Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Bridging Citation and Discovery: A Core Metadata Model for FAIR Research Data

**Michal Med**
CESNET, zájmové sdružení právnických osob
**Ondřej Zamazal**
National Library of Technology
**Petra Černohlávková**
petra.cernohlavkova@techlib.cz

National Library of Technology
**Michaela Capandová**
Masaryk University
**Matej Antol**
Masaryk University

**Article**

**Keywords:**

**Posted Date:** November 17th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11742125/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Bridging Citation and Discovery: A Core Metadata Model for FAIR Research Data

November 14, 2026

Michal Med[1,4], Ondřej Zamazal[2], Petra Černohlávková[2*], Michaela Capandová[3] and Matej Antol[3]

[1] CESNET, Prague, 160 00 Czech Republic; https://ror.org/050dkka69

[2] National Library of Technology, Prague, 160 80 Czech Republic; https://ror.org/028txef36

[3] Institute of Computer Science, Masaryk University, Brno, 628 00 Czech Republic; https://ror.org/02j46qs45

[4] Czech Technical University in Prague, Prague, 160 00 Czech Republic; https://ror.org/03kqpb082

* To whom correspondence should be addressed: petra.cernohlavkova@techlib.cz

**Abstract**

Research data can be combined and reused only if their metadata are understood consistently across research communities. While there is broad consensus on following general FAIR Principles, research communities have specific needs, particularly for publishing and citing datasets. Widely used metadata models fall short in delivery: DataCite supports the citation of research data, and DCAT-AP their discovery in data portals, but neither meets both needs. We present the Czech Core Metadata Model (CCMM), which bridges the gap between DCAT-AP (discoverability) and DataCite (citation), emerged from interoperability standards (i.e. metadata profiles and ontologies), FAIR Principles and the requirements of Czech research communities, and was agreed upon within the EOSC CZ Metadata Working Group. CCMM records can be transformed into metadata conforming to the DataCite Metadata Schema, DCAT-AP and relevant OpenAIRE Guidelines, enabling a single metadata record to support citation, discovery and aggregation across research infrastructures. Domain communities can extend it without breaking conformance. CCMM is the core metadata model of the Czech National Data Infrastructure and it is openly available.

# 1 Introduction

Over the past decade, the FAIR Guiding Principles [1] have become an integral part of research data management. Interoperability, the "I" of FAIR, allows research data to be combined, used and reused across research disciplines and communities. It requires data to be described by a common metadata model, which makes the choice of the model particularly important. Such a model has to balance two needs: a common set of elements shared across disciplines, and the specific requirements of individual research domains.

Numerous metadata standards, schemas and application profiles already exist, each developed for the requirements of a particular community. DataCite[1] [2] serves the citation of research data, while DCAT-AP[2] [3] serves the discovery of data in data portals. A new model is therefore justified only if a clear gap in the existing ones is identified. The aim is not to introduce another standard, but to establish a common metadata foundation that builds on the existing ones, connects research domains and supports interoperability without restricting their diversity.

To be adopted, such a model has to be created and agreed on by the research communities that will use it. In the Czech Republic, the need for a community-driven approach was channelled through the foundation of EOSC CZ [4], the national implementation of the European Open Science Cloud (EOSC) [5]. EOSC CZ brings together more than 560 members from 73 institutions, mainly researchers and research support staff, including librarians [6]. They are organised into eight disciplinary clusters, which broadly follow the ESFRI domains [7], and into thematic working groups addressing cross-disciplinary topics. One of them, the Metadata Working Group, brought together more than 120 experts from various institutions and disciplines [8] (Figure 1). Their consensus on the requirements for research data metadata is the basis of the model presented in this paper.

The resulting Czech Core Metadata Model (CCMM) is a cross-disciplinary application profile based on DataCite and DCAT-AP. It clearly defines how to describe research data, establishes a core common foundation, and simultaneously accounts for domain-specific requirements. It also provides support for the creation of community extensions based on the common core. CCMM does not replace existing standards or domain-specific profiles, but connects them through a shared core. It is the metadata model of the Czech National Data Infrastructure.

[1] DataCite is a global metadata standard focused on the persistent identification and citation of research datasets and other research outputs. Its schema defines a core set of properties that support reliable tracking and bibliographic referencing across repositories and research infrastructures.

[2] DCAT-AP represents the European standard for metadata exchange across public data portals and open data catalogues. Based on the W3C Data Catalog Vocabulary (DCAT) [22], DCAT-AP focuses on cross-domain discoverability, data set distributions and access conditions rather than scholarly citation alone.

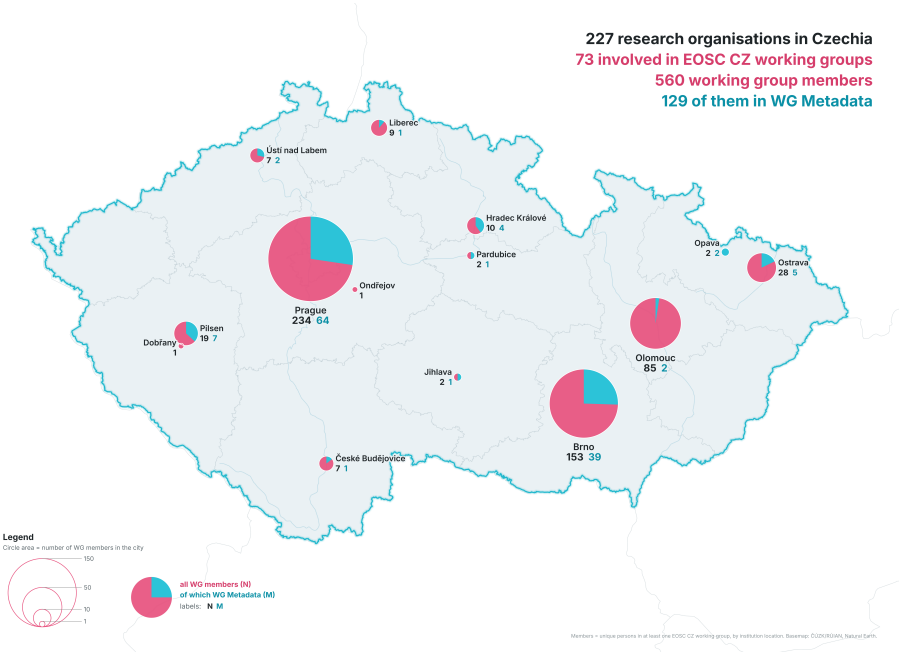



Figure 1: Members of the EOSC CZ working groups by city. The circle area shows the number of working group members in the city (N); the blue segment shows the members of the Metadata Working Group (M). Members are unique persons in at least one of the EOSC CZ working groups, located by their institution. Basemap: ČÚZK/RÚIAN, Natural Earth.

Although CCMM was developed for the Czech research environment, the problems it addresses are shared by research infrastructures and communities all over the world. This paper therefore describes not only the model, but also the method of its development and the requirements and considerations that shaped it, so that it can be reused or adapted elsewhere. The paper follows this development process: we begin by establishing the requirements for the model based on the FAIR Principles and a survey of Czech research communities, and examining how these requirements are addressed by DataCite and DCAT-AP. We then identify the gaps in existing metadata models and use them to formulate the design principles guiding the development of CCMM. On this basis, we develop and present the model. By making the complete model, its specification and technical implementation publicly available, we aim to make the Czech approach transparent, reusable and adaptable in other national and international research environments, seeking to establish a common core metadata model for FAIR research data.

# 2 Requirements and the gap in existing profiles

A metadata model for research data has to address two kinds of demands at once. On the one hand, the ambition to build high-quality, trustworthy data repositories requires the metadata to follow the FAIR Guiding Principles and the policies of EOSC; on the other hand, the model has to serve the researchers who describe, publish and cite their data in everyday scientific practice. This section therefore derives the requirements on the model from both perspectives. Section 2.1 translates the FAIR Principles and EOSC policies into requirements on a metadata model (RF1–RF12). Section 2.2 defines the requirements of Czech research communities (RC1–RC21), based on a questionnaire. The requirements are summarised in Tables 1 and 2, together with their fulfilment by the existing metadata models identified in an earlier analysis [9], namely DataCite and DCAT-AP. Section 2.3 identifies where the two sets of requirements overlap and shows that none of the profiles meets them sufficiently.

## 2.1 Requirements stemming from the FAIR Principles

In order to make research data compliant with the FAIR Principles, basic requirements for each FAIR principle are defined [1, 10]. Moreover, EOSC has adopted most of them into its policies [11]. Table 1 lists only the requirements that FAIR and EOSC place on a metadata model, together with their fulfilment by DataCite and DCAT-AP (Section 2.3).

Table 1: FAIR and EOSC requirements (RF) on the metadata model and their fulfilment by DataCite 4.7 and DCAT-AP 3.0.1. In the Source column, F1–R1.3 denote the FAIR Guiding Principles [1], and EOSC denotes European Open Science Cloud interoperability requirements.

| ID | Requirement on the model | Source | DataCite | DCAT-AP |
|---|---|---|---|---|
| RF1 | A persistent identifier for each data set, with its scheme. | F1 | Yes | Partial |
| RF2 | Rich, structured description beyond citation. | F2, R1 | Partial | Yes |
| RF3 | The metadata record links to the described data by its identifier. | F3 | Yes | Yes |
| RF4 | Records can be harvested into a searchable catalogue. | F4, EOSC | Yes | Yes |
| RF5 | Metadata records can be retrieved through a standard, open protocol. | A1 | Yes | Yes |
| RF6 | Metadata stay available even when the data are no longer available. | A2 | Yes | Partial |
| RF7 | Metadata use a formal, machine-readable language for knowledge representation. | I1 | No | Yes |
| RF8 | Metadata reuse values from shared vocabularies with resolvable IRIs. | I2 | Partial | Yes |
| RF9 | Records use qualified links to related resources, people and organisations. | I3 | Yes | Partial |
| RF10 | Metadata define explicit licence and access conditions for data. | R1.1 | Partial | Yes |
| RF11 | Metadata describe the provenance of the data. | R1.2 | Partial | Partial |
| RF12 | Metadata conform to community standards. | R1.3 | Yes | Partial |

Most of the requirements can be addressed directly by the implementation of the specific metadata attributes in the metadata model with defined data types, proper unambiguous description and methodological guidelines on the usage of the attributes, including recommended or required values from the controlled vocabularies and identifier schemes. Nevertheless, some requirements need to be addressed (at least partly) at the level of infrastructure (RF4, RF5), methodological guidelines (RF1, RF8) and the configuration of system processes (RF3, RF6, RF12).

