



Research Data Repository

<http://www.radar-projekt.org>

- Entwurf - Autorenrichtlinien für RADAR

Version 0.1

August 2014



Dieses Dokument wird unter den Lizenzbedingungen der Creative Commons-Lizenz „Namensnennung - Nicht kommerziell - Keine Bearbeitungen 4.0 International (CC BY-NC-ND 4.0)“ veröffentlicht.

gefördert durch



Autoren:	RADAR-Projektteam - Thomas Engel, Filipe Furtado, Matthias Hahn, Angelina Kraft, Jörn Martens, Janna Neumann, Andrea Porzel, Jan Potthoff, Frauke Ziedorn
Titel:	-Entwurf- Autorenrichtlinien für RADAR
Status, Version:	Version 0.1
Monat, Erscheinungsjahr:	August 2014
Sprache:	Deutsch
Format:	PDF/A
Lizenz:	CC BY-NC-ND 4.0
Förderung:	Deutsche Forschungsgemeinschaft LIS-Programm: Überregionale Informationsservices
Projektnummern:	BE 1042/7-1 HE 5985/4-1 NE 1352/2-1 RO 2273/4-1 WE 1467/14-1
Kontakt:	info@radar-projekt.org http://www.radar-projekt.org

Beschreibung:

Um die Autoren und Nutzer auf dem Gebiet des nachhaltigen wissenschaftlichen Forschungsdatenmanagements zu unterstützen bietet RADAR die Möglichkeit zur disziplinübergreifenden Archivierung und (optionalen) Publikation digitaler Forschungsdaten. Dieser Leitfaden soll RADAR-Kunden einen Überblick zu den Zielsetzungen und Funktionen des Datenarchivs geben und gleichzeitig eine Entscheidungshilfe zu den angebotenen Dienstleistungen anbieten. Darüber hinaus werden Empfehlungen zur Qualitätssicherung von wissenschaftlichen Daten auf der Basis aktueller informationstechnischer Standards dokumentiert. Die hier vorgestellten Standards umfassen Empfehlungen zur Datenauswahl, zur Wahl eines geeigneten Dateiformats und zur Wahl einer Lizenz für publizierte Daten. Dazu werden u.a. folgende Fragen für Autoren und Datengeber erörtert:

- Datenauswahl - Welche Daten sollten archiviert bzw. veröffentlicht werden?
- Dateiformat - Welche Software ist zum Lesen der Daten notwendig?
- Lizenz - Welche Bedingungen und Auflagen sind bei der Nachnutzung der Daten zu beachten?

Die Leitlinien sollen dazu beitragen, über einen langen Zeitraum eine möglichst nachhaltige Datenspeicherung zu garantieren und, im Fall von publizierten Daten, deren effektive Nachnutzbarkeit und Zitierfähigkeit sicherzustellen.

Inhalt:

1. RADAR – Grundlegende Leitlinien	- 1 -
1.1 Zielsetzung	- 1 -
1.2 Organisation und Betreiber	- 1 -
1.3 Was sind digitale Forschungsdaten?	- 2 -
1.4 Glossar & FAQ	- 2 -
2. RADAR Zielgruppen	- 2 -
3. RADAR Kundentypen	- 3 -
4. RADAR Dienstleistungen	- 4 -
4.1 Angebot Archivierung	- 4 -
4.2 Angebot Archivierung und Publikation	- 5 -
5. Empfehlungen zur Qualitätssicherung	- 7 -
5.1 Die Datenauswahl	- 7 -
5.2 Die Formatierung der Daten	- 7 -
5.3 RADAR-Lizenzmodelle für Forschungsdaten	- 8 -

Im nachfolgenden Text werden männliche oder weibliche Sprachformen verwendet. Dies schließt die jeweilige andere nicht aus und soll lediglich der verbesserten Lesbarkeit dienen. Symbolik: [Text] – Platzhalter interne Links [Text] – Platzhalter externe Links

1. RADAR – Grundlegende Leitlinien

RADAR, ein Akronym für Research Data Repository, ist eine disziplinübergreifende Infrastruktur für die Archivierung und Publikation von Forschungsdaten.

Die in RADAR archivierten digitalen Objekte umfassen wissenschaftliche [Forschungsdaten], die nach den Vorgaben der Datengeber archiviert und ggf. publiziert werden (Open oder Closed Access möglich). Alle archivierten digitalen Objekte sind eindeutig referenzierbar und werden mit strukturierten Metadaten versehen. RADAR bietet die Möglichkeit einer temporären oder einer zeitlich unbegrenzten Datenarchivierung und optionalen Datenpublikation.

RADAR wurde aufgebaut um die Verfügbarkeit und Nachnutzung wissenschaftlicher Forschungsdaten sicherzustellen. Zur Forschungsdatenspeicherung bietet RADAR ein zweistufiges Dienstleistungsmodell an. Hier kann zwischen einem generischen Einstiegsangebot, dem **Angebot Archivierung**, zur formatunabhängigen Datenarchivierung mit minimalen Metadatensatz und einem erweiterten Angebot, dem **Angebot Archivierung und Publikation** gewählt werden. Über [Schnittstellen] kann RADAR in Dienste weiterer Anbieter integriert werden und dient so als ein wichtiger Baustein der nationalen und internationalen Informationsinfrastruktur. Der Dienstleistungsumfang und der Qualitätsanspruch des Forschungsdatenarchivs werden in den nachfolgenden Abschnitten vorgestellt.

Die RADAR-Redaktion begrüßt ausdrücklich Hinweise, Fragen, Kommentare und Ergänzungen zu unserem Angebot. Bitte verwenden Sie hierfür unsere [Kontaktmöglichkeiten].

