

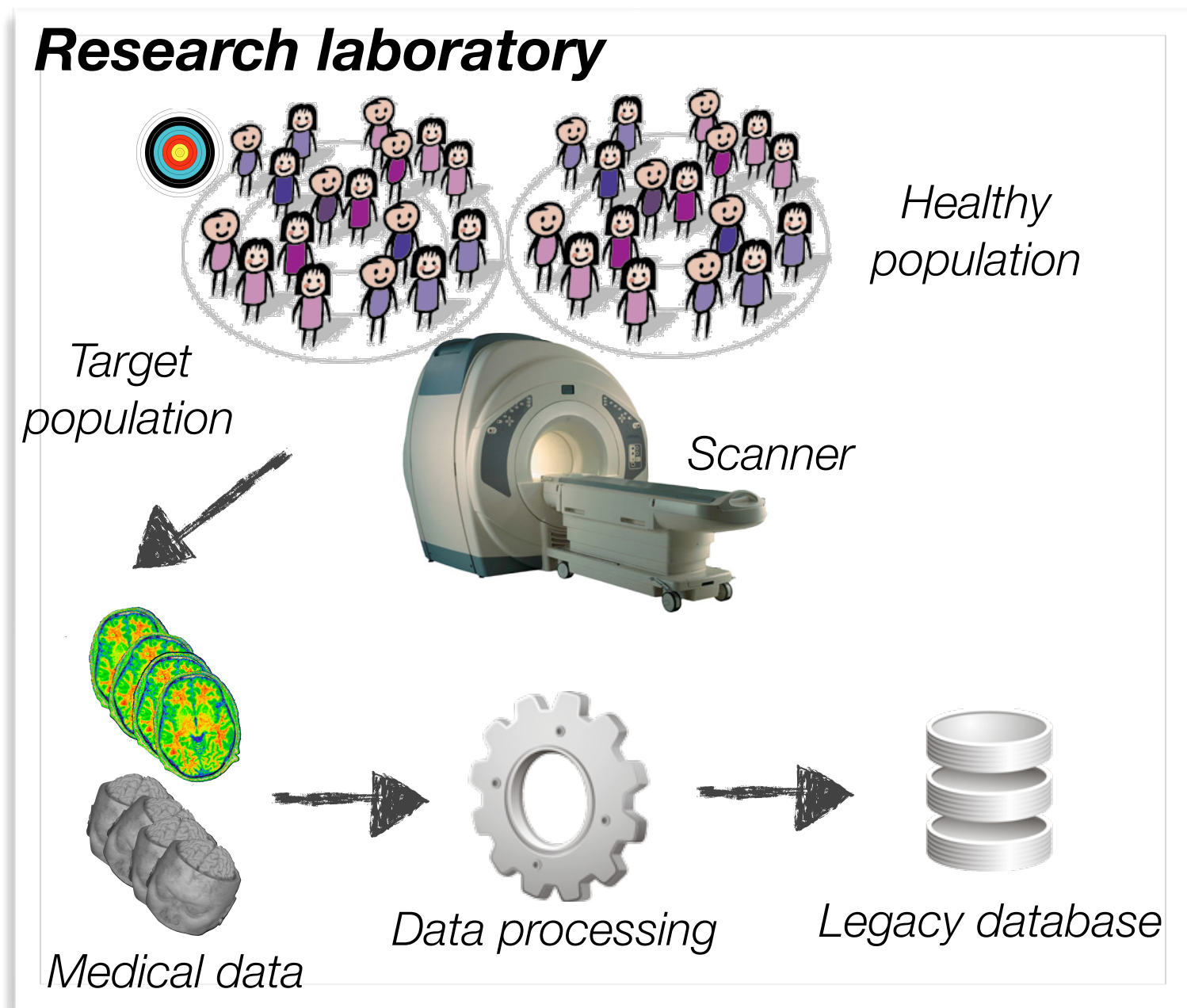
(semantic) Virtual Research Environment for Life Sciences

CARGO Day
19/11/2015

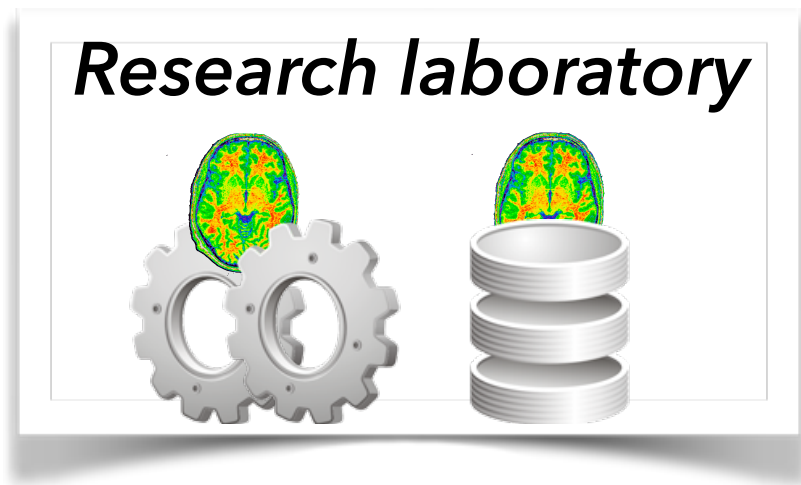
Alban Gaignard, CNRS - CHU Nantes
alban.gaignard@univ-nantes.fr

