

Crossing the Grid-Web “Boundary”: the LIGO Perspective

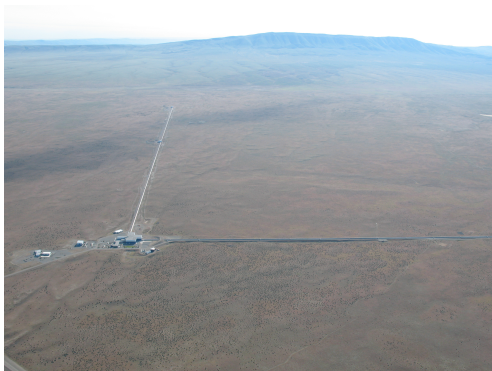
Scott Koranda for LIGO

LIGO and University of Wisconsin-Milwaukee

April 26, 2010
LIGO-G1000491-v1



LIGO Science Mission



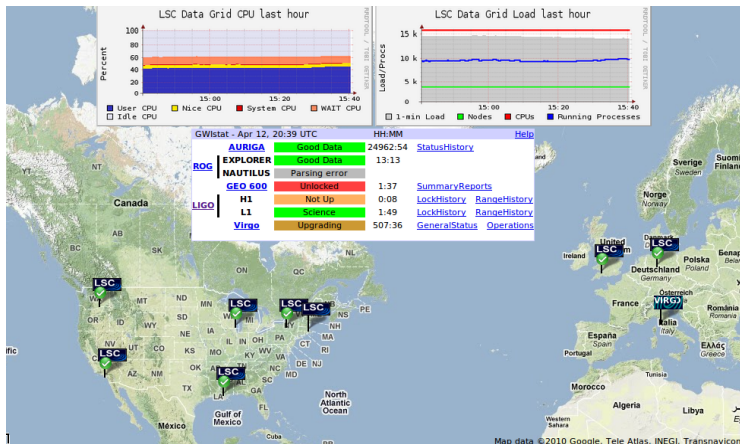
LIGO, the Laser Interferometer Gravitational-wave Observatory, seeks to detect gravitational waves – ripples in the fabric of spacetime. First predicted by Einstein in his theory of general relativity, gravitational waves are produced by exotic events involving black holes, neutron stars and objects perhaps not yet discovered.

LIGO Scientific Collaboration



The LIGO Scientific Collaboration (LSC) is a self-governing collaboration seeking to detect gravitational waves, use them to explore the fundamental physics of gravity, and develop gravitational wave observations as a tool of astronomical discovery. The LIGO Scientific Collaboration was founded in 1997 and currently has more than 800 members from 70 institutions worldwide.

LIGO Data Grid (LDG)



- ▶ 15000+ cores
- ▶ 10 sites
- ▶ Many flavors of data and metadata services
- ▶ > 200 users

LIGO Data Grid: Success



LIGO data replicated across all computing sites

Single X.509 certificate enables SSO across LDG

LIGO post-doc with a single command can...

- ▶ submit 1,120,600 job data analysis workflow
- ▶ query LIGO metadata and data services
- ▶ generate a terabyte of results

Many, many post-docs and students running these types of workflows all the time

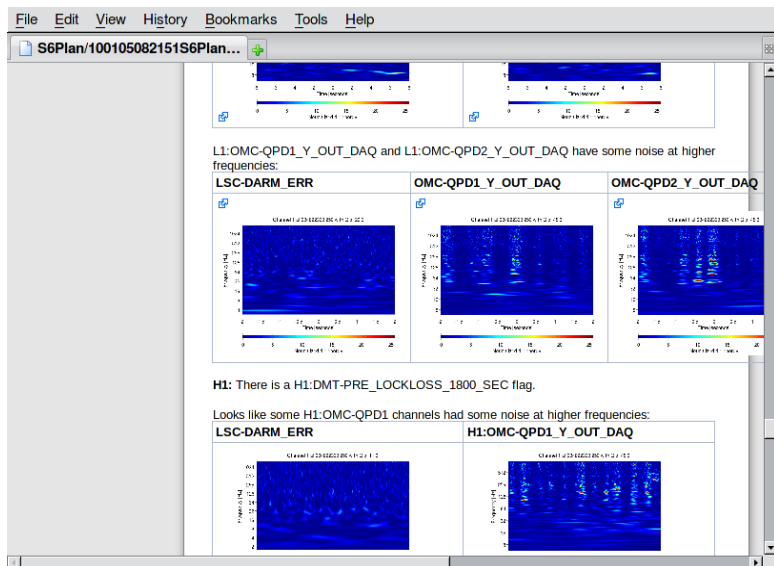
LIGO Data Grid: It's Not Enough

Sure, any single scientist use LDG to generate terabytes of output

Not the end of the “science” work

Scientists must collaborate, investigate, follow up results...

Wiki is tool used at this time



Seamless identity & credential management critical

- ▶ Scientists just want to use tools
- ▶ Don't care if “web” or “grid” (or “cloud”)
- ▶ Typical use case:
 - ▶ Submit large workflow to grid
 - ▶ Jobs run for week analyzing data
 - ▶ Workflow generates 1000's of summary images
 - ▶ Need to POST summary into analysis wiki (*not* possible now)
- ▶ Seamless identity & cred management across grid & web
- ▶ Delegation is important
 - ▶ Workflow management systems need to cache and refresh credentials during lifetime of workflow
 - ▶ LIGO working closely with UW Condor team
- ▶ Need I2 and grid communities to build together

Use case: GraCEDb

Gravitational-wave candidate event database

- ▶ “entity” submits event via client tool or web interface
- ▶ unique ID assigned to the candidate event
- ▶ submitter, search group, search type recorded
- ▶ web area created for details and later follow-up
- ▶ wiki page created for annotating candidate event
- ▶ alert published to corresponding node in LVAlert system
 - ▶ subscribers node receive alert and initiate follow-ups
 - ▶ XMPP based

Need better collaboration tools for data intensive science

Always prefer to always leverage existing tools

Are existing wikis the best collaboration tools?

- ▶ Tedious to link relevant information together
- ▶ Search is poor
- ▶ Version control is poor

“Search” (ala Google) is limited for collaborative science

Curation still largely ignored...need “e-Science librarians”?

What tools would a LIGO data and results librarian use?