



TW/C

Adoption* of RDA DTR and PIT in the Deep Carbon Observatory Data Portal

Xiaogang Ma, John Erickson, Patrick West,
Stephan Zednik, Peter Fox,
& the DCO-Data Science Team

Tetherless World Constellation
Rensselaer Polytechnic Institute

*Funded by RDA/US (NSF)

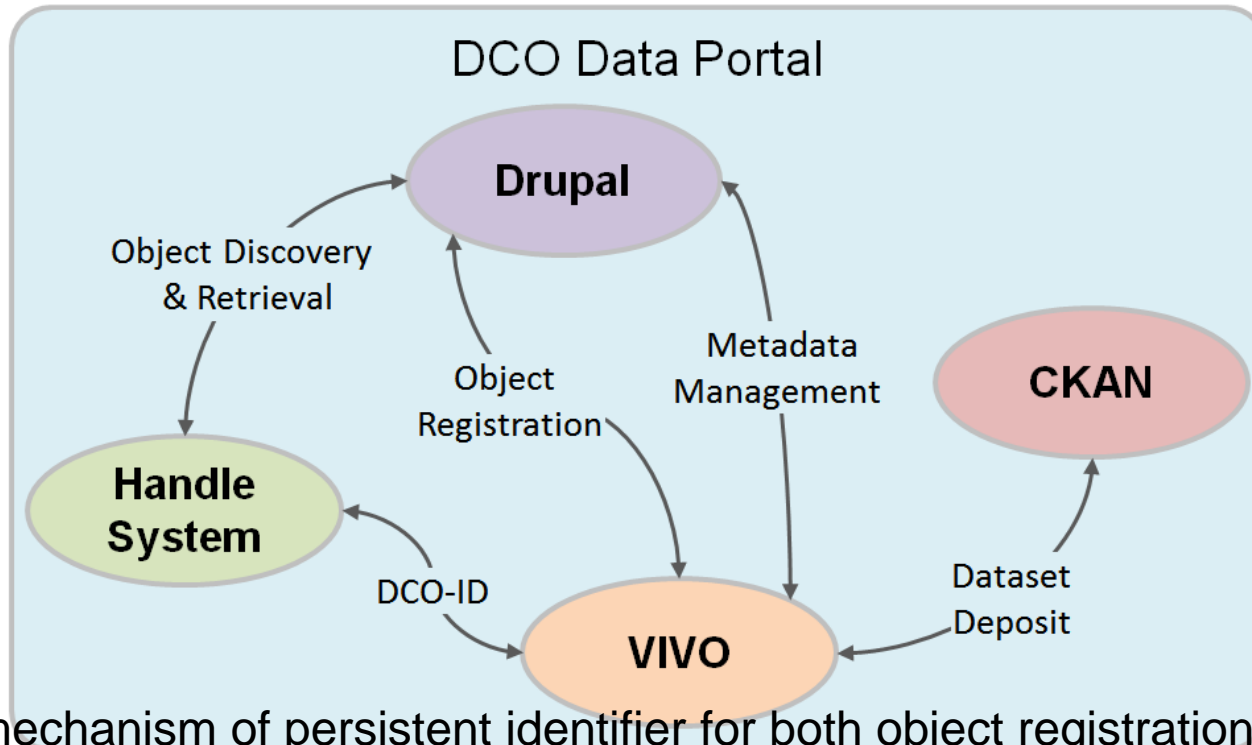
http://tw.rpi.edu/web/doc/20150921_slides_RDA_P6.pptx





DCO Research Requirements

- Each defined data type needs a stable and resolvable PID
- Provide semantics - meaning and context - to the defined data types
- Annotate datasets with one or more defined data types



