Improving findability on chemical resources with Bioschemas

Leyla Jael Castro,
ZB MED Information Centre for Life Sciences
Bioschemas

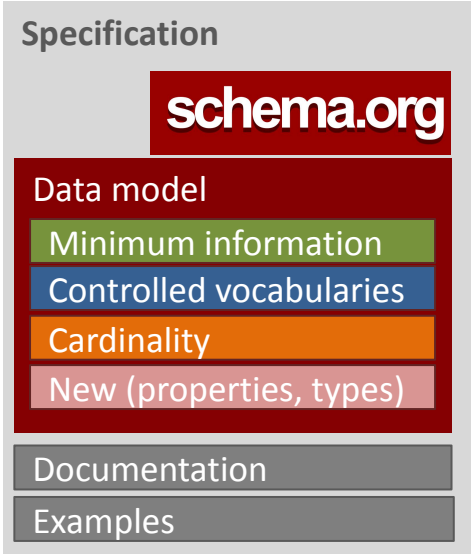
Egon Willighagen
Maastricht University
Bioschemas

ACS FALL 2023
13-17 August, San Francisco CA & Hybrid

doi:10.5281/zenodo.8369861

Bioschemas: metadata for (life) sciences websites and data feeds

- Community initiative built on top of **schema.org**
- Aim
  - Improve data discoverability and interoperability in Life Sciences
- Approach
  - Add Life Science types to schema.org
  - Provide usage guidelines and examples
    - Minimum, recommended, optional
    - Link to domain ontologies
  - Support software



Bioschemas

References: https://scholia.toolforge.org/topic/Q93995803

Slide modified from previous presentation

FAIRness support in Bioschemas

- F1, F3. Promote - use of unique identifiers
- F2. Provide - rich metadata specifications
- F4. Promote - use of registries

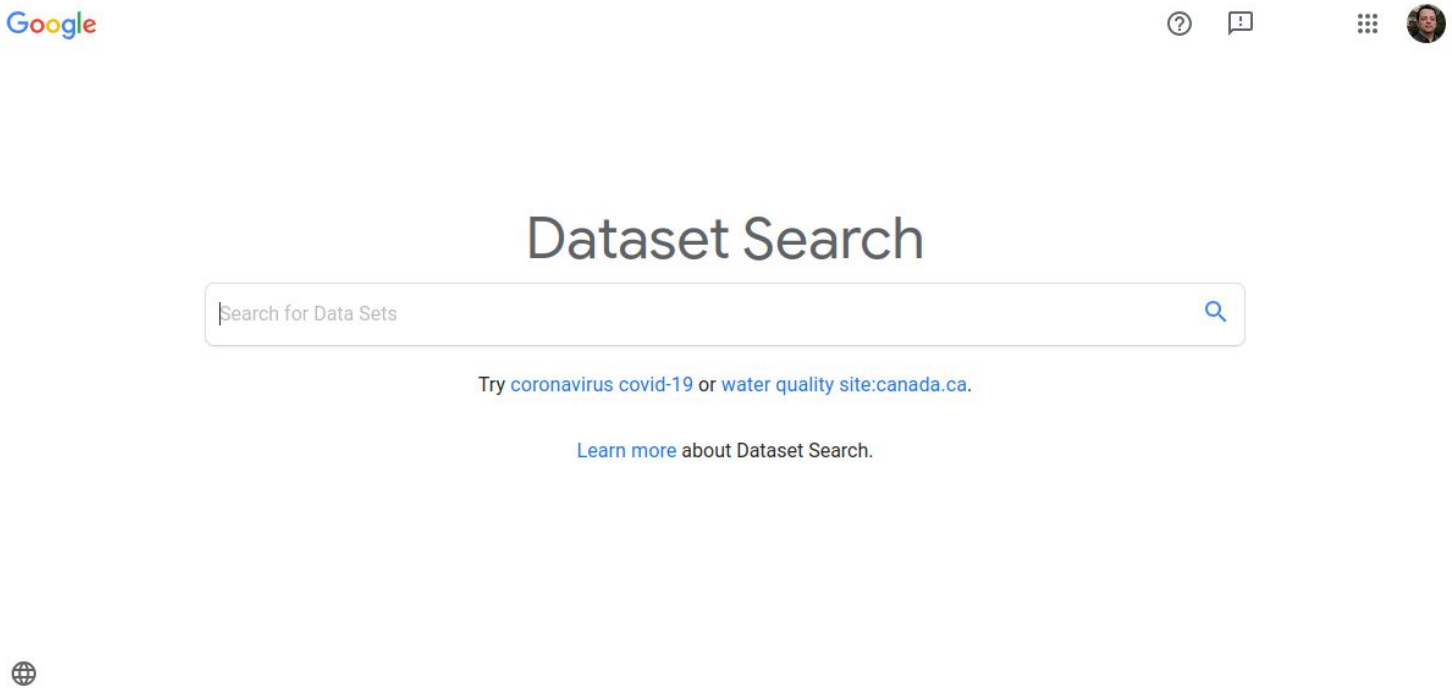
- A1. Use - HTTP(S)
- A2. Promote - use of registries

