

Drs. 5066-16
Berlin 22 01 2016

Empfehlungen zur Spezifikation des Kerndatensatz Forschung

INHALT

Vorbemerkung	5
Kurzfassung	8
A. Standardisierung in der Forschungsberichterstattung	9
A.I Verortung	9
A.II Steigender Informationsbedarf und heterogene Berichts Anfragen	10
II.1 Dezentrale Lösungsansätze	11
II.2 Koordinierte Lösungsansätze	11
A.III Ziele und Grundsätze des Kerndatensatz Forschung	13
III.1 Ziele	13
III.2 Grundsätze	14
III.3 Standardisierung in einem heterogenen Wissenschaftssystem	15
III.4 Kerndatensatz und Berichtswesen	18
III.5 Ziele und Aufwand der intendierten Nutzung	20
A.IV Bisheriger Prozess zur Erstellung einer Spezifikation	21
IV.1 Prozess	22
IV.2 Dokumentation	29
B. Zu den Projektergebnissen	31
B.I Hinweise zum Einsatz des Kerndatensatz Forschung	31
I.1 Prozessbegleitende Erfassung von Daten	31
I.2 Kontinuierliche Erfassung von Daten für flexible Nutzung	32
I.3 Qualitätssicherung	32
I.4 Datenschutz	32
I.5 Datenweitergabe	34
I.6 ID-Management	35
I.7 Interpretation der Daten	35
I.8 Datennutzerinnen und Datennutzer	36
B.II Die Spezifikation	36
II.1 Aggregation	36
II.2 Inhalte	37
C. Einführung und weitere Entwicklung des Kerndatensatz Forschung	44
C.I Einführung im Deutschen Wissenschaftssystem	44
I.1 Aufbau eines Kompetenznetzwerkes	45
I.2 Unterstützung der Einführung	46
I.3 Monitoring	47

I.4	Aufgaben bei der Unterstützung der Einführung	48
I.5	Anforderungen bei der Unterstützung der Einführung	49
C.II	Weiterentwicklung der Spezifikation des Kerndatensatz Forschung und seiner Schale	50
II.1	Inhalte der Weiterentwicklung	50
II.2	Gestaltung des Weiterentwicklungsprozesses	52
C.III	Überprüfung von Funktion und Nutzung	53
	Anhang	55
	Abkürzungsverzeichnis	56
	Glossar	57
	Übersicht über verschiedene Identifier	60
	Abbildungsverzeichnis	61
	Tabellenverzeichnis	62
	Anlage: Spezifikation des Kerndatensatz Forschung: Version 1.0	63

Vorbemerkung

Im Januar 2013 hat der Wissenschaftsrat Empfehlungen zu einem Kerndatensatz Forschung vorgelegt. |¹ Er folgte damit früheren Empfehlungen, die die herausragende Bedeutung einer qualitätsgesicherten Datengrundlage, z. B. für Begutachtungen durch *informed Peer review*, identifiziert hatten. |² Der vorgeschlagene Prozess zur Erarbeitung von Spezifikationen für Kerndaten (vgl. Glossar, S. 57 f.) zu Forschungsaktivitäten begann als Forschungsprojekt, gefördert vom Bundesministerium für Bildung und Forschung (BMBF), im Herbst 2013. Die Federführung des Projektes lag beim Institut für Forschungsinformation und Qualitätssicherung (iFQ, heute DZHW) in Berlin. Darüber hinaus waren das Fraunhofer-Institut für Angewandte Informationstechnik (FIT) in St. Augustin sowie die Geschäftsstelle des Wissenschaftsrates am Projekt beteiligt. Der Wissenschaftsrat hat die operative Arbeit der Projektpartner durch einen Beirat begleitet.

Zum Abschluss der Projektphase wurde eine konsolidierte Version zur öffentlichen Kommentierung vorgelegt. Nach der Auswertung der Rückmeldungen liegt nun die „Spezifikation des Kerndatensatz Forschung“ vor. Damit ist die Grundlage für die freiwillige Implementierung |³ an wissenschaftlichen Einrichtungen und die anschließende Weiterentwicklung der Spezifikation des Kerndatensatz Forschung gelegt. Der Wissenschaftsrat würdigt die im Projekt geleistete Arbeit.

Die Spezifikation des Kerndatensatzes ist ein Vorschlag, einige klar beschriebene Angaben zu Forschungsaktivitäten und den Rahmenbedingungen, unter denen sie erbracht werden, in standardisierten Datenformaten vorzuhalten. Damit soll eine Grundlage für eine datengestützte Berichterstattung über Forschungsaktivitäten und ihre Kontexte geschaffen werden, welche geeignet ist die interne Nutzung und Weitergabe qualitätsgesicherter Daten zu erleichtern,

|¹ Vgl. Wissenschaftsrat: Empfehlungen zu einem Kerndatensatz Forschung (Drs. 2855-13), Berlin Januar 2013. *Online verfügbare Dokumente und Quellen sind aus der PDF-Version dieser Empfehlungen heraus verlinkt.*

|² Vgl. Wissenschaftsrat: Empfehlungen zur Bewertung und Steuerung von Forschungsleistung (Drs. 1656-11), Halle November 2011.

|³ Für eine Erläuterung zentraler Begriffe vgl. Glossar, S. 57 f.

6 den Aufwand für wiederholte Erhebungen zu reduzieren und die Vergleichbarkeit und Anschlussfähigkeit der vorgehaltenen Daten sicher zu stellen. Der Kerndatensatz kann und will nicht den Anspruch erheben, alle Forschungsaktivitäten in allen Forschungskontexten – insbesondere nicht denen aller Fächer – umfassend oder auch nur in gleichem Umfang abzudecken. Der Wissenschaftsrat betont nachdrücklich, dass diese Angaben keinesfalls ausreichen, um wissenschaftliche Leistungen zu bewerten. Die Bewertung von Forschungsleistungen kann nur durch qualifizierte *Peers* erfolgen und benötigt immer zusätzliche qualitative und quantitative Informationen. Dies gilt umso mehr für die Bewertung spezifischer Leistungen, die für das besondere Profil einer Einrichtung oder eines Einrichtungstyps (z. B. der Fachhochschulen oder der Ressortforschung) charakteristisch sind. Der Kerndatensatz in der vorliegenden Fassung ist auch nicht dazu bestimmt und deshalb auch nicht geeignet, die Spezifika der Ressortforschung des Bundes abzubilden. Der Wissenschaftsrat bekräftigt nachdrücklich seine Empfehlungen zur Bewertung und Steuerung von Forschungsleistungen und die in seinen eigenen Leitfäden zur Bewertung von Forschungsleistung festgelegten Grundsätze. |⁴

Der Wissenschaftsrat empfiehlt Hochschulen |⁵ und außeruniversitären Forschungseinrichtungen (AUF) |⁶, diesen Standard anzuwenden. Ihre Bereitschaft dazu wird wesentlich davon abhängen, in welchem Umfang sich datenabfragende Stellen (Ministerien, Evaluationsagenturen u. a. m.) an diesem Standard orientieren. Der Wissenschaftsrat selbst beabsichtigt, seine Datenabfragen, jeweils in Absprache mit seinen Auftraggebern, im Laufe des Jahres 2016 an die Spezifikation des Kerndatensatz Forschung anzupassen. Er fordert auch andere Nutzerinnen und Nutzer von Daten, die in der Spezifikation des Kerndatensatz Forschung standardisiert werden, dazu auf, ihre Datenabfragen künftig an diesem Standard auszurichten.

In den vorliegenden Empfehlungen gibt der Wissenschaftsrat Hinweise zur Implementierung des Kerndatensatz Forschung an Hochschulen und außeruniversitären Forschungseinrichtungen auf Basis dieser Spezifikation. Weiterhin legt er einen Vorschlag zur Einführung dieses Standards im deutschen Wissenschaftssystem sowie zur Weiterentwicklung der Spezifikation des Kerndatensatz Forschung vor.

|⁴ Vgl. Wissenschaftsrat: Empfehlungen zur Bewertung und Steuerung von Forschungsleistung (Drs. 1656-11), Halle November 2011. Wissenschaftsrat: Aufgaben, Kriterien und Verfahren des Evaluationsausschusses des Wissenschaftsrates (Drs. 4205-14), Greifswald Oktober 2014.

|⁵ Unter dem Begriff „Hochschulen“ werden im Folgenden Universitäten, Fachhochschulen und sonstige Hochschulen in öffentlicher und privater Trägerschaft verstanden.

|⁶ Unter dem Begriff „außeruniversitäre Forschungseinrichtungen (AUF)“ werden im Folgenden alle nicht-universitären Forschungseinrichtungen verstanden. Neben den Einrichtungen der großen Wissenschaftsorganisationen sind dies u. a. die Forschungseinrichtungen des Bundes und der Länder.

Bei der Weiterentwicklung der Spezifikation sollen – der Idee dieses Standards treu – die Belange aller wichtigen Akteure des deutschen Wissenschaftssystems, darunter insbesondere auch, in Abstimmung mit dem Bund, die der Ressortforschung, Berücksichtigung finden. Solange die Spezifika der Ressortforschung nicht berücksichtigt sind, wird der Wissenschaftsrat bei der Erhebung von Daten aus der Ressortforschung – zum Beispiel bei Evaluationen – nicht auf den Kerndatensatz zurückgreifen.

Zur Begleitung des Projekts und zur Vorbereitung dieser Empfehlungen hat der Wissenschaftsrat im Sommer 2013 einen Beirat einberufen. Darin mitgewirkt haben auch Sachverständige, die nicht Mitglieder des Wissenschaftsrates sind. Ihnen weiß sich der Wissenschaftsrat zu besonderem Dank verpflichtet.

Die Ausarbeitung des vorliegenden Definitionsvorschlages in der vom Wissenschaftsrat empfohlenen Zeit hat einen Kraftakt und großes Engagement auch außerhalb der Gremien des Wissenschaftsrates erfordert. Der Wissenschaftsrat dankt allen Mitgliedern der befassten Projektgruppen, der Arbeitsgruppe der außeruniversitären Forschungseinrichtungen, dem Statistischen Bundesamt, dem Projektmanagement sowie besonders den Piloteinrichtungen, die schon während der Erarbeitung wertvolle Hinweise aus Anwendersicht beigesteuert haben. Als äußerst wichtig haben sich auch die teils sehr ausführlichen Rückmeldungen erwiesen, die das Projekt im Rahmen der Beta-Phase erreicht haben; der Wissenschaftsrat spricht den Antwortenden seinen Dank für ihre Mitwirkung aus.

Der Wissenschaftsrat hat die vorliegenden Empfehlungen am 22. Januar 2016 in Berlin verabschiedet.

Kurzfassung

In den vorliegenden Empfehlungen wird beschrieben, welche Angaben zu Forschungsaktivitäten und den Rahmenbedingungen, unter denen sie erbracht werden, Teil der Spezifikation des Kerndatensatz Forschung sind. Die Spezifikation legt fest, welche Datenformate hierbei verwendet werden sollten. Darüber hinaus formuliert der Wissenschaftsrat Vorschläge, wie die entwickelte Spezifikation im deutschen Wissenschaftssystem eingeführt werden kann, ebenso wie Anforderungen an ihre Weiterentwicklung, die die Anpassung dieses Standards an Entwicklungen im Wissenschaftssystem (z. B. zu Innovationsaktivitäten) langfristig sicherstellen sollen. Damit sind die „Empfehlungen zu einem Kerndatensatz Forschung“, die der Wissenschaftsrat 2013 verabschiedet hat, umgesetzt.

In der vorliegenden ersten Version besteht der Kerndatensatz einer Einrichtung aus Aggregatdaten in folgenden *sechs* Bereichen: Beschäftigte, Nachwuchsförderung, Drittmittel und Finanzen, Patente und Ausgründungen, Publikationen sowie Forschungsinfrastrukturen. Eine Übersicht über weitere vorgeschlagene Differenzierungen bietet Kapitel B.II.2. Detaillierte Hinweise zur Definition der Daten finden sich in der Spezifikation, die diesen Empfehlungen als Anlage beigefügt ist. Weitere Hinweise zur Dokumentation finden sich in Kapitel A.IV.2.

Zur Unterstützung der Einrichtungen, die sich für die Einführung des Kerndatensatz Forschung entscheiden, schlägt der Wissenschaftsrat die zeitnahe Einrichtung eines bundesweit ansprechbaren *Helpdesks* vor. Ebenso sollte ein Monitoring-Prozess angestoßen werden, der die Implementierung an einigen Einrichtungen intensiv begleitet sowie eine Überprüfung von Funktion und Nutzung ermöglicht, in der intendierte und ggf. nicht intendierte Auswirkungen erfasst und bewertet werden. Der Wissenschaftsrat bittet die Gemeinsame Wissenschaftskonferenz (GWK), sich über eine angemessene Finanzierung dieser Aufgaben zu verständigen. Nach einer Einführungsphase von ca. drei Jahren wird eine erste Anpassung der Spezifikation des Kerndatensatzes unter Berücksichtigung der gemachten Erfahrungen empfohlen. Spätere Weiterentwicklungen sollten regelmäßig, ungefähr alle fünf Jahre, unter Einbindung aller relevanten Akteure stattfinden.

A. Standardisierung in der Forschungsbericht- erstattung

A.1 VERORTUNG

Mit seinen „Empfehlungen zu einem Kerndatensatz Forschung“ aus dem Jahr 2013 hat der Wissenschaftsrat in Deutschland erstmalig einen Prozess zur Verständigung über die Verwendung von häufig genutzten Daten zu Forschungsaktivitäten und deren Kontext angestoßen. |⁷ Dieser gründet sich auf die Einsicht, dass Daten heute oft und in zunehmendem Maße für die Berichterstattung über Forschungsaktivitäten verwendet werden. Deswegen haben sie eine essentielle Bedeutung für verschiedenste Diskurse in Wissenschaft, Wissenschaftsmanagement und Wissenschaftspolitik, aber auch über die Wissenschaft hinaus. Häufig werden weit reichende Interpretationen zur Bewertung von Forschungsaktivitäten und -leistungen auf Basis von Angaben präsentiert, deren Herkunft und Verlässlichkeit nicht immer hinlänglich bekannt sind, so dass sich auch Schlussfolgerungen nur schwer hinterfragen lassen. Dies ist nicht im Sinne eines sach- und zielorientierten Diskurses. Der Wissenschaftsrat möchte mit dem Kerndatensatz daher eine sachangemessene Grundlage für die Dokumentation einiger Forschungsaktivitäten und ihres Kontextes schaffen.

Er schlägt vor, die an jeder wissenschaftlichen Einrichtung in gleicher oder ähnlicher Form bereits vorhandenen Angaben zu Forschungsaktivitäten und den Rahmenbedingungen, unter denen sie erbracht werden, in standardisierter Form vorzuhalten. Standardisierte Datenformate (vgl. Glossar, S. 57) sollen es ermöglichen, diese Angaben mit einem geringeren Aufwand und, aufgrund ihrer mehrfachen Verwendbarkeit und Vergleichbarkeit, auch qualitätsgesichert bereit zu stellen. Zur Frage des Aufwands betont der Wissenschaftsrat in die-

|⁷ Vgl. Wissenschaftsrat: Empfehlungen zu einem Kerndatensatz Forschung (Drs. 2855-13), Berlin Januar 2013.

sem Zusammenhang, dass das Gros der Angaben in der Spezifikation des Kerndatensatz Forschung so oder in ähnlicher Form bereits heute an Hochschulen und anderen Forschungseinrichtungen vorliegt und für verschiedene Zwecke häufig weitergegeben wird. Der Wissenschaftsrat hat den Kerndatensatz zudem dezentral konzipiert: Forschungseinrichtungen bleiben Dateneignerinnen und entscheiden bei jeder Anfrage, ob und in welchem Umfang sie Angaben weitergeben. Der Kerndatensatz schafft somit keine neue Verpflichtung zur Dokumentation oder Weitergabe von Daten, geschweige denn einen Automatismus.

A.II STEIGENDER INFORMATIONSBEDARF UND HETEROGENE BERICHTSANFRAGEN

Der vielfache Wunsch aus Gesellschaft und Politik, aber auch aus der Wissenschaft selbst, Informationen über Forschungsaktivitäten wissenschaftlicher Einrichtungen zu erhalten, hat zu einer Vielzahl unterschiedlicher Berichte geführt, die Hochschulen und andere Forschungseinrichtungen in Deutschland teils regelmäßig, teils auf Anfrage abgeben müssen oder aus eigener Initiative veröffentlichen. Datenlieferungen an die Statistischen Ämter sind ein Beispiel für Berichterstattung auf Basis einer gesetzlichen Grundlage. Auch Forschungsorganisationen wie die Fraunhofer Gesellschaft (FhG), Helmholtz-Gemeinschaft (HGF), Max-Planck-Gesellschaft (MPG) und die Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (WGL) berichten regelmäßig z. B. im Rahmen des Paktes für Forschung und Innovation sowie für ihre eigenen Verfahren der Qualitätssicherung. Hochschulen geben ferner u. a. Forschungsberichte ab und erstellen Berichte für Akkreditierungsagenturen. Weitere Berichte werden z. B. für Landesministerien angefertigt und auch der Wissenschaftsrat erfragt im Zuge von Evaluationen oder der Begutachtung von Landeshochschulsystemen Daten von Hochschulen und anderen wissenschaftlichen Einrichtungen. Nicht zuletzt werden solche Daten auch für die Drittmittelwerbung und die Rechenschaftslegung – zum Beispiel bei EU Projekten – abgefragt.

Die Zahl solcher datengestützten Berichte zu Forschungsaktivitäten ist in den letzten Jahren gewachsen und beansprucht die ohnehin knappen Ressourcen immer mehr. In Zeiten zunehmender Autonomie und Planungsverantwortung sind Hochschulen und die großen außeruniversitären Forschungsorganisationen FhG, HGF, MPG und WGL außerdem auch für interne Zwecke in immer stärkerem Maße, und über verschiedene Disziplinen und Einrichtungen hinweg, auf belastbare Daten, z. B. zur strategischen Planung und Steuerung ihrer

II.1 Dezentrale Lösungsansätze

Einer einzelnen Forschungseinrichtung stehen nur begrenzte Möglichkeiten zur Verfügung, auf die genannten Herausforderungen zu reagieren. Dabei ist eine in jüngerer Zeit immer häufiger gewählte Möglichkeit, die für die Berichtslegung notwendigen Daten zu organisieren, die Einführung von sogenannten Forschungsinformationssystemen (FIS, vgl. Glossar, S. 57). FIS sind Datenbankanwendungen, die verteilt vorgehaltene Informationen zu Forschungsaktivitäten sowie Kontextinformationen zusammenführen, eine strukturierte Sicht auf diese Daten ermöglichen und die Erstellung von Berichten unterstützen. Auf dieser Basis können auch für Forschende Mehrwertdienste (z. B. automatische Erzeugung tabellarischer Lebensläufe oder Aufbereitung von Informationen für die Homepage) bereitgestellt werden. Mitte 2014 betrieben 22 von 51 Hochschulen mit Promotionsrecht, die an einer entsprechenden Umfrage teilnahmen, Forschungsinformationssysteme. Bei weiteren 13 Hochschulen befanden sich solche Systeme im Aufbau bzw. waren bei zehn Hochschulen in Planung. |⁹

Ein FIS stellt einen hilfreichen und wichtigen technischen Lösungsansatz für die genannten Probleme dar. Der Wissenschaftsrat würdigt daher die Initiative Hessens (HeFIS) |¹⁰ und Nordrhein-Westfalens in Zusammenarbeit mit der Universität Münster, weitere landeseigene Hochschulen bei der Einführung solcher Systeme zu unterstützen.

II.2 Koordinierte Lösungsansätze

Ein FIS vermag aber nicht die organisatorisch-inhaltlichen Probleme zu lösen, die sich – einrichtungsübergreifend – daraus ergeben, dass Daten in unter-

|⁸ Vgl. DINI AG Forschungsinformationssysteme. Forschungsinformationssysteme in Hochschulen und Forschungseinrichtungen – Positionspapier, 2014, <http://dx.doi.org/10.5281/zenodo.14828> v. 14.01.2016. Vgl. Herwig, S.; Becker, J.: Einführung eines Forschungsinformationssystems an der Westfälischen Wilhelms-Universität Münster. Von der Konzeption bis zur Implementierung in: Bittner, S.; Hornbostel, S.; Scholze, F. (Hrsg.): Forschungsinformationen in Deutschland. Anforderungen, Stand und Nutzen existierender Forschungsinformationssysteme, Berlin, iFQ-Working Paper No. 10, 2012, S. 41-55.

|⁹ Sechs der antwortenden Hochschulen gaben an, keine hochschulweite Lösung zu betreiben. Vgl. Sticht, K.: Einsatz von Forschungsinformationssystemen an Universitäten und Hochschulen mit Promotionsrecht in Deutschland. Ergebnisbericht, 2014, <http://dx.doi.org/10.5281/zenodo.13841> v. 14.01.2016 und Sticht, K.: Einsatz von Forschungsinformationssystemen an Universitäten und Hochschulen mit Promotionsrecht in Deutschland. Dokumentation der Erhebung, 2014, <http://dx.doi.org/10.5281/zenodo.17169> v. 14.01.2016. Angeschrieben worden waren 88 Hochschulen mit Promotionsrecht. Eine weitere Erhebung ist dokumentiert bei: DINI AG Forschungsinformationssysteme. Forschungsinformationssysteme in Hochschulen und Forschungseinrichtungen – Request for Comments, 2014, <http://dx.doi.org/10.5281/zenodo.7697> v. 14.01.2016.

|¹⁰ <http://www.uni-giessen.de/hefis> v. 14.01.2016.

schiedlichen Abfragen anhand einer Vielzahl von nicht übereinstimmenden Definitionen erbeten werden. Zu einer nachhaltigen Reduzierung des Aufwands bei gleichzeitigem Gewinn an Systematik und Vergleichbarkeit kann deshalb nur ein standortübergreifender, koordinierter Ansatz führen. Einen solchen Ansatz hat der Wissenschaftsrat mit seinen Empfehlungen für einen Kerndatensatz Forschung initiiert. |¹¹

Die Standardisierung einer Zahl von Angaben und Datenformaten bei eigenverantwortlicher Datenverwaltung der Einrichtungen ist dabei eine Option. Eine andere, z. B. in Brasilien oder Norwegen praktizierte, Alternative ist eine zentrale, nationale Datenbank. Für eine solche Lösung spricht, dass sich im direkten Abgleich die Qualität der bereitgestellten Daten noch besser überprüfen und somit sichern ließe. Auch würden die Forschungseinrichtungen von der Bearbeitung von Anfragen und deren Prüfung, beispielsweise unter Gesichtspunkten des Datenschutzes, entlastet. Eine nationale Datenbanklösung entspräche jedoch nicht dem in Deutschland geltenden Verständnis von der Autonomie wissenschaftlicher Einrichtungen. Für die strategische Steuerung von wissenschaftlichen Einrichtungen ist es darüber hinaus zweckmäßig, Aggregatdaten durch Verknüpfung prozessbegleitend erzeugter Einzelinformationen zu erzeugen. Diese sogenannten Basisdaten sind in der Regel personenbezogen oder personenbeziehbar. Deshalb sprechen auch die besonders hohen Anforderungen an den Datenschutz, die in Deutschland gelten, gegen eine zentrale, nationale Datenbanklösung.

Der Ansatz, standardisierte Angaben dezentral vorzuhalten und auf Anfrage weiterzugeben, ermöglicht es dennoch, wesentliche Vorteile zu realisieren. Da zentrale Angaben zu Forschungsaktivitäten und deren Kontext einfacher abgefragt werden können, kann dieser Grundstock an Informationen den Bedarf an wiederholten Datenerhebungen und -abfragen innerhalb der einzelnen Einrichtung reduzieren.

Da Berichtserfordernisse nicht nur in Deutschland zunehmen, gibt es auch auf europäischer und internationaler Ebene Entwicklungen, die eine Harmonisierung erleichtern sollen. Soweit sich diese klar abzeichnen, hat der Kerndatensatz den Anspruch, anschlussfähig zu sein. |¹²

Der Wissenschaftsrat möchte mit dem Kerndatensatz keine Aussagen zur Interpretation und Bewertung der Daten machen. Dies wäre angesichts der Vielfalt institutioneller, aber auch fachbezogener Kontexte gar nicht möglich

|¹¹ Vgl. Wissenschaftsrat: Empfehlungen zu einem Kerndatensatz Forschung (Drs. 2855-13), Berlin Januar 2013.

|¹² Vgl. ebd., S. 11.

A.III ZIELE UND GRUNDSÄTZE DES KERNDATENSATZ FORSCHUNG

Die Spezifikation des Kerndatensatz Forschung standardisiert Angaben und zugehörige Datenformate, die für die Dokumentation oder Kontextualisierung von Forschungsaktivitäten bedeutsam sind. Dafür sind einheitliche Definitionen sowie Regeln für deren Handhabung nötig. Eine Standardisierung schafft die Grundlage für höhere Transparenz hinsichtlich der verwendeten Daten sowie eine bessere Qualität von Berichten. Gleichzeitig reduziert sie mittel- bis langfristig den Arbeitsaufwand. Abhängig vom Stand des Forschungsinformationssystems an den einzelnen Einrichtungen kann die Anpassung an den Standard zunächst allerdings einen Aufwand erzeugen.

Durch die Standardisierung soll in jeder Einrichtung ein einheitlichen Definitionen entsprechender Datenbestand entstehen, der sich in aggregierter Form und unter Berücksichtigung der Vorgaben des Datenschutzes zur Weitergabe eignet. Jede einzelne Forschungseinrichtung bleibt auch weiterhin Dateneignerin. Eine zentrale, einrichtungsübergreifende Datenbank zum Kerndatensatz ist nicht vorgesehen.

Die Spezifikation des Kerndatensatzes beschreibt verschiedene, vielfach schon heute administrativ erhobene und größtenteils seit Jahren für verschiedene Berichtsanlässe genutzte Daten von Forschungseinrichtungen. Diese zu standardisieren ist ein wichtiger Schritt zu einem gleichermaßen ressourcenschonenden wie die Nutzbarkeit vorhandener Daten erhöhenden Informationsmanagement.

III.1 Ziele

Mit dem Kerndatensatz Forschung verbindet der Wissenschaftsrat im Anschluss an seine Empfehlungen aus dem Jahr 2013 folgende weitere Ziele: |¹³

- _ Schaffen einer Grundlage für eine datengestützte Berichterstattung zu ausgewählten Aspekten von Forschungsaktivitäten und zu ihrer Kontextualisierung.
- _ Reduzierung des Aufwandes für wiederholte Erhebungen. Standardisierte Daten erlauben die Mehrfachnutzung für verschiedenste Anwendungen. Anlassunabhängig standardisierte Angaben können anlassbezogen auf unter-

|¹³ Wissenschaftsrat: Empfehlungen zu einem Kerndatensatz Forschung (Drs. 2855-13), Berlin Januar 2013, S. 13.

schiedliche Weise zusammengestellt und ausgewertet werden und sind so für verschiedene Berichtszwecke nutzbar.

- _ Erleichtern sowohl der internen Nutzung als auch der Weitergabe vergleichbarer Angaben zu Forschungsaktivitäten und ihrem Kontext.
- _ Sicherstellung der Anschlussfähigkeit der vorgehaltenen Daten auch im internationalen Rahmen. Daher hat der Wissenschaftsrat empfohlen, den Kerndatensatz am europäischen *Common European Research Information Format* (CERIF) zu orientieren. |¹⁴
- _ Erreichen einer verbesserten Datenqualität (vgl. B.I.1, S. 31). Klare und standardisierte Definitionen erhöhen die Aussagekraft der Daten und erleichtern ihre Nutzung. Durch mehrfache Nutzung werden Daten wiederholt überprüft und damit zusätzlich qualitätsgesichert.
- _ Standardisierung von Datenformaten als Grundlage für die vergleichbare Dokumentation von Forschungsaktivitäten und deren Kontext zwischen fachlich oder inhaltlich geeigneten Einheiten sowie für die Aggregation von Angaben, z. B zur Darstellung des Forschungsprofils einer Region.

III.2 Grundsätze

Bereits in seinen ursprünglichen Empfehlungen hat der Wissenschaftsrat wichtige Grundsätze des Kerndatensatz Forschung formuliert. |¹⁵ Sie haben die Erarbeitung der Spezifikation begleitet und sollen auch für die zukünftige Nutzung des Kerndatensatzes gelten:

- _ Die Spezifikation des Kerndatensatz Forschung schafft keine neuen Veröffentlichungspflichten.
- _ Die Spezifikation des Kerndatensatzes ist dem Prinzip der Datensparsamkeit verpflichtet.
- _ Jede Einrichtung erhebt und verwaltet ihre eigenen Informationen zu Forschungsaktivitäten selbst, sie bleibt Dateneignerin.
- _ Da die Spezifikation des Kerndatensatz Forschung Datenformate für die Weitergabe von Daten entwickelt, kann die einrichtungsinterne Organisation und Vorhaltung der den aggregierten Kerndaten zugrunde liegenden Basisdaten (vgl. Glossar, S. 57) auf unterschiedliche Art erfolgen.
- _ Für eine Weitergabe von Angaben gemäß der Spezifikation ist es somit auch nicht zwingend notwendig, ein Forschungsinformationssystem einzurich-

|¹⁴ Wissenschaftsrat: Empfehlungen zu einem Kerndatensatz Forschung (Drs. 2855-13), Berlin Januar 2013, S. 23.

|¹⁵ ebd., S. 16-19.

ten. |¹⁶ Der Wissenschaftsrat weist allerdings darauf hin, dass die Vorteile, die die Nutzung des Kerndatensatzes bietet, vor allem hinsichtlich der Qualitätssicherung und der Reduzierung des Arbeitsaufwandes, sich nur dann voll ausschöpfen lassen, wenn interne Daten prozessbegleitend und kontinuierlich erhoben und miteinander verknüpft werden.

_ Die Spezifikation des Kerndatensatz Forschung schafft keine eigenen rechtlichen Grundlagen zu Erhebung, Verarbeitung, Verwendung, Weitergabe und Übermittlung von Daten. Der Umgang mit den Daten muss im Einklang mit bestehenden rechtlichen Anforderungen, in Sonderheit denen des Daten-, Informations-, Geheimnis- und Persönlichkeitsschutzrechts, sowie ggf. weiterer vertraglicher Verpflichtungen stehen. Dies ist gängige Praxis und stellt insofern auch keine zusätzliche Aufgabe für forschende Einrichtungen dar.

_ Die Spezifikation des Kerndatensatz Forschung standardisiert Daten, die für die Dokumentation oder Kontextualisierung von Forschungsaktivitäten bedeutsam sind. Sie macht keine Aussagen hinsichtlich ihrer Interpretation und Bewertung (vgl. B.I.7, S. 35). Die Angaben im Kerndatensatz reichen keinesfalls aus, um wissenschaftliche Leistungen zu bewerten. Die Bewertung von Forschungsleistungen kann nur durch qualifizierte Peers erfolgen und erfordert zusätzliche qualitative und quantitative Informationen. Dies gilt z. B. für die Bewertung spezifischer Leistungen, die für das jeweils besondere Profil einer Einrichtung oder eines Einrichtungstyps charakteristisch sind. |¹⁷ Des Weiteren müssen auch in Zukunft weitere Angaben zusätzlich zu den im Kerndatensatz standardisierten Angaben erbeten werden.

III.3 Standardisierung in einem heterogenen Wissenschaftssystem

Ein Standard ist aufgrund historischer Entwicklungen und daraus resultierender heterogener Praktiken, vor allem auch mit Rücksicht auf den bestehenden Datenschutz, die beste Möglichkeit, die mit dem Kerndatensatz verfolgten Ziele Aufwandsreduktion und Qualitätssicherung zu erreichen (vgl. A.II.2, S. 11). Gleichwohl muss ein Vorschlag für eine Standardisierung berücksichtigen, dass

1 – auch bei Angaben, für die eine Verständigung über eine standardisierte Definition erreicht werden konnte, an einigen Einrichtungen zusätzlicher Erhebungsaufwand entstehen würde.

|¹⁶ Vgl. DINI AG Forschungsinformationssysteme. Forschungsinformationssysteme in Hochschulen und Forschungseinrichtungen – Positionspapier, 2014, <http://dx.doi.org/10.5281/zenodo.14828> v. 14.01.2016.

|¹⁷ Vgl. Wissenschaftsrat: Empfehlungen zur Bewertung und Steuerung von Forschungsleistung (Drs. 1656-11), Halle November 2011.

2 – viele wissenschaftliche Einrichtungen derzeit Forschungsinformationssysteme einführen und daher in der Anbindung des Kerndatensatzes an diese Systeme unterstützt werden sollten.

Die Architektur des Kerndatensatzes berücksichtigt diese beiden Aspekte mit der Unterscheidung von „Kern“ und „Schale“ sowie der Herstellung von Bezügen zwischen den für den Austausch vorgesehenen „Aggregatdaten“ und den innerhalb der Einrichtungen erfassten und verarbeiteten „Basisdaten“. Diese Elemente des Kerndatensatzes und Datenarten werden im Folgenden näher erläutert (vgl. auch Glossar, S. 57 f.).

III.3.a Kerndatensatz und Schale des Kerndatensatzes

Die Spezifikation des Kerndatensatzes im engen Sinn („Kern“) umfasst Datenformate für Angaben zu Forschungsaktivitäten und deren Kontext, die alle Einrichtungen im Wissenschaftssystem vorhalten und auf Anfrage und unter Beachtung rechtlicher und vertraglicher Einschränkungen weitergeben können sollten. Die Verwendung dieser Angaben ist z. B. für interne Planungszwecke und die Forschungsberichterstattung gängig.

Die in der Schale des Kerndatensatzes enthaltenen Angaben sind demgegenüber nur für spezifische Einrichtungen relevant bzw. aussagekräftig und ergänzen somit den eigentlichen Kerndatensatz. Auch für diese Elemente empfiehlt der Wissenschaftsrat, die vorgeschlagenen Definitionen zu verwenden, um den Datenaustausch zu erleichtern.

Für die Verortung in der Schale des Kerndatensatzes gibt es unterschiedliche Gründe: Beispielsweise gibt es für außeruniversitäre Forschungseinrichtungen bislang wenig Anlass, die Bezeichnung von Professorinnen und Professoren jenseits der „ordentlichen Professur“ systematisch zu dokumentieren (z. B. Stiftungsprofessuren), für Hochschulen hingegen sind diese Unterscheidungen zentral. Ein anderes Beispiel ist der Titel des übergeordneten Projektes bei Verbundvorhaben. Informationen darüber liegen an einer substantiellen Zahl von Einrichtungen nicht vor, so dass dieser Aspekt aus Gründen der Aufwandsvermeidung als Teil der Schale eingeordnet wurde.

Angaben zu laufenden Promotionen sind im Unterschied zu den genannten Beispielen als Teil des Kerndatensatzes ausgewiesen, obwohl Informationen hierzu in vielen Fällen noch nicht (lange) erfasst werden. Es besteht jedoch ein breiter Konsens in Wissenschaft und Wissenschaftspolitik, dass die Informationsdefizite in diesem wichtigen Bereich behoben werden sollten. Entsprechend

existieren auch Absichten, dies in der amtlichen Hochschulstatistik zu verankern. |¹⁸

Jenseits von Kern und Schale wurde in der Vorbereitung des Standards eine Vielzahl weiterer inhaltlicher Elemente diskutiert, für die aber zurzeit noch keine standardtauglichen Definitionen vorliegen (z. B. Mitarbeiterinnen und Mitarbeiter mit Weisungsbefugnis, Transfer jenseits der etablierten Indikatoreik, vgl. C.II.1, S. 50). Diese Desiderate sollten im Zuge der Weiterentwicklung der Spezifikation des Kerndatensatzes geprüft und ggf. bearbeitet werden.

