

Cyberinfrastructure for Community Health Research

Arun K. Datta*, Nitin Sukhija, Vineeth Jose, Bandhavi Nalajala

National University Community Research Institute (NUCRI), 11255 N. Torrey Pines Road, La Jolla, CA 92037

*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, 2] with a vision to offer it as a fully functional EHR¹ supported by a Decision Support System [3]. This system has been developed using open source portal technology with three-tiered Open Grid Services Architecture [1, 4]. This web portal is now deployed for the School Health Program of the Illinois Department of Human Services (DHS²) 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 research 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 [5] has been deployed for DHS to record video instructional materials and archive for training. Moreover, 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 for toxicogenomics and pharmacogenomics [6]. Data storage requirements are predicted to be in the range of terabyte to petabyte per patient [7]. 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 [8, for definition]. C-Grid, the Community Grid, is developed as a portlet of myCHOIS [9] 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 [10]. A GUI component based on PRODS has been created for C-Grid to interact with iRODS located in the data grid. C-Grid³ 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 [11, 12], a graphical workbench often used for biomedical informatics [11, 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

¹ http://www.himss.org/ASP/topics_ehr.asp; http://en.wikipedia.org/wiki/Electronic_health_record;
also see: http://healthit.hhs.gov/portal/server.pt/community/healthit_hhs_gov__home/1204

² <http://www.dhs.state.il.us/page.aspx?item=30318>

³ Manuscript under submission

⁴ <http://incubator.apache.org/airavata/>; <http://people.apache.org/~smarru/papers/airavata-gce11.pdf>

⁵ <http://www.taverna.org.uk/>

intervention yet achieving scientific fidelity to deliver reliable computational results despite unavoidable faults [13]. 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: Obesity, metabolic disease, Body-Mass-Index (BMI), OGCE, mobile communication, EHR, grid computing, workflow.

Acknowledgement

This project is supported in part through IDHS Coordinated School Health Technical Assistance grant.

References

- [1] Datta, A.K., Jackson, V., Nandkumar, R., Sproat, J., Zhu, W., and Kraehling, H. (2010). CHOIS: Enabling grid technologies for obesity surveillance and control. In 'Healthgrid Applications and Core Technologies (Eds. , T. Solomonides, I. Blanquer, V. Breton, T. Glatard, and Y. Legre), vol 159, p191-202, IOS Press, Washington D.C. ISBN 978-1-60750-582-2; Presented at the HealthGrid 2010, Paris, June 28-30, 2010.
- [2] Datta, A.K. (moderator), Jackson, V., and Rimmer, J. (2011). Informatics Challenges and Opportunities for Childhood Obesity Research. Public Health Informatics Conference (PHIN 2011), Atlanta (GA), August 20 - 24, 2011. Panel discussion, viewable at: <http://cdc.confex.com/cdc/phi2011/webprogram/Session12554.html>
- [3] Berner E.S. (2009). Clinical Decision Support Systems: State of the Art. AHRQ Publication No.09-0069-EF Rockville, Maryland: Agency for Healthcare Research and Quality.
- [4] Datta, A.K., Jackson, V., Nandkumar, R., and Zhu, W. (2010). Cyberinfrastructure for CHOIS - a Global Health initiative for obesity surveillance and control. Proceedings in the PRAGMA 18, San Diego (CA), March 3 -4, 2010.
- [5] Datta, A.K., Sumargo, A., Jackson, V., Dey, P.P. (2011). mCHOIS: An Application of Mobile Technology for Childhood Obesity Surveillance. Procedia Computer Science, vol. 5, p 653 - 660, 2011 (Elsevier). Presented at The 8th International Conference on Mobile Web Information Systems (MobiWIS), September 19-21, 2011, Niagara Falls, Ontario, Canada.
- [6] Weston, A.D., and Hood, L. (2004). Systems Biology, Proteomics, and the Future of Health Care: Toward Predictive, Preventative, and Personalized Medicine. Journal of Proteome Research 2004, 3, 179-196.
- [7] Hey, T. and Anne Trefethen, A. (2003). The Data Deluge: An e-Science Perspective. In 'Grid Computing – Making the Global Infrastructure a Reality' (Eds., F Berman, G C Fox and A J G Hey), Wiley, April 2003, ISBN: 0470853190..
- [8] Foster, I. and Kesselman, C. (editors). The Grid: Blueprint for a New Computing Infrastructure, Morgan Kaufmann, San Fransisco, 1999.
- [9] Sukhija, N. and Datta, A.K. (2012). Grid Technology for Community Health Research. XSEDE 2012: eXtreme Digital Discovery, July 17 - 20, Chicago (IL).
- [10] Rajasekar, A., Moore, R., et. al.(2010). Lectures on Information Concepts, Retrieval, and Services, 2010, Vol. 2, No. 1 , Pages 1-143. Morgan and Claypool Publishers.
- [11] Datta A.K., Tilson, J.L., Rendon, G., and Jakobsson, E. 2009. High-throughput sialylmotif analysis in the glycosyltransferase protein family. Proceedings of the TeraGrid '09, Arlington, Virginia, June 22 – 25, 2009. DOI= http://archive.teragrid.org/tg09/files/tg09_submission_43.pdf
- [12] 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.
- [13] Haupt, T., Sukhija, N., Zhuk, I.(2011). Autonomic Execution of Computational Workflows. Proceedings of the Federatedn Conference on Computer Science and Information Systems, p 965–972. ISBN 978-83-60810-22-4.