# *Monitoring User-Level Grid Functionality and Performance using*
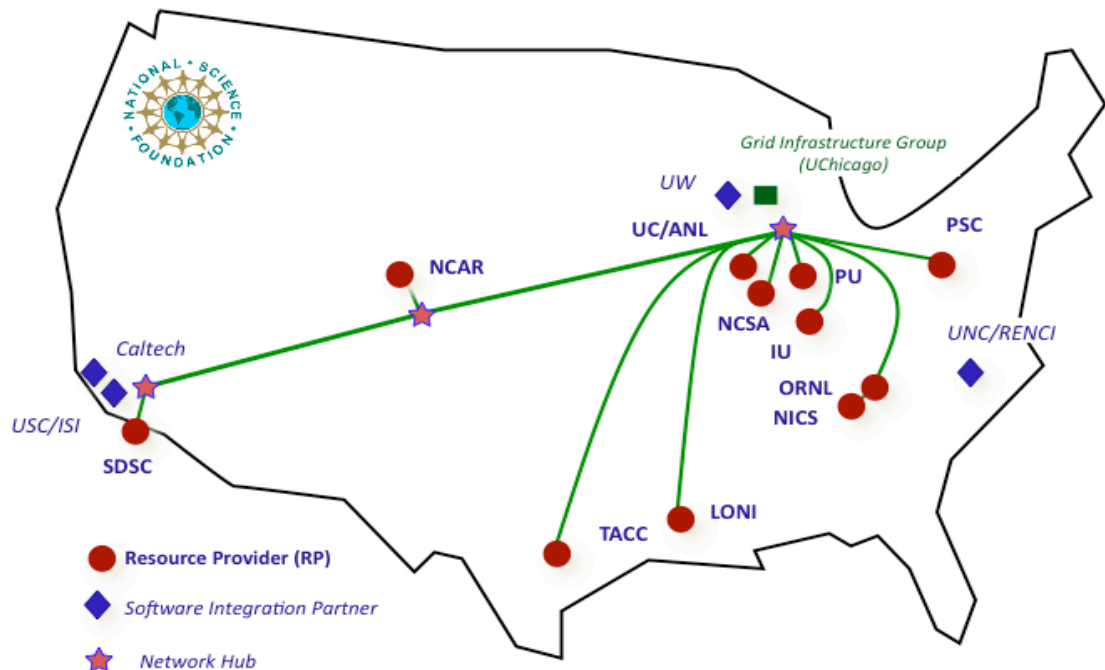


# Shava Smallen

ssmallen@sdsc.edu

May 15, 2008

**SDSC** **SAN DIEGO SUPERCOMPUTER CENTER**

# Goal: reliable grid software and services for users

- Over 750 TF

- Over 30 PB of online and archival data storage

- Connected via dedicated multi-Gbps links

- 30-63 software packages and 6-23 services per resource
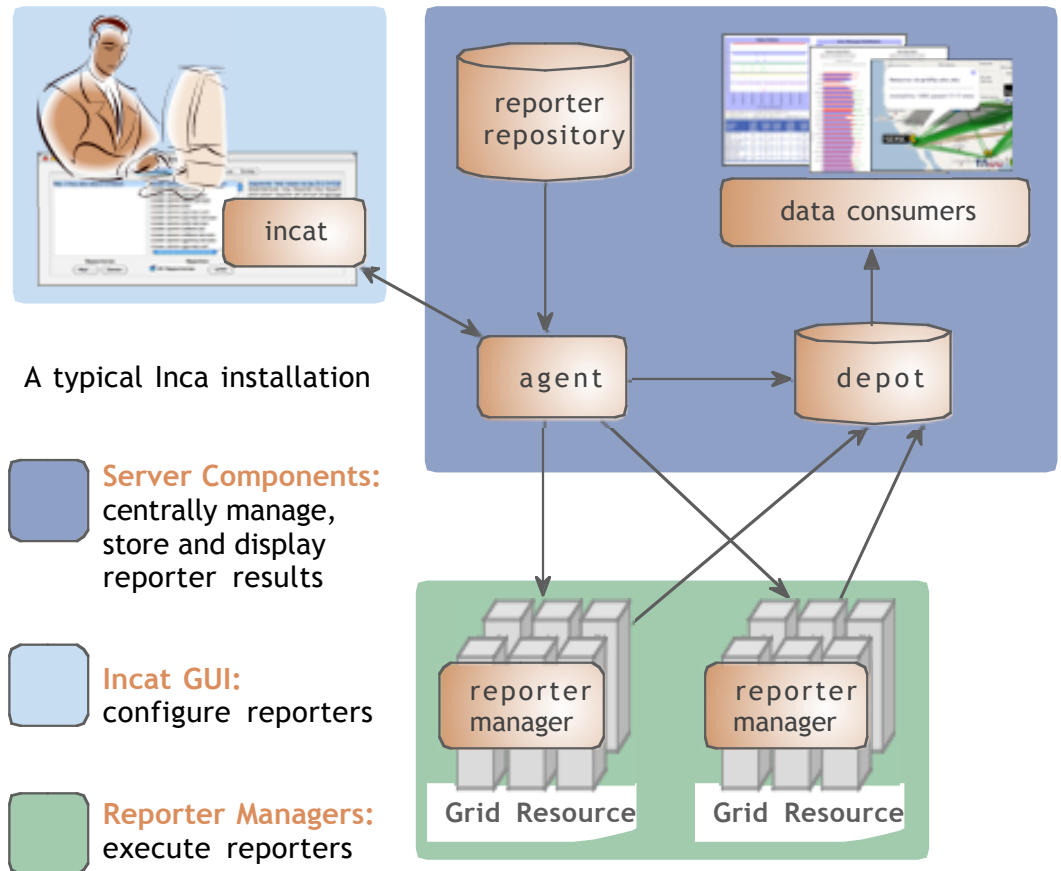


