

Design your e-infrastructure!

Use case: European Research Initiative in CLL (ERIC)

Proposers:

Kostas Stamatopoulos (*Greece*)

Anastasia Chatzidimitriou (*Greece*)

Paolo Ghia (*Italy*)

Fotis E. Psomopoulos (*Greece*)



Background and Users



www.egi.eu

EGI-Engage is co-funded by the Horizon 2020 Framework Programme of the European Union under grant number 654142



Who is the community/project the use case belongs to?

- ERIC (www.ericll.org) is a European organization devoted to improving the outcome of patients with chronic lymphocytic leukemia (CLL) and related diseases.
- ERIC is a Working Party of the European LeukemiaNet and a Scientific Working Group of the European Hematology Association.
- ERIC is open to all physicians and scientists working or interested in the field of chronic lymphocytic leukemia and related disorders.

Who will be the user? Can the users be characterised? How many are they?

- Scientists (physicians, biologists, bioinformaticians, biostatisticians, other) involved in research on CLL and related disorders
- ERIC has more than 680 members representing more than 46 countries.
- ERIC members are also active in other European and international scientific societies and consortia (e.g. European Haematology Association, European LeukemiaNet American Society of Hematology, Euroclonality-NGS, iwCLL), with large lists of members/partners who can be considered potential users.

What's your role in the use case? Any experience/link to EGI?

- Currently a board member of ERIC
- No formal involvement with EGI at the project level, however
 - Extensive experience as users of the EGI infrastructure as a Research Group (users of Chipster and Grid services)

What will be the value of the infrastructure for them? What will the system exactly deliver to them?

- Currently ERIC involves a loosely joined effort of geographically distributed computational efforts in CLL (services and datasets)
- Particular value of the system (and immediate step)
 - Persistent sensitive data repository (patient and clinical trial data)
 - Computational analysis services around the repository
- e-infras can facilitate the construction of a scalable virtual environment that will support:
 - Scalable, modular and persistent data repository
 - Biobanking data
 - NGS data
 - clinicobiological data
 - publication and guidelines
 - best practices guidelines
 - Centralized and scalable computational services

How should they use the system?

- Depending on user expertise:
 - non tech-savvy users: access to data and services through web-interface
 - data producers: command-line access to resources
 - project managers: orchestration of services towards specific solutions

What's the timeline for development, testing and large-scale operation? (Consecutive releases can/should be considered.)

- ERIC currently includes several mature services, however
 - only local infrastructure when available
 - No scalability
- Q1 2017: Development of the e-infra component
- Q3: 2017: Testing
- Q1: 2018: Large-scale deployment

Design and implementation plan



www.egi.eu

EGI-Engage is co-funded by the Horizon 2020 Framework Programme
of the European Union under grant number 654142



What should the first version include? - The most basic product prototype imaginable already bringing value to the users

(the so-called Minimal Viable Product - MVP)

- First version would be a virtual research environment on a single site that would allow:
 - Integration of the existing distributed services into a single end-point
 - Central repository of datasets
- Indicative KPIs:
 - Over 5 computational services that offer data analysis processes
 - Over 5 datasets that are currently across different sites
 - 1 integrated service combining 3 or more of the existing services on extended datasets

Which components/services already exist in this architecture?

- Several databases available to the community:
 - ERIC minimum dataset project
 - ERIC TP53 database
 - IMGT/CLL-DB
- Several tools available through ERIC:
<http://www.ericll.org/pages/services>
 - IGHV gene mutational analysis
 - MRD diagnostics
 - ARResT/AssignSubsets

Which components/services are under development (and by who)?

- An effort towards Data Integration is currently under way, involving the majority of ERIC members.
 - A centralized computational infrastructure would greatly facilitate this effort

Plans for the workshop



www.egi.eu

EGI-Engage is co-funded by the Horizon 2020 Framework Programme of the European Union under grant number 654142



What questions would you like to get answered today?

- Feasibility of constructing a dedicated ERIC VRE
- Possibility of combining multiple e-infrastructures:
 - EGI: computational aspects
 - EUDAT: data services
 - OpenAIRE: Best practices/guidelines/publications

What issues you would like to solve today?

- Technical aspects of transitioning existing services and datasets onto EGI's computational infrastructure
- User authentication aspects
- Data Privacy aspects

Other outcomes that you would like to get out from the workshop?

- Connection across multiple e-infrastructures?