

Domain Specific Metadata Profiles

Petra Černohlávková , Jakub Klímek , Hana Vyčítalová , Kristýna Vojtíšková

National EOSC CZ Conference 2025: Repo builders workshop

3 December 2025, Ostrava



Co-funded by
the European Union

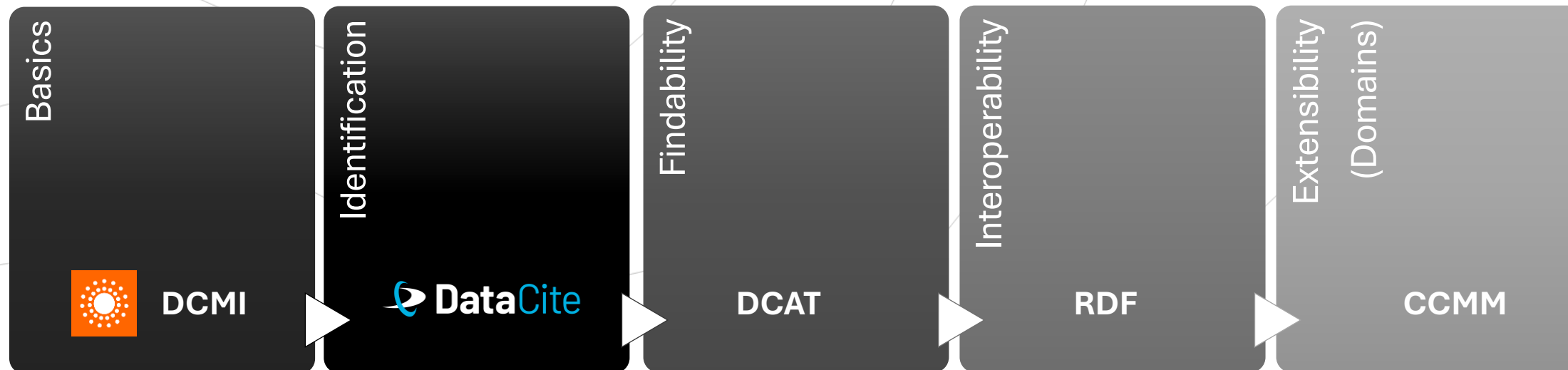


National Czech
Programme

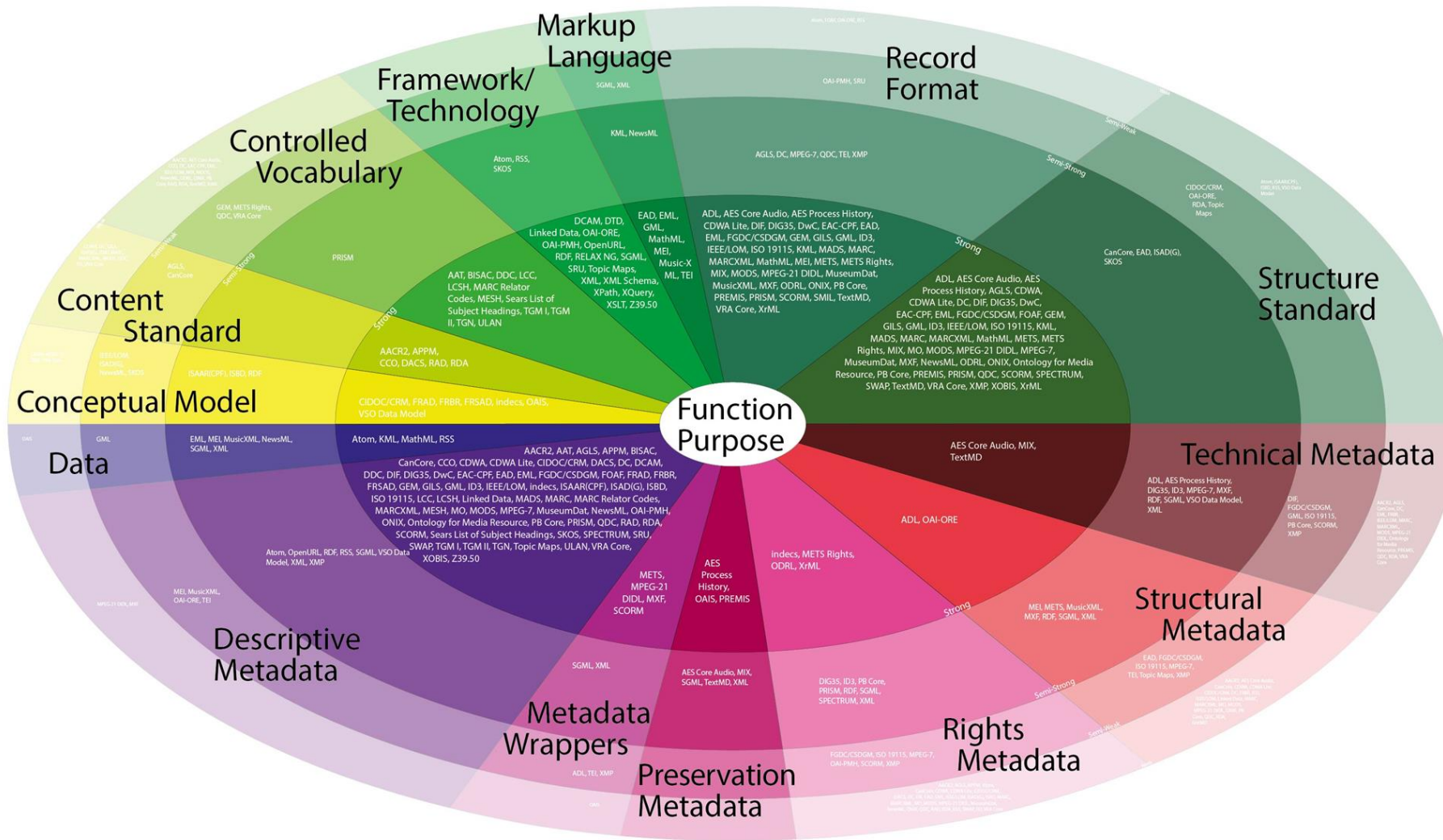


This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

- **Czech Core Metadata Model for Research Data (CCMM)**
 - universal schema for Czech research community
 - <https://www.ccmm.cz/>
 - recommended base for data description in the Czech Republic
 - stems from / Aligns with DCAT-AP, DataCite, Dublic Core
 - ensures national interoperability across **NRP** and **NMA** infrastructures



More metadata



What should the model do....

...fulfill your community needs
& the FAIR principles

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4: (Meta)data are registered or indexed in a searchable resource
- A2. Metadata are accessible, even when the data are no longer available
- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2.(Meta)data use vocabularies that follow FAIR principles
- I3.(Meta)data include qualified references to other (meta)data
- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
 - R1.1. (Meta)data are released with a clear and accessible data usage license
 - R1.2.(Meta)data are associated with detailed provenance
 - R1.3. (Meta)data meet domain-relevant community standards

What's the purpose of the model

- Data description and context
- Data identification
- Data management
- Searching
- Interoperability
- Archiving, preservation
- Re-usability
- Evaluation

Do you know what are your data?

Different objects might have different needs

Víno (láhev vína)

Název: Château Margaux 2018

Odrůda: Cabernet Sauvignon

Ročník: 2018

Obsah alkoholu: 13,5 %

Objem: 0,75 l

Region: Bordeaux, Francie

Certifikace: AOC (Appellation d'Origine Contrôlée)

