NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

Czech Academic and Research Discovery Services
CARDS

National Czech Programme | eosc

Metadata modelling principles

Michal Med, Cesnet

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

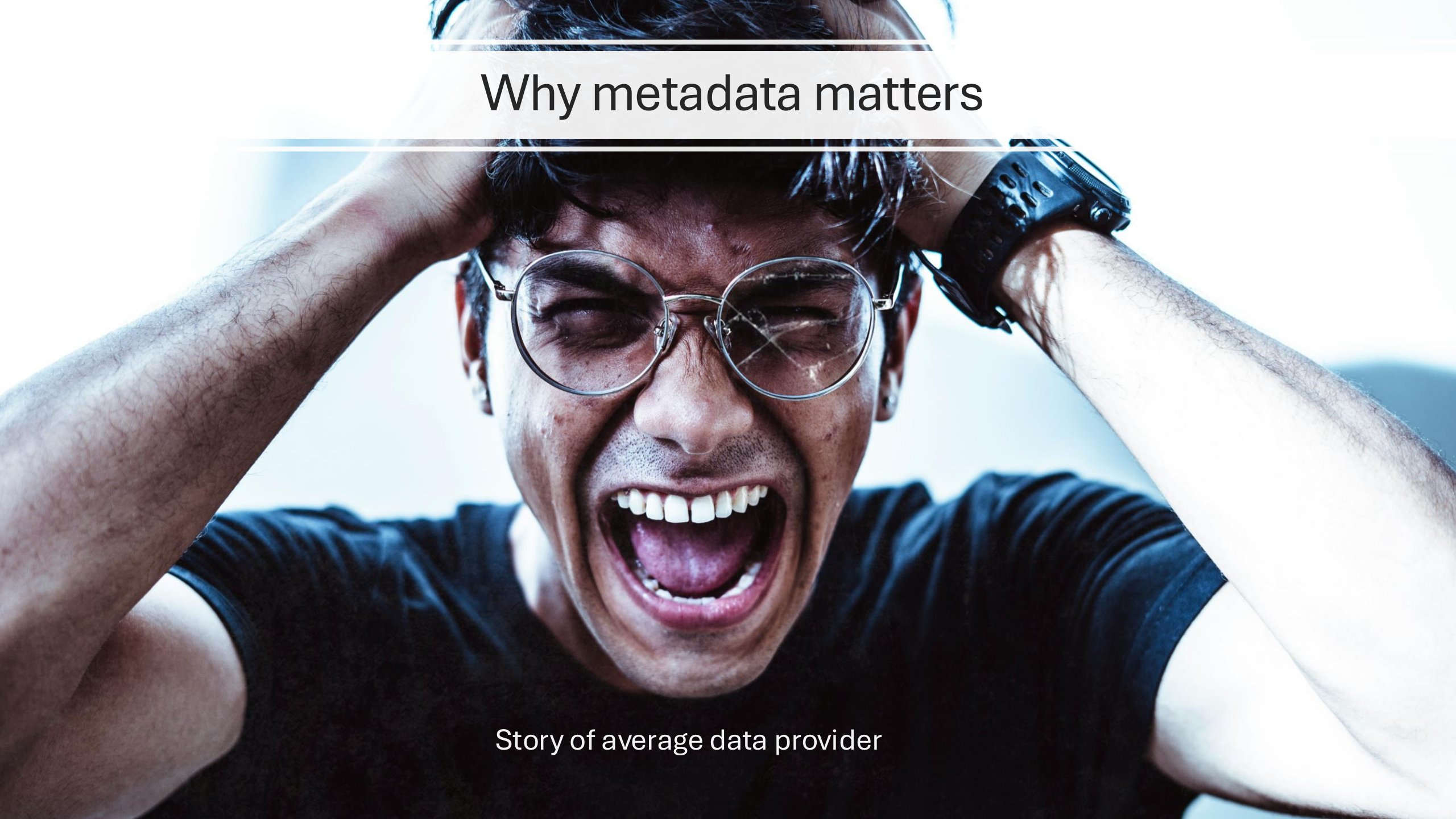
Seminar CCMM: Model, Principles, Tools

26th September 2025

Co-funded by the European Union

MINISTRY OF EDUCATION, YOUTH AND SPORTS

Why metadata matters



Story of average data provider

- No or bad metadata = Hard to find / interpret / reuse data
- Key in FAIR - https://www.go-fair.org/fair-principles/
  - Findability
  - Accessibility
  - Interoperability
  - Reuse
- Data lake vs. data swamp