An important part of the requirements concerns harvesting metadata into the OpenAIRE Graph [12]. A repository has to be registered, expose its records through an OAI-PMH endpoint [13] and pass the check of the metadata profile. Most of this is administrative and technical; the model only has to ensure that its records can be mapped unambiguously to DataCite, following the OpenAIRE Guidelines for Data Archive Managers [14], as further described in Section 4.3.2.

## 2.2 Requirements gathered from Czech research communities

The requirements of the research communities are based on a survey on the use of metadata models in Czech research institutions, carried out by the National Library of Technology (NTK) in 2024 and summarised in the *Evaluation of the questionnaire on needs for dataset description and discovery* [15]. The questionnaire was answered by 82 researchers from 26 institutions. It focused on the metadata attributes used to describe research data, the identifier schemes for persons, organisations and resources, and the required typologies.

The survey showed that respondents use metadata records mainly to publish and cite data alongside research articles. Discovery, access and reuse, which are the goals of EOSC and of making data FAIR, were not among their concerns. Most institutions publish their data in repository systems such as Invenio RDM,[3] CLARIN-DSpace[4] or ASEP-ARL,[5] each with its own internal metadata schema, and the same terms, e.g. publisher or provenance, are understood differently across the community. To make the data interoperable, at least at the national level and in their meaning, a common metadata model has to be set up.

Table 2 lists the resulting requirements, grouped by priority, together with their fulfilment by DataCite and DCAT-AP (Section 2.3). The priority is based on the mean rating of all respondents: high below 2.0, medium from 2.0 to 2.5, and low above 2.5. For agent identifiers (RC4), the priority is high for persons and medium for organisations.

Table 2: Research community requirements (RC) on the metadata model and their fulfilment by DataCite 4.7 and DCAT-AP 3.0.1.

| ID | Requirement | DataCite | DCAT-AP |
|---|---|---|---|
| *High priority* | | | |
| RC1 | Data set PID | Yes | Partial |
| RC2 | Description, documentation | Yes | Yes |
| RC3 | Creators, contributors, roles | Yes | Partial |
| RC4 | Agent PIDs | Yes | Partial |
| RC5 | Contact person | Yes | Yes |
| RC6 | Typed dates | Yes | Partial |
| RC7 | Versioning | Partial | Yes |
| RC8 | Multilingual metadata | Partial | Yes |
| RC9 | Access rights | No | Partial |
| RC10 | Licence | Yes | Yes |
| RC11 | File access, format | Partial | Yes |
| RC12 | Related resources | Yes | Partial |
| *Medium priority* | | | |
| RC13 | Publisher | Partial | Partial |

[3] https://inveniosoftware.org/products/rdm/
[4] https://github.com/ufal/clarin-dspace
[5] https://github.com/LIBCAS/ASEP-ARL

Table 2: (continued)

| ID | Requirement | DataCite | DCAT-AP |
|---|---|---|---|
| RC14 | Rights holder | Yes | Partial |
| RC15 | Subject, classification | Yes | Yes |
| RC16 | Location | Partial | Yes |
| RC17 | File integrity | No | Yes |
| RC18 | Provenance | Partial | Partial |
| *Low priority* | | | |
| RC19 | Title types | Yes | No |
| RC20 | Resource type | Yes | Partial |
| RC21 | Funding | Yes | No |

## 2.3 Shared needs and the gap in DataCite and DCAT-AP

Both sets of requirements share a common core: a persistent identifier for the data set with its scheme (RF1, RC1), typed links to resources, people and organisations by their identifiers (RF9; RC3, RC4, RC12), an explicit licence and access rights (RF10; RC9, RC10) and provenance (RF11, RC18). Beyond this core, the community asks mainly for what citation and the practical handling of data need, such as roles, contact persons and rights holders, both the publishing institution and the repository, versions, multilingual values and file-level information, while the FAIR Principles add machine-actionability, harvesting and persistence (RF3–RF7, RF12). Two priorities need interpretation: agent identifiers (RC4) are required for organisations as well as persons by RF9, and provenance (RC18) is required by RF11, although many respondents were unsure what the term means.

The requirements were evaluated against DataCite 4.7[6] and DCAT-AP 3.0.1, the metadata models recommended for further study by an earlier analysis of international metadata schemas [9]. Other models and profiles, such as Schema.org [16, 17], Dublin Core [18, 19] or OpenAIRE [14, 20], were not evaluated. Each requirement was assessed as fulfilled, partially fulfilled or not fulfilled by the model itself, i.e. without institution-specific extensions or local conventions (Tables 1 and 2).

The two models were designed for different purposes, and their gaps mirror each other. DataCite originates from scholarly communication and covers citation: identifiers, creators and typed contributors, typed dates, typed relations and funding. However, it is defined as an XML Schema without a formal semantic representation (RF7), its controlled values are closed lists without IRIs (RF8), it has no property for access rights (RC9) and it does not describe individual files and their checksums (RC11, RC17). DCAT-AP originates from public open data portals, including the Czech National Open Data Catalogue,[7] and covers discovery and reuse: it is based on DCAT [22, 23] and RDF, with distributions,

[6]The first evaluation was done against DataCite 4.5.

[7]The catalogue uses DCAT-AP-CZ [21], a profile of DCAT-AP for Czech open data.

checksums, access rights, versions and multilingual values. However, it does not require a persistent identifier with its scheme (RF1, RC1), it cannot guarantee that relations to agents and resources are qualified (RF9, RC3, RC4, RC12) and it prescribes no research-specific controlled vocabularies for date types, access rights or resource types (RC6, RC9, RC20). Each profile thus falls short on high-priority requirements that the other meets, and the partial ratings of DCAT-AP fall on the citation requirements that matter most to the respondents.

Still, neither profile sufficiently reflects requirement RC13 on the publisher – both allow only one publisher, while the community wants both the institution and the repository – or requirements RF11 and RC18 on provenance – both offer only free-text or loosely linked provenance. Both profiles could be used in parallel; however, 1) neither of them addresses these issues, and 2) managing two metadata records simultaneously is complex and inefficient. Another option may be to use CiteDCAT-AP [24] – a DataCite profile of DCAT-AP for citation data. There are several reasons not to use it: 1) it does not resolve RC13, RF11 and RC18 either, 2) it was never officially published and 3) it is based on the outdated versions DataCite 4.4 and DCAT-AP 2.0.1.

Since the requirements are not fully met by any existing metadata model, we decided to create a new core metadata model for describing research data – one that fulfils both FAIR and community requirements by leveraging features from existing standards.

# 3 Design principles for building an interoperable metadata model

The fundamental design principles emerged from the requirements and from the analysis of whether existing metadata models could be reused (Section 2). They translate the requirements RF1–RF12 and RC1–RC21 and the gaps identified in DataCite and DCAT-AP into rules that guided the modelling decisions. The principles are divided into three groups: the semantic foundation of the model (P1–P4), the structure of the model (P5–P8), and its use, interoperability and governance (P9–P12). Table 3 traces each principle back to the requirements it addresses.

Table 3: Design principles of the metadata model and the requirements they address.

| ID | Design principle | Addressed requirements |
|---|---|---|
| *Semantic foundation* | | |
| P1 | Reuse before invention | RF12 |
| P2 | Semantics first, serialisation second | RF7 |
| P3 | Identify everything that can be identified | RF1, RF3, RF9, RC1, RC4, RC12 |
| P4 | Controlled values over free text | RF8, RC9, RC15, RC16, RC19, RC20 |

Table 3: (continued)

| ID | Design principle | Addressed requirements |
|---|---|---|
| *Structure of the model* | | |
| P5 | Qualified, typed relations | RF9, RC3, RC5, RC12–RC14 |
| P6 | Separation of concerns | RF6, RF10, RC9–RC11, RC17 |
| P7 | Structured provenance and versioning | RF11, RC7, RC18 |
| P8 | Multilingual by design | RC8 |
| *Use, interoperability and governance* | | |
| P9 | Lossless alignment with both ecosystems | RF4, RF5, RF12 |
| P10 | Minimal mandatory core, rich optional extension | RF2; priority of RC1–RC21 |
| P11 | Unambiguous definitions and guidance | RC13, RC18; ambiguity of terms |
| P12 | Extensible core, versioned profile | RF12 |

**P1 Reuse before invention.** Classes and properties are taken from established vocabularies and ontologies, such as DCAT, Dublin Core Terms, PROV-O, ADMS and DataCite, whenever their meaning fits the requirement. Requirements fulfilled by DataCite are modelled with DataCite terms and those fulfilled by DCAT-AP with DCAT-AP terms. In other cases, a suitable term from another vocabulary or ontology is sought before a new term is defined. The model therefore stays close to community standards (RF12) and its records remain understandable to systems and communities that already know the reused terms.

