



Research Data Repository

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

Deskriptives Metadatenprofil für RADAR

Version 0.2

November 2014



Dieses Dokument wird unter den Lizenzbedingungen der Creative Commons-Lizenz „Namensnennung 4.0 International (CC BY 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:	Deskriptives Metadatenprofil für RADAR
Status, Version:	Version 0.2
Monat, Erscheinungsjahr:	November 2014
Sprache:	Deutsch
Format:	PDF/A
Lizenz:	CC BY 4.0
Förderung:	Deutsche Forschungsgemeinschaft LIS-Programm: Überregionale Informationsservices
Projektnummern:	BE 1042/7-1 HE 5985/4-1NE 1352/2-1 RO 2273/4-1 WE 1467/14-1
Kontakt:	info@radar-projekt.org http://www.radar-projekt.org

Beschreibung:

Eine zentrale Aufgabe im RADAR-Projekt ist die Erstellung eines Metadatenprofils, das sowohl einen interdisziplinären, zentralen Nachweis der in RADAR archivierten und publizierten Forschungsdaten erlaubt als auch die fachspezifischen Anforderungen zur Suche und zur Nachnutzung dieser Daten erfüllt. Dazu wurden geeignete generische Metadatenparameter, die disziplinspezifisch angepasst werden können identifiziert und ausgewählt. Das nachfolgende Metadatenprofil umfasst 9 Pflichtfelder, welche zusammen den allgemeinen, deskriptiven Teil des Profils darstellen, sowie 12 optionale Felder, welche auch die fachspezifischen Beschreibungen der Datensätze abbilden. Die Pflichtfelder des entwickelten Metadatenprofils enthalten die Grundanforderungen für eine DOI-Registrierung nach dem DataCite-Metadaten-Schema v 3.1¹. Um die heterogenen Ansprüche verschiedener wissenschaftlicher Fachgebiete und die generische Ausrichtung von RADAR zusammenzubringen wurde bei der Definition der 12 optionalen Metadatenfelder mit einer Kombination von kontrollierten Listen und Freitextfeldern gearbeitet. Bei der Definition der kontrollierten Listen wurde auf weltweit anerkannte, verständliche Standards (z.B. ISO-Normen für die Sprache und das Entstehungsland der Forschungsdaten) zurückgegriffen. Dieses Metadatenprofil soll, in Verbindung mit exemplarischen Beispielen den Wissenschaftlern der jeweiligen Fachgebiete darlegen, wie eine detaillierte Beschreibung eines Datensatzes aussehen kann.

Inhalt:

1. Übersicht -----	1 -
2. Neun Pflichtfelder -----	2 -
3. Zwölf optionale Felder -----	5 -

¹ <http://schema.datacite.org/meta/kernel-3>

Änderungshistorie des Dokuments:

<i>Versionsnummer</i>	<i>Datum</i>	<i>Inhalte</i>
0.1	04.08.2014	Erster Entwurf der RADAR-Metadatendokumentation, deutsch
0.2	11.11.2014	Erweiterter Entwurf mit Modifikation der Parameter 2. und 19.

1. Übersicht

9 PFLICHTFELDER (die ersten vier Pflichtfelder entsprechen den DataCite-Pflichtfeldern):

ID	Beschreibung des Elements	XML
1.	identifier	<identifier>
1.1	identifierType (Vergabe Handle oder DOI)	<identifierType>
2.	creator (Datenerheber, z.B. Autoren)	<creator>
2.1	creator name (Name(n) der Datenerheber)	<creatorName>
2.2	creator affiliation (Zugehöriges Institut/Unternehmen)	<creatorAffiliation>
3.	title	<title>
4.	publisher (Herausgeber der Daten, Institut)	<publisher>
5.	production year (Entstehungs-/ Erhebungsjahr der Daten)	<productionYear>
6.	subject area (Fachgebiet)	<subjectArea>
6.1	subject area detail (Zusätzliche Informationen zum Fachgebiet)	<subjectAreaDetail>
7.	resource (Art der Forschungsdaten)	<resource>
7.1	resource type	<resourceType>
8.	rights (Lizenz zur Nachnutzung)	<rights>
8.1	rights type	<rightsType>
9.	rightsholder (Inhaber der Nachnutzungsrechte)	<rightsholder>