1.1 Zielsetzung

Das Datenzentrum RADAR bietet Forschern, akademischen und kommerziellen Einrichtungen und wissenschaftlichen Verlagen (nachfolgend Datengeber genannt) die Möglichkeit zur disziplinübergreifenden Archivierung und optionalen Publikation digitaler wissenschaftlicher Forschungsdaten. Die Urheberrechte der Datengeber bzw. der jeweiligen Datenproduzenten (Autor(en)) bleiben dabei gewahrt.

Unsere Zielsetzung umfasst zwei wesentliche Grundsätze:

Archivierung von Forschungsdaten innerhalb von festen Haltefristen bis zu 15 Jahren oder unbegrenzt im Fall der **Publikation**.

Wissenschaftliche Forschungsdaten sollen vollständig, in sich verständlich, möglichst unverändert und mit ausreichenden Metadaten ausgestattet sein. Diese Qualitätsmerkmale ermöglichen ihre zukünftige Nachnutzbarkeit.

Im Unterschied zu fachspezifischen Forschungsdatenzentren ist RADAR **disziplinübergreifend** ausgerichtet und bietet seine Dienstleistungen grundsätzlich für alle wissenschaftlichen Disziplinen an. RADAR will dabei nicht in Konkurrenz zu etablierten Fachdatenzentren treten, sondern bietet offene [Schnittstellen] an, um externe Datenzentren bei der Dissemination und Nachnutzung von Forschungsdaten zu unterstützen.

1.2 Organisation und Betreiber

RADAR wurde als Infrastrukturprojekt von einem disziplinübergreifenden Team aus den Informations- und Naturwissenschaften entwickelt. Beteiligte Kooperationspartner sind:

[Leibniz-Institut für Informationsinfrastruktur (FIZ) Karlsruhe]
[Leibniz-Institut für Pflanzenbiochemie (IPB) Halle]
[Ludwig-Maximilians-Universität München (LMU) mit der Fakultät für Chemie und Pharmazie],
[Steinbuch Centre for Computing (SCC) des Karlsruher Instituts für Technologie (KIT)] und
[Technische Informationsbibliothek (TIB) in Hannover].

RADAR wird vom Leibniz-Institut für Informationsinfrastruktur (FIZ) Karlsruhe betrieben. Seine Infrastruktur ist in nationale und internationale Forschungsdatenstrukturen eingebunden. Diese umfassen u. a. das [DataCite] Konsortium und das wissenschaftliche Informationsportal [GetInfo] der TIB in Hannover. Die Kosten für die elektronische Archivierung und optionale Publikation richten sich nach Auftraggeber, Dauer der Haltefristen, Datenumfang und dem gewünschten Dienstleistungsmodell und können [hier] individuell berechnet werden. Die Bereitstellung digitaler wissenschaftlicher Forschungsdaten zur Archivierung und optionalen Publikation erfolgt über das RADAR-Portal [Uploadfunktion] oder ein herunterladbares Tool zur Metadatenerstellung und zum Datenimport [Tool].

1.3 Was sind digitale Forschungsdaten?

Im Kontext von RADAR werden unter ‚digitalen wissenschaftlichen Forschungsdaten‘, sowohl Rohdaten, Primärdaten (=intermediäre Arbeitsdaten), als auch Ergebnisdaten (=Sekundärdaten) verstanden, die im Rahmen wissenschaftlicher Studien aus Beobachtungen, Experimenten, numerischen Rechnungen, Arbeitsmethoden und Protokollen und bei der Auswertung der vorangegangenen Arbeitsschritte anfallen und digital erfasst werden können. Diese Definition ist konform zu den [Richtlinien der guten wissenschaftlichen Praxis der Deutschen Forschungsgemeinschaft (DFG)].

1.4 Glossar & FAQ

Zur Erklärung der hier verwendeten Terminologien und Fachbegriffe bieten wir Ihnen ein ausführliches [Glossar] und eine Liste häufig gestellter Fragen [FAQ] an. Sie enthalten neben allgemeinen Definitionen rund um die Datenspeicherung auch Angaben und Anwendungsbeispiele zu den von RADAR verwendeten datentechnischen und fachspezifischen Metadatenschemata.

2. RADAR Zielgruppen

RADAR ist ein interdisziplinäres Forschungsdatenzentrum. Die Basisdienstleistung umfasst ein breites Spektrum an wissenschaftlichem Datenmanagement, das in erster Linie auf Kunden aus den „small sciences“, den sogenannten wissenschaftlichen Spezialbereichen der großen Disziplinen der Natur-, Geistes-, Wirtschafts-, Sozial- und Kulturwissenschaften, zugeschnitten ist. RADAR stellt eine online erreichbare Infrastruktur zur Verfügung, die es Datengebern erlaubt, Forschungsdaten zu archivieren und digital zu publizieren. Zudem wählt der Datengeber eine entsprechende Lizenz aus, welche die Nachnutzung der Daten regelt. RADAR richtet sich mit seinem Dienstleistungsangebot sowohl an nationale Datengeber als auch an einen internationalen Datenproduzenten- und Datennutzerkreis. Alle in RADAR publizierten Daten und zugehörigen beschreibenden Metadaten sind über das RADAR-Portal und die an DataCite angeschlossenen Suchportale auffindbar (erweitertes Angebot Datenpublikation und -archivierung). Die Sichtbarkeit und Suchbarkeit der Forschungs- und Metadaten, die innerhalb der ausschließlichen Datenarchivierung (Angebot Archivierung) bei RADAR hochgeladen wurden, kann individuell durch den Datengeber festgelegt werden. Alle RADAR-Services sind in deutscher und englischer Sprache verfügbar. Die Suche nach publizierten Datensätzen ist jederzeit für alle Besucher des

Portals möglich; um die RADAR-Dienstleistungen, einschließlich des Service für einen Datenuploads und Datendownload nutzen zu können, ist eine einmalige Registrierung erforderlich.

