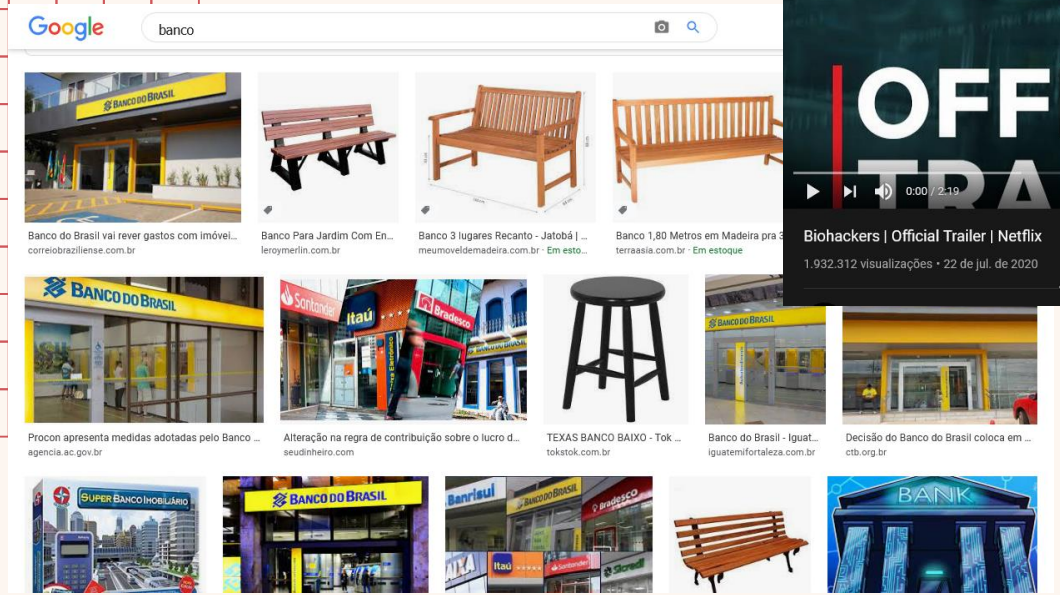
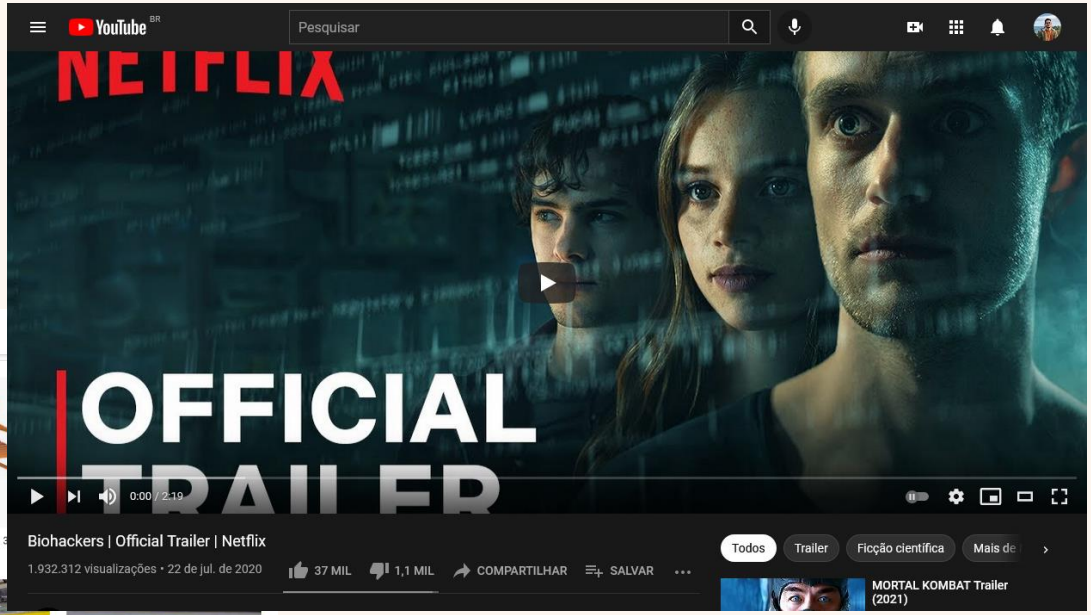
Web Semântica e Linked Data

Web na ótica humana

- Feita para seres humanos.
- Foco no reconhecimento visual.



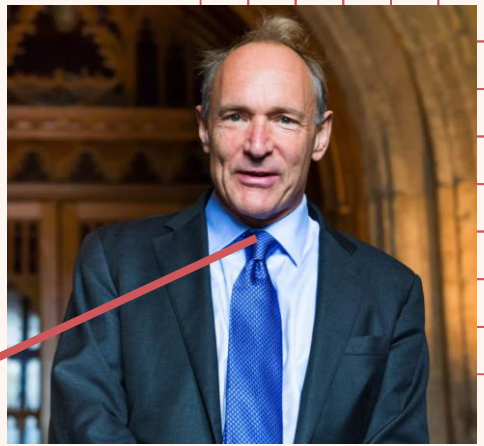


Web semântica

- Representar a web na ótica humana semanticamente para ferramentas inteligentes.
- Cada recurso é dado por um significado bem definido.

Como descrever que um determinado nó em uma rede é uma pessoa?

Com WEB SEMÂNTICA é possível!

