



@ontocommons



linkedin.com/company/ontocommons



info@ontocommons.eu

# Domain-level ontologies and the methodology to connect them to a top-level/middle-level ontology

The case of the CIF and crystallography domain ontology

---

*Emanuele Ghedini, UNIBO  
Jesper Friis, SINTEF*



# Existing standards

The International Union of Crystallography (IUCr) has long been working on standardising crystallographic data



CIF standard proposal: S.R. Hall, F.H. Allen & I.D. Brown (1991) Acta Cryst A47, 655-685

CIF is a acronym for both  
*Crystallographic Information File*  
and  
*Crystallographic Information Framework*



Data exchange file format

System of data dictionaries  
and relational rules

DDL

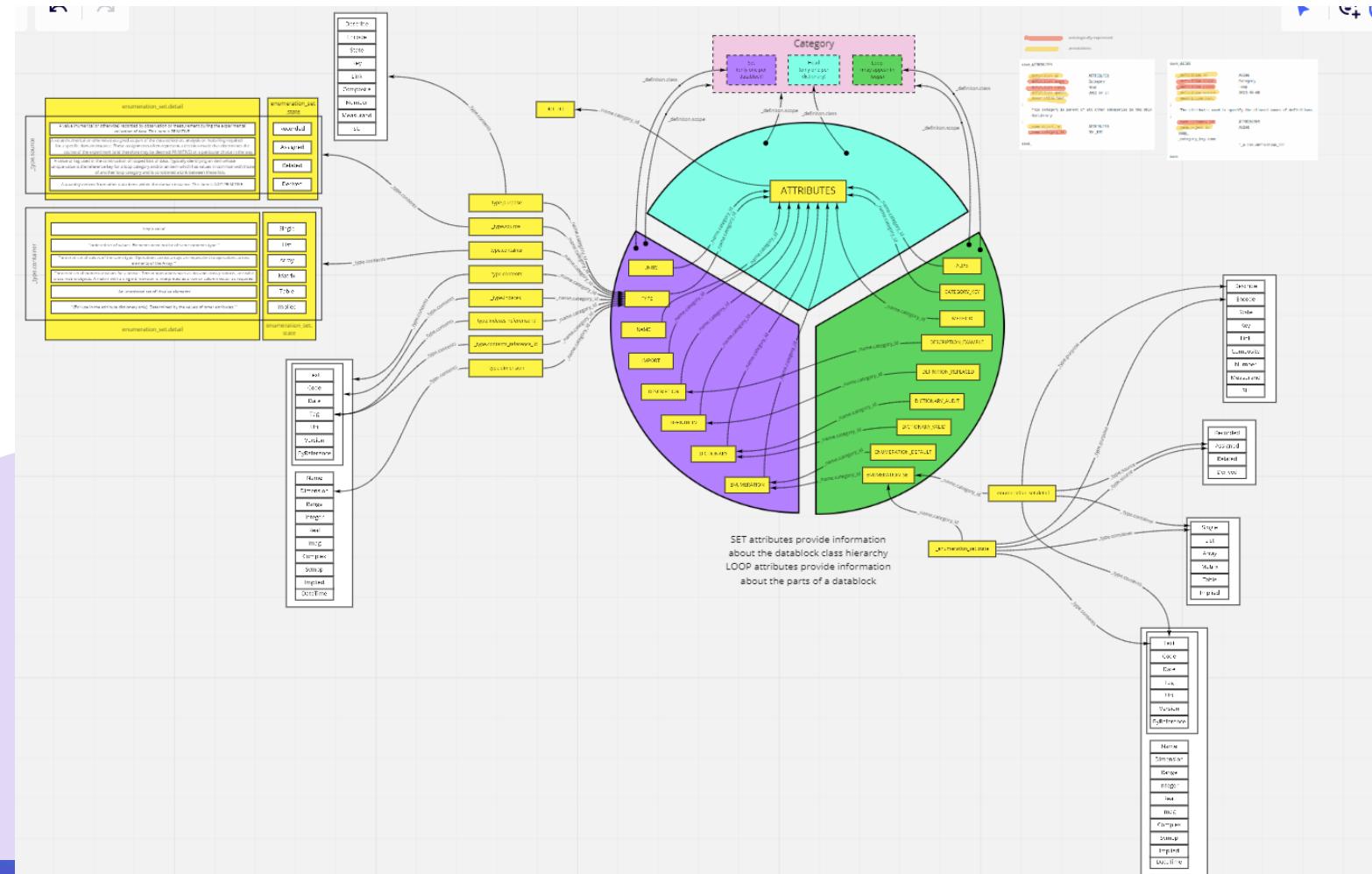
Dictionaries

Machine readable, but not formalised  
ontologically

- » Core CIF dictionary
- » Restraints dictionary
- » Powder CIF dictionary
- » Modulated structures CIF dictionary
- » Electron density CIF dictionary
- » Twinning CIF dictionary
- » Macromolecular CIF dictionary
- » Image CIF dictionary
- » Symmetry CIF dictionary
- » Magnetic CIF dictionary
- » Topology CIF dictionary
- » Suggested new CIF data items

