



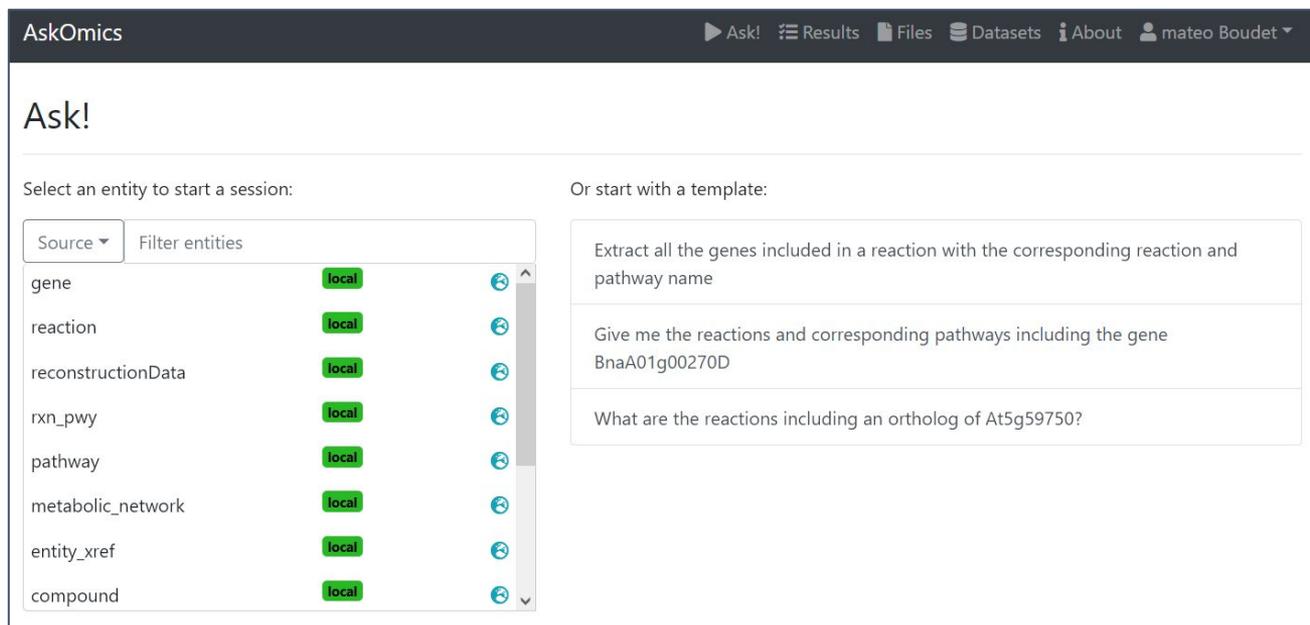
# AskOmics, a Semantic-Web application to integrate and query meaningful biological datasets