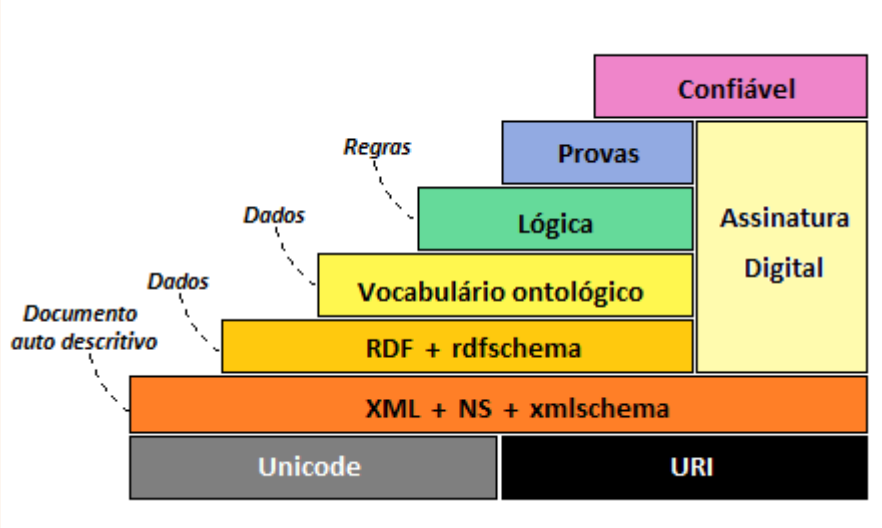


Tim Berners-Lee, criador do World Wide Web e Web Semântica

Web semântica

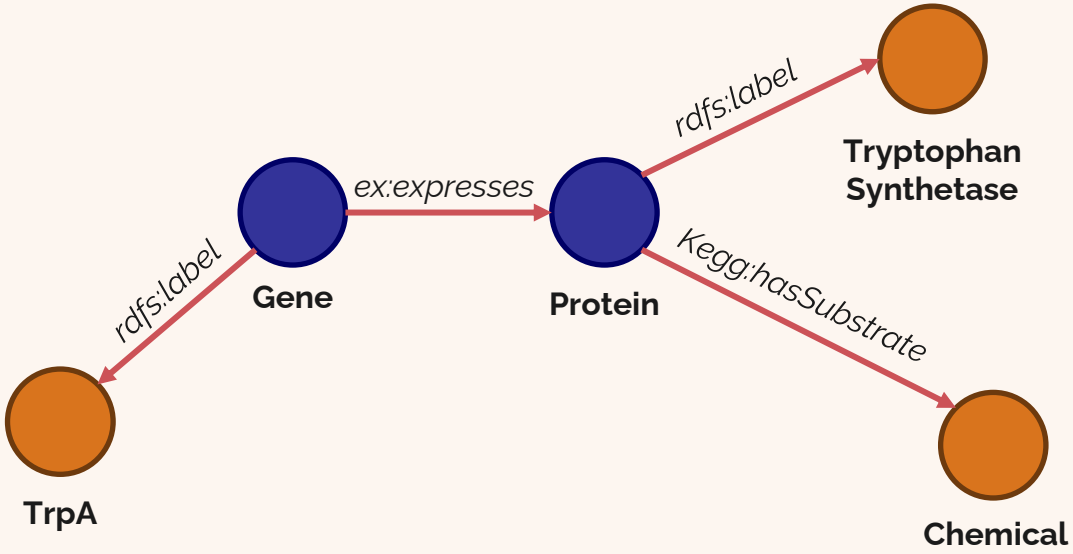
"... the idea of having data on the web defined and linked in a way that it ***can be used by machines not just for display purposes, but for automation, integration and reuse*** *of data across various applications."*

W3C Semantic Web Activity Group



James Miranda, Tableless

Web semântica

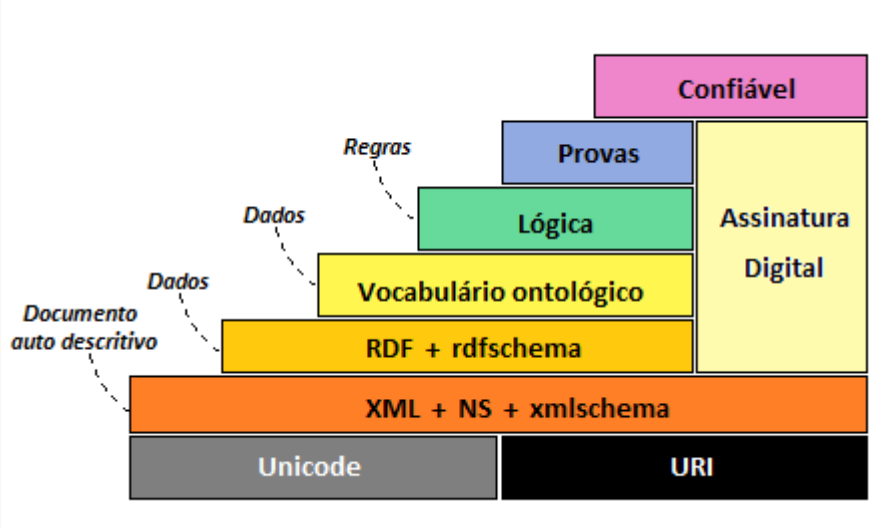


Web semântica - Benefícios

- Interoperabilidade
- Padronização e Organização
- Reuso de Informação
- Inferência
- Serendipidade

Web semântica -Tecnologias

- Linguagens de marcação estruturadas (XML, JSON)
- Linguagem RDF
- Ontologias
- Sparql
- Linked Data



Resource Description Framework

- RDF é um modelo padronizado para apresentação de um grafo dirigido.
- Apresenta de forma homogênea descrições associadas a recursos endereçáveis através da web.
- Foco no consumo por computadores.
- Atua sobre problemas de interoperabilidade
- É resumido na tripla Recurso, Propriedade, Valor.