# Dictionary definition language (DDL)

# A schema for the CIF standards



[https://github.com/COMCIFS/cif\\_core/blob/master/ddl.dic](https://github.com/COMCIFS/cif_core/blob/master/ddl.dic)

```

#\#CIF_2.0
#####
#                               DDLm REFERENCE DICTIONARY
#
#                               #
#####
data_DDL_DIC

_dictionary.title          DDL_DIC
_dictionary.class           Reference
_dictionary.version         4.0.1
_dictionary.date            2021-03-01
_dictionary.uri              https://raw.githubusercontent.com/COMCIFS/ci
_dictionary.ddl_conformance 4.0.1
_dictionary.namespace        DdlDic
_description.text

;
This dictionary contains the definitions of attributes that
make up the DDLm dictionary definition language. It provides
the meta meta data for all CIF dictionaries.
;

save_ATTRIBUTES

_definition.id             ATTRIBUTES
_definition.scope           Category
_definition.class           Head
_definition.update           2011-07-27
_description.text

;
This category is parent of all other categories in the DDLm
dictionary.
;

_name.object_id             ATTRIBUTES
_name.category_id           DDL_DIC

save_

#####

```

## CORE CIF DICTIONARY

The core dictionary is a set of data names designed to cover the requirements of archiving and exchanging raw and processed data and derived structural results for single-crystal small-molecule and inorganic crystal studies.

## CURRENT OFFICIAL RELEASE

- Version 2.4.5 of 21 November 2014
  - HTML version
- **Search** for a data name.

[https://www.iucr.org/resources/cif/dictionaries/cif\\_core](https://www.iucr.org/resources/cif/dictionaries/cif_core)



Definitions are arranged alphabetically by category and within category.

- Revision history
- \_atom\_site\_[]
  - \_atom\_site\_adp\_type
  - \_atom\_site\_aniso\_B\_11
  - \_atom\_site\_aniso\_B\_12
  - \_atom\_site\_aniso\_B\_13
  - \_atom\_site\_aniso\_B\_22
  - \_atom\_site\_aniso\_B\_23
  - \_atom\_site\_aniso\_B\_33
  - \_atom\_site\_aniso\_label
  - \_atom\_site\_aniso\_ratio
  - \_atom\_site\_aniso\_type\_symbol
  - \_atom\_site\_aniso\_U\_11
  - \_atom\_site\_aniso\_U\_12
  - \_atom\_site\_aniso\_U\_13
  - \_atom\_site\_aniso\_U\_22
  - \_atom\_site\_aniso\_U\_23
  - \_atom\_site\_aniso\_U\_33
  - \_atom\_site\_attached\_hydrogens
  - \_atom\_site\_B\_equiv\_geom\_mean
  - \_atom\_site\_B\_iso\_or\_equiv
  - \_atom\_site\_calc\_attached\_atom
  - \_atom\_site\_calc\_flag
  - \_atom\_site\_Cartn\_x
  - \_atom\_site\_Cartn\_y
  - \_atom\_site\_Cartn\_z
  - \_atom\_site\_chemical\_conn\_number
  - \_atom\_site\_constraints
  - \_atom\_site\_description
  - \_atom\_site\_disorder\_assembly
  - \_atom\_site\_disorder\_group
  - \_atom\_site\_fract\_x
  - \_atom\_site\_fract\_y
  - \_atom\_site\_fract\_z
  - \_atom\_site\_label
  - \_atom\_site\_label\_component\_0
  - \_atom\_site\_label\_component\_1
  - \_atom\_site\_label\_component\_2
  - \_atom\_site\_label\_component\_3

# Motivation

## IUCr community

### Formalise the CIF standards as an ontology

- enable precise description of experimental results
- enable use of all available data in structure determination

Domain experts

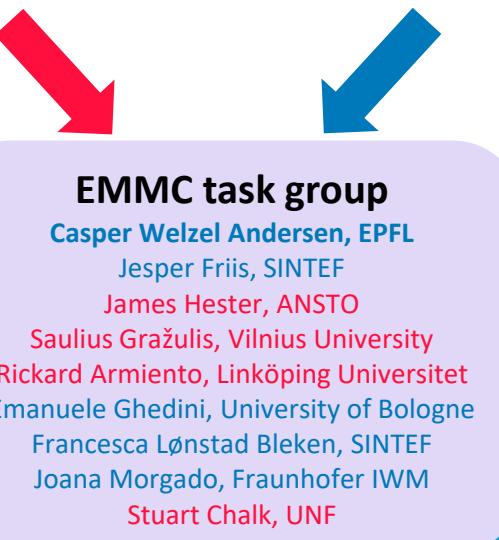
## EMMO community

### An endorsed standard for describing atomistic structures

Needed by several domain ontologies, like:

- structure characterisation
- atomistic modelling
- electronic modelling
- ...

