FAIR Data Stewardship

Erik Schultes, Barbara Magagna, Andrea Tarallo

5 December 2024

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 “Education and Research” - Component 2: “From research to business” - Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

Finanziato dall'Unione europea NextGenerationEU
Ministero dell’Università e della Ricerca
Italiadomani

Machine actionable Data Management Plans (maDMPs)

13:40-14:00
(Barbara)

FAIR DMPs – ma DMPs

- DMPs should be include aspects on how to make your data and metadata FAIR (mapping the FAIR principles (-> Alessa)
- DMPs should be understandable by machines to be FAIR
  - Should be standardized (F2) and adapted to community needs (R1.3)
  - Should be expressed with knowledge representation languages (I1) (machine-readable)
  - Should reuse vocabularies (I2) and should be structured based on an ontology (I3) (machine-interpretable)
  - Ideally they can be processed by other service because operations are formalized (machine-actionable)

Standardizing Data Management Plans

- RDA WG DMP Common Standards mainly driven by Tomasz Miksa, Marek Suchanék, Peter Neish
- Objective: Develop a common data model for machine-actionable DMPs
- Benefits:
  - Enable interoperability of systems
  - Facilitate automation of data collection and validation processes
  - Universal format for exchange
  - Investment in research will be safeguarded by assuring that scientific findings are trustworthy and reproducible and the data preserved