Matéo Boudet  
IGEPP/GOGEPP

# AskOmics (4.3.1)

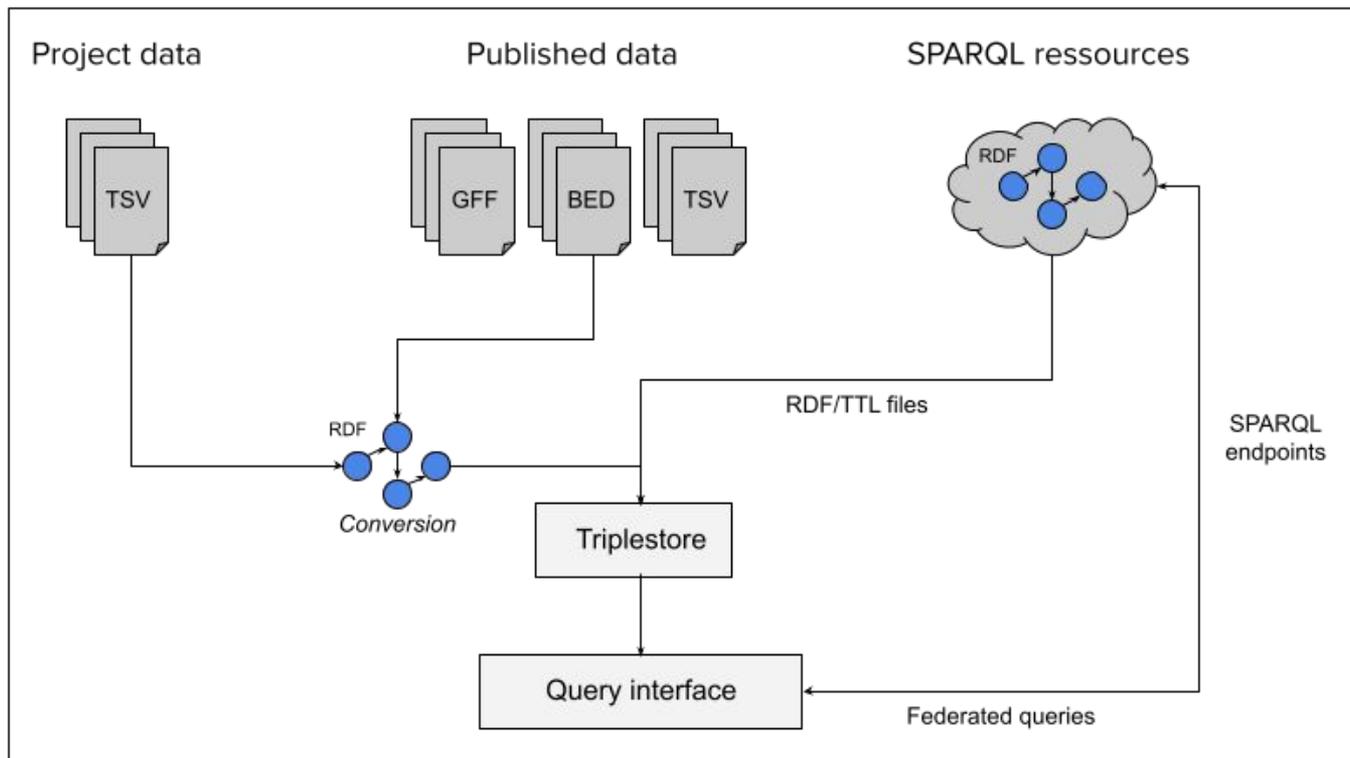
- Intégration de données hétérogènes (sans RDF)
- Interrogation de données locales et distantes (sans SPARQL)

*Les atouts du web sémantique sans les difficultés: AskOmics s'en charge!*



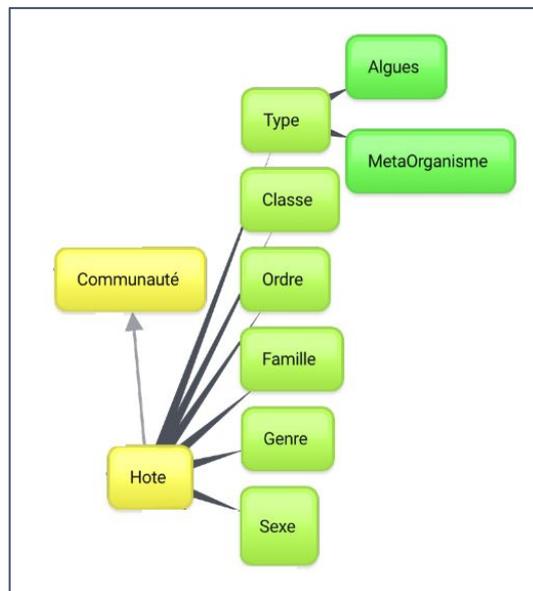
The screenshot shows the AskOmics web interface. At the top, there is a navigation bar with the text "AskOmics" on the left and "Ask!", "Results", "Files", "Datasets", "About", and "mateo Boudet" on the right. Below the navigation bar, the main content area is titled "Ask!". Underneath, there are two sections: "Select an entity to start a session:" and "Or start with a template:". The "Select an entity to start a session:" section features a "Source" dropdown menu and a "Filter entities" input field. Below this, a list of entities is displayed, each with a "local" label in a green box and a circular icon with a globe. The entities listed are: gene, reaction, reconstructionData, rxn\_pwy, pathway, metabolic\_network, entity\_xref, and compound. The "Or start with a template:" section contains three text boxes with pre-defined queries: "Extract all the genes included in a reaction with the corresponding reaction and pathway name", "Give me the reactions and corresponding pathways including the gene BnaA01g00270D", and "What are the reactions including an ortholog of At5g59750?".