Wissenschaftler und deren Forschungseinrichtungen können ihre Forschungsdaten sicher und nach deutschem Recht speichern, verwalten und wenn gewünscht online publizieren sowie mit einem DOI (Digital Object Identifier) versehen. Durch die DOI-Vergabe wird ein dauerhaftes Label (Persistent Identifier) geschaffen, der es erlaubt den Datensatz eindeutig zu identifizieren, zu zitieren und auffindbar zu machen. Die Forschungsdaten sind dann, wenn dies gewünscht wird, über die reine Archivierung hinaus über wissenschaftliche Datenbanken und Portale auffindbar und dadurch auch für andere sichtbar und nutzbar [[Angebot Datenpublikation und -archivierung](#)].

Wissenschaftliche Bibliotheken haben die Möglichkeit die von RADAR entwickelte Software als Vorlage zu nutzen, um potentiell eigene Angebote zum Forschungsdatenmanagement aufzubauen. Dies kann z.B. im Rahmen der Entwicklung eines eigenen ‚Portals‘ mit der Wahrnehmung der Aufgabe der Bitstream Preservation durch RADAR erfolgen. Das Datenzentrum übernimmt hierbei die dauerhafte und zuverlässige Datenspeicherung. Die [[RADAR-Software](#)] kann so als ein Baustein für weitere Informationsinfrastrukturen wie fachspezifische Datenzentren und institutionelle Systeme des Forschungsdatenmanagements eingesetzt werden.

Verlage und Herausgeber von wissenschaftlichen Zeitschriften haben ebenfalls die Möglichkeit, die ihnen zur Verfügung gestellten Forschungsdaten über RADAR zu archivieren und optional öffentlich zugänglich zu machen. Dies erhöht die Sichtbarkeit, Wiederauffindbarkeit und Zitierfähigkeit dieser Daten über digitale Informationsplattformen.

Datennutzer können nach einer einmaligen Registrierung direkt von ihrem Arbeitsplatz auf die publizierten Forschungsdaten zugreifen und diese als Referenz, unter Berücksichtigung der Lizenzbestimmungen, für ihre eigene Arbeit verwenden.

3. RADAR Kundentypen

Nutzer – Ein Nutzer kann sich bei RADAR mit einer gültigen E-Mail-Adresse registrieren & publizierte Datensätze downloaden; zudem kann er sich für einen Newsletter registrieren um z.B. über neue Datensätze aus von ihm gewählten Fachgebieten informiert zu werden.

Administrator – Der Administrator eines Arbeitsbereichs besitzt die meisten Rechte im RADAR-Kundenmodell. Er gibt bei der RADAR-Registrierung ausführliche Kontaktdaten einschließlich einer Rechnungsadresse an. Der Administrator kann Kuratoren zu einem Arbeitsbereich anlegen (Benutzerverwaltung), sich für Angebotsmodelle entscheiden, Angebotswechsel vornehmen und über den Arbeitsbereich Reports über den Datenbestand einsehen.

Kurator – Der Kurator wird vom Administrator als Bearbeiter eines Arbeitsbereichs bei RADAR angelegt und kann selbst der/ein Autor der Datensätze sein. Er ist für das Hochladen und Verwalten der/seiner Daten innerhalb RADARs zuständig; er kann Daten anlegen, updaten und löschen sowie zugehörigen Metadateninformationen ausfüllen und bekommt regelmäßige Informationen über Datenbestand (Checks, Backups, Downloads) mitgeteilt.

Autor – Der Autor der Datensätze ist der Daten-Erzeuger (Feld ‚creator‘ im RADAR-Metadatenschema). Er ist für die Erstellung und ggf. Aufarbeitung der Daten verantwortlich und stellt dem Kurator die allgemeinen und fachspezifischen beschreibenden Informationen zu den Daten (=Metadaten) zur Verfügung. Je nach Fall & Arbeitssituation kann der/ein (Co-) Autor auch selbst als Kurator der Datensätze fungieren.

4. RADAR Dienstleistungen

RADAR bietet die Datenkuration nach dem Grundsatz der nachhaltigen Archivierung und optionalen Publikation in zwei Angebotsformen an:

4.1 Angebot *Archivierung*

Grundstufe: *Archivierung* – Dieses Angebot beinhaltet eine sichere Datenspeicherung mit regelmäßiger Integritätsprüfung über eine frei wählbare Haltedauer zw. 5-15 Jahren. Mit diesem Angebot wird die Grundanforderung vieler Drittmittelgeber nach einer sicheren Datenspeicherung auch über das Ende der Projektlaufzeit hinaus erfüllt. Als Nachweis der Datenarchivierung vergibt RADAR einen Persistenten Identifier in Form eines Handles für jeden Datensatz. Der Kunde kann abgestufte Einsichts- und Zugriffsrechte auf die gespeicherten Daten & Metadaten innerhalb des Portals vergeben, z.B. für weitere Kuratoren oder bestimmte registrierte RADAR-Nutzer und so den (Meta)Datenzugriff individuell kontrollieren. Der Kunde wird 6 Monate vor Ablauf der Haltefrist kontaktiert, ob er

1. die Archivierung verlängern möchte (kostenpflichtig),
2. die Entität Publizieren möchte (kostenpflichtig) oder
3. Entität gelöscht werden soll (kostenlos).

Für welche Daten sollte das Angebot der I. Grundstufe: Archivierung genutzt werden?