Chutový profil: ovocné tóny, jemné taniny



Auto

Značka a model: Škoda Octavia Combi

Rok výroby: 2022

Typ motoru: 1.5 TSI benzín

Výkon: 110 kW

Barva: Modrá metalíza

Počet dveří: 5

VIN: unikátní identifikátor vozidla

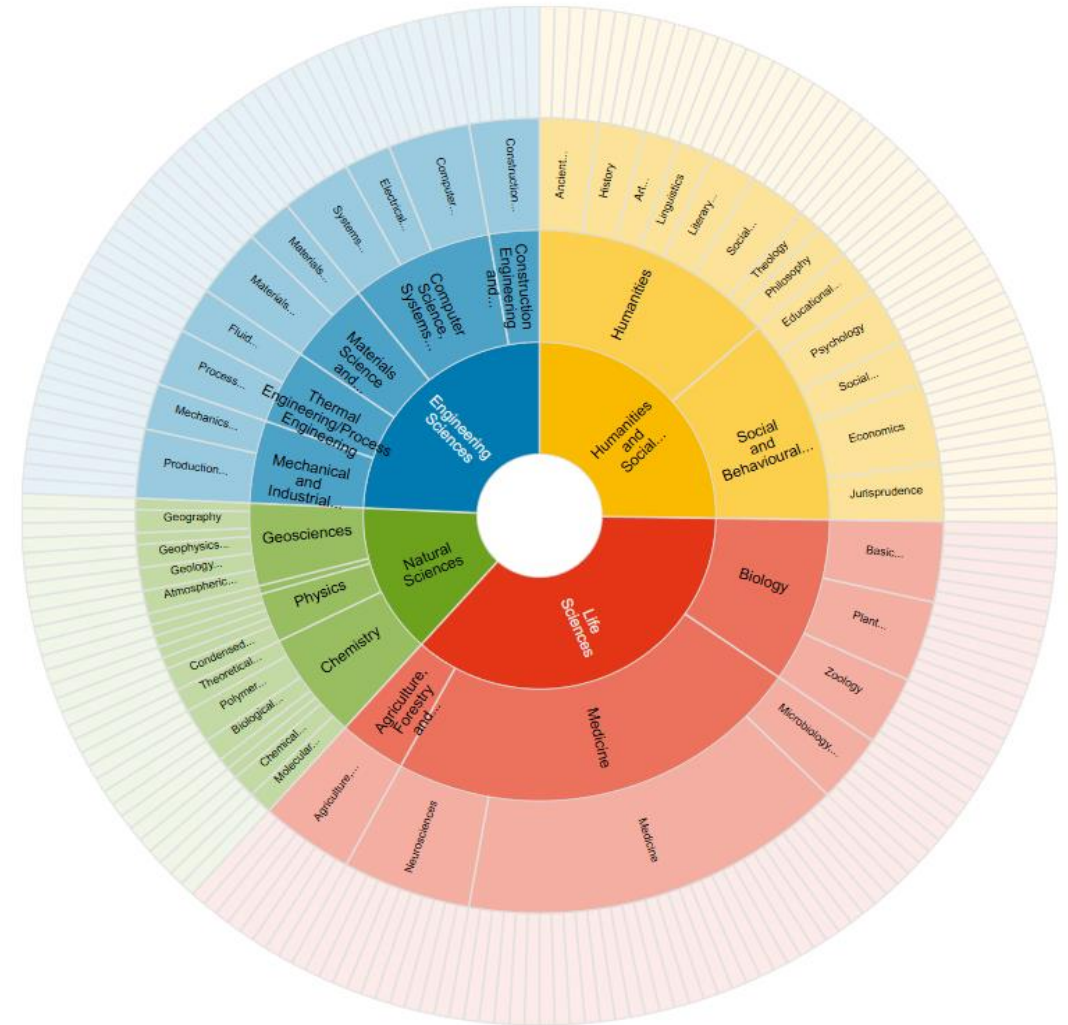
Spotřeba: 6,2 l/100 km

Výbava: Navigace, adaptivní tempomat

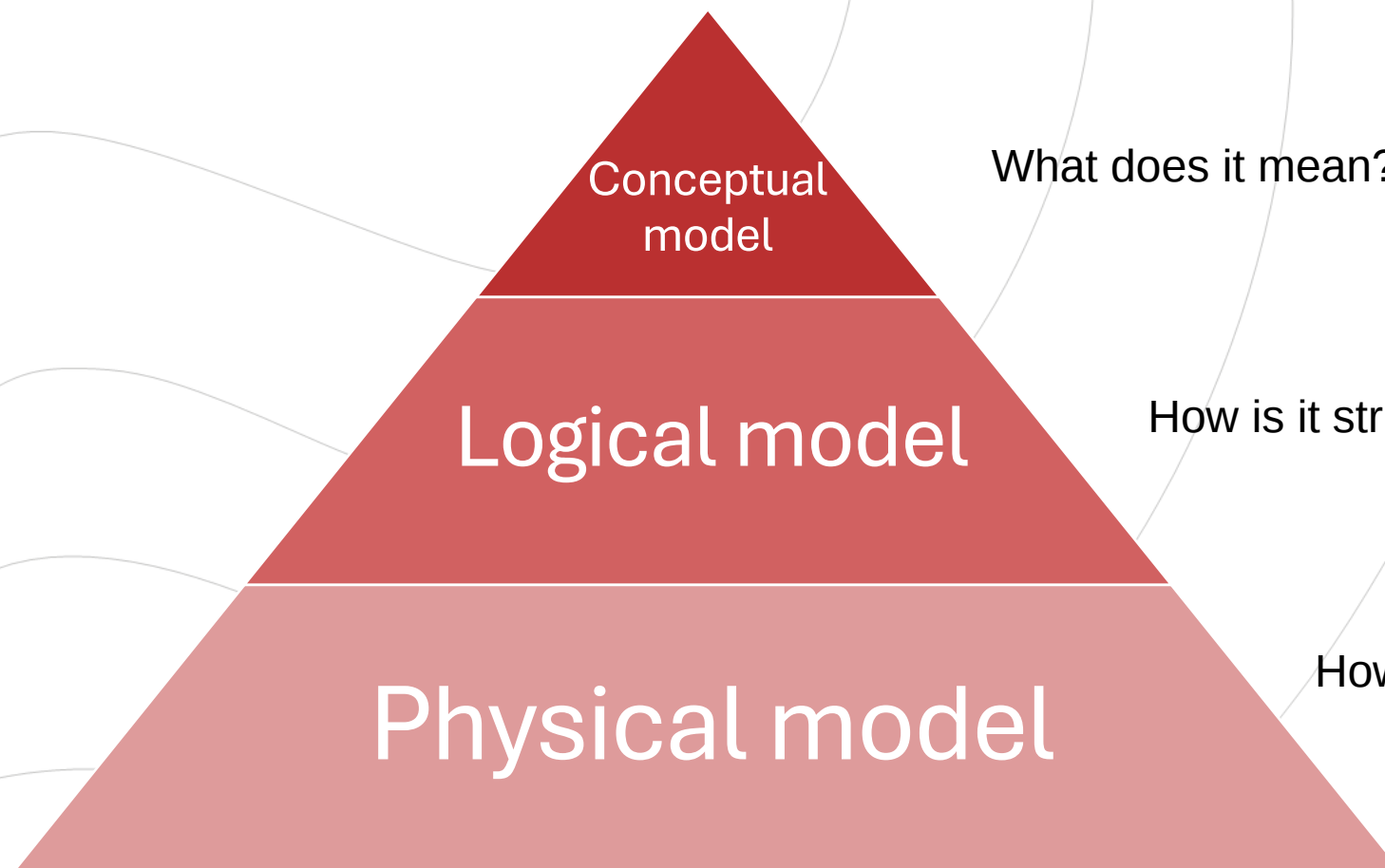


Understand the Domain Community

- What **disciplines** and **datatypes** are in scope?
- Who is the **target community**?
- Identify what the **community needs** for their data description
- Use interviews, domain-specific (meta)data examples
- Review existing repositories and their metadata models, e.g. via [Re3data.org](https://www.re3data.org)



(Meta)data modelling



What does it mean?

UML diagram, RDF classes, properties, vocabularies

How is it structured?

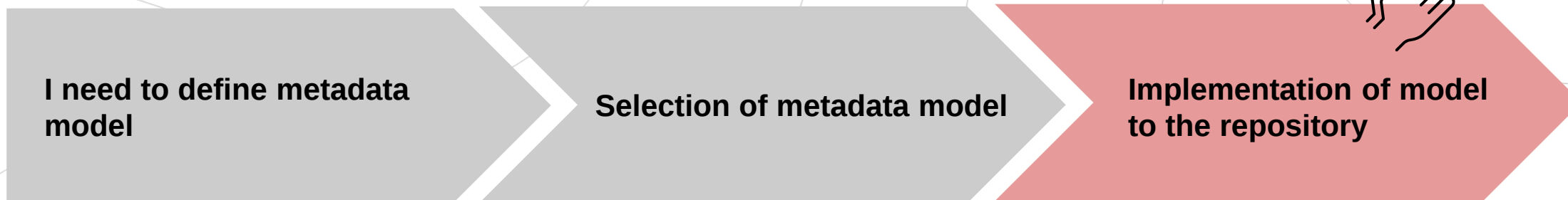
XML Schema, RDF/OWL, JSON schema, logical ER

How is it implemented?

SQL table, Turtle, XML file, JSON(-LD), YAML

Roadmap of metadata

About to build repository:



Next:



Selection of the metadata model

HOW STANDARDS PROLIFERATE: (SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)



What is the metadata/application profile?

- Selection of terms from other models (vocabularies and other profiles) to be used in a specific **context**