# AskOmics: vue d'ensemble



# AskOmics: en interne

- Convertit des fichiers *CSV*, *BED*, *GFF* en graphe de connaissance
  - Création d'*entités* (ex: Gène, mRNA...)
  - Les entités ont des attributs (ex: sexe, famille, référence, strand)
- 1 ligne d'un fichier = 1 *instance* d'une entité
- Les entités peuvent référer d'autres entités (ex: *mRNA* -> *Gène*)



# AskOmics: intégration de données : CSV/TSV

- Intégration du fichier basée sur l'en-tête du fichier (première ligne)
- Première colonne = URI (identifiant) de l'entité.
- Autres colonnes = Attributs & relations

Types de colonnes:

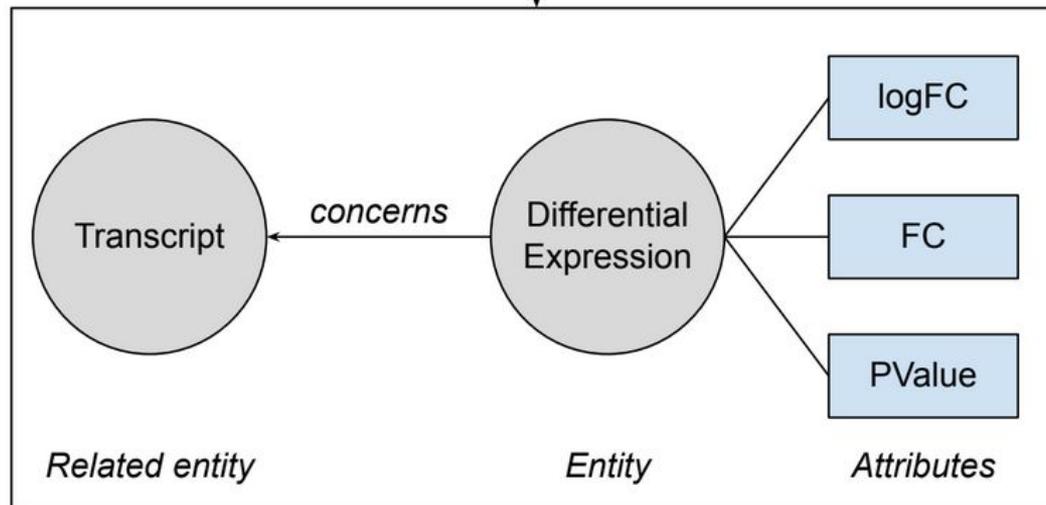
<i>Texte</i>	<i>Référence</i>
<i>Numérique</i>	<i>Strand</i>
<i>Booléen</i>	<i>Start</i>
<i>Catégorie</i>	<i>Stop</i>
<i>Date</i>	<i>Relation vers une autre entité</i>

Le type d'une colonne change l'interface de recherche

# AskOmics: intégration de données : CSV/TSV

	A	B	C	D	E
1	Differential Expression	concerns@transcript	logFC	FC	PValue
2	DE001	AT3G10490	-11.5741248327485	3049.00929422056	1.3364869977085e-09
3	DE002	AT3G22640	-11.2388325298478	2416.71650288007	8.0438336731249e-10
4	DE003	AT1G33615	8.35272150205517	326.903610768331	1.82953565833812e-09
5	DE004	AT1G57800	9.20759897504639	591.239552393754	4.99876947134223e-10
6	DE005	AT1G49500	-10.3470410152063	1302.47594721622	1.57559426781879e-09
7					