Für Forschungsdaten, die **sicher gespeichert** aber in erster Linie nicht öffentlich zugänglich sein sollen, z.B.:

Containerdaten, die zum Abschluss eines Projektes nach Vorgaben von Forschungsförderern archiviert werden sollen (z.B. über eine Haltefrist von 10 Jahren nach DFG-Vorgaben)- Daten projektinterner (Zwischen-)Schritte
Hintergrunddaten, die eine Geheimhaltung erfordern (z.B. personenbezogene Daten)
Daten zu vorläufigen Ergebnissen einer Studie
Daten mit Patenthintergrund.

Was zeichnet das **Angebot Archivierung** aus?

Jedes hochgeladene Datenpaket bekommt einen **Handle** als dauerhaften ‚Identifier‘ zugewiesen - dabei handelt es sich um einen aus einer Zahlen- und Buchstabenkombination existierenden Identifizierungscode, der eindeutig dem Datenpaket zugeordnet wird und auch bei einer Veränderung des Speicherortes dessen Identifizierbarkeit, Auffindbarkeit und Zitierbarkeit sicherstellt.

Als Datenarchiv übernimmt RADAR die **Verantwortung**, dass **gespeicherte Daten** für den Kunden (und ggf. einem von ihm definierten Nutzerkreis) **jederzeit abrufbar** sind, auch wenn zwischen Archivierung und Datenabruf mehrere Jahre (bis zu 15) liegen; dies wird durch regelmäßige Integritätsüberprüfungen der Daten sichergestellt.

Die Kosten für das Archivierungsangebot staffeln sich nach dem Datenvolumen. Da die Zielgruppe primär forschende Institutionen sind, bietet RADAR ausschließlich Einmalzahlungen abhängig von der tatsächlich zu archivierenden Datenmenge an; so sind die voraussichtlichen **Kosten im Vorfeld bekannt** und können z.B. in wissenschaftlichen Förderanträgen berücksichtigt werden.

RADAR speichert, überträgt und archiviert die von unseren Kunden übermittelten Forschungsdaten auf **Servern in Deutschland**.

Zusammenfassung des Daten- & Metadatenzugriffs im Angebot Archivierung

→ Abgestufte Einsichts- und Zugriffsrechte für gespeicherte Daten & Metadaten

→ Default: Einsichts- und Zugriffsrechte nur für Datengeber, auf weitere Kuratoren und RADAR-Nutzer erweiterbar.

4.2 Angebot Archivierung und Publikation

Grundstufe: Archivierung und Publikation – Dieses Angebot beinhaltet die Veröffentlichung wissenschaftlicher Forschungsdaten bei RADAR und deren maximale Nachnutzbarkeit durch die Sicherstellung ihrer dauerhaften Verfügbarkeit (Download) sowie ausführliche Beschreibung mit Hintergrundinformationen zu Datenerhebung (in Form von deskriptiven Metadaten). Bei diesem Angebotsmodell sind Forschungsdaten (ggf. nach Ablauf eines Embargos, siehe nachfolgende Optionen) & deren Beschreibung (=Metadaten) öffentlich zugänglich und mit einem DOI eindeutig zitierbar. Die Daten werden über einen unbegrenzten Zeitraum von RADAR zur Verfügung gestellt (mind. 15 Jahre) und der Kunde hat die Möglichkeit; die Metadaten zu editieren oder wenn gewählt das Embargo jederzeit zu verlängern. Eine Löschung von publizierten Daten ist nur nach gesonderter E-Mail Anfrage & Prüfung durch RADAR in möglich. Dieses Angebot umfasst **zwei Optionen**:

A. Direkte Publikation

B. Publikation mit Embargo

A. Im Angebot der **direkten Publikation** sind die Forschungsdaten sofort öffentlich zugänglich.

B. Das Angebot **Publikation mit Embargo** beinhaltet die Veröffentlichung wissenschaftlicher Forschungsdaten bei RADAR zu einem festgelegten Zeitpunkt in der Zukunft. Die zu RADAR hochgeladenen Daten sind nicht sofort öffentlich, sondern der Kunde bestimmt ein Datum in der Zukunft, wenn die Daten publiziert, also zum Download verfügbar sein sollen. Für den Nutzer entsteht in diesem Modell ein vorläufiger Mehrwert durch die Auffindbarkeit, die eindeutige Zitierbarkeit und damit verbundene Referenzierbarkeit der Daten (über DOI), auch wenn der eigentliche Datenabruf erst zu einem späteren, definierten Zeitpunkt über die Freischaltung des zugehörigen Downloadlinks erfolgen kann.

Für welche Daten sollte das Angebot der Grundstufe: **Direkte Publikation** genutzt werden?

Für **Forschungsdaten**, die **öffentlich zugänglich, zitierbar (DOI), nachnutzbar und sicher gespeichert** sein sollen, z.B.:

abgeschlossene Projektdaten

Datengrundlagen wissenschaftlicher Publikation (Journal)

Negativdaten, die Nachnutzungspotential besitzen, aber sich nicht oder nur eingeschränkt für ein Journal eignen

Datengrundlagen, welche bisher aufgrund von fehlenden (eigenen) Infrastrukturen nicht veröffentlicht wurden.

Für welche Daten sollte das Angebot der Grundstufe: **Publikation mit Embargo** genutzt werden?

Für Forschungsdaten, die nach Ablauf eines zeitlichen Embargos öffentlich zugänglich, zitierbar, nachnutzbar und sicher gespeichert sein sollen, z.B.:

Projektdaten und oder Negativdaten, die zu einem festen Datum in der Zukunft veröffentlicht werden sollen bzw. zeitgesteuert publiziert werden sollen

unveränderliche Datengrundlagen wissenschaftlicher Publikation (Journal), bei denen sich das zugehörige Paper noch im Review-Prozess befindet

Hintergrunddaten, die eine zeitlich befristete und absehbare Geheimhaltung erfordern

Daten mit Patenthintergrund, bei denen der Ablauf des Patentschutzes sich in der näheren Zukunft (< 3 Jahre) befindet.

