



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^{1,2,3*}, Alban Lermine^{1,2,3}, Camille Barette^{1,2,3}, [Christel Guillouf](#)^{1,4} and Emmanuel Barillot^{1,2,3}

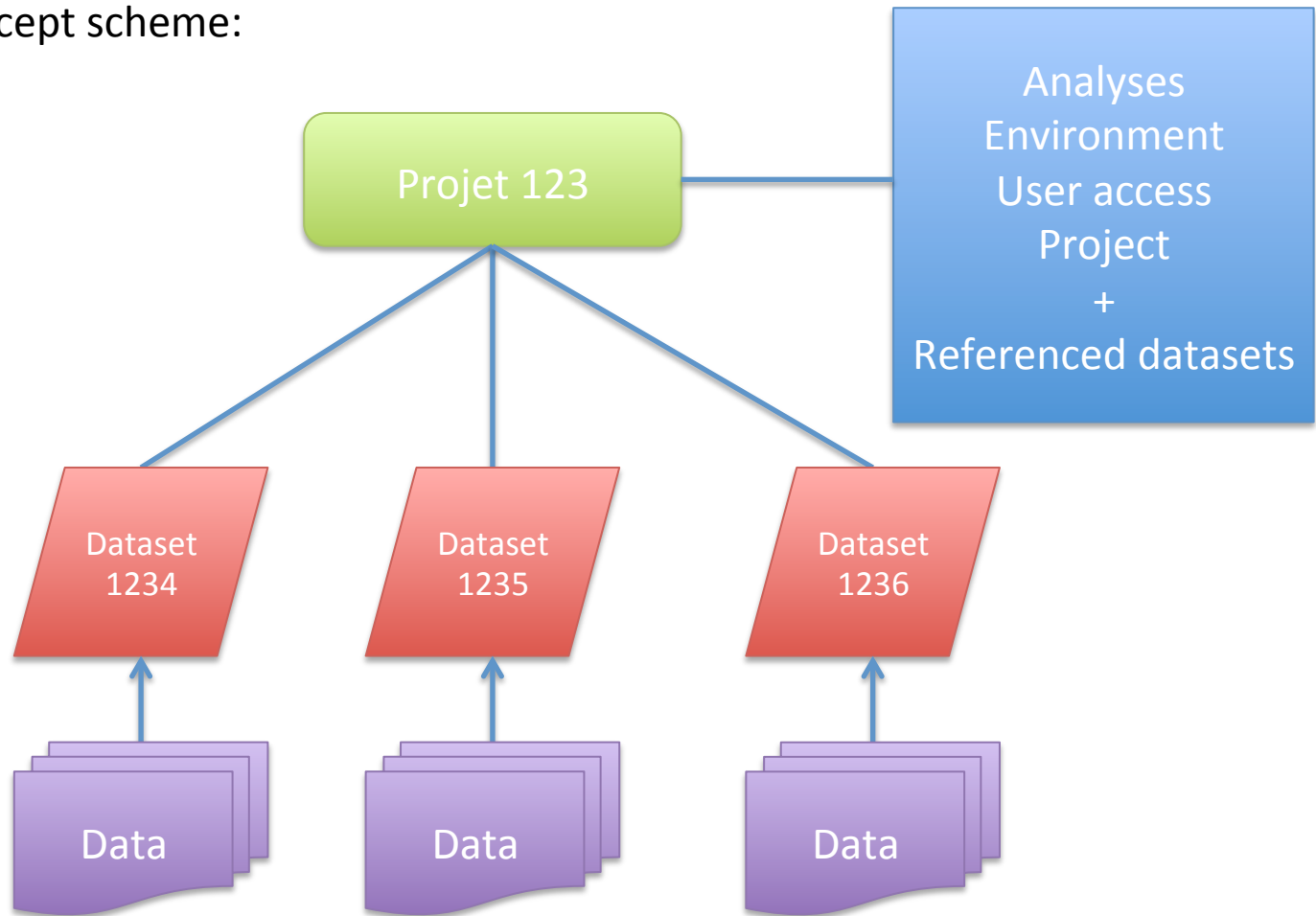
¹Institut Curie, Paris, F-75248 France, ²INSERM, U900, Paris, F-75248 France, ³Mines ParisTech, Fontainebleau, F-77300 France, ⁴INSERM, U830, Paris, F-75248 France