Audrey Bihouée, Université de Nantes
audrey.bihouee@univ-nantes.fr

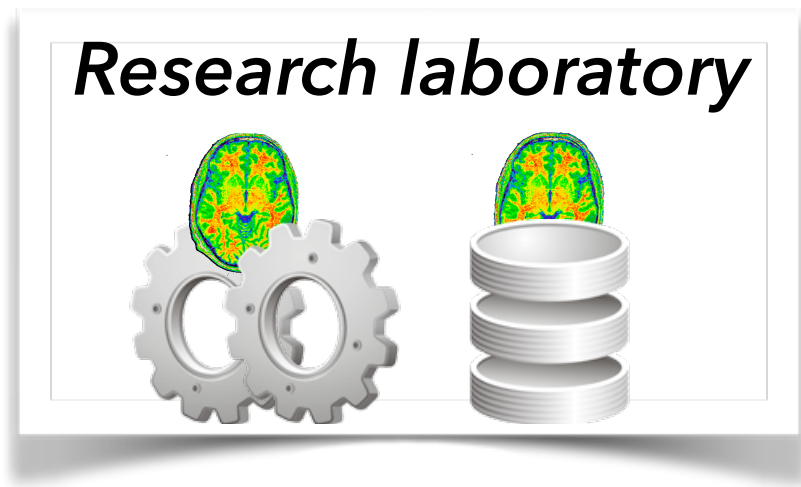
(big) Data-centric research



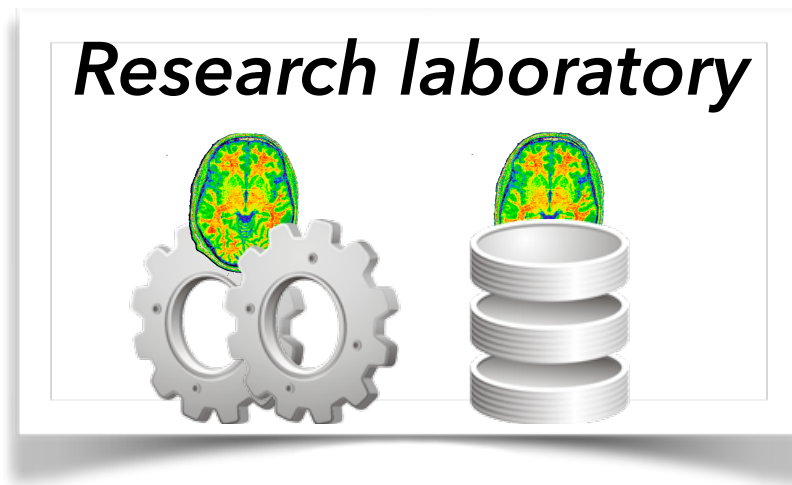
(big) Data-centric research



(big) Data-centric research

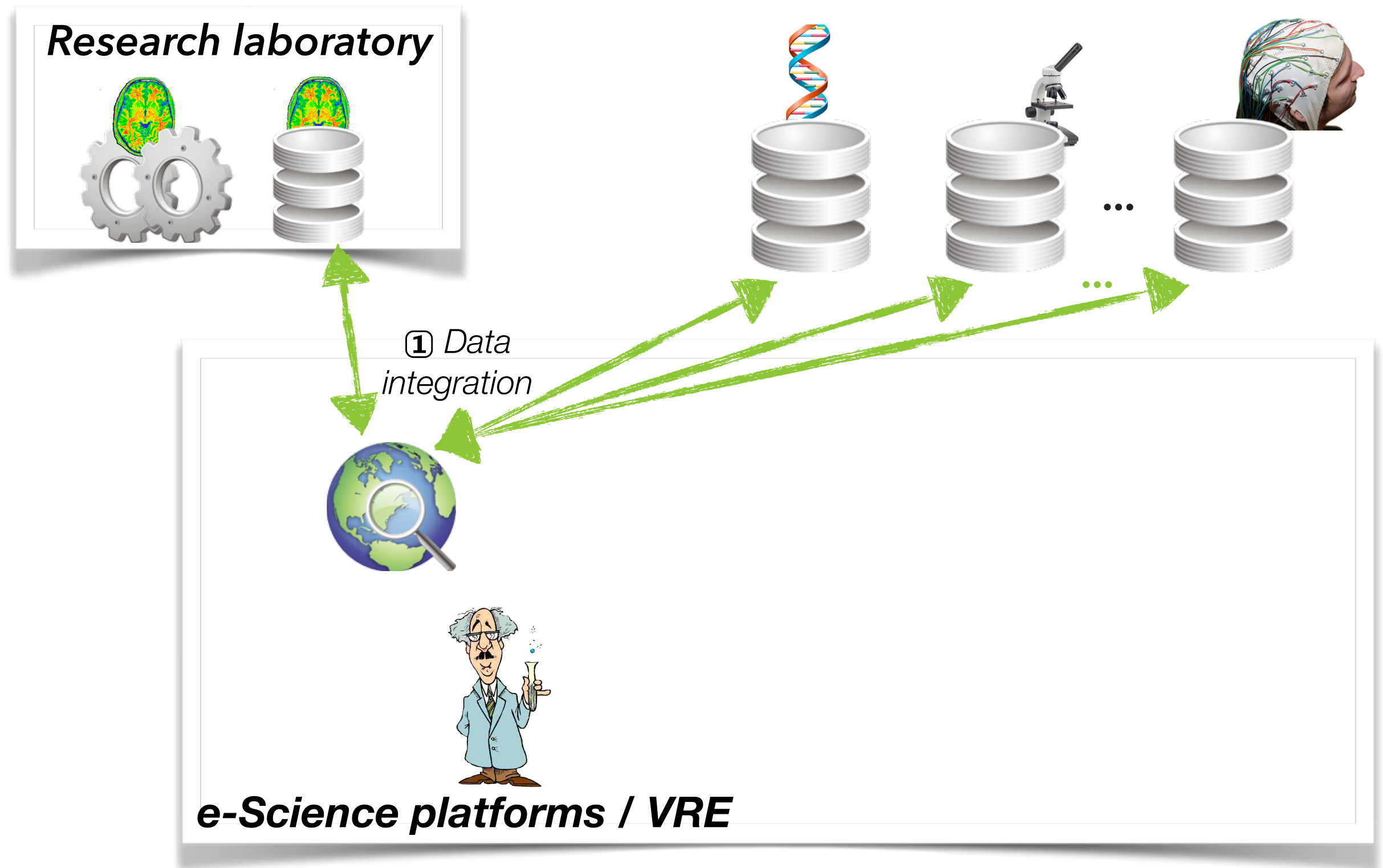


(big) Data-centric research

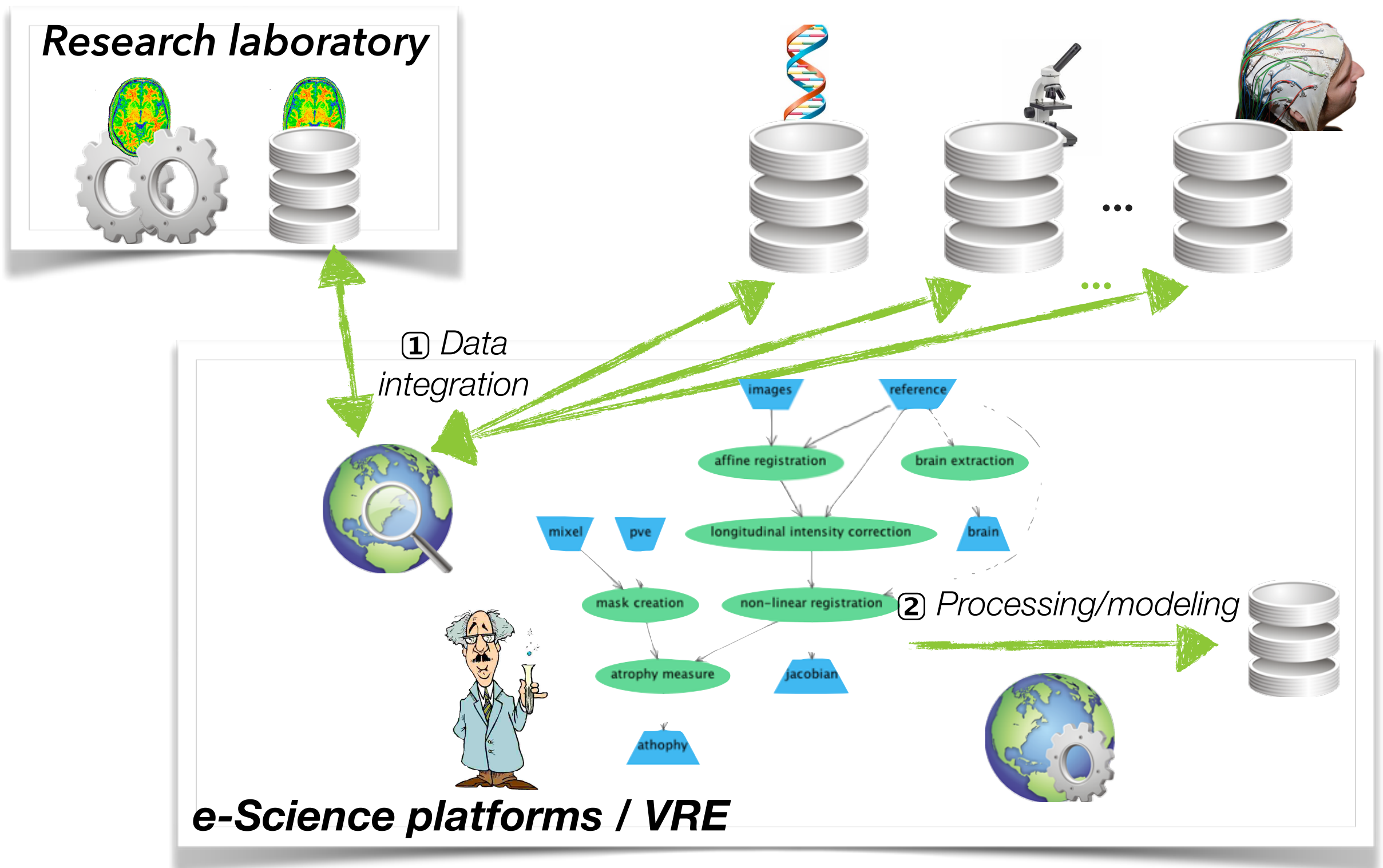


e-Science platforms / VRE

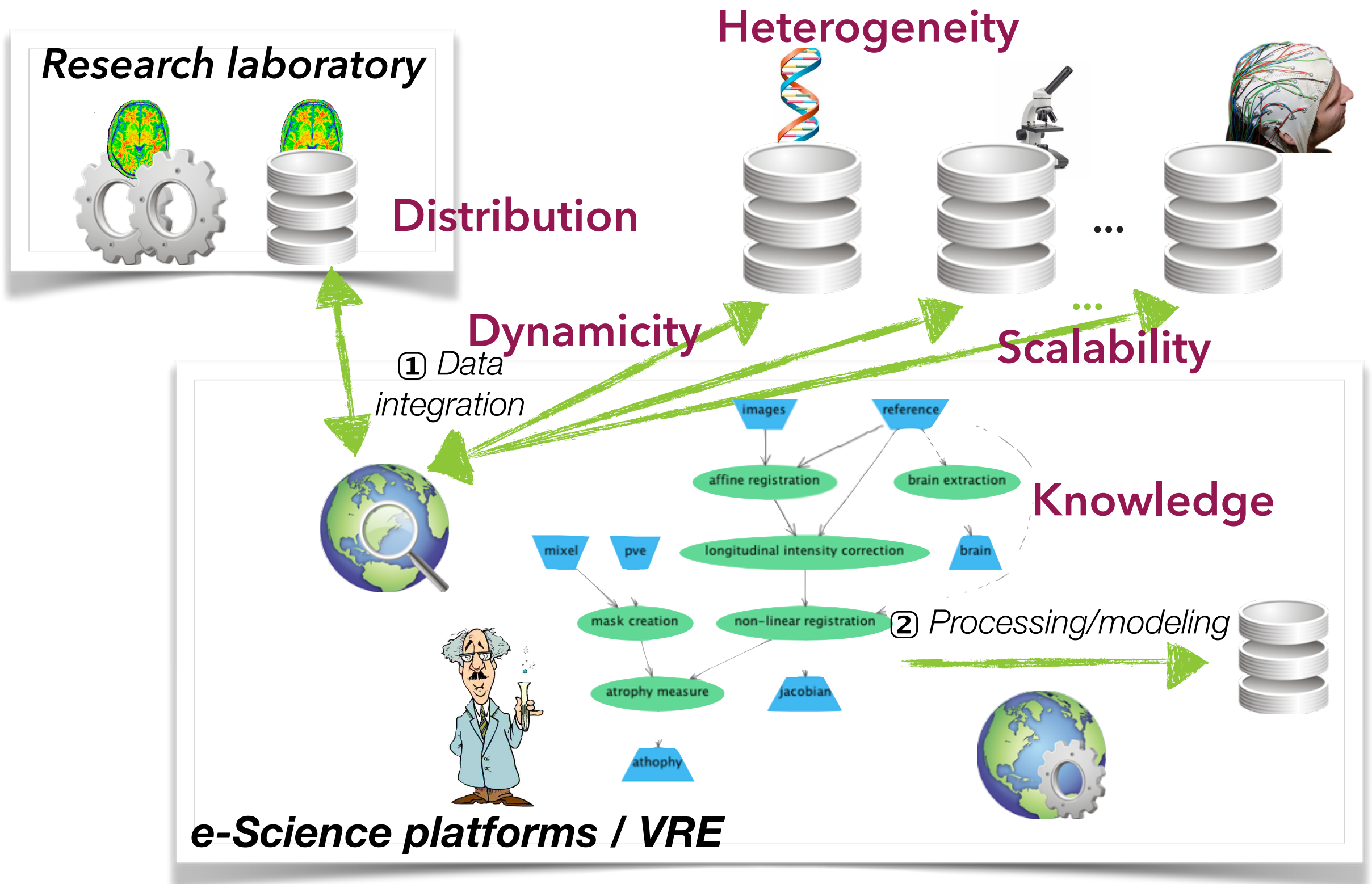
(big) Data-centric research



(big) Data-centric research



(big) Data-centric research



Biologiste



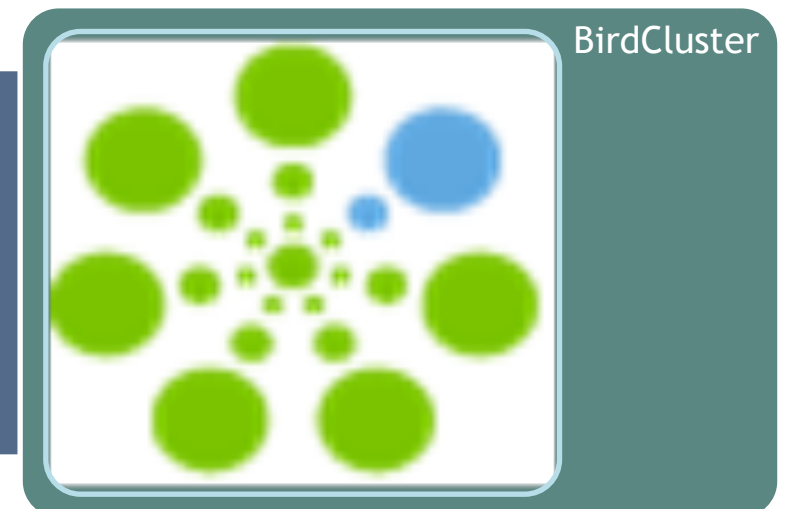
- Mise à disposition de logiciels bioinformatiques
- Mise à disposition de workflows d'analyse



Bioinformaticien



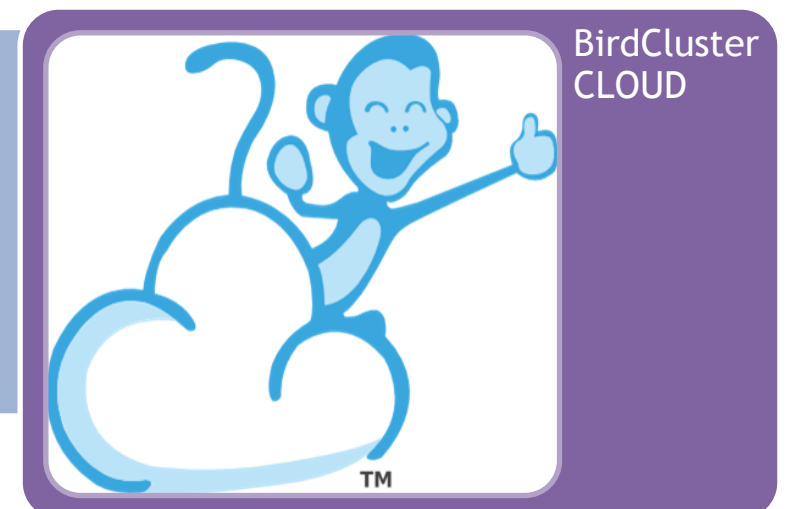
- Mise à disposition de ressources de calcul pour le développement d'outils
- Mise à disposition d'outils de développement



