

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

Das Versprechen der Vernetzung der NFDI

Dorothea Streckerⁱ Lukas C. Bossertⁱⁱ

Évariste Demandtⁱⁱⁱ

2021

Zitiervorschlag

Strecker, Dorothea, Lukas C. Bossert und Évariste Demandt. 2021. Das Versprechen der Vernetzung der NFDI. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 3/2021: S. 39-55. DOI: [10.17192/bfdm.2021.3.8336](https://doi.org/10.17192/bfdm.2021.3.8336).

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

ⁱHumboldt-Universität zu Berlin. ORCID: [0000-0002-9754-3807](https://orcid.org/0000-0002-9754-3807)

ⁱⁱRWTH Aachen University. ORCID: [0000-0003-3076-3968](https://orcid.org/0000-0003-3076-3968)

ⁱⁱⁱRWTH Aachen University. ORCID: [0000-0002-5705-0071](https://orcid.org/0000-0002-5705-0071)

1 Abstract

Das Versprechen der Vernetzung der NFDI ist ein Veranstaltungsbeitrag. Er stellt wesentliche Ergebnisse aus dem Workshop *Visualisierung von Netzwerken* auf dem 2. Community Meeting des Konsortiums NFDI4Ing vor und zeichnet nach, wie die Analyse und Visualisierung der Verbindungen zwischen (geplanten) Konsortien zu neuen Erkenntnissen und Fragestellungen führt.

The National Research Data Infrastructure's Promise to Partner is a workshop paper. It presents key findings from the *Visualization of Networks* workshop held at the 2nd Community Meeting of the NFDI4Ing consortium and traces how analyzing and visualizing connections between (planned) consortia leads to new insights.

2 Einleitung und Fragestellungen

Im Rahmen des 2. NFDI4Ing Community Meetings fand am 30.11.2020 ein anderthalbstündiger Online-Workshop zur Visualisierung und Analyse der Vernetzung geplanter Konsortien innerhalb der entstehenden Nationalen Forschungsdateninfrastruktur (NFDI) statt. Der Workshop war als Hands-on-Session konzipiert, bei der Teilnehmende einen Einblick in die Netzwerkanalyse mit R gewinnen konnten. Die Workshop-Materialien sind unter einer freien Lizenz veröffentlicht.¹

Dieser Beitrag folgt der Struktur des Workshops und beginnt mit einer Hinführung zum Begriff der Informationsinfrastruktur. Darauf aufbauend werden die Hintergründe der Netzwerkanalyse geplanter Kollaborationen innerhalb der NFDI erläutert. Mit einer Clusteranalyse wird im Hauptteil untersucht, inwiefern geplante Kollaborationen zu einer Vernetzung über Disziplingrenzen hinweg beitragen. Weiterhin werden Eigenschaften der Netzwerke basierend auf den Ausschreibungsrunden 2019 und 2020 verglichen, gefolgt von einer Prognose für die dritte Ausschreibungsrunde 2021. Das Workshopformat wird beschrieben und zum Abschluss reflektiert.

2.1 Infrastrukturen und Vernetzung

Mit der NFDI, die sich aktuell im Aufbau befindet, soll ein Angebot entstehen, das “[...] ein übergreifendes Forschungsdatenmanagement in Deutschland etabliert und fortentwickelt und die Effizienz des gesamten deutschen Wissenschaftssystems steigert.”² Um ein Verständnis davon zu bekommen, was die NFDI als eine auf Forschungsdaten spezialisierte Infrastruktur anstrebt und wie sie wirken kann, lohnt sich eine

¹Strecker, Dorothea, Lukas C. Bossert und Évariste Demandt, *Das Versprechen der Vernetzung: Workshop-Material* (Zenodo, 2020). DOI: <https://doi.org/10.5281/zenodo.4322757>.

²Satzung von Nationale Forschungsdateninfrastruktur (NFDI) e.V., <https://www.nfdi.de/wp-content/uploads/2021/05/Satzung-NFDI-eV.pdf>, §2(2).

genauere Betrachtung des Infrastrukturbegriffs allgemein. Geoffrey Bowker und seine Koautor:innen definieren Infrastrukturen als “pervasive enabling resources in network form”.³ In diesem Halbsatz werden bereits zentrale Eigenschaften von Infrastrukturen genannt:

- *Infrastrukturen sind allgegenwärtig und allgemein zugänglich.* Sie durchdringen bestimmte Lebensbereiche und werden so selbstverständlich benutzt, dass sie beinahe unsichtbar werden.⁴
- *Infrastrukturen ermöglichen bestimmte Aktivitäten.* Beispielsweise unterstützen Informationsinfrastrukturen Wissensarbeit, darunter auch Forschung.⁵ Infrastrukturen können dabei auf hochspezialisierte Tätigkeiten zugeschnitten sein, beispielsweise den Austausch bestimmter Datentypen innerhalb einer wissenschaftlichen Disziplin. Wird eine Infrastruktur in einer Disziplin stark genutzt, kann das Wissen über die Nutzung dieser Infrastruktur zu einer Mitvoraussetzung für die Zugehörigkeit zur Disziplin werden.⁶
- *Infrastrukturen umfassen Hardware, Menschen, Praktiken und Software.* Infrastrukturen verfügen zweifellos über materielle Komponenten, beispielsweise Speichermedien. Sie interagieren aber auch mit sozialen Welten der Wissensproduktion, beispielsweise wenn Daten aus einem bestimmten Kontext auch für Forschende anderer Disziplinen nachnutzbar gemacht werden sollen.⁷ Darum sind Infrastrukturen keine rein technischen Lösungen, sondern soziotechnische Systeme.⁸
- *Infrastrukturen ermöglichen Schichtung oder Vernetzung durch Standardisierung.* Standardisierung ist eine zwingende Voraussetzung für Austausch und Vernetzung von Infrastrukturen. Das verdeutlichen beispielsweise Menschen, die in Infrastrukturen arbeiten: sie gehen koordinierenden Aktivitäten nach und tragen dadurch wesentlich zu kollaborativen Strukturen in der Wissenschaft bei.⁹ Ein weiteres Beispiel für Vernetzung durch Standardisierung sind Schnittstellen, allerdings sind Schnittstellen unter Forschungsdateninfrastrukturen aktuell wenig

³Bowker, Geoffrey C., Karen Baker, Florence Millerand und David Ribes, “Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment”, in *International Handbook of Internet Research*, herausgegeben von Jeremy Hunsinger, Lisbeth Klastrup und Matthew Allen (Dordrecht: Springer Netherlands, 2010), 97-117, doi: https://doi.org/10.1007/978-1-4020-9789-8_5, 98.

⁴Vgl. Star, Susan Leigh und Karen Ruhleder, “Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces”, *Information Systems Research* 7, Nr. 1 (1996), 11-134, doi: <https://www.doi.org/10.1287/isre.7.1.111>.

⁵Vgl. Bowker et al., “Toward Information Infrastructure Studies”

⁶Vgl. Edwards, Paul N., *A vast machine: computer models, climate data, and the politics of global warming* (Cambridge, London: The MIT Press, 2013), 189

⁷Vgl. Bates, Jo, Yu-Wei Lin und Paula Goodale, “Data journeys: Capturing the socio-material constitution of data objects and flows”, *Big Data & Society* 3, Nr. 2 (2016), doi: <https://www.doi.org/10.1177/2053951716654502>.

