Building A Sustainable Cyberinfrastructure In the Campus and Across the Region

John-Paul Robinson
University of Alabama at Birmingham
<jpr@uab.edu>

Cyberinfrastructure

Cyberinfrastructure (CI) consists of computational systems, data and information management, advanced instruments, visualization environments, and ***people, all linked together*** by software and advanced networks to improve scholarly productivity and enable knowledge breakthroughs and discoveries not otherwise possible.

– Developing a Coherent CI, CASC/EDUCAUSE, February 2009

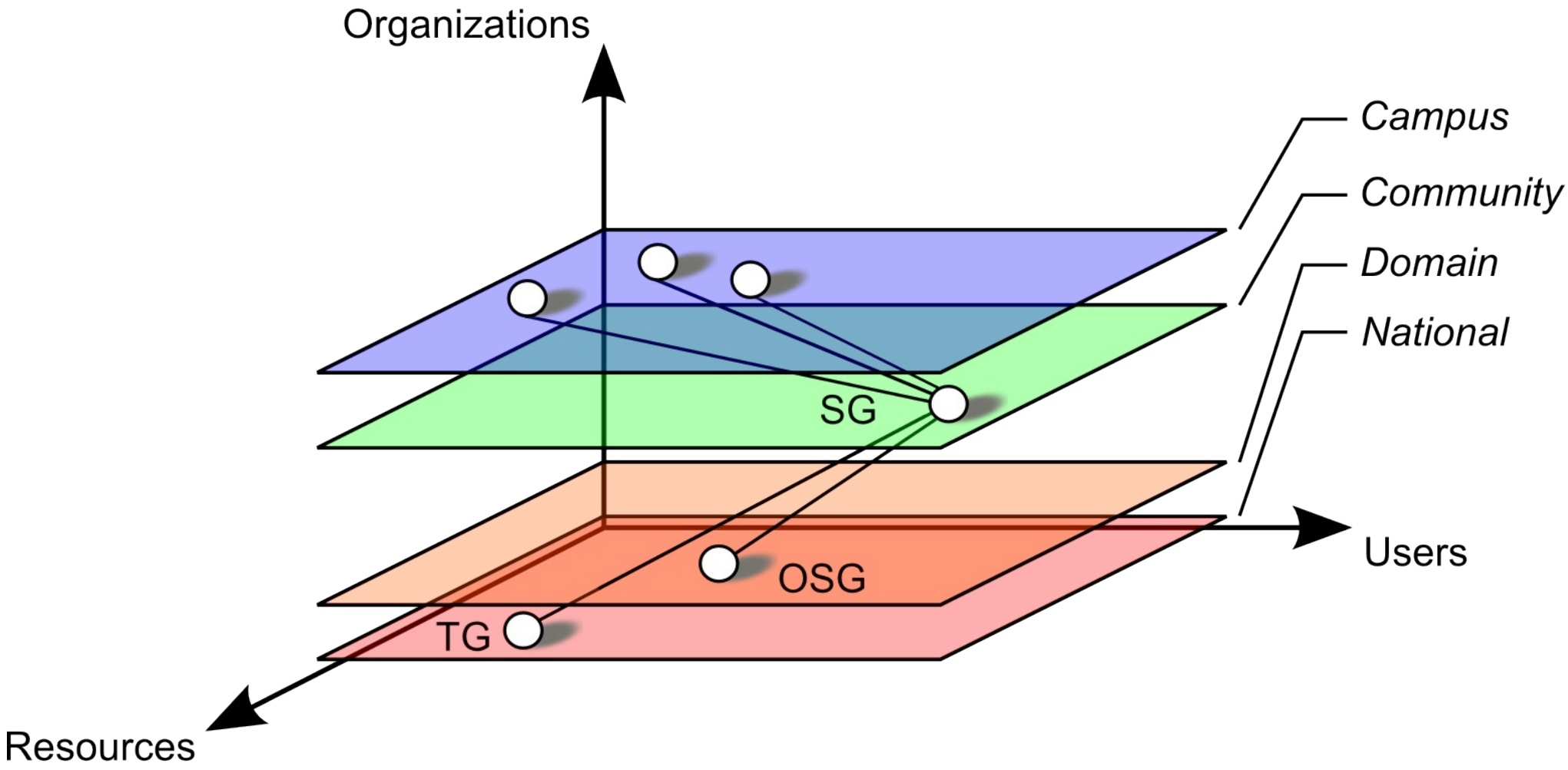
Cyberinfrastructure

- Confusing
- Complex
- Fragile
- Scary
- Myth
- ???

