

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

## Harmonisierung von Datenlebenszyklus-Modellen

Nutzung von Synergien für optimierte Anwendungen im FDM

Armin Harry Wolf<sup>i</sup>      Cindy Leppla<sup>ii</sup>

2020

### Zitiervorschlag

Wolf, Armin Harry und Cindy Leppla. 2020. Harmonisierung von Datenlebenszyklus-Modellen. Nutzung von Synergien für optimierte Anwendungen im FDM. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 2/2020: S. 1-19. DOI: [10.7192/bfdm.2020.2.8281](https://doi.org/10.7192/bfdm.2020.2.8281).

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

<sup>i</sup>Universität Vechta, Fakultät I/Universitätsbibliothek. ORCID: [0000-0001-7766-3089](https://orcid.org/0000-0001-7766-3089)

<sup>ii</sup>Universitätsbibliothek Vechta. ORCID: [0000-0001-9156-5637](https://orcid.org/0000-0001-9156-5637)

# 1 Einleitung

Es existiert eine Vielfalt und Vielzahl an Datenlebenszyklus-Modellen, die sich in der Anwendendenperspektive oder im Fokus auf bestimmte Lebensphasen unterscheiden<sup>1,2,3</sup>. Ursache hierfür sind einerseits die Heterogenität und Komplexität von Forschungsprozessen, andererseits die Entwicklung der Modelle durch verschiedene Anwendende wie Fachwissenschaften und Infrastruktureinrichtungen für verschiedene Zwecke. Diese reichen vom effizienten Datenmanagement bis hin zur Entwicklung von Infrastruktur- und Serviceangeboten. Der Bedarf an einem vereinheitlichten, allgemein anerkannten Modell hat sich bereits mehrfach herausgestellt etwa beim Management von Big Data<sup>4</sup>, bei der Ableitung von FDM-Workflows<sup>5</sup> und Datenmanagementplänen (DMPs)<sup>6</sup>, der Entwicklung gemeinsamer Kuratierungs- und Archivierungsstrategien für das FDM und die digitale Langzeitarchivierung<sup>7</sup> oder bei der Erarbeitung von Kompetenzmatrizen zur Vermittlung von Datenkompetenzen in der Hochschullehre<sup>8</sup>. Die gemeinsame Weiterentwicklung und Harmonisierung der Modelle durch die Anwendenden wäre wünschenswert, um deren Kooperation zu fördern und den Einsatz der Modelle in der Praxis zu optimieren. Ziel des Artikels ist es, auf Basis weit verbreiteter oder für bestimmte Anwendungszwecke empfohlener Datenlebenszyklus-Modelle, den Forschungsprozess unter Beachtung von Qualitätskriterien für Prozessmodellierungen<sup>9</sup> abstrakt abzubilden. Damit soll zugleich die theoretische Grundlage für ein breites FDM-Anwendungsspektrum geschaffen und der Diskurs zur Modellstandardisierung befördert werden. Ausgangspunkt der Modellentwicklung ist eine Evaluation von Modellen, die sich an den Aktivitäten des Forschungsprozesses oder an Dateneigenschaften orientieren: Zu den aktivitätsorientierten Modellen zählen der Research Data Lifecycle des UK Data Service und das für die Vermittlung von Datenkompetenzen erarbeitete, kompetenzbasierte Prozessmodell der Datenwertschöpfung, zu den dateneigenschaftsorientierten das Modell des Data Curation Continuum. Diese Mod-

<sup>1</sup>Vgl. Sinaeepourfard, Amir, Xavier Masip-Bruin, Jordi Garcia, und Eva Marín-Tordera. 2015. "A Survey on Data Lifecycle Models: Discussions toward the 6Vs Challenges", <https://www.ac.upc.edu/app/research-reports/html/RR/2015/18.pdf>, (letzter Zugriff: 01.07.2020).

<sup>2</sup>Vgl. Wissik, Tanja, und Matej Ďurčo. 2015. "Research Data Workflows: From Research Data Lifecycle Models to Institutional Solutions". Selected Papers, Nr. 123, <https://ep.liu.se/ecp/123/008/ecp15123008.pdf>, (letzter Zugriff: 01.07.2020).

<sup>3</sup>Vgl. Ball, Alex. 2012. "Review of Data Management Lifecycle Models", <https://purehost.bath.ac.uk/ws/portalfiles/portal/206543/redm1rep120110ab10.pdf>, (letzter Zugriff: 01.07.2020).

<sup>4</sup>Vgl. Sinaeepourfard, Masip-Bruin, Garcia, und Marín-Tordera. "A Survey on Data Lifecycle Models"

<sup>5</sup>Vgl. Wissik und Ďurčo. "Research Data Workflows"

<sup>6</sup>Vgl. Ball. "Review of Data Management Lifecycle Models"

<sup>7</sup>Vgl. Lindlar, Michelle, Pia Rudnik, Laurence Horton, und Sarah Jones. 2020. "You Say Potato, I Say Potato - Mapping Digital Preservation and Research Data Management Concepts Towards Collective Curation and Preservation Strategies", <https://doi.org/10.5281/zenodo.3664987>.

<sup>8</sup>Vgl. Schüller, Katharina, Paulina Busch, und Carina Hindinger. 2019. "Future Skills: Ein Framework für Data Literacy". Hochschulforum Digitalisierung Nr. 47 (August 2019). <https://doi.org/10.5281/zenodo.3349864>.

<sup>9</sup>Rauh, Otto und Eberhard Stickel. 1997. Konzeptuelle Datenmodellierung. Teubner-Reihe Wirtschaftsinformatik. Stuttgart [u. a.]: Teubner.

elle werden anhand anwendungs- und qualitätsbezogener Anforderungen evaluiert und aus den Evaluationsergebnissen ein Vorschlag für ein allgemeines Modell abgeleitet.

## 2 Datenlebenszyklus-Modelle im FDM

### 2.1 Anforderungen

Datenlebenszyklus-Modelle stellen den Lebensverlauf von Daten<sup>10</sup> beziehungsweise den Forschungsprozess generalisiert abstrakt als Abfolge von Phasen bzw. Aktivitäten dar, die sich durch die Erzeugung oder Manipulation von Daten auszeichnen<sup>11</sup>. Klassifiziert werden können die Modelle nach ihrer Struktur, z. B. linear, zyklisch oder nach ihren Entwicklern, z. B. infrastrukturorientiert, wissenschaftsorientiert. Konkrete Anwendungen der Modelle umfassen die Ableitung von DMP's<sup>12</sup> und Workflows für effizientes Datenmanagement<sup>13</sup>, die Entwicklung von Erhebungsinstrumenten für FDM-Bedarfsermittlungen<sup>14</sup>, die Identifikation von Infrastruktur- und Serviceangeboten<sup>15</sup>, die Planung und Evaluation technischer Infrastrukturen<sup>16</sup>, die Ableitung von Kompetenzmatrizen für Schulungsangebote<sup>17</sup> oder den Einsatz der Modelle in Beratungs- und Vermittlungskontexten. In der vorliegenden Arbeit soll der Forschungsprozess (Modellierungsgegenstand) für diese Anwendungen (Modellierungszwecke) als Abfolge von Aktivitäten dargestellt werden, die die Dateneigenschaften beeinflussen bzw. als Abfolge von Phasen, in denen Daten mit veränderten Eigenschaften vorliegen (Modellierungselemente). Hierfür werden die für Prozessmodellierungen verwendeten Qualitätskriterien Umfassendheit, Korrektheit, Redundanzfreiheit und Verständlichkeit herangezogen, die adaptiert und um weitere Kriterien wie Rückverfolgbarkeit auf zugrundeliegende Modelle<sup>18</sup> und Nachhaltigkeit ergänzt werden (vgl. [Tabelle 1](#) im Anhang).

<sup>10</sup>Vgl. Sinaeepourfard, Masip-Bruin, Garcia, und Marín-Tordera. "A Survey on Data Lifecycle Models"

<sup>11</sup>Humphrey, Charles. 2006. "E-science and the life cycle of research", <https://doi.org/10.7939/R3NR4V>.

