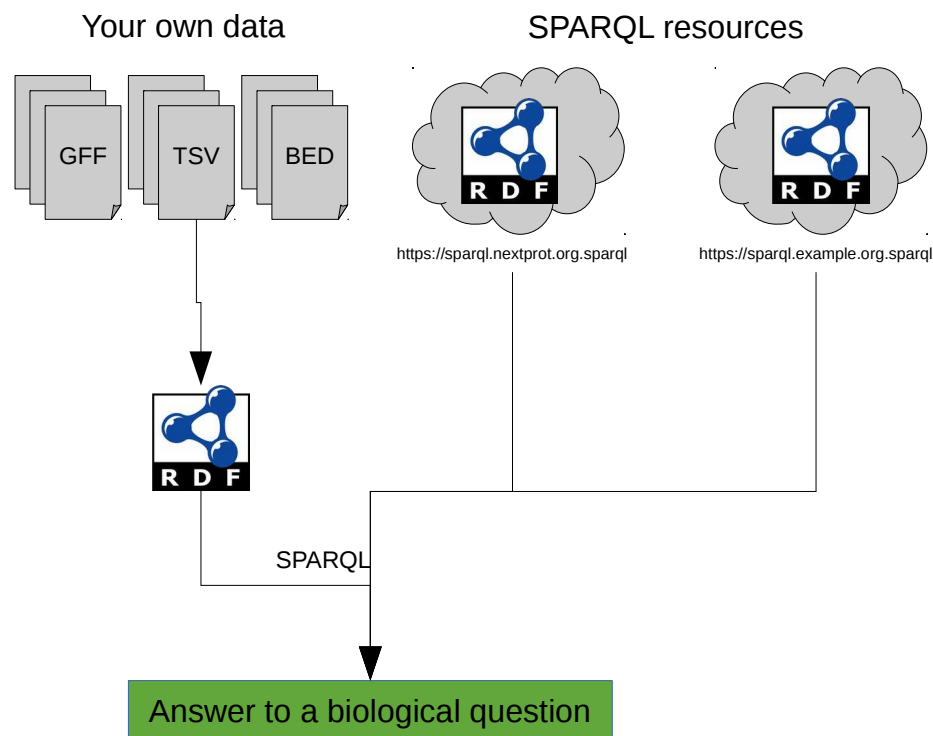


# Facilitating the connection between local datasets and neXtProt with Semantic Web technologies and AskOmicS

Xavier Garnier<sup>1</sup>, Anthony Bretaudeau<sup>1,2</sup>, Alain Gateau<sup>3</sup>, Lydie Lane<sup>3</sup>, Fabrice Legeai<sup>1,2</sup>, Pierre-André Michel<sup>3</sup>, Anne Siegel<sup>1</sup> and Olivier Dameron<sup>1</sup>

<sup>1</sup>Univ Rennes, CNRS, Inria, IRISA - UMR 6074, <sup>2</sup>IGEPP INRAE Institut Agro - Univ Rennes, <sup>3</sup>CALIPHO Group, SIB - Swiss Institute of Bioinformatics

- Study of complex biological mechanisms
  - Combine multiple data formats
  - Query unified data
- Linked open data (LOD)
  - Semantic web formats (RDF/SPARQL)
  - Biological databases (neXtProt) accessible via SPARQL endpoints
- AskOmicS<sup>1</sup>
  - Integrates multiple data formats into RDF
  - Performs federated queries over multiple endpoints