- I1. Use - JSON-LD
- I2. Use - schema.org
- I3. Provide - metadata specifications linking objects to each other

- R1. Provide - metadata specifications (minimum, recommended, optional)
- R1.1. Promote - use of licenses
- R1.2 Promote - provenance and attribution
- R.1.3 Provide - community standards

Slide modified from previous presentation

Making things findable



Bioschemas in use

Bioschemas defines domain-specific **profiles** to add structured metadata to Life Science resources on the Web by using and expanding **schema.org:** *ChemicalSubstance, Gene, MolecularEntity, Protein, ProteinStructure, Sample, Taxon*

General-purpose profiles are being lifted for use *beyond* life sciences (https://schemas.science/) including *Dataset, Course, ComputationalWorkflow, ComputationalTool, TrainingMaterial.*

Bioschemas markup is **deployed** at >90 websites

Bioschemas is developed and maintained by working groups in an active **community** (>150 members)

Name	Group	Use Cases	Cross Walk	Task & Issues	Exampl
ChemicalSubstance (v0.4-RELEASE) 07 April 2020	Chemicals				
ComputationalTool (v1.0-RELEASE) 11 October 2021	Tools				
ComputationalWorkflow (v1.0-RELEASE) 09 March 2021	Workflow				
Course (v1.0-RELEASE) 13 September 2022	Training				
CourseInstance (v1.0-RELEASE) 13 September 2022	Training				
DataCatalog (v0.3-RELEASE-2019_07_01) 01 July 2019	Data Repositories				
Dataset (v1.0-RELEASE) 12 July 2022	Datasets				
FormalParameter (v1.0-RELEASE) 09 March 2021	Workflow				
Gene (v1.0-RELEASE) 07 April 2021	Genes				
MolecularEntity (v0.5-RELEASE) 07 April 2020	Chemicals				
Protein (v0.11-RELEASE) 07 April 2020	Proteins				
Sample (v0.2-RELEASE-2018_11_10) 10 November 2018	Samples				
Taxon	Biodiversity				

https://bioschemas.org/profiles/

Slide modified from previous presentation

- Metadata Interoperability and Standardization across all website using Schema.org

3. Schema.org Formats

Schema.org markup can be embedded in your web page using Microdata, RDFa, or JSON-LD. In Bioschemas we favor the use of JSON-LD, as do most of the search engines.

3.1. Format examples

- JSON-LD

```
<script type="application/ld+json">
  {
    "@context": "https://schema.org",
    "@type": "SportsTeam",
    "name": "San Francisco 49ers",
    "member": {
      "@type": "OrganizationRole",
      "member": {
        "@type": "Person",
        "name": "Joe Montana"
      },
      "startDate": "1979",
      "endDate": "1992",
      "roleName": "Quarterback"
    }
  }
</script>
```

- Microdata

```
<div itemscope itemtype="http://schema.org/SportsTeam">
  <span itemprop="name">San Francisco 49ers</span>
```

Contributors:

- Leyla Garcia
- Victoria Dominguez del Angel
- Alasdair Gray

Version: 2.1

Last Modified: 23 April 2021

https://bioschemas.org/tutorials/what_why_schema

Dataset Profile

Version: 1.0-RELEASE (12 July 2022)

Property	Expected Type	Description	CD	Controlled Vocabulary	Example
Marginality: Minimum.					
@context	URL	Used to provide the context (namespaces) for the JSON-LD file. Not needed in other serialisations.	ONE		
@type	Text	Schema.org/Bioschemas class for the resource declared using JSON-LD syntax. For other serialisations please use the appropriate mechanism. While it is permissible to provide multiple types, it is preferred to use a single type.	MANY	Schema.org, Bioschemas	
@id	IRI	Used to distinguish the resource being described in JSON-LD. For other serialisations use the appropriate approach.	ONE		
dct:conformsTo	IRI	Used to state the Bioschemas profile that the markup relates to. The versioned URL of the profile must be used. Note that we use a CURIE in the table here but the full URL for Dublin Core terms must be used in the markup (http://purl.org/dc/terms/conformsTo), see example.	ONE	Bioschemas profile versioned URL	
description	Text	**Schema:** A description of the item. **Bioschemas:** A short summary describing a dataset.	ONE		
identifier	PropertyValue Text URL	**Schema:** The identifier property represents any kind of identifier for any kind of Thing, such as ISBNs, GTIN codes, UUIDs etc. Schema.org provides dedicated properties for representing many of these, either as textual strings or as URL (URI) links. See background notes for more details. **Bioschemas:** CURIEs that can be resolved using Identifiers.org should be used.	MANY		
keywords	DefinedTerm Text URL	**Schema:** Keywords or tags used to describe this content. Multiple entries in a keywords list are typically delimited by commas. **Bioschemas:** Keywords should be drawn from a controlled vocabulary, e.g. EDAM, and supplied as a DefinedTerm list.	MANY		
license	CreativeWork URL	**Schema:** A license document that applies to this content, typically indicated by URL. **Bioschemas:** A license under which the dataset is distributed.	ONE		
name	Text	**Schema:** The name of the item. **Bioschemas:** A descriptive name of the dataset.	ONE		
url	URL	**Schema:** URL of the item. **Bioschemas:** The location of a page describing the dataset.	ONE		