<sup>12</sup>Vgl. Ball. "Review of Data Management Lifecycle Models"

<sup>13</sup>Vgl. Wissik und Āurčo. "Research Data Workflows"

<sup>14</sup>Wolf, Armin Harry, und Cindy Leppla. 2020. "FDM-Services Einer Universitätsbibliothek zur Entfaltung von Regenbogenqualitäten im Forschungsprozess". O-Bib. Das Offene Bibliotheksjournal / Herausgeber VDB 7 (1), 1-16. <https://doi.org/10.5282/o-bib/5527>.

<sup>15</sup>Vgl. Universität zu Köln, Data Center for the Humanities, <https://dch.phil-fak.uni-koeln.de/dienste>, (letzter Zugriff: 01.07.2020).

<sup>16</sup>Z. B. innerhalb des BMBF-geförderten Projektes: UniV-FDM - "Bottom-up-Managementmodell zur Etablierung eines institutionellen Forschungsdatenmanagements", <https://bibliothek.uni-vechta.de/management-von-forschungsdaten/>, (letzter Zugriff: 01.07.2020).

<sup>17</sup>Whyte, Angus, und Kevin Ashley. 2017. "D7.1: Skills landscape analysis and competence model. EOSCpilot: The European Open Science Cloud for Research Pilot Project." <https://eoscipilot.eu/sites/default/files/eoscipilot-d7.1.pdf>, (letzter Zugriff: 01.07.2020).

<sup>18</sup>Vgl. Pitschke, Juergen. 2012. "Gute Modelle – Wie die Qualität von Unternehmensmodellen definiert und gemessen werden kann", [https://www.enterprise-design.eu/files/images/downloads-wissen/modelqualitaet\\_v2.0.pdf](https://www.enterprise-design.eu/files/images/downloads-wissen/modelqualitaet_v2.0.pdf), (letzter Zugriff: 01.07.2020).

Das Kriterium der Umfassendheit bezieht sich auf alle Sichten und Elemente des Forschungsprozesses, die für die Modellierungszwecke relevant sind: Dies umfasst die Aktivitätssicht mit wesentlichen Aktivitäten, Agierenden und Rollen sowie die Dateneigenschaftssicht mit vielfältigen Datentypen und -eigenschaften inklusive differenzierter Betrachtung von Daten und Forschungsergebnissen (Datenprodukte). Verbunden mit dem Merkmal der Datenauthenzizität ist die Unterscheidung von statischen und dynamischen Objekten zur Nachvollziehbarkeit von Inhaltsänderungen unter Sicherstellung der Urheberschaft. Datenintegrität lässt sich durch Vermeidung und Behebung von Datenfehlern und -verlusten erreichen, Datensicherheit durch Zugangskontrolle. Bei der Modellierung ist neben dem hypothesen- auch der datenbasierte Forschungsansatz mit datenbasiertem Entscheiden und Handeln zu berücksichtigen: Dieser stellt auf Daten als Ausgangspunkt des Forschungsprozesses ab, aus denen sich Forschungsfragen ableiten lassen und die die Basis für zu treffende Entscheidungen im Verlauf des Prozesses bilden. Den Standards und Richtlinien zur guten wissenschaftlichen Praxis, die u. a. auf die Vermeidung redundanter Datenerhebungen abzielen, wird Rechnung getragen, indem die Datenbeschaffung vor der Erhebung neuer Daten erfolgt. Im Fokus der übrigen Qualitätskriterien stehen die logische Abfolge und übersichtliche Anordnung von Modellelementen sowie eine sprachlich korrekte, verständliche, nachhaltige und redundanzfreie Terminologie. Auf Basis der beschriebenen Modellanforderungen werden Evaluationen der nachfolgenden drei Modelle durchgeführt und die Ergebnisse präsentiert.

## 2.2 Evaluation

### 2.2.1 Research Data Lifecycle des UK Data Service

Das vom UK Data Archive entwickelte zyklische Modell zielt auf den Forschungsprozess mit sozial- und wirtschaftswissenschaftlichen Daten ab und ist in den jeweiligen Fachwissenschaften und im Infrastrukturbereich weit verbreitet<sup>19,20,21</sup>. Eingesetzt wird es für die Entwicklung von Erhebungsinstrumenten<sup>22</sup>, DMPs<sup>23</sup>, Workflows und Kompetenzmatrizen<sup>24</sup>. Der Forschungsprozess wird als Abfolge wesentlicher verall-

<sup>19</sup>Vgl. Lindlar, Rudnik, Horton und Jones. "You Say Potato, I Say Potato"

<sup>20</sup>Vgl. RatSWD [Rat für Sozial- und Wirtschaftsdaten] (2018): "Forschungsdatenmanagement in den Sozial-, Verhaltens- und Wirtschaftswissenschaften – Orientierungshilfen für die Beantragung und Begutachtung datengenerierender und datennutzender Forschungsprojekte." RatSWD Output 3 (5). Berlin, Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://doi.org/10.17620/02671.7>.

<sup>21</sup>Der auf dem Informationsportal "forschungsdaten.info" aufgeführte Datenlebenszyklus orientiert sich am Research Data Lifecycle des UK Data Service. Vgl. [forschungsdaten.info](https://www.forschungsdaten.info/themen/informieren-und-planen/datenlebenszyklus/), <https://www.forschungsdaten.info/themen/informieren-und-planen/datenlebenszyklus/>, (letzter Zugriff: 09.09.2020).

<sup>22</sup>Wolf und Leppla. "FDM-Services Einer Universitätsbibliothek zur Entfaltung von Regenbogenqualitäten im Forschungsprozess"

<sup>23</sup>Vgl. RatSWD. "Forschungsdatenmanagement in den Sozial-, Verhaltens- und Wirtschaftswissenschaften"

<sup>24</sup>Whyte und Ashley. 2017. "D7.1: Skills landscape analysis and competence model."

gemeinerer Aktivitäten (vgl. Abb. 1) dargestellt, die semantisch korrekt zu Phasen gruppiert sind und in Einzelaktivitäten auffächern. Der Prozess beginnt mit dem Design der Forschungsfrage in der Planungsphase, berücksichtigt allerdings nicht den datenbasierten Forschungsansatz mit datenbasiertem Entscheiden und Handeln. Das Kriterium der syntaktischen Korrektheit ist für die Phasenabfolge durchgehend, für die Abfolge der Aktivitäten nur bedingt erfüllt, zumal die Beschaffung vorhandener Daten nicht explizit vor der Erzeugung neuer Daten erfolgt. Die Agierenden und Rollen werden nicht explizit benannt, sondern über die Aktivitäten ins Modell eingeführt: Dabei können die Rollen der Datenproduzierenden, -konsumierenden und -kuratierenden von Forschenden, Infrastruktureinrichtungen und nachnutzenden (forschenden) Dritten ausgeübt werden.