Was zeichnet das **Angebot Archivierung und Publikation** aus?

Erweiterter Metadatensatz erlaubt individuelle Beschreibung & Anpassung auf eigenen Fachbereich (Glossar, Beispiele)

RADAR ermöglicht durch seine Generik ein disziplinübergreifendes (& institutübergreifendes) Datenmanagement, z.B. über die erweiterte Vergabe von Zugriffsrechten auf die Datenverwaltung (individueller Arbeitsbereich) und den Datenzugriff (Embargo)

RADAR gewährleistet eine nachhaltige Datenspeicherung, bei dem die gespeicherten Daten durch den Kunden (und ggf. einem von ihm definierten Nutzerkreis) **jederzeit abrufbar** sind, auch wenn zwischen Archivierung und Datenabruf mehrere Jahre (bis zu 15) liegen; dies wird durch regelmäßige Integritätsüberprüfungen der Daten sichergestellt.

RADAR speichert, überträgt und archiviert die von unseren Kunden übermittelten Forschungsdaten auf **Servern in Deutschland**.

Als Datenzentrum übernimmt RADAR die **Verantwortung**, die **publizierten Daten** selbst nach einmaliger Registrierung für Nutzer offen zum Download zur Verfügung stehen

Der Kunde kann ein regelmäßiges Feedback zu Nutzungsstatistiken seiner Daten anfordern

RADAR bietet ein Geschäftsmodell mit Einmalzahlungen abhängig von der tatsächlich zu archivierenden Datenmenge an; so sind die voraussichtlichen **Kosten im Vorfeld bekannt** und **können z.B. in wissenschaftlichen Förderanträgen berücksichtigt werden**.

Zusammenfassung des Daten- & Metadatenzugriffs im **Angebot Archivierung mit Publikation**

→ Öffentlicher Datenzugriff & Download nach einmaliger Registrierung bei RADAR.

→ Optional ist die Vergabe eines zeitlichen Embargos durch den Datengeber möglich (Datum in der Zukunft), ab dem die Daten zum Download zur Verfügung stehen.

→ zugehörige Metadaten (Datenerzeuger, Titel, Inhalte, etc.) sind für alle Nutzer sichtbar.

→ Metadaten portalübergreifend sichtbar & auffindbar.

5. Empfehlungen zur Qualitätssicherung

Forschungsdaten bilden die Grundlagen aller wissenschaftlichen Erkenntnisse. Um die Daten schnell auffindbar und langfristig verfügbar zu machen wurden die nachfolgenden formalen und technischen Empfehlungen zur Qualitätssicherung erstellt. Obwohl grundsätzlich alle Formate für die Datenspeicherung akzeptiert werden, empfiehlt RADAR allen Datengebern, die nachfolgenden Standards so weit wie möglich zu befolgen. Diese Standards sollen dazu beitragen über einen langen Zeitraum eine möglichst nachhaltige Datenspeicherung, und soweit gewünscht eine effektive Nachnutzbarkeit und Zitierfähigkeit der publizierten Daten, sicherzustellen.

5.1 Die Datenauswahl

Im Rahmen der guten wissenschaftlichen Praxis sind die Datengeber und Datenproduzenten für die inhaltliche Vollständigkeit des Datensatzes zuständig. Dazu gehört die Bereitstellung ausreichender datentechnischer (z.B. Autor(en), Titel, Jahr) und fachspezifischer Metadaten (Datentyp, Messspezifikationen) und eine hinreichende Dokumentation möglicher Messfehler und -ausfälle, sowie Inkonsistenzen im Dateiojekt. RADAR kann als Dienstleister den Entscheidungsprozess, welche Daten verfügbar gemacht werden sollen, mit den folgenden Fragestellungen unterstützen:

Beinhaltet die Datei *Informationen, die den gesamten durchgeführten Forschungsprozess* einer wissenschaftlichen Studie oder wesentliche Teile hiervon für ein bestimmtes Zielpublikum *nachvollziehbar machen*?

Beinhaltet die Datei Informationen, die für die zukünftige Reproduzierbarkeit des wissenschaftlichen Forschungsprozesses von Bedeutung sind?

Werden die in der Datei enthaltenen *Informationen für praktische Zwecke*, beispielsweise die weitere Verbreitung und effektive Nachnutzung des wissenschaftlichen Forschungsprozesses benötigt?

Sobald eine dieser Fragestellungen mit ‚Ja‘ beantwortet wird, sollte die Aufnahme der Datei in das zu archivierende Datenpaket in Betracht gezogen werden.

Mit der Übertragung der Dateiojekte an das Repositorium erfolgt eine technische Qualitätssicherung der Forschungsdaten und zugehörigen Metadaten eines Uploads. Im Anschluss erhält der Datengeber ein Feedback zur Datenübermittlung und wird über sein Benutzerkonto und per E-Mail informiert, wenn die gewünschte Archivierung (und ggf. die optionale Datenpublikation) erfolgreich war. Nach einer erfolgreichen Datenarchivierung findet eine regelmäßige Überprüfung der Datenkonsistenz statt, deren Ergebnisse in einer Protokolldatei festgehalten werden.

5.2 Die Formatierung der Daten

Die Wahl des richtigen digitalen Formats ist essentiell für eine nachhaltige Aufbewahrung wissenschaftlicher Forschungsdaten. Die Formatierung sollte dabei nationalen und internationalen wissenschaftlichen Vorgaben der jeweiligen Fachdisziplin entsprechen. Für die Datenarchivierung in RADAR wird empfohlen, nach Möglichkeit die folgenden Qualitätsmerkmale zu berücksichtigen:

Wählen Sie die Formatierung der Datendateien nach Möglichkeit so, dass sie mit etablierter, handelsüblicher Software ausgelesen und dargestellt werden können.

