

Konzeptpapier Data Stewards

Im Folgenden wird das Konzept für das „**Rent a Data Steward**“-Programm vorgestellt.

Dem Pilotcharakter des Projektes entsprechend, bildet das Konzept die bisherigen Überlegungen für ein gutes Vorgehen im Rahmen des Programms ab. Es ist keinesfalls als starr anzusehen, sondern wird mit den im Laufe des Projektes gesammelten Erfahrungen stetig weiterentwickelt.

1 Was sind Data Stewards und wie werden sie in Thüringen eingesetzt?

Obwohl Forschungsdatenmanagement (FDM) tatsächlich immer öfter als selbstverständliche Kernaufgabe wissenschaftlichen Arbeitens erkannt wird, fehlen in den Forschungsgruppen¹ häufig die Erfahrung zur Verankerung der FDM-Prozesse in den eigenen Abläufen. Oft fehlen auch personelle Kapazitäten für diesen Teil eines jeden Forschungsvorhabens.

Neben verfügbaren, einfach erreichbaren und niederschweligen Beratungsangeboten zum FDM, benötigen Forschungsgruppen oftmals konkrete Unterstützung bei der Organisation ihres FDMs im Team.

An den Thüringer Universitäten gibt es etablierte Stellen zur Unterstützung des FDM, an den Hochschulen für angewandte Wissenschaften werden diese Angebote aufgebaut. Diese Stellen haben in erster Linie beratende Aufgaben. Für sie ist es selten möglich, konkrete Lösungen gemeinsam mit den Forschenden praktisch umzusetzen. Ebenso ist es für diese Stellen in der Regel nicht möglich, ganze Gruppen über einen längeren Zeitraum mit einer "Innensicht" auf den gruppeneigenen Umgang mit Forschungsdaten zu begleiten.

An dieser Stelle setzen Data Stewards an. Durch den zeitlich begrenzten Einsatz innerhalb einer Forschungsgruppe und den damit einhergehenden unmittelbaren Einblick in die fachspezifischen Gegebenheiten und Anforderungen haben sie die nötige Nähe und Einbindung, um die Herausforderungen "von innen" zu verstehen und konkrete Lösungen zur Verbesserung des FDM zu entwickeln. Bei auftretenden Hindernissen oder Problemen können sie gemeinsam und in der Gruppe geeignete Wege zur deren Beseitigung finden.

Ein Data Steward unterstützt in einer Forschungsgruppe bei der Analyse der Anforderungen, der Qualitätssicherung von Daten und Metadaten, der Auswahl geeigneter Speicherinfrastrukturen und -strategien, der Dokumentation der Datenquellen und Datenverarbeitungsschritte, sowie bei der Regelung von Rechten und Pflichten bei der Datenverarbeitung. Dabei greift er oder sie bevorzugt auf vorhandene Infrastrukturen zurück. Soweit im Einzelfall möglich, wird der Data Steward außerdem unterstützen, dass Forschungsdaten für andere im Sinne der FAIR-Prinzipien auffindbar, verfügbar, interoperabel und nachnutzbar gemacht werden. Die Begleitung einer

¹ Hierunter wird im Folgenden eine Gruppe von Forschenden verstanden, die in der Regel von einer Professorin oder einem Professor angeleitet oder von einer Nachwuchsgruppenleitung geführt wird, insbesondere Fachgebiete, Lehrstühle oder Arbeits- oder Nachwuchsgruppen.

Forschungsgruppe zielt darauf ab, konkrete Lösungen auf strategischer, technischer und sozialer Ebene zu entwickeln und diese zusammen zu implementieren und in gelebte Praxis zu überführen.

2 Einsatzgebiete, Aufgaben und Selbstverständnis der Thüringer Data Stewards

Das Ziel der Data Steward-Begleitungen ist es, ein effizientes und nachhaltiges FDM in die Forschungsprozesse einer Arbeitsgruppe zu integrieren. Die beteiligten Forschenden sollen nach dem Einsatz mit aktuellen Konzepten des FDM vertraut sein und die Umsetzung in Eigenverantwortung durchführen können.

Die Data Stewards unterstützen die Forschungsgruppe dabei, zu ihrem Bedarf passende Lösungen zu finden und umzusetzen. Zu erarbeitende Lösungen orientieren sich stärker an der Umsetzbarkeit als am theoretischen Idealzustand.