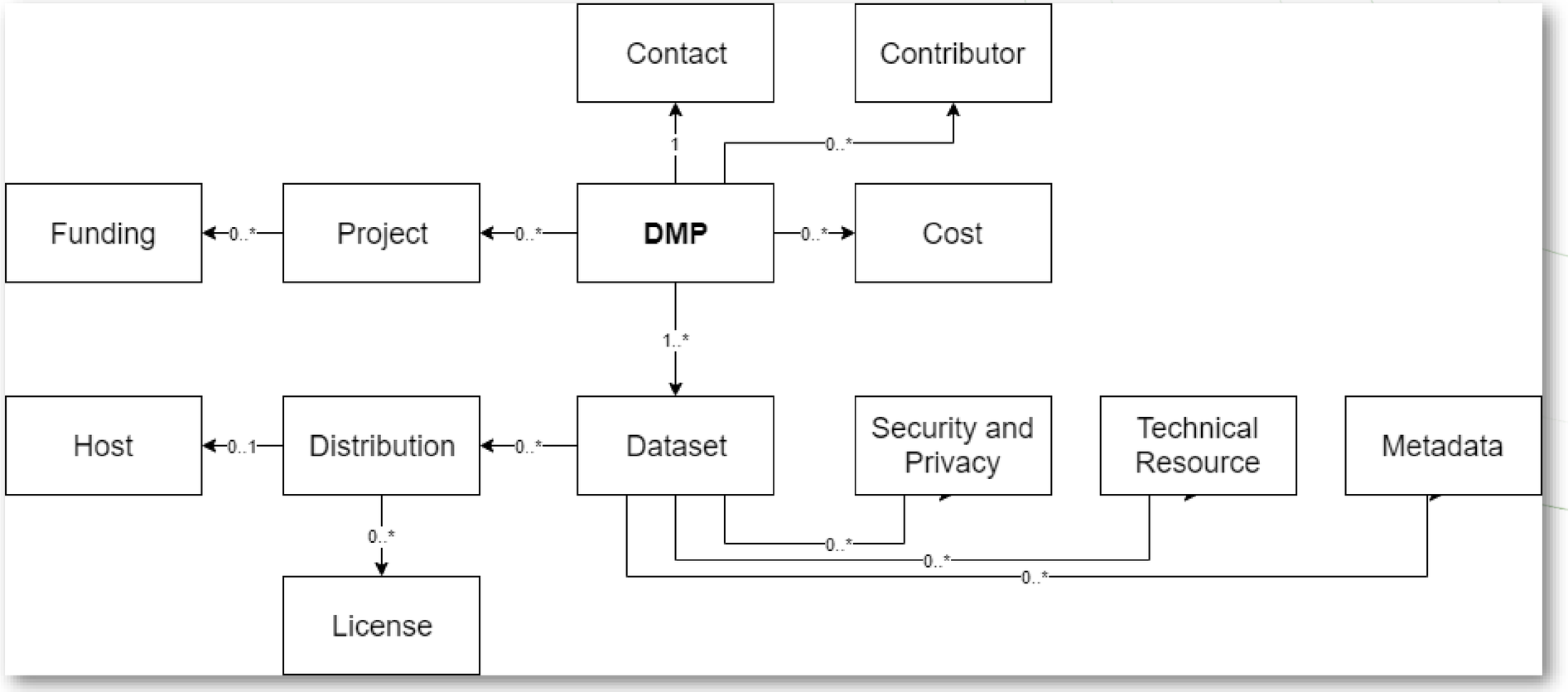
Standardizing Data Management Plans

maDMP hackathon organized involving many stakeholders, final slides

Outcome: https://zenodo.org/records/4036060#.X4VMvWgzY2w

- RDA endorsed recommendations
- Application profile that allows to express maDMPs
  https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/tree/master
- Find more associated papers and useful links here incl. toolings that initiate some automation

Underlying data model



https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard

DMP structure

Structure

- dmp
  - contact
    - contact_id
      - identifier
      - type
    - mbox
    - name
  - contributor
    - contributor_id
      - identifier
      - type
    - mbox
    - name
    - role
  - cost
    - currency_code
    - description
    - title
    - value
  - created
  - dataset
    - data_quality_assurance
    - dataset_id
      - identifier
      - type
    - description
    - distribution
      - access_url
      - available_until
      - byte_size
      - data_access

https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard

Properties in 'dataset'

Name	Description	Data Type	Cardinality	Example Value
data_quality_assurance	Data Quality Assurance	String	0..n	We use file naming convention.
dataset_id	Dataset ID	Nested Data Structure	1	
description	Description is a property in both Dataset and Distribution, in compliance with W3C DCAT. In some cases these might be identical, but in most cases the Dataset represents a more abstract concept, while the distribution can point to a specific file.	String	0..1	Field observation
distribution	To provide technical information on a specific instance of data.	Nested Data Structure	0..n	
issued	Issued. Encoded using the relevant ISO 8601 Date and Time compliant string	Date	0..1	2019-06-30
keyword	Keyword	String	0..n	keyword 1, keyword 2
language	Language of the dataset expressed using ISO 639-3	Term from Controlled Vocabulary	0..1	eng
metadata	To describe metadata standards used.	Nested Data Structure	0..n	
personal_data	Allowed Values: • yes • no • unknown	Term from Controlled Vocabulary	1	unknown

