Integration of HTC and HPC tools for solving complex problems in computational biology

Dragos Ciobanu-Zabet[1], Dragos Hont[2], George Necula[1], Ionut Vasile[1], Mihnea Dulea[1]

[1] Dept. of Computational Physics & Information Technology
Horia Hulubei National Institute for R&D in Physics and Nuclear Engineering (IFIN-HH)
[2] S.C. Totalsoft SA

Molecular modeling makes extensive use of:

- **HTC**, for database interrogations, massive data retrieval and conversion, high-throughput VLS, protein structure and function prediction, genomics, etc.
- **HPC**, for molecular dynamics simulations, chemical structure calculations, etc.

The molecular modeling of complex cellular subsystems implies procedures that require both sequential and parallel computing steps in order to obtain physically significant results, such as the description of drug-protein interaction in bacterial membranes.

The implementation of such a procedure often involves many different software tools for retrieval, conversion, processing and analysis of data, handled by experienced users.

We recently designed and currently implement a new integrated system for molecular modeling that automates the above-said procedures by means of workflows, in order to simplify the user's tasks.

The system consists of an extensible and scalable pool of **HTC and HPC resources** which are accessible to the user through a graphical frontend (**portal**).

The data handling and processing through various retrieval, conversion and modeling tools is managed by means of automatic, programmable and reusable workflows.

FUNCTIONALITY

PURPOSE: to empower the researchers with software tools allowing them to perform faster biological system setups, and numerical simulations.

The **portal** is currently customized for the modeling and simulation of subcellular structures of Gram-negative bacteria.

It provides access to software tools for microbiology and pharmacology researchers:

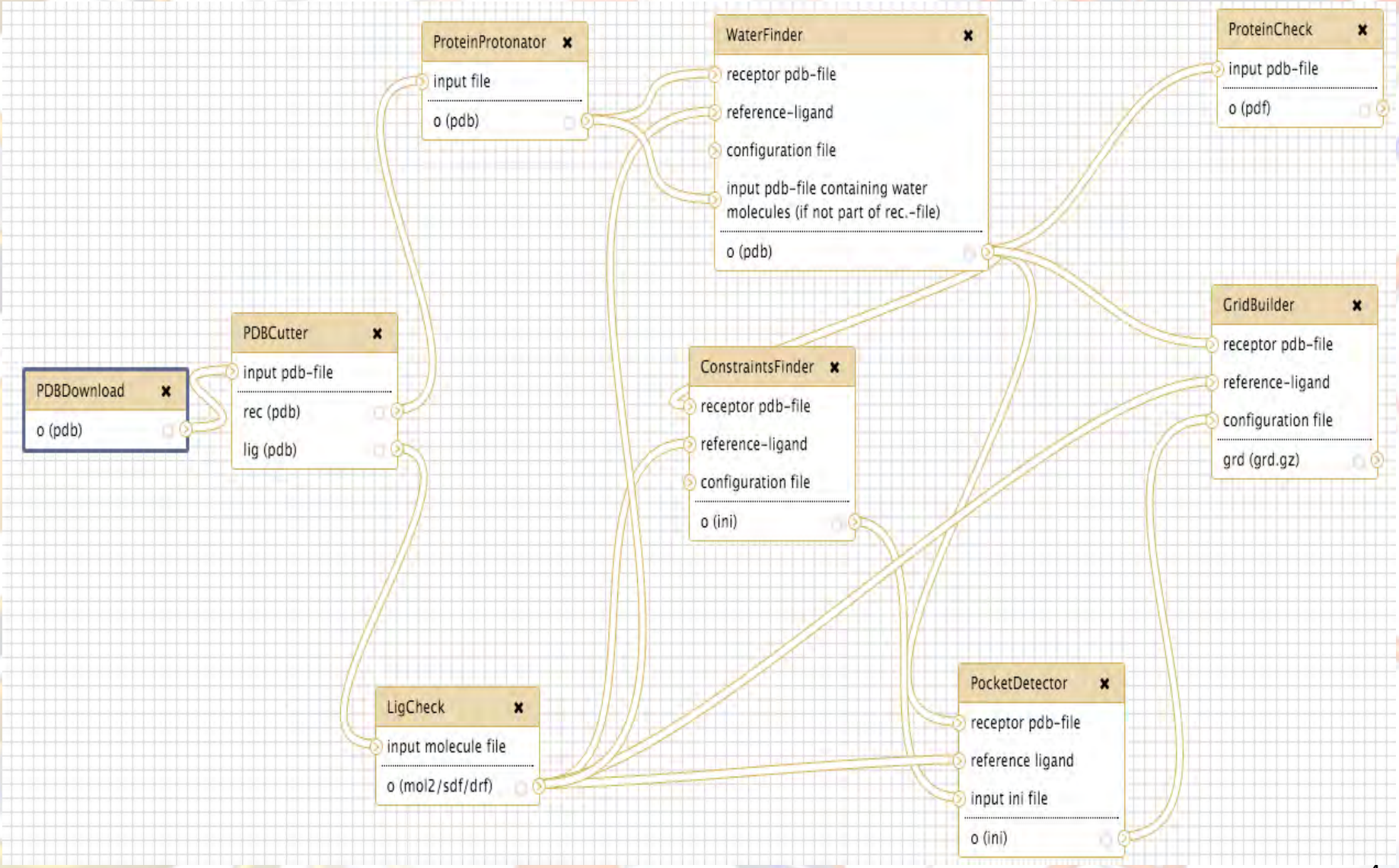
- ❑ computational modeling at molecular level
- ❑ workflows composed of tasks/modules that correspond to various modeling steps:

Examples:

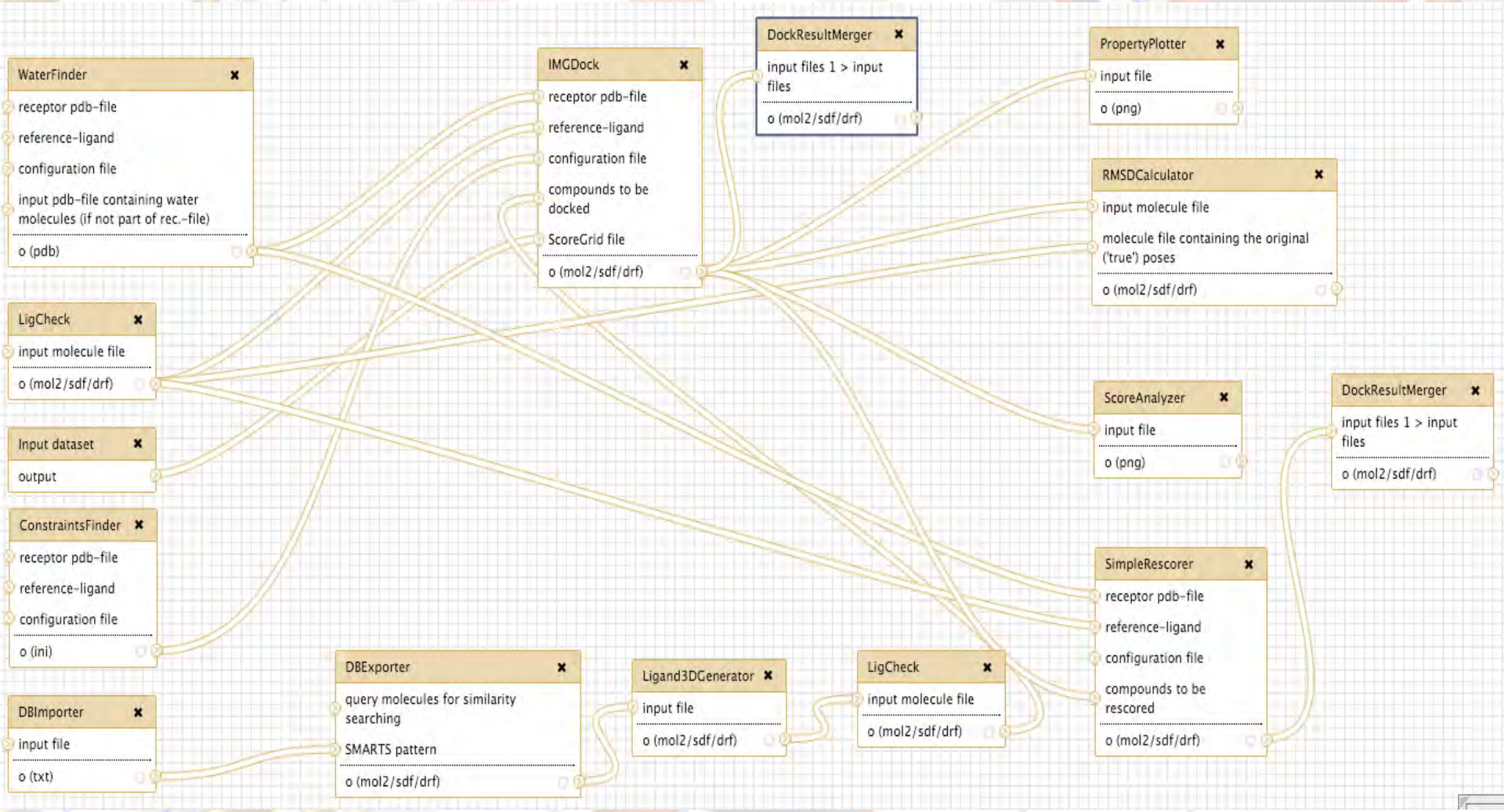
- Search of molecular structures in databases (LipidBank, RCSB)
- Data format conversion (OpenBabel)
- 3D structures construction
- Molecular modeling (CHARMM); molecular docking
- Quantum chemistry (Gaussian, GAMESS)

