SBOL Stack: The One-stop-shop to Storing and Publishing SBOL Data

Curtis Madsen[1], Goksel Misirli[1], Matthew Pocock[1,2], Jennifer Hallinan[1], and Anil Wipat[1]

[1]School of Computing Science, Newcastle University, Newcastle upon Tyne, UK, NE1 7RU
[2]Turing Ate My Hamster LTD, Newcastle upon Tyne, UK

Introduction

Synthetic biologists have developed the *Synthetic Biology Open Language* (SBOL) [1], a data exchange standard for descriptions of genetic parts, devices, modules, and systems. The goals of this standard are to allow researchers to exchange designs of biological parts and systems, to send and receive genetic designs to and from biofabrication centers, to facilitate storage of genetic designs in repositories, and to embed designs of genetic parts and systems in publications. In order to achieve these goals, it is necessary to develop an infrastructure to store, retrieve, exchange, and publish SBOL data.
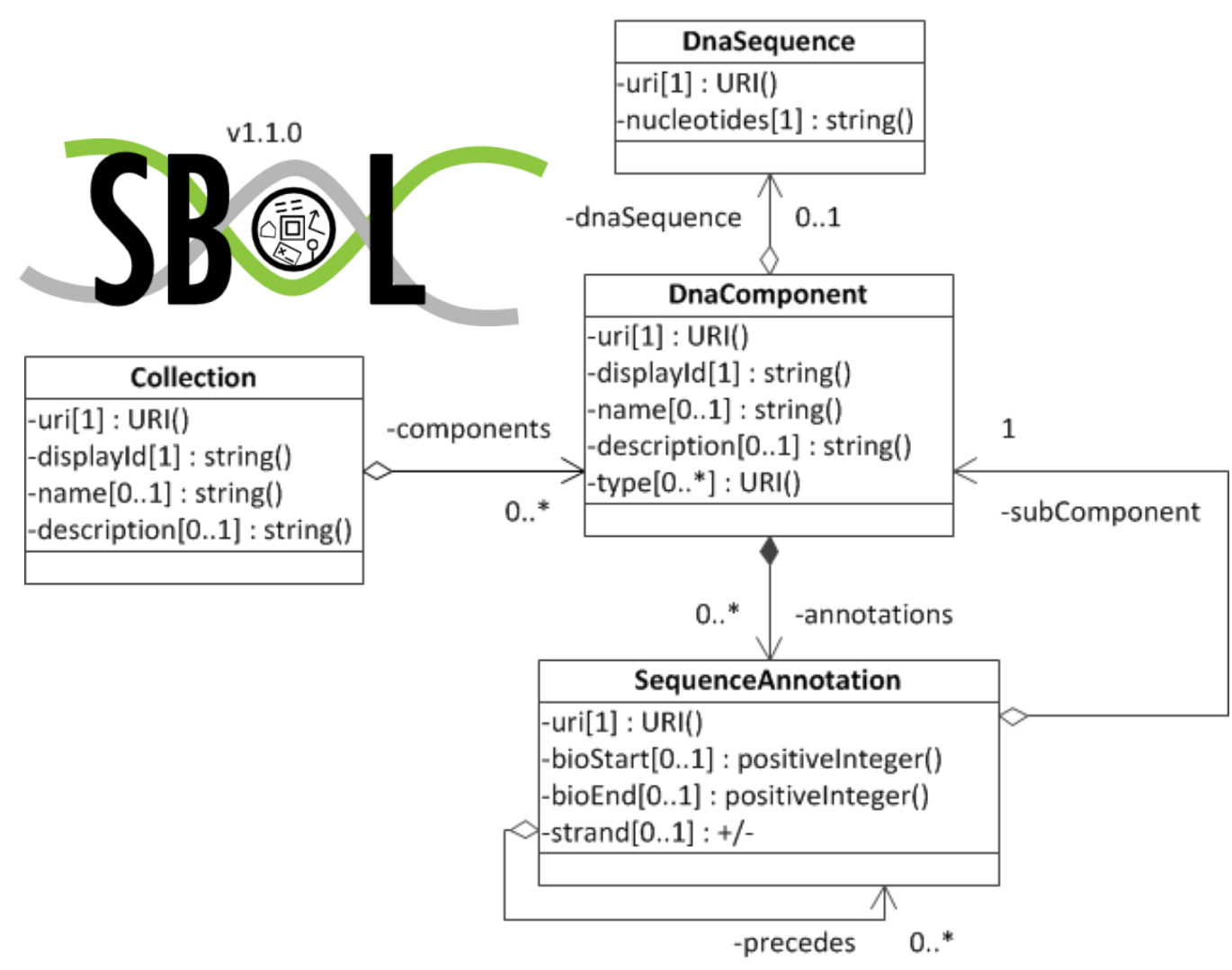
Repository

