



# 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

*Genome analysis*

## **Nebula – a web-server for advanced ChIP-Seq data analysis**

Valentina Boeva<sup>1,2,3\*</sup>, Alban Lermine<sup>1,2,3</sup>, Camille Barette<sup>1,2,3</sup>, Christel Guillouf<sup>1,4</sup> and Emmanuel Barillot<sup>1,2,3</sup>

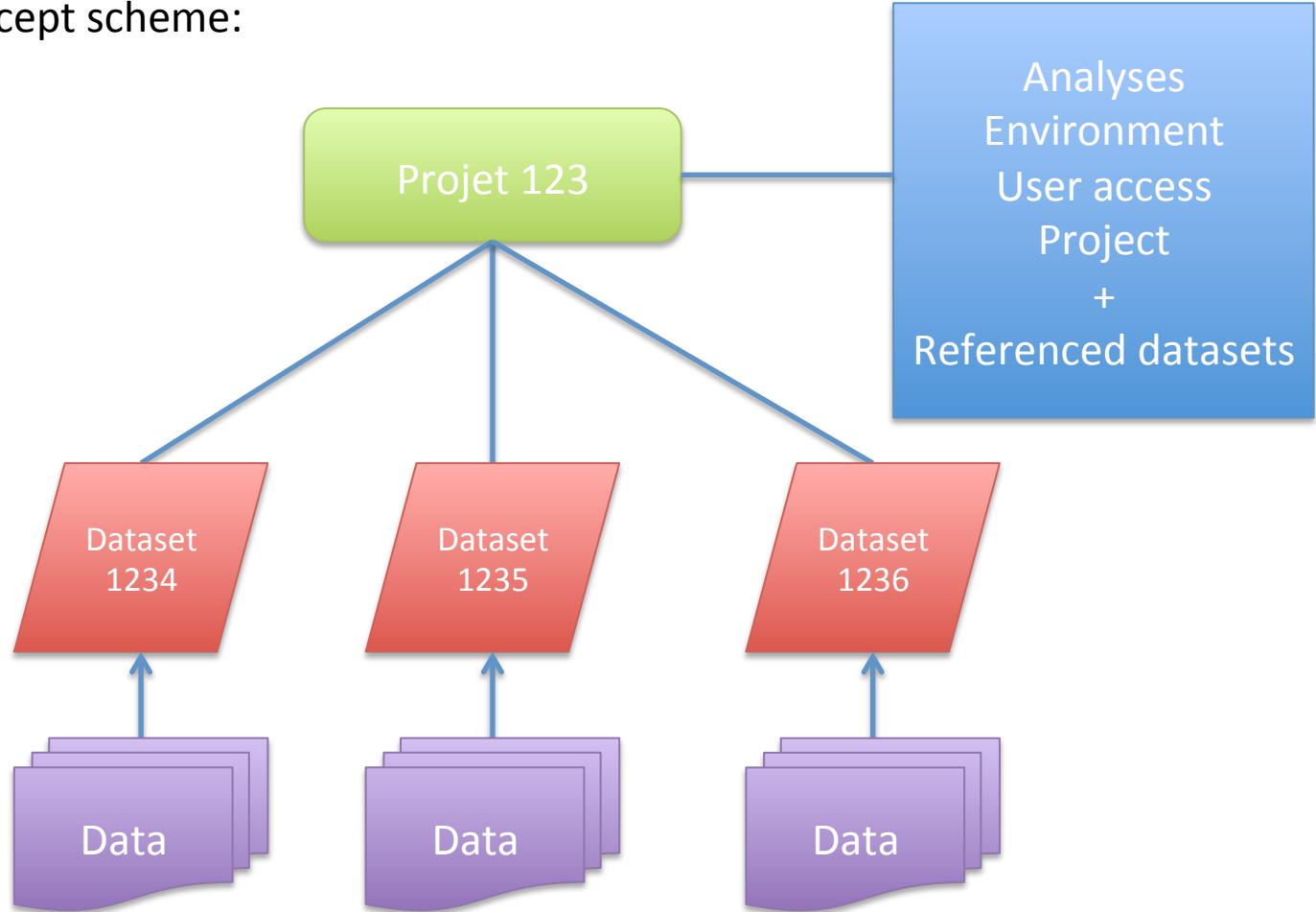
<sup>1</sup>Institut Curie, Paris, F-75248 France, <sup>2</sup>INSERM, U900, Paris, F-75248 France, <sup>3</sup>Mines ParisTech, Fontainebleau, F-77300 France, <sup>4</sup>INSERM, U830, Paris, F-75248 France