- 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



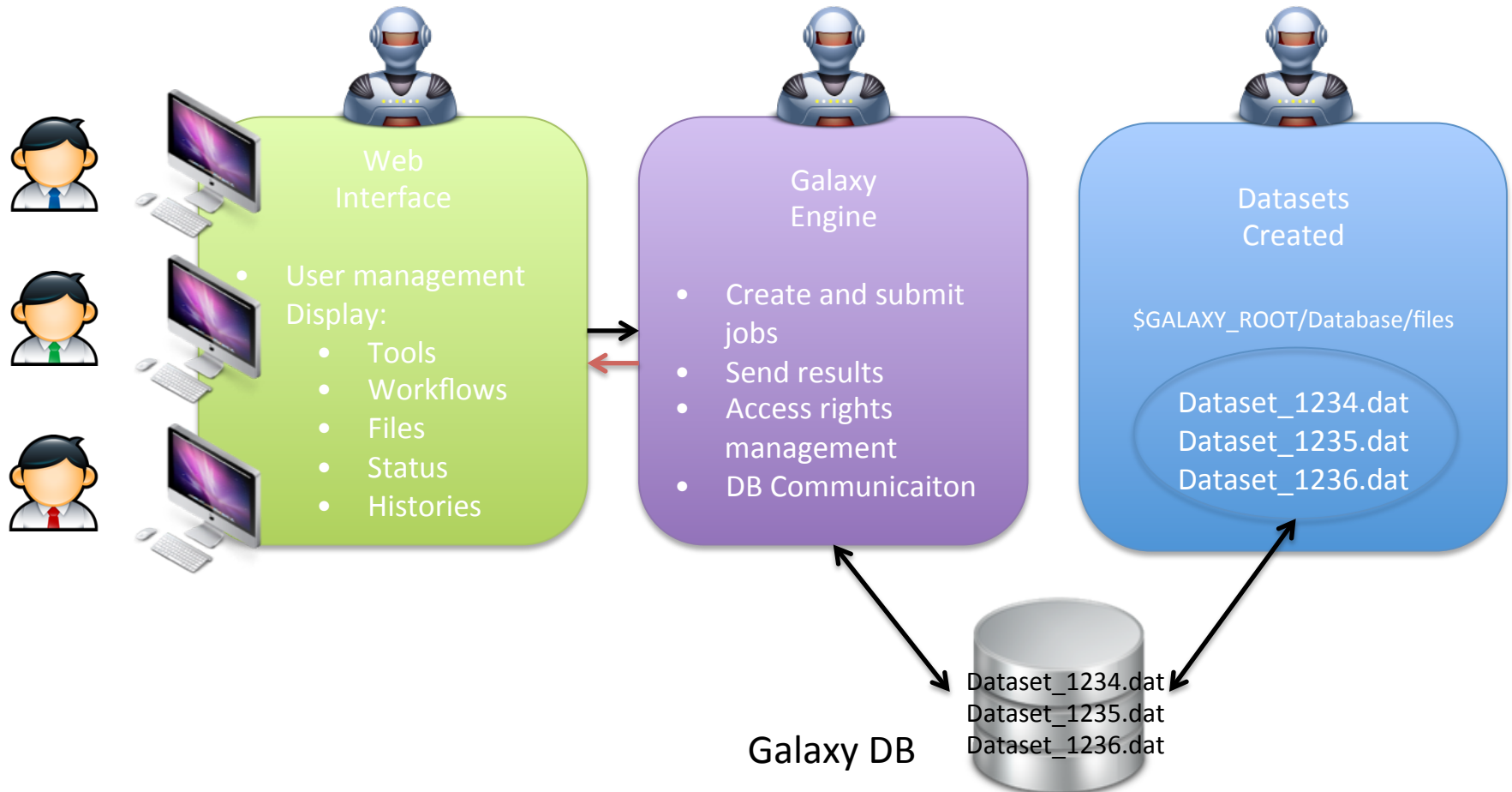
KDI – Institut Curie Information System

- 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



