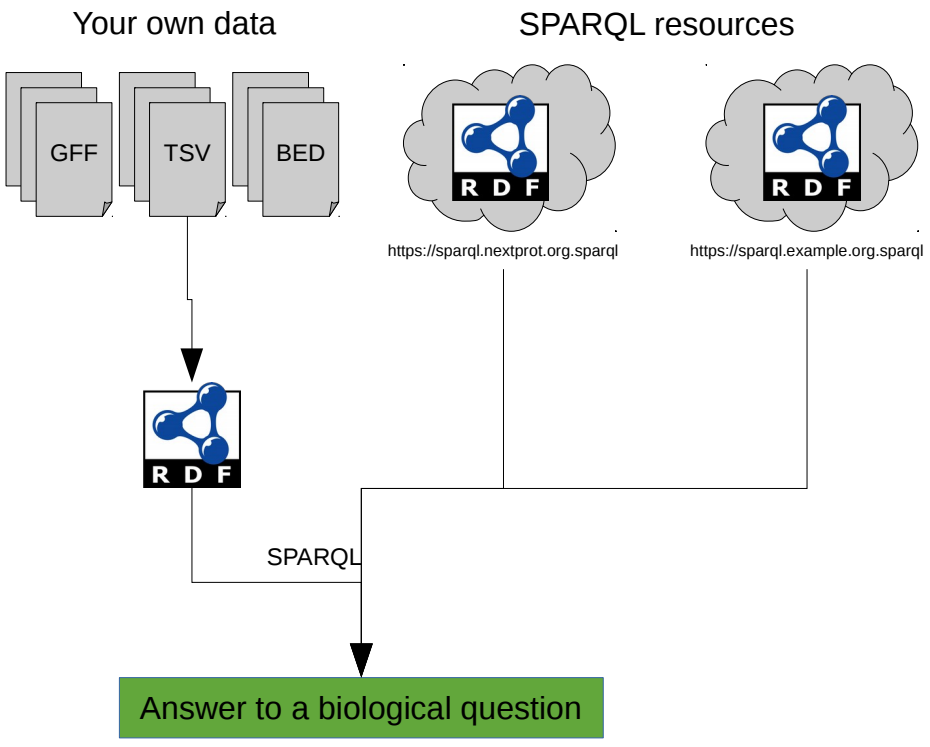
Facilitating the connection between local datasets and neXtProt with Semantic Web technologies and AskOmics

Xavier Garnier[1], Anthony Bretaudeau[1,2], Alain Gateau[3], Lydie Lane[3], Fabrice Legeai[1,2], Pierre-André Michel[3], Anne Siegel[1] and Olivier Dameron[1]

[1] Univ Rennes, CNRS, Inria, IRISA - UMR 6074, [2] IGEPP INRAE Institut Agro - Univ Rennes, [3] 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[1]
  - ➔ Integrates multiple data formats into RDF
  - ➔ Performs federated queries over multiple endpoints

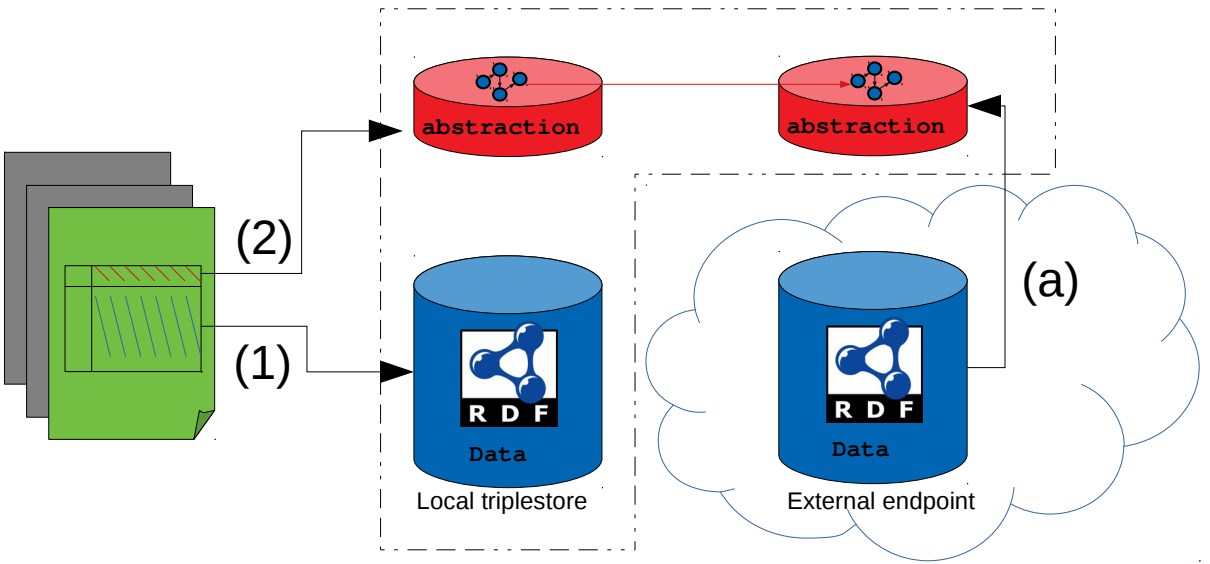


[1] https://github.com/askomics/flaskomics

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**[1] :
  (a) Generates an **RDF abstraction** for each external resource

