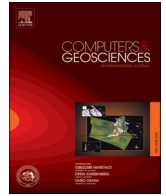




Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

The BonaRes metadata schema for geospatial soil-agricultural research data – Merging INSPIRE and DataCite metadata schemes



Xenia Specka^{a,*}, Philipp Gärtner^a, Carsten Hoffmann^a, Nikolai Svoboda^a, Markus Stecker^b, Udo Einspanier^b, Kristian Senkler^b, M.A. Muqit Zoarder^a, Uwe Heinrich^a

^a Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Str. 84, 15374, Müncheberg, Germany

^b con Terra GmbH, Martin-Luther-King-Weg 20, 48155, Münster, Germany

ARTICLE INFO

Keywords:

Metadata schema
Metadata mapping
Crosswalk
INSPIRE
DataCite 4.0
BonaRes Repository
FAIR data principles

ABSTRACT

A spatial data infrastructure (SDI) for the upload and provision of soil-agricultural research data in Germany was developed and launched in 2017. The precondition for the new SDI were to be compliant with the European initiative for spatial information (INSPIRE), to consider FAIR data principles, to be interoperable with other disciplinary national and international SDIs and to support dataset registrations with digital object identifiers (DOI). To meet these requirements, the new SDI had to support both the INSPIRE and DataCite metadata standards. As there was no metadata standard available that supported both metadata schemas from different disciplines, it became necessary to merge these two existing standards.

The objective of this paper is to describe the developmental steps of the newly created BonaRes metadata schema for geospatial soil-agricultural research data. First, we analyzed the INSPIRE and DataCite metadata schemas, identified and compared semantically equivalent metadata elements, including properties and allowed content values, for potential mapping. In the next step, we specified the new metadata model, whose elements were derived from INSPIRE or DataCite under consideration of the previously developed mapping. Third, additional metadata elements were defined to directly describe the underlying data model.

The advantages and disadvantages of the method that was applied to define the BonaRes metadata schema elements originating from INSPIRE and DataCite schemas are presented and discussed. By implementing the new metadata model, the SDIs can achieve cross-portal metadata interoperability with other INSPIRE-compliant SDIs. Furthermore, researchers who store their geospatial research data in SDIs to implement the BonaRes metadata schema can increase the visibility and findability of their research, as the data can be assigned a DOI, which is a prerequisite for data citations and data publications. The new metadata elements of the BonaRes metadata schema allows targeted dataset queries, e.g., soil pH values, yields or parameters, which were measured by a specific method, using the underlying data model, and improves the discovery and reusability of research data.

1. Introduction

There is an increasing trend to publish geospatial data in spatial data infrastructures (SDIs), both at the German national level (Simmons, 2018; Stein et al., 2019) and the international level (Masser, 2005; Christian et al., 2013; Trilles et al., 2017). According to state-of-the-art in international research data management, data should be findable (F) for the research community, easily accessible (A) by download functions, interoperable (I) with other data and repositories, and reusable (R), e.g., for modelers. These four requirements are described as FAIR data principles in data science (Wilkinson et al., 2016). The ability to

discover, analyze and reuse data that is managed in SDIs by other users, i.e. scientists, depends on the provided metadata that describe a) how the data were collected, processed or organized, and b) how the data are provided for reuse. Metadata standards were developed to facilitate data sharing and to define a common set of terms for metadata (Simmons, 2018). The use of standardized metadata ensures good documentation of the data and supports their interoperability and compatibility with other international infrastructures (Chan and Zeng, 2006; Zeng and Chan, 2006). There are many existing metadata standards that can be used to describe geospatial data (Laxton and Duffy, 2011), e.g., ISO 19115 or ISO 19139, INSPIRE metadata regulation (European Commission Joint

* Corresponding author.

E-mail addresses: specka@zalf.de (X. Specka), zoarder@zalf.de (M.A.M. Zoarder).

<https://doi.org/10.1016/j.cageo.2019.07.005>

Received 29 January 2019; Received in revised form 20 June 2019; Accepted 8 July 2019

Available online 9 July 2019

0098-3004/© 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Research Centre, 2007), INSPIRE Soil (INSPIRE Thematic Working Group Soil, 2013), FGDC (FGDC, 1998), EDMED (EU Marine Science and Technology, 2019), USGIN metadata profile (USGIN Standards and Protocols Drafting Team, 2009), UK AGMAP (Academic Geospatial Metadata Application Profile) and GeoSciML (Open Geospatial Consortium Inc., 2017). An overview of different available metadata standards can be found in RDA Metadata Standards Catalog Working Group (2019), RDA Metadata Standards Directory Working Group (2019) or Hoffmann et al., 2018.

Two open and widely accepted metadata standards are of great relevance for international geospatial data management: 1) The INSPIRE directive (2007) of the European Union forced the standardization of SDIs across member states (The European Parliament and the Council of the European Union, 2007). Until 2020, the directive will be implemented in a stepwise manner to allow the discovery, download and visualization of geospatial data across country boundaries. INSPIRE provides a metadata schema that focuses on the description of geospatial data or services and is based on the international ISO standards, ISO 19115, ISO 19119 and ISO 19139. INSPIRE metadata cross-portal interoperability between distributed SDIs can be achieved by implementing the OGC Catalogue Service of the Web (CSW) standard (Open Geospatial Consortium Inc., 2007b). 2) DataCite as an international nonprofit organization that aims to improve research data citations. If datasets are described by DataCite’s metadata elements (DataCite, 2016) a digital object identifier (DOI) can be assigned, which is a persistent identifier that is required for data citations and data publications. The DOI assures reliable and unambiguous access to data, supports its reuse and allows proper attribution (Chavan and Penev, 2011; Neumann and Brase, 2014). The use of the DOI system offers incentives for scientists to make this extra effort to provide detailed metadata for their data. Research data that are assigned a DOI can be referenced and cited, which is the basis for data publication.

Agricultural sciences comprise a broad and heterogeneous field of research disciplines such as soil science, plant production, applied genetics and physiology, agricultural economics and sociology. Due to the wide range of research questions among the various agricultural disciplines, a high diversity of data is generated, e.g., plant observational data, soil-profile information, land-use data, climate data, including geospatial information. A new SDI, the BonaRes Repository (BonaRes Data Centre, 2018; Hoffmann et al., 2018b), was established with a special focus on permanently storing geospatial soil-agricultural research data in Germany. This infrastructure aims to be an important access point when looking for national and international soil-agricultural research data.