Verwenden Sie eine eindeutige und ihrer Abfolge entsprechend fortlaufende Benennung für die Datendateien. Der Name sollte auf den Inhalt der einzelnen Datenpakete hinweisen.

Die Bezeichnung von Parametern in Datenpaketen sollte einem für Ihre Fachdisziplin anerkannten, internationalen Standard folgen. Wenn kein solcher Standard existiert

können auch institutsübliche Parameterbezeichnungen verwendet werden, welche dann mit einem eindeutigen, leichtverständlichen Begleittext erläutert werden.

Die Bezeichnung und Ausweisung von Einheiten für alle Parameter sollten einem für die jeweilige Fachdisziplin anerkannten, internationalen Standard folgen (z.B. dem [Internationalen Einheitensystem (SI) für physikalische Größen]¹).

Jede Datendatei oder, wenn es sich um ein Set von mehreren zusammengefassten einzelnen Dateien handelt, jedes Datenpaket, muss mit einem dazugehörigen XML-Metadatendokument bei RADAR eingereicht werden. Vorlagen und Beispiele der Metadatenchema(ta) für unsere Angebote finden Sie [\[hier\]](#).

RADAR akzeptiert alle Dateiformate zur Speicherung der Forschungsdaten. Um allerdings eine möglichst optimale Nachnutzbarkeit der publizierten Daten sicher zu stellen, sind die von RADAR bevorzugten Formate für das **Angebot Archivierung mit Publikation** in der nachfolgenden **Liste mit Formatempfehlungen** aufgeführt (Tab. 1).² Diese Empfehlungen werden in regelmäßigen Abständen aktualisiert und enthalten Formate, die nach Ansicht von RADAR die größten Chancen für eine langfristige Nachnutzbarkeit von Forschungsdaten liefern.

Tab. 1 RADAR – empfohlene Formate im Angebot 2: Datenpublikation und -archivierung

Datentyp	Empfohlen
Textdokumente	PDF/A (.pdf) sowie XML-basierte Formate (.docx)
Tabellen	Comma Separated Values (.csv) PDF/A (.pdf) sowie XML-basierte Formate (.xlsx)
Datenbanken	ANSI SQL (.sql) Comma Separated Values (.csv)
Statistische Angaben	SPSS Portable (.por) SAS transport (.sas) STATA (.dta)
Graphiken (Bilder oder rasterbasiert)	JPEG (.jpg, .jpeg) TIFF (.tif, .tiff)
Graphiken (vektorbasiert)	PDF/A (.pdf) Scalable Vector Graphics (.svg)
Video	Lossless AVI (.avi) MPEG-1, MPEG-2 (.mpg, .mpeg) MPEG-4, H264 (.mp4) FLV (.flv)
Audio	WAVE (.wav)
Computer Aided Design	AutoCAD DXF version R12 (.dxf)
Geographische Informationen	MapInfo Interchange Format (.mif, .mid)

5.3 RADAR-Lizenzmodelle für Forschungsdaten

Im Rahmen der Datenpublikation bei RADAR werden acht verschiedene Lizenzmodelle und ein optionales Freitextfeld, das zur Vergabe einer individuellen Lizenz dienen kann, angeboten.

A. Interessant für wen?

Datengeber: Um eine maximale Nachnutzbarkeit wissenschaftlicher Forschungsdaten zu gewährleisten, die -abhängig von ihrer geistigen Schöpfungshöhe - prinzipiell dem Urheberrecht unterliegen können, sollte die Vergabe von zusätzlichen Nutzungsrechten z.B. durch eine entsprechende Lizenzierung der Daten in Betracht gezogen werden. Daher

¹ http://www.bipm.org/fr/si/si_brochure/general.html

² Modifiziert nach DANS: www.dans.knaw.nl

empfiehlt RADAR die Verwendung liberaler Lizenzmodelle. Nachfolgend werden hierfür die weltweit anerkannten Creative Commons (CC)-Lizenzen vorgestellt. Ab Version 4.0 sind die CC-Lizenzen auch für Forschungsdaten geeignet. Die Wahl einer Lizenz bietet so dem Datengeber die Möglichkeit, die Nachnutzung der durch ihn publizierten Forschungsdaten und daran geknüpfte Bedingungen für Datennutzer nachvollziehbar darzustellen.

Datennutzer: Wenn die grundlegenden (Metadaten) Informationen (z.B. Autoren, Titel, Jahr, Dateninhalte) zu einem durch RADAR publizierten Forschungsdatensatz abgerufen werden, erhält jeder Nutzer gleichzeitig auch eine Information über die Nachnutzungsrechte dieser Daten.

Nachfolgend werden sieben Versionen der CC-Lizenzen und das deutsche Urheberrecht im Vergleich dazu vorgestellt.

B. Datenveröffentlichung in der Praxis

I. Publikation mit einigen vorgehaltenen Rechten

³ - **Creative Commons (CC)-Lizenzen Version 4.0 –**

Die Vergabe einer Creative Commons-Lizenz für digitale Objekte, die öffentlich verfügbar sein sollen, bietet eine weltweit verbreitete, standardisierte Methode die zugehörigen Nutzungsrechte des Objekts klar darzustellen. Im Rahmen dieser Lizenzen hat der Urheber, der selbst über die Nachnutzung seiner Daten bestimmen möchte und die Dienste anderer Verwertungsgesellschaften (z.B. wissenschaftlicher Verlage) nicht in Anspruch nimmt, die Wahl zwischen verschiedenen strengen Lizenz-Modellen. Durch die Wahl einer entsprechenden Lizenz kann der Datengeber so die Vervielfältigung und Verbreitung seiner Daten allen Interessierten erlauben, sofern diese ihn als Rechteinhaber nennen. Gleichzeitig kann der Datengeber ebenso eine gewerbliche Nutzung der Daten, oder die Veröffentlichung und Weitergabe von Bearbeitungen dieser Daten durch andere, ausschließen.

