Title: Scalability of data management in the KAMI bio-curation system

Supervisor: Russell Harmer

Duration: 3 months

Project Context:
I lead a small team (1 post-doc and 1 PhD student) developing a bio-curation tool for the development of rule-based models (in the Kappa language) and subsequent simulation and analysis of cellular signalling pathways and their deregulation in cancers.

The system, called KAMI, is built on top of a graph-based knowledge representation framework, built in Python on top of networkX, which provides a clear ‘audit trail’ of updates to our state of knowledge via the use of graph rewriting techniques.

Project Objectives:
In order for our approach to scale to very large data sets, we need to relocate much of the data structure (currently maintained in RAM) of the KR framework to a database from which we can reify small pieces to and from RAM as required.

The aim of the internship is to design and build (part of) the software infrastructure necessary to solve this scalability issue. This could involve work on designing an appropriate exchange format and/or on designing the database structure anf/or on providing input pipelines to the database. The exact objectives will be discussed with the successful candidate at the beginning of the internship as a function of their particular interests and technical competences.

A successful candidate would have experience with programming in Python, some knowledge of web development (JavaScript, Python back-ends) and/or databases (relational or semi-structured) and an interest in biology / bio-informatics.
Contact Information:
Russell Harmer
Phone: +33 4 72 72 89 01
russell.harmer@ens-lyon.fr

Location:
Laboratoire LIP,
ENS Lyon,
46 allée d’Italie,
69364 LYON Cedex 07
France

Institute:
ENS Lyon

Lab:
LIP, équipe Plume

Selected publications from the team
Bio-curation for cellular signalling: the KAMI project. Russ Harmer, Yves-Stan Le Cornec, Sébastien Légaré and Ievgeniia Oshurko.
CMSB 2017. LNCS volume 10545, pp. 3–19.