- ❑ workflows are submitted for execution on the available HTC/HPC resources

Example1: Workflow for preparing data for molecular docking



Example2: Molecular docking workflow



DISTRIBUTED SYSTEM OF HTC & HPC RESOURCES

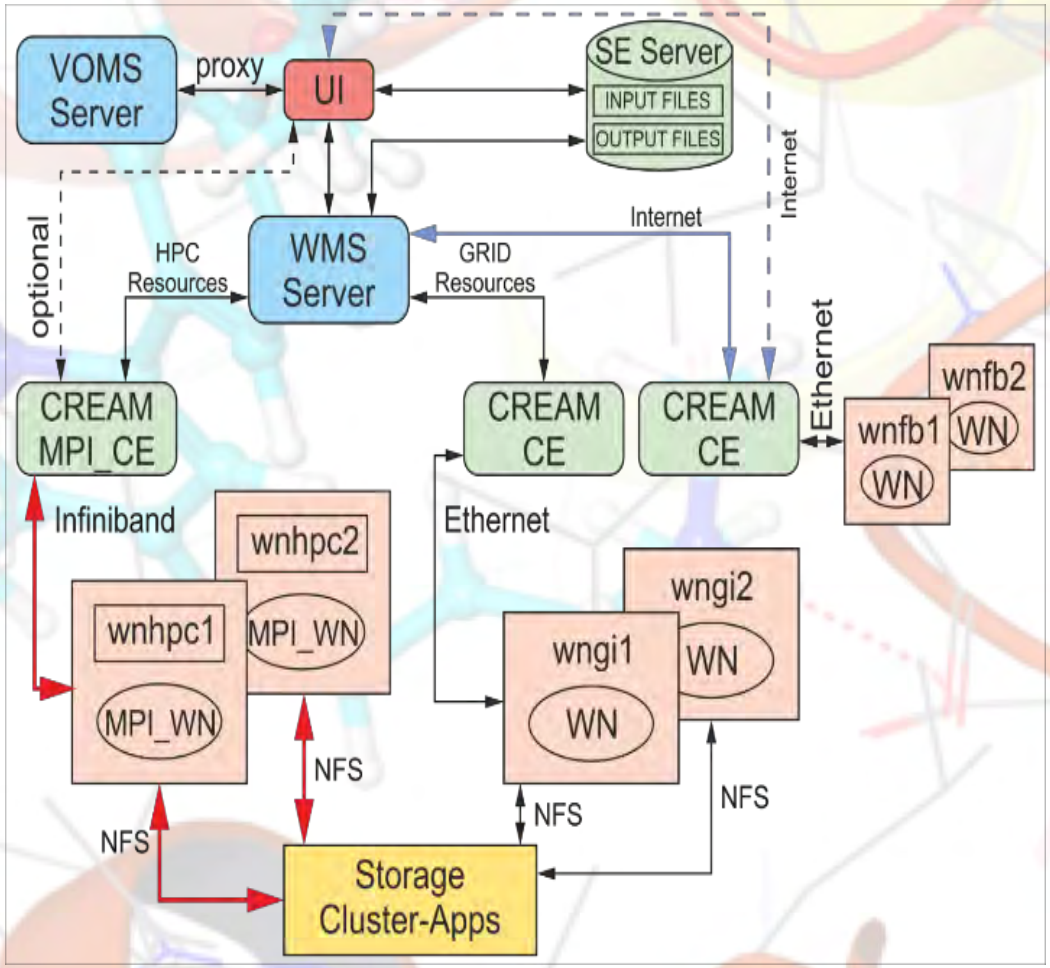
The system is currently using the resources of the GRIDIFIN site:

- HPC cluster
- HTC WNs
- Storage

GRIDIFIN: Two CEs (one for distributed and the other for parallel computing)

Multiple HTC/ HPC clusters can be attached to the system.

(The connection to another site is depicted in figure.)



IMPLEMENTATION

Grid jobs containing Taverna workflows are submitted and executed on the HTC/HPC clusters. The major advantage of Taverna is that it uses distributed services, optimizing the running of workflows, regardless of the local infrastructure.

The portal integrates graphical instruments for:

- Management of the input data

  Download/upload of input files, jdl files, shell files, etc. from user or databases
  Template files for creating jdl files, input for molecular dynamics, scripts
  Edit and save input files

- Preparing, submitting for execution and management of jobs

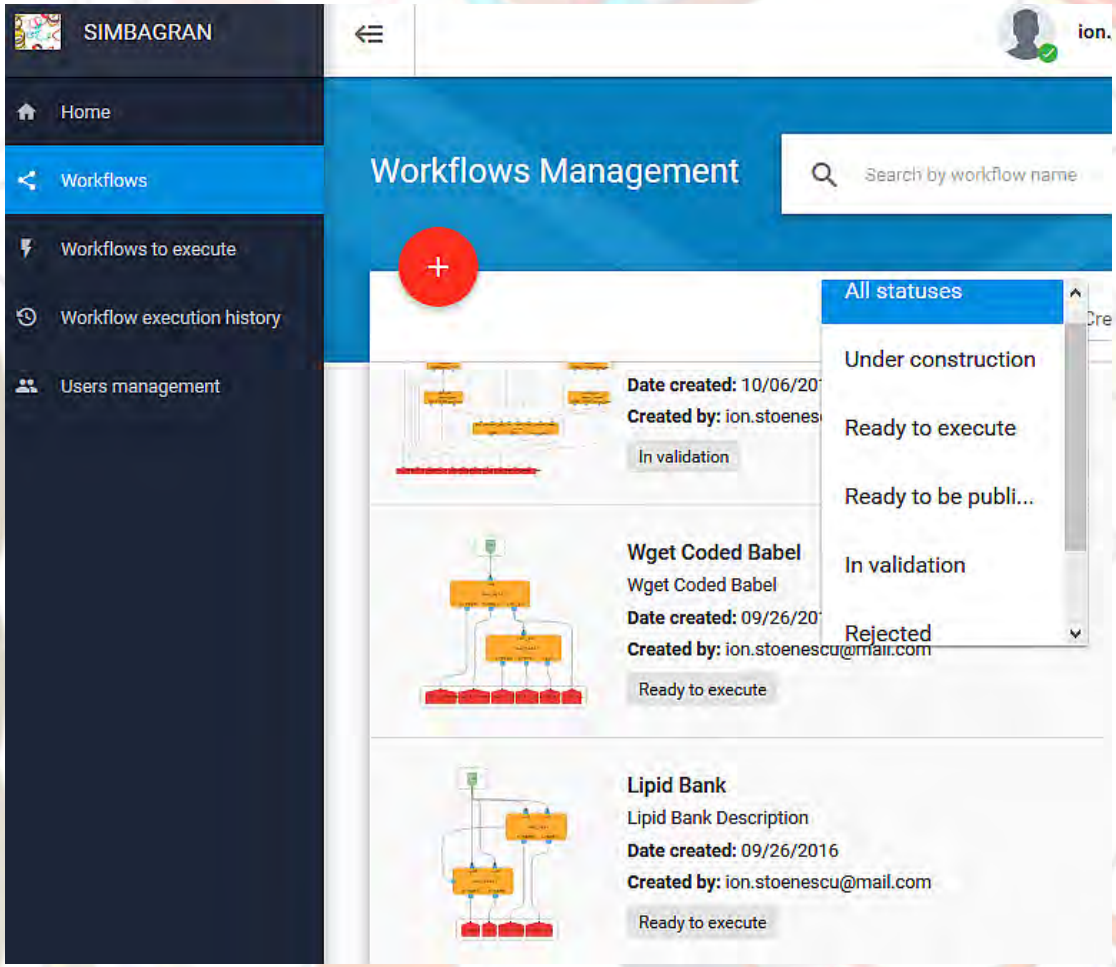
  Edit source files, jdl files, wrapper (for compiling), send jobs in execution