Die Vorteile von Creative Commons-Lizenzen zu anderen verfügbaren Lizenzen sind u. a.:

- Sie respektieren das Urheberrecht.
- Sie sind weltweit bekannt, in ihrer Geltung auch durch deutsche Gerichte grundsätzlich anerkannt und können daher auch gerichtlich durchgesetzt werden⁴.
- Werden kontinuierlich aktualisiert und an die Rechtslage angepasst.
- Die Nutzungsrechte werden klar und verständlich dargestellt.

Jede der nachfolgenden CC-Lizenzen beinhaltet die Kennzeichnung BY. BY erfordert die ‚Namensnennung‘ (Zitat) der Autoren, sobald das Werk eine Form der Nachnutzung durchläuft. Neben der Namensnennung kann der Datengeber zusätzlich folgende Nutzungsbeschränkungen festlegen:

- ‚keine kommerzielle Nutzung‘ (NC) – das Werk darf nur in einem nicht kommerziellen Rahmen genutzt werden.
- ‚keine Bearbeitungen‘ (ND) – die Weitergabe, Verbreitung und Nutzung des Werkes darf nur in Form einer exakten Kopie erfolgen. Die Weitergabe einer vom Original abweichenden Version ist nicht zulässig, im privaten Rahmen darf sie jedoch erstellt werden.
- ‚Weitergabe unter gleichen Bedingungen‘ (SA) – alle neuen Werke und Inhalte, die auf dem Original aufbauen, müssen mit derselben Lizenz (SA) oder einer mit dieser Lizenzversion kompatiblen Lizenz versehen werden. Diese Kennzeichnung stellt sicher, dass die neu entstandenen Werke den gleichen Nutzungsbedingungen wie denen des Originals unterliegen.

³ <http://creativecommons.org>

⁴ <http://creativecommons.org/Version4>



- Namensnennung CC BY⁵

Bedeutung: Diese Lizenz erlaubt Dritten, ein Werk bzw. den Inhalt zu kopieren, verbreiten und zugänglich zu machen, sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anzufertigen und zu verbreiten, auch kommerziell, solange der Urheber des Originals genannt wird. Diese Lizenz erlaubt eine maximale Verbreitung und Nutzung des lizenzierten Materials.



- Namensnennung - Keine Bearbeitung CC BY-ND⁶

Bedeutung: Diese Lizenz erlaubt Dritten, ein Werk bzw. den Inhalt zu kopieren, verbreiten und zugänglich zu machen, kommerziell wie nicht-kommerziell, solange dies ohne Veränderungen und vollständig geschieht und der Urheber genannt wird.



- Namensnennung - Weitergabe unter gleichen Bedingungen CC BY-SA⁷

Bedeutung: diese Lizenz erlaubt es Dritten, ein Werk bzw. den Inhalt zu kopieren, verbreiten und zugänglich zu machen, sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anzufertigen und zu verbreiten, auch kommerziell, solange der Urheber des Originals genannt wird und die auf diesem Werk oder Inhalt basierenden neuen Werke unter denselben Bedingungen veröffentlicht werden. Alle neuen Werke und Inhalte, die auf diesem aufbauen, werden unter derselben oder einer mit dieser Lizenz kompatiblen Lizenz stehen, also auch kommerziell nutzbar sein.



- Namensnennung - Nicht kommerziell CC BY-NC⁸

Bedeutung: Diese Lizenz erlaubt es Dritten, ein Werk bzw. den Inhalt zu kopieren, verbreiten und zugänglich zu machen, sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anzufertigen und zu verbreiten, allerdings nur nicht-kommerziell.



- Namensnennung - Nicht-kommerziell - Weitergabe unter gleichen Bedingungen CC BY-NC-SA⁹

Bedeutung: Diese Lizenz erlaubt es Dritten, ein Werk bzw. den Inhalt zu kopieren, verbreiten und zugänglich zu machen, sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anzufertigen und zu verbreiten, allerdings nur nicht-kommerziell und solange der Urheber des Originals genannt wird und die auf diesem Werk bzw. Inhalt basierenden neuen Werke unter denselben Bedingungen veröffentlicht werden.



- Namensnennung - Nicht-kommerziell - Keine Bearbeitung CC BY-NC-ND¹⁰

Bedeutung: Diese Lizenz erlaubt lediglich das Kopieren, Verbreiten und Zugänglichmachen des Werkes oder Inhaltes unter Nennung des Urhebers, jedoch keinerlei Bearbeitung oder kommerzielle Nutzung.

⁵ <http://creativecommons.org/licenses/by/4.0/>

⁶ <http://creativecommons.org/licenses/by-nd/4.0/>

⁷ <http://creativecommons.org/licenses/by-sa/4.0/>

⁸ <http://creativecommons.org/licenses/by-nc/4.0/>

⁹ <http://creativecommons.org/licenses/by-nc-sa/4.0/>

¹⁰ <http://creativecommons.org/licenses/by-nc-nd/4.0/>

II. Datenveröffentlichung in der „Public Domain“ in Deutschland



Creative Commons: - Lizenzmodell CC0 v1.0 - ¹¹

Bedeutung: Mit dieser Lizenz verzichtet der Datengeber (so weit dies nach deutschem Urheberrecht möglich ist) auf die Urheberrechte für die so lizenzierten Daten bzw. räumt vollumfassend Nutzungsrechte ein. Damit können die unter einer CC0-Lizenz stehenden Werke ohne Bedingungen kopiert, verändert und verbreitet werden. Bei diesen Werken besteht weiterhin auch keine Verpflichtung, den Urheber der Daten zu nennen.