*11 TeraGrid sites, 21 resources*

# *User-level grid monitoring*

- Runs from a standard user account
- Executes using a standard GSI credential
- Uses tests that are developed and configured based on user documentation
- Centrally manages monitoring configuration
- Automates periodic execution of tests
- Verifies user-accessible Grid access points
- Easily updates and maintains monitoring deployment
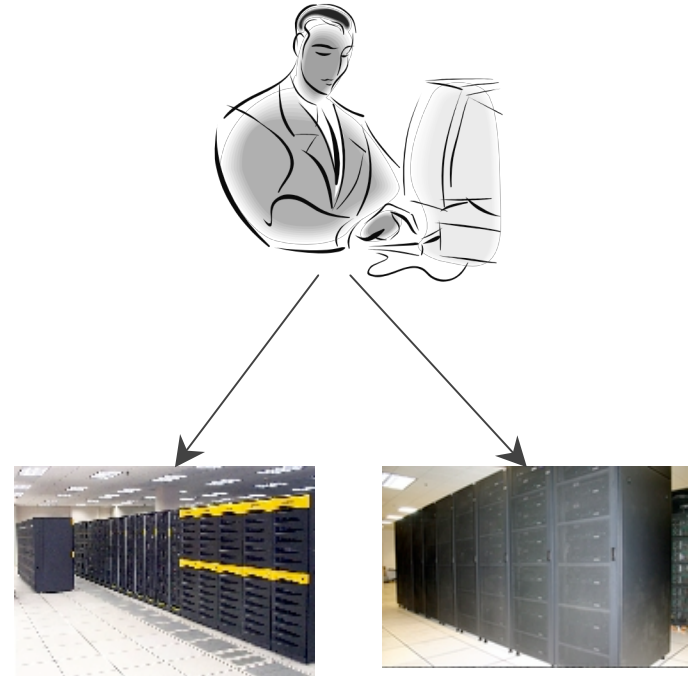
# *Inca features and architecture*