Dataset profiles in use

Gene / Protein ID mapping databases

JSON source

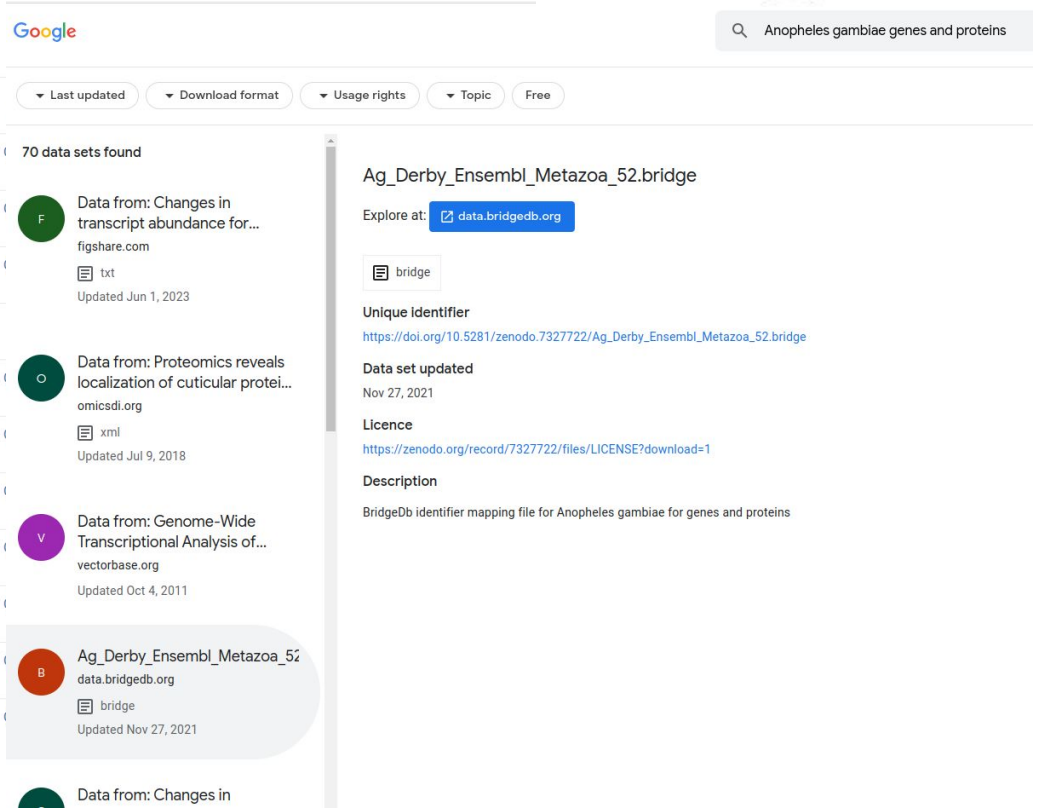
Species	BridgeDb Download	QC report	Size	License	Date	Tested with
Anopheles gambiae (ncbitaxon:7165)	Ag_Derby_Ensembl_Metazoa_52.bridge (doi:10.5281/zenodo.7327722)	QC	42.9 MB	license	2023-01-06	PathVisio 3.3 BridgeDb Webservice
Aspergillus niger (ncbitaxon:5061)	An_Derby_Ensembl_Fungi_52.bridge (doi:10.5281/zenodo.7781891)	QC	34.3 MB	license	2023-03-30	PathVisio 3.3 BridgeDb Webservice
Arabidopsis thaliana (ncbitaxon:3702)	At_Derby_Ensembl_Plant_52.bridge (doi:10.5281/zenodo.7781895)	QC	213.4 MB	license	2023-03-31	PathVisio 3.3 BridgeDb Webservice
Bacillus subtilis (ncbitaxon:1423)	Bs_Derby_Ensembl_91.bridge (doi:10.5281/zenodo.3667670)		8.2 MB	license	2020-02-14	BridgeDb Webservice
Bos taurus (ncbitaxon:9913)	Bt_Derby_Ensembl_108.bridge (doi:10.5281/zenodo.7781913)	QC	153.5 MB	license	2023-03-31	PathVisio 3.3 BridgeDb Webservice
Caenorhabditis elegans (ncbitaxon:6239)	Ce_Derby_Ensembl_108.bridge (doi:10.5281/zenodo.7781913)	QC	208.9 MB	license	2023-03-31	PathVisio 3.3 BridgeDb Webservice
Canis familiaris (ncbitaxon:9615)	Cf_Derby_Ensembl_108.bridge (doi:10.5281/zenodo.7781913)	QC	162.4 MB	license	2023-03-31	PathVisio 3.3 BridgeDb Webservice
Ciona intestinalis (ncbitaxon:7719)	Ci_Derby_Ensembl_108.bridge (doi:10.5281/zenodo.7781913)	QC	32.9 MB	license	2023-03-31	PathVisio 3.3 BridgeDb Webservice
Danio rerio (ncbitaxon:7955)	Dr_Derby_Ensembl_108.bridge (doi:10.5281/zenodo.7781913)	QC	272.0 MB	license	2023-03-31	PathVisio 3.3 BridgeDb Webservice
Daphnia magna (ncbitaxon:35525)	Da_Derby_Ensembl_Metazoa_52.bridge (doi:10.5281/zenodo.7327722)	QC	45.5 MB	license	2023-01-06	PathVisio 3.3 BridgeDb Webservice
Daphnia pulex (ncbitaxon:6669)	Dp_Derby_Ensembl_Metazoa_52.bridge	QC	60.8 MB	license	2023-01-	PathVisio 3.3 BridgeDb

