

**Bausteine Forschungsdatenmanagement**  
Empfehlungen und Erfahrungsberichte für die Praxis von  
Forschungsdatenmanagerinnen und -managern

**DataPLANT – Ein NFDI-Konsortium der  
Pflanzen-Grundlagenforschung<sup>i</sup>**

Dirk von Suchodoletz

Timo Mühlhaus

Jens Krüger

Björn Usadel

Cristina Martins Rodrigues

2021

**Zitiervorschlag**

von Suchodoletz, Dirk et al. 2021. DataPLANT – Ein NFDI-Konsortium der Pflanzen-Grundlagenforschung. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 2/2021: S. 46-56. DOI: [10.17192/bfdm.2021.2.8335](https://doi.org/10.17192/bfdm.2021.2.8335).

Dieser Beitrag steht unter einer  
[Creative Commons Namensnennung 4.0 International Lizenz \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

<sup>i</sup>Dirk von Suchodoletz (ORCID: [0000-0002-4382-5104](https://orcid.org/0000-0002-4382-5104)), Timo Mühlhaus (ORCID: [0000-0003-3925-6778](https://orcid.org/0000-0003-3925-6778)), Jens Krüger (ORCID: [0000-0002-2636-3163](https://orcid.org/0000-0002-2636-3163)), Björn Usadel (ORCID: [0000-0003-0921-8041](https://orcid.org/0000-0003-0921-8041)), Cristina Martins Rodrigues (ORCID: [0000-0002-4849-1537](https://orcid.org/0000-0002-4849-1537))

# 1 Zusammenfassung

Gemeinsam mit anderen Fachgebieten steht die Pflanzen-Grundlagenforschung vor der Herausforderung, ein Forschungsdatenmanagement zu etablieren, das sowohl den Bedürfnissen der Community entspricht als auch die Sammlung, die Verarbeitung, den Austausch und die Archivierung von Forschungsdatensätzen basierend auf den FAIR-Prinzipien erlaubt. Auf dieses Ziel ausgerichtet, ist die Arbeit des DataPLANT Konsortiums in vier Aufgabenbereiche gegliedert. In Task Area I werden Themen wie Datenqualität, Interoperabilität und Standardisierung vorangetrieben. Hierfür werden in einem ersten Schritt weit verbreitete Ontologien und deren Anwendung unterstützt. Weitergehend soll die Vervollständigung dieser durch die Ergänzung neuer Terme für den/die Anwender\*in möglichst reibungslos erfolgen. Durch Task Area II (Infrastruktur, Software und Service) wird die infrastrukturelle Basis für die entwickelten Tools geschaffen. Der DataPLANT-Hub soll dabei als zentraler Zugangspunkt dienen. Die personelle Unterstützung der Forschenden wird durch Data Stewards gewährleistet, deren Ausbildung, sowie die Koordination des Community-Kontakts, in den Aufgabenbereich der Task Area III fällt. Schließlich übernimmt Task Area IV die Kommunikation innerhalb des Projektes, ebenso wie die Einbindung in die Gesamt-NFDI und die Förderung gemeinsamer Bestrebungen. Ein erstes Konzept für die Realisierung eines der zentralen Anliegen, nämlich der nachhaltigen Gestaltung der NFDI, wurde bereits erstellt und in dem vorliegenden Text diskutiert.

## 2 NFDI Konsortium DataPLANT

In der modernen hypothesen-basierten Forschung sind Forschende zwingend auf Dienste und Infrastrukturen für Forschungsdatenmanagement (FDM) angewiesen, welche die Erfassung, die Verarbeitung, den Austausch und die Archivierung von Forschungsdatensätzen erleichtern. Dabei wird eine Verknüpfung der interdisziplinären Expertise mit dem Vergleich und der Integration diverser Analyseergebnisse erst durch ein zeitgemäßes FDM ermöglicht und bildet somit die Basis für zusätzliche Erkenntnisgewinne. Das Projekt DataPLANT<sup>1</sup> hat zum Ziel diesen Mehrwert für die Pflanzen Grundlagenforschung zu schaffen. Auf diesem Fachgebiet werden die (molekularen) Prinzipien des pflanzlichen Lebens untersucht, welche beispielsweise das Pflanzenwachstum, den Ernteertrag oder die Biomasseproduktion bestimmen. Die hierzu eingesetzten Methoden von umfassenden Genexpressions- (Transkriptomik), Proteinabundanz- (Proteomik) über Metabolitverteilungen (Metabolomik)-Analysen bis hin zu bildgebenden Verfahren erzeugen hochdimensionale, polymorphe Daten, die verarbeitet, fusioniert und interpretiert werden müssen. Eine erfolgreiche Nutzung von Daten unterschiedlicher Modalitäten – aus vielen Quellen und Experimenten, vorverarbeitet oder analysiert mit einer Vielzahl von Algorithmen – erfordert eine Kontextua-

