Architektura dużych projektów bioinformatycznych

Bartek Wilczyński

bartek@mimuw.edu.pl
http://www.mimuw.edu.pl/~bartek

Wykład 6. - O przepływach pracy – Kepler, Taverna i Galaxy

16. XI. 2015

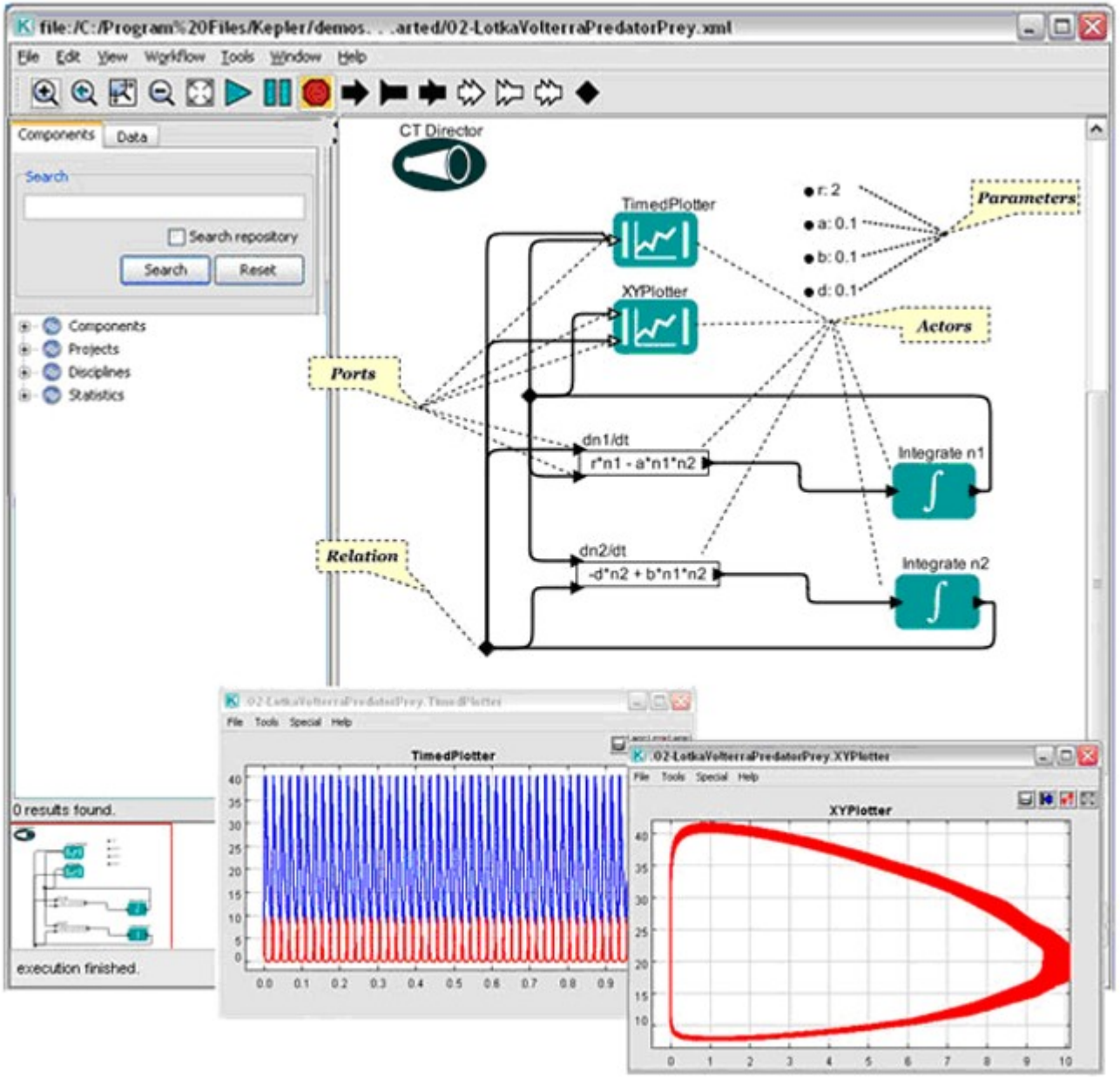
Plan na dziś

- Usługi sieciowe i projekty typu grid
- O pomyśle na przepływy pracy – Kepler (2002), Taverna workbench (2004), i wiele innych
- Pomysły na wymianę workflow'ów - myExperiment, i usług – biocatalogue
- Porównanie podejść – rozproszonego z użyciem aplikacji klienta(Taverna itp) i klient-serwer przez przeglądarkę (Galaxy)