Dataset profiles in use

Gene / Protein ID mapping databases

JSON source

Species	BridgeDb Download
Anopheles gambiae (ncbitaxon:7165)	Ag_Derby_Ensembl_Metazoa_52.bridge (doi:10.5281/zenodo.7327722)
Aspergillus niger (ncbitaxon:5061)	An_Derby_Ensembl_Fungi_52.bridge (doi:10.5281/zenodo.7781891)
Arabidopsis thaliana (ncbitaxon:3702)	At_Derby_Ensembl_Plant_52.bridge (doi:10.5281/zenodo.7781895)
Bacillus subtilis (ncbitaxon:1423)	Bs_Derby_Ensembl_91.bridge (doi:10.5281/zenodo.3667670)
Bos taurus (ncbitaxon:9913)	Bt_Derby_Ensembl_108.bridge (doi:10.5281/zenodo.7781913)
Caenorhabditis elegans (ncbitaxon:6239)	Ce_Derby_Ensembl_108.bridge (doi:10.5281/zenodo.7781913)
Canis familiaris (ncbitaxon:9615)	Cf_Derby_Ensembl_108.bridge (doi:10.5281/zenodo.7781913)
Ciona intestinalis (ncbitaxon:7719)	Ci_Derby_Ensembl_108.bridge (doi:10.5281/zenodo.7781913)
Danio rerio (ncbitaxon:7955)	Dr_Derby_Ensembl_108.bridge (doi:10.5281/zenodo.7781913)
Daphnia magna (ncbitaxon:35525)	Da_Derby_Ensembl_Metazoa_52.bridge (doi:10.5281/zenodo.7327722)
Daphnia pulex (ncbitaxon:6669)	Dp_Derby_Ensembl_Metazoa_52.bridge



Dataset profiles in use #2

datasets

DOI 10.5281/zenodo.7088624

Overview of open datasets released by NanoSafety Cluster projects

NanoCommons
Nano-Knowledge Community

This list provides an overview of archived datasets with an open license. Each one of them can be cited with DataCite and various datasets are not only downloaded from the archives but can also be interactively explored via databases.

Additional datasets and databases that provide interactive access to these datasets can be reported here.

Physico-chemical characterization of sterile Fe3O4 nanoparticles by XPS / HAXPES / SEM

- Date: 2023-05-31
- License: Creative Commons Attribution Non Commercial 3.0 Germany
- Project: NanoSolveIT
- URL: https://zenodo.org/record/7990302
- DOI: 10.5281/zenodo.7990302

Physico-chemical characterization of sterile citrated stabilized Au nanoparticles by XPS / HAXPES / SEM

- Date: 2023-05-31
- License: Creative Commons Attribution Non Commercial 3.0 Germany
- Project: NanoSolveIT
- URL: https://zenodo.org/record/7990251
- DOI: 10.5281/zenodo.7990251

Physico-chemical characterization of sterile ZnO nanoparticles by

https://nanocommons.github.io/datasets/

Scraping Bioschemas

sbd-data-nanocommons / extractBioschemas.groovy

Code | Blame | 70 lines (56 loc) · 2.79 KB

```
bioclipse = new net.bioclipse.managers.BioclipseManager(".");
rdf = new net.bioclipse.managers.RDFManager(".");
jsoup = new net.bioclipse.managers.JSoupManager(".");

htmlContent = bioclipse.download("https://nanocommons.github.io/datasets/")

htmlDom = jsoup.parseString(htmlContent)

// application/ld+json

bioschemasSections = jsoup.select(htmlDom, "script[type='application/ld+json']");

kg = rdf.createInMemoryStore()

for (section in bioschemasSections) {
  bioschemasJSON = section.html()
  rdf.importFromString(kg, bioschemasJSON, "JSON-LD")
}

turtle = rdf.asTurtle(kg);

println "#" + rdf.size(kg) + " triples detected in the JSON-LD"
//println turtle

sparql = """
PREFIX schema: <http://schema.org/>
SELECT ?dataset ?url ?name ?license ?description WHERE {
  ?dataset a schema:Dataset ;
    schema:url ?url .
  OPTIONAL { ?dataset schema:name ?name }
  OPTIONAL { ?dataset schema:license ?license }
  OPTIONAL { ?dataset schema:description ?description }
} ORDER BY ASC(?dataset)
"""
```

