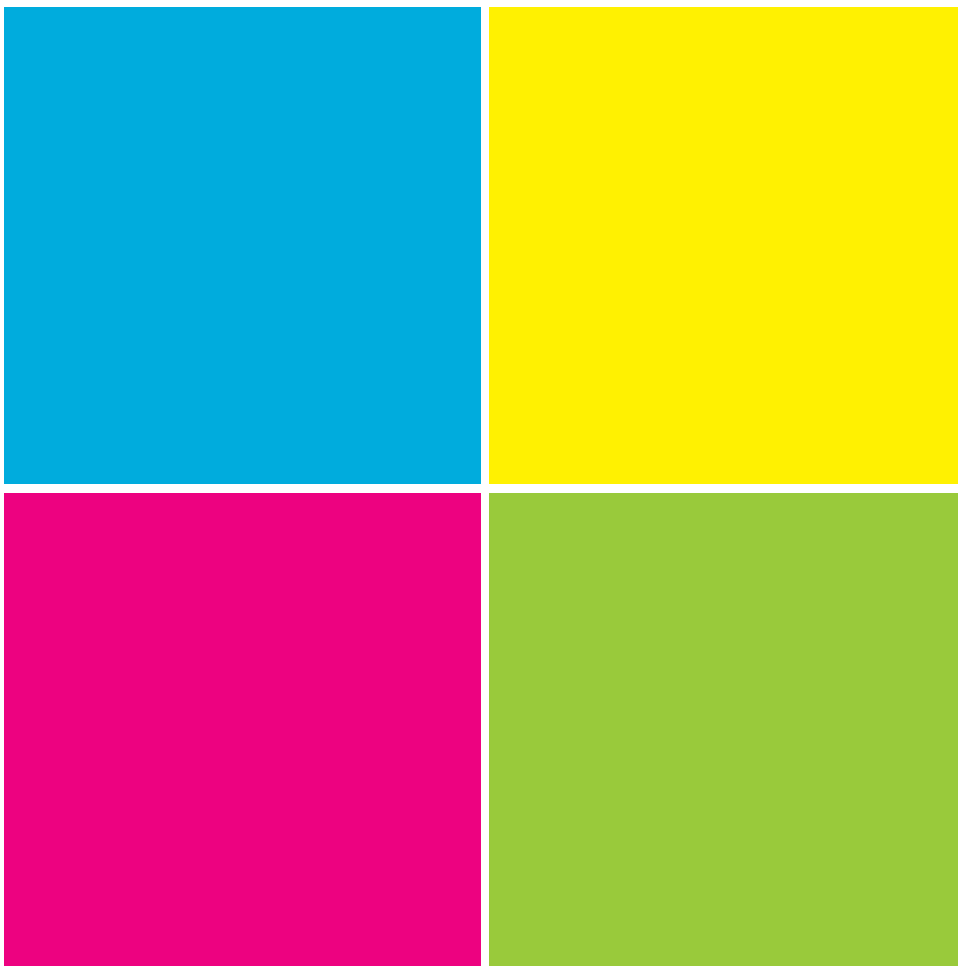


Forschungsdatenmanagement in den Sozial-, Verhaltens- und Wirtschaftswissenschaften

Orientierungshilfen für die Beantragung
und Begutachtung datengenerierender
und datennutzender Forschungsprojekte



GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Inhaltsverzeichnis

Vorbemerkungen	3
1 Basisinformationen zum Forschungsdatenmanagement in den Sozial-, Verhaltens-, und Wirtschaftswissenschaften	6
1.1 Datengenerierende Projekte (Primärdatenanalyse)	6
1.2 Datennutzende Projekte (Sekundärdatenanalyse)	7
2 Leitfragen für Datenmanagementpläne in den Sozial-, Verhaltens- und Wirtschaftswissenschaften	9
Fragebogen	10
A. Erfassung von Datenarten und Datennutzung	10
B. Dokumentation und Metadaten	11
C. Rechtliche und ethische Aspekte	12
D. Datenspeicherung und -sicherung	14
E. Archivierung und Nachnutzung der Forschungsdaten	15
F. Zuständigkeiten und Ressourcen	16
Quellen	17
Mitwirkende	18

Vorbemerkungen

■ Der RatSWD ist dem Ziel verpflichtet, Verbesserungen für die Generierung einschlägiger sozial-, verhaltens- und wirtschaftswissenschaftlicher Forschungsdaten aufzuzeigen und den Zugang der Wissenschaft zu diesen Daten herzustellen. Daher unterstützt der RatSWD Bemühungen um eine nachhaltige Sicherung und Bereitstellung qualitätsgesicherter Forschungsdaten – insbesondere aus öffentlicher Förderung – für die Wissenschaft. Solche Daten sind in vielen Forschungsfeldern eine Voraussetzung für wettbewerbsfähige und innovative Forschung.

Hierbei spielen die Entwicklung und Einhaltung methodenangemessener Standards zur Sicherstellung einer hohen Datenqualität – auch mit Blick auf die Wahrung geistigen Eigentums, die Innovationsfähigkeit des Forschungssystems, datenschutzrechtliche und forschungsethische Anforderungen – ebenso eine maßgebliche Rolle, wie die praktische Weiterentwicklung und Verbesserung entsprechender Forschungsinfrastrukturen.

Um sozial-, verhaltens- und wirtschaftswissenschaftliche Forschungsdaten nachhaltig sichern und ggf. auch bereitstellen zu können, sind verschiedene Vorkehrungen und Maßnahmen zu treffen, die bereits bei der Planung und Durchführung von Forschungsprojekten berücksichtigt werden müssen. Nationale und internationale Forschungsförderer erwarten deshalb zunehmend bereits bei der Beantragung von Forschungsprojekten die Vorlage von sogenannten Datenmanagementplänen. Die Anforderungen an die inhaltliche Gestaltung dieser Datenmanagementpläne in den einzelnen Disziplinen und bei den Institutionen der Forschungsförderung sind bislang sehr heterogen und unterscheiden sich insbesondere in der Erwartung der Detailgenauigkeit der jeweiligen Planungen.

Der RatSWD möchte mit dieser Empfehlung Antragstellenden und Gutachtenden eine Orientierungshilfe zu den grundlegenden Fragen eines modernen Forschungsdatenmanagements in den Sozial-, Verhaltens- und Wirtschaftswissenschaften geben.