Ontology experts



<https://emmc.eu/focus-areas/interoperability/>

<https://github.com/emmo-repo/CIF-ontology>

# Aim of the task group

- Create ontologies based on the existing CIF dictionaries
  - should be generated:
    - to ensure consistency
    - to ease maintenance
- Create an EMMO domain ontology for crystallography
  - ontological description of atomistic structures
  - based on EMMO and CIF Core dictionary
  - include crystallographic concepts not explicitly stated in CIF Core

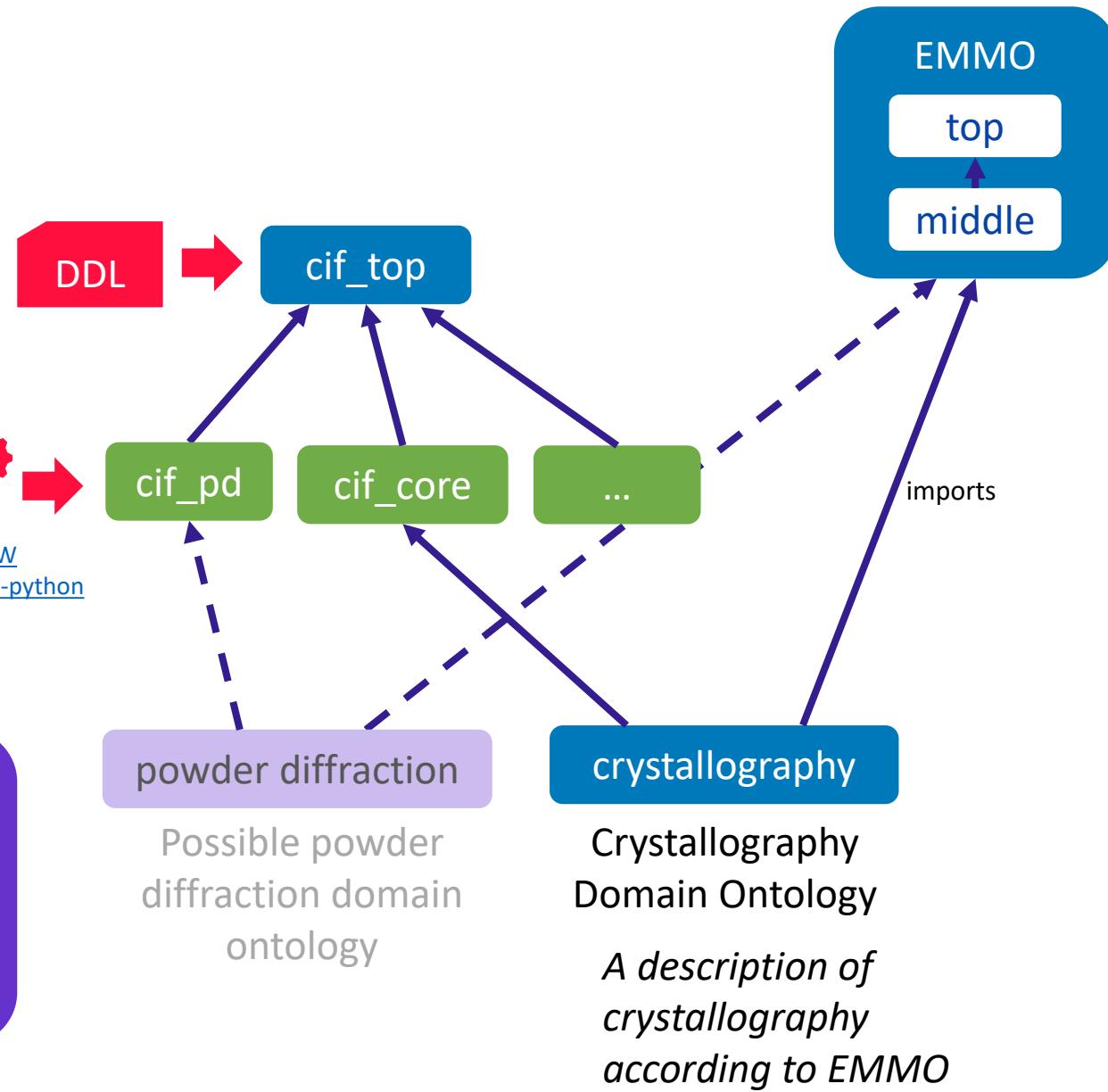
# Approach

Ontological description of the CIF data model

Ahead to the CIF DDL

Independent of any top-level ontology

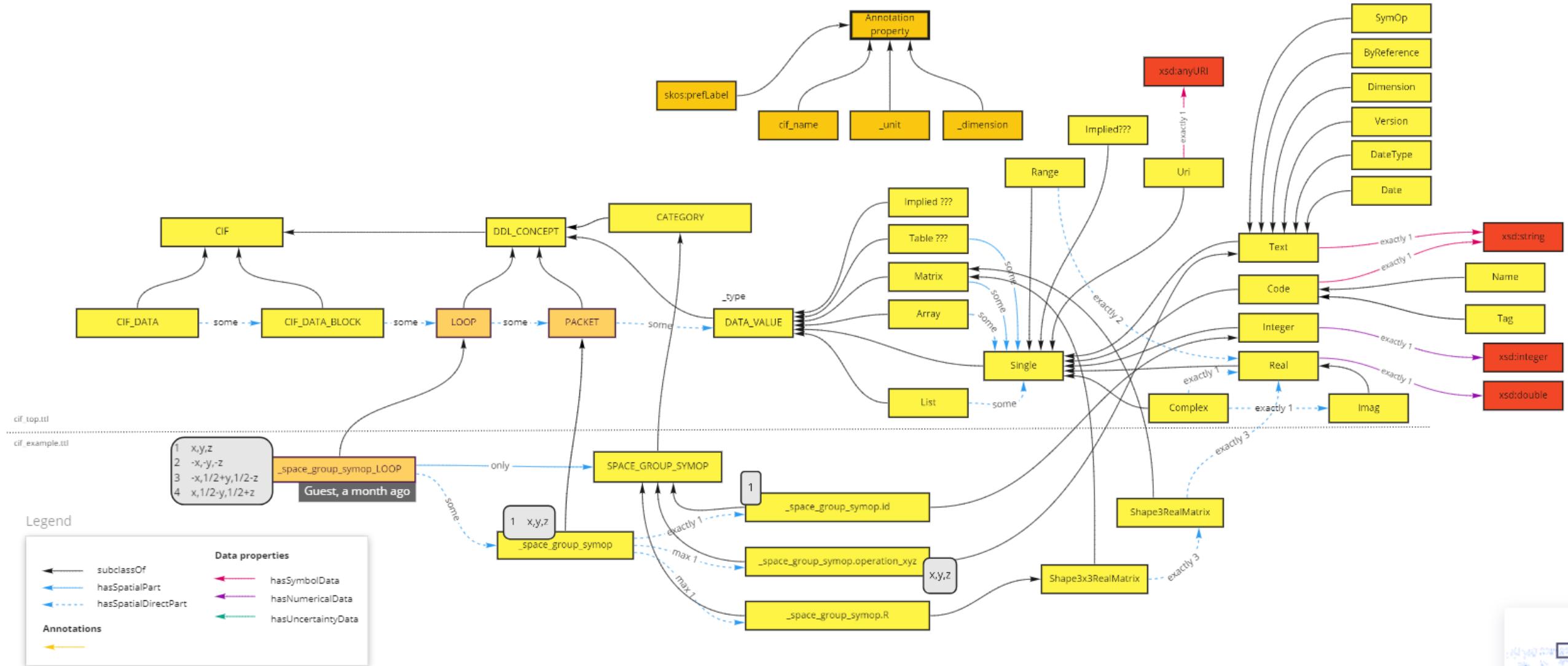
Generated ontologies



## Benefits

- Based on existing community standards
- Can be reused with any top-level ontology
- Ontological standard for atomic structures used by other EMMO domain ontologies

# The cif\_top ontology



File Edit View Reasoner Tools Refactor Window Help

< > cif\_core ([http://emmo.info/domain-crystallography/0.0.1/cif\\_core](http://emmo.info/domain-crystallography/0.0.1/cif_core))

