

# RDF形式で利用できる 生命科学RDFデータの紹介

情報システム研究機構  
データサイエンス共同利用基盤施設  
ライフサイエンス統合データベースセンター

川島秀一

2018/11/1

# UniProt RDF

- ・ UniProtは、2006年からRDF版を開発しており、その点でも生命科学データベースとしては先駆的な存在。データベースをRDF化する場合に非常に参考になります。
- ・ UniProt のRDF版は、ftpサイトからダウンロードできます。

[ftp://ftp.uniprot.org/pub/databases/uniprot/current\\_release/rdf/](ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/rdf/)



<http://www.uniprot.org/uniprot/O88559>

# RDFの各種フォーマット間の変換

- ・ RDFには、いくつかのフォーマットがあります（RDF/XML、N-Triples、Turtle、TriG、N3、JSON-LD）。
- ・ 前のスライドで紹介した、UniProtの各エントリーページからダウンロードできるRDFファイルはRDF/XMLフォーマットで、人間には読み難いので、Turtleに変換することで比較的読みやすくなります。
- ・ 変換は、Raptor（RDFをパース／シリアライズするためのライブラリ。<http://librdf.org/raptor/>）に含まれている rapper ユーティリティで行うことができます。

```
> rapper -i rdfoxml -o turtle 088559.rdf > 088559.ttl
```

# UniProt RDFの構造

UniProt

UniProtKB - Q88559 (MEN1\_MOUSE)

Display

Entry

Publications

Feature viewer

Feature table

Function

Protein: Men1

Gene: Men1

Organism: Mus musculus (Mouse)

Status: Reviewed - Annotation score: 100% - Experimental evidence at protein level

Function

Essential component of a RPL/SET1 histone methyltransferase (HMT) complex, a complex that specifically methylates 'lys-4' of histone H3 (H3K4). Functions as a transcriptional regulator. Binds to the TERT promoter and represses telomerase expression. Plays a role in TGF $\beta$ -mediated inhibition of cell-proliferation, positively regulating SMAD3 transcriptional activity. Represses RING-mediated transcriptional activation on AP1 sites, as well as that mediated by NFkB subunit RELA. Positively regulates HOK2 and HOK3 gene expression (by similarity). May be involved in normal hematopoiesis through the activation of HES6 expression. May be involved in DNA repair. [By similarity]

GO - Molecular function

- chromatin binding [Source: MGI]
- DNA binding [Source: MGI]
- double-stranded DNA binding [Source: MGI]
- four-way junction DNA binding [Source: MGI]
- protein binding, bridging [Source: MGI]
- protein-RNA binding [Source: MGI]
- R-SMAD binding [Source: MGI]
- sequence-specific DNA binding [Source: MGI]
- transcription regulatory region DNA binding [Source: MGI]
- T-form DNA binding [Source: MGI]

GO - Biological process

- cell cycle arrest [Source: MGI]
- cellular response to DNA damage stimulus [Source: MGI]
- chromatin remodeling [Source: MGI]
- embryonic skeletal system morphogenesis [Source: MGI]
- hemopoiesis [Source: MGI]
- histone lysine methylation [Source: MGI]
- histone methylation [Source: MGI]
- leukocyte homeostasis [Source: MGI]
- MAPK cascade [Source: MGI]
- maternal process involved in female pregnancy [Source: MGI]
- negative regulation of cell cycle [Source: MGI]
- negative regulation of cell proliferation [Source: MGI]
- negative regulation of cyclin-dependent protein serine/threonine kinase activity [Source: MGI]
- negative regulation of JAK cascade [Source: MGI]
- negative regulation of organ growth [Source: MGI]
- negative regulation of osteoblast differentiation [Source: MGI]
- negative regulation of protein phosphorylation [Source: MGI]
- negative regulation of sequence-specific DNA binding transcription factor activity [Source: MGI]
- negative regulation of telomerase activity [Source: MGI]
- negative regulation of transcription, DNA-templated [Source: MGI]
- negative regulation of transcription from RNA polymerase II promoter [Source: MGI]
- osteification [Source: MGI]
- osteoblast development [Source: MGI]
- osteoblast fate commitment [Source: MGI]
- osteoblast development [Source: MGI]
- positive regulation of apoptotic process [Source: MGI]
- positive regulation of cell differentiation [Source: MGI]
- positive regulation of cell division [Source: MGI]
- positive regulation of cysteine-type endopeptidase activity involved in apoptotic process [Source: MGI]
- positive regulation of gene expression [Source: MGI]
- positive regulation of histone methylation [Source: MGI]
- positive regulation of osteoblast differentiation [Source: MGI]
- positive regulation of protein binding [Source: MGI]
- positive regulation of transcription, DNA-templated [Source: MGI]
- positive regulation of transcription from RNA polymerase II promoter [Source: MGI]
- positive regulation of transforming growth factor beta receptor signaling pathway [Source: MGI]
- regulation of gene expression [Source: MGI]
- response to gamma radiation [Source: MGI]
- response to UV [Source: MGI]
- transcription, DNA-templated [Source: UniProtKB]