**P2 Semantics first, serialisation second.** The model is defined at the conceptual level in a formal, machine-readable language for knowledge representation (RF7). The technical representations (XML, JSON and RDF) are derived from the conceptual model, not the other way round. This avoids the limitation of DataCite, whose semantics are bound to an XML Schema and reach RDF only through crosswalks.[8]

**P3 Identify everything that can be identified.** The data set, the persons and organisations involved, the related resources and the data themselves are referenced by persistent identifiers, always together with the identifier scheme (RF1, RF3, RC1, RC4). Schemes required by the community, such as Handle and the Czech company registration number (IČO), are supported alongside DOI, ORCID and ROR. A free-text name is used only as a fallback when no identifier exists, or as a human-readable label.

**P4 Controlled values over free text.** Wherever a typology exists, its values are taken from shared controlled vocabularies identified by resolvable IRIs (RF8),

[8] In June 2026, TIB Hannover published DataCite Linked Data, an RDF representation of DataCite classes, properties and controlled terms.

e.g. COAR resource types and access rights, OECD FORD classification, or a gazetteer for places (RC9, RC15, RC16, RC19, RC20). Vocabularies are represented in SKOS. Where no suitable vocabulary exists, the maintainer of the profile provides one. Free text is kept for genuinely descriptive content, such as titles and abstracts.

**P5 Qualified, typed relations.** Relations of the data set to agents, resources and locations are expressed as qualified relations carrying an explicit type (RF9). Instead of separate elements for creators, contributors and publishers, agents are attached through a single relation bound to the Agent Roles vocabulary. This lets one record contain several creators, publishers and contributors with roles such as contact person and rights holder (RC3, RC5, RC13, RC14). Related resources are linked by their identifiers together with the relation type and the resource type (RC12).

**P6 Separation of concerns.** The model keeps apart entities that have different meaning and lifecycle. First, the data set as an abstract work is separated from its distributions in the form of downloadable files or data services, which carry the access URL, format, size and checksum (RC11, RC17). Second, the licence is separated from the access rights (RF10, RC9, RC10), as one does not imply the other. Third, the metadata record is described as a resource of its own, including information about its manager and original repository, so that it can outlive the data it describes (RF6).

**P7 Structured provenance and versioning.** Provenance and versions are modelled as structured, linked entities rather than as free text (RF11, RC18). A version carries its number and is linked to the previous and following versions (RC7).

**P8 Multilingual by design.** All human-readable values carry a language tag and may be given in several languages (RC8). English is expected for discoverability at the European and international level, while Czech is used for the national context (e.g. the Register of Information on Results, RIV). Unlike DataCite, where only selected properties are language-tagged, multilingualism applies across the whole model.

**P9 Lossless alignment with both ecosystems.** Every record must be transferable without loss of required information to DataCite and OpenAIRE, which serve citation and harvesting into the OpenAIRE Graph, and must be transformable to DCAT-AP, which serves discovery in data catalogues (RF4, RF5, RF12). Consequently, attributes required by OpenAIRE (publication date, resource type and COAR access rights, and description and funding when applicable) are mandatory or derivable in the model, and for every element there is an unambiguous mapping to its counterparts. The model thus meets both

needs identified in Section 2 – citation and accountability on one side, discovery and reuse on the other – within a single record.

**P10 Minimal mandatory core, rich optional extension.** The obligation of each element is set by the priority found among the research communities (Table 2). Requirements with high priority become mandatory or highly recommended, while those with low priority remain optional. This allows heterogeneous repository systems (such as Invenio RDM, CLARIN-DSpace and ASEP-ARL) to join the national infrastructure and, at the same time, supports the rich description required by RF2.

**P11 Unambiguous definitions and guidance.** Each element is a profile of a term from a top-level vocabulary, with a clear definition, the expected values and its representation as an RDF vocabulary term. This preserves the meaning of terms that researchers may understand differently; the survey showed this in particular for publisher (RC13) and provenance (RC18). The definitions are accompanied by methodological guidelines in the form of usage notes, which cover the requirements that cannot be fulfilled by the model alone (RF1, RF8).

**P12 Extensible core, versioned profile.** The model is a core shared across all research disciplines. Domain communities extend it with their specific elements without breaking conformance to the core (Section 4.4). The profile itself is versioned and its changes are documented, so that implementers can follow its evolution and adapt the metadata.

Section 2.3 thus shows that a new metadata model is needed, and this section sets out the design principles for building it. Designing it as an application profile [25] is the most suitable approach, as it ensures interoperability by reusing and combining established terms from existing standards. The principles are built on the gaps that DataCite and DCAT-AP leave in meeting the FAIR and community requirements. The resulting profile, which reuses elements from both models alongside other ontologies, is named the Czech Core Metadata Model (CCMM).

# 4 The Czech Core Metadata Model

This section gives the complete description of CCMM: its conceptual model (Section 4.1), its technical representations (Section 4.2), its alignment with other application profiles (Section 4.3) and its extension capabilities for various research domains (Section 4.4). The description refers to CCMM version 2.0.0, released in October 2026.[9]

[9] All releases with their release notes are available at https://github.com/techlib/CCMM/releases; previous versions are archived at https://github.com/techlib/CCMM-archive.

## 4.1 Conceptual model – the semantics matter

A common metadata structure alone does not make metadata interoperable [26]. A structured (typically XML or JSON) schema guarantees that records from different repositories have the same elements, but not that the receiving system understands each element in the same way as the sender meant it. The survey in Section 2 showed this risk in practice: researchers understand terms such as publisher (RC13) or provenance (RC18) differently, and a shared element name would only hide these differences. The comparison of DataCite and DCAT-AP showed the same problem between models, where the same concept is expressed differently or overlaps only partly. Records that are structurally valid but semantically ambiguous can be exchanged across repositories, but cannot be reliably aggregated, mapped to other models or processed by machines, which is what the FAIR Principles require (RF7–RF9).

CCMM therefore defines semantics first and structure second (P2). The model is specified as a conceptual model of classes, their attributes and relations between classes, independent of any serialisation; the technical representations described in Section 4.2 are derived from it. Each element of the model is a profile of a term from an established vocabulary or ontology (P1), with a definition, expected values and usage note (P11). Relations between the data set and agents, resources and locations are qualified by explicit relation types (P5), and their values are taken from controlled vocabularies identified by IRIs (P4). As a result, the meaning of every element is fixed by a resolvable term, not by its name or position in a record. The unambiguous mappings to DataCite, OpenAIRE and DCAT-AP (P9) described in Section 4.3 rely on this.

### 4.1.1 Model overview

CCMM is built as a profile of DCAT-AP 3.0.1 [3]. As the model describes research data, its central class is `Dataset`. The other classes are profiles of classes from DCAT-AP where applicable, or from other ontologies and vocabularies where more suitable, and are attached to `Dataset` through relationship profiles (Figure 2). To keep `Dataset` at the centre of the model, two relationships point in the opposite direction to their DCAT-AP counterparts: `is described by` from `Dataset` to `Metadata Record`, and `has original repository` from `Metadata Record` to `Repository`.

Most of the objects referenced in a record exist independently of the data set and can be shared by many records. In line with P3 and P6, they are therefore modelled as classes of their own: `Agent` with its subclasses `Person` and `Organization`; `Distribution` with its subclasses `Downloadable file` and `Data service`; `Repository`; `Metadata Record`; `Provenance activity` (P7); `Funding reference`; `Time reference`; `Location`; `Validation result`; and `Resource` with its subclass `Related resource`. Agents are attached to the data set through the qualified class `Resource attribution`, which carries the role of the agent from the Agent Roles vocabulary (P5). Multilingual-

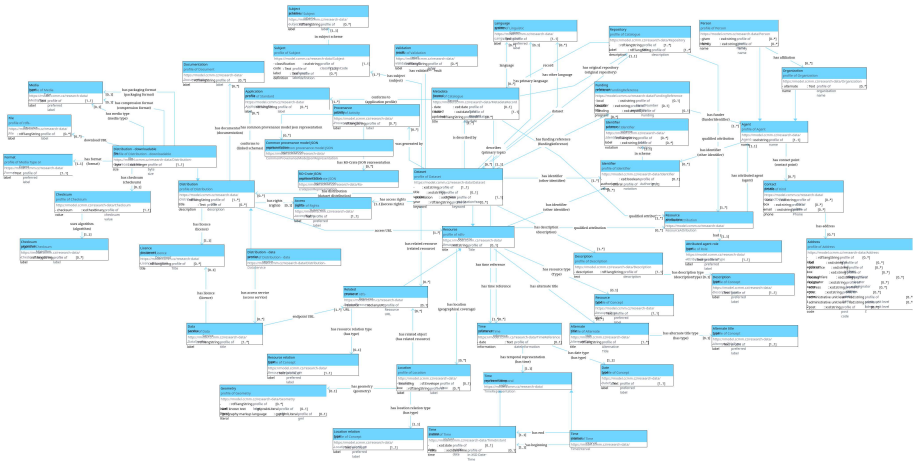


Figure 2: Overview of the CCMM 2.0.0 conceptual model as generated by Dataspecer. Each class is shown with the class it profiles and its attributes; arrows show relationships with their cardinalities. Source: https://github.com/techlib/CCMM.

ism (P8) is supported by the class `Language system`, by language-tagged strings (`rdf:langString`) and by `skos:prefLabel` in controlled vocabularies, which are preferred to free text (P4). The classes and properties are mostly reused from DCAT-AP, DataCite and Dublin Core, and for specific use cases from PROV-O, the Organization Ontology, FOAF, OWL-Time and others (P1). Reusing these terms also keeps the mappings to DataCite, OpenAIRE and DCAT-AP unambiguous (P9). Elements not covered by any existing vocabulary are defined in the CCMM Vocabulary.

The profile was created in Dataspecer [27]. Its conceptual editor was used to profile the classes, relationships and attributes, to set their definitions (P11), cardinalities and usage notes, and to generate the diagram in Figure 2 and a machine-readable description of the profile in RDF, using the Data Specification Vocabulary (DSV) [28] (P2). Its structural editor was then used to generate the technical representations described in Section 4.2. All outputs are published in the GitHub repository `techlib/CCMM` and at https://ccmm.cz/.