&gt; CIF &gt; DDL\_CONCEPT &gt; DATA\_VALUE &gt; Matrix &gt; Shape3x3RealMatrix &gt; space\_group\_symop.R

Active ontology x Entities x Individuals by class x DL Query x

Annotation properties Datatypes Individuals

Classes Object properties Data properties

Class hierarchy: space\_group\_symop.R

Asserted

Git: loop-and-packet

space\_group\_symop.R — <http://emmo.info/domain-crystallography>

OWLviz: space\_group\_symop.R

Annotations Usage

Asserted hierarchy Inferred hierarchy

Annotations +

skos:prefLabel [language: en]

space\_group\_symop.R

rdfs:comment [language: en]

A matrix containing the symmetry rotation operations of a space group

$$R = \begin{vmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{vmatrix}$$

skos:altLabel [language: en]

space\_group\_symop\_R

\_datatype

Real

\_dimension [type: xsd:string]

[3,3]

\_name [type: xsd:string]

\_space\_group\_symop.R

Description: space\_group\_symop.R

Equivalent To +

SubClass Of +

Shape3x3RealMatrix

SPACE\_GROUP\_SYMOP

General class axioms +

SubClass Of (Anonymous Ancestor)

hasSpatialPart some Single

hasSpatialDirectPart exactly 3

Shape3RealMatrix

Instances +

Target for Key +

Ontology metrics:

