



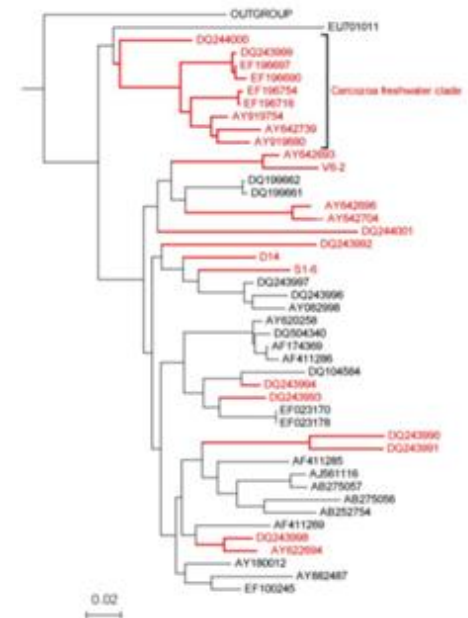
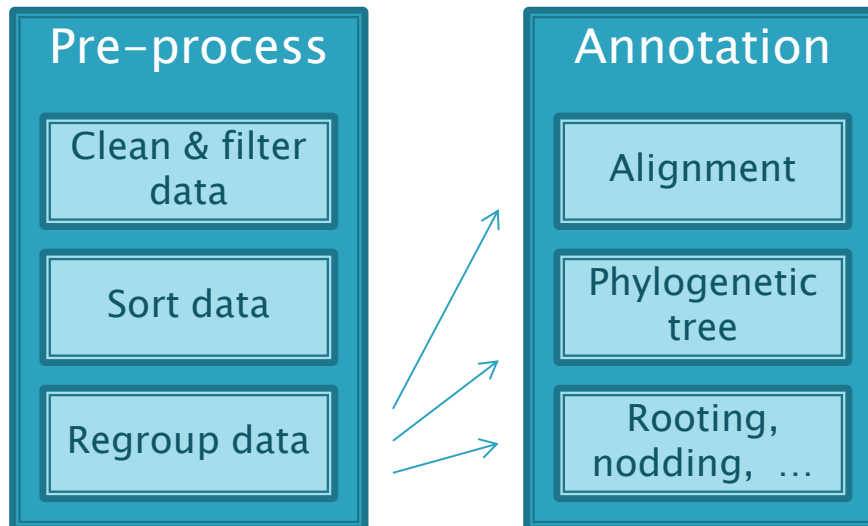
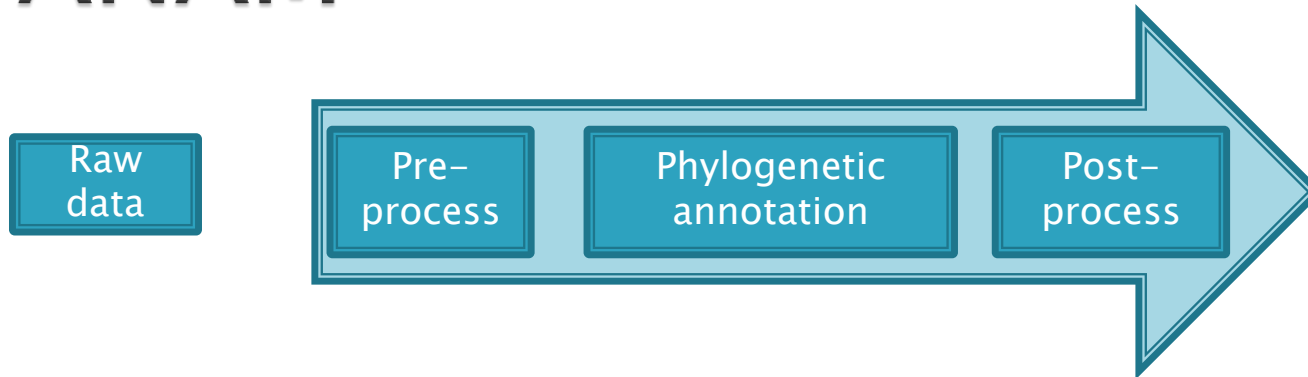
Gridification of PANAM, a phylogenetic pipeline for high throughput sequencing analysis, using the DIRAC framework

