

Projekt „Spezifikation Kerndatensatz Forschung“

Spezifikation des Kerndatensatz Forschung

Version 1.0

24. September 2015

Hinweis: Das vorliegende Dokument ist unverändert gültig auch für Version 1.1 des KDSF



Institut für Forschungsinformation und Qualitätssicherung (iFQ)

Fraunhofer-Institut für Angewandte Informationstechnik (FIT)

Geschäftsstelle des Wissenschaftsrates

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

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen M511300A gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

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

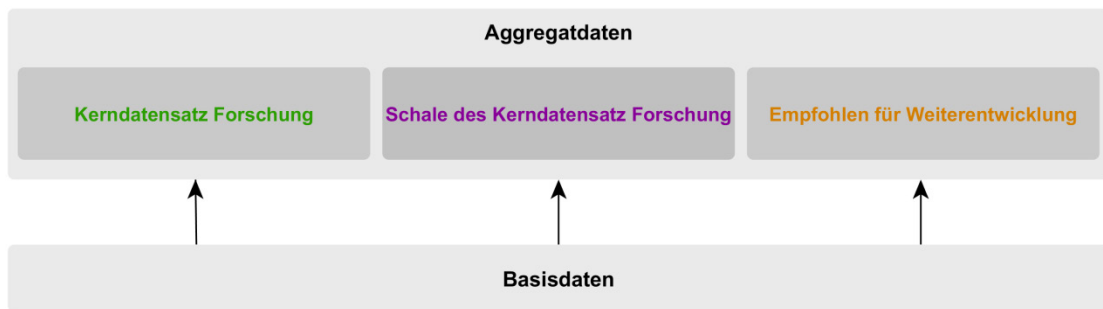
Abbildung 1 gibt einen Überblick zu den verschiedenen Kategorien von Aggregatdaten.

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 beispielsweise 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

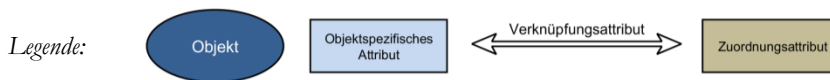
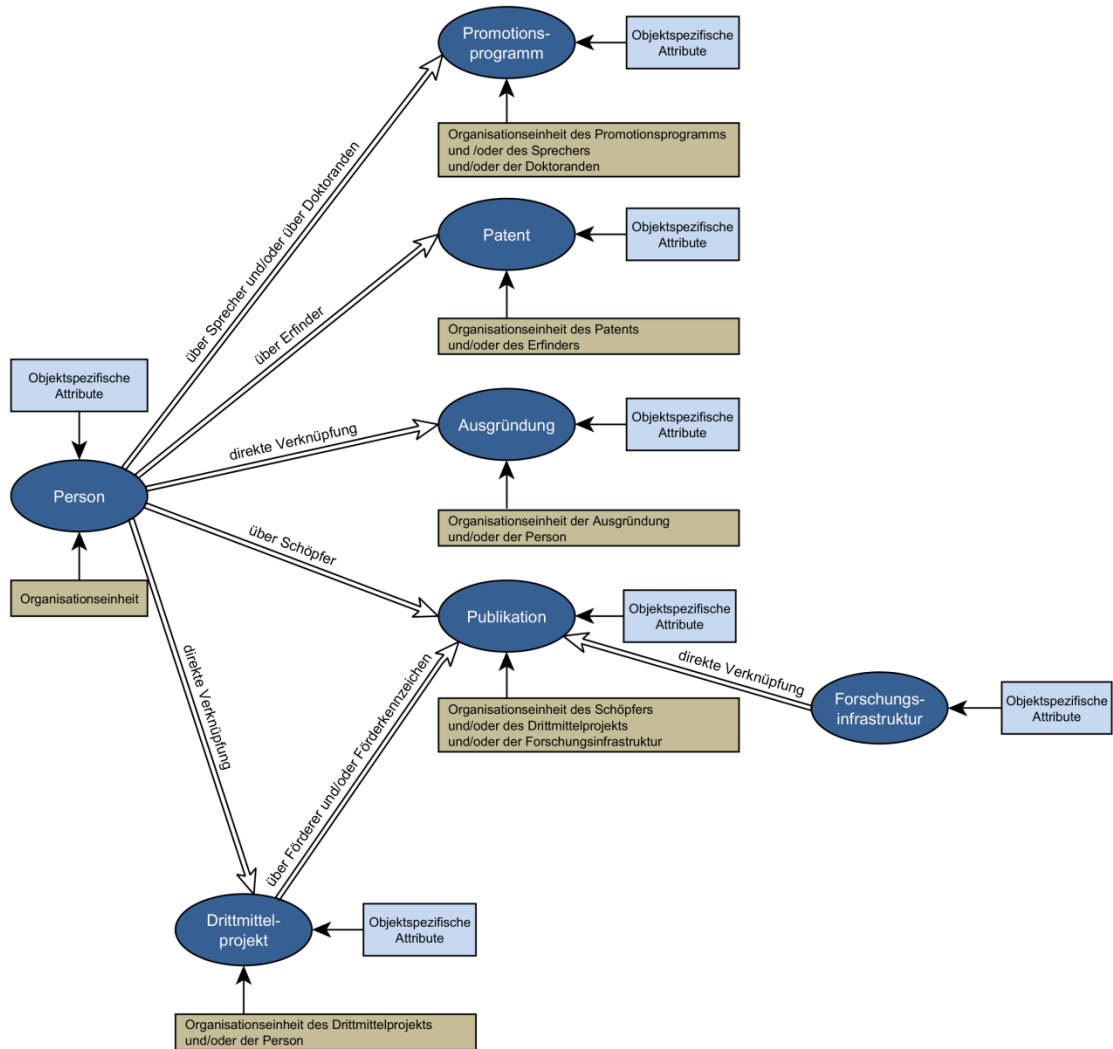