Der Begriff Datenmanagement bezeichnet alle im Zusammenhang mit Forschungsdaten durchzuführenden Tätigkeiten, die sich entlang des Datenlebenszyklus ergeben. Die Tätigkeiten des Datenmanagements reichen von der Datenerhebung im Forschungsprojekt über die Datenbearbeitung (insb. Auswertungen) bis hin zur Archivierung der Forschungsdaten, wie sie beispielsweise durch eine Einrichtung der Forschungsdateninfrastruktur erfolgt.

Als Bestandteile des Datenmanagements werden im Allgemeinen unterschieden:

- Datendokumentation
- Klärung der rechtssicheren und forschungsethisch unbedenklichen Verwendung von Forschungsdaten
- sichere Aufbewahrung der Daten
- Sicherung und Nachnutzung der im Projekt generierten Daten

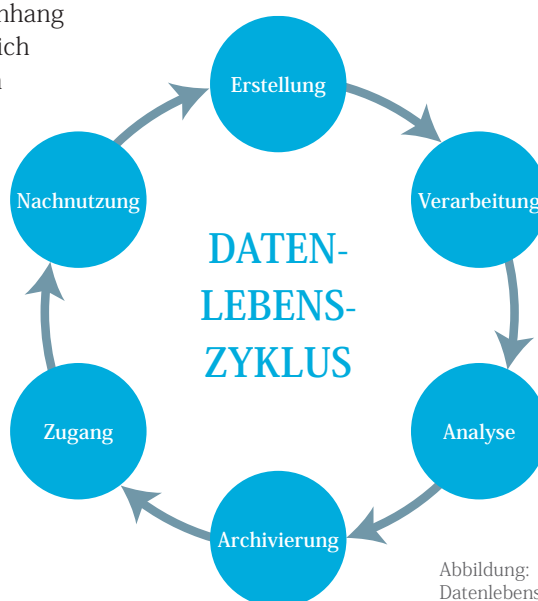


Abbildung:
Datenlebenszyklus
Quelle:
UK Data Archive

Die langfristige Sicherung von Forschungsdaten ist ganz im Sinne der „Regeln guter wissenschaftlicher Praxis“, wie der Senat der Deutschen Forschungsgemeinschaft zuletzt wieder am 30. September 2015 in den „Leitlinien zum Umgang mit Forschungsdaten“ (DFG 2015) festgestellt hat. Sie ist eine Voraussetzung für die Nachvollziehbarkeit und Replizierbarkeit von Forschungsergebnissen und trägt damit wesentlich zur Transparenz und zur Qualitätssicherung in der Wissenschaft bei. Darüber hinaus verbindet sich mit der Sicherung von – mit öffentlichen Mitteln – geförderten Forschungsdaten zunehmend auch die Erwartung einer Bereitstellung zur Nachnutzung durch die Scientific Community. Dies nicht nur deshalb, weil damit aufwändige Datenerhebungen dort vermieden werden können, wo bereits hinreichende Datenbestände existieren, sondern auch, weil eine wissenschaftliche Nachnutzung bestehender Datenbestände oftmals auch zur Beantwortung neuer inhaltlicher und methodischer Forschungsfragen sinnvoll und notwendig ist. Beides, die nachhaltige Sicherung ebenso wie die Bereitstellung der Daten tragen wesentlich zum Fortschritt empirischer Wissenschaft bei und fördern den kumulativen Erkenntnisgewinn.

Eine langfristige Sicherung (von der DFG wird beispielsweise eine Aufbewahrungsfrist von 10 Jahren genannt) oder Bereitstellung von Forschungsdaten ist jedoch in den Sozial-, Verhaltens-, und Wirtschaftswissenschaften nicht in gleicher Weise möglich wie in anderen empirischen Wissenschaften. In den meisten Fällen weisen diese Forschungsdaten einen Personenbezug auf, weil aus den Daten Rückschlüsse auf natürliche Personen und deren Identität möglich sind.

Deshalb verbinden sich mit der langfristigen Sicherung und Bereitstellung der Forschungsdaten aus den Sozial-, Verhaltens-, und Wirtschaftswissenschaften besondere datenschutzrechtliche Anforderungen, unabhängig z. B. von der Frage des Urheberrechts. Dabei geht es nicht nur um die Sicherung der Anonymität von Auskunftgebenden oder benannten Personen (z. B. in Interviews zusätzlich erwähnte Personen), sondern auch um die Beachtung und ggf. die Umsetzung ihres Rechts, ihre Einwilligung zur Mitwirkung in einem Forschungsprojekt wieder rückgängig zu machen und die dabei zur Verfügung gestellten Auskünfte und Informationen löschen zu lassen.

Eine nachhaltige Sicherung und ggf. Bereitstellung sozial-, verhaltens- und wirtschaftswissenschaftlicher Forschungsdaten im Rahmen von Open Access Policies ist deshalb in den meisten Fällen nicht möglich.

Mit den in den letzten Jahren aufgebauten und vom RatSWD akkreditierten Forschungsdatenzentren (FDZ) steht jedoch mittlerweile eine Forschungsdateninfrastruktur bereit, die eine datenschutzrechtskonforme, nachhaltige Sicherung und Nachnutzung für die meisten der in den Sozial-, Verhaltens- und Wirtschaftswissenschaften generierten Forschungsdaten ermöglicht. Die von den FDZ bereitgestellten Dienstleistungen reichen von der bloßen datenschutzrechtskonformen Archivierung und Sicherung der technischen Lesbarkeit bis hin zur Nutzerverwaltung und Bereitstellung von Scientific oder Public Use Files.

Ein wesentliches Ergebnis der Debatten über Archivierungs- und Nachnutzungsmöglichkeiten in den Sozial-, Verhaltens- und Wirtschaftswissenschaften der letzten Jahre besteht darin, dass nicht alle generierten Forschungsdaten unmittelbar einer Nachnutzung zugeführt werden können.

In manchen Fällen, so insbesondere in der qualitativen Sozialforschung und in der Bildungsforschung, ist eine ausreichende Anonymisierung der Daten überhaupt nicht oder nur durch einen Verzicht auf den wissenschaftlichen Aussagegehalt der Daten möglich. In anderen Fällen ist der Aufwand der Anonymisierung sehr hoch und steht in keinem Verhältnis zu dem wissenschaftlichen Ertrag, der von einer Nachnutzung zu erwarten wäre.