<sup>1</sup>Homepage und Community-Portal von DataPLANT, <https://www.nfdi4plants.de>, aufgerufen am 31.01.2021.

lisierung der Daten. Hierzu zählt die Annotation mit detaillierten Metadaten ebenso wie ein eindeutiges Referenzieren der jeweiligen Daten mit ihren Abhängigkeiten. Die FAIR-Data<sup>2</sup> and Linked-OpenData<sup>3</sup>-Prinzipien bieten entscheidende Richtlinien für den verantwortungsvollen Umgang mit Forschungsdaten.

Verschiedene Initiativen wie die Research Data Alliance oder die Verabschiedung von FDM-Policies an den Universitäten haben Vorschläge zu zielführenden Vorgehensweisen zur Erfüllung dieser Grundsätze gemacht, doch liegt es fast immer in der Verantwortung der Forschenden, diese auch umzusetzen. Daher stehen umfassende Informationen über die erforderliche Qualität der Datensätze nur in seltenen Fällen zur Verfügung. Es hat sich gezeigt, dass Forschende bei der Nutzung der fragmentierten und komplexen Ressourcenlandschaft praktische Unterstützung benötigen<sup>4</sup>. Dies erhöht die Notwendigkeit einer maßgeschneiderten Infrastruktur für das FDM. Durch den Zusammenschluss von technisch-fachlicher Expertise in den Bereichen Pflanzen-Grundlagenforschung, Informations- und Computerwissenschaften und Infrastrukturspezialisten werden Forschende der Pflanzenwissenschaften im Umgang mit Forschungsdaten individuell unterstützt. Dabei soll eine Serviceumgebung geschaffen werden, um Forschungsdaten nach den FAIR-Prinzipien und mit minimalem Zusatzaufwand zu kontextualisieren und den gesamten Forschungszyklus in der Pflanzenbiologie zu begleiten. Die angestrebte Servicelandschaft in DataPLANT wird sich aus technisch-digitaler Assistenz sowie personellem Vor-Ort-Support zusammensetzen. Auf diese Weise soll ein zentraler Einstiegspunkt und eine fachspezifische Daten- und Wissensressource geschaffen werden. In Verbindung mit Lehre und Trainingskonzepten soll das Sachverständnis im Umgang mit Daten gestärkt und eine Langzeitmotivation zur Erzeugung wohlannotierter Datenobjekte mit ihren jeweiligen Kontexten erzeugt werden. Durch die Integration der Pflanzenwissenschaft in die Gesamtstruktur der Nationalen Forschungsdateninfrastruktur (NFDI<sup>5</sup>), treibt DataPLANT den digitalen Wandel und die Demokratisierung der Forschungsdaten im Feld voran. Dies geschieht einerseits durch den Austausch mit den anderen Konsortien zu Standardisierung, Ontologien und Metadaten und andererseits durch das Vorantreiben des Konzepts der Datenpublikation. Diese professionelle Infrastruktur (Hardware + Data Stewards) erlaubt eine Unterstützung aller Gruppen, so dass ein gutes Forschungsdatenmanagement für alle ohne Zugangshürde gewährleistet werden kann.

In DataPLANT werden die wichtigsten FDM Kernpunkte aus Sicht der Pflanzenwissenschaften durch die Zusammenarbeit von vier Task Areas adressiert (**Abbildung 1**). Der Fokus in Task Area I liegt auf der Standardisierung in der Pflanzen-

<sup>2</sup>Vgl. Mark D. Wilkinson, Michel Dumontier, IJsbrand J. Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg *et al.* „The FAIR Guiding Principles for Scientific Data Management and Stewardship“ *Scientific Data* 3, Nr. 1 (2016): 1-9. <https://doi.org/10.1038/sdata.2016.18>

<sup>3</sup>Vgl. Christian Bizer *et al.* „Linked Data on the web (LDOW2008)“ *Proceedings of the 17th international conference on World Wide Web*, (2008): 1265-1266. <https://doi.org/10.1145/1367497.1367760>

<sup>4</sup>Vgl. Mark Hahnel, Briony Fane, Jon Treadway, Grace Baynes, Ross Wilkinson, Barend Mons, *et al.* „The State of Open Data Report 2018“ *Digital Science*, (2018). <https://doi.org/10.6084/m9.figshare.7195058.v2>

<sup>5</sup><https://www.nfdi.de>, aufgerufen am 31.01.2021.