- Why? I need to share how the repository will work with metadata
- To whom? Human (community) and Machine
- Artefacts making the profile = That's my language to do so
 - **human** readable specification
 - Diagram of the profile
 - **machine** readable specification
 - Relations to other specifications and terms
 - Xml schema (other schema depending on the needs)
 - Exmaples of values
 - Controlled vocabs

TABLE OF CONTENTS

- Abstract
- 1. Overview
- 2. Class profiles
 - 2.1 Access rights
 - 2.1.1 label
 - 2.2 Address
 - 2.2.1 address area
 - 2.2.2 administrative unit level 1
 - 2.2.3 administrative unit level 2
 - 2.2.4 full address
 - 2.2.5 label
 - 2.2.6 locator designator
 - 2.2.7 locator name
 - 2.2.8 post code
 - 2.2.9 post name
 - 2.2.10 post office box
 - 2.2.11 thoroughfare
 - 2.3 Agent
 - 2.3.1 has contact point
 - 2.3.2 has identifier
 - 2.3.3 name
 - 2.4 Alternate title
 - 2.4.1 has alternate title type
 - 2.4.2 title
 - 2.5 Alternate title type
 - 2.5.1 label
 - 2.6 Application profile
 - 2.6.1 label
 - 2.7 Checksum
 - 2.7.1 checksum value
 - 2.7.2 uses algorithm

Czech Core Metadata Model

15 September 2025

▼ More details about this document

Latest published version:
<https://www.w3.org/todo/>

Editor:
[DataSpecer](#)

Abstract

This file documents **Czech Core Metadata Model**.

§ 1. Overview



§ 2.3 Agent

Profiled class IRI(s)	foaf:Agent
IRI	https://model.ccm.cz/research-data/Agent
Label	Agent
Definition	Any entity carrying out actions with respect to the entities Catalogue and the Catalogued Resources.
Profiles	class profile Agent (https://mff-uk.github.io/specifications/dcat-ap#Agent) Definition: <i>Any entity carrying out actions with respect to the entities Catalogue and the Catalogued Resources.</i> class profile Agent (https://mff-uk.github.io/specifications/dcat-dap#Agent) class Agent (foaf:Agent)

KEEP ALL CONSISTENT!

Explore Existing Metadata Standards

Focus on:

1. Domain specific schemas
 - Full adoption
 - Adoption + editing, extending with domain elements
2. Universal schemas
 - Full adoption
 - Adoption + editing, extending with domain elements
 - CCMM
 - DataCite
 - DCAT-AP-CZ
 - DCAT-AP
 - Dublin Core

Explore Existing Metadata Standards

Sources to explore:

- Metadata Standards Catalog (available at <https://rdamsc.bath.ac.uk/>)
- Fairsharing.org (available at <https://fairsharing.org/>)
- Re3data.org (available at <https://www.re3data.org/>)
- MSCR Registry (available at <https://mscr-release.2.rahtiapp.fi/>)



Metadata „model“ analyses

Does it meet all your needs?

- Description granularity, rich metadata
- Community adoption/needs
- **Interoperability with CCMM** and other identified models/systems
- Crosswalk options
- SW technical options

YES	PARTIALLY	NO
• Reuse directly	• you found model to build your profile upon	• keep searching or extend general model

- Does it have the conceptual and logical layers?
- Does the license allow you to use it?



Source: <https://pixabay.com>

YES, fitting existing model

We choosed fitting existing model

- Analyse this metadata model
- Revise:
 - Cardinalities
 - Formats
 - Controlled vocabularies
- If the format is suitable with repository software
- If the format is **not** suitable with repository software



ready for implementation



convert to right format

YES

- Reuse directly

PARTIALLY

- you found model to build your profile upon

NO

- keep searching or extend general model

PARTIALY fitting existing model

**We choosed existing model,
but we need to do some minor extensions**

- Analyse selected metadata models
- Choose the main model
 - Focus on the choice of the elements from the analyzed models for extensions

YES

- Reuse directly

PARTIALLY

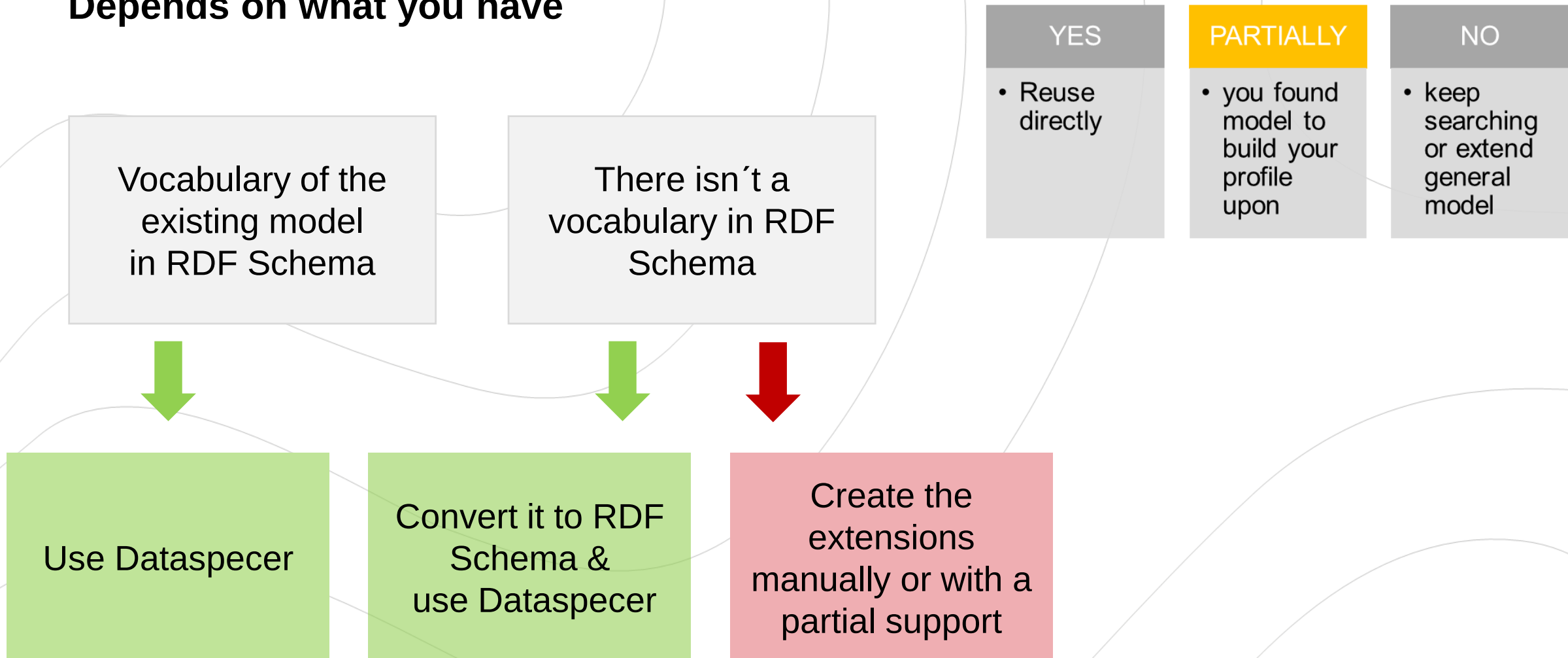
- you found model to build your profile upon

NO

- keep searching or extend general model

How can you do the extension?

Depends on what you have



A tool for management of semantic data specifications

Dataspecer is a tool for management of semantic data specifications, from vocabularies and application profiles to derived technical artifacts such as data schemas, validation rules, APIs and application prototypes.

[Read documentation](#)

[See specifications](#)

[Go to demo instance](#)

Reusable vocabularies

Define new classes, properties and their relationships as well as their

...

Application profiles for narrower contexts

Combine terms from existing

...

Semantically compliant technical artifacts

Use your domain ontology or

...

NO, lack of domain specific models

- Extend CCMM with domain specific attributes
- Use Dataspecer for creating the profile
- Use exports of CCMM on GitHub to start with
 - <https://www.ccmm.cz/>
- Protect your work by committing the progress on any git platform if working in a shared environment

YES

- Reuse directly

PARTIALLY

- you found model to build your profile upon

NO

- keep searching or extend general model



Metadata
výzkumných
dat



Základní model CCMM O nás Novinky

CCMM (Czech Core Metadata Model for Research Data) je základní metadatový model pro popis výzkumných dat v prostředí České republiky. Cílem CCMM je stanovit obecné prvky pro popis datových sad, které budou použitelné v různých datových repozitářích, a zajistit interoperabilitu základních metadat napříč obory.

Rychlé odkazy

Základní model
CCMM

Řízené slovníky

Tvorba a správa
metadatových
profilů

FAQ

Novinky

Workshop pro práci s CCMM a
nástrojem Dataspecer

Seminář CCMM: model,
principy, nástroje

Finalize the profile

- Export from Dataspecer  or create profile manually
- Model validation - with defined needs and community should be in place
- Publication: GitHub or similar platforms, www
 - Example: CCMM GitHub <https://github.com/techlib/CCMM>
- Define license

BUT...



Source: <https://pixabay.com>

Implementation to the repository

- Which **software** from the **NRP platform** do you use (Invenio, DSpace, ARL)?
 - Technical options (metadata format, hierarchy, vocabularies...)
 - Consultation with the NRP repository system specialists
- Affects physical model
- It may also affect logical model even though it shouldn't
- Testing and validation with defined needs and the community should be in place

Published your profile 1.0.0

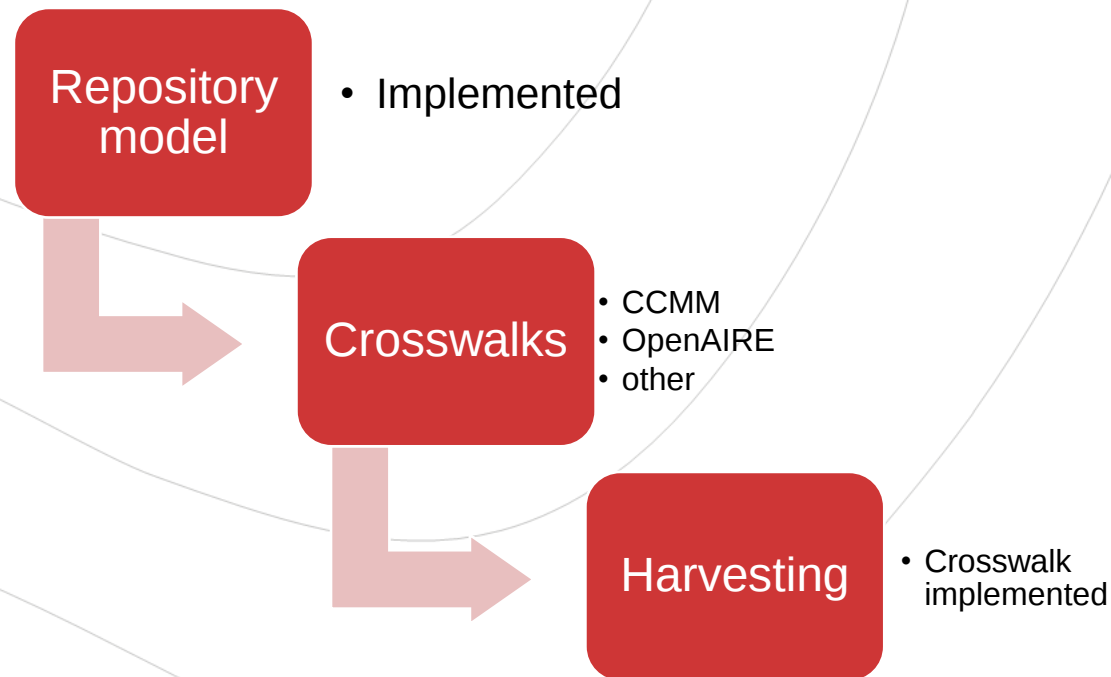
- Register to NKR
- Verify interoperability – export your metadata