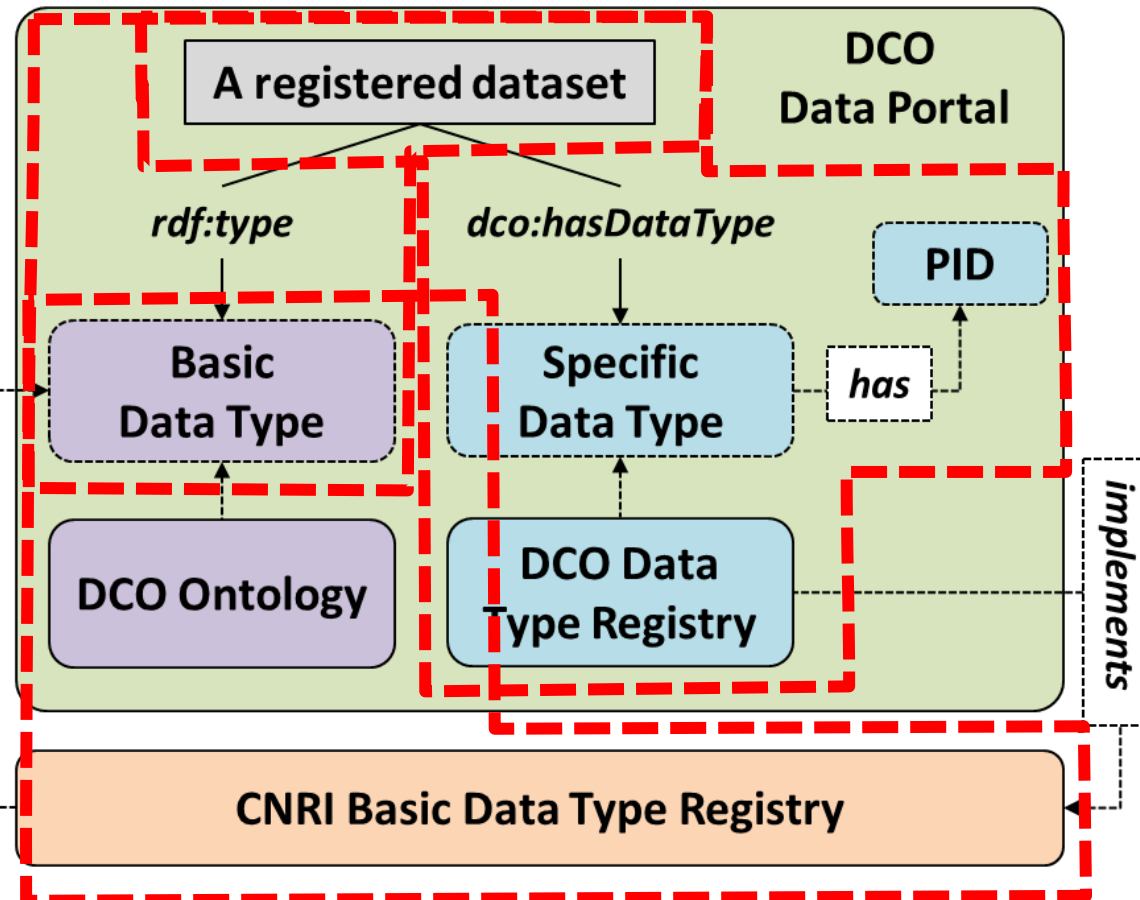
DCO-ID as a mechanism of persistent identifier for both object registration and retrieval²



Nature of efforts

The DTR primitives are comparable to a list of **BASIC DATA TYPE CLASSES** in the DCO ontology, e.g. Dataset, Image, Video, Audio, etc.

A registered DCO dataset is asserted as an instance of one of those basic data type classes.

