RatSWD Working Paper Series

www.ratswd.de

RatSWD
Rat für Sozial- und WirtschaftsDaten

242

da|ra 3.0/3.1

– Was hat sich bei der DOI-Registrierung für Sozial- und Wirtschaftsdaten geändert?

Brigitte Hausstein und Ralf Toepfer

Dezember 2014

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Working Paper Series des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

Die *RatSWD Working Papers* Reihe startete Ende 2007. Seit 2009 werden in dieser Publikationsreihe nur noch konzeptionelle und historische Arbeiten, die sich mit der Gestaltung der statistischen Infrastruktur und der Forschungsinfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften beschäftigen, publiziert. Dies sind insbesondere Papiere zur Gestaltung der Amtlichen Statistik, der Ressortforschung und der akademisch getragenen Forschungsinfrastruktur sowie Beiträge, die Arbeit des RatSWD selbst betreffend. Auch Papiere, die sich auf die oben genannten Bereiche außerhalb Deutschlands und auf supranationale Aspekte beziehen, sind besonders willkommen.

RatSWD Working Papers sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Alle Arbeiten können und sollen auch in fachlich, institutionell und örtlich spezialisierten Reihen erscheinen. Die *RatSWD Working Papers* können nicht über den Buchhandel, sondern nur online über den RatSWD bezogen werden.

Um nicht deutsch sprechenden Nutzer/innen die Arbeit mit der Reihe zu erleichtern, sind auf den englischen Internetseiten der *RatSWD Working Papers* nur die englischsprachigen Papers zu finden, auf den deutschen Seiten werden alle Nummern der Reihe chronologisch geordnet aufgelistet.

Einige ursprünglich in der *RatSWD Working Papers* Reihe erschienenen empirischen Forschungsarbeiten sind ab 2009 in der RatSWD Research Notes Reihe zu finden.

Die Inhalte der *RatSWD Working Papers* stellen ausdrücklich die Meinung der jeweiligen Autor/innen dar und nicht die des RatSWD. Das Bundesministerium für Bildung und Forschung hat die Publikationen nicht beeinflusst.

Herausgeber der RatSWD Working Paper Series:

Vorsitzender des RatSWD
(seit 2014 Regina T. Riphahn; 2009-2014 Gert G. Wagner; 2007-2008 Heike Solga)

The Registration Agency da|ra

da|ra operates as the registration agency for social science and economic data jointly run by GESIS (http://www.gesis.org) and ZBW (http://www.zbw.eu). da|ra pursues the goal of long-term, persistent identification and availability of research data via allocation of DOI names. In keeping with the ideals of good scientific practice there is a demand for open access to existing primary data so as to not only have the final research results but also be able to reconstruct the entire research process. GESIS and ZBW therefore offer a registration service for social and economic research data in cooperation with DataCite (http://www.datacite.org), an international consortium pursuing the goal of supporting the acceptance of research data as independent citable scientific objects. This infrastructure lays the foundation for long-term, persistent identification, storage, localization and reliable citation of research data.

Benefits of DOI names:

- **Permanent, persistent identification:** Each DOI name uniquely, unequivocally and permanently identifies the assigned object.
- **Availability of information on the web:** Via the Handle System, each DOI refers to one or more webpages assigned by the publication agent.
- **Semantic Interoperability:** The metadata associated with a DOI name enable direct, precise communicating – with each user, from every location, at every point in the production/distribution chain – with regard to every detail of the objects related with one another.

The DOI name is comprised of a unique alphanumeric character string; a prefix and suffix, whereby the prefix always begins with "10" and prefix and suffix are separated by a forward slash. Prefixes are assigned by the International DOI Foundation (IDF http://www.doi.org) via DataCite. Each data centre is assigned its own prefix thus permitting an unlimited number of DOI names. The suffix is agreed by the publication agent in conjunction with da|ra.

Each DOI name permanently identifies the assigned object as an entity regardless of whether the storage location changes. Updated, structured metadata is assigned to the resource using the DOI name. The allocation of DOI names to the objects transpires automatically following successful transmission of the metadata per object to be registered.

In 2013, da|ra began to develop value-added services. These included preparations for the extension of the registration service to other resource types. As social science and economic research produces not only datasets, but also other resource types it was a logical step. da|ra Metadata Schema Version 3.0 was therefore expanded for the registration of the resources Collection, Text, Image, Video, Audio and Interactive Resource.