Resource Description Framework

- Ex:

Breast Cancer → **Recurso**

Recurso

Câncer de Mama	
Tipo	Doença
Ocorre na	Glândula Mamária
Possui ontologia	DOID:1612
É mencionado em	PMID:28042876

Propriedade **Valor da Propriedade**

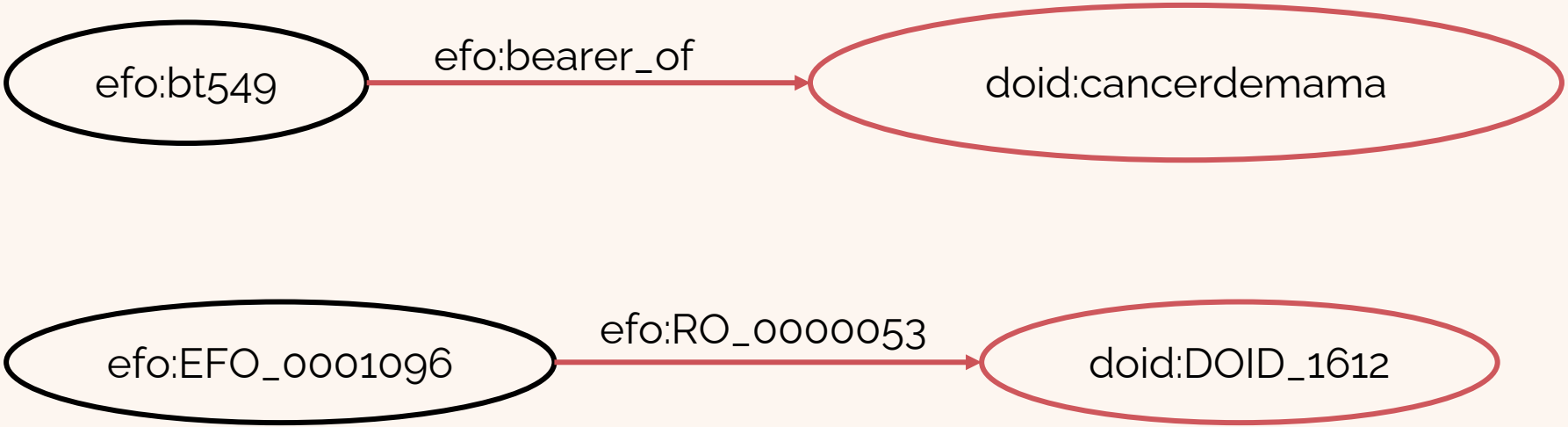
Resource Description Framework

- Grafo Direcionado e Rotulado (DLG)

- Sujeito: URI ou blank node
- Predicado: URI
- Objeto: URI, literal ou blank node

RDF – URI e predicado

- URI (Unified Resource Identifier)

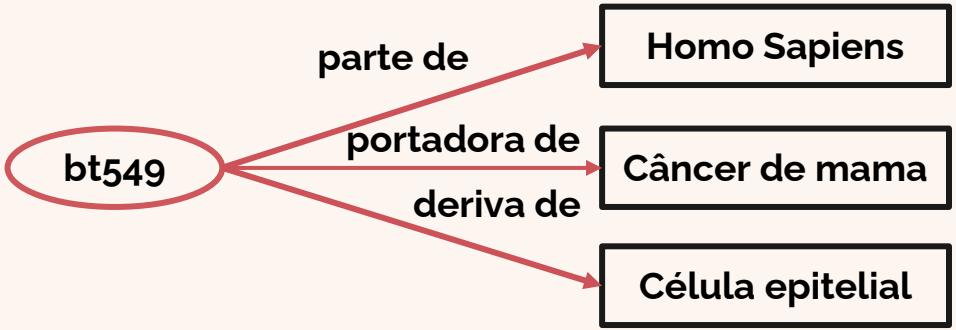


[scheme:][//authority][path][?query][#fragment]

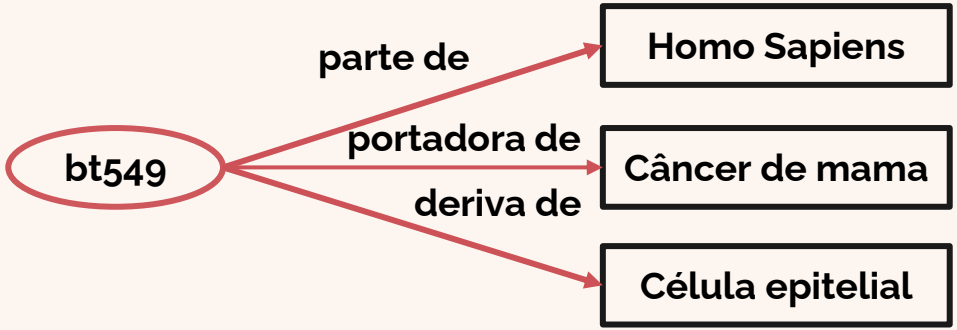
RDF – Exemplo

Sujeito	Predicado	Objeto
Linha celular bt549	Parte de	Homo Sapiens
Linha celular bt549	Portadora de	Câncer de mama
Linha celular bt549	Deriva de	Célula epitelial

http://www.ebi.ac.uk/efo/EFO_0001096



RDF – Exemplo


bt549
parte de
portadora de
deriva de
Homo Sapiens
Câncer de mama
Célula epitelial

http://www.ebi.ac.uk/efo/EFO_0001096
http://www.ebi.ac.uk/efo/efo.owlRO_0000053
http://www.ebi.ac.uk/efo/efo.owlBFO_0000050
http://www.ebi.ac.uk/efo/efo.owlRO_0001000
http://purl.obolibrary.org/obo/DOID_1612
http://purl.obolibrary.org/obo/NCBITaxon_9606
http://purl.obolibrary.org/obo/CL_0000066