Um vor diesem Hintergrund trotzdem den Regeln guter wissenschaftlicher Praxis und den datenschutzrechtlichen Anforderungen an den Umgang mit Forschungsdaten auf der einen Seite und den Forderungen nach der Bereitstellung öffentlich geförderter Forschungsdaten auf der anderen Seite gerecht werden zu können, schlägt der RatSWD im folgenden drei Varianten eines nachhaltigen Umgangs mit Forschungsdaten vor (vgl. Stanat 2014):

- (1) nachhaltige Sicherung von Forschungsdaten ohne nutzerfreundliche Dokumentation und ohne Nachnutzungsmöglichkeiten
- (2) nachhaltige Sicherung von Forschungsdaten mit nutzerfreundlicher Dokumentation und mit restriktiven Nutzungsmöglichkeiten
- (3) nachhaltige Sicherung und Bereitstellung von Forschungsdaten mit nutzerfreundlicher Dokumentation für die Datennutzung außerhalb des Aufbewahrungsortes

Je nachdem, welche Variante der Sicherung und Nachnutzung gewählt wird, müssen unterschiedliche ressourcenrelevante Maßnahmen und Arbeitsschritte bereits bei der Projektplanung berücksichtigt werden.

Ein entsprechend formulierter Datenmanagementplan erleichtert dabei die Projektplanung und insbesondere die Projektbudgetierung.

Mit den vorliegenden Leitfragen sollen Antragstellende bei der Planung und Beantragung datengenerierender und -nutzender Forschungsprojekte in den Sozial-, Verhaltens- und Wirtschaftswissenschaften und Gutachtende bei der Bewertung dieser Anträge unterstützt werden.

- Der erste Abschnitt des Leitfadens enthält eine Empfehlung zu den Basisinformationen, die bei der Beantragung von forschungsdatengenerierenden und -nutzenden Forschungsprojekten von Seiten der Antragstellenden gegenüber dem Forschungsförderer und involvierten Gutachtenden bereitgestellt werden sollten.
- Im zweiten Abschnitt werden die für die Erstellung detaillierter Datenmanagementpläne als zentral eingeschätzte Leitfragen vorgestellt und näher erläutert.

Sowohl die Empfehlung zu den Basisinformationen als auch die Leitfragen für die Erstellung von Datenmanagementplänen basieren zum einen auf den derzeit verfügbaren Datenmanagementplänen nationaler oder internationaler Forschungsförderer und Infrastruktureinrichtungen und zum anderen auf den operativen Erfahrungen mit der Datenarchivierung und -bereitstellung seitens der im Ständigen Ausschuss Forschungsdateninfrastruktur des RatSWD vertretenen Forschungsdatenzentren.

1. Basisinformationen zum Forschungsdatenmanagement in den Sozial-, Verhaltens-, und Wirtschaftswissenschaften

■ Im Folgenden werden die Basisinformationen für Projekte vorgestellt und dabei danach unterschieden, ob die Projekte selbst Daten generieren (1.1) oder im Rahmen einer Sekundärdatenanalyse nachnutzen (1.2).

1.1 Datengenerierende Projekte (Primärdatenanalyse):

A. Angaben zu Datensicherheit, Archivierung und Weitergabe der Forschungsdaten

Angaben zu den technischen und organisatorischen Maßnahmen zur Umsetzung der Datensicherheit (Sicherung vor Verlust, Manipulation/Verfälschung, unberechtigtem Zugang) bei Primärerhebungen und bei Sekundäranalysen sollten dokumentiert werden. Dies beinhaltet u.a. Angaben zur Einverständniserklärung, zur Privilegierung von Forschung wie auch zu Dokumentationspflichten.

Für alle Daten gilt eine in den Regeln guter wissenschaftlicher Praxis festgelegte 10-jährige Aufbewahrungsfrist (vgl. DFG 2015). Diese gilt sowohl für im Projekt generierte Primärdaten wie auch für – im Fall von sekundäranalytischen Projekten – an den Originaldaten vorgenommene Modifikationen (Syntaxfiles). Eine Sicherung kann in einer der drei folgenden Varianten erfolgen:

Variante 1:

Sicherung der Daten in einem Repositorium der Forschungseinrichtung des Datenproduzenten (Universität, Forschungsinstitut) oder Übergabe an eine Einrichtung der Forschungsdateninfrastruktur ohne Zugriffs- oder Nutzungsrechte für Dritte (<http://www.ratswd.de/forschungsdaten/fdz>).

Ziel ist es, die erhobenen Daten im Sinne der datenschutzrechtlichen Anforderungen und vor dem Hintergrund des informationstechnischen Wandels zu sichern. Ein Zugriff für andere Wissenschaftlerinnen und Wissenschaftler im Rahmen sekundäranalytischer Nachnutzungen ist aus datenschutzrechtlichen oder auch anderen, im Antrag darzustellenden, Gründen nicht vorgesehen. Die Dokumentation muss nicht nutzerfreundlich sein.

Variante 2:

Sicherung der Daten und des Datenzuganges auch für andere Wissenschaftlerinnen und Wissenschaftler zwecks sekundäranalytischer Nachnutzung unmittelbar beim Datenproduzenten oder einer Einrichtung der Forschungsdateninfrastruktur.

Die Forschungsdaten (Primärdaten) werden wie in Variante 1 archiviert, können jedoch vor Ort (an der Forschungseinrichtung selbst oder einer Infrastruktureinrichtung) auch von anderen Wissenschaftlerinnen und Wissenschaftlern im Rahmen von Gastaufenthalten genutzt werden. Aufgrund dieser Nutzungsmöglichkeiten entsteht ein höherer Anonymisierungs-, Aufbereitungs- und Dokumentationsaufwand, der in der Ressourcenausstattung des Projekts berücksichtigt werden sollte.

Variante 3:

Sicherung und Bereitstellung der Daten zwecks sekundäranalytischer Nachnutzung durch Übergabe an eine Einrichtung der Forschungsdateninfrastruktur

Die Daten werden an eine Infrastruktureinrichtung übergeben und dort zur Nutzung in Form eines Scientific Use Files (SUF), Public Use Files (PUF), einer Variante der Datenfernverarbeitung oder im Rahmen von Gastaufenthalten bereitgestellt. Hierzu sind besondere Maßnahmen der Anonymisierung, der Datensatzaufbereitung und der Dokumentation vorzusehen.

Falls eine Übergabe an eine bestehende Einrichtung der Forschungsdateninfrastruktur geplant ist, sollte diese im Förderantrag benannt und bereits vorab kontaktiert werden, um die Übergabemodalitäten und insbesondere den Ressourcenbedarf zu klären.

B. Angaben zu Datenschutz und Forschungsethik

Bei Primärerhebungen sollten zum Zeitpunkt der Antragstellung Angaben gemacht werden, auf welche der drei Sicherungsvarianten sich die verwendeten Einwilligungserklärungen für Studienteilnehmende beziehen. Angaben zu datenschutzrechtlichen Vorkehrungen während der Projektlaufzeit und nach Ablauf des Projektes sollten kurz benannt werden, ebenso wie die Information, ob das Votum einer Ethikkommission eingeholt wird.