Grundlagenforschung durch klar definierte Protokolle, erweiterbare Ontologien und Kommunikation mit den fachspezifischen Gremien. In Task Area II wird eine gemeinsame technische Infrastruktur geschaffen, die schrittweise im Projekt mit der Community abgestimmt und weiterentwickelt wird. Durch Task Area III wird die persönliche Unterstützung von Forschenden, Gruppen und Verbänden in allen Belangen des Datenmanagements gebündelt. Zusätzlich zu den auf fachliche Belange bezogenen Aktivitäten wird in Task Area IV die Abstimmung innerhalb des Projekts und mit der Gesamt-NFDI und die übergreifende Integration der gemeinsamen Aufgaben, die alle Konsortien betreffen, koordiniert.

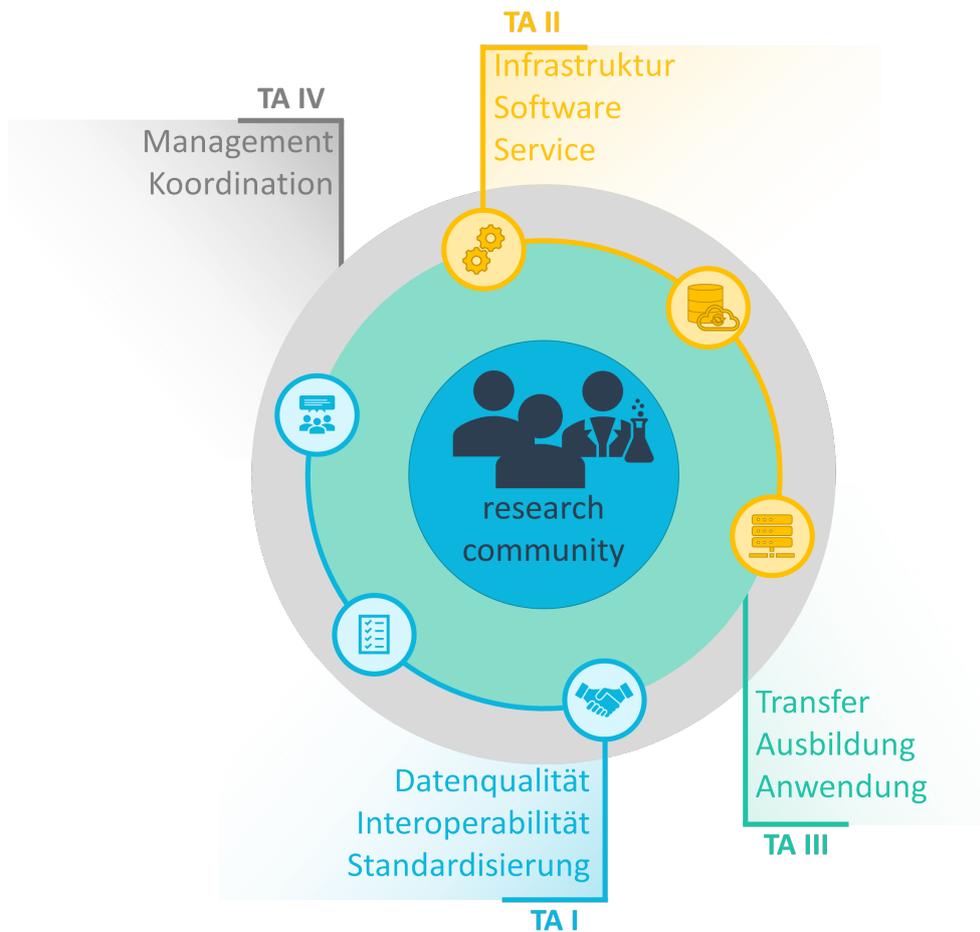


Abbildung 1: DataPLANT: Die treibende Kraft des digitalen Wandels in der Pflanzenwissenschaft.