RDF – Exemplo

Sujeito

Predicado

Objeto

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:efo="http://www.ebi.ac.uk/efo/efo.owl"
  xmlns:obo="http://purl.obolibrary.org/obo/"
>

  <rdf:Description rdf:about="http://www.ebi.ac.uk/efo/EFO_0001096">
     <efo:RO_0000053>http://purl.obolibrary.org/obo/DOID_1612</efo:RO_0000053>
   </rdf:Description>

   <rdf:Description rdf:about="http://www.ebi.ac.uk/efo/EFO_0001096">
     <efo:BFO_0000050>http://purl.obolibrary.org/obo/NCBITaxon_9606</efo:BFO_0000050>
   </rdf:Description>

   <rdf:Description rdf:about="http://www.ebi.ac.uk/efo/EFO_0001096">
     <efo:RO_0001000>http://purl.obolibrary.org/obo/CL_0000066</efo:RO_0001000>
   </rdf:Description>

 </rdf:RDF>
```

Linked Data

- Linked Data é uma materialização da Web Semântica.
- Melhores práticas para publicação e conexão de dados estruturados na web.
- O Linked Data é o meio para se tornar a web semântica.

Linked data: tornar a web sintática em web semântica

Linked Data

- Em 2006 Tim Berners-Lee estabelece os princípios para Linked Data.
- Usar URIs como nomes para coisas.
- Usar HTTP URIs para que pessoas possam pesquisar esses nomes.
- Quando alguém procurar um URI, forneça informações úteis, usando os padrões (RDF*, SPARQL)