- 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



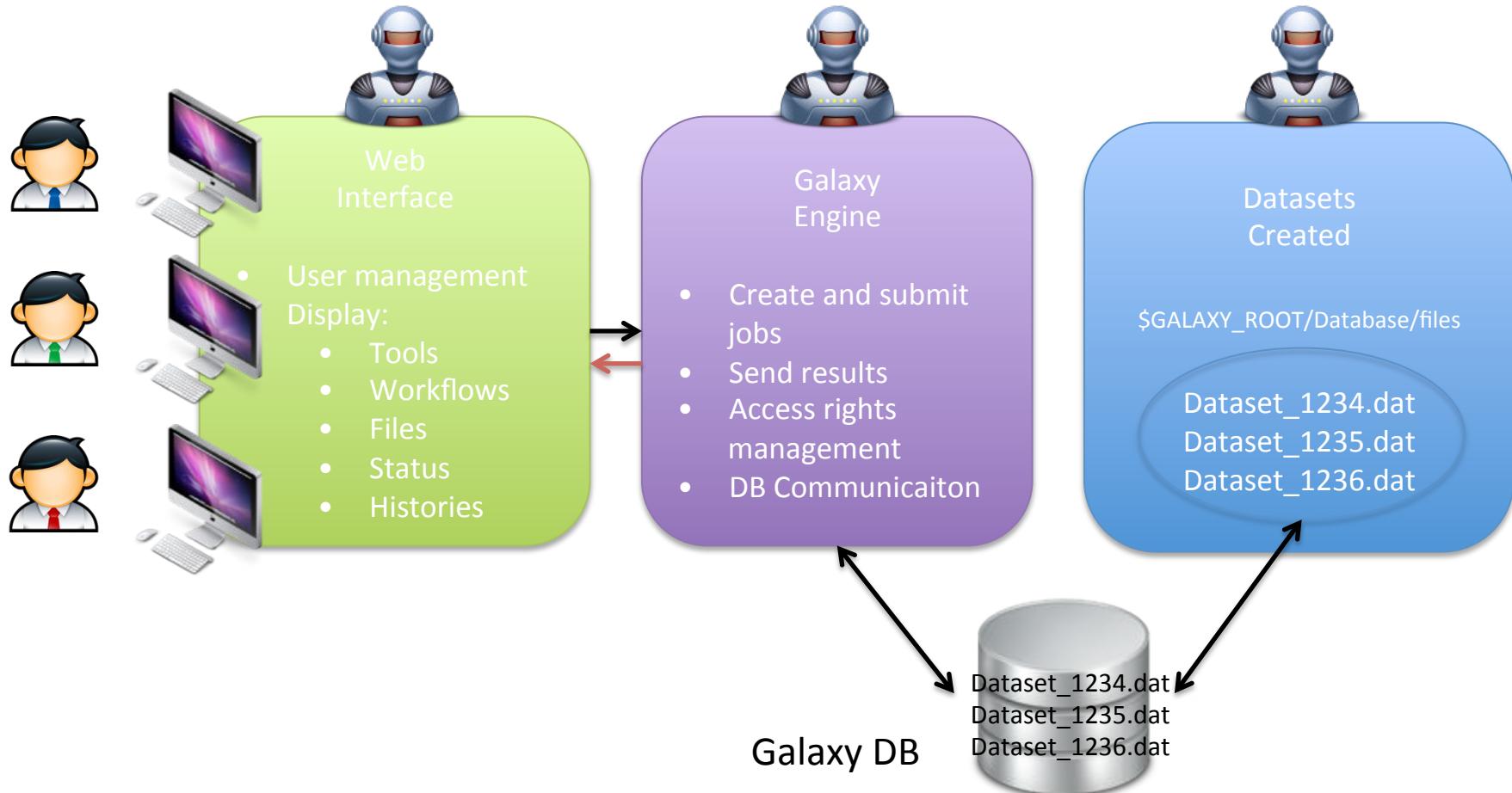
# KDI – Institut Curie Information System

