SGBC
SLU Global Bioinformatics Centre

SLU
Swedish University of Agricultural Sciences

Introduction to the eB3Kit

B3AFRICA - WORKPACKAGE 4

Erik Bongcam-Rudloff
Rafael Hernández de Diego
rafahdediego@gmail.com
2016-04-25

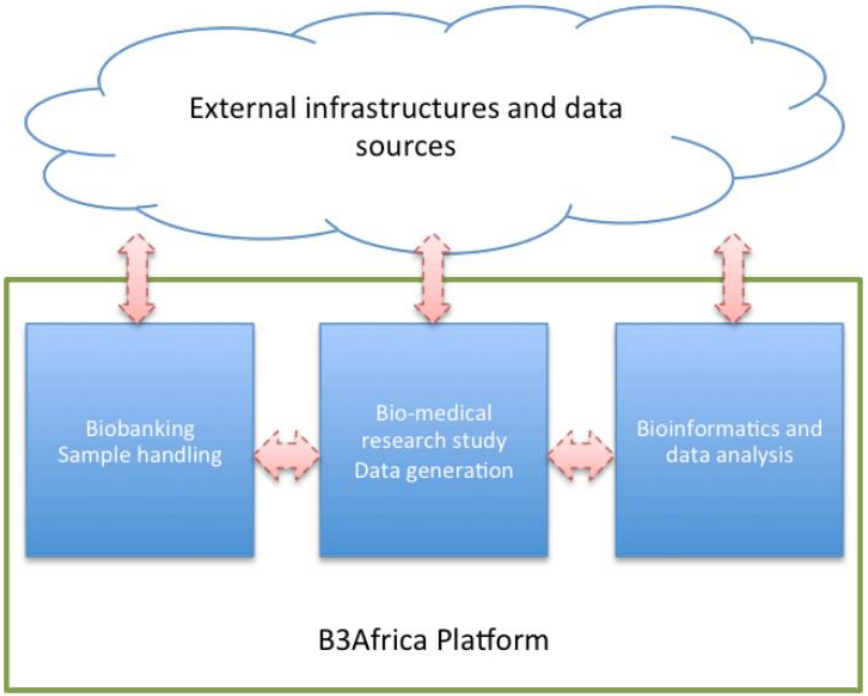
CONTENTS

- Introduction to the system
- Main concepts
- System overview
- Hardware
- Conclusions

EB3KIT: THE B3AFRICA PLATFORM

Main objectives:

- Integrate available open-source software, services and tools for biobanking, bioinformatics, ethics and regulations, and training.



SLU B3Africa SGBC SLU Global Bioinformatics Centre

EB3KIT: INSTALLED SERVICES

Laboratory Information Management Systems (LIMS)

- *BiBBox*
- *BIKA LIMS*
- *OpenSpecimen*
- *LabKey*

Bioinformatics tools

- *Galaxy*: web-based platform for data-intensive biomedical research. Includes many bioinformatics tools.

Experiment Management Systems (EMS)