III.3.b Basis- und Aggregatdaten

Auch Einrichtungen, die bislang nicht über ein Forschungsinformationssystem verfügen, können Daten entsprechend der Spezifikation des Kerndatensatzes austauschen. Gleichwohl spielt die Spezifikation für viele Einrichtungen, die solche Systeme bereits nutzen oder derzeit aufbauen bzw. beschaffen, eine wichtige Rolle. Bei der Erarbeitung der Spezifikation wurden daher auch Vorschläge erarbeitet, die die einrichtungsinterne Datenhaltung unterstützen sollen. Aus diesem Bestreben resultiert die Unterscheidung zwischen Basisdaten und Aggregatdaten, die im Folgenden erläutert wird. |¹⁹

Entscheidet sich eine Einrichtung, Daten gemäß dem Kerndatensatz-Standard an Datennutzer weiter zu geben, so gibt sie nur diese Aggregatdaten bzw. zu Listen gruppierte Daten weiter (vgl. Abbildung 1, S. 19). |²⁰ Die Aggregatdaten fassen Sachverhalte für eine Organisationseinheit wie ein Institut, ein Zentrum oder eine Abteilung oder aber für ein Fach (vgl. Glossar, S. 57) innerhalb der jeweiligen Einrichtung zusammen. Falls aggregierte Daten zur Weitergabe (z. B. Anzahl der Professorinnen der Geographie) nicht manuell (z. B. einmal jährlich) zusammengestellt werden sollen, ist einrichtungsintern die Erfassung und Verarbeitung von Basisdaten erforderlich (z. B. für jede beschäftigte Person Merkmale wie Personalkategorie, Geschlecht, Fach). Die Spezifikation enthält entsprechende Vorschläge.

|¹⁸ Vgl. Bundesrat. Entwurf eines Gesetzes zur Änderung des Hochschulstatistikgesetzes (Drs. 394/15 vom 4. September 2015).

|¹⁹ Diese hatte der Wissenschaftsrat in seinen ursprünglichen Empfehlungen noch nicht getroffen; Die Unterstützung der einrichtungsinternen Datenhaltung (von Basisdaten) war nicht vorgesehen. In der ursprünglichen Terminologie entsprechen „Kerndaten“ den Aggregatdaten.

|²⁰ Im Kerndatensatz sind zwei Typen von zur Weitergabe vorgesehenen Aggregatdaten angelegt: zum einen Summenaggregate wie z. B. zehn Professorinnen W1 und acht Professorinnen W3. Zum anderen werden Daten in Listenform zusammengefasst. So können z. B. Drittmittelprojekte mit Titel, Beginn und Ende pro Institut oder Fach aufgelistet werden. Insbesondere für Evaluationen und andere qualitative Begutachtungen stellen Listen elementare Information dar.

Für die Erhebung und Verarbeitung von Basisdaten sowie ihre Verknüpfung muss eine rechtliche Basis gegeben sein. |²¹ Basisdaten oder Daten, die auch nach Aggregation Rückschlüsse auf einzelne Personen erlauben, dürfen nur weitergegeben werden, wenn es dafür eine gesetzliche Grundlage gibt oder die betroffenen Individuen zustimmen (zu datenschutzrechtlichen Aspekten des Kerndatensatz Forschung vgl. B.I.3).

III.4 Kerndatensatz und Berichtswesen

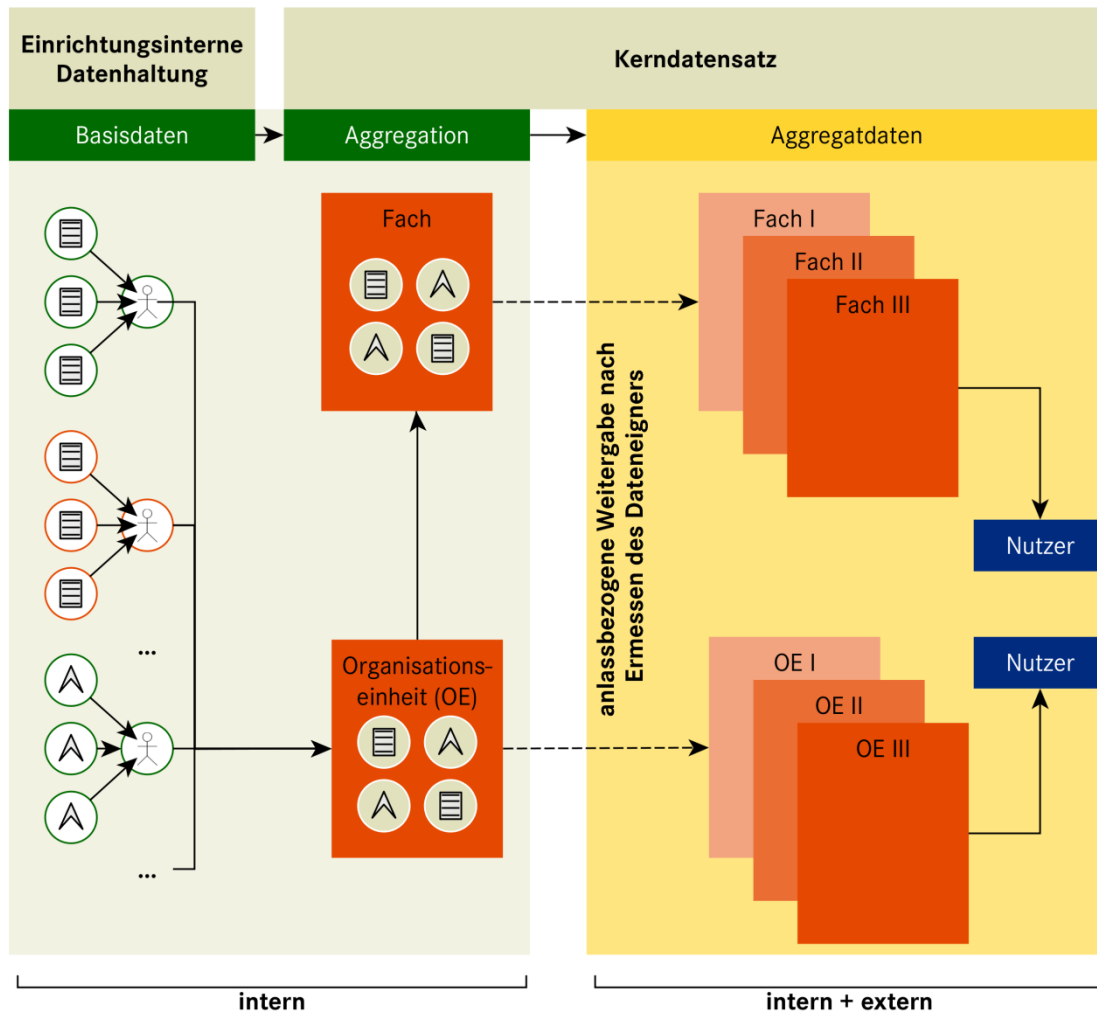
Abbildung 1 stellt das Zusammenspiel der genannten Elemente des Kerndatensatz Forschung sowie der Schale des Kerndatensatzes schematisch dar. Eine detailliertere Erläuterung bietet die Spezifikation. |²²

In Abbildung 1 ist links die einrichtungsinterne Haltung der Basisdaten dargestellt (z. B. Angaben zu Publikationen oder Drittmittelprojekten; Begriffe vgl. Glossar; S. 57 sowie Kapitel A.III.3.b, S. 17). Sie findet – teils datenbankgestützt, teils nicht – auch unabhängig vom Kerndatensatz Forschung statt. Die Spezifikation des Kerndatensatzes macht Vorschläge für die Organisation dieser einrichtungsinternen Prozesse, die als Service für implementierende Einrichtungen gedacht sind (z. B. welche Basisdaten für die Aggregation im Kerndatensatz vorliegen sollten). Solche Basisdaten werden bspw. für Zwecke der Personalverwaltung oder der Buchführung ohnehin erhoben, werden von den Einrichtungen in der Regel jedoch nicht weiter gegeben (z. B. Name des *Principal Investigator* eines Drittmittelprojektes) |²³. Eine Ausnahme bilden allerdings Publikationen, bei denen es von Seiten der Autorinnen und Autoren in der Regel gewünscht und durch das Urheberrecht auch gefordert ist, dass ihr Name mit dem Produkt ihrer Arbeit in Verbindung gebracht wird.

|²¹ Aufgrund der Heterogenität der gesetzlichen Grundlagen in den Bundesländern können diese nicht Gegenstand der vorliegenden Empfehlungen sein. Dass Hochschulen in verschiedenen Bundesländern bereits mit Zustimmung der jeweiligen Datenschutzbeauftragten Forschungsinformationssysteme eingeführt haben, legt jedoch nahe, dass die rechtlichen Voraussetzungen für die Erfassung, Verarbeitung und Verknüpfung von Basisdaten vielfach gegeben sind.

|²² Vgl. Spezifikation des Kerndatensatz Forschung, Version 1.0, in der Anlage.

|²³ Die individuelle Datenweitergabepaxis einzelner Wissenschaftlerinnen und Wissenschaftler ist nicht Gegenstand dieser Empfehlungen.



Quelle: Wissenschaftsrat.

Hauptzweck der Spezifikation ist die Standardisierung der für den Austausch bestimmten Daten. Vor der Weitergabe müssen daher intern vorliegende Basisdaten aggregiert (bzw. gruppiert) werden. Für die Aggregation werden mehrere mögliche Verfahren unterschieden. Für nahezu alle Kerndaten sind

- a) die Organisationseinheit (d. h. Institut, Zentrum, Abteilung ...),
- b) das Fach (Zuordnung gemäß der Organisationseinheit auf Basis der etablierten Fächersystematik des Statistischen Bundesamtes) und
- c) das Forschungsfeld (disziplinunabhängige, thematisch abgrenzbare, problemorientierte Forschung, vgl. auch Glossar, S. 57)

mögliche Aggregationen. |²⁴

|²⁴ Die Zuordnung zu Forschungsfeldern sollte nicht nach selbstgewählten Bezeichnungen, sondern anhand einer standardisierten Liste erfolgen, die noch nicht vorliegt. Einrichtungen, die ihre Berichtssysteme am Kerndatensatz Forschung ausrichten, sollten diese Klassifikationsmöglichkeit dennoch schon vorsehen. Vgl. Projektgruppe „Fächerklassifikation und Thesauri“: Interdisziplinarität im Forschungsberichtswesen – Emp-

Als Aggregatdaten stehen dann Informationen wie z. B. die Anzahl der Wissenschaftlerinnen in der Geographie, die Summe der Drittmittel im Institut für Astrophysik oder eine Liste der laufenden Drittmittelprojekte im Forschungsfeld Sicherheitsforschung für die interne oder externe Nutzung bereit. Eine Weitergabe nach außen erfolgt dabei grundsätzlich nicht automatisch, sondern nur anlassbezogen. Die erhebende Einrichtung ist Dateneignerin und entscheidet im Rahmen der rechtlichen Vorgaben über die Weitergabe der angefragten Daten (vgl. B.I.4, S. 32).

III.5 Ziele und Aufwand der intendierten Nutzung

Die heutige Praxis der Forschungsberichterstattung ist äußerst heterogen. Dies führt zu einer verzerrten Dokumentation. Die Einführung des Kerndatensatz Forschung verspricht nach Auffassung des Wissenschaftsrates, zur Lösung dieses Problems beizutragen.

Die Einführung eines Standards in eine heterogene Praxis hinein kann ihren Nutzen aber nur entfalten, wenn die beteiligten Akteure Anpassungen vornehmen. Daher kann zu Beginn für die Einführung des Kerndatensatz Forschung ein einmaliger Aufwand entstehen. Dieser wird – je nach Einrichtung – unterschiedlich ausfallen. Dies hängt insbesondere vom jeweiligen Stand der Entwicklung des Berichtswesens einer Einrichtung und seiner technischen Ausgestaltung ab. Die meisten Angaben, die der Kerndatensatz einer Einrichtung gemäß Spezifikation enthält, liegen heute bereits im Routinebetrieb von Einrichtungen vor. Die Standardisierung kann dennoch die Anpassung der internen Erfassung dieser Daten erfordern. Der wesentliche Aufwand für eine Einrichtung bei der Implementierung des Kerndatensatzes liegt somit darin, Prozesse umzustellen, für die auch bisher schon Mittel aufgewendet werden. Nur soweit der Kerndatensatz Definitionen für Sachverhalte enthält, die in einer Einrichtung bisher nicht erhoben werden, entsteht dauerhaft ein neuartiger Aufwand.

Die Spezifikation des Kerndatensatz Forschung enthält einen Grundbestand an Angaben, der sich einrichtungsübergreifend standardisieren lässt. Aus Sicht des Wissenschaftsrates sollte jede Einrichtung, die forschend tätig ist, auch unabhängig von externen Berichtspflichten sich selbst mindestens über diesen Grundbestand im vorgeschlagenen Umfang Rechenschaft ablegen können. Für die interne Steuerung dürften in der Regel weitere, über den Kerndatensatz hinausgehende Daten nötig sein.

Es ist darüber hinaus zu berücksichtigen, dass der Kerndatensatz vorrangig ein Angebot zur Vereinheitlichung von Datendefinitionen im Bereich der Forschung ist. Er enthält Elemente (z. B. Angaben zu Beschäftigten), die auch in die Beschreibung der Rahmenbedingungen anderer Bereiche eingehen können, z. B. der Lehre.

Die folgende Übersicht fasst die wichtigsten Ziele bzw. die intendierte Nutzung des Kerndatensatz Forschung zusammen und grenzt sie von anderen, bisweilen auf das Vorhaben projizierten Nutzungsweisen ab, für die sich der Kerndatensatz nicht eignet.

Tabelle 1: Gegenüberstellung von Zielen und intendierter Nutzung versus zweckwidrigen und nicht intendierten Nutzungsweisen

Ziele / intendierte Nutzung	Zweckwidrige und nicht intendierte Nutzung
Aufwandsreduktion bei Datenbereitstellung	Bewertungsinstrument
Standardisierung von Datenformaten	Einrichtungsübergreifende Datenbank
Freiwilligkeit auf Basis von Nutzenabwägungen der Einrichtungen	Pflicht zur Einführung eines Forschungsinformationssystems
Basis für <i>Peer-review</i> Prozesse	Ersatz von <i>Peer-review</i> Prozessen
Unterstützung der internen Steuerung (auch mithilfe von Benchmarks)	Entscheidungsgrundlage für indikatorgestützte Mittelverteilung
Erhalt der Dateneignerschaft der Forschungseinrichtungen	Veröffentlichungspflicht
Dokumentation fachspezifischer Besonderheiten durch Aggregationsmöglichkeit nach Fächern	Homogenisierung von Indikatorik über alle Fächer hinweg
Anlassbezogene Dokumentation aggregierter Angaben (z. B. regional, deutschlandweit)	Rankings

Quelle: Wissenschaftsrat.

Ohne einen Standard stünde zu erwarten, dass schlecht vergleichbare Daten immer häufiger, intensiver und in zunehmend heterogener Weise genutzt werden. Das Fehlen einer qualitätsgesicherten Dokumentation würde zudem die Aussagekraft weiterhin begrenzt halten.

A.IV BISHERIGER PROZESS ZUR ERSTELLUNG EINER SPEZIFIKATION

Der Wissenschaftsrat hat in seinen Empfehlungen aus dem Jahr 2013 wichtige Eckpunkte für den Prozess zur Erarbeitung der Spezifikation des Kerndatensatzes empfohlen. |²⁵ Er betrachtet die Umsetzung dieser Empfehlungen durch das Projekt als sehr gelungen und sieht in den nun vorliegenden Ergebnissen eine gute Grundlage für die breite Verwendung des Kerndatensatzes. Im Fol-

|²⁵ Vgl. Wissenschaftsrat: Empfehlungen zu einem Kerndatensatz Forschung (Drs. 2855-13), Berlin Januar 2013.

genden wird der bisherige Prozess dargestellt. Diese Übersicht soll zur Einordnung der in Kapitel B vorgestellten Projektergebnisse und als Hintergrund für die in Kapitel C formulierten Empfehlungen zur Einführung und Weiterentwicklung des Kerndatensatzes im deutschen Wissenschaftssystem dienen.

IV.1 Prozess

Nach der Veröffentlichung der Empfehlungen des Wissenschaftsrates im Januar 2013 förderte das BMBF ab dem Herbst desselben Jahres ein Projekt unter Federführung des Instituts für Forschungsinformation und Qualitätssicherung (iFQ) in Zusammenarbeit mit dem Fraunhofer-Institut für Angewandte Informationstechnik (FIT) sowie der Geschäftsstelle des Wissenschaftsrates. Das Projekt wurde durch einen vom Wissenschaftsrat eingesetzten Beirat begleitet und bestand aus vier Projektgruppen, die von einem Projektmanagement koordiniert wurden.

Als Piloteinrichtungen stellten zwei Universitäten, eine Fachhochschule und ein Institut der Helmholtz Gemeinschaft dem Projekt ihre Expertise hinsichtlich der praktischen Umsetzung der in Erarbeitung befindlichen Vorschläge zur Verfügung. Zudem haben sich FhG, HGF, MPG sowie WGL intensiv in den Prozess eingebracht. |²⁶

IV.1.a Projektgruppe „Definitionen“

Die Projektgruppe „Definitionen“ erarbeitete auf Basis der Empfehlungen des Wissenschaftsrates Vorschläge für die Definition der Elemente des Kerndatensatzes. |²⁷ Im Zuge dessen sichtete und diskutierte die Gruppe bestehende und teilweise widersprüchliche Definitionsansätze umfassend. Daraus resultierten Vorschläge für die Definition von Kerndaten und zu deren Aggregation. In die Bestandsaufnahme wurden u. a. die 16 Landesstatistikgesetze, das *Frascati Manual*, die Hochschul-, Personal- und Finanzstatistik sowie das *Common European Research Information Format (CERIF)* einbezogen.

In einer knapp viermonatigen Konsultation wurden die Pilot- und Kooperationseinrichtungen um Rückmeldungen zum Diskussionsstand sowie zu den Definitionsentwürfen gebeten. Abweichende bzw. ablehnende Stellungnahmen sollten begründet sowie ggf. Alternativvorschläge gemacht werden. Die Rückmeldungen aus dieser Konsultation wurden anschließend in der Projektgruppe diskutiert und in die Spezifikation integriert. Zusätzlich zu den für den Aus-

|²⁶ Vgl. Fraunhofer Gesellschaft, Helmholtz Gemeinschaft, Max-Planck Gesellschaft, Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz. Kerndatensatz Forschung. Vorschlag der außeruniversitären Forschungseinrichtungen, 2015.

|²⁷ Vgl. Wissenschaftsrat: Empfehlungen zu einem Kerndatensatz Forschung (Drs. 2855-13), Berlin Januar 2013.

tausch vorgesehenen Daten erarbeitete die Gruppe auch einen Vorschlag, der Einrichtungen dabei unterstützen soll, Datenbanken aufzubauen, die Berichte nach der Kerndatensatzspezifikation erzeugen können. Die Arbeit dieser Projektgruppe stellt das Fundament des Kerndatensatzes dar.

IV.1.b Projektgruppe „Bibliometrie“

Die Projektgruppe „Bibliometrie“ war mit der Erarbeitung eines Vorschlages zur Erfassung von Publikationsdaten im Rahmen des Kerndatensatzes befasst. Der von dieser Gruppe vorgelegte Vorschlag unterscheidet sich von den anderen Bereichen im Kerndatensatz Forschung: für den Austausch von Publikationsdaten sieht er auch Basisdaten, die eine qualitative Auswertung der Publikationen durch *Peers* ermöglichen sollen, vor. Die Zahl der zu erfassenden Deskriptoren ist hoch, da im Interesse einer breiten fachlichen Abdeckung eine Vielzahl verschiedener Publikations- und Dokumenttypen berücksichtigt wurde.

Zur Spezifikation der Publikationstypen wurde eine Bestandsaufnahme aller im *Registry of Open Access Repositories* genannten deutschen Publikations-Repositorien durchgeführt, anhand derer eine konsolidierte Liste der bestehenden Publikationstypen erstellt wurde. Eine erste Empfehlung wurde an 259 Fachgesellschaften zur Kommentierung verschickt. Die Prüfung der prinzipiellen Kompatibilität bestehender Publikationsdaten aus Systemen der Piloteinrichtungen erfolgte schriftlich und war aus der Sicht der Piloteinrichtungen im Wesentlichen gegeben.

Nach Sichtung und Diskussion bestehender Vorlagen, u. a. bei *Web of Science*, beim Norwegischen Forschungsinformationssystem CRISTin sowie bei EuroCRIS (CERIF), wurde ein Metadatenmodell erstellt, das Publikationen im Kerndatensatz Forschung abbildet. Die Auseinandersetzung mit kommerziellen Anbietersystemen wie beispielsweise *Web of Science* zeigte, dass allein über Importe aus diesen Systemen keine vollständige Erfassung von Publikationen der Einrichtungen erreicht werden kann. Die Projektgruppe empfiehlt daher zunächst eine Mischung aus Datenbankimporten, Importen aus lokalen Systemen und manueller Eingabe.

IV.1.c Projektgruppe „Fächerklassifikation und Thesauri“

Die Einrichtung einer Projektgruppe „Fächerklassifikation und Thesauri“ hatte der Wissenschaftsrat empfohlen, um Fragen der Aggregation anhand von Fächersystematiken sowie Möglichkeiten der Klassifikation interdisziplinärer

Forschungsaktivitäten zu prüfen. |²⁸ Die inhaltliche Zusammenfassung von Forschungsaktivitäten nach Fächern ist etabliert. Als Vorschlag für die Standardisierung wurde daher eine Klassifikation gesucht, die es einerseits ermöglicht, Forschung adäquat z. B. Fächern zuzuordnen, gleichzeitig bei den Einrichtungen aber möglichst keine Anpassungskosten verursacht. Die Klassifikation des Statistischen Bundesamtes erfüllt diese Maßgaben weitgehend. Für eine forschungsadäquatere Zuordnungsmöglichkeit wurden Modifikationsvorschläge entwickelt, die inzwischen weitestgehend in die amtliche Statistik übernommen wurden. Für die gute Zusammenarbeit dankt der Wissenschaftsrat dem Statistischen Bundesamt, insbesondere seinem Ausschuss für die Hochschulstatistik. Die enge Zusammenarbeit hat die großen Vorteile einer engen Anbindung an die amtliche Statistik deutlich gemacht: Die Verbindlichkeit der dort etablierten Prozesse garantiert deren breite Bekanntheit und Verwendung und somit in hohem Maße qualitätsgesicherte Daten.

Zu den unterschiedlichen Aspekten der Projektgruppenarbeit wurden insgesamt 52 Nutzerinnen und Nutzer, Anbieter von Fächersystematiken sowie Fachgesellschaften konsultiert. Auch für zukünftige – vom Kerndatensatz unabhängige – regelmäßige Revisionsprozesse der Klassifikation unter maßgeblicher Beteiligung der Wissenschaft wurden im Rahmen des Projektes Vorschläge erarbeitet.

Die Erfassung – auch von missionsorientierten – Forschungsaktivitäten quer zu den Grenzen etablierter Fächer wird immer wichtiger. Der Wissenschaftsrat regte daher 2013 an, die hierarchische Fächerklassifikation durch eine Möglichkeit zu ergänzen, die es zulässt, interdisziplinäre und/oder sich dynamisch entwickelnde Forschungsfelder besser dokumentieren zu können. Die Projektgruppe sprach sich gegen die Erarbeitung eines Thesaurus aus, wie es der Wissenschaftsrat ursprünglich vorgeschlagen hatte. Der hierfür notwendige Aufwand bei Erstellung und Pflege wurde als zu groß bewertet.

Als praktikablere Lösung empfiehlt die Projektgruppe für gegenstands- oder problembezogene Forschung, die quer zu den Grenzen etablierter Fächer liegt, die Erstellung einer zusätzlichen Klassifikation nach Forschungsfeldern. Listen solcher Forschungsfelder enthalten Schlagworte wie „Energieforschung“, „Sicherheitsforschung“ oder „*Global Ethics*“ und sind schon heute etabliert (z. B. für *Horizon 2020* oder die Liste „Forschungsthemen Fraunhofer-Gesellschaft“). Sie zeichnen sich derzeit aber durch eine große, teils organisationspezifische, Vielfalt aus. Um eine gemeinsame Berichtsgrundlage zu erzeugen, die gleichzeitig auch die nötige Differenzierung aufweist, gilt es daher, bestehende

|²⁸ Vgl. Projektgruppe „Fächerklassifikation und Thesauri“: Interdisziplinarität im Forschungsberichtswesen – Empfehlungen zur Abbildung; http://www.kerndatensatz-forschung.de/version1/PGK_Empfehlungen_zur_Abbildung_von_Interdisziplinaritaet.pdf, abgerufen am 14.01.2016.

Schlagworte zu harmonisieren und in ein kombinatorisches System zusammenzuführen. |²⁹

IV.1.d Projektgruppe „Technik“

Die Arbeit der Projektgruppe „Technik, Schnittstellen und Datenformate“ konzentrierte sich auf die Datenmodellierung und die Umsetzung der Vorgaben und Definitionen der anderen Projektgruppen. Das Datenmodell enthält eine formale Beschreibung der zum Kerndatensatz einschließlich der Schale gehörenden Objekte mit ihren Eigenschaften und Beziehungen. Es kann somit den Aufbau bzw. eine Anpassung von Forschungsinformationssystemen und einzelnen Datenbanken unterstützen. Für die Aggregatdaten existiert zudem ein XML-Schema, das als Basis für den Datenaustausch dienen kann. Die Projektgruppe „Technik“ hat damit auch im Wortsinn eine Schnittstellenfunktion für die Umsetzung des Kerndatensatz Forschung in den Alltag von Forschungseinrichtungen übernommen. Im Vorlauf der Modellierung wurden existierende Forschungsinformationssysteme der Piloteinrichtungen und an den Forschungseinrichtungen der Projektgruppenmitglieder untersucht. Die Verbindung zum europäischen CERIF-Standard wurde durch Verknüpfungen von einzelnen Elementen des Datenmodells mit den entsprechenden Elementen in CERIF realisiert. Entwickelt wurde auch ein technisches Datenmodell, dessen Elemente auf den europäischen CERIF-Standard und damit auch auf gängige Forschungsinformationssysteme abgebildet werden können. |³⁰

IV.1.e Alpha-Version und Abnahme

Nachdem die Vorschläge in den Projektgruppen erarbeitet worden waren, befasste sich der Beirat Mitte Mai 2015 mit einem vom Projekt vorgelegten ersten Entwurf der Spezifikation („Alpha-Version“). Dabei wurden auch Vorschläge einer AG von FhG, HGF, MPG und WGL beraten. Diese hatten viele Ähnlichkeiten zu den im Projekt erarbeiteten Vorschlägen, wodurch sich beide gut zur sogenannten „Beta-Version“ verbinden ließen.

Im Rahmen dieser Beratungen verständigten sich die Beteiligten, einen neuen Bereich zu Forschungsinfrastrukturen in die Schale des Kerndatensatzes aufzunehmen. Der Vorschlag greift die vielfältigen Entwicklungen auf, die in den

|²⁹ Im Unterschied zu einer Zuordnung zu Fächern, lassen Forschungsfelder eine überschneidungsfreie Zuordnung nicht zu. Vgl. Projektgruppe „Fächerklassifikation und Thesauri“: Interdisziplinarität im Forschungsberichtswesen – Empfehlungen zur Abbildung; http://www.kerndatensatzforschung.de/version1/PGK_Empfehlungen_zur_Abbildung_von_Interdisziplinaritaet.pdf, abgerufen am 14.01.2016.

|³⁰ Das technische Datenmodell ist online hinterlegt. www.kerndatensatzforschung.de/version1/technisches_datenmodell. Es bezieht sich auf die Basisdaten (vgl. A.III.3.b; S. 16 und Glossar, S. 51).

letzten Jahren in diesem Bereich stattgefundenen haben. Er bietet so die Möglichkeit, (überregional bedeutsame) außeruniversitäre und hochschulische Forschungsinfrastrukturen vom Forschungsschiff über verteilte Datensammlungen bis hin zu sozialen Forschungsinfrastrukturen einheitlich zu dokumentieren. |³¹ Der Beirat befürwortete außerdem die Aufnahme eines kleinen Sets an wichtigen Preisen und Auszeichnungen in die Beta-Version des Kerndatensatzes.

Trotz weit gehender Verständigung erschweren einige Unterschiede zwischen hochschulischer und außeruniversitärer Forschung das Finden gemeinsamer Berichtsgrößen. Wichtige Angaben wie v. a. zum Gesamtbudget einer Einrichtung werden daher hinsichtlich ihrer Belastbarkeit und Vergleichbarkeit zunächst noch Wünsche offen lassen – insbesondere im Hochschulsektor. Dennoch scheinen sie als Ausgangsbasis für weitere Entwicklungen geeignet und sind daher Teil der Spezifikation.

IV.1.f Öffentliche Beta-Phase

Die vom Beirat angepasste Fassung der Spezifikation wurde Anfang Juni 2015 für eine knapp zweimonatige Beta-Phase zur öffentlichen Kommentierung bereitgestellt. Diese Konsultation sollte helfen, mögliche Probleme und Missverständnisse zu identifizieren, die trotz größter Sorgfalt bei der Erarbeitung nicht ausgeschlossen werden konnten. Ziel war es zu eruieren, welche Vorschläge u. U. nicht für eine Aufnahme in die erste Version des Kerndatensatzes reif waren. Darüber hinaus sollten Anregungen für die zukünftige Weiterentwicklung des Kerndatensatzes gesammelt werden.

Im Rahmen der Beta-Phase fand Ende Juli 2015 ein fachöffentliches „Forum Kerndatensatz“ statt, das Gesprächsmöglichkeiten zu Schwerpunkten der Spezifikation sowie zu Nutzungsoptionen des Kerndatensatz Forschung bot. Dadurch trug es dazu bei, den konstruktiven Austausch zu den Vorschlägen der Spezifikation zu vertiefen. Auch hierbei wurden Rückmeldungen zu den Vorschlägen eingeholt und systematisch ausgewertet.

Insgesamt gingen schriftliche Rückmeldungen von 100 Hochschulen, AUF, Ministerien und anderen Einrichtungen oder Organisationen ein. Das eingesetzte standardisierte Rückmeldetool ermöglichte jeder registrierten Person, einen übergeordneten Kommentar abzugeben. Davon haben 81 Rückmeldende Ge-

|³¹ Vgl. u. a. Wissenschaftsrat: Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020 (Drs. 2359-12), Berlin Juli 2012. Vgl. auch: Wissenschaftsrat: Bericht zur wissenschaftsgeleiteten Bewertung umfangreicher Forschungsinfrastrukturvorhaben für die Nationale Roadmap (Pilotphase), (Drs. 2841-13), Köln April 2013. Vgl. auch: Wissenschaftsrat: Konzept für eine wissenschaftsgeleitete Bewertung umfangreicher Forschungsinfrastrukturvorhaben für eine Nationale Roadmap (Pilotphase), (Drs. 1766-11), Köln Dezember 2011. Wissenschaftsrat: Perspektiven des deutschen Wissenschaftssystems (Drs. 3228-13), Braunschweig Juli 2013.

brauch gemacht. |³² Dazu kommen insgesamt 1820 Rückmeldungen zu einzelnen Elementen bzw. Definitionsvorschlägen. Alle Kommentare, die mit Erläuterungen versehen waren, wurden dabei ausgewertet. Diese bezogen sich auf 289 Elemente oder Ausdifferenzierungen, die meisten davon (95) im Bereich Publikationen. Obwohl es das erklärte Ziel der Beta-Phase gewesen war, mögliche Probleme und Missverständnisse zu identifizieren, war ein großer Teil der Rückmeldungen zustimmend oder kommentierte ohne Bewertung. |³³

Da in den Rückmeldungen vielfach ähnliche Anliegen behandelt wurden und Kommentare sich auf mehrere Definitionsvorschläge bezogen, befasste sich der Beirat bei seiner Prüfung mit 149 Elementen der Beta-Version, bei denen entweder überwiegend kritische Kommentierungen eingegangen oder inhaltliche Gründe für Anpassungsbedarf genannt worden waren. Die vorgenommenen Anpassungen konnten Anliegen aus 615 elementspezifischen Kommentaren |³⁴ und weiteren übergreifenden Kommentaren aufgreifen.

Insgesamt wurde deutlich, dass verschiedene Einrichtungen bei der Anwendung neuer Formatvorgaben für die Bereitstellung ihres Kerndatensatzes zunächst einen Mehraufwand erwarten. Dieser soll und kann nicht verneint werden (vgl. A.III.5, S. 20). Sofern er sich auf die Erfassung laufender Promotionen bezieht, sei darauf verwiesen, dass auch die amtliche Statistik diese Angaben in Bälde – verpflichtend – erheben wird. |³⁵

Sofern vor allem ein Aufwand bei der Verknüpfung von derzeit an verschiedenen Stellen bereits existierenden Daten gesehen wird, erscheint dieser dem Wissenschaftsrat mit Blick auf den mittel- und langfristigen Nutzen für ein gleichermaßen ressourcenschonendes wie an Gehalt gewinnendes Berichtswesen für interne wie externe Zwecke gerechtfertigt.

IV.1.g Anpassungen im Anschluss an die Beta-Phase

Bei der Sichtung und der Entscheidung über die Berücksichtigung der Rückmeldungen aus der Beta-Phase standen die Zielsetzungen des Kerndatensatzes im Mittelpunkt: Die vorzuschlagenden Definitionen sollten in ihrer Substanz in möglichst vielen Verwendungszusammenhängen aussagekräftig sein. Gleichzeitig sollte ihre Bereitstellung möglichst keinen zusätzlichen Aufwand

|³² Weitere Rückmeldungen wurden außerhalb des standardisierten Tools eingereicht.

|³³ So gingen neben 626 zustimmenden und 469 Kommentierungen „ohne Bewertung“ 725 kritischen Einschätzungen ein.

|³⁴ Dies entspricht 48 Prozent der 1275 Kommentare, bei denen ein Erläuterungstext vorliegt und es somit möglich ist zu beurteilen, ob ein Kommentar berücksichtigt wurde.

|³⁵ Vgl. Bundesrat. Entwurf eines Gesetzes zur Änderung des Hochschulstatistikgesetzes (Drs. 394/15 vom 4. September 2015).

bedeuten. Im Folgenden sind die wichtigsten aus der Beta-Phase resultierenden Anpassungen dokumentiert.

Da die Spezifikation des Kerndatensatzes keine von der gängigen Praxis abweichende Erfassung von Personalkategorien einführen soll, wurden die Bezeichnungen in diesem Bereich gegenüber der Beta-Version an die amtliche Statistik angeglichen.

Für Drittmittel, die in Deutschland ein wichtiger Deskriptor für Forschungsaktivität sind, ist die Verständigung auf eine belastbare Definition besonders wichtig. In der amtlichen Statistik, aber auch in einer Vielzahl anderer Berichte werden Drittmiteleinahmen bzw. -erträge erfasst. Der Wissenschaftsrat hatte hingegen im Jahr 2013 empfohlen, im Kerndatensatz verausgabte Drittmittel zu standardisieren, da sich damit aus damaliger Sicht der Zeitraum des Mitteleinsatzes für Forschungsaktivitäten am präzisesten beschreiben ließe. Die Rückmeldungen in der Beta-Phase haben aber deutlich gemacht, dass an einer Vielzahl von Einrichtungen ein solches Datum erst nach langer Übergangszeit und mit hohem Aufwand qualitätsgesichert berichtet werden könnte. Durch eine veränderte Zuwendungspraxis bei wichtigen Mittelgebern hat sich das Problem, das durch einmalige Zuwendungen zu Beginn eines Drittmittelprojektes geschaffen wurde, in der Zwischenzeit allerdings ohnehin teilweise gelöst. Auch erfassen Erträge bei kaufmännischer Rechnungslegung weitgehend dasselbe wie Verausgabungen. Daher wurde die Spezifikation der gängigen Praxis entsprechend angepasst.