Conversion to RDF



# AskOmics: intégration de données : GFF3s et BED

Formats *plus stricts* = intégration simplifiée

gene.gff3 (preview)

- gene
- transcript
- five\_prime\_UTR

gene.bed (preview)

Entity name

gene.bed

- Identifiant
- Paramètres
  - Référence
  - Strand
  - Start
  - End
  - (+ Score (BED))
- Relations (GFF)

# AskOmics: après l'intégration

## Interface de gestion des *Datasets*

- Gestion de la visibilité (pour les admins)

Datasets						
<input type="checkbox"/>	Dataset name ↑↓	Creation date ↑↓	Exec time ↑↓	Public ↑↓	Triples ↑↓	Status ↑↓
<input type="checkbox"/>	gene.bed	Wed, 06 Oct 2021 13:56:41 GMT	<1s	<input type="checkbox"/>	220	Success

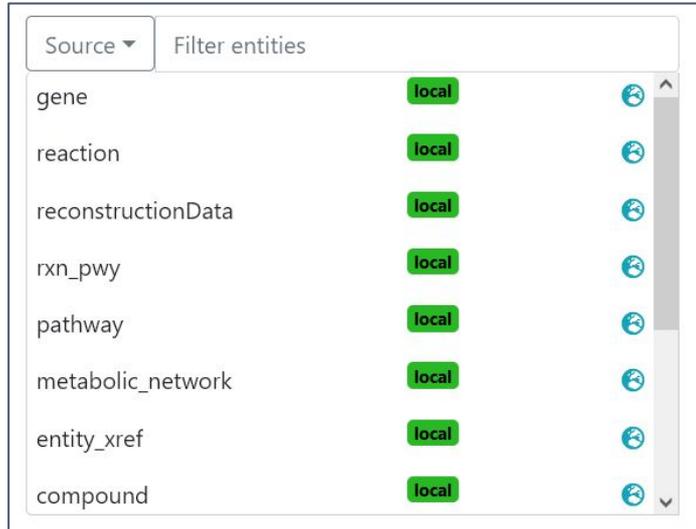
10 ▾

1

AskOmics génère l'interface de requête en fonction des datasets accessibles (personnels & publics)

# AskOmics: Interface de requêtage : Point de départ

## Choix de l'entité de départ



The screenshot shows a web interface for selecting an entity type. At the top, there is a 'Source' dropdown menu and a 'Filter entities' input field. Below this, a list of entity types is displayed, each with a green 'local' label and a globe icon. The list includes: gene, reaction, reconstructionData, rxn\_pwy, pathway, metabolic\_network, entity\_xref, and compound.

Entity Type	Label	Icon
gene	local	Globe
reaction	local	Globe
reconstructionData	local	Globe
rxn_pwy	local	Globe
pathway	local	Globe
metabolic_network	local	Globe
entity_xref	local	Globe
compound	local	Globe

Liste tous les gènes impliqués dans une réaction, avec le nom de la réaction et la voie métabolique associée

Liste les réactions et les voies métaboliques dans lesquelles le gène BnaA01g00270D est impliqué

Liste les réactions impliquant un orthologue de At5g59750

*“Donne-moi tous les entités XXXX qui valident la condition Y”*

# AskOmics: Interface d'interrogation

Entités liées à l'entité sélectionnée

Attributs de l'entité sélectionnée

The screenshot displays the AskOmics interface. On the left, a network diagram shows a central pink node labeled 'transcript\_1'. It is connected to five other nodes: 'CDS' (grey), 'three\_prime\_UTR' (yellow), 'five\_prime\_UTR' (blue), 'gene' (purple), and 'exon' (orange). All connections are labeled 'Parent'. A red box highlights this network diagram, with an arrow pointing to the text 'Entités liées à l'entité sélectionnée'. On the right, a configuration panel for the selected entity 'transcript\_1' is shown. It contains four sections: 'Uri' with a dropdown set to 'exact' and an equals sign; 'Label' with a dropdown set to 'exact' and an equals sign; 'reference' with a dropdown set to '1'; and 'strand' with a dropdown set to '+'. A red box highlights this configuration panel, with an arrow pointing to the text 'Attributs de l'entité sélectionnée'.

