fachgebieten beim Thema FDM abzuholen. Das wäre für Modelle 1 wohl eher übertrieben aber passend für Modell 2 und 3. Diese fachspezifischen Inhalte können aber auch gleichzeitig in der Lehre bei unterschiedlichen Fachrichtungen eingesetzt werden. Gleichzeitig braucht aber ein Data Steward nicht alle Forschungsrichtungen im Detail kennen, obwohl ein Überblick, auch im eigenen Bereich, sehr hilfreich sein kann.

Modul 4 ist auch wieder für Modell 2 und 3 geeignet, auch wenn es sich bei 30 Stunden Zeitaufwand nur um einen Überblick handelt, ist es bei Modell 1 nicht unbedingt notwendig. Modul 5 ist auch wieder für alle Modelle relevant, aber bei Modul 6 wird das Modell 1 nicht vorgesehen, da es sowieso Beauftragte für technische Infrastruktur geben sollte und somit auf diese verwiesen werden kann. Für Modell 2 und Modell 3 sollte man allerdings schon Kenntnisse in diesem Bereich haben.

Modul 7 wird hier für keines der Modelle als geeignet angesehen und hat auch einen etwas irreführenden Titel. Anders als dieser glauben lassen würde, geht es nämlich darum einen Projektantrag mit Hinblick auf das Forschungsdatenmanagement zu erstellen. Alternativ wird für dieses Modul der Name „Wie schreibe ich einen Antrag?“ vorgeschlagen. Auch wenn ein Data Steward dabei Unterstützung leisten kann, handelt es sich nicht um einen DMP und somit nicht um eine direkte Aufgabe eines Data Steward.

Modul 8 kann sicher für Modell 2 und 3 verwendet werden aber wieder werden die Ressourcen in Modell 1 nicht reichen um ein adäquates Beratungs- und Schulungsangebot sicher zu stellen. Daher wird Modul 8 für dieses Modell nicht vorgeschlagen.

Modul 9 ist für alle Modelle besonders wichtig, da erst bei einem eigenen Projekt klar wird was man aus dem Kurs wirklich mitgenommen hat und wo die Herausforderungen und Schwierigkeiten liegen (DMP schreiben, Forschungsdatenkonzept ausarbeiten, etc.).

Zusammenfassend wird festgehalten, dass die Kursinhalte als durchwegs positiv wahrgenommen werden und dabei Modul 3 besonders hervorgehoben wird. Ein großer Nachteil ist, dass nur Personen aus Nordrhein-Westfalen an diesem Kurs teilnehmen können.

Masterstudium Digitales Datenmanagement

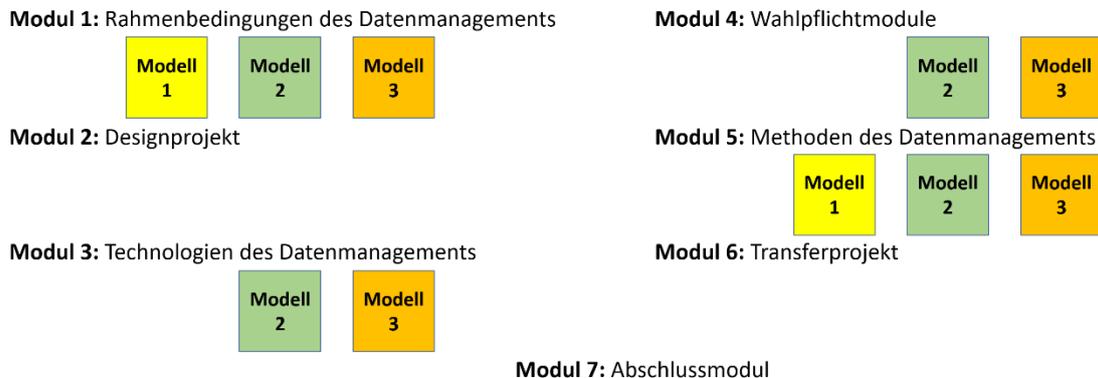


Abbildung 2: Masterstudium Digitales Datenmanagement mit den beurteilten Modulen in Hinblick auf die jeweiligen Data Steward Modelle