Auch im Bereich Nachwuchsförderung konnte dank der Rückmeldungen eine wichtige Verbesserung vorgenommen werden: so wird der Kerndatensatz zwischen Doktorandinnen und Doktoranden an titelvergebenden Einrichtungen und jenen, die an nicht-titelvergebenden Einrichtungen (mit)betreut werden, unterscheiden. Damit entsteht einerseits eine eindeutige Statistik, andererseits aber auch eine Dokumentation des Ausbildungsbeitrags z. B. von Hochschulen für Angewandte Wissenschaften und außeruniversitären Forschungseinrichtungen.

Durch eine Vielzahl vorwiegend kritischer Rückmeldungen wurde deutlich, dass sich der Bereich „Forschungspreise und Auszeichnungen“ derzeit noch nicht für die Standardisierung von Angaben und Datenformaten eignet. Er ist daher nicht Teil der Spezifikation in Version 1.0. Der Wissenschaftsrat erachtet es dennoch als wertvoll, wenn Einrichtungen Preise und Auszeichnungen, die durch Dritte vergeben werden, schon jetzt zu erfassen beginnen. Mit zunehmender Erfahrung kann ggf. zukünftig ein Standard entwickelt werden, um Informationen zu erhaltenen Preisen und Auszeichnungen systematisch in ein fachspezifisches wie fachübergreifendes Berichtswesen zu integrieren.

Deutlich betont der Wissenschaftsrat, dass der Bereich „Patente und Ausgründungen“ nicht beansprucht, den wesentlich umfangreicheren und durch stan-

dardisierte Berichtsgrößen derzeit kaum beschreibbaren Begriff „Transfer“ abzudecken. Hier führte die Berücksichtigung der Rückmeldungen insgesamt zu einer Reduzierung der ursprünglich vorgeschlagenen Kerndaten. Um hinsichtlich dieser, für alle Akteure des Wissenschaftssystems zunehmend wichtigen Dimension von Innovationsaktivitäten und Transfer eine weitergehende Verständigung auf Grundsätze einer angemessenen Dokumentation zu erreichen, bedarf es weiterer Anstrengungen aller beteiligten Akteure (vgl. C.II.1, S. 50).

Im Bereich Forschungsinfrastrukturen konnten aufgrund der Rückmeldungen mehrere Definitionen nachgeschärft werden.

Die meisten Rückmeldungen – kritisch wie auch zustimmend – entfielen auf den Bereich Publikationen. Auch hier konnten dank des Feedbacks der Leserinnen und Leser der Beta-Spezifikation etliche Vereinfachungen und Klarstellungen vorgenommen werden.

Sollen Angaben von Forschungseinrichtungen einrichtungsübergreifend zusammengestellt oder für ein Benchmark aufbereitet werden, bedarf es einer geeigneten Aggregationsmöglichkeit. Traditionell wird dafür auf die Einteilung in Disziplinen zurückgegriffen, auch wenn viele wissenschaftliche Aktivitäten unstreitig auf Kooperationen über Fächergrenzen hinweg beruhen. Der Wissenschaftsrat hat sich daher dafür ausgesprochen, eine etablierte Fächersystematik auch künftig als eines von mehreren alternativen Aggregationsverfahren zu verwenden und im Kerndatensatz entsprechend zu verankern: Die Fächersystematik des Statistischen Bundesamtes. Das Hochschulstatistikgesetz verpflichtet schon heute alle Hochschulen zu einer Berichterstattung nach Fächern. In der Finanz- und Personalstatistik müssen auch alle außeruniversitären Einrichtungen nach (gröberen) Fachzuordnungen berichten. Hinsichtlich der Aggregation nach Fächern wurden im Zuge der Berücksichtigung der Rückmeldungen daher vor allem die Vorschläge zur Zuordnung von Objekten (z. B. Drittmittelprojekte) und Personen zu Fächern vereinfacht.

Bei allen Verbesserungen bestärkt die Erfahrung aus der intensiven Auseinandersetzung mit Rückmeldungen zur Beta-Version den Wissenschaftsrat darin, den Kerndatensatz als lernendes und anpassungsfähiges System auszulegen. Gleichzeitig verweist er darauf, dass die Einführung eines Standards notwendigerweise mit gewissen Anpassungen verbunden ist. Aufwände in der Umstellungsphase werden sich im Wirkbetrieb aber rasch amortisieren.

IV.2 Dokumentation

Diese Empfehlungen beziehen sich auf eine Reihe von Dokumenten, die im Kontext des Projektes „Spezifikation Kerndatensatz Forschung“ erstellt wurden. Diese richten sich an unterschiedliche Zielgruppen:

_ Die Spezifikation des Kerndatensatz Forschung in Version 1.0 adressiert all jene, die eine Berichterstattung gemäß dem Kerndatensatz-Standard bereit-

stellen wollen. Für Datennutzer enthält sie Übersichten über die Aggregatdaten, die standardisiert wurden und im Rahmen der rechtlichen Vorgaben weiter gegeben werden können. Sie richtet sich aber auch an einschlägige Expertinnen und Experten, die z. B. ein Forschungsinformationssystem entlang den Vorschlägen des Kerndatensatzes aufbauen möchten.

Der Ergebnisbericht des Projektes „Spezifikation Kerndatensatz Forschung“ bietet eine Übersicht über alle Ergebnisse des Projektes und ergänzt somit die Spezifikation. Hier werden Prozesse und Hintergründe erläutert, die für ein Verständnis der Ergebnisse hilfreich sind. Der Bericht enthält auch ein Kapitel, das sich speziell dem Thema Datenschutz widmet. Ein weiteres Kapitel diskutiert Schnittmengen und Unterschiede verschiedener Definitionsvorschläge im Kerndatensatz und z. B. beim Statistischen Bundesamt (Harmonisierung der Definitionen und Klassifikationen sowie des Datenmodells mit bestehenden Standards). Ferner sind die Empfehlungspapiere der Projektgruppe „Fächerklassifikation und Thesauri“ enthalten – z. B. zur Revision der Fächersystematik des Statistischen Bundesamtes.

Darüber hinaus eignet sich zur Erschließung des Kerndatensatzes besonders die Online-Dokumentation: |³⁶ Im Zuge der Erarbeitung der Spezifikation wurden eine Vielzahl von Entscheidungen bezüglich möglicher Alternativen getroffen. Die Dokumentation der Argumente erschien angesichts der Erarbeitung eines neuen Standards als besonders wichtig für die Nachvollziehbarkeit und somit die Akzeptanz der Ergebnisse. Im Projekt lag so ein besonderes Augenmerk auf der transparenten Darstellung der Argumente für und gegen bestimmte Definitionsmöglichkeiten im Rahmen des Kerndatensatz Forschung. |³⁷ In den online verfügbaren Materialien lassen sich daher auch alle erwogenen aber nicht gewählten Alternativdefinitionen mit den jeweiligen Entscheidungsgründen nachvollziehen. Auch die Gründe für Entscheidungen des Beirates sind in der Online-Dokumentation hinterlegt, ebenso wie Ergebnisse aus der öffentlichen Beta-Phase.

Der Wissenschaftsrat begrüßt diese transparente Art der Dokumentation des Arbeitsprozesses im Projekt Kerndatensatz Forschung.

|³⁶ Übersicht über die verfügbare Online-Dokumentation des Kerndatensatz Forschung. <http://www.kerndatensatz-forschung.de/>. Auch das technische Datenmodell ist online hinterlegt. www.kerndatensatz-forschung.de/version1/technisches_datenmodell/

|³⁷ Vgl. Riechert, M.; Herwig, S.; Sirtes, D.: Towards using Design Rationale as an Acceptance Strategy in large-scale Information System Projects. 12. International Conference on Wirtschaftsinformatik, Osnabrück 2015.

B. Zu den Projektergebnissen

Mit dem Abschluss des Projektes „Spezifikation Kerndatensatz Forschung“ liegt eine Basis für die Implementierung des Kerndatensatz Forschung an deutschen Hochschulen und außeruniversitären Forschungseinrichtungen vor. Im Folgenden gibt der Wissenschaftsrat zunächst Hinweise zum Einsatz des Kerndatensatz Forschung und stellt anschließend die wesentlichen Inhalte der Version 1.0 der Spezifikation des Kerndatensatz Forschung dar. Diese Ausführungen richten sich vor allem an diejenigen, die mit organisatorischen oder technischen Fragen der Implementierung an ihren Einrichtungen befasst sind.

B.1 HINWEISE ZUM EINSATZ DES KERNDATENSATZ FORSCHUNG

Die Spezifikation des Kerndatensatz Forschung standardisiert Aggregatdaten, die bei Bedarf weitergegeben werden können. Einrichtungintern können solche Angaben entweder händisch erhoben oder anhand definierter Regeln aus Basisdaten rechnerisch aggregiert werden (z. B. Drittmiteleinahmen pro Institut). Einige Angaben werden nicht rechnerisch aggregiert, sondern nach spezifischen Regeln in Listenform zusammengestellt (z. B. Publikationsliste im Fach Geschichte). Um eine möglichst aufwandsarme und gleichwohl qualitativ anspruchsvolle Generierung dieser Daten zu gewährleisten, gibt der Wissenschaftsrat im Folgenden Hinweise für den Umgang mit dem Kerndatensatz. Er appelliert an alle Nutzerinnen und Nutzer dieses Standards, nötigenfalls auch andere Nutzerinnen und Nutzer auf sie aufmerksam zu machen und ihre Umsetzung zu befördern.

I.1 Prozessbegleitende Erfassung von Daten

Soweit möglich, sollten Basisdaten zu Forschungsaktivitäten generell im Rahmen standardisierter (administrativer) Prozesse erhoben werden. Bei vielen Daten ist dies schon heute der Fall. Angaben zu Forschungsprojekten können so beispielsweise direkt bei der Bearbeitung von Zuwendungsbescheiden oder mit der Eintragung in Projektdatenbanken erfasst werden. Auf diese Weise lassen sich Routinen etablieren, die dazu beitragen, die Qualität der erfassten Daten

zu sichern. Viele der Daten des Kerndatensatzes werden heute schon im Rahmen der internen Steuerungssysteme, z. B. im Rahmen der sog. leistungsorientierten Mittelvergabe (LOM), erhoben. Die Einrichtungen haben nicht zuletzt aus diesem Grund ein starkes Interesse daran, eine hohe Qualität dieser Daten sicherzustellen.

I.2 Kontinuierliche Erfassung von Daten für flexible Nutzung

Aggregatdaten im Kerndatensatz Forschung beziehen sich in der Regel auf ein Kalenderjahr (können sich aber bei Hochschulen analog zu den Vorgaben des Statistischen Bundesamtes auch auf Prüfungsjahre beziehen). Der Wissenschaftsrat weist darauf hin, dass sich bei kontinuierlicher Fortschreibung der Basisdaten gegenüber einer lediglich anlassbezogenen Erfassung zusätzliche Nutzungsmöglichkeiten ergeben. Z. B. lassen sich so für interne Zwecke auch Quartalsübersichten erzeugen oder Daten für unterschiedliche Zeiträume zusammenfassen. |³⁸

I.3 Qualitätssicherung

Für die Qualitätssicherung sollten Einrichtungen – auch in ihrem eigenen Interesse – regelhafte Prozesse etablieren und hierfür Ressourcen einplanen. Der Wissenschaftsrat weist ferner Dateneigner ebenso wie Datennutzer darauf hin, dass es im Sinne der Qualitätssicherung wichtig ist, vorgelegte Informationen auf Stimmigkeit zu prüfen. Sofern Inkonsistenzen auffallen, sollten diese den Dateneignern zurückgemeldet werden. Der Wissenschaftsrat empfiehlt darüber hinaus Datennutzern und Dateneignern, sich im Zuge der Einführungsphase gemeinsam in einen Austausch über die gegenseitigen Erwartungen an Daten einzubringen und somit die Qualitätssicherung zu befördern.

I.4 Datenschutz

Mit dem Kerndatensatz hat der Wissenschaftsrat ein Instrument zum Austausch standardisierter Daten zu Forschungsaktivitäten und deren Kontext empfohlen. Er hat dabei betont, dass der Umgang mit den Daten im Einklang mit bestehenden Anforderungen des Daten-, Geheimnis- und Informationsschutzes stehen muss. |³⁹ Es ist dabei nochmals hervorzuheben, dass der Kerndatensatz keine neuen Grundlagen schafft, um auf Informationen zuzugreifen. Der Bedeutung des Schutzes von Informationen, insbesondere des Datenschut-

|³⁸ Vgl. Spezifikation des Kerndatensatz Forschung, Version 1.0, in der Anlage, Kapitel 3.

|³⁹ Wissenschaftsrat: Empfehlungen zu einem Kerndatensatz Forschung (Drs. 2855-13), Berlin Januar 2013, S. 17.

zes, trägt ein eigenes Kapitel im Ergebnisbericht des Projektes Rechnung. |⁴⁰ Dort wird dokumentiert, dass die mit dem Datenschutz verbundenen Herausforderungen an vielen Stellen im Projekt thematisiert wurden. Ausgewählte Ergebnisse sollen hier wiedergegeben werden:

- _ Der Kerndatensatz erzeugt keine neuen Berichtspflichten. Es werden keine eigenständigen Rechtsgrundlagen für die Abfrage, Verarbeitung und Weitergabe von Daten geschaffen.
- _ Forschungseinrichtungen sind, wie bei anderen Datenanfragen auch, bei der Weitergabe und Übermittlung von Daten an die jeweils geltenden rechtlichen Vorgaben, insbesondere des Daten-, Informations- und Geheimnisschutzrechts, gebunden.
- _ Grundsätzlich gilt im Kerndatensatz wie auch generell, dass die Erhebung und Verarbeitung aller Daten im Einklang mit dem Datenschutz stehen muss. Personenbezogene Daten dürfen nur auf der Basis einer Einwilligung der bzw. des Betroffenen oder einer gesetzlichen Grundlage, bspw. des jeweiligen Landeshochschulgesetzes, erhoben, verarbeitet, gespeichert und übermittelt werden. Dies ist gängige Praxis.
- _ Die Spezifikation des Kerndatensatz Forschung legt lediglich das Format der Daten fest, die für die Weitergabe empfohlen werden.
- _ Dies bedeutet mit Blick auf die Weitergabe von Aggregatdaten bzw. von Informationen in Listenform, dass bei datenschutzrechtlich einschlägigen Informationen die Rückidentifizierung von Personen ausgeschlossen sein muss, sofern keine Einwilligung oder gesetzliche Grundlage vorliegt. |⁴¹ Weiterhin sind sonstige geheimnis- und informationsrechtliche Schutzrechte zu beachten. Hierfür tragen die übermittelnden Einrichtungen die Verantwortung.
- _ Der Wissenschaftsrat weist alle Datennutzer darauf hin, dass sie in der Verantwortung stehen, ihre Datenanfragen ausreichend zu legitimieren.
- _ Der Wissenschaftsrat regt an, dass die Forschungseinrichtungen die Einführung des Kerndatensatzes auch zum Anlass zu nehmen, ihre Selbstauskunftsfähigkeit durch Einführung eines integrierten Informationsmanagements zu verbessern und – wie bei der Datenverarbeitung in den Bereichen Personal, Finanzen und Lehre bereits üblich – zur Vorhaltung der Basisdaten geeignete Forschungsinformationssysteme einzusetzen. Zur Erfassung der Basisdaten gibt der Projektbericht viele Hinweise, betont aber auch, dass angesichts einer Vielzahl von Regelungen innerhalb der Bundesrepublik die datenschutz-

|⁴⁰ Vgl. Projekt „Spezifikation Kerndatensatz Forschung“. Dokumentation der Ergebnisse. 2015. http://www.kerndatensatz-forschung.de/version1/Ergebnisbericht_Projekt_KDSF.pdf.

|⁴¹ Bei Publikationslisten ist die Identifizierbarkeit allerdings gewollt.

rechtlichen Voraussetzungen für die Erfassung von Forschungsinformationen in den Ländern und im Bund derzeit unterschiedlich gestaltet sind. Die Rechtsgrundlagen für die Datenverarbeitung in den Einrichtungen können daher erheblich variieren. |⁴²

1.5 Datenweitergabe

Neben einer vereinfachten internen Nutzung ist es das wesentliche Ziel des Kerndatensatz Forschung, durch Standardisierung die Weitergabe von Informationen zu Forschungsaktivitäten an externe Nutzerinnen und Nutzer zu erleichtern. Daher sollte jede Datenlieferung deutlich machen, auf welcher Version des Kerndatensatz Forschung sie basiert.

Vor allem für die Weitergabe von Daten an externe Datennutzer bedarf es eines gemeinsamen Verständnisses hinsichtlich der Bedingungen der Datenweitergabe auf beiden Seiten des Austausches. Sofern keine Verpflichtung zur Weitergabe besteht, empfiehlt der Wissenschaftsrat, einen Datenweitergabevertrag abzuschließen, der mindestens folgende Punkte umfasst:

- _ Datenanfragen erklären immer den Zweck der Datenverwendung. Die Verwendung der übermittelten Daten erfolgt zweckgebunden. Eine Weiterverwendung und Weitervermittlung unterliegt gleichfalls rechtlichen Bindungen.
- _ Dies ist nicht zuletzt notwendig, um die angefragten Einrichtungen bei ausnahmsweise vorliegenden personenbeziehbaren Daten (z. B. in Listen) oder sonstigen geheimnisgeschützten Informationen, in die Lage zu versetzen, zu beurteilen, ob sie zur Weitergabe der Daten befugt sind.
- _ Datenanfragen sollten genau spezifizieren, welche Kerndaten für welche Einheiten (z. B. Institute oder Fächer) und welchen Zeitraum angefordert werden.
- _ Die herausgebende Einrichtung überprüft, ob die Daten herausgegeben werden können. Sofern Anfragen abschlägig beurteilt werden, sollte dies der oder dem Anfragenden gegenüber nachvollziehbar begründet werden.
- _ Ggf. wird eine Löschfrist vereinbart.

Der Wissenschaftsrat befürwortet Offenheit im Umgang mit Kerndaten und Daten aus der Schale des Kerndatensatzes zu Forschungsaktivitäten. Er ist der Überzeugung, dass das Wissenschaftssystem von Transparenz profitiert. Gleichzeitig betont er erneut, dass durch den Kerndatensatz keine neuen Berichtspflichten entstehen. Jenseits ihrer verbindlichen Berichts- und Aus-

|⁴² Vgl. Projekt „Spezifikation Kerndatensatz Forschung“. Dokumentation der Ergebnisse. 2015. http://www.kerndatensatz-forschung.de/version1/Ergebnisbericht_Projekt_KDSF.pdf.

kunftspflichten entscheiden Einrichtungen im Rahmen des rechtlich und vertraglich Zulässigen selbst über die Weitergabe ihrer Daten.

1.6 ID-Management

Im Kerndatensatz leiten sich die zur Weitergabe geeigneten Aggregatdaten aus Basisdaten ab. Diese sollten idealerweise in einem die verschiedenen einrichtungsinternen Datenbanken (Personaldatenbank, Haushaltsdatenbank, etc.) verbindenden Informationssystem organisiert werden. Für die eindeutige Erfassung, das Zusammenführen verschiedener einrichtungsinterner Quellen, aber auch für die Weitergabe objektbezogener Daten ist das Management sog. *Identifizier* von zentraler Bedeutung. Buchungsnummern in Haushaltsdatenbanken und Personen-IDs in Personalsystemen sind gängige Beispiele für solche internen IDs.

Im Zusammenhang mit dem Kerndatensatz kommt vor allem solchen IDs eine wichtige Rolle zu, die den einrichtungsübergreifenden Austausch von Informationen unterstützen können. Im Anhang (S. 60) findet sich eine exemplarische Liste, die auch einige weniger etablierte, für den Kontext von Forschungsinformationssystemen aber potentiell sehr hilfreiche, ‚externe‘ IDs vorstellt. Es bietet sich an, beim Aufbau von Forschungsinformationssystemen solche IDs einzusetzen.^{|43} Aus Sicht des Wissenschaftsrats spricht, da es sich um einen offenen Standard handelt, vieles für den Einsatz von ORCID (*Open Researcher and Contributor ID*)^{|44} für die Zuordnung von Personen sowie von DOIs (*Digital-Object-Identifizier*) zur eindeutigen Identifikation von Forschungsausput.

1.7 Interpretation der Daten

Für sich genommen bieten Kerndaten keine hinreichende Basis zur Interpretation und damit zur Bewertung von Forschung. Der Wissenschaftsrat hat bereits mehrfach darauf verwiesen, dass die Bewertung wissenschaftlicher Leistung nur durch *informed Peer review*, die Beurteilung durch *Peers* auf der Basis belastbarer, also qualitätsgesicherter, Daten erfolgen kann.^{|45} Hintergrund ist, dass Daten keine Indikatoren sind und immer in ihrem Kontext betrachtet werden müssen. Deutlich wird dies beispielsweise bei Informationen zu Innovationsbeiträgen bzw. zum Transfer. Ausgründungen und Patente sind nur zwei von vielen Möglichkeiten, zum Technologietransfer beizutragen, deren

^{|43} Vgl. auch Spezifikation des Kerndatensatz Forschung, Version 1.0, in der Anlage.

^{|44} Eine gleichlautende Empfehlung gaben jüngst Wilson, et al. für Großbritannien. Vgl. <http://doi.org/10.13140/RG.2.1.4929.1363>. Vgl. auch: Science Europe Working Group on Research Policy and Programme Evaluation (2015). Survey on Data Collection and Use in Research Funding and Performing Organisations – Preliminary Report.

^{|45} Vgl. Wissenschaftsrat: Empfehlungen zur Bewertung und Steuerung von Forschungsleistung (Drs. 1656-11), Halle November 2011.

Nutzung stark vom Institutionentyp sowie der Fachdisziplin abhängt. Der Wissenschaftsrat weist somit ausdrücklich darauf hin, dass der hier definierte Kern an Daten noch nicht für alle Analysezwecke in gleichem Maße geeignet ist.

I.8 Datennutzerinnen und Datennutzer

Der Rückgriff auf standardisierte Daten liegt im Interesse der Nutzerinnen und Nutzer innerhalb wie auch außerhalb einer Einrichtung. Sie erhalten auf diese Weise besser qualitätsgesicherte Informationen. Der Wissenschaftsrat fordert die Nutzerinnen und Nutzer daher dazu auf, die Implementierung des Kerndatensatz Forschung zu befördern und eigene Abfragen der im Kerndatensatz enthaltenen Angaben nach den dort vorgenommenen Definitionen vorzunehmen. Der Wissenschaftsrat selbst beabsichtigt, seine Datenabfragen im Laufe des Jahres 2016, jeweils in Absprache mit seinen Auftraggebern, an den Kerndatensatz Forschung anzupassen.

B.II DIE SPEZIFIKATION

Wichtige Ergebnisse des Spezifikationsprozesses (vgl. Anlage, S. 63) sind im Folgenden knapp zusammen gefasst: sie können und sollen eine Lektüre der Spezifikation aber nicht ersetzen.

II.1 Aggregation

Die Empfehlungen des Wissenschaftsrates zu einem Kerndatensatz Forschung haben dem „Fach“ als Aggregationseinheit eine wichtige Rolle zugebilligt. |⁴⁶ Nach Auffassung des Wissenschaftsrates soll durch die Zuordnung von Forschungsaktivitäten zu Fächern eine nach Disziplinen differenzierte Betrachtung ermöglicht werden, sei es innerhalb von fachlich breit aufgestellten Organisationen, sei es organisationsübergreifend. Dabei sind bei der Interpretation der Angaben immer auch Fachspezifika zu betrachten. |⁴⁷

Eine Einrichtung sollte ihre Kerndaten gemäß Spezifikation nach Organisationseinheiten und ggf. nach Fächern aggregiert weitergeben können. Die Zuordnung nach Fächern soll dabei so detailliert wie möglich entlang der Fächer-systematik des Statistischen Bundesamtes erfolgen (s. auch A.IV.1.c, S. 23). |⁴⁸

|⁴⁶ Vgl. Wissenschaftsrat: Empfehlungen zu einem Kerndatensatz Forschung (Drs. 2855-13), Berlin Januar 2013.

|⁴⁷ Vgl. Wissenschaftsrat: Empfehlungen zur Bewertung und Steuerung von Forschungsleistung (Drs. 1656-11), Halle November 2011.

|⁴⁸ Wünschenswert ist damit eine Zuordnung zu Fachgebieten („Viersteller“) gemäß der Systematiken des Statistischen Bundesamtes, minimal erforderlichlich eine Zuordnung zu einer Fächergruppe („Zweisteller“).

Der Wissenschaftsrat empfiehlt für die Einführungsphase des Kerndatensatz Forschung eine enge Orientierung an der Erhebungspraxis der amtlichen Statistik. Dort erfolgt die fachliche Zuordnung von Daten primär über das Fach des (universitären) Instituts oder entsprechender kleinerer (universitärer) Organisationseinheiten.

Vor allem bei den häufig eher nach interdisziplinären Forschungsfeldern organisierten außeruniversitären Forschungseinrichtungen ist eine solche Zuordnung von Organisationseinheiten zu Fächern derzeit nicht realisiert. Sie wäre nach Angabe der Dach- und Trägerorganisationen inhaltlich nicht sinnvoll. In solchen Fällen beschränkt sich der Austausch auf Daten, die nach Organisationseinheiten aggregiert sind (z. B. Instituten, Zentren, Abteilungen). |⁴⁹ Um den Vorschlag der Projektgruppe „Fächerklassifikation und Thesauri“ für eine adäquate Zuordnung zu interdisziplinären Forschungsfeldern umzusetzen, muss zunächst eine Liste von Forschungsfeldern erstellt und abgestimmt werden.

II.2 Inhalte

Dieses Kapitel gibt eine knappe Zusammenfassung der Inhalte der Kerndatensatz-Spezifikation. In der vorliegenden ersten Version besteht der Kerndatensatz aus Aggregatdaten in folgenden *sechs* Bereichen: Beschäftigte, Nachwuchsförderung, Drittmittel und Finanzen, Patente und Ausgründungen, Publikationen sowie Forschungsinfrastrukturen.

Diese Bereiche beinhalten Anpassungen gegenüber den ursprünglichen Empfehlungen des Wissenschaftsrates, die im Folgenden erläutert werden sollen. Die Reihenfolge ihrer Darstellung folgt der Spezifikation. |⁵⁰

II.2.a Beschäftigte

Das Personal stellt eine zentrale Ressource für die Forschung dar. Verschiedene Gruppen von Beschäftigten sind Träger der wissenschaftlichen Aktivität von Forschungseinrichtungen. Der Kerndatensatz erfasst nur Beschäftigte – also solche Angehörigen, die einen Arbeitsvertrag mit der berichtenden Einrichtung haben. Andere Angehörige der Einrichtung – v. a. Stipendiatinnen und Stipendiaten oder auch Honorarprofessorinnen und Honorarprofessoren – werden aus pragmatischen Gründen zunächst nicht berücksichtigt. Während sie zweifellos wichtige Beiträge leisten können und an einigen Einrichtungen auch in großer Zahl arbeiten, würde ihre zuverlässige Zählung doch in weiten

|⁴⁹ Vgl. Spezifikation des Kerndatensatz Forschung, Version 1.0, in der Anlage, Kapitel 3.2.3.

|⁵⁰ Vgl. Spezifikation des Kerndatensatz Forschung, Version 1.0, in der Anlage.

Teilen des Wissenschaftssystems völlig neue Erhebungsprozesse erfordern. Für die erste Version des Kerndatensatzes wurde daher darauf verzichtet.

In der Forschung spielen Nachwuchsgruppenleitungen eine zunehmend wichtige Rolle. Daneben gibt es außerdem eine Reihe von sich entwickelnden Differenzierungen im Bereich der wissenschaftlichen Mitarbeiterinnen und Mitarbeiter, die eigenständige Karrierepfade jenseits der Professur eröffnen. |⁵¹ So wichtig diese Aspekte sind, konnte bei der Erarbeitung der Spezifikation keine Definition gefunden werden, die eine zuverlässige Erfassung dieser Tätigkeiten sowohl im Kontext z. B. von FhG, HGF, MPG und WGL als auch an den Hochschulen zuließe. Dieser Punkt bleibt somit ein wichtiges Desiderat für zukünftige Weiterentwicklungen der Spezifikation des Kerndatensatzes.

Die folgende Tabelle bietet eine Übersicht über die Inhalte, die für den Kerndatensatz bzw. die Schale des Kerndatensatzes vorgeschlagen werden.

Tabelle 2: Inhalte - Beschäftigte

Kerndatensatz [Schale des Kerndatensatzes]	Ausdifferenzierungen	Aggregierbar nach
Gesamtpersonal		Gesamteinrichtung
Anzahl Personen	Geschlecht, Personalkategorie, Finanzierungsform, Qualifikation, Befristung, Staatsangehörigkeit, [Altersgruppe, Tätigkeitsart]	Org.-einheit, Fach
Vollzeitäquivalente	Geschlecht, Personalkategorie, Finanzierungsform, Qualifikation, Befristung, Staatsangehörigkeit, [Altersgruppe, Tätigkeitsart]	Org.-einheit, Fach

Quelle: Spezifikation des Kerndatensatz Forschung (vgl. Anlage). Elemente des Kerndatensatzes sind fett gedruckt, Elemente der Schale des Kerndatensatzes normal gesetzt. Diese Tabelle enthält aus Gründen der Übersichtlichkeit Vereinfachungen, ausführliche Angaben finden sich in der Spezifikation.

II.2.b Nachwuchsförderung

Meist beschränkt sich der Kerndatensatz auf die Standardisierung von bereits etablierten Größen. Die Erfassung von Doktorandinnen und Doktoranden stellt eine wichtige Ausnahme von diesem Grundprinzip dar. In der Praxis steht sie noch am Anfang. Es besteht jedoch ein breiter Konsens in der Wissenschaft wie auch in der Wissenschaftspolitik, dass es unverantwortlich wäre, die Informationsdefizite in diesem wichtigen Bereich weiter bestehen zu lassen. Daher bestehen auch Absichten, Daten zu Doktorandinnen und Doktoranden in

|⁵¹ Vgl. Wissenschaftsrat: Empfehlungen zu Karrierezielen und -wegen an Universitäten (Drs. 4009-14), Dresden Juli 2014.

der amtlichen Statistik zu verankern. |⁵² Der Kerndatensatz soll helfen, die derzeit an vielen Orten laufenden Bemühungen zu fokussieren.

Die folgende Tabelle bietet eine Übersicht über die Inhalte, die für den Kerndatensatz bzw. die Schale des Kerndatensatzes vorgeschlagen werden.

Tabelle 3: Inhalte - Nachwuchsförderung

Kerndatensatz [Schale des Kerndatensatzes]	Ausdifferenzierungen	Aggregierbar nach
Anzahl Doktorandinnen und Doktoranden an titelverleihenden Einrichtungen	Geschlecht, Staatsangehörigkeit, Start der Promotion, Erstbetreuer an Einrichtung, Strukturiertes Promotionsprogramm, Kooperation, [Land der Promotionsberechtigung]	Org.-einheit, Fach
Anzahl Doktorandinnen und Doktoranden an nicht-titelverleihenden Einrichtungen	Geschlecht, Staatsangehörigkeit, Start der Promotion, Erstbetreuer an Einrichtung, Strukturiertes Promotionsprogramm, Kooperation, [Land der Promotionsberechtigung]	Org.-einheit, Fach
Anzahl abgeschlossener Qualifizierungsverfahren	Geschlecht, Staatsangehörigkeit, Promotion, Habilitation ⁵³, Erstbetreuer an Einrichtung, Zeitpunkt des Abschlusses, [Altersgruppe bei Abschluss]	Org.-einheit, Fach
Strukturiertes Promotionsprogramm (Listenform)	Titel, Zahl Teilnehmer, Geschlecht, Zahl Teilnehmer, Staatsbürgerschaft, Beteiligte Institutionen, [Finanzierung, Sprecher]	Org.-einheit, Fach, Strukt. Promotionsprogramm

Quelle: Spezifikation des Kerndatensatz Forschung (vgl. Anlage). Elemente des Kerndatensatzes sind fett gedruckt, Elemente der Schale des Kerndatensatzes normal gesetzt. Diese Tabelle enthält aus Gründen der Übersichtlichkeit Vereinfachungen, ausführliche Angaben finden sich in der Spezifikation.

II.2.c Finanzen und Drittmittel

Finanzmittel stellen ebenfalls eine elementare Größe für die Ermöglichung von Forschung dar. Der Wissenschaftsrat empfiehlt, im Kerndatensatz Drittmittel-einnahmen oder Drittmittelerträge zu erfassen (Vgl. zur Erläuterung A.IV.1.g, S. 27). Der Kerndatensatz standardisiert darüber hinaus auch Angaben zu Projekten, die für die qualitative Darstellung von Forschung eine wichtige Rolle einnehmen. |⁵⁴

Bereits in den Rückmeldungen der Piloteinrichtungen war deutlich geworden, dass die Beschränkung auf Drittmittel ein partielles Bild vom Ressourceneinsatz für die Forschungsaktivitäten einer Einrichtung zur Folge hätte, was insbesondere mit Blick auf die Verwendung von Mitteln der Grundfinanzierung

|⁵² Noch offene Fragen, z. B. zur Dauer des Doktorandenstatus bei letztlich nicht abgeschlossenen Promotionen, sind im Zuge dieses Prozesses noch zu klären. Vgl. auch Bundesrat: Entwurf eines Gesetzes zur Änderung des Hochschulstatistikgesetzes (Drs. 394/15 v. 4. September 2015).

|⁵³ Der Wissenschaftsrat weist darauf hin, dass mit der Etablierung neuer Karrierewege an Universitäten und Fachhochschulen auch für diese belastbare Daten definiert werden müssen. Vgl. Wissenschaftsrat: Empfehlungen zu Karrierezielen und -wegen an Universitäten (Drs. 4009-14), Dresden Juli 2014.

|⁵⁴ Es ist möglich, die Namen von Drittmittelgebern nicht offen zu legen, sofern dem rechtliche oder vertragliche Vereinbarungen entgegenstehen.

zu Verzerrungen führen würde. Daher erschien eine Ergänzung um Angaben zum Gesamtbudget sinnvoll. Dabei ist dem Wissenschaftsrat bewusst, dass Zahlen zum Gesamtbudget insbesondere bei Hochschulen schwer vergleichbar sind. Sie werden allerdings auch für das Statistische Bundesamt erhoben und stellen unzweifelhaft eine wichtige Bezugsgröße dar. Der Wissenschaftsrat verbindet diese Diagnose mit der Bitte an die statistischen Ämter, Regeln zur Verbesserung der Belastbarkeit und Vergleichbarkeit dieser wichtigen Größe zu entwickeln. Einstweilen weist er darauf hin, dass die Daten zum Gesamtbudget aus methodischen Gründen keinen Vergleich der an verschiedenen Einrichtungen für Forschung verfügbaren Ressourcen ermöglichen, sondern lediglich eine grobe Orientierung zulassen.

Die folgende Tabelle bietet eine Übersicht über die Inhalte, die für den Kerndatensatz bzw. die Schale des Kerndatensatzes vorgeschlagen werden.