Complete GO annotation...

Keywords

Molecular function: Chromatin regulator, DNA-binding, Repressor

Biological process: Transcription, Transcription regulation

Enzyme and pathway databases

Reaction

- 6-HHO-211122: Formation of the beta-catenin/TCF transactivating complex.
- 6-HHO-217756: SMAD2/SMAD3-SMAD4 heterotrimer regulates transcription.
- 6-HHO-219402: Deactivation of the beta-catenin transactivating complex.
- 6-HHO-281436: Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs).
- 6-HHO-342467: RHO GTPases activate IQGAPs.
- 6-HHO-4057173: Post-translational protein phosphorylation.

Names & Taxonomy

Protein name: Men1

Gene name: Men1

Organism: Mus musculus (Mouse)

Taxonomic identifier: 10090 [NCBI]

Taxonomic lineage: Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Euarchontoglires > Glires > Rodentia > Myomorpha > Muridae > Murinae > Mus > Mus

@prefix : <http://purl.uniprot.org/core/> .

<http://purl.uniprot.org/uniprot/Q88559> a :Protein .

<http://purl.uniprot.org/uniprot/Q88559> のようなタンパク質エントリーのURIが、:Protein と定義されていて、基本的にはその他の情報はそこから辿れる

# UniProt RDFの構造

UniProtKB - O88559 (MEN1\_MOUSE)

Display

Entry

Publications

Feature viewer

Feature table

BLAST

Align

Format

Add to basket

History

Feedback

Help video

Other tutorials and videos

View this entry in a different format

Protein | **Menin**

Gene | **Men1**

Organism | *Mus musculus (Mouse)*

Status | Reviewed - Annotation score: - Experimental evidence at protein level<sup>1</sup>

All None

@prefix : <http://purl.uniprot.org/core/> .

```
<http://purl.uniprot.org/uniprot/088559> a :Protein .  
    :encodedBy <http://purl.uniprot.org/uniprot/088559#gene-MD5948FF492D4C49D3B58EB3BDC162CFE26>  
    :recommendedName <http://purl.uniprot.org/uniprot/088559#SIPC75DF59E860493C8> ;
```

```
<http://purl.uniprot.org/uniprot/088559#SIPC75DF59E860493C8>  
    :fullName "Menin" ;  
    a :Structured_Name .
```

```
<http://purl.uniprot.org/uniprot/088559#gene-MD5948FF492D4C49D3B58EB3BDC162CFE26>  
    a :Gene ;  
    skos:prefLabel "Men1" .
```

# UniProt RDFの構造

UniProtKB - O88559 (MEN1\_MOUSE)

Display

Entry

Publications

Feature viewer

Feature table

BLAST

Align

Format

Add to basket

History

Feedback

Help video

Other tutorials and videos

View this entry in a different format

Protein | **Menin**

Gene | **Men1**

Organism | *Mus musculus (Mouse)*

Status | Reviewed - Annotation score: ●●●●●● Experimental evidence at protein level

