



---

# **LEVERAGING GRID TECHNOLOGY TO ENHANCE LOCAL BIOSURVEILLANCE CAPACITY**

**NCPHI COLLABORATION WITH RODS AND OGSA-DAI**

**NCPHI COLLABORATION WITH HARVARD ELECTRONIC SUPPORT  
FOR PUBLIC HEALTH (ESP)**

## **PROJECT CHARTER**

---

Version 1.1

June 9, 2008

---

## VERSION HISTORY

Version #	Implemented By	Revision Date	Reason	Approved By	Approval Date
0.1	PMO	04/28/08	Initial draft	Crystal Watson	05/05/08
0.2	William Duck	05/08/08	Input by Tom Savel, Crystal Watson, William Duck, Ken Hall, Moses Miles, and Brian Lee	Crystal Watson	05/14/08
0.3	Ken Hall	05/22/08	Input from Ken Hall, Crystal Watson and William Duck	Crystal Watson	05/23/08
1.0	Ken Hall and Crystal Watson	05/27/08	Final Draft Edits	Crystal Watson	05/27/08
1.1	Brian Lee	06/09/08	Added Harvard ESP		

# TABLE OF CONTENTS

<b>1 INTRODUCTION.....</b>	<b>4</b>
1.1 Purpose of LITE Project Charter .....	4
<b>2 PROJECT AND PRODUCT OVERVIEW.....</b>	<b>4</b>
<b>3 JUSTIFICATION .....</b>	<b>5</b>
3.1 Business Need .....	5
3.2 Public Health and Business Impact .....	6
<b>4 SCOPE .....</b>	<b>7</b>
4.1 Objectives .....	7
4.2 High-Level Requirements.....	8
4.3 Major Deliverables/Milestones.....	8
<b>5 DURATION .....</b>	<b>8</b>
5.1 Duration.....	8
5.2 Resources.....	9
<b>6 ASSUMPTIONS, CONSTRAINTS AND RISKS .....</b>	<b>9</b>
6.1 Assumptions .....	9
6.2 Constraints.....	9
6.3 Risks .....	9
<b>7 PROJECT ORGANIZATION.....</b>	<b>10</b>
7.1 Roles and Responsibilities .....	10
<b>8 PROJECT CHARTER APPROVAL.....</b>	<b>12</b>
<b>APPENDIX A: RESEARCH LAB REQUIREMENTS.....</b>	<b>13</b>

## **1 INTRODUCTION**

### **1.1 PURPOSE OF LITE PROJECT CHARTER**

The Leveraging GRID Technology to Enhance Local Biosurveillance Capacity Project Charter documents and tracks the necessary information required by decision maker(s) to approve the project for funding.

The intended audience of the Leveraging GRID Technology to Enhance Local Biosurveillance Capacity Project Charter is the project sponsor and senior leadership.

## **2 PROJECT AND PRODUCT OVERVIEW**

Since the inception of the Public Health Research GRID proof of concept (PoC) in December of 2007, both the intramural and extramural projects have made significant progress in building developer capacity regarding application of GRID technologies towards public health practice. The NCPHI computer research lab in collaboration with the Real-time Outbreak and Disease Surveillance (RODS) Open Source Project and the OGSA-DAI group from the University of Edinburgh, UK continue to work on the following tasks:

- Establishing successful communication between GRID nodes using the Globus GRID toolkit
- Build developer proficiency in using the Open GRID Services Architecture-Database Access and Integration (OGSA-DAI) toolkit to construct federated queries across multiple biosurveillance datasets
- Construct a JSP front-end application to execute a federated query

The following charter represents the convergence of these efforts towards building relevant public health services to be consumed within a GRID framework.

This charter defines one PoC project that is a direct application of the ongoing distributed development efforts regarding the Phase I GRID PoC. Specifically, the PoC will focus on building a GRID-enabled federated query service that can be integrated into a deployment of RODS. The outcome of this project will demonstrate how GRID technologies, specifically Globus, and distributed database processing of federated data sources (OGSA-DAI) can be used to develop services that enhance local biosurveillance capacity to detect events across multiple jurisdictions. This project will inform how public practice surrounding early event detection and situational awareness can be supported using distributed queries over federated biosurveillance data in a multi-regional GRID framework.