In RDF, a CCMM record is not bound to a rigid hierarchical structure, but is represented as a graph that directly reflects the conceptual model (P2). A record uses the IRIs of the reused classes and properties, such as `dcat:Dataset`, `dcterms:title`, `adms:identifier`[10] or `prov:qualifiedAttribution` with `dcat:hadRole`, together with terms of DataCite Linked Data,[11] such as `publicationYear`, and of the CCMM Vocabulary,[12] such as `hasTimeReference`,

[10] ADMS: https://www.w3.org/TR/vocab-adms/
[11] https://tibhannover.github.io/datacite/
[12] https://model.ccmm.cz/vocabulary/ccmm

isDescribedBy or originalRepository (Figure 3). Values from controlled vocabularies are IRIs and labelled with skos:prefLabel (P4). Any RDF consumer that understands DCAT [22], PROV-O or DataCite Linked Data can therefore interpret a CCMM record without knowing CCMM itself, except for the terms of the CCMM Vocabulary (P1, P9, RF7, RF8). Each record states the profile it conforms to by dcterms:conformsTo of its metadata record, with the IRI https://model.ccmm.cz/research-data/2.0.0. The RDF representation can be written in any RDF syntax, such as Turtle, RDF/XML or JSON-LD.[13] The structural representations described in Section 4.2 are mapped to it: JSON records through the JSON-LD context and XML records through the lifting transformation.

```
@prefix dcat:    <http://www.w3.org/ns/dcat#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix adms:    <http://www.w3.org/ns/adms#> .
@prefix prov:    <http://www.w3.org/ns/prov#> .
@prefix org:     <http://www.w3.org/ns/org#> .
@prefix foaf:    <http://xmlns.com/foaf/0.1/> .
@prefix xsd:     <http://www.w3.org/2001/XMLSchema#> .
@prefix dcp:     <https://w3id.org/tib/datacite/property/> .
@prefix ccmm:    <https://model.ccmm.cz/vocabulary/ccmm#> .
@prefix coar:    <http://purl.org/coar/access_right/> .
@prefix role:    <https://vocabs.ccmm.cz/registry/codelist/AgentRole/> .

<urn:uuid:c5f64068-3c77-4598-a9ca-f9fed5cfe4a8> a dcat:Dataset ;
  dcterms:title "Data matrix for Data on Health literacy ..." ;
  adms:identifier <https://doi.org/10.48700/datst.xd12h-dfz24> ;
  dcterms:accessRights coar:c_f1cf ;
  dcp:publicationYear "2026"^^xsd:gYear ;
  prov:qualifiedAttribution [ a prov:Attribution ;
      dcat:hadRole role:Publisher ;
      prov:agent [ a org:Organization ;
                   foaf:name "University of South Bohemia" ] ] ;
  ccmm:isDescribedBy [ a dcat:CatalogRecord ;
      dcterms:conformsTo <https://model.ccmm.cz/research-data/2.0.0> ;
      ccmm:originalRepository <https://data.narodni-repozitar.cz> ] .
```

Figure 3: Shortened, translated and adapted excerpt of the minimal sample CCMM record in Turtle (_metadata-samples/rdf/dataset-mini.ttl in techlib/CCMM); prefixes were renamed for readability.

CCMM is designed as a core profile that is shared by all research fields and can be extended to meet the needs of individual communities (P10, P12). How the core can be extended for specific domains is described in Section 4.4.

### 4.1.2 Employed controlled vocabularies

To enhance semantic interoperability and ensure consistency, values from controlled vocabularies are preferred over free text (P4). Table 4 provides an

[13] Turtle: https://www.w3.org/TR/turtle/; RDF/XML: https://www.w3.org/TR/rdf-syntax-grammar/; JSON-LD: https://www.w3.org/TR/json-ld11/.

overview of the employed controlled vocabularies, grouped by their metadata context, with their registering authorities and the community requirements they serve.[14]

Table 4: Controlled vocabularies and ontologies used in CCMM, their registering authorities and the research community requirements (RC) they serve.

| **Vocabulary or ontology** | **Registering authority** | **Requirements** |
|---|---|---|
| *Descriptive and subject context* | | |
| OECD FORD Classification | OECD | RC15 |
| EU Languages | EU Publications Office | RC8 |
| Description Types | DataCite | RC2 |
| Alternate Title Types | DataCite | RC19 |
| Time Reference Types | DataCite | RC6 |
| *Agents, relations and location* | | |
| Agent Roles | NTK (DataCite) | RC3, RC5, RC13, RC14 |
| Relation Types | DataCite | RC12 |
| Location Relation Types | NTK | RC16 |
| *Resources, formats and access* | | |
| COAR Resource Types | COAR | RC20 |
| COAR Access Rights | COAR | RC9 |
| IANA Media Types | IANA | RC11 |
| EU File Types | EU Publications Office | RC11 |
| SPDX Checksum Algorithm | Linux Foundation (SPDX) | RC17 |

The Agent Roles controlled vocabulary is administered by NTK. Although it is derived from the DataCite Contributor Type controlled vocabulary, the CCMM profile required the inclusion of explicit Publisher and Creator concepts, which are defined as separate top-level elements in the core DataCite schema. Additionally, the Location Relation Types vocabulary is managed directly by NTK, as no existing standardised controlled vocabulary adequately covered spatial relation semantics for data set description.

The usage notes of the CCMM specification state which vocabularies are required. Some properties require at least one specific value or at least one value from a specific vocabulary (COAR Access Rights, OECD FORD, Agent Roles), while others recommend values from a specific vocabulary where applicable (IANA Media Types, COAR Resource Types). Most properties allow communities to use any controlled vocabulary.

Technically, within the Czech National Data Infrastructure, controlled vocabularies are represented using the Simple Knowledge Organization System (SKOS) data model [29]. While several standardised vocabularies are natively available in SKOS (such as the COAR vocabularies), others lack an official SKOS representation. To ensure semantic consistency across the infrastructure, NTK prepared

[14] Direct links and resolvable IRIs for all vocabularies are available in the online registry at https://www.ccmm.cz/en/controlled-vocabularies/.

SKOS representations for vocabularies such as the OECD FORD Classification and IANA Media Types. However, NTK does not act as their registering authority. Furthermore, formal ontologies can also serve as controlled vocabularies within CCMM; a prime example is the adoption of the SPDX ontology for defining checksum algorithms.

#### 4.1.3 Core elements of the model

The elements of the model are grouped by their purpose into seven groups (Table 5). For each group, the following paragraphs explain what its elements mean in CCMM and the design decisions behind them, with reference to the requirements of Section 2 and the design principles of Section 3. Cardinalities refer to the `Dataset` class unless stated otherwise.

Table 5: Groups of CCMM elements and the requirements and design principles they address.

| Group | Main classes and properties | Requirements | Principles |
|---|---|---|---|
| Dataset | `Dataset`; title, alternate title, description, publication year, resource type | RF2, RC2, RC19, RC20 | P1, P4, P10 |
| Identification, agents and relations | `Identifier`; `Resource attribution`; `Agent` (`Person`, `Organization`); `Contact details`; `Related resource` | RF1, RF9, RC1, RC3–RC5, RC12–RC14 | P3, P5, P11 |
| Descriptive context | `Time reference`; `Subject`; keyword; primary and other language; `Location` | RF8, RC6, RC8, RC15, RC16 | P4, P8 |
| Distributions and access | `Downloadable file`; `Data service`; `Checksum`; access rights; `Licence document` | RF3, RF10, RC9–RC11, RC17 | P4, P6 |
| Funding, provenance and quality | `Funding reference`; `Provenance activity`; `Validation result` | RF11, RC18, RC21 | P7, P9 |
| Metadata about metadata | `Metadata Record`; `Repository` | RF6, RF12, RC13 | P6, P9 |
| Versioning and maintenance | version; `Related resource` with version relation types | RC7 | P5, P7 |

**Dataset.** `Dataset` is the central class of the model and the subject of every record. It is a profile of the DCAT-AP `Dataset` and a subclass of `Resource`, and represents the data set as an abstract work, independent of the files in which it is distributed (P6). Its own properties give the minimal description needed for citation and discovery (RF2): a title, at least one description and the publication year, optionally complemented by alternate titles and a resource type (usually *dataset*). Alternate titles carry a type from the DataCite Alternate Title Types,

so that translated titles and subtitles can be told apart (RC19). Descriptions carry a type from the DataCite Description Types, so that an abstract can be distinguished from methods or technical information (RC2). The resource type is taken from the COAR Resource Types (RC20, P4). The title, description and publication year are mandatory, while the resource type is optional, reflecting its low priority in the survey (P10). In addition, every data set must be linked to at least one identifier, resource attribution, time reference and subject, to its access rights and to at least one metadata record; these are described in detail in the following paragraphs.

**Identification, agents and relations.** This group identifies the data set and the agents and resources connected to it (RF1, RF9). A data set has at least one `Identifier`, always stated with its identifier scheme, so that DOI, Handle and other schemes can be distinguished and resolved (P3, RC1). Agents – a `Person` or an `Organization` – are attached through a `Resource attribution` that carries a role from the Agent Roles vocabulary (P5, RC3). At least one attribution is mandatory: the creator role is mandatory, and the publisher role is highly recommended and given preferably to an organisation. In CCMM, the publisher of a data set is always an agent; the repository is not an agent and is therefore linked through the metadata record (see *Metadata about metadata*). This gives the publisher a single, clear meaning (P11) while keeping both the institution and the repository the survey respondents asked for (RC13). Contact person and rights holder are roles as well (RC5, RC14); a contact is reached through `Contact details`, such as e-mail, phone, postal address or Czech data box. Persons are preferably identified by ORCID [30] and organisations by ROR, but any identifier scheme is supported, with a name as a fallback (RC4). Related resources are linked by their identifier, together with a relation type from the DataCite Relation Types and a resource type (RC12, P4).