- Input data pre-processing and data analysis (MMTSB / AmberTools)
- Graphic representation and analysis of results (JSmol – in progress)

PORTAL: WORKFLOW MANAGEMENT

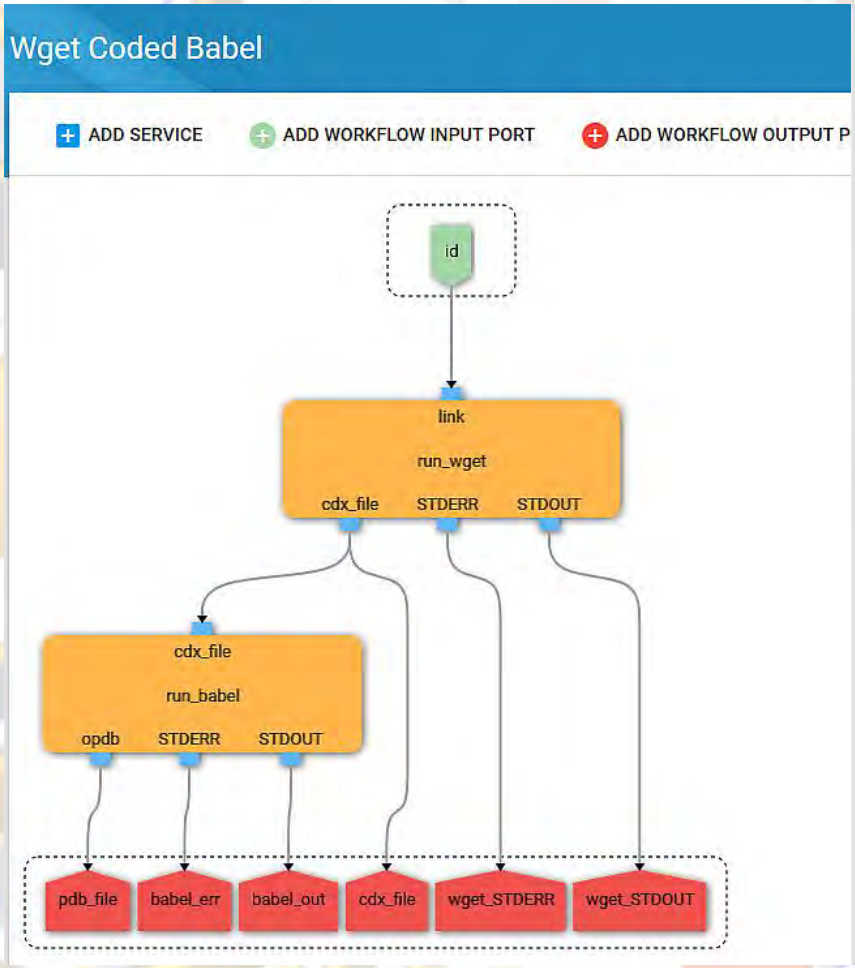
Example of simple workflow composed of two parts:

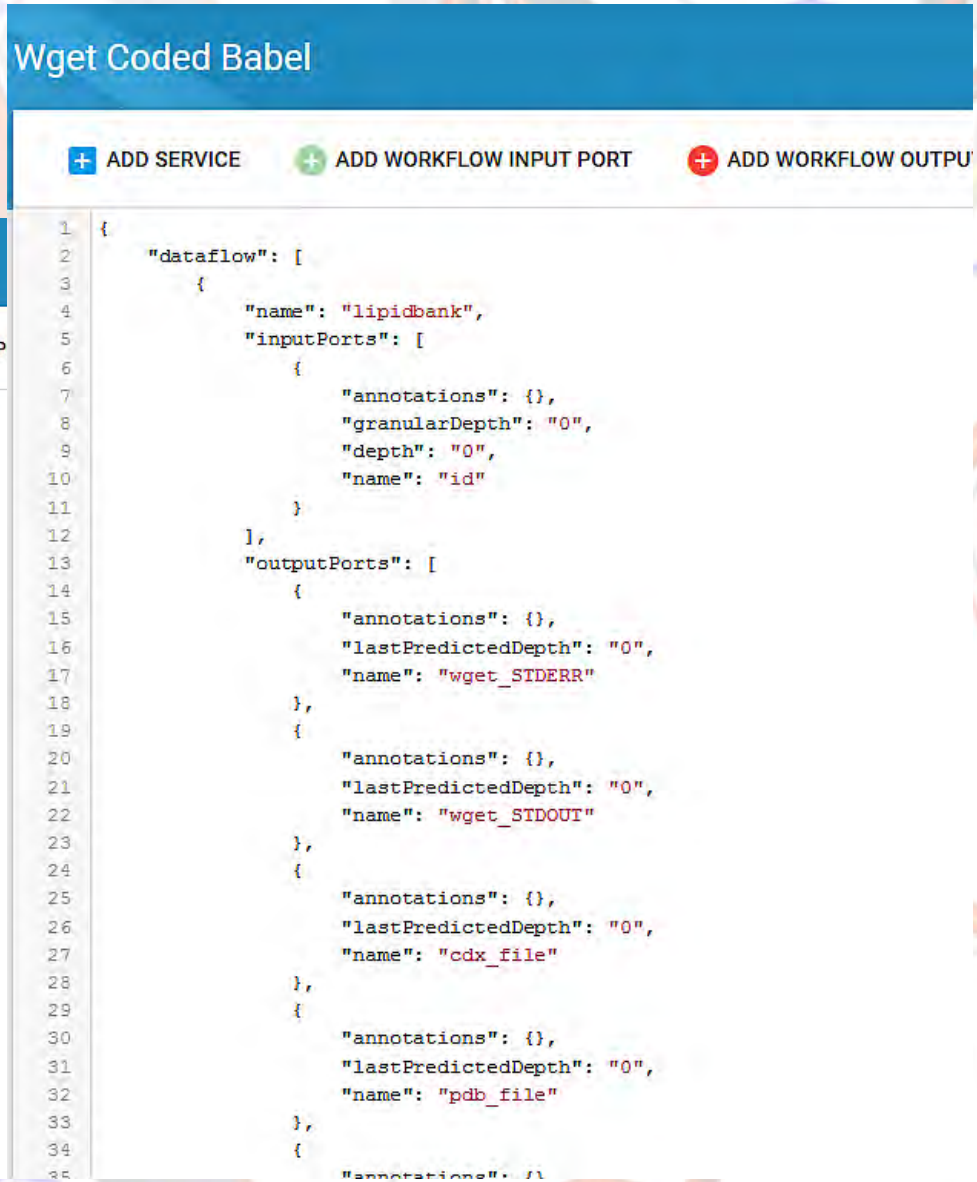
- file download from LipidBank
- file conversion from cdx format to pdb format using OpenBabel