- Stores and archives a wide variety of monitoring results

- Captures context of monitoring result as it is collected

- Eases the writing, deploying, and sharing of new tests or benchmarks

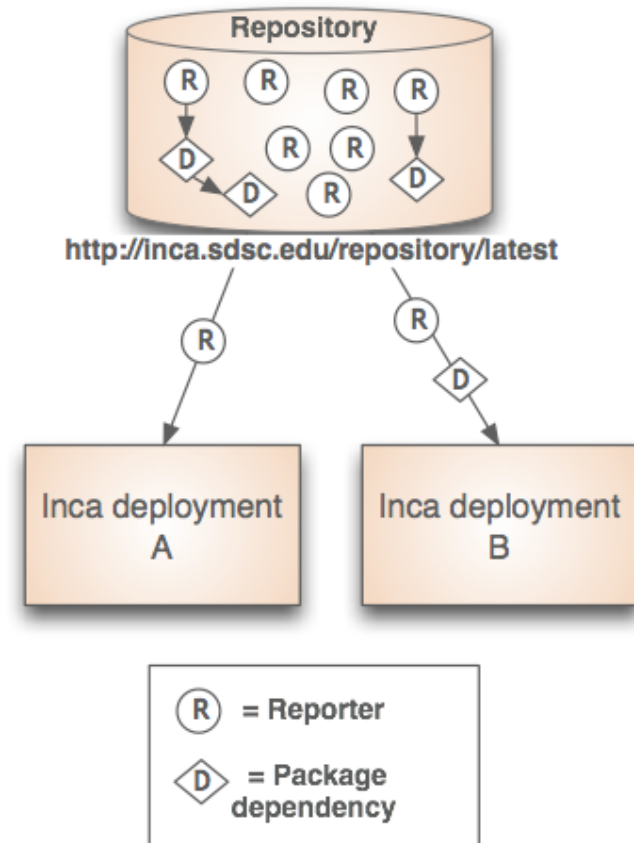- Flexible and comprehensive web status pages

- Secure

A typical Inca installation

**Server Components:** centrally manage, store and display reporter results

**Incat GUI:** configure reporters

**Reporter Managers:** execute reporters

reporter repository

data consumers

incat

agent

depot

reporter manager

reporter manager

Grid Resource

Grid Resource

SDSC

inca

# *Reporters* collect monitoring data

- Executable programs that measure some aspect of the system or installed software

- Supports a set of command-line options and writes XML to stdout

- Schema supports multiple types of data

- Extensive library support for perl and python scripts (most reporters < 30 lines of code)