All None

```
<http://purl.uniprot.org/uniprot/O88559> a :Protein .  
      :organism <http://purl.uniprot.org/taxonomy/10090> ;
```

O88559.ttl には、Taxonomy ID (10090) までしか記述されていません

```
<http://purl.uniprot.org/taxonomy/10090>  
  a :Taxon ;  
  :commonName "Mouse" ;  
  :mnemonic "MOUSE" ;  
  :rank :Species ;  
  :scientificName "Mus musculus" ;  
  rdfs:subClassOf <http://purl.uniprot.org/taxonomy/862507> ;  
  skos:narrowerTransitive <http://purl.uniprot.org/taxonomy/10091> ,  
                          <http://purl.uniprot.org/taxonomy/10092> ,  
                          <http://purl.uniprot.org/taxonomy/116058> ,  
                          :
```

taxonomy:10090  
の詳細情報は、  
taxonomy.ttl に記述  
されている

# UniProt RDFの構造

## Display

Entry

Publications

Feature viewer

Feature table

All None

✓ Function

✓ Names & Taxonomy

✓ Subcell. location

✓ Pathol./Biotech

✓ PTM / Processing

✓ Expression

✓ Interaction

## Function<sup>i</sup>

Essential component of a MLL/SET1 histone methyltransferase (HMT) complex, a complex that specifically methylates 'Lys-4' of histone H3 (H3K4). Functions as a transcriptional regulator. Binds to the TERT promoter and represses telomerase expression. Plays a role in TGFB1-mediated inhibition of cell-proliferation, possibly regulating SMAD3 transcriptional activity. Represses JUND-mediated transcriptional activation on AP1 sites, as well as that mediated by NFkB subunit RELA. Positively regulates HOXC8 and HOXC6 gene expression (By similarity). May be involved in normal hematopoiesis through the activation of HOXA9 expression. May be involved in DNA repair. ⓘ By similarity ⓘ 1 Publication ▼

### GO - Molecular function<sup>i</sup>

- chromatin binding ⓘ Source: MGI ▼
- DNA binding ⓘ Source: MGI ▼
- double-stranded DNA binding ⓘ Source: MGI
- four-way junction DNA binding ⓘ Source: MGI
- protein binding, bridging ⓘ Source: MGI
- protein N-terminus binding ⓘ Source: MGI
- R-SMAD binding ⓘ Source: MGI
- sequence-specific DNA binding ⓘ Source: MGI ▼
- transcription regulatory region DNA binding ⓘ Source: MGI
- Y-form DNA binding ⓘ Source: MGI

GO annotation の chromatin binding は、どのように記述されているか

```
<http://purl.uniprot.org/uniprot/088559>
  a :Protein .
  :classifiedWith <http://purl.obolibrary.org/obo/GO_0000122>,
                  :
                  <http://purl.obolibrary.org/obo/GO_0003682>,
                  :
obo:GO_0003682
  a owl:Class ;
  rdfs:label "chromatin binding" ;
  rdfs:subClassOf obo:GO_0005488 .
```

go:GO\_0003682の  
詳細情報は、go.ttl  
に記述されている

# UniProt RDFの構造

☒ Function

☒ Names & Taxonomy

☒ Subcell. location

**GO - Molecular function<sup>i</sup>**

- chromatin binding Source: MGI
- DNA binding Source: MGI
- double-stranded DNA binding Source: MGI