C. Angaben zur methodischen, technischen und bibliographischen Dokumentation der Daten

Entsprechend der geplanten Sicherungsvariante sollten Angaben zur methodischen, technischen und bibliographischen Dokumentation und den dabei ggf. verwendeten Standards (z.B. den technischen Formaten, den Metadatenformaten) gemacht werden. Bei der Projektplanung sollten die dafür notwendigen Arbeitspakete spezifiziert und der Ressourcenbedarf berücksichtigt werden. Ebenfalls sollte kurz erläutert werden, ob im Rahmen des Projekts ein Datenmanagementplan erstellt wird und auf welche Vorlagen hierfür zurückgegriffen wird.

1.2 Datennutzende Projekte (Sekundärdatenanalyse):

In Forschungsprojekten, in denen selbst keine Daten generiert werden, sondern auf bereits existierende Forschungsdaten zurückgegriffen wird, sollten die im Projekt vorgenommenen Arbeiten zur Datenaufbereitung und Datenmodifikation dokumentiert und nach Projektabschluss bereitgestellt werden (z.B. mit Programmfiles zur Datenaufbereitung).

Die Bereitstellung der für die Projektergebnisse relevanten datenverändernden Maßnahmen ist für eine spätere Rekonstruktion der Datengrundlage und der erlangten Forschungsergebnisse und deren Replikation unabdingbar. Bereits jetzt erwarten einzelne Fachzeitschriften die Bereitstellung der kompletten Programmcodes, um die einzelnen Arbeitsschritte von der Aufbereitung der Grunddaten bis hin zu den in den Publikationen präsentierten Tabellen und Grafiken nachvollziehbar dokumentieren zu können.

In den Forschungsdatenzentren werden gesonderte Repositorien aufgebaut, um den Datennutzenden die Archivierung und Bereitstellung ihrer Programmcodes zu ermöglichen (z.B. datorium bei der GESIS). Projektanträge sollten deshalb Aussagen darüber enthalten, wie die Aufbereitung der Sekundärdaten dokumentiert wird und in welcher Form und in welchem Repository diese Arbeiten nach Projektabschluss bereitgestellt werden. Analog zur Sicherung bzw. Nachnutzung von Primärdaten sind hier drei Varianten möglich:

Variante 1:

Mindestens 10-jährige Aufbewahrung der Programmcodes, der datenmodifizierenden Arbeiten und abgeleiteten Daten in einem Repository der Forschungseinrichtung des Sekundärdatennutzers (Universität, Forschungsinstitut) oder Übergabe an eine Einrichtung der Forschungsdateninfrastruktur ohne Zugriffs- oder Nutzungsrechte Dritter. Die Dokumentation muss dabei nicht zwingend nutzerfreundlich sein.

Variante 2:

Mindestens 10-jährige Aufbewahrung der Programmcodes, der datenmodifizierenden Arbeiten und abgeleiteten Daten in einem Repository der Forschungseinrichtung des Sekundärdatennutzers (Universität, Forschungsinstitut) oder Übergabe an eine Infrastruktureinrichtung mit restriktiven Nutzungsrechten für Dritte nach expliziter Genehmigung durch den Sekundärdatennutzer. Die Dokumentation sollte nutzerfreundlich sein.

Variante 3:

Sicherung und mindestens 10-jährige Aufbewahrung in einem Repository der Forschungseinrichtung des Sekundärdatennutzers (Universität, Forschungsinstitut) bzw. Übergabe an eine Einrichtung der Forschungsdateninfrastruktur mit erweiterten Nutzungsrechten für Dritte. Die Programmcodes, datenmodifizierenden Arbeiten und abgeleiteten Daten, die nutzerfreundlich dokumentiert sein sollten, sind von den Sekundärdatennutzern für die Nutzung auch durch andere Wissenschaftlerinnen und Wissenschaftler unter Einhaltung der üblichen Zitationsregeln freigegeben.

Auch hier sind unterschiedliche Dokumentationsumfänge in den einzelnen Varianten bei der Ressourcenplanung zu berücksichtigen. Bei einer geplanten Übergabe an eine Infrastruktureinrichtung sollten im Zuge der Antragstellung vorab die Möglichkeiten und Bedingungen der Sicherung bzw. Nachnutzung geklärt werden.

2. Leitfragen für Datenmanagementpläne in den Sozial-, Verhaltens- und Wirtschafts- wissenschaften

■ Die vorliegenden Leitfragen behandeln – so umfassend wie möglich und so detailliert wie nötig – die verschiedenen Aspekte eines Datenmanagementplans in den Bereichen Datendokumentation, rechtssichere Verwendung, Aufbewahrung, Archivierung und Nachnutzung von Forschungsdaten. Dazu wurden zu jedem der genannten Bereiche entsprechende Leitfragen entwickelt und um Hinweise ergänzt. Grundlage für dieses Dokument bilden bestehende nationale und internationale Vorlagen für Datenmanagementpläne und die Erfahrungen von Forschungsdatenzentren mit der Archivierung und Bereitstellung von Forschungsdaten aus den Sozial-, Verhaltens- und Wirtschaftswissenschaften in Deutschland. Diese Leitfragen können als Orientierungshilfe für die Erstellung und Bewertung von Datenmanagementplänen sowie für die Erarbeitung disziplinspezifischer Regularien unterschiedlicher Fachrichtungen verstanden werden.

Bei der Erstellung eines Datenmanagementplans (DMP) geht es darum, die hierfür notwendigen Tätigkeiten möglichst zielorientiert, systematisch und effizient zu bestimmen. Der Nutzen eines derartigen, formalisierten DMP unter Zuhilfenahme von Leitlinien besteht darin, durch eine frühzeitige und umfassende Planung Beschränkungen in der späteren Datennutzung zu vermeiden. Es geht darum,

- eine effiziente, rechtssichere und qualitätsorientierte Nutzung der Daten in den datengenerierenden und datennutzenden Forschungsprojekten selbst zu ermöglichen und
- die Replizierbarkeit bzw. Transparenz und Nachvollziehbarkeit der Forschungsergebnisse, die auf diesen Daten beruhen, anhand der Möglichkeit der Nachnutzung der Daten zu gewährleisten.

Schließlich ist ein systematisches Datenmanagement eine notwendige Voraussetzung für die Nachhaltigkeit der Forschungsdaten und den Erhalt des kulturellen Wertes, den diese darstellen.