Informaticien



- Test de nouveaux algorithmes de calcul
- Test de performance des infrastructures avec des données réelles



Projets bioinformatiques 2014-2015

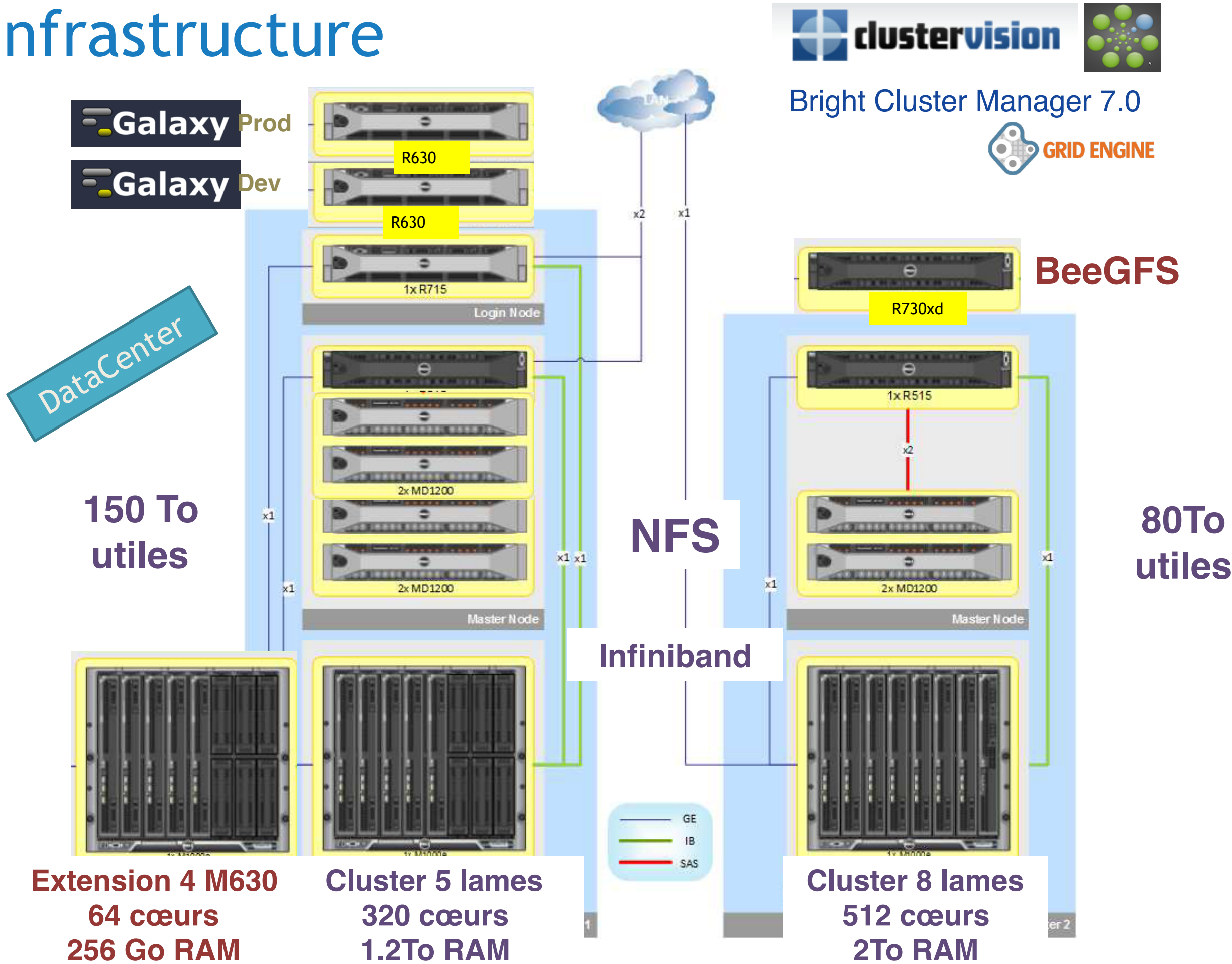
41 projets "NGS" dont :

- **Exome** (génomique - ADN)
projet HugoDims : 180 exomes + développement d'un workflow dédié,
- **RNAseq** (expression - ARN)
projet ERRATA : 200 échantillons à séquencer en oncologie

18 projets "puces ADN"

Temps CPU (jobs) : 6.3 années Stockage: 64 To
--

Infrastructure



"Volume" en génomique

TopHat : alignement d'une séquence avec un génome de référence

RNA-seq : séquençage de tous les ARNs d'un échantillon pour mesurer différents niveaux d'expression des gènes

1 échantillon	
Données brutes	2x 17Go
Données produites	12 Go
Temps CPU (1 core)	170 h
Temps CPU (32 cores)	32 h

"Volume" en génomique

TopHat : alignement d'une séquence avec un génome de référence

RNA-seq : séquençage de tous les ARNs d'un échantillon pour mesurer différents niveaux d'expression des gènes

	1 échantillon	300 échantillons
Données brutes	2x 17Go	10.2 To
Données produites	12 Go	3.6 To
Temps CPU (1 core)	170 h	5.9 ans
Temps CPU (32 cores)	32 h	13 mois

"Volume" en génomique

TopHat : alignement d'une séquence avec un génome de référence

RNA-seq : séquençage de tous les ARNs d'un échantillon pour mesurer différents niveaux d'expression des gènes

	1 échantillon	300 échantillons	30 projets
Données brutes	2x 17Go	10.2 To	306 To
Données produites	12 Go	3.6 To	108 To
Temps CPU (1 core)	170 h	5.9 ans	177 ans
Temps CPU (32 cores)	32 h	13 mois	32 ans

"Volume" en génomique

TopHat : alignement d'une séquence avec un génome de référence

RNA-seq : séquençage de tous les ARNs d'un échantillon pour mesurer différents niveaux d'expression des gènes

	1 échantillon	300 échantillons	30 projets
Données brutes	2x 17Go	10.2 To	306 To
Données produites	12 Go	3.6 To	108 To
Temps CPU (1 core)	170 h	5.9 ans	177 ans
Temps CPU (32 cores)	32 h	13 mois	32 ans

- **Performance** (algorithmique, parallélisme)
- **Stockage** (et préservation ?)
- **Utilisation secondaire** des données traitées (indexation sémantique)
- **Reproductibilité** (qualité) des analyses de données

Semantic Web

(yet another) introduction

Initially a **Web** of documents ...

Initially a **Web of documents** ...

*From **untyped** HTML pages and links ...*

*... towards more **"intelligent"** applications.*

Initially a **Web of documents** ...

*From **untyped** HTML pages and links ...*

*... towards more **"intelligent"** applications.*

How **machine and humans** can **better interpret** diverse data structures and content ?

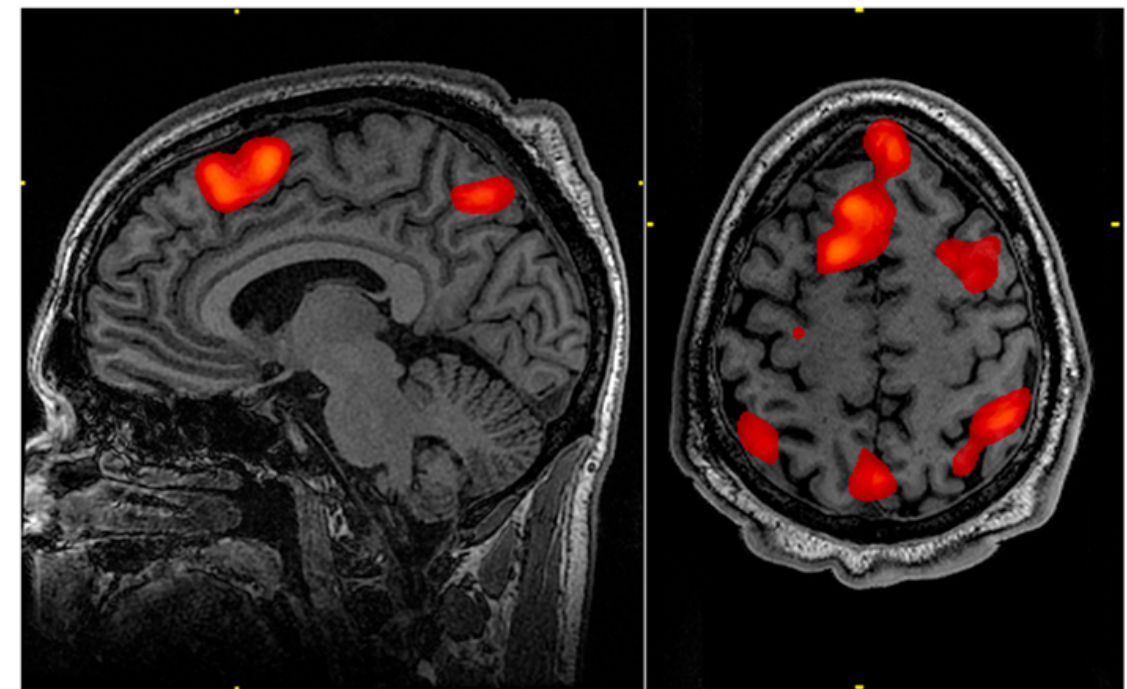
Initially a **Web of documents** ...

From *untyped* HTML pages and links ...

... towards more “*intelligent*” applications.

How **machine and humans** can **better interpret** diverse data structures and content ?

```
@SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
GGGTGATGGCCGCTGCCGATGGCGTCAAATCCCACC
+SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII9IG9IC
```



Initially a **Web of documents** ...

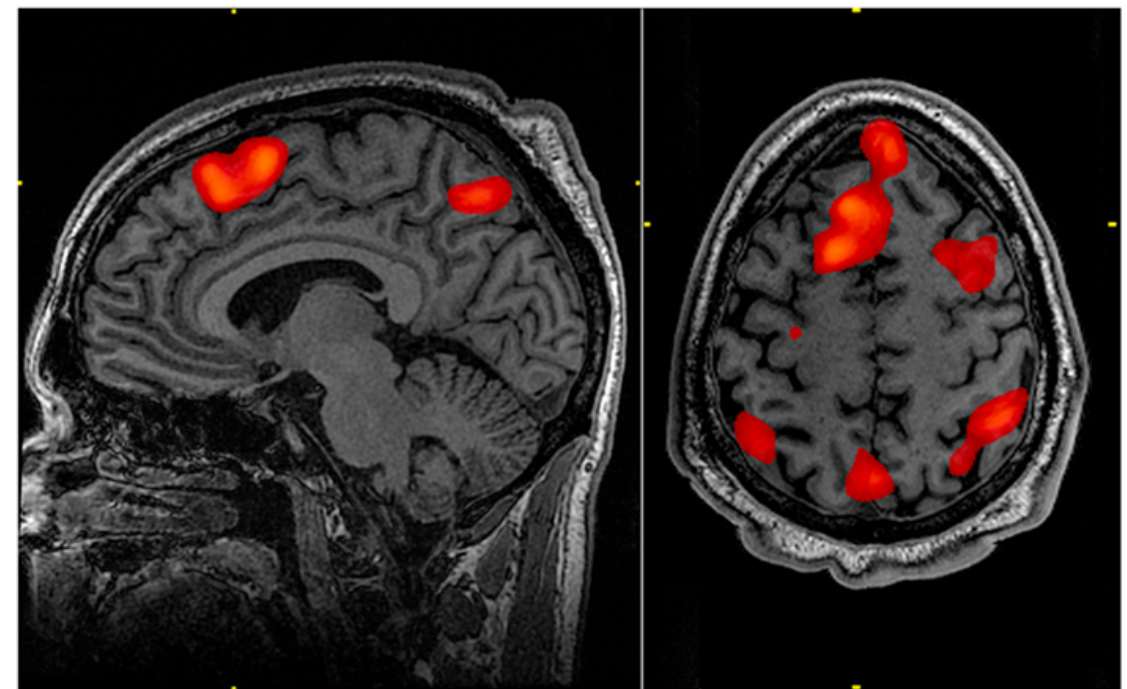
*From **untyped** HTML pages and links ...*

*... towards more **"intelligent"** applications.*

How **machine and humans** can **better interpret** diverse data structures and content ?

```
@SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36  
GGGTGATGGCCGCTGCCGATGGCGTCAAATCCCACC  
+SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36  
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII9IG9IC
```

meta-data !!