**Descriptive context.** This group places the data set in time and space and describes its subject and language. Dates are expressed as `Time reference`s, each with a date type from the DataCite Time Reference Types and a temporal representation that is either an instant or an interval (RC6). One element thus covers the dates of creation, publication and modification as well as the period the data cover. At least one time reference, with the type *created*, is mandatory. Subjects are taken from a classification identified by a subject scheme and a code (RC15, P4), and at least one subject from the OECD FORD classification (Frascati Manual), defining the scientific field, is mandatory; free keywords are kept apart from them as plain strings, so that controlled and uncontrolled values do not mix. The primary language and other languages of the data set are taken from the EU Languages vocabulary (RC8, P8). A `Location` can be given by a name, a geometry or a bounding box, or by a link to a related object such as a gazetteer entry, and its relation type states how the place relates to the data set (RC16). Language, keywords and location are optional.

**Distributions and access.** A distribution is a concrete form in which the data set is available, kept apart from the data set itself (P6). A `Downloadable file` has a mandatory access URL, format and byte size, and optionally a download URL, a media type, packaging and compression formats and a `Checksum` with its algorithm (RC11, RC17). Formats are taken from the EU File Types, media types from IANA and checksum algorithms from SPDX (P4). A `Data service` gives access through an endpoint, with the standards it conforms to and its documentation. Every distribution has a mandatory licence (RC10), as different distributions of the same data set may be licensed differently, while the access rights are stated once for the data set and are mandatory, with values from the COAR Access Rights (RC9). Licence and access rights are therefore separate properties (RF10, P6). Distributions themselves are optional, allowing a record to describe data that is not available online. When present, their access URLs link the metadata to the data (RF3).

**Funding, provenance and quality.** A `Funding reference` names at least one funder, which is an agent, and optionally the award title, the grant number and the funding programme (RC21). Its structure follows the DataCite funding reference, so that it can be transferred to DataCite and OpenAIRE (P9). Provenance is described by a `Provenance activity`, a profile of the PROV-O activity, that generated the data set (RF11, RC18, P7). Instead of reproducing the full provenance in the record, the activity links to its machine-readable representation in the Common Provenance Model [31] JSON or RO-Crate [32]. A `Validation result` records the outcome of a quality check of the data set. All three are optional, reflecting the medium or low priority of provenance and funding in the survey (P10).

**Metadata about metadata.** The record itself is described by a `Metadata Record`, a resource separate from the data set (P6). It states the date of the last update (mandatory) and of creation, the language of the record, the agents responsible for it (with the role *data manager*), and the standards it conforms to, such as the version of CCMM (RF12). Each data set has at least one metadata record, and each metadata record has exactly one original repository. A `Repository` represents the storage where the data are deposited. It is a profile of the DCAT-AP `Catalogue`, not an agent: it is identified by an IRI, has a name and a description, and has a mandatory attribution to the organisation that operates it as publisher. Together with the publisher of the data set, this records both the institution and the repository (RC13) without mixing their meaning. As the record is a resource of its own, it can be kept when the data are no longer available (RF6). Harvesting through standard protocols (RF4, RF5) is provided by the infrastructure that publishes the records, not by the model.

**Versioning and maintenance.** The version of the data set is stated as an optional string (RC7). Relations between versions are expressed as related resources with version relation types from the DataCite Relation Types, such as

*is new version of* and *is previous version of* (P5, P7), in line with the principles of data versioning proposed by the Research Data Alliance [33]. The profile itself is versioned, so that each record states the version of CCMM it conforms to (P12). A note on changes, which the community also requested (RC7), is deliberately not part of the model: the metadata describe the data, not the changes between their versions. The version is therefore only a number, and the history of changes is kept in the repository.

## 4.2 Technical representations

The conceptual model defines what the elements of CCMM mean, but not how a record is written down. This is defined by structural models, created in the structural editor of Dataspecer on top of the conceptual model. A structural model selects the profiled classes and properties, sets their order and nesting and gives them technical names; its root is the `Dataset` class. Each class has its own structural model, which can be reused wherever the class occurs, e.g. `Agent` in attributions, funding references and repositories. As all technical representations are generated from the same structural models, they express the same content and differ only in syntax (P2). The RDF representation is not one of them: it instantiates the conceptual model directly (Section 4.1.1), and the structural representations are mapped to it.

Table 6 lists the artefacts generated for CCMM 2.0.0.

Table 6: Technical artefacts of CCMM 2.0.0 generated by Dataspecer and published in the GitHub repository `techlib/CCMM`.

| Artefact | Purpose |
|---|---|
| *For each class* | |
| XML Schema | Structure and validation of XML records |
| JSON Schema | Structure and validation of JSON records |
| JSON-LD context | Maps JSON keys to the IRIs of the profiled terms, so that a JSON record is also an RDF graph |
| Lifting and lowering | Transformation of XML records into their RDF representation and back |
| Example | Example record of the class |
| *For the whole profile* | |
| Profile description | Machine-readable description of the profile in DSV |
| SHACL shapes | Validation of RDF records |
| Specification | Human-readable specification in English and Czech |

The representations serve different users. XML is the primary exchange format: the XML Schema with the root element `Dataset` defines a CCMM record, and the mapping to DataCite and OpenAIRE described in Section 4.3 works on XML. Since version 2.0.0, JSON Schemas and JSON-LD contexts are generated as well, for systems that work with JSON. Thanks to the JSON-LD context, a JSON

record carries the same semantics as the conceptual model (RF7). XML records gain their semantics through the lifting transformation into RDF, generated together with the other technical artefacts. RDF records can be validated against the SHACL shapes.

The GitHub repository also contains sample records in XML and RDF, including a minimal valid record with only the mandatory elements, and the Dataspecer project, which can be imported into another Dataspecer instance as a starting point for a domain profile (Section 4.4). Supporting tools are provided in the separate repository `techlib/CCMM-utils` (see Code Availability).

## 4.3 Alignment with other application profiles

The previous sections defined the requirements coming from both research communities and the FAIR Principles, justified the creation of a new application profile and described its content in full detail. Only a few metadata classes and properties were newly defined, as most of them were profiled from existing vocabularies and ontologies. This section describes how CCMM records can be transferred to other metadata profiles, as the FAIR Principles call for interoperability and reusability.

### 4.3.1 DataCite, OpenAIRE and DCAT-AP

To achieve seamless interoperability across diverse infrastructure layers, CCMM is aligned with three primary metadata frameworks: DataCite to support DOI registration and citation of research data, OpenAIRE to enable metadata aggregation and linking of research outputs, and DCAT-AP to facilitate interoperability with European public-sector data catalogues.

### 4.3.2 How the model is transferable to DataCite and OpenAIRE

**Transferability to DataCite** Ensuring the interoperability and transferability of CCMM to globally adopted metadata schemas such as DataCite requires addressing complex semantic and structural differences. Because the national research infrastructure needs to export CCMM-compliant metadata into a DataCite-compatible format (e.g. to obtain DOIs) and conversely harmonise DataCite records back into the National Metadata Directory to collect metadata from repositories outside the National Data Infrastructure, a bidirectional transformation was developed, as described in [34].

The complete mapping process and its operational realisation are documented and made available in a dedicated GitHub repository `techlib/CCMM-DataCite-mappings` (see Code Availability). The mappings and transformations cover CCMM versions 1.1.0 and 2.0.0 alongside the DataCite Metadata Schema version 4.7 and 4.6.

The transformation resolves differences at three levels: *syntactic*, such as element nesting, attributes and formatting, handled by XPath expressions and XSLT

functions; *structural*, as DataCite has separate elements for creators, contributors and publishers, whereas CCMM attaches all agents through attributions with roles, so that the creator and publisher roles map to the dedicated DataCite elements and other roles to contributors; and *semantic*, as DataCite uses free-text rights, whereas CCMM uses IRIs from controlled vocabularies and separates licences from access rights.

**Transferability to OpenAIRE** While CCMM provides the baseline structural mapping for DataCite, exporting to OpenAIRE requires additional constraints defined by the OpenAIRE Guidelines for Data Archive Managers. Since OpenAIRE harvests metadata via OAI-PMH in DataCite XML format[14], the transformation builds on the CCMM-to-DataCite mapping with three extensions:[15]

- **Mandatory fields:** OpenAIRE mandates attributes that are optional in base DataCite, i.e. the publication date, the resource type and, when applicable, descriptions such as abstracts. As every CCMM record describes a data set, the resource type, which is optional in CCMM, is set to *dataset* when a record does not state it. The transformation workflow incorporates automated validation rules to ensure that no mandatory OpenAIRE attributes are omitted during export.
- **Funding references:** CCMM captures grant funding structures and serialised them into OpenAIRE-compliant `fundingReference` blocks containing normalised funder identifiers (e.g. Crossref Funder ID, ROR) and grant agreement numbers to support automated indexing in the OpenAIRE Graph.
- **Access Rights encoding:** OpenAIRE requires access rights to use the COAR Access Rights vocabulary. Since CCMM natively binds its `accessRights` property to COAR IRIs, the mapping transfers these values directly (e.g. *open access*, *embargoed access*) without requiring heuristic conversion.

Consequently, CCMM records can be seamlessly exported to both DataCite and OpenAIRE schemas from a single source, enabling harvesting by European research infrastructures without manual post-processing.

#### 4.3.3 CCMM as a DCAT-AP profile

CCMM reuses the classes of DCAT-AP and follows its structure and cardinalities, but it is deliberately not a strict profile of DCAT-AP, as it doesn't fully comply with the DCAT AP reuse guidelines [42]. The differences are concentrated in the elements where CCMM replaces a fixed DCAT-AP property with a general, typed element:

[15]While the CCMM-to-DataCite transformation is already available, the CCMM-to-OpenAIRE implementation is currently in progress and is expected to be ready for the camera-ready version of this paper.