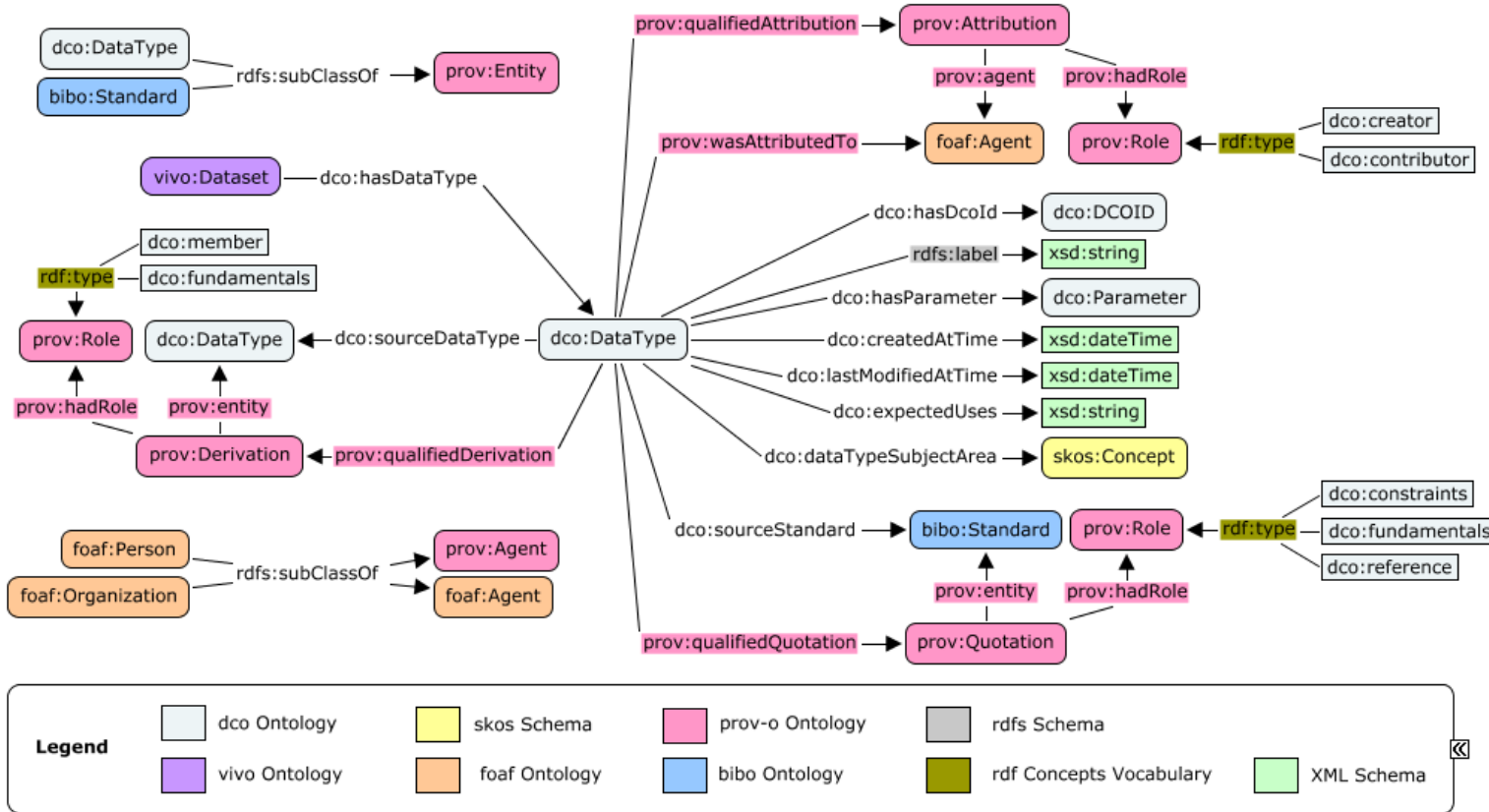


It is possible to further annotate the dataset with the **SPECIFIC DATA TYPES** defined within a DTR, and each data type has a unique PID.



Results of data type specification

- Updates to the DCO Ontology:
 - A new class `dco:DataType`. Each specific data type is an instance of it
 - An object property `dco:hasDataType` linking a dataset and a data type
 - A collection of other classes and properties associated with `dco:DataType`





Implementation of data type and DCO-ID



Admin Panel [Edit this individual](#)

Verbose property display is **off** | [Turn on](#)

Resource URI: <http://info.deepcarbon.net/individual/n5989>

DCO Ontology [Dataset](#)

DCO ID [11121/1490-1829-4848-8082-CC](#)

Websites

The basic data type

```
dco:dcoOntology
rdf:type vivo:Dataset .
```

Overview

Research

Contact

Related Documents

Time

Identity

Additional Document Info

Bib Mapping

Other

View All

Has data type [+](#)

[Resource Description Framework](#)


keywords [+](#)

The specific data type

```
dco:dcoOntology
dco:hasDataType dco:RDF .
```



Profile of a registered data type

 [Admin Panel](#) [Edit this individual](#) Verbose property display is **off** | [Turn on](#)

Resource URI: <http://info.deepcarbon.net/individual/n47205>

Resource Description Framework | [Data Type](#)

DCO ID [11121/1305-5365-5550-2605-CC](#)