Um für die Forschung hilfreich sein zu können, sollten DMP frühzeitig erstellt werden, d. h. am besten bereits bei Projektplanung und Antragstellung. DMP sind keine statischen, sondern dynamische Dokumente, die an den laufenden Forschungsprozess anzupassen sind. Das vorliegende Dokument stellt hierfür entsprechende allgemeine Leitlinien vor, die zur Orientierung für das spezifische Forschungsdatenmanagement dienlich sind.

Je nach Forschungsdesign, -frage und -schwerpunkt und das jeweils am Ende stehende Datenprodukt (Scientific Use File, Public Use File etc.) sind die vorgestellten Leitlinien flexibel auf das individuelle Forschungsvorhaben übertragbar, anpassbar und entfalten unterschiedlich starke Relevanz.

Fragebogen

A. Erfassung von Datenarten und Datennutzung

Bitte beachten Sie im Folgenden, dass abhängig von Forschungsdesign, -frage und -schwerpunkt (z.B. bei einer Sekundäranalyse) unter Umständen nicht alle Leitfragen relevant sein müssen.

A1. Welche Daten erheben oder verwenden Sie?

Hinweis: Bitte geben Sie eine kurze Beschreibung aller Daten, die im Projekt erhoben oder sekundäranalytisch genutzt werden (also auch bereits vorhandener oder von Dritten erhobener Daten), an. Eine Übersicht zu Einstiegspunkten für die Suche nach nutzbaren Forschungsdaten findet sich auf der Webseite des RatSWD (<http://www.ratswd.de/forschungsdaten/fdz>).

Leitfragen:

- Welche Art von Daten (aus Primärerhebung/Sekundärdaten; quantitativ/qualitativ/Mixed Methods) werden im Projekt erhoben oder verwendet?
- Können bereits existierende Daten genutzt werden und welche Nutzungsbedingungen sind daran gebunden?
- Auf welchen Wegen werden bereits existierende Daten genutzt (Gastaufenthalt, Scientific Use File, Datenfernverarbeitung, Remote-Access, direkte Datenlieferung durch den Datenproduzenten)?
- Werden Daten aus unterschiedlichen Quellen verknüpft?
- Welche Datenerhebungsverfahren werden eingesetzt?
- Welche Maßnahmen zur Qualitätssicherung werden im Rahmen der Datenerhebung ergriffen?

A2. Wie werden die Forschungsdaten innerhalb und nach Ablauf des Projekts genutzt?

Hinweis: Die Formen der Nutzung sind vorab zu klären, weil sich daraus datenschutzrechtliche, technische und organisatorische Anforderungen ergeben. Bitte bedenken Sie die Konsequenzen und Aufwände für die Datenaufbereitung und -dokumentation, die je nach Nachnutzungsvariante variieren können. Die Dokumentation sollte so ausgestaltet sein, dass eine Nachnutzung der Daten möglich ist (z.B. entsprechende Formulierung in der Einverständniserklärung).

Leitfragen:

- Welche Nachnutzungsmöglichkeit der Daten planen Sie für die Zeit nach dem Projektende?
- Werden Daten an mehreren Projektstandorten genutzt?

Variante 1:

Die Daten werden lediglich mit einer Minimaldokumentation in einem Repository der Forschungseinrichtung oder einer Einrichtung der Forschungsdateninfrastruktur archiviert. Es besteht keine Möglichkeit der Nachnutzung.

Variante 2:

Die Daten werden archiviert und nutzerfreundlich dokumentiert und es besteht ein kontrollierter Datenzugang für Replikationen unmittelbar vor Ort beim Datenproduzenten oder in einer Einrichtung der Forschungsdateninfrastruktur.

Variante 3:

Die Daten werden für sekundäranalytische Nachnutzungen nutzerfreundlich dokumentiert an eine Einrichtung der Forschungsdateninfrastruktur übergeben

B. Dokumentation und Metadaten

Forschungsdaten sollten so dokumentiert werden, dass für Dritte langfristig nachvollziehbar und transparent ist, wie die Ergebnisse und Interpretationen des Primärprojekts auf Basis dieser Daten zustande kamen. Darüber hinaus soll eine vollständige Dokumentation der Daten Dritte dazu befähigen, die Analysepotentiale des Datenmaterials sowie die methodische Beschaffenheit desselben zu beurteilen und sie in angemessener Weise für eigene Forschungszwecke zu nutzen. Der Umfang und die Nutzerfreundlichkeit der Dokumentation sind abhängig von der geplanten Variante der Sicherung bzw. Nachnutzung.

Um die Nachvollziehbarkeit der Forschungsdaten zu gewährleisten, sollte die Dokumentation Informationen dazu enthalten, wer die Daten wie generiert hat. Die Dokumentation sollte, soweit möglich, Einzelheiten zur Untersuchungspopulation, zur Stichprobenziehung bzw. zum Feldzugang und zum Untersuchungsraum enthalten und die angewandten qualitativen und quantitativen Methoden sowie den Prozess der Datenerhebung, -verarbeitung und -analyse beschreiben (ggf. inklusive verwendeter Syntax-Dateien/Skripte).

Hinweis: Überlegen Sie, wie Sie diese Informationen erfassen können und wo sie in welcher Form gespeichert werden sollen. Wenn möglich, legen Sie vorhandene und etablierte Standards zur Dokumentation von Daten zugrunde.

Leitfragen:

- Welche methodischen und technischen Informationen werden benötigt, damit die Daten – in Abhängigkeit zur gewählten Sicherungsvariante – zukünftig verwendet werden können?
- Primärerhebung: Wie werden Datenerhebung, Datenaufbereitung, Datenauswertung/-analyse und Datenspeicherung/Datenarchivierung dokumentiert?
- Sekundäranalyse: Wie werden die Programmcodes dokumentiert, die für die Datenaufbereitung und Datenanalyse in den jeweils genutzten Statistikprogrammen (z.B. SAS, SPSS, STATA, R) erzeugt werden? Wie wird sichergestellt, dass diese auch nach Projektende für den Antragsteller oder Dritte verfügbar sind?
- Werden bestehende disziplinäre Dokumentations- und Metadatenstandards (z.B. DDI, Dublin Core, ISO etc.) angewendet und wenn ja welche?
- Wie wird – insbesondere bei qualitativen Erhebungsprojekten – die Erhebungssituation dokumentiert?

C. Rechtliche und ethische Aspekte

Rechtliche Richtlinien und ethische Aspekte, die zum Teil von Wissenschaftsorganisationen formalisiert sind, nehmen Einfluss darauf, in welcher Form welche Daten erhoben werden dürfen, wer diese Daten verarbeiten darf und wie lange sie gespeichert werden dürfen.