In addition to the work above, local biosurveillance capacity can be extended by developing a syndromic surveillance module for the Harvard Center of Excellence's Electronic Support of Public Health (ESP) project and using Globus as the means of secure, reliable data transfer and Globus services as the means of secure access to the collected syndromic data. The full time frame for this module is 12 months, but given the duration of this proof of concept project, the initial

steps can be completed within this project with additional steps to be completed over the following 8 months.

The new ESP module will generate daily reports of syndromic surveillance segmented into three levels of data: aggregate syndrome counts by zip code partially de-identified records containing location, age in years and gender; identified records containing name, address and date of birth and essential clinical details.

The NCPHI computer research lab in collaboration with the Harvard Center of Excellence and CDC Biosurveillance resources can accomplish the following tasks:

- Install Globus node at Harvard site
- Determine the Syndrome Definitions to be used for classification
- Determine the data fields to be used in generated syndromic surveillance reports
- Determine the means of transport of the generated syndromic surveillance reports

An additional phase to this proof of concept project is required to accomplish the following tasks:

- Complete secure data transfer of generated syndromic surveillance reports.
- Development of grid services to provide access to data extracted from generated syndromic surveillance reports.
- Development of demonstration JSP user interface to invoke grid services based on user access level.

The NCPHI computer research lab will continue its role as the principal collaboration and development site of the services developed within this PoC. The NCPHI Office of the Director (OD) Project Management Office (PMO) will provide oversight, governance and PM services for the PoC. The NCPHI Associate Director of Science will provide guidance for the PoC. Coordinated efforts between the following organizations are required: CDC (OD & PMO team); BearingPoint; Tarrant County (TX) Public Health; the RODS Open Source Project: OSGA-DAI group from the University of Edinburgh, UK; Harvard CoE. The PoC will document results of building GRID-enabled services and their integration into existing biosurveillance applications.

### **3 JUSTIFICATION**

#### **3.1 BUSINESS NEED**

NCPHI is responsible for a significant portion of CDC's public health informatics investments, and must manage these investments for maximizing benefits for CDC, the public health system, and the public at large. As stewards of these investments, NCPHI has announced that it will continue its architectural transition

towards a highly interoperable, federated architecture, promoting data accessibility and analysis, in a highly secure environment. PoC's that adequately demonstrate how GRID architectures can support the development of public health services will continue to solidify NCPHI's evaluation role of emerging technologies in support of this strategy.

Key to NCPHI's strategy will be leveraging the tools, methodologies, and expertise of academic partners with proven track records in public health practice and biosurveillance. The RODS laboratory of the University of Pittsburgh is a recognized leader in the field, having developed the open source RODS application, as well as publishing several academic articles in biosurveillance. The RODS laboratory encourages open contribution to its software while fostering collaboration and innovation. Finally, with at least 19 RODS instances in production, there is a natural laboratory that the Public Health Research GRID can use to test the performance, security, usability, and other aspects of distributed queries. A partnership between NCPHI and RODS has already been established in the early phases of the Public Health Research GRID, and the innovation proposed within this PoC charter will prove to be beneficial not only to NCPHI, but to the entire community that RODS supports.

Additionally, the Harvard Center of Excellence Electronic Support for Public Health (ESP) has been operational since 2006 in detecting notifiable disease cases among electronic medical records from a 600,000 patient multisite, multispecialty practice. This portable, open-source ESP infrastructure offers an ideal platform for generating comprehensive syndromic surveillance data streams. This project will collaborate with modifications to the ESP project to generate syndromic surveillance reports based on their existing data stream. NCPHI will be able to leverage research into grid technology to provide low cost infrastructure for data transfer and data access.

### **3.2 PUBLIC HEALTH AND BUSINESS IMPACT**

The public health impact will be to inform, but not define, the long-term target architecture, services, and standards that will support nationwide public health informatics capacity, and specifically biosurveillance. This PoC will help NCPHI and RODS evaluate potential technologies and methodologies that can leverage the community and services of the Public Health Research Grid and RODS Open Source project. Additionally, this PoC will allow NCPHI to evaluate the effort involved to extend existing functionality in the public health community using the Public Health Research Grid. This will solidify NCPHI's role as a steward and principal catalyst to seed the larger public health community with resources necessary to support collaborative, open source development of GRID-enabled public health services. Federal public health partners will assume a key in role driving the emergence of sustainable development standards for these services while ownership of their implementation would be shared by the entire public health community.