### 3 Vortreiben der Standardisierung in Task Area I

Standardisierung schafft die Grundlage für den Austausch von Daten und Erkenntnissen im Fachgebiet und über dieses hinaus sowie die Vergleichbarkeit von Messungen und Ergebnissen. Die dazu notwendige Koordination der Etablierung neuer und der Weiterentwicklung bestehender Standards ist insbesondere von Nachwuchsfororschenden und kleinen Arbeitsgruppen kaum leistbar. In der Tat können kleine Gruppen alleine kaum einen Überblick über alle Werkzeuge, Initiativen und Standards halten. Es erfordert einen guten Überblick im Fachgebiet und über dessen Grenzen hinaus, sowie eine gute Vernetzung mit allen relevanten Gremien und Akteuren. In Task Area I sollen neben der Festlegung von Metadatenstandards vor allem bestehende Ontologie benutzt werden. Als Beispiel dienen die oftmals genutzten Gene<sup>6</sup>-, MapMan<sup>7</sup>-, KEGG<sup>8</sup>-Ontologien, welche eine Genfunktion beschreiben, wie auch z. B. die Plant Ontology (PO) mit der z. B. beschrieben werden kann, dass Material aus der „Wurzel“ entnommen wurde.

Diese bestehenden Ontologien sollen dann iterativ an die Anforderungen der Forschenden und deren spezifische Experimente angepasst werden, um Experimente umfassend zu beschreiben. In Zusammenarbeit mit den Data Stewards aus Task Area III werden diese in Form von Metadaten Templates an die Forschenden wechselseitig kommuniziert. Sich weiterentwickelnde Ontologien stehen innerhalb eines Tools automatisch zur Auswahl zur Verfügung. Um die Ziele von FAIRen Daten und Open Access Prinzipien zu erreichen, wird auf Data Management Pläne gesetzt, die durch in Task Area I erarbeitete Richtlinien für die Metadaten-Annotation, die Archivierung und Lizenzmodelle vorgeben.

Der oft lange Weg von der Feststellung eines fehlenden Terms oder Parameters bis zur endgültigen Übernahme in die beschreibenden Referenzontologien soll deutlich verkürzt und komfortabler gestaltet werden. Die Forschenden sollen von der zeitraubenden direkten Interaktion mit den Organisationen für Standardisierung durch geeignete Workflows und Tools entlastet werden. Ein erster Ansatz hierfür wird mit dem von DataPLANT entwickelten SWATE (**S**wate **W**orkflow **A**notation **T**ool for **E**xcel)-Tool<sup>9</sup> verfolgt. Es übernimmt die Aufgabe der Vermittlung durch Integration in die Protokolle der Forschenden direkt auf seinem digitalen Arbeitsplatz. Es übernimmt den Vorschlag neuer Terme, so dass Wissenschaftler fehlende Beschreibungen zunächst in Freiform angeben können und leitet sie über das für die Nutzer transparente Datenbank-Backend direkt an das Standardisierungsteam von DataPLANT weiter. Dieses gewährleistet zudem, dass nach einer Aufnahme neuer Terme in internatio-

<sup>6</sup>Vgl. Michael Ashburner, Catherine A. Ball, Judith A. Blake *et al.* „Gene Ontology: tool for the unification of biology“ *Nat Genet*, Nr. 25 (2000): 25-29. <https://doi.org/10.1038/75556>

<sup>7</sup>Vgl. Rainer Schwacke R. *et al.* „MapMan4: A Refined Protein Classification and Annotation Framework Applicable to Multi-Omics Data Analysis“ *Molecular plant* 12, Nr. 6 (2019): 879-892. <https://doi.org/10.1016/j.molp.2019.01.003>

<sup>8</sup>Vgl. Minoru Kanehisa und Susumu Goto. „KEGG: kyoto encyclopedia of genes and genomes“ *Nucleic acids research* 28, Nr. 1 (2000): 27–30 <https://doi.org/10.1093/nar/28.1.27>

<sup>9</sup><https://github.com/nfdi4plants/Swate>, aufgerufen am 15.04.2021.

nale Ontologien eine nachträgliche Kuratierung mit diesen Termen durch DataPLANT stattfinden kann und diese dann automatisch auch für neue Annotation via SWATE zur Verfügung stehen.

Daher werden Metadatenstandards und ihre Ontologien ständig im engen Austausch mit anderen Konsortien diskutiert und weiterentwickelt werden, um gemeinsame Lösungen und Verfahren zu forcieren. Dabei steht das Ziel im Zentrum innerhalb bestehender (internationaler) Standards wie beispielsweise dem „Minimal Information about a Plant Phenotyping Experiment“<sup>10</sup> zu arbeiten, diese zu verknüpfen und zu erweitern, statt konkurrierende Standards ins Leben zu rufen.