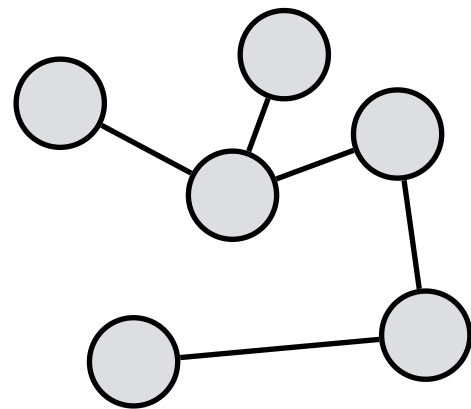
... moving towards a **Web of Data**

... moving towards a **Web of Data**

Internet + (deep) Web

... moving towards a **Web of Data**

Information



**Identify
& Link Data**

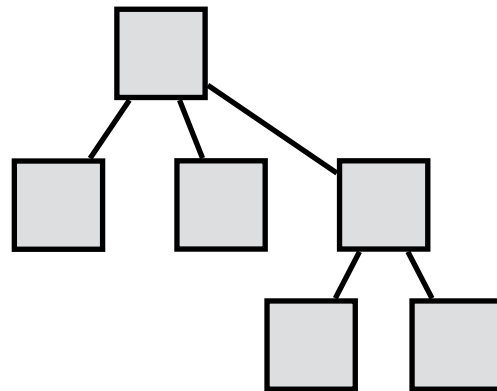


Querying

Internet + (deep) Web

... moving towards a **Web of Data**

Knowledge

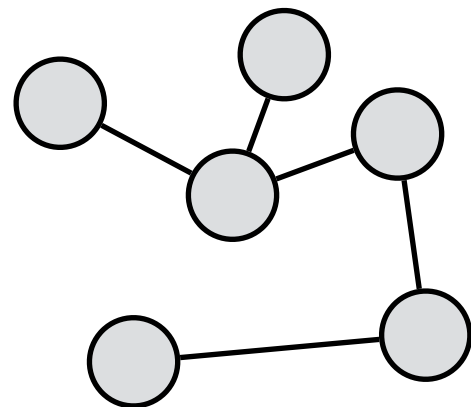


**Vocabularies
Ontologies**

Reasoning



Information



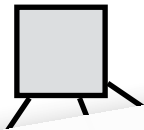
**Identify
& Link Data**

Querying

Internet + (deep) Web

... moving towards a **Web of Data**

Knowledge



T. R. Gruber. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199-220, 1993

1. Introduction

A body of formally represented knowledge is based on a *conceptualization*: the objects, concepts, and other entities that are presumed to exist in some area of interest and the relationships that hold them (Genesereth & Nilsson, 1987). A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualization, explicitly or implicitly.

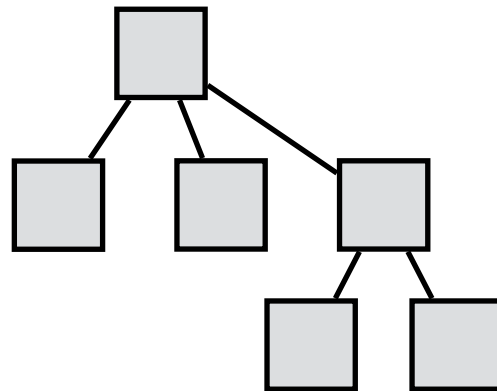
An ontology is an explicit specification of a conceptualization. The term is borrowed from philosophy, where an ontology is a systematic account of Existence. For knowledge-based systems, what "exists" is exactly that which can be represented.