⁸Vgl. Edwards, *A vast machine*.

⁹Vgl. Paine, Drew und Charlotte P. Lee, “Coordinative Entities: Forms of Organizing in Data Intensive Science”, *Computer Supported Cooperative Work (CSCW)* 29, Nr. 3 (2020), 335-380, doi: <https://www.doi.org/10.1007/s10606-020-09372-2>.

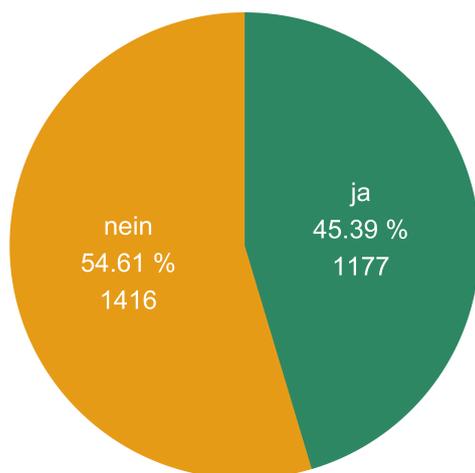


Abbildung 1: In re3data verzeichnete Forschungsdatenrepositorien mit und ohne Schnittstelle.

verbreitet. Das lässt sich beispielsweise mittels re3data zeigen: Das globale und disziplinübergreifende Nachweisinstrument für Forschungsdatendienste wies am 20.11.2020 nur für 45.39,% der verzeichneten Infrastrukturen mindestens eine Schnittstelle nach (s. Abb. 1).

Diese zentralen Eigenschaften von Infrastrukturen lassen sich auch auf die NFDI übertragen. Im Folgenden wird besonders der Aspekt der Vernetzung in Bezug auf die NFDI beleuchtet. Ein Netzwerk wird dabei verstanden als System, das als Graph modelliert werden kann, also als Menge von Objekten (Knoten) und ihren Verbindungen (Kanten).¹⁰

2.2 Konzeption der NFDI als Netzwerk

Die NFDI wird etabliert, nachdem sich bereits eine Vielzahl spezialisierter Angebote für den Umgang mit Forschungsdaten herausgebildet hat, beispielsweise Repositorien, welche die Publikation und Archivierung von Forschungsdaten ermöglichen. Aufgrund dieser heterogenen Ausgangslage ist im Kontext der NFDI die Vernetzung bestehender Infrastrukturen von herausragender Bedeutung. Das zeigt sich auch daran, dass die NFDI bereits als Netzwerk konzipiert wurde.

2013 richtet die Gemeinsame Wissenschaftskonferenz (GWK) den Rat für Informationsinfrastrukturen (RfII) als Beratungsgremium für die Weiterentwicklung wissen-

¹⁰<https://de.wikipedia.org/wiki/Netzwerk>.

schaftlicher Informationsinfrastrukturen ein.¹¹ Der RfII empfiehlt 2016 die Gründung einer nationalen Forschungsdateninfrastruktur, die “als bundesweites, verteiltes und wachsendes Netzwerk arbeitsteilig angelegt” sein soll.¹² Bund und Länder beschließen 2018 die Gründung der NFDI. In der Bund-Länder-Vereinbarung ist die NFDI als “digitaler, regional verteilter und vernetzter Wissensspeicher” angelegt.¹³

Die NFDI soll sich laut dieser Vereinbarung nach drei Ausschreibungsrunden aus bis zu 30 Konsortien zusammensetzen, die in den jeweiligen Fachgemeinschaften verankert sind und dafür geeignete Dienste etablieren. In der ersten Ausschreibungsrunde 2019 gingen bei der Deutschen Forschungsgemeinschaft (DFG) 24 verbindliche Letters of Intent ein. 22 eingereichte Anträge wurden in einer statistischen Übersicht dargestellt.¹⁴ Die verbindlichen Letters of Intent von zwei Konsortien (NFDI4Crime und RWD-DHCt) wurden zurückgezogen. Seit Oktober 2020 werden 9 Konsortien aus der ersten Ausschreibungsrunde gefördert.

Auch nach der ersten Ausschreibungsrunde wird die Notwendigkeit der Vernetzung der NFDI auf mehreren Ebenen betont: So hebt die DFG im Programmmerkleblatt 2020 den Anspruch auf nationale und internationale Vernetzung hervor, die mehrere Stakeholder und Disziplinen einschließt.¹⁵ 2020 wird mit der *Leipzig-Berlin-Erklärung zu NFDI-Querschnittsthemen der Infrastrukturentwicklung* ein Diskussionsimpuls zu konsortienübergreifenden Handlungsfeldern veröffentlicht. Das Dokument betont die Bedeutung der disziplinübergreifenden Vernetzung von Diensten für den kulturellen Wandel innerhalb des gesamten Wissenschaftssystems.¹⁶ Vorausgegangen war die *Berlin Declaration on NFDI Cross-Cutting Topics*.¹⁷ Auch das NFDI-Expertengremium der DFG

¹¹Vgl. Gemeinsame Wissenschaftskonferenz, *Informationsinfrastrukturen / NFDI*, <https://www.gwk-bonn.de/themen/weitere-arbeitsgebiete/informationsinfrastrukturen-nfdi/>.

¹²Vgl. Rat für Informationsinfrastrukturen, *Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland* (Göttingen, 2016), urn: <http://nbn-resolving.de/urn:nbn:de:101:1-201606229098>, 2.

¹³Vgl. Gemeinsame Wissenschaftskonferenz, *Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI) vom 26. November 2018*, <https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/NFDI.pdf>, 1.

¹⁴Vgl. Deutsche Forschungsgemeinschaft, *Nationale Forschungsdateninfrastruktur. Statistische Übersichten zum Antragsengang* (2019), https://www.dfg.de/download/pdf/foerderung/programme/nfdi/nfdi_auswertung_2019.pdf.

¹⁵Vgl. Deutsche Forschungsgemeinschaft, *Guidelines for Consortia – National Research Data Infrastructure (NFDI)* (2020), https://www.dfg.de/formulare/nfdi100/nfdi100_en.pdf, 2.

¹⁶Vgl. Bierwirth, Maik, 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), doi: <https://www.doi.org/10.5281/zenodo.3895209>, 3.

¹⁷Glöckner, Frank Oliver, Michael Diepenbroek, Janine Felden, Jörg Overmann, Aletta Bonn, Birgit Gemeinholzer, Anton Güntsch, Birgitta König-Ries, Bernhard Seeger, Annette Pollex-Krüger, Juliane Fluck, Iris Pigeot, Toralf Kirsten, Timo Mühlhaus, Christof Wolf, Uwe Heinrich, Christoph Steinbeck, Oliver Koepler, Oliver Stegle, Jochaim Weimann, Thomas Schörner-Sadenius, Christian Gutt, Florian Stahl, Kurt Wagemann, Torsten Schrade, Robert Schmitt, Chris Eberl, Frank Gauterin, Martin Schultz und Lars Bernard, *Berlin Declaration on NFDI Cross-Cutting Topics* (Zenodo, 2019), doi: <https://www.doi.org/10.5281/zenodo.3457213>.