Research Identity Other View All

Has source standard +

[W3C - Resource Description Framework \(RDF\)](#)

source data type +

- [JSON-LD](#)
- [N-Quads](#)
- [N-Triples](#)
- [N3](#)

```
dco:RDF
  a dco:DataType .
dco:DataType
  a owl:Class .
```

Each registered object, such as a data type, has a unique **DCO-ID**, which is resolvable by the global Handle System



A faceted browser for registered data types

▼ Creation Year

- 1999 (1) ⓘ
- 2011 (1) ⓘ
- 2014 (4) ⓘ
- 2015 (1) ⓘ

► Creator

▼ Parameter

- Data Source (1) ⓘ
- Entropy Increment (1) ⓘ
- Heat Content (1) ⓘ
- Mineral Name (1) ⓘ
- Molecular Formula (1) ⓘ
- Molecular Weight (1) ⓘ
- Temperature (1) ⓘ
- Temperature Change (1) ⓘ

► Source Standard

▼ Annotations

- Computer Modeling (1) ⓘ
- Deep Energy (1) ⓘ
- File Format (5) ⓘ
- Geochemistry (1) ⓘ
- Knowledge representation (1) ⓘ
- Mineralogy (1) ⓘ

1-7 of 7 < > 10 ▾

JSON-LD

DCO ID: [11121/3245-1505-7270-4505-CC](#)

Author: [Kellogg, Gregg](#) ; [Lanthaler, Markus](#) ; [Sporny, Manu](#)

Source Standard: [W3C - JSON-LD 1.0](#)

N-Triples

DCO ID: [11121/2785-3389-1903-1770-CC](#)

Author: [Carothers, Gavin](#) ; [Seaborne, Andy](#)

Source Standard: [W3C - RDF 1.1 N-Triples](#)

N3

DCO ID: [11121/8064-6347-8072-9328-CC](#)

Author: [Berners-Lee, Tim](#) ; [Connolly, Dan](#)

Source Standard: [W3C - Notation3 \(N3\)](#)

RDF/XML

DCO ID: [11121/7331-3225-6985-5846-CC](#)

Author: [Gandon, Fabien](#) ; [Schreiber, Guus](#)

Source Standard: [W3C - RDF 1.1 XML Syntax](#)

Resource Description Framework

DCO ID: [11121/1305-5365-5550-2605-CC](#)

Source Standard: [W3C - Resource Description Framework \(RDF\)](#)

Thermodynamics of chemicals and minerals

DCO ID: [11121/2224-7054-5955-3791-CC](#)

Author: [Ma, Xiaogang \(Marshall\)](#)

Parameters: [Mineral Name](#); [Heat Content](#); [Temperature](#); [Data Source](#); [Entropy Increment](#); [Molecular Formula](#); [Molecular Weight](#); [Temperature Change](#)

Turtle

DCO ID: [11121/9814-7548-3804-9868-CC](#)

Author: [Prud'hommeaux, Eric](#) ; [Carothers, Gavin](#)

Source Standard: [W3C - RDF 1.1 Turtle](#)



Using Data Type as a facet in DCO dataset browser

► Years

► DCO Authors

► Communities

► Groups

► Projects

▼ Data Types

Comma-Separated Values (1) ⓘ

Keyhole Markup Language (1) ⓘ

Resource Description Framework (1) ⓘ

Thermodynamics of chemicals and minerals (1) ⓘ

1-2 of 2 < > 10 ▼

DCO Ontology

DCO ID: [11121/1490-1829-4848-8082-CC](#)

Communities: [Data Science Team](#)

Groups: [Data Science Community](#)

Authors: [Wang, Han](#); [Erickson, John](#); [Ma, Xiaogang \(Marshall\)](#); [Fox, Peter](#)

Data Types: [Resource Description Framework](#)

Distributions:

[DCO Ontology Distribution - 201404](#) (Direct access: [DCO Ontology Turtle File - 201404](#))

[DCO Ontology Distribution - 201408](#) (Direct access: [DCO Ontology Turtle File - 201408](#))

[DCO Ontology Distribution - 201405](#) (Direct access: [DCO Ontology Turtle File - 201405](#))

The Heat Capacities of Magnesium, Zinc, Lead, Manganese and Iron Carbonates at Low Temperatures

DCO ID: [11121/4589-9713-2502-3685-CC](#)

Communities: [Data Science Team](#); [Extreme Physics and Chemistry Community](#)

Groups: [Data Science Community](#)