Das Masterstudium Digitales Datenmanagement (DDM) ist als viersemestriges Blended Learning-Studium konzipiert. Es ist jedoch auch möglich gewisse Einzelkurse eines Moduls zu absolvieren. Auch die Absolvierung eines gesamten Moduls (inkl. Modulabschlussprüfung) ist möglich, um ein Hochschulzertifikat zu erhalten. Der Fokus der Bearbeitung dieses Studiums in der Breakout Session lag bei jenen Modulen, die auch einzeln zu absolvieren sind, da sie inhaltlich die interessantesten Module darstellen und ohnehin die meisten ECTS enthalten.

Das Modul 1 „Rahmenbedingungen des Datenmanagements“ vermittelt Grundlagenwissen zu den Themenbereichen Datenmanagement, Data Literacy, Forschungs- und Informationsinfrastrukturen, Open Access, Open Data, Open Science sowie Metadaten, Standards und Interoperabilität. Es eignet sich daher für alle drei Data Stewardship Modelle. Besonders positiv hervorzuheben ist, dass das Thema Metadaten schon früh behandelt wird, genauso wie Forschungs- und Informationsinfrastrukturen.

In Modul 3 „Technologien des Datenmanagements“ ist der Modulkurs „Digitale Repositorien“ besonders gut aufbereitet und eine echte Empfehlung. Die anderen Themenbereiche sind sehr technisch und geben einen guten Überblick über IT-Grundlagen (Internet- und Webtechnologien, Datenmanagementsysteme, Einführung in Algorithmen und Datenstrukturen). An sich sind die Themen passend, jedoch nicht für Modell 1 zu empfehlen, sondern nur für Modell 2 und 3.

Modul 5 „Methoden des Datenmanagements“ ist schwierig zu bewerten, da hier Grundlagenthemen (Forschungsdatenmanagement, Datenmanagementpläne) mit sehr spezifischen Themen wie z.B. statistischen Methoden zur Datenaufbereitung und -auswertung, Datenanalyse und Datenvisualisierung vermischt werden. FDM-Grundlagen und DMPs gehören natürlich inhaltlich in alle Modelle, aber die anderen Inhalte gehören je nach Disziplin und Ausrichtung in Modell 2 und 3 oder sogar nur in Modell 3. Modul 5 deckt viele sehr wichtige Bereiche ab, allerdings müssen je nach Modell die passenden Themen gewählt werden.

Die Wahlpflichtmodule (Reallabor Daten, Reallabor Technologien, Reallabor Methodik, ausgewählte Themen des Datenmanagements) in Modul 4 können nicht einzeln belegt werden (da sie nur bei Absolvierung des Studiums absolviert werden können) und behandeln Best Practice Beispiele sowie Herausforderungen in der Praxis. Thematisch sehr interessant und zu empfehlen, jedoch erst ab Modell 2 sinnvoll.

Modul 2 „Designprojekt“ und Modul 6 „Transferprojekt“ beschäftigen sich mit dem Thema Projektmanagement. Die Aneignung dieser Methodenkompetenz kann auch in einer anderen Weiterbildung, z.B. in spezifischen Projektmanagementkursen, erfolgen. Modul 7 „Abschlussmodul“ beinhaltet gute wissenschaftliche Praxis sowie die Masterarbeit und ist selbstverständlich obligatorisch für den Abschluss des Studiums.