The new metadata schema that was implemented with the BonaRes Repository should meet the following requirements:

1. **INSPIRE interoperability:** To ensure interoperability and connectivity with existing national and international SDIs by applying national and international standards for metadata, network services and data specifications and to guarantee data accessibility for the community. Thus, the new SDI must be INSPIRE-conformant for disseminating metadata based on the OGC CSW interface (Open Geospatial Consortium Inc., 2007b).
2. **DOI registration:** Provide all of the required metadata information and technical infrastructure to allow for the registration of research data with a DOI.
3. **Description of the data model:** Increase the usability of the research data by describing the underlying data model, e.g., methods used, units and data relations.
4. **AGROVOC integration:** Use keywords from the AGROVOC thesaurus (FAO, 2018; Rajbhandari and Keizer, 2012) to describe and align soil-agricultural research data.

To be compliant with INSPIRE and to be able to register data with a DOI, the BonaRes Repository had to support the INSPIRE and the

DataCite metadata schemas. As there were no metadata schemas available that served both standards, a new metadata model needed to be developed. All specifications and regulations of the INSPIRE and DataCite metadata schemas, with regard to metadata elements and their properties, had to be integrated into the new BonaRes metadata schema. All mandatory/obligatory metadata elements of INSPIRE and DataCite must be assigned by this model, and, at the same time, the number of elements should be kept as small as possible to avoid redundancies and to foster user-friendliness. The new metadata schema had to include all metadata elements of INSPIRE and DataCite within a single XML schema. Metadata in the new BonaRes format could be either derived into the INSPIRE (ISO 19139) or DataCite XML format.

This paper’s objective is to introduce the new BonaRes metadata schema and the method we applied to integrate the INSPIRE and DataCite metadata schemas into the new model. First, an overview of the INSPIRE and DataCite metadata schemas is given. In the second section, we describe how we adopted the method for metadata crosswalks, including the mapping between the INSPIRE and DataCite metadata schemas to develop the new metadata model. Section 3 presents the results of the developed metadata crosswalk and introduces the new BonaRes metadata schema. This is followed by a chapter that discusses the results of the metadata mapping and the new metadata model (Section 4). Section 5 presents the conclusions and planned future work.

2. Materials and methods

2.1. INSPIRE metadata model

The INSPIRE metadata model includes 27 main metadata elements consisting of 16 mandatory and 11 conditional elements (Fig. 1) (European Commission Joint Research Centre, 2007). Each main element is detailed by different subelements. In INSPIRE, three different resource types (spatial dataset, spatial dataset series and service) are distinguished. Depending on the resource type, conditional elements may become mandatory. More detailed information about the INSPIRE metadata model can be found in (European Commission Joint Research Centre, 2007; Reznik, 2013; Trilles et al., 2017).

2.2. DataCite metadata schema 4.0

The DataCite consortium was founded in 2009 by leading research libraries and information centers (Brase, 2009; Petritsch, 2017). The main objective was to provide easy online access to research data and improve its citability (Petritsch, 2017). The DataCite metadata schema is

INSPIRE					
Resource title	Resource abstract	Resource type	Topic category	Unique resource identifier	Responsible organization (Point of contact)
Metadata language	Resource language	Resource locator	Coupled resource	Spatial resolution	Lineage
Metadata date	Temporal extent	Date of publication	Date of last revision	Date of creation	Conformity
Metadata point of contact	Coordinate reference system	Access constraints	Other constraints		Keyword value
Temporal reference	Limitations on public access		Conditions applying to access & use		Geographic bounding box

Mandatory
 Conditional

Fig. 1. Main elements of the INSPIRE metadata model.

generally based on Dublin Core (Weibel, 1997). It contains 19 main metadata elements that consist of six mandatory fields, six recommended fields and seven optional fields (Fig. 2). Three different levels of obligation are distinguished: mandatory (M), recommended (R) and optional (O). A detailed description of elements together with information on the cardinality and allowed values can be found in DataCite, 2016.

2.3. Metadata crosswalk

A metadata crosswalk is a specification of how elements, semantics and syntax from one metadata scheme can be mapped to another schema (Chan and Zeng, 2006; St Pierre and LaPlant, 1998). We used this technique to define metadata elements and properties of the BonaRes metadata schema in such a way that 1) metadata records for INSPIRE and DataCite can be derived and 2) the complexity of the resulting model is as small as possible. Considering the study of St Pierre and LaPlant, (1998) in which various aspects and problems for setting up the metadata crosswalk are discussed, we have taken the following steps to develop the specifications of the new metadata model:

1. *Define commonly used terminology to deal with the content and elements of all considered standards.* As there is no common terminology among different metadata standards, it is necessary to define a common set of terminology that is used in the specification of metadata standards.
2. *Identify similarities of both used metadata standards and generalize similar concepts.* Similar properties of the used metadata standards, e. g., a unique identifier or multiplicity, which are named differently in both standards, needs to be identified to simplify the metadata mapping specification.
3. *Determine the semantic mapping of elements from both standards.* Specify a mapping of each element of the INSPIRE standard to a semantically equivalent element of the DataCite standard.
4. *Mapping of properties:* When mapping each element of one standard to another, all properties of the metadata elements, e.g., the obligational level or multiplicity, must be considered, too. If semantically mapped elements of different standards have identical properties, the resulting mapped property will be the same. If the properties differ, then a mapping of all possible combinations of the properties' values of both standards must be specified.

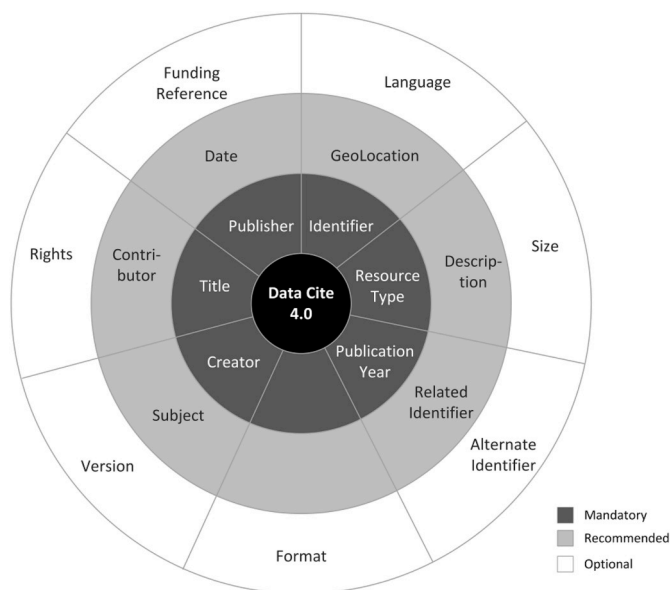


Fig. 2. DataCite 4.0 properties with different obligation levels. Detailed descriptions of the properties can be found in DataCite, 2016.

