Czech Core Metadata Model for research data

Michal Med, Cesnet

https://orcid.org/0000-0002-3844-426X

EOSC CZ

3rd December 2025

Ostrava

HAZY GALAXY

SESSION IPA

FERMENTATION:

Top

STYLE:

India Pale Ale (IPA)

ABV: 4,9 %

PLATO: 12

WATER:

Groundwater – infiltration from Jizera river

MALTS:

Wheat

HOPS:

Cascade, Galaxy, Sabro

VINOHRADSKÝ PIVOVAR

PRAHA · KARANT

HAZY GALAXY

SESSION IPA

CRAFT BEER · 4,9% ABV

Metadata usage

 END USERS (DRINKERS)

OTHER PROVIDERS (BREWERS)

GOVERNANCE (TRADE INSPECTION, HYGIENE)

Specific needs

Groceries

Drinks

Alcoholic bevarages

Beers

Wines

Soft drinks

Ales

Lagers

Dry wines

Red wines

Needs to understand (at least something) each other

Whats new in **1.1.0**

Agent ✎

Data structure for Agent. Any entity carrying out actions with respect to the entities Catalogue and the Catalogued Resources.

⌄ (agent) **OR**

∧ [refers to] **Organization** (`organization`)

∧ [refers to] **Person** (`person`)

Distribution ✎

Data structure for Distribution. A specific representation of a dataset. A dataset might be available in multiple serializat including natural language, media-type or format, schematic organization, temporal and spatial resolution, level of detai any or all of the above).

⌄ (distribution) **OR**

∧ [refers to] **Distribution - data service** (`distribution_data_service`)   +   🗑   ✎   ⋮

∧ [refers to] **Distribution - downloadable file** (`distribution_downloadable_file`)

Time reference ✎

Data structure for Time reference. An object representing time reference relevant to some event related to the resource has happened.

⌄ **Time reference** (`time_reference`)

⌄ **has temporal representation** (`temporal_representation`): [refers to] **OR** (`time_representation`) [1..1]

∧ [refers to] **Time interval** (`time_interval`)

∧ [refers to] **Time instant** (`time_instant`)

∧ **has date type** (`date_type`): [refers to] **Date type** (`date_type`) [1..1]

— **date information** (`date_information`) : Text [0..1]

Definition of publicationYear contains Item #96

⊘ Closed

MichalMed opened on Oct 13

Replace word item with dataset.

[Create sub-issue | ▾] 😊

Provenance statement - definition #67

⊘ Closed

6tova opened on Aug 7

Could this class definition be adjusted to cover more than just changes in ownership and custody, specifically the history of the creation of the dataset. Description of the origin. If it could be aligned with the ISO 19115 Lineage definition maybe?

consistency: title on distribution service #36

⊘ Closed   Task

mesemus opened on May 28

We have a single string title on a dataset (without a language), with additionalTitles for translations etc.
On distribution service, the title is i18nStr (language aware) but we can not specify any translations etc.

Should be unified.

Should original_repositories have maxoccurs=unbounded #78

✓ Closed

mesemus opened on Aug 30

In the context of repositories, a dataset should be stored in one repository (where it gets its primary persistent identifier). Do we need to have original_repository as an array then? What are the use cases?

- production repositories: these should contain the primary copy of the dataset, so no need there
- NMA: should harvest primary repositories, not aggregators. It might even be misleading - if NMA would somehow merge metadata of a single dataset from multiple repositories, we can not capture from which repository a single piece of metadata comes (such a relation is not present in ccmm) and thus we might show misleading information.

Description text and description type label are not language tagged #90

✓ Closed

MichalMed opened on Sep 25   `Collaborator`

Both descrption text and description type label shall nbe language specific with mandatory usage of xml:lang attribute.

Setup logical order of elements in XML #69

✓ Closed

MichalMed opened on Aug 19

> order of elements in some cases does not make much sense, e.g. subject definition bef

Originally posted by @MichalMed in #60

Design order of elements:

1. by importance
2. group them by topic (distribution info, identification of dataset, etc...)