SbD Nano 4
SAFE BY DESIGN FOR NANO

sbd-data-book

allDatasources.rq

Code examples: curl

SPARQL

```
PREFIX void:    <http://rdfs.org/ns/void#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dc:      <http://purl.org/dc/elements/1.1/>

SELECT ?resource ?resourceLabel ?img WHERE {
  ?void a void:DatasetDescription ;
    dcterms:title|dc:title ?resourceLabel .
  OPTIONAL { ?void dc:source ?resource }
  OPTIONAL { ?void foaf:img ?img }
}
```

run

Output

resource	img
AOP-Wiki Key Event Relationships	A-O-P-Wiki
NanoWiki v6 causal relationships	eNM eNanoMapper
SmartNanoTox Transcriptomics datasets	SmartNanoTox Smart Tools for Gauging Nano Hazards
Overview of open datasets released by NanoSafety Cluster projects	NanoCommons Nano-Knowledge Community
NanoInformaTIX Tools	nix

https://h2020-sbd4nano.github.io/sbd-data-book/

Chemical profiles in use

Sites: 2 ChemicalSubstance (0.4-RELEASE)

- NanoCommons v0.4
- RiskGONE v0.4

Sites: 9 MolecularEntity (0.5-RELEASE)

ChEMBL v0.2-DRAFT		Page	SMV	BMUSE
ChEMBL (Mirror) v0.5-RELEASE	Over 1.9 million molecular entities			
iPPI-DB v0.5	2,035 compounds	Page	SMV	BMUSE
IUPHAR/BPS Guide to Pharmacology v0.2-DRAFT	Over 9,800 ligands	Page	SMV	BMUSE
MassBank Europe v0.5		Page	SMV	BMUSE
Metabolic Atlas v0.5-RELEASE	Over 3'000 metabolites.	Page	SMV	BMUSE
MetaNetX v0.5-RELEASE	Over 1M chemical compound pages	Page	SMV	BMUSE
Rhea (Unofficial) v0.5-RELEASE	Over 12,000 chemicals			
Scholia v0.5	More than 1.2M pages annotated with chemical compounds. Live statistics from https://scholia.toolforge.org/chemical/.	Page	SMV	BMUSE

MolecularEntity Profile

Version: 0.5-RELEASE (07 April 2020)

Bioschemas profile describing a MolecularEntity

If you spot any errors or omissions with this type, please file an issue in our GitHub.

Description | Contributors | Links

Schema.org hierarchy

This Profile fits into the schema.org hierarchy as follows:

Thing > BioChemEntity > MolecularEntity

Description

Any constitutionally or isotopically distinct atom, molecule, ion, ion pair, radical, radical ion, complex, conformer etc., identifiable as a distinguishable entity. (Source: ChEBI:23367)

Property	Expected Type	Description	CD	Controlled Vocabulary	Example
Marginality: Minimum.					
@context	URL	Used to provide the context (namespaces) for the JSON-LD file. Not needed in other serialisations.	ONE		
@type	Text	Schema.org/Bioschemas class for the resource declared using JSON-LD syntax. For other serialisations please use the appropriate mechanism. While it is permissible to provide multiple types, it is preferred to use a single type.	MANY	Schema.org, Bioschemas	
@id	IRI	Used to distinguish the resource being described in JSON-LD. For other serialisations use the appropriate approach.	ONE		
dct:conformsTo	IRI	Used to state the Bioschemas profile that the markup relates to. The versioned URL of the profile must be used. Note that we use a CURIE in the table here but the full URL for Dublin Core terms must be used in the markup (http://purl.org/dc/terms/conformsTo), see example.	ONE	Bioschemas profile versioned URL	
identifier	PropertyValue Text URL	**Schema:** The identifier property represents any kind of identifier for any kind ofThing, such as ISBNs, GTIN codes, UUIDs etc. Schema.org provides dedicated properties for representing many of these, either as textual strings or as URL (URI) links. See background notes for more details.	ONE		
name	Text	**Schema:** The name of the item.	ONE		
url	URL	**Schema:** URL of the item.	MANY		

MolecularEntity Profile

Version: 0.5-RELEASE (07 April 2020)

Bioschemas profile describing a MolecularEntity

If you spot any errors or omissions with this type, please file an issue in our GitHub.

Description | Contributors | Links

Schema.org hierarchy

This Profile fits into the schema.org hierarchy as follows:

Thing > BioChemEntity > MolecularEntity

Description

Any constitutionally or isotopically distinct atom, molecule, ion, ion pair, radical, radical ion, complex, conformer etc., identifiable as a distinguishable entity. (Source: ChEBI:23367)

Marginality: Recommended.				
inChI	Text	**Schema:** Non-proprietary identifier for molecular entity that can be used in printed and electronic data sources thus enabling easier linking of diverse data compilations.	ONE	
inChIKey	Text	**Schema:** InChIKey is a hashed version of the full InChI (using the SHA-256 algorithm).	ONE	
iupacName	Text	**Schema:** Systematic method of naming chemical compounds as recommended by the International Union of Pure and Applied Chemistry (IUPAC).	ONE	
molecularFormula	Text	**Schema:** The empirical formula is the simplest whole number ratio of all the atoms in a molecule.	ONE	
molecularWeight	QuantitativeValue Text	**Schema:** This is the molecular weight of the entity being described, not of the parent. Units should be included in the form ' ', for example '12 amu' or as '.	ONE	
smiles	Text	**Schema:** A specification in form of a line notation for describing the structure of chemical species using short ASCII strings. Double bond stereochemistry \ indicators may need to be escaped in the string in formats where the backslash is an escape character.	MANY	