The Metadata Schema

The da|ra Metadata Schema is a list of core metadata properties chosen for the identification of data and retrieval purposes. Each DOI name is linked to a set of metadata, a collection of bibliographical and content information, which describe in detail the registered resources (title, author, publication date, copyright etc.) and present the properties of resources, their structure and contextual relations.

The da|ra Metadata Schema provides a determined number of mandatory elements – core properties –, that have to be submitted by the publication agent at the time of data registration. Publication agents may also choose to use optional properties to identify their data more clearly.

For all metadata properties the respective names, definitions, attributes, conditions, cardinality (maximum occurrence) as well as value domains are defined. Some properties comply with ISO norms. These norms determine e. g. which code for a language or geographic coverage has to be applied. Controlled vocabularies such as thesauri and classifications are applicable. These vocabularies are complemented by da|ra controlled terms.

Although da|ra complies with the official DataCite Metadata Schema, it has broadened the DataCite metadata by adding some specific properties related to the social sciences and economics.

Please note that da|ra reserves the right to share metadata with information indexes and other entities.[1] da|ra metadata are subject to Creative Commons CC0 license.

Version 3.1 Update

The current Version 3.1 is based on the Version 3.0 of the da|ra Metadata Schema and has been further developed in line with the DataCite Metadata Schema, Version 3.1, to achieve greater exposure of social and economic research data outputs for all resource types.

Version 3.1 of the da|ra Metadata Schema introduces some notable changes:

Addition of:

10.1	Internal Resource Identifier *
23.2.3	Contributor Affiliation *
23.2.4	Contributor Affiliation ID *
23.2.4.1	Vocabulary of Contributor Affiliation ID *
23.2.4.2	URI Name Authority Record *
38	Publication Place (optional for all resource types) *

Addition of new values to controlled lists:

23.1	Contributor Type: *DataCurator*
24	Collection Mode
31.1	Relation Type: *IsReviewedBy, Reviews, IsDerivedFrom, IsSourceOf*
31.2	Identifier Type: *arXiv, bibcode*

Documentation:

The documentation of Version 3.1 of the da|ra Metadata Schema includes the following changes:

- Documentation for
  - the sub-property Internal Resource Identifier
  - the affiliation attributes for Contributor
  - the new Contributor Type DataCurator

[1] See da|ra Service Level Agreement and Policy in its respective current form http://www.da-ra.de/en/about-us/da-ra-policy/

* Properties and sub-properties marked with a star (*) were already implemented in the previous Version 3.0, which was unfortunately forgot to list in the Technical Report 2014|07. Therefore, they are still listed as an update here.

    - the new Collection Modes
    - the four new Relation Types IsReviewedBy, Reviews, IsDerivedFrom, IsSourceOf
    - the two new Identifier Types arXiv and bibcode
    - version changes
- Better indication of recommended metadata
- Addition of more detailed XML examples on the da|ra website
- Deletion of documentation for administrative properties DOI name, publication agent and registration agency, as they are not part of the metadata set submitted by the publication agent

Please note that da|ra Metadata Schema Version 3.0 is still valid and can also be used.

A Note about da|ra DOI Registration

da|ra obtains the DOI names via the GESIS membership in DataCite. DataCite is accredited as an official DOI registration agency within the DOI foundation (IDF http://www.doi.org).

TIB Hanover (http://www.tib.uni-hannover.de) acts as a managing agent of DataCite and organizes the control of prefixes and the connection to IDF. The figuration of the suffixes is done by the publication agents and is determined in the Service Level Agreement.

da|ra governs the assignment of DOI names. It functions as the DOI allocation agency and is not commercially oriented. Besides the DOI allocation, da|ra is responsible for the elaboration of the service agreement (Service Level Agreement) together with the publication agents as well as for the administration of metadata. For both the maintenance and the storage of metadata the data centres are responsible.

Ensuring that metadata is persistent does not exclude its modifiability: data producers have the opportunity to amend the metadata whenever and as often as needed.