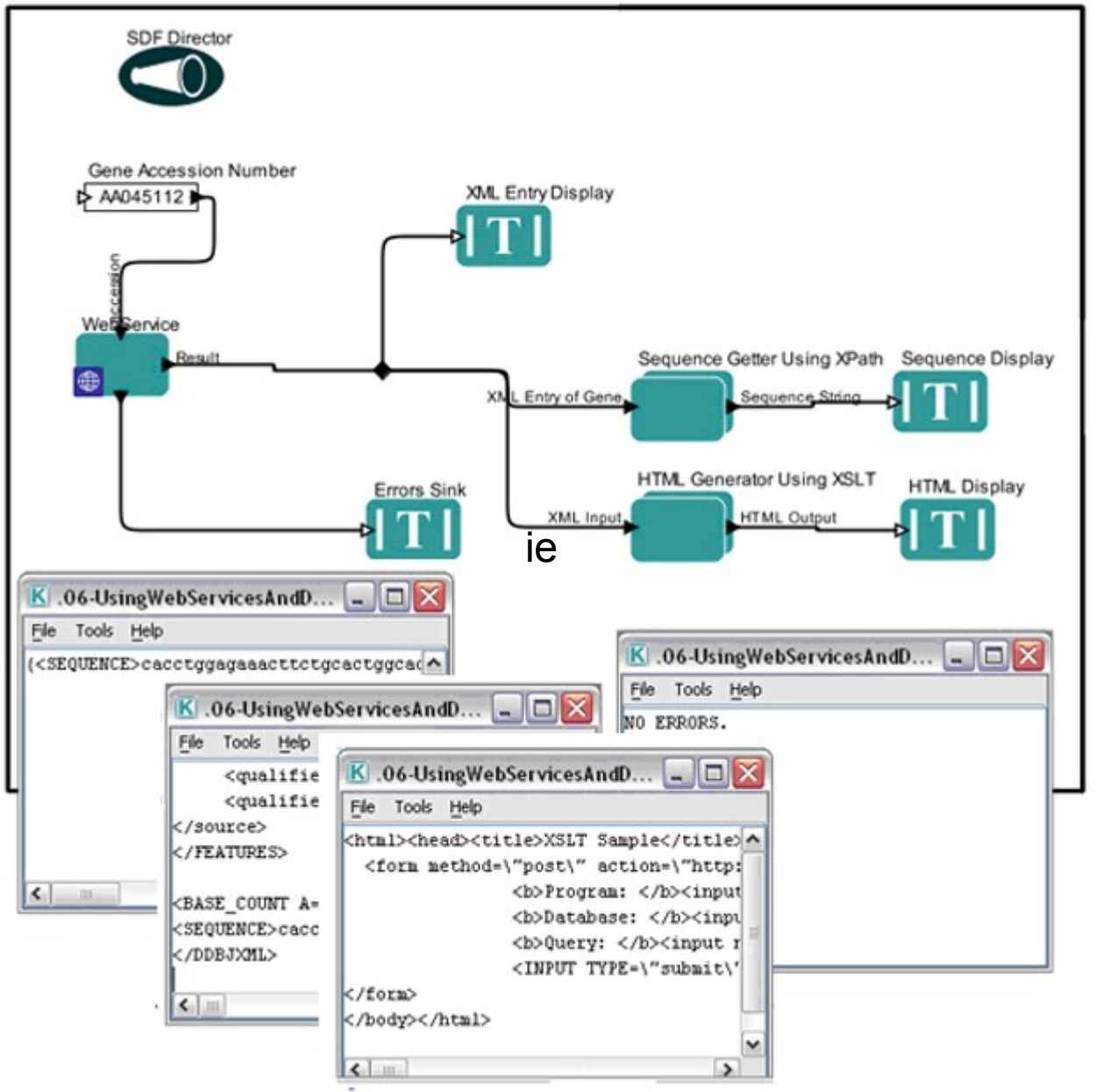
Projekt Kepler

- Pomyślany jako interfejs do pakietu Ptolemy (1999), służącego do złożonego modelowania matematycznego procesów fizycznych
- Napisany w Javie
- Pierwsze wersje w roku 2002
- Obecnie w wersji stabilnej (ostatnie wydanie w połowie 2013)
- Napisany w Javie
- Open Source(BSD), wspierany przez granty NSF