12 OPTIONALE FELDER zur fachspezifischen Anpassung:

ID	Beschreibung des Elements	XML
10.	additional title (z.B. Untertitel, Untersuchungsobjekt)	<additionalTitle>
10.1	additional title type	<additionalTitleType>
11.	description	<description>
11.1	description type (z.B. Abstract, Method)	<descriptionType>
12.	keyword (Schlagwörter zum Forschungsschwerpunkt)	<keyword>
13.	contributor (Mitwirkende/r)	<contributor>
13.1	contributorType (z.B. Sponsor, Funder)	<contributorType>
14.	language	<language>
15.	alternate identifier (z.B. institutseigener Identifier)	<alternateIdentifier>
15.1	alternate identifier type (z.B. interne Nummer eines Instituts)	<alternateIdentifierType>
16.	related identifier (z.B. Hinweis auf wiss. Paper)	<relatedIdentifier>
16.1	related identifier type (z.B. Angabe eines related DOI od. URL)	<relatedIdentifierType>
16.2	relation type (Beziehung zu verw. Ressource, z.B. isCitedBy)	<relationType>
17.	geo location (Ortsangaben)	<geoLocation>
17.1	geo location country (Entstehungs-/Bezugsland)	<geoLocationCountry>
17.2	geo location region (z.B. geographisches Gebiet)	<geoLocationRegion>
17.3	geo location point (Koordinatenset)	<geoLocationPoint>
17.4	geo location box (räumliche Grenzen, zwei Koordinatenpaare)	<geoLocationBox>
18.	data source (Prozedur/Quelle der Datenerhebung)	<dataSource>
18.1	data source detail (Prozedur/Quelle der Datenerhebung)	<dataSourceDetail>
19.	software type (Kategorisierung benutzter Software)	<softwareType>
19.1	software name (Name/Informationen zur Software)	<softwareName>
19.1.1	software name version (Angabe zu Software-Version(en))	<softwareNameVersion>
19.2	software alternative (Name/Informationen alternativer Software)	<softwareAlternative>
19.2.1	software alternative version (Version(en) alternativer Software)	<softwareAltVersion>
20.	data processing (Angabe weiterer Modifikationen, z.B. Statistik)	<dataProcessing>
21.	related information (z.B. Datenkomponenten, Substanzen, etc.)	<relatedInformation>
21.1	related information type ((z.B. Datenbankidentifikator, CAS-Nr.)	<relatedInformationType>

Erläuterungen zur nachfolgenden Tabelle:

Die Spalte A/C gibt an, ob ein Element Attribut (A) oder Child (C) des übergeordneten Elements ist. Spalte Occurrence (Occ) gibt an, wie häufig ein Element auftreten kann. Beispiel: „Occ 1-n“ heißt, dass ein Element mindestens einmal (1) auftreten muss, und beliebig oft wiederholt werden kann (n).

2. Neun Pflichtfelder

ID	Element	Definition	A/C	Occ	Erlaubte Werte, Beispiele, Beschränkungen
1	identifizier	Ein einzigartiger String, der ein Objekt eindeutig identifiziert.		1	Angebot Archivierung: Vergabe Handle Angebot Archivierung mit Publikation: Vergabe DOI (Digital Object Identifier)
1.1	identifizier type	Handle oder DOI.	A	1	Kontrolliertes Vokabular: - Handle - DOI
2	creator	Personen, die für den Inhalt der Forschungsdaten verantwortlich sind, z.B. die Ersteller der Daten (Autoren).		1-n	
2.1	creator name		A	1	Der Name eines Unternehmens/einer Institution oder Personennamen. Personen sollen im Format „Nachname, Vorname“ angegeben werden. Namen, die nicht im lateinischen Alphabet geschrieben sind, sollen nach ALA-LC Schema transkribiert werden. ²
2.2	creator affiliation		A	0-n	Name des zugehörigen Unternehmens/der Institution.
3	title	Titel des Datensets.		1	Das Format ist offen. Zeichen sollen im Rahmen von UTF-8 ³ erlaubt sein.
4	publisher	Personen oder Organisation, die dafür verantwortlich ist, dass Daten in der gegebenen Form bei RADAR archiviert werden, z.B. Herausgeber, Institut.		1-n	Der Name eines Unternehmens/einer Institution oder ein Personennamen. Personennamen sollten im Format „Nachname, Vorname“ angegeben werden.
5	production year	Entstehungsjahr der Daten oder Jahr, in dem sie erhoben wurden. Wenn der User bei dieser Angabe unsicher ist und das Feld bewusst freilassen möchte, soll er hier die Möglichkeit erhalten, alternativ ein Feld ‚unknown‘ auszuwählen. Der User wird in diesem Fall darauf hingewiesen, dass auf der späteren LandingPage das Uploadjahr der Daten angezeigt wird		1	Format: YYYY Alternative: unknown Für Datengeber wird ‚unknown‘ angeboten, wenn z.B. Entstehungsjahr der Daten unklar ist oder er das Feld bewusst freilassen möchte. Es wird Hinweis gegeben, dass auf der späteren RADAR-LandingPage dann alternativ das Uploadjahr der Daten angezeigt wird (productionYear=publicationYear); dies muss der externe Nutzer