## Metrics

Axiom	10238
Logical axiom count	1559
Declaration axioms count	1255
Class count	1236
Object property count	2
Data property count	7
Individual count	0
Annotation Property count	14

## Class axioms

SubClassOf	1457
EquivalentClasses	0
DisjointClasses	1
GCI count	0
Hidden GCI Count	0

## Object property axioms

SubObjectPropertyOf	1
EquivalentObjectProperties	0
InverseObjectProperties	0
DisjointObjectProperties	0
FunctionalObjectProperty	0
InverseFunctionalObjectProperty	1
TransitiveObjectProperty	1
SymmetricObjectProperty	0
AsymmetricObjectProperty	1
ReflexiveObjectProperty	0
IrreflexiveObjectProperty	1
ObjectPropertyDomain	0
ObjectPropertyRange	0
SubPropertyChainOf	0

## Data property axioms

SubDataPropertyOf	5
EquivalentDataProperties	0
DisjointDataProperties	0
FunctionalDataProperty	0
DataPropertyDomain	0
DataPropertyRange	5

## Individual axioms

ClassAssertion	0
ObjectPropertyAssertion	0
DataPropertyAssertion	0
NegativeObjectPropertyAssertion	0
NegativeDataPropertyAssertion	0
SameIndividual	0
DifferentIndividuals	0

## Annotation axioms

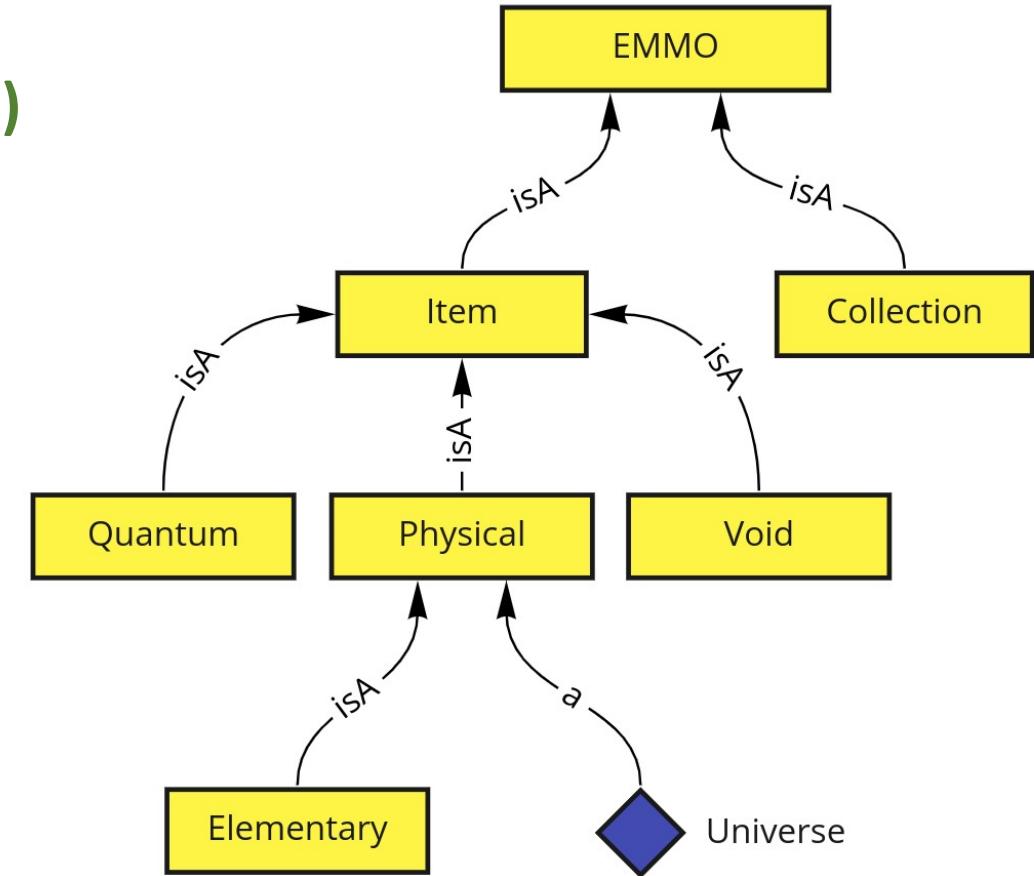
AnnotationAssertion	2
AnnotationPropertyDomain	2
AnnotationPropertyRangeOf	2

# Crystallography domain ontology

Based on EMMO

**European Materials Modelling Ontology (EMMO)**  
was initiated by EMMC in 2017

An ontology able to express concepts and  
constraints coming from materials modelling and  
**applied sciences**

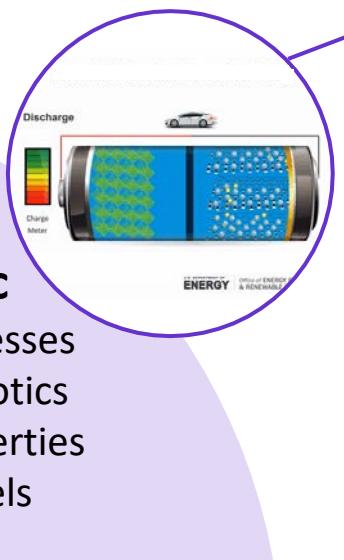


In EMMO there are no absolute definitions of the ontological nature of objects except for the Universe and the single indivisible quantum elements of which it is made up.