- Incluir links para outros URIs, para que eles possam descobrir mais coisas.

https://www.w3.org/DesignIssues/LinkedData.html

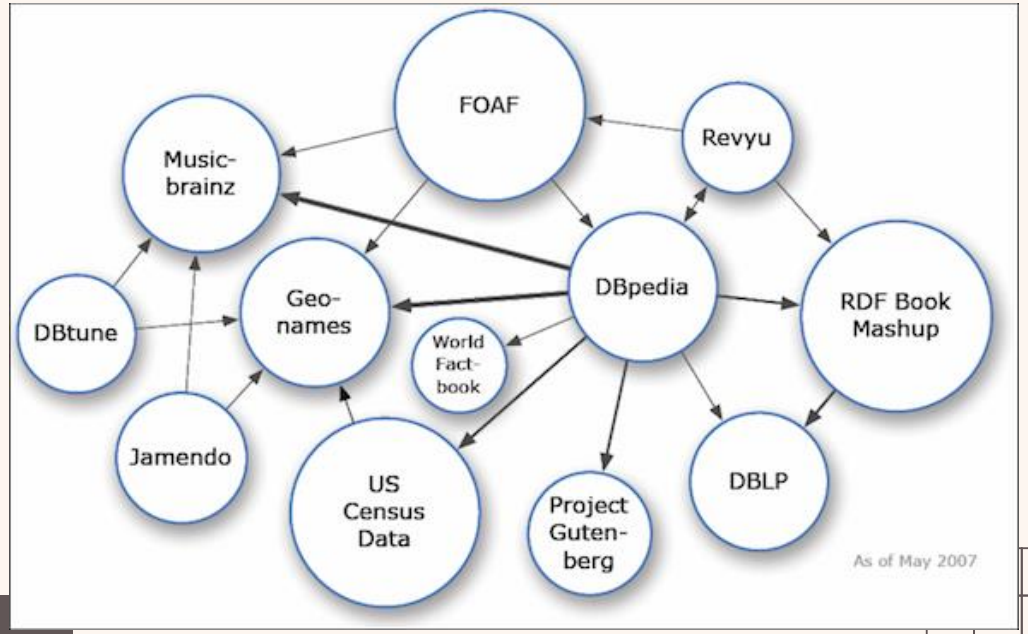
Linked Data

- Em 2006 Tim Berners-Lee estabelece os princípios para Linked Data.
- Usar URIs como nomes para coisas.
- Usar HTTP URIs para que pessoas possam pesquisar esses nomes.
- Quando alguém procurar um URI, forneça informações úteis, usando os padrões (RDF*, SPARQL)
- Incluir links para outros URIs, para que eles possam descobrir mais coisas.

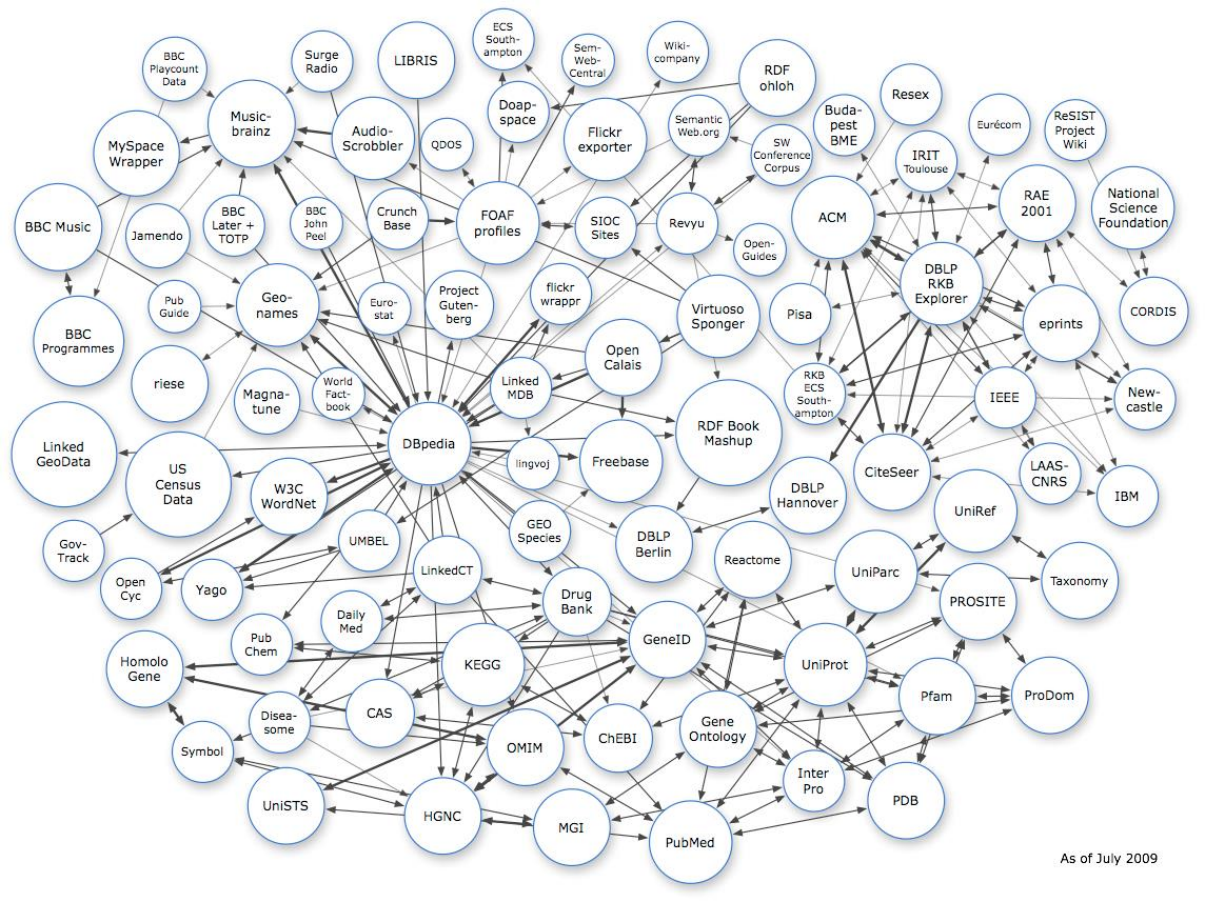
https://www.w3.org/DesignIssues/LinkedData.html

Linked Open Data (LOD)

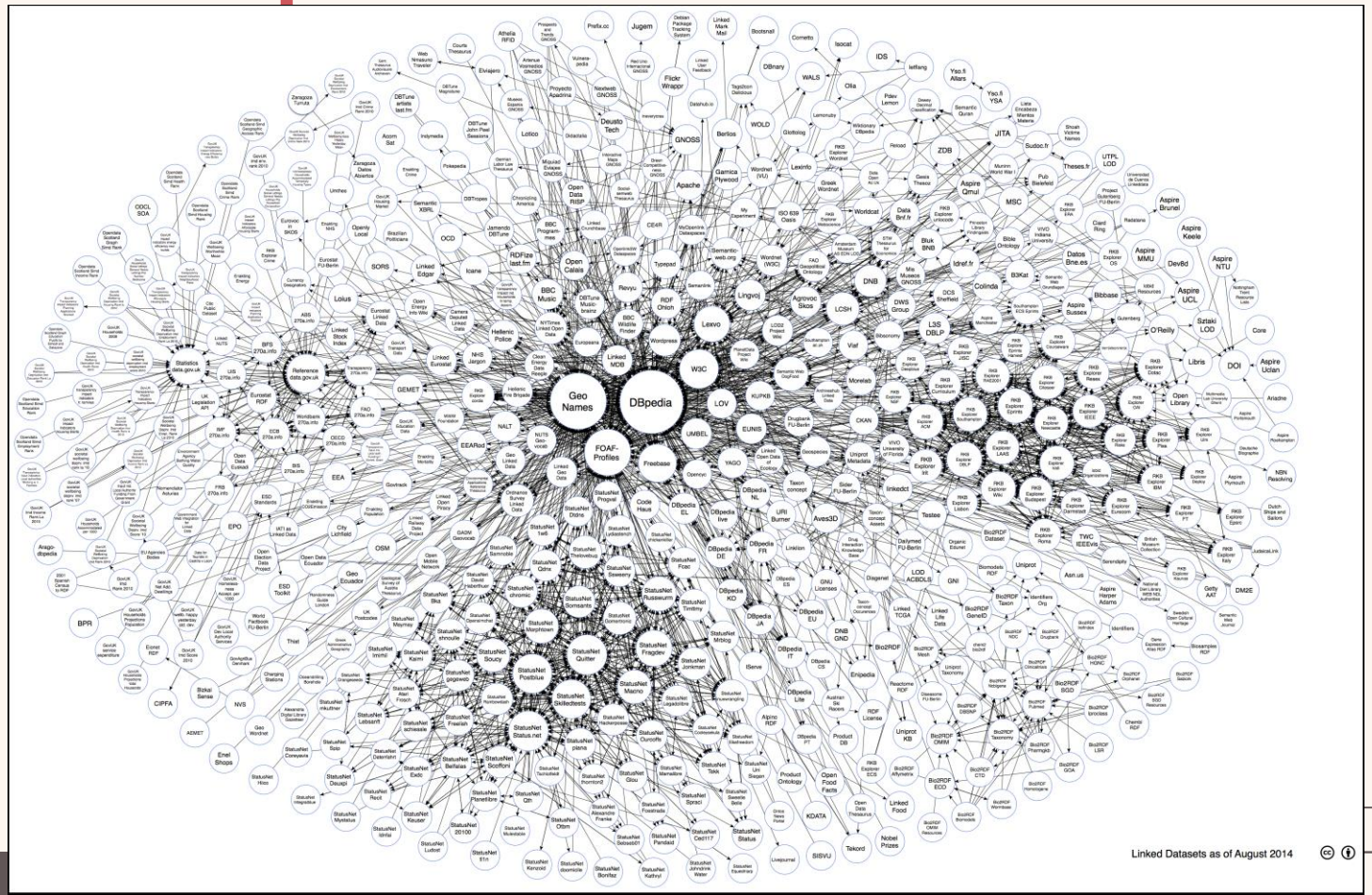
- Criado em 2007 pela comunidade W3C SWEO Linking Open Data.
- Maior exemplo de utilização de Linked Data.
- Gerar uma nuvem de dados composta por datasets estruturados em RDF.



Linked Open Data (LOD)



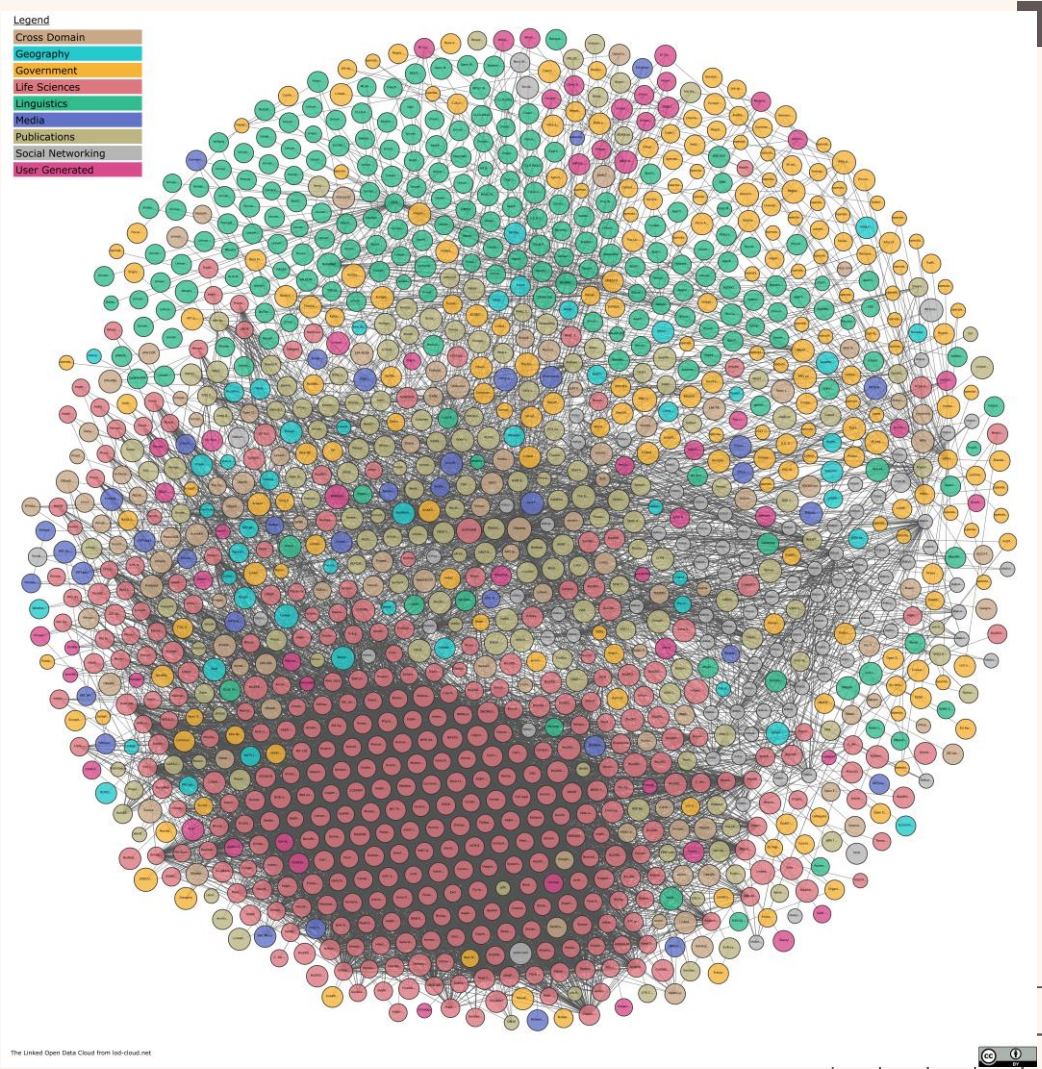
Linked Open Data (LOD)



LOD

Legend

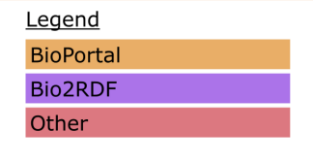
- Cross Domain
- Geography
- Government
- Life Sciences
- Linguistics
- Media
- Publications
- Social Networking
- User Generated

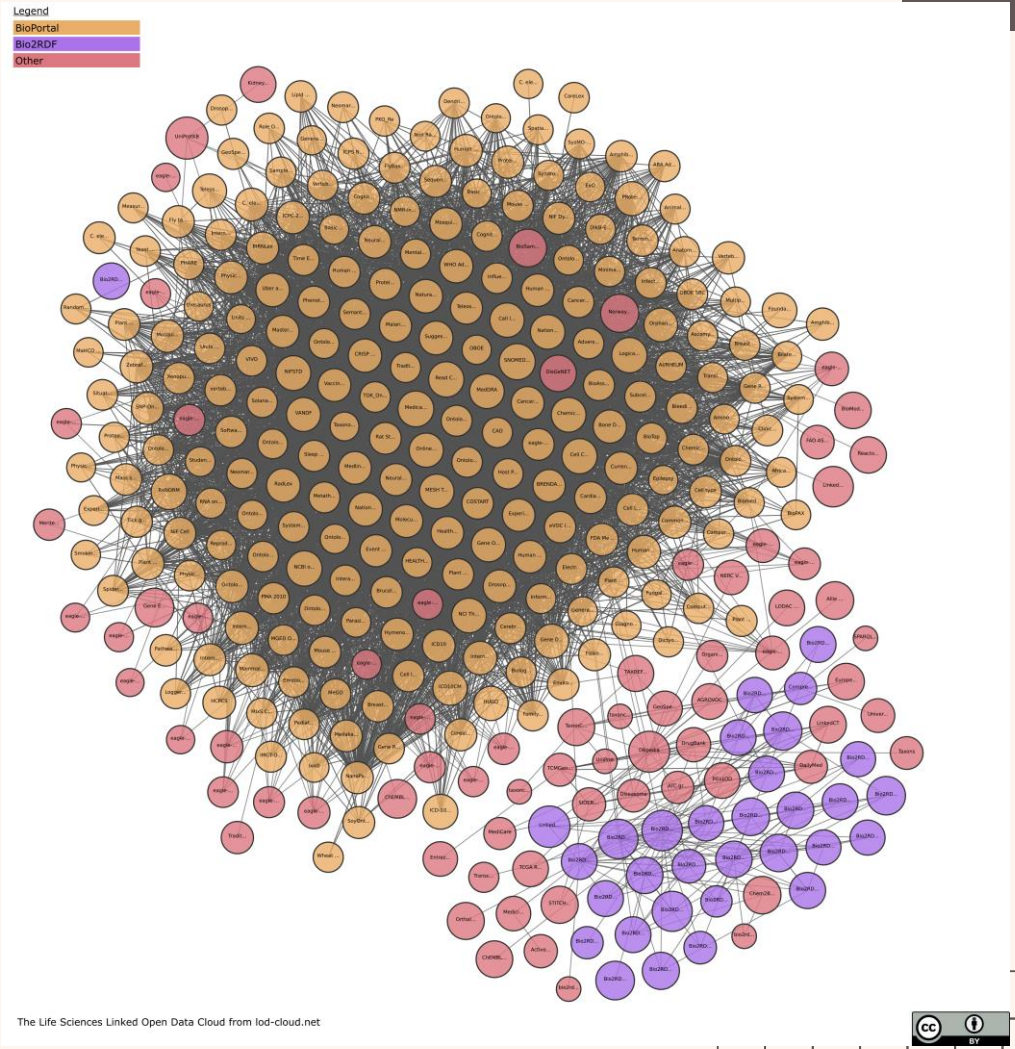


1.255 datasets