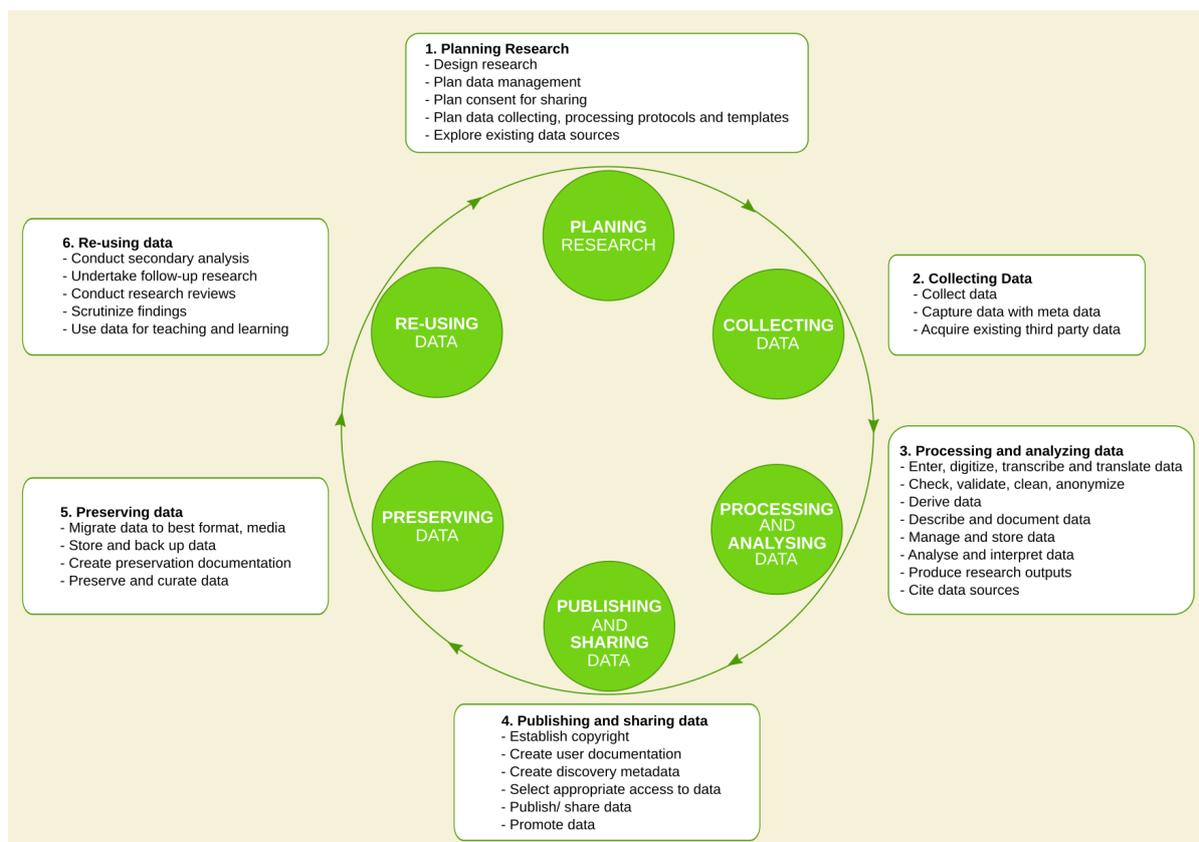


Abbildung 1: Research Data Lifecycle des UK Data Service (eigene Darstellung)<sup>25</sup>

Im Fokus des Modells stehen die den Forschungsprozess charakterisierenden Aktivitäten. Daten und deren Eigenschaften lassen sich nur indirekt daraus ableiten. Das Modell ist aufgrund seiner Übersichtlichkeit und Ausdrucksfähigkeit gut verständlich

<sup>25</sup>Vgl. UK Data Service Research data lifecycle, <https://www.ukdataservice.ac.uk/manage-data/lifecycle.aspx>, (letzter Zugriff: 01.07.2020).

und lässt sich aufgrund der überwiegend breiten Terminologie auf verschiedene Anwendungskontexte übertragen. Die in der Planungs- und Nachnutzungsphase mehrfach implizit auftretenden Kernaktivitäten “Recherchieren, Beschaffen, Prüfen/Interpretieren von Daten” weisen auf Redundanzen im Modell hin. Details zu den Evaluationsergebnissen sind [Tabelle 2](#) im Anhang zu entnehmen.

### 2.2.2 Kompetenzbasiertes Prozessmodell der Datenwertschöpfung

Das zyklische Modell von Schüller et al. ist Bestandteil der vom Hochschulforum Digitalisierung beauftragten Studie “Future Skills: Ein Framework für Data Literacy”<sup>26</sup>, die vom Stifterverband im Kontext der Data Literacy Education empfohlen wird<sup>27</sup> und geht zurück auf den Cycle of Encoding and Decoding nach Münster<sup>28</sup> und den Zyklus der datenbasierten Entscheidungsfindung<sup>29</sup>. Das Modell (vgl. [Abb. 2](#)) betrachtet Systeme in der realen Welt mit ihren messbaren Objekten, die in einem Kodierungsprozess zweimal verschlüsselt werden: in Daten und deren visuelle und verbalisierte Produkte. Im anschließenden Dekodierungsprozess werden die Datenprodukte zweimal entschlüsselt, so dass über die Daten Rückschlüsse auf die Systeme mit ihren Objekten gezogen werden können. Die Ergebnisse der Modellevaluation lassen sich wie folgt zusammenfassen (vgl. [Tabelle 2](#) im Anhang): Aufgebaut ist das Modell aus aufeinanderfolgenden Phasen mit forschungsbezogenen Aktivitäten: Der Forschungsprozess beginnt mit der Identifikation der Forschungsgegenstände und endet mit der Nachnutzung, in der Ergebnisse und Daten interpretiert und datenbasiert entschieden und gehandelt wird. Wesentliche Aktivitäten und Inhalte mit Ausnahme des Aufbewahrens und Bereitstellens von Daten und deren Produkten zur Nachnutzung und dem datenbasierten Forschungsansatz sind enthalten. Die Aktivitäten sind semantisch und syntaktisch korrekt zu Phasen gruppiert, allerdings wird in der Abfolge nicht zwischen Datenbeschaffung und Datenneuerzeugung getrennt. Das Modell unterscheidet explizit zwischen kodierenden bzw. dekodierenden Agierenden und Aktivitäten, die auf die Rollen der Datenproduzierenden und -kuratierenden bzw. auf die der Datenkonsumierenden entfallen. Mit Blick auf den Forschungsprozess liegt der Fokus auf den datenproduzierenden und -nachnutzenden Forschenden, weniger auf den kuratierenden Infrastruktureinrichtungen, zumal Aktivitäten wie Datenaufbewahrung und -bereitstellung zur Nachnutzung im Modell fehlen. Daten und Eigenschaften lassen sich indirekt über Aktivitäten ableiten. Das Kriterium der Verständlichkeit im Sinne von Über-

<sup>26</sup>Vgl. Schüller, Busch und Hindinger. “Future Skills: Ein Framework für Data Literacy”

<sup>27</sup>Vgl. Stifterverband, <https://www.stifterverband.org/data-literacy-education>, (letzter Zugriff: 01.07.2020).

<sup>28</sup>Münster, Evelyn. 2019. “Hilfe, wieso versteht niemand meine Datenvisualisierung?”, <https://www.designation.eu/2019/01/30/wieso-versteht-niemand-meine-datenvisualisierung/>, (letzter Zugriff: 01.07.2020).

<sup>29</sup>The Association of Independent Schools of New South Wals (Australia). Data Informed Decision Making Cycle, <https://www.aisnsw.edu.au/Resources/WAL%204%20%5BOpen%20Access%5D/Data%20Informed%20Decision-Making%20Cycle.pdf>, (letzter Zugriff: 01.07.2020).

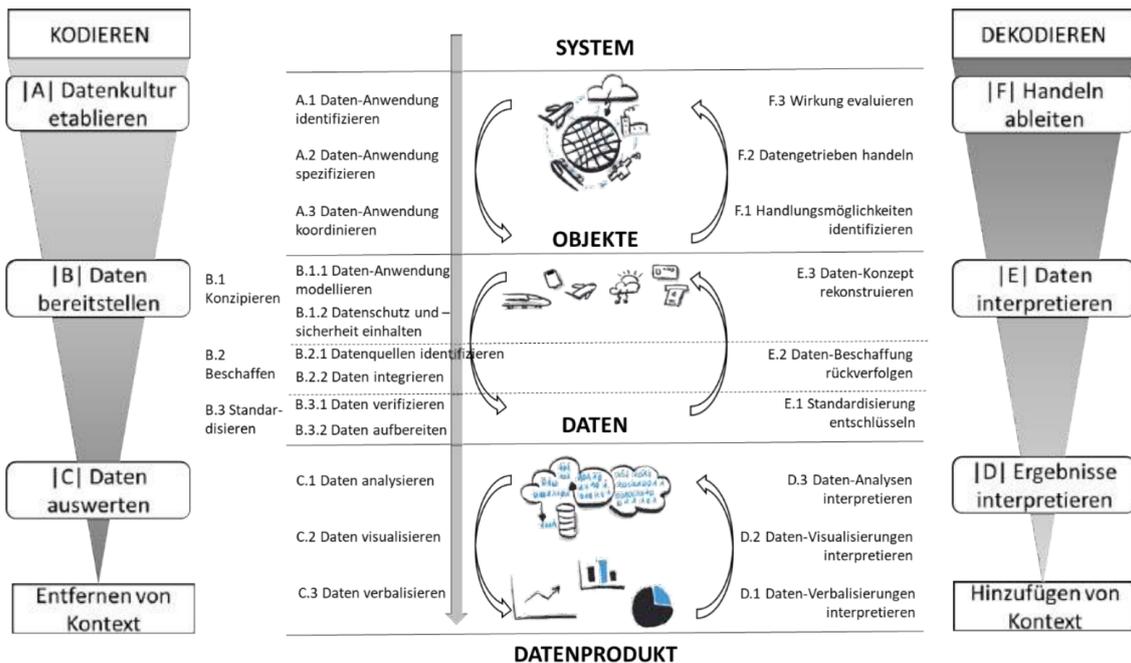


Abbildung 2: Kompetenzbasiertes Prozessmodell der Datenwertschöpfung<sup>30</sup>

sichtigkeit und Ausdrucksfähigkeit ist ebenso erfüllt wie das der Nachhaltigkeit, Redundanzfreiheit und Rückverfolgbarkeit auf zugrundeliegende Modelle.