**Perspectives** allow categorizing things according to different views.

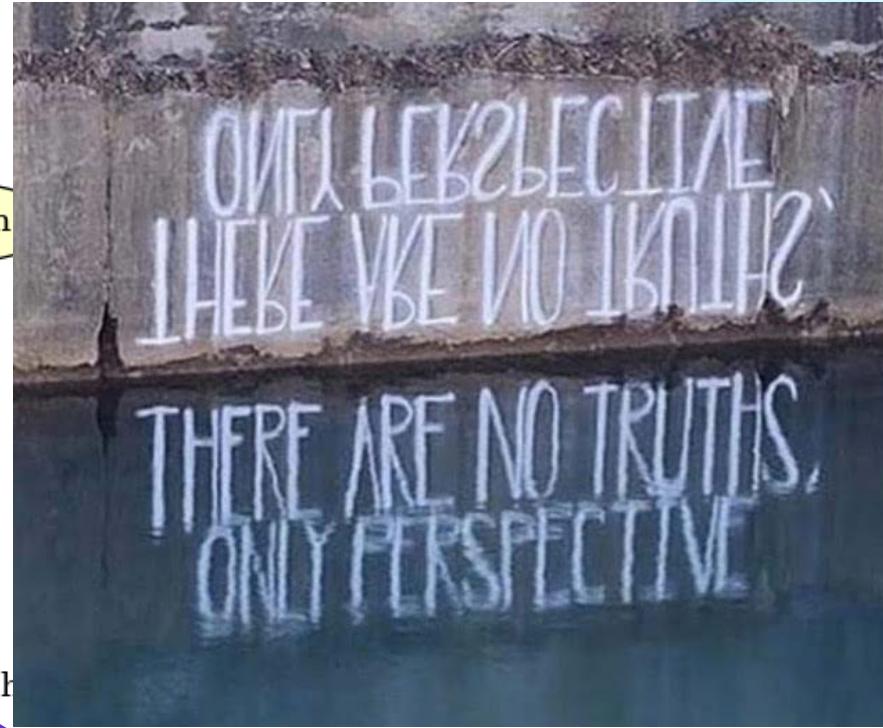
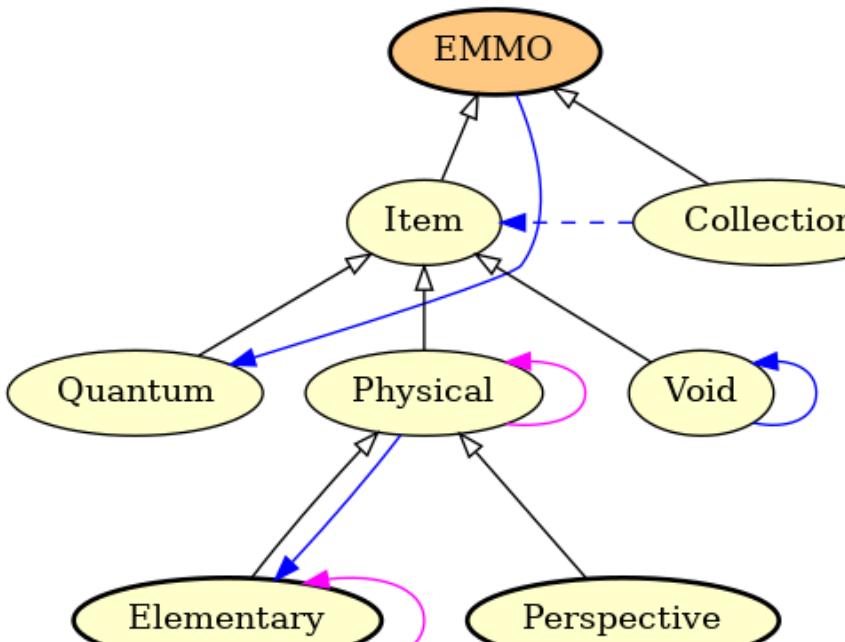
### Holistic