GO annotation の Source:MGIは、どのように記述されているか

```
<http://purl.uniprot.org/uniprot/088559>
  a :Protein .
  :classifiedWith <http://purl.obolibrary.org/obo/GO_0000122>,
                  :
                  <http://purl.obolibrary.org/obo/GO_0003682>,
                  :
<http://purl.uniprot.org/uniprot/#_kb.088559_up.classifiedWith_obo.GO_0003682>
  :attribution <http://purl.uniprot.org/uniprot/088559#attribution-0020AFA47BB288231A49330BFA13AADF>,
               <http://purl.uniprot.org/uniprot/088559#attribution-05EC9312E990B877EC8BC3ADF07EA7D6>,
               <http://purl.uniprot.org/uniprot/088559#attribution-AF533202130C9732B5E56B6FB50FCC79>;
  rdf:object <http://purl.obolibrary.org/obo/GO_0003682> ;
  rdf:predicate :classifiedWith ;
  rdf:subject <http://purl.uniprot.org/uniprot/088559> ;
  a rdf:Statement .

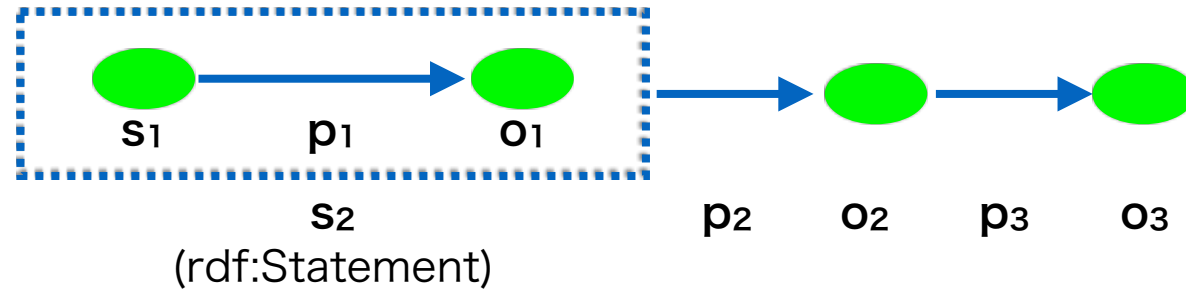
<http://purl.uniprot.org/uniprot/088559#attribution-0020AFA47BB288231A49330BFA13AADF>
  dcterms:creator <http://purl.uniprot.org/goa-projects/MGI> ;
  :evidence <http://purl.obolibrary.org/obo/ECO_0000314> ;
  :manual true ;
  :source <http://purl.uniprot.org/citations/16415155> .
```

ECO\_0000314は、direct assay evidence used in manual assertion



# RDF Reification

- ・ RDF のあるトリプルに対して言及したい場合に、RDF reification (具体化) という手法が使えます。



```
s2
  p2 o2;
  rdf:object    o1 ;
  rdf:predicate p1 ;
  rdf:subject   s1 ;
  a rdf:Statement .

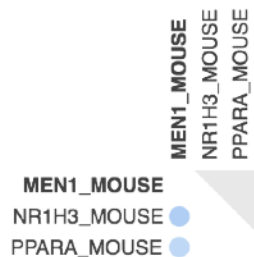
o2 p3 o3 .
```

# UniProt RDFの構造

## Binary interactions<sup>i</sup>

O88559 has binary interactions with 2 proteins

Subcellular location ▼



[Show more details](#)