## **4 SCOPE**

### **4.1 OBJECTIVES**

The diverse landscape of locally deployed and maintained biosurveillance systems presents unique challenges to the growing public health informatics community. The demands for leveraging existing biosurveillance infrastructure and workforce capacity to build a common services platform available to the global public health community continue to grow. However, given the fragmented evolution of public health software infrastructure and applications, reconciliation of these technologies continues to be a challenge. Achieving semantic interoperability has been unapproachable without a secure, fault-tolerant framework that can protect sensitive public health information transactions. The GRID technology work performed in previous PoCs by the NCPHI research lab and others have paved the way to begin open, distributed development of services relevant to all users of biosurveillance applications.

Based on the collaborative development efforts of this growing public health community, a common, extensible informatics platform for federating heterogeneous data and expanding biosurveillance capacity is emerging. The ability to execute and visualize distributed queries over a GRID as outlined in this PoC represent the first of many services to be explored. These services will become the foundation of distributed surveillance applications that can be used for reportable condition surveillance, case detection, outbreak management, and chronic disease surveillance. Continued research into GRID-enabled public health services will encourage innovation, discovery, and participation on the behalf of local, federal, and academic public health partners.

The objectives of the Leveraging GRID Technology to Enhance Local Biosurveillance Capacity charter are as follows:

- Use existing Globus nodes and GRID infrastructure within NCHPI research lab for developing a distributed query service
- Federation of three RODS databases at Tarrant county, TX, Dallas county, TX, and the RODS lab at University of Pittsburgh will be accomplished using the OGSA-DAI API
- Federated queries will be created to select syndrome counts by date, age, sex, and location from pre-classified RODS data
- GRID-enabled service interface will be built using JSP and integrated within a deployment of RODS to allow end users to easily execute distributed queries
- Resultant datasets will be visualized using existing RODS AVR functionality
- Install new Globus node and GRID infrastructure within the Harvard CoE
- Facilitate the development of a new ESP module by the Harvard CoE

- Design the interfaces for transmitting aggregate and unit record data streams generated by ESP module using GRID infrastructure

## 4.2 HIGH-LEVEL REQUIREMENTS

The following table presents the requirements that the project's product, service or result must meet in order for the project objectives to be satisfied.

Req. #	Requirement Description
01	Develop and demonstrate capability to federate biosurveillance data repositories over Globus GRID nodes using the OGSA-DAI toolkit
02	Develop and demonstrate distributed query capability against multiple Globus GRID nodes using the OGSA-DAI toolkit
03	Develop and demonstrate GRID-enabled RODS deployment with integrated federated query service to return syndrome counts by date, location, age, and sex across Globus GRID nodes
04	Develop and demonstrate visualization of distributed query results in a table, time-series, and geo-spatially using existing RODS analysis features
05	Install and test Globus GRID node at Harvard CoE site
06	Facilitate the development of ESP module by coordinating syndrome definition, data field definition and transport definition.

## 4.3 MAJOR DELIVERABLES/MILESTONES

Major Deliverable	Deliverable Description
Project Kick-off	Meet with Project Sponsors to confirm the: <ul style="list-style-type: none"> <li>• Project Scope</li> <li>• Project Approach</li> <li>• Project Deliverables</li> </ul>
PoC Assessment	Level of effort and success rate for meeting above PoC objectives
PoC results Documented	Results intended for use for presentation to NCPHI, Centers of Excellence, Advanced Practice Centers, other partners and submission of paper(s) to peer-reviewed journals
PoC Demonstrated	Provide a live, hands-on demonstration of the distributed query service within a RODS deployment at PHIN conference (optional)
PoC Future Phase Design	Design and charter for additional phases necessary to complete support for Harvard activities in development of new ESP module

## 5 DURATION

### 5.1 DURATION

The project has a defined end date of September 1, 2008.

A Project Schedule will be created upon confirmation of the project scope and resources available to complete the project.

## 5.2 RESOURCES

**Project Core Team:** Current members of the NCPHI OD, NCPHI PMO and an EA Technical Resource will be used to staff this project.

**NCPHI Lab Team:** Division of Shared Services will provide resources for technical and infrastructure services for the NCPHI Lab.

**NCPHI Biosurveillance Team:** Division of Emergency Preparedness and Response will provide resources for public health support and evaluation.

**External Partners:** BearingPoint; Tarrant County (TX) Public Health; the RODS Open Source Project; OSGA-DAI group from the University of Edinburgh, UK

**Project Execution Support:** This project will require active participation by the system owner and their project teams.