- *STATegra EMS*

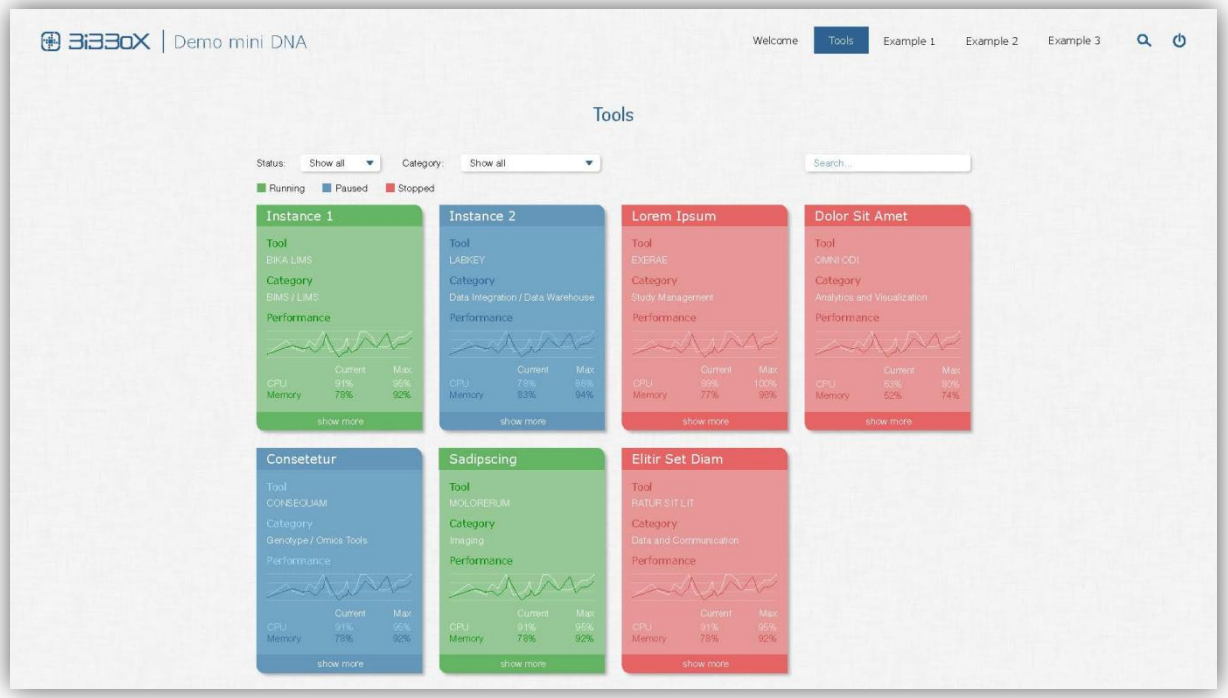
… and many more! LifeRay, PhenoTips, Molgenis,

EB3KIT: SYSTEM OVERVIEW

- **3 main concepts:**
  - The system uses the **BiBBoX platform to integrate all services** and provide an unified user experience.
  - **All services** running in the eB3Kit are "**dockerized**".
  - **All files uploaded** or produced by users are controlled using **iRODS**. iRODS allows the storage and sharing of large data sets, with the ability to tag/add meta data.

EB3KIT: BIBBOX

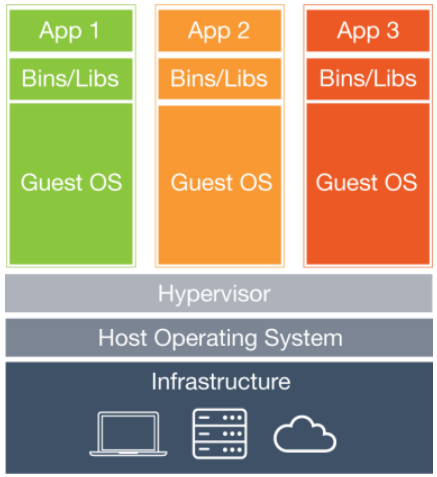
- **Unified user interface for all services.**
  - Provides access to all installed services
  - Service management, user and session management, …



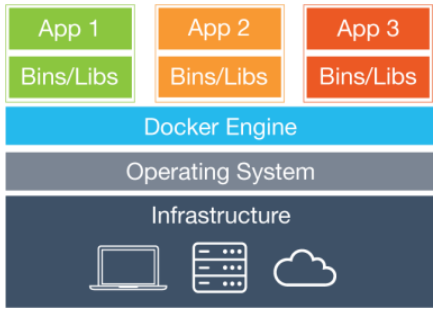
SLU B3 Africa SGBC SLU Global Bioinformatics Centre

EB3KIT: DOCKERS

- A virtual machine "*reduced to the minimum*"
- Docker containers wrap up a piece of software in a complete filesystem that contains everything it needs to run: code, runtime, system libraries, etc.
- ***Virtual Machines* vs. *Dockers***
  - Containers include the application and all of its dependencies, but share the kernel with other containers.
  - They start instantly, use less space and make more efficient use of RAM.