Querying

.....
Internet + (deep) Web

... moving towards a **Web of Data**

Knowledge

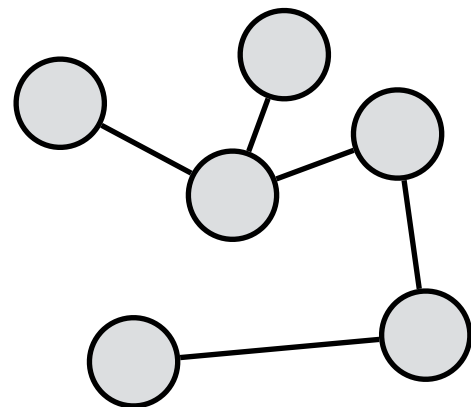


**Vocabularies
Ontologies**

Reasoning



Information



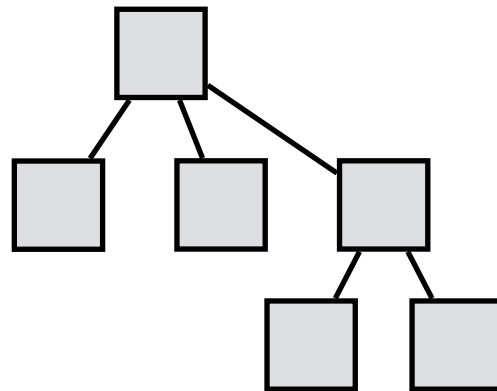
**Identify
& Link Data**

Querying

Internet + (deep) Web

... moving towards a **Web of Data**

Knowledge

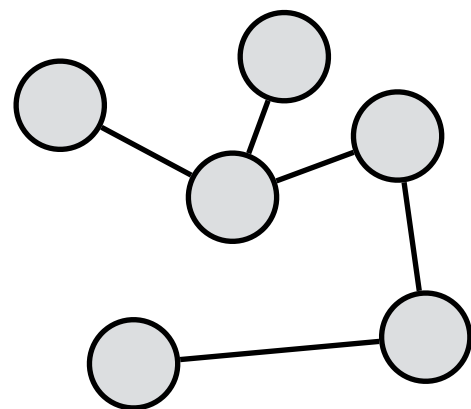


Vocabularies
Ontologies

Reasoning



Information



URI

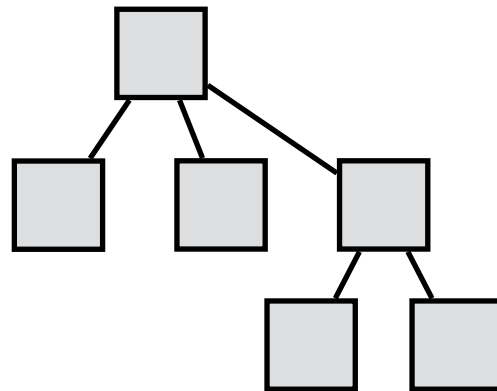
Identify
& Link Data

Querying

Internet + (deep) Web

... moving towards a **Web of Data**

Knowledge

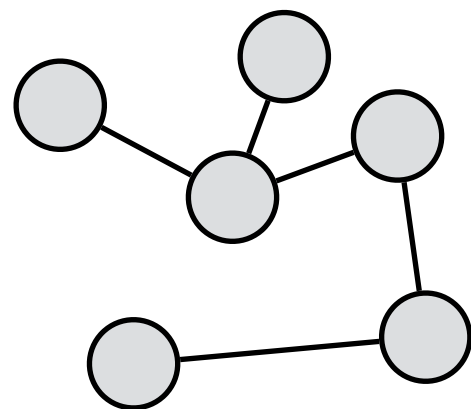


Vocabularies
Ontologies

Reasoning



Information



URI
RDF

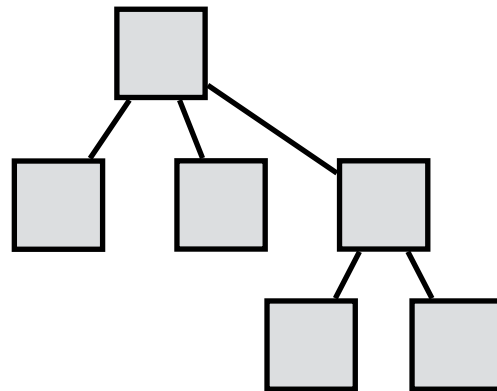
Identify
& Link Data

Querying

Internet + (deep) Web

... moving towards a **Web of Data**

Knowledge

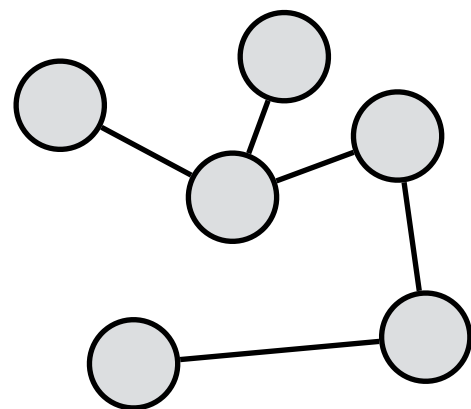


Vocabularies
Ontologies

Reasoning



Information



URI
RDF

Identify
& Link Data

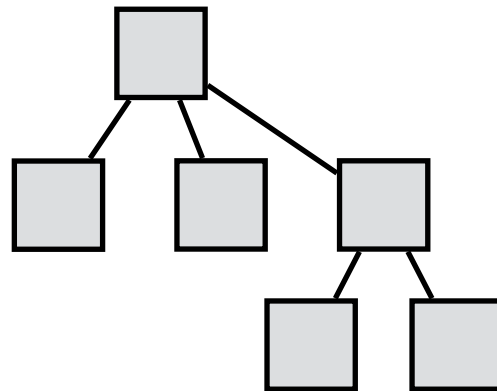
Querying

SPARQL

Internet + (deep) Web

... moving towards a **Web of Data**

Knowledge

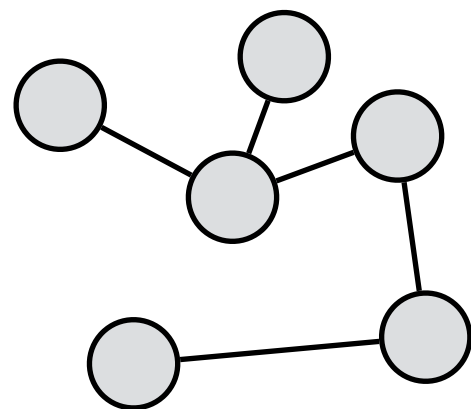


Vocabularies
Ontologies

Reasoning



Information



URI
RDF
JSON-LD

Identify
& Link Data

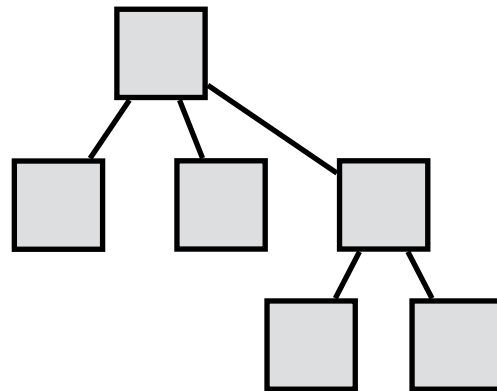
Querying

SPARQL

Internet + (deep) Web

... moving towards a **Web of Data**

Knowledge



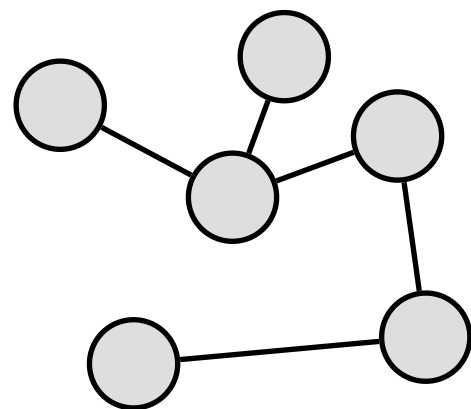
Vocabularies
Ontologies

RDFS

Reasoning



Information



URI
RDF
JSON-LD

Identify
& Link Data

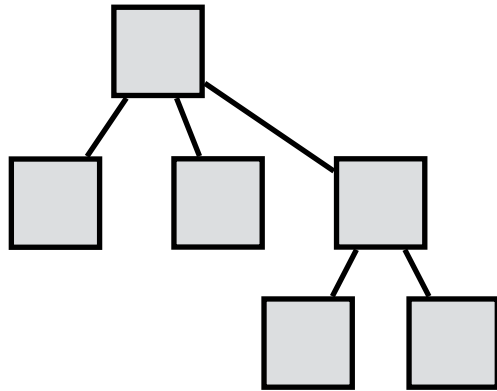
Querying

SPARQL

Internet + (deep) Web

... moving towards a **Web of Data**

Knowledge



Vocabularies
Ontologies

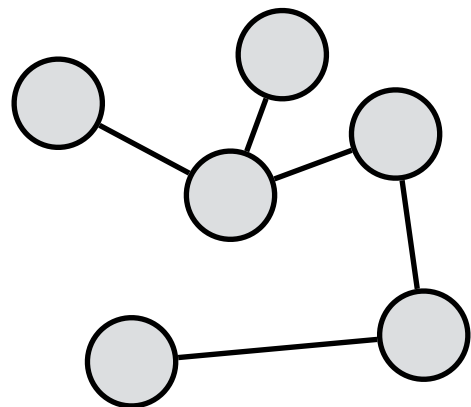
RDFS

OWL

Reasoning



Information



URI

RDF

JSON-LD

Identify
& Link Data

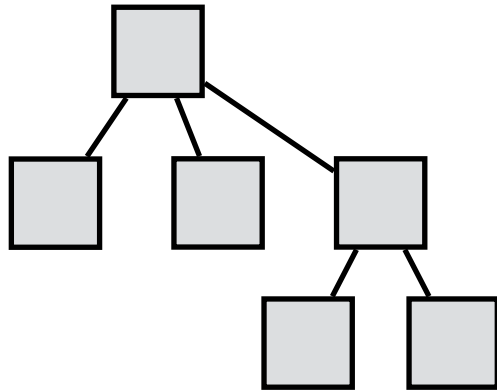
Querying

SPARQL

Internet + (deep) Web

... moving towards a **Web of Data**

Knowledge



Vocabularies
Ontologies

RDFS

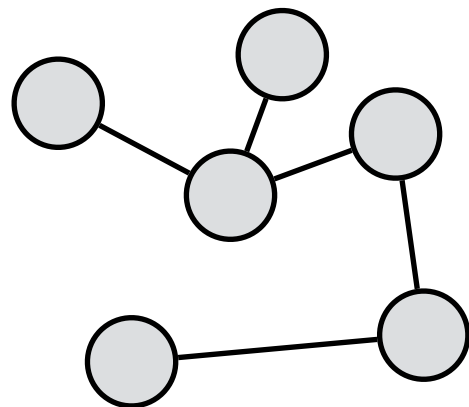
RIF

OWL

Reasoning



Information



URI

RDF

JSON-LD

Identify
& Link Data

Querying

SPARQL

Internet + (deep) Web



WIKIPEDIA
The Free Encyclopedia

Article **Talk**

Read

Edit

View history

Search



BRCA1

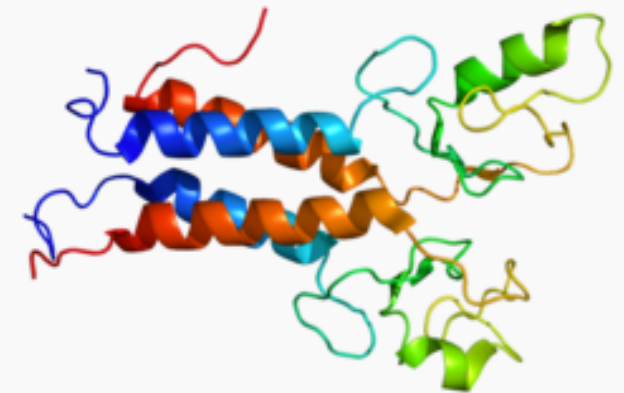
From Wikipedia, the free encyclopedia

BRCA1 and **BRCA1** (/ˈbrækəˈwʌn/^[1]) are a human [gene](#) and its [protein](#) product, respectively. The official symbol (BRCA1, italic for the gene, nonitalic for the protein) and the official name (**breast cancer 1, early onset**) are maintained by the [HGNC](#). [Orthologs](#), styled *Brca1* and *Brca1*, are common in other [mammal](#) species.^[2] *BRCA1* is a human [tumor suppressor gene](#)^{[3][4]} (to be specific, a [caretaker gene](#)), found in all humans; its [protein](#), also called by the [synonym](#) **breast cancer type 1 susceptibility protein**, is responsible for repairing DNA.^[5]

BRCA1 and [BRCA2](#) are normally expressed in the cells of [breast](#) and other tissue, where they help repair damaged [DNA](#) or destroy cells if DNA cannot be repaired. They are involved in the repair of [chromosomal](#) damage with an important role in the error-free [repair of DNA](#) double-strand breaks.^{[6][7]} If BRCA1 or BRCA2 itself is damaged by a [BRCA mutation](#), damaged DNA is not repaired properly, and this increases the risk for [breast cancer](#).^{[8][9]} Thus, although the terms "breast cancer susceptibility gene" and "breast cancer susceptibility protein" (used frequently both in and outside the [medical literature](#)) sound as if they describe an [oncogene](#), *BRCA1* and *BRCA2* are normal; it is [their mutation](#) that is abnormal.

BRCA1 combines with other tumor suppressors, DNA damage sensors, and signal transducers to form a large multi-subunit protein complex known as the BRCA1-associated genome surveillance complex (BASC).^[10] The BRCA1 protein associates with [RNA polymerase II](#), and through the [C-terminal](#) domain, also interacts with [histone deacetylase](#) complexes. Thus, this protein plays a role in transcription, [DNA](#)

Breast cancer 1, early onset



PDB rendering based on 1jm7.

Available structures

PDB

Ortholog search: [PDBe](#) [RCSB](#)

List of PDB id codes

[\[show\]](#)

Identifiers

Symbols

[BRCA1](#) ; [BRCAI](#); [BRCC1](#); [BROVCA1](#); [IRIS](#); [PNCA4](#); [PPP1R53](#); [PSCP](#); [RNF53](#)

External IDs

[OMIM: 113705](#) [MGI: 104537](#)
[HomoloGene: 5276](#) [ChEMBL: 5990](#) [GeneCards: BRCA1 Gene](#)

[Gene ontology](#)

[\[show\]](#)

“BRCA1 is a human gene”
“a BRCA1 mutation increases breast cancer risk ”

[Main page](#)

[Contents](#)

[Featured content](#)

[Current events](#)

[Random article](#)

[Donate to Wikipedia](#)

[Wikimedia Shop](#)

Interaction

[Help](#)

[About Wikipedia](#)

[Community portal](#)

[Recent changes](#)

[Contact page](#)

Tools

[What links here](#)

[Related changes](#)

[Upload file](#)

[Special pages](#)

[Permanent link](#)

[Page information](#)

[Wikidata item](#)

[Cite this page](#)

Print/export

[Create a book](#)

RDF to link data and represent knowledge

RDF to link data and represent knowledge

"BRCA1 – is a – human gene"