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



# 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



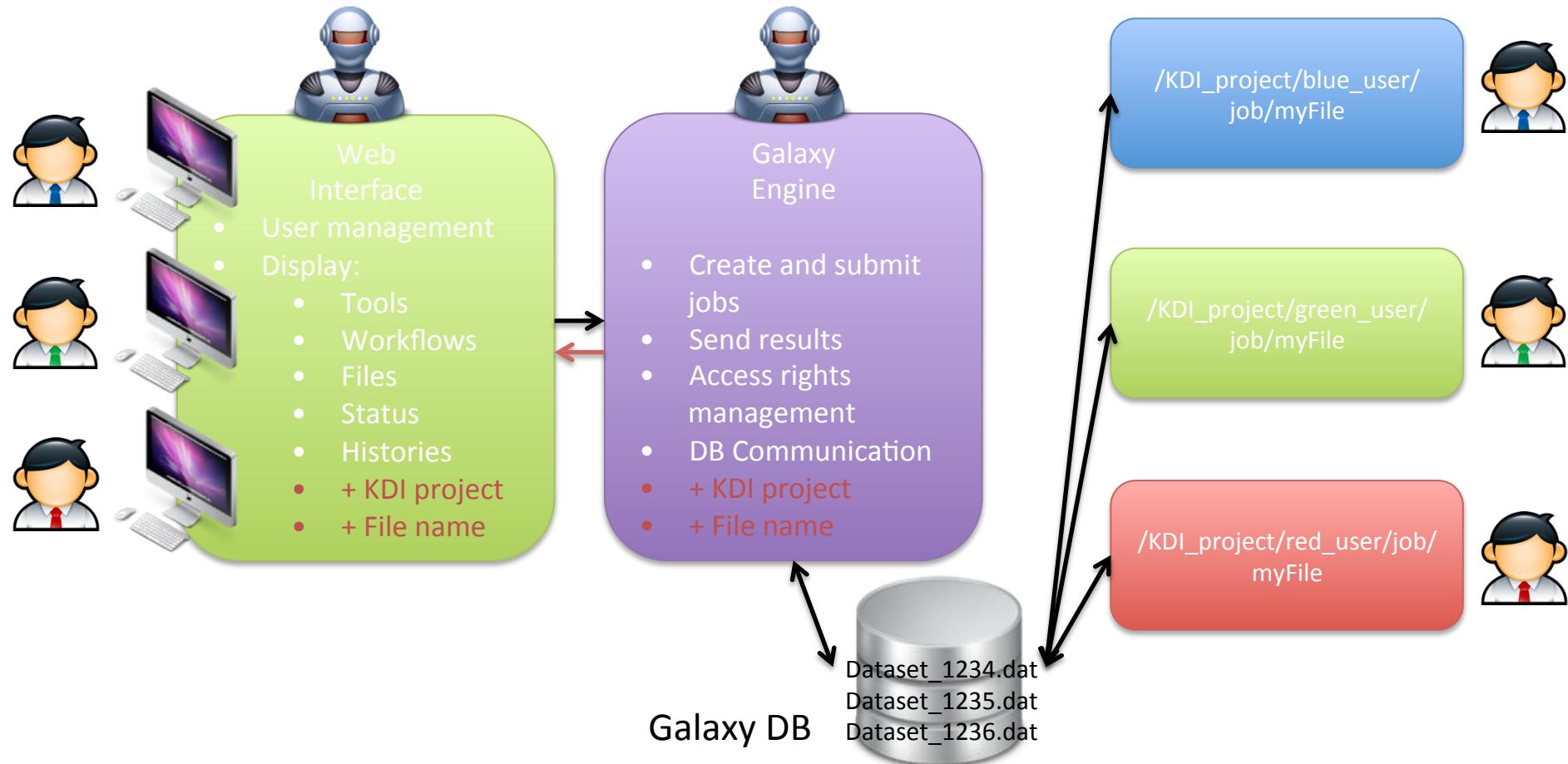
# Needs for Galaxy Integration into KDI

- 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)