We have developed the SBOL Stack, a Sesame *Resource Description Framework* (RDF) [2] database specifically designed for storing and publishing SBOL data. The SBOL Stack can be used to publish a library of synthetic parts and designs as a service, to share SBOL with collaborators, and to store designs of biological systems locally. It includes a Web client that allows users to upload new biological data to the database and to perform SPARQL queries to access desired parts in SBOL format. When the desired parts are found, a user can download an SBOL document containing the SBOL data.
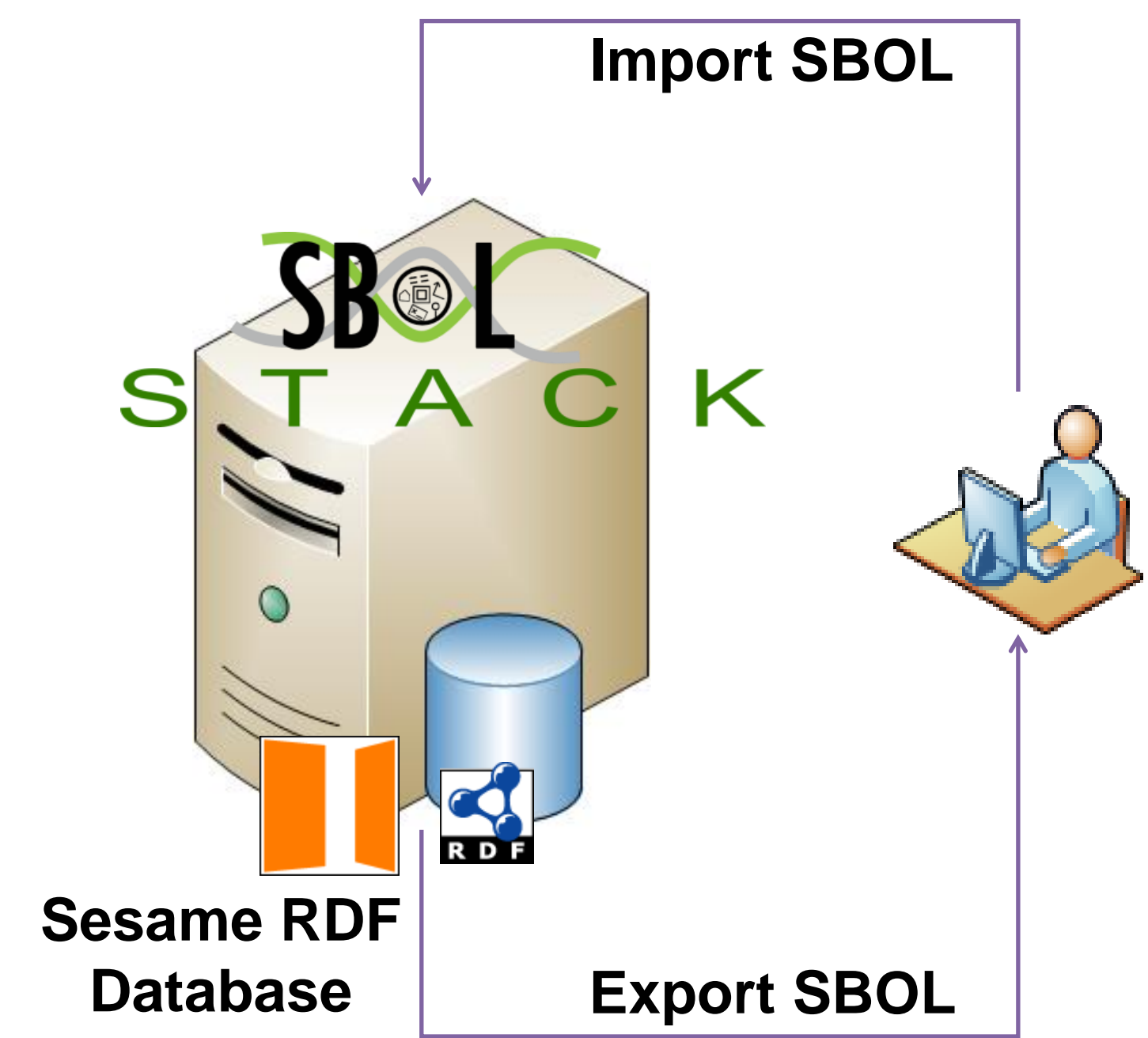
The SBOL Stack is available at:
www.sbolstack.org

