Open Science Grid

Staff Retreat

May 14, 2014

LATBauerdick
OSG Executive Director

Open Science Grid

desired output of retreat

- OSG needs a detailed plan of work with milestones for coming years
  - ★ document our plans and resource needs, agency review ~ in August
- this retreat is following up on internal review and AHM
- “roadmaps” for the different areas as input for

a document outlining the year 3-5 work plan

- ★ like
  - operation center roadmap
  - software roadmap
  - security roadmap
  - network monitoring roadmap
  - technology evaluation roadmap
  - etc

Open Science Grid

starting from a strong basis

- In developing the OSG year 3-5 work plan our starting position is a well established and well working project with a number of forward looking initiatives that we already have started, like opportunistic “Open Facility” enabling OSG to provide value to campus researchers and being part of XD, the OSG Connect approach to connect researchers and campuses, the network monitoring, the work on trust relationships and replacing certificates with traceability, the software factory, and many more. Here are guidelines for developing the year 3-5 plan

★ Continue doing brilliantly where we are doing brilliantly now!
=> Don't loose track of what we do well

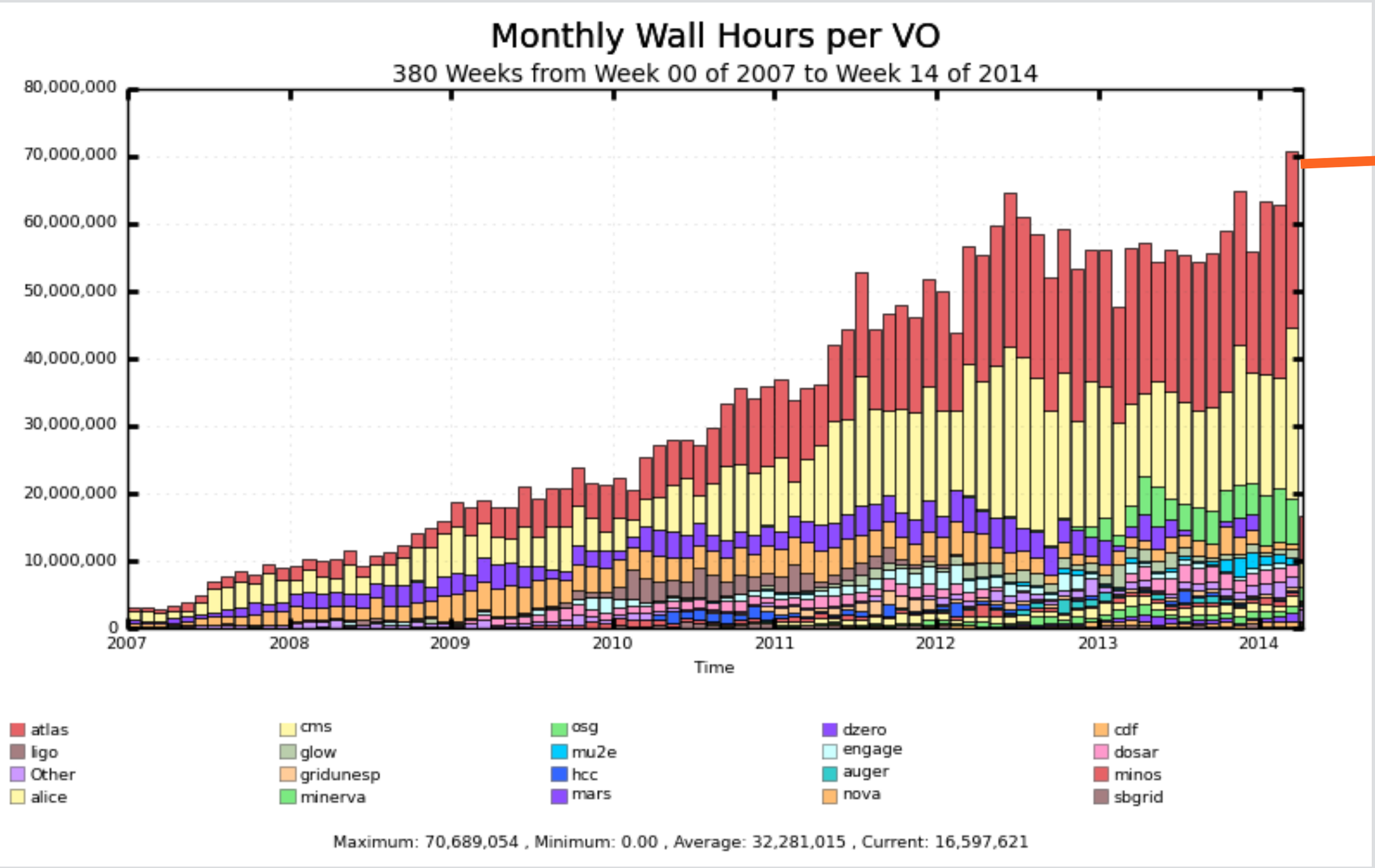
★ Expand on opportunities we see emerging from work we already have started, and solidify them towards doing brilliantly.

