

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

George A. Komatsoulis, Ph.D.<sup>1</sup>, Hemant Darbari, Ph.D.<sup>2</sup>, Anil Srivastava<sup>1</sup>, Ravi Madduri, MS<sup>3</sup>, John G. Williams, Ph.D.<sup>4</sup> and Ganesh Shankar, MS<sup>4</sup>

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



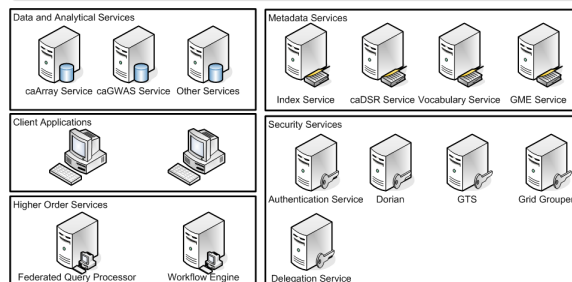
## Overview

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 Applications



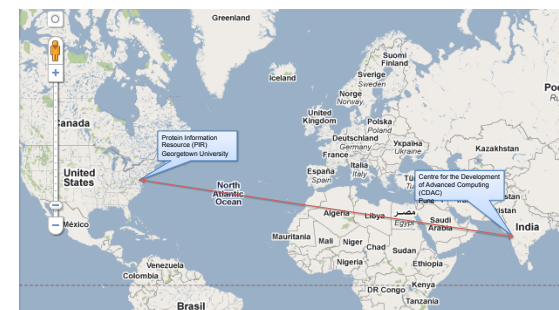
The caBIG program has developed a library of 70+ software tools and services to support biomedical research. All caBIG tools provide APIs in a variety of technologies (Java, Web Services, etc.) and most can connect to the caGrid infrastructure. All are released under a non-viral open source license that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

## 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



CDAC and caBIG's initial collaborations utilize the strengths supplied by both organizations. The first is to provide access to CDAC High Performance Computing resources and protein folding algorithms. Protein sequences from the Protein Information Resource (<http://www.pir.org/>) will be accessed through the GridPIR caGrid resource and transferred to CDAC for protein folding analysis, using caGrid workflow, transport and security technology. The development of this workflow will enable both parties to set up the necessary connectivity, security and trust infrastructure to enable more complicated workflows with higher sensitivity data in the future. Researchers at the University of Chicago are leading the work to develop the specific workflow.

The second collaboration is in the area of multi-lingual support for controlled biomedical terminologies. In this case, caBIG will be providing a well curated biomedical terminology and technology to provide runtime access to that terminology and CDAC will provide advanced translation tools for medical terminology. The test case will involve the creation of a Spanish language translation of the Common Terminology Criteria – Adverse Events (CTC-AE), a specialized vocabulary used for recording information about Adverse Events; undesirable changes in health or side effects that occur during the course of a clinical trial. If successful, this work could lead to more efficient generation of local translations of other biomedical terminologies.

## Acknowledgements

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