5. *Convert the content of elements:* The content of elements is sometimes restricted by the underlying data type, range of values or controlled vocabulary. When mapping elements of different standards, a mapping of allowed values has to be developed for those elements that differ in their content's restriction.

While the described technique of using metadata crosswalks to develop the specifications of the new metadata model, each of the metadata standards were not simply mapped onto one another. Moreover, the crosswalk method was used to combine INSPIRE and DataCite elements and properties into the new metadata model in such a way that it would then be possible to derive metadata that completely conform to both metadata schemas.

3. Results

3.1. Metadata crosswalk

3.1.1. Step 1: Definition of a common terminology

Table 1 contains definitions of the terms that we used in this study to develop the new metadata model.

3.1.2. Step 2: Identification of the similarities of INSPIRE and DataCite

When analyzing the formal specification of INSPIRE and DataCite, we identified concepts similar in both standards. Table 2 gives an overview of the similar concepts of both standards.

3.1.3. Step 3: Semantic mapping of elements from INSPIRE and DataCite

When developing semantic mapping, we analyzed each element of DataCite and attempted to find a semantic equivalent in INSPIRE (Table 3).

3.1.4. Step 4: Mapping of Properties

Obligation level: In INSPIRE, the three different obligation levels (1) mandatory, (2) conditional and (3) optional are defined. Conditional elements in INSPIRE are either mandatory or optional, depending on the resource type of the element. DataCite distinguishes three levels: (1) mandatory, (2) recommended and (3) optional. Recommended elements are optional but should be provided to improve interoperability (DataCite, 2016). In the new metadata model, only two obligation levels (1) mandatory and (2) optional are allowed. As part of the harmonization process, a mapping of INSPIRE and DataCite obligation levels to the allowed levels of the new metadata model was developed (Fig. 3). As the conditional elements of INSPIRE are either mandatory or optional, the

Table 1
Definition of terms that were used in this study.

Term	Description
ID	Unique numerical identifier
Element	Name of a metadata element. In INSPIRE, elements are referred to as 'Metadata element name'. In DataCite, elements are called 'Properties'.
Definition	Semantic definition or description of a metadata element. Information about the metadata element and its possible content.
Obligation	Level of obligation of the metadata element. (1) Mandatory: This element is required. (2) Optional: The element is optional. In INSPIRE, elements can be 'Conditional', which means that the element is either mandatory or optional, depending on the resource type of the element. DataCite also distinguishes the 'Recommended' level, in which elements are optional, but should be provided for interoperability.
Multiplicity	Indicates the quantity constraints for the metadata elements. In INSPIRE, the multiplicity of an element is named 'INSPIRE multiplicity'. In DataCite, the multiplicity is called '0 cc'.
Data type	Constraints on the value of an element caused by the data type.
Domain	Element constraints. Specification of the allowed values for the metadata element.

Table 2
Overview of similarities of INSPIRE and DataCite 4.0

INSPIRE	DataCite	Definition
Reference	ID	Unique identifier for each metadata element
Metadata element name	DataCite-Property	Name of a metadata element
Definition	Definition	Description/Definition
INSPIRE Obligation	Obligation	Information about whether a metadata element is mandatory, recommended or optional
Multiplicity	Occ	Information about whether a metadata element may occur multiple times
Data type + Domain	Allowed values, examples, other constraints	Constraints imposed on the value of elements, such as the data type or controlled list values

Table 3
Crosswalk of DataCite and INSPIRE elements. As the INSPIRE standard is more complex and contains more metadata elements, this table focuses on the main elements of DataCite and how they have been mapped to INSPIRE elements. If there was no semantic equivalent in INSPIRE, this element was left empty.

DataCite Element	INSPIRE Element
Title	Resource title
Creator	Responsible organization (responsible party role: author)
Publisher	Responsible organization (responsible party role: publisher)
Identifier	Unique resource identifier
Resource Type	Resource Type
Publication Year	Date of publication
Date (dateType: Updated)	Date of last revision
Date (dateType: Created)	Date of creation
GeoLocation	Geographic bounding box
Description	Resource abstract
Related Identifier	
Alternate Identifier	
Subject	Keyword value
Contributor	Responsible organization
Language	Resource language
Size	
Format	
Version	
Funding Reference	
Rights	Limitations on public access (subelement 'other Constraints')

developed mapping considers only the two mandatory and optional levels for INSPIRE elements.

Multiplicity: Multiplicity indicates the quantity of a metadata element. Both INSPIRE and DataCite distinguish four similar multiplicity values with identical meanings. Within the BonaRes metadata schema, the following four multiplicity values are defined.

- [0..n]: Elements are optional and are repeatable
- [0..1]: Elements are optional but not repeatable
- [1..n]: Elements are required and are repeatable
- [1]: Elements are required but are not repeatable

Fig. 4 gives an overview of the multiplicity values of each metadata standard and how they have been mapped to the BonaRes metadata schema multiplicity.