- Independent of other Inca components



**SDSC** **SAN DIEGO SUPERCOMPUTER CENTER**
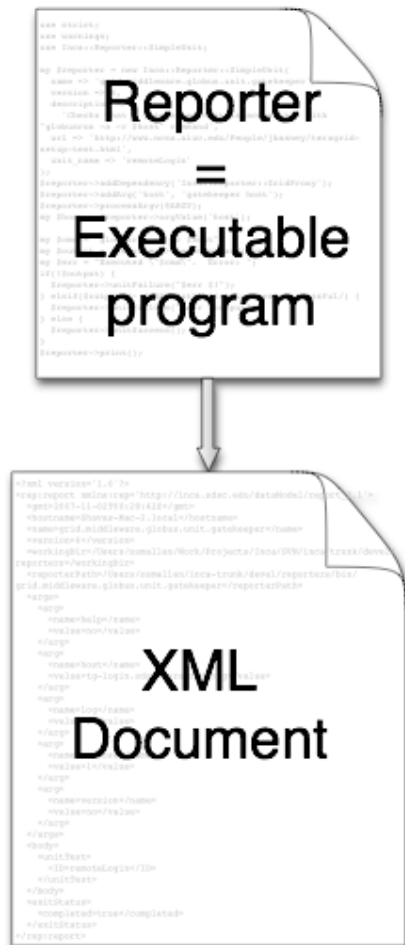
inca

# *Repositories support sharing*

- Collection of reporters available via a URL

- Supports package dependencies

- Packages versioned to allow for automatic updates

- Inca project repository contains 150+ reporters

  - Version, unit test, performance benchmark reporters

  - Grid middleware and tools, compilers, math libraries, data tools, and viz tool



**Repository**

http://inca.sdsc.edu/repository/latest

Inca deployment A

Inca deployment B

(R) = Reporter

(D) = Package dependency

SDSC

inca

# *Agent* provides centralized configuration and management

- Implements the configuration specified by Inca administrator

- Stages and launches a reporter manager on each resource

- Sends package and configuration updates

- Manages proxy information

- Administration via GUI interface (incat)
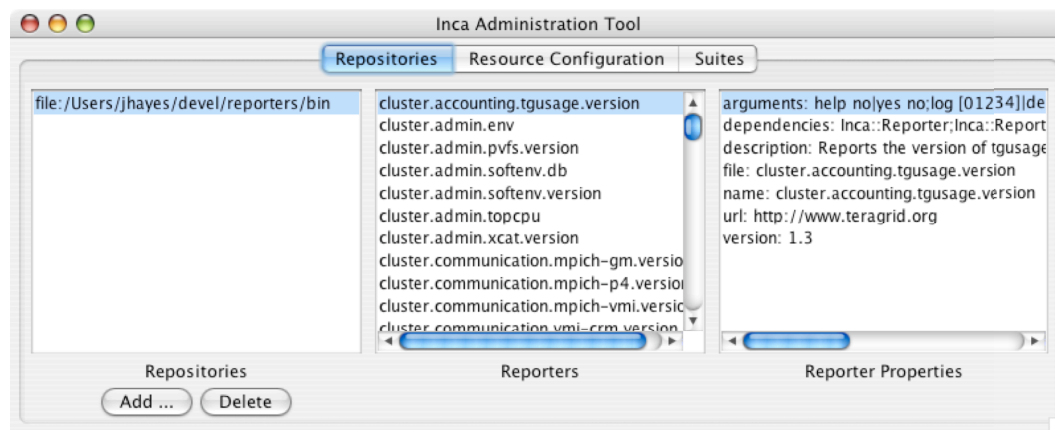


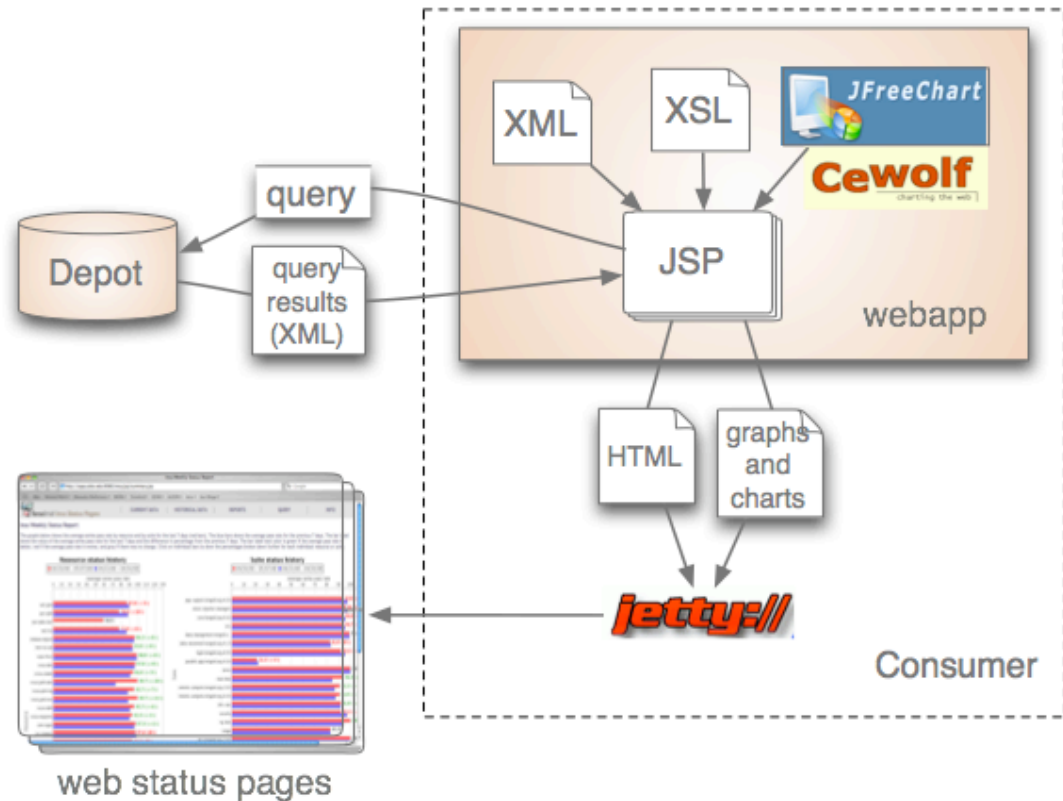*Screenshot of Inca GUI tool, incat, showing the reporters that are available from a local repository*

**SDSC** **SAN DIEGO SUPERCOMPUTER CENTER**

# *Depot* stores and publishes data

- Stores configuration information and monitoring results

- Provides full archiving of reports

- Uses relational database backend via Hibernate

- Supports HQL and predefined queries

- Supports plug-in customization (e.g., email notifications, downtimes)