# AskOmics: Filtrage sur les attributs

- Filtrage sur les valeurs des attributs
  - Regex, <, >, =, etc...
  - Choix de catégories
  - Multiples filtres
- Choix des attributs à afficher dans les résultats

The image shows a vertical stack of filter panels in the AskOmics interface. Each panel has a title, a set of control icons (a crossed-out square, a question mark, and a refresh arrow), and a list of values. The 'Label' panel has a dropdown menu with 'exact' and an equals sign operator. The 'chromosome' panel shows 'Chrc05' and 'Chra08'. The 'ancestor' panel shows 'A' and 'C'. The 'organism' panel shows 'Brassica napus' (highlighted in blue) and 'Brassica rapa'. The 'end' panel has an equals sign operator and a plus sign button.

# AskOmics: Filtrage sur les relations

Requête sur l'existence d'une relation

- Je veux lister les transcripts liés à un gène

Requête sur les attributs des entités liées

- Je veux les transcripts liés à un gène dont la position START est XXX

La requête se construit de manière itérative, entité par entité

The screenshot displays the AskOmics query builder interface. On the left, a graph shows a sequence of entities connected by relationships: a green circle (pathway) is connected to a pink circle (rxn\_pwy) via 'is\_included\_in', the pink circle to a purple circle (reaction) via 'concerns', the purple circle to a cyan circle (rxn\_gene\_recData) via 'concerns', the cyan circle to a red circle (gene) via 'is\_linked\_to', and the red circle to a light pink circle (gene) via 'included in'. On the right, a filter panel is visible with the following sections:

- Uri**: exact = [ ]
- Label**: exact = BnaA01g00270D
- chromosome**: Chrc05, Chra08
- ancestor**: A, C

# AskOmics: Les résultats

- Possibilité d'avoir un aperçu des résultats (< 30 lignes)
- Sauvegarde de la requête et de ses résultats dans une page dédiée

<input type="checkbox"/>	Id ↑↓	Description ↑↓	Creation date ↑↓	Exec time ↑↓	Template ↑↓	Form ↑↓	Public ↑↓	Status ↑↓	Rows ↑↓	Size ↑↓
<input type="checkbox"/>	1	Query	Tue, 05 Oct 2021 08:50:35 GMT	<1s	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Success	8	69 B

- Téléchargement des résultats (csv)
- Possibilité de rejouer la requête (en changeant des paramètres)
- Accès au code SPARQL généré par AskOmics

# Les fonctionnalités ++ : Requêtes fédérées

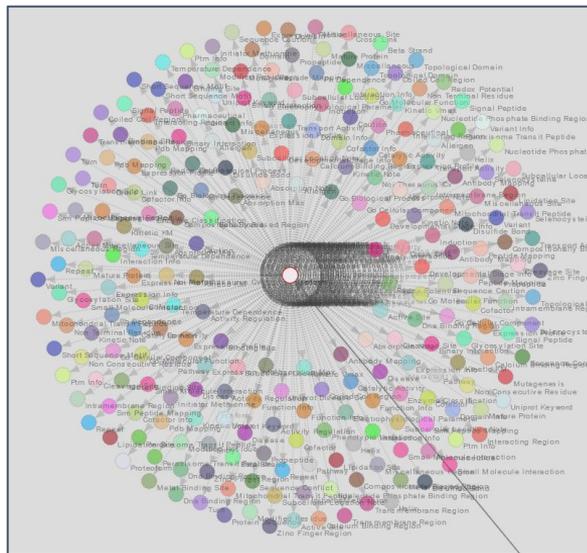
Intégration des endpoints SPARQL distants

