



The DO-KB is expanding the disease information ecosystem

- expanding the representation of the DO diseaseome
- a comprehensive network of disease to disease relationships
- a disease feature similarity network for differential diagnosis exploration

<https://disease-ontology.org/do-kb/>

[DO-KB SPARQL Sandbox](#)

- A data playground for querying the Human Disease Ontology SPARQL endpoint and to perform federated SPARQL queries across connected resources for disease knowledge discovery.
- A linked data representation (RDF triple store) of the Human Disease Ontology's diseases, disease features and mechanisms.

[DO-KB Faceted Search Interface](#)

- A new way to explore human disease features and mechanisms, represented here as data facets.
- The DO-KB Faceted Search Interface enables exploration of connectivity across diseases.

The DO-KB's SPARQL endpoint, explore DO data and query disease, genomic, proteomic resources through federated queries.

Mine the DO-KB disease-data connections through faceted query and retrieval of DO diseases sharing features or mechanisms, including phenotypes, environmental or genetic drivers, anatomy, variant type, and age of onset.

<https://disease-ontology.org/do-kb/sparql>



DO-KB SPARQL Sandbox

Select one of the provided SPARQL queries or paste in a novel query to retrieve bespoke ML-ready datasets. Searches may be performed against the DO's primary release file, `doid.owl` or the `doid-merged.owl` file for exploring annotations, e.g. ECO codes, and axiom relationships.

- The DO's SPARQL endpoint (<https://sparql.disease-ontology.org>) is available for building federated queries.
- Visit our [SPARQL Resources](#) page to learn more about using the SPARQL Sandbox and endpoint.
- Our SPARQL Sandbox and endpoint are currently in **beta**. If you discover any issues, please [reach out](#) and let us know!

Query

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX obo: <http://purl.obolibrary.org/obo/>
PREFIX obolnOwl: <http://www.geneontology.org/formats/obolnOwl#>
```