Abbildung 4 gibt einen graphischen Überblick über empfohlene Zuordnung von Basisdaten (Objekte und deren Attribute) zu den unterschiedlichen Bereichen des Kerndatensatz *Forschung*.²⁷

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 Drittittelgeber) - Titel des übergeordneten Projekts 	<p>Verknüpfungen:</p> <ul style="list-style-type: none"> - Drittittelgeber / Förderer (s. Publikation) - Förderkennzeichen je Drittittelgeber (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 / Drittittelgeber (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> <ul style="list-style-type: none"> - Organisationseinheit - Fach - Forschungsfeld

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 **einrichtungsinterne**, 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

The screenshot shows the 'KDSF Konzeptuelles Datenmodell' interface. The top navigation bar includes 'Basismodell', 'Aggregationen', 'Liste aller Objekte', and 'OWL-Dateien'. The left sidebar, titled 'Liste der Objekte', has a dropdown menu with 'Person' selected. Below the dropdown is a list of objects: 'hat Fach', 'hat Forschungsfeld', 'hat Qualifikation', 'Beschäftigte / Beschäftigter', 'Projekt bzw. Programm', 'Patent', 'Publikation', and 'Forschungsinfrastruktur'. The main content area displays the details for the 'Person' object. It includes the ID 'http://kerndatensatz-forschung.de/owl/Basis#Person', a description 'Zusammenfassendes Konzept für die Darstellung aller personenbezogenen Eigenschaften und Attribute', and a list of relationships: 'hat Fach zu Fach', 'hat Forschungsfeld zu Forschungsfeld', and 'hat Qualifikation zu Qualifikation'. The 'Eigenschaften' section lists: 'Alter : positivelnteger', 'Geschlecht : { "Maennlich", "Ohne Angabe", "Weiblich" }', 'Name : string', 'Staatsangehoerigkeit : string', and 'hat Promotionsberechtigung aus : string'. The 'Spezialisierungen' section lists 'Beschäftigte / Beschäftigter'. The 'CERIF Mapping' section lists 'Abgeschlossen' and 'urn:xmllns.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.

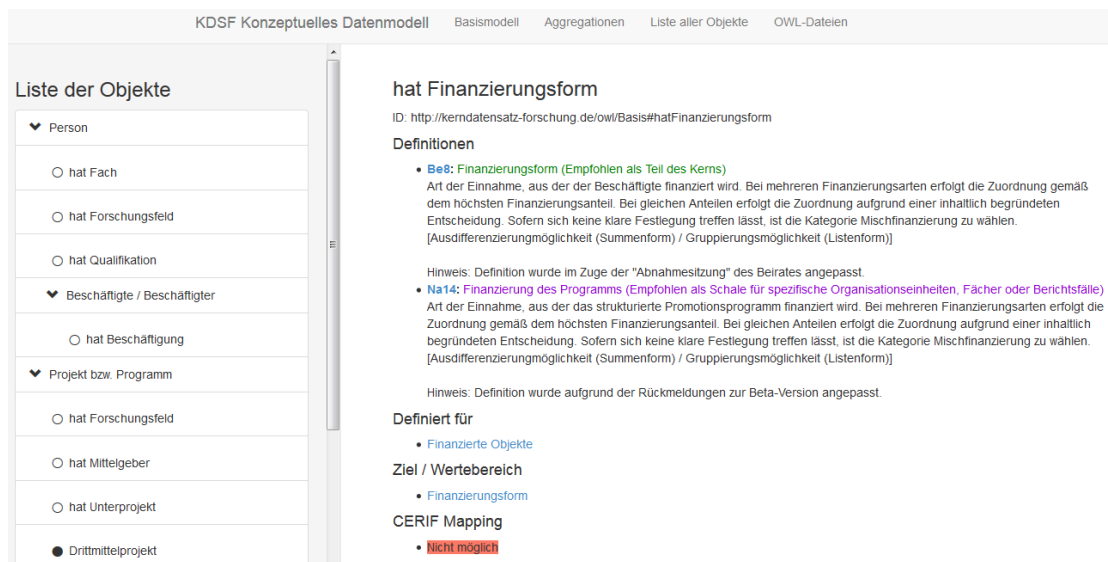
⁴¹ Die Liste der Beziehungen eines Objekts zu anderen Objekten (Verknüpfungs- und Zuordnungsattribute) lässt sich darüber hinaus auf der linken Seite des technischen Datenmodells in Teilbäumen unter dem jeweiligen Objekt aufklappen. Dort sind die jeweiligen Beziehungen als *Icons* aufgeführt.

