

Grid Technology for Community Health Research

Nitin Sukhija[#]
National University (NUCRI)
11255 N. Torrey Pines Road
La Jolla, CA 92037 (USA)
Ph. (001) 619-955-7866
nitin@cavs.msstate.edu

Arun K. Datta*
National University (NUCRI)
11255 N. Torrey Pines Road
La Jolla, CA 92037 (USA)
Ph: (001) 858-642-8535
adatta@nu.edu
*Corresponding author

ABSTRACT

National University Community Research Institute (NUCRI) is engaged in developing *myCHOIS* or maternity and Child Health Obesity Informatics System as a health-IT solution to address the childhood obesity and related chronic diseases [1] with a vision to offer it as a fully functional EHR [2]. This system has been developed using open source portal technology with three-tiered Open Grid Services Architecture [1, 3]. This web portal is now deployed for the School Health Program of the Illinois Department of Human Services (IDHS) after successful pilot testing. In 2010 alone, health related data from about 120,000 cases have been collected into this myCHOIS database that has become a valuable resource for research¹ on childhood obesity. Moreover, as a part of the preventive care, school nurses often arrange for educational training to provide valuable instructions on a variety of topics including overweight/obesity prevention & control. Recently, a mobile application based on mCHOIS technology [4] has been deployed for IDHS to record video instructional materials and archive for training.

Personalized medicine can be a reality by capturing all relevant data that can define a patient's health profile. If that becomes the trend, a huge amount of data will be generated on any individual patient from next generation of experiments including microarray data [5] for toxicogenomics and pharmacogenomics². Data storage requirements are predicted to be in the range of terabyte to petabyte per patient [6]. Given the situation where any large healthcare organization handles thousands of patients in a year, it can easily be understandable that such enormous volumes of data can easily exceed the capacity of any local databases. Therefore, in reality, the practice will be that most of these data will be stored in a distributed fashion in a Grid [7, for definition].

C-Grid, the Community Grid, is developed³ as a portlet of myCHOIS [1] to store, manage and share large amounts of distributed community health related data in a data grid for further analysis by the health researchers in a collaborative environment. Remote management of this data grid is performed using iRODS, the Integrated Rule-Oriented Data System, which is a middleware developed by the Data Intensive Cyber Environments (DICE)

research group, and collaborators [8]. A GUI component based on PRODS has been created for C-Grid to interact with iRODS located in the data grid. C-Grid has been designed to create and manage "virtual data collection" that can be stored in heterogeneous data resources across distributed network.

For data analysis utilizing High Performance Computing and related resources from XSEDE, we have deployed neoGRID in Quarry, a virtual hosting environment, for working with workflows. A number of workflow management tools are available in open source that includes Airavata⁴ and Taverna⁵. Earlier, we used Taverna-based workflows [9, 10], a graphical workbench often used for biomedical informatics [10, and the references therein]. The neoGRID is designed to offer a HPC-supported collaborative environment for the researchers from multidisciplinary scientific fields to gather community health related data from the repository (such as, C-Grid and other data resources), and analyze those heterogeneous data using workflows developed by using various workflow management tools. Workflows specifically designed for analysis of clinical data for decision support will be available in neoGRID. Autonomic execution of these workflows will be performed without any human intervention yet achieving scientific fidelity to deliver reliable computational results despite unavoidable faults [11].

This web-based myCHOIS is under continuing development utilizing grid technology to meet the demand of the healthcare workforce and scientific collaborators. Further details of this project will be presented.

Keywords

iRODS, Data grid, community health data Repository, Virtual Organization, Workflow

ACKNOWLEDGMENTS

The C-Grid project was initiated with full technological support of Reagan Moore of SDSC (now at RENC) and his group including Wayne Schroeder and Sheau-yen Chen. Thanks to Thomas MacCalla of National University for his support.

REFERENCES

- [1] Datta, AK., Jackson, V., Nandkumar, R., Sproat, J., Zhu, W., and Kraehling, H. (2010). In 'Healthgrid Applications and Core

[#]Present address: Mississippi State University (CAVS), MS 39762 (USA)

¹http://uic-chp.org/CHP_A6_DRRP_01.html

²<http://www.mayoclinic.com/health/personalized-medicine/CA00078>

³Manuscript in submission

⁴<http://airavata.org/>

⁵<http://www.taverna.org.uk/>

- Technologies (Eds. , T. Solomonides, I. Blanquer, V. Breton, T. Glatard, and Y. Legre), Proceedings of HealthGrid 2010 (Paris), pp 191-202. IOS Press, Netherlands. ISBN 978-1-60750-582-2.
- [2] Berner E.S. (2009). AHRQ Publication No.09-0069-EF. Rockville, Maryland: Agency for Healthcare Research and Quality.
- [3] Datta, A.K., Jackson, V., Nandkumar, R., and Zhu, W. (2010). Proceedings of the PRAGMA 18, San Diego (CA), March 3 -4, 2010.
- [4] Datta, A.K., Sumargo, A., Jackson, V., Dey, P.P. (2011). mCHOIS: Procedia Computer Science, vol. 5, p 653 - 660, 2011 (Elsevier).
- [5] Weston, AD., and Hood, L. (2004). Journal of Proteome Research 2004, 3, 179-196 179.
- [6] Hey, T. and Anne Trefethen, A. (2003). in Grid Computing – Making the Global Infrastructure a Reality”, Wiley, January 2003.
- [7] Foster, I. and Kesselman, C. (editors) The Grid: Blueprint for a New Computing Infrastructure, Morgan Kaufmann, San Francisco 1999.
- [8] Rajasekar, A., Moore, R., et. al.(2010). Lectures on Information Concepts, Retrieval, and Services, 2010, Vol. 2, No. 1 , Pages 1-143 (Publisher: Morgan and Claypool Publishers)
- [9] Datta, A.K., Eller, B., and Sinai, A. (2011). neoGRID: Enabling access to teragrid resources - Application for sialylmotif analysis in the protozoan Toxoplasma gondii. TeraGrid '11: Extreme Digital Discovery, July 18-21, Salt Lake City, Utah.
- [10] Datta A.K., Tilson, J.L., Rendon, G., and Jakobsson, E. 2009. Proceedings of the TeraGrid '09 (Arlington, Virginia, June 22 – 25, 2009). DOI=http://archive.teragrid.org/tg09/files/tg09_submission_43.pdf
- [11] Haupt, T., Sukhija, N., Zhuk, I.(2011). Autonomic Execution of Computational Workflows. Proceedings of the Federated Conference on Computer Science and Information Systems pp. 965–972. ISBN 978-83-60810-22-4.