3.1.5. Step 5: Content conversion of elements

Some metadata elements are restricted to a list of controlled values. For elements whose contents are restricted to controlled lists in both standards, a mapping of the allowed values in both lists was created. First, semantically equivalent values of the allowed code lists of INSPIRE

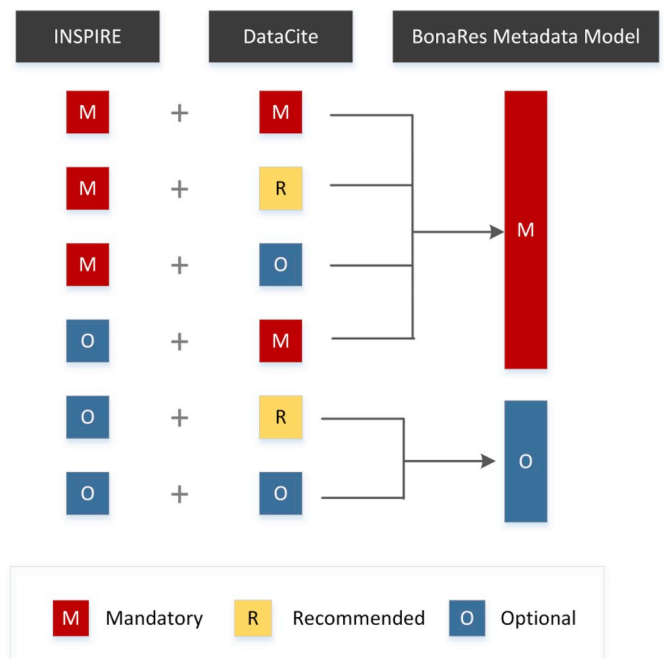


Fig. 3. Mapping of the INSPIRE and DataCite obligational levels.

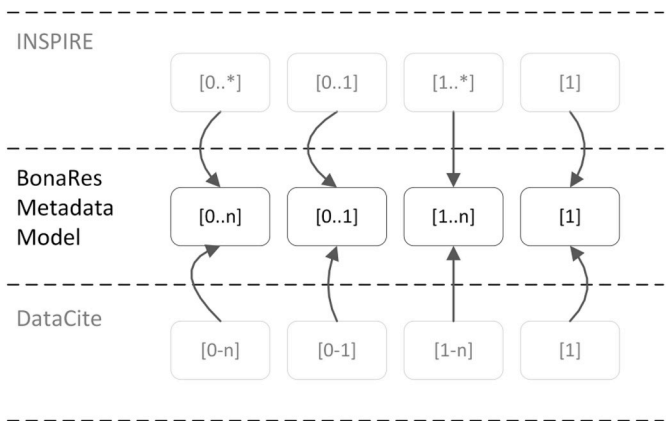


Fig. 4. Mapping of the multiplicity values of INSPIRE and DataCite to the BonaRes metadata schema.

or DataCite were identified and mapped. If the code list values from both standards could not be mapped, they were integrated additionally into the list of allowed values of the BonaRes metadata schema. For example, both standards provide different date elements, e.g., ‘Date of publication’, ‘Date of creation’ or ‘Date of last revision’ from INSPIRE and ‘Date’ with different ‘dateType’ values from DataCite. In both schemas, the date type is distinguished by an attribute that is constrained to the values in a controlled list. Table 4 shows the developed mapping for the code list values of the elements ‘CI_DateTypeCode’ (INSPIRE) and ‘dateType’ (DataCite).

For elements whose contents were specified as free text in one schema but were restricted to a controlled list in the other schema, the characteristics of the more restrictive standard were adopted in the BonaRes metadata schema.

3.2. The BonaRes metadata schema

3.2.1. General description

Fig. 5 provides a schematic overview of the composition of metadata

Table 4
Mapping of attribute values of date types from INSPIRE and DataCite and the resulting date type in the BonaRes metadata schema.

INSPIRE	DataCite	BonaRes metadata schema	
@CI_DateTypeCode	@dateType	@Date types	Definition
creation	Created	Created	The date the resource was created.
publication	Available	Available	The date the resource will be publicly available.
revision	Updated	Updated	The date when the resource was last updated or revised.
	Accepted	Accepted	The date the resource was accepted by the publisher into their system.
	Collected	Collected	The date in which the dataset content was collected.
	Copyrighted	Copyrighted	The date the resource receives copyright status.
	Issued	Issued	The date the resource is published or distributed in the publisher's data infrastructure.
	Submitted	Submitted	The date the author submits the resource to the publisher.
	Valid	Valid	The date during which the resource is accurate.

elements of the BonaRes metadata schema that originate either from the INSPIRE or the DataCite standard or have been newly specified. The newly developed metadata model consists of 22 mandatory elements and 19 optional main elements (Fig. 6). Detailed information regarding the main elements and their respective subelements can be found in (Gärtner et al., 2017). The complete mapping from the BonaRes metadata schema to INSPIRE or DataCite can be found in (Specka et al., 2019).

3.2.2. Integration of INSPIRE and DataCite specifications

After creating the metadata mapping from INSPIRE and DataCite, the BonaRes metadata schema was developed. First, all elements from both standards that could be successfully mapped were integrated into the BonaRes metadata schema. Regarding the mapping of properties and the allowed content values of the mapped elements, we adopted the most restrictive property content values for the mapped elements in the BonaRes metadata schema. Afterwards, the remaining elements from both standards that did not have an equivalent in the respective other standard were integrated with the originating properties and content values.

3.2.3. Specification of additional elements

New metadata elements (Table 5) were defined to meet additional requirements, i.e., to support a bilingual data description (English and German) and to include important scientific descriptions for research data, which were both not fulfilled by INSPIRE or DataCite elements. Two metadata elements were specified, which were to be used to specifically to provide a German translation of the title and summary. In addition to the INSPIRE constraint to provide at least one keyword based on the GEMET thesaurus (General Multilingual Environmental Thesaurus, EIONET, 2018), additional keywords must be specified based on the AGROVOC thesaurus (FAO, 2018; Rajbhandari and Keizer, 2012).

Three new metadata elements (metadata identifier, metadata standard, and metadata character set) add detailed information about the metadata itself. The ‘category’ element was added to classify the datasets, e.g., organizational units, research projects or other structural units. As funding references are an important meta-information in the research community, this element was made mandatory in the BonaRes metadata schema, whereas in the originating schema from DataCite, it was optional.