Zusammenfassend gibt dieses Studium einen guten Überblick über alle relevanten Aspekte des Datenmanagements. Das Reallabor behandelt sehr praxisnah die Herausforderungen im Alltag und ist daher eine sehr gute Idee. Auch die Teilnehmerzahl (kleine Gruppen) ist gut geeignet, um die Themen auszuarbeiten. Positiv anzumerken ist auch, dass Recht und Ethik in diesem Master enthalten sind (zu finden in Modul 1, Modulkurs „Open Access, Open Data und Open Science“ und in Modul 4, Modulkurs „Ausgewählten Themen des Datenmanagements“). Allerdings ist das Studium an gewissen Stellen zu spezifisch (siehe Modul 5) und daher nur für eine eingeschränkte Personengruppe voll nutzbar. Kompetenzen, die für Data Stewards unbedingt notwendig sind (soziale Kompetenzen, Didaktik und Awareness Building für Forschungsdatenmanagement), kommen in diesem Master nicht vor und müssen daher bereits vorhanden sein oder von anderer Stelle gefördert werden. Leider kann das Reallabor nicht einzeln belegt werden, da es für Teilnehmer*innen mit bereits gutem theoretischem Hintergrund eine optimale Ergänzung wäre. Trotzdem ist die Belegbarkeit und Zertifizierung einzelner Module sehr praktisch und auch der Anteil an E-Learning besonders hoch. Für Interessierte, die eine Ausbildung mit hohem Präsenzanteil suchen, ist dieses Angebot allerdings ungeeignet.

Data Train: Training in Research Data Management & Data Science

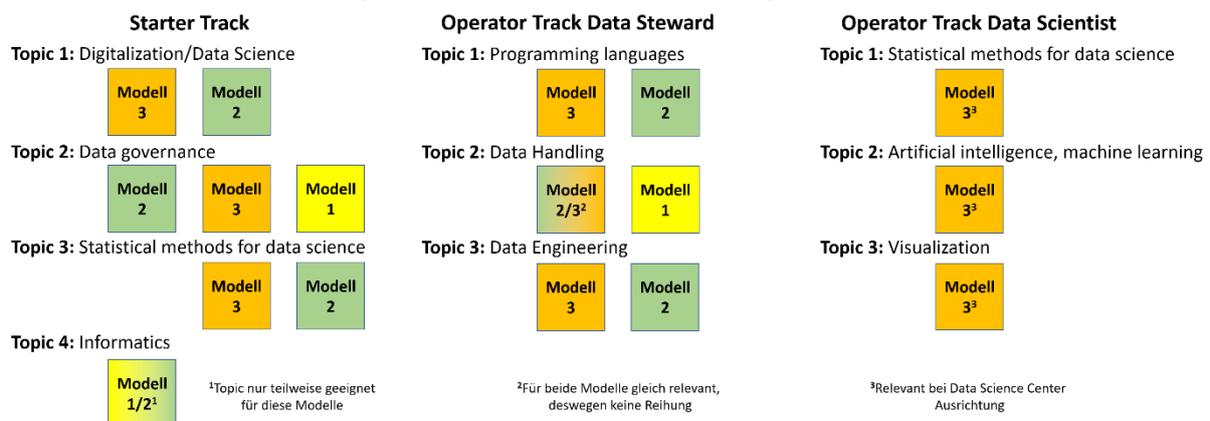


Abbildung 3: Data Train mit den beurteilten Topics in Hinblick auf die jeweiligen Data Steward Modelle

Das Data Train Programm als Ganzes, aber besonders der Starter Track, legt eine sehr gute und breite Basis für das Thema Forschungsdatenmanagement und Data Science und spart auch das Thema Awareness Building nicht aus.

Prinzipiell geben alle drei Tracks einen guten Überblick über die verschiedensten Themen und es ist positiv hervorzuheben, dass das Programm zu Teilen auch für organisationsfremde Personen verfügbar ist. Beim Starter Track reicht eine einfache Anmeldung aus und bei den beiden Operator Tracks werden nach verfügbaren Plätzen auch externe Personen ohne zusätzliche Kosten aufgenommen. Die Qualität der Beschreibung der Lehreinheiten ist sehr hoch, sowohl für den Starter als auch den Data Steward Track¹⁵. Bei den Vortragenden handelt es sich um Experten, die Großteiles aus der Uni Bremen Research Alliance stammen und die Kurse interaktiv gestalten, es können also jederzeit Fragen gestellt

¹⁵Zum Zeitpunkt des Workshops waren die Inhalte des Data Science Tracks nicht bekannt, aber es wird erwartet, dass das Niveau gehalten wird.