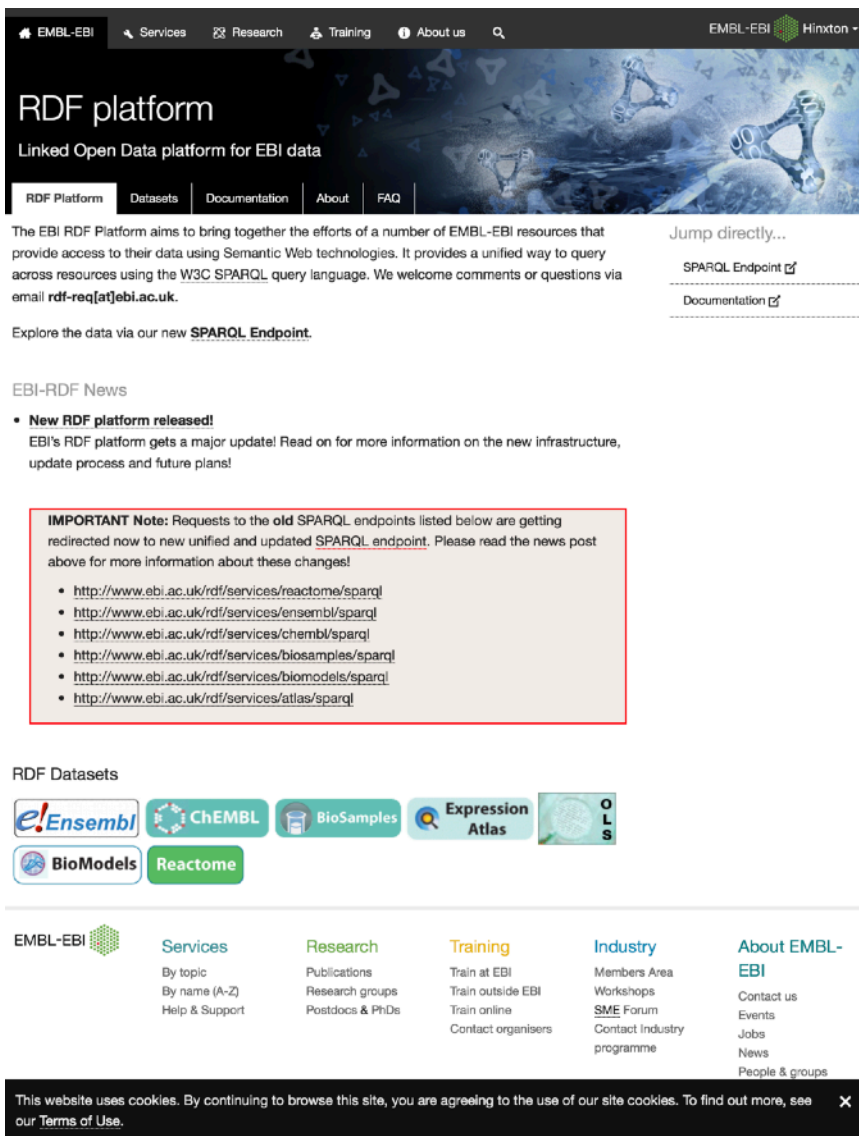
タンパク間相互作用の記述はどうなっているか

```
<http://purl.uniprot.org/uniprot/O88559>
  a :Protein .
  :interaction <http://purl.uniprot.org/uniprot/O88559#interaction-3990176-5273083>,
               <http://purl.uniprot.org/uniprot/O88559#interaction-3990176-5276764> ;

<http://purl.uniprot.org/uniprot/O88559#interaction-3990176-5273083>
  :experiments "2"^^<http://www.w3.org/2001/XMLSchema#int> ;
  :participant <http://purl.uniprot.org/intact/EBI-3990176>,
               <http://purl.uniprot.org/intact/EBI-5273083> ;
  :xeno false ;
  a :Interaction, :Non_Self_Interaction .

<http://purl.uniprot.org/intact/EBI-5273083>
  a :Participant ;
  rdfs:label "Ppara" ;
  owl:sameAs <http://purl.uniprot.org/uniprot/P23204> .
```

# EBI RDF platform



The screenshot shows the EBI RDF platform website. The header includes navigation links for Services, Research, Training, and About us. The main heading is "RDF platform" with the subtitle "Linked Open Data platform for EBI data". Below this is a navigation bar with links to RDF Platform, Datasets, Documentation, About, and FAQ. The main content area describes the platform's purpose and provides a "Jump directly..." section with links to the SPARQL Endpoint and Documentation. A news section titled "EBI-RDF News" features a "New RDF platform released!" announcement. A red-bordered box contains an "IMPORTANT Note" about redirected SPARQL endpoints, listing several URLs. Below this is a "RDF Datasets" section with logos for Ensembl, ChEMBL, BioSamples, Expression Atlas, BioModels, and Reactome. The footer includes a grid of links for Services, Research, Training, Industry, and About EMBL-EBI, along with a cookie consent banner at the bottom.