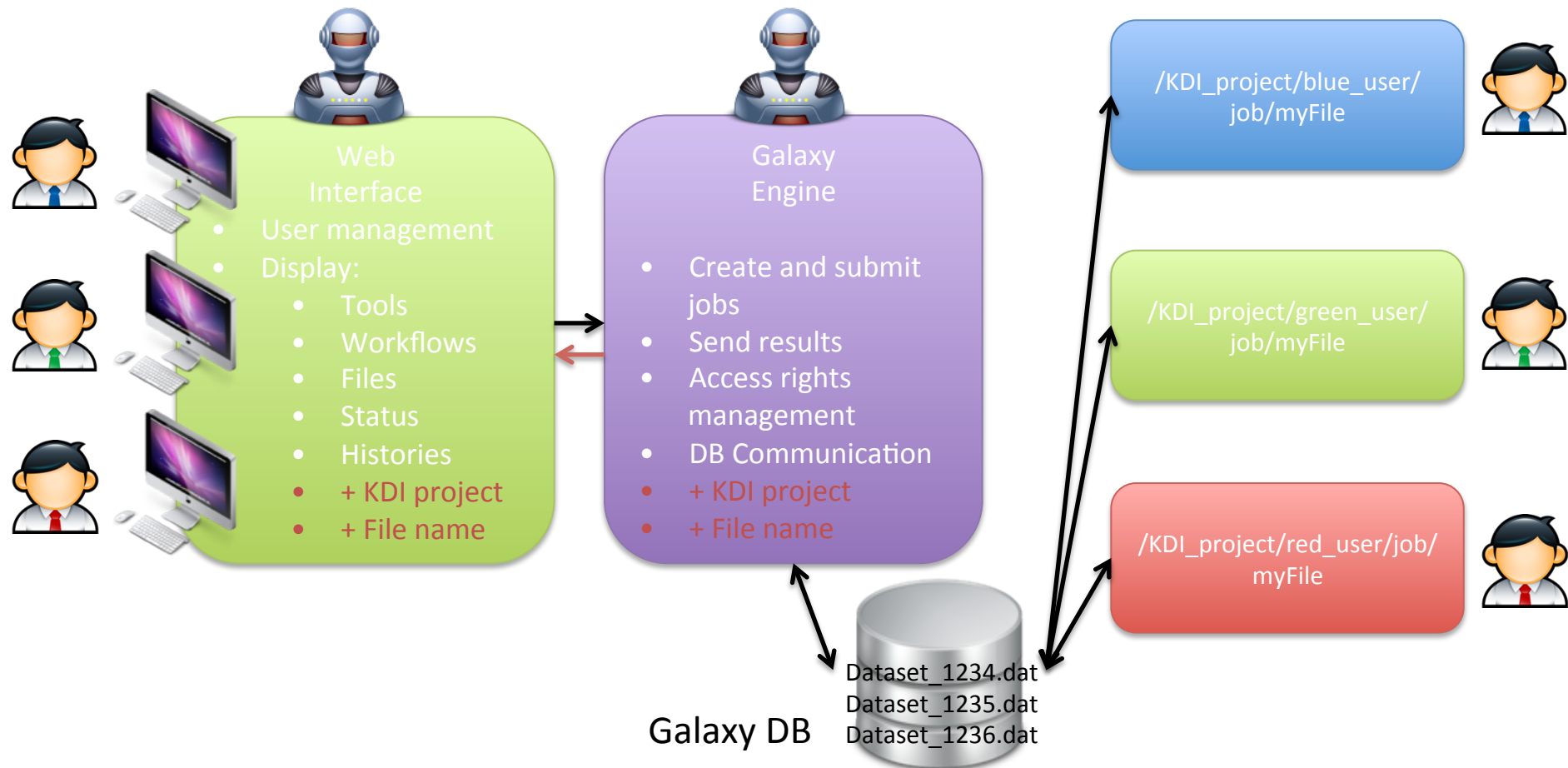
Needs for Galaxy Integration into KDI

- 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 whitout using network upload (avoid data copy)

