The automation of bioinformatics processes through workflow management systems

Paolo Romano[1], Chiara Rasi[1,2] and Domenico Marra[1]
{paolo.romano,chiara.rasi,domenico.marra}@istge.it

[1] Bioinformatics and Structural Proteomics, National Cancer Research Institute (IST), Genova, Italy
[2] Institute of Electronics, Information Engineering and Telecommunications, National Research Council, Genova, Italy

Extended abstract

There is a huge and increasing amount of biological information distributed over the Internet. The data structure and contents of data banks are extremely heterogeneous. A lot of bioinformatics software tools are also available on the net through many different interfaces. This implies that nowadays data integration on the network is a must for biomedical research [1].

Current integration tools, mainly represented by data warehouses and integration software (like SRS [4]), have strong limitations due to the quantity of data to be managed and the frequency of related updates affecting both the data itself and the structure of databases.

This makes the adoption of new flexible and extensible data integration and analysis network tools a necessity in bioinformatics. Information and Communication Technology (ICT) standards and tools, like XML, Web Services and Workflow Management Systems (WMS), can support the creation and deployment of such systems. An effort should also be done to the adherence to standards and the use of open source, that is increasingly adopted for the development of new tools [2].

Many XML languages and Web Services for bioinformatics have already been designed and implemented, together with development and support tools [4,7]. Consequently, some WMS have been proposed [5,6] and are now under careful testing aimed at the verification of their actual ability to cope with the data integration issue. While their potentiality is clear, some limitations are now arising. These include both network issues (e.g., quality of service, speed, access restrictions) and practical issues (e.g., long running jobs, huge input/output).

WMS assume that researchers know which bioinformatics resources can be reached through a programmatic interface. They also presume that users are skilled in programming and building workflows. Therefore, they are not viable to the majority of unskilled researchers. Portals enabling these to take profit from new technologies by executing predefined workflows in a user-friendly environment can therefore be useful and are also under development and testing [8].

In this paper, we present a methodology for data integration in biomedical research that is based on the adoption of XML and Web Services for data interchange, on WMS for the automation of processes and on portals for their exploitation by interested researchers. We shortly describe available WMS, by comparing them on the basis of their functions, adherence to standards and availability. We finally introduce the biowep workflow enactment portal for bioinformatics [8] and discuss about current limitations of this methodology and perspectives for overcoming them.

References

[1] L. Stein, **Creating a bioinformatics nation**. Nature, 417:119-120, 2002

[2] D.C. Jamison, **Open Bioinformatics** (editorial). Bioinformatics, 19(6):679-680, 2003

[3] M.D. Wilkinson and M: Links, **BioMOBY: an open-source biological web services proposal.** Briefings in Bioinformatics 3(4):331-341, 2002.

[4] T. Etzold, A. Ulyanov and P. Argos, (1996) **SRS: information retrieval system for molecular biology data banks.** Meth. Enzymol., 266, 114-128

[5] T. Oinn, M. Addis, et al, **Taverna: a tool for the composition and enactment of bioinformatics workflows**. Bioinformatics, 20(17):3045-3054, 2004

[6] R. Stevens, A. Robinson and C. Goble, ᵐʸ**Grid: personalised bioinformatics on the information grid**, Bioinformatics, 19(1):i302-i304, 2003, doi:10.1093/bioinformatics/btg1041

[7] P. Romano, D. Marra and L. Milanesi, **Web services and workflow management for biological resources,** BMC Bioinformatics 2005, 6(Suppl 4):S24 (doi:10.1186/1471-2105-6-S4-S24)

[8] P. Romano, E. Bartocci et al, **biowep: a workflow enactment portal for bioinformatics applications**, BMC Bioinformatics (accepted)