



8º FÓRUM  
GESTÃO DE DADOS  
DE INVESTIGAÇÃO

25 e 26 DE NOVOBRO 2021  
UNIVERSIDADE DE COIMBRA

# Uma Representação Semântica para Planos de Gestão de Dados

26 de novembro de 2021

João Cardoso

Leyla J. Garcia

Fajar J. Ekaputra

Marie C. Jacquemot

Tomasz Miksa

José Borbinha

# O que é um PGD?

Um **PGD** é um documento formal que tem como objectivo fomentar o **planeamento** e **servir de referência durante a execução** de actividades de gestão de dados.

Um PGD deve descrever **políticas** e **métodos** que são aplicados aos dados durante as múltiplas etapas do **ciclo de vida dos dados (de investigação)**.



# PGD Accionável por Máquina

O conceito de **PGD accionável por máquina** (PGDam, maDMP) mitiga alguns dos problemas dos PGD tradicionais.

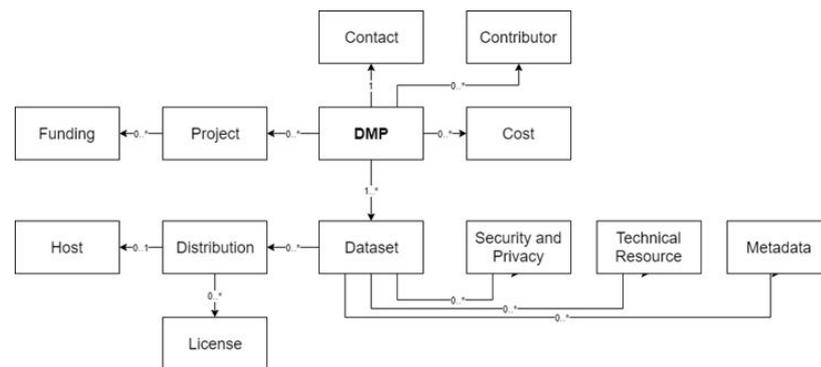
- Permite a troca de informação entre sistemas
- Permite a integração dos PGD em workflows de GDI
- Obriga ao uso de identificadores persistentes
- Facilita o processo de actualização de um PGD
- Facilita o processo de criação de um PGD



# DMP Common Standards WG

.O grupo de trabalho **DMP Common Standards (DCS)** dedicou-se em criar um standard que define **conjunto mínimo e universal de conceitos que definem um PGD.**

O standard tem como objectivo garantir a **interoperabilidade entre sistemas** que utilizem PGDs.



Miksa, T., Walk, P., Neish, P., Oblasser, S., Murray, H., Renner, T., Jacquemot-Perbal, M.C., Cardoso, J., Kvamme, T., Praetzellis, M. and Suchánek, M., 2021. Application Profile for Machine-Actionable Data Management Plans. Data Science Journal, 20(1).

# Porquê uma Ontologia?

O DCS fornece **serializações de referência** para o perfil aplicacional.

O nosso objectivo era criar uma nova **serialização com características distintas das serializações existentes.**

- Tecnologia Semantica
- A DMP Common Standard Ontology (DCSO)

As ontologias permitem uma representação de uma **conceptualização partilhada de conhecimento, através do uso de semântica formal.**

- Fácil extensão
- Permite a inferência de conhecimento não explícito
- Permite a integração de estruturas de terceiros



A DCSO está organizada em **três blocos principais**:

- DCSO Core
- DCSO Extension
- DCSO Constraints Validation Layer

João Cardoso, Leyla J. García, Fajar J. Ekaputra, Marie C. Jacquemot, Tomasz Miksa, José Borbinha:  
Towards semantic representation of machine-actionable Data Management Plans. 1-9



# DCSO Core

A DCSO Core **representa o conjunto base dos elementos** que compõem o perfil aplicacional do DCS.

É composta por **26 classes**.

**13 correspondem directamente** a termos no perfil aplicacional DCS.

13 dividem-se entre **classes de identificação e classes externas**.

Vocabulários controlados utilizados:

- W3C DCAT Specification
- FOAF Vocabulary Specification 0.99

```
dcso:Cost
dcso:DMP
dcso:Funding
dcso:Host
dcso:Licence
dcso:Metadata
dcso:Project
dcso:SecurityPrivacy
dcso:TechnicalResource
dcso:Id
  dcso:ContactId
  dcso:ContributorId
  dcso:DatasetId
  dcso:DMPId
  dcso:FunderId
  dcso:GrantId
  dcso:MetadataStandardId
dcat:Dataset
  dcso:Dataset
dcat:Distribution
  dcso:Distribution
dcsx:Country
dcsx:CurrencyCode
dcsx:Language
foaf:Agent
  dcso:Contact
  dcso:Contributor
```



# DCSO Extension

A DCSX foi criada para suportar o uso de **vocabulários controlados** na DCSO Core.

Cada classe representa um vocabulário controlado::

- A classe **dcsx:Country** representa o ISO 3166-1
- A classe **dcxs:CurrencyCode** representa o ISO 4217
- A classe **dcsx:Language** representa o ISO 639-3

```
dcsx:Country  
dcsx:CurrencyCode  
dcsx:Language
```

# DCSO Constraints Validation Layer



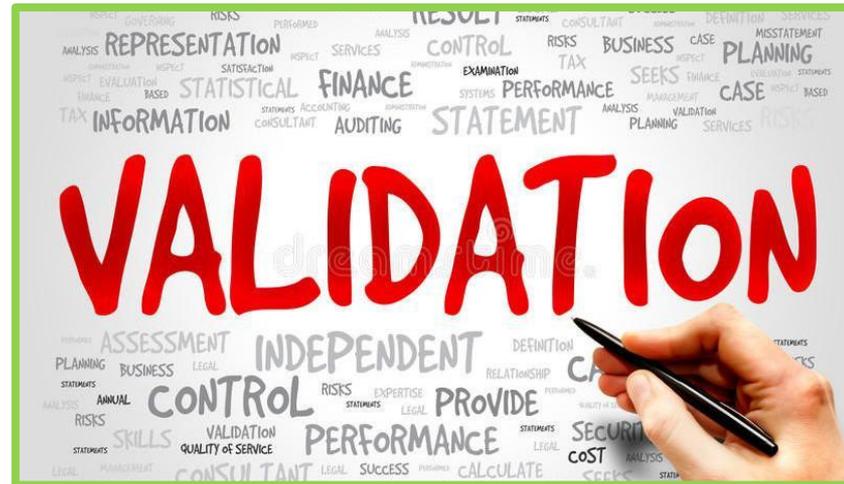
8º FÓRUM  
GESTÃO DE DADOS  
DE INVESTIGAÇÃO  
25 a 26 DE NOVEMBRO 2021  
UNIVERSIDADE DE COIMBRA

A DCSO Constraints Validation Layer facilita a **validação e força o cumprimento** do modelo subjacente.

É representada utilizando a **linguagem ShEx**.

Os esquemas ShEx criados seguem as **diretrizes estabelecidas no perfil aplicativo DCS**. Em particular:

- Existência
- Cardinalidade
- Relações entre elementos



# Pontos a reter

O DCSO é uma forma de **representar PGDs de forma accionável por máquina.**

Permite a **inferência de conhecimento não explicitamente descrito.**

Tem o potencial de servir de **base para a criação de serviços** que utilizem o conhecimento descrito nos PGD.

Tem o potencial de permitir a **automação** de actividades GDI.



THANK  
YOU!