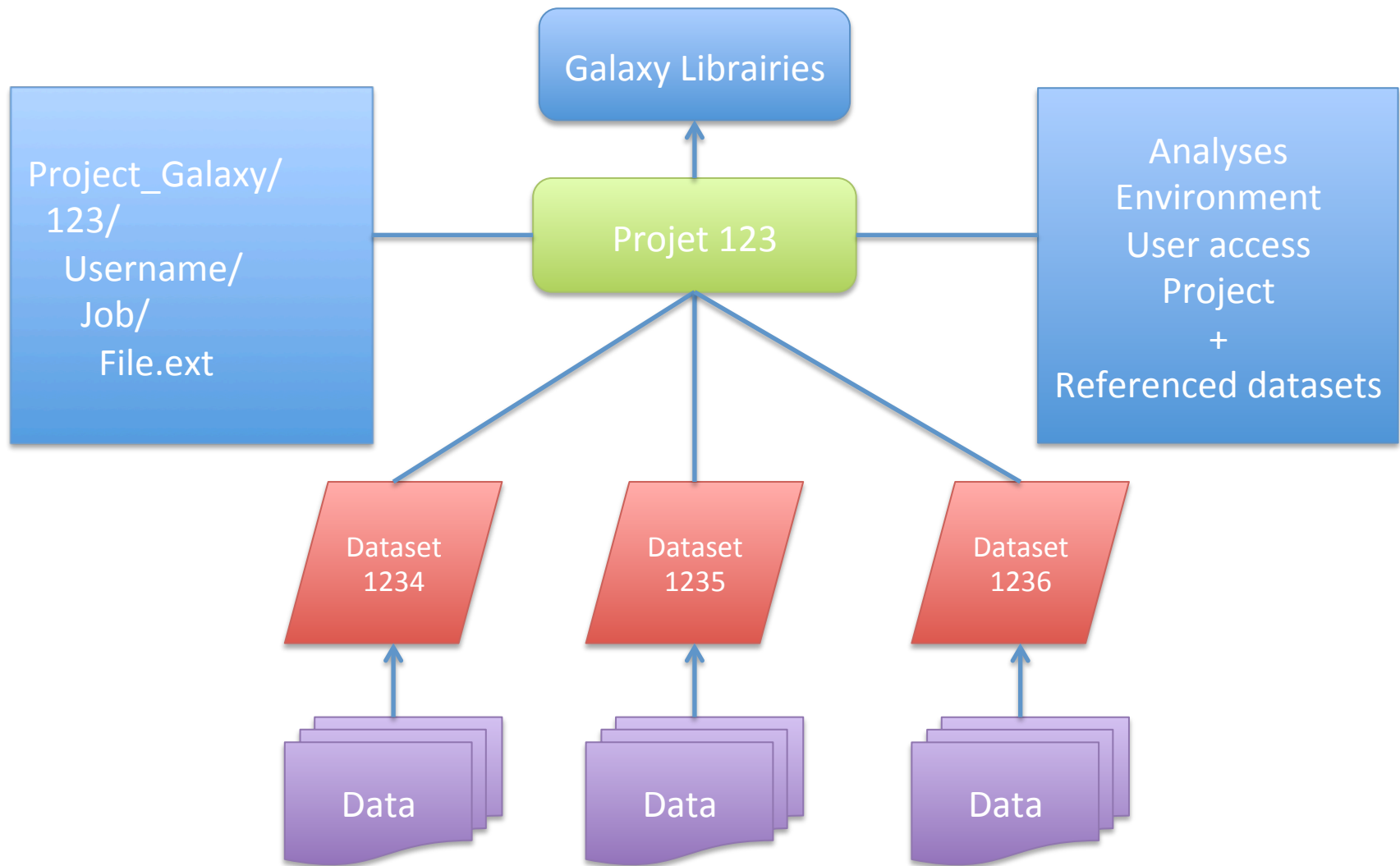
What I have done

- 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