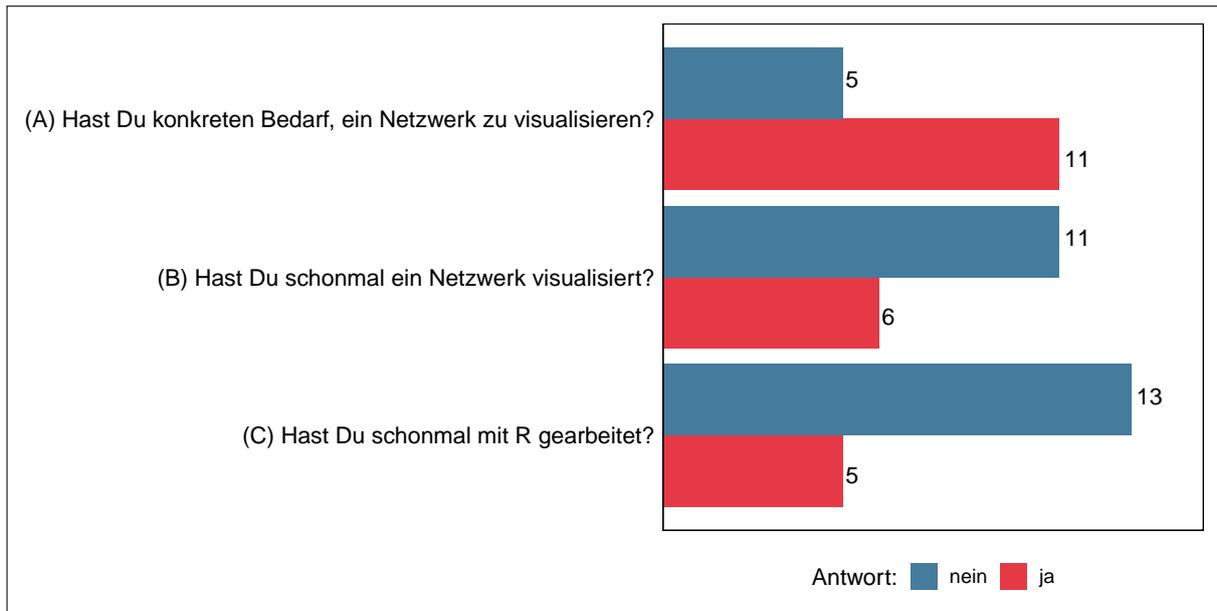


Abbildung 2: Drei Fragen, welche von den Teilnehmenden des Workshops beantwortet wurden.

veröffentlicht 2020 eine Stellungnahme. Das Gremium sieht die Notwendigkeit für Vernetzung auf drei Ebenen, darunter auch die Vernetzung mit bestehenden Strukturen außerhalb des NFDI-Kontexts.¹⁸

3 Workshopformat

Der Workshop *Visualisierung von Netzwerken* beim 2. NFDI4Ing Community Meeting am 30.11.2020 war auf eine Dauer von 90 Minuten angelegt. Insgesamt beteiligten sich 23 Teilnehmende am Workshop. In einer kurzen Umfrage (vgl. Abb. 2) äußerte die Mehrzahl Bedarf an der Visualisierung von Netzwerken. Demgegenüber hatten die meisten Teilnehmenden noch nie ein Netzwerk visualisiert. Auch mit der Statistiksoftware R hatten die meisten Teilnehmenden bisher noch nicht gearbeitet. Der Workshop deckte damit einen inhaltlichen Bedarf und ermöglichte einen Einblick in die Netzwerkanalyse und -visualisierung mit R.

Durchgeführt wurde der Workshop über eine Kamerverbindung zu den Teilnehmenden auf der Konferenzplattform Zoom. Zum Programmieren wurden Jupyter Notebooks empfohlen, also webbasierte interaktive Entwicklungsumgebungen, über die sich Softwarecode mit erläuternden Texten kombinieren lassen. Dieses Verfahren ist auch als

¹⁸Deutsche Forschungsgemeinschaft, *Der Aufbau einer Nationalen Forschungsdateninfrastruktur (NFDI)*. Zweite Stellungnahme des NFDI-Expertengremiums (2020), https://www.dfg.de/download/pdf/foerderung/programme/nfdi/stellungnahme_nfdi_201112_de.pdf, 3.

*literate programming*¹⁹ bekannt. Teilnehmende, die noch nie mit R gearbeitet hatten, konnten online ein Jupyter Notebook anlegen²⁰, damit programmieren und dem Workshop folgen. Einige Teilnehmende, die bereits Erfahrung mit R hatten, nutzten die ihnen vertraute Entwicklungsumgebung und graphische Benutzeroberfläche RStudio.²¹

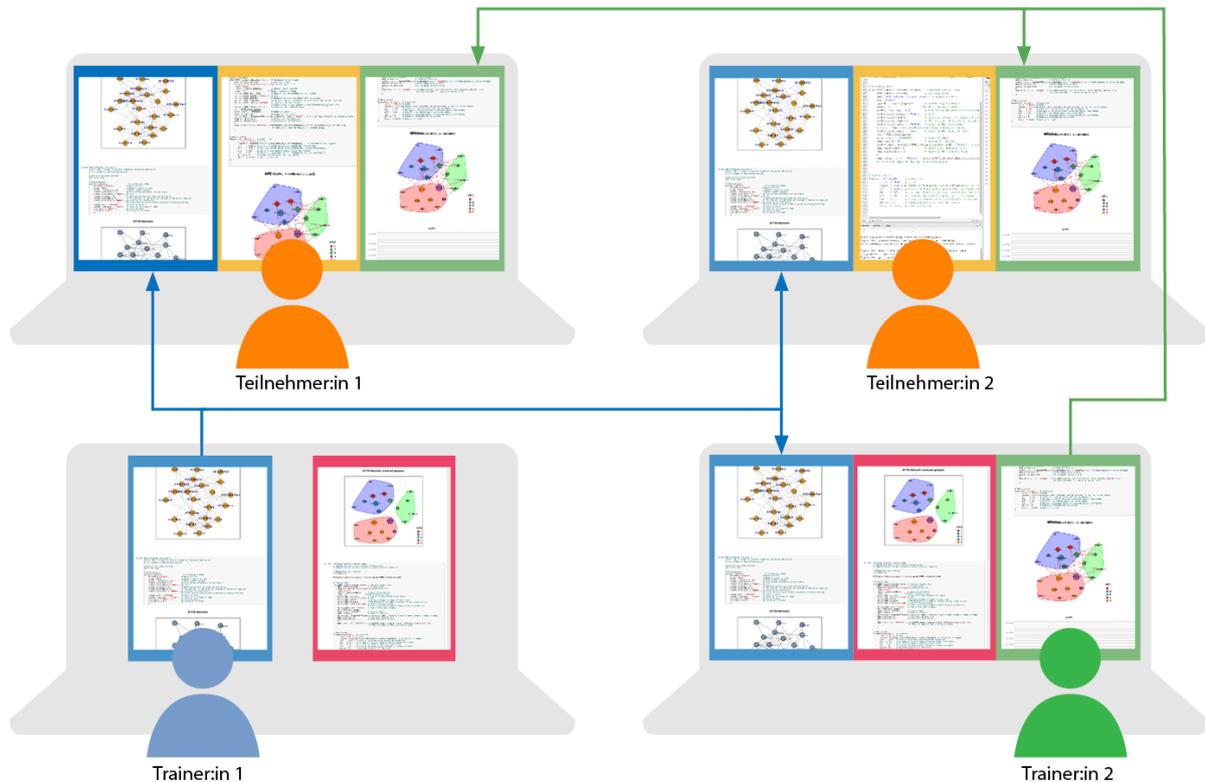


Abbildung 3: Schematische Darstellung des Workshopformats. Trainer:innen und Teilnehmer:innen sitzen jeweils vor der Bildschirmpartition, die sie befüllen. Trainer:in 1 spricht und befüllt dabei ein Jupyter Notebook (blau), welches read-only an alle übertragen wird. Trainer:in 2 befüllt ein geteiltes Jupyter Notebook (grün) sukzessive mit den Inhalten des blauen. Die Teilnehmer:innen können sich aus dem geteilten grünen Jupyter Notebook mit *copy & paste* bedienen, oder aber aus dem blauen abtippen. Trainer:in 2 beantwortet zudem die Fragen im Chat.