EMBL-EBI Hinxton

## RDF platform

Linked Open Data platform for EBI data

RDF Platform | Datasets | Documentation | About | FAQ

The EBI RDF Platform aims to bring together the efforts of a number of EMBL-EBI resources that provide access to their data using Semantic Web technologies. It provides a unified way to query across resources using the W3C SPARQL query language. We welcome comments or questions via email [rdf-req\[at\]ebi.ac.uk](mailto:rdf-req[at]ebi.ac.uk).

Explore the data via our new [SPARQL Endpoint](#).

Jump directly...

[SPARQL Endpoint](#)

[Documentation](#)

### EBI-RDF News

- New RDF platform released!**  
EBI's RDF platform gets a major update! Read on for more information on the new infrastructure, update process and future plans!

**IMPORTANT Note:** Requests to the old SPARQL endpoints listed below are getting redirected now to new unified and updated SPARQL endpoint. Please read the news post above for more information about these changes!

- <http://www.ebi.ac.uk/rdf/services/reactome/sparql>
- <http://www.ebi.ac.uk/rdf/services/ensembl/sparql>
- <http://www.ebi.ac.uk/rdf/services/chembl/sparql>
- <http://www.ebi.ac.uk/rdf/services/biosamples/sparql>
- <http://www.ebi.ac.uk/rdf/services/biomodels/sparql>
- <http://www.ebi.ac.uk/rdf/services/atlas/sparql>

### RDF Datasets

Ensembl | ChEMBL | BioSamples | Expression Atlas | OLS | BioModels | Reactome

EMBL-EBI

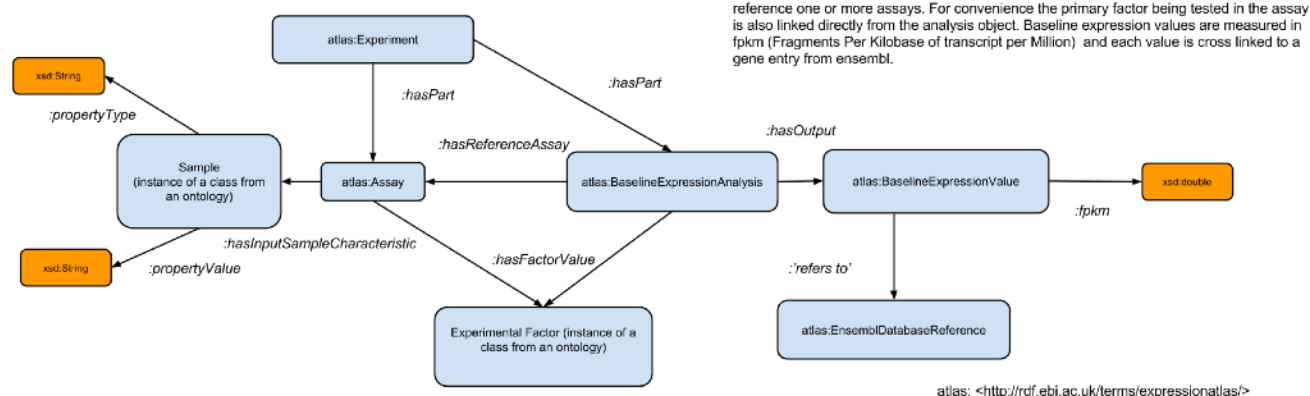
- Services**
  - By topic
  - By name (A-Z)
  - Help & Support
- Research**
  - Publications
  - Research groups
  - Postdocs & PhDs
- Training**
  - Train at EBI
  - Train outside EBI
  - Train online
  - Contact organisers
- Industry**
  - Members Area
  - Workshops
  - SME Forum
  - Contact Industry programme
- About EMBL-EBI**
  - Contact us
  - Events
  - Jobs
  - News
  - People & groups

This website uses cookies. By continuing to browse this site, you are agreeing to the use of our site cookies. To find out more, see our [Terms of Use](#).