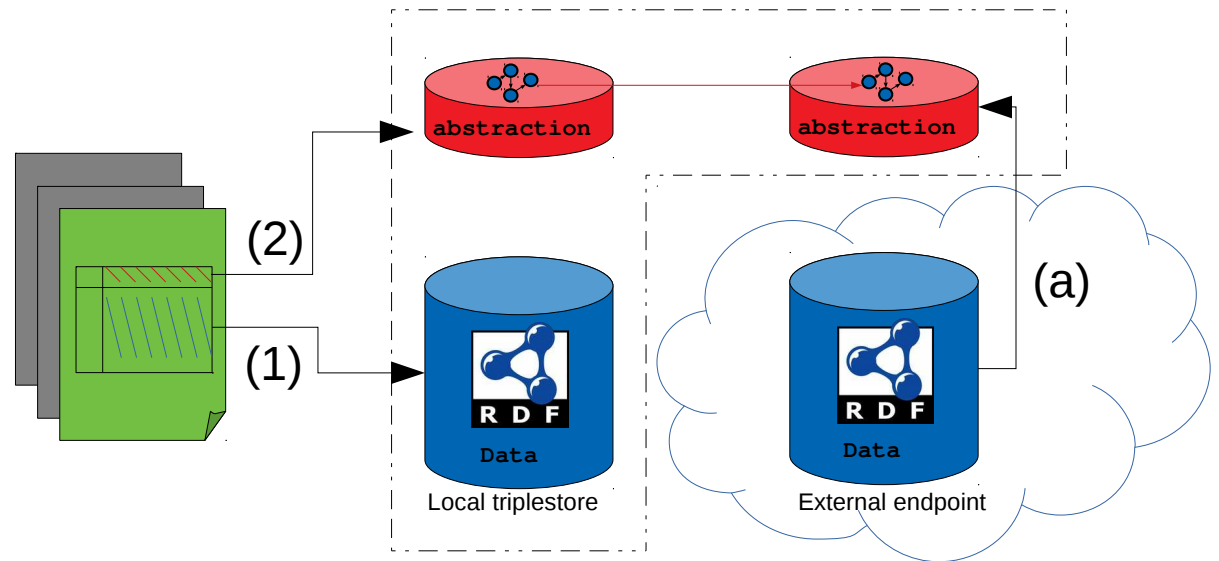


<sup>1</sup> <https://github.com/askomics/flaskomics>

# Integrate easily local data and external resources

- From input files (TSV, GFF, BED), **AskOmics** :
  - (1) Generates **RDF data**
  - (2) Creates a representation of the structure of the data: the **RDF abstraction**, based on the file header
- From external resources (already in **RDF format**) **abstractor**<sup>1</sup>:
  - (a) Generates an **RDF abstraction** for each external resource

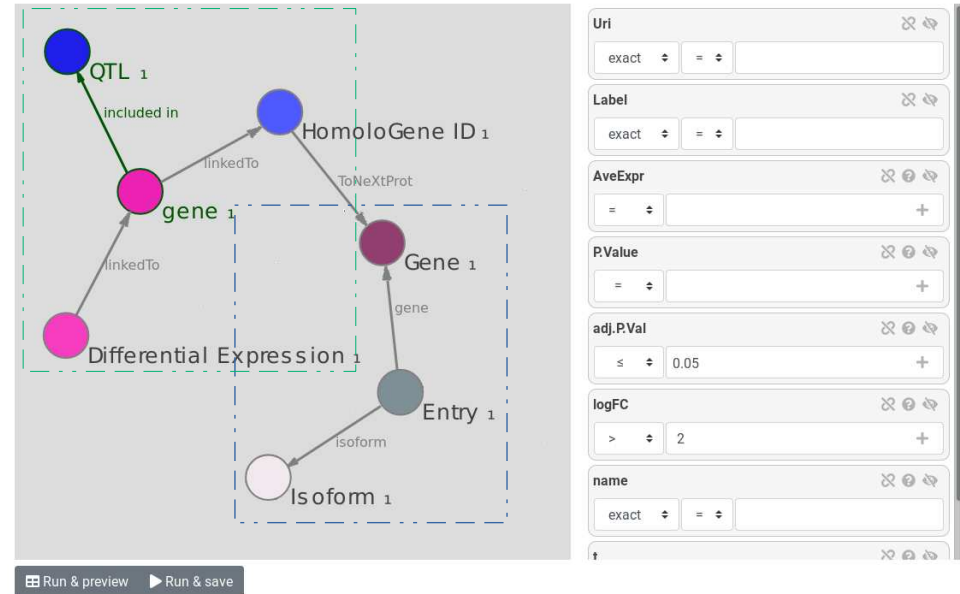
Only the **local data**, **local abstraction** and **external abstractions** are stored on the embedded triplestore of AskOmics



<sup>1</sup> <https://github.com/askomics/abstractor>

# Query easily your own data and external resources

- (1) Traversal of the **abstractions** is used to build a **query** that covers **local** and **distant** endpoints
- (2) **AskOmics** converts the **query** into SPARQL code
- (3) A **federated query engine** (Corese<sup>1</sup>) splits the SPARQL query and dispatches it to the endpoints
- (4) Results are displayed and downloadable

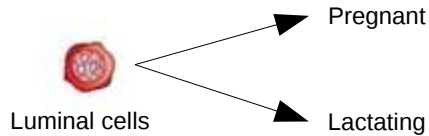


gene1_Label	QTL1_Label	QTL1_Name	Uniprot_Subcellular_Location_Cv1_Label
ENSMUSG00000008136	W10q6	weight 10 weeks QTL 6	Cell membrane
ENSMUSG000000025969	W10q6	weight 10 weeks QTL 6	Cell membrane
ENSMUSG00000008136	W10q7	weight 10 weeks QTL 7	Cell membrane
ENSMUSG000000025969	W10q7	weight 10 weeks QTL 7	Cell membrane
ENSMUSG000000026271	W10q7	weight 10 weeks QTL 7	Cell membrane
ENSMUSG000000049608	W10q7	weight 10 weeks QTL 7	Cell membrane