## 6 ASSUMPTIONS, CONSTRAINTS AND RISKS

### 6.1 ASSUMPTIONS

This section identifies the statements believed to be true and from which a conclusion was drawn to define this project charter.

1. Resources and end date are fixed.
2. System owners and project teams will participate and comply to requests necessary for completion of this project.
3. This proof of concept represents research. The knowledge and experience generated by these projects should inform other projects leading to future GRID-enabled public health services for consumption by NCPHI and the public health community.

### 6.2 CONSTRAINTS

This section identifies any limitation that must be taken into consideration prior to the initiation of the project.

1. There are time constraints that will limit the amount and depth of analysis than can be performed.
2. There are a limited number of developer resources available to complete the project.

### 6.3 RISKS

Risk	Mitigation
------	------------

Risk	Mitigation
Unfunded initiative with limited timeframe and resources	Some high-level objectives are noted as “time permitting”
Lack of formal buy-in from internal stakeholders	Sponsorship from NCPHI Director and a Project Charter authorizing the project.
Use of external partners	Assessment nature of PoC does not require hard deadlines for external partners
No hands-on experience with developing GRID-enabled services for existing biosurveillance applications	Developer knowledge base of OSGA-DAI and Globus toolkits accelerating. Rudimentary interface design and federated query execution already underway.
Scalability of GRID-enabled services across a larger nodal network	If scalability is an issue, develop method to prioritize available GRID service bandwidth.

## 7 PROJECT ORGANIZATION

### 7.1 ROLES AND RESPONSIBILITIES

This section describes the key roles supporting the project.

#### RODS Proof of Concept

Name & Organization	Project Role	Project Responsibilities
Dr. Leslie Lenert NCPHI OD	Project Sponsor	Person responsible for acting as the project’s champion and providing direction and support to the team. In the context of this document, this person approves the request for funding, approves the project scope represented in this document, and sets the priority of the project relative to other projects in his/her area of responsibility.
Dr. Tom Savel Crystal Watson NCPHI OD	Project Officers	A program representative responsible for coordinating with acquisition officials on projects for which contract support is contemplated. This representative is responsible for technical monitoring and evaluation of the contractor's performance after award.
Ken Hall BearingPoint	Technical Steward	The Technical Stewards are responsible for providing technical direction to the project.

Name & Organization	Project Role	Project Responsibilities
PMO/EA Team  NCPHI         BearingPoint	Project Team	Performs the day to day activities of the projects.  PMO Team <ul style="list-style-type: none"> <li>• Crystal Watson &amp; TBD PMO Resource</li> </ul> EA Technical Resource <ul style="list-style-type: none"> <li>• Brian Lee</li> </ul> * Other members of the Project Team may include NCPHI staff and contractors who may serve in a Subject Matter Expert role.
External Partners	Advisory Group	Provide technical resources and direction. <ul style="list-style-type: none"> <li>• RODS Open Source Group</li> <li>• Tarrant County, TX Public Health</li> <li>• OGSA-DIA group from University of Edinburgh, UK</li> </ul>

## 8 PROJECT CHARTER APPROVAL

The undersigned acknowledge they have reviewed the Leveraging GRID Technology to Enhance Local Biosurveillance Capacity **Project Charter** and authorize and fund the project. Changes to this **Project Charter** will be coordinated with and approved by the undersigned or their designated representatives.

Signature: \_\_\_\_\_ Date: \_\_\_\_\_  
Print Name: Dr. Leslie Lenert  
Title: NCPHI Director  
Role: Executive Sponsor

Signature: \_\_\_\_\_ Date: \_\_\_\_\_  
Print Name: Dr. Tom Savel  
Title: NCPHI Director of Science  
Role: Project Officer

Signature: \_\_\_\_\_ Date: \_\_\_\_\_  
Print Name: Crystal Watson  
Title: NCPHI PMO Director  
Role: Project Officer

## APPENDIX A: Research Lab Requirements

<b>Server Instances</b>	<b>2</b>
<b>Server OS Installations</b>	<b>Redhat Linux (latest version)</b>
<b>Server RAM</b>	<b>2 – 4 GB</b>
<b>Processors</b>	<b>Dual core (2 GHz per core)</b>
<b>Workstation locations</b>	<b>1</b>
<b>Project End-date</b>	<b>September 1, 2008</b>
<b>Internet Access</b>	<b>T1</b>
<b>3<sup>rd</sup> Party Software</b>	<b>Java 1.5 JDK</b>
<b>Access Mode</b>	<b>Root access</b>