### 2.2.3 Modell des Data Curation Continuum

Das aus der Informationswissenschaft stammende Modell wird zur Entwicklung von Forschungsdatenrepositorien eingesetzt und beschreibt das Datenobjekt mit seinen Eigenschaften. Darauf aufbauend erfolgt eine Aufteilung der Repositorieninfrastruktur in Domänen mit Migrationsgrenzen<sup>31</sup>. Das Modell unterscheidet drei Kontinua mit spezifischen Merkmalen, die kontinuierlich zwischen zwei Endpunkten verlaufen: Während das objektbezogene Kontinuum gekennzeichnet ist durch eine geringe bis hohe Anzahl an Datenobjekten und an Metadaten, ein geringes bis hohes Datenvolumen sowie dynamische und statische Datenobjekte aufweist, werden im Managementkontinuum die Zuständigkeiten für das Datenmanagement zwischen Forschenden und Infrastruktureinrichtung und der Aufbewahrungszeitraum von kurzfristiger Sicherung bis zur Langzeitarchivierung flexibel festgelegt. Merkmale des Zugangskontinuums sind Datenzugänglichkeit und –auffindbarkeit in verschiedenen Abstufungen. Die Be-

<sup>30</sup>Vgl. Schüller, Busch, und Hindinger. "Future Skills: Ein Framework für Data Literacy". Die Abbildung ist Seite 34 entnommen.

<sup>31</sup>Treloar, Andrew, und Cathrine Harboe-Ree. 2008. "Data Management and the Curation Continuum: How the Monash Experience Is Informing Repository Relationships", 2008. <https://www.vala.org.au/vala2008-proceedings/vala2008-session-6-treloar/>, (letzter Zugriff: 01.07.2020).

ziehung des Data Curation Continuum zur Repositorieninfrastruktur wird hergestellt durch Zuordnung der Merkmale zum Forschungsprozess. Merkmalsänderungen an Migrationsgrenzen in Form von Datenselektion, -anreicherung mit Metadaten, Datenanpassung z. B. mittels Formatänderung ermöglichen dessen Aufteilung in Domänen mit unterschiedlichen infrastrukturellen Anforderungen. Treloar et al. arbeiten mit einer privaten und einer kollaborativen Forschungsdomäne sowie einer öffentlichen Domäne, die jeweils links, mittig und rechts im Data Curation Continuum verortet sind (vgl. Abb. 3).

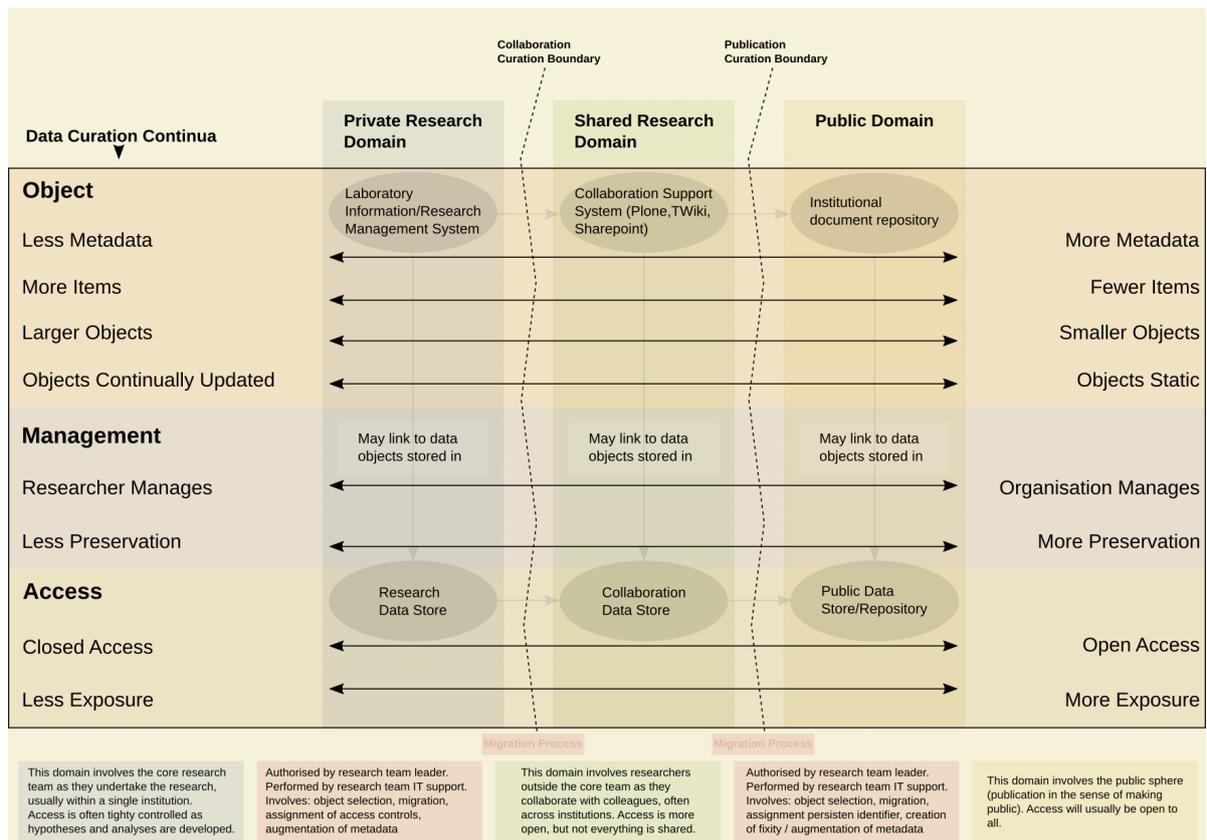


Abbildung 3: Modell des Data Curation Continuum (eigene Darstellung)<sup>32</sup>

Die Modellevaluation liefert folgende Erkenntnisse (vgl. Tabelle 2 im Anhang): Der Forschungsprozess wird als alternierende, lineare Abfolge von Domänen mit spezifischen Eigenschaften von Datenobjekten und Migrationsgrenzen mit datenmanipulierenden Aktivitäten semantisch und syntaktisch korrekt dargestellt. Im Fokus des Modells steht das Datenobjekt mit seinen Eigenschaften. Der Datenbegriff deckt eine

<sup>32</sup>Treloar und Harboe-Ree. "Data Management and the Curation Continuum". Die Abbildung 3 ist eine eigene Darstellung: Die von Treloar et al. separat dargestellten Data Curation Continua sowie die Domains, Data Stores und Curation Boundaries wurden in einer einheitlichen Grafik integrativ dargestellt. Die einzelnen Abbildungen bei Treloar et al. sind auf den Seiten 3 und 6 zu finden.