² <http://www.loc.gov/catdir/cpsol/roman.html>

³ <http://www.utf8-zeichentabelle.de/>

		(productionYear=publicationYear); dies muss der externe Nutzer ausdrücklich erkennbar sein. DataCite-Pflichtfeld publication year soll Einstellungsjahr in RADAR sein & automatisiert aus technischen MD übernommen werden.			ausdrücklich erkennbar sein.
6	subject area	Einordnung in die Fachgebiete. Als Orientierung diene die englischsprachige GEPRIS-Liste der DFG ⁴ .		1-n	Auswahl eines oder mehrerer Fachgebiete, dem die Daten am ehesten zuzuordnen sind. Hier sollte (auch für vorgesehene Suchfunktion) die Auswahl über eine kontrollierte Liste erfolgen. Kontrollierte Liste: <ul style="list-style-type: none"> - Agriculture - Architecture - Arts and Media - Astrophysics and Astronomy - Biochemistry - Biology - Behavioural Sciences - Chemistry - Computer Science - Economics - Engineering - Environmental Science and Ecology - Ethnology - Geological Science - Geography - History - Horticulture - Information Technology - Linguistics - Materials Science - Mathematics - Medicine - Philosophy - Physics - Psychology - Social Sciences - Software Technology - Sports - Theology - Veterinary Medicine - Other
6.1	subject area detail	Nähere Beschreibung des Fachgebiets.	C	0-n	Das Format ist offen. Hier wird die Option geboten, die subjectArea in einem Freitext-Feld ggf. näher zu spezifizieren.
7	resource	Kategorisierung zur Art der Forschungsdaten.		1	Das Format ist offen. Hier sollen Angaben zur Art der Forschungsdaten gemacht werden.
7.1	resource type	Zusätzliche Angaben zur Art der Forschungsdaten.	A	1	Hier muss eine Angabe zur Kategorisierung des resourceType erfolgen. Kontrollierte Liste:

⁴ <http://gepris.dfg.de>

					<ul style="list-style-type: none"> - Audiovisual - Collection - Dataset - Image - Model - Software - Sound - Text - Workflow - Other
8	rights	Nachnutzungsrechte für Datensätze; wenn der Datengeber sich für einen Service entschieden hat.		1	<p>Kontrollierte Liste:</p> <ul style="list-style-type: none"> - <i>CC BY 4.0 Attribution</i> - <i>CC BY-ND 4.0 Attribution-NoDerivs</i> - <i>CC BY-SA 4.0 Attribution-ShareAlike</i> - <i>CC BY-NC 4.0 Attribution-NonCommercial</i> - <i>CC BY-NC-SA 4.0 Attribution-NonCommercial-ShareAlike</i> - <i>CC BY-NC-ND 4.0 Attribution-NonCommercial-NoDerivs</i> - <i>CC0 1.0 Universal Public Domain Dedication</i> - <i>All rights reserved</i> - <i>Other</i>
8.1	rights type	Weitere optionale Angaben zu Nutzungsrechten.	A	0-n	Hier wird die Option geboten, die rights in einem Freitext-Feld ggf. näher zu spezifizieren. Das Format ist offen.
9	rightsholder	Name des Inhabers der Nachnutzungs-rechte der Forschungsdaten.		1-n	Name eines Unternehmens/einer Institution oder Personennamen. Personen sollen im Format „Nachname, Vorname“ stehen. Namen, die nicht im lateinischen Alphabet geschrieben werden, sollten nach dem ALA-LC Schema transkribiert werden. ⁵

