Public Health Informatics Research Grid Initiative

Executive Summary
Fall 2008

Overview

This document provides an overview of the Public Health Informatics Research Grid Initiative, coordinated at CDC's National Center for Public Health Informatics (NCPHI). Specifically, the issues addressed in this document include why this initiative was started, what are its goals, who is involved, what are some of the technical issues that need to be considered, what are examples of some of its achievements, and how others can become involved.

Why was the Public Health Informatics Research Grid initiated?

The public health community faces significant challenges in achieving efficient and effective electronic exchange of data, information, and knowledge. These challenges include:

1. Wide distribution of public health data.
2. Rapid growth of public health data.
3. Cultural, social and political impediments to data sharing.
4. Significant and chronic financial constraints.
5. A dynamic and complex environment - global in scale.
6. An environment containing many redundant systems, as well as application and data silos.
7. An environment with a wide variety of complex requirements (disease surveillance, alerting, event detection, etc).

New research in the field of public health informatics is underway to find innovative solutions to solve this significant information exchange challenge. Specifically, it is the hope that research in the domain of grid computing will afford this critical solution.

Grid computing is a paradigm that proposes aggregating geographically-distributed, heterogeneous computing, storage and network resources to provide unified, secure and pervasive access to their combined capabilities (Foster and Kesselman, 1999).

What are the overall goals of this initiative?

The ultimate goal is to provide the public health community a secure, cost-effective, high-value, and intuitive technical and social infrastructure for addressing public health challenges.

Specifically the goal is to achieve this infrastructure through the use of grid technologies. Our research will help define and simplify how exactly to achieve this goal. Other core principles include: Long-term sustainability; low barrier to entry (technically, financially & socially); 100% standards-based; reusability of components; usage of open source software and open collaboration; leveraging best practices; and facilitate a federated environment.

In the near future, we hope to continue to mature and refine our understanding and use of grid technology for public health impact, increase community participation in projects, and increase our education and outreach activities (publications, presentations, etc).

Who is involved in this initiative?

There are a wide variety of partners involved this initiative. These include academic partners, corporate partners, public health partners, and grid computing partners. Each partner brings a unique set of skills and expertise to the initiative.

What are some of the technical issues that need to be considered?

Grid technology provides a secure software infrastructure for federating databases and services utilizing a service-oriented architecture (SOA). The Grid is "distinguished from conventional distributed computing by its focus on large-scale resource sharing, innovative applications, and in some cases, high performance orientation (Foster & Kesselman 2002). Globus has been selected as the preferred software infrastructure (Grid Middleware). Globus is an open source toolkit that is used for building grid computing systems. Globus enables users to securely share computing power, files, and services without compromising the remote node's local security policies. In order to use grid services, each user is required to have a valid x.509 user certificate mapped to a valid user account on the remote node. This feature allows system administrators to locally maintain the security of a node connected to the Public Health Grid. The installation process for a Grid node has been optimized, and continues to be streamlined.
The installation of a node is performed on a Linux Operating System, specifically in a virtual environment (e.g., VM Ware). Once a digital certificate has been installed on a new node, specific firewall ports must be opened. The connection is then validated via GridFTP. Once active on the grid, data and services can be shared and consumed, based, of course, on agreed security controls.

What are examples of achievements, thus far, in this initiative?

In the relatively short time this initiative has been in existence, our team in collaboration with our partners has:

- Established communication with existing grid based communities: caBIG, Argonne National Lab, Ohio State Department of Biomedical Informatics, and the Globus alliance.

- Created node installation documentation, procedures and packages allowing installation in less than 30 minutes.

- Established nodes with public health partners including: University of Pittsburgh, Mayo Clinic, Columbia University, Tarrant County Department of Health, Dallas County Department of Health, and the Indiana Hemophilia and Thrombosis Center.

- Successfully performed distributed secure data transfers between the NCPHI Lab node and partner nodes.

- Developed distributed data queries to combine sample biosurveillance data stored on partner nodes.

- Developed grid services to provide access to computational and data resources stored on partner nodes.

- 2008 JAMIA Publication: In Response to: What is a Grid? Thomas G. Savel MD, Jonathan C. Silverstein MD, Kenneth E. Hall MDIV, and Leslie Lenert MD, MS. Volume 15, Number 5. Page(s) 705-706.

- Presentations and posters at national conferences (HIMSS, HealthGrid.US.

How can others become involved?

Please feel to contact the core members of our team:

Executive: Tom Savel, CDC, NCPHI Lead – Grid Research: tsavel@cdc.gov

Ken Hall, Strategic Science & Research: kha6@cdc.gov

Development:

Dan Washington, Grid Infrastructure Engineer: Dan.Washington@bearingpoint.com

Peter White, Grid Software Engineer: Peter.White@bearingpoint.com

When you contact us, please let us know what type of partner you are (academic, public health, etc), as well as your area(s) of interest / expertise, and your availability of time to volunteer in this rapidly growing initiative. Do you want to have a public health informatics research grid node installed at your location? Do you want to try using a service? Do you want to work with other partners to create a new service? Do you simply want to learn more about Grid? Please understand that this is a research initiative that is leveraging open-source solutions, and an open-collaborative development framework. We look forward to hearing from you.