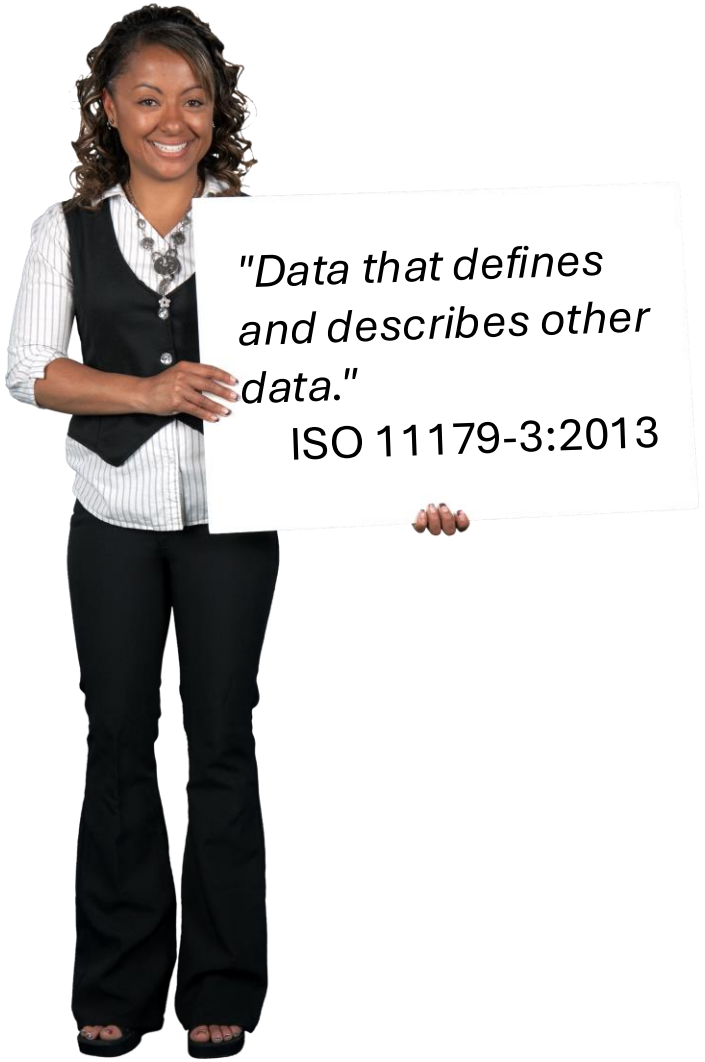
What is

HAZY GALAXY

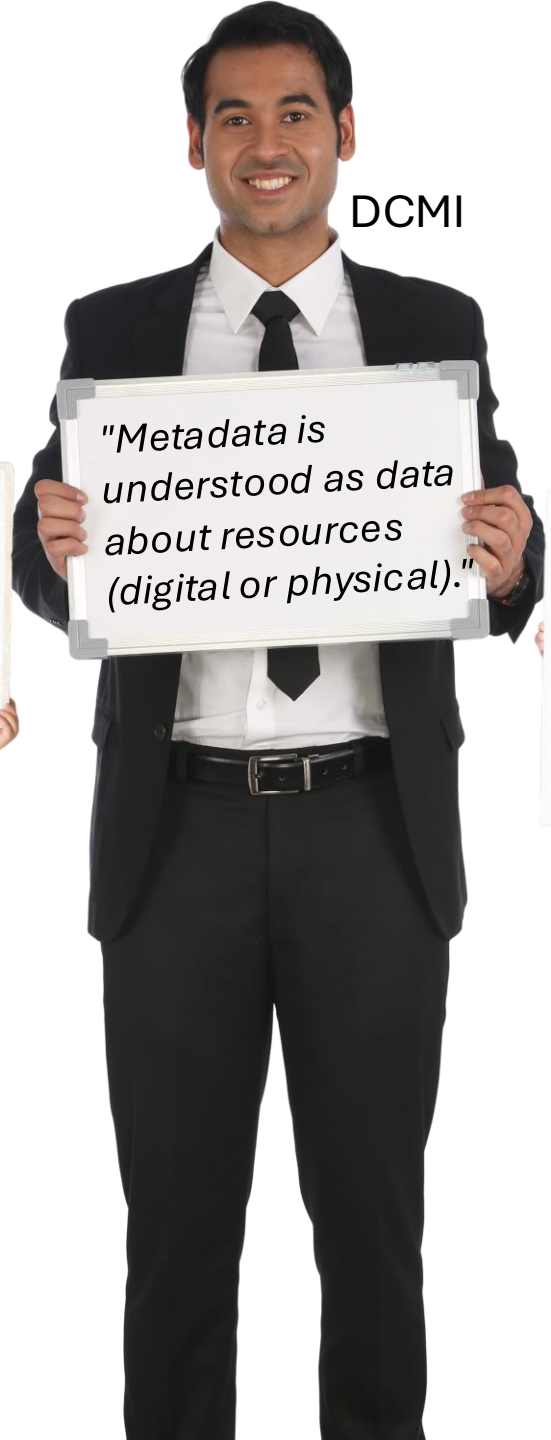
SESSION IPA

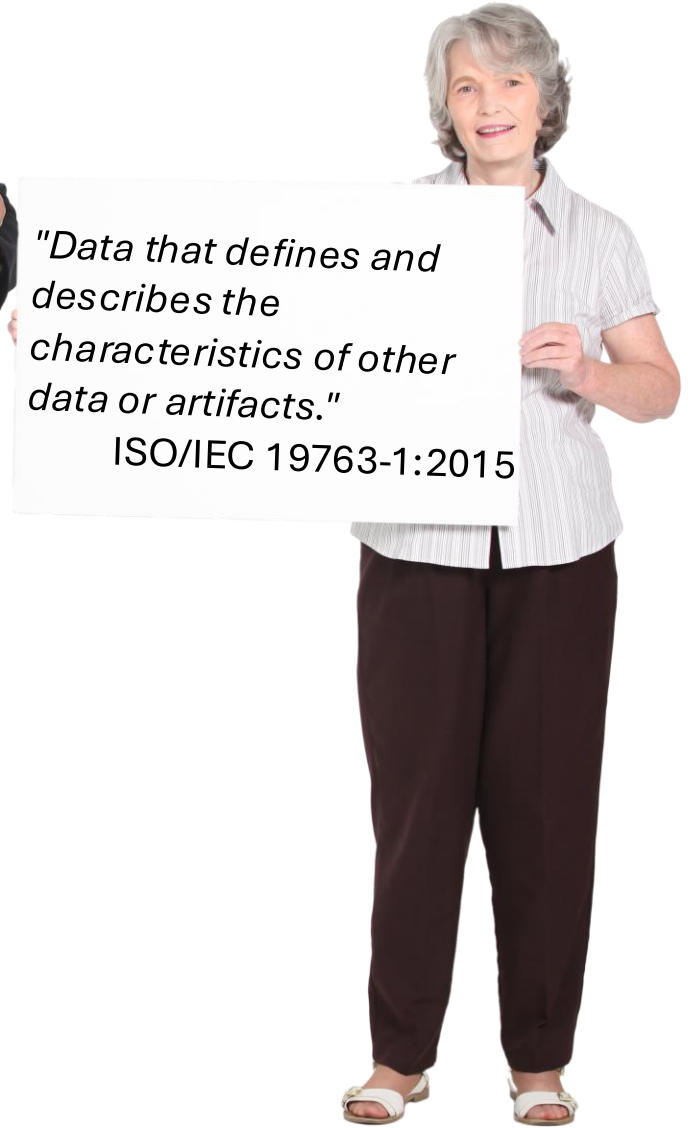
FERMENTATION:
Top
STYLE:
India Pale Ale (IPA)
ABV: 4,9 %
PLATO: 12

WATER:
Groundwater – infiltration from Jizera river
MALTS:
Wheat
HOPS:
Cascade, Galaxy, Sabro


"Data that defines and describes other data."
ISO 11179-3:2013


"Data about data."
ISO/IEC 2382-1:1993


DCMI
"Metadata is understood as data about resources (digital or physical)."


"Data that defines and describes the characteristics of other data or artifacts."
ISO/IEC 19763-1:2015

Metadata types

Descriptive	Structural	Administrative
Title	Formats	Rights
Author	Checksums	Provenance
Description	Relationships	Technical

Metadata Profile	Use Case (with Examples)
Dublin Core	Describing digital objects in **institutional repositories** (DSpace, Omeka), **archives**, and **digital libraries**.
MARC / MARC21	Bibliographic metadata in **library catalogs** (OCLC WorldCat, Koha, national libraries).
MODS	Rich bibliographic records for **digital libraries** and **archives** (e.g., Library of Congress collections).
PREMIS	**Digital preservation** metadata to track provenance, fixity, rights, and preservation actions (Archivematica, Rosetta).
DCAT	Metadata for **open data portals** and **research data catalogues** (e.g., data.gov, European Data Portal, Dataverse).
DataCite	Describing **research datasets** and assigning DOIs (Zenodo, Figshare, institutional repositories).
Schema.org	**Structured data for the web**, enabling dataset discovery by search engines (Google Dataset Search, JSON-LD in websites).
VRA Core	Describing **artworks, architecture, and visual resources** (museum and art databases like CONTENTdm, CollectiveAccess).
ISO 19115	**Geospatial metadata** for datasets, maps, satellite imagery, and GIS services (GeoNetwork, INSPIRE directive portals).

Examples

filmroll	2025_07_27_Plzeň
image id	1138
group id	1138
filename	IMG_4558.CR3
version	0
full path	/home/michal/Pictu... Plzeň/IMG_4558.CR3
local copy	no
import timestamp	Po 15.9.2025 20:47:14
change timestamp	-
export timestamp	-
print timestamp	-
flags	0...r..a....l
model	EOS R10
maker	Canon
lens	Canon RF-S 55-210mm F5-7.1 IS STM
aperture	f/6.3
exposure	1/125
exposure bias	+0.00 EV
focal length	55.0 mm
35mm equiv focal length	88.7 mm
crop factor	1.6
focus distance	9.10 m
ISO	400
datetime	Ne 27.7.2025 11:48:28
width	6000 (6192)
height	4000 (4060)
export width	4000
export height	6000
title	-
description	-
creator	-
publisher	-
rights	-
notes	-
version name	-
latitude	-
longitude	-
elevation	-
tags	-
categories	-

Sakrální objekty

Institut plánování a rozvoje hlavního města Prahy

§

Využití území | Urbanismus/Urban planning

Datová vrstva obsahuje data o sakrálních objektech na území Prahy, jak je IPR Praha identifikoval vlastním sběrem dat z veřejně dostupných zdrojů. Bodová vrstva je vztažena na adresní body sakrálních objektů.