<sup>1</sup> <https://github.com/Wimmics/corese>

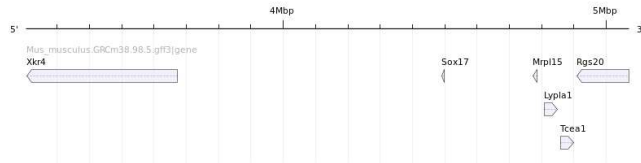
# neXtProt use case

## RNA-Seq analysis of Mouse mammary gland<sup>1</sup> (TSV)



ENTREZID	SYMBOL	GENENAME	logFC	adj.P.Val
12992	Csn1s2b	casein alpha s2-like B	-8.603611114762	6.05395889659601e-11
13358	Slc25a1	solute carrier family 25	-4.12417532129173	1.38964155864574e-09
11941	Atp2b2	ATPase, Ca++ transporting	-7.38698638678659	2.43279979019347e-09
20531	Slc34a2	solute carrier family 34	-4.17781242057656	2.43279979019347e-09
100705	Acacb	acetyl-Coenzyme A	-4.3143199499725	4.74112875360987e-09

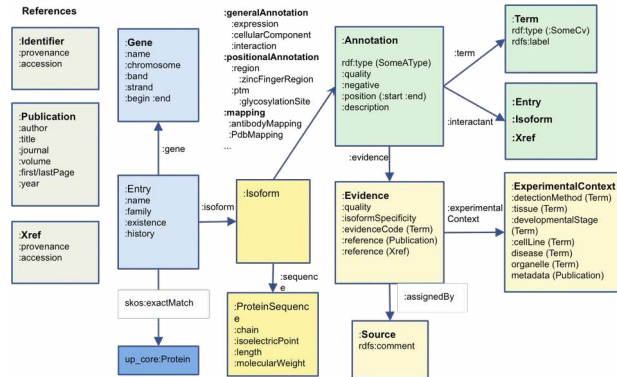
## Mus musculus annotation<sup>2</sup> (GFF)



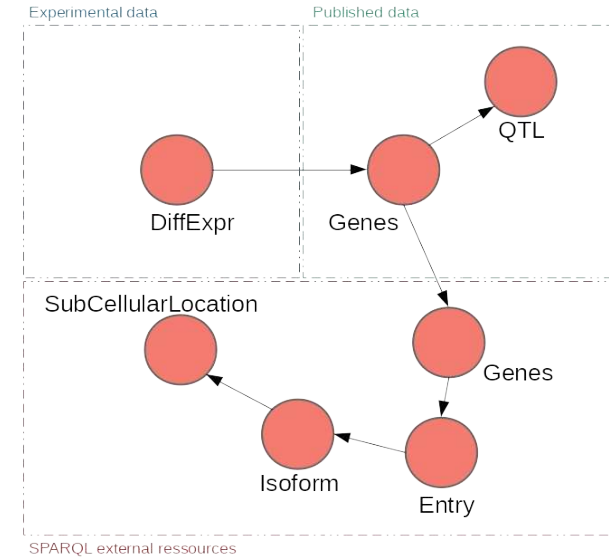
## Mouse QTL<sup>3</sup> (TSV)

Input	Name	Chr	Start	End
Hbtq	habituation QTL	15	68288859	68288984
Adq1	aortic diameter QTL 1	9	32838331	32838331
Adq2	aortic diameter QTL 2	9	32838331	32838331
Ahrq1	airway hyperresponsiveness QTL 1	12	54649125	82619165

## NeXtProt SPARQL endpoint<sup>4</sup>



Try it at [jobim2020.askomics.org](http://jobim2020.askomics.org)!



## Homology groups<sup>3</sup> (TSV)

HomoloGene ID	Common Organism Name	Symbol
3	mouse_laboratory	Acadm
3	human	ACADM
5	mouse_laboratory	Acadvl
5	human	ACADVL

- Which genes are over-expressed in the pregnant mouse compared to the lactating mouse ?
- Are these genes associated with a known phenotype (included in a QTL)?
- Do these genes have human homologs ? Where the proteins coded by these homologs are located?

# Use with your own data

---

## Use our dedicated AskOmics instance to query neXtProt with local data

- Visit <https://nextprot.askomics.org>
- Create a free account
- Add your own data and compare them with neXtProt

## Usefull links

Website: [askomics.org](https://askomics.org)

Documentation: [flaskomics.readthedocs.io](https://flaskomics.readthedocs.io)

Github: [github.com/askomics](https://github.com/askomics)

Contact: [askomics@inria.fr](mailto:askomics@inria.fr)

## Install your own instance

- Easy deploy AskOmics with our docker-compose files<sup>1</sup>
- Use **abstractor** to build external endpoint **abstraction**
- Integrate your data and build complex queries over mutiple endpoints



Swiss Institute of  
Bioinformatics

---

<sup>1</sup> <https://github.com/askomics/flaskomics-docker-compose>