JSON schema 1.1

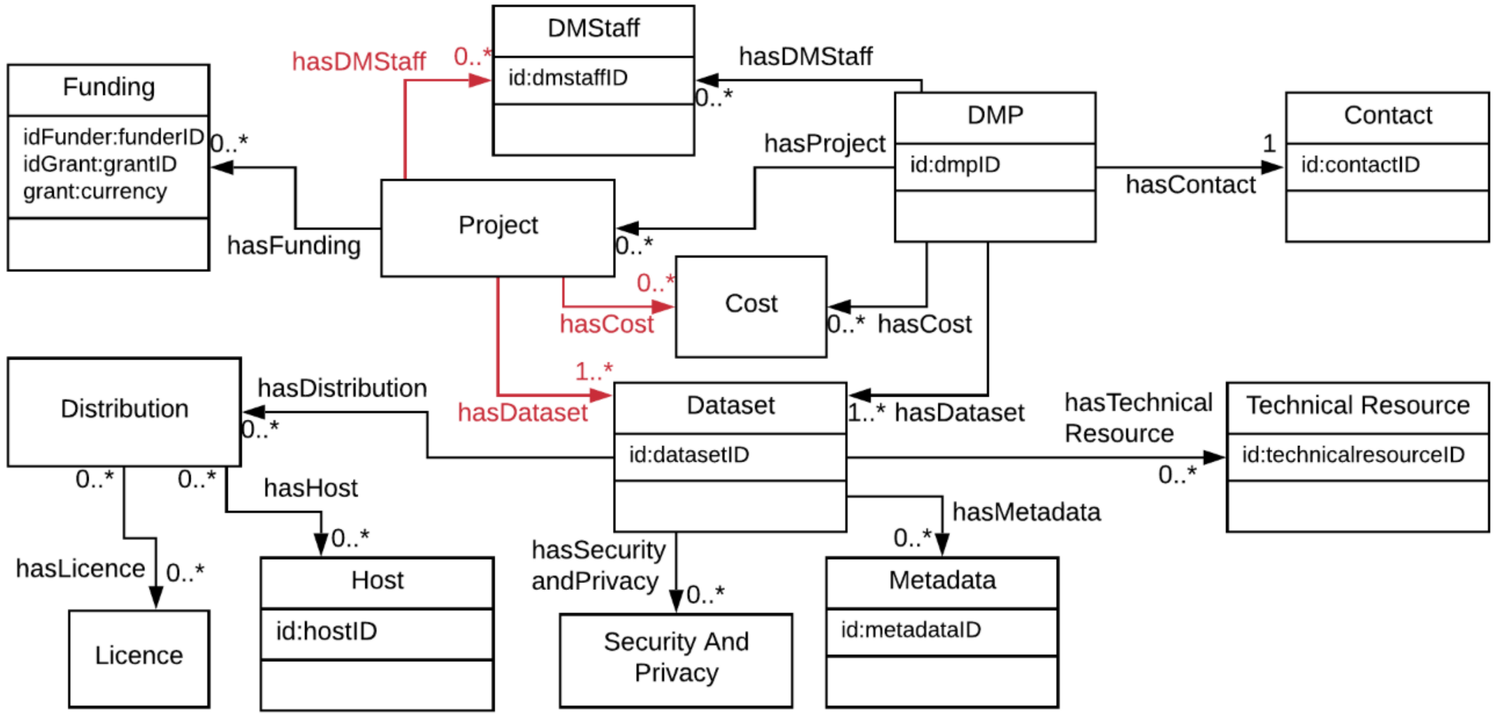

```
{
  "$schema": "http://json-schema.org/draft-07/schema#",
  "$id": "https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/tree/master/examples/JSON/JSON-schema/1.1",
  "title": "RDA DMP Common Standard Schema",
  "description": "JSON Schema for the RDA DMP Common Standard",
  "type": "object",
  "properties": {
    "dmp": {
      "$id": "#/properties/dmp",
      "type": "object",
      "title": "The DMP Schema",
      "properties": {
        "contact": {
          "$id": "#/properties/dmp/properties/contact",
          "type": "object",
          "title": "The DMP Contact Schema",
          "properties": {
            "contact_id": {
              "$id": "#/properties/dmp/properties/contact/properties/contact_id",
              "type": "object",
              "title": "The Contact ID Schema",
              "properties": {
                "identifier": {
                  "$id": "#/properties/dmp/properties/contact/properties/contact_id/properties/identifier",
                  "type": "string",
                  "title": "The DMP Contact Identifier Schema",
                  "examples": ["https://orcid.org/0000-0003-0644-4174"]
                },
                "type": {
                  "$id": "#/properties/dmp/properties/contact/properties/contact_id/properties/type",
                  "type": "string",
                  "enum": [
                    "orcid",
                    "isni",
                    "openid",
                    "other"
                  ],
                  "title": "The DMP Contact Identifier Type Schema",
                  "description": "Identifier type. Allowed values: orcid, isni, openid, other",
                  "examples": ["orcid"]
```


The schema provides a

- knowledge representation language (I1)

