

Turning Software Projects into Production Solutions

Dan Fraser, PhD
Production Coordinator
Open Science Grid

OU Supercomputing Symposium
October 2009

● Four Questions

- What gives some software projects “staying power” while others stagnate ?
- What is the most important thing for a sustainable software project ?
- What is the difference between a software project and a software solution ?
- What is needed to create a “production” quality solution ?



My background

- LANL, Air Force Supercomputing, Thinking Machines, NEC, Sun, Paremus Ltd., Globus, Open Science Grid
- Software development, software support, user and application support, software scalability, production system management...
- I have never fully appreciated the answers to these questions until I worked on Production



Open Science Grid

And I'm still learning ...



Open Science Grid

● Software with “Staying Power”

Microsoft Windows

How many of you
run linux as the
primary os on your
laptop ?

Mac does not count
It solves the same
“problem” as windows

GridFTP, GRAM2,
MyProxy, Nagios,
Ganglia, Rocks,
CFEngine, SSH ...



Open Science Grid



Open Science Grid

Question 1

- What gives some software projects “staying power” while others stagnate ?



Question 1

- What gives some software projects “staying power” while others stagnate ?
- The answer is NOT:
 - Funding
 - Microsoft did not get big because it had lots of cash
 - MyProxy has not been funded for over 5 years.
 - But the team still fixes bugs on weekends...
- Some answers include:
 - It solves a problem that users need solved
 - It “works” in the user environment
 - The community supports it

● Solutions that “work”

When software solutions “work”, even though they are not perfect, it is hard to get users to migrate away from them

Anyone remember
plot10 ?



Open Science Grid



Open Science Grid

Question 2

- What is the most important thing for a sustainable software project ?



Question 2

- What is the most important thing for a sustainable software project ?
 - Answer: Users.
 - Production \equiv Users
 - The “users” are not always who we think

GridFTP

- A well engineered production code
- Used in all the High Energy Physics experiments + many other places
- Easy to install (make install gridftp)
- We did not always understand end user difficulties.
 - “Works great one day, try again one week later and it fails”
 - Problem with CRLs (security issues)
 - GSI is the software “environment”

GridFTP is most successful

- In places where the security environment is well maintained
- Where end users don't necessarily see GridFTP directly.
 - TeraGrid developed TGCP, a friendly user interface to GridFTP
 - UberFTP and other interfaces widely used
 - GridFTP is an *embedded* component at all the HEP site
 - OSG packages GridFTP as part of a complete solution (a Storage Element) including security

The main “user” is often the “integrator”



● Integration is Critical

Software integration is not a dirty word, but it can be dirty work.

Integration is a critical component in production solution development



Open Science Grid

Find an Integrator to work with !



Open Science Grid

Question 3

- What is the difference between a software project and a software solution?
- A solution is something greater than the sum of its software components
- Examples:
 - GridFTP is a software project
 - UI + GridFTP + Authz framework + Credential mgmt = a solution
 - GRAM2 is a software project
 - Condor-G + Gram2 + GridMon = a solution
 - Ganglia is a software project that does monitoring
 - Ganglia + something that acts on the data = a solution



Solutions create new problems

- Multiple components must work together
 - Which components do I need?
 - Which versions work together?
 - I need a new feature. Can I upgrade one component, or will it break everything ?
 - When a solution breaks, which component is to blame ?
 - How do I get resolution when component providers are blaming each other ?

● Question 4

What is needed to
create a “production”
quality solution ?

Answer: A good process

Putting components together
that work and solve a problem
for a community

Processes can be
reliable even with weak
components !



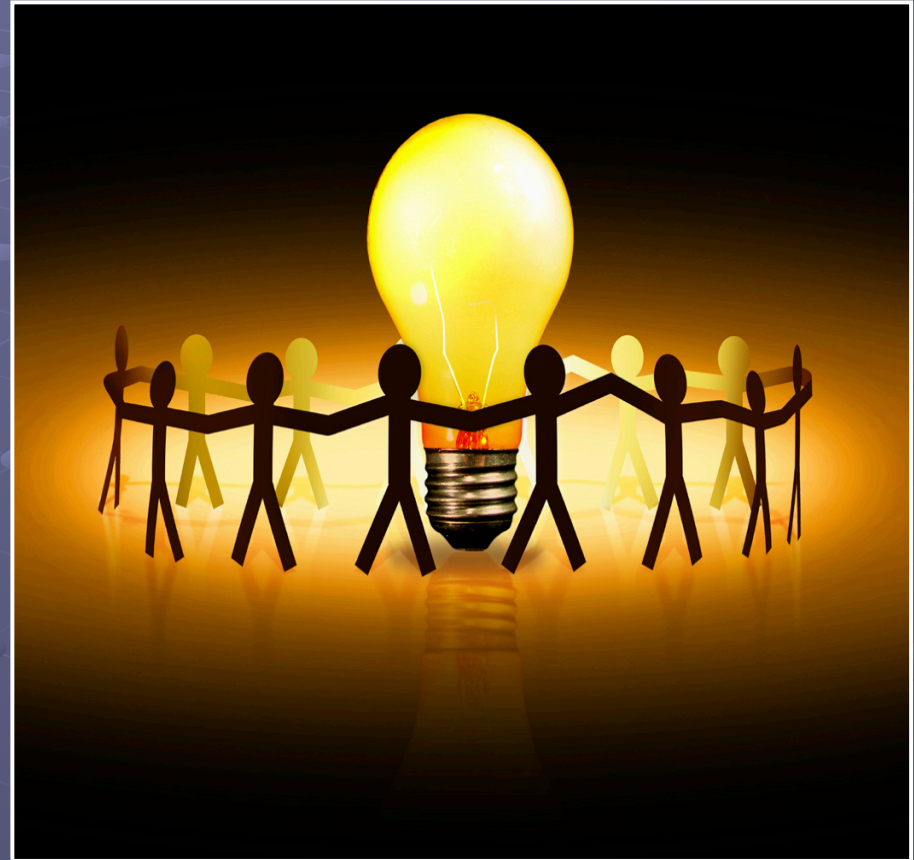
Open Science Grid



Open Science Grid

● What is needed for a Production process?

- Coordination with software providers and communities
- Integration
- Testing
- Monitoring
- Solution Releases
- Operations
- ...



● OSG – A good process example

- Projects in the OSG software stack have good “staying power”
- Projects get into OSG by being demanded by the user community.
- OSG provides the “production” process



OSG Enabled Production Solutions

- *HEP Analysis: HEP, CMS, D0, CDF, ...*
- *Gravitational Wave: #2 in the world*
- *Protein structure prediction: Toyota Institute*
- *Weather Research Forecasting: UNC, U Nebraska*
- *Structural Biology Predictions: Harvard Medical*
- *Nanotechnology Simulation and Modelling: Purdue, Nanohub collaboration.*
- *Molecular Dynamics: U of Buffalo, Argentina*
- *Theoretical Nuclear Physics: Duke University..*
- *Text Mining: U. North Carolina*
- *Exploring Mathematical algorithms: U. Colorado*



On creating production solutions

- Understand what the problems are
 - Understand what the users really want to do
 - Make sure there are enough likely users
 - Understand the barriers to adoption
- Understand the *environment* where the solution will live
- Work with integrators (or be one)
- Know and nurture your user community