16.174 links

LOD

- LOD classificado pelo domínio de ciências da vida





LOD - DBPedia

- Uma das maiores bases e de maior destaque no LOD.
- Extrair informação estruturada da Wikipédia e tornar esta informação disponível semanticamente.

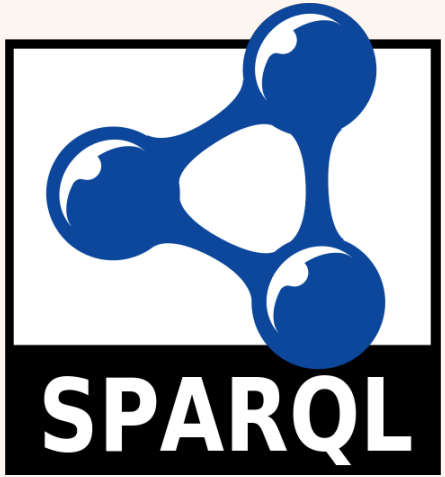
Formato em texto

DBpedia

Web semântico

SPARQL

- **S**PARQL **P**rotocol **a**nd **R**DF **Q**uery **L**anguage
- Linguagem de consulta e protocolo para acesso a RDF.
- Consultas podem ser realizadas via endpoint ou localmente com o uso do GraphDB por exemplo (https://graphdb.ontotext.com/)



SPARQL – Exemplo de Consulta

- DisGeNET – Coleção de genes e variantes associadas a doenças humanas.
- 1.134.942 de associações gene-doença.
- Pode ser usado em diversas pesquisas, como:
  - Análise das propriedades de genes de doenças.
  - Geração de hipóteses sobre ação terapêutica de drogas.
  - Investigação das bases moleculares de doenças humanas e suas comorbidades.
  - Avaliação de desempenho de métodos de mineração de texto
  - ....

- Acesso ao SPARQL endpoint

http://rdf.disgenet.org/sparql/

DisGeNET

SPARQL – Exemplo de Consulta

- Ao acessar o endpoit do DisGeNET uma consulta de exemplo é fornecida.

Query Text