- neXtProt
- UniProt

On ne récupère que *l'abstraction* en local (la structure du graphe)

Les requêtes sont automatiquement envoyées sur les bases distantes

-> Possibilité de lier les données locales aux données distantes



Abstraction partielle de neXtProt

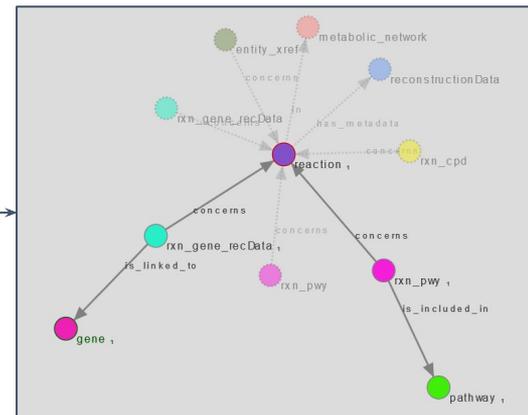
# Les fonctionnalités ++ : Partager ses requêtes

Possibilité de *partager une requête*

- Rejouer la requête et la modifier depuis le graphe

Or start with a template:

Extract all the genes included in a reaction with the corresponding reaction and pathway name



Possibilité de *partager un formulaire*

- Vue simplifiée de la requête (formulaire)
- Seul les attributs sont modifiables par l'utilisateur

Wild Population

Humidity

Average  
Wet

Color

Brown  
Red/Brown

# Les fonctionnalités ++ : Galaxy

Intégration avec Galaxy :

- Import de datasets Galaxy
- Export vers Galaxy des requêtes et résultats
- Lancement d'un AskOmics comme *Interactive Tool*

AskOmics a visual SPARQL query builder (Galaxy Version 4.3.1)

☆ Favorite ▾ Options

Datasets to load into AskOmics

- 72: nextprot\_abstraction.ttl
- 71: HOM\_MouseHumanSequence.rpt
- 70: QTL
- 69: Symbol.tsv
- 68: Mus\_musculus.GRCm38.98.subset.gff3
- 67: DE results

Email notification

No

Send an email notification when the job completes.

Execute

74: AskOmics on data 7 2, data 71, and others

This job is currently running

72: nextprot\_abstraction.ttl

71: HOM\_MouseHumanSequence.rpt

70: QTL

69: Symbol.tsv

68: Mus\_musculus.GRCm3 8.98.subset.gff3

67: DE results

# AskOmicS: Les utilisations

<https://bbip.askomics.org/>

Agrégation d'information sur les réseaux métaboliques des Brassicacées

### Ask!

Select an entity to start a session:

Source  Filter entities

gene	local	↕
metabolic_network	local	↕
reaction	local	↕
rxn_pwy	local	↕
pathway	local	↕
reconstructionData	local	↕
pathway_rate	local	↕
entity_xref	local	↕

Or start with a template:

Extract all the genes included in a reaction with the corresponding reaction and pathway name

Give me the reactions and corresponding pathways including the gene BnaA01g00270D

What are the reactions including an ortholog of At5g59750?

gene1_Label ↑↓	reaction1_Label ↑↓	pathway1_Label ↑↓	pathway1_COMMON_NAME ↑↓
BnaA01g00270D	4.4.1.14-RXN	ETHYL-PWY	ethene biosynthesis from methionine
BnaA01g00270D	4.4.1.14-RXN	ETHYL-PWY	ethylene biosynthesis I (plants)

# AskOmics: Les utilisations

<https://braskomics.askomics.org/>

Agrégat automatique (+ *validation*) des données du projet BrasExplor

- Données d'observation
- Données génomiques
- (Soon) Données phénotypiques

### Ask!

Select an entity to start a session:

Source ▾ Filter entities

Landrace Population	local	🔒
Botanical density	local	🔒
Picture	local	🔒
Sequence	local	🔒
Population	local	🔒
Wild Population	local	🔒