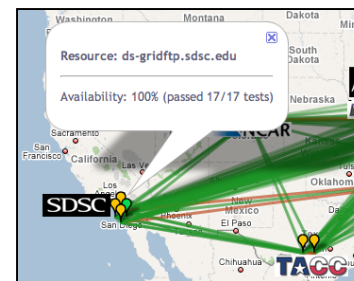- Web services - Query data from depot and return as XML
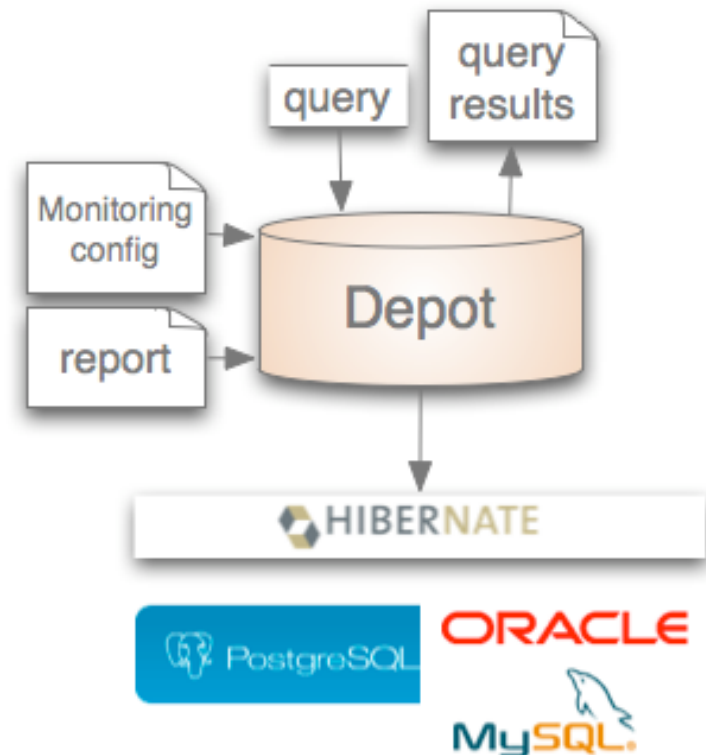
# *Consumer* displays data

- Current and historical views

- Web application packaged with Jetty

- JSP 2.0 pages/tags to query data and format using XSLT

- CeWolf/JFreeChart to graph data



web status pages

**Tests Summary**

*Average test pass rate*

*Cumulative test status by resource*

Resource: ds-gridftp.sdsc.edu
Availability: 100% (passed 17/17 tests)

*Summary of errors this week*

**Suite status history**

■ 05/01/08 – 05/08/08   ■ 04/24/08 – 05/01/08

Average series pass rate

| Suite | pass rate |
|---|---|
| app-support.teragrid.org-4.0.0 | 93.1% (−2%) |
| check-reporter-managers | 96.3% |
| core.teragrid.org-4.0.0 | 93.6% (−1%) |
| ctss | 93.9% (−1%) |
| data-management.teragr... | 99.4% |
| data-movement.teragrid.o... | 85.0% (+3%) |
| login.teragrid.org-4.0.0 | 91.3% (−4%) |
| parallel-app.teragrid.org-4.0.0 | 23.5% (+1%) |
| proxy | 100.0% |
| real-time | 92.9% (+8%) |
| remote-compute.teragrid.o... | 90.6% (+4%) |
| remote-compute.teragrid.o... | 90.0% (+3%) |
| sdsc-ops | 91.9% (+0%) |

■ indiana-bigred (gridftp-auth)

globus-url-copy -len 1280 file:///dev/zero gsiftp:/
error: an end-of-file was reached

*Test status by package and resource*

*Related test histories*

*Individual test history*

*Individual test result details*

**Test Details**

*Historical*        *Current status*
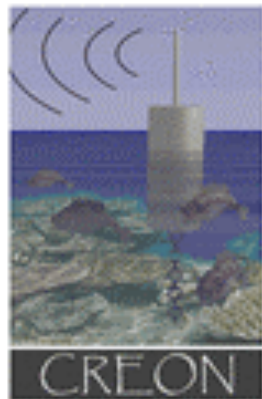
# *Software status and deployments*