Sich mit rechtlichen und ethischen Fragen auseinanderzusetzen beinhaltet

- die Einhaltung forschungsethischer Aspekte, welche die Voraussetzung für die Publikation der Ergebnisse ist,
- das Einholen von informierten Einwilligungen der Betroffenen,
- Angaben zur Anonymisierung der Daten, zur Löschung von personenbezogenen Informationen bzw. deren Aufbewahrung sowie
- die Regulierung der Zugriffsbedingungen zu den Daten.

Hinweis: Sie sollten zeigen, dass Sie sich ethischer Aspekte des Forschungsprozesses und rechtlicher Regelungen zum Datenschutz bewusst sind und den Umgang mit den Daten entsprechend geplant haben. Zur Festlegung der projektspezifischen Besonderheiten und zur Klärung von Fragen sollte ggf. ein Datenschutzbeauftragter oder Justiziar zur Verfügung stehen. Weitere Erörterungen zum Datenschutz in den Sozialwissenschaften finden sich beispielsweise im Working Paper Nr. 90 (Häder 2009) und Nr. 238 (Arbeitsgruppe Datenschutz und qualitative Sozialforschung 2014) des RatSWD sowie in der Mustervorlage für Einwilligungserklärungen auf der Website des Projektes „Verbund Forschungsdaten Bildung“.

C.1 Wie gewährleisten Sie die Einhaltung des Datenschutzes?

Leitfragen:

- Werden personenbezogene Daten erhoben?
- Werden personenbezogene Daten (dauerhaft) gespeichert?
- Wie wird die Einwilligungserklärung zur Teilnahme an der Untersuchung von den Informanten eingeholt, wie sieht diese aus (Inhalte) und wie werden die Einwilligungserklärungen langfristig gespeichert?
- Sind sämtliche geplanten Verwendungszwecke der Daten Bestandteil der Einwilligung (z. B. Grund der Datenerhebung, Datenübermittlung, Datennachnutzung, Langzeitarchivierung)?
- Wie sieht Ihr Anonymisierungskonzept aus, um die Identität der Teilnehmenden zu schützen?
- Wurde das Anonymisierungskonzept durch die zuständigen Datenschutzbeauftragten bzw. eine vergleichbare Instanz geprüft?
- Sind die Daten aufgrund gesetzlicher oder vertraglicher Verpflichtungen zu einem bestimmten Zeitpunkt zu löschen? Wie wird das Löschen dieser Daten technisch sichergestellt?

Personenbezogene Daten sind Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmaren natürlichen Person. Um Angaben über eine bestimmte Person handelt es sich, wenn die Daten mit dem Namen der betroffenen Person verbunden sind oder sich aus dem Inhalt bzw. dem Zusammenhang der Bezug unmittelbar herstellen lässt. Bestimmbar ist eine Person, wenn ihre Identität unmittelbar oder mittels Zusatzwissen festgestellt werden kann.

Die **Anonymisierung** ist das Verändern personenbezogener Daten derart, dass diese Daten nicht mehr einer Person zugeordnet werden können.

Bei der **Pseudonymisierung** wird der Name oder ein anderes Identifikationsmerkmal durch ein Pseudonym (zumeist eine mehrstellige Buchstaben- oder Zahlenkombination, auch Code genannt) ersetzt, um die Identifizierung des Betroffenen auszuschließen oder wesentlich zu erschweren.

C.2 Wie berücksichtigen Sie relevante ethische Aspekte?

Leitfragen:

- Wie wurden forschungsethische Aspekte beim Datenmanagementplan (DMP) berücksichtigt?
- Ist der Antrag einer Ethikkommission vorgelegt worden? Welche war dies?
- Warum war die Vorlage vor einer Ethikkommission gegebenenfalls nicht notwendig?

C.3 Wie gehen Sie mit Urheberrechtsfragen und dem Recht am geistigen Eigentum um?

Hinweis: Erläutern Sie, wer das Urheberrecht an den Daten hat und wer Verwertungsrechte an den Daten besitzt, die Sie generieren. Falls Sie nicht Urheber der verwendeten Daten sind, welche Regelungen und Lizenzmodelle/-vereinbarungen für die Nutzung und Nachnutzung bestehen?

Für Verbundprojekte empfiehlt es sich, Urheberschaft und Verwertungsrechte in einem Abkommen der Verbundpartner festzuhalten. Berücksichtigen Sie alle relevanten vertraglichen Regelungen (inkl. Arbeitsverträge) und institutionellen Richtlinien der Geldgeber, Abteilungen oder beteiligten Gruppen im Hinblick auf das Urheberrecht.

Falls Sie die Sekundärnutzung von Daten planen, beachten Sie mögliche Restriktionen z. B. hinsichtlich des Datenzugangs, deren Verwendung bzw. Nachnutzung sowie der Ergebnisse Ihrer Forschung mit diesen Daten.

Leitfragen:

- Wer wird nach Abschluss der Erhebung im Sinne des Urheberrechts Urheber der Daten sein und wer besitzt Verwertungsrechte?
- Falls eine Nachnutzung möglich ist, wie wird diese vertraglich geregelt? Wird dabei auf ein vorliegendes Lizenzmodell zurückgegriffen?
- Werden durch andere generierte Forschungsdaten, Software, Erhebungsverfahren (insb. Fragebögen, Tests), Materialien (z. B. Bilder) etc. verwendet, die dem Urheberrecht oder anderen geistigen Eigentumsrechten unterliegen?

D. Datenspeicherung und -sicherung

Hinweis: Zu empfehlen ist der Einsatz stabiler und administrierter Speichermöglichkeiten mit Unterstützung durch Ihre IT-Abteilung. Es ist in der Regel besser, automatische Sicherungsservices der IT zu verwenden, als sich auf manuelle Prozesse zu verlassen. Wenn Sie einen Drittanbieterdienst für die Speicherung und/oder Sicherung nutzen, sollten Sie sicherstellen, dass dies nicht gegen Datenschutzbestimmungen oder andere institutionelle Regelungen verstößt.

D.1 Wie werden die Daten während des Forschungsprozesses gespeichert und gesichert?

Leitfragen:

- Werden Versionierungen der Dateien vorgenommen und wie erfolgt dies?
- Wie werden die Daten gesichert, d.h. welche Art von Sicherung wird in welchen Intervallen durchgeführt?
- Ist sichergestellt, dass ausreichende Kapazitäten für Speicherung und Sicherung zur Verfügung stehen?