Soil-agricultural research data are mostly provided in tables in which the columns include values of a specific measurement parameter, collected by a specific method and provided in a defined unit. Two new metadata elements (data model thumbnail and data model attributes)

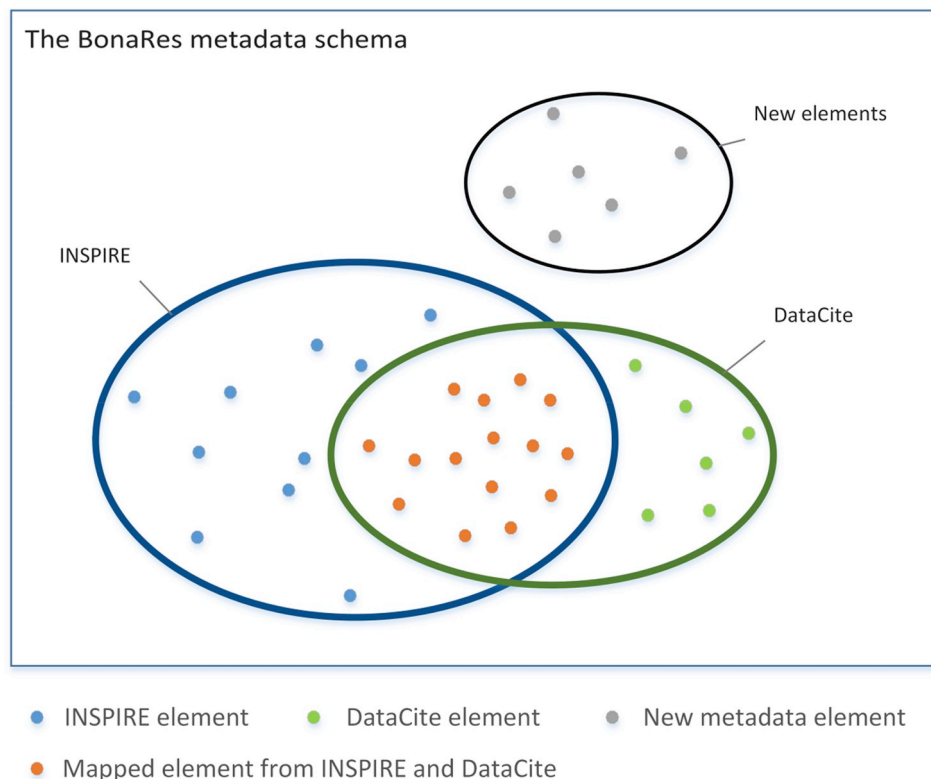


Fig. 5. Schematic concept of the newly created BonaRes metadata schema by different types of metadata elements.

BonaRes Metadata Model 1.0						
Title	German title	Summary	German summary	Dates (available, created, updated)	Responsible party (author, publisher, projectLeader, pointOfContact)	Resource type
Metadata language	Alternate title	Alternate summary	Graphic overview	Resource language	Resource character set	Resource identifier
Metadata date	Temporal extent	Supplemental information	Reference system	Related identifier	Alternate identifier	Funding reference
Metadata identifier	Online resource	Equivalent scale	Distance	Format	Size	Geographic bounding box
Other constraints	Metadata standard	Metadata charset	Data model thumbnail	Data model attributes		Keywords
Use limitations	Access constraints	Use constraints	Lineage statement	Conformity result	Categories	Topic category

Mandatory
 Optional

Fig. 6. Elements of the BonaRes metadata schema. Detailed information on the metadata elements can be found in Gärtner et al. (2017).

were used to describe the underlying data model of the research dataset and the possible relationships to other datasets. The 'data model attributes' element is further detailed by 14 subelements (element IDs 43.0–43.10.3, Table 5) that provide important scientific background information about measured parameters in the respective columns, e.g., yield or soil organic carbon (SOC), in data tables. These are descriptions of the measured parameters, data types and units in which the measured parameter is provided, the applied method for data collection and quality measures. Three further subelements can be used to describe possible relationships between different datasets or tables, i.e., similar to relationships that occur between tables in a relational database.

4. Discussion

A new SDI, the BonaRes Repository (BonaRes Data Centre, 2018; Hoffmann et al., 2018b), was developed with a special focus on storing and describing geospatial soil-agricultural research data in Germany. Soil-agricultural research data is highly diverse both in format and content, e.g., table information, spatial maps, time series and gene sequence data. First, a survey of existing metadata schemes was performed to identify possible metadata schemes that a) could be used to describe diverse soil-agricultural research data and b) could fulfill the two main requirements of being compliant with INSPIRE and allow the assignment of a DOI. Currently, there are many metadata schemes available that are compliant with INSPIRE or ISO 19115 and can be used for the description of geospatial data. These schemas often define additional metadata elements to describe data of a specific disciplines, e.g., SoilML (ISO 28258) for the exchange of soil-related data, GeoSciML (Open Geospatial Consortium Inc., 2017) or INSPIRE geology (INSPIRE Thematic Working Group Geology, 2013) for describing geological data or INSPIRE Soil (INSPIRE Thematic Working Group Soil, 2013) for describing upper parts of the earth's crust. On one hand, these schemas are profiles of ISO 19115 and are compliant with INSPIRE, but on the other hand, either their area of application is too specific to describe

soil-agricultural research data or they are not compliant with DataCite.

One objective when developing a new model was to be able to derive metadata that are compliant with the INSPIRE or DataCite metadata schema. A straightforward solution would have been to transfer all elements and their properties from both standards, 1-to-1, into a new model. However, this would have led to a very complex model with a high number of metadata elements in which much information would have been redundantly requested (Chan and Zeng, 2006; Chen, 2015). For this reason, we applied and adopted a technique for metadata crosswalks to combine the elements of INSPIRE and DataCite into the new model (Chan and Zeng, 2006). First, a mapping between INSPIRE and DataCite metadata elements was realized to consider the element properties as well as their data types or domain values. The mapping reduced the number of metadata elements in the new model that originate from INSPIRE or DataCite, compared to the total sum of elements of both standards, which, in turn, facilitates the creation and editing of metadata.

When merging the properties of the mapped INSPIRE and DataCite elements, we always used the most restrictive property of both standards and applied it to the derived element in the new model (St Pierre and LaPlant, 1998). With regard to the obligational level, this led to a high number of mandatory elements derived from INSPIRE or DataCite in the new model (17) compared to the number of mandatory elements in INSPIRE (16) or in DataCite (6). The same effect occurred when merging the allowed content values of semantically equivalent elements. If the content of the INSPIRE or DataCite elements was restricted to different code list values, we did merge the allowed content values of both standards. If not all code list values from INSPIRE and DataCite could be mapped, the remaining values, either from INSPIRE or DataCite, were inserted into the allowed code list values of the element in the new model. As a result, the number of values in such code lists of the new model was higher than the respective number of values from either INSPIRE or DataCite. However, by merging code list values we reduced number of available values to a possible minimum which in turn reduces

Table 5

Overview of new metadata elements and their subelements in the BonaRes metadata schema. Additional information can be found in [Gärtner et al. \(2017\)](#).