- processes
- semiotics
- properties
- models



### Reductionistic

- direct parthood
- countability
- ordering



### Physicalistic

- matter
- field
- material



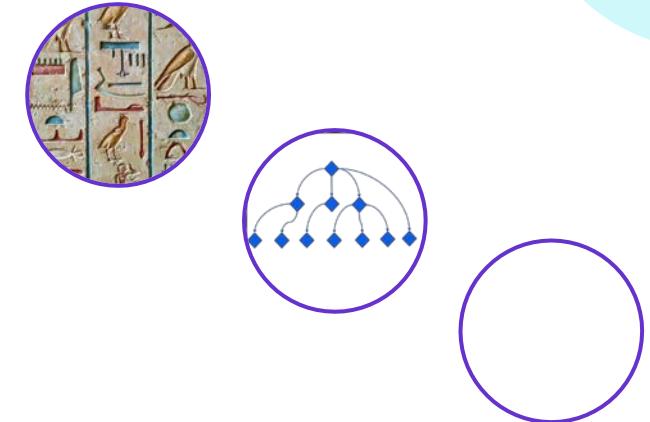
### Perceptual

- symbols
- languages
- metrology
- mathematics

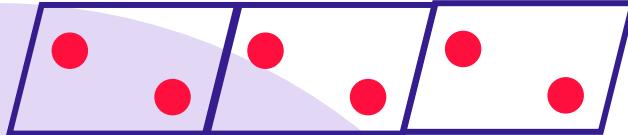
# Crystallography domain ontology

A semantic description of crystallography

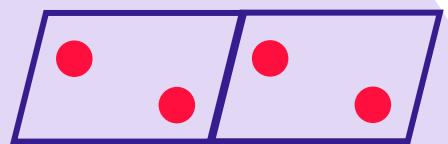
- Formulated as a formal language (Perceptual perspective)
- Utilised strict granularity and ordering (Reductionistic perspective)
- Refers to physical objects, like atoms (Physicalistic perspective)



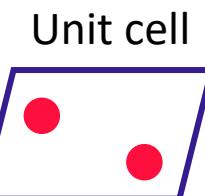
Infinitely repeated unit cells (crystal)



Finite structures

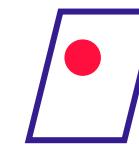


hasSpatialDirectPart



Unit cell

hasSpatialDirectPart



Asymmetric unit

hasSpatialDirectPart



Site, lattice vector

# Crystallography domain ontology

File Edit View Reasoner Tools Refactor Window Help

< > ⚖ crystallography (<http://emmo.info/domain-crystallography/0.0.1/crystallography>) Search...

EMMO > Item > Physical > Perspective > Perceptual > Graphical > Symbolic > Language > Crystallographical > UnitCell

Active ontology x Entities x Individuals by class x DL Query x

Annotation properties Datatypes Individuals  
Classes Object properties Data properties

Class hierarchy: UnitCell Inferred

Perspective  
  Holistic  
  Perceptual  
    Acoustical  
    Graphical  
      Geometrical  
      Pictorial  
      Representation  
    Symbolic  
      Language  
        Chemical  
      Crystallographical  
        CellVolume  
          CIF  
            CIF\_DATA  
            CIF\_DATA\_BLOCK  
            DDL\_CONCEPT  
              CATEGORY  
              DATA\_VALUE  
              LOOP  
              PACKET  
        CrystallographicSymbol  
          CrystalStructure  
            FractionalPositionA  
            FractionalPositionB  
            FractionalPositionC  
          LatticeParameter  
            Site  
          Spacegroup  
          SymmetryOperationString  
          UnitCell  
            ConventionalUnitCell  
            PrimitiveUnitCell

Annotations: UnitCell

Annotations +  
 skos:prefLabel [language: en]  
 UnitCell  
 rdfs:comment [language: en]  
 A unit cell is ca be described by either specifying the three lattice vectors that it is spanned by, or by 6 the lattice parameters (the length of and angles between the 3 lattice vectors).  
 cifLabel  
 \_cell\_[]  
 elucidation  
 A periodic unit in a crystal structure. It is spanned by three lattice vectors.