Model Lotki-Volterry w Keplerze



Pobieranie sekwencji DNA z usługi sieciowej



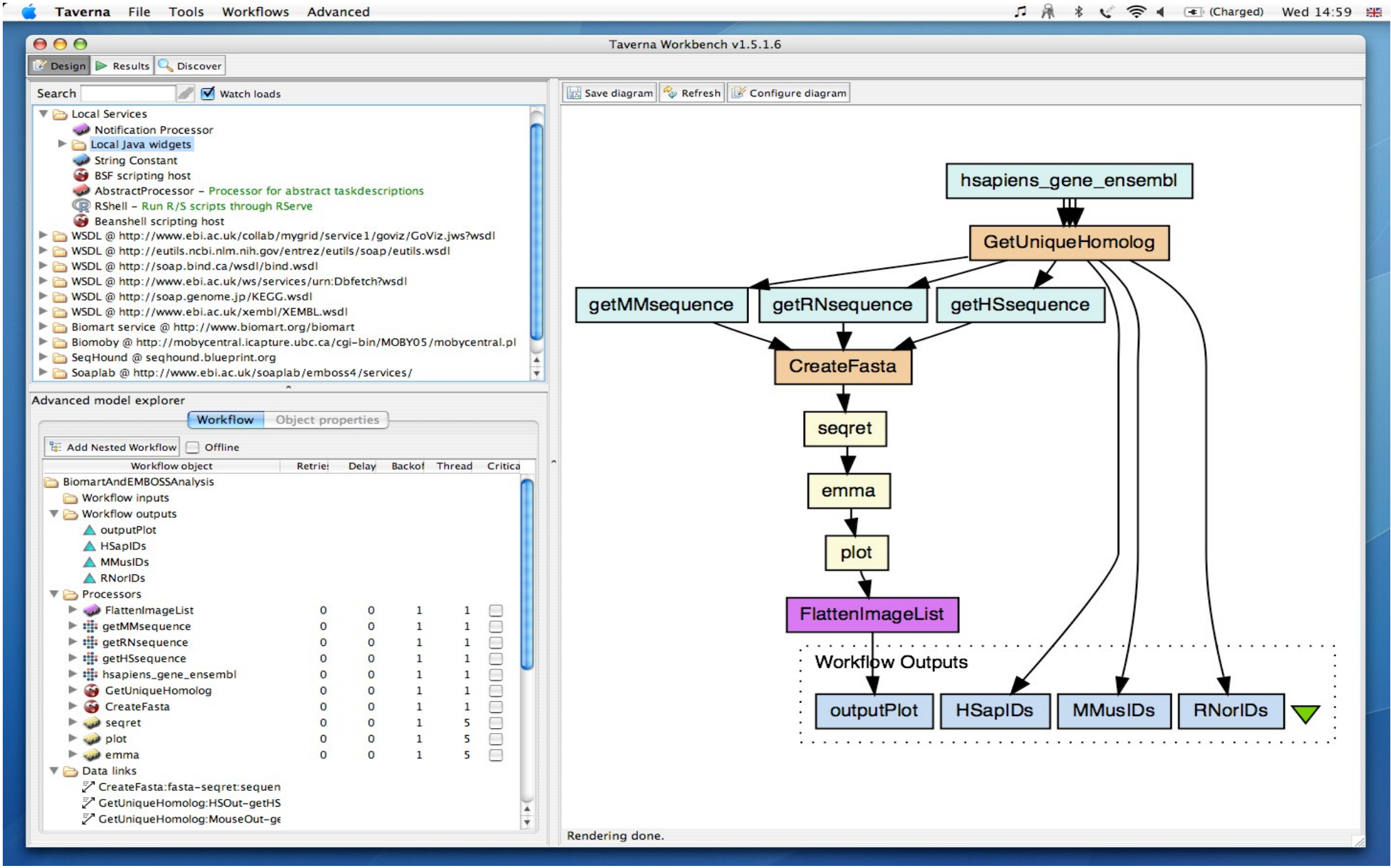
Cechy Kepler'a

- Środowisko uruchomieniowe Ptolemy (równolegle działające systemy)
- Interfejsy do czujników eksperymentalnych
- Wersje na GRID
- BioKepler dla bioinformatyków
- Raczej dla eksperymentatorów
- Silne wsparcie w środowisku fizyków doświadczalnych