Current software version: 2.3
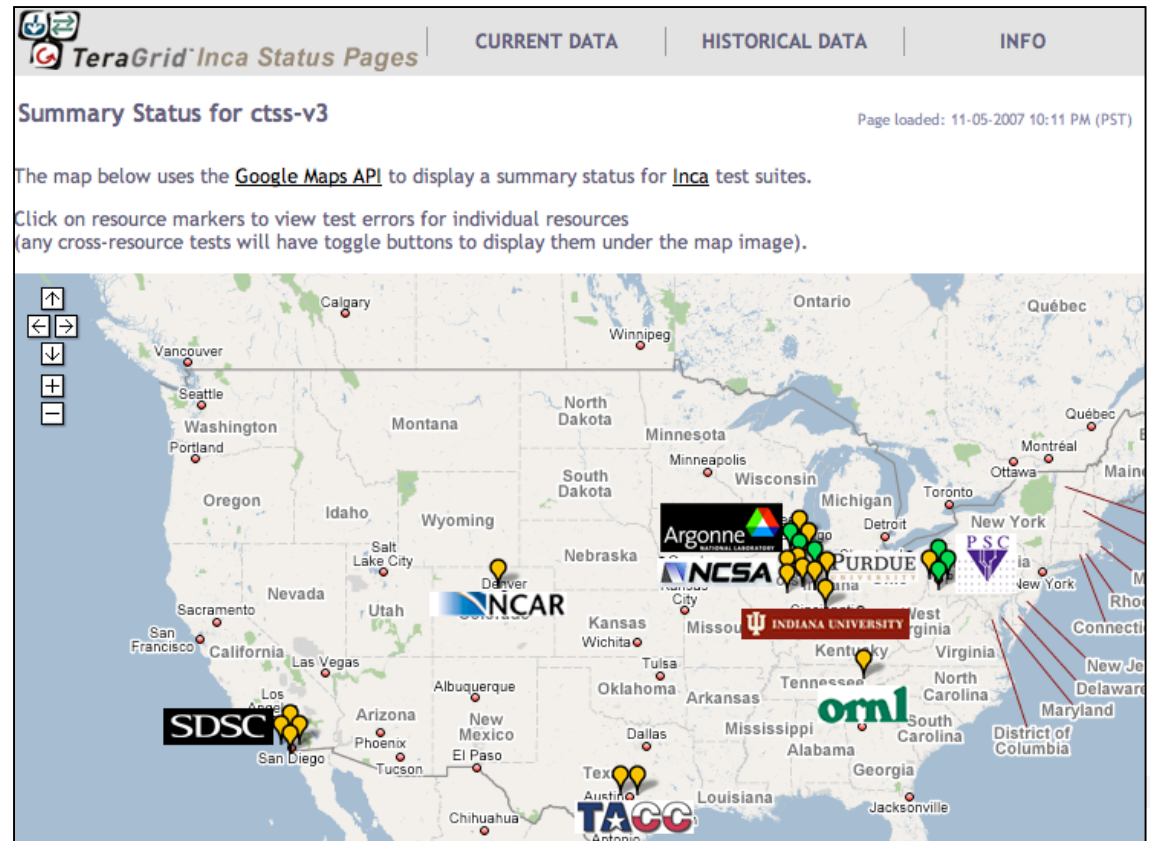
(available from Inca website)