SBOL Meets Semantic Web

To facilitate exchange, instances of the SBOL Stack can be installed locally by researchers at various organisations.

Data Integration

One of the problems in selecting a biological part for use in a genetic circuit design is identifying the mapping between a biological function and the relevant nucleotide sequences. Information about genetic features and their biological constraints is usually spread among many databases and should be integrated in order to constrain the designs. The SBOL language is based on RDF, which is ideal for data integration. SBOL data can be linked to integrated data in RDF format through the use of standard terms from ontologies such as the Sequence Ontology and the Gene Ontology. Semantic Web technologies can therefore be used to effectively store, query, and retrieve SBOL-related data.

An example of semantically-enriched integrated data is the BacillOndex [3] dataset, which includes information about genetic features, gene products and their annotations (such as molecular functions, orthology assignments, and enzymatic classifications), gene regulatory networks, metabolic pathways, and so on. Here, we used the RDF version of the BacillOndex dataset to map information about biological entities to SBOL objects. The data model from this dataset is then used to automate the identification of biological parts via standard SPARQL queries.
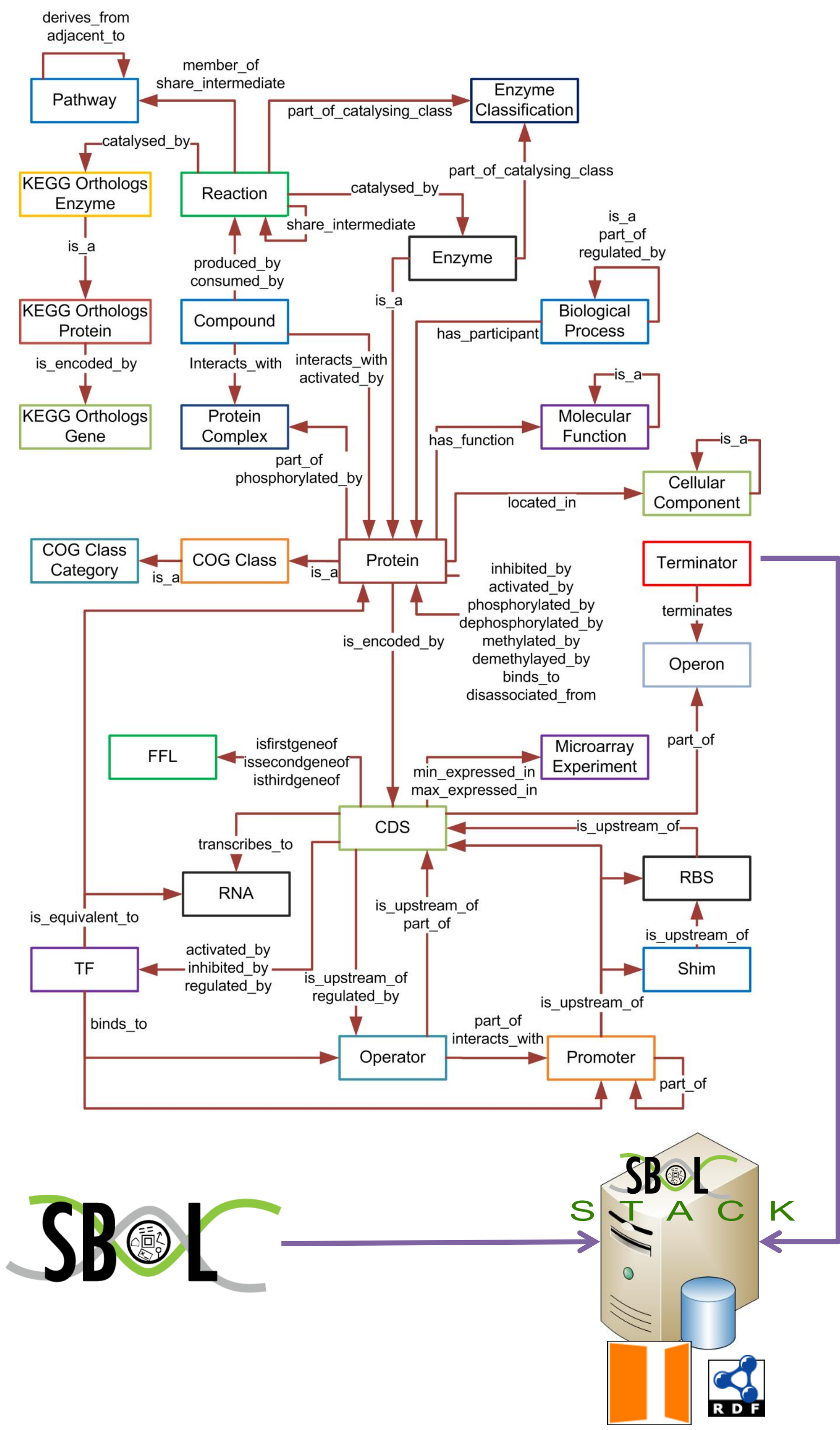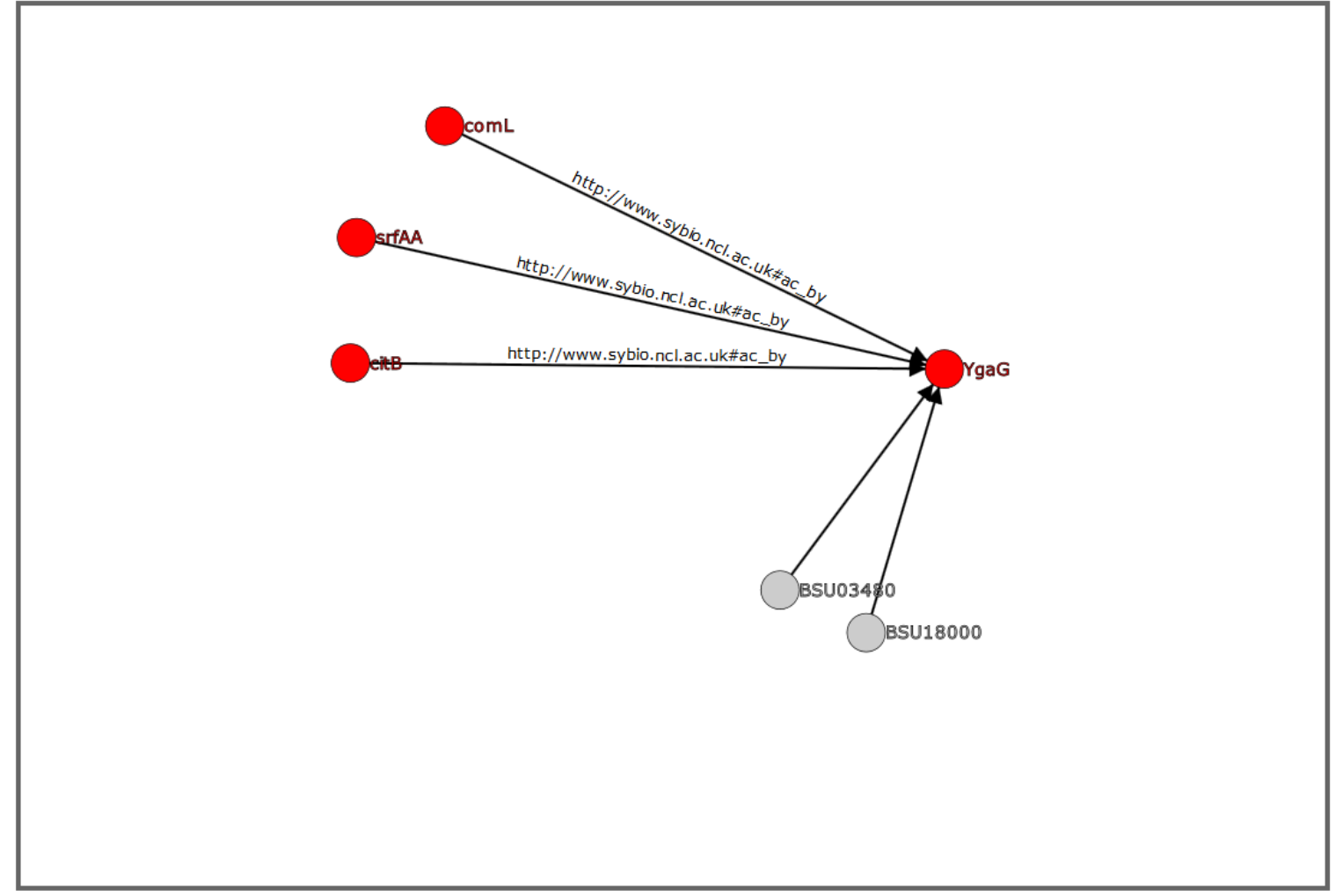