Create sub-issue ▾ ☺

ⵔ **MichalMed** self-assigned this on Aug 19

⊟ **MichalMed** added this to the 1.1.0 milestone on Aug 19

MichalMed on Aug 19 · edited by MichalMed

Following attribut groups as in ISO MD:

MD identification - described by metadata (language, responsible person, date, used meta

data identification - identifier, title, alternate title, description, version, publication year, pr
resource type, agent relationship, time reference, subject, terms of use, location, funding

distribution info - distribution

data quality - validation, provenance

☺

Required items in ccmm spec #72

✓ Closed

dvoriik opened on Aug 21

Add "minimal" example of mandatory items (1-1), (1-n) of the ccmm specification with possibility to filter them in datas

Create sub-issue ▾ ☺

⊟ **MichalMed** added this to the 1.1.0 milestone on Sep 4

MichalMed on Sep 4   Collabora

@Pavlki

☺

ⵔ **MichalMed** assigned Pavlki on Oct 15

Pavlki on Oct 24 · edited by Pavlki   Edits ▾   Contribu

The README in the 'main' branch has been updated to include a reference and link to 'dataset-mini.xml' (currently loca
the 'examples' branch, not yet merged).

The reference with a link to dataset-mini.xml has been added to the README in the '1.1.0' branch as well.
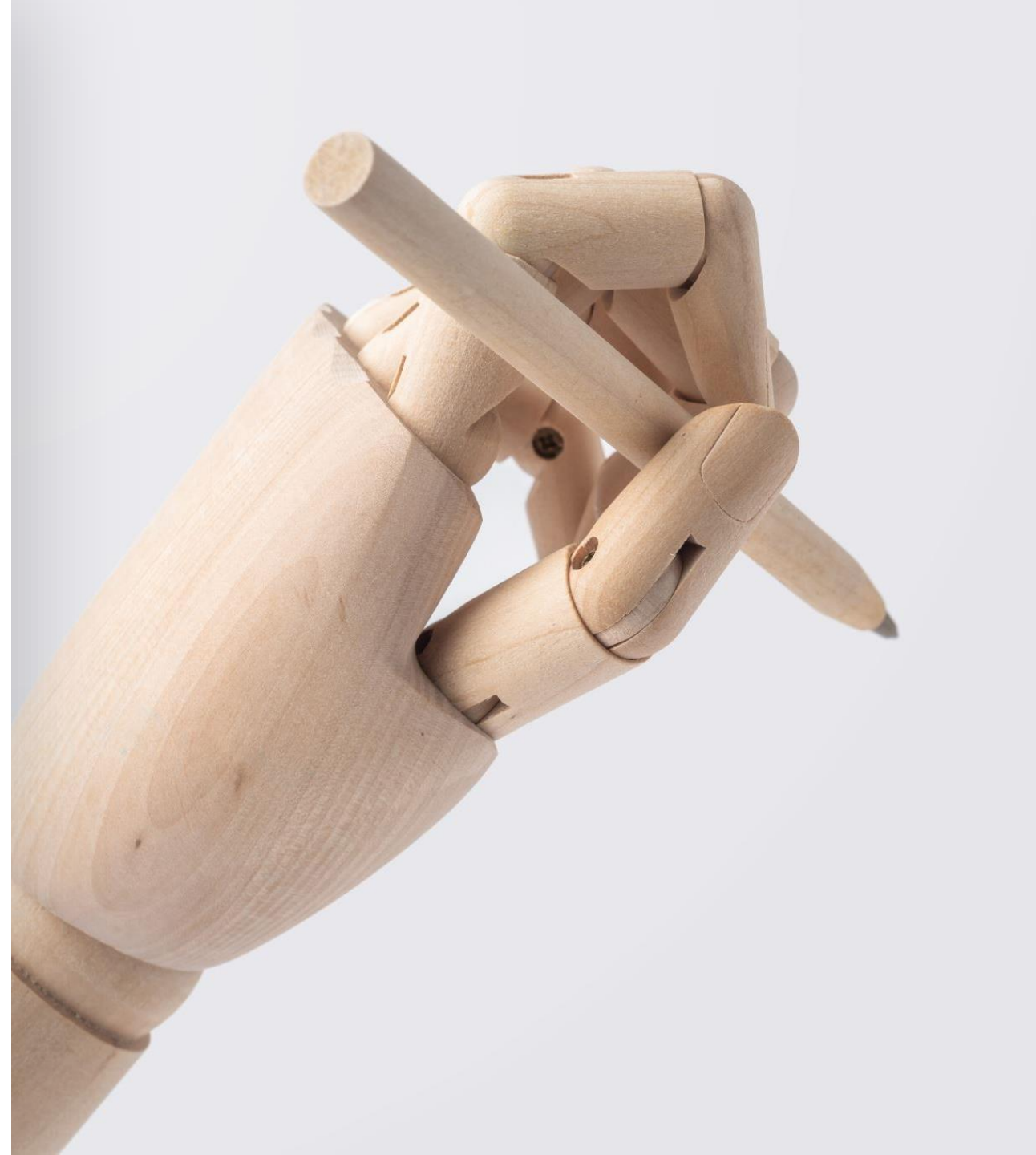
☺