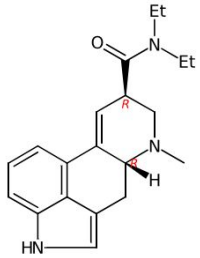
MolecularEntity in use: Scholia

SCHOLIA Author Work Organization Location Event Project Award Topic Tools Help Search...

chemical / Q23118 Improve data

lysergic acid diethylamide (Q23118)

Lysergic acid diethylamide, commonly known as LSD, and known colloquially as acid, is a potent psychedelic drug. Effects typically include intensified thoughts, emotions, and sensory perception. At sufficiently high dosages LSD manifests primarily mental, visual, and auditory hallucinations. Dilated pupils, increased blood pressure, and increased body temperature are typical. Effects typically begin within half an hour and can last for up to 20 hours. LSD is also capable of causing mystical experiences and ego dissolution. It is used mainly as a recreational drug or for spiritual reasons. LSD is both the prototypical psychedelic and one of the "classical" psychedelics, being the psychedelic with the greatest scientific and cultural significance. LSD is synthesized as a solid compound, typically in the form of a powder or a crystalline material. This solid LSD is then dissolved in a liquid solvent, such as ethanol or distilled water, to create a solution. The liquid serves as a carrier for the LSD, allowing for accurate dosage and administration onto small pieces of blotter paper called tabs. LSD is typically either swallowed or held under the tongue. (Read more on English Wikipedia)



Related: D-methamphetamine · nitrobenzene · biuret · dibutyl phthalate · N,N-dimethylformamide

SCHOLIA

chemical / Q23118 Improve data

lysergic acid diethylamide (Q23118)

Lysergic acid diethylamide, commonly known as LSD, and known colloquially as acid, is a potent psychedelic drug. Effects typically include intensified thoughts, emotions, and sensory perception. At sufficiently high dosages LSD manifests primarily mental, visual, and auditory hallucinations. Dilated

Elements Console Sources Network Performance Memory Application Security 8

```
<!DOCTYPE html>
<html>
▶ <head> … </head>
▼ <body> flex
  ▶ <nav class="navbar navbar-expand-lg" style="max-width: 1200px;"> … </nav> flex
  ▼ <div class="content">
    ▶ <div class="container d-flex justify-content-between"> … </div> flex
    ▼ <div class="container" data-protonpass-cluster data-protonpass-form>
      ▶ <h1 id="h1"> … </h1>
      ▼ <script type="application/ld+json" id="bioschemas"> == $0
        {"@context":"https://schema.org","@type":"MolecularEntity","http://purl.org/dc/terms/conformsTo":
        {"@type":"CreativeWork","@id":"https://bioschemas.org/profiles/MolecularEntity/0.5-
        RELEASE/"},"identifier":"Q23118","inChIKey":"VAYOSLLFUXYJDT-RDTXWAMCSA-
        N","url":"http://www.wikidata.org/entity/Q23118","name":"lysergic acid
        diethylamide","inChI":"InChI=1S/C20H25N3O/c1-4-23(5-2)20(24)14-9-16-15-7-6-8-17-19(15)13(11-21-17)10-
        18(16)22(3)12-14/h6-9,11,14,18,21H,4-5,10,12H2,1-3H3/t14-,18-/m1/s1","molecularFormula":"CCN(CC)C(=O)
        [C@H]1CN([C@@H]2CC3=CNC4=CC=CC(=C34)C2=C1)C"}
      </script>
    ▶ <div class="card-deck mb-3"> … </div> flex
    ▶ <div id="wembedder"> … </div>
```

ChemicalSubstance Profile

Version: 0.4-RELEASE (07 April 2020)

Bioschemas profile describing a ChemicalEntity

If you spot any errors or omissions with this type, please file an issue in our GitHub.

Description | Contributors | Links

Schema.org hierarchy

This Profile fits into the schema.org hierarchy as follows:

Thing > BioChemEntity > ChemicalSubstance

Description

This profile describes a ChemicalSubstance which is 'a portion of matter of constant composition, composed of molecular entities of the same type or of different types' (source: ChEBI:59999).

Property	Expected Type	Description	CD	Controlled Vocabulary	Example
Marginality: Minimum.					
@context	URL	Used to provide the context (namespaces) for the JSON-LD file. Not needed in other serialisations.	ONE		
@type	Text	Schema.org/Bioschemas class for the resource declared using JSON-LD syntax. For other serialisations please use the appropriate mechanism. While it is permissible to provide multiple types, it is preferred to use a single type.	MANY	Schema.org, Bioschemas	
@id	IRI	Used to distinguish the resource being described in JSON-LD. For other serialisations use the appropriate approach.	ONE		
dct:conformsTo	IRI	Used to state the Bioschemas profile that the markup relates to. The versioned URL of the profile must be used. Note that we use a CURIE in the table here but the full URL for Dublin Core terms must be used in the markup (http://purl.org/dc/terms/conformsTo), see example.	ONE	Bioschemas profile versioned URL	
identifier	PropertyValue Text URL	**Schema:** The identifier property represents any kind of identifier for any kind ofThing, such as ISBNs, GTIN codes, UUIDs etc. Schema.org provides dedicated properties for representing many of these, either as textual strings or as URL (URI) links. See background notes for more details.	ONE		
name	Text	**Schema:** The name of the item.	ONE		
url	URL	**Schema:** URL of the item.	MANY		
Marginality: Recommended.					
chemicalComposition	Text	**Schema:** The chemical composition describes the identity and relative ratio of the chemical elements that make up the substance. **Bioschemas:** For substances this often cannot be accurately determined, an approximation is acceptable.	ONE		
hasBioChemEntityPart	BioChemEntity	**Schema:** Indicates a BioChemEntity that (in some sense) has this BioChemEntity as a part. Inverse property:isPartOfBioChemEntity.	MANY		