– DEMO –

By Doan Trung Tung et al.
EGI Technical Forum 2013



PANAM

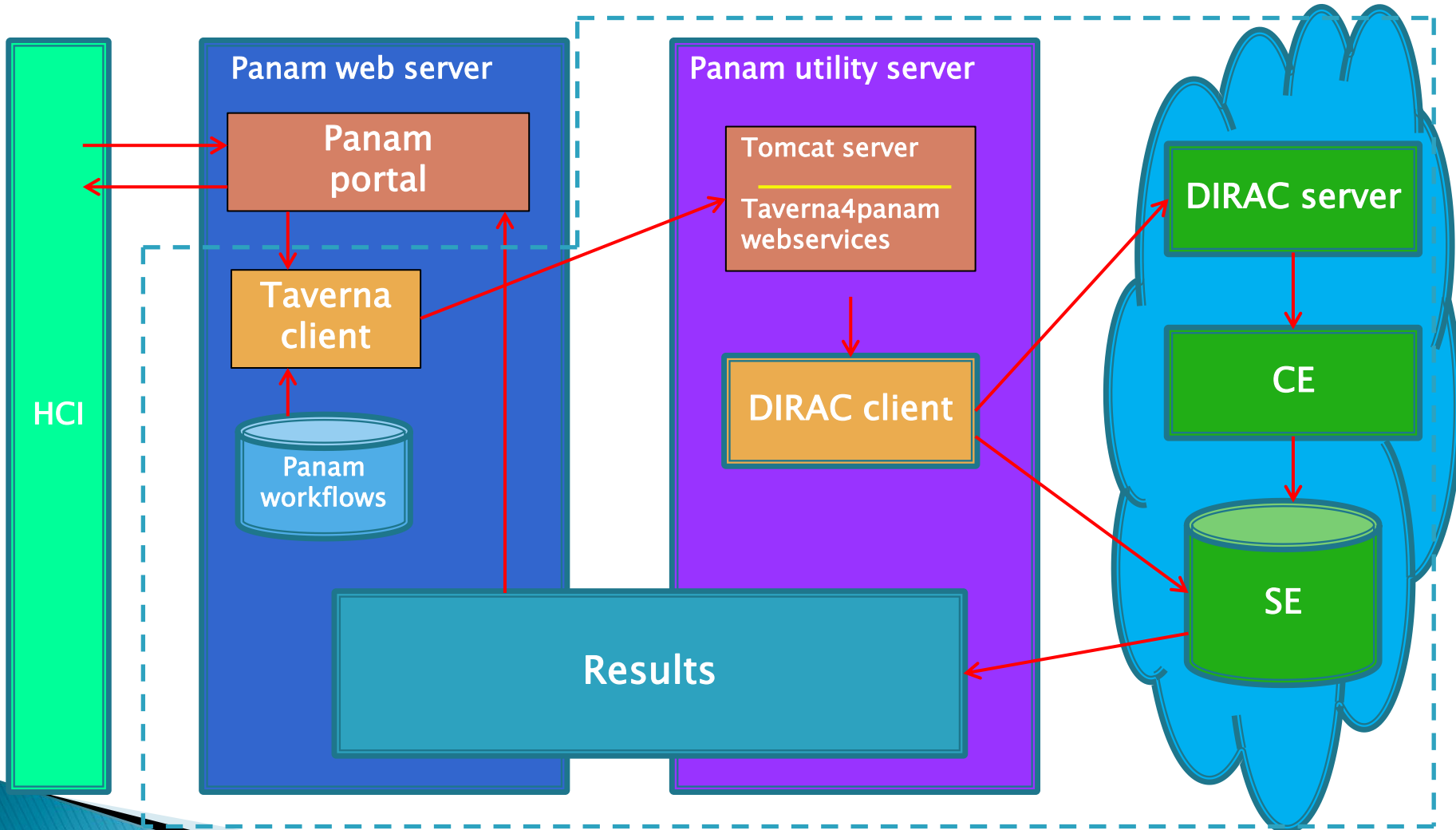


Gridification PANAM (1)