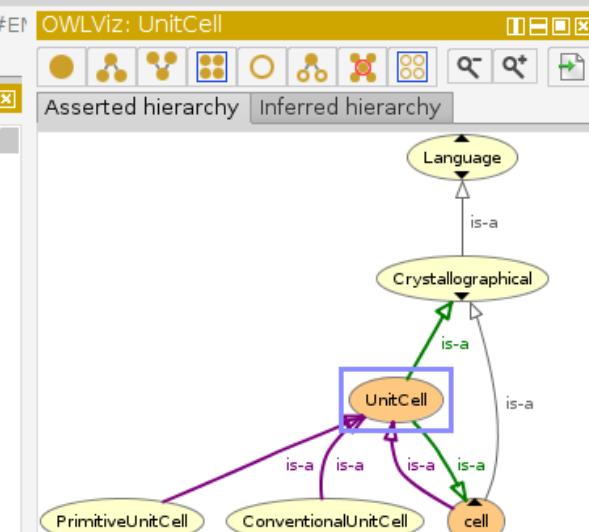
Description: UnitCell

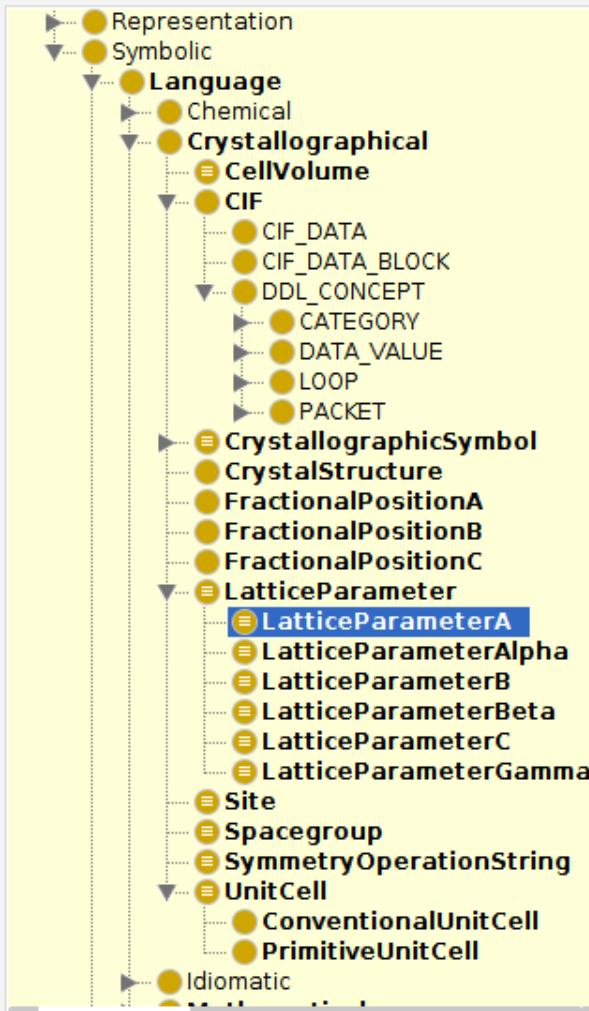
Equivalent To +  
 cell

SubClass Of +  
 Crystallographical  
 hasProperty exactly 6 LatticeParameter  
 hasProperty some CellVolume  
 hasSpatialDirectPart exactly 3 LatticeVector  
 Matter  
 Object  
 State

OWLviz: UnitCell

Asserted hierarchy Inferred hierarchy





skos:prefLabel [language: en]

LatticeParameterA

cifLabel

\_cell\_length\_a

cifMatch

[https://www.iucr.org/\\_data/iucr/cifdic\\_html/1/cif\\_core.dic/\\_cell\\_length\\_.html](https://www.iucr.org/_data/iucr/cifdic_html/1/cif_core.dic/_cell_length_.html)

elucidation

The length of lattice vectors 'a', where lattice vectors 'a', 'b' and 'c' defines the unit cell,

physicalDimension

T0 L+1 M0 I0 θ0 NO JO

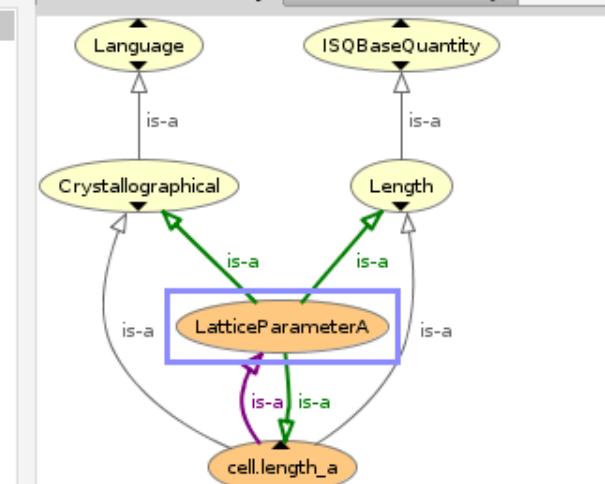
cell.length\_a

Crystallographical

hasQuantityValue some Real

Length

LatticeParameter



# Summary

A practical example of development of a domain ontology

- reuse of existing standards
- fits the needs of different communities
- showed how it is connected to EMMO
- the approach allows to connect to other TLOs as well



@ontocommons



linkedin.com/company/ontocommons



info@ontocommons.eu

# Acknowledgement

## Projects

- OntoCommons (2020-2023) GA no 862136
- MarketPlace (2018-2022) GA no 760173
- OntoTrans (2020-2024) GA no 862136
- BIG-MAP (2020-2023) GA no 957189



## Contributors

- Casper Welzel Andersen, EPFL
- James Hester, ANSTO
- Saulius Grazulis, Vilnius University
- Rickard Armiento, Linköping Universitet
- Francesca Lønstad Bleken, SINTEF
- Joana Morgado, Fraunhofer IWM
- Stuart Chalk, UNF