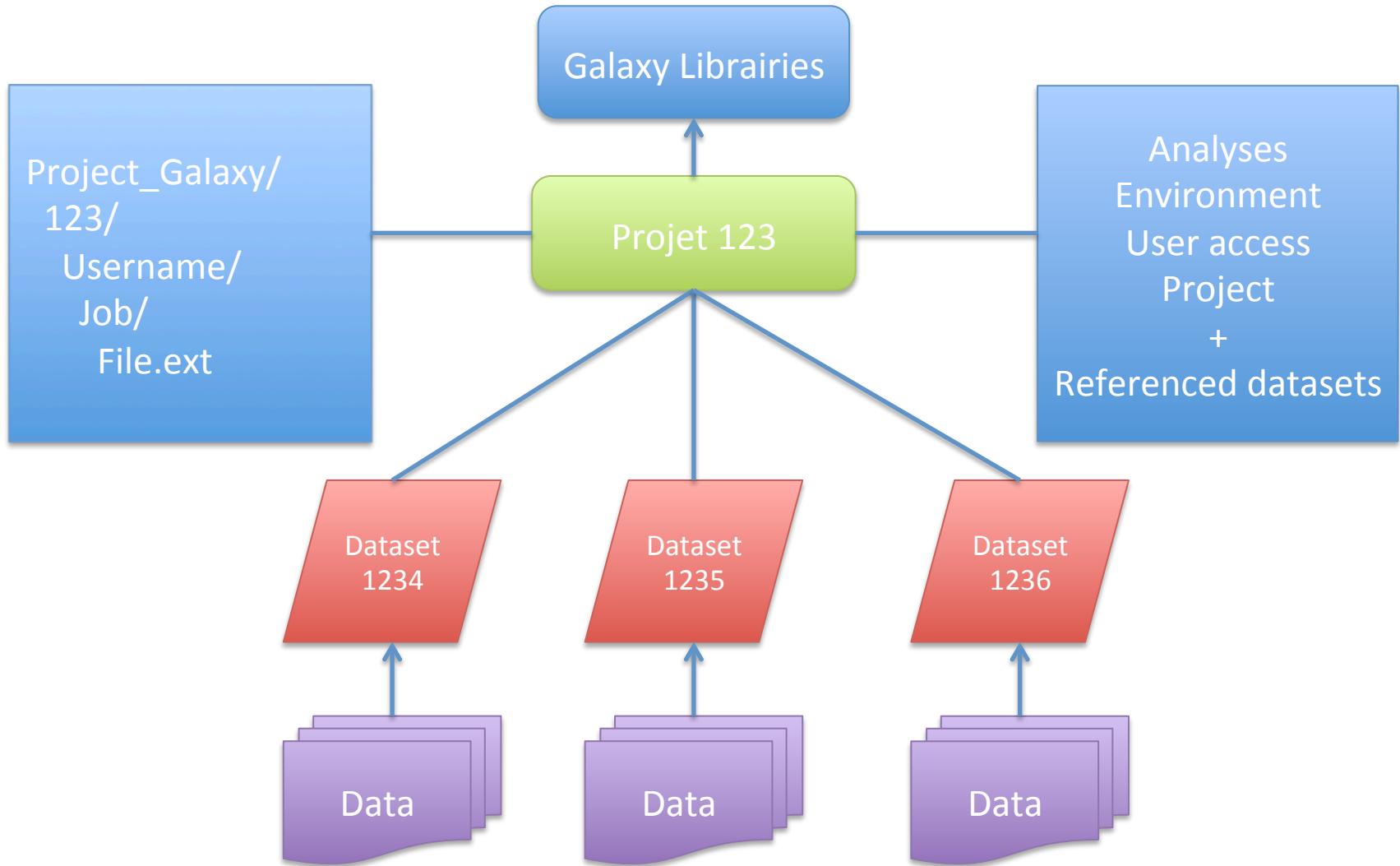
# What I have done

- 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



# Galaxy Integration into KDI scheme