✓ **Pavlki** closed this as completed on Oct 24

The publication

- All Changes are in the standalone branch stable since the beginning of November.
- Currently implemented in Invenio.
- Will be releases as the implementation is done.

Best practices

- CCMM is often referred as too wide for core model

- In the meantime it allows users to express a lot of information using just the core model

- Let's look at two specific use cases

FAQs

- Dataset was created according to the European/national legislation. I need to add this property to the model.

- I have used this concrete microscope and it has some specific properties. **I need to add this property to the model.**

- The dataset was created following this specific methodical document. **I need to add this property to the model.**

§ 2.35 Resource

Profiled class IRI(s)	rdfs:Resource
IRI	https://model.ccmm.cz/research-data/Resource
Label	Resource
Definition	Resource represents any resource, physical or digital, that is related to the described dataset.
Profiles	class Resource (rdfs:Resource) Definition: *The class resource, everything.*
Usage note	Class is used for description of the resource together with the type of relation, for which is used value from codelist https://vocabs.ccmm.cz/registry/codelist/RelationType/. Distributions of the datasets *SHALL NOT* be modeled as related resource, but as a distribution. Use qualified resource property to link to the dataset class. Examples of resources: other related dataset, used methodology, instrument used for measurement, conference or conference paper etc...

Backwards associations

- from domain 2.10 Data service → 2.10.1 endpoint URL
- from domain 2.27 Location → 2.27.4 has related object
- from domain 2.11 Dataset → 2.11.10 has related resource

https://model.ccmm.cz/research-data/en/#Resource

Data structure editor

Resource 🖉

Data structure for Resource. Resource represents any resource, physical or digital, that is related to the described dataset.

⌄ **Resource** (resource)

　⬚ ∧ **has identifier** (identifier): [refers to] **Identifier** (identifier) [0..*]

　— **title** (title) : String [0..1]

　⬚ ∧ **has alternate title** (alternate_title): [refers to] **Alternate title** (alternate_title) [0..*]

　— **resource URL** (resource_url) : URI, IRI, URL [0..1]

　⬚ ∧ **qualified relation** (qualified_relation): [refers to] **Resource to agent relationship** (resource_to_agent_relationship) [0..*]

　⬚ ∧ **has time reference** (time_reference): [refers to] **Time reference** (time_reference) [0..*]

　⬚ ∧ **has resource type** (resource_type): [refers to] **Resource type** (resource_type) [0..1]

　⬚ ∧ **has resource relation type** (resource_relation_type): [refers to] **Resource relation type** (resource_relation_type) [0..1]

Data structure editor

ARTIFACTS ⌄   SET ROOT ELEMENT ▾

Resource ✎

Data structure for Resource. Resource represents any resource, physical or digital, that is related to the described dataset.