Der Zugang zu personellen Ressourcen, Ausstattung, Verfügbarkeit von Diensten und Infrastrukturanangeboten ist an den Thüringer Hochschulen und in den einzelnen Forschungsgruppen mitunter sehr unterschiedlich ausgeprägt. Das Angebot der Data Stewards muss diese unterschiedlichen Ausgangspunkte berücksichtigen.

Neben der Erschließung der bestehenden Infrastrukturanangebote für die Forschungsgruppen sollen Prozesse und Strukturen zum guten Umgang mit Forschungsdaten im Team und in den Projekten etabliert werden. Konsensuale Einigungen beispielsweise hinsichtlich Verantwortlichkeiten, Ansprechpersonen, Ordnerstrukturen und Namenskonventionen oder Onboarding für neue Mitarbeitende begünstigen die Akzeptanz und Umsetzung im ganzen Team.

Durch ihre Kenntnis und ihren Austausch mit den institutionellen, hochschuleigenen und hochschulübergreifenden Stellen, die jeweils bei verschiedenen Aspekten des FDM weiterhelfen können, können Data Stewards die richtigen Verbindungen vermitteln.

Mit einer Haltung des gegenseitigen Lernens von einander, werden diese Verbindungen aber nicht nur in Richtung der Forschenden geknüpft, sondern werden Bedarfe der Forschenden auch an die infrastrukturegebenden Institutionen wie etwa die Bibliotheken, die Rechtsberatungsstellen, die Rechenzentren oder das IT-Zentrum der Thüringer Hochschulen (HS-ITZ) zurückgekoppelt. Die Erkenntnisse aus den Einsätzen inklusive der innerhalb der Einsätze entwickelten Lösungen und Werkzeuge sollen dabei als thüringenweite Best-Practice dienen.

Die Arbeit mit den Data Stewards wird zum großen Teil in Workshops stattfinden. Auf diese Workshops wird in Abschnitt 5.3 „Arbeitsphase“ eingegangen.

3 Anbindung an IT/ Rechenzentren/ HS-ITZ

Da die Data Stewards auf bestehende oder geplante IT-Infrastruktur der Thüringer Hochschulen, ihrer Rechenzentren und des HS-ITZ zurückgreifen sollen, ist es erforderlich, dass sie in regelmäßigem und engem Austausch mit diesen Einrichtungen stehen, um die Rahmenbedingungen der Angebote zu kennen, um über aktuelle Entwicklungen im Bilde zu sein und schließlich um identifizierte Bedarfe der Forschenden an neuen Diensten mit den genannten Einrichtungen zu diskutieren und ggf. benannte Veränderungswünsche der Forschenden zu

kommunizieren. Die Data Stewards sind Schnittstelle und Beförderer dieses Austausches zwischen Infrastrukturanbietern und Forschenden.

4 Inhalt des Einsatzes eines Data Stewards in einer Forschungsgruppe und formale Voraussetzungen

Der Einsatz eines Data Stewards ist nur möglich, wenn die folgenden Kriterien erfüllt sind:

- Die anfragende Forschungsgruppe ist an einer Thüringer Hochschule verortet.
- Als gemeinsames Ziel wird die nachhaltige Verbesserung des FDM in einer Forschungsgruppe vereinbart. Diese absehbare Nachhaltigkeit eines Einsatzes ist ein wesentliches Einsatzkriterium.
Die Begleitung durch die Data Stewards ist daher nicht zuerst für die Unterstützung von Forschungsprojekten mit kurzer oder sehr begrenzter Laufzeit vorgesehen, aber durchaus für die Begleitung von Forschungsgruppen, in denen solche Projekte durchgeführt werden.
- Die Anfrage muss von einem Professor oder einer Professorin an einer öffentlichen Hochschule stammen. Berechtig sind ebenfalls Juniorprofessor_innen und Nachwuchsgruppenleiter_innen an diesen Hochschulen.
- Die Anfrage kann nicht von einem An-Institut gestellt werden.