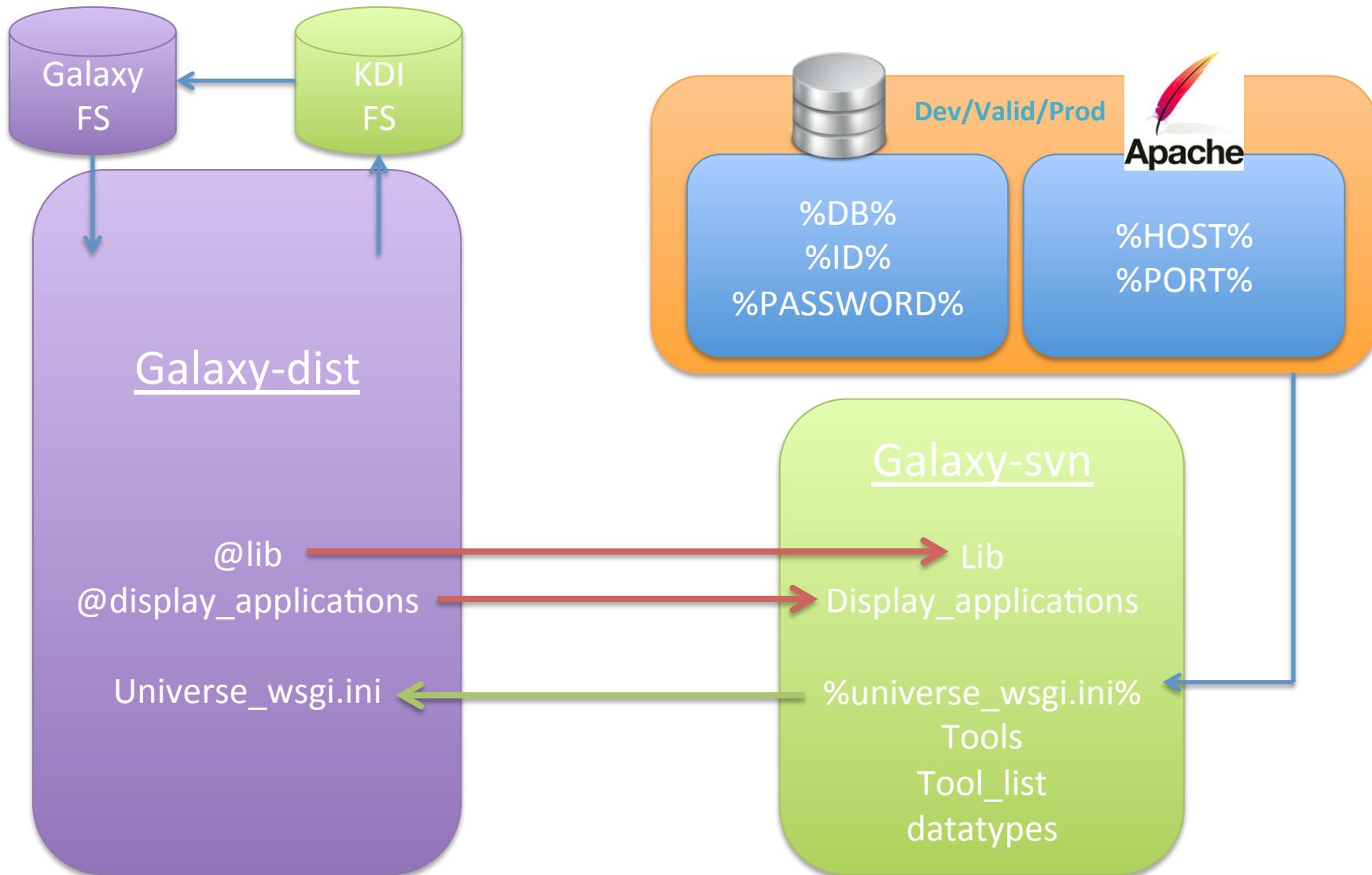


# LIVE DEMO

# 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



# 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