ChemicalSubstance in use: NanoCommons

Working draft

Ontology IRIs for the JRC representative industrial nanomaterials

NanoCommons Working Draft 06 June 2023

This version:
http://nanocommons.github.io/specifications/2023/WD-jrc-20230606/
Latest published version:
http://nanocommons.github.io/specifications/jrc/
Previous version:
http://nanocommons.github.io/specifications/2023/WD-jrc-20230529/
Editor:
Egon Willighagen, Maastricht University

Abstract

The JRC representative industrial nanomaterials are a series of nanomaterials used in the European nanosafety community for research [*Totaro2016*]. NanoCommons is developing an ontology to be used as common language by this community. This document links the JRC materials to specific terms in the ontology and provides identifiers for each of them. It is based on an earlier specification by eNanoMapper [*Jeliazkova2015*], available at https://enanomapper.github.io/specifications/jrc/. The sources can be found at https://github.com/nanocommons/specifications.

Status of This Document

This document is a specification by NanoCommons. It has no official standing of any kind and does not represent the support or consensus of any standards organisation.

ChemicalSubstance

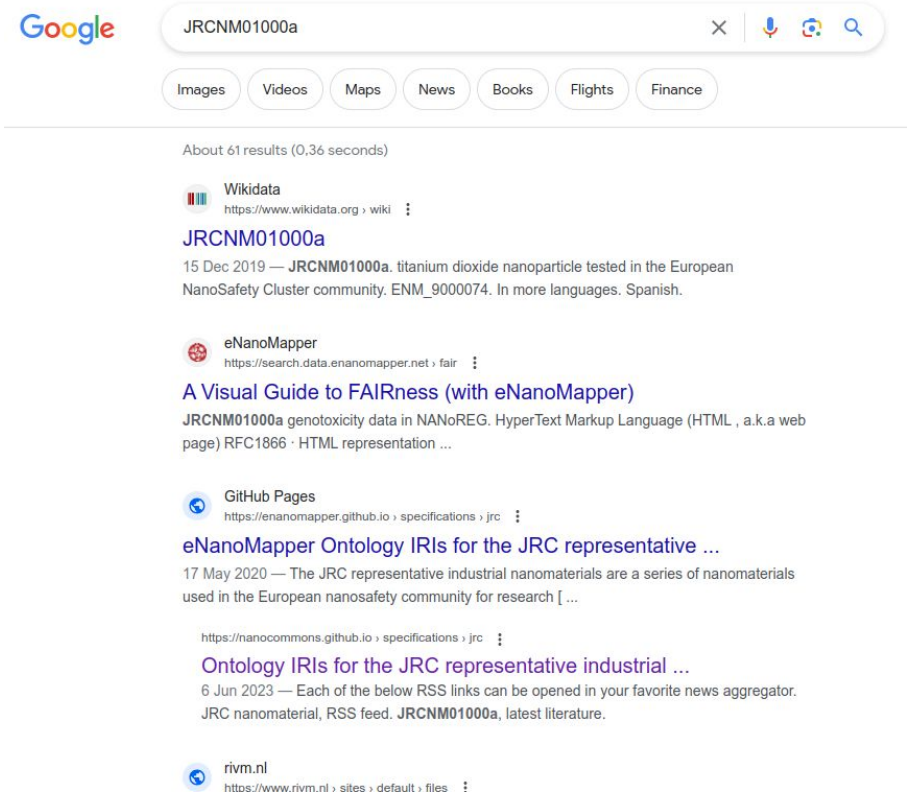
JRC nanomaterial Code	Ontology IRI	Wikidata	Compact Identifier	ERM Identifier
JRCNM01000a ENM_9000074	http://purl.enanomapper.org/onto/ENM_9000074	Q27918612	wikidata:Q27918612	
JRCNM01001a ENM_9000075	http://purl.enanomapper.org/onto/ENM_9000075	Q47461406	wikidata:Q47461406	
JRCNM01002a ENM_9000076	http://purl.enanomapper.org/onto/ENM_9000076	Q47461416	wikidata:Q47461416	
JRCNM01003a ENM_9000083	http://purl.enanomapper.org/onto/ENM_9000083	Q47461418	wikidata:Q47461418	
JRCNM01004a ENM_9000084	http://purl.enanomapper.org/onto/ENM_9000084	Q47461419	wikidata:Q47461419	
JRCNM01005a ENM_9000077	http://purl.enanomapper.org/onto/ENM_9000077	Q47461422	wikidata:Q47461422	erm:ERM00000064
JRCNM01100a ENM_9000078	http://purl.enanomapper.org/onto/ENM_9000078	Q47462004	wikidata:Q47462004	
JRCNM01101a ENM_9000086	http://purl.enanomapper.org/onto/ENM_9000086	Q47462008	wikidata:Q47462008	
JRCNM02000a ENM_9000087	http://purl.enanomapper.org/onto/ENM_9000087	Q47462022	wikidata:Q47462022	
JRCNM02001a ENM_9000088	http://purl.enanomapper.org/onto/ENM_9000088	Q47468470	wikidata:Q47468470	
JRCNM02002a ENM_9000089	http://purl.enanomapper.org/onto/ENM_9000089	Q47468473	wikidata:Q47468473	
JRCNM02003a ENM_9000090	http://purl.enanomapper.org/onto/ENM_9000090	Q78642247	wikidata:Q78642247	
JRCNM02004a ENM_9000091	http://purl.enanomapper.org/onto/ENM_9000091	Q47468478	wikidata:Q47468478	

