



APPLIED STATISTICS WORKSHOP

Karim CHINE,
Cambridge, United Kingdom

*"Towards a Federative, Collaborative, User-Centric and
Grid-Enabled Computational Open Platform"*

Friday, November 7, 2008

14:30

Room : c 115 (STAT)

Abstract

R is becoming the lingua franca of data analysis and statistical computing. It has a very powerful graphics system as well as cross-platform capabilities for packaging any computational code. Hundreds of available R packages implement the most up-to-date computational methods and reflect the state-of-the-art of research in various fields. R packages are foreseen as a reproducible research enabler. There is no obstacle to a large-scale deployment of R on public grids since it is a GPL software. However R is not multithreaded, doesn't operate as a server and has only a low-level non-object oriented API. GUIs development for R remains nonstandardized. R's potential as a computational back end engine for applications has yet to be fully exploited. While its user base is growing at a high rate, this growth rate would be significantly higher in the presence of a userfriendly and rich workbench Biocep is a general unified open source solution for integrating and virtualizing the access to R engines/servers and aims to become a federative user-friendly computational e-platform for research, finance and education. The Biocep virtual workbench enables the plugability for all the elements of a computational environment: the computational resource whether it is a local machine, a cluster, a grid or a cloud server via a simple URL, the computational components via the import of R packages and the computational GUIs via the import of plugins from repositories or the design of new views with a drag&drop GUI editor. Several dockable built-in views allow users to work interactively with R engines running anywhere. The views include a console, highly interactive remote graphic devices, a workspace explorer, PDF and SVG viewers, R data inspectors, linked plots and spreadsheets fully integrated with R functions and data. Biocep enables collaborative R sessions: multiple web users can connect simultaneously to an R server running anywhere and analyse data collaboratively via a set of broadcasted views.

A Biocep based R virtualization infrastructure has been successfully deployed on the British National Grid Service and the result leaves no doubt on the usability and usefulness of such a service for researchers. If the new platform was adopted, it would greatly enhance the usability of existing HPC infrastructures and would increase their usage. It would be an important step forward in the direction of the interoperability, reusability and seamless integration of research resources and therefore a reproducible research enabler. Biocep may also work as an enabler of a new computing business model that would synergize the utility computing model (resources) and the pay-per-use software model (components/GUIs).

You are welcome to the coffee break after the seminar (room : c 105)

Visit our page at: <http://www.uclouvain.be/72906.html>