```
SELECT DISTINCT ?gda ?disease ?gene
FROM <http://rdf.disgenet.org>
WHERE {
        ?gda sio:SIO_000628 ?disease,?gene .
        ?disease rdf:type ncit:C7057 .
        filter regex(?disease, "umls/id") .
        ?gene rdf:type ncit:C16612 .
        filter regex(?gene, "ncbigene") .
} LIMIT 20
```

O que queremos que retorne

De onde?

Em que condição?

SPARQL – Exemplo de Consulta

gda	disease	gene
http://rdf.disgenet.org/resource/gda/DGN6dc7c3597c1de43e46ba75dec6b40761	http://linkedlifedata.com/resource/umls/id/C0678202	http://identifiers.org/ncbigene/339145
http://rdf.disgenet.org/resource/gda/DGNcfa20e7b4de7e5d42a35f8d26059a05e	http://linkedlifedata.com/resource/umls/id/C0678202	http://identifiers.org/ncbigene/219771
http://rdf.disgenet.org/resource/gda/DGNca73fee7d721d58f0611394f44905b03	http://linkedlifedata.com/resource/umls/id/C0678202	http://identifiers.org/ncbigene/163486
http://rdf.disgenet.org/resource/gda/DGNfb14f7dc0edbfa4829ce273b25ee4725	http://linkedlifedata.com/resource/umls/id/C4225407	http://identifiers.org/ncbigene/1475
http://rdf.disgenet.org/resource/gda/DGNd91c8743be0a2353356e9547b64d096d	http://linkedlifedata.com/resource/umls/id/C4225407	http://identifiers.org/ncbigene/1475
http://rdf.disgenet.org/resource/gda/DGN3f0f82dae6a0319ecf0caa14c72850d9	http://linkedlifedata.com/resource/umls/id/C4225407	http://identifiers.org/ncbigene/1475
http://rdf.disgenet.org/resource/gda/DGN79d0d1ef1d572d0ffda6560ddd1e6404	http://linkedlifedata.com/resource/umls/id/C4225407	http://identifiers.org/ncbigene/1475
http://rdf.disgenet.org/resource/gda/DGNc4185052f39ed8a1d4f0e26006d2c104	http://linkedlifedata.com/resource/umls/id/C4540327	http://identifiers.org/ncbigene/2186
http://rdf.disgenet.org/resource/gda/DGN30f78e4ac417d54545fa146bf2d02242	http://linkedlifedata.com/resource/umls/id/C4540327	http://identifiers.org/ncbigene/2186
http://rdf.disgenet.org/resource/gda/DGN9d2679e6a6d1fd0da6a1e8f330af69d3	http://linkedlifedata.com/resource/umls/id/C4540327	http://identifiers.org/ncbigene/2186
http://rdf.disgenet.org/resource/gda/DGNbcff9614a7de394c8223a9f52f605e79	http://linkedlifedata.com/resource/umls/id/C0678202	http://identifiers.org/ncbigene/345611
http://rdf.disgenet.org/resource/gda/DGN73c3915530c50ec44a9fb199a64eb462	http://linkedlifedata.com/resource/umls/id/C0678202	http://identifiers.org/ncbigene/345611
http://rdf.disgenet.org/resource/gda/DGN967a7628e7c3f0fca4af8ec7e6da69fd	http://linkedlifedata.com/resource/umls/id/C0678202	http://identifiers.org/ncbigene/345611
http://rdf.disgenet.org/resource/gda/DGN26063453fcd479d596132ae4a8638f6f	http://linkedlifedata.com/resource/umls/id/C0678202	http://identifiers.org/ncbigene/345611
http://rdf.disgenet.org/resource/gda/DGNc4b1c8f460b9fc1caf70eb442e6f508f	http://linkedlifedata.com/resource/umls/id/C0029810	http://identifiers.org/ncbigene/3658
http://rdf.disgenet.org/resource/gda/DGNc860b71dcd16458cce481013c8b47141	http://linkedlifedata.com/resource/umls/id/C0029810	http://identifiers.org/ncbigene/3658
http://rdf.disgenet.org/resource/gda/DGN7f14f020a0e7b2f88e9deae5e6f68e62	http://linkedlifedata.com/resource/umls/id/C0678202	http://identifiers.org/ncbigene/4689
http://rdf.disgenet.org/resource/gda/DGNea67de178fded50ede1d8d44b0acdada	http://linkedlifedata.com/resource/umls/id/C0678202	http://identifiers.org/ncbigene/159296
http://rdf.disgenet.org/resource/gda/DGNdac98a0d2ab9c130177d7943fa5fd88f	http://linkedlifedata.com/resource/umls/id/C0678202	http://identifiers.org/ncbigene/159296
http://rdf.disgenet.org/resource/gda/DGNa43f397d79061a29c32a697e7dd10876	http://linkedlifedata.com/resource/umls/id/C0678202	http://identifiers.org/ncbigene/159296

SPARQL – Exemplo de Consulta

Query Text