Vielfalt an Datentypen und wichtige Dateneigenschaften differenziert ab, wobei Dateauthentizität über die Verknüpfung der Urheberschaft mit statischen Objektversionen, Datenintegrität über Qualitätssicherung an den Migrationsgrenzen sowie bereinigte Objektversionen und Datensicherheit durch Zugangsbeschränkungen über alle Domänen hinweg gesichert ist. Das Modell benennt explizit Forschende und Infrastruktureinrichtungen als Agierende in der gemeinsamen Rolle der Kuratierenden mit sich ändernder Zuständigkeit über das Data Curation Continuum hinweg. Aussagen zu Forschungsansätzen, zu domänenbezogenen Aktivitäten und deren Abfolge sowie zur Rolle der Datenproduzierenden und -konsumierenden lassen sich nur implizit über die Verlinkung mit dem Forschungsprozess treffen. Das Modell zeichnet sich aufgrund seiner übersichtlichen, flexiblen Modellstruktur und der einfachen, breiten Terminologie durch Verständlichkeit und Nachhaltigkeit aus und ist zugleich redundanzfrei.

### 3 Aktivitäts- und eigenschaftsorientierter Datenlebenszyklus

Aus der vorausgegangenen Modellevaluation folgt, dass keines der drei evaluierten Modelle die gestellten Anforderungen vollumfänglich erfüllt. Für das Kriterium der Umfassendheit ergibt sich dabei folgendes Bild: Die wesentlichen Aktivitäten werden vom Research Data Lifecycle des UK Data Service abgedeckt, die wichtigsten Dateneigenschaften vom Data Curation Continuum-Modell. Bezogen auf die geforderten Rollen und Agierende unterscheiden sich die drei Modelle in ihrem Fokus: Während der Research Data Lifecycle des UK Data Service auf datenproduzierende Forschende und das kompetenzbasierte Prozessmodell der Datenwertschöpfung zusätzlich auf datennachnutzende (forschende) Dritte fokussiert, stehen beim Modell des Data Curation Continuum die datenkuratierenden Forschenden und Infrastruktureinrichtungen im Vordergrund. Von den beiden genannten Forschungsansätzen wird lediglich der hypothesenbasierte vom Research Data Lifecycle des UK Data Service und dem kompetenzbasierten Prozessmodell der Datenwertschöpfung adressiert. Letzteres beinhaltet zusätzlich das datenbasierte Entscheiden und Handeln bei der Nachnutzung und führt diese auf die wesentlichen Aktivitäten des Prüfens/Interpretierens von Ergebnissen und Daten zurück. Beim Research Data Lifecycle des UK Data Service sind das Recherchieren, Beschaffen, Prüfen/Interpretieren von Daten implizite Bestandteile der Planungs- und Nachnutzungsphase, wobei letztere diese Aktivitäten noch auf Ergebnisse ausweitet. Das Kriterium der Korrektheit ist im Modell des Data Curation Continuum für die Inhalte und Abfolge der Domänen erfüllt. Der Research Data Lifecycle des UK Data Service und das Prozessmodell der Datenwertschöpfung gruppieren die Aktivitäten semantisch korrekt zu Phasen mit logischer Phasenabfolge, berücksichtigen aber nicht explizit, dass die Datenbeschaffung vor der Datenneuerzeugung erfolgen sollte. Alle drei Modelle zeichnen sich durch Verständlichkeit, Flexibilität und Erweiterbarkeit aus. Die Rückverfolgbarkeit auf andere Modelle ist gegeben, sofern diese bei der Modellentwicklung verwendet wurden. Im nächsten Schritt werden die Modelle so

zusammengeführt, dass sie sich hinsichtlich der gestellten Anforderungen gegenseitig ergänzen, wobei das fehlende Modellelement des datenbasierten Forschungsansatzes hinzugefügt, aktivitätsbezogene Redundanzen in der Planungs- und Nachnutzungsphase vermieden und die korrekte Abfolge "Datenbeschaffung vor Datenneuerzeugung" eingehalten wird. Daraus resultiert ein aktivitäts- und eigenschaftsorientierter Datenlebenszyklus als Modell für den Forschungsprozess bestehend aus einer Aktivitätssicht und einer Eigenschaftssicht.

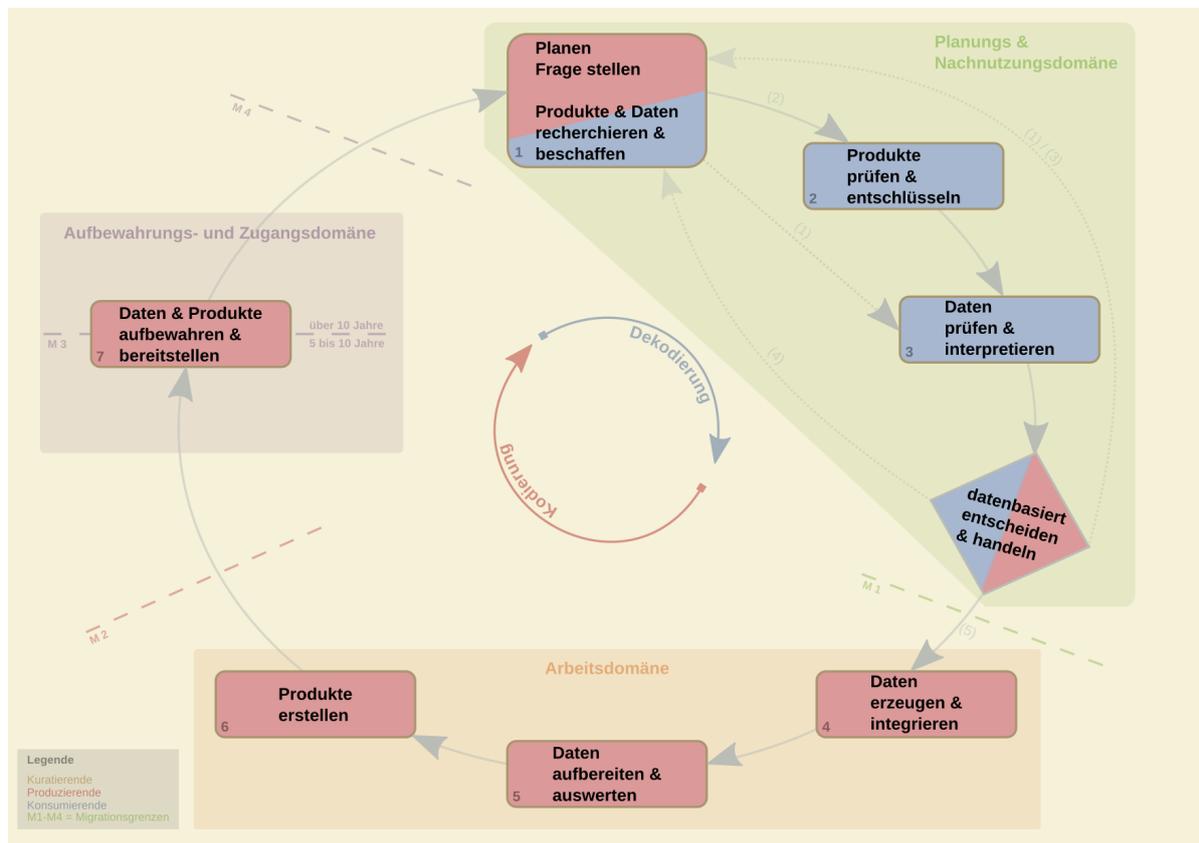


Abbildung 4: Aktivitäts- und eigenschaftsorientierter Datenlebenszyklus

### 3.1 Aktivitätssicht

Die Aktivitätssicht stellt den Forschungsprozess als zyklische Abfolge von Aktivitäten mit Hilfe der beiden Modellelemente "Phase/Aktivität (Rechteck)" und "Entscheidung (Raute)" dar. Zusätzlich zum hypothesenbasierten Forschungsansatz ("Frage stellen") geht auch der datenbasierte Ansatz in das Modell ein und das datenbasierte Entscheiden und Handeln wird zu Beginn des Forschungsprozesses berücksichtigt. Das zu Prozessbeginn stattfindende Recherchieren, Beschaffen, Prüfen/Interpretieren von Daten wird um Datenprodukte wie Literatur ergänzt, um eine umfassende Prüfung des