## 4 Infrastrukturelle Community-Unterstützung durch Task Area II

Die Task Area II Service, Software und Infrastruktur fokussiert sich auf die technisch-infrastrukturelle Unterstützung der Forschenden der Pflanzen-Grundlagenforschung. Dabei steht ein agiles Servicekonzept im Zentrum, das passgenau auf die Bedürfnisse der Community aufgebaut und kontinuierlich angepasst werden kann. Zentraler Anlaufpunkt ist das DataPLANT Hub, welches, basierend auf HUBZero<sup>11</sup>, einen uniformen Zugang zu Daten, Compute, Annotation und weiteren Services bietet. Komplementiert wird dieser Ansatz durch eine erste Generation an Tools, wie etwa SWATE, die zusätzlich direkten Zugang zur zugrundeliegenden Infrastruktur ermöglichen.

DataPLANT kann auf umfangreiche technische Infrastrukturen für Compute und Storage, wie der de.NBI Cloud<sup>12</sup> in Tübingen und Freiburg, der bwCloud<sup>13</sup>, dem bwForCluster BinAC<sup>14</sup>, EOSC-Life<sup>15</sup>, dem bwSFS<sup>16</sup> in Tübingen und Freiburg, sowie weitere verteilte Ressourcen zurückgreifen. Somit stehen umfangreiche Kapazitäten für die Verwendung durch die Pflanzenforschungscommunity zur Verfügung, die unter anderem über den DataPLANT Hub zugänglich gemacht werden.

Für DataPLANT wird, wo immer möglich, auf ein Single-Sign-On Konzept gesetzt, welches auf der ELIXIR AAI<sup>17</sup> basiert. Über die Nutzung von ELIXIR IDs in Kombination mit ORCID als persistentem Identifier ist stets eine eindeutige und nachhaltige Assoziation der Nutzenden mit ihren Forschungsdaten möglich<sup>18</sup>. In Verbindung mit den Funk-

<sup>10</sup><https://www.miappe.org/>, aufgerufen am 21.04.2021.

<sup>11</sup><https://hubzero.org>, aufgerufen am 31.01.2021.

<sup>12</sup><https://www.denbi.de>, aufgerufen am 31.01.2021.

<sup>13</sup><https://www.bw-cloud.org>, aufgerufen am 31.01.2021.

<sup>14</sup>vgl. <https://www.binac.uni-tuebingen.de>, aufgerufen am 31.01.2021.

<sup>15</sup><https://eosc-life.eu>, aufgerufen am 31.01.2021.

<sup>16</sup>vgl. <https://www.alwr-bw.de/bwsfs>, aufgerufen am 31.01.2021.

<sup>17</sup><https://www.elixir-europe.org>, aufgerufen am 31.01.2021.

<sup>18</sup>Vgl. von Suchodoletz, Dirk, Elisabeth Böker, Peter Brettschneider, and Franziska Rapp. „Entwicklung in Baden-Württemberg: ORCID und ROR IDs als Standard für langfristige Personen- und Institutionen-Identifier“ *Bausteine Forschungsdatenmanagement*, Nr. 2 (2020): 80-88. <https://doi.org/10.17192/bfdm.2020.2.8272>

tionalitäten des DataPLANT Hubs und der zugrundeliegenden Infrastruktur ist somit eine feingranulares Rollen- und Rechtemanagement möglich.

Task Area I passt existierende Standards und Definitionen den Anforderungen aus der Pflanzengrundlagenforschung an. Diese Anpassungen müssen auf eine effiziente und zuverlässige Art gesammelt, prozessiert und gespeichert werden. Für diesen Zweck werden maßgeschneiderte Lösungen implementiert, die das Indizieren der Metadaten, die Handhabung von Ontologien und umfangreiche Suchfunktionalitäten verfügbar machen. Um den vollen Umfang dieser Information nutzbar zu machen, werden die verschiedenen Teilinformationen, wie beispielsweise die umweltabhängige oder gewebsspezifische Expressionsänderung eines bestimmten Gens, kontextualisiert und miteinander verknüpft. Hiermit wird die Grundlage für den Einsatz von Methoden des maschinellen Lernens geschaffen, die Querbeziehungen und die inhärente Logik in den Forschungsdaten fassbar machen. Hierauf aufbauend ist vorgesehen die Annotations- und Suchfunktionalitäten um ein Vorschlagssystem zu erweitern, welches bei der Erstellung von Datensätzen wahrscheinliche Kontextualisierungen empfiehlt und somit die Forschenden unterstützt.