Federated Querying

Users can register different instances of the SBOL Stack with their own and perform federated queries [4] over all registered databases. These queries allow users to retrieve and compile more complete data from multiple databases without the need to manually query each repository individually. The SBOL Stack can register any Sesame RDF database, so other repositories that contain information about biological parts can be included in the federated queries.

Discussion

The automatic retrieval and integration that the SBOL Stack supports make it a must-have tool for researchers working on the design of systems in synthetic biology. In the future, we plan to add computational access to the SBOL Stack, which will allow synthetic biology tools to query and download SBOL data directly instead of requiring the data to be accessed via the web client.

References

[1] M. Galdzicki et al., "Synthetic Biology Open Language (SBOL) Version 1.0.0," 2011.
[2] F. Manola and E. Miller, editors. "RDF Primer," W3C Recommendation, World Wide Web Consortium, 2004.
[3] G. Misirli, A. Wipat, J. Mullen, K. James, M. Pocock, W. Smith, N. Allenby, and J. Hallinan, "BacillOndex: An Integrated Data Resource for Systems and Synthetic Biology," Journal of Integrative Bioinformatics, 2013.
[4] P. Jacsó, "Thoughts About Federated Searching," Information Today, 2004, Vol. 21, Issue 9.

School of Computing Science
Newcastle University

EPSRC
Engineering and Physical Sciences
Research Council