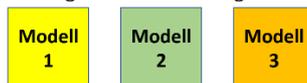
und Probleme behandelt werden. Die Tracks sind modular aufgebaut und können zusammengestellt werden wie man möchte, auch weil es keine echten/bindenden Voraussetzungen gibt, nur Empfehlungen. Fehlende Inhalte des Programms sind allerdings die Bereiche Fördergeber/Förderrichtlinien und Didaktik. Generell sind die Inhalte nicht für Modell 1 geeignet oder bestenfalls nur nachrangig.

Selbst im Starter Track steht Modell 1 bei keinem Themenblock an erster Stelle und wird auch nur beim „Data Governance“ Block als gerade noch passend eingestuft. Die grundlegenden Kurse orientieren sich inhaltlich trotzdem mehr an Modell 2 und 3 und ein ganzer Themenblock namens „Informatics“ wird überhaupt als ungeeignet für alle Modelle angesehen. Ausgenommen davon sind die ersten beiden Themen in diesem Block, die sich mit Grundlagen des Programmierens auseinandersetzen und damit weiterhin relevant bleiben. Kompetenzen die im vorangegangenen Workshop² Modell 3 zugeordnet wurden, sind durch das gesamte Data Train Programm gut abgedeckt. Der Operator Track Data Steward befasst sich im ersten Block mit dem Thema Programmieren (MATLAB, R, Python) und ist somit sicher wichtig für Modell 3. Teilweise ist er auch für Modell 2 relevant aber sicher nicht mehr für Modell 1. Der Themenblock „Data Handling“ wird inhaltlich bei Modell 2 und 3 als gleichwertig angesehen und auch Modell 1 hat seinen Anteil beim Unterpunkt „Tools for FAIR data handling: DMP tool“. Modell 1 ist auch dritten Themenblock (Data Handling) nicht mehr vertreten und hauptsächlich für Modell 3 aber auch für Modell 2 interessant.

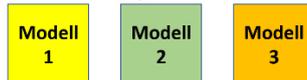
Der Operator Track Data Scientist ist für keines der Data Stewardship Modelle direkt einsetzbar, allerdings bedingt in Modell 3 wenn es bereits eine gewisse Ausrichtung zu einem Data Science Center gibt (z.B: University of Arizona) oder es gewünscht ist, dass die Stewards diese Spezialisierung besitzen. Modell 1 wird vom Data Train Programm kaum abgedeckt und wird im Self-Assessment Tool wohl hier kaum vorkommen. Am ehesten ist das Programm für Modell 2 geeignet und aufgrund des modularen Aufbaus kann auch das Modell 3 bedient werden.

Zertifikatskurs Data Steward

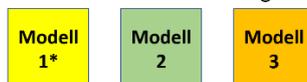
Modul 1: Grundlagen des Forschungsdatenmanagements und Open Science



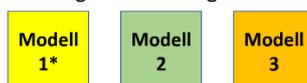
Modul 2: FAIRe Forschungsdaten im Lebenszyklus



Modul 3: IT und Data Science Grundlagen



Modul 4: Forschungsdatenmanagement-Support



*Optionales bzw. weniger relevantes Modul für dieses Modell

Abbildung 4: Zertifikatskurs Data Steward mit den beurteilten Modulen in Hinblick auf die jeweiligen Data Steward Modelle

Basiert auf der Gruppendiskussion eignet sich der geplante Zertifikatskurs Data Steward für alle drei Data Stewardship Modelle, indem es die Grundlagen von den wesentlichen Themen vermittelt und eine berufsbegleitende Durchführung geplant ist.

Als positiv bewertete die Gruppe die Zielgruppe des Kurses, eine Kombination aus Forschenden und Forschungssupport, die über Berufserfahrung in relevanten Bereichen verfügen. Die Entwicklung des Angebots anhand des bestehenden Kurses Data Librarian sowie ähnlichen Aus- und Fortbildungsangeboten aus anderen Ländern wurde ebenso gut angenommen. Der Kurs könnte einen

schnellen Start in die Thematik für neue Data Stewards in Österreich bieten. Mit einer ersten Durchführung in 2022 hätten österreichische Forschungseinrichtungen bereits nach einem Jahr ausgebildete Data Stewards zur Verfügung.