Dimensions of Complexity in CI

- Resources
  - Big, Small, Local, Remote, Instruments, Computers, Data, Oh My!
  - In short, it's everything
- Users
  - Scientists, Teachers, Students
  - Everyone (Implementors, Administrators, Secretaries)
- Organizations
  - Research Group, Campus, Regional, National, International
  - All of 'em
- No wonder CI is so complex!
- Greater Complexity = Greater Cost + Greater Risk
- We Need Order

A Coherent Cyberinfrastructure



TG=TeraGrid, OSG=Open Science Grid, SG=SURAgrid

Success = Less Complexity

- Make it Simple
  - “...but no simpler than it needs to be...” – Einstein
- Organizations Can Scope *Internal* Complexity
  - Set #Organizations = 1
    - (i.e. you make the rules)
  - Deal only with #Users and #Resources
  - Problem Solved!
- Organizations Cannot Scope *External* Complexity (*...or can they?*)

Scoping Complexity

- Sources of External Complexity
  - Established but Divergent Procedures
  - Non-Uniform Talent Distribution
  - Different User Requirements and Support Needs
  - Governance
  - Operational Transparency
  - Trust
- Solutions to External Complexity
  - Operational Transparency
    - (Trust and Autonomy)
  - Consistent Interface to Remote (and Local) Resources
    - (Sustainability)
- We need an Organizing Framework

The Open Source Model

- Overcomes barriers in complex environments
  - Distance, skill, and the many people, many organizations, *and* many resources problem
- Organization around vexing problems
  - Bring experts together to solve real use-cases and identify common needs
- Order through similar project interfaces
  - Overview, Documentation, User and Developer Forums, Bug Tracking, Downloads, About
- Sustainability through open solutions
  - Contributions, derivative works, and autonomy

Open Source Patterns

- Access to accumulated understanding of a process
- Community dialog: mailing lists, forums, wikis
- Use-case discovery through user engagement
- Issue tracking
- Governance
- Freedom to Engage (and Disengage)

Familiar Open Source Communities

- Linux Kernel
  - Good example of competing vendors with self interest to support a common infrastructure on which to build value for their customers
- Apache Software Foundation (ASF)
  - Good model for community governance plus the above
- Mozilla Firefox
  - Good example of an open code base that encourages continued innovation in closed-source competitor products
- SourceForge
  - Good example of common project framework that offers a familiar user experience across projects and open reporting of development metrics
- Globus Toolkit
  - Good example of adopting OSS model to build a shared infrastructure

Applying the Model to Build CI

- Build Open Communities of Practice
- Document CI deployment and implementation best practices
- Software is just codified knowledge of process
- Open implementations support collaboration
- Open development builds trust and collaboration
- Transparency protects autonomy

CI Practices are the Source

- Campus bridging is software too!
  - We're just not accustomed to “seeing” the code at such high levels: instantiate project wiki, add authorization zone (myCollaborators), post updates
- Program CI in high-level business process abstractions not assembly languages
  - This is the level at which we should be programming – instead we get stuck at the technology point of “assembly languages” (check this, move that, jump here, jump there).