⁵ <http://www.loc.gov/catdir/cpsd/roman.html>

3. Zwölf optionale Felder

ID	Element	Definition	A/C	Occ	Erlaubte Werte, Beispiele, Beschränkungen
10	additional title	Titelergänzungen, z.B. Angaben zu alternativen Untertiteln, experimentellen Methoden, etc.		0-n	Das Format ist offen; Zeichen sollen im Rahmen von UTF-8 ⁶ erlaubt sein. → Wenn diese Information durch Datengeber gegeben wird, muss er die die Beschreibungen der nachfolgenden kontrollierten Liste zuordnen:
10.1	additional title type		A	1	Kontrollierte Liste zur Auswahl des additional title types; der Inhalt ist Freitext: - Subtitle - TranslatedTitle - AlternativeTitle
11	description	Abstract/Anmerkung zu Daten.		0-n	Format ist offen.
11.1	description type	Art der Angaben.	A	1	Kontrollierte Liste: - Abstract - TableOfContents - TechnicalRemarks - Object - Method - Other
12	keyword	Vergabe von Schlagwörtern.		0-n	Format ist offen. Hier soll explizit der Hinweis gegeben werden, dass bereits zuvor Begriffe aus dem Titel und ggf. Untertitel nicht noch mal wiederholt, sondern stattdessen Synonyme oder weitere, neue Begriffe verwendet werden sollen.
13	contributor	Alle weiteren Personen und Organisationen, die an der Erhebung des Datensatzes beteiligt waren.		0-n	Der Name eines Unternehmens/einer Institution oder ein Personennamen. Personennamen sollen im Format „Nachname, Vorname“ angegeben werden. Namen, die nicht im lateinischen Alphabet geschrieben werden, sollten nach dem ALA-LC Schema transkribiert werden. ⁷
13.1	contributor type	Mitwirkungsart der beitragenden Person/Organisation	A	1	Kontrollierte Liste: - ContactPerson - DataCollector - DataCurator - DataManager - Distributor - Editor - Funder - HostingInstitution

⁶ <http://www.utf8-zeichentabelle.de/>

⁷ <http://www.loc.gov/catdir/cpso/roman.html>

					<ul style="list-style-type: none"> - Producer - ProjectLeader - ProjectManager - ProjectMember - RegistrationAgency - RegistrationAuthority - RelatedPerson - Researcher - ResearchGroup - Sponsor - WorkPackageLeader - Other
14	language	Vorherrschende Sprache, in der die Daten dokumentiert sind.		0-1	Kontrolliertes Vokabular: ISO 639-3 Example: eng, fre, ger
15	alternate identifier	Alternativer Identifier.		0-n	Das Format ist offen. Der Datengeber kann angeben, wenn bereits ein anderer Identifier (auch z.B. ein institutsinterner Identifier) für den zu archivierenden/publizierenden Datensatz existiert. Beispiel: Lokale Nummer E-GEOD-34814
15.1	alternate identifier type	Beschreibung des alternativen Identifiers	A	1	Das Format ist offen. Wenn ein alternativer Identifier angegeben wird, muss hier die Art des Identifiers angegeben werden. Beispiel: interne Identifikationsnummer
16	related identifier	Hinweis/ Angaben zum ergänzenden Quellen.		0-n	Das Format ist offen. Der Datengeber kann hier entsprechende angaben zu ergänzenden Quellen/Materialien zum Originaldatensatz in RADAR machen.
16.1	related identifier type	Art der Quelle/des Materials.	A	1	Wenn 17. angegeben wird, muss hier der related identifier type ausgewählt werden. Kontrollierte Liste: <ul style="list-style-type: none"> - ARK - arXiv - bibcode - DOI - EAN13 - EISSN - Handle - ISBN - ISSN - ISTC - LISSN - LSID - PMID - PURL - UPC - URL - URN Eine Option „Other“ wird <u>nicht</u> angeboten, da im Zweifel bei digitalen Inhalten immer eine URL