Dataset theme
Education, culture and sport

Spatial coverage
Praha
Česká republika
https://publications.europa.eu/resource/authority/place/CZE_PRG
https://sws.geonames.org/3067695/

Documentation
Show documentation

Contact point
Institut plánování a rozvoje hl. m. Prahy

Frequency
other

Distributions of the dataset

ArcGIS Hub Dataset
HTML

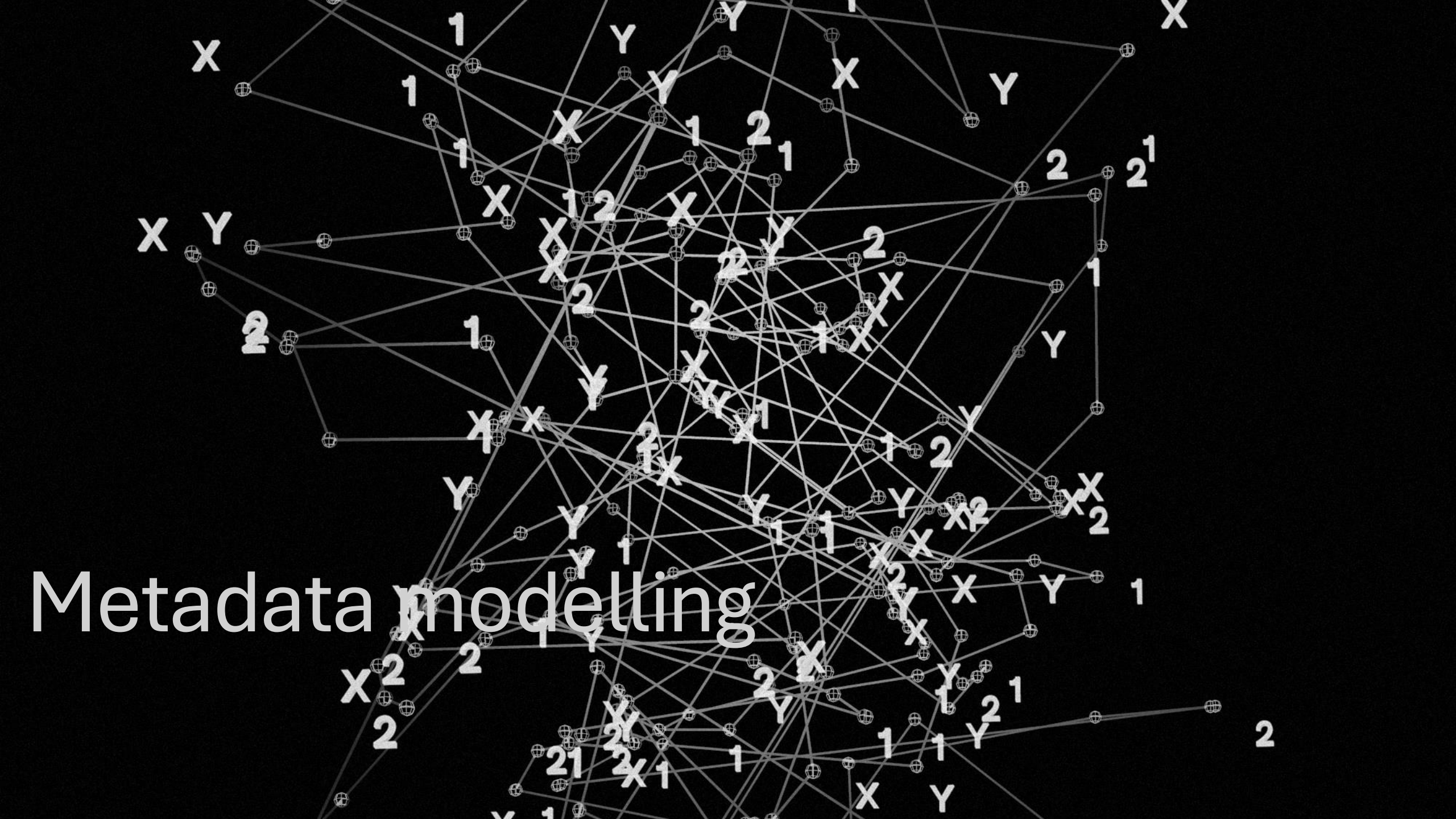
Terms of use	Downloadable file
CC BY 4.0 Institut plánování a rozvoje hl. m. Prahy	Download text/html Media type
Not copyrighted Copyrighted database	

ZIP
ZIP

Terms of use	Downloadable file
CC BY 4.0 Institut plánování a rozvoje hl. m. Prahy	Download application/zip Media type
Not copyrighted Copyrighted database	

ArcGIS GeoService
JSON

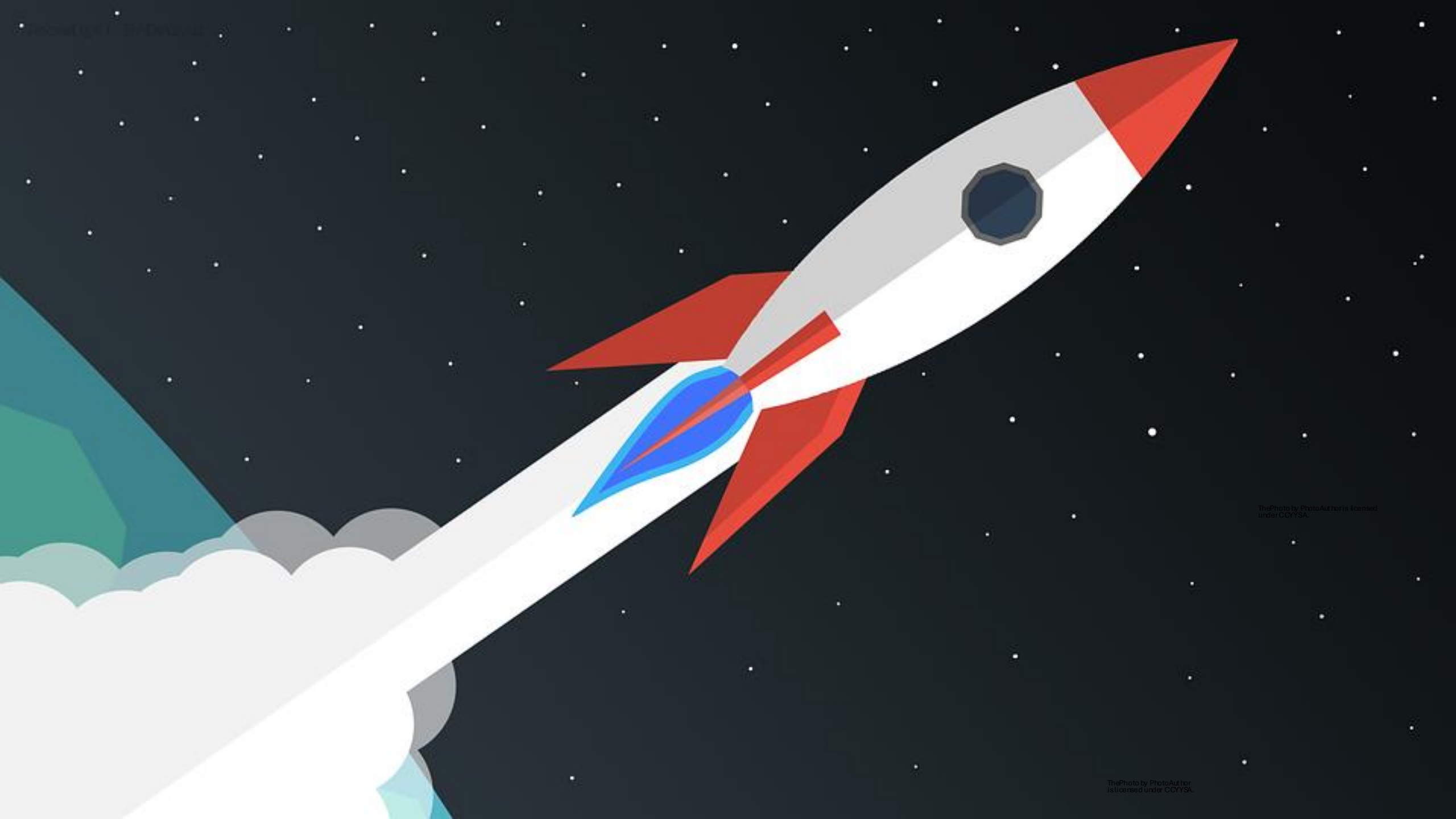
Terms of use	Downloadable file
CC BY 4.0 Institut plánování a rozvoje hl. m. Prahy	Download application/json Media type
Not copyrighted Copyrighted database	