Source: <https://pixabay.com>

Interoperability in practice

- Map to CCMM for NMA indexing (aggregation and harvesting)
- Create and test export formats for OAI-PMH (XML) and other API
- Some crosswalks will be provided by metadata support



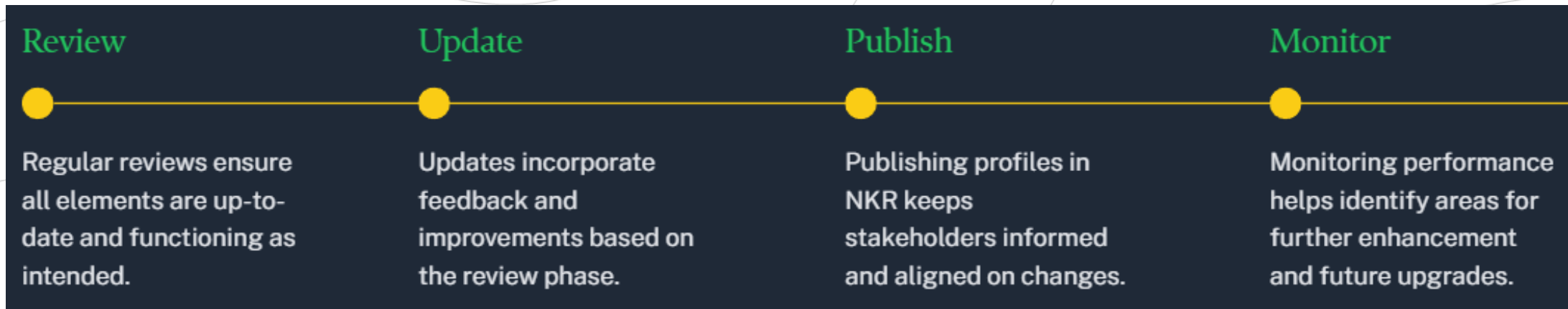
Mapping (crosswalks)

- Looking for existing mappings <https://mscr-release.2.rahtiapp.fi/>
 - Mappings from NTK will be created
 - CCMM x OpenAIRE
 - CCMM x RIV
 - CCMM x DataCite
- New crosswalks register to NKR that will push them to MSCR
 - <https://mscr-release.2.rahtiapp.fi/>



Maintenance, Versioning, and Documentation

- Community feedback
- Keep version control (e.g. SemVer „MAJOR.MINOR.PATCH.“ 1.0.0., 1.1.0)
- Document changes and rationale (e.g. using Github or other systems)
- Plan for regular updates (and publish information)
 - Watch the changes and new versions in used schemas!



You are not alone

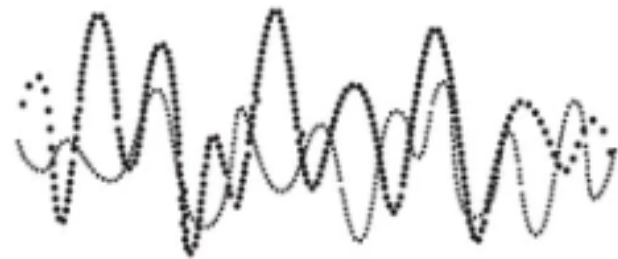
- **Consultations**
 - managers of existing repositories
 - NRP repository system specialists
 - NTK metadata department (metadata@techlib.cz)
 - Repository specialist of the used SW