```
SELECT ?class ?id ?label
WHERE {
  ?class owl:Class ;
    obolnOwl:id ?id ;
    rdfs:label ?label ;
    rdfs:subClassOf* obo:D0ID.225 .
}
```

Submit Query

Download Results

Choose a Query:

- All diseases with their parents (identified by label)
- Count of diseases in each branch (diseases may be in multiple branches)
- Syndromes (branch search; returns IRI, ID and label)
- Diseases with MeSH cross-references
- Diseases with OMIM cross-references
- Diseases with PubMed sources (limited to first 10 results)
- Diseases in `DO_infectious_disease_slim` with their parents and branches
- Diseases with their definitions and count of logical axioms
- Diseases with their Evidence and Conclusion Ontology (ECO) reference types (queries the `doid-merged.owl` graph)
- Count of classes imported from other ontologies (queries the `doid-merged.owl` graph)
- All diseases with equivalent class axioms (axioms are formatted for readability and may not be displayed in full; queries the `doid-merged.owl` graph)
- Extract the disease and supporting import class hierarchies using `subClassOf` relationships (executes a CONSTRUCT query that returns RDF triples)

Request New SPARQL Queries

<https://disease-ontology.org/do-kb/facet>

DO-KB Faceted Search Interface

Identify diseases with similar characteristics and explore relationships between diseases through common characteristics or features.

Faceted Search

anatomy(747) filter:

<input type="checkbox"/> anatomy	8760
<input type="checkbox"/> material anatomical entity	8760
<input type="checkbox"/> anatomical structure	8375
<input type="checkbox"/> anatomical system	7686
<input type="checkbox"/> multicellular anatomical	

symptom(360) filter:

<input type="checkbox"/> symptom	4037
<input type="checkbox"/> general symptom	2293
<input type="checkbox"/> inflammation	1808
<input type="checkbox"/> abdominal symptom	833

ncbitaxon(978) filter:

<input type="checkbox"/> ncbitaxon	486
<input type="checkbox"/> cellular organisms	378
<input type="checkbox"/> Eukaryota	220
<input type="checkbox"/> Bacteria	197

Search Results

Show 50 entries

DOID	Name
D0ID_0001816	angiosarcoma
D0ID_0002116	ptyerygium
D0ID_0014667	disease of metabolism
D0ID_0040001	shrimp allergy
D0ID_0040002	aspirin allergy
D0ID_0040003	benzylpenicillin allergy
D0ID_0040004	amoxicillin allergy
D0ID_0040005	ceftriaxone allergy
D0ID_0040006	carbamazepine allergy
D0ID_0040007	abacavir allergy
D0ID_0040008	isoniazide allergy
D0ID_0040009	lidocaine allergy
D0ID_0040010	mepivacaine allergy
D0ID_0040011	phenobarbital allergy
D0ID_0040012	phenytoin allergy
D0ID_0040013	ranitidine allergy
D0ID_0040014	corticosteroid allergy
D0ID_0040015	sulfonamide allergy
D0ID_0040016	sulfamethoxazole allergy
D0ID_0040017	suprofen allergy
D0ID_0040018	thiopental allergy
D0ID_0040019	D-mannitol allergy
D0ID_0040020	cefotaxime allergy
D0ID_0040021	cephalosporin allergy
D0ID_0040022	amodiaquine allergy
D0ID_0040023	cefactor allergy
D0ID_0040024	ceftazidime allergy

What is SPARQL?



- SPARQL: **S**PARQL **P**rotocol **A**nd **R**df **Q**uery **L**anguage
- An RDF (Resource Description Framework) Query Language
- A semantic query language for databases

What can you do with SPARQL?

- Enables knowledge to be represented in a machine-readable way
- Used to retrieve and manipulate data stored in RDF format.

- Codifies (stores) relationships between semantic data, in the form of subject-predicate-object 'triples'. E.g., Baltimore is in Maryland
-
- Use SPARQL to create complex queries with SELECT statements

- A SPARQL endpoint which allows you to query against the data set

service	endpoint
PRO	< https://sparql.proconsortium.org/virtuoso/sparql >
UniProt	< http://sparql.uniprot.org/sparql >
DisGeNet	< http://rdf.disgenet.org/sparql/ >
MeSH	< http://id.nlm.nih.gov/mesh/sparql >
Orphanet	< https://www.orpha.net/sparql >


DO-KB SPARQL Sandbox

Select one of the provided SPARQL queries or paste in a novel query to retrieve bespoke ML-ready datasets. Searches may be performed against the DO's primary release file, `doid.owl` or the `doid-merged.owl` file for exploring annotations, e.g. ECO codes, and axiom relationships.

- The DO's SPARQL endpoint (<https://sparql.disease-ontology.org>) is available for building federated queries.
- Visit our [SPARQL Resources](#) page to learn more about using the SPARQL Sandbox and endpoint.
- Our SPARQL Sandbox and endpoint are currently in **beta**. If you discover any issues, please [reach out](#) and let us know!

Query

Submit Query

Download Results 

The DO-KB's SPARQL Sandbox provides programmatic access to the DO-KB Knowledgebase for exploring DO diseases, disease features, cross references and mechanisms, and exploring disease-data connections to other SPARQL endpoints.

Choose a Query:

1. All diseases with their parents (identified by label)
2. Count of diseases in each branch (diseases may be in multiple branches)
3. Syndromes (branch search; returns IRI, ID and label)
4. Diseases with MeSH cross-references
5. Diseases with OMIM cross-references
6. Diseases with PubMed sources (limited to first 10 results)
7. Diseases in `DO_infectious_disease_slim` with their parents and branches
8. Diseases with their definitions and count of logical axioms
9. Diseases with their Evidence and Conclusion Ontology (ECO) reference types (queries the `doid-merged.owl` graph)
10. Count of classes imported from other ontologies (queries the `doid-merged.owl` graph)
11. All diseases with equivalent class axioms (axioms are formatted for readability and may not be displayed in full; queries the `doid-merged.owl` graph)
12. Extract the disease and supporting import class hierarchies using `subClassOf` relationships (executes a CONSTRUCT query that returns RDF triples)

Request New SPARQL Queries

- Select a SPARQL query to retrieve ML-ready datasets
- SELECT (table formatted results) or CONSTRUCT (RDF results) queries
- Edit queries to refine your search
- Suggest new SPARQL queries to DO-KB
- Utilize the DO-KB SPARQL Endpoint to build your own federated queries

- Coming Soon:
 - Federated queries to UniProt, MeSH, Protein Ontology

Query execution time: 2.22 second(s).

id	label	
DOID:0050120	hemophagocytic lymphohistiocytosis	OMIM:PS267700
DOID:0050156	idiopathic pulmonary fibrosis	OMIM:178500
DOID:0050158	desquamative interstitial pneumonia	OMIM:263000
DOID:0050167	autoimmune polyendocrine syndrome type 1	OMIM:240300

Structuring a SPARQL Query

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>  
PREFIX owl: <http://www.w3.org/2002/07/owl#>  
PREFIX obolInOwl: <http://www.geneontology.org/formats/obolInOwl#>
```

```
SELECT ?id ?label ?parent  
WHERE {  
  ?class a owl:Class .  
  ?class obolInOwl:id ?id .  
  ?class rdfs:label ?label .  
  ?class rdfs:subClassOf ?parent_class .  
  
  ?parent_class rdfs:label ?parent .  
}
```

Building a SPARQL Query

} Abbreviations used as shortcuts in the query

◆ SELECT : output variables

} Conditions that define the query

OWL file, data elements

[ID]	obolInOwl:id ?id
[label]	rdfs:label ?label
[Synonym]	obolInOwl:hasExactSynonym
[Namespace]	obolInOwl:hasOBONamespace
[Slim/Subset]	obolInOwl:inSubset
[Alt ID (merges)]	obolInOwl:hasAlternativeId