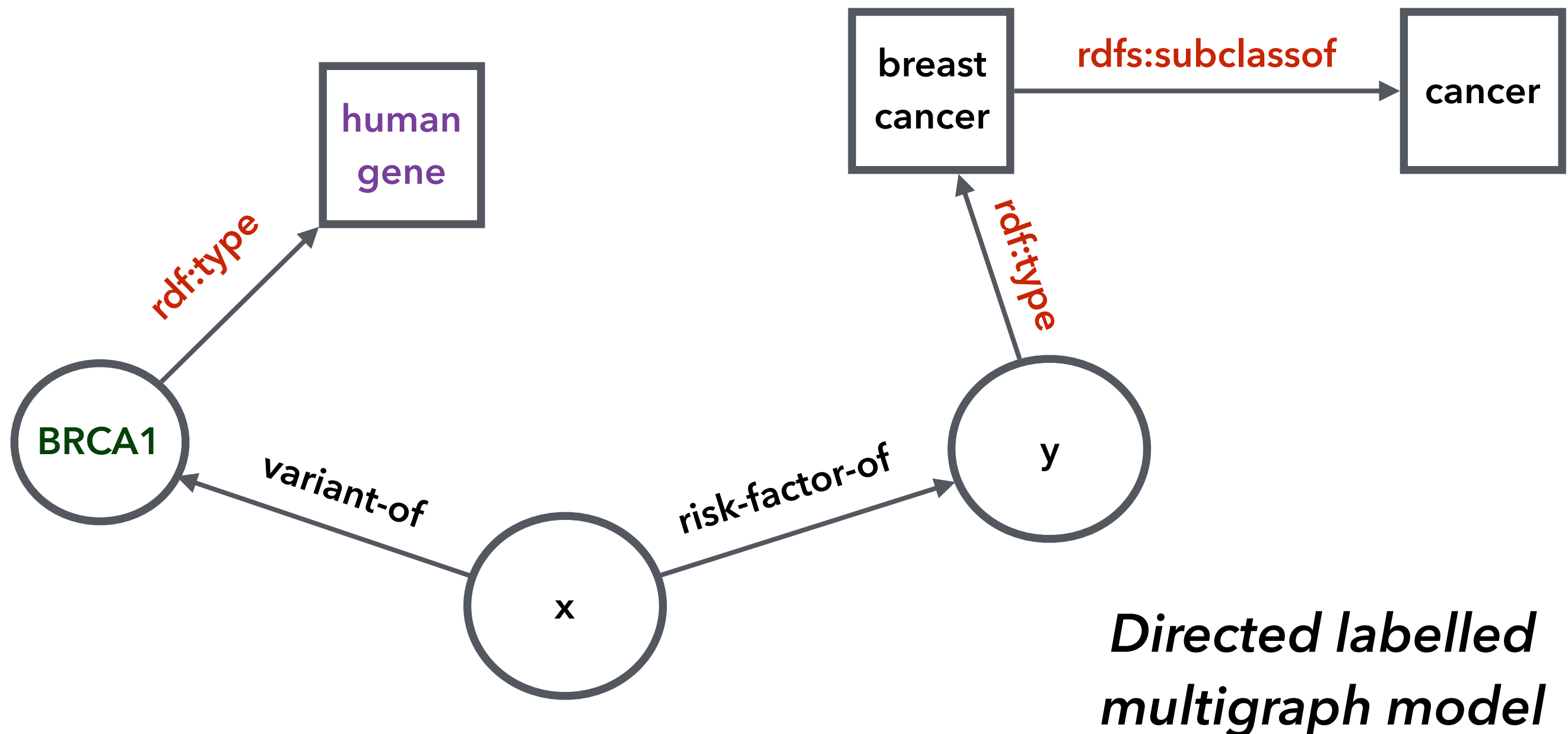
subject – *predicate* – *object*

Triple model

RDF to link data and represent knowledge

"BRCA1 – is a – human gene"
subject – predicate – object

Triple model



SPARQL to query and reason on linked data

SPARQL to query and reason on linked data

Triple Patterns & Basic Graph Patterns

```
BGP {  
  TP1  ?pathology rdf:type  cancer .  
  TP2  ?r    risk-factor-of ?pathology .  
  TP3  ?v    variant-of  ?r .  
}
```

SPARQL to query and reason on linked data

Triple Patterns & Basic Graph Patterns

```
BGP {  
  TP1 ?pathology rdf:type cancer .  
  TP2 ?r risk-factor-of ?pathology .  
  TP3 ?v variant-of ?r .  
}
```

SPARQL "SELECT" Query

```
SELECT ?v WHERE {  
  ?pathology rdf:type cancer .  
  ?r risk-factor-of ?pathology .  
  ?v variant-of ?r .  
}
```

SPARQL to query and reason on linked data

Triple Patterns & Basic Graph Patterns

```
BGP {  
  TP1  ?pathology rdf:type cancer .  
  TP2  ?r risk-factor-of ?pathology .  
  TP3  ?v variant-of ?r .  
}
```

SPARQL "SELECT" Query

```
SELECT ?v WHERE {  
  ?pathology rdf:type cancer .  
  ?r risk-factor-of ?pathology .  
  ?v variant-of ?r .  
}
```

Several query kinds : SELECT, DESCRIBE, ASK, CONSTRUCT

Several features : union, optional, aggregates, service, etc.

Semantic Web is working !

About: [BRCA1](#) [Goto](#) [Sponge](#) [NotDistinct](#) [Permalink](#)

An Entity of Type : [dbpedia-owl:Protein](#), within Data Space : [dbpedia.org](#) associated with source [document\(s\)](#)

Type: Command:

BRCA1 and BRCA1 (/brækə'wʌn/) are a human gene and its protein product, respectively. The official symbol (BRCA1, italic for the gene, nonitalic for the protein) and the official name (breast cancer 1, early onset) are maintained by the HGNC. Orthologs, styled Brca1 and Brca1, are common in other mammal species.

Attributes

[type](#)

Values

[Thing](#)
[dul:BiologicalObject](#)
[Biomolecule](#)
[protéine](#)

[sameAs](#)

[fbase:m.0256rl](#)
[http://fr.dbpedia.org/resource/BRCA1](#)
[http://de.dbpedia.org/resource/BRCA1](#)
[BRCA1](#)
[http://es.dbpedia.org/resource/BRCA1](#)
[http://it.dbpedia.org/resource/BRCA1](#)
[http://ja.dbpedia.org/resource/BRCA1](#)
[http://nl.dbpedia.org/resource/BRCA1](#)
[http://pl.dbpedia.org/resource/BRCA1](#)
[http://pt.dbpedia.org/resource/BRCA1](#)
[http://wikidata.org/entity/Q227339](#)
[http://ca.dbpedia.org/resource/BRCA1](#)
[http://fi.dbpedia.org/resource/BRCA1](#)
[http://sl.dbpedia.org/resource/BRCA1](#)
[http://ta.dbpedia.org/resource/பிரகாந்தி](#)
[http://ur.dbpedia.org/resource/برسا](#)

[label](#)

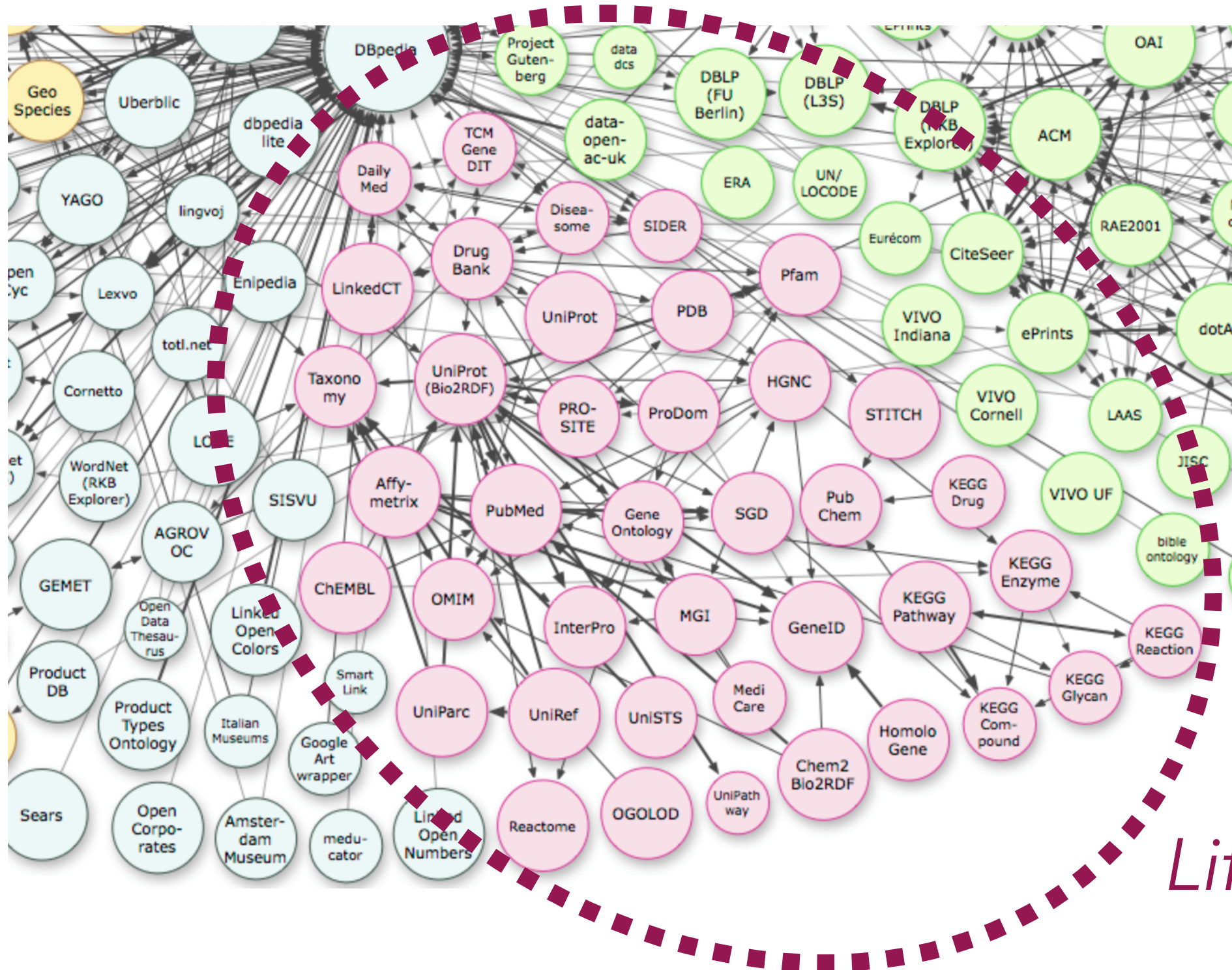
[BRCA1](#)

[Subject](#)

[Tumor markers](#)
[Tumor suppressor genes](#)
[Genes on chromosome 17](#)
[Breast cancer](#)

Linked Data Cloud

inter-connected data silos : better interpreting, explaining, reusing



**31 billions
RDF triples
(sept. 2011)**

Life Sciences

Linked Vocabularies, Ontologies

<http://lov.okfn.org>



Category Tag Cloud

- Methods
- Metadata
- Catalogs
- Geography
- API
- Quality
- Society
- RDF
- Support
- People
- Vocabularies
- General & Upper
- Time
- Events
- Multimedia
- Geometry
- Biology
- SPAR
- Academy
- Environment
- Industry
- PLM
- FRBR
- Government
- W3C Rec
- Services
- Tag
- Security
- eBusiness
- Press

The image shows the logos for three major search engines: Bing (blue), Yahoo! (purple), and Google (multi-colored). They are arranged diagonally from top-left to bottom-right.

<https://schema.org>

<http://biportal.bioontology.org>

(472 LS Ontologies + mappings)


[Browse](#)
[Search](#)
[Mappings](#)
[Recommender](#)
[Annotator](#)
[Resource Index](#)
[Projects](#)

[Recently Viewed](#)
[Sign In](#)
[Help](#)
[Feedback](#)

[Follow @bioontology](#)
1,792 followers

[Cite Us](#)

Welcome to BioPortal, the world's most comprehensive repository of biomedical ontologies.

Current Release: [4.22 \(November 2015\)](#)

For help using BioPortal, click on this icon: 

Search all ontologies

[Advanced Search](#)

Find an ontology

[Browse Ontologies >](#)

Search resources

[Advanced Resource Search](#)

Ontology Visits (October 2015)

Current Procedural Terminology (CPT)	30710
RxNORM (RXNORM)	17578
Systematized Nomenclature of Medicine – Clinical Terms (SNOMEDCT)	14871
Medical Dictionary for Regulatory Activities (MEDDRA)	13611
National Drug Data File (NDDF)	6447
More	

Latest Notes

[Grade definition \(Common Terminology Criteria for Adverse Events\)](#)
about 1 month ago by hpotier
Hi, We would like for users to be able navigate down to the complete definition for each grade i...

[Grade definition \(Common Terminology Criteria for Adverse Events\)](#)
about 1 month ago by hpotier
Hi, We would like for users to be able navigate down to the complete definition for each grade i...