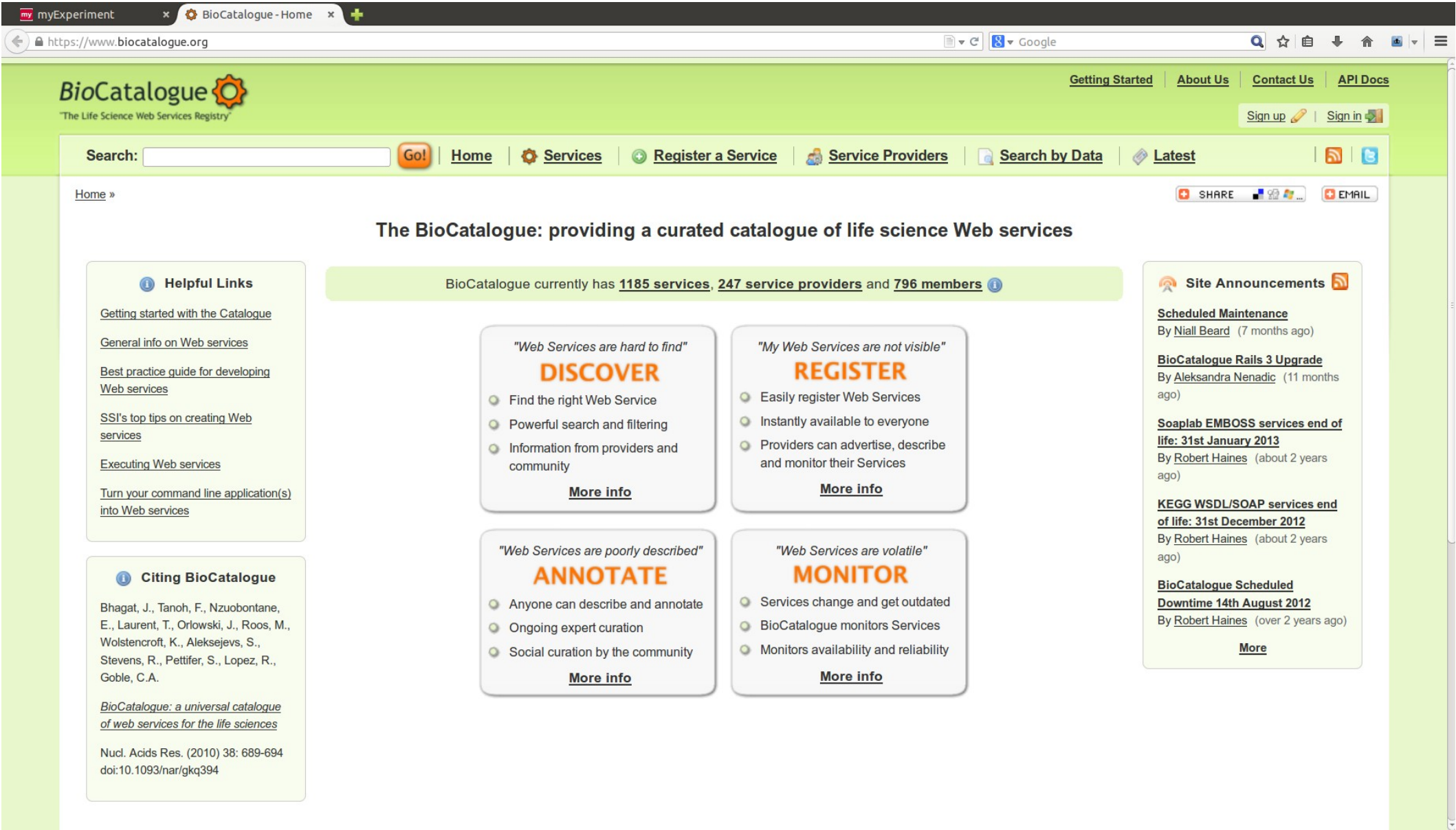
Taverna Workbench

- Rozpoczęta jako oddzielny projekt w roku 2002
- Wsparcie grantami badawczymi, obecnie konsorcjum
- Licencja LGPL
- Zasadniczo oparta na pomyśle reproducible research
- Oryginalnie pomyślana jako interfejs do webserwisów (bez obliczeń lokalnych)
- Wersje dla astronomów, bioinformatyków, itp.
- Obecnie również do obliczeń lokalnych, interfejs do R, Galaxy itp

Pobieranie sekwencji z Ensembl



BioCatalogue



MyExperiment

myExperiment

Switch to tab: www.myexperiment.org

Google

my experiment

myExperiment makes it easy to **find**, **use** and **share scientific workflows** and other **Research Objects**, and to build **communities**.

All Search

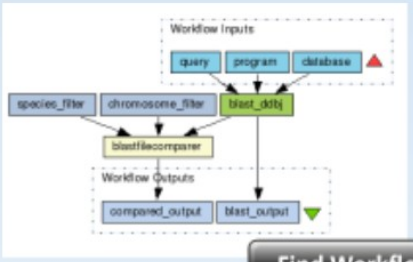
First time visitor? Try these videos:

- Project Introduction
- Bioinformatics Case Study

Use myExperiment to...

- Find Workflows
- Share Your Workflows and Files
- Create and Find Packs of Items
- Find People and Make Friends
- Create and Join Groups
- Build your Profile and Reputation
- Tag and Rate things
- Write Reviews and Comments

Explore



Find Workflows

About myExperiment

Join the Mailing List

myExperiment Publications

For Developers

Give us Feedback

The BioCatalogue Project

Register

or Login:

Username or Email:

Password:

Remember me:

Or use OpenID:

(eg: name.myopenid.com)