Forschungsstandes zu gewährleisten. Mit dieser Aktivitätsabfolge findet die Datenbeschaffung vor der Datenneuerzeugung statt, so dass den Regeln zur guten wissenschaftlichen Praxis Rechnung getragen wird. Der Prozess beginnt entweder mit einer Hypothese bzw. vorgegebenen Frage oder mit einer auf Basis beschaffter Daten formulierten Fragestellung (Daten recherchieren und beschaffen, Daten prüfen und interpretieren, datenbasiert entscheiden und Frage stellen = Pfad 1). Zur Beantwortung der Frage werden Datenprodukte (Literatur, Tabellen, Diagramme etc.) recherchiert und beschafft, anschließend geprüft und entschlüsselt, wobei ein Abgleich mit der Fragestellung und die Zurückführung auf die zugrundeliegenden Daten erfolgt (Pfad 2). Die beschafften Daten werden nach Qualitätsgesichtspunkten geprüft und interpretiert, um den Bezug zur eigenen Fragestellung und zu vorhandenen Datenprodukten herzustellen. Datenbasiert wird entschieden, ob die Frage bereits beantwortet und eine Änderung/Neuformulierung der Frage erforderlich ist (Pfad 3), ob zusätzliche Daten über weitere Recherche-, Beschaffungs- und Prüfungsschritte in den Prozess einzubeziehen sind (Pfad 4) oder ob sich die Frage mit den beschafften Daten bereits beantworten lässt (Pfad 5). Liegen geeignete Daten zur Beantwortung der Frage vor, werden diese im Anschluss an die datenbasierte Entscheidungsfindung integriert, d.h. aus verschiedenen Datenquellen und im Bedarfsfalle auch mit selbst erzeugten Daten zusammengeführt. Die Erzeugung neuer Daten ist dann alternativlos, wenn die vorgeschalteten Prozessschritte erfolglos verlaufen sind. Die nachfolgende Phase der Datenaufbereitung und –auswertung beinhaltet die Verarbeitung der Daten inklusive Qualitätskontrolle (prüfen, validieren, bereinigen) und weiterer gegebenenfalls erforderlicher Transformationsschritte (digitalisieren, transkribieren, übersetzen, anonymisieren, normalisieren, strukturieren) sowie die Datenanalyse und Ergebnisableitung. Die Ergebnisse werden in der Phase “Produkte erstellen” in Form von Datenprodukten visuell und/oder verbalisiert dargestellt, bevor sie zusammen mit den zugrundeliegenden Daten und sonstigen zur Nachvollziehbarkeit der Ergebnisse benötigten Materialien (Software, Datenbanken, Dokumentation etc.) aufbewahrt und nach Möglichkeit für Dritte zur Nachnutzung bereitgestellt werden. Diese nutzen die Ergebnisse und Daten für eigene Fragestellungen nach, indem sie mindestens die ersten drei Phasen inklusive des datenbasierten Entscheidens und Handelns durchlaufen. Das Modell unterscheidet zwischen kodierenden und dekodierenden Aktivitäten und Agierenden. Die kodierenden Aktivitäten in den Phasen 1 und 4 bis 7 sind der Rolle der Datenproduzierenden und der Rolle der Kuratierenden zugeordnet, die von Forschenden bzw. Infrastruktureinrichtungen ausgeübt werden können. Die dekodierenden Aktivitäten finden sich in den Phasen 1 bis 3 wieder und entfallen zum einen auf die konsumierenden Forschenden und zum anderen auf die kuratierenden Forschenden bzw. Infrastruktureinrichtungen. Zur Vermeidung von Redundanzen wurden die für Forschende und nachnutzende Dritte identischen Aktivitäten “Daten und Datenprodukte recherchieren”, “Datenprodukte prüfen und entschlüsseln”, “Daten prüfen und interpretieren” sowie das datenbasierte Entscheiden und Handeln zu Beginn des Prozesses verortet. Diesem Sachverhalt wird auch durch Zusammenführung der entsprechenden Phasen in einer Planungs- und Nachnutzungsdomäne (vgl. 3.2 Eigenschaftssicht) Rechnung getragen.

## 3.2 Eigenschaftssicht

Die Eigenschaftssicht basiert auf dem Data Curation Continuum und stellt den Forschungsprozess als zyklische Abfolge von Domänen mit unterschiedlichen Dateneigenschaften dar. Zur Domänenbildung werden die Merkmale des Kontinuums mit den Aktivitäten des Forschungsprozesses korreliert. Das Ergebnis sind drei Domänen mit charakteristischen Dateneigenschaften und Migrationsgrenzen, an denen arbeitsteilig zwischen Forschenden und Infrastruktureinrichtung Daten ausgewählt, angepasst (z. B. Formatkonversion), mit Metadaten angereichert und migriert werden: Die Planungs- und Nachnutzungsdomäne, die Arbeitsdomäne und die Aufbewahrungs- und Zugangsdomäne. Letztere lässt sich bei Bedarf in zwei Subdomänen aufteilen, die dem Kriterium des Aufbewahrungszeitraums von 5 bis 10 Jahren für Repositorien und über 10 Jahre hinaus für die digitale Langzeitarchivierung Rechnung tragen. Die Planungs- und Nachnutzungsdomäne zeichnet sich durch eine Vielzahl und Vielfalt an (statischen) Datenprodukten und Datenobjekten, große Datenvolumina und (umfassende, standardisierte) Metadaten aus. Der Zugriff ist beschränkt auf einzelne Forschende, die Datenprodukte und Daten im Rahmen der Projektplanung beschaffen und für deren Kuratierung und temporäre Sicherung zuständig sind und dabei von Infrastruktureinrichtungen unterstützt werden können. An der Migrationsgrenze M1 werden geeignete Datenobjekte mit ihren Metadaten ausgewählt, Formate konvertiert und die Daten in die Arbeitsdomäne migriert. In der Arbeitsdomäne werden Daten aus verschiedenen Quellen, darunter auch selbst erzeugte, zusammengeführt, aufbereitet und ausgewertet. Damit liegen eine Vielzahl digitaler, dynamischer Objekte in verschiedenen Versionen und wenige, heterogene Metadaten (Provenienz-, rudimentäre Erhebungsmetadaten) vor. Der Zugriff wurde ausgeweitet auf das Team der Forschenden, das sich mit Unterstützung der Infrastruktureinrichtung um die Kuratierung und Backups kümmert. An der Migrationsgrenze M2 erfolgt die Selektion von Datenprodukten und -objekten zur Archivierung und Bereitstellung, die Anreicherung mit Verarbeitungs- und Bereitstellungsmetadaten (u. a. Angaben zur Verarbeitungshistorie, Lizenzen, Persistent Identifier) und Standardisierung, Formatkonversion und Migration in die Aufbewahrungs- und Zugangsdomäne. Merkmale dieser Domäne sind die deutlich reduzierte Anzahl an Datenobjekten bzw. -produkten, die in statischen Versionen mit umfassenden, standardisierten Metadaten vorliegen und ein verringertes Datenvolumen. Eingeschränkte bis freie Zugänglichkeit bei optimaler Auffindbarkeit der Daten bzw. Datenprodukte und deren langfristige Sicherung werden durch Infrastruktureinrichtungen sichergestellt. Je nach Anforderungen und lokalen Gegebenheiten lässt sich die Domäne in zwei Subdomänen über die Migrationsgrenze M3 aufteilen, womit eine weitere Auswahl an Datenobjekten bzw. -produkten für die digitale Langzeitarchivierung vorgesehen und den damit verbundenen Anforderungen an Technik und Organisation Rechnung getragen wird. An der Migrationsgrenze M4 werden Dritten Zugriffs- und Nutzungsrechte eingeräumt und Datenobjekte bzw. -produkte zur Nachnutzung für die weitere Forschung ausgewählt.

