The Indo-US Cancer Research Grid: Enabling global collaboration through Information Technology

George A. Komatsoulis, Ph.D.1, Hemant Darbari, Ph.D.2, Anil Srivastava2, Ravi Madduri, MS3, John G. Williams, Ph.D.4 and Ganesh Shankar, MS4

1Center for Biomedical Informatics and Information Technology, NCI, NIH. 2Centre for the Development of Advanced Computing (CDAC), India. 3University of Chicago, 4Indiana University

Overview

In 2004, the National Cancer Institute launched the cancer Biomedical Informatics Grid (caBIG®) to create a “world wide web of cancer research”. The caBIG program has created more than 70 software applications and an interoperable infrastructure that supports data semantics, aggregation and workflows.

The Centre for the Development of Advanced Computing (CDAC) is a Research and Development Organization in the Department of Information Technology, Ministry of Communications and Information Technology, Government of India. CDAC is a global leader in the area of High Performance Computing, Multi-lingual Computing and Health Informatics.

CDAC and caBIG are working to enable collaborations between the US and India by creating the Indo-US Cancer Research Grid (IUCRG). This Grid will provide access to useful biomedical research capabilities to scientists and clinicians in both countries using the underlying caGrid technology developed by caBIG, leveraging the capabilities that are developed by both groups.

caGRID Architecture

The caGrid (http://www.cagrid.org) is the technical implementation of caBIG’s interoperability paradigm. It is architected as a series of services that provide discovery, semantics, federated query and workflow, and a federated security infrastructure. caGrid infrastructure is derived from the Open Source Globus Toolkit (http://www.globus.org) with extensions to support the interoperability needs of the caBIG community. caGrid can be used to integrate systems locally at institutions as well as across institutions. Its modular nature supports the concept of a “Grid of Grids”, a group of networks that are locally managed but that can interoperate via a common trust framework.

caBIG US Deployment

The initial production release of caGrid was in the fourth quarter of 2007 with 11 systems. Since then, the national caBIG grid has grown to include systems at more than 50 US cancer centers and other academic institutions across the US. This number does not include independent systems that are installed behind institutional firewalls (that could connect to the caGrid if desired) and systems connected to local caGrid instances (such as those at the Ohio State University, University of Alabama, Birmingham and Washington University, St. Louis).

Initial IUCRG Collaborations

The caBIG program has developed a library of 70+ software tools and services to support biomedical research. All caBIG tools provide APIs in a variety of technologies (Java, Web Services,etc.) and most can connect to the caGrid infrastructure. All are released under a non-viral open source license that explicitly allows for both commercial and non-commercial derivative works.

The authors wish to acknowledge the National Knowledge Network, Government of India, Internet2, ERNET India and the National Science Foundation.