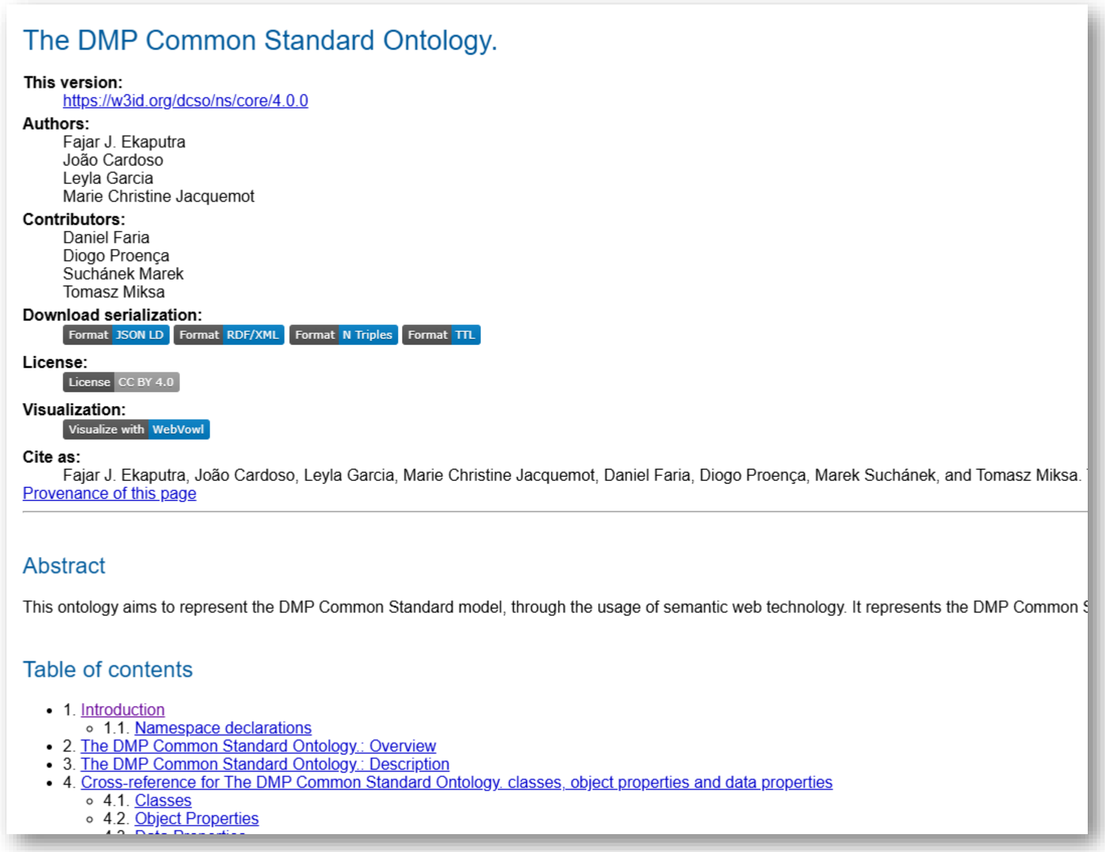
https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/blob/master/examples/JSON/JSON-schema/1.1/maDMP-schema-1.1.json

DMP Common Standard Ontology (DCSO)



https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/tree/master/ontologies

DMP Common Standard Ontology (DCSO)



The DMP Common Standard Ontology.

This version:
https://w3id.org/dcso/ns/core/4.0.0

Authors:
Fajar J. Ekaputra
João Cardoso
Leyla Garcia
Marie Christine Jacquemot

Contributors:
Daniel Faria
Diogo Proença
Suchánek Marek
Tomasz Miksa

Download serialization:
Format JSON LD | Format RDF/XML | Format N Triples | Format TTL

License:
License CC BY 4.0

Visualization:
Visualize with WebVowl

Cite as:
Fajar J. Ekaputra, João Cardoso, Leyla Garcia, Marie Christine Jacquemot, Daniel Faria, Diogo Proença, Marek Suchánek, and Tomasz Miksa.
Provenance of this page

Abstract

This ontology aims to represent the DMP Common Standard model, through the usage of semantic web technology. It represents the DMP Common S

Table of contents

- 1. Introduction
  - 1.1. Namespace declarations
- 2. The DMP Common Standard Ontology.: Overview
- 3. The DMP Common Standard Ontology.: Description
- 4. Cross-reference for The DMP Common Standard Ontology. classes, object properties and data properties
  - 4.1. Classes
  - 4.2. Object Properties

https://w3id.org/dcso

The ontology provides (I3):