## 4 Fazit und Ausblick

Die zentralen Erkenntnisse aus der Evaluation der drei Modelle "Research Data Lifecycle des UK Data Service", "Kompetenzbasiertes Prozessmodell der Datenwertschöpfung" und "Data Curation Continuum" stellen sich wie folgt dar: Die den Forschungsprozess modellhaft beschreibenden Datenlebenszyklen lassen sich je nach Anwendungszweck den beiden Kategorien "aktivitätsorientiert" bzw. "dateneigenschaftsorientiert" zuordnen und unterscheiden sich in der Gewichtung der Rollen der Datenproduzierenden, -konsumierenden und -kuratierenden, für deren Ausübung Agierende aus der Wissenschaft und dem Infrastrukturbereich in Frage kommen. Beim Research Data Lifecycle des UK Data Service und dem kompetenzbasierten Prozessmodell der Datenwertschöpfung handelt es sich um aktivitätsorientierte Modelle, die Dateneigenschaften nur implizit über die Aktivitäten berücksichtigen. Während der Research Data Lifecycle des UK Data Service aufgrund der Fokussierung auf die datenproduzierenden Forschenden gut für die Ableitung von DMPs, Workflows und für FDM-Bedarfserhebungen geeignet ist, erweitert das kompetenzbasierte Prozessmodell der Datenwertschöpfung zur Entwicklung von Schulungs- und Beratungsangeboten den Fokus auf die Datenkonsumierenden. Demgegenüber trägt das Modell des Data Curation Continuum aufgrund seiner Fokussierung auf Dateneigenschaften und kuratierende Forschende bzw. Infrastruktureinrichtungen der Planung und Entwicklung technischer Infrastrukturen Rechnung, benötigt hierzu aber Kenntnisse zum Ablauf des Forschungsprozesses. Die vergleichende Betrachtung der Modelle ermöglicht Rückschlüsse auf wesentliche forschungsbezogene Aktivitäten, die der Nachnutzung zugrunde liegen und auch unmittelbar bei bzw. nach der Prozessplanung Berücksichtigung finden sollten: Das Recherchieren, Beschaffen von Daten bzw. Forschungsergebnissen und deren Interpretation mit datenbasiertem Entscheiden und Handeln. Daraus resultieren gemeinsame Aktivitäten bzw. Phasen für Forschungsplanung und Nachnutzung in einer Domäne mit spezifischen Dateneigenschaften. Durch Kombination der drei Modelle entsteht ein aus zwei Sichten bestehender aktivitäts- und eigenschaftsorientierter Datenlebenszyklus, der die jeweiligen Modellqualitäten vereint, ein breites Anwendungsspektrum abdeckt, den Regeln zur guten wissenschaftlichen Praxis Rechnung trägt und die Modellanwendung in der Praxis optimiert: DMPs und Workflows profitieren von der Berücksichtigung der Dateneigenschaften, u. a. von der differenzierten Betrachtung der Datenaufbewahrung und -zugänglichkeit sowie der Festlegung von Zuständigkeiten, der Auf- und Ausbau technischer Infrastrukturen dagegen von der Zuordnung der Dateneigenschaften zu Aktivitäten in Verbindung mit einer flexiblen Aufteilung der Infrastrukturlandschaft. Das vorgestellte kombinierte Modell kann in Anschlussarbeiten für Use Cases aus dem betrachteten Anwendungsspektrum adaptiert werden und einen Beitrag zur kooperativen Weiterentwicklung und Standardisierung von Datenlebenszyklen leisten.

## 5 Anhang

**Tabelle 1:** Evaluationskriterien für Prozessmodellierungen am Beispiel des Forschungsprozesses

Kriterium	Beschreibung	Beispiel
Umfassendheit	Vorhandensein aller für den Modellierungszweck relevanten Sichten und Elemente	<p>1) Aktivitätssicht bestehend aus:</p> <p><u>a) Aktivitäten:</u></p> <ul style="list-style-type: none"> <li>• Datenbeschaffung mit Recherche, Prüfung/Interpretation und Datenneuerzeugung;</li> <li>• Datenauswertung mit Aufbereitung, Analyse und Ergebnisdarstellung;</li> <li>• Datenaufbewahrung und –bereitstellung;</li> <li>• Datennachnutzung;</li> <li>• hypothesenbasierter Forschungsansatz;</li> <li>• datenbasierter Forschungsansatz</li> </ul> <p><u>b) Akteure:</u></p> <ul style="list-style-type: none"> <li>• Forschende;</li> <li>• Infrastruktureinrichtungen</li> </ul> <p><u>c) Rollen:</u></p> <ul style="list-style-type: none"> <li>• Datenproduzierende;</li> <li>• Datenkonsumierende;</li> <li>• Datenkuratierende</li> </ul> <hr/> <p>2) Dateneigenschaftssicht bestehend aus:</p> <p><u>a) Datentypen:</u></p> <ul style="list-style-type: none"> <li>• Text;</li> <li>• Audio;</li> <li>• Video;</li> <li>• numerische Daten;</li> <li>• Unterscheidung von Daten und Datenprodukten</li> </ul> <p><u>b) Dateneigenschaften:</u></p> <ul style="list-style-type: none"> <li>• Datenvolumen;</li> <li>• Versionierung;</li> <li>• Datenformate;</li> <li>• Metadaten/Dokumentation;</li> </ul>

		<ul style="list-style-type: none"> <li>• Aufbewahrungszeitraum;</li> <li>• Datenauthentizität;</li> <li>• Datenintegrität;</li> <li>• Datensicherheit</li> </ul>
syntaktische und semantische Korrektheit	logische Abfolge inhaltlich zusammenhängender Modellelemente mit sprachlich korrekter und standard- bzw. regelkonformer Darstellung des Modellierungsgegenstandes	Anordnung der Aktivitäten „Recherchieren“, „Beschaffen“, „Prüfen/Interpretieren“ von Daten und Datenprodukten in logischer Abfolge zu Beginn des Forschungsprozesses und vor der Erzeugung neuer Daten
Verständlichkeit	Übersichtlichkeit, schnelle Erfassbarkeit und Ausdrucksfähigkeit des Modells	Unterteilung des Forschungsprozesses in mehrere Abstraktionsebenen unter Verwendung weniger aussagekräftiger Begriffe für Aktivitäten bzw. Dateneigenschaften beginnend mit Ebene aus generalisiert abstrakten Oberbegriffen und deren Konkretisierung durch Auffächerung in Unterebenen mit spezifischen Begriffen
Nachhaltigkeit	flexible Erweiterbarkeit und Übertragbarkeit des Modells auf verschiedene Anwendungskontexte (diverse Zielgruppen, Praktiken, Standards, Bedarfe)	Ergänzung von Aktivitäten bzw. Dateneigenschaften und/oder deren Zusammenfassung unter Verwendung kontextualisierter Begriffe
Redundanzfreiheit	Verwendung einer eindeutigen und in ihrer Bedeutung unterscheidbaren Terminologie für Modellelemente	Verwendung identischer Begriffe für Aktivitäten bzw. Dateneigenschaften, die in ihrer Definition übereinstimmen und vice versa
Rückverfolgbarkeit	Herstellung von Zusammenhängen zwischen einzelnen Modellen durch Identifikation übereinstimmender Modellelemente und -strukturen	Identifizierbarkeit übereinstimmender Aktivitäten bzw. Dateneigenschaften in einzelnen Datenlebenszyklus-Modellen (in Bezug auf Inhalte, Anordnung, Abfolge)

Tabelle 2: Evaluationsmatrix der untersuchten Datenlebenszyklus-Modelle

Kriterium	Datenlebenszyklus-Modell		
	Research Data Lifecycle des UK Data Service	Kompetenzbasiertes Prozessmodell der Datenwertschöpfung	Data Curation Continuum
Umfassendheit			
a) Aktivitäten (Phasen)	Datenplanung, Sammeln von Daten, Datenverarbeitung und -analyse, Veröffentlichen und Teilen von Daten, Datenaufbewahrung und -nachnutzung	Identifikation von Forschungsgegenständen ( <i>Phase A</i> ), Datenbeschaffung, -erzeugung und -aufbereitung ( <i>Phase B</i> ), Datenauswertung und -darstellung ( <i>Phase C</i> ), Datennachnutzung mit datenbasiertem Entscheiden und Handeln ( <i>Phasen D bis F</i> ); <u>Es fehlt:</u> Aufbewahren und Bereitstellen von Daten und deren Produkten zur Nachnutzung	Aktivitäten an Migrationsgrenzen ( <i>select, modify, augment objects</i> ); domänenbezogene, wesentliche Aktivitäten nur via Verlinkung mit Forschungsprozess
b) Forschungsansätze	rein hypothesengetrieben	rein hypothesengetrieben	nur via Verlinkung mit Forschungsprozess
c) Agierende	Forschende, nachnutzende (forschende) Dritte, Infrastruktureinrichtungen	<ul style="list-style-type: none"> <li>· kodierend: Forschende, Infrastruktureinrichtungen (<i>Phasen A bis C</i>);</li> <li>· dekodierend: nachnutzende (forschende) Dritte (<i>Phasen D bis F</i>)</li> </ul>	Forschende, Infrastruktureinrichtungen
d) Rollen	<ul style="list-style-type: none"> <li>· Datenproduzierend (<i>Phasen 1 bis 3: planning research, collecting data, processing and analyzing data</i>);</li> <li>· datenkuratierend (<i>Phase 4 und 5: publish/share data, promote data, preserve and curate data</i>);</li> </ul>	<ul style="list-style-type: none"> <li>· datenproduzierend (<i>Phasen A bis C</i>);</li> <li>· datenkuratierend (<i>Phasen A bis C</i>);</li> <li>· datenkonsumierend (<i>Phasen D bis F</i>)</li> </ul>	<ul style="list-style-type: none"> <li>· datenkuratierend</li> <li>· datenproduzierend (via Verlinkung mit Forschungsprozess);</li> <li>· datenkonsumierend (via Verlinkung mit Forschungsprozess)</li> </ul>

<p>e) Datentypen und -eigenschaften</p>	<ul style="list-style-type: none"> <li>· datenkonsumierend (z.B. <i>conduct research reviews, scrutinize findings</i>)</li> </ul>		
	<p>ableitbar aus Aktivitäten:</p> <ul style="list-style-type: none"> <li>· <i>digitize, transcribe</i> → digitale, analoge Daten;</li> <li>· <i>produce research outputs</i> → Daten und Datenprodukte;</li> <li>· <i>translate, anonymize, check, validate, clean data</i> → Versionierungen/ Datenintegrität;</li> <li>· <i>capture data with metadata, describe and document data</i> → Metadatentypen/Dokumentation;</li> <li>· <i>migrate to best format</i> → Datenformate;</li> <li>· <i>select appropriate access to data</i> → Datensicherheit bei der Datenbereitstellung</li> </ul>	<p>ableitbar aus Aktivitäten:</p> <ul style="list-style-type: none"> <li>· <i>Daten visualisieren, verbalisieren</i> → visuelle, verbalisierte Daten;</li> <li>· <i>Daten verifizieren, aufbereiten</i> → Versionierung/Datenintegrität;</li> <li>· <i>Daten aufbereiten, integrieren</i> → Datenformate/Metadaten;</li> <li>· Datensicherheit bereits bei der Prozessplanung</li> </ul>	<ul style="list-style-type: none"> <li>· digitale Datenobjekte;</li> <li>· Datenvolumen;</li> <li>· Datenformate;</li> <li>· Metadaten;</li> <li>· Aufbewahrungszeitraum;</li> <li>· Versionierungen;</li> <li>· Zugänglichkeit;</li> <li>· Auffindbarkeit;</li> <li>· Datenauthentizität;</li> <li>· Datenintegrität;</li> <li>· Datensicherheit</li> </ul>
<p>Korrektheit</p> <p>a) semantisch</p>	<p>Gruppierung inhaltlich zusammenhängender Aktivitäten zu Phasen</p>	<p>Gruppierung inhaltlich zusammenhängender Aktivitäten zu Phasen</p>	<p>Gruppierung von Dateneigenschaften zu Domänen anhand quantitativer und qualitativer Eigenschaftsänderungen entlang des Forschungsprozesses:</p> <ul style="list-style-type: none"> <li>· abnehmende Zahl an Datenobjekten;</li> <li>· zunehmende Zahl an Metadaten;</li> <li>· steigender Anteil statischer Objekte;</li> <li>· Verlängerung des Aufbewahrungszeitraums;</li> <li>· Verbesserung der Auffindbarkeit;</li> </ul>

			<ul style="list-style-type: none"> <li>· Erhöhung der Zugänglichkeit;</li> <li>· Änderung der Zuständigkeit für Datenmanagement von Forschenden zu Infrastruktureinrichtung</li> </ul>
b) syntaktisch	für Phasenabfolge durchgehend, für Aktivitätsabfolge nur bedingt ( <i>Datenbeschaffung, Erzeugen und Sammeln von Daten in der Phase „collecting data“ ohne explizite Differenzierung der Abfolge</i> )	für Phasenabfolge durchgehend, für Aktivitätsabfolge nur bedingt ( <i>Beschaffung vorhandener und Erzeugung neuer Daten ohne Differenzierung der Abfolge</i> )	chronologische Wiedergabe des Forschungsprozesses durch lineare, alternierende Abfolge von Domänen (Dateneigenschaften) und Migrationsgrenzen (Aktivitäten)
Verständlichkeit	2 Abstraktionsebenen: <ul style="list-style-type: none"> <li>· Ebene 1: Phasen aus verallgemeinerten Aktivitäten;</li> <li>· Ebene 2: Auffächerung der Phasen in Einzelaktivitäten</li> </ul>	<ul style="list-style-type: none"> <li>· mehrere Abstraktionsebenen mit zunehmender Ausdifferenzierung der Aktivitäten;</li> <li>· Verwendung einer einfachen, breiten Terminologie</li> </ul>	2 Abstraktionsebenen: <ul style="list-style-type: none"> <li>· Ebene 1: Domänen aus verallgemeinerten Oberbegriffen für die Dateneigenschaft „Zugänglichkeit“ (<i>private, kollaborative, öffentliche Domäne</i>);</li> <li>· Ebene 2: Konkretisierung der Domänen durch Dateneigenschaften;</li> <li>· Verwendung einer einfachen, breiten Terminologie</li> </ul>
Nachhaltigkeit	flexible Erweiterbarkeit und Übertragbarkeit auf verschiedene Anwendungskontexte durch überwiegend breite Terminologie	flexible Erweiterbarkeit und Übertragbarkeit auf verschiedene Anwendungskontexte durch einfache, breite Terminologie	Übertragbarkeit auf verschiedene Anwendungskontexte durch <ul style="list-style-type: none"> <li>· flexible Adaptierbarkeit der Kontinuaufteilung an lokale Gegebenheiten und Anforderungen;</li> <li>· flexible Erweiterbarkeit bzw. Reduzierbarkeit der Kontinua um Domänen bzw. kontextualisierte Dateneigenschaften/ Aktivitäten</li> </ul>

Redundanzfreiheit	Hinweis auf Redundanzen in Form mehrfach implizit auftretender Kernaktivitäten „Recherchieren, Beschaffen, Prüfen/Interpretieren von Daten“ als Bestandteile der Planungs- und Nachnutzungsphase („ <i>explore existing data sources</i> “; „ <i>conduct secondary analysis</i> “, „ <i>undertake follow-up research</i> “, „ <i>conduct research reviews</i> “, „ <i>scrutinize findings</i> “)	eindeutige Beschreibung von Aktivitäten mit unterschiedlicher Bedeutung	eindeutige Beschreibung von Dateneigenschaften mit unterschiedlicher Bedeutung
Rückverfolgbarkeit	unzutreffend	Übereinstimmung mit Modellen „Cycle of Encoding and Decoding nach Münster“, „Zyklus der datenbasierten Entscheidungsfindung“ in zyklischer Abfolge von Phasen/Aktivitäten, wesentlichen Kernaktivitäten („Forschungsgegenstand identifizieren“, „Daten sammeln“, „Daten auswerten/analysieren“, „Daten/Ergebnisse interpretieren“, „datenbasiertes Entscheiden/Handeln“)	unzutreffend

Legende:

- Kriterium voll erfüllt
- Kriterium teilweise erfüllt
- Kriterium implizit erfüllt
- Kriterium implizit nicht erfüllt
- Kriterium unzutreffend