- EBIが公開しているデータベースの中で、RDF化されたものを一覧できるサービス。
- 現状では、Ensembl, ChEMBL, BioSample, ExpressionAtlas, BioModels, Reactome が公開されている。

# EBI RDF platform

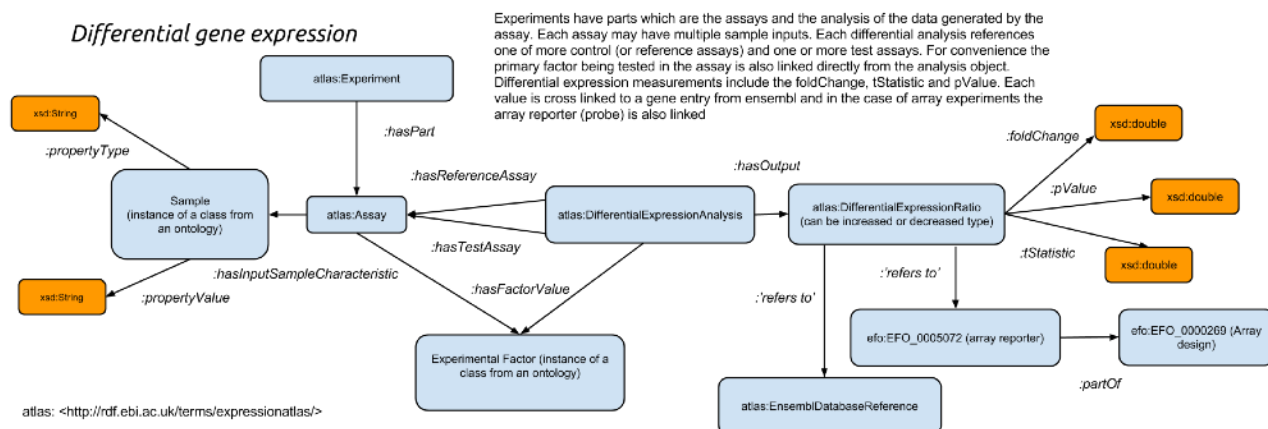
## Baseline expression schema



atlas: <<http://rdf.ebi.ac.uk/terms/expressionatlas/>>

- EBI RDF platform では、RDFスキーマ図が提供されているので、SPARQL問い合わせを記述する際に役に立ちます。

## Differential gene expression



atlas: <<http://rdf.ebi.ac.uk/terms/expressionatlas/>>

ExpressionAtlasのスキーマ図

# PubChem RDF

The screenshot shows a web browser window with the address bar displaying "pubchem.ncbi.nlm.nih.gov/rdf". The main heading is "PubChemRDF Release Notes". Below it, the version "V1.6.1 beta" is noted, along with a link to the release notes and a total triple count of 103,280,532,587 as of 09-25-2017. A blue box highlights the triple count. The left sidebar contains a navigation menu with links to Introduction, Getting Started, Querying, Downloading, and About. The right sidebar lists a detailed table of contents.

# PubChemRDF Release Notes

V1.6.1 beta (See the [V1.6 beta Release Notes](#))

Total number of triples: **103,280,532,587** (last update on: 09-25-2017) for more details, please see [Table 2](#)

## 1. Introduction

Semantic Web technologies are emerging as an increasingly important approach to distribute and integrate scientific data. These technologies include the trio of the [Resource Description Framework \(RDF\)](#), [Web Ontology Language \(OWL\)](#), and [SPARQL query language](#). The [PubChemRDF project](#) provides RDF formatted information for the PubChem [Compound](#), [Substance](#), and [Bioassay](#) databases.

### 1.1. What is RDF?

RDF constitutes a family of World Wide Web Consortium (W3C) specifications for data interchange on the Web. RDF breaks down knowledge into machine-readable discrete pieces, called “triples.” Each “triple” is organized as a trio of ‘subject-predicate-object’. For example, in the phrase “atorvastatin *may treat* hypercholesterolemia,” the subject is “atorvastatin”, the predicate is “*may treat*”, and the object is “cholesterol.” RDF uses a Uniform Resource Identifier ([URI](#)) to name each part of the “subject-predicate-object” triple. A URI looks just like a typical web [URL](#). RDF is a core part of [semantic web standards](#). As an extension of the existing [World Wide Web](#), the semantic web attempts to make it easier for users to find, share, and combine information. Semantic web leverages the following technologies: extensible markup language ([XML](#)), which provides syntax for RDF; web ontology language ([OWL](#)), which extends the ability of RDF to encode information; resource description framework (RDF), which expresses knowledge; and RDF query language ([SPARQL](#)), which enables query and manipulation of RDF content.