- ⌄ **Resource** (resource)
  - ∧ **has identifier** (identifier): [refers to] **Identifier** (identifier) [0..*]
  - — **title** (title) : String [0..1]
  - ∧ **has alternate title** (alternate_title): [refers to] **Alternate title** (alternate_title) [0..*]
  - — **resource URL** (resource_url) : URI, IRI, URL [0..1]
  - ∧ **qualified relation** (qualified_relation): [refers to] **Resource to agent relationship** (resource_to_agent_relationship) [0..*]
  - ∧ **has time reference** (time_reference): [refers to] **Time reference** (time_reference) [0..*]
  - ∧ **has resource type** (resource_type): [refers to] **Resource type** (resource_type) [0..1]
  - ∧ **has resource relation type** (resource_relation_type): [refers to] **Resource relation type** (resource_relation_type) [0..1]

Profiled relationship IRI(s)	dcterms:type
IRI	https://model.ccmm.cz/research-data/Resource.type
Label	has resource type
Definition	A resource type of the related resource.
Domain	Resource
Range	Resource type [0..1]
Profiles	property Type (dcterms:type) Definition: *The nature or genre of the resource.*
Usage note	Use IRI identifier of a value from the register https://vocabularies.coar-repositories.org/resource_types/.

C O **AR** Controlled Vocabularies for Repositories

Home   Resource Types   Access Rights   Version Types   About   Give Feedback

Resource Types 3.2

Concepts

- artistic work
- cartographic material
  - map
- collection
  - archival collection
  - court documents
- dataset
  - aggregated data
  - clinical trial data
  - compiled data
  - encoded data
  - experimental data
  - genomic data
  - geospatial data
  - laboratory notebook
  - measurement and test data
  - observational data
  - recorded data

This is the latest version of this vocabulary

Description

The Resource Type vocabulary defines concepts to identify the genre of a resource. Such resources, like publications, research data, audio and video objects, are typically deposited in institutional and thematic repositories or published in ejournals. This vocabulary supports a hierarchical model that relates narrower and broader concepts. Multilingual labels regard regional distinctions in language and term. Concepts of this vocabulary are mapped with terms and concepts of similar vocabularies and dictionaries.

Released

2024-12-03

Namespace URI

http://purl.org/coar/resource_type/

Vocabulary PID

https://vocabularies.coar-repositories.org/resource_types/

Versions

- 3.2 (latest version)

Creators

- Isabel Bernal, CSIC, Spain (Editorial Board Chair)
- Joseph Cera, University of California-Berkeley, US
- Mick Eadie, University of Glasgow, United Kingdom
- Benedetta Gandolini, 4Science, Italy
- Sebastiano Giorgi, UOC, Spain
- Gültekin Gürdal, Izmir Institute of Technology Library, Turkey
- Tomoko Kataoka, JPCOAR, Japan
- Pierre Lasou, Université Laval, Canada
- Yoo Young Lee, University of Ottawa, Canada
- Ku (Alan) Liping, Chinese Academy of Science

```xml
<related_resource>
    <title>ENVI LVS1 Sampler pro odběr prašného aerosolu</title>
    <resource_url>https://www.envitech-bohemia.cz/p/264/envi-lvs1-sampler-pro-odber-prasneho-aerosolu</resource_url>
    <resource_type>
        <iri>http://purl.org/coar/resource_type/8KJG-QS0Y</iri>
        <label xml:lang="en">research instrument</label>
    </resource_type>
    <!-- has no representative in given codelist -->
</related_resource>
```

```xml
<related_resource>
        <iri>http://data.europa.eu/eli/dir/2008/50/oj</iri>
        <title>Směrnice Evropského parlamentu a Rady 2008/50/ES ze dne 21.
  května 2008 o kvalitě
            vnějšího ovzduší a čistším ovzduší pro Evropu</title>
        <resource_url>https://eur-lex.europa.eu/legal-
  content/CS/TXT/HTML/?uri=CELEX:32008L0050"%26"qid=1754039487879</resource
  _url>
        <resource_type>
            <iri>http://purl.org/coar/resource_type/c_18cf/</iri>
            <label xml:lang="en">text</label>
            <label xml:lang="cs">text</label>
        </resource_type>
        <resource_relation_type>
            <iri>https://vocabs.ccmm.cz/registry/codelist/RelationType/IsR
  eferencedBy</iri>
            <label xml:lang="cs">je na něj odkazováno z (čeho)</label>
        </resource_relation_type>
    </related_resource>
```

