Galaxy Integration into an External Information System

Alban Lermine – Galaxy IFB Day – 4th December 2013
Galaxy @ Institut Curie

- We start using Galaxy since half 2011
- 3 production instances
  - Public server (January 2012)
    - **Nebula** (http://nebula.curie.fr)
    - Dedicated to ChIP-seq data analyses
    - 29 tools, 558 users (worldwide)
    - Analysis tutorial and toy data accessible on web site
  - Internal access server (July 2013 and November 2013)
    - **Galaxy**
      - 383 tools, 45 users, 23 projects
      - Many tutorials and toy data available (e-learning)
    - **GalaxyDX**
      - Dedicated to diagnostic analyses
KDI – Institut Curie Information System

- KDI = Knowledge Data Integration
- KDI also managed system rights access to projects
- Simplified Concept scheme:

```
Projet 123
  Dataset 1234
  Dataset 1235
  Dataset 1236

Analyses
Environment
User access
Project
+ Referenced datasets

Data
Data
Data
```
Default Galaxy - Data organisation

- One applicative user - Owner of all files created
- Creates incremented named files in one directory (ex: dataset_1234.dat)
- Access rights managed at the web interface level using metadata

Web Interface
- User management
- Display:
  - Tools
  - Workflows
  - Files
  - Status
  - Histories

Galaxy Engine
- Create and submit jobs
- Send results
- Access rights management
- DB Communication

Datasets Created
- Dataset_1234.dat
- Dataset_1235.dat
- Dataset_1236.dat

Galaxy DB
- Dataset_1234.dat
- Dataset_1235.dat
- Dataset_1236.dat
Needs for Galaxy Integration into KDI

- Created files have to be stored under corresponding project and username directory
- Created files have to be owned by the user who launch the analysis
- Galaxy has to be able to test system right access on:
  - Input files
  - Project directory
- Simple user should be able to upload files without using network upload (avoid data copy)
What I have done

- At the web interface level
  - Add 2 mandatory fields in all tools:
    - KDI project number
    - Output file name
  - Script to add Automatically this 2 new fields
- At the engine level
  - Automatic interpretation of the new fields
    - Add 2 new values in the job param_dict (use for command line creation)
    - No need to rewrite xml for each tool
  - Create automatically the new output path – new output files
    - /data/<kdi_version>/project_galaxy/<KDI_project>/<user>/job/
    - Link new output files to the datasets created by Galaxy in DB
- At the system level
  - Check input and output right access for user
  - The web applicative user gives output files to the user who launch the analysis
New Galaxy Data organisation

Web Interface
- User management
- Display:
  - Tools
  - Workflows
  - Files
  - Status
  - Histories
  - + KDI project
  - + File name

Galaxy Engine
- Create and submit jobs
- Send results
- Access rights management
- DB Communication
- + KDI project
- + File name

Galaxy DB
- Dataset_1234.dat
- Dataset_1235.dat
- Dataset_1236.dat

/KDI_project/blue_user/job/myFile

/KDI_project/green_user/job/myFile

/KDI_project/red_user/job/myFile
Galaxy Integration into KDI scheme

Project_Galaxy/123/
Username/
Job/
File.ext

Galaxy Libraries

Projet 123

Analyses
Environment
User access
Project
+ Referenced datasets

Dataset 1234

Data

Dataset 1235

Data

Dataset 1236

Data
LIVE DEMO
Subversion management

- SVN
- Manage 3 distincts environments
  - Dev
  - Valid
  - Prod
- Directory galaxy-svn containing all files needing subversion control:
  - Tools
  - Tool list
  - Galaxy engine
  - Some configuration files (Annotations, datatypes)
- Template for universe_wsgi.ini used for Dev, Valid & Prod instance
  - IP and port
  - DB name, ID and password
- Deployment script
  - Starting from fresh galaxy-dist distribution
  - Edit universe_wsgi.ini template (depending on environment)
  - Create symbolic link for lib/ & display_applications/
Subversion management scheme

Galaxy-dist

@lib
@display_applications
Universe_wsgi.ini

Galaxy-svn

Lib
Display_applications
%universe_wsgi.ini%
Tools
Tool_list
datatypes

KDI FS

Apache

Dev/Valid/Prod

%DB%
%ID%
%PASSWORD%

%HOST%
%PORT%
Acknowledgments

- NGS Team:
  
  Nicolas Servant
  Séverine Lair
  Valentina Boeva
  Vivien Deshaies
  Bruno Zeitouni

- KDI Team:
  
  Philippe Hupé
  Eric Viara
  Stuart Pook

- System team:
  
  François Prud’homme
  Camille Barette