Der strukturierte Umgang mit Daten und Metadaten, wie auch die Kontrolle der Qualität der Daten macht den Einsatz von modernen Workflowkonzepten unabdingbar, da erst dadurch eine weitreichende Analyse und Verarbeitung der generierten Daten erreicht wird. Diese bioinformatischen Workflows werden von DataPLANT im Forschungsdatenkontext gleichwertig behandelt. Demnach will DataPLANT auch Workflows mit existierenden Metadatenstandards beschreiben und darauf basierende Weiterentwicklungen, kollaborative Nutzung und Ausführung unterstützen. Das Workflowhandling in DataPLANT wird unter anderem durch Galaxy<sup>19</sup> erreicht und durch den Einsatz weiterer Container-basierter Technologien zusätzliche flexibilisiert und komplettiert. Hierbei wird über den DataPLANT Hub die Orchestrierung der Workflows abgebildet, um das volle Potential der zur Verfügung stehenden Infrastruktur auszuschöpfen.

Um die wohlannotierten Datensätze der Forschenden, die workflowbasiert prozessiert und analysiert wurden, angemessen auswerten zu können, wird ein Service zur interaktiven Visualisierung aufgebaut. Auf Basis der in der Community üblichen Best-Practices und häufig genutzter Techniken wird es möglich sein, entlang dieser Workflows einzelne Schritte direkt und ohne zusätzliche Datentransfers zu visualisieren.

Weiterhin wird ein Service eingerichtet, über den die Versionierung von Daten und Metadaten, die Kollaboration unter Forschenden und das Teilen von Daten innerhalb laufender Projekte unterstützt wird. Um die Forschenden bei der Publikation ihrer Arbeiten zu unterstützen, wird außerdem ein Publikationsservice auf Basis von InvenioRDM<sup>20</sup> aufgebaut. Für die Vergabe eines Digital Object Identifier (DOI) für wohlannotierte Datensätze in Invenio, wird auf etablierte Dienste der in DataPLANT beteiligten Universitätsbibliotheken zurückgegriffen. Für die persistente Identifikation von Forschenden auf ORCID.

<sup>19</sup><https://usegalaxy.eu>, aufgerufen am 15.04.2021.

<sup>20</sup><https://invenio-software.org>, aufgerufen am 31.01.2021.

Im Zentrum all dieser Bemühungen steht der Mehrwert für die Forschenden aus der grundlegenden Pflanzenforschung. Für eine Community getriebene Interaktion und die Implementation neuer Lösungen wird Task Area I durch ein Technical Board begleitet.

## 5 Personelle Community-Unterstützung durch Task Area III

Task Area III kümmert sich um die direkte personelle Unterstützung der einzelnen Forschungsgruppen im Fachgebiet. Dies geschieht personell durch Data Stewards als Kernelement der DataPLANT Forschungsdatenmanagement-Strategie. Data Stewards werden hierbei zunehmend als valides Konzept zur fachnahen Unterstützung der Forschenden wahrgenommen.<sup>21</sup> Diese unterstützen die Community fortlaufend bei ihren alltäglichen Datenmanagement-Aufgaben. Durch das in DataPLANT vereinbarte Modell einer Präsenz von Data Stewards in den Forschungsgruppen, profitieren letztere von direkter und bedarfsgerechter Unterstützung. Dadurch bilden Data Stewards eine entscheidende Brücke zwischen technischen Lösungen, Infrastruktur und den Forschenden. Durch Unterstützung und Beratung im Daten- und Workflow-Management fördern sie den Prozess der Standardisierung von Metadaten, Nutzung von Ontologien und Provenienz in der Datenverarbeitung. Durch die direkte Interaktion vor Ort, können sie unverzüglich auf Anforderungen und Bedürfnisse der Forschenden reagieren.

Eine große Herausforderung besteht – wie bei den anderen NFDI-Konsortien auch – in der Ausbildung und im Know-How-Transfer. Ein spezifisch auf die Bedürfnisse von Data Stewards ausgerichtetes Ausbildungsprogramm, in den Bereichen Data Sciences und Data Literacy, das in seiner inhaltlichen Ausgestaltung maßgeblich durch Task Area III bestimmt wird, befindet sich gegenwärtig in der Konzeptionsphase. Mittel- bis langfristig ist intendiert, die Data Stewards in die Gestaltung und Durchführung des Ausbildungsprogramms zu integrieren. Dadurch soll sichergestellt werden, dass das Programm sich an evolvierende Randbedingungen anpassen und aktuell Entwicklungen mit einbeziehen kann. Auf diese Weise erhöht der koordinierte Einsatz von Data Stewards die Akzeptanz und Verbreitung guter wissenschaftlicher Praxis im Umgang mit Daten in der von DataPLANT anvisierten Gemeinschaft von Pflanzenforschern und Pflanzenforscherinnen.