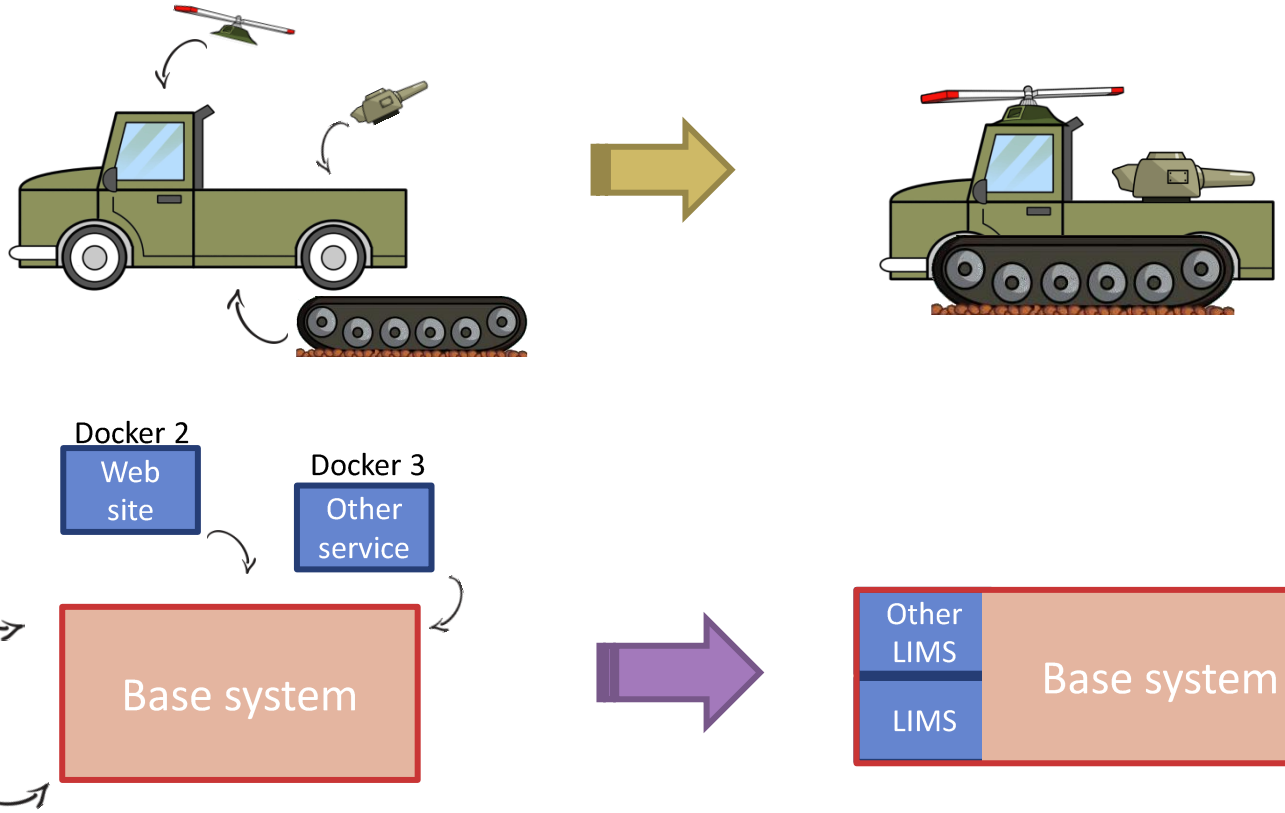
V.M.