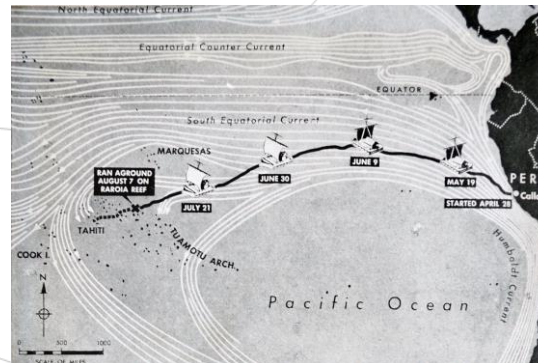
Source: <https://pixabay.com>

Goal Seeking balance

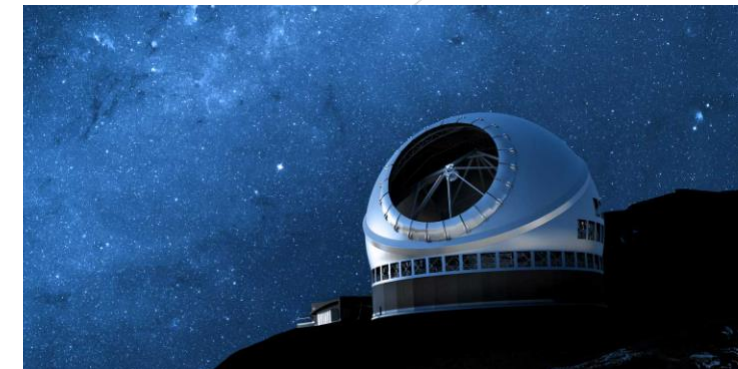
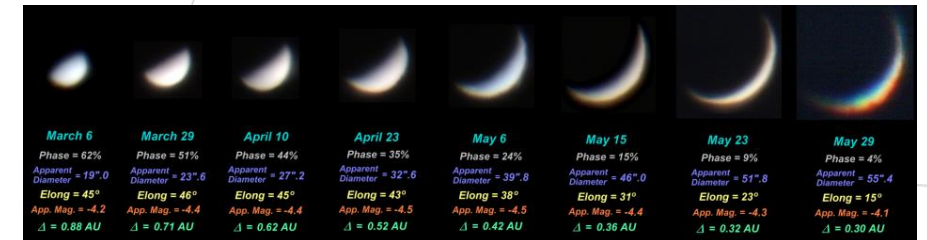
- Detailed but still possible to fill
- Vocabularies – help or limit
- What have to part of the model, when reference is good enough
- Interoperability with disciplinary precision



Gustav Holst, Opus 32, 2nd Movement
 Source: Sydney Symphony Orchestra, 2023



Kon Tiki Expedition and Polynesian Navigation,
 Source: Kontiki Museum in Oslo, 2016



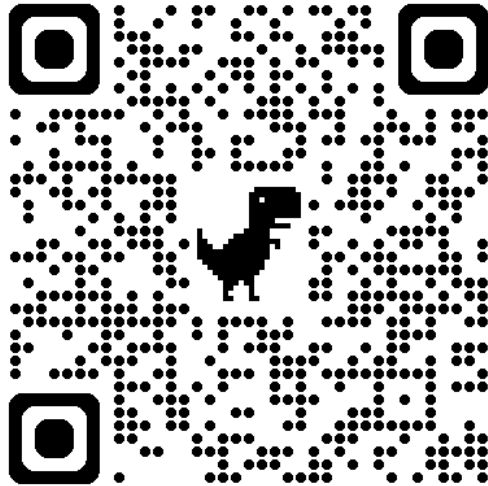
TMT Laser Guide Star Facility, Hawai'i
 Source: www.tmt.org, 2025

HOW STANDARDS PROLIFERATE: (SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)

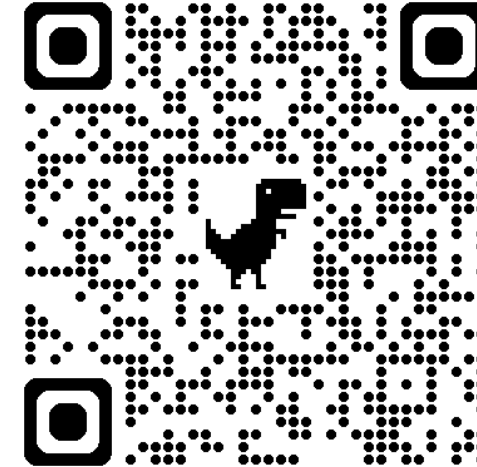


Resources and Support

- **NTK Metadata Unit:** metadata@techlib.cz
- **CCMM documentation and issue submissions:** <https://github.com/techlib/CCMM>
- **Dataspecer:** <https://dataspecer.com/>



www.ccmm.cz/en/



<https://techlib.github.io/CCMM/en/>

Thank you!