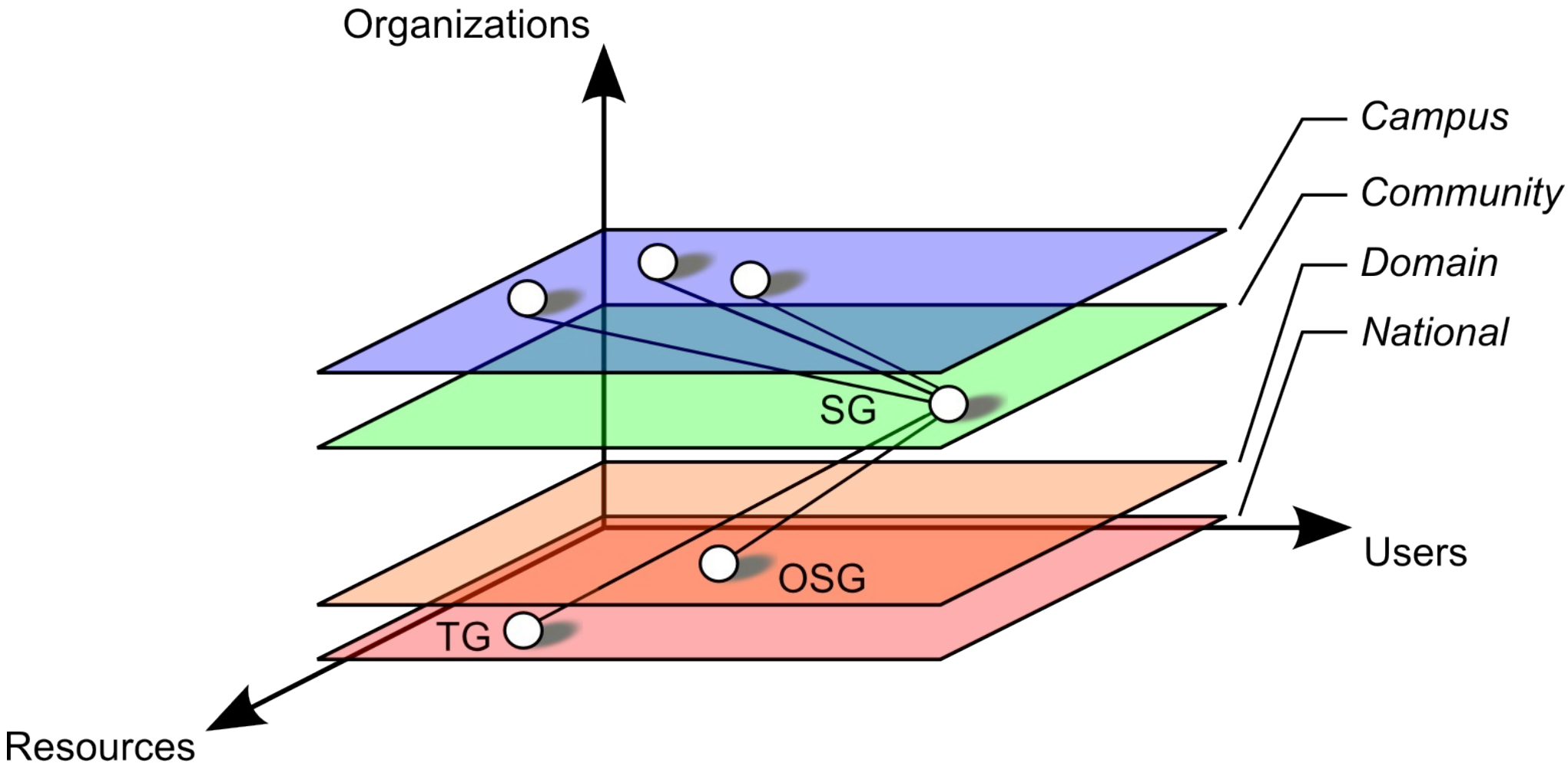
Benefits of Model

- Modularity
  - Supports Extensibility
  - Address Local Needs
- Sustainability
  - Internalized Services Support
  - Self-interest in maintaining operating infrastructure
- Reproducibility
  - Codified Environments
  - Codified deployment supportl software provenance
- Natural selection of effective solutions

Structure for Communities of Practice

- Open Interfaces to Campus
  - UABgrid – an open development environment for campus CI development
  - Projects fill campus needs by addressing community use-cases
- Open Cooperation Across Campuses
  - SURAgrid – an open community of practice for CI deployment at campus
  - A success of the NSF Middleware Initiative (circa 2003)
  - Governance by community members (circa 2006)
  - Developed 4 year Strategic Plan for Deploying CI across member sites (August 2008)
    - Strategic Goals: Outreach, Infrastructure, Communications, Sustainability, Corporate Partnership

A Coherent Cyberinfrastructure



TG=TeraGrid, OSG=Open Science Grid, SG=SURAgrid

Challenges

- Need to grow community
  - We have supported this because we believe in the vision and have experienced the benefits (we have been building with shoe-string budgets)
  - We need growth capital to bridge the gap to maturity – where community operations are self-sustained because they are internal to the business operations of the member organizations
- Significant barriers in established operating models
- Requires willingness to work with new organizing models – *collaborate* not *control*

What's Needed?

- Expand the community of implementors who will integrate CI into campus and collaboratively build solutions
- Support for this community and the process
- Measure success and growth of community
- Adjust & Repeat
- Core CI “comes out in the wash”...it's the foundation that everyone builds on

Further Reading

- SURAgrid Strategic Plan
- UABgrid Development Project
- Developing A Coherent Cyberinfrastructure (CASC/Educause)
- NSF Software Sustainability Workshop
- Coming Soon: SURAgrid white paper on building CI communities of practice

Acknowledgments

- SURAgrid Collaborators
  - Gary Crane, SURA
  - Phil Smith, Texas Tech University
  - Art Vandenberg, Georgia State University
  - Mary Fran Yafchak, SURA
- University of Alabama at Birmingham Office of the Vice President for Information Technology