The following presentation (in German) was held on the DataCite Workshop "Bessere Zitierbarkeit und Auffindbarkeit von Forschungsdaten durch DataCite-DOIs" in Hamburg on 16 January 2014
http://zbw.eu/de/ueber-uns/veranstaltungen/calendar/2014/01/16/event/tx_cal_phpicalendar/2_datacite_workshop_bessere_zitierbarkeit_und_auffindbarkeit_von_forschungsdaten_durch_datacite_do/

References

Helbig, Kerstin, Brigitte Hausstein, Ute Koch, Jana Meichsner, and Andreas Oskar Kempf. 2014. *da|ra Metadata Schema: Version 3.1*. GESIS - Technical Reports 2014/17. Köln: GESIS. doi: http://dx.doi.org/10.4232/10.mdsdoc.3.1.

Helbig, Kerstin, and Brigitte Hausstein. 2014. *Best practice guide for the registration of resources with da|ra*. GESIS - Technical Reports 2014/18. doi: http://dx.doi.org/10.4232/10.bpg.1.0.

Hausstein, Brigitte, Natalija Schleinstein, Ute Koch, Jana Meichsner, Kerstin Becker, and Lena-Luise Stahn. 2014. *da|ra Metadata Schema: Version 3.0*. GESIS - Technical Reports 2014/07. doi: http://dx.doi.org/10.4232/10.mdsdoc.3.0.

DataCite Metadata Schema Repository. http://schema.datacite.org/

da|ra Registrierungsagentur für Sozial- und Wirtschaftsdaten

da|ra 3.0/3.1 – Was hat sich bei der DOI-Registrierung für Sozial- und Wirtschaftsdaten geändert?

Brigitte Hausstein (GESIS- Leibniz Institut für Sozialwissenschaften)
Ralf Toepfer (ZBW)

2. DataCite Workshop am 16. Januar 2014 in Hamburg
(einschließlich update da|ra 3.1. Berlin/Hamburg 20.12.2014)

DataCite

da|ra 3.0/3.1 – was ist neu?

- Öffnung des da|ra-Systems für andere Ressourcentypen:
  Sammlung, Bild, Video, Audio, Interaktive Ressource
- Eingabemaske und Webservice für alle Ressourcentypen
- Anpassung der Anzeige
- Metadatenschema 3.0 (Release März 2014)
- Metadatenschema 3.1 (Release November 2014)
- Plug-In für das Open Journal System (OJS)

DataCite

Updates Metadatenschema da|ra 3.0

	da|ra 2.2.1		da|ra 3.0/3.1	Bemerkungen
0	Ressourcentyp	0	Genereller Ressourcentyp	Bezeichnung des generellen Ressourcentyps (da|ra kontrollierte Liste). **Pflichtelement**.
		0.1	Ressourcentyp	Beschreibung der Ressource. Format ist offen, es sollte aber ein einzelner, ergänzender Begriff zum Generellen Ressourcentyp sein, damit ein Paar gebildet werden kann (Text/Artikel; Bild/Tabelle). **Optionales Element**.
18	Geographischer Raum	18	Geographischer Raum	Oberbegriff
18.1	Geographischer Raum (kontrolliert)	18.1	Geographischer Raum (kontr.)	Kontrollierte Liste (ISO 2-3)
18.2	Geographischer Raum (frei)	18.2	Geographischer Raum (frei)	Geographische Koordinaten (nach WGS 84 - World Geodetic System)
		18.3	geoLocationPoint	Angaben zu Länge und Breite, getrennt durch Leerzeichen. Beispiel: <geoLocationPoint> 31.233 - 67.302 </geoLocationPoint>
		18.4	geoLocationBox	Geographische Ausdehnung nach WGS 84 Eine Box enthält zwei getrennte Paarangaben zu Breite und Länge. Die ersten Paarangaben beziehen sich auf die niedrigere Ecke, die zweiten auf die obere Ecke. Beispiel: <geoLocationBox>42.893 -71.032 41.090 -68.211 </geoLocationBox