Funding

§ 2.21 Funding reference

Profiled class IRI(s)	https://model.ccmm.cz/vocabulary/datacite#FundingReference
IRI	https://model.ccmm.cz/research-data/FundingReference
Label	Funding reference
Definition	Information about financial support (funding) for the resource being registered.
Profiles	class Funding reference (https://model.ccmm.cz/vocabulary/datacite#FundingReference) Definition: *Information about financial support (funding) for the resource being registered.*

Backwards associations

- from domain 2.11 Dataset → 2.11.4 has funding reference

https://model.ccmm.cz/research-data/en/#FundingReference

Data structure editor

Funding reference ✎

Data structure for Funding reference. Information about financial support (funding) for the resource being registered.

∨ **Funding reference** (`funding_reference`)

   — **local identifier** (`local_identifier`) : String [0..1]

   — **award title** (`award_title`) : String [0..1]

   — **funding program** (`funding_program`) : URI, IRI, URL [0..1]

   ∧ **has funder** (`funder`): [refers to] **OR** [1..*]

Grant funding

```xml
<funding_reference>
    <local_identifier>EH23_021/0008433</local_identifier>
    <award_title>Development of special wire semiproducts for welding and 3D printing</award_title>
    <funding_program>Operační program Jan Amos Komenský</funding_program>
    <funder>
      <organization>
        <identifier>
            <iri>https://ror.org/037n8p820</iri>
            <value>037n8p820</value>
          <scheme>
            <iri>https://ror.org/</iri>
            <label xml:lang="">ROR</label>
          </scheme>
        </identifier>
        <identifier>
          <iri>http://dx.doi.org/10.13039/501100001823</iri>
          <value>10.13039/501100001823</value>
          <scheme>
            <iri>https://doi.org/</iri>
            <label xml:lang="">DOI</label>
          </scheme>
        </identifier>
        <name>Ministerstvo školství, mládeže a tělovýchovy</name>
        <alternate_name xml:lang="en">Ministry of Education, Youth and Sports of the Czech
Republic</alternate_name>
      </organization>
    </funder>
  </funding_reference>
```