Authors: [Zhong, Hao](#); [Ma, Xiaogang \(Marshall\)](#); [Ghiorso, Mark](#)

Data Types: [Thermodynamics of chemicals and minerals](#)

Distributions:

[2015-03-22-1532](#)



Thermodynamics of chemicals and minerals | [Data Type](#)

DCO ID [11121/9177-8600-7213-5328-CC](#)

- Overview
- Identity
- Other
- View All

expected uses

The data type is for thermodynamics of chemicals and minerals. Records cover two major topics: Enthalpy & Entropy. Detailed items include but are not limited to:

- Mineral Name
- Molecular Formula
- Molecular Weight
- Temperature (T, °K)
- Temperature Change
- Heat Content (Cp, calorie/mole)
- Entropy Increment
- Data Source

Additionally, information about materials used in the test of thermodynamic features can also be recorded. Detailed items can include:

- Material
- Source
- Sample, g.
- Density Density temp., °C
- Purity, %
- Impurities
- Impurities corrected for

has parameter

[Data Source](#)



Notable

- Machine accessibility and readability
 - Given the DCO-ID of a data type, SPARQL queries can be sent to the triple store of the DCO data portal to retrieve information about the data type
 - Such SPARQL queries can be derived from a query template that is tailored for data types
 - To have a such a query template, further work may be needed to identify the metadata kernel of data type
 - These also show the vision of a API for data type information



Conclusions

- The methodology of RDA DTR and PIT is highly implementable, especially in the environment of the Semantic Web.
- The technical framework in the current demonstration systems of DTR and PIT can be adapted or further extended for production uses.
- Initial good researcher response (they recognize their data types)
- Slides with backup detail at:
http://tw.rpi.edu/web/doc/20150921_slides_RDA_P6.pptx
- Contact Marshall at max7@rpi.edu

Thank you!



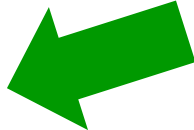


Outline

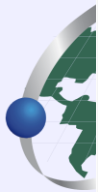
- Background & Research questions
 - RDA-DTR, RDA-PIT, DCO Data Portal
- Nature of efforts
 - Basic data type and Specific data type
- Approaches
 - Integration architecture vs. Self-contained architecture
 - DCO-ID
- Results and recommendation



Possible DCO-DTR Approaches

- An integration architecture
 - DCO Data Portal is built on the VIVO platform
 - DTR and DCO-VIVO as separate knowledge bases
 - DCO-VIVO uses DTR API to access data type information
- A self-contained architecture  *We have worked on this approach*
 - To have the functionality of DTR completely within the DCO Data Portal
 - Need to modify the DCO Ontology, e.g. add a class `dco:DataType` and collect properties associated with it

DCO Datatype Summary



DATATYPE TITLE

Thermodynamics of chemicals and minerals

DCO ID

<http://dx.deepcarbon.net/11121/9177-8600-7213-5328-CC>

SOURCE DATA TYPE

SOURCE DATA TYPE URI

SOURCE STANDARD

ATTRIBUTED TO

Ma, Xiaogang (Marshall)

DATATYPE SUBJECT AREA

Geochemistry
Thermodynamics
Deep Energy
Mineralogy

PARAMETERS

Molecular Weight
Temperature
Heat Content
Temperature Change
Molecular Formula
Mineral Name
Data Source
Entropy Increment

DATATYPE EXPECTED USES

The data type is for thermodynamics of chemicals and minerals. Records cover two major topics: Enthalpy & Entropy. Detailed items include but are not limited to:



Thermodynamics of chemicals and minerals | Data Type [🔗](#)

DCO ID [11121/9177-8600-7213-5328-CC](#)

- Overview
- Identity
- Other
- View All

expected uses

The data type is for thermodynamics of chemicals and minerals. Records cover two major topics: Enthalpy & Entropy. Detailed items include but are not limited to:

- Mineral Name
- Molecular Formula
- Molecular Weight
- Temperature (T, °K)
- Temperature Change
- Heat Content (Cp, calorie/mole)
- Entropy Increment
- Data Source

Additionally, information about materials used in the test of thermodynamic features can also be recorded. Detailed items can include:

- Material
- Source
- Sample, g.
- Density Density temp., °C
- Purity, %
- Impurities
- Impurities corrected for

has parameter

[Data Source](#)