Tabelle 4: Inhalte - Finanzen und Drittmittel

Kerndatensatz [Schale des Kerndatensatzes]	Ausdifferenzierungen	Aggregierbar nach
Drittmiteleinahmen / Drittmittelerträge	Jahr, Drittmittelgeber	Org.-einheit, Fach, Drittmittelprojekt, [Koordinator]
Drittmittelprojekt (Liste- form)	Titel des Projekts, Titel des übergeordneten Projektes, Koordinationsrolle (Einrichtung), Name der Sprechereinrichtung / des Konsortialführers, Projektbeginn, Projektende, Drittmittelgeber, Förderkennzeichen, [Bewilligungssumme]	Org.-einheit, Fach, Drittmittelprojekt
[Sonstige Einnahmen]		Org.-einheit, Gesamteinrichtung
[Gesamtbudget]		Org.-einheit, Gesamteinrichtung

Quelle: Spezifikation des Kerndatensatz Forschung (vgl. Anlage). Elemente des Kerndatensatzes sind fett gedruckt, Elemente der Schale des Kerndatensatzes normal gesetzt. Diese Tabelle enthält aus Gründen der Übersichtlichkeit Vereinfachungen, ausführliche Angaben finden sich in der Spezifikation.

II.2.d Patente und Ausgründungen

Zu den in diesem Bereich spezifizierten Daten gehören neben Angaben zu prioritätsbegründeten Patenten auch solche zur Anzahl der Patentfamilien, zu Erträgen aus Schutzrechten sowie zu Ausgründungen. Der Bereich deckt damit solche Daten ab, die sich – gemäß der Zielsetzung des Kerndatensatz Forschung – klar definieren lassen und die häufig abgefragt werden. In den Rückmeldungen zur Beta-Version wurde vielfach geäußert, dass für die Dokumentation von Transfer eine größere Zahl an Angaben erforderlich wäre, auch um der steigenden Bedeutung von Innovationsbeiträgen im Wissenschaftssystem insgesamt gerecht zu werden. Für andere Transferleistungen der Wissenschaften (z. B. Wissenschaftskommunikation, Politikberatung, Beiträge zu Normierungsverfahren) kann jedoch zurzeit kein vergleichbar klarer Definitionsstandard vorgeschlagen werden. Dies ist bei der Interpretation dieser Daten

zu beachten. Sofern sich ein klarer Trend für weitere Wege zur standardisierten Dokumentation von Transferaktivitäten abzeichnet, könnte dies in einer Weiterentwicklung des Kerndatensatzes berücksichtigt werden. Entsprechend wurde vorerst auch darauf verzichtet, diesen Bereich mit „Transfer“ zu überschreiben.

Die folgende Tabelle bietet eine Übersicht über die Inhalte, die für den Kerndatensatz bzw. die Schale des Kerndatensatzes vorgeschlagen werden.

Tabelle 5: Inhalte - Patente und Ausgründungen

Kerndatensatz (Schale des Kerndatensatzes)	Ausdifferenzierungen	Aggregierbar nach
Patent (Listenform)	Titel des prioritätsbegründenden Patents, Datum der prioritätsbegründenden Erstanmeldung, Veröffentlichungsnummer, [Erfinder]	Fach, Org.-einheit, Patent, Gesamteinrichtung
Anzahl Patentfamilien		Fach, Org.-einheit, Patent, Gesamteinrichtung
Anzahl prioritätsbegründender Patentanmeldungen		Fach, Org.-einheit, Patent, Gesamteinrichtung
Anzahl der erteilten Patente		Fach, Org.-einheit, Patent, Gesamteinrichtung
Anzahl Ausgründungen		Fach, Org.-einheit, Patent, Gesamteinrichtung
[Erträge aus Schutzrechten]		Fach, Org.-einheit, Patent, Gesamteinrichtung

Quelle: Spezifikation des Kerndatensatz Forschung (vgl. Anlage). Elemente des Kerndatensatzes sind fett gedruckt, Elemente der Schale des Kerndatensatzes normal gesetzt. Diese Tabelle enthält aus Gründen der Übersichtlichkeit Vereinfachungen, ausführliche Angaben finden sich in der Spezifikation.

II.2.e Publikationen

Abweichend von den sonstigen Bereichen macht die Spezifikation im Bereich Publikationen keine Vorschläge, welche Angaben zu Veröffentlichungen besonders wichtig sind und daher in aggregierten Listen zur Weitergabe empfohlen werden. Vielmehr beschreibt der Vorschlag, welche Metadaten zu Publikationen verschiedener Typen vorgehalten werden sollen.

Trotz der vergleichsweise großen Zahl an Angaben wird für jede einzelne Publikation nur ein Ausschnitt davon und davon wiederum nur ein kleiner Teil manuell zu erfassen sein; viele Angaben können aus Publikationsdatenbanken importiert werden. Der Wissenschaftsrat erachtet es in jedem Fall als zentral, dass Einrichtungen Bibliographien führen und somit über die bei ihnen entstandenen Publikationen auskunftsfähig sind. Damit kann eine Basis geschaffen werden, durch einrichtungsinterne Repositorien die Fehlstellen in den kommerziellen Datenbanken besonders in den Bereichen Ingenieur-, Sozial-

und Geisteswissenschaften auszugleichen und die wissenschaftlichen Einrichtungen über alle Disziplinen hinweg auskunftsfähig zu machen.

Die folgende Tabelle bietet eine Übersicht über die Inhalte, die für den Kerndatensatz bzw. die Schale des Kerndatensatzes vorgeschlagen werden.

Tabelle 6: Inhalte - Publikationen

Kerndatensatz [Schale des Kerndatensatzes]	Ausdifferenzierungen	Aggregierbar nach
Publikation (Listenform)	Schöpfer, Titel des Werks, Veröffentlichungsjahr, Verlag, Quelle, Identifier, Format, Sprachcode, Zugangsrechte, Peer-Reviewed, Qualifikationsschrift, Dokumenttyp, Publikationstyp , [Förderer, Ressource, Förderkennzeichen]	Fach, Org.-einheit, Publikation, Publikationstyp, Dokumenttyp, Schöpfer, Peer-reviewed, Veröffentlichungsjahr

Quelle: Spezifikation des Kerndatensatz Forschung (vgl. Anlage). Elemente des Kerndatensatzes sind fett gedruckt, Elemente der Schale des Kerndatensatzes normal gesetzt. Diese Tabelle enthält aus Gründen der Übersichtlichkeit Vereinfachungen, ausführliche Angaben finden sich in der Spezifikation.

II.2.f Forschungsinfrastrukturen

Eine wichtige Ergänzung erfährt der Kerndatensatz durch die Aufnahme von Forschungsinfrastrukturen. Der Wissenschaftsrat hat die Bereitstellung und Pflege von Forschungsinfrastrukturen als eine zentrale Leistungsdimension von Wissenschaft benannt, |⁵⁵ die, wie auch an dem Interesse an einer Nationalen Roadmap für Forschungsinfrastrukturen sichtbar wird, in allen Bereichen des Wissenschaftssystems an Bedeutung zunimmt. |⁵⁶ Der Vorschlag, der die Entwicklungen in diesem Bereich in den letzten Jahren aufnimmt, bietet eine Möglichkeit (überregional bedeutsame) außeruniversitäre und universitäre Forschungsinfrastrukturen vom Forschungsschiff bis hin zur verteilten Datensammlung einheitlich zu dokumentieren. Der Beitrag, den unterschiedliche Einrichtungen somit für ihre, vor allem aber auch für die Forschung anderer erbringen, kann somit sichtbarer gemacht werden. Hervorzuheben ist, dass die vorgeschlagene Definition nicht die Kosten einer Infrastruktur zum Kriterium für ihre Aufnahme in die Schale des Kerndatensatzes macht, sondern auf ihre Bedeutung für ein Wissenschaftsgebiet abstellt. |⁵⁷ Sie ist damit geeignet, auch solche Ressourcen und Dienstleistungen abzubilden, die, wie z. B. Soziale Inf-

|⁵⁵ Vgl. Wissenschaftsrat: Perspektiven des deutschen Wissenschaftssystems (Drs. 3228-13), Braunschweig Juli 2013, S. 25.

|⁵⁶ Vgl. Wissenschaftsrat: Bericht zur wissenschaftsgeleiteten Bewertung umfangreicher Forschungsinstrukturvorhaben für die Nationale Roadmap (Pilotphase) (Drs. 2841-13), Köln April 2013.

|⁵⁷ Definition: Forschungsinfrastrukturen im Sinne des Kerndatensatzes sind umfangreiche / aufwendige Instrumente, Ressourcen oder Serviceeinrichtungen für die Forschung in allen Wissenschaftsgebieten, die sich durch eine mindestens überregionale Bedeutung für das jeweilige Wissenschaftsgebiet sowie durch eine mittel- bis langfristige Lebensdauer (mehr als 5 Jahre) auszeichnen und für eine externe Nutzung zur Verfügung stehen für die Zugangs- bzw. Nutzungsregelungen etabliert sind.

rastrukturen |⁵⁸, im Vergleich zu naturwissenschaftlichen Großgeräten deutlich geringere Anschaffungs- und Betriebskosten verursachen. |⁵⁹

Die folgende Tabelle bietet eine Übersicht über die Inhalte, die für den Kerndatensatz bzw. die Schale des Kerndatensatzes vorgeschlagen werden.

Tabelle 7: Inhalte - Forschungsinfrastrukturen

Kerndatensatz [Schale des Kerndatensatzes]	Ausdifferenzierungen	Aggregierbar nach
Forschungsinfrastruktur [Listenform]	[Bezeichnung, Beschreibung, Typ, Art, Art des Zugangs, Nutzung, Betreiber, Koordination, Wissenschaftliches Betriebspersonal, Publikationen]	Infrastruktur

Quelle: Spezifikation des Kerndatensatz Forschung (vgl. Anlage). Elemente des Kerndatensatzes sind fett gedruckt, Elemente der Schale des Kerndatensatzes normal gesetzt. Diese Tabelle enthält aus Gründen der Übersichtlichkeit Vereinfachungen, ausführliche Angaben finden sich in der Spezifikation.

Im Gegensatz zu anderen Bereichen des Kerndatensatzes existieren bislang wenige Erfahrungen mit der systematischen Beschreibung von Forschungsinfrastrukturen. |⁶⁰ Da die Angaben zu diesem Bereich Teil der Schale des Kerndatensatzes sind, erlauben sie es, mit diesem Bereich Erfahrungen zu sammeln und ggf. in einer späteren Version Anpassungen vorzunehmen. Angesichts einer eher geringen Zahl überregionaler Infrastrukturen ist ihre Erfassung zudem ohne nennenswerten zusätzlichen Aufwand möglich.

|⁵⁸ Soziale Infrastrukturen verstehen sich als „Begegnungsräume des diskursiven Austauschs von aktuellen und der Entwicklung von neuen Forschungsfragen“. Vgl. Wissenschaftsrat: Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften (Drs. 10465-11), Berlin Januar 2011, S. 21.

|⁵⁹ Vgl. Wissenschaftsrat: Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020 (Drs. 2359-12), Berlin Juli 2012.

|⁶⁰ Auf europäischer Ebene wird eine Datenbank großer Forschungsinfrastrukturen, die seit 2010 unter dem Titel „Mapping of the European Research Infrastructure Landscape“ (MERIL, <http://portal.meril.eu> (v. 14.01.2016)) firmiert, seit 2013 von der *European Science Foundation* fortgeführt. Die DFG hat 2014 ein breiter angelegtes Verzeichnis von Infrastrukturen aufgebaut, in das nach festgelegten Regeln weitere Infrastrukturen aufgenommen werden können, vgl. <http://resources.dfg.de> (v. 14.01.2016).

C. Einführung und weitere Entwicklung des Kerndatensatz Forschung

Mit dem Abschluss des Projektes liegt eine erste Spezifikation des Kerndatensatz Forschung vor. Damit ist die Grundlage für die Implementierung an den wissenschaftlichen Einrichtungen und eine anschließende Weiterentwicklung des Kerndatensatz Forschung hergestellt. Der Wissenschaftsrat erwartet, dass die Berücksichtigung des Standards durch berichtsanzfordernde Stellen bei den berichtenden Einrichtungen zu Zeit- und Aufwandsersparnis führt. Ebenso rechnet er mit einer Steigerung der Qualität und Aussagekraft der Daten auch für die interne Steuerung, nicht zuletzt weil die Spezifikation es auch möglich macht, einrichtungsübergreifende Benchmarks zu erstellen.

Während der Einführung des Kerndatensatz Forschung wird es aus Sicht des Wissenschaftsrates vor allem darauf ankommen, Erfahrungen im Umgang mit diesem auf einem Standard basierenden System zu sammeln. Diese Erfahrungen müssen anderen Akteuren weitergegeben werden, um ein möglichst schnelles Lernen sicherzustellen. Mittel- bis langfristig muss der Kerndatensatz beweisen, dass er sich an dynamische Entwicklungen im Wissenschaftssystem anpassen kann. Dies ist Bestandteil seiner Weiterentwicklung, bei der darauf zu achten ist, dass die Bedarfe und Bedürfnisse sowohl von Dateneignern als auch von Datennutzern berücksichtigt werden.

Im Folgenden legt der Wissenschaftsrat Empfehlungen dazu vor, wie sowohl die Einführung als auch die spätere Anpassung an das dynamische Wissenschaftssystem gestaltet werden können.

C.1 EINFÜHRUNG IM DEUTSCHEN WISSENSCHAFTSSYSTEM

Der erreichte Stand des Kerndatensatz Forschung nach der fast zweijährigen Entwicklungsphase rechtfertigt es, mit der Implementierung des Kerndatensatz Forschung an den wissenschaftlichen Einrichtungen zu beginnen. Die vorliegenden Empfehlungen und die beiliegende Version 1.0 der Spezifikation

können die Grundlage dafür bilden. Datennutzer sind aufgefordert, ihre Anfragen wo immer möglich am Kerndatensatz auszurichten.

Obwohl der Kerndatensatz vor allem auf der Verbindung administrativ generierter Daten aufbaut, wird die Implementierung an den einzelnen Einrichtungen mit unterschiedlichem Vorbereitungsaufwand verbunden sein und somit unterschiedlich viel Zeit und Ressourcen in Anspruch nehmen. Unter Einführung wird hier der Zeitraum verstanden, in dem die Implementierung des Kerndatensatzes an Forschungseinrichtungen begleitet und unterstützt wird. Diese Einführungsphase sollte nach ca. drei Jahren abgeschlossen sein. Dies bedeutet voraussichtlich nicht, dass zu diesem Zeitpunkt alle Forschungseinrichtungen entsprechend der Spezifikation des Kerndatensatzes berichten können werden. Gleichwohl sollte die angestrebte Vereinfachung der Berichterstattung durch eine standardisierte Weitergabe beginnen sichtbar zu werden.

Obschon während der Projektphase zur Erarbeitung der vorliegenden Spezifikation vier Piloteinrichtungen die Vorschläge auf ihre Anwendbarkeit überprüft haben, |⁶¹ konnte der Standard vor der Entscheidung über seine abschließende Ausgestaltung noch an keiner Einrichtung implementiert werden. Aufgrund der Vielfalt im deutschen Wissenschaftssystem, die auch in den zahlreichen Rückmeldungen zur Beta-Version sichtbar wird, spricht sich der Wissenschaftsrat daher nachdrücklich für folgende Maßnahmen aus, die die Einführung unterstützen.

1.1 Aufbau eines Kompetenznetzwerkes

Bei einer flächendeckenden Einführung des Kerndatensatzes sollten die dezentral gewonnenen Erkenntnisse möglichst schnell möglichst vielen Anwenderinnen und Anwendern zugänglich gemacht werden. So können Lösungen für Probleme in der Umsetzung zeitnah ausgetauscht und „gelernt“ werden. Daher gilt es, die Vernetzung der Expertinnen und Experten an den Einrichtungen, die den Kerndatensatz nutzen oder nutzen wollen, zu unterstützen. Hierzu sollten Informationen und Austauschplattformen auf Webseiten sowie Workshops und Konferenzen angeboten und bestehende Netzwerke eingebunden werden.

Der Wissenschaftsrat appelliert an die Forschungsreferentinnen und Forschungsreferenten, die Bibliothekarinnen und Bibliothekare, die Rechen- und Medienzentren sowie die Mitarbeiterinnen und Mitarbeiter in den Controlling-

| ⁶¹ Die Piloteinrichtungen umfassen die Westfälische Wilhelms-Universität Münster (WWU), die Technische Universität Dresden (TUD), das Helmholtz-Zentrum für Umweltforschung Leipzig (UFZ) sowie zu Projektbeginn die Technische Hochschule Mittelhessen (THM).

und anderen administrativen Abteilungen der wissenschaftlichen Einrichtungen: Gegenseitiger Austausch – auch über Bundesländer hinweg – hilft, das gemeinsame Lernen stark zu beschleunigen und damit auch, eigene Herausforderungen besser zu bewältigen. Den Instituts- und Hochschulleitungen empfiehlt der Wissenschaftsrat, die Vernetzung ihrer Mitarbeiterinnen und Mitarbeiter als Gewinn für die eigene Einrichtung anzuerkennen und gezielt zu unterstützen.

Alle einschlägigen Akteure ermuntert der Wissenschaftsrat, sich auch unabhängig von den empfohlenen Unterstützungsstrukturen über ihre Erfahrungen mit dem Kerndatensatz und die damit verbundenen Fragestellungen auszutauschen. Engagement seitens kompetenter und engagierter Akteure und zentrale Unterstützung können und sollten einander ergänzen.

1.2 Unterstützung der Einführung

Darüber hinaus ist es im Sinne eines deutschlandweiten Standards essentiell, dass ein einheitlicher Umgang mit den Definitionen sichergestellt ist. Es ist absehbar, dass im Zuge der Einführung des Kerndatensatz-Standards Verständnisfragen beantwortet und Präzisierungen vorgenommen werden müssen. Da solche Fragen bei allen Einrichtungen auftreten können, sollte ein *Helpdesk* allen implementierenden Einrichtungen Unterstützung anbieten.

Dieser *Helpdesk* soll eine Serviceeinrichtung für Forschungseinrichtungen sein. Er soll Hilfestellung bei der Interpretation der Spezifikation bieten und seine Hinweise öffentlich dokumentieren, um sie anderen zugänglich zu machen. Für diese Aufgaben muss eine hohe Erreichbarkeit gewährleistet sein.

Seine Aufgaben müssen aus Kapazitätsgründen auf die Unterstützung bei der Einführung des Kerndatensatzes beschränkt sein. Entsprechend sollten von der Betreuung durch den *Helpdesk* jene Aspekte ausgenommen sein, die auch unabhängig von der Einführung des Kerndatensatzes Beratungsbedarf auslösen können. Dies sind:

- _ rechtliche Fragen, einschließlich übergreifender Fragen des Datenschutzes,
- _ Fragen der Gestaltung administrativer Prozesse und / oder Qualitätssicherungssysteme (Organisationsberatung),
- _ IT-technische Probleme, die über die reine Umsetzung des Kerndatensatzes Forschung hinausgehen und sich beispielsweise bei der Einführung eines FIS ergeben.

Für diese Punkte verweist der Wissenschaftsrat auf bestehendes Know-how an Hochschulen und anderen Forschungseinrichtungen, das in Kompetenzverbänden auf Länderebene wie z. B. HeFIS zum Selbstkostenpreis verfügbar gemacht werden kann, sowie ggf. auf kommerzielle Anbieter.

Das Ziel der Einführungsphase muss es sein, die selbständige und unterstützungsfreie Nutzung des Kerndatensatz Forschung durch wissenschaftliche Einrichtungen und Datennutzer zu erreichen. Dafür erscheint es dem Wissenschaftsrat elementar, über die Sicherstellung einer einheitlichen Interpretation der Spezifikation des Kerndatensatzes hinaus Schwierigkeiten und Erfolge systematisch zu dokumentieren, um deren Ursachen sowie intendierte und nicht intendierte Wirkungen analysieren zu können (vgl. Kapitel C.III, S. 53). Dazu gehört es, zu differenzieren, welche Fragen die Spezifikation des Kerndatensatzes selbst und welche die mit ihrer Nutzung ggf. verbundenen Aspekte (Forschungsinformationssystem, organisationsinterne Datenerfassung, etc.) aufwerfen. Als Serviceeinrichtung kann der *Helpdesk* das Gros der Einrichtungen allerdings nicht in dem Maß aktiv begleiten wie dies für ein solches Monitoring erforderlich ist.

Das Monitoring erfordert die Mitarbeit der implementierenden Einrichtungen. Aus diesem Grund spricht sich der Wissenschaftsrat für die Auswahl einiger Einrichtungen in einem transparenten Prozess aus. Die Implementierungsprozesse des Kerndatensatzes an diesen Einrichtungen sollten dann – unter Rückbindung an den *Helpdesk* – intensiv begleitet und analysiert werden.

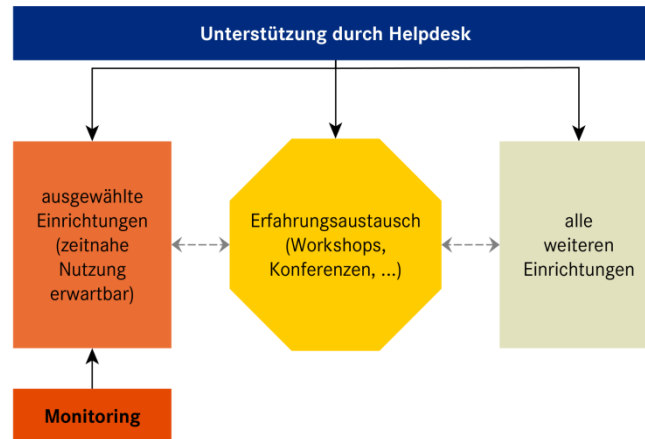
Ein daraus entstehender Monitoring-Bericht sollte am Ende der Einführungsphase konkrete Hinweise zu möglichen Anpassungsbedarfen des Kerndatensatz Forschung enthalten.

Der Wissenschaftsrat empfiehlt, besonders solche Einrichtungen zur Teilnahme am Monitoring zu ermuntern, bei denen eine zeitnahe Nutzung des Kerndatensatzes erwartbar ist. Ein guter Anhaltspunkt hierfür ist die Anzahl der derzeit im Aufbau befindlichen Forschungsinformationssysteme. Forschungseinrichtungen, die ein solches System aufbauen, integrieren dabei ihre Berichtsprozesse und schaffen somit gute Voraussetzungen dafür, zeitnah Daten gemäß der Spezifikation des Kerndatensatzes bereitzustellen. Nach einer Studie zählen derzeit etwa 20 Hochschulen zu dieser Gruppe |⁶², darüber hinaus eine Zahl außeruniversitärer Forschungseinrichtungen. Am Monitoring-Prozess sollte ein Teil dieser Einrichtungen teilnehmen können. Teilnehmende Einrichtungen stehen für Rückfragen hinsichtlich Fehlinterpretationen, Verständnisschwierigkeiten oder Implementierungshindernissen zur Verfügung und geben, falls erforderlich, auch Einblick in ihre Datenerhebungs- und Qualitätssicherungsprozesse.

|⁶² Vgl. Sticht, K.: Einsatz von Forschungsinformationssystemen an Universitäten und Hochschulen mit Promotionsrecht in Deutschland. Ergebnisbericht, 2014, <http://dx.doi.org/10.5281/zenodo.13841> v. 14.01.2016.

Abbildung 2 veranschaulicht das empfohlene Zusammenspiel des Monitoring-Prozesses mit der Unterstützung durch den *Helpdesk*. Ebenso verdeutlicht die Abbildung, dass die Kompetenzen und Erfahrungen der implementierenden Einrichtungen in Netzwerken miteinander verbunden werden sollten.

Abbildung 2: Einführung des Kerndatensatz Forschung



Quelle: Wissenschaftsrat.

Die hier vom Wissenschaftsrat empfohlene Unterstützung und Begleitung der Einführung kann für einzelne Einrichtungen nur dann gewinnbringend sein, wenn sie ihrerseits die Voraussetzungen dafür schaffen, Berichte nach dem Kerndatensatz-Standard erstellen zu können. Der Aufwand dafür wird – je nach Einrichtung – unterschiedlich groß ausfallen. Von denjenigen Einrichtungen, die am Monitoring-Prozess teilnehmen, erwartet der Wissenschaftsrat, dass sie auch selbst bereit sind, ihre Erfahrungen weiterzugeben. Ein in diesem Prozess entstehender Monitoring-Bericht sollte wichtige Hinweise auf Anpassungsbedarfe an der Spezifikation des Kerndatensatzes nach der Einführungsphase geben können (vgl. Abbildung 3, S. 53).

I.4 Aufgaben bei der Unterstützung der Einführung

Der Wissenschaftsrat spricht sich dafür aus, den Aufbau eines Kompetenznetzwerkes, die Bereitstellung eines *Helpdesks* und das Monitoring einem, nach Möglichkeit bereits etablierten, Akteur zu übertragen. Für diesen stellen sich folgende Aufgaben:

I.4.a Aufbau eines Kompetenznetzwerkes

– Förderung horizontaler Vernetzung, z. B. durch Konferenzen, Workshops und Möglichkeiten zum virtuellen Erfahrungsaustausch. Moderation eines Erfahrungsaustausches zwischen den Einrichtungen und mit anderen Orga-

I.4.b Helpdesk

- _ Hilfestellung für die Interpretation der Spezifikation in der Anwendung auf die konkrete Einrichtung.
- _ Beratung von Datennutzern bei der Formulierung spezifikationskonformer Datenabfragen.
- _ Öffentlich zugängliche Dokumentation der Interpretationshilfen und des erarbeiteten Erfahrungswissens.
- _ Bereitstellung von Informationen zu (ggf. auch kommerziellen) Beratungsangeboten zu verwandten Themen wie der Einführung von Forschungsinformationssystemen.

I.4.c Monitoring

- _ Systematische Auswertung und Analyse der Ursachen von Fehlinterpretationen, Verständnisschwierigkeiten oder Implementierungshindernissen in enger Zusammenarbeit mit ausgewählten Einrichtungen.
- _ Erarbeitung von Lösungsvorschlägen.
- _ Berücksichtigung der Erfahrungen des *Helpdesks*.
- _ Dokumentation für die Anpassung der Kerndatensatz-Spezifikation nach der Einführungsphase in einem Monitoring-Bericht.

I.5 Anforderungen bei der Unterstützung der Einführung

Für die Begleitung der Einführung in der Breite stellen sich folgende Anforderungen an den zu befassenden Akteur:

- _ Kenntnisse der verschiedenen Berichtsprozesse und Forschungsinformationssysteme im Hochschul- und im außeruniversitären Bereich,
- _ Erfahrungen mit Monitoring-Prozessen,
- _ hohe Erreichbarkeit bei Klärungsbedarf hinsichtlich der Spezifikation,
- _ Ressourcen für eine transparente, zugängliche Dokumentation, |⁶⁴
- _ Neutralität hinsichtlich der verschiedenen Berichtsbedürfnisse im Wissenschaftssystem.

|⁶³ Vgl. <https://www.dini.de/> und <https://www.forschungsreferenten.de/>

|⁶⁴ Hierbei kann die Dokumentation des Projektes „Spezifikation Kerndatensatz Forschung“ Vorbild sein.

Die dargelegten Aufgaben erfordern eine Anbindung an eine Einrichtung und eine angemessene Finanzierung. Der Wissenschaftsrat bittet daher die Gemeinsame Wissenschaftskonferenz (GWK), zeitnah nach der Veröffentlichung dieser Empfehlungen einen geeigneten Akteur mit der Durchführung zu beauftragen.

Damit es während der Einführungsphase nicht zu Unsicherheiten und divergenten Interpretationen kommt, die die einführenden Einrichtungen zusätzlich belasten, ist eine angemessene personelle und sachliche Ausstattung des *Helpdesk* erforderlich. Der Wissenschaftsrat bittet die GWK, sich zeitnah auf eine entsprechende Finanzierung zu verständigen.

C.II WEITERENTWICKLUNG DER SPEZIFIKATION DES KERNDATENSATZ FORSCHUNG UND SEINER SCHALE

Das Wissenschaftssystem entwickelt sich dynamisch. Ein Standard, der die Dokumentation von Forschungsaktivitäten unterstützen will, muss daher ebenfalls anpassungsfähig angelegt sein. Da es die Anlage des Kerndatensatz Forschung als freiwilliger Standard voraussetzt, dass seine Definitionen breit akzeptiert werden, schlägt der Wissenschaftsrat einen offenen und transparenten Prozess für eine systematische Weiterentwicklung vor. Der Wissenschaftsrat erwartet, dass ein solcher Prozess, der zu einer neuen Version des Kerndatensatz Forschung führt, etwa alle fünf Jahre erforderlich sein wird.

Während die Pflege der forschungsbezogenen Daten jeweils dezentral an den verschiedenen Forschungseinrichtungen erfolgt, bedarf es für Anpassungen und die Weiterentwicklung der Spezifikation einer Instanz, die die möglicherweise veränderten Bedürfnisse ihrer Nutzerinnen und Nutzer berücksichtigt. So kann auch sichergestellt werden, dass zu jedem Zeitpunkt nur eine verbindliche Version des Kerndatensatzes existiert.

II.1 Inhalte der Weiterentwicklung

Bereits in seinen Empfehlungen 2013 hat der Wissenschaftsrat darauf hingewiesen, dass einige Inhalte, die die Kriterien für eine Aufnahme in den Kerndatensatz oder seine Schale erfüllen, nicht schon mit der ersten Version vorliegen würden. Die folgende Liste nennt einige besonders wichtige Inhalte, auf deren künftige Berücksichtigung in der Spezifikation von verschiedenen Wissenschaftsorganisationen besonders Wert gelegt wurde:

– Mitarbeiterinnen und Mitarbeiter mit Weisungsbefugnis (z. B. Nachwuchsgruppenleitungen): Hier wird ein besonderes Augenmerk darauf zu richten sein, wie sich diese Gruppen hinsichtlich ihrer Kompetenzen und Aufgaben von Professorinnen und Professoren und anderem Leitungspersonal abgrenzen lassen, so dass sie ihrer wichtigen Rolle gemäß sowohl in den verschiede-

nen außeruniversitären Forschungseinrichtungen als auch an Hochschulen einheitlich erfasst werden können.

- _ Stipendiatinnen und Stipendiaten leisten häufig wichtige Beiträge zur Forschung, sind aber keine Beschäftigten, wodurch ein verlässlicher Bestand an Daten für Berichte über sie fehlt.
- _ Klärung des Umgangs mit dem Doktorandenstatus bei abgebrochenen Promotionen.
- _ Die Aufbereitung nach Personenjahren, wie z. B. bei einigen außeruniversitären Forschungseinrichtungen gängig, ermöglicht eine präzisere Betrachtung der Personalkapazitäten als Zählungen von Köpfen und VZÄ dies tun, da sowohl Zeit- als auch Vertragsanteile abgebildet werden können.⁶⁵ Die Erfassung der Basisdaten sollte daher in einer Form erfolgen, die mittelfristig einen Umstieg ermöglicht.
- _ PostDoc-Phase: Für diese unbestreitbar wichtige Karrierephase von Nachwuchswissenschaftlerinnen und Nachwuchswissenschaftlern liegt bislang keine einheitliche Definition vor.
- _ Umsetzung des Modells zur Zuordnung zu interdisziplinären Forschungsfeldern bzw. gegenstands- oder problembezogener Forschung, die quer zu den Grenzen etablierter Fächer liegt, entsprechend den Empfehlungen der Projektgruppe „Fächerklassifikation und Thesauri“ und Erstellung einer Liste von Forschungsfeldern (vgl. A.IV.1.c, S. 23).
- _ Verständigung über die Erfassung weiterer Aspekte des Transfers in alle gesellschaftlichen Bereiche, wobei die vielfältigen Dimensionen dieses Begriffes berücksichtigt werden sollten.
- _ Die derzeit geführte Diskussion um die Dokumentation veröffentlichter Forschungsdaten sollte mit dem Ziel aufgegriffen werden, die bestehende Definition im Bereich Publikationen weiter auszuarbeiten.
- _ Prüfung der Erfahrungen bei der Erfassung von Forschungsinfrastrukturen.
- _ Forschungspreise und Auszeichnungen, ggf. auf Basis der Erfahrungen mit der Erhebung an den Einrichtungen.

Jenseits der genannten Inhalte sollte die Frage, welche weiteren Elemente für den Kerndatensatz in Erwägung gezogen werden, auch weiterhin anhand transparenter Kriterien abgewogen werden. Besonders wichtig sind hierbei Da-

| ⁶⁵ Würde für eine Person in den Basisdaten beispielsweise Beginn und Enddatum eines Vertrags sowie die je Zeitabschnitt gültige Wochenstundenzahl erfasst, könnten auf dieser Basis sowohl VZÄ als auch Personenjahre und Kopffzahlen berechnet werden.

tensparsamkeit und die Einschlägigkeit der Daten für möglichst viele Institutionen sowie Fächer bzw. Wissenschaftsgebiete.

Der Wissenschaftsrat empfiehlt ferner, nach Möglichkeit Anpassungen abwärtskompatibel zu gestalten. |⁶⁶ Nach Möglichkeit sind Überleitungsregeln anzugeben. Weiterhin empfiehlt der Wissenschaftsrat mit Nachdruck, die Weiterentwicklung des Kerndatensatzes so anzulegen, dass es auch die Möglichkeit gibt, einzelne Elemente des Kerndatensatzes (z. B. wegen geringer Nutzung) zu entfernen.

II.2 Gestaltung des Weiterentwicklungsprozesses

Im Folgenden gibt der Wissenschaftsrat einige Hinweise zur Gestaltung des Weiterentwicklungsprozesses. Er erkennt dabei an, dass konkrete Erfahrungen im Umgang mit dem Kerndatensatz erst gemacht werden müssen. Entsprechend erheben die folgenden Hinweise keinen Anspruch auf Vollständigkeit, sondern legen Verfahrensgrundsätze fest.

Der Prozess der Weiterentwicklung muss so gestaltet sein, dass er neutral gegenüber den Interessen verschiedener Akteure wie Hochschulen, außeruniversitären Forschungseinrichtungen, Förderorganisationen, Ministerien, der amtlichen Statistik und regelmäßigen Datennutzern ist.

Zu Beginn einer Phase der Weiterentwicklung sollte zunächst der Bedarf an Anpassungen ermittelt werden. Dazu sollten die Dateneignerinnen – Hochschulen wie auch andere Forschungsorganisationen – aber auch Datennutzer begründete Definitionsvorschläge einreichen können. |⁶⁷ Sinnvoll wäre, dass schon in dieser Phase die verschiedenen Vorschläge allgemein einsehbar sind.

Die Vorschläge sollten anschließend in einem transparent gestalteten und vorab dokumentierten Prozess unter Beteiligung einschlägiger Expertinnen und Experten, darunter den Fachgesellschaften, beraten und entschieden werden. Die Erfahrungen des Projektes „Spezifikation Kerndatensatz“ können dabei hilfreich sein. Der Wissenschaftsrat empfiehlt aufgrund dieser Erfahrungen vor allem, erarbeitete Vorschläge zur Änderung des Kerndatensatzes in einer Konsultation auf Tragfähigkeit und Akzeptanz zu prüfen. Auf Basis dieser Konsultation kann abschließend eine Entscheidung getroffen werden, ob eine Definition in den Kern aufgenommen, ggf. als Schale des Kerndatensatzes eingeführt, nicht berücksichtigt oder entfernt werden soll. Begleitend zu einer so entstehenden neuen Version des Kerndatensatzes sollten *Release Notes* veröffentlicht werden, die weitere Entwicklungen klar projektieren.

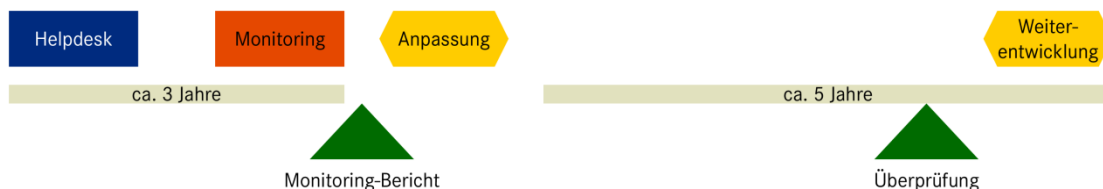
|⁶⁶ Eine Anpassung darf nicht dazu führen, dass bereits i. S. des Kerndatensatzes angelegte Basisdatenbestände nachträglich angepasst werden müssen.

|⁶⁷ Eine Möglichkeit dafür wäre z. B. ein entsprechendes Online-System.