Metadata modelling

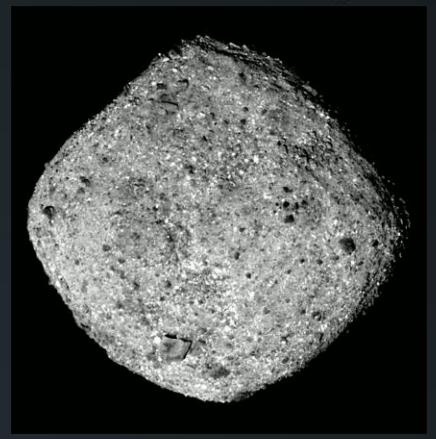
Purpose of metadata modelling

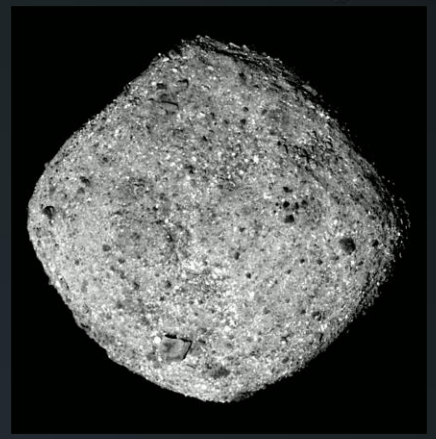
- Consistency of resource description
- Interoperability across systems
- Extensibility & scalability



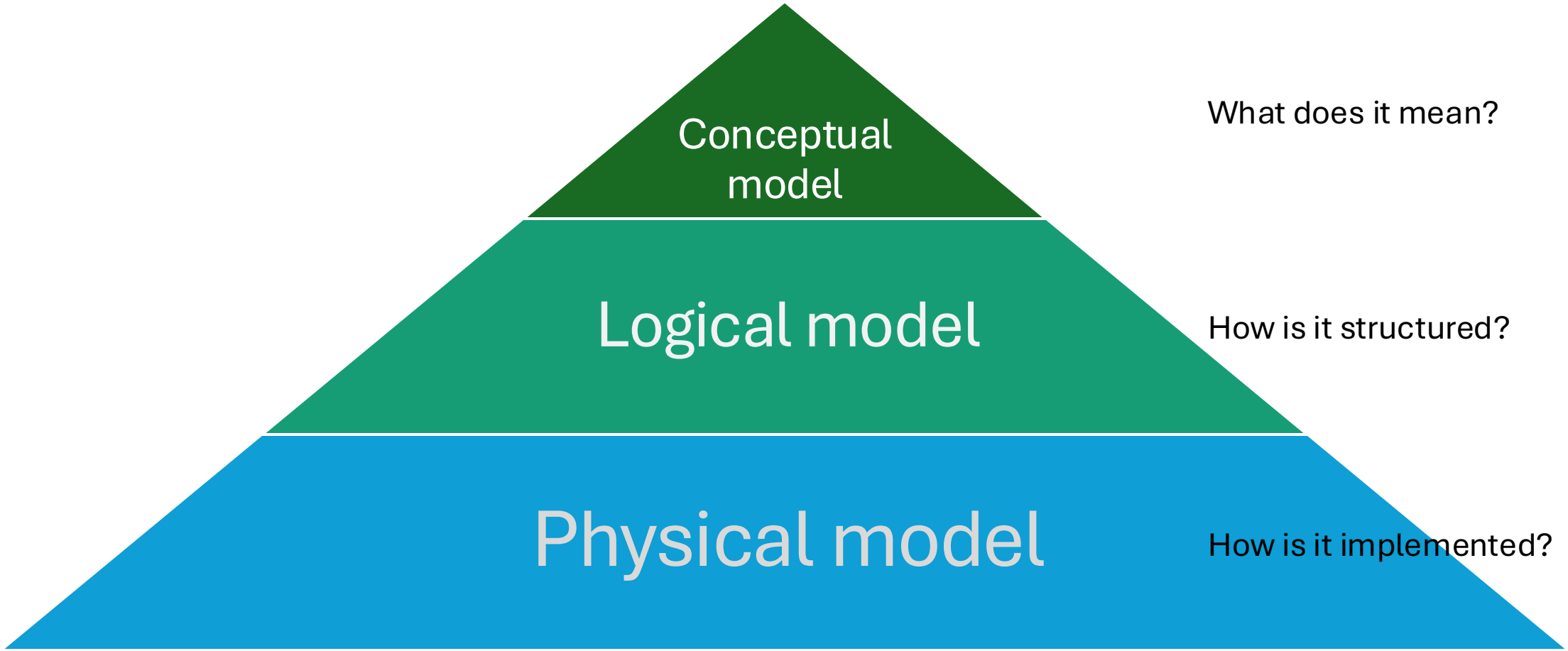
ThePhoto by PhotoAuthor is licensed under CCYSA.

ThePhoto by PhotoAuthor is licensed under CCYSA.