D.2 Wie wird der Zugriff auf die Daten verwaltet und werden die Daten vor Zugriffen Unbefugter geschützt?

Forschungsdaten sind während des Forschungsprozesses durch geeignete Sicherungsmaßnahmen vor dem Zugriff und der Manipulation durch Unbefugte zu schützen.

Hinweis: Beschreiben Sie, welche Sicherungsmaßnahmen (Firewall, Virenskan, Verschlüsselung, Authentifizierung) angewendet werden. Besondere Vorsichtsmaßnahmen sind dabei für personenbezogene Daten zu berücksichtigen. Bei Unklarheiten ist eine Konsultation des Datenschutzbeauftragten oder eines Justizars ratsam.

Leitfragen:

- Wie wird verhindert, dass Unbefugte auf die Daten zugreifen können?
- Wie werden Daten bei der Übermittlung (z.B. zwischen den im Feld eingesetzten Systemen und den Systemen am Arbeitsplatz oder zwischen Projektmitarbeitenden unterschiedlicher Einrichtungen) geschützt?

E. Archivierung und Nachnutzung der Forschungsdaten

E.1 Welche Daten müssen gelöscht werden?

Hinweis: Bedenken Sie, welche Vorgaben für die Datenarchivierung existieren (durch Drittmittelgeber, Universitäten, Forschungsinstitute und -abteilungen). Berücksichtigen Sie dabei auch die 10-jährige Aufbewahrungsempfehlung der DFG zur „Sicherung guter wissenschaftlicher Praxis“. Erläutern Sie auch, welche Daten nicht erhalten werden sollen und warum nicht.

Leitfragen:

- Welche Daten können oder müssen von langfristigen Erhaltungsmaßnahmen ausgenommen und dauerhaft gelöscht werden (z. B. bestimmte Pretestdaten aufgrund ihrer lediglich temporären Bedeutung. Bitte kurz begründen.)? Wie werden diese Daten vernichtet?

E.2 Wie werden die Daten für die Nachnutzung zur Verfügung gestellt?

Hinweis: Denken Sie darüber nach ob, wo, wie und wem Daten zur Nachnutzung zur Verfügung gestellt werden sollen. Möglichkeiten der Nachnutzung der Daten hängen von einer Reihe von Faktoren ab, wie beispielsweise der Art, der Größe, der Komplexität und der Vertraulichkeit der Daten.

Leitfragen:

- Wie werden die Daten zugänglich gemacht (z. B. Repositorium, Forschungsdatenzentrum)?
- Über welche Zugangswege sollen die Daten für Sekundäranalysen nutzbar sein (z. B. Gastaufenthalt, Scientific Use File, Public Use File)?
- Zu welchem Zeitpunkt werden Sie die Daten voraussichtlich zur Verfügung stellen?
- Ist sichergestellt, dass die Daten unter einem digitalen Identifikator (z. B. DOI) registriert und dauerhaft auffindbar sind?
- In welchen Dateiformaten stehen die Daten für die Nachnutzung zur Verfügung?
- Wie soll die (nutzerfreundliche) Dokumentation gewährleistet werden?

E.3 Sind Beschränkungen für die Nachnutzung erforderlich?

Hinweis: Falls dies aufgrund der Art der erhobenen Daten bereits absehbar ist: Benennen Sie die Restriktionen, Zugangsbeschränkungen und Bedingungen der Nutzung Ihrer Daten durch andere Wissenschaftlerinnen und Wissenschaftler, insbesondere geplante Sperrfristen.

Leitfragen:

- Wer kann die Daten unter welchen Bedingungen bzw. Restriktionen nachnutzen?
- Bestehen forschungsethische, datenschutzrechtliche, anonymisierungstechnische oder methodologische Einschränkungen der Nachnutzung der Daten?

F. Zuständigkeiten und Ressourcen

Datenmanagement ist ein dynamischer Prozess, der an den laufenden Forschungsprozess anzupassen ist. Dies erfordert ggf. auch eine Aktualisierung des Datenmanagementplans.

Hinweis: Skizzieren Sie die Rollen und Verantwortlichkeiten für alle Aktivitäten im Rahmen des DMP. Erörtern Sie, wer für die Einhaltung entsprechender Vereinbarungen verantwortlich ist. Beschreiben Sie, in welchen Abständen dieser Datenmanagementplan aktualisiert wird. Bedenken Sie alle personellen, technischen und finanziellen Ressourcen, die nötig sind, um Ihren Datenmanagementplan umzusetzen (z.B. Software, Hardware, technische Expertise, datenschutzrechtliche oder ethische Beratung, zusätzliche Personalmittel zur Anonymisierung und Dokumentation). Listen Sie die benötigten Ressourcen auf und begründen Sie deren Erfordernis. Denken Sie daran, notwendige Personal- und Sachmittel für alle Schritte des Datenmanagements bei der Antragstellung mit zu berücksichtigen.

F.1 Welche personellen und materiellen Ressourcen werden benötigt, um den Datenmanagementplan umzusetzen?

Leitfragen:

- Wer ist verantwortlich für die Umsetzung des Datenmanagementplans und gewährleistet dessen Überprüfung und Aktualisierung?
- Wird zusätzliche Fachkompetenz und Expertise benötigt? Wie kann diese beschafft werden (z. B. Inanspruchnahme von Beratungsleistungen durch Forschungsdatenzentren oder Schulungen der Mitarbeiterinnen und Mitarbeiter)?
- Welche zusätzlichen Kosten fallen für das Datenmanagement an?
- Sind die Aufwendungen und Kosten, die im Rahmen des Datenmanagement anfallen werden, im beantragten Projektbudget berücksichtigt?

