

## caBIG Architecture Workspace Participant Form

Please complete this form in advance of the caBIG kickoff meeting and return by e-mail to [adamsm@mail.nih.gov](mailto:adamsm@mail.nih.gov). Completed forms will be made available to all participants in advance of the meeting to enhance workspace discussions. You should also use this worksheet to structure your presentation. This presentation should be kept to 20 minutes, so be brief! Additional materials you would like reviewed but will not have time to present can be submitted in advance to the email address above and we will see that it is made available to the other Workspace participants.

1. Sponsoring Cancer Center
  1. Lombardi Cancer Center, Georgetown University
2. Workspace
  1. Architecture
3. Relevant architecture experience with similar efforts and the technologies that were implemented
  - a. System design
    - i. Georgetown University Performance Processing Initiative (GUPPI) (<http://www.clusters.arc.georgetown.edu/>)
    - ii. Beowulf Clusters
    - iii. IBM Information Integrator
    - iv. MySQL Databases with Java web front ends
  - b. Component details
    - i. Globus Toolkit
    - ii. OSCAR Beowulf
    - iii. Condor Job Scheduler
    - iv. Ganglia monitoring tools
    - v. Linux
    - vi. MySQL
    - vii. Java
    - viii. Perl
  - c. Relevant standards
    - i. Open Grid Services Infrastructure v1.0
    - ii. Other standards have been historically addressed in the tools selected
  - d. Size of project installed software base
    - i. Concentration on infrastructure and the tools to connect to the infrastructure. Selected PI's currently using GU architecture:
      1. The UniProt Protein Sequence and Function Knowledgebase U01 HG02712-01, NIH) PI: Dr. Cathy Wu
      2. Title: Electronic Structure of Conducting Polymers and Organic Materials; National Science Foundation, DMR-0331710; PI: Miklos Kertesz, Professor
      3. Biodemography of Health, Social Factors & Life Challenge (5R01 - AG16661 NIA(NIH) PI: DR. Maxine Weinstein
      4. "The impacts of tropical forest fragmentation and population structure, seed and pollen gene flow and future genetic diversity in the tropical tree *Corythophora alta*." (DEB-9983014, NSF) PI: Dr. Mathew Hamilton
  - e. Development Environment (tools, languages, bug tracking, etc.)
    - i. Java
    - ii. SQL (MySQL, PostgreSQL and Oracle)

iii. Perl

4. Lessons learned that could be applied to caBIG
  1. The architecture requires implementation as early in the project as possible to allow API's to be tested against it.
  2. One architecture with one login should be available for all tools.
    1. A portal should provide access to tools, applications and data.
    2. This portal should house user certificates.
    3. This portal should be web based.
  3. Metadata creation must be automated.
  4. Certificate exchange as described is a temporary solution until access control based on attributes is achieved.
  
5. Role you see yourself playing in supporting the domain workspace implementation of Architecture Workspace guidelines
  1. Local test bed implementation.
  2. First office remote test bed implementation.
  3. Training
  4. Local support for installation of Grid services tools at all sites.
  5. Ongoing maintenance of Certificate Server.
  6. Development of Web Portal to access data grid, computational resources and tools.
  7. Ongoing maintenance of Web Portal
  8. Hosting software repository.
  9. Porting existing databases to open source tools as indicated.
  10. Assisting centers that wish to expand their high performance computational resources.
  11. Assisting in the development of tools to utilize NCI API's to authenticate and utilize databases across a data grid.
  
6. Mechanisms for providing architecture standards and guidance throughout the caBIG project
  1. Utilizing the OGSi allows NCI the confidence that future grid expansion will coincide with progressive developments in the grid space. It also allows for the rapid dissemination of existing standards to members of the project. NCI should not have to reinvent the wheel.
  2. Using Globus will allow for a rapid roll-out of the architecture itself. This speed in roll-out will make test beds available for developers to work on in a real-life scenario.
  3. Globus is a toolkit, not a complete solution. In addition to Globus, essential authentication development, e.g., Shibboleth, will be required for NCI tools to interact with disparate databases
  4. Lombardi is in a position to set up the basic Globus grid and host the certificate server in a comparatively short period of time. Lombardi representatives can then expand grid connections via telephone and web support and local site visits to Cancer Centers. Once this grid is running, it becomes a common platform for developers to complete their work.