JSON-LD | [Data Type](#)

DCO ID [11121/3245-1505-7270-4505-CC](#)

- [Research](#)
- [Identity](#)
- [Provenance](#)
- [View All](#)

has source data type

[Resource Description Framework](#)

has source standard

[W3C - JSON-LD 1.0](#)

has subject area

[File Format](#)

[Semantic Web](#)

[Web Science](#)



Background

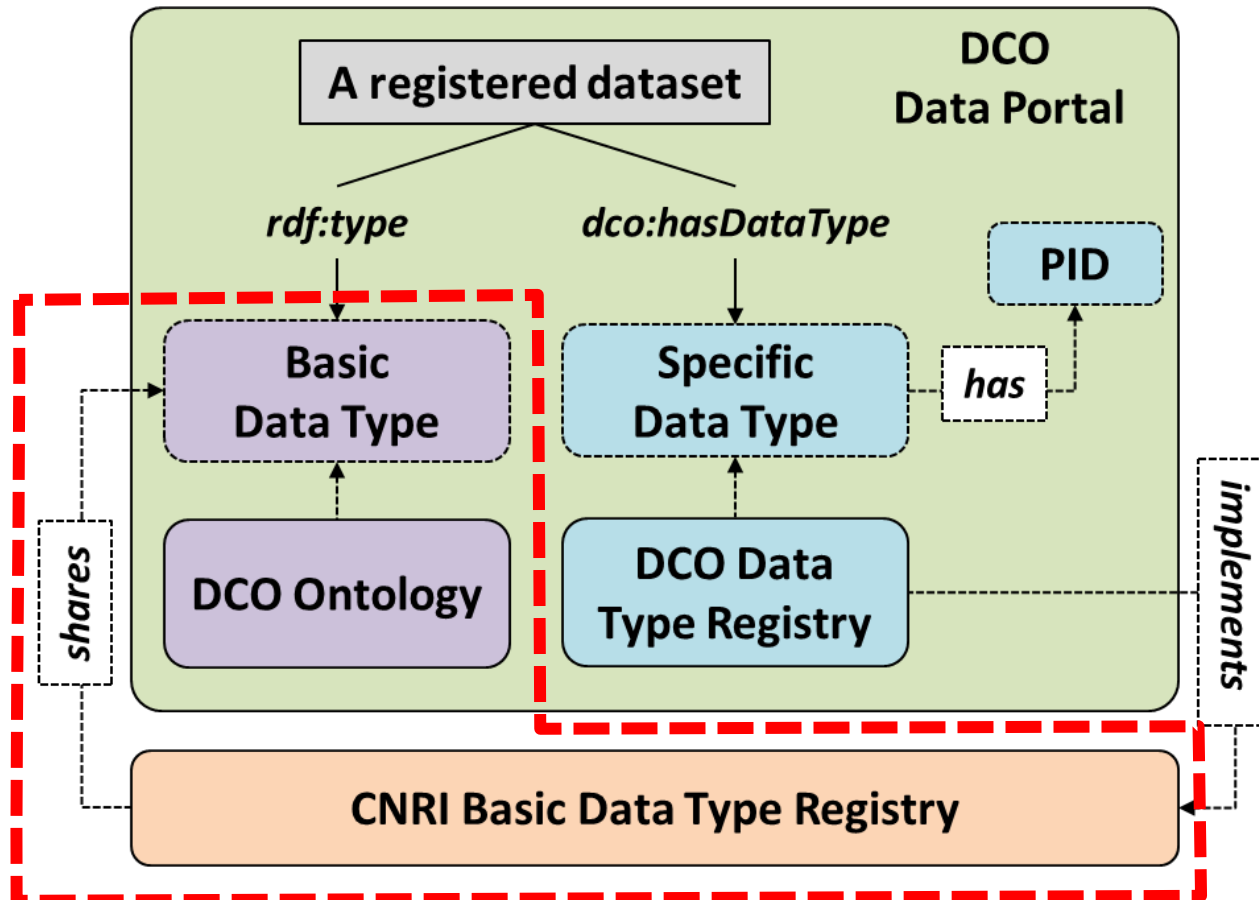
- RDA - Data Type Registry (**DTR**) working group
 - Addressed a core issue of data interoperability: to parse, understand, and reuse data retrieved from others
- RDA - Persistent Identifier Information Types (**PIT**) working group
 - Addressed the essential types of information associated with persistent identifiers (PID)
- Deep Carbon Observatory (**DCO**) Data Portal
 - Centrally-managed digital object identification, object registration, metadata management and knowledge graph curation.
 - <http://deepcarbon.net>