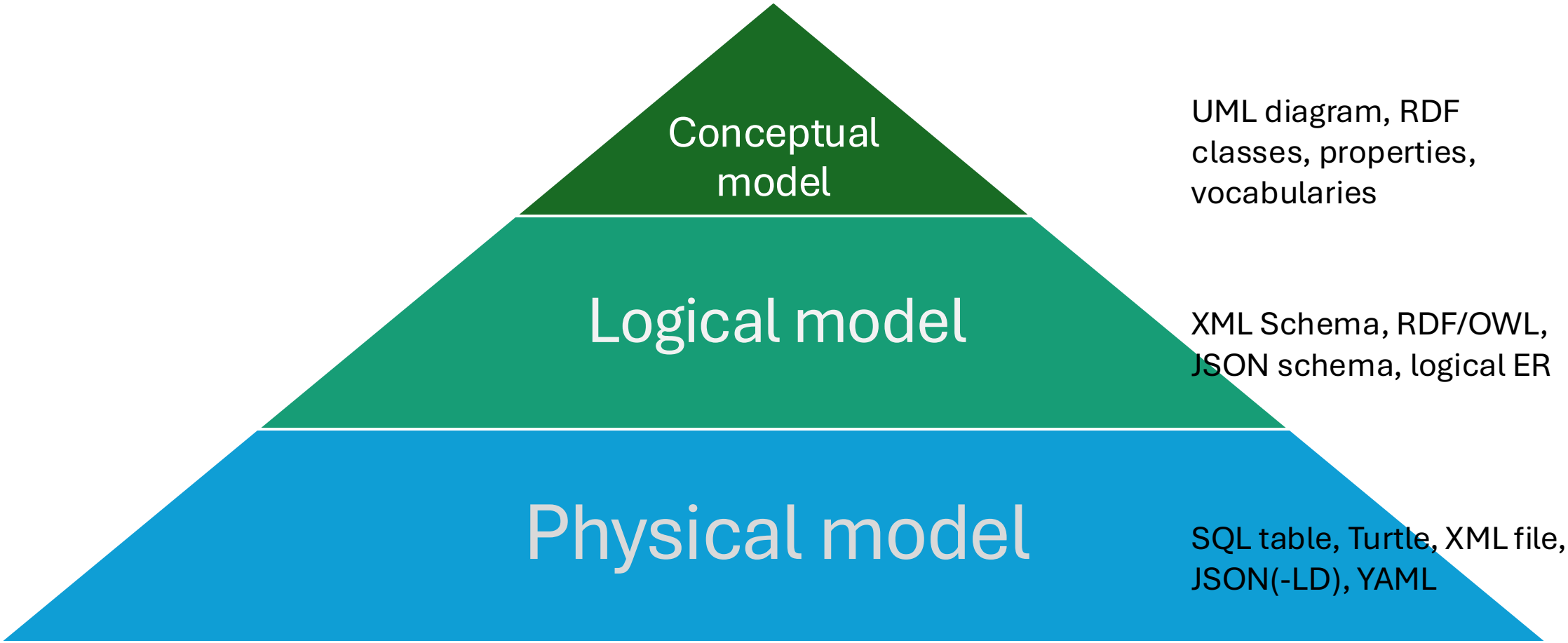


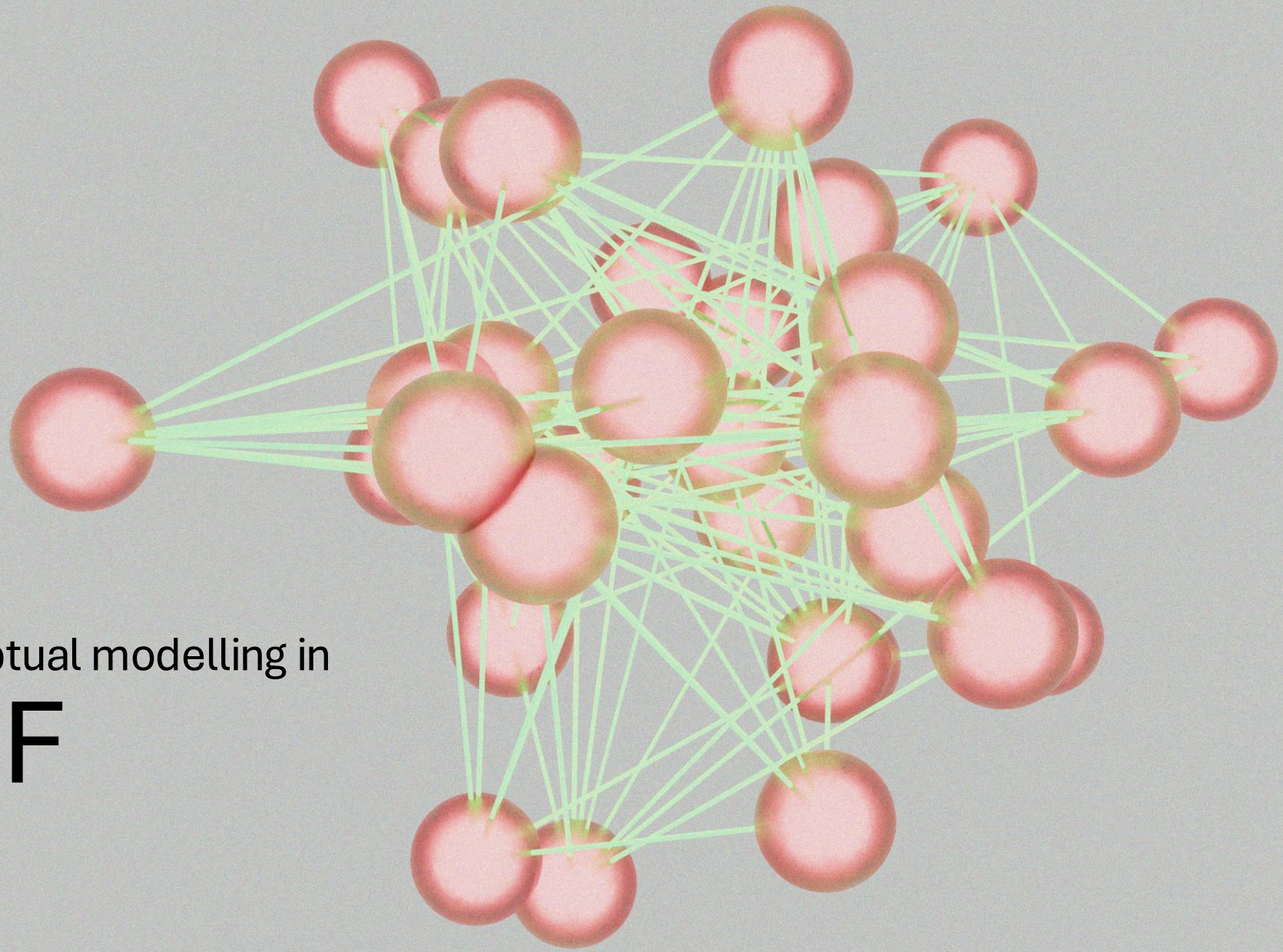


(Meta)data modelling



(Meta)data modelling



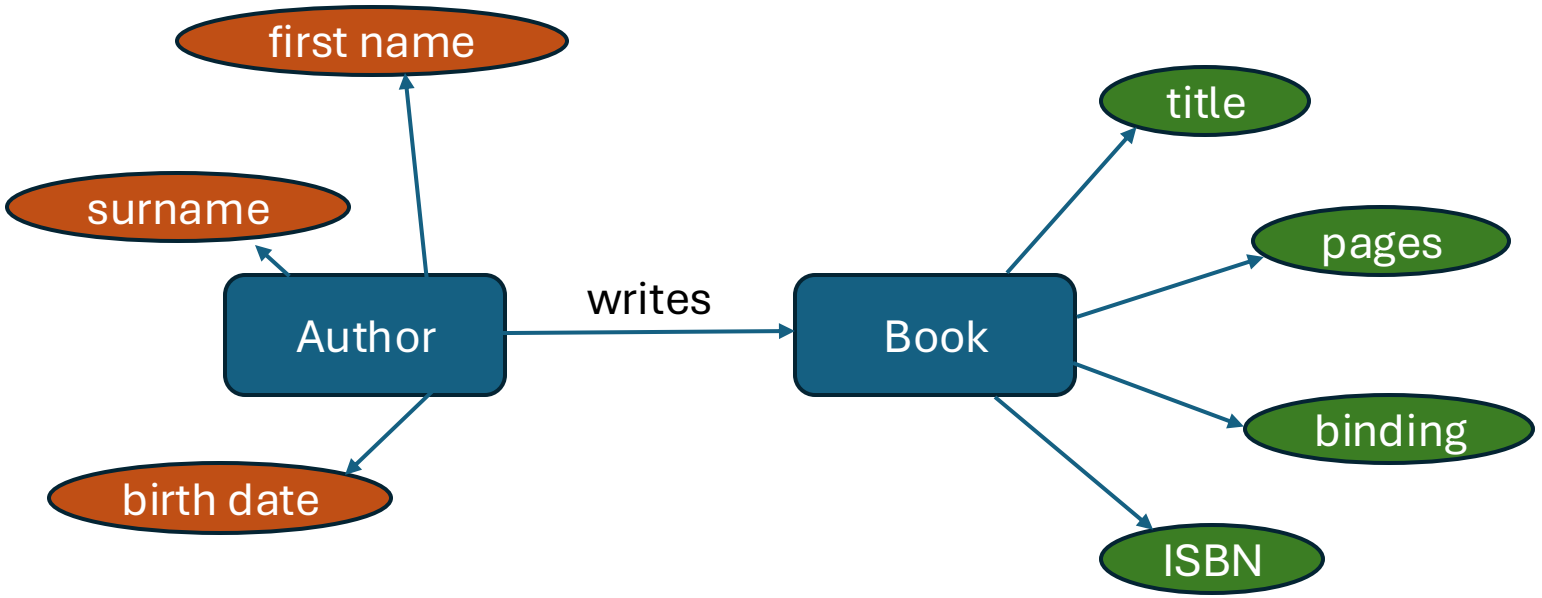


Conceptual modelling in

RDF

Conceptual model

- Captures meaning and semantics
- Brings common understanding among stakeholders