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



[1] https://github.com/askomics/abstractor

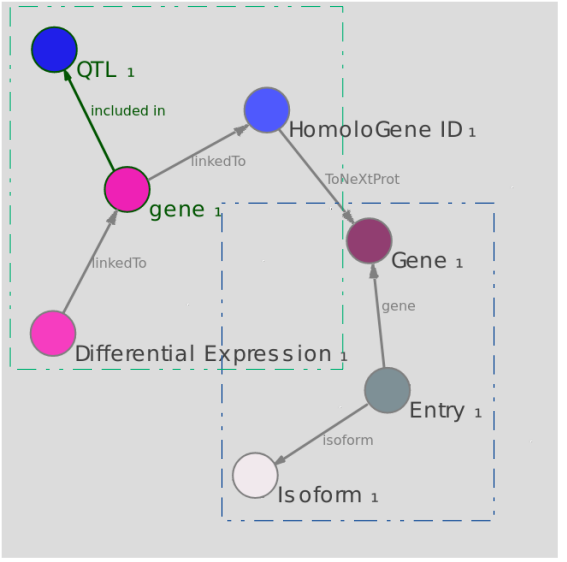
Query easily your own data and external resources

(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[1]) splits the SPARQL query and dispatches it to the endpoints

(4) Results are displayed and downloadable

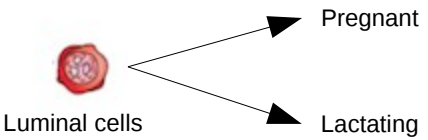


gene1_Label	QTL1_Label	QTL1_Name	Uniprot_Subcellular_Location_Cv1_Label
ENSMUSG00000008136	W10q6	weight 10 weeks QTL 6	Cell membrane
ENSMUSG00000025969	W10q6	weight 10 weeks QTL 6	Cell membrane
ENSMUSG00000008136	W10q7	weight 10 weeks QTL 7	Cell membrane
ENSMUSG00000025969	W10q7	weight 10 weeks QTL 7	Cell membrane
ENSMUSG00000026271	W10q7	weight 10 weeks QTL 7	Cell membrane
ENSMUSG00000049608	W10q7	weight 10 weeks QTL 7	Cell membrane

[1] https://github.com/Wimmics/corese

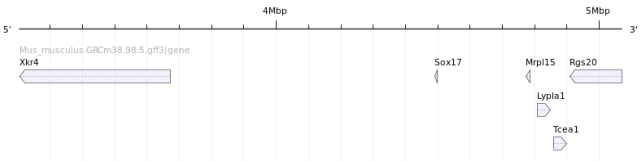
neXtProt use case

RNA-Seq analysis of Mouse mammary gland[1] (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[2] (GFF)



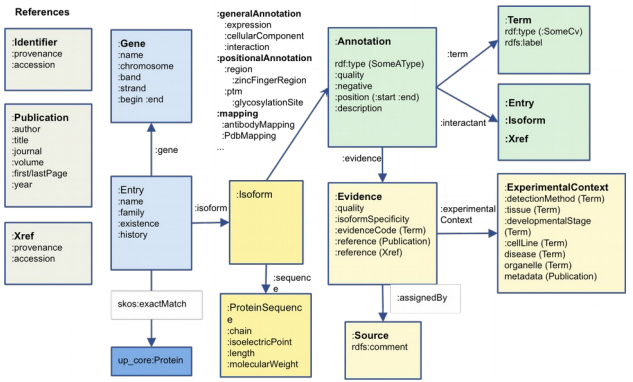
Mouse QTL[3] (TSV)

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

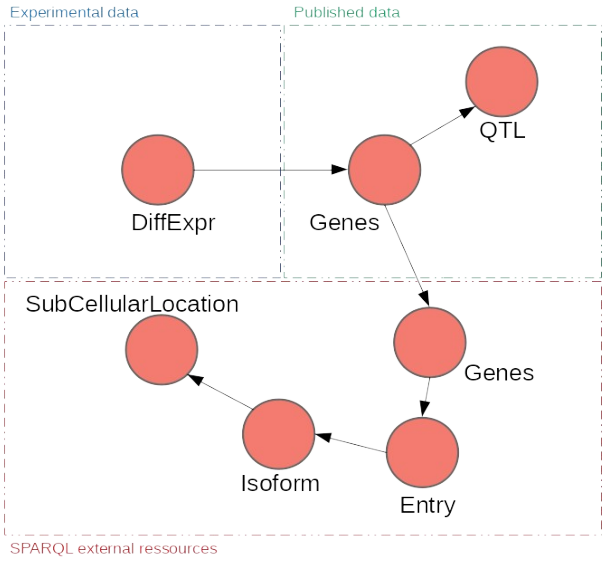
Homology goups[4] (TSV)

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

NeXtProt SPARQL endpoint



Try it at nextprot.askomics.org!



- 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

Install your own instance

- Easy deploy AskOmics with our docker-compose files[1]
- Use **`abstractor`** to build external endpoint **abstraction**
- Integrate your data and build complex queries over mutliple endpoints

Usefull links

Website: askomics.org
Documentation: flaskomics.readthedocs.io
Github: github.com/askomics
Contact: askomics@inria.fr

[1] https://github.com/askomics/flaskomics-docker-compose

Swiss Institute of Bioinformatics