Dockers

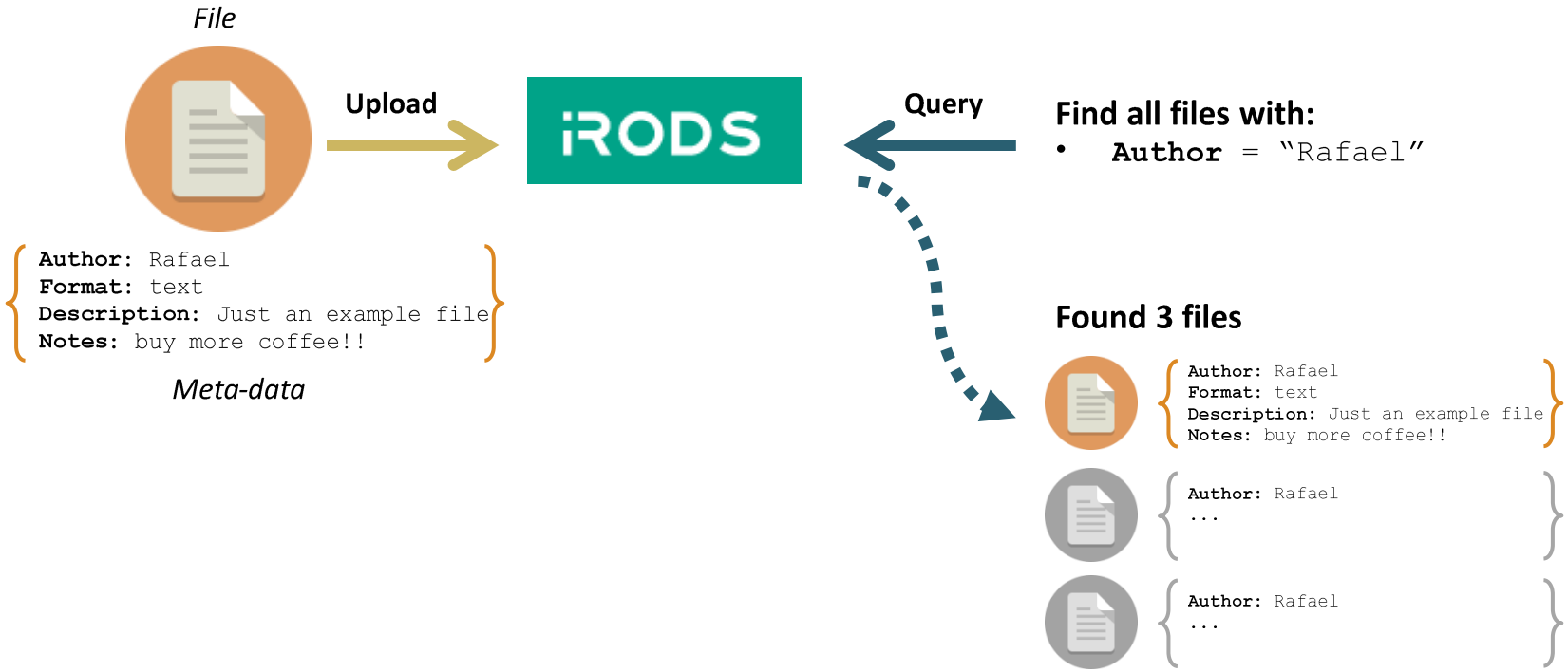
EB3KIT: «DOCKERIZED» SYSTEM

- **Using dockers we can easily add/remove tools and services to the system**

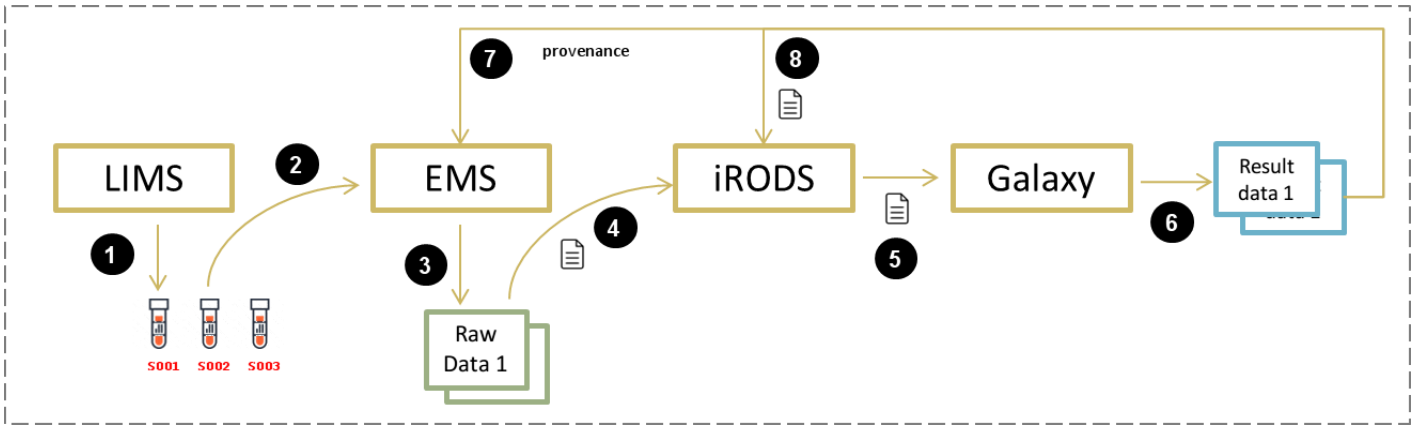


EB3KIT: IRODS

- Storage of large datasets + meta data.
- Includes tools for querying the system, file permissions, data virtualization, APIs for different programming languages,…



EB3KIT: SYSTEM OVERVIEW



1. Information for samples is stored in the different LIMS systems
2. Experiments (studies) are registered in the EMS.
3. Raw data (e.g. a FASTQ file obtained after sequencing a sample) are registered in the EMS (meta-information) for an experiment.
4. The files for the Raw data are stored in iRODS keeping the minimum meta-data to identify the associated entries in the EMS and the LIMS.
5. Galaxy can import files from iRODS for analysis
6. The user analyses the loaded data and produces some result files.
7. Galaxy can save the results in the system. For each result file, the description for the analysis workflow is stored in the EMS, and the files are stored in iRODS.

EB3KIT: HARDWARE

Mac Pro

Hardware configurable

- 4-8 Cores Intel Xeon processor (3,7 GHz)