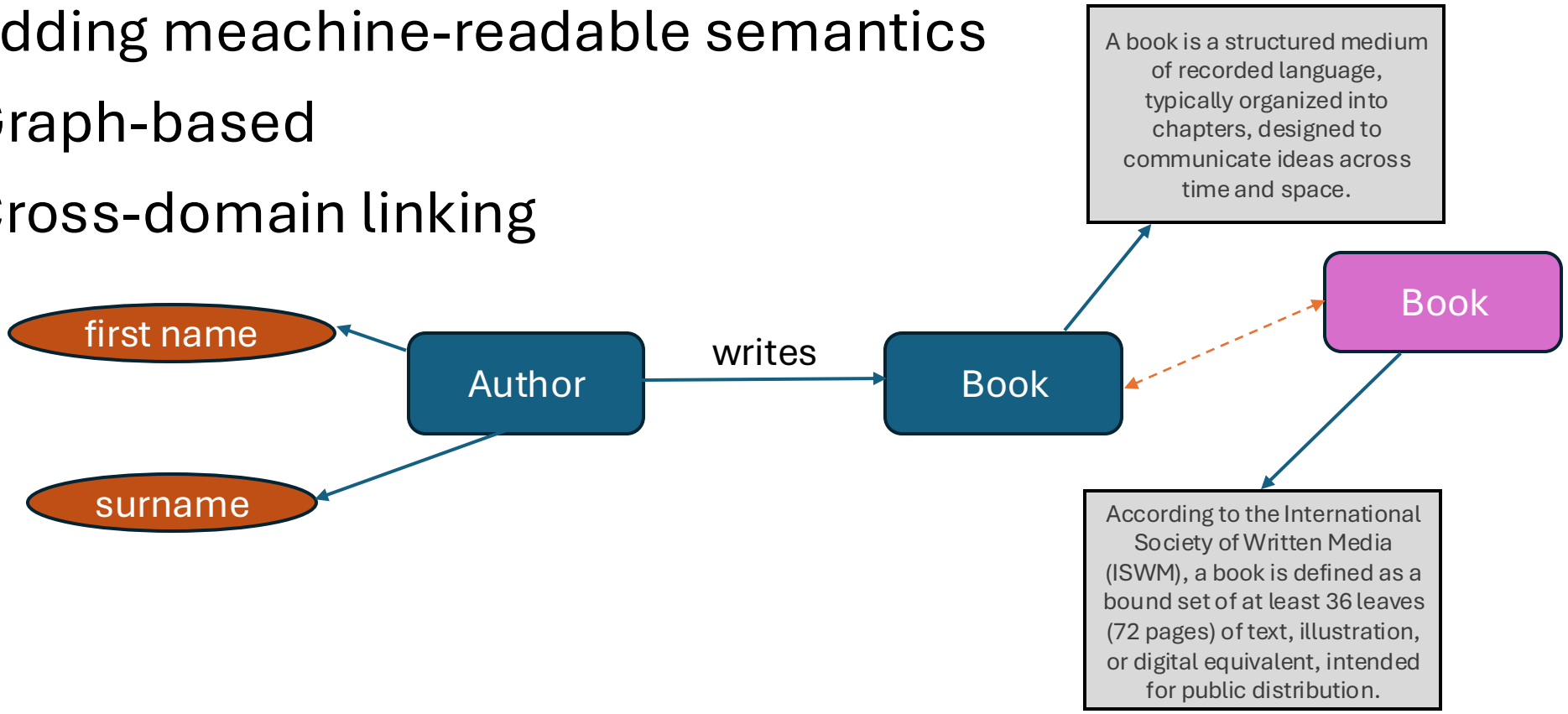


What is RDF

- A W3C standard for representing data on the web
- Based on triples:
  - Author writes a book
  - Author has first name Stephen
  - Author has surname King
  - Pet Sematary is a book
  - Author Stephen King wrote a book Pet Sematary

Using RDF for conceptual modelling

- Adding meachine-readable semantics
- Graph-based
- Cross-domain linking



RDF principles

✅ **Interoperability** — shared standards (Dublin Core, Schema.org, SKOS, DCAT).

✅ **Extensibility** — extend vocabularies without breaking existing data.

✅ **Reusability** — align with existing ontologies rather than reinventing.

RDF principles

✅ **Linked Data** — easy integration across domains (libraries, archives, research data).

✅ **Reasoning** — semantic queries (SPARQL) and inference.

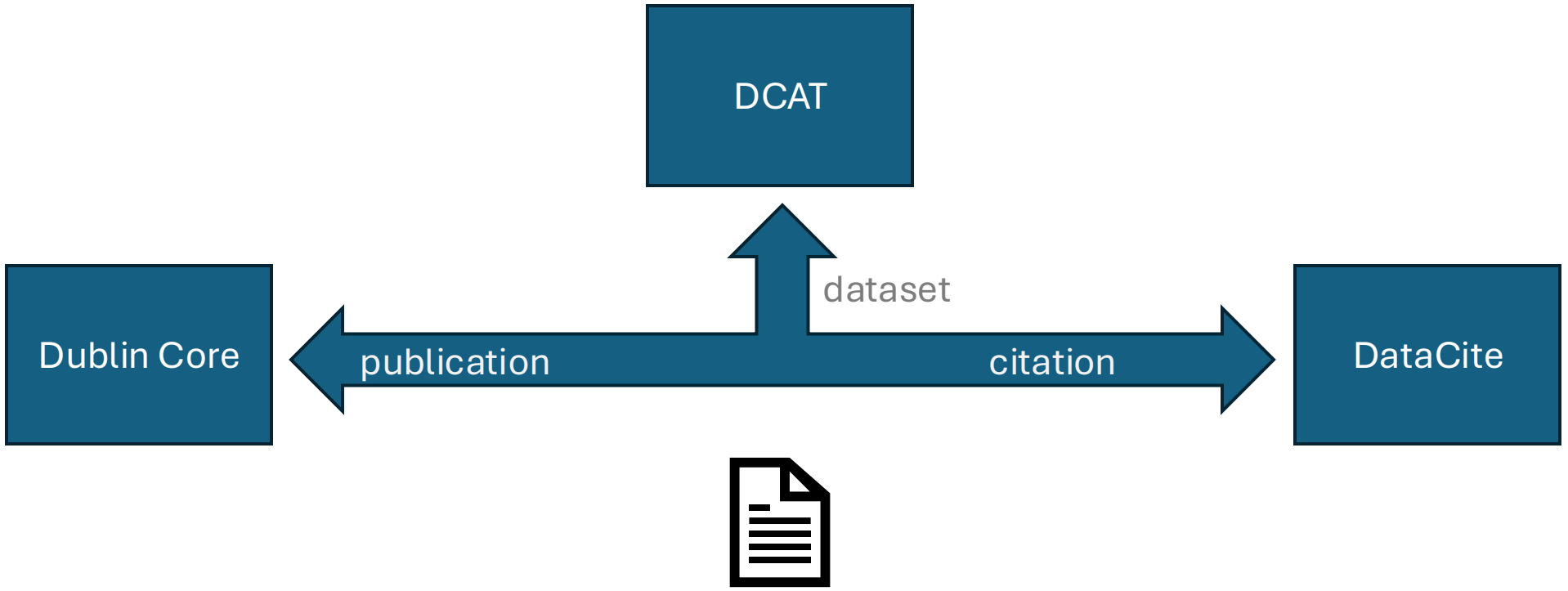
✅ **Technology independence** — RDF can be implemented in databases, APIs, or documents.