Die Einführung eines Standards für den Austausch von Daten zu Forschungsaktivitäten stellt ein Novum im deutschen Wissenschaftssystem dar. Der Wissenschaftsrat spricht sich nachdrücklich dafür aus, nach einer gewissen Zeit zu prüfen, inwieweit die Ziele erreicht wurden, sich der erwartete Nutzen des Kerndatensatz Forschung für einzelne Forschungseinrichtungen – aber auch für das Wissenschaftssystem als Ganzes – eingestellt hat und das Aufwand-Nutzenverhältnis angemessen ist.

Dem Wissenschaftsrat ist aber auch bewusst, dass ein Standard, wie ihn der Kerndatensatz darstellt, im Laufe der Zeit auch andere als die intendierten Effekte hervorrufen kann. |⁶⁸ Dies sollte erkannt und reflektiert werden können. Erste wichtige Hinweise darauf kann ein Monitoring-Bericht (vgl. C.I.4.c, S. 49) geben. Um jedoch auch eine etwas längerfristige Einschätzung über die Effekte und den Nutzen des Kerndatensatzes zu erhalten, bietet es sich an, nach einer gewissen Zeit der Nutzung die Effekte des Kerndatensatzes zu überprüfen. Eine solche Überprüfung sollte in der Lage sein, durch den Kerndatensatz ausgelöste Veränderungen im Sinne eines Vorher-Nachher-Vergleichs über einen längeren Zeitraum zu beurteilen. Erst im Anschluss sollte die nächste Version der Spezifikation des Kerndatensatzes entwickelt werden.

Abbildung 3: Stufen der Einführung und Weiterentwicklung



Quelle: Wissenschaftsrat.

Abbildung 3 stellt die vorgeschlagenen Stufen der Einführung, Weiterentwicklung und Überprüfung des Kerndatensatzes Forschung auf einer denkbaren Zeitschiene schematisch dar.

Die Überprüfung der Effekte des Kerndatensatzes Forschung sollte einem anderen Akteur übertragen werden als demjenigen, der die Begleitung der Einführung übernimmt.

|⁶⁸ Vgl. Wissenschaftsrat: Empfehlungen zur Bewertung und Steuerung von Forschungsleistung (Drs. 1656-11), Halle November 2011.

Anhang

Abkürzungsverzeichnis	56
Glossar	57
Übersicht über verschiedene Identifier	60
Abbildungsverzeichnis	61
Tabellenverzeichnis	62

AG	Arbeitsgruppe
AUF	Außeruniversitäre Forschungseinrichtungen
BMBF	Bundesministerium für Bildung und Forschung
CERIF	<i>Common European Research Information Format</i>
CRIS	<i>Current Research Information System</i> (engl. für → FIS, s. Glossar,)
DINI	Deutsche Initiative für Netzwerkinformation e. V.
DOI	<i>Digital Object Identifier</i>
DZHW	Deutsches Zentrum für Hochschul- und Wissenschaftsforschung
FhG	Fraunhofer Gesellschaft
FIS	Forschungsinformationssystem
FIT	Fraunhofer-Institut für Angewandte Informationstechnik
GWK	Gemeinsame Wissenschaftskonferenz
HeFIS	Zusammenschluss sieben hessischer Hochschulen zur koordinierten Einführung eines → FIS
HGF	Helmholtz-Gemeinschaft Deutscher Forschungszentren
ID	<i>Identifier</i>
iFQ	(ehemaliges) Institut für Forschungsinformation und Qualitätssicherung, Berlin. Teil des → DZHW
LOM	Leistungsorientierte Mittelvergabe (z. B. an Hochschulen)
MPG	Max-Planck Gesellschaft
ORCID	<i>Open Researcher and Contributor ID</i> (orchid.org)
VZÄ	Vollzeitäquivalent(e)
WGL	Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz
WR	Wissenschaftsrat

Dieses Glossar definiert zentrale Konzepte der vorliegenden Empfehlungen.

Aggregatdaten	Entscheidet sich eine Einrichtung, Daten gemäß dem Kerndatensatz-Standard weiter zu geben, so gibt sie Aggregatdaten oder Daten in Listenform weiter. Diese Daten fassen Sachverhalte z. B. für ein Institut oder ein Fach zusammen. In der Terminologie des Wissenschaftsrates aus den „Empfehlungen zu einem Kerndatensatz Forschung“ entsprechen Aggregatdaten den „Kerndaten“.
Ausdifferenzierung	Eine Ausdifferenzierung ist Teil eines → Kerndatums und beschreibt relevante Merkmale desselben, z. B. bei Beschäftigten das Geschlecht. Eine Ausdifferenzierung hat verschiedene → Ausprägungen .
Ausprägung	Ausprägungen sind die Werte, die eine Ausdifferenzierung annehmen kann (z. B. Geschlecht: <i>weiblich, männlich</i> ; ISBN: <i>entsprechende Nummer</i>).
Basisdaten	Einzeldaten bzw. Daten über die einzelnen Fälle, die einrichtungsintern erfasst werden (z. B. Personen, Drittmittelvereinnahmungen, Patenterteilungen). Zur Weitergabe fasst eine Einrichtung die Basisdaten zu → Aggregatdaten zusammen.
Datenformat	Ein Datenformat bezeichnet die Verbindung einer inhaltlichen Definition mit syntaktischen Vorgaben für ein Element des → Kerndatensatzes oder seiner → Schale .
Definition	Die mit dem Kerndatensatz vorgelegten Definitionen machen für den Datenaustausch nach diesem Standard verbindliche Setzungen hinsichtlich der Inhalte oder der Aggregation von → Basisdaten .
Einführung	Die Einführung des Kerndatensatzes bezeichnet die Begleitung und Unterstützung der → Implementierung an mehreren Forschungseinrichtungen.
Fach	Ein Fach bezeichnet im Kontext des Kerndatensatz Forschung einen thematisch durch die Fächersystematiken des Statistischen Bundesamtes abgegrenzten Gegenstandsbereich. Technisch präziser sollte von Fächergruppen, Lehr- und Forschungsbereichen sowie

Fachgebieten gesprochen werden.

Forschungsfeld	Ein Forschungsfeld umfasst eine Menge von Forschungsaktivitäten, die sich problemorientiert mit einem thematisch abgrenzbaren Gegenstandsbereich, der quer zu Disziplinen liegt, befassen.
Forschungs- informationssystem	Ein Forschungsinformationssystem (FIS) verbindet eine Vielzahl von Datenquellen zu Forschungsaktivitäten und Kontextinformationen und stellt eine Schnittstelle für standardisierte Abfragen zur Verfügung.
Implementierung	Die Implementierung des Kerndatensatz Forschung bezeichnet den Prozess, in dem die Definitionen und Formate in die Berichtsprozesse einer Forschungseinrichtung übernommen werden, um Daten gemäß der Spezifikation austauschen zu können.
Indikatoren	Unter Indikatoren sind direkt beobachtbare Merkmale zu verstehen, die auf die Ausprägung theoretischer Konstrukte bzw. unbeobachtbarer Variablen schließen lassen. Sie können auf der Basis mehrerer → Aggregatdaten gebildet werden Die in der Spezifikation definierten Aggregatdaten sind selbst nicht als Indikatoren zu verstehen und es werden auch keine Hinweise zu ihrer Bildung formuliert.
Kerndatensatz	Der Kerndatensatz umfasst aggregierte Angaben zu Forschungsaktivitäten und deren Kontext in standardisierten Datenformaten, die jede Einrichtung im Wissenschaftssystem selbst vorhalten und auf Anfrage und unter Beachtung rechtlicher und vertraglicher Einschränkungen weitergeben können sollte.
Kerndatum	Ein Kerndatum ist ein Element des → Kerndatensatz Forschung oder seiner → Schale (z. B. Zahl der angemeldeten Patente einer Organisationseinheit der betreffenden Einrichtung in einem bestimmten Jahr).
Schale (des Kerndatensatzes)	Ergänzend zum → Kerndatensatz sind die in der Schale des Kerndatensatzes enthaltenen Angaben nur für spezifische Einrichtungen relevant bzw. aussagekräftig Auch für diese Elemente empfiehlt der Wissenschaftsrat, die getroffenen Definitionen zu verwenden, um den Datenaustausch zu erleichtern. Wissenschaftliche Einrichtungen halten die entsprechenden Angaben zu ihren Forschungsaktivitäten und deren Kontext jeweils

dezentral vor.

- Spezifikation** Die Spezifikation des Kerndatensatz Forschung ist die Menge der Definitionen zu den standardisierten
→ **Kerndaten** mit ihren → **Ausdifferenzierungen** und
→ **Ausprägungen** sowie den zugehörigen Datenformaten.
- Standardisierung** Standardisierung bezeichnet im Kontext des Kerndatensatz Forschung die Verständigung auf gemeinsame, normierende Definitionen von Datenformaten. Diese sind in der Spezifikation des Kerndatensatz Forschung dokumentiert.
- Weiterentwicklung** Die Weiterentwicklung der Spezifikation des Kerndatensatzes bezeichnet den Prozess zur Anpassung und Ergänzung der nun vorgelegten Definitionen mit dem Ziel, den Standard auch an zukünftigen Bedarfen seiner Nutzerinnen und Nutzer ausrichten zu können. Als Weiterentwicklungsempfehlung des Kerndatensatzes gelten solche Elemente für die im Zuge der nun vorliegenden Spezifikation noch keine abschließende Definition vorgeschlagen werden konnte und die daher im Zuge einer Weiterentwicklung standardisiert werden sollten (zu Inhalten vgl. C.II.1).

Diese Übersicht listet einige gängige *Identifier*, die sich zur einrichtungsinternen Organisation von Basisdaten anbieten.

ID	Entität	Kurzbeschreibung
ORCID / ResearcherID	Personen	Nicht-proprietärer, überwiegend numerischer Code zur eindeutigen Identifizierung von in der Wissenschaft aktiven Personen. Neben den meist üblichen einrichtungsinternen Personen- <i>Identifiern</i> etabliert sich die ORCID-ID auch als Austauschformat, das es ermöglicht, personenbezogene Informationen (Autorenschaften, Projektbeteiligungen, Produzentin bzw. Produzent von Forschungsdaten, Anmelderrinnen und Anmelder von Patenten etc.) zu vernetzen. Insbesondere Thomson Reuters verwendet hierzu alternativ die <i>ResearcherID</i> . Die Integration einer Personen-ID mit einrichtungsübergreifender Gültigkeit schafft perspektivisch die Möglichkeit, dass mobile Wissenschaftlerinnen und Wissenschaftler selbstbestimmt Informationen zu ihrer Person von einer Einrichtung zur nächsten mitnehmen (bzw. mitbringen) können. Die so unterstützte Wiederverwendung von Daten ist ein Beitrag zur Aufwandsreduzierung. www.orcid.org ; www.researcherid.com
DOI	Dokumente	Der <i>Digital Object Identifier</i> (DOI) wird vor allem für online veröffentlichte Aufsätze in wissenschaftlichen Zeitschriften verwendet, setzt sich aber zunehmend für andere Objekte (z. B. online bereitgestellte Forschungsdaten) durch. http://www.doi.org/
Fach-ID	Fächer	Die mehrstufige Fachsystematik des Statistischen Bundesamtes verwendet neben den textlichen Fachbezeichnungen (die Änderungen unterworfen sind) persistente Identifikationsnummern. https://www.destatis.de/DE/Methoden/Klassifikationen/BildungKultur/PersonalStellenstatistik.pdf
FUNDRF	Forschungsförderer	Ein international eindeutiger Schlüssel für Fördereinrichtungen mit dem ursprünglichen Verwendungszweck, Angaben zu <i>funding acknowledgements</i> in Publikationsdatenbanken zu vereinheitlichen. http://www.crossref.org/fundref/
Institution-ID	Institute	Von der DFG eingeführter Standard zur eindeutigen Kennzeichnung von Instituten an deutschen Hochschulen und außeruniversitären Forschungseinrichtungen. Der Schlüssel ist Teil des online zugänglichen Informationssystems <i>Research Explorer</i> . http://www.research-explorer.de
ISO-Code 3166	Länder	Der ISO-Code 3166 ist ein lang etablierter Schlüssel zur eindeutigen Kennzeichnung von Ländern. In Forschungsinformationssystemen kann er z. B. zur Erfassung der Herkunft von (Gast-)Wissenschaftlern sowie zur Kennzeichnung internationaler Kooperationen verwendet werden. http://laendercode.net/
ROR-ID	Regionen in Deutschland	Vergleichbar zum ISO-Code 3166 kennzeichnet die ROR-ID sog. „Raumordnungsregionen“ (derzeit 96) in Deutschland. In Forschungsinformationssystemen kann die ROR z. B. zur Erschließung regionenspezifischer Partnerschaften und Kooperationen genutzt werden. http://www.bbsr.bund.de/

Abbildung 1:	Funktionsweise des Kerndatensatz Forschung	19
Abbildung 2:	Einführung des Kerndatensatz Forschung	48
Abbildung 3:	Stufen der Einführung und Weiterentwicklung	53

Tabelle 1:	Gegenüberstellung von Zielen und intendierter Nutzung versus zweckwidrigen und nicht intendierten Nutzungsweisen	21
Tabelle 2:	Inhalte - Beschäftigte	38
Tabelle 3:	Inhalte - Nachwuchsförderung	39
Tabelle 4:	Inhalte - Finanzen und Drittmittel	40
Tabelle 5:	Inhalte - Patente und Ausgründungen	41
Tabelle 6:	Inhalte - Publikationen	42
Tabelle 7:	Inhalte - Forschungsinfrastrukturen	43

Anlage: Spezifikation des Kerndatensatz Forschung:
Version 1.0

2016

Berlin 22.01.2016

Projekt „Spezifikation Kerndatensatz Forschung“

Spezifikation des Kerndatensatz Forschung

Version 1.0

24. September 2015



Institut für Forschungsinformation und Qualitätssicherung (iFQ)

Fraunhofer-Institut für Angewandte Informationstechnik (FIT)

Geschäftsstelle des Wissenschaftsrates

<http://www.kerndatensatz-forschung.de/>



Dieses Dokument (Spezifikation des Kerndatensatz Forschung: Version 1.0), das darauf basierende Datenmodell des Kerndatensatz Forschung sowie das Logo des Kerndatensatz Forschung auf der Titelseite dieses Dokuments sind lizenziert unter einer [Creative Commons Namensnennung - Weitergabe unter gleichen Bedingungen 4.0 International Lizenz](https://creativecommons.org/licenses/by-sa/4.0/).

Inhalt

1	Einleitung und Vorbemerkungen.....	5
1.1	Bereiche und Objekte des Kerndatensatz Forschung.....	7
1.2	Unterscheidung von Aggregat- und Basisdaten.....	7
1.3	Unterscheidung der Kategorien Kerndatensatz Forschung – Schale des Kerndatensatz Forschung – Empfohlen für die Weiterentwicklung	8
2	Aggregatdaten.....	10
2.1	Unterscheidung der Elemente Kerndatum – Ausdifferenzierung – Aggregationsniveau... ..	10
2.2	Regeln der Aggregation und Gruppierung von Kerndaten über Aggregationsniveaus und Ausdifferenzierungen	10
2.3	Aggregatdatenübersicht	11
2.4	Lesehinweise für die Online-Aggregatdatenübersicht	13
3	Basisdaten.....	14
3.1	Unterscheidung der Basisdaten nach Objekt – Attribut – Ausprägung	14
3.2	Erfassung von Attributen und Ausprägungen	15
3.2.1	Objektspezifische Attribute	16
3.2.2	Verknüpfungsattribute	17
3.2.3	Zuordnungsattribute: Zuordnung von Objekten zu Fächern, Organisationseinheiten und Forschungsfeldern	17
3.3	Basisdatenübersicht	19
3.4	Lesehinweise für die Übersicht der Basisdaten im Anhang.....	23
4	Implementierung des Kerndatensatz Forschung.....	24
4.1	Lesehinweise für das technische Datenmodell (Basis- und Aggregatdatenmodell)	24
4.2	Abgleich zwischen dem technischen Datenmodell des Kerndatensatz Forschung und CERIF	28
5	Glossar.....	30
6	Anhang: Übersicht der Basisdaten (<i>Kerndatensatz Forschung</i> und <i>Schale des Kerndatensatz Forschung</i>).....	34

1 Einleitung und Vorbemerkungen

Das Projekt „Kerndatensatz Forschung“ hat in seiner zweijährigen Laufzeit eine Spezifikation für eine standardisierte Weitergabe und Vorschläge für die Vorhaltung von Forschungsinformationen erarbeitet, auf deren Grundlage Forschungseinrichtungen unterschiedliche Berichtsanlässe bedienen können. Die vorgelegte Spezifikation basiert auf den Empfehlungen des Wissenschaftsrats zu einem Kerndatensatz Forschung,¹ expertengeleiteten Diskussionen in verschiedenen Projektgruppen,² auf Konsultationen mit einer großen Zahl von Akteuren des deutschen Wissenschaftssystems, der anschließenden „Abnahme“ und Anpassung der vorgelegten Projektergebnisse durch einen vom Wissenschaftsrat eingesetzten Beirat sowie einer *Open-Beta-Phase*. Im Rahmen dieser zweimonatigen Open-Beta-Phase wurde Vertreterinnen und Vertretern des Wissenschaftssystems die Gelegenheit gegeben, Kommentare und Anpassungsvorschläge für die einzelnen Elemente und Definitionen der Spezifikation einzubringen. Die insgesamt mehr als 1800 eingegangenen elementspezifischen Kommentare sowie deren Auswertung stellten die Grundlage für weitere Anpassungsschritte dar und sind in der nun vorliegenden Version 1.0 des Kerndatensatzes berücksichtigt. Der Kerndatensatz Forschung ist so gestaltet, dass er in zukünftigen Weiterentwicklungsprozessen um weitere Bereiche ergänzt werden kann.

Ziel des Kerndatensatz Forschung ist es, eine Grundlage für den standardisierten Austausch von **aussagekräftigen Informationen** über Forschungsaktivitäten, Forschungspersonal und dessen Produkte unter der Berücksichtigung des Prinzips der Datensparsamkeit zu schaffen. Auf diese Weise soll sichergestellt werden, dass der Kerndatensatz Forschung nur auf Daten beruht, „die eindeutig definiert und für eine große Anzahl an Fächern erhoben werden können“.³ Im Rahmen des Projekts stand die Definition möglichst vielseitig einsetzbarer Kerndaten im Vordergrund; explizit nicht die Bildung und Bewertung von Indikatoren und Kennzahlen.

Aufbauend auf den Empfehlungen des Wissenschaftsrats besteht der Kerndatensatz Forschung aus **aggregierten oder gruppierten Daten** (im Folgenden Aggregatdaten genannt). Diese Aggregatdaten sollten Forschungseinrichtungen auf Anfrage an Datenabfrager weitergeben können (siehe Kapitel 2). Um den Einrichtungen eine Abschätzung in Hinblick auf den damit einhergehenden Verarbeitungs- und Erhebungsaufwand zu erleichtern, hat das Projekt über die Empfehlungen des Wissenschaftsrats hinausgehend eine aus dem Kerndatensatz Forschung **abgeleitete Basisdatenübersicht** entwickelt. Diese Basisdaten sind *nicht für die Weitergabe* vorgesehen (siehe Kapitel 3). Die Aggregat- und Basisdaten werden ebenfalls in einem **technischen Datenmodell** (siehe auch Kapitel 4) dargestellt, dessen Elemente auf den europäischen CERIF-Standard und damit z. B. gängigen Forschungsinformationssystemen abgebildet werden können.

Die Aggregatdaten des Kerndatensatz Forschung (d. h. die *Kerndaten* oder *Daten aus der Schale* des Kerndatensatzes mit bestimmten *Ausdifferenzierungen* und *Aggregationsniveaus*; siehe Abschnitt 2.1 oder Glossar für Erläuterungen) folgen allgemeinen Anforderungen an Berichte. Sie eignen sich damit nicht für alle konkreten Berichtsanlässe in gleicher Weise. Für einzelne Berichtsanlässe können weitere Daten sinnvoll sein. Die Bewertung von Forschungsaktivitäten (wie sie z. B. bei der Erarbeitung von Kennzahlen und Indikatoren erforderlich ist) ist nicht Gegenstand der vorgelegten Spezifikation. Die Konstruktion sinnvoller und tragfähiger Kennzahlen oder Indikatoren auf Basis des Kerndatensatz Forschung obliegt den berichtsnachfragenden Einrichtungen.

Die Basisdaten lassen sich einerseits zu den Aggregatdaten des *Kerndatensatz Forschung* und seiner *Schale* zusammenfassen. Andererseits sind eine Reihe weiterer, über den Kerndatensatz Forschung hinausgehender Aggregationsmöglichkeiten denkbar, die durch die jeweiligen Berichtsanlässe bestimmt werden. Grundsätzlich sieht das hier vorgelegte Modell keine gebündelte Speicherung und Verarbeitung von Daten über verschiedene Forschungseinrichtungen vor. Vielmehr

¹ Wissenschaftsrat: Empfehlungen zu einem Kerndatensatz Forschung (Drs. 2855-13), Berlin, Januar 2013.

² *Ibid.*, S. 40 ff.

³ *Ibid.*, S. 16.

obliegt die Pflege und Qualitätssicherung der Daten den Dateneignern. Dies gilt auch für die Entscheidung über die Weitergabe von Aggregatdaten.

Die Verarbeitung von personenbezogenen Basisdaten sowie die Weitergabe von personenbezogenen Aggregatdaten (z. B. in der Form von Publikationslisten) müssen dabei immer im Rahmen der jeweils gültigen datenschutzrechtlichen Regelungen stattfinden. Es empfiehlt sich daher, insbesondere die Einführung von Softwaresystemen zur Unterstützung des Kerndatensatz Forschung mit einer systematischen Sichtung der Verfahrensverzeichnisse sowie der institutionellen Rechtslage in Zusammenarbeit mit den institutionellen Datenschutzbeauftragten zu verbinden (siehe Ergebnisbericht des Projekts „Spezifikation Kerndatensatz Forschung“).⁴

Die in diesem Dokument („Spezifikation des Kerndatensatz Forschung: Version 1.0“⁵) vorgelegten Ergebnisse des Projekts „Spezifikation Kerndatensatz Forschung“ stecken einen Rahmen für die Vorhaltung und Verarbeitung von Forschungsinformationen in Hochschulen und außeruniversitären Forschungseinrichtungen ab.

Im Sinne der Transparenz und um den zugrundeliegenden Entscheidungsprozess abzubilden, wurden die Argumente für und gegen die Aufnahme und Definition der Kerndaten sowie ihrer Ausdifferenzierungen dokumentiert. Die Erarbeitung und Dokumentation der Definitionen erfolgte auf Basis eines Argumentationsmodells, das Lösungsansätze, deren Alternativen sowie Argumente in einem Netzgraph in Beziehung setzt. Als Modellierungswerkzeug kam die freie Software *Compendium* zum Einsatz.⁵ Sie bietet neben einer intuitiven Benutzeroberfläche vor allem Exportfunktionalitäten. Weitere Hinweise finden sich im Ergebnisbericht. Die Argumente für und gegen die Aufnahme und Definition der Kerndaten sowie ihrer Ausdifferenzierungen lassen sich in der beigelegten Online-Aggregatdatenübersicht detailliert nachverfolgen.⁶

Aufbau des Dokuments und Lesehinweise

Das vorliegende Dokument ist wie folgt aufgebaut: Der nächste Abschnitt (1.1) gibt einen Überblick zu den Bereichen und Objekten des Kerndatensatz Forschung. Abschnitt (1.2) stellt die der Spezifikation zugrunde liegenden Perspektiven dar und erläutert den Unterschied zwischen Basis- und Aggregatdaten. Der anschließende Abschnitt 1.3 beschreibt die grundsätzliche Architektur des Kerndatensatz Forschung und definiert die unterschiedlichen Kategorien von Aggregatdaten.

Kapitel 2 geht danach auf die einzelnen Aggregatdaten des Kerndatensatz Forschung und der Schale des Kerndatensatz Forschung ein. Zuerst werden die für die Aggregatdaten verwendeten Begrifflichkeiten definiert (Abschnitt 2.1). Im Anschluss werden die Regeln zur Aggregation der Kerndaten (Abschnitt 2.2) dargestellt, gefolgt von der Aggregatdatenübersicht (Abschnitt 2.3). Das Kapitel schließt mit Lesehinweisen für die online hinterlegte Aggregatdatenübersicht⁷ (siehe auch Abschnitt 2.4).

Kapitel 3 beschreibt die Basisdaten, die einrichtungsintern für eine Bereitstellung des Kerndatensatz Forschung vorgehalten werden können. Es erläutert die Terminologie (Abschnitt 3.1), expliziert Regeln zur Erfassung von Basisdaten (Abschnitt 3.2), gibt eine Basisdatenübersicht (Abschnitt 3.3) und schließt mit Lesehinweisen für die im Anhang befindliche komplette Basisdatenübersicht (Abschnitt 3.4).

Im Anschluss daran wird in Kapitel 4 die Implementierung des Kerndatensatz Forschung thematisiert. Es umfasst eine Beschreibung und Erläuterung des technischen Datenmodells (4.1) sowie

⁴ http://www.kerndatensatz-forschung.de/version1/Ergebnisbericht_Projekt_KDSE.pdf

⁵ <http://www.compendiumng.org>

⁶ http://www.kerndatensatz-forschung.de/version1/Spezifikationstabelle_KDSE_v1.html sowie http://www.kerndatensatz-forschung.de/version1/Spezifikationstabelle_KDSE_v1_komplett.html

⁷ *Ibid.*

eine Zusammenfassung des Abgleichs zwischen dem technischen Datenmodell des Kerndatensatz Forschung und dem europäischen CERIF-Standard (Abschnitt 4.2; siehe auch Glossar).

Dem Dokument sind in Kapitel 5 ein Glossar und in Kapitel 6 ein Anhang beigelegt, der die Basisdatenübersicht enthält. Die Definitionen, die den in dieser Spezifikation beschriebenen Basis- und Aggregatdaten zugrunde liegen, sind gemeinsam mit den Aggregatdaten online hinterlegt.⁸

Aus Gründen der besseren Lesbarkeit und Übersicht wird im vorliegenden Dokument und in den Anhängen ausschließlich die männliche Form für die Bezeichnung und Darstellung von Typen oder Rollen von Personen verwendet. Die gewählten Formulierungen sind nicht geschlechtsspezifisch zu verstehen und gelten für Personen jeden Geschlechts.

1.1 Bereiche und Objekte des Kerndatensatz Forschung

Die inhaltlichen **Bereiche** (siehe Glossar) des Kerndatensatz Forschung wurden bereits in den Empfehlungen des Wissenschaftsrats zu einem Kerndatensatz Forschung identifiziert und benannt.⁹ Sie umfassen **Beschäftigte, Nachwuchsförderung, Drittmittel und Finanzen, Patente und Ausgründungen, Publikationen** sowie **Forschungsinfrastrukturen**.

Um die Abbildung der in den Empfehlungen zu einem Kerndatensatz Forschung vorgeschlagenen Kerndaten (also Aggregatdaten wie z. B. die Anzahl von Beschäftigten oder die Anzahl von abgeschlossener Qualifizierungsverfahren, siehe auch Kapitel 2) nach den empfohlenen Ausdifferenzierungen (wie z. B. Geschlecht) zu erleichtern, wurden Vorschläge dazu entwickelt, welche grundlegenden Informationen eine Forschungseinrichtung hierfür in einem internen System vorhalten sollte. Diese sind damit auf der Ebene der nicht für die Weitergabe vorgesehenen Basisdaten (siehe Glossar). Sie umfassen **Personen** (darunter fallen Beschäftigte, Doktoranden, Promovierte und Habilitierte), **Strukturierte Promotionsprogramme, Drittmittelprojekte, Patente, Ausgründungen, Publikationen** sowie **Forschungsinfrastrukturen**.

1.2 Unterscheidung von Aggregat- und Basisdaten

In der vorliegenden Spezifikation für einen Kerndatensatz Forschung wird grundsätzlich zwischen Aggregatdaten und Basisdaten unterschieden.

Aufbauend auf den Empfehlungen des Wissenschaftsrats besteht der Kerndatensatz Forschung aus **Aggregatdaten (oder gruppierten Daten)** (siehe Glossar). Die möglichen Kombinationen der Kerndaten nach Ausdifferenzierungen und Aggregationsniveaus sind in der online hinterlegten Aggregatdatenübersicht dargestellt.¹⁰ Zur Erzeugung der verschiedenen Aggregatdaten durch Aggregationsregeln ist die einrichtungsinterne Zusammenfassung von **Basisdaten** (siehe Glossar) für unterschiedliche Objekte erforderlich. Die zentrale bzw. dezentrale Erhebung, Verarbeitung und Vorhaltung von Basisdaten in den jeweiligen Forschungseinrichtungen ist nicht zu verwechseln mit der Übertragung von anonymisierten Aggregatdaten an unterschiedliche Berichtsnachfrager.

Die **Aggregatdaten** (siehe Kapitel 2) im Kerndatensatz Forschung sind nach unterschiedlichen *Bereichen* (Beschäftigte, Nachwuchsförderung, Drittmittel und Finanzen, Patente und Ausgründungen, Publikationen, Forschungsinfrastrukturen) gegliedert und umfassen *Kerndaten* jeweils mit *Ausdifferenzierungen* und *Aggregationsniveaus* (siehe Abschnitt 2.1 für weitere Erläuterungen). Ein Beispiel für Aggregatdaten ist die *Anzahl von Personen (Kerndatum)* eines bestimmten Personentyps (etwa von Beschäftigten oder Promovierten eines bestimmten Geschlechts) nach *Geschlecht* (Ausdifferenzierung) und *Fach* (Aggregationsniveau). Andere Aggregationsregeln (siehe Abschnitt 2.2)

⁸ http://www.kerndatensatz-forschung.de/version1/Spezifikationstabelle_KDSF_v1.html

⁹ Wissenschaftsrat: Empfehlungen zu einem Kerndatensatz Forschung (Drs. 2855-13), Berlin, Januar 2013. Empfehlungen zur einrichtungsinternen Datenhaltung (von Basisdaten) waren nicht vorgesehen. In der ursprünglichen Terminologie entsprechen „Kerndaten“ den Aggregatdaten.

¹⁰ http://www.kerndatensatz-forschung.de/version1/Spezifikationstabelle_KDSF_v1.html

beziehen sich beispielsweise auf die Ermittlung von *Vollzeitäquivalenten von Beschäftigten* auf der Ebene einer *Organisationseinheit*.

Basisdaten (siehe Kapitel 3) beziehen sich auf die zur Abbildung der unterschiedlichen *Aggregatdaten* des Kerndatensatzes bzw. der Schale benötigten *Objekte* (Personen, Strukturierte Promotionsprogramme, Drittmittelprojekte, Patente, Ausgründungen, Publikationen und Forschungsinfrastrukturen) und deren Eigenschaften (*Attribute* und jeweilige *Ausprägungen*, siehe Abschnitt 3.1). Attribute von Personen sind beispielsweise das *Geschlecht* oder die *Staatsangehörigkeit* mit den Ausprägungen „männlich“, „weiblich“ oder „ohne Angabe“¹¹ bzw. der Angabe des konkreten „Landes der Staatsangehörigkeit“ der Person. Diese Basisdaten sind nicht Gegenstand des Austauschformates Kerndatensatz Forschung. Sie werden nicht zur Weitergabe empfohlen.

Die Vorhaltung der Basisdaten stellt sicher, dass die in der Aggregatdatenübersicht spezifizierten Aggregatdaten erzeugt werden können und in der Berichtslegung weitergegeben werden können. Basisdaten wie auch die daraus abgeleiteten Aggregatdaten unterliegen, wie auch in schon bestehenden Berichtslegungsprozessen, der Datenhoheit der Forschungseinrichtung.

1.3 Unterscheidung der Kategorien Kerndatensatz Forschung – Schale des Kerndatensatz Forschung – Empfohlen für die Weiterentwicklung

Der **Kerndatensatz Forschung** (siehe Glossar) umfasst die Bereiche und Elemente, die der Wissenschaftsrat für alle Forschungseinrichtungen im Wissenschaftssystem als relevant erachtet.¹²

Über die Empfehlungen des Wissenschaftsrats hinausgehend hat sich das Projekt auch mit Bereichen und Elementen befasst, die nur für spezifische Forschungseinrichtungen berichtsrelevant sind. Diese sind in der **Schale des Kerndatensatz Forschung** (siehe Glossar) aufgeführt. Auch für diese Elemente wird empfohlen, die vereinbarten Definitionen zu verwenden, damit die Einrichtungen in vergleichbarer Form dazu auskunftsfähig sind. Teilweise werden im Rahmen der Definitionen auch Vorschläge gemacht, für welche Einrichtungen diese Elemente für sinnvoll erachtet werden. Der *Kerndatensatz Forschung* umfasst Elemente aus den Bereichen *Beschäftigte*, *Nachwuchsförderung*, *Drittmittel und Finanzen*, *Patente und Ausgründungen* sowie *Publikationen*. Einige Elemente dieser Bereiche sind Bestandteil der Schale des Kerndatensatz Forschung. Der Bereich *Forschungsinfrastrukturen* ist in Gänze Teil der *Schale des Kerndatensatz Forschung*. Im vorliegenden Dokument werden die Projektempfehlungen in Hinblick auf den *Kerndatensatz Forschung* grün und hinsichtlich der *Schale des Kerndatensatz Forschung* violett dargestellt.

Schließlich wurden im Rahmen des Projekts auch Bereiche und Elemente diskutiert, die in einem zukünftigen Weiterentwicklungsprozess behandelt und ausdefiniert werden sollten (**Empfohlen für die Weiterentwicklung**; siehe Glossar).¹³

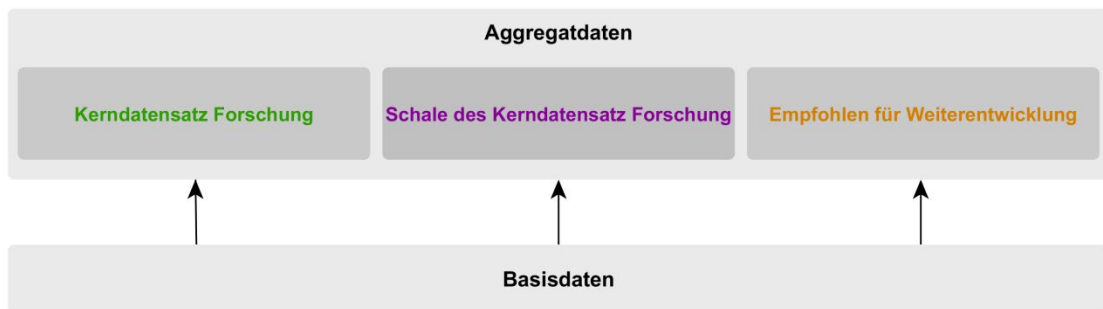
Abbildung 1 gibt einen Überblick zu den verschiedenen Kategorien von Aggregatdaten.

¹¹ Mit diesen Ausprägungen folgt der Kerndatensatz Forschung dem derzeit gültigen deutschen Personenstandsgesetz (PStG).

¹² Wissenschaftsrat: Empfehlungen zu einem Kerndatensatz Forschung (Drs. 2855-13), Berlin, Januar 2013.

¹³ Diese Elemente sind nicht Teil des Kerndatensatzes. Sie finden sich online dokumentiert (dargestellt in orangefarbener Schrift) unter: http://www.kerndatensatz-forschung.de/version1/Spezifikationstabelle_KDSF_v1_komplett.html

Abbildung 1: Kategorien von Aggregatdaten



Anmerkungen: Die Aggregatdaten und ihre Kategorien werden in Kapitel 2 thematisiert. Die zugrundeliegenden Basisdaten werden in Kapitel 3 erläutert.

2 Aggregatdaten

Der Kerndatensatz Forschung standardisiert Aggregatdaten z. B. für die Weitergabe im Rahmen von Berichten. Um Berichtsfragen zu beantworten, werden die innerhalb der Forschungseinrichtung vorgehaltenen Basisdaten zu Aggregatdaten zusammengefasst (siehe auch Abschnitt 1.2). Die Berichtslegung von Aggregatdaten erfolgt entweder zu bestimmten Stichtagen oder für ein Kalenderjahr (siehe Abschnitt 2.2).¹⁴

2.1 Unterscheidung der Elemente Kerndatum – Ausdifferenzierung – Aggregationsniveau

Die Aggregatdaten des *Kerndatensatzes* sowie der *Schale des Kerndatensatzes* umfassen verschiedene Elemente (siehe Glossar). Ein Bericht umfasst **Kerndaten** mit verschiedenen **Ausdifferenzierungen** (z. B. *Anzahl der Beschäftigten nach Personalkategorie*). Dabei können die Kerndaten unterschiedlich zusammengefasst werden. Hierfür sind verschiedene **Aggregationsniveaus** vorgesehen.

Gemäß den Empfehlungen des Wissenschaftsrats repräsentiert ein **Kerndatum** „eine einzelne Klasse von Informationen über Forschungsaktivitäten“¹⁵ (siehe auch Glossar). Kerndaten können entweder aggregierte Informationen (z. B. die Anzahl von Personen, Drittmittelträge bzw. -einnahmen, die Anzahl erteilter Patente) oder beschreibende Informationen über einzelne Objekte (z. B. Strukturierte Promotionsprogramme, Drittmittelprojekte, Patente, Publikationen) enthalten. Im ersten Fall ist das Kerndatum eine **Summe**, etwa die *Anzahl der Beschäftigten*. Im zweiten Fall repräsentiert ein Kerndatum eine **Liste**, etwa die *Liste aller Publikationen*. Die Information, dass sich ein Kerndatum auf diese Listenform bezieht, ist in der Aggregatdatenübersicht (siehe Abbildung 2) in eckigen Klammern mit angegeben.

Ausdifferenzierungen (siehe Glossar) erlauben es, die relativ groben Kerndaten je nach Berichts Anlass weiter aufzuschlüsseln. So kann z. B. die *Anzahl von Beschäftigten* einer Forschungseinrichtung nach *Geschlecht* und/oder *Qualifikation* aufgeschlüsselt werden.

Kerndaten werden nur auf bestimmten **Aggregationsniveaus berichtet** (siehe Glossar). Das Aggregationsniveau bestimmt, welche Elemente zusammengefasst werden. So kann beispielweise die *Anzahl der Beschäftigten* eines *Fachs* (z. B. Chemie) oder einer *Organisationseinheit* (z. B. Institut für Prozesstechnik) berichtet werden. Hier würden dann alle Beschäftigten dieses Faches (siehe 2.2) aufsummiert bzw. eine Liste der Publikationen des Instituts zusammengestellt. Die möglichen Aggregationsniveaus sind zusammen mit den Kerndaten in der Aggregatdatenübersicht in Abschnitt 2.3 gelistet. Aggregations- bzw. Gruppierungsregeln werden im folgenden Abschnitt 2.2 beschrieben.

2.2 Regeln der Aggregation und Gruppierung von Kerndaten über Aggregationsniveaus und Ausdifferenzierungen

Eine Aggregation von **Kerndaten in Listenform** wird durch die Weitergabe aller an den Datenabfrager weitergebbaren Daten in Form einer einfachen Liste erreicht. Für Aggregationen von **Kerndaten in Summenform** werden alle Einheiten eines Objekts (je nach zu berichtenden Aggregationsniveau und Ausdifferenzierungen) für den Berichtszeitraum aufsummiert.

Die **Berichtslegung** (siehe Glossar) gemäß Kerndatensatz bezieht sich je nach Bereich und Kerndatum auf den Stichtag oder das Kalenderjahr. Die Kerndaten der Bereiche *Beschäftigte* und *Forschungsinfrastrukturen* sowie die Mehrzahl der Kerndaten des Bereichs *Nachwuchsförderung* sollten stichtagsbezogen (31.12. bzw. 01.12. für Daten, die auch für die Personalstatistik des Statistischen Bundesamts geliefert werden müssen) berichtet werden. Für die meisten Kerndaten des Bereichs *Drittmittel und Finanzen* sowie für die Bereiche *Patente und Ausgründungen* und *Publikationen* gilt das

¹⁴ http://www.kerndatensatz-forschung.de/version1/Berichtszeitraum_KDSF_v1.pdf

¹⁵ Wissenschaftsrat: Empfehlungen zu einem Kerndatensatz Forschung (Drs. 2855-13), Berlin, Januar 2013, S.57.

Kalenderjahr als Zeitraum für die Berichtslegung.¹⁶ Für eine Einrichtung, die ihre Daten auf der Ebene der Basisdaten vorliegen hat, lassen sich – je nach Anwendungsfall – Berichte auch für andere Zeiträume erstellen (z. B. nach Prüfungsjahren wie in der Prüfungsstatistik des Statistischen Bundesamts). Eine Aufsummierung erfolgt dann entsprechend für den gewählten Zeitraum. Bei der Meldung abgeschlossener Promotionen besteht – analog zum Statistischen Bundesamt (Destatis) – die Möglichkeit, im vergangenen Berichtszeitraum nicht mehr eingegangene Promotionsmeldungen an den aktuellen Berichtszeitraum anzufügen. Eine rückwirkende Nachtragung in das Vorjahr ist, äquivalent zu Destatis, nicht vorgesehen, da dies die Integrität des Berichtsverfahrens verringern und die Datenpflegeanforderungen erheblich erhöhen würde.

Die Aufsummierung erfolgt immer nach der kleinsten Kombinationssumme der zu berichtenden Kategorien. So ist z. B. zur Erstellung eines Berichts, der die *Anzahl von Beschäftigten* nach Geschlecht und Qualifikation erfragt, die Summe aller Beschäftigten nach den jeweiligen Ausprägungen von *Geschlecht* („männlich“, „weiblich“, „ohne Angabe“) und *Qualifikation* („promoviert“, „habilitiert“) zu bilden. In diesem Beispiel werden also $3 * 2 = 6$ Einzelsummen berechnet und an den Berichtsnachfrager übertragen. Ein Berichtsnachfrager kann diese dann flexibel zusammenfassen, ohne dass für verschiedene Fragestellungen einzelne Abfragen notwendig wären. So sind unterschiedlich detaillierte Auswertungen möglich.

Bei einer hohen Anzahl von im Bericht gewünschten Ausdifferenzierungen und Aggregationsniveaus ergibt sich eine Vielzahl von Aufschlüsselungsmöglichkeiten und entsprechend kleine Einzelsummen. Aus Gründen des Datenschutzes sollte darauf geachtet werden, die abgefragten Kombinationen von Aggregationsniveaus und Ausdifferenzierungen nur so detailliert wie nötig zu wählen. Weiterhin ist die Trennung der Verarbeitung und Übertragung bislang separat vorgehaltener Forschungsinformationen durch organisatorisch-technische Maßnahmen innerhalb der Forschungseinrichtung sicherzustellen (siehe Ergebnisbericht des Projekts „Spezifikation Kerndatensatz Forschung“).¹⁷

2.3 Aggregatdatenübersicht

Die folgende Übersicht (Abbildung 2) führt die Kerndaten für den *Kerndatensatz Forschung* und die *Schale des Kerndatensatz Forschung* sowie die empfohlenen Aggregationsniveaus auf. Empfehlungen zu den Ausdifferenzierungen, nach denen die jeweiligen Aggregatdaten bereitgestellt werden sollten, finden sich in Spalte zwei der entsprechenden Übersichtstabelle¹⁸ (siehe auch Abschnitt 2.4).

Elemente sind grün hervorgehoben, sofern sie Teil des *Kerndatensatz Forschung* sind. Violett hervorgehobene Elemente sind Teil der *Schale des Kerndatensatz Forschung*.

Die Spezifikation umfasst Kerndaten unterschiedlichen Typs. Diese umfassen

- Summenangaben unterschiedlicher Objekte (z. B. Summen von Beschäftigten oder Doktoranden) hinsichtlich einer (z. B. *Geschlecht*) oder mehrerer Ausdifferenzierungen (z. B. *Geschlecht* und *Staatsangehörigkeit*),
- Listen von Objekten (wie z. B. der Strukturierten Promotionsprogramme, Drittmittelprojekte oder Publikationen einer Forschungseinrichtung).

¹⁶ http://www.kerndatensatz-forschung.de/version1/Berichtszeitraum_KDSF_v1.pdf

¹⁷ http://www.kerndatensatz-forschung.de/version1/Ergebnisbericht_Projekt_KDSF.pdf

¹⁸ http://www.kerndatensatz-forschung.de/version1/Spezifikationstabelle_KDSF_v1.html

Abbildung 2: Aggregatdatenübersicht mit Aggregationsniveaus

Beschäftigte	Vollzeitäquivalente	Anzahl Personen (Kopfzahlen)	Gesamtpersonal
	- Organisationseinheit - Fach - Forschungsfeld	- Organisationseinheit - Fach - Forschungsfeld	- Gesamteinrichtung
Nachwuchsförderung	Anzahl Doktoranden an titelvergebenden Einrichtungen	Anzahl Doktoranden an nicht-titelvergebenden Einrichtungen	
	- Organisationseinheit - Fach - Forschungsfeld	- Organisationseinheit - Fach - Forschungsfeld	
	Anzahl abgeschlossener Qualifizierungsverfahren	Strukturiertes Promotionsprogramm [Listenform]	
	- Organisationseinheit - Fach - Forschungsfeld	- Organisationseinheit - Fach - Strukturiertes Promotionsprogramm - Forschungsfeld	
Drittmittel und Finanzen	Drittmiteleinnahmen	Drittmittelerträge	Drittmittelprojekt [Listenform]
	- Organisationseinheit - Fach - Drittmittelprojekt - Forschungsfeld	- Organisationseinheit - Fach - Drittmittelprojekt - Forschungsfeld	- Organisationseinheit - Fach - Drittmittelprojekt - Forschungsfeld
	Sonstige Einnahmen	Gesamtbudget	
	- Organisationseinheit - Gesamteinrichtung	- Gesamteinrichtung	
Patente	Patent [Listenform]	Anzahl Patentfamilien	Anzahl prioritätsbegründender Patentanmeldungen
	- Organisationseinheit - Fach - Patent - Gesamteinrichtung - Forschungsfeld	- Organisationseinheit - Fach - Gesamteinrichtung - Forschungsfeld	- Organisationseinheit - Fach - Gesamteinrichtung - Forschungsfeld
	Anzahl der erteilten Patente	Anzahl Ausgründungen	Erträge aus Schutzrechten
	- Organisationseinheit - Fach - Gesamteinrichtung - Forschungsfeld	- Organisationseinheit - Fach - Gesamteinrichtung - Forschungsfeld	- Organisationseinheit - Fach - Gesamteinrichtung - Forschungsfeld
Publikationen	Publikation [Listenform]		
	- Organisationseinheit - Fach - Publikation - Publikationstyp - Dokumenttyp - Schöpfer - Peer-reviewed - Veröffentlichungsjahr - Forschungsfeld		
Forschungsinfrastrukturen	Forschungsinfrastruktur [Listenform]		
	- Infrastruktur		

2.4 Lesehinweise für die Online-Aggregatdatenübersicht

Die online hinterlegte Tabelle mit Definitionen der Aggregatdaten (Aggregatdatenübersicht,¹⁹ siehe Glossar) bildet verschiedene mögliche Berichtsabfragen ab. Die Tabelle bietet zunächst eine hierarchische Übersicht in Listenform. Weiterhin enthält sie einen **Definitionskatalog**, der die Definitionen aller Elemente des Kerndatensatzes sowie der Schale des Kerndatensatzes enthält. Die Elemente des *Kerns* sind dabei **grün** hervorgehoben, **violett** hervorgehobene Elemente sind Teil der *Schale des Kerndatensatzes Forschung*.

In der hierarchischen Übersicht (**Überblicksverzeichnis**) sind die Bereiche *Beschäftigte*, *Nachwuchsförderung*, *Drittmittel und Finanzen*, *Patente und Ausgründungen*, *Publikationen* sowie *Forschungsinfrastrukturen* dargestellt. Zu den einzelnen Bereichen sind jeweils in der ersten Spalte die Kerndaten, in der zweiten Spalte die Ausdifferenzierungen mit ihren jeweiligen Ausprägungen und in der dritten Spalte die zugeordneten Aggregationsniveaus angeführt. Abfragen werden sich in der Regel nur auf eine begründete Auswahl dieser Daten beziehen (siehe Abschnitt 2.2).

Eine Ausnahme in der Darstellungsform ergibt sich für die Bereiche *Publikationen* und *Forschungsinfrastrukturen*. Im Bereich *Publikationen* wurde eine weitere Ausdifferenzierung „Format“ gebildet. Das „Format“ einer *Publikation* lässt sich nur über die darunter aufgeführten Ausdifferenzierungen (*Band*, *Heft*, *Name der Konferenz*, *Seitenbereich*) darstellen. Im Bereich *Forschungsinfrastrukturen* betrifft diese Ausnahme die Ausdifferenzierungen „Nutzung/Nutzungsintensität“ sowie „Betriebspersonal“. Diese Kerndaten können ebenfalls ausschließlich nach den jeweils darunter aufgeführten Ausdifferenzierungen dargestellt werden

Alle Elemente im Überblicksverzeichnis sind in Klammern mit einem eindeutigen Identifier versehen.²⁰ Dieser dient zum Auffinden der Definition des Elements im Definitionskatalog. Sofern ein Element eines Bereichs in einem weiteren Bereich ebenfalls vorkommt, ist eine Verknüpfung mit der betreffenden Definition durch ein Pfeilsymbol (z. B. Be72 → Na58) dargestellt.

Der **Definitionskatalog** im Anschluss an das Überblicksverzeichnis enthält die Definitionen aller Elemente des *Kerndatensatzes Forschung* der *Schale des Kerndatensatzes Forschung*. In der ersten Spalte ist der Identifier aufgeführt. Die Bezeichnung des Elements steht in der zweiten Spalte. Zusätzlich wird die Information, dass es sich bei einem Element um ein Kerndatum oder ein Aggregationsniveau handelt, in dieser Spalte in eckigen Klammern mit angegeben. Die dritte Spalte enthält die Definition des Elements. Die möglichen Ausdifferenzierungen eines Elements sind in Spalte vier dargestellt. In der fünften Spalte finden sich die Argumente, die von den Projektbeteiligten im Verlauf der Diskussionen für bzw. gegen die Aufnahme des Elements geäußert wurden. Schließlich sind in Spalte sechs Verweise zu Definitionen in anderen Bereichen des Kerndatensatzes aufgeführt.

¹⁹ http://www.kerndatensatz-forschung.de/version1/Spezifikationstabelle_KDSF_v1.html

²⁰ Die Identifier reflektieren den Diskussionsverlauf im Projekt. Sie sind daher nicht chronologisch oder hierarchisch.

3 Basisdaten

Dieses Kapitel bietet Vorschläge für die einrichtungsinterne Haltung von **Basisdaten**. Diese sind nicht Teil des Kerndatensatzes als Standard für die Weitergabe von Informationen. Die Basisdaten stellen die Grundlage für die Weitergabe von Aggregatdaten dar. Einrichtungen organisieren die Datenhaltung ihrer Basisdaten autonom; dieses Kapitel will hierfür eine Unterstützung anbieten.

Alle hier genannten Basisdaten sind aus den Aggregatdaten des Kerndatensatz Forschung abgeleitet. Das heißt, dass sie unter Umständen nicht dafür ausreichen, über die Aggregatdaten des Kerndatensatzes hinausgehende Informationen bereitzustellen.

Die Vorhaltung von Basisdaten sollte **kontinuierlich** und **im Rahmen administrativer** Prozesse erfolgen. Eine berichtsspezifische Erfassung von Basisdaten ausschließlich zu bestimmten Stichtagen ist nicht zu empfehlen. So sollten z. B. Arbeitsverträge von Beschäftigten über den jeweiligen *Anteil an Vollzeitstelle* sowie den *Beginn* und das *Ende des Arbeitsvertrags* (mit Datumsangaben) erhoben werden. Eine solche Vorhaltung erlaubt zum einen eine gewisse Flexibilität in Hinblick auf die Erfüllung unterschiedlicher Vorgaben für die Berichtslegung (es lässt sich damit z. B. sowohl auf Kalenderjahre als auch Prüfungsjahre aggregieren; siehe auch Abschnitt 2.2). Zum anderen ist damit generell eine flexible Aggregation der Basisdaten möglich (z. B. ließen sich so neben dem Aggregatdatum *Vollzeitäquivalente* von Beschäftigten auch Personenjahre abbilden).

Es ist darüber hinaus denkbar, dass Basisdaten des Kerndatensatz Forschung in Teilen aus **externen Datenbanken** gewonnen werden. So können Informationen zu *Publikationen* bereits jetzt aus unterschiedlichen Publikationsdatenbanken gewonnen und zusammengeführt werden. Ebenso wäre es aus Sicht des Projekts sinnvoll, dass Drittmittelgeber Angaben zu *Drittmittelprojekten* in Zukunft im Kerndatensatzformat aufbereiten und diese Hochschulen und außeruniversitären Forschungseinrichtungen zur weiteren Verwertung bereitstellen.

Für die Kerndaten *Drittmiteleinnahmen*, *Drittmittelerträge*, *Sonstige Einnahmen* und *Gesamtbudget* (Bereich *Drittmittel und Finanzen*) sowie *Erträge aus Schutzrechten* (Bereich *Patente und Ausgründungen*) schien es nicht sinnvoll Basisdaten abzuleiten, da die Festlegung auf bestimmte Objekte nicht möglich ist.

3.1 Unterscheidung der Basisdaten nach Objekt – Attribut – Ausprägung

Um eine vergleichbare Datenbasis zu schaffen, sollte die Erfassung der **Objekte** möglichst vollständig erfolgen. Es werden dabei die folgenden Objekte (detaillierte Definitionen sind der online hinterlegten Übersicht der Aggregatdaten²¹ zu entnehmen) unterschieden:

- I. **Person**: Personen stehen nicht nur in einer Arbeitsbeziehung zu einer Forschungseinrichtung, sie werden in Teilen auch durch sie qualifiziert. Sollen die Aggregatdaten des Kerndatensatz Forschung gebildet werden, kann eine Person z. B. modelliert werden als ein Beschäftigter der Einrichtung und/oder ein Angehöriger der Gruppe des wissenschaftlichen Nachwuchses, die Doktoranden, Promovierte sowie Habilitierte umfasst.
- II. **Strukturiertes Promotionsprogramm**: Ein strukturiertes Promotionsprogramm ist ein Programm der Doktorandenförderung, das ein strukturiertes Veranstaltungsprogramm für alle Teilnehmer vorsieht und welches darüber hinaus mindestens zwei der folgenden drei Bedingungen erfüllt:
 1. gemeinsame Verantwortung für die Betreuung der Doktoranden durch die beteiligten Hochschullehrer,

²¹ http://www.kerndatensatz-forschung.de/version1/Spezifikationstabelle_KDSF_v1.html

2. offenes, wettbewerbliches Aufnahmeverfahren mit Ausschreibung,
 3. Stipendien oder Stellen für zumindest einen Teil der teilnehmenden Doktoranden.
- III. Drittmittelprojekt: Drittmittelprojekte im Sinne des Kerndatensatz Forschung umfassen Vorhaben mit einem Forschungsbezug, deren Finanzierung nicht aus Grund- oder Sondermitteln erfolgt. Unterschiedliche Förderphasen eines übergeordneten Projekts (z. B. eines Sonderforschungsbereichs) gelten hierbei als separate Drittmittelprojekte.
 - IV. Patent: Patente im Sinne des Kerndatensatz Forschung sind prioritätsbegründete Patente nach den Definitionen in den jeweiligen Gesetzen (Europäisches Patentübereinkommen, Patentgesetz – PatG, Gebrauchsmustergesetz – GebrMG, Gesetz über den Schutz von Marken und sonstigen Kennzeichen – MarkenG).
 - V. Ausgründung: Ausgründungen im Sinne des Kerndatensatz Forschung bezeichnen jene Unternehmensgründungen die zur Verwertung von geistigem Eigentum oder Know-how der Einrichtung unter Abschluss einer formalen Vereinbarung (Nutzungs-, Lizenz- und/oder Beteiligungsvertrag) erfolgt sind.
 - VI. Publikation: Eine wissenschaftliche Publikation im Sinne des Kerndatensatz Forschung ist ein manifestes und veröffentlichtes Produkt, das aus der Forschungsarbeit einer oder mehrerer Personen resultiert.
 - VII. Forschungsinfrastruktur: Forschungsinfrastrukturen im Sinne (der *Schale*) des Kerndatensatz Forschung sind umfangreiche/aufwendige Instrumente, Ressourcen oder Serviceeinrichtungen für die Forschung in allen Wissenschaftsgebieten, die sich durch eine mindestens überregionale Bedeutung für das jeweilige Wissenschaftsgebiet sowie durch eine mittel- bis langfristige Lebensdauer von mindestens fünf Jahren auszeichnen und für eine externe Nutzung zur Verfügung stehen.

Für die genannten Objekte sollte eine Datenbankstruktur angelegt werden, die die unterschiedlichen **Attribute** (siehe Glossar) mit ihren verschiedenen **Ausprägungen** (siehe Glossar) aufnehmen kann. So sollte eine Einrichtung beispielsweise für alle Personen die Attribute *Geschlecht* (mit den Ausprägungen „männlich“, „weiblich“ oder „ohne Angabe“), *Staatsangehörigkeit* (Angabe des Lands/der Länder der Staatsangehörigkeit) sowie *Geburtsdatum* (exakte Angabe) aus ihren Beständen zusammenführen. Bei Drittmittelprojekten sollten u. a. die Attribute *Titel des Projekts* oder *Drittmittelgeber* (mit unterschiedlichen Ausprägungen wie z. B. „DFG“, „Bund“, „Bundesländer“) bereitgestellt werden.

Um diese Struktur abbilden zu können, sind die Basisdatenübersichten (siehe Abschnitt 3.3 sowie Anhang in Kapitel 6) nach drei unterschiedlichen Typen von Attributen (siehe Glossar) gegliedert:

1. **Objektspezifische Attribute**,
2. **Verknüpfungsattribute**, die Verknüpfungen des Objektes zu anderen Objekten beinhalten,
3. **Zuordnungsattribute**, die die Zuordnung des Objekts zu Organisationseinheit, Fach sowie zukünftig Forschungsfeld spezifizieren.

3.2 Erfassung von Attributen und Ausprägungen

Die folgenden Abschnitte (3.2.1, 3.2.2, 3.2.3) typisieren die Regeln für die Erfassung der Attribute: objektspezifische Attribute, Verknüpfungsattribute und Zuordnungsattribute.

Maßgeblich für den Kerndatensatz als Austauschformat ist die definitionsgemäße Bereitstellung der aggregierten bzw. gruppierten Daten. Insofern schließt der Kerndatensatz Forschung trotz

der im Folgenden detaillierten Regelungsvorschläge alternative und parallele Erfassungswege und -regeln für die Aggregation nicht aus.

Objektspezifische Attribute (Abschnitt 3.2.1) sind in der Mehrzahl so angelegt, dass für ein Attribut eines Objekts nur eine Ausprägung vorliegt. Fälle, in denen logisch mehrere Ausprägungen denkbar sind (so kann z. B. ein Beschäftigter mehrere *Staatsangehörigkeiten* oder mehrere Arbeitsverträge mit unterschiedlichen *Personalkategorien* und Befristungsangaben haben), werden in 3.2.1 explizit als solche ausgewiesen.

Die Erfassung von **Verknüpfungsattributen** (Abschnitt 3.2.2) ist mit vergleichbar wenigen Vorgaben hinsichtlich der Zuordnung von Ausprägungen zu Attributen versehen.

Bei der **Zuordnung** von Objekten zu Organisationseinheit, Fach und Forschungsfeld schließlich werden unterschiedliche Regeln vorgeschlagen (ausführlich siehe Abschnitt 3.2.3). Dabei gilt, dass alle Objekte primär der Organisationseinheit zugeordnet werden, der sie angehören. Diese Zuordnung sollte zu einer Organisationseinheit erfolgen.

3.2.1 Objektspezifische Attribute

Objektspezifische Attribute sind separat für jedes Objekt aufgelistet. Für die Anlage von Basisdaten empfehlen sich die folgenden Regeln für die Erfassung von objektspezifischen Attributen.

Die Ausprägungen sind nicht zwangsläufig für alle Objekte zutreffend. So kann z. B. für Beschäftigte (Objekt *Person*) in Bezug auf das Attribut (höchste) *Qualifikation* „promoviert“ oder „habilitiert“ genannt werden. Sollte für einen Beschäftigten keine der beiden Ausprägungen zutreffen (der Beschäftigte also weder promoviert noch habilitiert sein), so ist das entsprechende Feld leer zu lassen.

Manche Attribute können auch mehrere Ausprägungen annehmen (z. B. mehrere *Staatsangehörigkeiten* von Personen). Für Attribute, bei denen solche Mehrfachausprägungen logisch möglich sind, liefern die Definitionen des Kerndatensatz Forschung Regeln für eine eindeutige Festlegung zur Ableitung der **Aggregatdaten**.

In Hinblick auf die **Basisdaten** gelten bei Mehrfachausprägungen die folgenden Regeln:

1. Ein Beschäftigter (Objekt *Person*) kann mehrere parallele Arbeitsverträge haben und über diese verschiedenen *Personalkategorien* zuzuordnen sein. Auch kann sich die Befristung über parallele Arbeitsverträge unterscheiden. Der für den Beschäftigten ebenfalls zu erhebende *Anteil an Vollzeitstelle* sollte in diesem Fall ebenfalls **für parallele Arbeitsverträge** separat erhoben werden, um auf dieser Grundlage z. B. das Aggregatdatum *Vollzeitäquivalente* (Bereich *Beschäftigte*) bestimmen zu können. Die Attribute

- *Personalkategorie* (ebenso wie mögliche Unterattribute),
- *Anteil an Vollzeitstelle*,
- *Beginn und Ende des Arbeitsvertrags*

sollen für einen Beschäftigten (Objekt *Person*) vertragsspezifisch für jeden einzelnen parallelen Arbeitsvertrag erhoben werden. Diese Attribute können somit **bei einer Person mehrere (Unter-)Ausprägungen** haben.

2. Eine weitere Ausnahme bezieht sich auf die Erfassung von Drittmittelprojekten (Objekt *Drittmittelprojekt*, Bereich *Drittmittel und Finanzen*). Die Attribute

- *Drittmittelgeber*,
- *Förderkennzeichen*,
- *Bewilligungssumme*

können für von mehreren Gebern finanzierte Drittmittelprojekte **mehrere Ausprägungen** haben und sollen daher drittmittelgeberspezifisch, d. h. für jeden einzelnen parallelen Drittelgeber, erhoben werden.

3. Es sind darüber hinaus Mehrfacherfassungen von Ausprägungen für bestimmte Attribute möglich. Neben den oben genannten Attributen ist die objektbezogene Erhebung mehrerer Ausprägungen bei Vorhandensein in Hinblick auf die folgenden Attribute vorgesehen:
 - *Staatsangehörigkeit* (Objekt *Person*)
 - *Kooperation* (Objekt *Person*)
 - *Land des zur Aufnahme einer Promotion berechtigenden Studienabschlusses* (Objekt *Person*)
 - *Beteiligte Institutionen* (Objekt *Strukturiertes Promotionsprogramm*)
 - *Land der Institution* (Objekt *Strukturiertes Promotionsprogramm*)
 - *Schöpfertyp* (Objekt *Publikation*)
 - *Zugangsrechte* bei Publikationen (Objekt *Publikation*)
 - *Ressource* bei Publikationen (Objekt *Publikation*)

3.2.2 Verknüpfungsattribute

Attribute, die Objekte verknüpfen, also zwei oder mehrere Basisdaten zueinander in Beziehung setzen, lassen sich einteilen in solche, die Mehrfacherfassung von Ausprägungen nicht ausschließen, und solche, bei denen die Festlegung einer Ausprägung vorgesehen ist. In der Regel ist die Mehrfacherfassung möglich.

Mehrfacherfassung von Ausprägungen möglich:

- (mehrere) *Sprecher* eines Strukturierten Promotionsprogramms (Objekt *Person* – Objekt *Strukturiertes Promotionsprogramm*) [Schale]
- (mehrere) *Doktoranden* oder *Promovierte* eines Strukturierten Promotionsprogramms (Objekt *Person* – Objekt *Strukturiertes Promotionsprogramm*)
- (mehrere) *Erfinder* eines Patents (Objekt *Person* – Objekt *Patent*)
- (mehrere) *Schöpfer* einer Publikation (Objekt *Person* – Objekt *Publikation*)
- (mehrere) *Forschungsinfrastrukturen* einer Publikation (Objekt *Forschungsinfrastruktur* – Objekt *Publikation*)
- (mehrere) *Publikationen* einer Forschungsinfrastruktur (Objekt *Publikation* – Objekt *Forschungsinfrastruktur*)
- (mehrere) *Drittmittelprojekte* einer Publikation (Objekt *Drittmittelprojekt* – Objekt *Publikation*)
- (mehrere) *Publikationen* eines Drittmittelprojekts (Objekt *Publikation* – Objekt *Drittmittelprojekt*)

Mehrfacherfassung von Ausprägungen nicht möglich:

- (ein) *Erstbetreuer* des Doktoranden oder Promovierten (Objekt *Person* – Objekt *Person*)
- (ein) Strukturiertes Promotionsprogramm (Objekt *Strukturiertes Promotionsprogramm*), an dem der *Doktorand* oder *Promovierte* (Objekt *Person*) teilnimmt

3.2.3 Zuordnungsattribute: Zuordnung von Objekten zu Fächern, Organisationseinheiten und Forschungsfeldern

Da der Kerndatensatz Forschung den Datenaustausch zwischen Datenabfragern und Dateneignern durch standardisierte Vorgaben vereinfachen soll, spielt die Zuordnung von Objekten zu **Organisationseinheiten, Fächern** und **Forschungsfeldern** (Zuordnungsattribute) eine zentrale Rolle. Für sie gelten folgende Hinweise:

Eine Einrichtung sollte ihre Kerndaten nach Organisationseinheiten *und* nach Fächern aggregiert liefern können. Sofern eine Aggregation nach Fächern inhaltlich nicht sinnvoll ist, reicht es aus, Kerndaten nach Organisationseinheiten aggregiert bereit zu halten.

Primär für die Aggregation im Rahmen des Kerndatensatz Forschung ist die Zuordnung zu einer **Organisationseinheit**. Als Organisationseinheit gilt dabei, Destatis folgend, jeweils die kleinste, an der Forschungseinrichtung bestehende organisatorische Einheit, der Personal/Stellen zugeordnet ist/sind.²² Es wird empfohlen, bei der Benennung der Organisationseinheiten die Bezeichnungen des *DFG-Research Explorers* zu verwenden.²³ Davon abweichende, in Einrichtungen etablierte, Bezeichnungen sind ebenfalls zulässig. Sie sollten im Sinne einer einheitlichen Benennung der DFG gemeldet werden.

Objekte werden der Organisationseinheit zugeordnet, mit der sie verbunden sind. Personen beispielsweise sind mit einer Organisationseinheit über Arbeitsverträge, Doktoranden über Betreuungsvereinbarungen, Prüfungsordnungen etc. verbunden. Für die meisten Objekte des Kerndatensatzes ist zumindest auch eine indirekte Zuordnung über Personen möglich. Eine Publikation würde so z. B. aufgrund der Zugehörigkeit ihrer Autoren einer Organisationseinheit zugeordnet werden. Wie die Zuordnung zu einer Organisationseinheit vorgenommen wird, entscheidet eine Einrichtung selbst. Eine Zuordnung darf allerdings innerhalb eines Berichtszeitraumes nicht auf unterschiedliche Arten und Weisen vorgenommen werden. Eine *Mehrfachzuordnung* zu Organisationseinheiten kann nur anteilig (prozentual) und nur dann erfolgen, wenn sie rechnerisch und inhaltlich nachvollziehbar ist, beispielsweise also wenn für ein Projekt Drittmittel an mehrere Institute geflossen sind.

Sofern ein Berichtssystem aufgebaut werden soll, empfiehlt sich dringend die Möglichkeit der indirekten Zuordnung über Personen vorzusehen, da diese die größte Flexibilität auch für andere Aggregationen bietet.

Die Zuordnung zu einem **Fach** oder zu mehreren Fächern erfolgt für eine Organisationseinheit. Alle zu ihr gehörenden Objekte ‚erben‘ die Fachzuordnung der Organisationseinheit. Grundlage für die Fachzuordnung sind die Fächersystematiken des Statistischen Bundesamtes.²⁴ Die Zuordnung der Organisationseinheiten erfolgt in höchstmöglicher Detaillierung. Diese sollte dem Fachgebiet („Viersteller“ gemäß Destatis, z. B. 3710 – Anorganische Chemie), in jedem Fall jedoch mindestens der Fächergruppe („Zweisteller“) entsprechen. Sobald Objekte zueinander in Bezug gesetzt werden sollen, müssen sie derselben Ebene zugeordnet sein (z. B. beide als „Viersteller“), um Aggregationsprobleme zu vermeiden. Der *DFG-Research Explorer* enthält bereits eine Zuordnung der dort erfassten Organisationseinheiten gemäß den Destatis-Systematiken.

Eine Organisationseinheit *kann* (anteilig) mehreren Fächern (z. B. zu 60 Prozent Mathematik und zu 40 Prozent Biologie) zugeordnet werden. Die jeweiligen Anteile sollen in Prozent ausgewiesen werden und in der Summe 100 Prozent ergeben. Die anteilige Zuordnung sollte über mehrere Berichtsperioden hinweg konstant bleiben und rechnerisch nachvollziehbar sein (z. B. anhand der Zahl der beschäftigten Personen mit einem bestimmten fachlichen Hintergrund). Auch die Mehrfachzuordnung sollte auf der niedrigsten Ebene der Destatis-Fächersystematiken erfolgen. Zuordnungen zu dieser Ebene nehmen z. B. auch außeruniversitäre Forschungseinrichtungen im Rahmen ihrer Berichtspflichten für das Finanz- und Personalstatistikgesetz vor.

Für *einrichtungsinterne* Zwecke kann es darüber hinaus sinnvoll sein, Produkte wie Publikationen auch direkt einem Fach oder mehreren Fächern zuzuordnen.

Die Zuordnung zu **Forschungsfeldern**, die quer zu Fächern liegen, kann vorgenommen werden, sobald eine Liste mit Forschungsfeldern vorliegt. Diese Liste ist noch nicht fertig gestellt und wird bei Verfügbarkeit online bereitgestellt (vergleiche Empfehlungen der Projektgruppe

²² Siehe Statistisches Bundesamt: Schlüsselverzeichnisse für die Personalstatistiken (H201/32134000 und H201/32135100), Stand 2015.