DataCite

Updates Metadatenschema da|ra 3.0

	da|ra 2.2.1		da|ra 3.0/3.1	Bemerkungen
23	Datenerhebung	23	Contributor Name/Mitwirkender	Die Erfassung Person/Institution analog 4 (Primärforscher) Kontrollierte Liste (DataCite). Wenn das Element Mitwirkender ausgefüllt ist, dann ist die Angabe zu "Typ der Mitwirkender" Pflicht.
23.1	ID Datenerhebung	23.1	Contributor Type	
23.1.1	Vokabular ID Datenerhebung	23.2	Contributor Name ID	
23.1.2	URI zum Normdatensatz	23.2.1	Vokabular Contributor Name ID	Datenerhebung wird unter **contributorType DataCollector** erfasst
		23.2.2	URI zum Normddatensatz	
		23.2.3	Contributor Affiliation	Die Affiliation des Contributors. **Optionales Element**.
		23.2.4	ID Contributor Affiliation	Unique Identifier der Affiliation in Anlehnung an unterschiedliche Schemata. Kann von da|ra ergänzt werden, wenn nicht vorhanden. **Optionales Element**.
		23.2.4.1	Vokabular ID Contributor Affiliation	Name des Vokabulars, aus dem der Affiliation-Identifier stammt. Pflicht, wenn 23.2.4 (ID Contributor Affiliation) ausgefüllt ist
		23.2.4.2	URI zum Normdatensatz	Persistenter Identifier des Normdatensatzes

DataCite

Updates Metadatenschema da|ra 3.0

	da|ra 2.2.1		da|ra 3.0/3.1	Bemerkungen
26 **26.1** **26.2** **26.3** **26.4** **27** **27.1** **27.2** **27.3** **27.4** **27.5**	Daten Typ der Einheiten Anzahl der Einheiten Anzahl der Variablen Typ der Daten Technische Beschreibung der Daten Dateiname Datenformat Größe Daten-Fingerprint Verfahren Fingerprint	26 26.1 26.2 26.3 26.4 26.5 26.6 26.7 26.8 26.9	Daten Typ der Einheiten Anzahl der Einheiten Anzahl der Variablen Typ der Daten Dateiname Datenformat Größe Daten-Fingerprint Verfahren Fingerprint	**Neue Zuordnung** Alle Subelemente von (26) Daten können unabhängig voneinander erfasst werden. **Alle optional**. Ausnahme: 26.1 (Typ der Einheiten) ist Pflicht wenn 26.2 (Anzahl der Einheiten) angegeben ist. Und die 26.2 (Anzahl der Einheiten) ist Pflicht wenn 26.1 (Typ der Einheiten) angegeben ist.
38	Herausgeber			**Removal (wird unter contributorType Herausgeber erfasst)**

DataCite

Updates Metadatenschema da|ra 3.0

	da|ra 2.2.1		da|ra 3.0/3.1	Bemerkungen
31 **31.1** **31.2**	Relation Art der Relation Typ des Identifiers	31 31.1 31.2 31.3 31.4 31.5	Relation Art der Relation Typ des Identifiers Vokabular der Metadaten (relatedMetadataScheme) URI des Vokabulars (schemeURI) Typ des Vokabulars (schemeType)	**Neue sub elements:** Werden benutzt nur in Verbindung mit den Deskriptoren "Hat Metadaten"/"Ist Metadaten für" (Art der Relation)
38	Herausgeber			**Removal (wird unter contributorType Herausgeber erfasst)**

DataCite

Updates der kontrollierten Vokabulare

	Element	da|ra 2.2.1	da|ra 3.0/3.1	Bemerkungen
0	Genereller Ressourcentyp	1.Sammlung 2. Datensatz 3. Graue Literatur 4. Film 5. Bild 6. Text	1. Sammlung 2. Datensatz 3. Text 4. Bild 5. Video 6. Audio 7. Interaktive Ressource	Graue Literatur wird in Text umbenannt. **Neue Typen**: Video, Audio, Interaktive Ressource
17.1	Art der Beschreibung	1. Zusammenfassung 2. Information zur Schriftenreihe 3. Inhaltsverzeichnis 4. Sonstiges	1. Zusammenfassung 2. Information zur Schriftenreihe 3. Inhaltsverzeichnis 4. Methoden 5. Sonstiges	**Neuer Deskriptor**: Methoden

