

Development of Information Management System for Bioinformatics Laboratory

Chang F. Quo¹, Bin Wu¹, May Wang^{1*}

Keywords: bioinformatics, information management, semantics

1 Introduction.

The quantifiable bio-information keeps growing exponentially due to constantly refined and optimized experimental technologies such as microarray chips to mine genomic data, mass spectrometry for proteomic data and nanotechnology such as molecular beacons for metabolic data. Furthermore, with burgeoning interest in bioscience, the fields of researchers contributing to bio-information steadily grow in size and diversity. Researchers for example in traditional engineering fields such as mechanical and electrical engineering are mapping their expertise to the biomedical science for seeking a convergence between these fields. Such diversity, however, brings non-uniformity; as such, researchers might not be able to share bio-information readily as intended. To overcome this issue, the World Wide Web Consortium (W3C) recommends standards for publishing Web information. While in the bioinformatics field, the Microarray Gene Expression Data (MGED) Society describes the Minimum Information about a Microarray Experiment (MIAME) ^[1] standard to facilitate sharing of microarray data, and the Protein Experiment Data Repository (PEDRo) ^[2] suggests a schema for annotating proteomic experiments. The key in unifying this spectrum of information however, bioscience related or otherwise, lies in managing the user's activity through metadata and semantics. Effective metadata leads to efficient indexing of data as well as intuitive semantics to the user.

In this poster, we present a preliminary solution for the information management within bioinformatics laboratory using PHP and MySQL. The current solution examines the network and group dynamics of user interactions within a bioinformatics laboratory. The system ultimately aims to achieve synergy in bioinformatics data and process management by presenting an integrated approach to analyze, visualize and present bio-information.

2 System Development.

The goal of this project is to implement an Information Management System for Bioinformatics Laboratory (IMSBL). A bioinformatics laboratory deals not only with data management for genomic, proteomic, and metabolic data, but furthermore, process management with the aim of achieving synergy from amongst these data. To meet these needs, we model the network of users' activities and specially group events within a bioinformatics laboratory by virtue of the form of discussion forums. By the traditional form of discussions by Email the threads of thoughts are easily lost unless previous related discussions are religiously forwarded in replies. While the emerging GmailTM preserves the context of email discussions by grouping relevant messages in a conversation format, we believe discussion forums not only track the progress of a discussion but also extend the discussion effortlessly to multiple users. As shown in

¹ Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, 313 Ferst Dr., Atlanta, Georgia, USA. Email: maywang@bme.gatech.edu

* Authors to which correspondence should be addressed.

Figure 1, the IMSBL comprises 4 modules. Group goals are first identified based on clinical demands; these demands are focused into research objectives and further specific deliverables. Personalized event management helps members track individual research progress. Discussion forums facilitate collaboration and sharing of resources such as databases and tools.

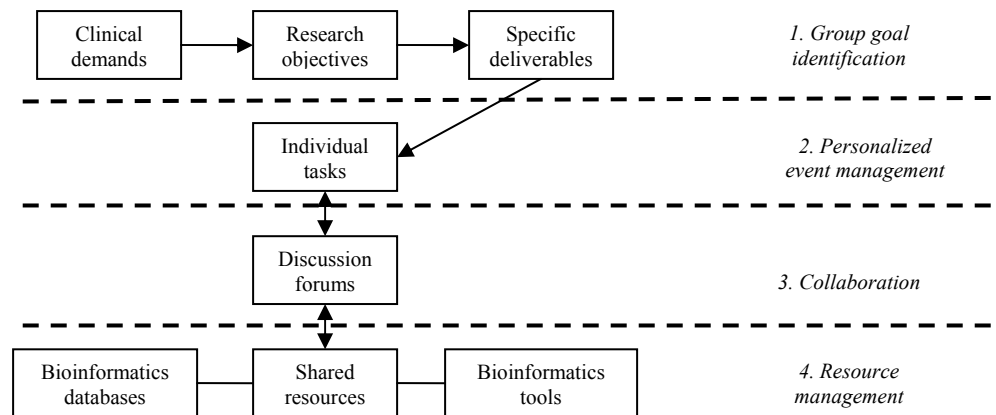


Figure 1: IMSBL Process Management.

The information related metadata is exploited behind the scenes to assist the group members to receive information, for instance, microarray data, journal papers, conference information, etc, which is specifically targeted to their research interests.

3 Future Work.

Our implementation is currently being tested by group members with the different engineering backgrounds and is under improvement based on their feedback. Interaction between the various aspects of bio-information (genomic, proteomic, and metabolic) can be modeled in a similar fashion while incorporating users' demands. Furthermore, integration with existing bioinformatics tools such as GoMiner^[3] for data management and analysis will provide a one-stop solution for bioinformatics. Finally, future work will also focus on employing semantics^[4] to achieve the automatic bio-information and processes management.

References

- [1] Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., et al. 2001. Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. *Nature Genetics*. 29(4): 365-371.
- [2] Taylor, C.F., Paton, N.W., Garwood, K.L., Kirby, P.D., Stead, D.A., Yin, Z., et al. 2003. A systematic approach to modeling, capturing, and disseminating proteomics experimental data. *Nature Biotechnology*. 21(3): 247-254.
- [4] Berners-Lee, T., Hendler, J. and Lassila, O. 2001. The semantic web. *Scientific American*.
- [3] Zeeberg, B.R., Feng, W., Wang, G., Wang, M.D., Fojo, A.T., Sunshine, M., et al. 2003. GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biology*. 4(4): R28.