Ontologies & Vocabularies

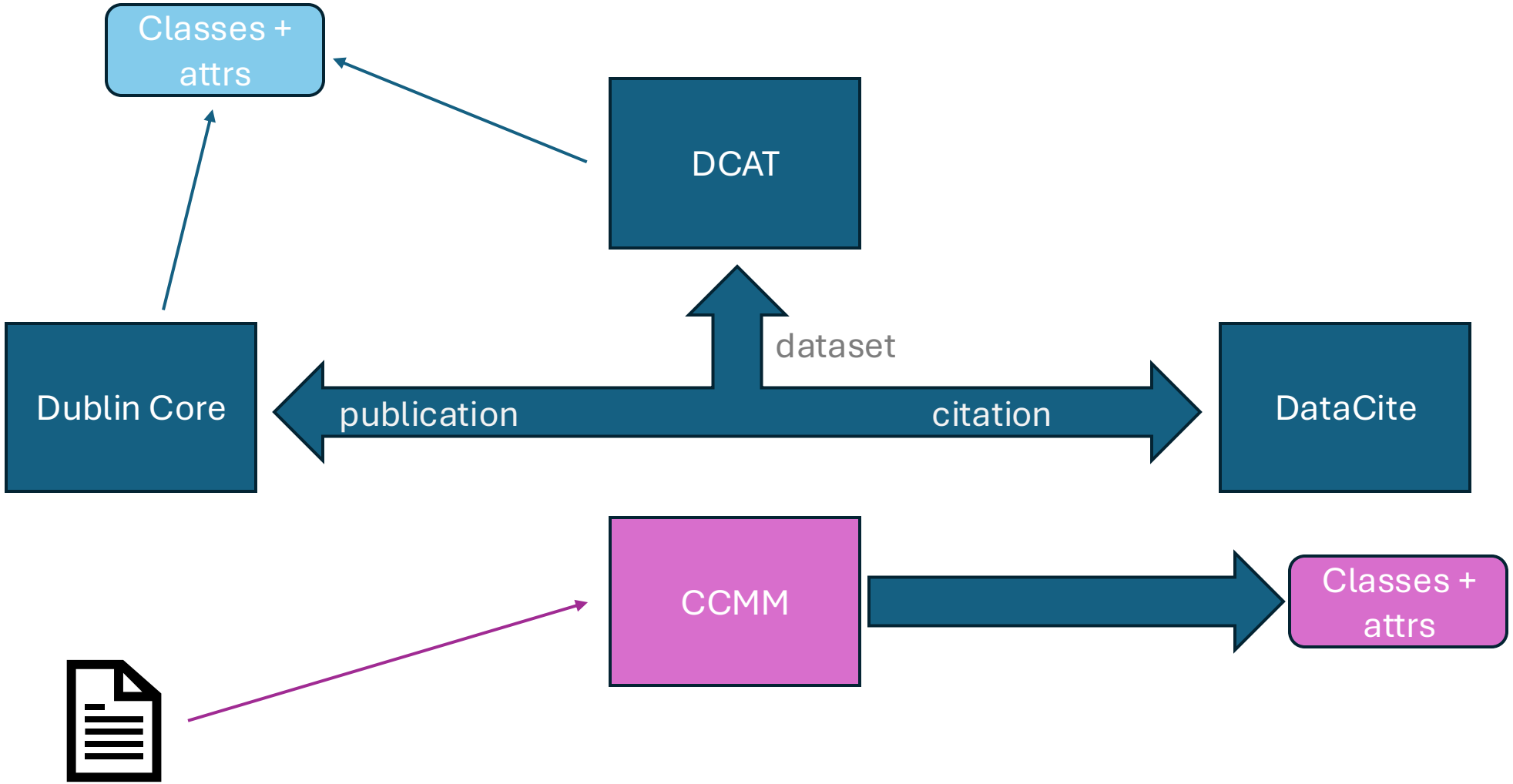
- Provides shared semantic anchors
- Facilitate cross domain mapping and linking

- **Dublin Core** – general resource description
- **FOAF** – people and organizations
- **SKOS** – controlled vocabularies
- **DCAT** – data catalogues and data sets
- **PROV-O** – provenance metadata

Interoperability & Extensibility



Interoperability & Extensibility



Comparison

Aspect	RDF / conceptual model	Pure technical model
Flexibility	Graph based, open-world, extendable	Rigid, schema-bound, closed-world
Interoperability	Uses shared ontologies	Siloed, difficult to cross-domain mapping
Evolution	Can add new predicates and classis without breaking	Schema changes disruptive
Semantics	Rich semantics meaning using ontologies	Semantics implicit in system design
Queries	SPARQL, federated queries across datasets	SQL/XPath limited to one system
Use case	Ideal for cross-domain metadata, linked data	Suitable for performance-intensive structured storage
Performance	Slower, complex queries costly	Optimized, mature query engines
Learning curve	Requires ontology knowledge, RDF syntax	Familiar to most developers
Adoption	Strong in libraries, archives, research, gov data	Dominant in enterprise

RDF Coneptual model summary

Strengths of RDF modelling

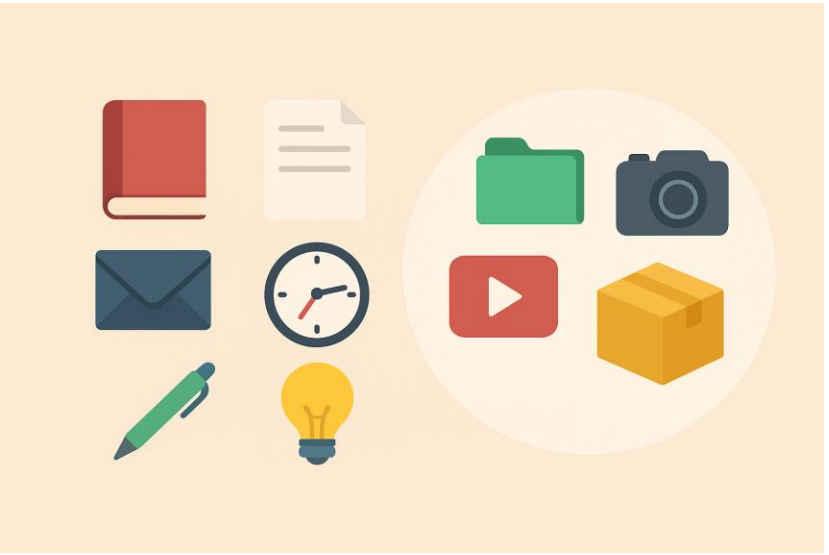
- Semantic richness
- Flexibility & Extensibility
- Interoperability & Linking
- Technology independence

Disadvantages of RDF in practical use

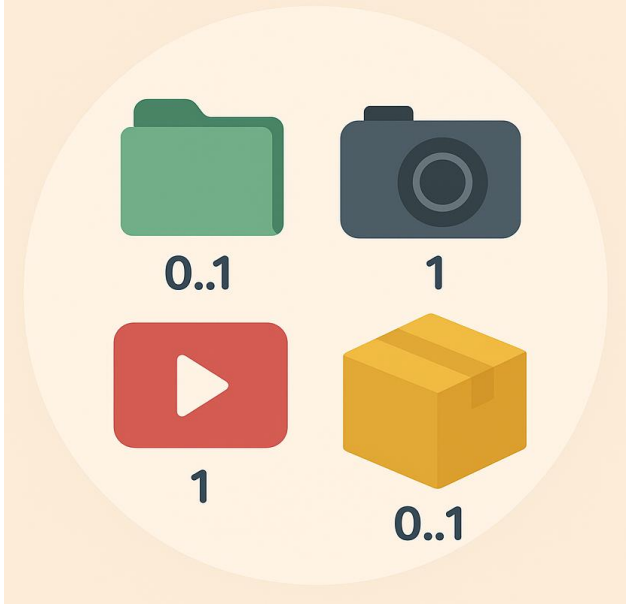
- Complexity of standards
- Performance limitations
- Steep learning curve
- Tooling fragmentation
- Limited industry adoption

Logical & technical models

Metadata Profile

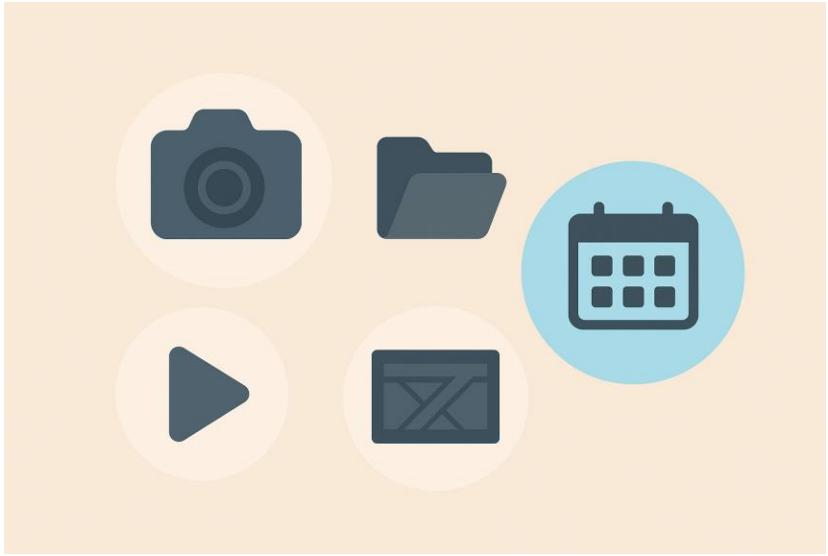


Selected elements



Cardinalities and requirements

Constraints and rules



Community extensions

Building profile

Define scope

- Which concepts are relevant?
- Are the definitions narrow enough?