- concepts
- object properties

used to FAIRify the structured DMP standard (JSON-LD, RDF)

What is still needed:

Name	Description	Data Type	Cardinality	Example Value
identifier		String	1	
type	Identifier type Allowed Values: • orcid • isni • openid • other	Term from Controlled Vocabulary	1	orcid

Properties in 'contributor'

Name	Description	Data Type	Cardinality	Example Value
contributor_id		Nested Data Structure	1	
mbox	Mail address	String	0..1	john@smith.com
name	Name	String	1	John Smith
role	Type of contributor	String	1..n	Data Steward

To make the DMP ma-interpretable we also need:

- FAIR vocabularies (I2)
- e.g. for identifier types
- contributor roles

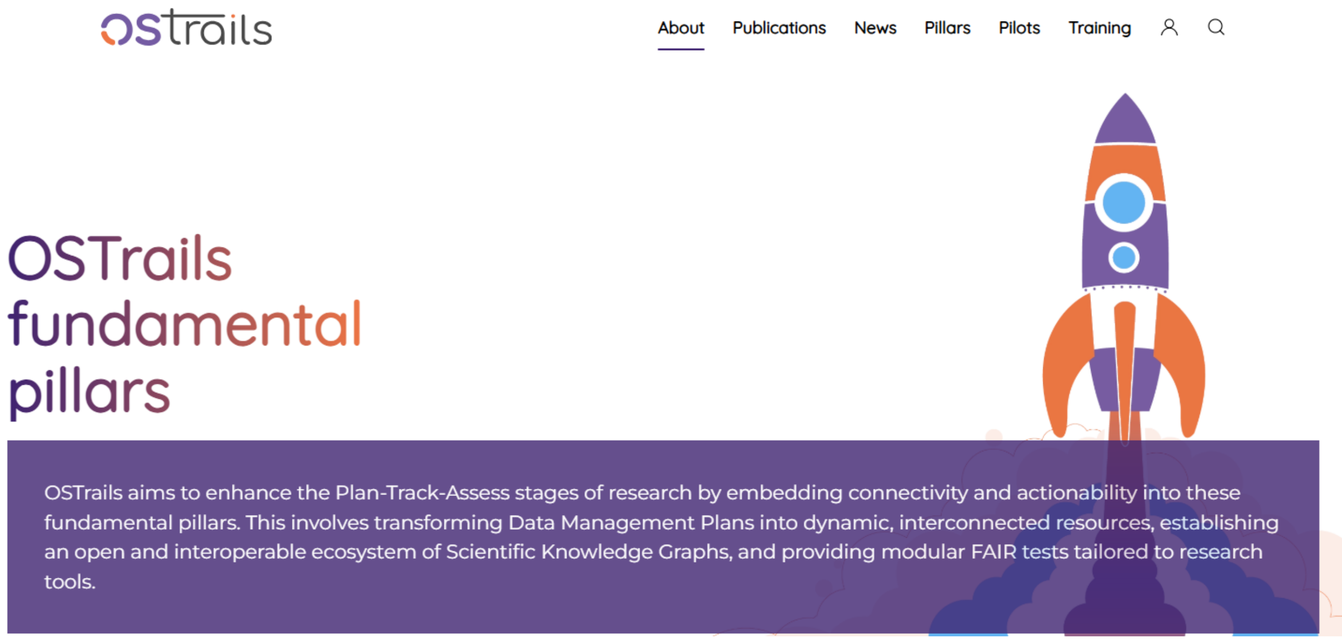
What‘s next?

OSTrails INFRA EOSC project

- DMP into Knowledge Graphs

PARC project:

- FIP2DMP mapping resources used to make Data FAIR to the DMP questions
- find here an article describing the process
- We will hear more in the FIP course



FAIR Well!

THANKS!

Erik Schultes, Barbara Magagna, Andrea Tarallo
27 November 2024

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 "Education and Research" - Component 2: "From research to business" - Investment 3.1: "Fund for the realisation of an integrated system of research and innovation infrastructures"

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani