



The Open Science Grid

Talking Points

OSG Executive Board
All Hands Plenary
Wednesday 5th March 2008



Supported by the Department of Energy Office of Science SciDAC-2 program from the High Energy Physics, Nuclear Physics and Advanced Software and Computing Research programs, and the National Science Foundation Math and Physical Sciences, Office of CyberInfrastructure and Office of International Science and Engineering Directorates.





Why We are Here Today?

Have a dialogue about how the Open Science Grid is organized, managed and operates.

Have a dialogue on what this means in the wider sense as a contributor to local, national and international cyberinfrastructures, scientific and research communities.

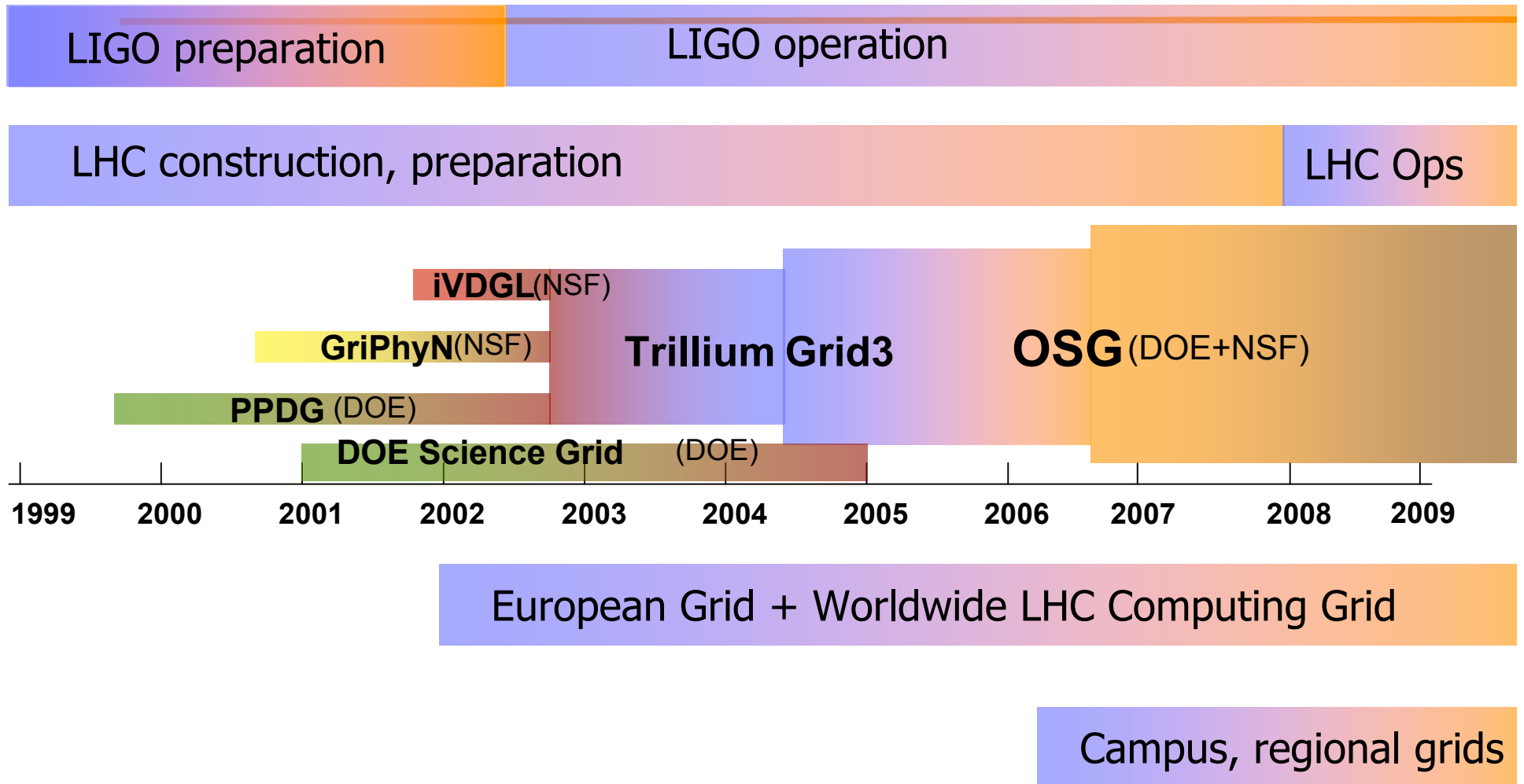
What can OSG learn from your experience?

How should we go about identifying and improving “best practices” in our organization, management and operation?



Open Science Grid

The Evolution of the OSG





The Open Science Grid vision

Transform processing and **data** intensive science through a cross-domain self-managed **national distributed** cyber-infrastructure that **brings together** campus and community infrastructure and facilitating the needs of **Virtual Organizations (VO)** at all scales



The Three Building Blocks

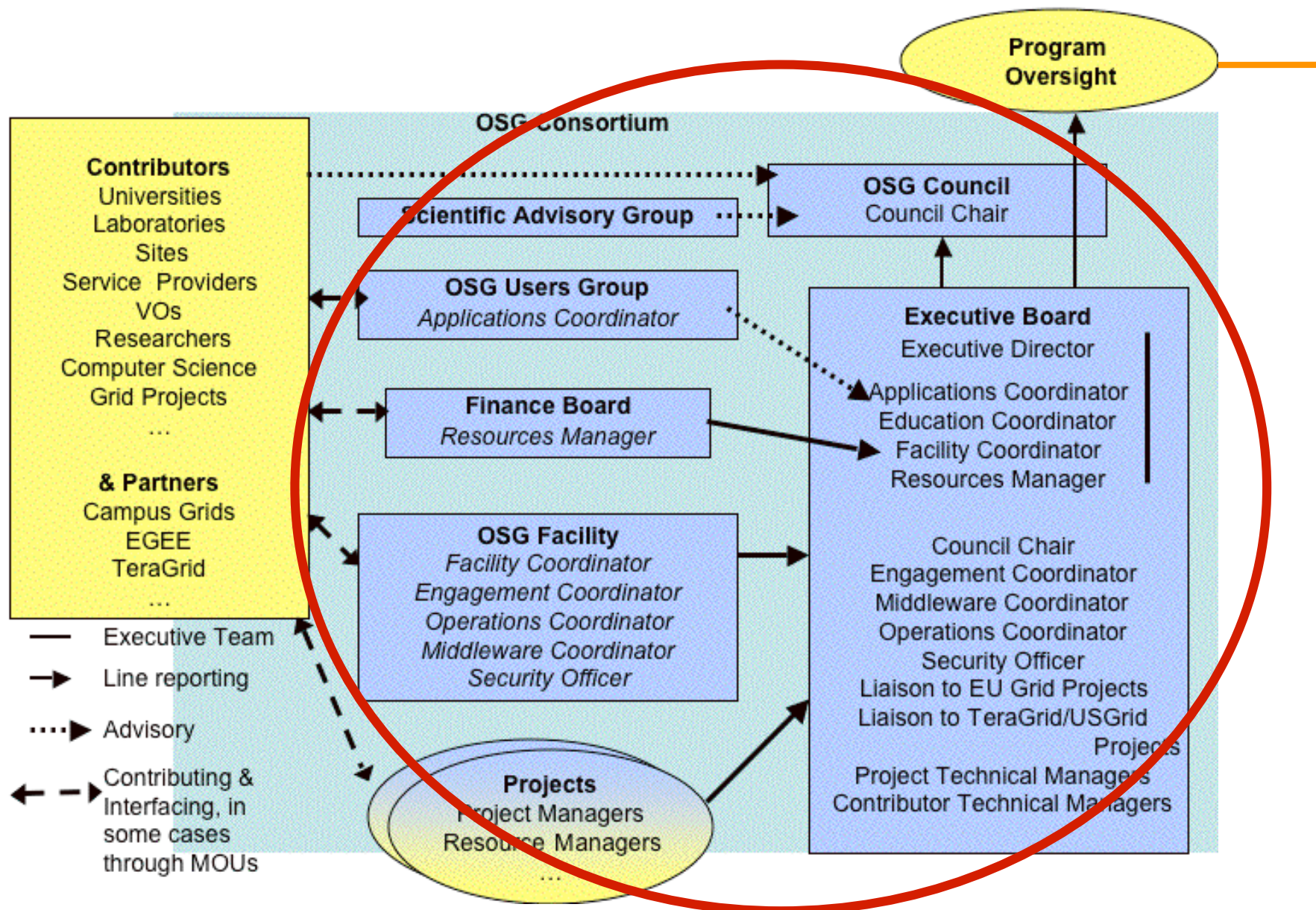
The OSG organization, management and operation is structured around three components:

the **Consortium**

the **Project**

the **Facility**

Structure of the Consortium



Academia Sinica	Kyungpook National University	University of Arkansas
Argonne National Laboratory (ANL)	Laser Interferometer Gravitational Wave Observatory (LIGO)	Universidade de São Paulo
Boston University	Lawrence Berkeley National Laboratory (LBL)	Universidade do Estado do Rio de Janeiro
Brookhaven National Laboratory (BNL)	Lehigh University	University of Birmingham
California Institute of Technology	Massachusetts Institute of Technology	University of California, Davis
Center for Advanced Computing Research	National Energy Research Scientific Computing Center (NERSC)	University of California, Riverside
Center for Computation & Technology at Louisiana	National Taiwan University	University of California, San Diego
The State University of New York at Buffalo	New York University	University of Chicago
Center for High Performance Computing at the University of New Mexico	Notre Dame University	University of Connecticut Health Center
Clemson University	Oak Ridge National Laboratory	University of Florida
Collider Detector at Fermilab (CDF)	OSG Grid Operations Center (GOC)	University of Illinois at Chicago
Columbia University	Pennsylvania State University	University of Michigan
Condor Project	Purdue University	University of Nebraska - Lincoln
Cornell University	Renaissance Computing Institute	University of New Mexico
DZero Collaboration	Rice University	University of North Carolina
Fermi National Accelerator Laboratory (FNAL)	Rochester Institute of Technology	University of Northern Iowa
Florida International University	São Paulo Regional Analysis Center (SPRACE)	University of Oklahoma
Georgetown University	Sloan Digital Sky Survey (SDSS)	University of Rochester
The Globus Alliance	Solenoidal Tracker at RHIC (STAR)	University of South Florida
Hampton University	Southern Methodist University	University of Texas at Arlington
Harvard University	Stanford Linear Accelerator Center (SLAC)	University of Virginia
Indiana University	State University of New York at Buffalo	University of Wisconsin-Madison
Indiana University-Purdue University,	Syracuse University	University of Wisconsin-Milwaukee Center for Gravitation and Cosmology
Wayne State University	Texas Tech University	US ATLAS
	Thomas Jefferson National Accelerator Facility	US CMS
		Vanderbilt University



OSG Challenges

- Develop the **organizational** and management structure of a consortium that drives such a Cyber Infrastructure
- Develop the **organizational** and management structure for the project that builds, operates and evolves such Cyber Infrastructure
- Maintain and evolve a **software stack** capable of offering powerful and dependable capabilities that meet the science objectives of the NSF and DOE scientific communities
- Operate and evolve a dependable and well managed **distributed facility**



OSG Sites and VOs

- A **Site** is a collection of computing and/or storage **resources** and **software**. Sites are autonomous and manage their own policies, security and usage.
- A **Virtual Organization** (VO) is a collection of **people** (VO members), **resources** and **software**.
- OSG brings together many VOs to
 - Allow for more **effective use** of their collective resources
 - Allow easier use of **shared guaranteed** resources that are distributed.
 - **Opportunistic sharing** of distributed resources.



The Project

The Facility, organized in six groups:

- **Engagement** – Identify and support new groups
- **Integration** – Transition the OSG software stack to deployment
- **Operation** – Monitor activities and support VOs and sites
- **Security** – define, implement and monitor the security plan of the OSG
- **Software** – evolve, package and support the VDT
- **Troubleshooting** – work with sites and VOs to resolve “problems” in end-to-end functionality

Education and Training.

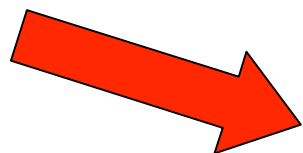
Extensions and Users Group.

Communication and administration.

Becoming a Full OSG Citizen

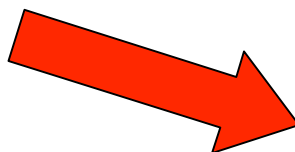
Join the OSGEDU VO:

Run small applications after learning how to use OSG from schools



Be part of the Engagement program and Engage VO:

Support within the Facility to bring applications to production on the distributed infrastructure



Be a standalone VO and a Member of the Consortium:

Ongoing use of OSG & participate in one or more activity groups.



Open Science Grid

OSG Software Release

Domain science requirements

Condor, Globus,
Privilege,
EGEE, ...

OSG stakeholders and middleware
developer (joint) projects.

Test on “VO specific grid”

Integrate into VDT Release.
Deploy on OSG integration grid
Interoperability testing

Provision in OSG release &
deploy on OSG sites.

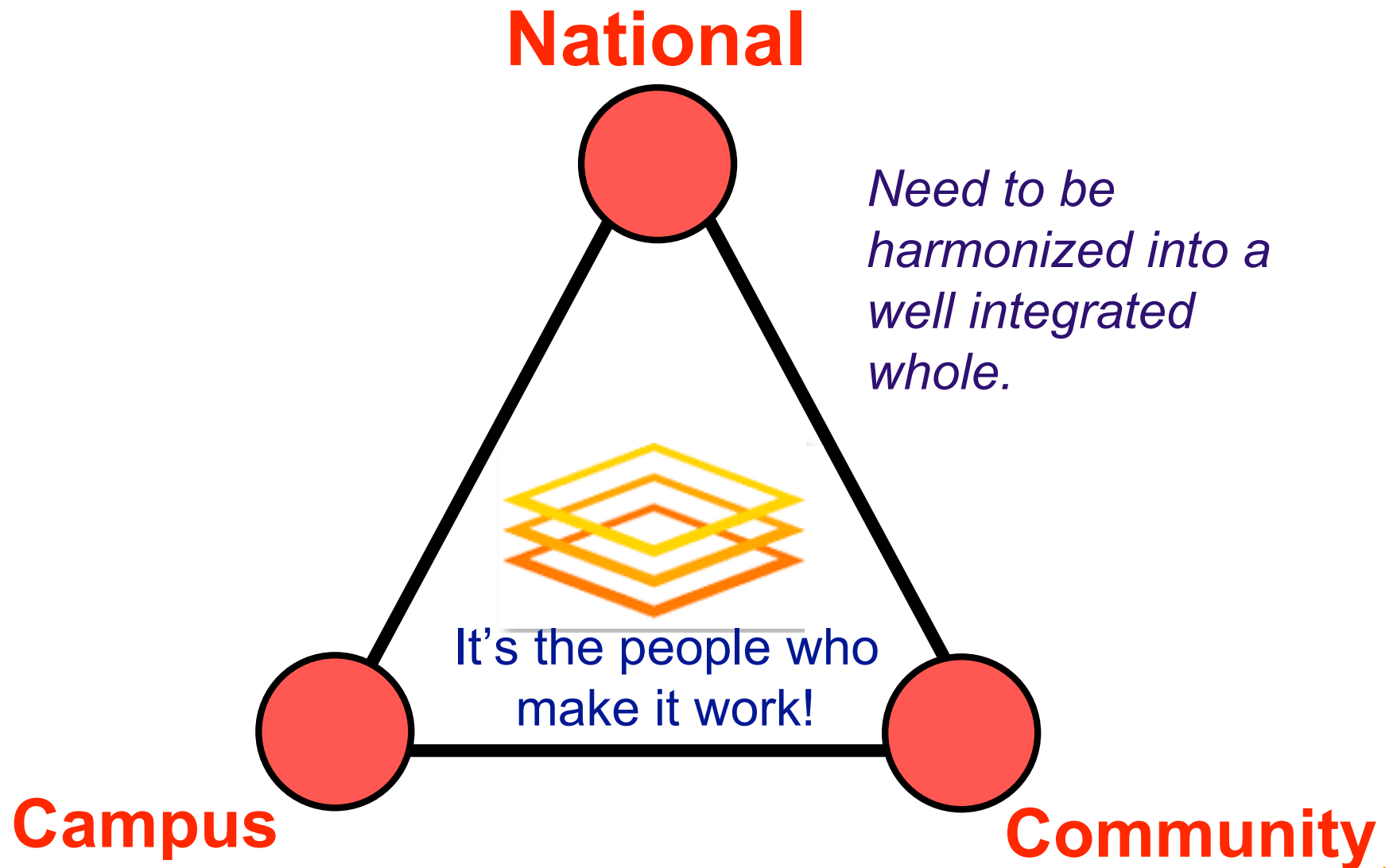


The OSG Virtual Data Toolkit

- A collection of software initially packaged by GriPhyN, iVDGL, PPDG. The VDT doesn't write software, but gets it from providers
 - Condor, Globus, EGEE Components, Community Software (e.g. Fermilab accounting), open source utilities (Apache, MySQL etc).
 - Composition driven by stakeholders and controlled by the OSG project
 - An easy installation.
 - A support infrastructure.
 - Provides a middleware foundation for the software stack of several production Grids - including the OSG.



The Three Cornerstones





Questions ...

What are we not doing that we should do?

What are we doing wrong?

Are we doing anything right?

What are the tough and/or critical problems?