- **Description.** DCAT-AP requires a plain `dcterms:description` of the data set, whereas CCMM uses a `Description` with a description type (RC2).
- **Dates.** DCAT-AP uses separate properties for the dates of issue and modification and for the temporal coverage, whereas CCMM expresses all dates as `Time reference`s with a date type (RC6). The publication year is taken from DataCite.
- **Agents.** DCAT-AP uses fixed properties such as `dcterms:publisher`, `dcterms:creator` and `dcat:contactPoint`, whereas CCMM attaches all agents through a `Resource attribution` with a role (RC3, RC5, RC13, RC14).
- **Subjects.** DCAT-AP recommends themes from the EU Data Themes, whereas CCMM requires a subject from a classification identified by a scheme and a code, such as OECD FORD (RC15).
- **Access rights.** DCAT-AP recommends the EU Access Rights vocabulary, whereas CCMM uses the COAR Access Rights (RC9); the specification recommends a mapping between the two.

Some of these properties are mandatory in DCAT-AP, e.g. the description of a data set, or the publisher of a catalogue, which CCMM expresses as an attribution of the repository to an organisation with the publisher role. A CCMM record does not contain these mandatory properties in the form DCAT-AP requires, and therefore it is not, as it stands, a valid DCAT-AP record. It can, however, be made valid by an unambiguous mapping.

Making CCMM a strict profile of DCAT-AP would mean replacing these typed elements with the fixed DCAT-AP properties. This would break the extensibility of the model described in Section 4.4: domain communities could no longer add their own roles, date types or description types as values of controlled vocabularies, but would have to define new properties, and their records would no longer share one structure with the core (P10, P12). CCMM therefore keeps the typed elements and stays as close to DCAT-AP as this allows. As every typed element states its type explicitly, the corresponding DCAT-AP properties can be derived from a CCMM record, e.g. `dcterms:publisher` from an attribution with the publisher role or `dcterms:issued` from a time reference of the type *issued*. The mapping is unambiguous, as the type of each element determines the DCAT-AP property it maps to. Its result is a valid DCAT-AP record, so that CCMM records can be published in DCAT-AP-based catalogues (P9).

Even without this derivation, a CCMM record can largely be read by systems that understand DCAT-AP, although it does not validate against it. Its classes and many of its properties are taken from DCAT-AP, such as `dcat:Dataset`, `dcterms:title`, `adms:identifier` or the distributions with their access URLs and formats, and they keep their DCAT-AP meaning. Any system that understands DCAT-AP therefore understands these parts of a CCMM record. It only

does not interpret the CCMM specifics, such as the typed time references and resource attributions or the terms of the CCMM Vocabulary, which it can ignore without misreading the rest of the record.

## 4.4 Extending the core for domains

CCMM is shared by all scientific fields, but individual research communities often need more detail to describe their data (P12). The model supports this in two ways. First, roles, resource types, date types, formats and other types are values from controlled vocabularies rather than fixed properties (Section 4.1.1), so a community can add its own values, such as specific contributor roles or data services, without changing the structure of the model. Second, a community can create a domain profile of CCMM, in the same way as CCMM profiles DCAT-AP (P1). The core thus stays simple, while the domain profiles describe the data as their communities understand them best.

A domain profile is built on top of CCMM, not next to it: it takes over all classes, properties, cardinalities and vocabulary bindings of CCMM and adds the classes and properties its community needs. In practice, the community imports the Dataspecer project of CCMM (Section 4.2), adds its own elements and generates the technical representations in the same way as for CCMM. A domain profile should follow these rules:

- **Keep the core.** All elements of CCMM keep their meaning; their constraints can be made stricter, e.g. an optional element can become mandatory, but never relaxed.
- **Extend, do not redefine.** New elements are new classes and properties attached to the classes of CCMM, such as `Dataset` or `Distribution`; no CCMM element is reused with a different meaning.
- **Reuse before invention.** As in CCMM (P1), new elements profile terms from established domain vocabularies and ontologies wherever possible; terms defined by the community itself have IRIs in its own namespace.
- **State the base.** The domain profile states the version of CCMM it builds on and follows new releases of CCMM.

A record of a domain profile is a single record containing both the CCMM core and the domain extension. The community uses all of it, while systems that know only CCMM, such as the National Metadata Directory or the transformations to DataCite, OpenAIRE and DCAT-AP (Section 4.3), process the core and leave the extension uninterpreted. In RDF, this follows from the use of IRIs: a CCMM consumer can ignore the unknown terms of the extension without misreading the rest of the record, similarly to the relation between CCMM and DCAT-AP (Section 4.3.3). The core metadata thus remain understandable in the same way at the national and international levels, the domain community keeps all the metadata it needs, and no separate CCMM record has to be maintained next to the domain record, a drawback of using several profiles in parallel (Section 2.3).

No extension of CCMM has been published yet, but at least two are being drafted, for Czech-BioImaging and for materials technologies.

This is how CCMM fulfils P12: domain communities extend the core with their specific elements, every record of a domain profile remains a valid CCMM record, and the core can evolve in versioned releases that the domain profiles follow.

# 5 Discussion

The goal of CCMM is to answer the need for a common metadata model for research data that follows the FAIR Principles and fulfils the requirements of the research communities for the description of datasets.

## 5.1 How the model addresses FAIR Guiding Principles

CCMM translates the concepts embodied in the FAIR Guiding Principles [1] into practice within the Czech National Data Infrastructure, with a particular emphasis on interoperability. Table 7 revisits the requirements derived from the FAIR Principles and endorsed by EOSC (Section 2.1) and shows how CCMM addresses each of them. Most of the requirements are met by the model itself. RF4 and RF5 depend on the infrastructure that publishes the records, and RF3, RF6 and RF11 are supported by the model, but their fulfilment depends on how repositories use it.

Table 7: How CCMM addresses the FAIR and EOSC requirements (RF).

| ID | Requirement | How CCMM addresses it |
|---|---|---|
| RF1 | Persistent identifier | At least one `Identifier` per data set, always with its identifier scheme |
| RF2 | Rich metadata | Mandatory descriptive core, complemented by optional descriptive context, distributions, funding and provenance |
| RF3 | Link to the data | Access URLs of distributions; distributions are optional |
| RF4 | Harvestable records | Provided by the infrastructure; the model provides the mappings to DataCite and OpenAIRE |
| RF5 | Open retrieval protocol | Provided by the infrastructure |
| RF6 | Metadata outlives data | `Metadata Record` as a resource separate from the data set; retention depends on the infrastructure |
| RF7 | Formal representation | RDF representation of every record; JSON-LD contexts for JSON and lifting transformation for XML records |
| RF8 | Vocabularies with IRIs | Controlled vocabularies represented in SKOS and identified by IRIs |
| RF9 | Qualified references | `Resource attribution` with roles; related resources and locations with relation types |

Table 7: (continued)

| ID | Requirement | How CCMM addresses it |
|---|---|---|
| RF10 | Licence and access | Licence of every distribution; access rights of the data set from the COAR Access Rights |
| RF11 | Provenance | Optional `Provenance activity` linking to a Common Provenance Model or RO-Crate representation |
| RF12 | Community standards | Profile of DCAT-AP reusing DCAT, DataCite and other vocabularies; conformance stated in the metadata record |

The EOSC Interoperability Framework distinguishes technical, semantic, organisational and legal interoperability [11]. CCMM contributes mainly to the first two, on three levels:

1. **Within the national infrastructure.** CCMM provides a common baseline for describing data from heterogeneous institutional and domain repositories, so that their records can be aggregated and searched in the National Metadata Directory [4].
2. **Across infrastructures.** The alignment with DataCite, OpenAIRE and DCAT-AP (Section 4.3) allows CCMM records to be exported to these ecosystems, including the European Open Science Cloud and open data portals.
3. **Semantic interoperability.** Controlled vocabularies represented in SKOS and identified by IRIs (Section 4.1.2) reduce the ambiguity of free text, and the RDF representation (Section 4.1.1) makes the meaning of every element machine-interpretable.

The model alone does not make data FAIR. Persistent identifiers are mandatory for datasets, but optional for agents, which may be identified by name only, and distributions and provenance are optional. The extent to which records are FAIR therefore also depends on the repositories and researchers that create them and on the infrastructure that publishes them.

## 5.2 How the model addresses community requirements

CCMM was created because neither DataCite nor DCAT-AP sufficiently met the requirements of the Czech research communities (Section 2.3). Table 8 revisits these requirements and shows how CCMM addresses each of them. All of them are addressed by elements of the model; three are met only partly. Creators are not ordered (RC3), as the order is not expressed in DCAT-AP and RDF. A contact cannot be non-public (RC5), as the metadata are public, so a contact is either public or not given. The note on changes is deliberately left out of the model (RC7, Section 4.1.3). The end of an embargo needs no date type of its own, as it is represented by the date type *available* set in the future (RC6).