Im Folgenden sind beispielhaft mögliche Gegenstände der Data Steward-Einsätze aufgezählt:

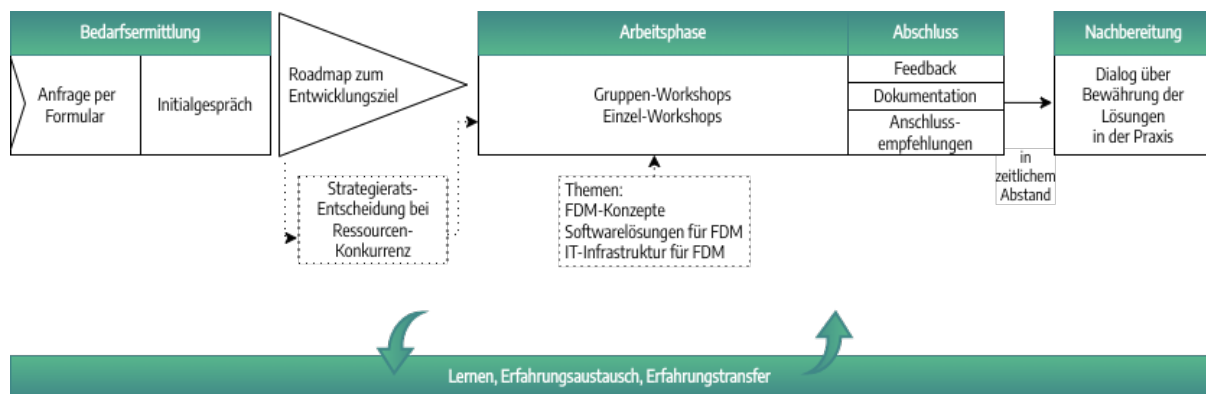
- FDM-Diagnose, Analyse des gegenwärtigen Standes
 - dient zur Bestandsaufnahme, um geeignete Maßnahmen der Begleitung ableiten und vereinbaren zu können
- Begleitung bei der Erstellung und Einführung von FDM-Konzepten in einer Gruppe (Ordnerstrukturen, Namenskonventionen, Übereinkünfte zum Publizieren, ...)
 - In Zusammenarbeit und im direkten Austausch mit den Forschenden
 - angepasst, an den bereits etablierten Umgang der Gruppe mit ihren Forschungsdaten
- Etablierung von Softwareanwendungen aus der IT-Infrastruktur und Verankerung von Prozessen
 - Einschließlich passender Workshops zu ausgewählten Fokusthemen, wie etwa Namenskonventionen, GitLab, Langzeitarchivierung, ...
- Evaluation der Etablierung erarbeiteter FDM-Lösungen
 - Fokus auf eigenständige Umsetzbarkeit in der begleiteten Forschungsgruppe
 - Gemeinsame Bewertung des Etablierungsstandes von Prozessen
 - ggf. nachsorgender Austausch in größer werdenden zeitlichen Abständen

Die Data Stewards sollen nicht für die folgenden Aufgaben eingesetzt werden:

- Antragsberatungen (wie etwa zu Datenmanagementplänen in DFG-Projektanträgen)
- Einzelfallberatungen (etwa zu Datenschutz, Veröffentlichungen, ...)
- Einführung und Dauerbetreuung von Anwendungen, die besser von einem zentralen Infrastrukturanbieter bereitgestellt werden sollte (bspw. Elektronische Laborbücher/ ELNs). Sollten doch spezielle Anwendungen eingeführt werden, muss eine Verantwortungsübergabe erfolgen.
- Datenmanager oder Data Steward über die gesamte Laufzeit eines Verbundprojektes (z.B. SFB, SPP). Die dafür notwendigen Mittel sind im Rahmen der Antragstellung einzuwerben.

Im ersten Jahr des "Rent a Data Steward"-Programms arbeiten die Data Stewards exklusiv lokal in der eigenen Hochschule, um das vorliegende Konzept zu testen und Erfahrungen zu sammeln.

In der folgenden Grafik ist der Ablauf eines Einsatzes in einer Forschungsgruppe in allgemeiner Form und vereinfacht dargestellt. Die Dauer einzelner Termine im Laufe eines Betreuungsverhältnisses richtet sich nach den Gegebenheiten und Anforderungen in der Forschungsgruppe.



5 Ablauf eines Data Steward-Einsatzes

5.1 Anfrage

Im Vorfeld eines Data Steward-Einsatzes in einer Forschungsgruppe bedarf es der Anfrage eines Professors oder einer Professorin über ein Formular, welches auf der TKFDM-Webseite² dafür bereitgestellt wird.

Neben den Grunddaten wie z.B. Name der Einrichtung, Kontaktpersonen und der Leitung der Gruppe, werden hier als Grundlage für die Bedarfsermittlung die nachfolgenden Angaben erfragt:

- Wie groß ist die zu begleitende Gruppe?
- Gibt es in der Gruppe eine Person, die sich um das Datenmanagement kümmert

²www.forschungsdaten-thueringen.de

- Wurden bereits Daten publiziert oder sind Daten-Publikationen geplant?
- Welche Erwartungen haben Sie an den Data Steward?
- Wieviel Zeit wird Ihrer Einschätzung nach für die Data Steward-Begleitung benötigt?
- Gibt es Geheimhaltungsobliegenheiten im Projekt (Datenschutz, Interessen Dritter)?