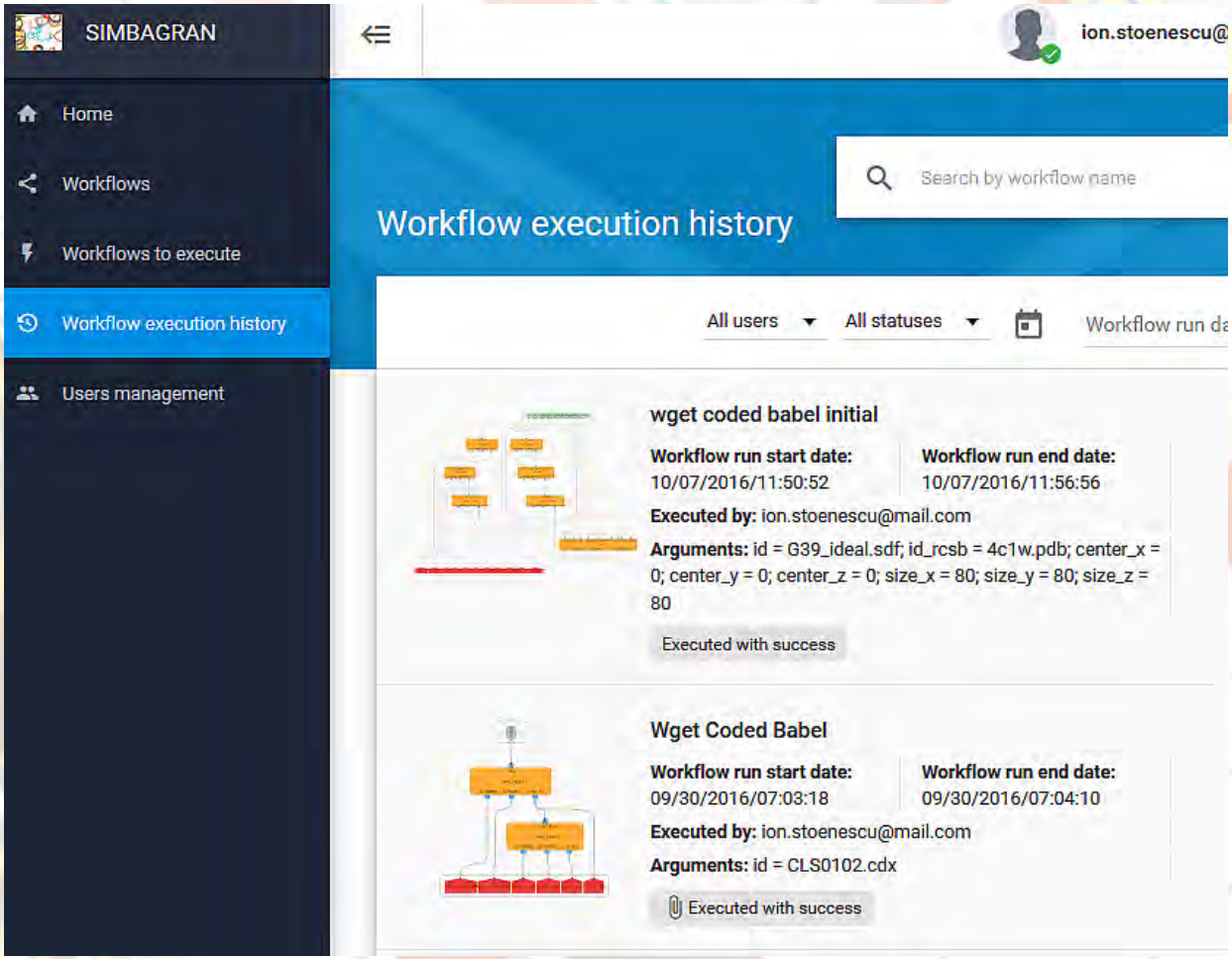
WORKFLOW CREATION

Graphical representation and code





WORKFLOW EXECUTION HISTORY



Summary and conclusions

Graphical frontend (portal)

- Makes use of advanced (HTC/HPC) computing resources
- Zero user knowledge of HTC/HPC management
- Based on Taverna workflow management system
- Many tools in a single place
- Ease of use
- Graphical simplicity (user friendly)

THANK YOU FOR YOUR ATTENTION !