Table 8: How CCMM addresses the research community requirements (RC).

| ID | Requirement | How CCMM addresses it |
|---|---|---|
| *High priority* | | |
| RC1 | Data set PID | Mandatory `Identifier` with its identifier scheme, e.g. DOI or Handle |
| RC2 | Description, documentation | At least one description, typed by the DataCite Description Types |
| RC3 | Creators, contributors, roles | `Resource attribution` with a role from the Agent Roles vocabulary; the creator role is mandatory; creators are not ordered, as the order is not expressed in DCAT-AP and RDF (partly met) |
| RC4 | Agent PIDs | Identifiers of persons and organisations with any scheme, preferably ORCID and ROR, also IČO; name as a fallback |
| RC5 | Contact person | Optional contact role with `Contact details`, such as e-mail, phone, address or data box; as the metadata are public, a contact is either public or not given (partly met) |
| RC6 | Typed dates | `Time reference`s typed by the DataCite Time Reference Types, as an instant or interval; the creation date is mandatory; the end of an embargo as the date type *available* set in the future |
| RC7 | Versioning | Optional version and version relations; no note on changes (partly met) |
| RC8 | Multilingual metadata | Language-tagged strings throughout the model; primary and other languages from EU Languages |
| RC9 | Access rights | Mandatory access rights from the COAR Access Rights |
| RC10 | Licence | Mandatory licence of every distribution |
| RC11 | File access, format | `Downloadable file` with mandatory access URL, format and size; `Data service`; distributions are optional |
| RC12 | Related resources | Optional `Related resource` with identifier, DataCite relation type and resource type |
| *Medium priority* | | |
| RC13 | Publisher | Publisher role of an organisation for the data set; repository linked through the metadata record, with its own publisher |
| RC14 | Rights holder | Optional rights holder role |
| RC15 | Subject, classification | Mandatory subject from the OECD FORD classification; other classifications by scheme and code; free keywords kept apart |
| RC16 | Location | Optional `Location` by name, geometry, bounding box or gazetteer entry, with a relation type |
| RC17 | File integrity | `Checksum` with an SPDX algorithm; packaging and compression formats |

Table 8: (continued)

| ID | Requirement | How CCMM addresses it |
|---|---|---|
| RC18 | Provenance | Optional `Provenance activity` linking to a Common Provenance Model or RO-Crate representation |
| *Low priority* | | |
| RC19 | Title types | Optional alternate titles typed by the DataCite Alternate Title Types |
| RC20 | Resource type | Optional resource type from the COAR Resource Types |
| RC21 | Funding | Optional `Funding reference` with funder, award title, grant number and programme |

CCMM thus closes the gaps of both analysed profiles identified in Section 2.3, apart from the limits named above, including the two requirements that neither profile met: the publisher (RC13), which is an agent with the publisher role while the repository is linked through the metadata record, and provenance (RC18), which is a structured activity rather than free text (P7).

The cardinalities follow the community needs (P10). Elements needed for citation, such as the identifier, title, description, creator, creation date and access rights, are mandatory. The subject from the OECD FORD classification is mandatory as well, although its priority was medium, because it is required by the Czech Information System for Research, Development and Innovation (IS VaVaI) [35].

Some high-priority requirements that do not apply to every data set, such as versioning (RC7), file access (RC11) and related resources (RC12), are met by optional elements.

## 5.3 How the model solves the problem

The Introduction (Section 1) called for a common metadata model for research data that is shared across disciplines, leaves room for domain-specific requirements and is agreed on by the communities, without adding another competing standard. The gap that justifies the creation of CCMM was shown: neither DataCite nor DCAT-AP, alone or in parallel, meets the FAIR and community requirements (Section 2.3). CCMM does not compete with these standards, but connects them: it reuses their terms (P1) and its records can be transformed to DataCite, OpenAIRE and valid DCAT-AP (P9). Its core is minimal and domain communities can extend it without breaking conformance (P10, P12). Its requirements and obligations come from the survey of Czech research communities (Section 2.2) and from the EOSC CZ Metadata Working Group, and it was adopted as the metadata model of the Czech National Data Infrastructure.

The national specifics of CCMM are implemented as values of controlled vocabularies rather than being part of its structure. This makes the model, as

well as the method of its development, easy to reuse elsewhere. However, the requirements were collected from Czech communities only, so whether CCMM fits other environments without changes remains to be shown by its use there.

### 5.4 How the model fits within the European research data landscape

The Czech research data community is not the only one facing fragmented requirements on metadata profiles. Several European initiatives have developed more explicit generic metadata models or application profiles for research data management. Belgium has developed the Flemish Application Profile for Research [37]; Finland has developed the Metax V3 data model for research datasets, fully compatible with DCAT 3;[16] and Sweden uses a common metadata structure in its DORIS system, based on DDI Lifecycle 3.3 and DataCite.[17]

For example, within Germany's National Research Data Infrastructure (NFDI), the need emerged to represent the broader experimental context of research data, including provenance and the circumstances of data generation. In response, they developed a generic application profile called DCAT-AP+ (DCAT Application Profile Linking to Use-case Specific Context). DCAT-AP+ extends DCAT-AP 3.0 with a generic provenance and contextual layer based primarily on PROV-O (classes Entity, Activity and Agent), complemented by concepts from QUDT and Dublin Core Terms. Like CCMM, it is designed to be extensible into domain-specific application profiles. The first such specialisation is the Chemistry DCAT Application Profile (ChemDCAT-AP), developed jointly by NFDI4Chem and NFDI4Cat. [36]

## 6 Conclusion

This paper presented the Czech Core Metadata Model (CCMM), a core application profile for describing research data, together with the process of its development. The requirements on the model were derived from the FAIR Principles and EOSC policies on one side and from a survey of Czech research communities, complemented by continuous input from the communities, on the other, and evaluated against DataCite and DCAT-AP. Neither profile met them on its own: where DataCite fails, DCAT-AP mostly succeeds, and the other way around. From these gaps, twelve design principles were set out, and CCMM was built on them as a profile of DCAT-AP that reuses DataCite terms and meets the requirements that neither of them did.

CCMM addresses all FAIR and EOSC requirements (RF1–RF12) that a metadata model can address, and all community requirements (RC1–RC21), three of them only partly. Its records can be transformed to DataCite and OpenAIRE and,

[16] https://metax.fairdata.fi/v3/docs/user-guide/differences-from-dcat/

[17] https://snd.se/en/doris-researchers/describe-data-doris/metadata-and-subject-specific-profiles-doris

by an unambiguous mapping, to valid DCAT-AP, so that a single record serves both citation and discovery. Typed elements bound to controlled vocabularies and domain profiles built on top of the core let research communities describe their data in more detail without breaking conformance. Developed with the EOSC CZ Metadata Working Group and adopted as the metadata model of the Czech National Data Infrastructure, CCMM shows that a community-agreed core can connect existing standards rather than compete with them.

The model alone does not make data FAIR; this also depends on how repositories and researchers use it and on the infrastructure that publishes the records. Even though the requirements were collected only in the Czech Republic, the principles on which CCMM was built and the method of its development were intentionally designed to be reusable in other national and international environments, and our aim is to extend the use of CCMM to the international community.

# Declarations

## Declaration on Generative AI

During the preparation of this work, the authors used Generative Artificial Intelligence (Google Gemini 2.5 Flash, Claude Opus 5.5, OpenAI GPT-6) in order to check grammar and spelling, improve text readability, and assist in text editing. After using the tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.


## Funding

The design of CCMM and related work were supported by the EOSC CZ Initiative through the project IPs CARDS (CZ.02.01.01/00/22_004/0004342) and through the project IPs EOSC-CZ (CZ.02.01.01/00/22_004/0007682), both funded by the Programme Johannes Amos Comenius (P JAC) of the Ministry of Education, Youth and Sports of the Czech Republic (MEYS).

Funding for open access charge: National Library of Technology.


## Data Availability

CCMM is shared as an application profile and an ontology in two GitHub repositories, both published under the CC BY 4.0 licence. The repository https://github.com/techlib/CCMM [38] contains release 2.0.0 of the profile, whose XML namespace is https://schema.ccmm.cz/research-data/2.0. For every class of the model, it has a folder (e.g. `dataset/`, `agent/` or `distribution/`) with the XML Schema (`schema.xsd`), the JSON Schema (`schema.json`), the JSON-LD context (`context.jsonld`), the transformations between XML and RDF (`lifting.xslt`, `lowering.xslt`), the structural model (`structure.ttl`) and an example record (`example.json`) (Table 6). At the root of the repository, `dsv.ttl` contains the machine-readable description of the profile in DSV and

`shacl.ttl` the SHACL shapes; the folders `en/` and `cs/` contain the human-readable specification in English and Czech, `_metadata-samples/` contains sample records in XML and RDF, and the ZIP file contains the Dataspecer project for import into another Dataspecer instance. The repository https://github.com/techlib/CCMM-V [39] contains the CCMM Vocabulary, which defines the terms not taken from existing vocabularies, as an OWL ontology (`model.owl.ttl`) with its documentation in English and Czech (`en/`, `cs/`). The summary of the questionnaire on needs for dataset description and discovery, from which the community requirements were derived (Section 2.2), is available at https://doi.org/10.48813/6ytw-be20 [15], also under the CC BY 4.0 licence.

### Code Availability

All custom code is openly available without restrictions under the CC BY 4.0 licence. The repository https://github.com/techlib/CCMM-utils [40] contains tools supporting the use of CCMM: `flattenCCMM/`, a Python tool that merges the multi-file XML Schema into a single schema with `Dataset` at the root; `ccmm2rdf/`, which transforms XML records into RDF (Turtle, RDF/XML and JSON-LD) using the lifting transformation; `version-transformation/`, with XSLT templates for transforming records between CCMM 1.1.0 and 2.0.0; and `ceCCMM/`, which visualises the XML Schema with mandatory elements in bold. The repository https://github.com/techlib/CCMM-DataCite-mappings [41] contains the mappings and transformations between CCMM 2.0.0 and DataCite 4.7, with test records (`metadata-samples-test/`); the version described in this paper is the release `release-ccmm-2.0.0-datacite-4.7-alpha1`. While the CCMM-to-DataCite transformation is already available, the CCMM-to-OpenAIRE implementation is currently in progress and is expected to be ready for the camera-ready version of this paper.

## References


[1] Mark D. Wilkinson et al. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3:160018, 2016. `doi:10.1038/sdata.2016.18`.

[2] DataCite Metadata Working Group. DataCite metadata schema documentation for the publication and citation of research data and other research outputs. version 4.7. Technical report, DataCite e.V., 2026. `doi:10.14454/qdd3-ps68`.