DataCite

Updates der kontrollierten Vokabulare

	Element	da|ra 2.2.1	da|ra 3.0/3.1	Bemerkungen
23.1	contributorType		1. ContactPerson/Kontaktperson 2. DataCollector/Datenerhebung 3. DataManager/Datenmanager 4. Distributor/Vertrieb 5. Editor/Herausgeber 6. Funder/Geldgeber 7. HostingInstitution/Anbieter 8. Producer/Produzent 9. ProjectLeader/Projektleiter 10. ProjectManager/Projektmanager 11. ProjectMember/Projektmitarbeiter 12. RegistrationAgency/Registrierungsagentur 13. RegistrationAuthority/Registrierungsbehörde 14. RelatedPerson/Verbundene Person 15. Researcher/Forscher 16. ResearchGroup/Forschungsgruppe 17. RightsHolder/Rechtsinhaber 18. Sponsor/Sponsor 19. Supervisor/Administrator 20. WorkPackageLeader/Leiter des Arbeitspakets 21. Other/Sonstiges 22. DataCurator/Datenkurator	**neue kontrollierte Liste contributorType und bei 3.1 neuer Typ** Datenkurator

DataCite

Updates der kontrollierten Vokabulare

	Element	da|ra 2.2.1	da|ra 3.0/3.1	Bemerkungen
24	collectionMode		Erweiterung der kontrollierten Liste um: 25. Interview: PAPI/Interview: PAPI 26. Interview: Web-based/Interview: Web-basiert 27. Self-completed questionnaire: Fixed form/ Selbst ausgefüllter Fragebogen: standardisierter 28. Self-completed questionnaire: E-mail/ Selbst ausgefüllter Fragebogen: E-Mail 29. Self-completed questionnaire: SMS/MMS/ Selbst ausgefüllter Fragebogen: SMS/MMS 30. Self-completed questionnaire: Interactive/ Selbst ausgefüllter Fragebogen: Interaktiv 31. Self-completed questionnaire: VCASI/ Selbst ausgefüllter Fragebogen: VCASI 32. Self-completed questionnaire: TACASI/ Selbst ausgefüllter Fragebogen: TACASI 33. …	**Neue Deskriptoren:** 25 – 53 (Anpassung an DDI)

Updates der kontrollierten Vokabulare

	Element	da|ra 2.2.1	da|ra 3.0/3.1	Bemerkungen
31.1	Art der Relation	1. Wird zitiert von 2. Zitiert 3. Ist Ergänzung zu 4. Wird ergänzt durch 5. Wird fortgesetzt von 6. Setzt fort 7. Ist neue Version von 8. Ist vorherige Version von 9. Ist Teil von 10. Enthält Teil von 11. Wird referenziert von 12. Verweist auf / Referenziert 13. Wird dokumentiert von 14. Dokumentiert 15. Erstellt von 16. erstellt 17. Ist Variante von 18. Ist Original von	1. Wird zitiert von 2. Zitiert 3. Ist Ergänzung zu 4. Wird ergänzt durch 5. Wird fortgesetzt von 6. Setzt fort 7. Ist neue Version von 8. Ist vorherige Version von 9. Ist Teil von 10. Enthält Teil von 11. Wird referenziert von 12. Verweist auf / Referenziert 13. Wird dokumentiert von 14. Dokumentiert 15. Erstellt von 16. erstellt 17. Ist Variante von 18. Ist Original von 19. Hat Metadaten 20. Ist Metadaten für 21. Ist identisch zu 22. Wird rezensiert von 23. Rezensiert 24. Ist abgeleitet von 25. Ist Quelle von	**Neue Deskriptoren**: 19 - 25

Updates der kontrollierten Vokabulare

	Element	da|ra 2.2.1	da|ra 3.0/3.1	Bemerkungen
31.2	Typ des Identifiers	ARK, DOI, EAN13, EISSN, Handle, ISBN, ISSN, LISSN, LSID, PURL, UPC, URL, URN	ARK, DOI, EAN13, EISSN, Handle, ISBN, ISSN, LISSN, LSID, PURL, UPC, URL, URN, ISTC, PMID, arXiv, bibcode	**Neue Deskriptoren**
32.1.14	Dokumententyp	10. Magazine	10. Newspaper	**Übersetzungskorrektur**

DataCite

Mehr Informationen:

da|ra Metadatenschema:

http://doi.org/10.4232/10.mdsdoc.3.1

da|ra Best Practice Guide:

http://doi.org/10.4232/10.bpg.1.0

DataCite