<sup>21</sup>Vgl. Daniela Hausen, Jessica Rosenberg, Ute Trautwein-Bruns und Annett Schwarz. „Data Stewards an Der RWTH Aachen University – Aufbau Eines Flexiblen Netzwerks“. *Bausteine Forschungsdatenmanagement*, Nr. 2 (2020): 20-28. <https://doi.org/10.17192/bfdm.2020.2.8278>

## 6 Erhöhung der Reichweite und nachhaltige Entwicklung in Task Area IV

Eine zentrale Aufgabe in DataPLANT besteht in der Ausweitung auf die Gesamtheit der Forschenden im Bereich der Pflanzen-Grundlagenforschung. Nur so kann das Konsortium glaubwürdig das Fachgebiet vertreten und gemeinsame Standards und Konzepte etablieren. Eines der Erfolgskriterien für die geförderten Konsortien besteht in der möglichst vollständigen Durchdringung der eigenen Fach-Community und in der Vernetzung mit angrenzenden Forschungsgebieten. Hierfür soll ein koordinierter Onboarding-Prozess abgestimmt werden, durch welchen neue Mitglieder formal einem Konsortium beitreten und in die Governance-Struktur aufgenommen werden.

Zu den strategischen Überlegungen zählen Empfehlungen, die an zukünftige Forschungs- und Verbundprojekte mit Anteilen an Pflanzenforschung gegeben werden sollten. Solche Projekte benötigen typischerweise Unterstützung in den verschiedenen Aspekten des Datenmanagements und der technischen Infrastruktur. DataPLANT hat einen erheblichen Personalanteil für den Community-Support durch Data Stewards vorgesehen. Diese Unterstützung wird zunächst auf die ursprünglichen Mittragstellenden verteilt. Jedoch sollten zusätzlich auch später hinzukommende Forschergruppen von der NFDI profitieren.

Eine denkbare Herangehensweise wäre, dass bei neu eingereichten Forschungsanträgen für Supportleistungen in personeller oder infrastruktureller Form Mittel beantragt werden können, vergleichbar zu INF-Teilprojekten in SFBs oder GRKs. Das erlaubt insbesondere mit Teilstellen – kleine Projekte können vom Umfang her beispielsweise für das Datenmanagement nicht immer volle Stellen beantragen – gut umzugehen. Ebenso lassen sich Nachhaltigkeit und Stetigkeit bei in der NFDI beschäftigten Experten sicherstellen. Das Finden von geeignetem Personal und Anschlussverwendung nach Projektende sind eine oft ungelöste Herausforderung im Wissenschaftsbetrieb. Eine Beteiligung der Forschenden und Involvierung der Communities an der Ausgestaltung lässt sich über die Gremien in den einzelnen Konsortien und des NFDI-Vereins sicherstellen. Auf diese Weise kann ein klares Vorgehen ohne den Aufbau von Doppelstrukturen etabliert werden. So hat DataPLANT bereits unmittelbar nach dem Start begonnen in Richtung einer engen Kooperation und Abstimmung zu arbeiten. Hier konnte schon eine enge Verzahnung mit dem INF-Projekt des SFB TRR175<sup>22</sup> bzw. den Data Stewards von CEPLAS<sup>23</sup> eingerichtet werden. Potenziell wären diese Überlegungen schrittweise für die gesamte NFDI auszubauen und weiterzuentwickeln.

In DataPLANT können später hinzustoßende Forschende bei der initialen Mittelverteilung nicht direkt berücksichtigt werden. Bestimmte Basisaufgaben, wie eine gewisse Versorgung mit Support durch Data Stewards, Beratung in Antrags- und Rechtsfragen sowie Infrastruktur ist vorgesehen. Für größere Bedarfe oder „außerhalb“ liegende (Nicht-Kern-)Bereiche sollte sich nach dem vorgeschlagenen Konzept eine Versorgung via zusätzlich zu beantragende Mittel realisieren lassen. Um die NFDI als Ganzes

<sup>22</sup><https://www.tr175.bio.lmu.de>, aufgerufen am 31.01.2021.

<sup>23</sup><https://www.ceplas.eu>, aufgerufen am 31.01.2021.

zu stärken, sollten unabhängig davon alle Fachanträge an die DFG durch das jeweilige Konsortium befürwortet werden.

## 7 Mitgestaltung der Gesamt-NFDI

DataPLANT möchte die Ausformung der NFDI von Anfang an aktiv mitgestalten. Für das langfristige Ziel der „Einen NFDI“ wird es wichtig, die noch nicht involvierten Forschenden auf verschiedenen Ebenen einzubinden. Neben eigenständigen Anträgen in der zweiten und dritten Förderrunde der GWK kann dieses in der angesprochenen Verbreiterung der bereits etablierten Konsortien geschehen. Hier werden jeweils überzeugende Konzepte benötigt. So stehen geförderte Konsortien, Antragstellende und Reviewer vor der Herausforderung einer multidimensionalen Optimierung: Sehr kreative und innovative Konsortien mögen eben nicht passgenau zur benötigten Abdeckung der Wissenschaftslandschaft beitragen, während breit aufgestellte Konsortien potenziell nicht individualisiert genug auf die Bedürfnisse der einzelnen Fachgebiete eingehen könnten. Ein weiterer Weg der stärkeren Einbindung der gesamten wissenschaftlichen Community könnte über den Antragsprozess und die gemeinsame Ausgestaltung von Data Management Plänen laufen, indem eine Konsultation mit der NFDI eingefügt wird<sup>24</sup>.

Während bei den Cross-Cutting Topics bereits eine erste Abstimmung der Konsortien untereinander erfolgt ist<sup>25,26</sup> und diese Themen sich beispielsweise in Sektionen des NFDI-Vereins widerspiegeln werden, gibt es system- oder themenbedingt Überschneidungen bei FDM-Diensten, Methoden, Datentypen sowie genutzten Programmen und Workflow-Systemen. Ein zentraler Berührungspunkt aller Konsortien besteht in der Ausbildung von Data Stewards, die in ihren Profilen neben fachspezifischen Aspekten viele gemeinsame Anteile aufweisen.

Die Ebene der Cross-Cutting Topics versammelt übergreifende, gemeinsame Herausforderungen und Bedarfe aller Konsortien. Diese zeichnen sich durch einen großen gemeinsamen und eher schmalen fachspezifischen Anteil aus. Demgegenüber stehen Konsortien-übergreifende Interessen und Gemeinsamkeiten, die deutlich spezialisierter nur bestimmte Gruppen adressieren. Der fachspezifische Anteil ist vielfach höher als in den bereits in den Deklarationen identifizierten Themen. Durch eine engere Abstimmung und Verteilung der Aufgaben kann sicherlich sowohl ein Effizienzgewinn als auch eine erhöhte Standardisierungswirkung erzielt werden.

<sup>24</sup>NFDI: NFDI InfraTalk 01.03.2021, 2021, [YouTube] <https://www.youtube.com/watch?v=VTGhtQmn2p4>.

<sup>25</sup>Vgl. Frank Oliver Glöckner, Michael Diepenbroek, Janine Felden, Jörg Overmann, Aletta Bonn, Birgit Gemeinholzer, Anton Güntsch *et al.* „Berlin Declaration on NFDI Cross-cutting Topics“. *Zenodo* (2019). <http://doi.org/10.5281/zenodo.3457213>

<sup>26</sup>Vgl. Maik Bierwirth, Frank Oliver Glöckner, Christian Grimm, Sonja Schimmler, Franziska Boehm, Christian Busse, Andreas Degkwitz, Oliver Koepler und Heike Neuroth. „Leipzig-Berlin-Erklärung zu NFDI-Querschnittsthemen der Infrastrukturentwicklung“. *Zenodo* (2020). <http://doi.org/10.5281/zenodo.3895209>

Solche gemeinsamen Themen sollten daher bereits im Formierungsprozess des NFDI-Vereins zügig identifiziert werden, sodass eine Matrixstruktur entsteht, in der sich die geförderten Konsortien abstimmen und antragstellende Konsortien eingliedern können, um insbesondere neuen Fachkonsortien die Möglichkeit zu bieten, an bestimmten Themen lediglich „anzudocken“ ohne diese erneut erarbeiten zu müssen.

## **Danksagung**

Die vorliegende Arbeit wurde gefördert durch die Deutsche Forschungsgemeinschaft (DFG) (NFDI 7/1 – Projekt Nr. 442077441) auf Basis der Bund-Länder-Vereinbarung zum Aufbau und Förderung einer nationalen Forschungsdateninfrastruktur (NFDI) vom 26. November 2018.