Or start with a simplified form:

List reads and pictures for specific population attributes

Or start with a template:

Get all pictures from NW-oriented populations with wet climate

List all reads from landrace population with seeds from a seed merchant

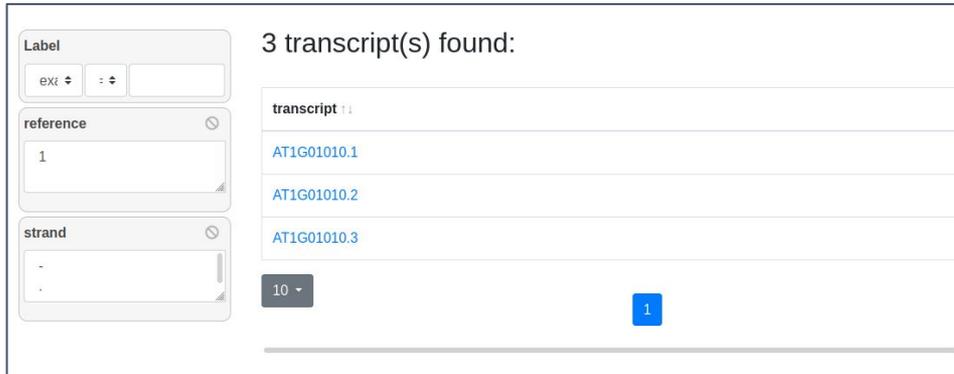
Recherche + Téléchargement des fichiers liés (Reads & Photos)

Population1_Label ↑ ↓	Sequence1_Read1 ↑ ↓	Sequence1_Read2 ↑ ↓	Picture1_File ↑ ↓
BO_F_ETRE_W_A	1fcf18ec-de4d-4942-8cd3-41be048d0e8c	fd03c977-d46b-4c6a-9a5a-8afac89a86fb	88d149d1-af29-4567-a3dc-3616480c

# AskOmics: Mode catalogue de données?

- Ajout des fichiers de métadonnées dans AskOmics
  - Métadonnées du fichier
  - Chemin interne du fichier de donnée
  - Lien direct externe vers la donnée?

Une interface de recherche simplifiée?



The screenshot shows a search interface with three filter sections on the left: 'Label' (with 'exl' and ':'), 'reference' (with '1'), and 'strand' (with '-' and '-'). The main area displays '3 transcript(s) found:' followed by a list of three transcripts: AT1G01010.1, AT1G01010.2, and AT1G01010.3. Below the list is a pagination bar with a '10' dropdown and a blue square containing the number '1'.

Information about entity AT1G01010.1

Property	Value
ID	transcript:AT1G01010.1
source	tair
biotype	protein_coding
type	transcript
Parent	AT1G01010
label	AT1G01010.1
transcript_id	AT1G01010.1
Name	ANAC001
end	5899
start	3630

# AskOmics: Les projets de GOGEPP

- **Projet BrasExplor: Intégration des données hétérogènes**
  - Données de population (observation)
  - Données génomiques
  - Données phénotypiques
- **Projet DeepImpact: Intégration des données hétérogènes**
  - Données génomiques
  - Données phénotypiques
  - Données climatiques
  - Microbiote du sol
- **Projet Metabolomic Semantic Datalake**
  - Intégration d'AskOmics à la solution



# AskOmics: Pour le moment, et a venir

- Développement actif (<https://github.com/askomics/flaskomics>)
- Documentation disponible (<https://flaskomics.readthedocs.io>)
- Déploiement All-in-one via docker-compose
- Une API pour l'ajout de fichiers + intégration

## A venir:

- Intégration d'ontologies (recherche par ancêtres/descendants)
- Gestion de requêtes modulaires (pouvoir réutiliser une partie d'une requête)
- Autres améliorations de l'UI & bugfixes



**Merci de votre attention!**



**INRAE**

12-10-21

Matéo Boudet