[3] SEMIC – Semantic Interoperability Community. DCAT application profile for data portals in Europe (DCAT-AP v3.0.1). Technical report, Publications Office of the European Union, 2024. URL: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe.

[4] Matej Antol, Jiří Marek, Michaela Capandová, Jaroslav Juráček, and Luděk Matyska. EOSC CZ: Towards the development of Czech national ecosystem for FAIR research data. arXiv:2402.13343, 2024. doi:10.48550/arXiv.2402.13343.

[5] European Commission. European Open Science Cloud (EOSC). Research and innovation, European Commission. URL: https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science/european-open-science-cloud-eosc_en.

[6] EOSC CZ. Working Groups in Data. EOSC CZ. URL: https://www.eosc.cz/en/working-groups/working-groups-in-data.

[7] ESFRI. Thematic Areas. ESFRI Roadmap 2021, Part 2: Landscape Analysis, Section 1, 2021. URL: https://roadmap2021.esfri.eu/landscape-analysis/section-1/.

[8] EOSC CZ. Metadata Working Group. EOSC CZ. URL: https://www.eosc.cz/en/working-groups/metadata.

[9] Hana Vyčítalová, Petra Černohlávková, and Jakub Klímek. Analysis of international metadata schemas for describing research data. Analysis/study, National Library of Technology, Praha, 2023. Project Czech Academic and Research Discovery Services. Licence CC BY. Handle: 20.500.14391/2921. doi:10.48813/6kjx-cr80.

[10] Annika Jacobsen et al. FAIR principles: Interpretations and implementation considerations. *Data Intelligence*, 2(1–2):10–29, 2020. doi:10.1162/dint_r_00024.

[11] Oscar Corcho, Magnus Eriksson, Krzysztof Kurowski, Milan Ojsteršek, Christine Choirat, Mark van de Sanden, and Frederik Coppens. EOSC interoperability framework: Report from the EOSC executive board working groups FAIR and architecture. Technical report, Publications Office of the European Union, 2021. doi:10.2777/620649.

[12] Paolo Manghi, Alessia Bardi, Claudio Atzori, Miriam Baglioni, Natalia Manola, Jochen Schirrwagen, Pedro Principe, et al. The OpenAIRE research graph data model, 2019. doi:10.5281/zenodo.2643199.

[13] Carl Lagoze and Herbert Van de Sompel. The Open Archives Initiative: Building a low-barrier interoperability framework. In *Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '01)*, pages 54–62, 2001. doi:10.1145/379437.379449.

[14] OpenAIRE. OpenAIRE guidelines for data archive managers v3 (draft). OpenAIRE Guidelines, 2020. Based on DataCite Metadata Schema 4.3. URL: https://openaire-guidelines-for-data-archive-managers.readthedocs.io/.

[15] Hana Vyčítalová, Alexandr Kuchynka, and Petra Černohlávková. Shrnutí výsledků dotazníku k potřebám popisu a vyhledávání datasetů, 2024. URL: https://hdl.handle.net/20.500.14391/3032, doi:10.48813/6ytw-be20.

[16] Ramanathan V. Guha, Dan Brickley, and Steve Macbeth. Schema.org: Evolution of structured data on the web. *Communications of the ACM*, 59(2):44–51, 2016. doi:10.1145/2844544.

[17] Schema.org Community Group. Schema.org. Vocabulary specification, March 2026. URL: https://schema.org/version/30.0/.

[18] DCMI Usage Board. DCMI metadata terms. DCMI Recommendation, January 2020. Issued 2020-01-20. URL: http://dublincore.org/specifications/dublin-core/dcmi-terms/2020-01-20/.

[19] ISO. ISO 15836-2:2019 information and documentation – the Dublin Core metadata element set – part 2: DCMI properties and classes. Technical report, International Organization for Standardization, Geneva, 2019. URL: https://www.iso.org/standard/71341.html.

[20] Sandro Fabrizio La Bruzzo, Michele Artini, Claudio Atzori, Alessia Bardi, Miriam Baglioni, Michele De Bonis, Andrea Mannocci, Paolo Manghi, and Gina Pavone. Data model description of the OpenAIRE research graph. ISTI Technical Report ISTI-TR-2022/031, ISTI-CNR, Pisa, 2022. doi:10.32079/ISTI-TR-2022/031.

[21] Digitální a informační agentura. DCAT-AP-CZ: Rozhraní katalogů otevřených dat. Otevřená formální norma, May 2024. Verze 2024-05-28, založeno na DCAT-AP 3.0.0. URL: https://ofn.gov.cz/dcat-ap-cz-rozhranÃŋ-katalogÅŕ-otevÅŹenÃ¡ch-dat/2024-05-28/.

[22] Riccardo Albertoni, David Browning, Simon Cox, Alejandra Gonzalez-Beltran, Andrea Perego, and Peter Winstanley. Data catalog vocabulary (DCAT) – version 3. W3C Recommendation, 2024. URL: https://www.w3.org/TR/vocab-dcat-3/.

[23] Riccardo Albertoni, David Browning, Simon Cox, Alejandra N. Gonzalez-Beltran, Andrea Perego, and Peter Winstanley. The W3C data catalog vocabulary, version 2: Rationale, design principles, and uptake. *Data Intelligence*, 6(2):457–487, 2024. doi:10.1162/dint_a_00241.

[24] European Commission, Joint Research Centre. CiteDCAT-AP: DataCite profile of DCAT-AP. URL: https://w3id.org/citedcat-ap.

[25] Rachel Heery and Manjula Patel. Application profiles: mixing and matching metadata schemas. *Ariadne*, (25), 2000. URL: http://www.ariadne.ac.uk/issue/25/app-profiles/.

[26] Alon Halevy. Why your data won't mix. *ACM Queue*, 3(8):50–58, 2005. doi:10.1145/1103822.1103836.

[27] Štěpán Stenchlák, Martin Nečaský, Petr Škoda, and Jakub Klímek. DataSpecer: A model-driven approach to managing data specifications. In *The Semantic Web: ESWC 2022 Satellite Events*, volume 13384 of *Lecture Notes in Computer Science*. Springer, 2022. doi:10.1007/978-3-031-11609-4_10.

[28] Jakub Klímek, Štěpán Stenchlák, and Petr Škoda. Data specification vocabulary (DSV): Representation of application profiles of semantic data specifications. In *Information Integration and Web Intelligence*, pages 221–236, Cham, 2026. Springer Nature Switzerland. doi:10.1007/978-3-032-11976-6_16.

[29] Alistair Miles and Sean Bechhofer. SKOS simple knowledge organization system reference. W3C Recommendation, 2009. URL: https://www.w3.org/TR/skos-reference/.

[30] Laurel L. Haak, Martin Fenner, Laura Paglione, Ed Pentz, and Howard Ratner. ORCID: a system to uniquely identify researchers. *Learned Publishing*, 25(4):259–264, 2012. doi:10.1087/20120404.

[31] Rudolf Wittner, Cecilia Mascia, Matej Gallo, Francesca Frexia, Heimo Müller, Markus Plass, Jörg Geiger, and Petr Holub. Lightweight distributed provenance model for complex real-world environments. *Scientific Data*, 9:503, 2022. doi:10.1038/s41597-022-01537-6.

[32] Stian Soiland-Reyes, Peter Sefton, Mercè Crosas, et al. Packaging research artefacts with RO-Crate. *Data Science*, 5(2):97–138, 2022. doi:10.3233/DS-210053.

[33] Jens Klump, Lesley Wyborn, Mingfang Wu, Julia Martin, Robert R. Downs, and Ari Asmi. Versioning data is about more than revisions: A conceptual framework and proposed principles. *Data Science Journal*, 20:12, 2021. doi:10.5334/dsj-2021-012.

[34] Ondřej Zamazal. Aligning the Czech Core Metadata Model with DataCite: A real-world matching case study. In *Proceedings of the Ontology Matching Workshop (OM 2026)*, 2026. URL: https://om.ontologymatching.org/2026/.

[35] Rada pro výzkum, vývoj a inovace. Popis údajů dodávaných do IS VaVaI – RIV v roce 2026. Version 3.5.0, Rada pro výzkum, vývoj a inovace, Praha, 2025. Description of data submitted to the Czech Information System for Research, Development and Innovation, Register of Information on Results (RIV), 2026. In Czech. URL: https://www.isvavai.cz/dokumenty/RIV26_3.5.0_v5.pdf.

[36] Philip Stroemert, Hendrik Borgelt, David Linke, Mark Doerr, Bhavin Katabathuni, Oliver Koepler, and Norbert Kockmann. ChemDCAT-AP: Enabling semantic interoperability with a contextual extension of DCAT-AP. arXiv:2602.01822, 2026. doi:10.48550/arXiv.2602.01822.

[37] Evy Neyens and Sadia Vancauwenbergh. Towards a semantic interoperable Flemish research information space: Development and implementation of a Flemish application profile for research datasets. *International Journal of Digital Curation*, 16(1):17, 2021.

[38] National Library of Technology. Czech Core Metadata Model (CCMM), release 2.0.0. GitHub repository, 2026. URL: https://github.com/techlib/CCMM.

[39] National Library of Technology. CCMM Vocabulary, release ccmm-2.0.0. GitHub repository, 2026. URL: https://github.com/techlib/CCMM-V.

[40] National Library of Technology. CCMM Utilities. GitHub repository, 2026. URL: https://github.com/techlib/CCMM-utils.

[41] National Library of Technology. CCMM–DataCite mappings, release release-ccmm-2.0.0-datacite-4.7-alpha1. GitHub repository, 2026. URL: https://github.com/techlib/CCMM-DataCite-mappings.

[42] SEMIC. *How to Create Your DCAT-AP Profile*. European Commission, 2026. Available at: https://semiceu.github.io/DCAT-AP-reuse-guidelines/ (Accessed: 14 November 2026).