[maklon kosmetik \(Radiology Lexicon\)](#)
2 months ago by radiasioim
http://www.smart-edu.web.id/2015/06/maklon-kosmetik.html

[maklon kosmetik \(Radiology Lexicon\)](#)
2 months ago by radiasioim
http://www.smart-edu.web.id/2015/06/maklon-kosmetik.html

[Change Property Value Proposal: Synonym proposed for use for "viral hemagglutination inhibition assay" OBI:0000875 \(Ontology for Biomedical Investigations\)](#)
6 months ago by sbour

Latest Mappings

[Health Entities \(HEIO\) <=> Site of care \(SNOMEDCT\)](#)
REST Mapping 04/09/2015 by pigjr

[Site of care \(SNOMEDCT\) <=> Health Entities \(HEIO\)](#)
REST Mapping 04/09/2015 by pigjr

[Emergency Medical Services Entities \(HEIO\) <=> Land ambulance \(SNOMEDCT\)](#)
REST Mapping 04/09/2015 by pigjr

[Land ambulance \(SNOMEDCT\) <=> Emergency Medical Services Entities \(HEIO\)](#)
REST Mapping 04/09/2015 by pigjr

[Primary Care Entities \(HEIO\) <=> Primary care clinic \(SNOMEDCT\)](#)
REST Mapping 04/09/2015 by pigjr

Statistics

Ontologies	472
Classes	6,449,217
Resources Indexed	48
Indexed Records	39,359,542
Direct Annotations	95,468,433,792
Direct Plus Expanded Annotations	144,789,582,932



THE NATIONAL CENTER FOR
BIOMEDICAL ONTOLOGY

The National Center for Biomedical Ontology is one of the National Centers for Biomedical Computing supported by the NHGRI, the NHLBI, and the NIH Common Fund under grant U54-HG004028. Copyright © 2005-2015, The Board of Trustees of Leland Stanford Junior University. All rights reserved.

[NCBO Website](#)
[Release Notes](#)
[Terms of Use](#)
[Privacy Policy](#)
[How to Cite](#)

SyMeTRIC

a Systems Medicine
e-Science platform (VRE)

Connect Talent - Région Pays de la Loire
(2015-2017, 4 eng. + M2 internships)

"Systems Medicine"

Développer des **modèles prédictifs**,

| **intégrer** différentes sources d'information pour construire et valider des **marqueurs** et **modèles** bio-pathologiques.

pour une **médecine personnalisée**.

| identifier le bon traitement pour le bon groupe de patient, au bon moment, et à la bonne dose.

Acquisition, analyse, intégration de données massives, hétérogènes, cloisonnées.

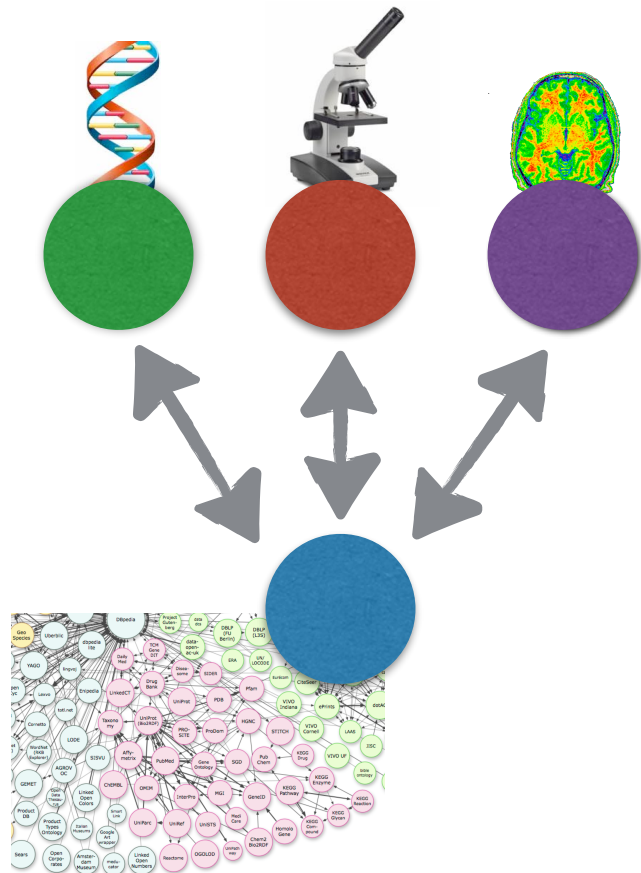
Demain, une plateforme "données et modèles bio-médicaux"

- Incitation forte (Europe, gouvernements, société civile) à **publier les données de la recherche** (H2020 data management plans, agence européenne du médicament EMA)
- Contexte "Hôpital Numérique" ; "Conseil National du Numérique" ; projet de loi de santé (SNDS)
- "Open Data" = données brutes
≠ **données inter-opérables**
≠ **données intelligibles**

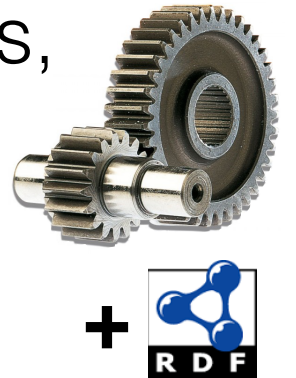
Metadonnées () en Médecine Systémique

pour améliorer la confiance et la reproductibilité

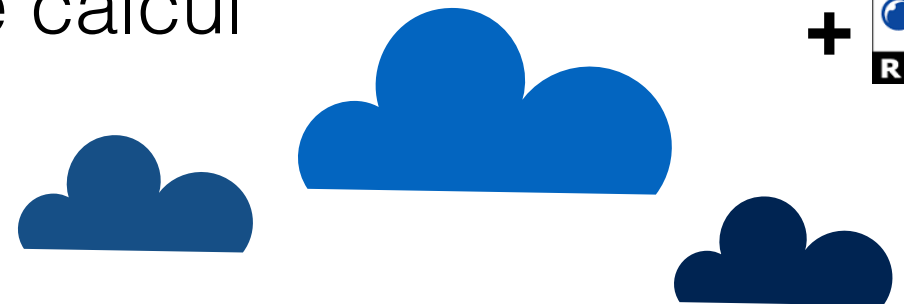
1. Intégration de données (locales, ouvertes)



2. Analyse de données, modélisation, validation *in-silico*.



3. Infrastructures distribuées de calcul



↓

★ **Validation
datasets**

↓

★ **Scientific
publication**



↓

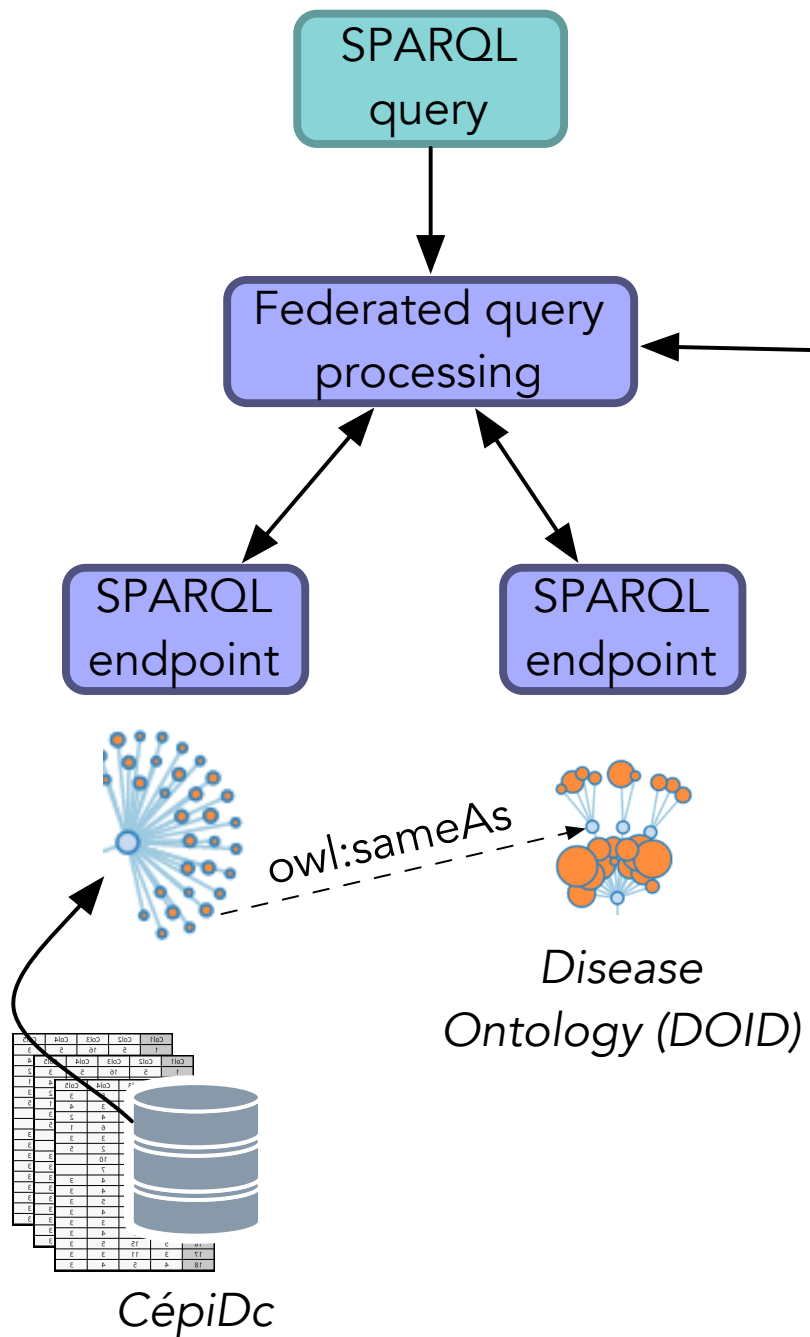
★ **Public-Private
partnerships**

↓

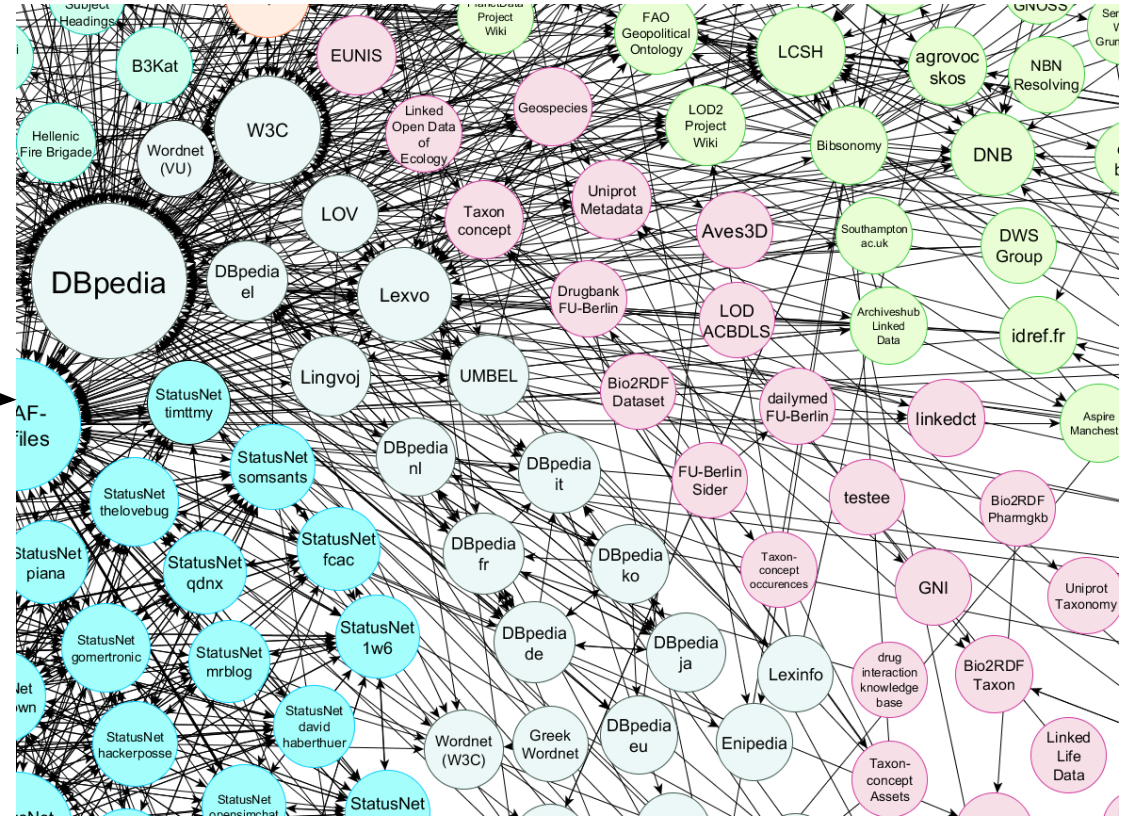
★ **Predictive &
Preventive medicine**

Démonstrateur Web Sémantique

Questions involving combining and filtering distributed and heterogeneous datasets