					existiert.
16.2	relation type	Art der Beziehung zwischen den in RADAR aufzunehmenden Forschungsdaten und der verwandten Ressource.	A	1	Wenn 17.1 angegeben wird, muss auch die Art der Beziehung angegeben werden: Kontrollierte Liste: - IsCitedBy - Cites - IsSupplementTo - IsSupplementedBy - IsContinuedBy - Continues - HasMetadata - Is MetadataFor - IsNewVersionOf - IsPreviousVersionOf - IsPartOf - HasPart - IsReferencedBy - References - IsDocumentedBy - Documents - IsCompiledBy - Compiles - IsVariantFormOf - IsOriginalFormOf - IsIdenticalTo - IsReviewedBy - Reviews - IsDerivedFrom - IsSourceOf
17	geo location	Geographische Region, Land oder Ort, an dem Daten erhoben wurden oder auf welchen sich die Daten beziehen.		0-n	
17.1	geo location country	Beschreibung eines Landes, in dem Daten erhoben wurden bzw. auf welches sich die Daten beziehen.	C	0-1	Kontrolliertes Vokabular: englischer Ländername nach ISO 3166 ⁸ Example: GERMANY
17.2	geo location region	Beschreibung einer Region.	C	0-1	Das Format ist offen. z.B. geographische Region oder Gebiet, in dem die Daten erhoben wurden.
17.3	geo location point	Ein definierter Ort im Raum.	C	0-1	Ein Ortspunkt enthält ein einziges Koordinatenset bestehend aus Breiten- und Längengradangabe, die durch ein Leerzeichen getrennt sind (nach dem Schema des World Geodetic System 84 ⁹). Ort: Dezimal(-90 bis 90Süd/Nord) LEERZEICHEN Dezimal(-180 bis 180West/Ost) Beispiel: 31.233 -67.302
17.4	geo location box	Beschreibt die Grenzen eines definierten Orts im Raum.	C	0-1	Die räumlichen Grenzen werden durch zwei Koordinatensets jeweils bestehend aus einer Breiten- und einer Längengradangabe definiert. Das

⁸ <ftp://ftp.fu-berlin.de/doc/iso/iso3166-countrycodes.txt>

⁹ http://www.oosa.unvienna.org/pdf/icg/2012/template/WGS_84.pdf

					erste Koordinatenpaar beschreibt die untere linke Ecke, das zweite Koordinatenpaar die obere rechte Ecke des Raums. Beispiel: 42.893 -71.032 41.090 -68.211
18	data source	Beschreibung Prozedur oder Quelle für Datenerhebung.		0-n	Das Format ist offen, Freitext.
18.1	data source detail	Detaillierte Angaben zur Art der Datenquelle/Grundlage der Datenerhebung.	A	1	Wenn 18 angegeben wird, muss auch die Art der Quelle angegeben werden: Kontrollierte Liste: - Instrument - Media - Observation - Trial - Organism - Tissue - Other
19	software type	Beschreibung der zugehörigen Software.		0-n	Spezifiziert die Art der Software, die bei der Erzeugung, Bearbeitung, Ansicht oder anderweitigen Verwertung der Datenquelle Anwendung fand bzw. findet. Kontrollierte Liste: - Resource Production - Resource Processing - Resource Viewing - Other
19.1	software name	Bezeichnung der Software.	C	1-n	Angabe zur Bezeichnung und ggf. weitere Informationen zur verwendeten Software.
19.1.1	software name version	Version(en) der Software.	A	1	Pflichtfeld, sofern 19.1 angegeben wurde. Das Format ist offen.
19.2	software alternative	Angabe alternativer Software.	C	0-n	Optionale Angabe, welche weitere alternative Software für die zuvor spezifizierte Software-Art (Erzeugung, Bearbeitung, Ansicht oder anderweitigen Verwertung) Anwendung finden kann, z.B. alternative Open Source Software.
19.2.1	software alternative version	Version(en) der alternativen Software.	A	1	Pflichtfeld, sofern 19.2 angegeben wurde. Das Format ist offen.
20	data processing	Angabe von weiteren, ggf. sekundären Modifikationen an Forschungsdaten.		0-n	Das Format ist offen, Freitext. Hier kann der Datengeber Angaben machen, wenn z.B. Rohdaten weiter bearbeitet wurden (z.B. Statistik).
21	related information	Angabe von bestimmten, wichtigen Komponenten die in den Forschungsdaten ggf. enthalten sind.		0-n	Das Format ist offen, Freitext.
21.1	related information type	Art der enthaltenen Komponente.	C	0-n	Freitext zur Auswahl des relatedInformationTypes Beispiele für mögliche Inhalte: Database ID, Registrierungsnummer, GenBank,

				IntEnz, PubChem, MedGen, PMID, PDB, Molecular Formula
--	--	--	--	--