★ Pick a few areas that we think are completely new to us, and where we can have a large impact with modest effort

OSG "size" in terms of CPU resources

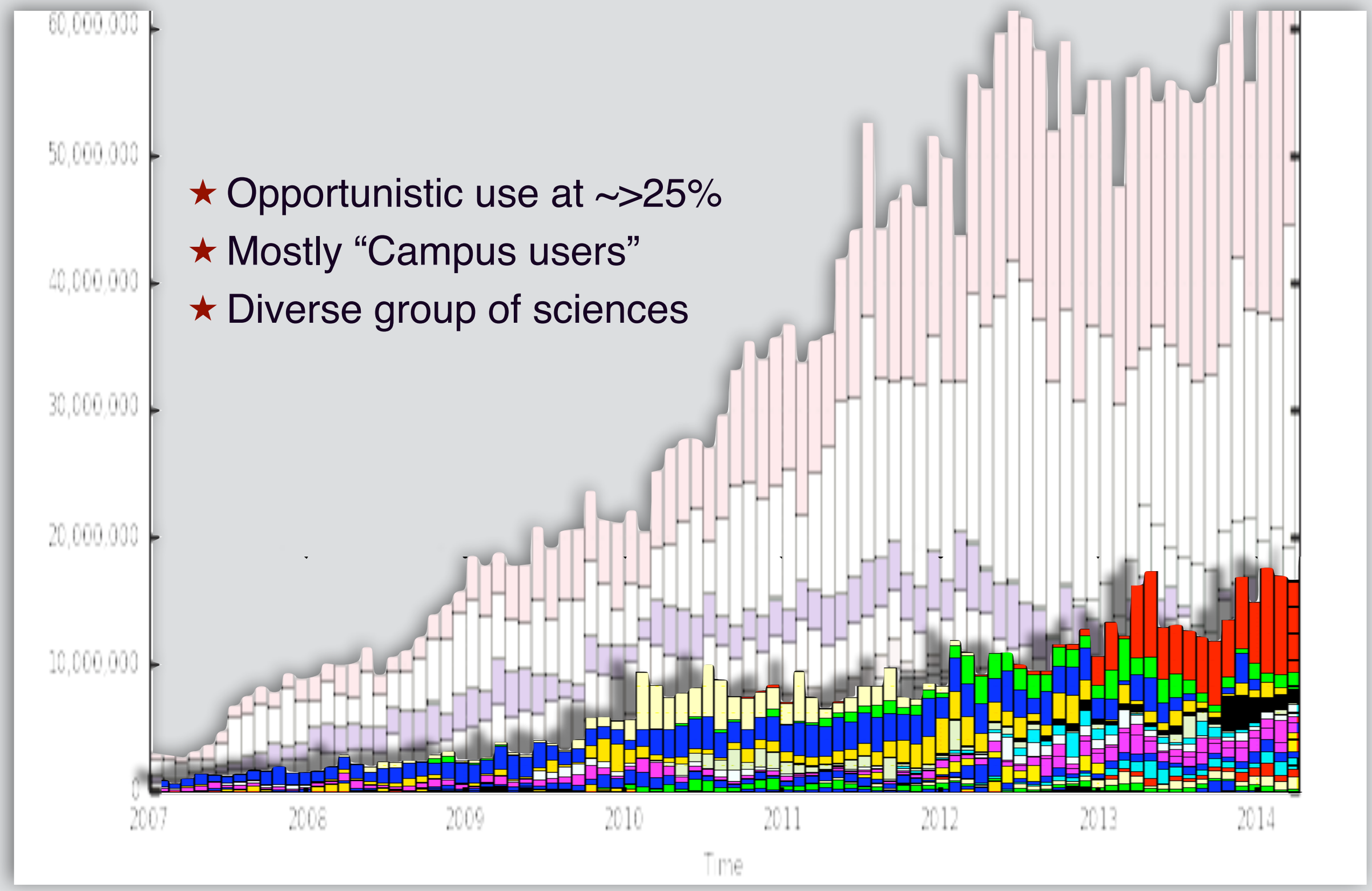
Open Science Grid

★ LHC plans to ~double CPU resources over coming 2 years



~proportional increase of opportunistic use

- ★ Opportunistic use at ~>25%
- ★ Mostly “Campus users”
- ★ Diverse group of sciences