Der angemessene Arbeitsaufwand sowie die Konzeption des Kurses als eine hybride berufsbegleitende Fortbildung im Sinne der Digitalisierung, wurden als besonders positiv angesehen. Die Präsenzveranstaltungen sollten vor allem auf praktische Workshops beschränkt werden. Möglichst viele Unterrichtseinheiten sollten online abgehalten werden, um Berufstätigen die Teilnahme zu erleichtern.

Das Niveau eines Masterabschlusses (Stand Juli 2021) als Voraussetzung wird als zu hoch angesehen. Es sollten mehr Anreize geschaffen werden, um Interesse für die Belegung des Kurses zu generieren. Langjährige Forschungserfahrung, die nicht zwingend mit einem akademischen Titel verbunden sein muss, sollte als Hauptkriterium für die Aufnahme gelten. Stand September 2021: Die Teilnahmevoraussetzungen wurden in einem Stufenmodell ausgearbeitet. Es könnten demnach bei nachgewiesener einschlägiger Berufserfahrung auch Personen mit einem abgeschlossenen Lehrberuf, einer Matura oder einem Bachelorabschluss den Zertifikatskurs absolvieren. Die Teilnehmer*innenzahl sollte noch reguliert werden, da dies große Auswirkungen auf die Gestaltung mit sich bringt (Stand Juli 2021: Anzahl der Teilnehmenden noch offen). Stand September 2021: Die Anzahl wird auf ca. 20 Personen beschränkt.

Generell fehlt eine genaue Definition der Rolle von Data Stewards, da sich diese an österreichischen Forschungseinrichtungen derzeit erst etabliert. Dies sollte in der Konzeption und bei der Durchführung des Kurses berücksichtigt werden und auch mit den Teilnehmenden kritisch diskutiert werden. Die Vorteile und Karrierewege sollten für die Kursteilnehmenden klar sichtbar sein, um Kosten und Zeitaufwand zu rechtfertigen. Es sollten Wege gezeigt werden, wie die gelehrten Inhalte in Praxis umgesetzt werden können, z. B. über eine abschließende Projektarbeit. Data Stewardship Aufgaben, die derzeit von verschiedenen Personen an Forschungseinrichtungen ausgeübt werden, werden meistens nicht anerkannt und/oder entlohnt. Dies ist für die Sichtbarkeit und Akzeptanz des Berufs nicht förderlich. Formalisierte Ausbildungsangebote wie der geplante Zertifikatskurs können dazu beitragen, diesen Umstand zu ändern.

Der Kurs selbst besteht aus vier obligatorischen Modulen (siehe Einführung). Im Grunde sind alle Module für alle Modelle relevant, besonders die Module 1 und 2. Modul 3 (IT und Data Science Grundlagen) scheint für Modell 2 und 3 besser geeignet, für Modell 1 sollte dies eher als optional angeboten werden. Auch Modell 4 (Forschungsdatenmanagement-Support) ist für alle Modelle relevant. Obwohl bei Modell 1 verstärkt Anfragen an die richtigen Stellen weitergeleitet und nicht selbst bearbeitet werden, so sollte es trotzdem in Betracht gezogen werden, um sich auf mögliche Änderungen des Modells an der eigenen Einrichtung vorzubereiten.

Das Modul 3 und 4 (IT und Data Science Grundlagen und Forschungsdatenmanagement-Support) werden für das Modell 1 (1 Person/Einrichtung) als nur bedingt relevant betrachtet. Hierbei wurde betont, dass vor allem technisches Know-how von einer einzigen Person an einer Forschungseinrichtung neben der Vielfalt anderer notwendigen Kenntnisse und Fähigkeiten realistisch nicht verlangt werden kann. Weiters sind Data Stewards im Modell 1 laut den Teilnehmenden wenig in technische Entwicklungen oder die aktive Konzeption und Abhaltung von Trainings eingebunden. Es wird empfohlen, die Module auch einzeln als Fortbildungen anzubieten, da sich der Aufbau des Kurses hierfür gut eignet.