²³ Der *DFG-Research Explorer* ist „das Forschungsverzeichnis der Deutschen Forschungsgemeinschaft (DFG) und des Deutschen Akademischen Austauschdienstes (DAAD) in Zusammenarbeit mit der Hochschulrektorenkonferenz (HRK). <http://www.research-explorer.de/>

²⁴ <https://www.destatis.de/DE/Methoden/Klassifikationen/BildungKultur/Content75/SystematikenHochschulen.html>

„Fächerklassifikation und Thesauri“: *Interdisziplinarität im Forschungsberichtswesen – Empfehlungen zur Abbildung*).²⁵

1. Die Zuordnung zu **Forschungsfeldern** erfolgt *direkt* für einzelne Elemente wie Strukturierte Promotionsprogramme, Drittmittelprojekte, Patente und Publikationen. Beispiel: Eine Publikation wird dem Forschungsfeld *Energieforschung* zugeordnet.
2. Eine Mehrfachzuordnung ist möglich. Beispiel: *Energieforschung* und *Sicherheit*.
3. Im Gegensatz zur fachlichen Zuordnung können die Forschungsfelder, auf denen ein Forscher tätig ist, aus den Angaben zu den ihm zugeordneten Objekten abgeleitet werden. Damit können für jede Person mehrere Forschungsfelder angegeben werden. Die Zuordnung von Personen zu Forschungsfeldern erfolgt somit indirekt.
4. Eine indirekte („vererbte“) Zuordnung der Objekte über Organisationseinheiten ist möglich, wird aber nicht empfohlen.
5. Von einer freien Zuordnung zu Forschungsfeldern – unabhängig bzw. vor Vorliegen der geplanten standardisierten Liste – wird dringend abgeraten.

3.3 Basisdatenübersicht

Der folgende Abschnitt gibt eine nach Objekten strukturierte Übersicht zu den Basisdaten, wie Einrichtungen sie intern vorhalten können – einschließlich Attributen und Ausprägungen. Eine vollständige Tabelle der Basisdaten findet sich im Anhang (Kapitel 6).

Die Listen in Abbildung 4 auf S. 21 umfassen Angaben zu **drei Typen von Attributen**:

1. **Objektspezifische Attribute der obersten Ebene** für das jeweilige Objekt des Kerndatensatzes bzw. der Schale.
2. **Verknüpfungen des Objekts zu einem anderen Objekt** (Verknüpfungsattribute).
Solche Verknüpfungen entstehen z. B., wenn ein Beschäftigter einer Forschungseinrichtung (Objekt *Person*) der *Erstbetreuer* eines Doktoranden (Objekt *Person*) oder *Schöpfer* einer Publikation (Objekt *Publikation*) ist.
3. **Zuordnung des Objekts zu Organisationseinheit, Fach sowie zukünftig Forschungsfeld** (Zuordnungsattribute).

Objekte werden der **Organisationseinheit** zugeordnet mit der sie verbunden sind. Personen beispielsweise sind mit einer Organisationseinheit über Arbeitsverträge, Doktoranden über Betreuungsvereinbarungen, Prüfungsordnungen etc. verbunden. Für die meisten Objekte des Kerndatensatzes ist zumindest auch eine indirekte Zuordnung über Personen möglich. Die Zuordnung zu einem **Fach** oder zu mehreren Fächern erfolgt für eine Organisationseinheit.

Die Ausprägungen aller aufgeführten Attribute sowie Angaben zu möglichen Unterattributen und -ausprägungen sind der vollständigen Übersicht der Basisdaten im Anhang (Kapitel 6) zu entnehmen. Die den Objekten, Attributen und Ausprägungen jeweils zugrundeliegenden Definitionen sind Teil der Übersicht der Aggregatdaten.²⁶

Die in der Basisdatenübersicht (siehe Anhang in Kapitel 6) dargestellten Attribute bedeuten nicht, dass eine Forschungseinrichtung nicht auch darüber hinausgehende Attribute erfassen kann. Da sich diese Basisdaten am Kerndatensatz ausrichten, steht zu erwarten, dass verschiedene themen- oder einrichtungsspezifische Berichtsinteressen nicht abgedeckt sind.

²⁵ http://www.kerndatensatz-forschung.de/version1/PGK_Empfehlungen_zur_Abbildung_von_Interdisziplinaer.pdf

²⁶ http://www.kerndatensatz-forschung.de/version1/Spezifikationstabelle_KDSF_v1.html

Abbildung 3 veranschaulicht beispielhaft den Zusammenhang zwischen unterschiedlichen Objekten, objektspezifischen Attributen, Verknüpfungsattributen und Zuordnungsattributen für den *Kerndatensatz Forschung*. Nicht dargestellt ist der Übersichtlichkeit halber die Zuordnung von Objekten zu Forschungsfeld sowie Fach, da letzteres durch die die Zuordnung zur Organisationseinheit bestimmt wird.

Abbildung 3: Objekte des Kerndatensatz Forschung und jeweilige Attributtypen

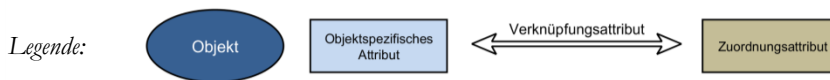
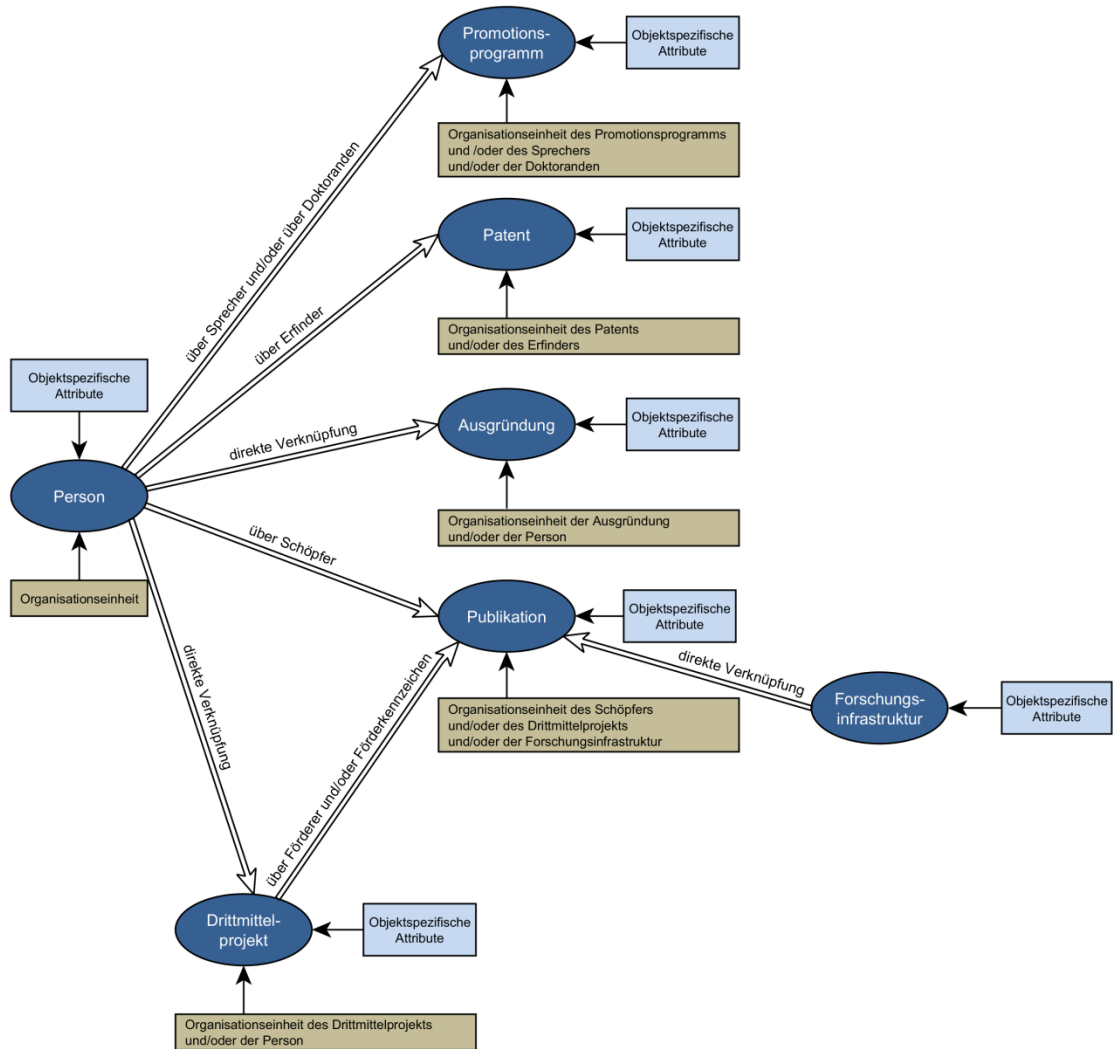


Abbildung 4 gibt einen graphischen Überblick über empfohlene Zuordnung von Basisdaten (Objekte und deren Attribute) zu den unterschiedlichen Bereichen des Kerndatensatz Forschung.²⁷

²⁷ Andere Wege der Erfassung sind möglich, sofern sie reproduzierbar die definitionsgemäße Bereitstellung der Aggregatdaten sicherstellen.

Abbildung 4: Objekte des Kerndatensatzes Forschung und Attribute nach Bereichen

Beschäftigte	Person		
	<ul style="list-style-type: none"> - Geschlecht - Staatsangehörigkeit - Geburtsdatum <p>Verknüpfungen:</p> <ul style="list-style-type: none"> - Betreuer des Doktoranden / Promovierten - Doktorand / Promovierter des Promotionsprogramms - Sprecher des Promotionsprogramms - Erfinder des Patents - Schöpfer der Publikation 	<p>Zuordnungen:</p> <ul style="list-style-type: none"> - Organisationseinheit - Fach - Forschungsfeld 	<p>Beschäftigter</p> <ul style="list-style-type: none"> - Personalkategorie (je Arbeitsvertrag) - Besoldung, Bezeichnung, Angaben zu gemeinsamer Berufung inkl. Name des weiteren Einrichtung (für Professoren) - Anteil an Vollzeitstelle (je Arbeitsvertrag) - Vertragsbeginn und -ende (je Arbeitsvertrag) - Finanzierungsform - Qualifikation - Tätigkeitsart
Nachwuchsförderung	Doktorand und Promovierter		Habilitierter
	<ul style="list-style-type: none"> - Start der Promotion - Erstbetreuer an Einrichtung - Strukturiertes Promotionsprogramm - Kooperation - Land des zur Aufnahme einer Promotion berechtigenden Studienabschlusses <p>Zusätzlich für Promovierten:</p> <ul style="list-style-type: none"> - Zeitpunkt des Abschlusses des Promotionsverfahrens 	<ul style="list-style-type: none"> - Zeitpunkt des Abschlusses des Habilitationsverfahrens 	
	Strukturiertes Promotionsprogramm		
	<ul style="list-style-type: none"> - Titel - Beteiligte Institutionen und jeweils Land der Institution - Finanzierung des Programms 	<p>Verknüpfungen:</p> <ul style="list-style-type: none"> - Sprecher (s. Person) - Doktorand / Promovierter (s. Person) 	<p>Zuordnungen:</p> <ul style="list-style-type: none"> - Organisationseinheit - Fach - Forschungsfeld
Drittittel und Finanzen	Drittittelprojekt		
	<ul style="list-style-type: none"> - Titel des Projekts - Koordinationsrolle (Einrichtung) - Name der Sprechereinrichtung / des Konsortialführers (bei Verbundprojekten) - Projektbeginn - Projektende - Bewilligungssumme (je Drittmittelgeber) - Titel des übergeordneten Projekts 	<p>Verknüpfungen:</p> <ul style="list-style-type: none"> - Drittmittelgeber / Förderer (s. Publikation) - Förderkennzeichen je Drittmittelgeber (s. Publikation) 	<p>Zuordnungen:</p> <ul style="list-style-type: none"> - Organisationseinheit - Fach - Forschungsfeld
Patente und Ausgründungen	Patent		
	<ul style="list-style-type: none"> - Titel des prioritätsbegründenden Patents - Datum der prioritätsbegründenden Erstanmeldung - Veröffentlichungsnummer - Patentfamilie - Datum der Erteilung 	<p>Verknüpfungen:</p> <ul style="list-style-type: none"> - Erfinder (s. Person) 	<p>Zuordnungen:</p> <ul style="list-style-type: none"> - Organisationseinheit - Fach - Forschungsfeld
	Ausgründung		
	<ul style="list-style-type: none"> - Name der Ausgründung - Datum der Ausgründung 		<p>Zuordnungen:</p> <ul style="list-style-type: none"> - Organisationseinheit - Fach - Forschungsfeld
Publikationen	Publikation		
	<ul style="list-style-type: none"> - Schöpfer typ - Titel des Werks - Veröffentlichungsjahr - Verlag - Verlagsort - Quelle - Identifizier inkl. Bezeichnung - Sprachcode - Zugangsrechte 	<ul style="list-style-type: none"> - Format (Band, Heft, Name der Konferenz, Seitenbereich, falls zutreffend) - Peer-Reviewed - Art der Qualifikationsschrift - Dokumententyp - Publikationstyp - Ressource 	<p>Verknüpfungen:</p> <ul style="list-style-type: none"> - Schöpfer (s. Person) - Förderer / Drittmittelgeber (s. DM-Projekt) - Förderkennzeichen (s. DM-Projekt) <p>Zuordnungen:</p> <ul style="list-style-type: none"> - Organisationseinheit - Fach - Forschungsfeld
Forschungsinfrastrukturen	Forschungsinfrastruktur		
	<ul style="list-style-type: none"> - Bezeichnung - Beschreibung - Typ der Forschungsinfrastruktur - Art der Forschungsinfrastruktur - Art des Zugangs 	<ul style="list-style-type: none"> - Nutzung/Nutzungsintensität (nach Kategorie) - Betreiber - Koordinator - Betriebspersonal (nach Kategorie) 	<p>Verknüpfungen:</p> <ul style="list-style-type: none"> - Publikation(en) <p>Zuordnungen:</p>

Verwendung von Identifikatoren in den Basisdaten

Es wird empfohlen, bei der Erfassung und Zusammenführung der Basisdaten für **alle Objekte** des Kerndatensatz Forschung eindeutige Identifikatoren (sogenannte *Identifizier*) zu verwenden. Dabei können sowohl **einrichtunginterne**, als auch **externe Identifizier** verwendet werden. Ein Beispiel für einen auch einrichtungsübergreifend einsetzbaren Identifizier und offenen Standard zur Zuordnung von **Personen** stellt ORCID (Open Researcher and Contributor ID) dar, dessen Einsatz empfohlen wird.²⁸

Für **Publikationen** und Forschungsoutputs gibt es eine Reihe von verschiedenen externen Identifikatoren, die von unterschiedlichen Organisationen vergeben werden. Es wird empfohlen, alle Identifizier im Rahmen der Basisdaten zu erfassen und zu verarbeiten, zu denen der Forschungseinrichtung für den jeweiligen Publikationstyp Informationen vorliegen. Diese umfassen z. B.

- DOI (Digital Object Identifier), dessen Verwendung als weitverbreiteter und publikationstypenübergreifender Standard besonders empfehlenswert ist,²⁹
- ISSN (International Standard Serial Number) [Online und/oder Print],³⁰
- ISBN (International Standard Book Number),³¹
- Scopus EID (durch die bibliographische Datenbank *Scopus* vergebener Identifizier),³²
- WOSut (durch die bibliographische Datenbank *Web of Science* vergebener Identifizier),³³
- URI (Uniform Resource Identifier),³⁴
- ISMN (International Standard Music Number),³⁵
- Artikelnummer (Nummer des Artikels/Beitrags im Journal oder Sammelband),
- PMID (durch die bibliographische Datenbank *PubMed* vergebener Identifizier).³⁶

Erfassung und Verarbeitung von Informationen zu Publikationen in den Basisdaten

Die für das Objekt *Publikation* vorgeschlagenen Attribute lassen sich nicht für alle Publikationen gleichermaßen erheben. Zum einen sind einige Attribute auf bestimmte Publikationstypen beschränkt (z. B. lässt sich *Heft* oder *Name der Konferenz* nicht für die Publikationstypen „Buch“ oder „Software“ erfassen). Zum anderen gibt es Attribute, die mit nur einem bestimmten Erfassungsverfahren erhoben werden können. So sollen die unter *Identifizier* aufgeführten Attribute beispielsweise nicht alle parallel erhoben werden. Sofern eine Einrichtung die Angaben für bestimmte Publikationen allerdings aus Publikationsdatenbanken bezieht, sollen alle Identifizier erhoben werden, die in dieser Datenbank bereitgestellt werden. Es ist generell nicht vorgesehen, diese Informationen manuell zu erfassen (siehe oben).

Für die Erfassung eines wichtigen Teils von gängigen Publikationen (insgesamt mindestens 75 Prozent der Publikationen) reichen die folgenden Merkmale zur Erfassung aus:

²⁸ <http://orcid.org/>

²⁹ <http://www.doi.org/>

³⁰ <http://www.issn.org/>

³¹ <http://www.isbn.org/>

³² <http://www.elsevier.com/solutions/scopus>

³³ <http://thomsonreuters.com/en/products-services/scholarly-scientific-research/scholarly-search-and-discovery/web-of-science.html>

³⁴ <http://www.iana.org/assignments/uri-schemes/uri-schemes.xhtml>

³⁵ <http://www.ismn-international.org/>

³⁶ <https://www.ncbi.nlm.nih.gov/pubmed/>

Beispiel Journalartikel	Beispiel Monographie	Beispiel Sammelbandbeitrag
- Name des Schöpfers	- Name des Schöpfers	- Name des Schöpfers
- <u>Schöpfertyp</u> : Autor	- <u>Schöpfertyp</u> : Autor	- <u>Schöpfertyp</u> : Autor
- Titel des Werks	- Titel des Werk	- Titel des Werk
- Veröffentlichungsjahr	- Veröffentlichungsjahr	- Veröffentlichungsjahr
- Quelle	- Verlag	- Verlag
- Identifier (z. B. DOI)	- Identifier (z. B. ISBN)	- Quelle
- Band, Heft, Seitenbereich	- Sprachcode	- Identifier (z. B. DOI)
- Sprachcode	- Zugangsrechte	- Seitenbereich
- Zugangsrechte	- Qualifikationsschrift	- Sprachcode
- Peer-Reviewed	- <u>Publikationstyp</u> : Monographie	- Zugangsrechte
- <u>Dokumenttyp</u> : Wissenschaftl. Artikel	- [Förderer] ³⁸	- Peer-Reviewed
- <u>Publikationstyp</u> : Journalartikel	- [Förderkennzeichen] ³⁸	- <u>Publikationstyp</u> : Sammelbandbeitrag
- [Förderer] ³⁷		- [Förderer] ³⁸
- [Förderkennzeichen] ³⁸		- [Förderkennzeichen] ³⁸

3.4 Lesehinweise für die Übersicht der Basisdaten im Anhang

In Kapitel 6 findet sich die vollständige Liste der Basisdaten, die auch die Ausprägungen zu den genannten Attributen der obersten Ebene aufführt sowie Angaben zu Unterattributen und -ausprägungen enthält. Die Liste ist ebenfalls **nach den genannten Objekten strukturiert** und hebt Verknüpfungen zu anderen Objekten (gleichen und unterschiedlichen Typs)³⁸ sowie die Zuordnungen zu Organisationseinheiten, Fächern und Forschungsfeldern gesondert hervor.

1. In einem ersten Unterabschnitt sind jeweils die **objektspezifischen Attribute** (bei dem Objekt *Person* sind diese differenziert nach Personentyp – Beschäftigter, Doktorand/Promovierter, Habilitierter) aufgeführt.
2. Im zweiten Unterabschnitt sind die **Verknüpfungen des Objektes** zu einem anderen Objekt (Verknüpfungsattribute) dargestellt.
3. Im dritten Unterabschnitt sind die **Zuordnungen des Objekts** zu Organisationseinheit, Fach sowie zukünftig Forschungsfeld (Zuordnungsattribute) beschrieben.

Die den aufgeführten Objekten, Attributen und Ausprägungen jeweils zugrundeliegenden Definitionen sind Teil der Übersicht der Aggregatdaten.³⁹

³⁷ Die in eckigen Klammern angegebenen Merkmale beziehen sich auf Attribute der *Schale des Kerndatensatz* Forschung.

³⁸ Wie z. B. zwischen zwei Personen (*Doktorand* und *Erstbetreuer*, der z. B. Beschäftigter der Forschungseinrichtung sein kann) oder zwischen Person und Publikation (über z. B. den *Schöpfer*).

³⁹ http://www.kerndatensatz-forschung.de/version1/Spezifikationstabelle_KDSF_v1.html

4 Implementierung des Kerndatensatz Forschung

Die folgenden Abschnitte liefern eine Beschreibung und Erläuterung des technischen Datenmodells (4.1) und generelle Hinweise zur Umsetzung und praktischen Machbarkeit des Kerndatensatz Forschung an Forschungseinrichtungen (Abschnitt 4.2). Das technische Datenmodell ist unter der *Creative Commons* Lizenz CC BY-SA veröffentlicht.⁴⁰

4.1 Lesehinweise für das technische Datenmodell (Basis- und Aggregatdatenmodell)

Die folgende Darstellung versteht sich als exemplarisch: Basis- und Aggregatdaten können in einem einrichtungsinternen Datenmodell repräsentiert werden. Das vorgeschlagene technische Datenmodell (siehe Glossar) ist unterteilt in eine Basisdatenebene und eine Aggregationsebene. Es bildet damit sowohl die Basisdaten, die zur Bereitstellung der definierten Aggregatdaten des Kerndatensatz Forschung herangezogen werden können, als auch die Aggregatdaten selbst ab. Das Modell besteht im Wesentlichen aus der Definition von **Objekten** (entsprechend den Definitionen für Basisdaten und Aggregatdaten) sowie deren **Eigenschaften** (*objektspezifische Attribute*) und **Beziehungen** untereinander (*Verknüpfungs- und Zuordnungsattribute*). Das technische Datenmodell wurde in der Modellierungssprache OWL (Web Ontology Language) erstellt und ist in einem XML-Dokument definiert.

Um die Lesbarkeit zu erleichtern, wurde für das technische Datenmodell eine Online-Dokumentationsplattform entwickelt, die die Darstellung der Definitionen in einem Webbrowser ermöglicht. Die Adresse der Plattform lautet:

www.kerndatensatz-forschung.de/version1/technisches_datenmodell/

Basismodell

Wählt man im Menü der Online-Dokumentationsplattform das *Basismodell* aus, so sind auf der linken Seite unter *Liste der Objekte* die Objekte des technischen Datenmodells dargestellt. Die Objekte entsprechen den *Objekten* der Basisdaten (siehe Abschnitt 3.3), wobei die beiden Objekte *Strukturiertes Promotionsprogramm* und *Drittmittelprojekt* zum Objekt *Projekt bzw. Programm* zusammengefasst wurden. Die folgende Abbildung (Abbildung 5) zeigt ein Beispiel.

Abbildung 5: Darstellung von Beziehungen und Eigenschaften im Basismodell des technischen Datenmodells

KDSF Konzeptuelles Datenmodell Basismodell Aggregationen Liste aller Objekte OWL-Dateien

Liste der Objekte

- Person
- hat Fach
- hat Forschungsfeld
- hat Qualifikation
- Beschäftigte / Beschäftigter
- Projekt bzw. Programm
- Patent
- Publikation
- Forschungsinfrastruktur

Person

ID: <http://kerndatensatz-forschung.de/owl/Basis#Person>

Beschreibung

Zusammenfassendes Konzept für die Darstellung aller personenbezogenen Eigenschaften und Attribute

Beziehungen

- hat Fach zu Fach
- hat Forschungsfeld zu Forschungsfeld
- hat Qualifikation zu Qualifikation

Eigenschaften

- Alter : positivelnteger
- Geschlecht : { "Maennlich", "Ohne Angabe", "Weiblich" }
- Name : string
- Staatsangehoerigkeit : string
- hat Promotionsberechtigung aus : string

Spezialisierungen

- Beschäftigte / Beschäftigter

CERIF Mapping

- Abgeschlossen
- urn:xmlns.org:eurocris:cerif-1.6-2#ctfPers

⁴⁰ <http://creativecommons.org/licenses/by-sa/4.0/deed.de>

Wird im Menü auf der linken Seite ein Objekt ausgewählt, so erscheint im Hauptbereich rechts die Beschreibung des Objekts mit den Beziehungen und Eigenschaften.

Unter **Eigenschaften des Objekts** im Hauptbereich werden die Attribute des jeweiligen Objektes (*objektspezifische Attribute*) aufgelistet. Auch sind dort die **Beziehungen zu anderen Objekten** (*Verknüpfungs- und Zuordnungsattribute*) aufgeführt.⁴¹ Im Beispiel aus Abbildung 5 werden die Details zum Objekt *Person* angezeigt. Wurde ein Objekt explizit definiert, wie etwa *Patent*, wird anstelle der Beschreibung die spezifizierte Definition (siehe auch Aggregatdatenübersicht)⁴² angezeigt mit einem Verweis auf die entsprechende Definitionstabelle.⁴³

Sobald die Angaben zu Eigenschaften oder Beziehungen eines Objekts angeklickt werden, erscheinen die entsprechenden detaillierten Informationen mit entsprechenden Definitionen.

Im Hauptbereich finden sich darüber hinaus Angaben zu den **Spezialisierungen** bzw. **Generalisierungen** des Objekts. Spezialisierungen bzw. Generalisierungen geben an, welche Unter- bzw. Oberklassen es zu dem jeweiligen Objekt gibt (eine Spezialisierung von *Person* ist z. B. ein *Beschäftigter*; umgekehrt ist *Person* eine Generalisierung von *Beschäftigter*). Ebenfalls angegeben ist, sofern vorhanden, das entsprechende Element aus dem CERIF-Datenmodell.

Abbildung 6: Darstellung von Spezialisierungen und Generalisierungen im Basismodell des technischen Datenmodells

The screenshot shows a web interface for the 'KDSF Konzeptuelles Datenmodell'. The top navigation bar includes 'Basismodell', 'Aggregationen', 'Liste aller Objekte', and 'OWL-Dateien'. On the left, a sidebar titled 'Liste der Objekte' contains a tree view with items like 'Person', 'Projekt bzw. Programm', 'hat Forschungsfeld', 'hat Mittelgeber', 'hat Unterprojekt', 'Drittmittelprojekt', 'Strukturiertes Promotionsprogramm', 'Patent', 'Publikation', and 'Forschungsinfrastruktur'. The 'Projekt bzw. Programm' item is selected and expanded. The main content area displays details for 'Projekt bzw. Programm' with the following sections:

- ID:** <http://kerndatensatz-forschung.de/owl/Basis#ProjektProgramm>
- Beschreibung:** Oberklasse für Projekte und Programme, da beide über gemeinsame Eigenschaften (z.B. Sprecher und Antragsteller) verfügen
- Beziehungen:**
 - hat Forschungsfeld zu Forschungsfeld
 - hat Mittelgeber zu Mittelgeber
 - hat Unterprojekt zu Projekt bzw. Programm
- Eigenschaften:**
 - Bewilligungssumme : integer
 - Drittmiteleinahmen : float
 - Drittmittelträge : float
 - Förderkennzeichen : string
 - Koordinationsrolle (Einrichtung) : boolean
 - KoordinatorEinrichtung : string
 - Projektbeginn : dateTime
 - Projektende : dateTime
 - Titel : string
- Spezialisierungen:**
 - Drittmittelprojekt
 - Strukturiertes Promotionsprogramm
- Generalisierungen:**
 - Finanzierte Objekte

In dem in Abbildung 6 dargestellten Beispiel ist ein Teil der Objekthierarchie auf der linken Seite zu erkennen. Die einzelnen Elemente sind durch unterschiedliche *Icons* markiert:

- kennzeichnet ein Objekt mit weiteren Spezialisierungen bzw. Beziehungen zu anderen Objekten. Der Teilbaum der Hierarchie kann durch einen Klick eingeblendet werden. Die jeweiligen Spezialisierungen des Objekts werden dann in der aufgeklappten Liste durch die Symbole ➤ und ● gekennzeichnet.

⁴¹ Die Liste der Beziehungen eines Objekts zu anderen Objekten (Verknüpfungs- und Zuordnungsattribute) lässt sich darüber hinaus auf der linken Seite des technischen Datenmodells in Teilbäumen unter dem jeweiligen Objekt aufklappen. Dort sind die jeweiligen Beziehungen als *Icons* aufgeführt.

⁴² http://www.kerndatensatz-forschung.de/version1/Spezifikationstabelle_KDSF_v1.html

⁴³ Der Identifier wird zum Abschluss des Projekts durch eine URL ersetzt werden, die auf die Dokumentation des Objekts verweist. Der Identifier dient aber gleichzeitig zur Identifikation der Objekte und kann in anderen Datenmodellen oder Systemen als Verweis auf den Kerndatensatz Forschung zur genauen Charakterisierung eines dort benutzten Objekts eingesetzt werden.

- ▼ gibt an, dass die Spezialisierungen des Objekts bzw. Beziehungen des Objekts zu anderen Objekten sichtbar sind.
- stellt ein Objekt ohne weitere Spezialisierungen bzw. Beziehungen zu anderen Objekten dar.
- kennzeichnet eine Beziehung zu anderen Objekten (üblicherweise in der Form „hat ...“). Bei einem Klick erhält man zunächst weitere Informationen zu der Beziehung, wie in der folgenden Abbildung (Abbildung 7) dargestellt ist.

Abbildung 7: Darstellung von Beziehungen im Basismodell des technischen Datenmodells

The screenshot shows the 'KDSF Konzeptuelles Datenmodell' interface. On the left, a 'Liste der Objekte' sidebar shows a hierarchy: Person (expanded), hat Fach, hat Forschungsfeld, hat Qualifikation, Beschäftigte / Beschäftigter (expanded), hat Beschäftigung, Projekt bzw. Programm (expanded), hat Forschungsfeld, hat Mittelgeber, hat Unterprojekt, and Drittmittelprojekt (selected). The main area displays details for the selected relationship 'hat Finanzierungsform' (ID: http://kerndatensatz-forschung.de/owl/Basis#hatFinanzierungsform). It includes definitions for 'Be8: Finanzierungsform' and 'Na14: Finanzierung des Programms', a note about the definition being adapted for the beta version, and sections for 'Definiert für' (Finanzierte Objekte), 'Ziel / Wertebereich' (Finanzierungsform), and 'CERIF Mapping' (Nicht möglich).

Wie in Abbildung 7 dargestellt bezeichnet die Angabe *Definiert für* das Objekt, von dem die jeweilige Beziehung ausgeht, und *Ziel/Wertebereich* das Objekt, auf das verwiesen wird. Im obigen Beispiel werden dazu die entsprechenden Definitionen sowie ein Verweis auf die Definitionstabelle angezeigt.

Aggregatdatenmodell

Wählt man im oberen Menü das Aggregatdatenmodell (im Datenmodell *Aggregationen* genannt), listet die Objekthierarchie auf der linken Seite nur eine Ebene, da keine Ausprägungen bzw. Spezialisierungen definiert werden, sondern nur die Kerndaten (siehe Abschnitt 2.3). Das folgende Beispiel (Abbildung 8) bezieht sich auf das Objekt *Personen*, zu dem die Bereiche *Beschäftigte* und *Nachwuchsförderung* gehören.

Abbildung 8: Darstellung des Aggregationsmodells des technischen Datenmodells

KDSF Konzeptuelles Datenmodell - Aggregationsebene Basismodell Aggregationen Liste aller Objekte OWL-Dateien

Liste der Objekte

- Personen
- Strukturierte Promotionsprogramme
- Drittmittel
- Patente
- Publikationen
- Forschungsinfrastrukturen

Personen

ID: <http://kerndatensatz-forschung.de/owl/Aggregationen#Personen>

Definitionen

- **Be0: Beschäftigte (Empfohlen als Teil des Kerns)**
Beschäftigte umfassen im Sinne des Kerndatensatzes alle Personen, die zum Berichtszeitpunkt in einem Arbeitsvertragsverhältnis/Beamtenverhältnis mit mindestens einer Stunde Wochenarbeitszeit mit der berichtenden Einrichtung stehen.

Die Erfassung erfolgt zum Stichtag gemäß der Hochschulpersonalstatistik von Destatis (1.12.). Alternativ kann auch zum 31.12. erhoben werden, um die aggregierten Informationen zu Beschäftigten mit anderen Bereichen vergleichbar zu machen.

Eine Übersicht von Erhebungszeitpunkten und -zeiträumen ist unter http://www.forschungsinform.de/kerndatensatz/version1/Berichtszeitraum_KDSF_v1.pdf einsehbar.

Eigenschaften

- Anzahl Absolventinnen / Absolventen : nonNegativeInteger
- Anzahl Doktorandinnen / Doktoranden : nonNegativeInteger
- Anzahl Personen (Kopfzahlen) : nonNegativeInteger
- Vollzeitäquivalente : float
- Zahl der Doktoranden mit ausländischer Staatsbürgerschaft : nonNegativeInteger
- Zahl der Doktoranden mit inländischer Staatsbürgerschaft : nonNegativeInteger
- Zahl der männlichen Doktoranden : nonNegativeInteger
- Zahl der weiblichen Doktorandinnen : nonNegativeInteger

Generalisierungen

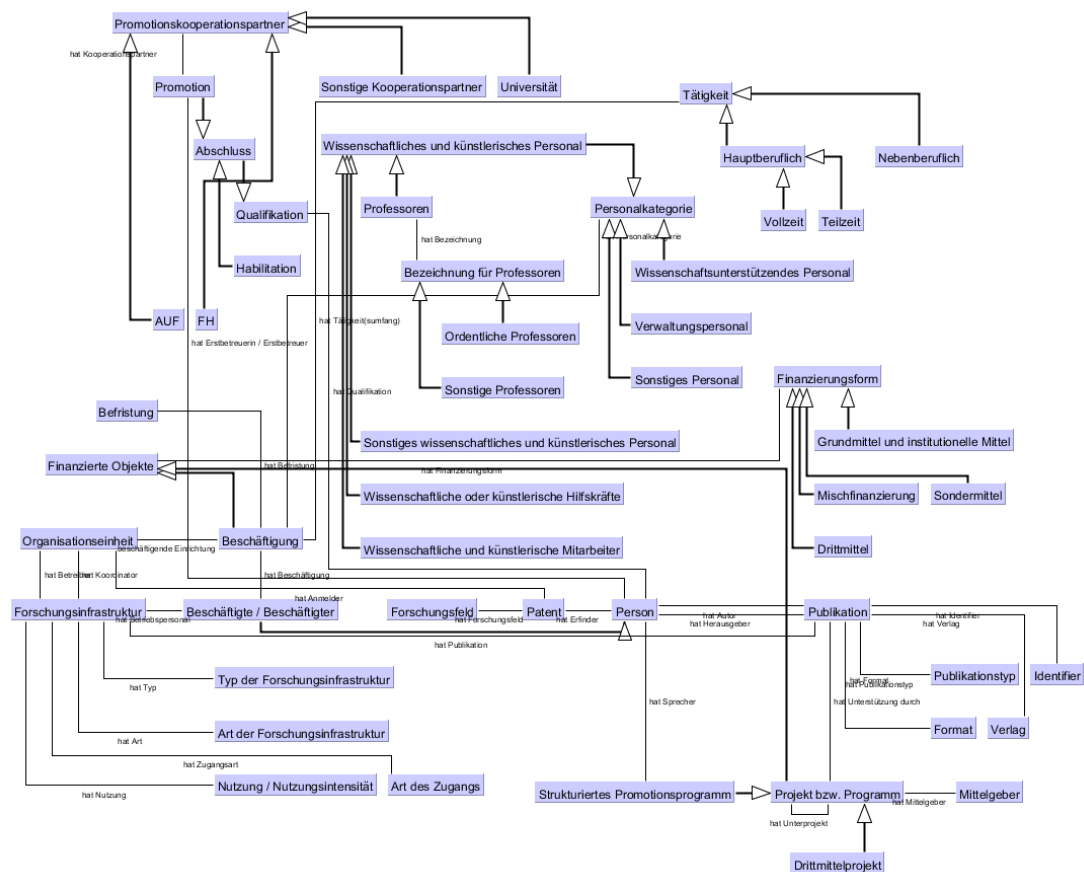
- Kerndaten

Unter *Beziehungen* finden sich in dieser Darstellung die zum Objekt gehörenden Kerndaten in Listenform. Unter *Eigenschaften* wiederum sind die jeweiligen Kerndaten in Summenform aufgeführt.