ID	Element	Definition	Obligation	Multiplicity	Data type
1.1	German title	German translation of the title	Mandatory	[1]	Free text
3.1	German summary	German translation of the summary.	Mandatory	[1]	Free text
17	Supplemental information	Supplemental information such as field notes, images or code.	Optional	[0..1]	Link to a compressed file (.zip)
14	Resource character set	The character encoding used in the dataset.	Optional	[0..1]	Controlled List Value
31	Distance	The level of detail of the dataset expressed as ground sampling distance. Mandatory if a resolution distance can be specified (e.g., 0.25 [degrees])	Optional	[0..1]	Numeric
	Metadata identifier	Unique identified for this metadata set.	Mandatory	[1]	Free text
	Metadata standard	The name of the metadata standard used.	Optional	[0..1]	Free text
	Metadata character set	The character encoding used in the dataset.	Optional	[0..1]	Free text
42	Data model thumbnail	A graphical representation, e.g., an entity-relationship diagram of the provided data resources.	Optional	[0..1]	Image
43	Data model attributes	Description of the attribute values e.g., columns included in the dataset.	Mandatory	[1..n]	
43.0	@Table name	Name of the table, the column belongs to.	Mandatory	[1]	Free text
43.1	@Column name	Name of the column.	Mandatory	[1]	Free text
43.2	@Long name	Column name without abbreviations.	Optional	[0..1]	Free text
43.3	@Description	Detailed description of the column data.	Mandatory	[1]	Free text
43.4	@Unit	Unit of the data, ideally an SI-Unit.	Optional	[0..1]	Free text
43.5	@Methods	If standard methods were applied, the standard used should be specified here. Provide information about used instruments.	Optional	[0..1]	Free text
43.6	@Quality	Measures taken to check the quality of the data.	Optional	[0..1]	Free text
43.7	@Data type	A classification of data which tells how to use the data. The type defines the operations that can be done on the data, the meaning of the data, and the way values of that type can be stored.	Mandatory	[1]	Free text
43.8	@Domain	Refers to all the values which a data element may contain.	Optional	[0..1]	Free text
43.9	@Missing value	Codes for missing values, no specification, e.g., 'NA' or 'NULL'	Mandatory	[1]	Free text
43.10	@Foreign Key	Table relationships containing primary key and foreign key constraints. A foreign key is a key used to link two tables together. A foreign key in a table points to a primary key in another table.	Optional	[0..n]	
43.10.1	@@Identifier	A file Identifier of the metadata to which this metadata is a subset (child). (e.g., metadata identifier)	Optional	[0..1]	Free text
43.10.2	@@Table name	The table name to which this table is relating to.	Optional	[0..1]	Free text
43.10.3	@@Column name	The column name of the foreign key.	Optional	[0..1]	Free text

the input effort for our users in the metadata description.

[St Pierre and LaPlant, \(1998\)](#) point out that metadata values may be lost when converting metadata records from a detailed metadata schema into a simpler one. Because the BonaRes metadata schema integrates both standards, the new schema is more extensive than those of either INSPIRE or DataCite. Information that is based on an element that originates from one schema may be lost if the metadata is derived from the other one. For example, the BonaRes metadata schema allows the specification of multiple instances of the 'Title' element to be distinguished by different title types that originate from the DataCite element 'Title' and its subelement 'title Type'. INSPIRE, on the other hand, allows the specification of only one instance of 'Resource title'. If multiple titles are defined in the BonaRes metadata schema, only one instance of title information can be mapped to the INSPIRE element, while other information will be lost. Thus, the crosswalk describing how the BonaRes metadata schema can be mapped to INSPIRE must include information about which title information will be used to map the 'Resource title' element.

As the realized mapping of INSPIRE and DataCite requires enormous intellectual effort, an automated procedure is not possible. Additionally, it requires some effort to be compliant with future versions of both standards ([St Pierre and LaPlant, 1998](#)). The new metadata model needs to be updated whenever the originating metadata models of INSPIRE or DataCite are updated. If there are changes in the original metadata such as the addition or removal of new elements or a change of properties or content values, the realized mapping between INSPIRE and DataCite must also be modified; subsequently, all the affected elements in the BonaRes metadata schema must also be updated.

In addition to the combination of INSPIRE and DataCite elements, the strengths of the BonaRes metadata schema are the new elements that

focus on the description of soil-agricultural research data. Two mandatory metadata elements have been specified to provide specifically German translations for the title and description of the datasets. This increases the findability of research data for German researchers. With the integration of AGROVOC keywords, in addition to the GEMET keywords as required by INSPIRE, datasets can be described with more subject-specific keywords, which, in turn, improves professional data discovery and retrieval. Additional elements can be used to describe the underlying data model and its relationships to other datasets. By providing information on the data model in such a structured way as in the BonaRes metadata schema, the findability and, consequently, reusability of research data can be improved. Data model elements and their respective subelements can interlink columns or attributes from different tables by describing relationships between different datasets, such as in a relational data model, which in turn facilitates the reuse of complex datasets from other researchers.

Compared with other metadata schemes for geospatial data such as GeoSciML ([Open Geospatial Consortium Inc., 2017](#)) or INSPIRE Geology, the BonaRes metadata schema is a more generic one. The additional data model elements provide special support for the description of table or attribute data without, however, being restricted to a specific discipline such as soil or geology. When exchanging metadata between different infrastructures via metadata harvesting by OGC CSW ([Open Geospatial Consortium Inc., 2007a](#)), information specified by the newly added metadata elements will not be exchanged as the information exchange is limited to INSPIRE/ISO 19139 elements. Infrastructures will benefit from the additions to the metadata schema only when visualizing information of the further elements in their user interface as well as integrating the information into the search tools. However, the new metadata schema bridges the gaps that arise when research data has to

be found (INSPIRE incentives), cited (DataCite) and reused (incentives of the BonaRes metadata schema). It is thus a significant and crucial component to fulfill the FAIR principles for soil-agricultural research data.