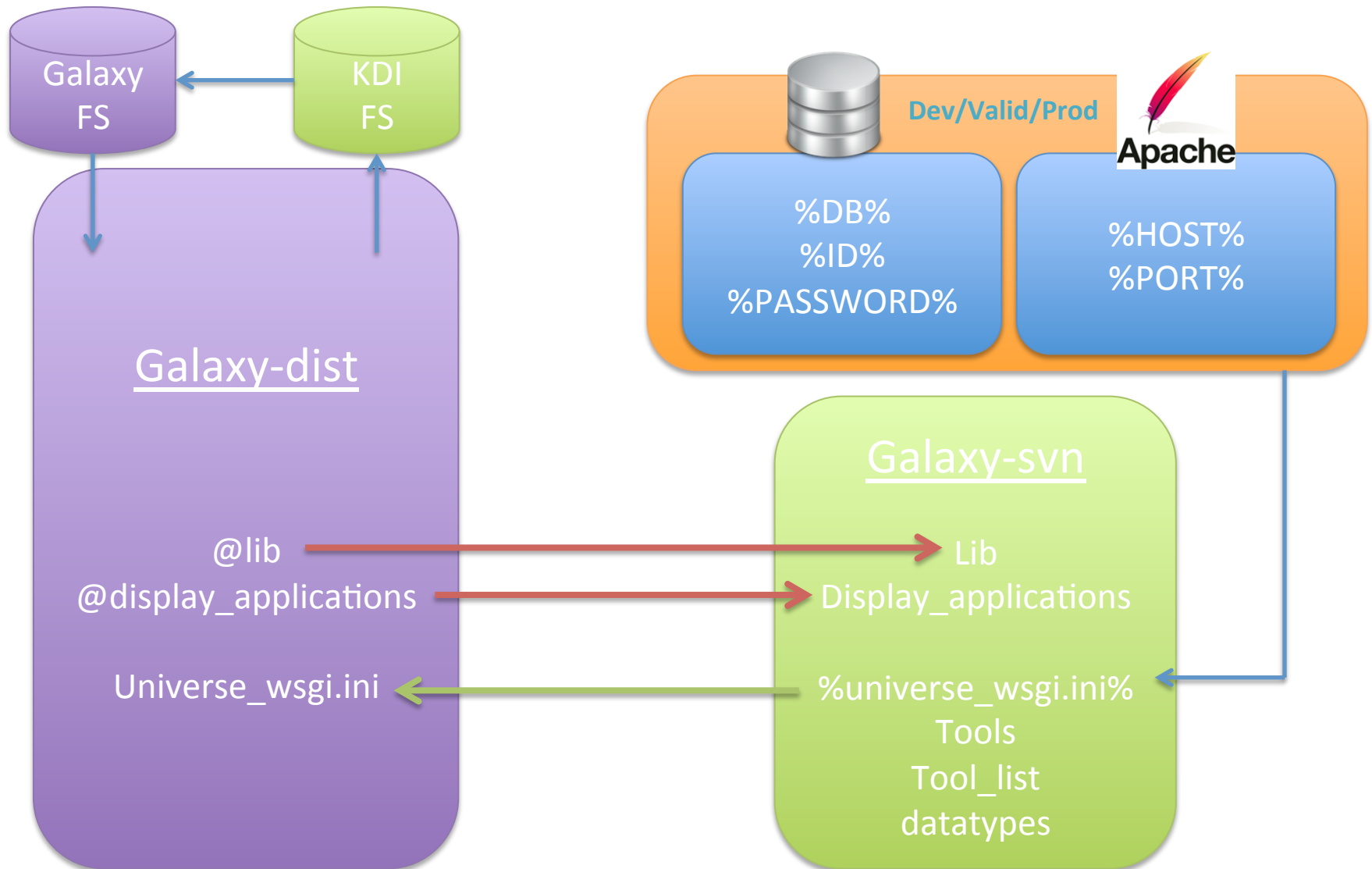
LIVE DEMO



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