- 12 – 64 GB RAM
- Flash storage 256GB – 1TB
- 6 Thunderbolt port, wireless, …
- Portability: 25 cm height, 16,7 cm ∅, ~ 5kg

LaCie 2big Thunderbolt™ 2

- 6TB – 16TB
- Speeds up to 360MB/s, 7200RPM
- Dual Thunderbolt 2 ports, 1 USB 3.0 port
- Portability: 11cm x 22cm x 13cm, ~ 3kg

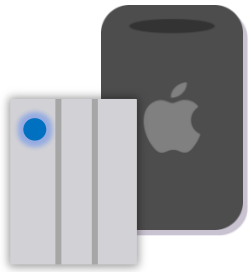
SLU B3 Africa SGBC SLU Global Bioinformatics Centre

EB3KIT: SCALABILITY

- **Configurable hardware**
- **Virtualization (dockers):** installable in other platforms (future).

Performance

Cost



Installation in servers

- Running as a virtual machine in the Centre's infrastructure

Advance configuration

- 8 Cores Intel Xeon processor
- 64 GB RAM
- 1TB Flash storage
- 16TB External storage1

Basic configuration

- 4 Cores Intel Xeon processor
- 12 GB RAM
- 256GB Flash storage
- 6TB External storage

SLU B3 Africa SGBC SLU Global Bioinformatics Centre

EB3KIT: CONCLUSIONS

- Many services installed as a modular system using dockers
- Unified entry-point using BiBBox
- Simple management and maintenance
- Powerful, Portable and affordable hardware

QUESTIONS?

BRIDGING BIOBANKING AND BIOMEDICAL RESEARCH
ACROSS EUROPE AND AFRICA