



HAL
open science

AskOmics: a user-friendly interface to Semantic Web technologies for integrating local datasets with reference resources

Xavier Garnier, Anthony Bretaudeau, Fabrice Legeai, Anne Siegel, Olivier Dameron

► To cite this version:

Xavier Garnier, Anthony Bretaudeau, Fabrice Legeai, Anne Siegel, Olivier Dameron. AskOmics: a user-friendly interface to Semantic Web technologies for integrating local datasets with reference resources. JOBIM 2019 - Journées Ouvertes Biologie, Informatique et Mathématiques, Jul 2019, Nantes, France. pp.1. hal-02401750

HAL Id: hal-02401750

<https://hal.science/hal-02401750v1>

Submitted on 10 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AskOmics: a user-friendly interface to Semantic Web technologies for integrating local datasets with reference resources

Xavier GARNIER¹, Anthony BRETAUDEAU^{1,2}, Fabrice LEGEAI^{1,2}, Anne SIEGEL¹ and Olivier DAMERON¹

¹ Univ Rennes, CNRS, Inria, IRISA - UMR 6074, F-35000 Rennes, France

² Institut de Génétique, Environnement et Protection des Plantes (IGEPP) - Institut national de la recherche agronomique (INRA) : UMR1349, Agrocampus Ouest - Agrocampus Ouest, UMR1349 IGEPP, F-35 042 Rennes, France

Corresponding author: xavier.garnier@irisa.fr

1 Introduction

The study of biological mechanisms require the production of large and heterogeneous datasets. These *omics* datasets are obtained routinely into labs, and are also available from public databases. Each of them have their own format and linking them require lot of time to link them. To ease integration of this data, we have developed AskOmics, a web tool designed to integrate heterogeneous biological data, and query them using a user-friendly interface. The software uses the semantic web technologies in order to homogenize data. AskOmics is used by several research team to analyze genomics, transcriptomics and pathways data. Software development is still ongoing and new features are added as new versions are released. Next version (AskOmics 3) will brings a new set of features.

2 AskOmics, integration and query using the semantic web technologies

AskOmics is a web software that uses the semantic web technologies (RDF/SPARQL) to integrate multiple data formats, and query them through a user-friendly interface. During data integration, user provides input files in common formats (CSV, GFF and BED). AskOmics internally generates the corresponding RDF triples and load them into a triplestore. Two kinds of information are generated, the *content*, corresponding to the raw data, and the *abstraction*, which describe how the raw data are organized and interlinked.

The query interface is composed of a dynamic graph that uses the generated abstraction to represent the entities integrated. Users interact with the graph to build a complex query linking several datasets integrated in AskOmics. When the query is built, AskOmics internally converts the graph into a SPARQL query and use it to interrogate the triplestore. Results are returned to the web interface and can be downloaded by the user.

AskOmics source code is available under AGPL3 licence at <https://github.com/askomics/askomics>. The GenOuest bioinformatics platform hosts a sandbox instance at <https://askomics.genouest.org>.

3 Ongoing work

We are currently developing version 3 of AskOmics (<https://github.com/xgaia/flaskomics>). This version will bring new features: the possibility to generate an AskOmics abstraction from RDF data, the hierarchy management between entities, and the implementation of federated queries against external endpoints, such as uniprot. Users will be able to link their data to large existing databases without having to import them locally.

This version has a new graphical user interface, build with React, a Javascript library made for building user interface. this change will provide a more modern and maintainable interface. The Python API is being refactored using Flask framework and Celery task queue for better performance by reducing the number of calls and making asynchronous call for long tasks such as data integration.

This new version is splitted into several micro services, *AskOmics*, a Python API and a Javascript interface. *Celery*, a task queue, to execute long task asynchronously, *Redis*, a worker, dependency of Celery, *Virtuoso*, a triplestore, for storing RDF and accepting SPARQL queries, and *Nginx*, a web proxy for url redirecting to the services. All these microservices are provided as docker containers. AskOmics deployment and upgrading is therefore very easy.