Erläuterung: In vielen Ländern des englischsprachigen Raums einschließlich der USA unterliegen Daten nicht dem Urheberrechtsschutz und Forschern und Institutionen ist es möglich auf ihnen zustehende Urheberrechte an von ihnen geschaffenen Werken und wissenschaftlichen Arbeiten gänzlich zu verzichten und sie innerhalb der sog. „Public Domain“ zu veröffentlichen. Die (Rechts-)Bezeichnung „Public Domain“ bedeutet im angelsächsischen Raum, dass solche Werke „frei von Urheberrechten“ sind. Rechtlich gesehen verzichtet der Dateneigentümer bzw. Urheber dabei auf sämtliche seiner Urheberrechte. Im deutschen Rechtsraum ist ein solcher vollständiger Verzicht auf ein Urheberrecht im Hinblick auf das Urheberpersönlichkeitsrecht nicht möglich (§ 29 UrhG)¹². Daher behält z.B. auch ein Angestellter das Urheberrecht über seine Arbeiten welche er im Rahmen eines Dienstverhältnisses geschaffen hat, auch wenn er die zugehörigen Nutzungsrechte vollständig an seinen Arbeitgeber übertragen hat bzw. übertragen muss. Dies wird aber im Rahmen der CC-Lizenz CC0 v1.0 Universal, welche eine uneingeschränkte Weitergabe und Nutzung zu jedem Zweck (auch kommerziell) der so lizenzierten Daten erlaubt und damit dem rechtlichen Kontext der Gemeinfreiheit am nächsten kommt, berücksichtigt.

III. Datenveröffentlichung unter Vorbehalt aller Rechte

- Nutzung des deutschen Urheberrechts -

In Deutschland gilt das deutsche Urheberrechtsgesetz für die Nutzung von literarischen, künstlerischen und wissenschaftlichen Werken, die die Schutzvoraussetzungen des Urheberrechtsgesetzes erfüllen. Sofern den Nutzern nicht weitere Nutzungsrechte durch eine über die in den Schrankenregelungen hinausgehende Lizenz eingeräumt werden, ist eine Nachnutzung nur im Rahmen der Schrankenregelungen des deutschen Urheberrechtsgesetzes möglich. Danach ist die Nachnutzung z.B. auf die Ansicht und die Speicherung einer Kopie ohne Modifikation für den privaten oder eigenen Gebrauch beschränkt, wobei der Urheber namentlich genannt werden muss. Eine Weitergabe von veränderten Daten oder erneute Veröffentlichung der Daten in einem anderen Kontext ist ausgeschlossen.

Grundsätzlich unterliegen wissenschaftliche Primärdaten, also die Ausgangsdaten, als reine Fakten nicht dem Schutz des Urheberrechtsgesetzes. Dieser Schutz liegt erst dann vor, wenn eine persönliche geistige Schöpfungsleistung gegeben ist, oder eine wesentliche Investition für die Datensammlung erforderlich war. Da diese Investition bei Forschungsdaten meist im Rahmen der Datenerhebung erfolgt, deren Kosten nach §87a UrhG nicht berücksichtigt werden dürfen, unterliegen die so geschaffenen Datensammlungen nicht dem Urheberrecht¹³. Dies muss jedoch im jeweiligen Einzelfall geprüft werden, z.B. wenn es sich bei den ermittelten Daten um intermediäre Arbeitsdaten

¹¹ http://wiki.creativecommons.org/CC0_use_for_data

¹² <http://www.gesetze-im-internet.de/urhgb/BjNR012730965.html>

¹³ Schulze/Dreier, § 87a, Rn. 11 ff. in Dreier, T./Schulze, G., *Urheberrechtsgesetz: Kommentar*, 3. Aufl. 2009.

oder Ergebnisdaten handelt. Auch hier sollte im Zweifelsfall die Beratung durch einen Fachanwalt erfolgen.

Um eine maximale Nachnutzbarkeit wissenschaftlicher Forschungsdaten zu gewährleisten, die prinzipiell dem Urheberrecht unterliegen können, sollte die Vergabe von zusätzlichen Nutzungsrechten z.B. durch eine entsprechende Lizenzierung der Daten (z.B. durch CC-Lizenzen) in Betracht gezogen werden. Die Vergabe solcher Lizenzen führt meist zu einer höheren Nutzung der Daten in der wissenschaftlichen Forschung und kann so zu einem Reputationsgewinn des Wissenschaftlers, auch über die Grenzen der jeweiligen Fachcommunity hinaus, beitragen.

IV. Vor der endgültigen Wahl der Lizenz:

Unabhängig davon, welche Lizenz der letztendlich gewählt wird, sollte der Datengeber sicher stellen, dass

- der Eigentümer der Forschungsdaten eindeutig identifiziert ist, und er das Recht hat, die Daten zu öffentlich oder für einen bestimmten Nutzerkreis zugänglich zu machen,
- der (potentielle) kommerzielle Wert der Daten bedacht wurde. Forschungsdaten, die bereits veröffentlicht wurden können z.B. in der Regel nicht mehr im Rahmen einer Patentanmeldung verwendet werden

und

- eine möglichst weitreichende Nachnutzung der Forschungsdaten nach dem Sinne des Datengebers und Datenautors erreicht werden kann.