Institutional funding

Data structure editor

Funding reference ✏️

Data structure for Funding reference. Information about financial support (funding) for the resource being registered.
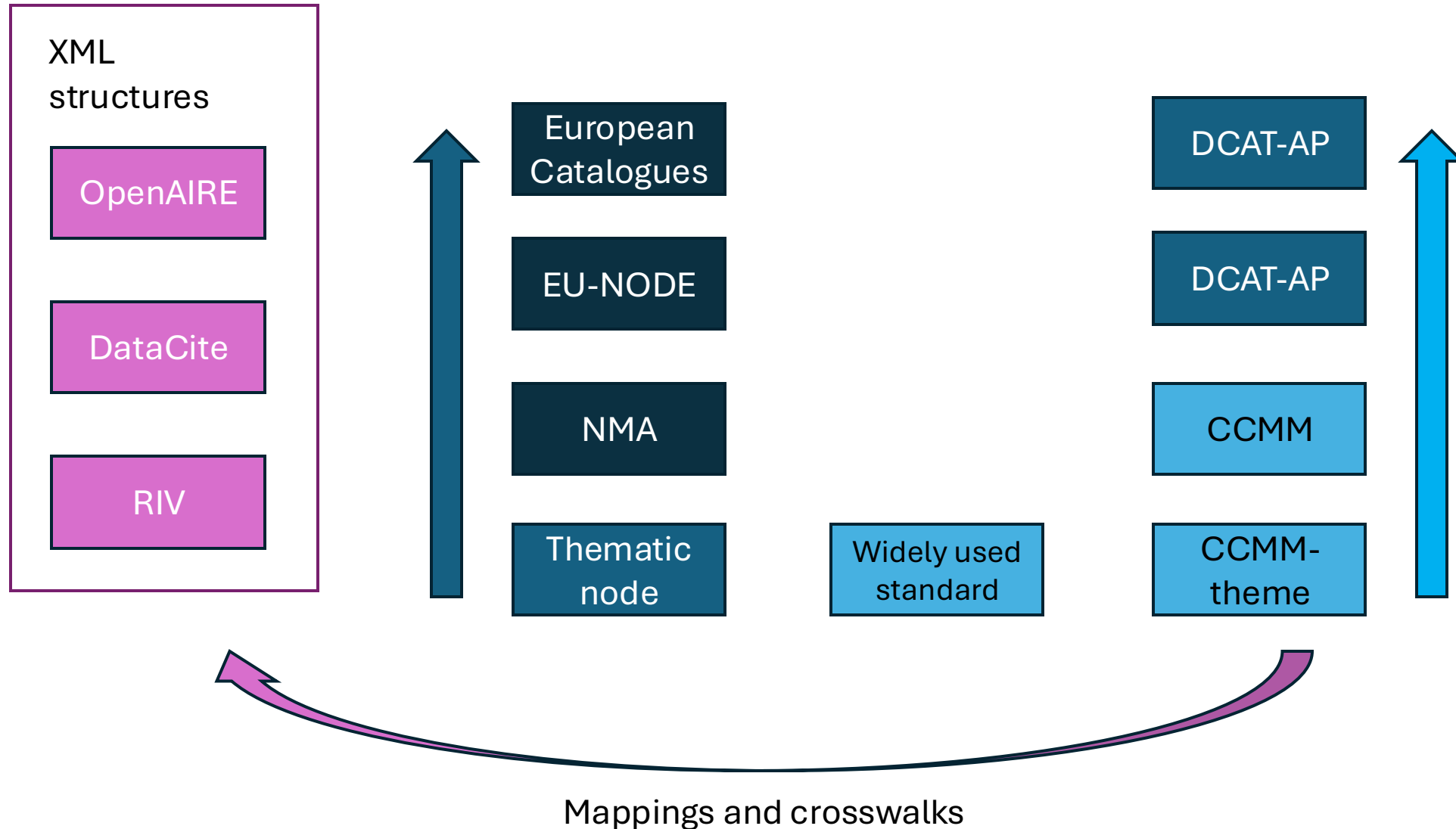
⌄ **Funding reference** (`funding_reference`)

    — **local identifier** (`local_identifier`) : String [0..1]

    — **award title** (`award_title`) : String [0..1]

    — **funding program** (`funding_program`) : URI, IRI, URL [0..1]

🔲 ⌃ **has funder** (`funder`): [refers to] OR [1..*]

```xml
<funding_reference>
    <funder>
      <organization>
        <identifier>
          <iri>https://ror.org/03nvfe914</iri>
          <value>03nvfe914</value>
          <scheme>
            <iri>https://ror.org/</iri>
            <label xml:lang="">ROR</label>
          </scheme>
        </identifier>
        <name>Masarykův ústav a Archiv Akademie věd ČR</name>
        <alternate_name xml:lang="en">Czech Academy of Sciences, Masaryk Institute and Archives</alternate_name>
      </organization>
    </funder>
  </funding_reference>
```

Interoperability of the metadata model

XML
structures

OpenAIRE

DataCite

RIV

European
Catalogues

EU-NODE

NMA

Thematic
node

Widely used
standard

DCAT-AP

DCAT-AP

CCMM

CCMM-
theme

Mappings and crosswalks

Thank you for attention

michal.med@cesnet.cz