```
SELECT DISTINCT ?gda ?diseaseName ?geneSymbol
FROM <http://rdf.disgenet.org>
WHERE {
        ?gda sio:SIO_000628 ?disease,?gene .
        ?disease rdf:type ncit:C7057 ;
                dcterms:title ?diseaseName .
        filter regex(?disease, "umls/id") .
        ?gene rdf:type ncit:C16612 .
        ?symbolUri dcterms:title ?geneSymbol .
        filter regex(?gene, "ncbigene") .
} LIMIT 20
```

http://rdf.disgenet.org/resource/gda/DGN6dc7c3597c1de43e46ba75dec6b40761	"Regional enteritis"@en	"RNY3"@en
http://rdf.disgenet.org/resource/gda/DGN6dc7c3597c1de43e46ba75dec6b40761	"Regional enteritis"@en	"RNY4"@en
http://rdf.disgenet.org/resource/gda/DGN6dc7c3597c1de43e46ba75dec6b40761	"Regional enteritis"@en	"RNY5"@en
http://rdf.disgenet.org/resource/gda/DGN6dc7c3597c1de43e46ba75dec6b40761	"Regional enteritis"@en	"SYPL1"@en
http://rdf.disgenet.org/resource/gda/DGN6dc7c3597c1de43e46ba75dec6b40761	"Regional enteritis"@en	"ZNF3"@en
http://rdf.disgenet.org/resource/gda/DGN6dc7c3597c1de43e46ba75dec6b40761	"Regional enteritis"@en	"ZNF12"@en
http://rdf.disgenet.org/resource/gda/DGN6dc7c3597c1de43e46ba75dec6b40761	"Regional enteritis"@en	"ZKSCAN1"@en
http://rdf.disgenet.org/resource/gda/DGN6dc7c3597c1de43e46ba75dec6b40761	"Regional enteritis"@en	"ZSCAN21"@en