5.2 Vorgespräch zur Bedarfsermittlung

Das Formular ist der erste Schritt in einer mehrstufigen Ermittlung des Unterstützungsbedarfs, dessen Umfang sich im Verlauf des Initialgesprächs konkretisiert. In diesem Gespräch werden im direkten Dialog diejenigen Informationen zusammengetragen, die das Team des TKFDM benötigt, um einschätzen zu können, welcher Bedarf existiert und welcher zeitliche Aufwand dafür eingeplant werden muss. Bei Bekanntwerden von formalen Gründen im Sinne der o.g. Einsatz- und Nichteinsatzszenarien muss die Betreuung abgelehnt werden.

Sollte der erkannte Bedarf im Rahmen einer initialen Beratung bedient und das zu begleitende Anliegen so geklärt werden können, sind keine weiteren Maßnahmen nötig.

Anderenfalls wird in gemeinsamer Arbeit die Bestandsaufnahme fortgeführt und eine Roadmap der notwendigen Schritte zum Entwicklungsziel erstellt. Bei Anfragen, die nicht in Konkurrenz mit anderen Anfragen stehen, setzt das TKFDM die vorgeschlagenen Maßnahmen im vollen Umfang um.

Beansprucht ein Anliegen einen erheblichen Anteil der Ressourcen des TKFDM, entscheidet der Strategierat Digitale Forschung über die Umsetzung der vorgeschlagenen Maßnahmen und deren Umfang.

5.3 Arbeitsphase

Workshops zu Fokusthemen sind ein wichtiger Bestandteil der Begleitung. Diese ergänzen das bestehende Workshop-Angebot der lokalen FDM-Kontaktstellen und leisten den Transfer der Inhalte aus dem Schulungsangebot der regionalen, nationalen und internationalen FDM-Netzwerke in die gruppeneigene Forschungspraxis.

Eigene Workshops z.B. zu Infrastruktur, der Gestaltung von FDM-Prozessen oder auch zu einzelnen Softwarelösungen können sich an Gruppen oder Einzelpersonen richten. In der Arbeit mit solchen Fokusthemen können dann inhaltliche Aspekte aus der folgenden Aufzählung einfließen:

- FDM-Konzepte
 - Forschungsdaten-Lebenszyklus, Leitlinien, Rechtliche Fragen
 - Datenorganisation, Ordnerstrukturen, Namenskonventionen
 - Datendokumentation, Metadaten(-standards)
 - Datenqualität
 - Datenspeicherung, Backup

- Dateiformate
- Datenpublikation nach FAIR-Prinzipien
- Datenarchivierung
- Softwarelösungen für FDM
 - Elektronische Laborbücher (ELN), z.B. elabFTW, Chemotion
 - Versionsverwaltung, z.B. mit GIT/GitLab
 - Skripten/Dokumentieren, z.B. mit JupyterHub
 - Dateiverwaltung, z.B. mit Nextcloud
 - IT-Infrastruktur für FDM
 - Speicherung von Projektdaten
 - Veröffentlichen von Daten
 - Langzeitarchivierung

Die Auswahl der Fokusthemen erfolgt im Hinblick darauf, wie einfach und effektiv Verbesserungen im alltäglichen Umgang mit Forschungsdaten erreicht werden können.

Der tatsächliche zeitliche Umfang der Begleitung richtet sich nach den identifizierten Problemstellungen und nach dem identifizierten Unterstützungsbedarf. Die Begleitung durch die Data Stewards wird für kurze und begrenzte Dauer vereinbart, die drei Monate in der Regel nicht überschreiten sollte.

5.4 Abschluss, Nachbereitung, Evaluation

Mit Abschluss der Begleitung wird die Dokumentation vervollständigt und ein strukturiertes Feedback eingeholt.

In zeitlichem Abstand zum Abschluss einer Begleitung soll wiederum im Dialog mit der Gruppe geschaut werden, ob das gemeinsam erarbeitete Konzept den tatsächlichen Anforderungen der Arbeitsgruppe entspricht und sich in der Praxis bewährt hat, inwieweit Anpassungen vorgenommen werden müssen und ob ggf. neue begleitende Impulse von Seiten eines Data Stewards hilfreich sind.