4.1. Recent development and outlook

Until recently, it was difficult for soil and agricultural scientists to store, describe and provide their specific research data in a standardized, findable, and citable format, and thus to make it accessible and reusable to other scientists. When the BonaRes Repository was established in 2017, the BonaRes metadata scheme was developed and implemented in the infrastructure as a mandatory step to describe the research data. While until 2018 primarily BonaRes project data were managed in the BonaRes infrastructure, increasingly, research data from external data owners are uploaded and described using the new metadata schema. Due to the number of mandatory elements, the description of soil-agricultural research data with the BonaRes metadata schema requires some effort for the researchers. However, researchers, who store research data in the BonaRes Repository, can benefit significantly and increase their scientific output by creating data publications of datasets with the DOI stored in the BonaRes Repository. Researchers can gain more credit for their work if their data is reused by others and is cited accordingly via the DOI. As of May 2019, 18 research datasets managed in the BonaRes Data Portal have been assigned a DOI. By supporting the INSPIRE scheme, other SDIs applying the BonaRes metadata schema and implementing the OGC-CSW interface ([Open Geospatial Consortium Inc., 2007a](#)) can exchange metadata with other INSPIRE-compliant SDIs. Currently, the BonaRes Repository harvests metadata from two national INSPIRE compliant data infrastructures, the Geodata Catalog Produktcenter ([Federal Institute for Geosciences and Natural Resources, 2019](#)) and the Edaphobase ([Burkhardt et al., 2014](#)). Furthermore, we are currently working on transferring the metadata from the BonaRes Repository into the metadata catalogue of the Spatial Data Infrastructure Germany (GDI-DE, [German Federal Government \(2019\)](#)) via OGC CSW. After developing the BonaRes metadata schema and implementing it in the BonaRes Repository, the new DataCite version 4.1 was published ([DataCite, 2017](#)). For this reason, one of the next steps will be to update the BonaRes metadata schema according to the specifications of the new DataCite standard.

5. Conclusions

A new metadata model for soil-agricultural research data was developed that is compliant with the INSPIRE and DataCite metadata schemas and meets the requirements of modern research data management which are described as FAIR principles ([Wilkinson et al., 2016](#)). By developing a mapping between the INSPIRE and DataCite metadata models, we demonstrated that it is possible and practicable to combine models of different disciplines with different applications. The mapping helped us to identify the minimum number of necessary metadata elements that needed to be integrated in the BonaRes metadata schema to support both standards. Additional metadata elements specified in the new metadata model support the targeted search for specific parameters, such as soil pH or organic carbon values, in research data, and improve the retrievability and usability of research data. The BonaRes metadata scheme has proven itself in practice. It is planned to further extend the metadata elements of the data model to open the schema to a larger variety of agricultural-related data, which fits to the BonaRes Repository. The information of the BonaRes metadata schema ([Gärtner et al., 2017](#)) is freely available and operators from other SDIs with similar research data and from other national data infrastructures are able to adapt the schema or to transfer parts of it to the demands of their infrastructures.

Author contributions

Xenia Specka was responsible for preparing and writing the draft, choosing the methodology and visualization of results. **Philipp Gärtner** contributed to analyzing existing metadata schemas, developing the mapping and editing the manuscript draft. **Carsten Hoffmann** contributed to developing the mapping and editing the manuscript draft. **Nikolai Svoboda** contributed to developing the mapping and editing the manuscript draft. **Markus Stecker** was mainly involved in developing the metadata mapping. **Udo Einspanier** was mainly involved in developing the metadata mapping. **Kristian Senkler** was mainly involved in developing the metadata mapping. **Muqit Zoarder** was involved in the first analysis of the metadata schemas. **Uwe Heinrich** contributed by developing the concept of the BonaRes metadata schema and supervising the metadata mapping.

Declarations of interests

None.

Computer code availability

No code was being developed as part of this research.

Acknowledgements

This work was funded by the German Federal Ministry of Education and Research (BMBF) in the framework of the funding measure “Soil as a Sustainable Resource for the Bioeconomy – BonaRes”, project “BonaRes (Module B): BonaRes Centre for Soil Research, subproject B” (grant 031B0511B).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cageo.2019.07.005>.

References

- BonaRes Data Centre, 2018. In: The BonaRes Data Portal. BonaRes Centre for Soil Research. Available from: <https://maps.bonares.de/mapapps/resources/apps/bonares/index.html?lang=en>.
- Brase, J., 2009. European initiative to facilitate access to research data. D-Lib Mag. 15. <https://dx.doi.org/10.1045/may2009-inbrief>.
- Burkhardt, U., Russell, D.J., Decker, P., Döhler, M., Höfer, H., Lesch, S., Rick, S., Römbke, J., Trog, C., Vorwald, J., Wurst, E., Xylander, W.E.R., 2014. The Edaphobase project of GBIF-Germany—a new online soil-zoological data warehouse. Appl. Soil Ecol. 83, 3–12. <https://doi.org/10.1016/j.apsoil.2014.03.021>.
- Chan, L.M., Zeng, M.L., 2006. Metadata interoperability and standardization - a study of methodology part I: achieving interoperability at the schema level. D-Lib Mag. 12, 23–41. <https://dx.doi.org/10.1111/j.1468-0084.2006.00151.x>.
- Chavan, V., Penev, L., 2011. The data paper: a mechanism to incentivize data publishing in biodiversity science. BMC Bioinf. 12, S2. <https://doi.org/10.1186/1471-2105-12-S15-S2>.
- Chen, Y.N., 2015. A RDF-based approach to metadata crosswalk for semantic interoperability at the data element level. Libr. Hi Tech 33, 175–194. <https://dx.doi.org/10.1108/LHT-08-2014-0078>.
- Christian, O., Freddy, N., Ronald Vargas, R., 2013. State of the Art Report on Global and Regional Soil Information: where Are We? where to Go? Food and Agriculture Organization of the United Nations, Rome.
- DataCite, 2016. DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. DataCite Metadata Working Group. Version 4.0. <http://doi.org/10.5438/0012>.
- DataCite, 2017. DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. DataCite Metadata Working Group. Version 4.1. <http://doi.org/10.5438/0014>.
- EIONET, 2018. About GEMET - General Multilingual Environmental Thesaurus. European Environment Information and Observation Network. <https://www.eionet.europa.eu/gemet/en/about/>.
- EU Marine Science and Technology, 2019. European Directory of Marine Environmental Data (EDMED). <https://www.bodc.ac.uk/resources/inventories/edmed/>.
- European Commission Joint Research Centre, 2007. INSPIRE Metadata Implementing Rules: Technical Guidelines Based on EN ISO 19115 and EN ISO 19119, p. 75.

