Making use of the GRID for Microarray Data Analysis

Rob Allan[2], Dan Grose[3], Mark Hayes[1], John Kewley[2] and Lorna Morris[1]

[1]Cambridge e-Science Centre, University of Cambridge

[2]e-Science Centre, CCLRC Daresbury Laboratory

[3]Centre for eScience, University of Lancaster

Email address of corresponding author: ljm59@cam.ac.uk

Microarray data analysis is notoriously resource intensive and requires the manipulation of large sets of data in a variety of different formats. There is a need for analysis to exploit multi-node computer systems to speed up processing, enable calculations to be performed more quickly and aid seamless automation of performing multiple sub-tasks. However many developers and users of these applications are not ideally catered for by the current Grid middleware and usually lack the programming skills to use the current infrastructure to the full.

The Grid Technology Group at CCLRC Daresbury Laboratory have created a prototype library called GROWL: *Grid Resources on Workstation Library*. GROWL is a lightweight client toolkit that is easily installable, yet provides extensible access mechanisms to Grid resources, A client-server model is used to interface to existing Grid Services via Web Service calls. GROWL is currently being further developed as part of the JISC-funded Virtual Research Environment (VRE) Programme by a collaboration from CCLRC Daresbury Laboratory and the Universities of Lancaster and Cambridge.

The three institutions are focusing on three different user communities: Chemistry and Physics at the Daresbury Laboratory, Social Sciences at the Univeristy of Lancaster and Bioinformatics at the University of Cambridge. Our main area of interest in the Bioinformatics field, is microarray data analysis. Algorithms for the analysis of microarray data make use of a plethora of statistical methods, and the R system is rapidly becoming the method of choice for this purpose. The GROWL library is written in C, but additional wrapper support will be provided for interfacing to applications written in a variety of other languages, including C++, Fortran and R.