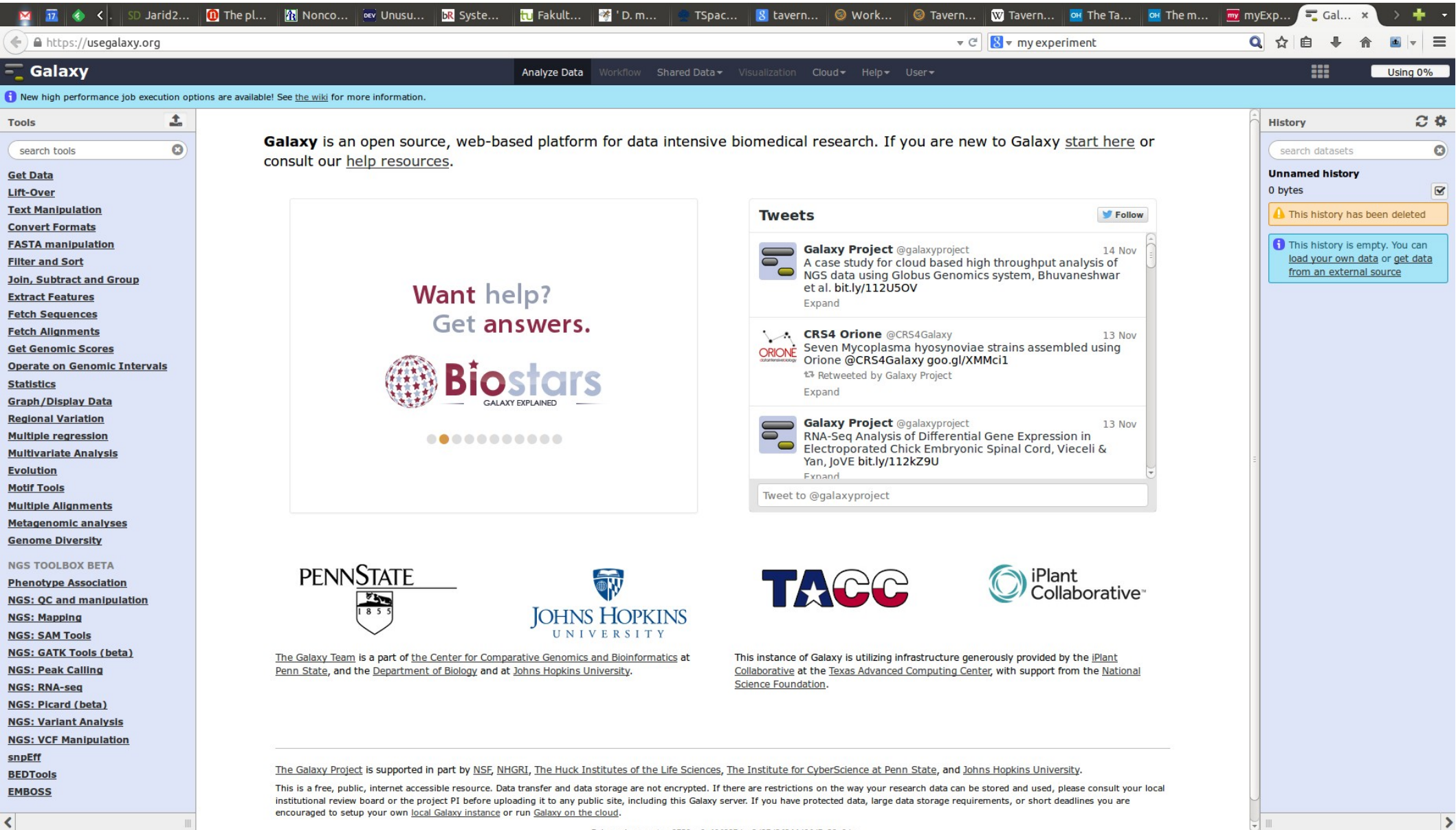
Login

Forgot Password?