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



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

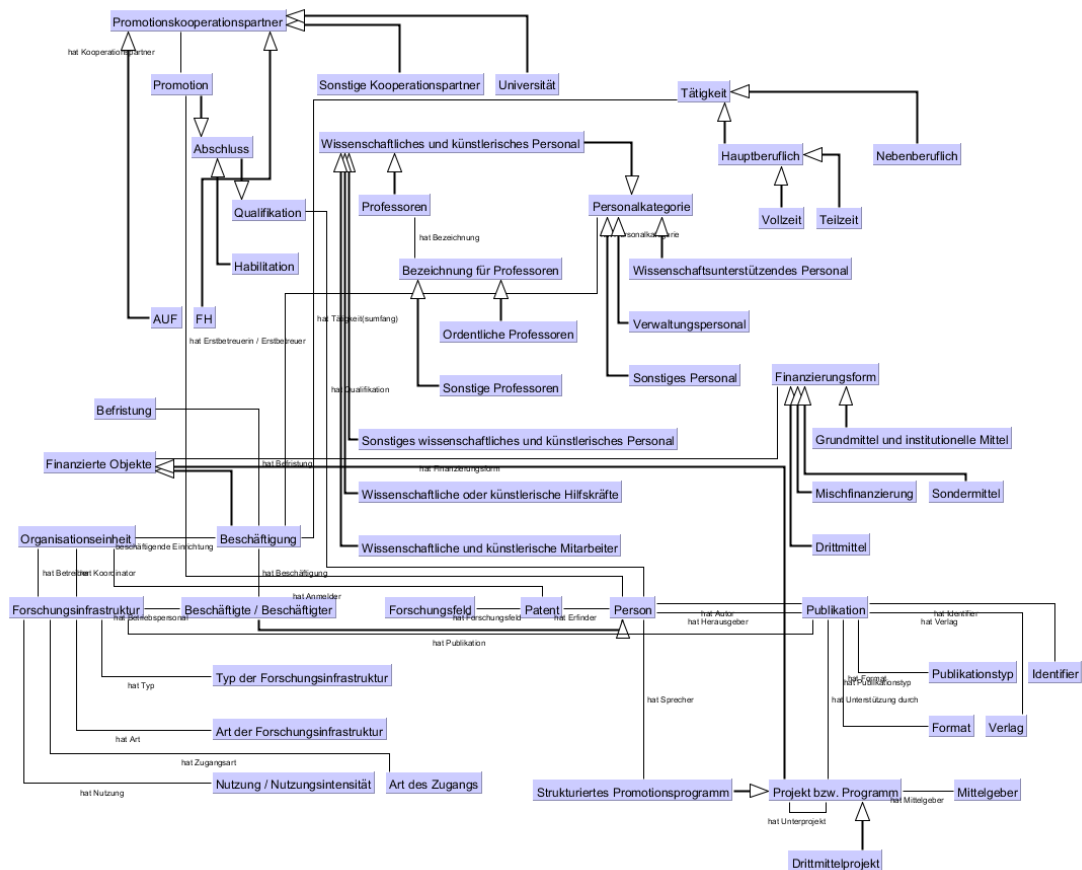
- Kerndaten

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 „Drittmiteleinahmen (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 Identifizierers (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 (→ V. Publikation) über einrichtungsinterne Forschungsinfrastruktur- bzw. Publikations-ID (s. u.):*
 - *Titel Publikation(en)* (Fi14 bzw. Pu4a, s. o.)