ChemicalSubstance

JRC nanomaterial Code	Ontology IRI	Wikidata	Compact Identifier	ERM Identifier
JRCNM01000a ENM_9000074	http://purl.enanomapper.org/onto/ENM_9000074	Q27918612	wikidata:Q27918612	
JRCNM01001a ENM_9000075	http://purl.enanomapper.org/onto/ENM_9000075	Q47461406	wikidata:Q47461406	
JRCNM01002a ENM_9000076	http://purl.enanomapper.org/onto/ENM_9000076	Q47461416	wikidata:Q47461416	
JRCNM01003a ENM_9000083	http://purl.enanomapper.org/onto/ENM_9000083	Q47461418	wikidata:Q47461418	
JRCNM01004a ENM_9000084	http://purl.enanomapper.org/onto/ENM_9000084	Q47461419	wikidata:Q47461419	
JRCNM01005a ENM_9000077	http://purl.enanomapper.org/onto/ENM_9000077	Q47461422	wikidata:Q47461422	erm:ERM00000064
JRCNM01100a ENM_9000078	http://purl.enanomapper.org/onto/ENM_9000078	Q47462004	wikidata:Q47462004	
JRCNM01101a ENM_9000086	http://purl.enanomapper.org/onto/ENM_9000086	Q47462008	wikidata:Q47462008	
JRCNM02000a ENM_9000087	http://purl.enanomapper.org/onto/ENM_9000087	Q47462022	wikidata:Q47462022	
JRCNM02001a ENM_9000088	http://purl.enanomapper.org/onto/ENM_9000088	Q47468470	wikidata:Q47468470	
JRCNM02002a ENM_9000089	http://purl.enanomapper.org/onto/ENM_9000089	Q47468473	wikidata:Q47468473	
JRCNM02003a ENM_9000090	http://purl.enanomapper.org/onto/ENM_9000090	Q78642247	wikidata:Q78642247	
JRCNM02004a ENM_9000091	http://purl.enanomapper.org/onto/ENM_9000091	Q47468478	wikidata:Q47468478	

```
<script type="application/ld+json">
{
  "@context": "https://schema.org/",
  "@type": "ChemicalSubstance",
  "name": "JRCNM01000a",
  "@id":"http://purl.enanomapper.org/onto/ENM_9000074",
  "identifier": "ENM_9000074",
  "url": [ "https://scholia.toolforge.org/Q27918612" ]
}
</script>
```

ChemicalSubstance: does Google index them?



Proteins

Sites: 10 | Protein (0.11-RELEASE)

ChEMBL (Mirror)	v0.12-DRAFT	Over 8.5 thousand proteins		
DisProt	v0.11	~2k proteins	Page	SMV BMUSE
HmmerWeb	v0.1	Result of searches are annotated (results last 1 week)	Page	SMV BMUSE
IUPHAR/BPS Guide to Pharmacology	v0.11	Proteins associated with COVID-19	Page	SMV BMUSE
MobiDB	v0.11	~2k proteins	Page	SMV BMUSE
PDBe	v0.1	Over 215,000 proteins	Page	SMV BMUSE
Protein Ensemble (PED)	v0.11		Page	SMV BMUSE
PSnpBind	v0.11	Bioschemas annotation of 26 protein structures. For these 0.6 million docking results are found in the database	Page	SMV BMUSE
Scholia	v0.11	More than 100,000 proteins.	Page	SMV BMUSE
STRING	v0.9-DRAFT	25 million+ protein networks	Page	SMV BMUSE

What is next?

Getting indexed

- MolecularEntity
- ChemicalSubstance

But indexed on what?

More providers

- Databases
  - ELIXIR Toxicology Community
- Scientific journals
  - Or is this just a unicorn?
- Integration with existing tools
  - E.g. Markdown/Jekyll

PubChem

WIKIDATA

Thank you!

Leyla Jael Castro,
ZB MED Information Centre for Life Sciences
Bioschemas

Egon Willighagen
Maastricht University
Bioschemas

ACS FALL 2023
13-17 August, San Francisco CA & Hybrid

Bioschemas

https://bioschemas.org/