Darüber hinaus kann man auf der Titelleiste der Webseite unter *Liste aller Objekte* eine Druckversion ansehen, in der alle Objekte unsortiert aufgeführt sind. Unter *OWL-Dateien* kann ferner auf die XML-Dateien mit der formalen Definition der Datenmodelle zugegriffen werden.

Mit Version 1.0 der Spezifikation des Kerndatensatz Forschung wurde aus dem Datenmodell für die Aggregationsebene auch ein XML-Schema erstellt, das als Basis für den Datenaustausch zwischen Forschungseinrichtungen und abfragenden Stellen eingesetzt werden kann.

Abbildung 9: Ausschnitt des Basisdatenmodells



4.2 Abgleich zwischen dem technischen Datenmodell des Kerndatensatz Forschung und CERIF

Ziel des Abgleichs des technischen Datenmodells für den Kerndatensatz Forschung mit dem europäischen CERIF-Standard ist es, die Implementierung des Kerndatensatzes zu unterstützen, da auch aktuelle Forschungsinformationssysteme eine Unterstützung für CERIF als Datenmodell anbieten bzw. mit CERIF kompatibel sind. Daher ist eine Verbindung des Kerndatensatz Forschung mit CERIF eine wichtige Voraussetzung für die Akzeptanz des Kerndatensatz Forschung insgesamt.

Ein Ziel des Projekts Kerndatensatz Forschung war die Erarbeitung eines mit CERIF kompatiblen Datenmodells. Jedoch ist der Kerndatensatz Forschung in weiten Bereichen wesentlich spezifischer als CERIF, da z. B. die (spezifisch deutschen) Ausprägungen in Hinblick auf Personalkategorien oder Publikationstypen in diesem Detailierungsgrad in CERIF nicht abgebildet werden. Der Kerndatensatz Forschung ist also detaillierter als CERIF. Dies war insofern zu erwarten, da CERIF als europäischer Standard nicht die Spezifika aller EU-Länder aufnehmen kann. Insbesondere sind in CERIF die für das deutsche Forschungssystem spezifischen Ausprägungen nicht vorhanden (z. B. die unterschiedlichen Drittmittelgeber wie *Bund*, *Bundesländer*, *DFG*, etc.). Diese Informationen können zwar in CERIF dargestellt werden, da nahezu alle Objekte und Beziehungen in CERIF klassifiziert werden können.⁴⁴ Die Ausprägungen der Klassifikationsschemata sind

⁴⁴ In CERIF ist dafür das Konzept „cfClass“ bzw. „cfClassScheme“ vorgesehen. Diese Klassifikationen sind nicht zu verwechseln mit der im Kerndatensatz Forschung vorgesehenen Fächerklassifikation, da in CERIF Klassifikationsschemata für beliebige Bereiche definiert werden können. So wäre es auch möglich, Klassifikationsschemata für Personalkategorien, Publikationstypen oder Fördertypen entsprechend der deutschen Forschungslandschaft zu definieren.

jedoch nicht standardisiert, es gibt lediglich Vorgaben für einige Bereiche. In diesem Punkt wird ein wesentliches Problem bei der Verwendung eines CERIF-kompatiblen Datensatzes für den Kerndatensatz Forschung bestehen. Für eine Übertragung der Daten aus einem „CERIF-kompatiblen“ Forschungsinformationssystem in den Kerndatensatz Forschung müsste das Forschungsinformationssystem die gleichen Klassifikationsschemata verwenden wie sie im Kerndatensatz vorgesehen sind. Eine solche Annahme ist aber aufgrund der heterogenen IT-Landschaft in Forschungseinrichtungen und der vorhandenen Altdaten nicht realistisch. Um die Verwendung des Kerndatensatzes in CERIF-kompatiblen Systemen zu vereinfachen, wurden im Rahmen des Projekts dennoch entsprechende Klassifikationsschemata definiert. Diese sind als Ergänzung zum technischen Datenmodell verfügbar und auf der Webseite zum technischen Datenmodell veröffentlicht.⁴⁵ Sie erleichtern den Forschungseinrichtungen das *Mapping* ihrer bestehenden Klassifikationsschemata auf die im Kerndatensatz Forschung benutzten Klassifikationsschemata.

Neben den Klassifikationsschemata müssen aber auch die grundlegenden Elemente der Datenmodelle miteinander verglichen werden. Der Vergleich des Kerndatensatz Forschung mit dem Datenmodell von CERIF erfolgte durch die Projektgruppe „Technik, Schnittstellen und Datenformate“ in mehreren Schritten. Das *Mapping* steht ebenfalls als Ergänzung zum technischen Datenmodell auf der entsprechenden Webseite⁴⁶ zur Verfügung.

⁴⁵ www.kerndatensatz-forschung.de/version1/technisches_datenmodell/

⁴⁶ *Ibid.*

5 Glossar

Aggregatdaten

Entscheidet sich eine Einrichtung, Daten gemäß dem Kerndatensatz-Standard weiter zu geben, so gibt sie Aggregatdaten oder Daten in Listenform weiter (zur Vereinfachung der Darstellung wird im gesamten Dokument für beide Arten von Daten übergreifend der Begriff „Aggregatdaten“ verwendet). Diese Daten fassen Sachverhalte z. B. für ein Institut oder ein Fach zusammen. In der Terminologie des Wissenschaftsrates aus den „Empfehlungen zu einem Kerndatensatz Forschung“ entsprechen Aggregatdaten den „Kerndaten“.

Aggregatdatenübersicht

Die hinterlegte Aggregatdatenübersicht (siehe http://www.kerndatensatzforschung.de/version1/Spezifikationstabelle_KDSF_v1.html) ist unterteilt in ein Überblicksverzeichnis in hierarchischer Listenform und einen Definitionskatalog, der die Definitionen aller empfohlenen → **Elemente** enthält. Die Aggregatdatenübersicht enthält Informationen zu den → Elementen des → **Kerndatensatzes** und der → **Schale (des Kerndatensatzes)**. Die Elemente sind grün hervorgehoben, sofern eine Aufnahme als Teil des *Kerndatensatzes* empfohlen wird. Violett hervorgehobene Elemente sind empfohlen als *Schale des Kerndatensatzes*.

Aggregationsniveau

Die Kerndaten beziehen sich immer auf bestimmte Aggregationsniveaus (in Bezug auf Kerndaten in Summenform) bzw. → **Gruppierungsniveaus** (in Bezug auf Kerndaten in Listenform). Das Aggregations- bzw. Gruppierungsniveau bestimmt, auf welcher Ebene die Aufsummierung oder Auflistung der Daten erfolgen soll. So ist beispielsweise die Anzahl der Beschäftigten eines → **Fachs** (z. B. Chemie) oder einer → **Organisationseinheit** (z. B. Institut für Prozesstechnik) zu berichten. In einem weiteren Berichtsbeispiel könnten Publikationen nach *Fach* oder *Publikationstyp* (Bücher, Artikel etc.) gelistet werden. Die möglichen Aggregationsniveaus sind zusammen mit den Kerndaten in der Aggregatdatenübersicht gelistet. Zur Vereinfachung der Darstellung wird im vorliegenden Dokument übergreifend der Begriff „Aggregationsniveau“ verwendet, auch wenn er sich auf die Gruppierung von Kerndaten in Listenform bezieht.

Attribut

Für die → **Objekte** (Basisdaten) sollen unterschiedliche Attribute erhoben werden, die jeweils verschiedene Ausprägungen haben können. So werden beispielsweise für alle Personen die Attribute *Geschlecht* (mit den Ausprägungen „männlich“, „weiblich“ oder „ohne Angabe“), *Staatsangehörigkeit* (Angabe des Lands/der Länder der Staatsangehörigkeit) sowie *Geburtsdatum* (exakte Angabe) erfasst. Die Basisdatenübersicht (ist nach drei unterschiedlichen Typen von Attributen strukturiert: (1) → **Objektspezifische Attribute**, (2) → **Verknüpfungsattribute**, die Verknüpfungen des Objektes zu anderen Objekten beinhalten, (3) → **Zuordnungsattribute**, die die Zuordnung des Objekts zu Organisationseinheit, Fach sowie zukünftig Forschungsfeld spezifizieren.

Ausdifferenzierung

Ein Attribut ist Teil eines → **Kerndatums** und beschreibt relevante Merkmale desselben, z. B. bei Personal das Geschlecht. Ein Attribut hat verschiedene → **Ausprägungen**.

Ausprägung

Ausprägungen sind die Werte, die ein → **Attribut** oder eine → **Ausdifferenzierung** annehmen kann (z. B. *Geschlecht*: „männlich“, „weiblich“, „ohne Angabe“; *ISBN*: entsprechende Nummer).

Basisdaten

Zur Weitergabe der verschiedenen → **Aggregatdaten** ist die einrichtungsinterne Erfassung von Basisdaten erforderlich.

Bereich

Die Bereiche des → **Kerndatensatzes** umfassen Beschäftigte, Nachwuchsförderung, Drittmittel und Finanzen, Patente sowie Publikationen. Die → **Schale (des Kerndatensatzes)** umfasst darüber hinaus den Bereich Forschungsinfrastrukturen.

Berichtslegung

Für die Ermittlung von → **Kerndaten** werden alle Einheiten eines → **Objekts** (je nach zu berichtenden → **Aggregationsniveaus** bzw. → **Gruppierungsniveaus** und → **Ausdifferenzierungen**) für den Berichtszeitraum aufsummiert bzw. aufgelistet. Die Berichtslegung gemäß Kerndatensatz erfolgt je nach Bereich und Kerndatum zum Stichtag oder nach Kalenderjahr (siehe http://www.kerndatensatz-forschung.de/version1/Berichtszeitraum_KDSF_v1.pdf).

CERIF

CERIF (kurz für *Common European Research Information Format*) ist ein europäisches Datenaustauschformat für Forschungsinformationen, das eine formale Syntax und Semantik für die Abbildung unterschiedlicher Elemente (Forschungsinformationen) bereitstellt.

Definition

Die mit dem Kerndatensatz vorgelegten Definitionen machen für den Datenaustausch nach diesem Standard verbindliche Setzungen hinsichtlich der Inhalte oder der Aggregation von → **Basisdaten**.

Einführung

Die Einführung des Kerndatensatzes bezeichnet die Begleitung und Unterstützung der → **Implementierung** an mehreren Forschungseinrichtungen.

Element

Die Aggregatdaten des → **Kerndatensatzes** oder der → **Schale (des Kerndatensatzes)** umfassen verschiedene Elemente unterschiedlichen Typs. Ein Bericht umfasst die Darstellung von (1.) → **Kerndaten** nach (2.) verschiedenen → **Ausdifferenzierungen**. Dabei können die Kerndaten nach (3.) unterschiedlichen → **Aggregationsniveaus** aggregiert bzw. gruppiert werden.

Empfohlen für die Weiterentwicklung

Über den → **Kerndatensatz** und die → **Schale (des Kerndatensatzes)** hinaus hat sich das Projekt auch mit Bereichen und Elementen befasst, die in der zukünftigen → **Weiterentwicklung** behandelt und ausdefiniert werden sollten. Im vorliegenden Dokument werden die Projektempfehlungen allerdings ausschließlich in Hinblick auf den *Kerndatensatz* (grün gekennzeichnet) und die *Schale des Kerndatensatzes* (violett gekennzeichnet) dargestellt.

Fach

Ein Fach bezeichnet im Kontext des Kerndatensatz Forschung einen thematisch durch die Fächersystematiken des Statistischen Bundesamtes abgegrenzten Gegenstandsbereich. Technisch präziser sollte von Fächergruppen, Lehr- und Forschungsbereichen sowie Fachgebieten gesprochen werden.

Forschungsfeld

Ein Forschungsfeld umfasst eine Menge von Forschungsaktivitäten, die sich problemorientiert mit einem thematisch abgrenzbaren Gegenstandsbereich, der quer zu Disziplinen liegt, befassen.

Forschungsinformationssystem

Ein Forschungsinformationssystem verbindet eine Vielzahl von Datenquellen zu Forschungsaktivitäten und Kontextinformationen und stellt eine Schnittstelle für standardisierte Abfragen zur Verfügung.

Gruppierungsniveau

Die Kerndaten beziehen sich immer auf bestimmte → **Aggregationsniveaus** (in Bezug auf Kerndaten in Summenform) bzw. Gruppierungsniveaus (in Bezug auf Kerndaten in Listenform). Das Aggregations- bzw. Gruppierungsniveau bestimmt, auf welcher Ebene die Aufsummierung oder Auflistung der Daten erfolgen soll. So ist beispielweise die Anzahl der Beschäftigten eines → **Fachs** (z. B. Chemie) oder einer → **Organisationseinheit** (z. B. Institut für Prozesstechnik) zu berichten. In einem weiteren Berichtsbeispiel könnten Publikationen nach *Fach* oder *Publikationstyp* (Bücher, Artikel etc.) gelistet werden. Die möglichen Aggregationsniveaus sind zusammen mit den Kerndaten in der Aggregatdatenübersicht gelistet. Zur Vereinfachung der Darstellung wird im vorliegenden Dokument übergreifend der Begriff „Aggregationsniveau“ verwendet, auch wenn er sich auf die Gruppierung von Kerndaten in Listenform bezieht.

Implementierung

Die Implementierung des Kerndatensatz Forschung bezeichnet den Prozess der Übernahme der Definitionen und Formate in die Berichtsprozesse einer Forschungseinrichtung mit dem Ziel, Daten gemäß der Spezifikation auszutauschen.

Kennzahl

Eine Kennzahl kann auf der Basis mehrerer → **Aggregatdaten** gebildet werden (z. B. Publikationen pro Wissenschaftlerin pro Jahr). Kennzahlen sind nicht Gegenstand des Kerndatensatzes. Es werden auch keine Hinweise zu ihrer Bildung formuliert.

Kerndatensatz

Der Kerndatensatz umfasst die Elemente, die der Wissenschaftsrat für alle Einrichtungen im Wissenschaftssystem als relevant erachtet. Der Wissenschaftsrat empfiehlt allen Einrichtungen, den Kerndatensatz nach den spezifizierten Definitionen vorzuhalten. Diese Daten sind für die Forschungsberichterstattung relevant und können auf Anfrage in geeigneter Aggregation sowie unter Beachtung rechtlicher und vertraglicher Einschränkungen weitergegeben werden.

Kerndatum

Ein Kerndatum bezeichnet ein → **Element** des → **Kerndatensatzes** oder seiner → **Schale** (z. B. Patente).

Objekt

Die Basisdaten des → **Kerndatensatzes** und seiner → **Schale** sind nach Objekten strukturiert. Sie umfassen Personen (darunter fallen Beschäftigte, Doktoranden, Promovierte und Habilitierte), Strukturierte Promotionsprogramme, Drittmittelprojekte, Patente, Ausgründungen, Publikationen sowie Forschungsinfrastrukturen.

Objektspezifisches Attribut

Objektspezifische Attribute stellen einen Typ von → **Attributen** für die → **Objekte** der → **Basisdaten** dar. Ihre Erfassung unterscheidet sich je nach Objekt.

Organisationseinheit

Als Organisationseinheit gilt im Kontext des Kerndatensatz Forschung, Destatis folgend, jeweils die kleinste, an der Einrichtung bestehende organisatorische Einheit, der die Kerndaten zugeordnet sind.

Schale (des Kerndatensatzes)

Ergänzend zum → **Kerndatensatz** sind die in der Schale des Kerndatensatzes aufgeführten → **Elemente** nur für spezifische Einrichtungen relevant. Auch für diese Elemente empfiehlt der Wissenschaftsrat, die getroffenen Definitionen zu verwenden, damit die Einrichtungen in vergleichbarer Form dazu auskunftsfähig sind. Teilweise macht der Wissenschaftsrat im Rahmen der Definitionen auch Vorschläge, für welche Einrichtungen er diese Elemente für sinnvoll erachtet.

Technisches Datenmodell

Das technische Datenmodell ist unterteilt in eine Basisdatenebene und eine Aggregationsebene und bildet damit die → **Basisdaten** und die → **Aggregatdaten** des → **Kerndatensatzes** sowie der → **Schale (des Kerndatensatzes)** ab. Es besteht im Wesentlichen aus der Definition von → **Objekten** sowie deren Eigenschaften (→ **Objektspezifische Attribute**) und Beziehungen untereinander (→ **Verknüpfungsattribute** und → **Zuordnungsattribute**). Das technische Datenmodell wurde in der Modellierungssprache OWL (*Web Ontology Language*) erstellt und ist in einem XML-Dokument definiert.

Verknüpfungsattribut

Verknüpfungsattribute stellen einen Typ von → **Attributen** für die → **Objekte** der → **Basisdaten** dar. Sie beinhalten die Verknüpfungen eines Objektes zu anderen Objekten.

Weiterentwicklung

Die Weiterentwicklung des Kerndatensatzes bezeichnet den Prozess zur Anpassung und Ergänzung der vorgelegten Definitionen mit dem Ziel, den Standard auch an zukünftigen Bedarfen seiner Nutzer ausrichten zu können. Als Weiterentwicklungsempfehlung des Kerndatensatzes gelten solche Elemente für die im Zuge der nun vorliegenden Spezifikation noch keine abschließende Definition vorgeschlagen werden konnte und die daher im Zuge einer Weiterentwicklung standardisiert werden sollten.

Zuordnungsattribut

Zuordnungsattribute stellen einen Typ von → **Attributen** für die → **Objekte** der → **Basisdaten** dar. Sie spezifizieren die Zuordnung des Objekts zu → **Organisationseinheit**, → **Fach** sowie zukünftig → **Forschungsfeld**.

6 Anhang: Übersicht der Basisdaten (*Kerndatensatz Forschung und Schale des Kerndatensatz Forschung*)

Lesehinweise

- Die im Folgenden dargestellten Basisdaten für unterschiedliche Objekte ermöglichen die Berechnung der in der Aggregatdatenübersicht (siehe Kapitel 2; siehe auch http://www.kerndatensatz-forschung.de/version1/Spezifikationstabelle_KDSF_v1.html) dargestellten Kerndaten mit Ausnahme der im Bereich *Drittmittel und Finanzen* vorgesehenen Kerndaten „Drittmitteleinnahmen (Dr1a)“, „Drittmittelerträge (Dr1b)“, „Sonstige Einnahmen (Dr139)“ und „Gesamtbudget (Dr137)“ sowie das im Bereich *Patente und Ausgründungen* vorgesehenen Kerndatum „Erträge aus Schutzrechten (Pa34)“.
- *Ausprägungen* für die jeweiligen *Attribute* sind nicht in allen Fällen disjunkt. Es sind für bestimmte Attribute Mehrfacherfassungen vorgesehen (siehe Abschnitt 3.2).
- Attribute sowie Ausprägungen für die Objekte (Personen, Strukturierte Promotionsprogramme, Drittmittelprojekte, Patente, Ausgründungen, Publikationen, Forschungsinfrastrukturen) sind in Bezug auf das jeweilige Objekte zu erfassen. Abweichungen von dieser Regel (z. B. Erfassung von „Personalkategorie (Be11)“, Anteil an Vollzeitstelle sowie Beginn und Ende des Arbeitsvertrags in Bezug auf den *Arbeitsvertrag* von *Personen* oder Erfassung von „Förderkennzeichen (Dr13)“ und „Bewilligungssumme (Dr88)“ in Bezug auf den „*Drittmittelgeber (Dr12)*“ bei *Drittmittelprojekten*) sind explizit als solche gekennzeichnet und hervorgehoben.

I. Person

0. Personentyp

- Beschäftigter (Be0)
- Wissenschaftlicher Nachwuchs (Na0): Person in oder mit Abschluss einer Qualifikationsphase im Berichtszeitraum [Doktorand (Na46 und Na101), Promovierter (Na43 und Na4a) oder Habilitierter (Na44 und Na4a)]

1. Objektspezifische Attribute für Personen aller Personentypen [Beschäftigte (Be0), Doktoranden (Na46 und Na38a), Promovierte (Na43 und Na4a) sowie Habilitierte (Na44 und Na4a)]

- (einrichtungsinterne) ID für Personen
- Geschlecht (Be7 bzw. Na7) (Ausprägungen: Männlich, Weiblich, ohne Angabe)
- Staatsangehörigkeit (Be72 bzw. Na58) (*Land/Länder aus Liste*)
- Geburtsdatum (Be52 bzw. Na69 bzw. Na71) (*exakte Angabe*)

1.1 Zusätzliche objektspezifische Attribute für Beschäftigte (Be0)

- **Personalkategorie** (Be11) **je parallelen Arbeitsvertrag** (Ausprägungen: Wissenschaftliches und künstlerisches Personal, Wissenschaftsunterstützendes Personal, Verwaltungspersonal, Sonstiges Personal)
 - Unterausprägungen für Wissenschaftliches und künstlerisches Personal (Be68) (Ausprägungen: Professoren, Wissenschaftliche und künstlerische Mitarbeiter, Sonstiges wissenschaftliches und künstlerisches Personal, Wissenschaftliche oder künstlerische Hilfskräfte)
 - Attribute für Professor (Be19):

- Besoldung / Vergütung (Be40) (Ausprägungen: W1, W2/C3, W3/C4, C2, B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11),
- Bezeichnung (Be41) (Ausprägungen: Professoren, Juniorprofessoren, Sonstige Professoren),
- Gemeinsame Berufung (Be4a) (Ausprägungen: ja, nein)
 - Attribute für Gemeinsame Berufung (Be4a):
 - Name der weiteren Einrichtung (Be12) (Name der Einrichtung *aus Liste*)
 - Unterausprägungen für Sonstige Professoren (Be88) (Ausprägungen: Stiftungsprofessoren, Vertretungsprofessoren / Lehrstuhlvertreter, Außerplanmäßige Professoren, Seniorprofessoren, Hauptberufliche Gastprofessoren)
- **Anteil an Vollzeitstelle und Beginn sowie Ende des Arbeitsvertrags** (Datum) **je parallelen Arbeitsvertrag** (*Angabe in Prozent bzw. exakte Datumsangabe*)
- **Finanzierungsform** (Be8) (Ausprägungen: Grundmittel / Institutionelle Mittel, Drittmittel, Mischfinanzierung, Sondermittel)
- **Qualifikation** (Be9) (Ausprägungen: Habilitiert, Promoviert)
- **Tätigkeitsart** (Be32) (Ausprägungen: Hauptberuflich / Im Hauptamt, Nebenberuflich / Im Nebenamt)
 - Unterausprägungen für Hauptberuflich / Im Hauptamt (Be35) (Ausprägungen: Vollzeit, Teilzeit)

1.2 Zusätzliche objektspezifische Attribute für Doktoranden und Promovierte (Na46 und Na101 bzw. Na43 und Na4a)

- **Start der Promotion** (Na38a bzw. Na102) (*Datum*)
- **Zeitpunkt des Abschlusses** des Promotionsverfahrens (Na70) (*Datum*)
- **Kooperation** (Na86) (Ausprägungen: AUF, FH, Universität, Sonstige)
- **Land des zur Aufnahme einer Promotion berechtigenden Studienabschlusses** (Na95) (*Land/Länder aus Liste*)

1.3 Zusätzliche objektspezifische Attribute für Habilitierte (Na44 und Na4a)

- Zeitpunkt des Abschlusses des Habilitationsverfahrens (Na72) (*Datum*)

2. Verknüpfungen mit anderen Objekten

- Verknüpfung von Doktorand sowie Promovierten (Na46 und Na101 bzw. Na43 und Na4a) mit Beschäftigtem (Be0) (→ I. Person) über einrichtungsinterne Personen-ID:
 - **Name(n) Erstbetreuer** (Na47) [nur eine Person als Erstbetreuer je Doktorand und Promovierten zu erfassen]
- Verknüpfung von Doktorand sowie Promovierten (Na46 und Na101 bzw. Na43 und Na4a) mit Strukturiertem Promotionsprogramm (→ II. Strukturiertes Promotionsprogramm) über einrichtungsinterne Personen- bzw. Programm-ID:
 - **Titel Strukturiertes Promotionsprogramm** (Na48) [nur ein Strukturiertes Promotionsprogramm je Doktorand und Promovierten zu erfassen]
- Verknüpfung Beschäftigter (Be0), Doktorand (Na46 und Na101), Promovierter (Na43 und Na4a) sowie Habilitierter (Na44 und Na4a) und Strukturiertes Promotionsprogramm, Drittmittelprojekt, Patent sowie Publikation über einrichtungsinterne Personen-, Programm-, Projekt- Patent- bzw. Publikations-ID (Details siehe jeweils unten):
 - **Sprecher** (Na17) des Strukturierten Promotionsprogramms (→ II: Strukturiertes Promotionsprogramm)
 - **Erfinder** (Pa5) des Patents (→ IV. Patent)

- Schöpfer (Pu2) der Publikation (→ V. Publikation)

3. Zuordnung zu Organisationseinheit, Fach sowie Forschungsfeld

- Organisationseinheit (Be75 bzw. Na92)
- Fach (Be38 bzw. Na2)
- Forschungsfeld (Be77 bzw. Na66)

II: Strukturiertes Promotionsprogramm (Na6)

1. Objektspezifische Attribute für Strukturiertes Promotionsprogramm

- (einrichtungsinterne) ID für Strukturierte Promotionsprogramme
- **Titel** (Na13) (*Freitextfeld*)
- **Beteiligte Institutionen** (Na60) (*Name(n) der deutschen Einrichtung(en) aus Liste, Name der ausländischen Einrichtung(en) über Freitextfeld*)
 - Attribut für Beteiligte Institutionen (Na60): Land der Institution (Na82) (*Land/Länder aus Liste*)
- **Finanzierung des Programms** (Na14) (Ausprägungen: Grundmittel / Institutionelle Mittel, Drittmittel, Mischfinanzierung, Sondermittel)

2. Verknüpfungen mit anderen Objekten

- Verknüpfung mit Person (Beschäftigter (Be0), Doktorand (Na46 und Na38a), Promovierter (Na43 und Na4a) oder Habilitierter (Na44 und Na4a)) (→ I. Person über einrichtungsinterne Personen- bzw. Programm-ID (s. o.):
 - *Name(n) Sprecher* (Na17)
 - *Name(n) zugehöriger Doktorand oder Promovierter* (Na46, Na101, Na43, Na4a) [mehrere Doktoranden und Promovierte je Strukturiertes Promotionsprogramm erfassbar]

3. Zuordnung zu Organisationseinheit, Fach sowie Forschungsfeld

- Organisationseinheit (Na92)
- Fach (Na2)
- Forschungsfeld (Na66)

III. Drittmittelprojekt (Förderphasen als eigene Projekte) (Dr30)

1. Objektspezifische Attribute für Drittmittelprojekt (Förderphasen als eigene Projekte)

- (einrichtungsinterne) ID für Drittmittelprojekte
- **Titel des Projekts** (Dr9) (*Freitextfeld*)
- **Koordinationsrolle** (Einrichtung) (Dr142) (Ausprägungen: ja, nein)
- *Bei Verbundprojekten: Name der Sprechereinrichtung / des Konsortialführers* (Dr19a) (*Name der deutschen Einrichtung aus Liste, Name der ausländischen Einrichtung über Freitextfeld*)
- **Datum des Projektbeginns** (Dr15a) (*exakte Angabe*)
- **Datum des Projektendes** (Dr14) (*exakte Angabe*)
- **Drittmittelgeber** (Dr12) (Ausprägungen: Drittmittel EU und sonstige internationale Organisationen, DFG, Bund, Bundesländer, Sonstige Drittmittelgeber, Gewerbliche Wirtschaft und sonstige private Bereiche, Nicht erklärt)
 - Unterausprägungen für Drittmittel EU und sonstige internationale Organisationen (Dr20) (Ausprägungen: EU, Sonstige internationale Organisationen)
 - Unterausprägungen für Nicht erklärt (Dr134) (Ausprägungen: Nicht erklärt (Private Mittelgeber), Nicht erklärt (Öffentliche Mittelgeber))

- Förderkennzeichen (Dr13) **je Drittmittelgeber** (Freitextfeld)
- Bewilligungssumme (Dr88) **je Drittmittelgeber** (Angabe in EUR)
- Titel des übergeordneten Projektes (Dr10) (Freitextfeld)

2. Verknüpfungen mit anderen Objekten

- Verknüpfung mit Publikation (→ V. Publikation) über einrichtungsinterne Projekt- bzw. Publikations-ID (s. u.) sowie
 - Drittmittelgeber (Dr12) bzw. Förderer (Pu86, s. u.)
 - Förderkennzeichen (Dr13 bzw. Pu90, s. u.)

3. Zuordnung zu Organisationseinheit, Fach sowie Forschungsfeld

- Organisationseinheit (Dr121)
- Fach (Dr28)
- Forschungsfeld (Dr124)

IV. Patent (Pa25)

1. Objektspezifische Attribute für Patent

- (einrichtungsinterne) ID für Patente
- Titel des prioritätsbegründenden Patents (Pa7) (Freitextfeld)
- Datum der prioritätsbegründenden Erstanmeldung (Pa10) (exakte Angabe)
- Veröffentlichungsnummer (Pa33) (Freitextfeld)
- Patentfamilie (Pa38) (Freitextfeld)
- Datum der Erteilung (Pa4) (exakte Angabe)

2. Verknüpfungen mit anderen Objekten

- Verknüpfung mit Person (Beschäftigter (Be0), Doktorand (Na46 und Na38a), Promovierter (Na43 und Na4a) oder Habilitierter (Na44 und Na4a)) (→ I. Person) über einrichtungsinterne Personen- bzw. Patent-ID (s. o.):
 - Name(n) Erfinder (Pa5)

3. Zuordnung zu Organisationseinheit, Fach sowie Forschungsfeld

- Organisationseinheit (Pa35)
- Fach (Pa1)
- Forschungsfeld (Pa36)

V. Ausgründung (Pa39)

1. Objektspezifische Attribute für Patent

- (einrichtungsinterne) ID für Ausgründungen
- Name der Ausgründung (Freitextfeld)
- Datum der Ausgründung (exakte Angabe)

3. Zuordnung zu Organisationseinheit, Fach sowie Forschungsfeld

- Organisationseinheit (Pa35)
- Fach (Pa1)
- Forschungsfeld (Pa36)

VI. Publikation (Pu1a)

1. Objektspezifische Attribute für alle Publikationen

- **Schöpfertyp** (Pu55) (Ausprägungen: Autor, Herausgeber, Körperschaft mit Autorenfunktion, Gruppe mit Autorenfunktion, Körperschaft mit Herausgeberfunktion, Gruppe mit Herausgeberfunktion, Interviewter)
- **Titel des Werks** (Pu5) (*Freitextfeld*)
- **Veröffentlichungsjahr** (Pu84) (*exakte Angabe*)
- **Verlag** (Pu67) (*Freitextfeld*)
 - *Attribut für Verlag (Pu67)*: Verlagsort (Pu153) (*Freitextfeld*)
- **Quelle** (Pu143) (*Freitextfeld; Zeitschriftentitel, Sammelwerk etc.*)
- **Identifizier** (Pu132)
 - *Attribut für Identifizier (Pu132)*: Bezeichnung des Identifiziers (Pu133) (*aus Liste*)
- **Format**: Band (Pu123) (*Freitextfeld*), Heft (Pu91) (*Freitextfeld*), Name der Konferenz (Pu94) (*Freitextfeld*), Seitenbereich (Pu145) (*Freitextfeld*)
- **Sprachcode** (Pu95) (*aus Liste*)
- **Zugangsrechte** (Pu19) (*Freitextfeld, Lizenzen bzw. Zugriffsrechte*)
- **Peer-Reviewed** (Pu104) (Ausprägungen: ja, nein)
- **Qualifikationsschrift** (Pu146) (Ausprägungen: Dissertation, Habilitation)
- **Dokumenttyp** (101) (Ausprägungen: Editorial, Wissenschaftlicher Artikel, Review, Bibliographie, Rezension, Quellenedition, Letter to the Editor, Meeting Abstract)
- **Publikationstyp** (Pu6) (Ausprägungen: Buch, Artikel, Software, Sonderheft einer Zeitschrift, Forschungsdaten, Konferenzposter, Beiträge/Interviews in nicht-wissenschaftlichen Medien, Wissenschaftliche Vortragsfolien, Arbeitspapier/Forschungsbericht, Neuer Publikations- oder Dokumenttyp)
 - *Unterausprägungen für Buch (Pu22)* (Ausprägungen: Monographie, Sammelband, Quellenedition)
 - *Unterausprägungen für Artikel (Pu25)* (Ausprägungen: Journalartikel, ePaper, Sammelbandbeitrag)
- **Ressource** (Pu102) (Ausprägungen: Text, Bild, Audio, Bewegte Bilder)

2. Verknüpfungen mit anderen Objekten

- *Verknüpfung mit Person (Beschäftigter (Be0), Doktorand (Na46 und Na101), Promovierter (Na43 und Na4a) oder Habilitierter (Na44 und Na4a))* (→ I. Person) über einrichtungsinterne Personen- bzw. Publikations-ID (s. o.):
 - *Name(n) Schöpfer* (Pu2)
- *Verknüpfung mit Drittmittelprojekt* (→ III. *Drittmittelprojekt*) über einrichtungsinterne Projekt- bzw. Publikations-ID (s. o.) sowie
 - **Förderer** (Pu86) bzw. Drittmittelgeber (Dr12, s. o.)
 - **Förderkennzeichen** (Pu90 bzw. Dr13, s.o.)
- *Verknüpfung mit Forschungsinfrastruktur* (→ VII. *Forschungsinfrastruktur*) über einrichtungsinterne Publikations- bzw. Forschungsinfrastruktur-ID (s. u.):
 - *Name(n) Forschungsinfrastrukturen* (Fi2)

3. Zuordnung zu Organisationseinheit, Fach sowie Forschungsfeld

- Organisationseinheit (Pu140)
- Fach (Pu138)
- Forschungsfeld (Pu141)

VII. Forschungsinfrastruktur (Fi2)

1. Objektspezifische Attribute für alle Forschungsinfrastrukturen

- (einrichtungsinterne) ID für Forschungsinfrastrukturen
- **Bezeichnung** (Fi3) (*Freitextfeld*)
- **Beschreibung** (Fi27) (*Freitextfeld*)
- **Typ der Forschungsinfrastruktur** (Fi4) (Ausprägungen: Großgeräte und Instrumente, Wissensressourcen, Informations- und Kommunikationsinfrastrukturen, Sonstiger Typ)
- **Art der Forschungsinfrastruktur** (Fi5) (Ausprägungen: Lokal, Verteilt, Virtuell)
- **Art des Zugangs** (Ausprägungen: User Access, Shared Access, Open Access)
- **Nutzung/Nutzungsintensität**: Anzahl Nutzer (Fi7) (*exakte Angabe*), Anzahl genutzte Stunden (Fi28) (*exakte Angabe*), Anzahl Nutzungszugriffe (Fi29) (*exakte Angabe*)
- **Betreiber** (Fi11) (*Freitextfeld*)
- **Koordinator** (Fi12) (*Freitextfeld*)
- **Betriebspersonal**: Wissenschaftsunterstützendes Betriebspersonal (Fi26) (*Vollzeitäquivalente*), Sonstiges Betriebspersonal (Fi30) (*Vollzeitäquivalente*)

2. Verknüpfungen mit anderen Objekten

- Verknüpfung mit Publikation (\rightarrow V. Publikation) über einrichtungsinterne Forschungsinfrastruktur- bzw. Publikations-ID (s. u.):
 - **Titel Publikation(en)** (Fi14 bzw. Pu4a, s. o.)