- ▶ DIRAC
 - Grid platform for WMS and DMS
 - A service provided by French Grid Initiative
- ▶ Taverna
 - Created by Univ. Manchester
 - Very common workflow engine, especially for bioinformatics
 - Support: webservices, command-line, desktop tool, etc.
- ▶ VO biomed
 - A large scale international and multi-disciplinary VO supporting Life Sciences communities
 - Supported by many different sites, delivering access to tens of thousands CPU cores to its users



Gridification PANAM (2)



Deployment (1)

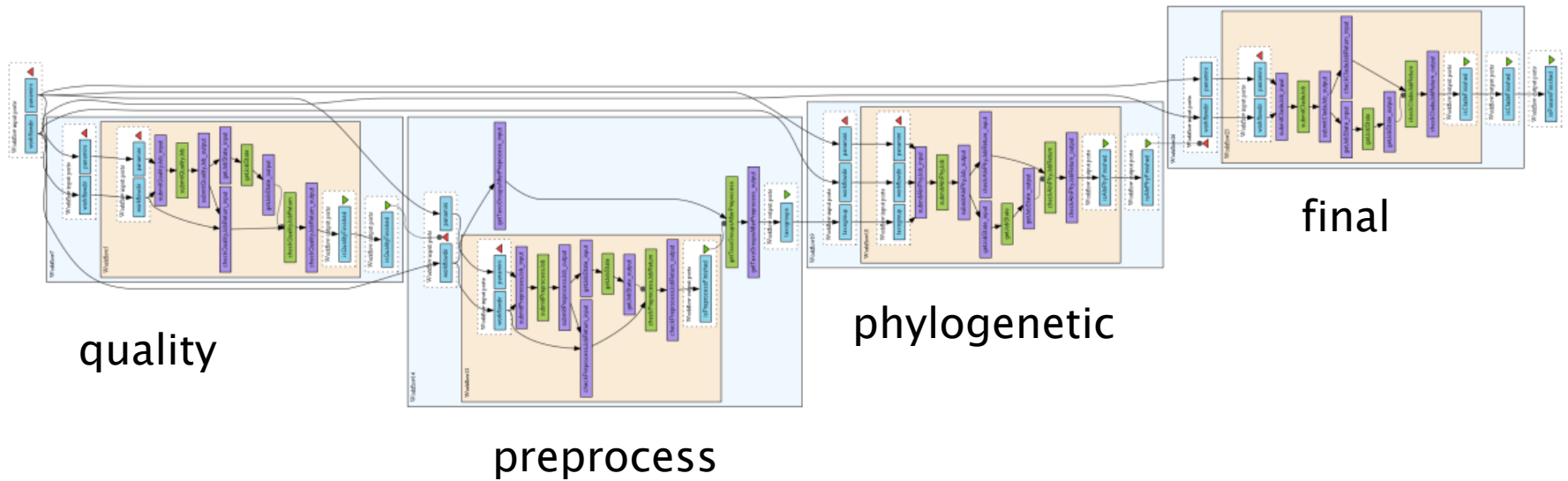
- ▶ File structure on the grid
 - /biomed/-----/panam/
 - inputs
 - Reference.tar.gz
 - TestNNN-pre.tar.gz, TestNNN-qual.tar.gz
 - outputs
 - TestNNN
 - final.tar.gz
 - profiles
 - Profiles_NSF573_NSR951.tar.gz
 - tools
 - qualitytool.tar.gz / preprocesstool.tar.gz / alnphytool.tar.gz

Deployment (2)

▶ Taverna Webservices

- <http://auvergridpf03.univ-bpclermont.fr/PanamWS/panamws?wsdl>
- submitQualityJobs / submitPreprocessJobs / ...
- checkQualityJobsReturn / ...
- getJobState
- getTaxoGroupsAfterPreprocess

Deployment (3)



Running PANAM

- ▶ Prepare data
 - Upload to `/biomed/.../panam/inputs/`
- ▶ Execute workflow
 - Using taverna desktop tool
 - Using taverna command line
- ▶ Monitor the workflow
 - Using taverna desktop tool
 - Using DIRAC site
- ▶ Get the result
 - `/biomed/.../panam/outputs/TestNNN`