- https://inspire.ec.europa.eu/documents/Metadata/INSPIRE_MD_IR_and_ISO_v1_2_20100616.pdf.
- FAO, 2018. AGROVOC. FAO. <http://aims.fao.org/vest-registry/vocabularies/agrovoc>.
- Federal Institute for Geosciences and Natural Resources, 2019. Geodata Catalog Produktcenter. <https://produktcenter.bgr.de/terraCatalog/Start.do>.
- FGDC, 1998. FGDC-STD-001-1998: Content Standard for Digital Geospatial Metadata. Federal Geographic Data Committee, Washington, DC.
- Gärtner, P., Svoboda, N., Kühnert, T., Zoarner, M.M.A., Heinrich, U., 2017. The BonaRes metadata schema. In: BonaRes Series. BonaRes Centre of Soil Research. <https://doi.org/10.20387/BonaRes-5PGG-8YRP>.
- German Federal Government, 2019. Spatial Data Infrastructure Germany (GDI-DE). <http://www.geoportal.de/EN/GDI-DE/gdi-de.html?lang=en>.
- Hoffmann, C., Schulz, S., Eberhardt, E., Grosse, M., Russell, D.J., Kühnert, T., Stein, S., Zoarner, M.A.M., Specka, X., Svoboda, N., Heinrich, U., 2018. Overview of relevant standards for the BonaRes-Program. In: BonaRes Series. BonaRes Centre of Soil Research.
- Hoffmann, C., Svoboda, N., Stein, S., Eberhardt, E., Schulz, S., Russell, D., Heinrich, U., 2018. The BonaRes Data Infrastructure: Easy Access to Soil Research Data. Geophysical Research Abstracts, Vienna. <https://meetingorganizer.copernicus.org/EGU2018/EGU2018-8323-1.pdf>.
- INSPIRE Thematic Working Group Geology, 2013. D2.8.II.4 INSPIRE Data Specification on Geology – Technical Guidelines.
- INSPIRE Thematic Working Group Soil, 2013. D2.8.III.3 INSPIRE Data Specification on Soil – Draft Guidelines. European Commission Joint Research Centre.
- Laxton, J.L., Duffy, T.R., 2011. Developing and Implementing International Geoscience Standards - A Domestic Perspective. In: Special Paper of the Geological Society of America. Geological Society of America.
- Masser, I., 2005. GIS Worlds - Creating Spatial Data Infrastructures. ESRI Press, Redlands, CA, USA.
- Neumann, J., Brase, J., 2014. DataCite and DOI names for research data. Journal of Computer-Aided Molecular Design 28, 1035–1041. <https://dx.doi.org/10.1007/s10822-014-9776-5>.
- Open Geospatial Consortium Inc, 2007. OpenGIS® Catalogue Services Specification. Open Geospatial Consortium Inc.
- Open Geospatial Consortium Inc, 2007. OpenGIS® Catalogue Services Specification 2.0.2 - ISO Metadata Application Profile.
- Open Geospatial Consortium Inc, 2017. OGC Geoscience Markup Language 4.1 (GeoSciML). <https://docs.opengeospatial.org/is/16-008/16-008.html>.
- Petritsch, B., 2017. Metadata for research data in practice. Mitt. Ver. Österreichischer Bibliothekarinnen und Bibliothekare 70, 200–207.
- Rajbhandari, S., Keizer, J., 2012. The AGROVOC concept scheme – a walkthrough. Journal of Integrative Agriculture 11, 694–699. [https://doi.org/10.1016/S2095-3119\(12\)60058-6](https://doi.org/10.1016/S2095-3119(12)60058-6).
- RDA Metadata Standards Catalog Working Group, 2019. Metadata Standards Catalog. <https://rdamsc.bath.ac.uk/>.
- RDA Metadata Standards Directory Working Group, 2019. Metadata Directory. <http://rd-alliance.github.io/metadata-directory/>.
- Reznik, T., 2013. Geographic information in the age of the INSPIRE Directive: discovery, download and use for geographical research. Geografie 118, 77–93.
- Simmons, S., 2018. 1.09 - Metadata and Spatial Data Infrastructure A2 - Huang, Bo, Comprehensive Geographic Information Systems. Elsevier, Oxford, pp. 110–124.
- Specka, X., Gärtner, P., Svoboda, N., 2019. Semantic mapping of the BonaRes metadata schema to INSPIRE and DataCite. BonaRes Data Portal. <https://doi.org/10.20387/BonaRes-3E47-R97N>.
- St Pierre, M., LaPlant Jr., W.P., 1998. Issues in Crosswalking Content Metadata Standards, NISO - White Papers. National Information Standards Organization. https://groups.niso.org/publications/white_papers/crosswalk/.
- Stein, S., Eberhardt, E., Grosse, M., Helming, K., Hierold, W., Hoffmann, C., Kühnert, T., Ließ, M., Russell, D.J., Schulz, S., Specka, X., Svoboda, N., Zoarner, M.A.M., Heinrich, U., 2019. Report on Available Soil Data for German Agricultural Areas (Submitted). In: BonaRes Series. BonaRes Centre for Soil Research.
- The European Parliament and the Council of the European Union, 2007. Directive 2007/2/EC of the European parliament and of the Council of 14 March 2007 establishing an infrastructure for spatial information in the European community (INSPIRE). Official Journal of the European Union L-14.
- Trilles, S., Diaz, L., Huerta, J., 2017. Approach to facilitating geospatial data and metadata publication using a standard geoservice. ISPRS Int. J. Geo-Inf. 6, 15. <https://dx.doi.org/10.3390/ijgi6050126>.
- USGIN Standards and Protocols Drafting Team, 2009. USGIN Metadata Profile: use of ISO metadata specifications to describe geoscience information resources. In: Richard, S. M., Grunberg, W. (Eds.), Arizona Geological Survey. http://usgin.github.io/usgin_specs/USGIN_ISO_Metadata.htm.
- Weibel, S., 1997. The Dublin Core: a simple content description model for electronic resources. Bull. Am. Soc. Inf. Sci. 24, 9–11. <http://doi.org/10.1002/bult.70>.
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., t Hoen, P.A.C., Hoof, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.
- Zeng, M.L., Chan, L.M., 2006. Metadata interoperability and standardization - a study of methodology part II: achieving interoperability at the record and repository levels. D-Lib Mag. 12, 43–62. <https://dx.doi.org/10.1045/june2006-zeng>.