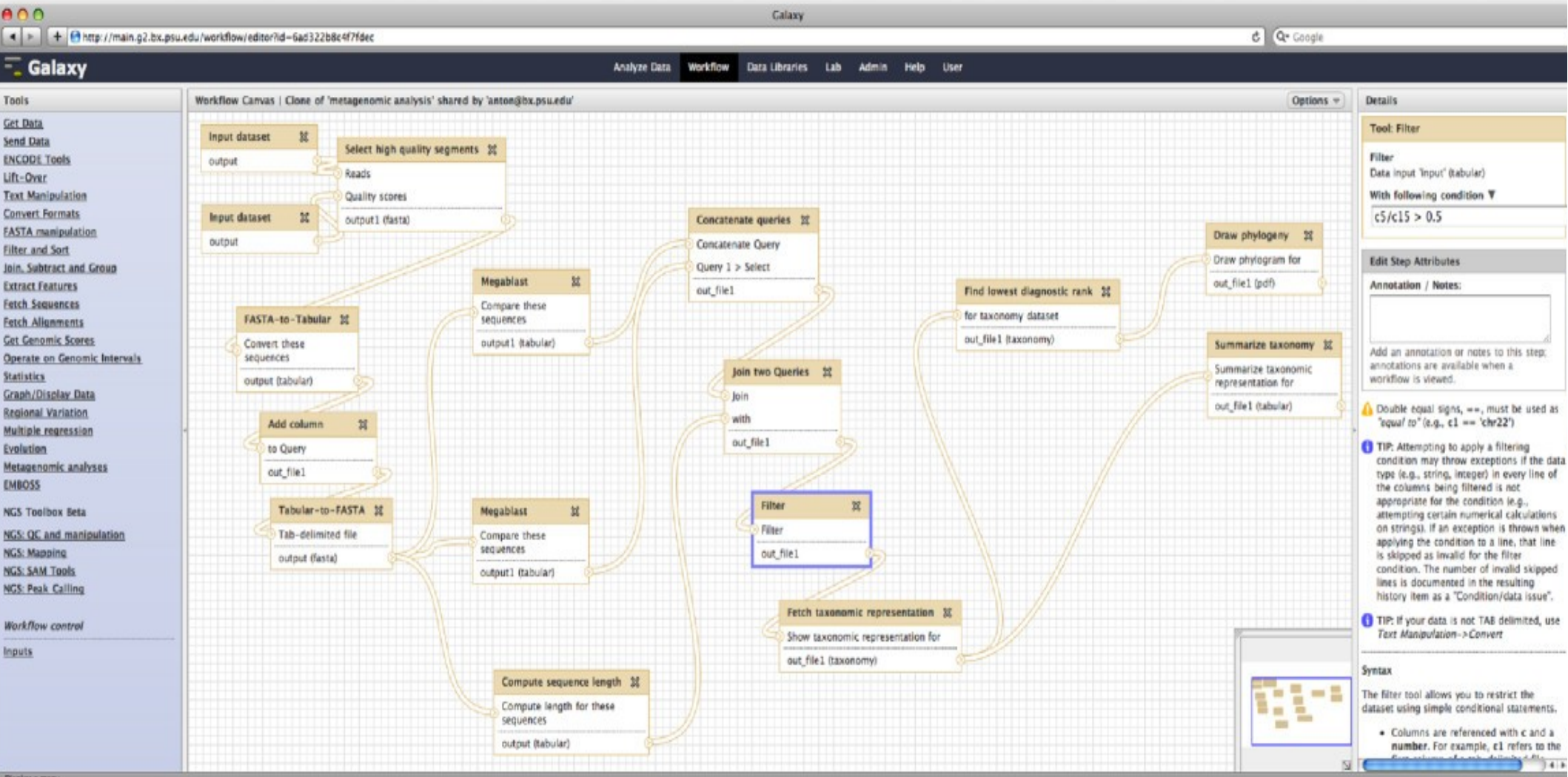
Galaxy

- Środowisko do uruchamiania narzędzi bioinformatycznych
- Architektura klient-serwer, ze wsparciem dla historii danych i analiz
- Pozwala na zapisywanie historii i dzielenie się nimi w ramach serwera
- Edytor “przepływów”, pozwalający na uruchamianie tych samych analiz na wielu zbiorach danych
- Zachowuje elementarne zabezpieczenia danych
- Pozwala na publikację wyników analiz