Select vocabularies

- Dublin Core
- Data Cite
- DCAT
- PROV-O

Specify constraints

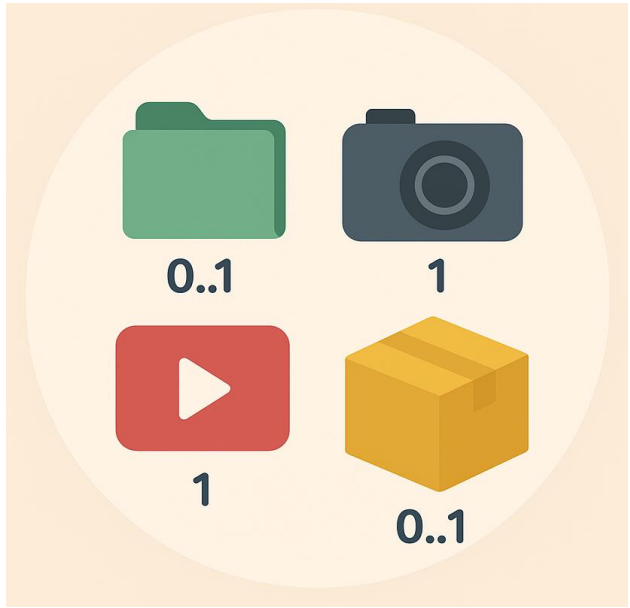
- Cardinality
- Data type
- Controlled vocabularies

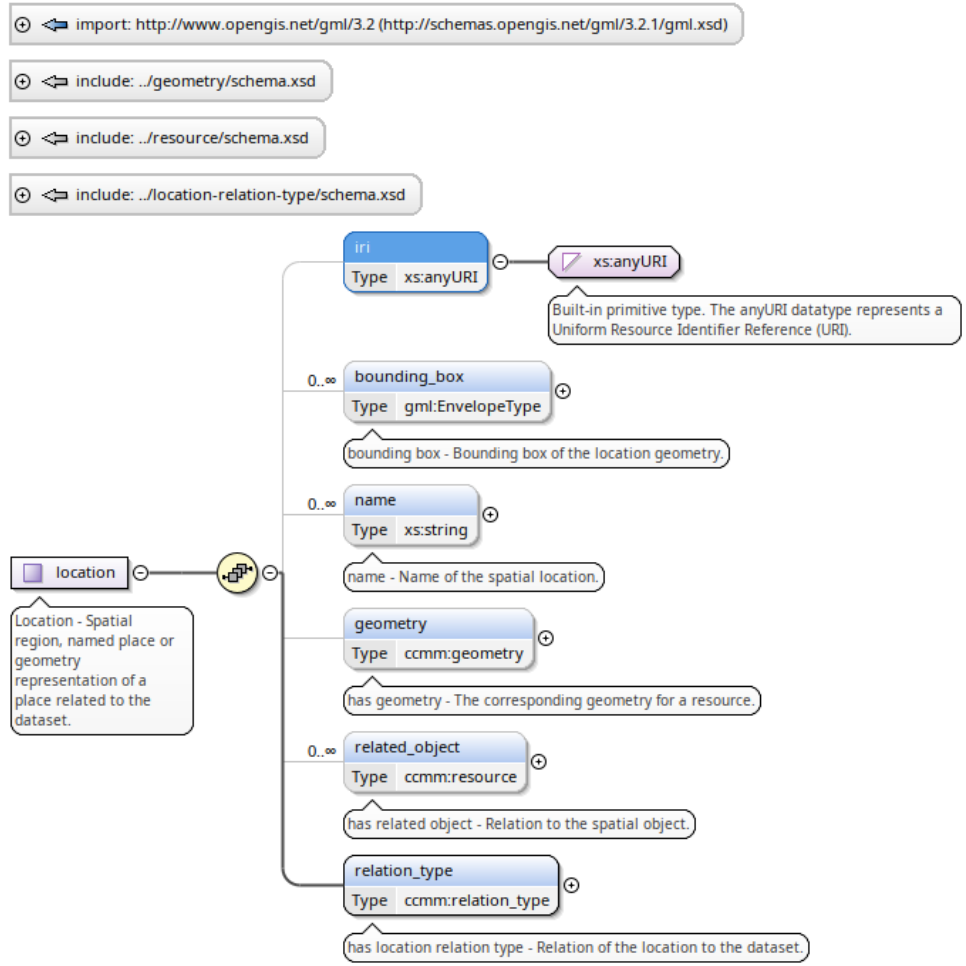
Output as formal profile document

- SHACL
- HTML
- Etc..

Technical Schema

```
"required": [
  "metadata_identification",
  "identifier",
  "title",
  "qualified_relation",
  "publication_year",
  "time_reference",
  "terms_of_use",
  "subject"
],
"properties": {
  "id": {
    "type": "string",
    "format": "iri"
  },
  "type": {
    "oneOf": [
      {
        "const": "Datová sada"
      },
      {
        "type": "array",
        "contains": {
          "const": "Datová sada"
        },
        "items": {
          "type": "string"
        }
      }
    ]
  },
  "metadata_identification": {
    "title": "is described by",
    "description": "Inverse relation between dataset and metadata record.",
    "type": "array",
    "minItems": 1,
    "items": {
      "$ref": "../metadata-record.json"
    }
```





Different serializations, but same conceptual semantics.

Detail of: **label**

https://model.ccmm.cz/research-data/File.label

Direct profile of:	label (http://www.w3.org/2000/01/rdf-schema#label)
The original profiled entity:	label (http://www.w3.org/2000/01/rdf-schema#label)
Type:	relationship profile (attribute)
Definition:	File label
Domain:	File : [0..*]
Datatype:	rdf:langString (http://www.w3.org/1999/02/22-rdf-syntax-ns#langString)

```
"label": {
  "title": "label",
  "description": "File label",
  "type": "array",
  "items": {
    "title": "Language tagged string",
    "type": "object",
    "required": [
      "@value",
      "@language"
    ],
    "properties": {
      "@value": {
        "title": "Text v daném jazyce",
        "type": "string"
      },
      "@language": {
        "title": "Jazyk textu",
        "type": "string"
      }
    }
  }
}
```

File (file)

— **label** (label) : Language tagged string [0..*]

```
<xs:element minOccurs="0" maxOccurs="unbounded" name="label"
  <xs:annotation>
    <xs:documentation xml:lang="en">
      label - File label
    </xs:documentation>
  </xs:annotation>
  <xs:complexType>
    <xs:simpleContent>
      <xs:extension base="xs:string">
        <xs:attribute ref="xml:lang" use="required"/>
      </xs:extension>
    </xs:simpleContent>
  </xs:complexType>
</xs:element>
```

Same metadata in various formats

- Conceptual element (RDF):
  - `my:metadata dct:title “Metadata Modelling Presentation”@en`
- Logical profile:
  - Title: mandatory, single value, string.
- Technical implementation:
  - XML (XSD): `<dc:title>Metadata modelling presentation</dc:title>`
  - JSON (Schema): `"title": "Metadata modelling presentation"`


Q&A's