Linked Open Data



Public data sources (31 billion graph edges) :

*DBPedia, LinkedCT, DrugBank, Disasome,
UniProt, Bio2RDF, LinkedTCGA, OMIM, etc.*



CépiDc is an epidemiology center providing statistical data on medical death causes in France, since 1968
32k+ RDF triples



DOID is a comprehensive hierarchical controlled vocabulary for human disease representation
9k+ OWL classes, 200k+ RDF triples.



OMIM contains information on all known mendelian disorders and over 12,000 genes. OMIM focuses on the relationship between phenotype and genotype
1.3 Billion RDF triples.

Select year

2011

Phenotype

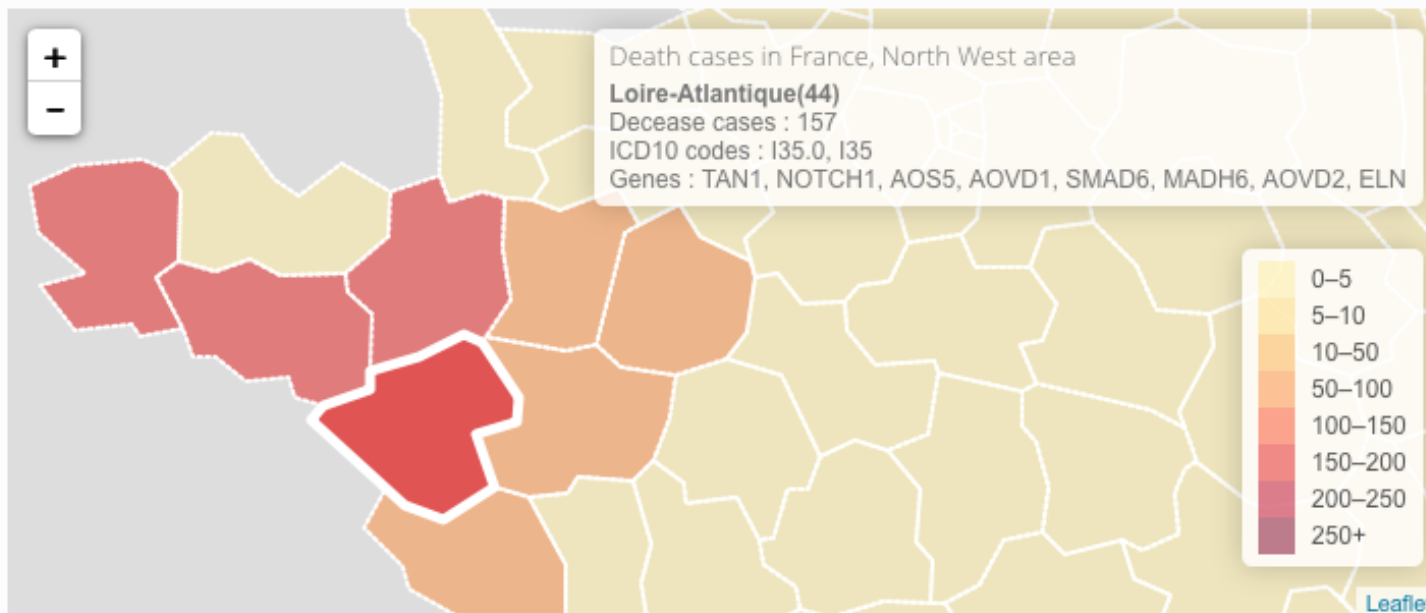
stenosis

☒ Results as table & map

☐ Results as graph

SPARQL query (editable)

Query



?area	?cases	?population	?ICD	?labels	?omimRefs	?genelds
29	88, 111	886500	I35.0, I35	aortic valve stenosis, supravulvular aortic stenosis	OMIM:109730, OMIM:614823, OMIM:185500	TAN1, NOTCH1, AOS5, AOVD1, SMAD6, MADH6, AOVD2, ELN

Démonstrateur Web :

- **génération** de requêtes SPARQL
- interrogation de **sources locales** et de **sources distantes**
- **visualisation** (carte, graphes)
- **authentification** (sécurité)

PREFIX obolnOwl: <http://www.geneontology.org/formats/obolnOwl#>
 PREFIX om: <http://bio2rdf.org/omim_vocabulary:>
 PREFIX b: <http://bio2rdf.org/bio2rdf_vocabulary:>
 PREFIX dbp: <http://dbpedia.org/property/>
 PREFIX dbo: <http://dbpedia.org/ontology/>
 PREFIX dcterms: <http://purl.org/dc/terms/>

SELECT DISTINCT

(?reg as ?area)
 (group_concat(distinct ?nb; separator=", ") AS ?cases)
 (?pop as ?population)
 (group_concat(distinct ?icd10; separator=", ") AS ?ICD)
 (group_concat(distinct ?label; separator=", ") AS ?labels)
 (group_concat(distinct ?ncit; separator=", ") AS ?omimRefs)
 (group_concat(distinct ?gs; separator=", ") AS ?genelds) WHERE {

SERVICE <http://live.dbpedia.org/sparql> {
 ?dpt dbp:subdivisionName <http://dbpedia.org/resource/France> .
 ?dpt dbp:blankInfoSec ?dptCode .
 ?dpt dbp:populationTotal ?pop .
 }

DBpedia

?doidClass rdfs:label ?label .
 FILTER (regex(?label, "stenosis")) .

 ?doidSubClass rdfs:subClassOf* ?doidClass .
 ?doidClass rdfs:subClassOf* ?doidSuperClass .

 ?doidSubClass obolnOwl:hasDbXref ?ncit .
 FILTER strstarts(str(?ncit), "OMI") .

DOID

{ {?t owl:sameAs ?doidClass} UNION {?t owl:sameAs ?doidSubClass} } UNION {?t owl:sameAs ?doidSuperClass}
 ?x rdf:type ?t .
 ?x <http://cepidc.data.symetric.org/has-icd10-cause> ?icd10 .
 ?x <http://cepidc.data.symetric.org/refers-to-gender> "T" .
 ?x <http://cepidc.data.symetric.org/refers-to-period> "2011" .
 ?x <http://cepidc.data.symetric.org/has-value-for-all-ages> ?nb .
 FILTER (?nb > 0) .
 ?x <http://cepidc.data.symetric.org/refers-to-geographical-area> ?reg .

FILTER (regex(xsd:string(?dptCode), ?reg)) .

CepiDC

BIND(URI(LCASE(CONCAT("http://bio2rdf.org/",?ncit))) as ?s) .
 SERVICE <http://omim.bio2rdf.org/sparql> {
 ?s om:phenotype-map ?map .
 ?map om:gene-symbol/b:identifier ?gs .
 }

OMIM

}

GROUP BY ?reg

ORDER BY ?reg

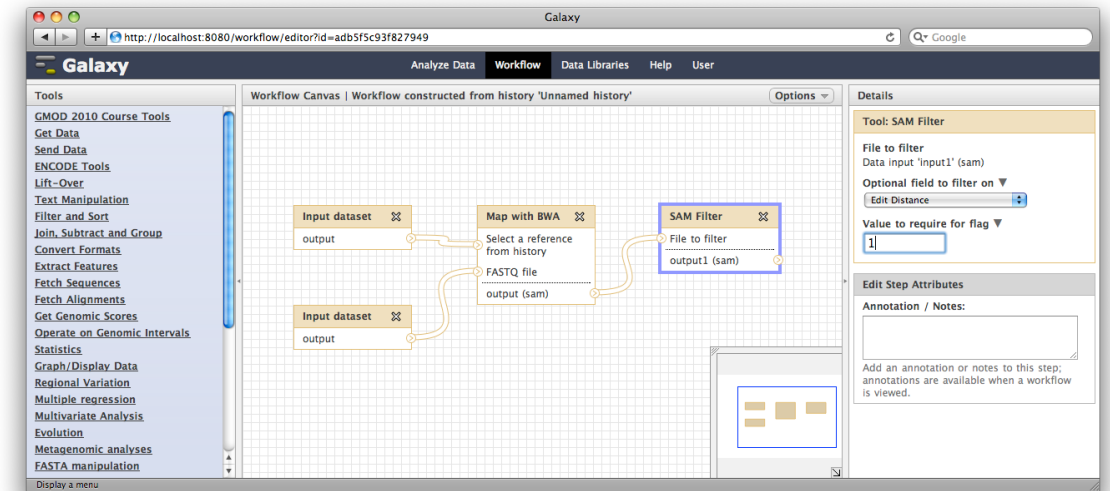
Démonstrateur Web :

- **génération** de requêtes SPARQL
- interrogation de **sources locales** et de **sources distantes**
- **visualisation** (carte, graphes)
- **authentification** (sécurité)

Démonstrateur workflows RNA-seq

Déployer un portail Web : publication et partage de **méthodes** d'analyse RNA-seq

- ➔ **abstraire hardware + software !!**
- ➔ **reproductibilité !!**



- Myélome RNA-seq Chip-seq : TopHat, CuffLinks, MACS
- Onco-hématologie (ERRATA) RNA-seq : TopHat/Star, HTseqCount, DeSeq2
- Leucémie : BWA/TopHat, CuffLinks

A venir : projet collaboratif IFB (BiRD + GenOuest) sur l'annotation automatique de données.

Perspectives

Summary :

Link your data ! (use W3C standards / vocabularies / ontologies for sharing, reusing, repurposing, preserving, etc.)

Workflows environments to analyze raw data & **produce Linked Data !**

Future works :

- **BioDataRing** : multi-site replication for highly available and long-term preserved datasets
- **Provenance** to bridge **data querying** and **data processing** (IFB project)
- **Federated** / standard **SPARQL** query identification from shared logs (M2 internship)
- **VM placement** in **multi-cloud** infrastructure (M2 internship)