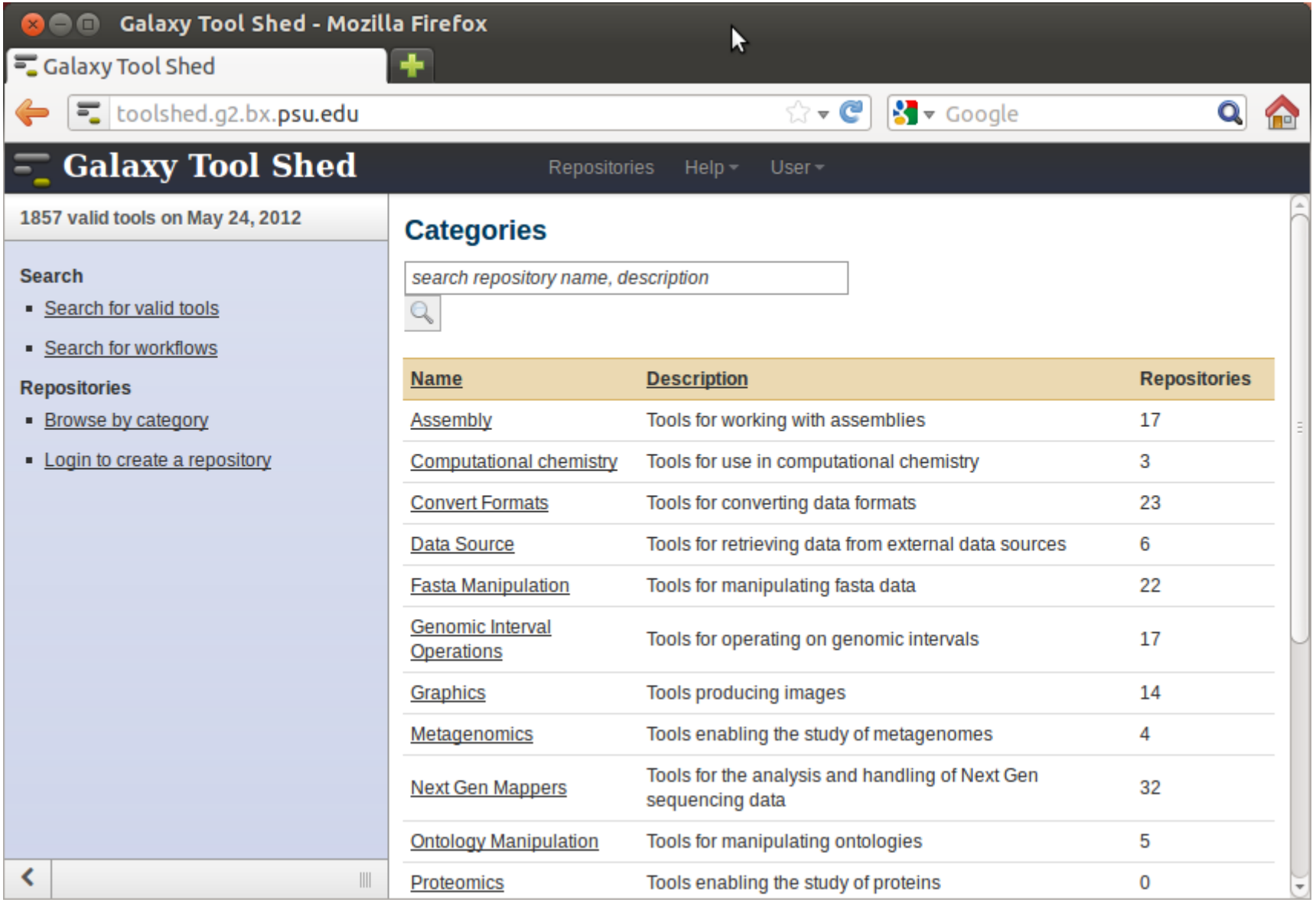
Publiczny serwer galaxy



Workflow editor



Repozytorium narzędzi - toolshed



Możliwość tworzenia własnych “opakowań” dla istniejących narzędzi

- Pisząc dość krótki plik “opakowujący” (Ang. Wrapper) można dołączyć dowolny program działający na serwerze do własnego galaxy
- Łatwo też można zmodyfikować istniejące wrappery, aby dostosować istniejący program do własnych potrzeb
- Jest to dość proste, choć jeśli dotyczy aplikacji graficznych, może być bardziej skomplikowane

Podsumowanie

- Wiele projektów pomyślanych jako narzędzie do “reproducible research”
- W zasadzie jedyny model finansowania to granty badawcze
- Bardzo powolna adopcja rozwiązań, które wymagają dodatkowego wysiłku, lub są ograniczone
- Duży sukces galaxy w środowisku bioinformatycznym, bo tu może stanowić ułatwienie
- Nadal niełatwe zdobywanie popularności wśród specjalistów