Quellen

- Arbeitsgruppe Datenschutz und qualitative Sozialforschung des RatSWD (2014): Datenschutzrechtliche Anforderungen bei der Generierung und Archivierung qualitativer Interviewdaten. http://www.ratswd.de/dl/RatSWD_WP_238.pdf [31.05.2016]
- CESSDA Training, <http://cessda.net/CESDDA-Training> [31.05.2016]
- DFG (2015): Leitlinien zum Umgang mit Forschungsdaten. http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf [31.05.2016]
- Data Management Plans Online, <https://dmponline.dcc.ac.uk> [31.05.2016]
- ESRC (2014): ESRC Research Data Policy. <http://www.esrc.ac.uk/files/about-us/policies-and-standards/esrc-research-data-policy/> [31.05.2016]
- ESRC (2013): Data management plan – guidance for peer reviewers. <http://www.esrc.ac.uk/files/about-us/policies-and-standards/data-management-plan-guidance-for-per-reviewers/> [31.05.2016]
- Forschungsdaten Bildung (2015): Checkliste zur Erstellung eines Datenmanagementplans in der empirischen Bildungsforschung. <http://www.forschungsdaten-bildung.de/datenmanagementplan> [31.05.2016]
- Horizon 2020 (2013): Guidelines on Data Management in Horizon 2020. Version 1.0. http://www.gsrt.gr/EOX/files/h2020-hi-oa-data-mgt_en.pdf [31.05.2016]
- Häder, Michael (2009): Der Datenschutz in den Sozialwissenschaften. Anmerkungen zur Praxis sozialwissenschaftlicher Erhebungen und Datenverarbeitung in Deutschland. http://www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_90.pdf [31.05.2016]
- ICPSR (2012): Guidelines for Effective Data Management Plans. <http://www.icpsr.umich.edu/files/datamanagement/DataManagementPlans-All.pdf> [31.05.2016]
- Jensen, Uwe (2012): Leitlinien zum Management von Forschungsdaten. Sozialwissenschaftliche Umfragedaten. GESIS-Technical Reports 2012|07. http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-07.pdf [31.05.2016]
- National Science Foundation (2010): Data Management for NSF SBE Directorate Proposals and Awards. http://www.nsf.gov/sbe/SBE_DataMgmtPlanPolicy.pdf [31.05.2016]
- Stanat, Petra (2014): Bereitstellung und Nutzung quantitativer Forschungsdaten in der Bildungsforschung: Memorandum des Fachkollegiums „Erziehungswissenschaft“ der DFG, S. 4-5, http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten_bildungsforschung.pdf [31.05.2016]
- UK Data Archive, Research Data Life Cycle, <http://www.data-archive.ac.uk/create-manage/life-cycle> [31.05.2016]

Mitwirkende bei der Erstellung

Prof. Dr. Stefan Liebig (Vorsitz)

Universität Bielefeld

Dr. Laura Castiglioni

pairfam

Tobias Gebel

Universität Bielefeld

Prof. Dr. Monika Jungbauer-Gans

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), RatSWD

Dr. Poldi Kuhl

Institut zur Qualitätsentwicklung im Bildungswesen (IQB)

Alexia Meyermann

Deutsches Institut für Internationale Pädagogische Forschung (DIPF)

Dr. Sebastian Netscher

GESIS – Leibniz Institut für Sozialwissenschaften

Dr. Sophie Rosenbohm (bis Sept. 2015)

Universität Bielefeld

Dr. Thomas Schmidt

Institut für Deutsche Sprache (IDS)

Prof. Dr. Jörg Strübing

Eberhard Karls Universität Tübingen, RatSWD

PD Dr. Erich Weichselgartner

Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID)

Impressum

Herausgeber:

Rat für Sozial- und Wirtschaftsdaten (RatSWD)

Chausseestraße 111

10115 Berlin

office@ratswd.de

www.ratswd.de

Redaktion:

Prof. Dr. Stefan Liebig

Gestaltung/Satz:

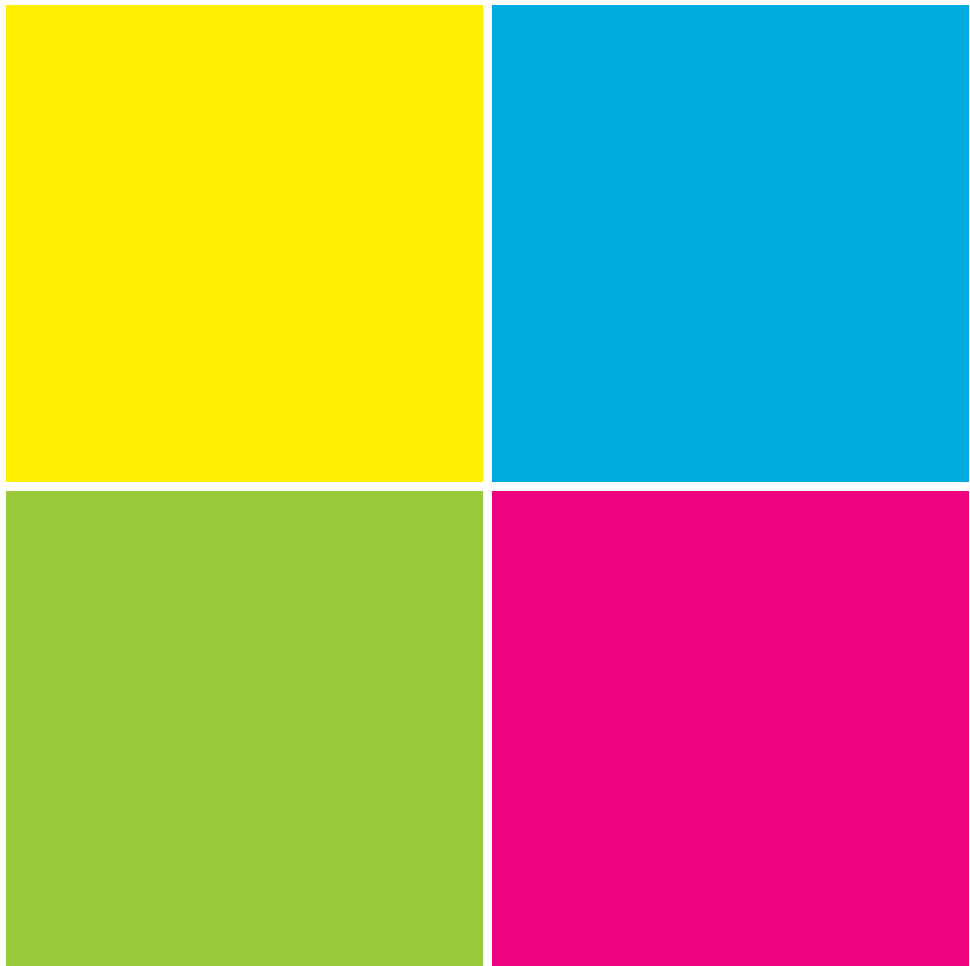
Markus Weiß

Berlin, Juni 2016

RatSWD Output:

Die RatSWD Output Series dokumentiert die Arbeit des RatSWD in seiner 5. Berufungsperiode (2014–2017). In ihr werden seine Stellungnahmen und Empfehlungen veröffentlicht und auf diesem Weg einer breiten Leserschaft zugänglich gemacht.

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01UW1402 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt, sofern nicht anders ausgewiesen, beim RatSWD.



www.ratswd.de