http://inca.sdsc.edu

# *Inca TeraGrid deployment*

- Running since 2003

- Testing for CTSS

- Cross-site tests

- GRAM usage

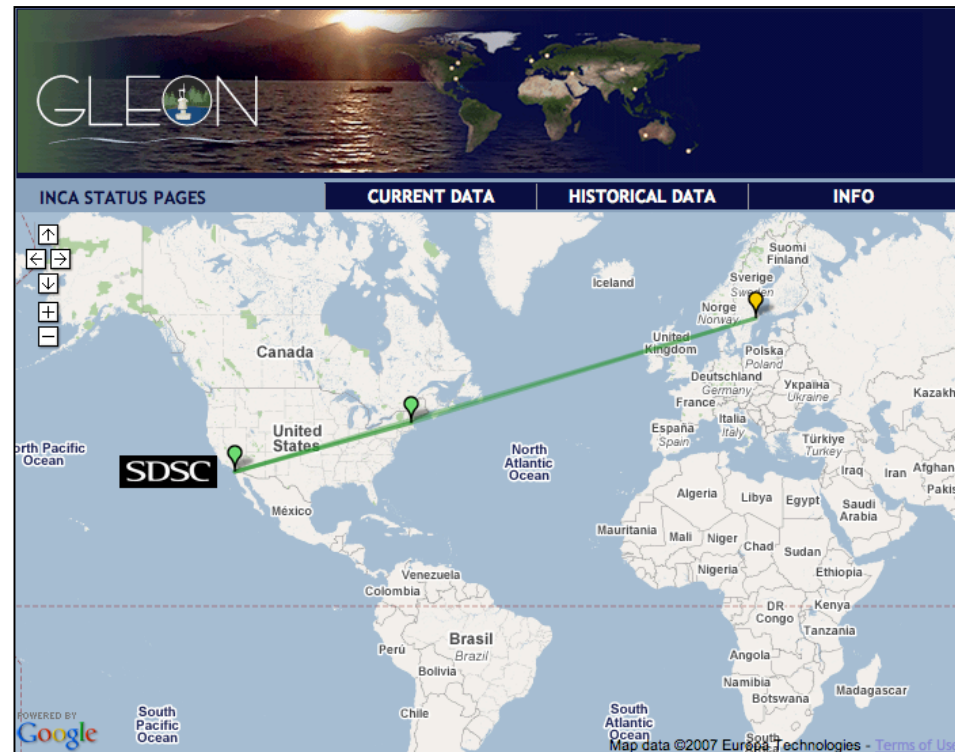- CA certificate and CRL checking

- Resource registration in MDS



*Screenshot of Inca status pages for TeraGrid*

http://inca.teragrid.org/

# *Inca GLEON deployment*

- Sensors in lake: dissolved oxygen level, temperature, velocity (some), etc.

- Monitoring Data Turbine deployments since Oct. 24

- Currently deployed for Lake Sunapee and Lake Erken



OPEN SOURCE DATA TURBINE INITIATIVE

*Empowering the Scientific Community with Streaming Data Middleware*

# *Benefits of using Inca*

- Detect problems before the users notice them

- Easy to write and share tests and benchmarks

- Easy to deploy and maintain

- Flexible and comprehensive displays
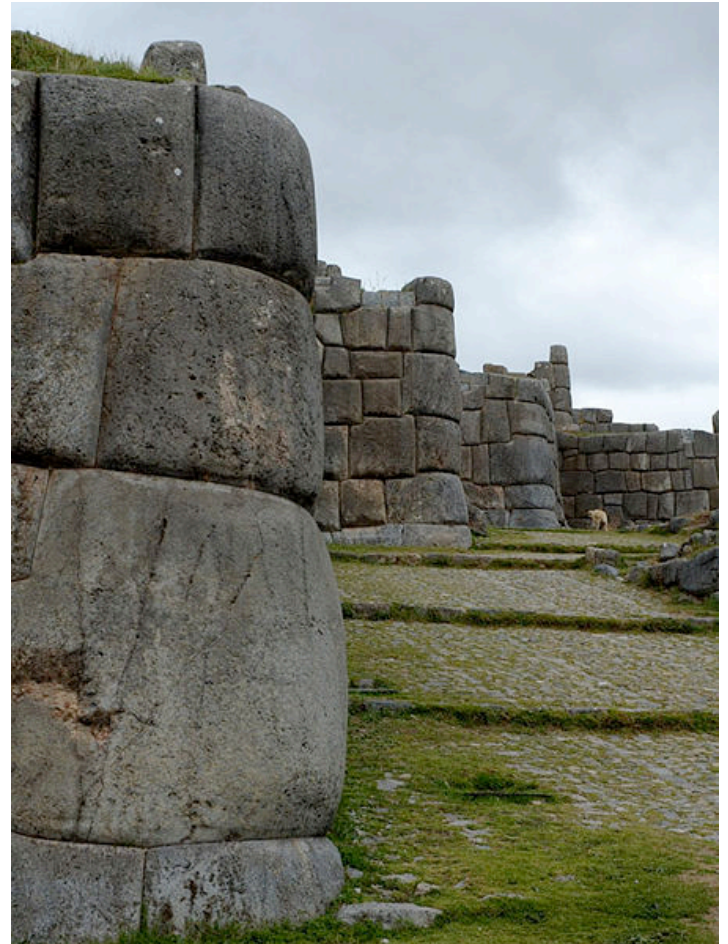
# *More information*

**Website:**
http://inca.sdsc.edu

**Announcements:**
inca-users@sdsc.edu

**Email:**
inca@sdsc.edu

**Funded by:**