### 1. Introduction

- 1.1. What is RDF?
  - 1.2. How can PubChemRDF help your research?
2. Ontology-based Data Integration
3. PubChemRDF URI Constructions
4. PubChemRDF Subdomains
  - 4.1. PubChem Compound
  - 4.2. PubChem Substance
  - 4.3. PubChem Descriptors
  - 4.4. PubChem InChIKey
  - 4.5. PubChem Synonym
  - 4.6. PubChem BioAssay
  - 4.7. PubChem MeasureGroup
  - 4.8. PubChem Endpoint
  - 4.9. PubChem Protein
  - 4.10. PubChem ConservedDomain
  - 4.11. PubChem Gene
  - 4.12. PubChem Biosystem
  - 4.13. PubChem Neighbor
  - 4.14. PubChem Source
  - 4.15. PubChem Reference
  - 4.16. PubChem Concept
5. RESTful INTERFACE
  - 5.1. URI Dereferencing
  - 5.2. Query RESTful Interface
  - 5.3. HTTP Response Status
6. RDF FTP Download Directory Layout
  - 6.1. PubChem Compound

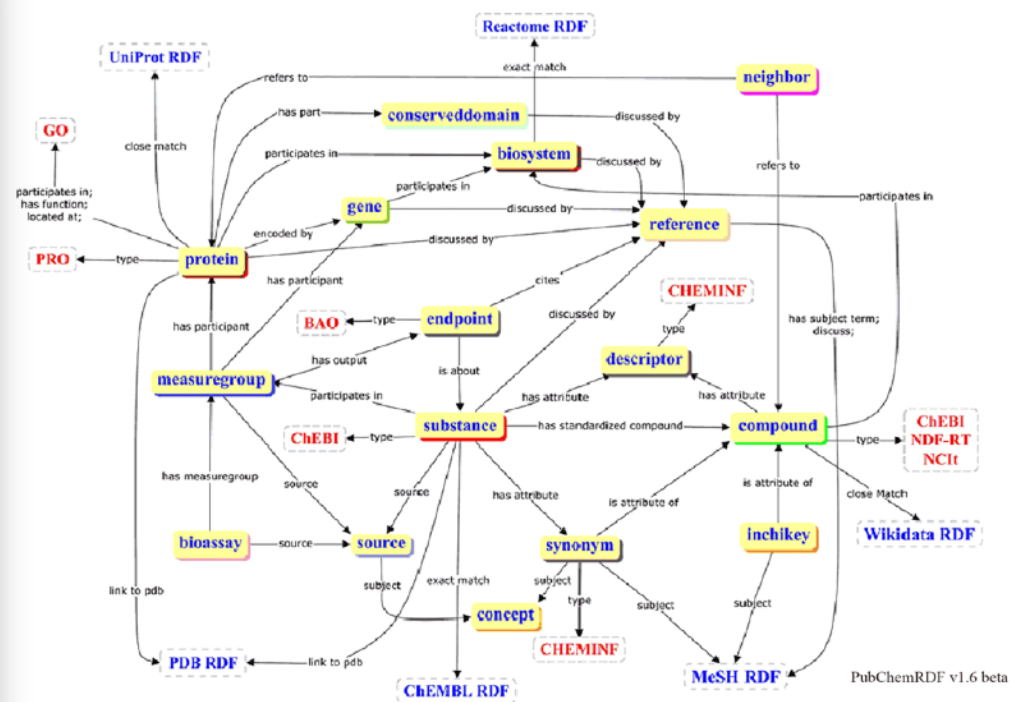


Figure 1. Color-coded diagram showing a high-level overview of the PubChemRDF semantic relationships.

<https://pubchem.ncbi.nlm.nih.gov/rdf/>