Nature of efforts

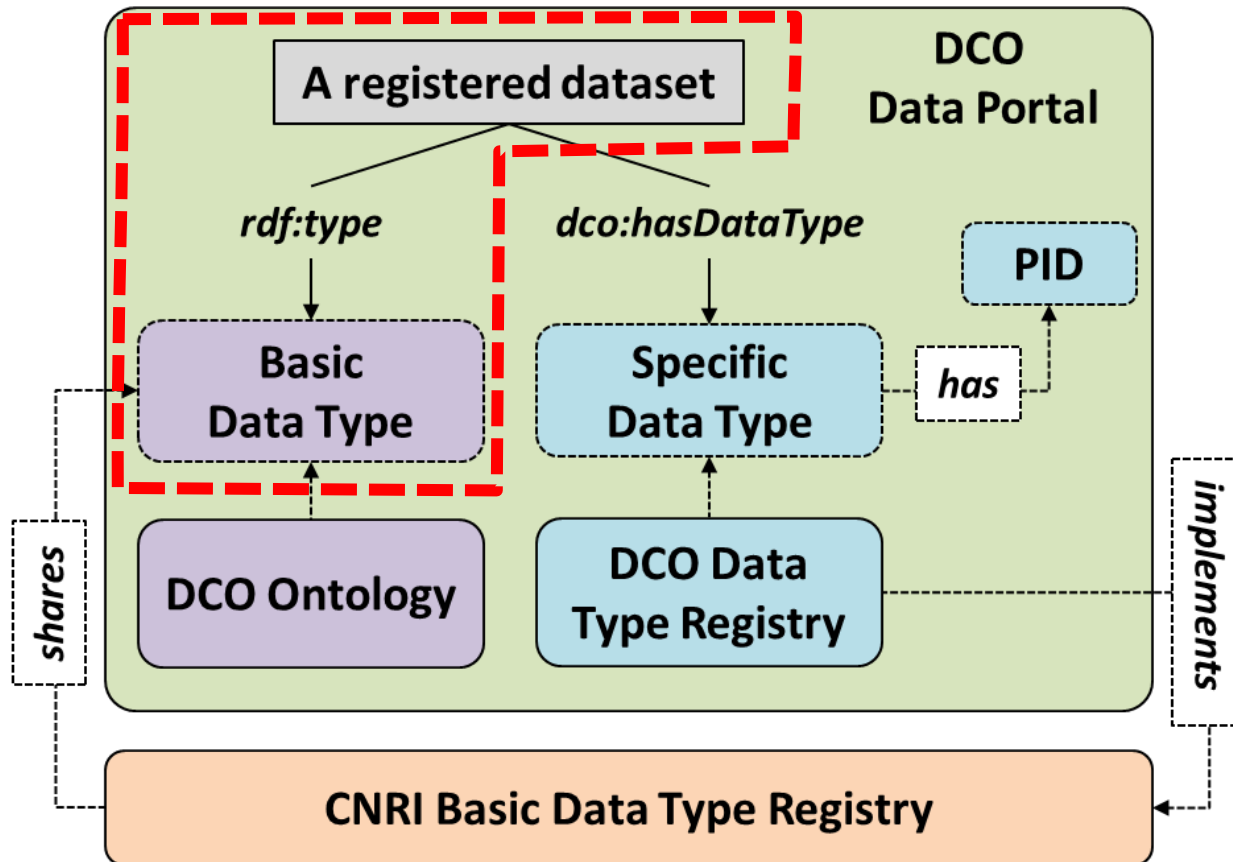
The DTR primitives are comparable to a list of **BASIC DATA TYPE CLASSES** in the DCO ontology, e.g. Dataset, Image, Video, Audio, etc.





Nature of efforts (cont.)

A registered DCO dataset is asserted as an instance of one of those basic data type classes.





Nature of efforts (cont.)

It is possible to further annotate the dataset with the **SPECIFIC DATA TYPES** defined within a DTR, and each data type has a unique PID.