An der Durchführung des Workshops waren zwei Trainer:innen mit unterschiedlichen Jupyter Notebooks beteiligt (Abb. 3). Trainer:in 1 schrieb in einem Jupyter Notebook (blau), teilte ihren Bildschirm und erläuterte dabei die durchgeführten Analyseschritte. Dabei

¹⁹Knuth, Donald E., "Literate Programming", *The Computer Journal* 27 (1984), 97-111; <https://doi.org/10.1093/comjnl/27.2.97>.

²⁰<https://rnotebook.io/>

²¹<https://www.rstudio.com/about/>

orientierte sich Trainer:in 1 an einem vorher fertig ausgearbeiteten Lösungs-Jupyter-Notebook (rot). Trainer:in 2 beantwortete aufkommende Fragen im Chat und verfolgte den Verlauf der Präsentation von Trainer:in 1, indem sie auch den geteilten Bildschirm (blau) vor sich hatte. Da nicht immer alle Teilnehmenden gleich schnell im abtippen und nachverfolgen des Codes waren, gab es ein über den zugehörigen Link geteiltes Lösungs-Jupyter-Notebook (grün), das sukzessive von Trainer:in 2 befüllt wurde. Die Code-Blöcke entnahm Trainer:in 2 aus dem fertigen Lösungs-Jupyter-Notebook (rot). Damit wurde sichergestellt, dass die Live-Demonstration mit den geteilten Codeblöcken übereinstimmte.

Während also die Teilnehmenden der Präsentation von Trainer:in 1 (blau) folgten, konnten sie in einem eigenen Jupyter Notebook, oder ggf. RStudio (beides orange), den Code direkt ausprobieren. Über den geteilten Link konnten sie zu jeder Zeit in das sich füllende Lösungs-Jupyter-Notebook (grün) schauen und der Analyse jederzeit folgen, unabhängig von eventuell auftretenden Verbindungsproblemen, der Schreibgeschwindigkeit von Trainer:in 1 oder anderen Faktoren.

Durch die Nutzung von Jupyter Notebooks entfiel für Teilnehmende die zwingende Notwendigkeit, Software lokal zu installieren. Gegenüber den Ansätzen anderer Workshopformate, insbesondere Software Carpentry²², verließen die Teilnehmenden die Veranstaltung nicht mit einem fertig eingerichteten Gerät. Im Rahmen der Veranstaltung war technischer Support für alle Teilnehmenden nicht umsetzbar. Für diese interaktive Online-Veranstaltung erwiesen sich Jupyter Notebooks darum als geeignete Option.

Im Folgenden werden die Datenerhebung sowie Ergebnisse der Netzwerkanalyse beschrieben. Die Workshopteilnehmenden visualisierten das Netzwerk und führten ein Clustering des Netzwerks durch. Dieser Beitrag beschreibt darüber hinaus weitere Eigenschaften des Netzwerks.

3.1 Datenerhebung

Inwiefern die NFDI der angestrebten Vernetzung unter Konsortien gerecht wird, wird im Folgenden mit Methoden der Netzwerkanalyse untersucht. Netzwerkanalysen der NFDI-Konsortien sind grundsätzlich möglich, da die DFG die Letters of Intent geplanter Konsortien öffentlich zugänglich macht. Die Daten für die vorliegende Netzwerkanalyse wurden aus den für die **erste** und **zweite** Ausschreibungsrunde eingereichten Binding Letters of Intent, jeweils Abschnitt 3, extrahiert. Hervorzuheben ist, dass sowohl verbindliche (binding) Letters of Intent für die aktuelle als auch unverbindliche (non-binding) für die jeweils folgenden Antragsrunden eingereicht werden können. In der folgenden Analyse werden nur die Binding Letters of Intent für eine Antragstellung 2019 beziehungsweise 2020 berücksichtigt. Für jedes Binding Letter of Intent wurde erhoben, mit welchen Konsortien Kollaborationen geplant sind, sofern diese selbst auch eine verbindliche Absichtserklärung abgegeben haben. Abbildung 5 bietet eine

²²<https://software-carpentry.org/>.

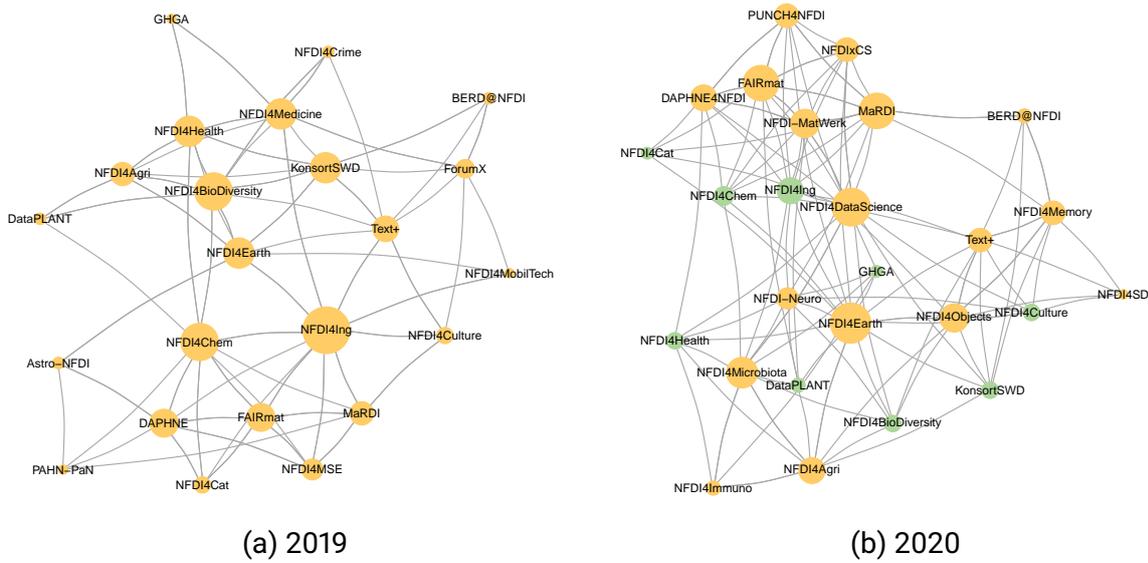


Abbildung 4: Visualisierung geplanter Verbindungen zwischen Konsortien der jeweiligen Antragsrunde. Im Netzwerk der zweiten Antragsrunde sind bereits geförderte Konsortien in grün hervorgehoben.

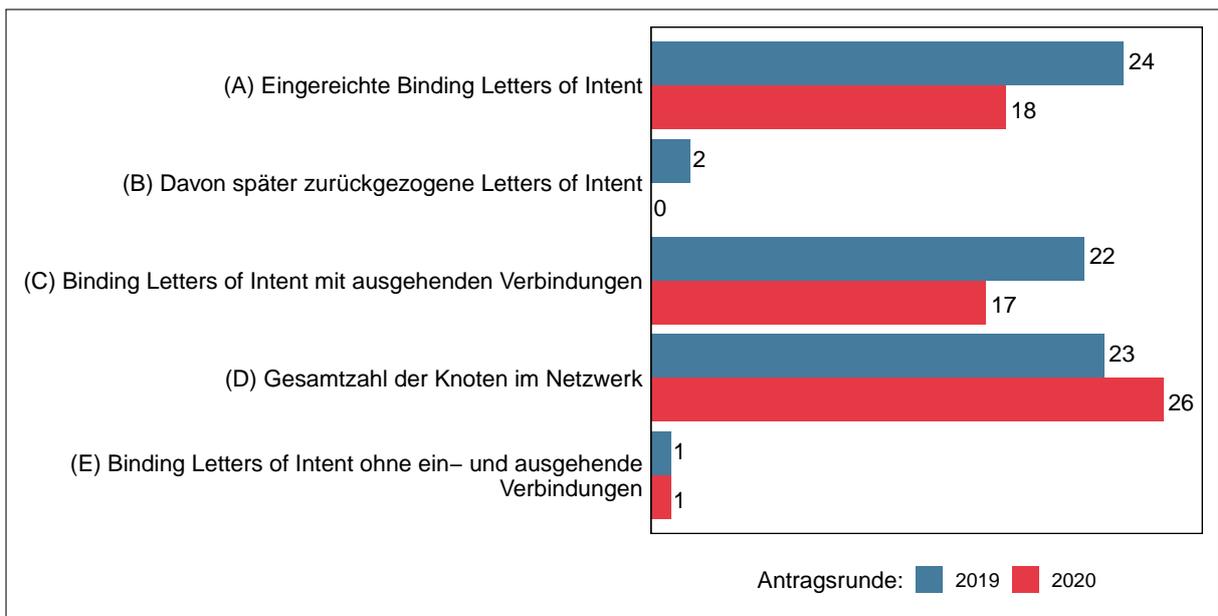


Abbildung 5: Übersicht der Einreichungen von Binding Letters of Intent in den ersten beiden Antragsrunden.

Übersicht. Nicht im Netzwerk enthalten sind Konsortien ohne ein- und ausgehende Verbindungen. 2019 und 2020 betrifft dies jeweils ein Konsortium (RWD-DHCt bzw. CoDiRPS).

Anzumerken ist, dass die Kollaborationen in den Binding Letters of Intent nicht weiter qualifiziert werden. Auf Grundlage der Netzwerkanalyse können also keine Aussagen über Form oder Erfolg der Zusammenarbeit getroffen werden. Darüber hinaus stellt das Netzwerk nur geplante Kollaborationen zum Zeitpunkt der Einreichung der verbindlichen Absichtserklärung dar. Es ist davon auszugehen, dass das Netzwerk mit zunehmender Ausgestaltung der NFDI wächst.

Abbildung 4 stellt jeweils geplante Kollaborationen aus der ersten und zweiten Ausschreibungsrunde dar. Die Größe der Knoten skaliert mit der Anzahl ein- und ausgehender Verbindungen. Im Netzwerk von 2020 sind bereits geförderte Konsortien aus der ersten Ausschreibungsrunde in grün hervorgehoben.

Sowohl die der Analyse zugrundeliegenden Daten als auch das verwendete Skript werden mit dieser Publikation verfügbar gemacht.

4 Das Netzwerk 2020 im Lichte der Konferenzsystematik

4.1 NFDI-Konferenzsystematik

Im Zuge des Ausschreibungsprozesses organisiert die DFG jedes Jahr eine NFDI-Konferenz im Vorfeld der Einreichung der Letters of Intent. Auf der ersten NFDI-Konferenz am 13. und 14. Mai 2019 wurden 57 Konsortialinitiativen in fünf parallelen Sessions vorgestellt.²³ Die Zuordnung zu den Sessions der ersten NFDI-Konferenz hat sich während des Antragsverfahrens als geeignete Fächerzuordnung erwiesen und kam beispielsweise bei der Verteidigung der Anträge zum Einsatz. Sie wird im Folgenden als NFDI-Konferenzsystematik bezeichnet und wurde auch für die zweite NFDI-Konferenz am 8. und 9. Juli 2020 wieder genutzt:

1. Medizin
2. Lebenswissenschaften
3. Geistes- und Sozialwissenschaften
4. Informatik, Mathematik und Ingenieurwissenschaften
5. Chemie und Physik

Auf den ersten Blick sind alle Begrifflichkeiten der NFDI-Konferenzsystematik aus der DFG-Fachsystematik²⁴ bekannt, wenngleich sie auch aus drei unterschiedlichen Ebenen stammen. Ein gutes Beispiel dafür bietet die Fächergruppierung vier der NFDI-Konferenzsystematik: Die Informatik ist in der DFG-Fachsystematik auf der Ebene

²³Eickhoff, Ulrike, *Overview of Extended Abstracts* (2019), https://www.dfg.de/download/pdf/foerderung/programme/nfdi/welcome_speech_dfg_head_office_eickhoff.pdf.

²⁴Deutsche Forschungsgemeinschaft, *DFG-Fachsystematik der Wissenschaftsbereiche, Fachgebiete, Fachkollegien und Fächer für die Amtsperiode 2016–2019* (2017), https://www.dfg.de/download/pdf/dfg_im_profil/gremien/fachkollegien/amtsperiode_2016_2019/fachsystematik_2016-2019_de_grafik.pdf.

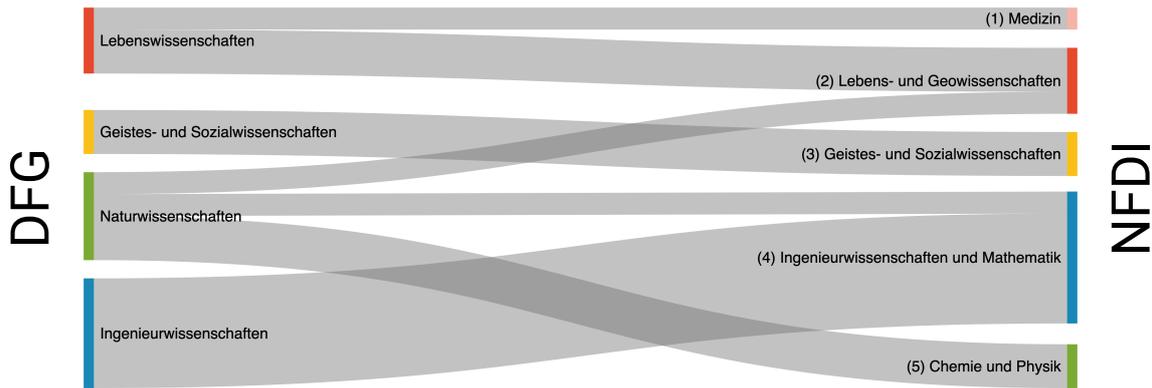


Abbildung 6: Sankey-Diagramm mit dem Wechsel der Fächerzugehörigkeit zwischen DFG-Fachsystematik und NFDI-Konferenzsystematik.

der Fachkollegien, die Mathematik hingegen eine Ebene darüber in den Fachgebieten verortet, die Ingenieurwissenschaften schließlich sind als Wissenschaftsbereich auf der höchsten Ebene angesiedelt. Das Sankey-Diagramm veranschaulicht, inwiefern die Begriffe der NFDI-Konferenzsystematik und der DFG-Fachsystematik als inhaltlich deckungsgleich angesehen werden können.

Links zeigt das Sankey-Diagramm die Wissenschaftsbereiche aus der DFG-Fachsystematik für die Amtsperiode 2016-2019, die Länge der Balken links ist proportional zu der Anzahl der Fachgebiete pro Wissenschaftsbereich.²⁵ Rechts zeigt das Sankey-Diagramm die NFDI-Konferenzsystematik, welche sich aus den entsprechenden Flusspfeilen aus der DFG-Fachsystematik links speist. Für die Fächer (1) Medizin, (3) Geistes- und Sozialwissenschaften, sowie (5) Chemie und Physik werden rechts keine begrifflichen Änderungen vorgenommen, da diese als inhaltlich deckungsgleich mit Fachgebieten aus der DFG-Fachsystematik angesehen werden können. Für die Fächer (2) und (4) werden rechts begriffliche Änderungen vorgenommen. Diese Änderungen werden im Folgenden begründet. Die Lebenswissenschaften rechts werden im Vergleich zur DFG-Fachsystematik nicht nur um die separat geführte Medizin verkleinert, sondern gleichzeitig durch die Geowissenschaften vergrößert. Dies wird mit einem Flusspfeil von den Naturwissenschaften links zu den Lebens- und Geowissenschaften rechts abgebildet und spiegelt sich in Konsortialinitiativen wie NFDI4Earth und NFDI4Natural-Resources wieder, die von der DFG hier eingeordnet werden. Die zweite Änderung betrifft die Informatik, welche laut DFG-Fachsystematik als Fachkollegium innerhalb der Ingenieurwissenschaften angesiedelt ist. Es erscheint daher irreführend, sie in der NFDI-Konferenzsystematik in einem Konglomerat aus "Informatik, Mathematik und Ingenieurwissenschaften" auszuweisen. Darum wird die Informatik im Sankey-Diagramm rechts nicht genannt. Entsprechend der Flusspfeilgewichtung werden zudem die Ingenieurwissenschaften in der neuen Bezeichnung rechts vor der Mathematik genannt.

²⁵Deutsche Forschungsgemeinschaft, *DFG-Fachsystematik*.

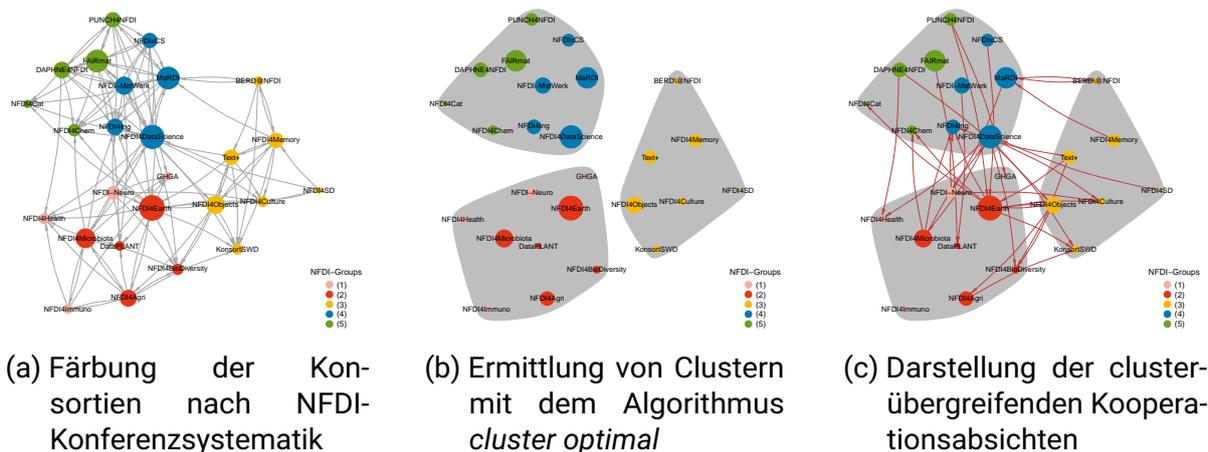


Abbildung 7: Visualisierung der NFDI-Konsortien der zweiten Antragsrunde (2020) mit Fokus auf die Clustereinteilung.

Zusammenfassend lässt sich sagen, dass die NFDI-Konferenzsystematik eine teilweise Neuordnung auf der Ebene der Fachgebiete aus der DFG-Fachsystematik darstellt, die insbesondere die Naturwissenschaften betrifft.

4.2 Clusteranalyse

Der Einsatz der NFDI-Konferenzsystematik im Antragsverfahren ermöglicht das Einfärben der Netzwerkknoten nach Fächern. Damit lässt sich nachzuvollziehen, inwiefern geplante Kollaborationen zu einer Vernetzung auch über Disziplingrenzen hinweg beitragen. In Verbindung mit dem hier verwendeten Layout-Algorithmus *graphopt*²⁶ werden bereits Häufungen von Knoten sichtbar (Abb. 7a). Diese erste Beobachtung legt die Vermutung nahe, dass Kollaborationen vorwiegend unter Konsortien mit ähnlicher inhaltlicher Ausrichtung geplant werden. Um diese Vermutung zu überprüfen, wird in einem zweiten Analyseschritt das Netzwerk mit dem Algorithmus *cluster optimal* visualisiert²⁷. Es können weitere Verfahren zur Community Detection auf das Netzwerk angewendet werden, wie zum Beispiel Verfahren basierend auf *edge betweenness*, der *walktrap* Algorithmus und hierarchische Verfahren basierend auf dem ungerichteten Netzwerk. Alle genannten Verfahren liefern identische Cluster. Mit der diesem Artikel zugehörigen Datenpublikation können über das Jupyter Notebook weitere Verfahren getestet werden. Im Einzelnen setzen sich die Cluster wie folgt zusammen (Abb. 7b):

- Cluster 1: Gruppe 1 (Medizin), Gruppe 2 (Lebenswissenschaften)
- Cluster 2: Gruppe 3 (Geistes- und Sozialwissenschaften)
- Cluster 3: Gruppe 4 (Informatik, Mathematik und Ingenieurwissenschaften), Gruppe 5 (Chemie und Physik)

²⁶https://igraph.org/r/doc/layout_with_graphopt.html.

²⁷https://igraph.org/r/doc/cluster_optimal.html.

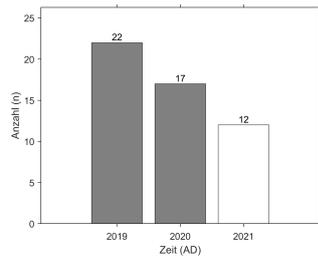
Bei einem genaueren Blick auf die Konsortien der drei Cluster ist diese Gruppierung wenig überraschend. Cluster 1 setzt sich aus den Konsortien zusammen, die nach der DFG-Fachsystematik zu den Lebens- und den Naturwissenschaften zählen, wobei die Lebenswissenschaften nach der NFDI-Konferenzsystematik bereits einen Teil der Naturwissenschaften enthalten (Abb. 6). Cluster 2 isoliert die Geistes- und Sozialwissenschaften, sowohl nach DFG-Systematik als auch nach der NFDI-Konferenzsystematik. Das dritte Cluster führt Fachgebiete der Ingenieur- und Naturwissenschaften zusammen, wobei die Ingenieurwissenschaften nach der NFDI-Konferenzsystematik wiederum bereits einen Teil der Naturwissenschaften enthalten (Abb. 6). Das Clustering bestätigt die Vermutung, dass Kollaborationen vor allem mit fachnahen Konsortien beabsichtigt werden.

Das Clustering ist allerdings nicht mit einer Silobildung gleichzusetzen. Hier wird ein weiterer Analyseschritt notwendig, der clusterübergreifende Kooperationsabsichten sichtbar macht. Abbildung 7c zeigt ausschließlich die Verbindungen von Knoten, die über Clustergrenzen hinweg reichen. Mit diesen Verbindungen kann potentiell eine tiefgreifende Vernetzung erzielt werden. Es zeigt sich, dass bis auf eine Ausnahme alle Konsortien entweder als Kooperationspartner aus einem anderen Cluster heraus genannt werden oder selbst ein Konsortium in einem anderen Cluster nennen.

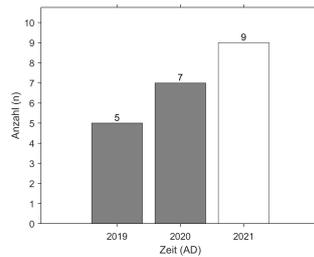
4.3 Kennzahlen der Netzwerke im Jahresvergleich und Prognose für die dritte Ausschreibungsrunde

Das Antragsverfahren für Konsortien der NFDI sieht drei Förderrunden vor. Nachdem zum Zeitpunkt dieses Manuskriptes die Nennungen aus den Binding Letters of Intent von 2019 und 2020 vorliegen, lassen sich nicht nur Eigenschaften der Netzwerke beider Ausschreibungsrunden vergleichen, sondern auch Vorhersagen für die letzte Ausschreibungsrunde von 2021 treffen. In der folgenden Abbildung sind die Balken mit den bekannten Werten von 2019 und 2020 grau eingefärbt, die Balken mit einer linearen Extrapolation der Werte für 2021 weiß. Bei den Abb. 8a bis 8c liegt die Gesamtheit der Binding Letters of Intent mit ausgehenden Verbindungen zu Grunde (vgl. Abb. 5, Zeile C). Bei den Abb. 8d bis 8f liegt die Gesamtheit der jeweiligen Knoten im Netzwerk zu Grunde (vgl. Abb. 5, Zeile D).

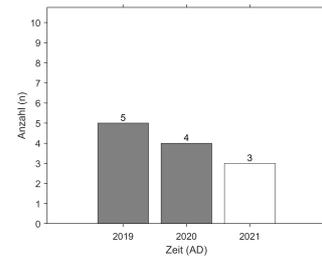
Abbildung 8a zeigt die Werte aus Abb. 5, Zeile C, und extrapoliert die Anzahl der Binding Letters of Intent mit ausgehenden Verbindungen in die dritte Antragsrunde, wo eine weitere Abnahme zu erwarten ist. Für diese Abnahme gibt es mindestens zwei mögliche Erklärungen. Einerseits fallen mit jeder Antragsrunde die geförderten Konsortien aus dem Verfahren heraus, andererseits ist keineswegs gesichert, dass Konsortialinitiativen in mehreren Antragsrunden antreten. In diesem Zusammenhang ist auch darauf hinzuweisen, dass die jährlich stattfindenden NFDI-Konferenzen im Vorfeld des Einreichungszeitpunktes der Binding Letters of Intent einen Rahmen bilden, in dem über Zusammenschlüsse von Konsortialinitiativen diskutiert wird. Abbildung 8b zeigt



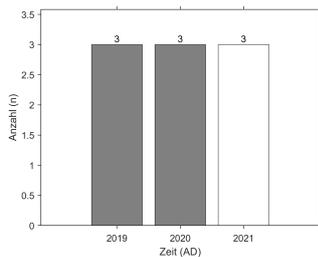
(a) Binding Letters of Intent mit ausgehenden Verbindungen



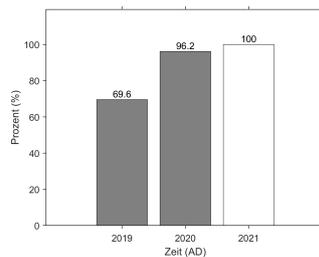
(b) Median ausgehende Verbindungen pro Binding Letter of Intent



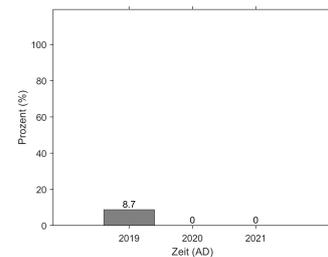
(c) Median ausgehende Verbindungen zu Konsortien der gleichen Antragsrunde



(d) Cluster im Netzwerk



(e) Konsortien in clusterübergreifenden Verbindungen



(f) Disziplinfremde Zuordnungen

Abbildung 8: Kennzahlen der Netzwerke im Jahresvergleich (grau) und Prognose für die dritte Ausschreibungsrunde (weiß)

den Median an ausgehenden Verbindungen pro Konsortium an. Hier wird für 2021 eine weitere Zunahme erwartet, da anzunehmen ist, dass sich der Netzwerkgedanke in den Binding Letters of Intent zunehmend niederschlägt. Gleichzeitig wird eine weitere Abnahme ausgehender Verbindungen zu Konsortien der gleichen Antragsrunde erwartet, wie Abb. 8c zeigt. Diese Abnahme erklärt sich möglicherweise dadurch, dass in der Antragsrunde 2020 erstmals ausgehende Verbindungen zu bereits geförderten Konsortien auftreten. Der Trend, sich mit bereits geförderten Konsortien zu vernetzen, könnte die Abnahme an Verbindungen zu Konsortien der gleichen Antragsrunde erklären und wird sich 2021 voraussichtlich verstärken.

Abbildung 8d zeigt die Anzahl der Cluster, die ein Clustering mit dem Algorithmus cluster optimal²⁸ erzeugt. Es ist zu erwarten, dass diese Anzahl unverändert bei drei Clustern bleibt. Abbildung 8e zeigt den Anteil an Knoten im Netzwerk mit Verbindungen über Clustergrenzen hinweg, ungeachtet der Richtung dieser Verbindungen. Hier ist eine weitere Zunahme zu erwarten, die sich zumindest teilweise mit der gesteigerten Zahl ausgehender Verbindungen pro Konsortium begründen lässt (vgl. Abb. 8b). Abbildung 8f zeigt Knoten, die im Clustering gruppenfremd zugeordnet werden. Im

²⁸https://igraph.org/r/doc/cluster_optimal.html.

Netzwerk 2019 betrifft das die Konsortialinitiativen KonsortSWD und NFDI4MobilTech. Vermutlich lässt sich die gruppenfremde Zuordnung mit der Spärlichkeit ausgehender Verbindungen im Netzwerk 2019 begründen. 2020 werden keine Knoten gruppenfremd zugeordnet, und 2021 voraussichtlich auch nicht.

5 Diskussion

5.1 Workshopformat

Das Workshopformat erwies sich als geeignet zur Durchführung interaktiver Online-Veranstaltungen. Die Nutzung von Jupyter Notebooks senkte technische Hürden für Teilnehmende. Durch den Support im Chat und das schrittweise Teilen von Codeblöcken wurden Teilnehmende bei der Durchführung der Analyseschritte unterstützt. Allerdings reichte der zeitliche Rahmen des Workshops nicht aus, um Teilnehmende Fragestellungen eigenständig bearbeiten zu lassen; zu diesem Zweck erscheint rückblickend eine Dauer von drei Stunden angemessener.

5.2 Netzwerkanalyse

Dieser Beitrag folgt der Struktur des Workshops auf dem 2. Community Meeting des Konsortiums NFDI4Ing und beginnt mit einer Reflexion des Infrastrukturbegriffs, die verdeutlicht, dass Vernetzung eine zentrale Eigenschaft von Infrastrukturen ist. Das gilt auch für die NFDI, die bestehende Infrastrukturen zusammenführen, koordinieren und erweitern soll und dadurch die Vernetzung auf verschiedenen Ebenen fördert. Bereits in ihrer Konzeption war die NFDI als Netzwerk verteilter Strukturen angelegt. In diesem Beitrag wird darum vor allem die geplante Vernetzung unter Konsortien der NFDI untersucht. Zukünftig könnte auch die Vernetzung auf anderen Ebenen berücksichtigt werden, beispielsweise der Einfluss der NFDI auf die Vernetzung individueller Dateninfrastrukturen wie Repositorien. Zudem könnten die Verbindungen zwischen Konsortien unter Berücksichtigung gemeinsamer Projekte oder Publikationen weiter qualifiziert werden. Beispielsweise fand am 26.08.2020 ein Workshop zu Querschnittsthemen der NFDI statt, an dem die neun geförderten Konsortien aus der ersten Ausschreibungsrunde teilnahmen.²⁹ Der Workshop stellt die Weichen für konsortienübergreifende Kollaborationen in Bereichen, die für alle Konsortien der NFDI relevant sind. Zukünftige Analysen könnten diese Kollaborationen im Bereich von Querschnittsthemen untersuchen.

²⁹Ebert, Barbara, Juliana Fluck, Frank Oliver Glöckner, Oliver Koepler, Bernhard Miller, Robert Schmitt, Torsten Schrade, Oliver Stegle, Christoph Steinbeck, Dirk von Suchodoletz, Kurt Wagemann, Jennifer Knebes, Sophie Kraft, Hendrik Seitz-Moskaliuk, York Sure-Vetter und Elena Wössner, *NFDI Cross-cutting Topics Workshop Report* (Zenodo 20210), doi: <https://www.doi.org/10.5281/zenodo.4593770>.

Inwiefern die Vernetzung innerhalb der NFDI geplant wird, untersucht der vorliegende Beitrag mit Methoden der Netzwerkanalyse basierend auf den bei der DFG eingereichten Binding Letters of Intent. Im ersten Teil der Analyse steht die Vernetzung von Konsortien innerhalb wissenschaftlicher Disziplinen und über Disziplingrenzen hinweg im Fokus. Ein Clustering unterteilt das Netzwerk von 2020 in drei Cluster, die jeweils ein bis zwei Fachgruppierungen aus der NFDI-Konferenzsystematik entsprechen. Daraus lässt sich schließen, dass Kollaborationen vorwiegend unter Konsortien geplant sind, die ähnliche wissenschaftliche Disziplinen vertreten. Allerdings sind im Netzwerk 2020 bis auf ein Konsortium alle Initiativen auch in clusterübergreifenden Kollaborationen.

In der Vorbereitung auf die Analyse zeigte sich außerdem, dass sich die DFG-Fachsystematik nur bedingt zur Analyse der Netzwerke eignet. Die DFG-Fachsystematik spiegelt die interne Organisation der DFG wider und dient primär der Förderung von Forschungsprojekten. Demgegenüber erwies sich die NFDI-Konferenzsystematik als geeigneter, um die Selbstorganisation der Disziplinen im Rahmen des NFDI-Ausschreibungsprozesses abzubilden.

In der Analyse werden darüber hinaus Kennzahlen der Netzwerke 2019 und 2020 verglichen, gefolgt von einer Vorhersage für die dritte Ausschreibungsrunde 2021. Der Vergleich der Kennzahlen macht in der Tendenz eine Zunahme ausgehender Verbindungen deutlich, es wird also eine Steigerung der Nennungen geplanter Kollaborationen in den Binding Letters of Intent für 2021 erwartet. Das gilt auch für geplante Kollaborationen über Clustergrenzen hinweg, was für eine breitere Durchdringung des Netzwerkgedankens unter antragstellenden Initiativen sprechen könnte. Die beobachteten Veränderungen von 2019 zu 2020 sowie die Extrapolation für 2021 deuten jedoch darauf hin, dass sich geplante Kollaborationen zunehmend auf bereits geförderte Konsortien beschränken. Dadurch, dass sich die bereits geförderten Konsortien der ersten Antragsrunde gleichmäßig auf die fünf Fächergruppen nach der NFDI-Konferenzsystematik verteilen, kommt ein weiterer Umstand ins Spiel, der vermutlich zu einer in Bezug auf Fächergruppen gleichmäßigen Verteilung geplanter Kollaborationen führt.

Das hier vorgestellte Verfahren ist geeignet, Vernetzungsabsichten geplanter Konsortien der NFDI zum Zeitpunkt der Einreichung der Binding Letters of Intent zu analysieren. Das Clustering der Konsortien hat sich im Vergleich von 2019 und 2020 als stabil erwiesen und wird sich erwartungsgemäß wenig verändern, das heißt, thematisch nah verwandte Konsortien werden in der Tendenz weiterhin vermehrt Kollaborationen planen. Darauf aufbauende Untersuchungen könnten genauer aufzeigen, wie Konsortien zusammenarbeiten und wie sich das Netzwerk im Laufe der Zeit verändert.

6 Danksagung

Die Autor:innen bedanken sich beim NFDI4Ing Community Cluster Wärmetechnik/ Verfahrenstechnik (DFG-Fachgebiet 42) für die Ausrichtung des Workshops im Rahmen

des 2. NFDI4Ing Community Meetings und bei den Teilnehmenden für den angeregten Austausch. Der Autor Évariste Demandt bedankt sich bei Bund, Ländern und bei der Gemeinsamen Wissenschaftskonferenz (GWK) für die Förderung und Unterstützung im Rahmen des Konsortiums NFDI4Ing. Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 442146713. Der Autor Évariste Demandt bedankt sich außerdem für eine Anschubfinanzierung aus dem Kooperationsprojekt nfdi4ing.nrw.