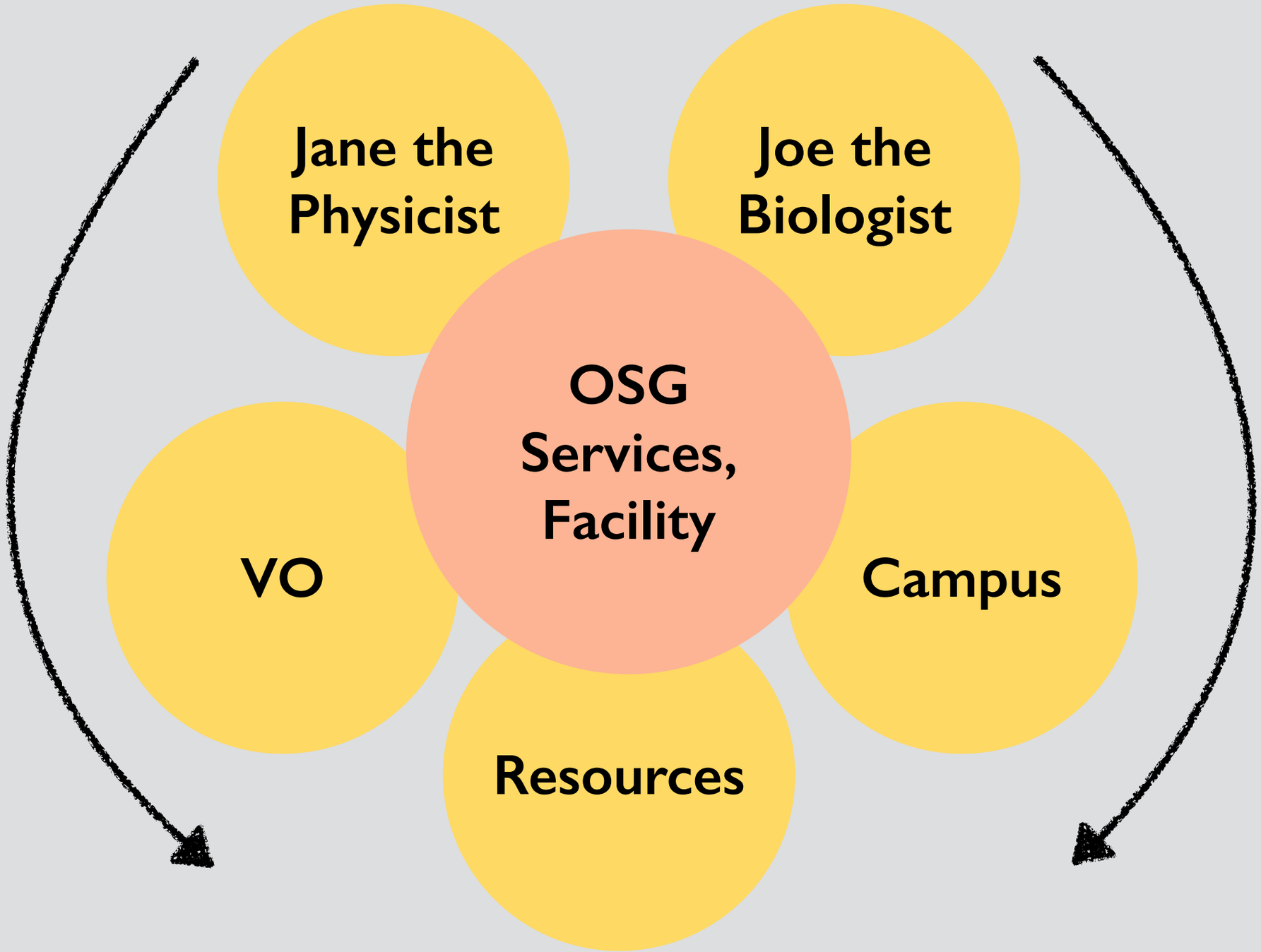


OSG services and facility ecosystem

Open Science Grid

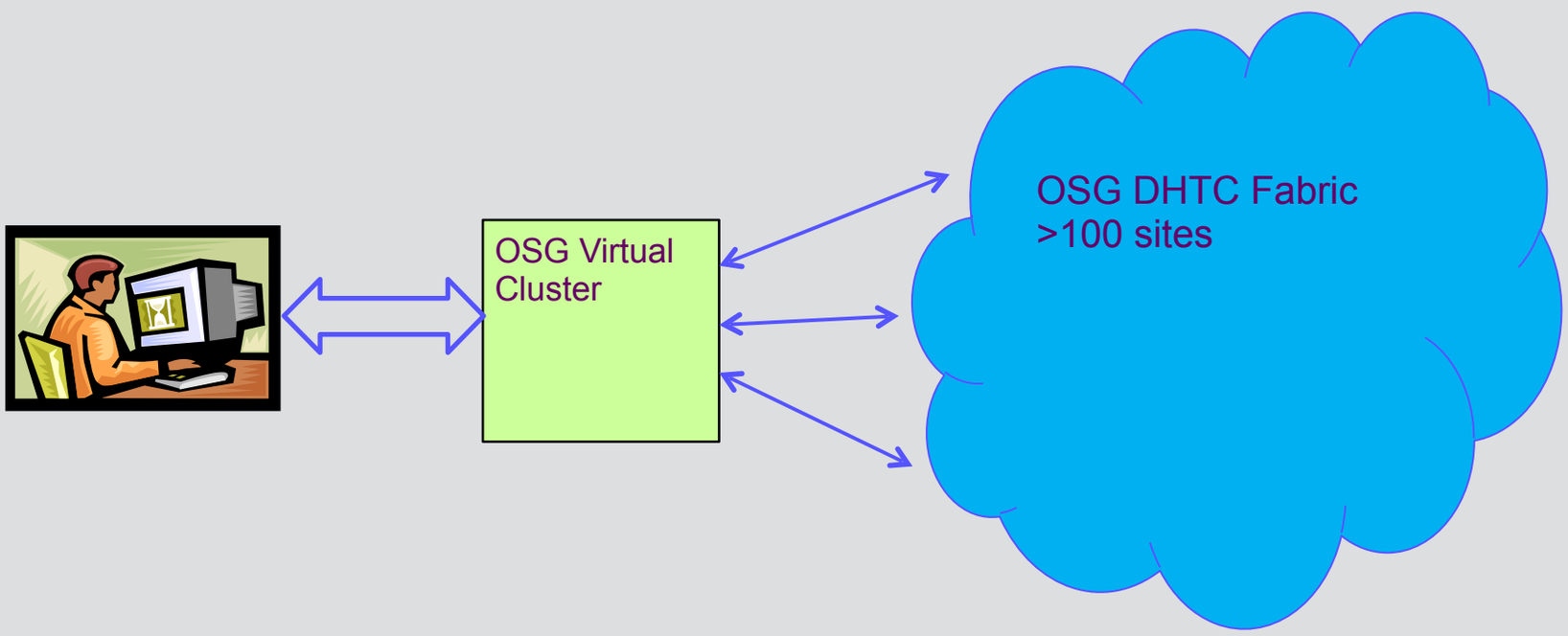
VO-centric Model

Campus User Model



OSG "Open Facility"

- A Distributed High-Throughput Computing Facility based on harvesting otherwise idle resources
  - ★ used by a quite large and diverse community of researchers and groups
  - ★ is being provided across the fabric, with resources harvested opportunistically
  - ★ the "OSG-VO" setup is now replacing engage and the focus of opportunistic cycles
  - ★ this facility is also the basis for OSG as an XD Service Provider



Bringing in New Types of Resources

Open Science Grid

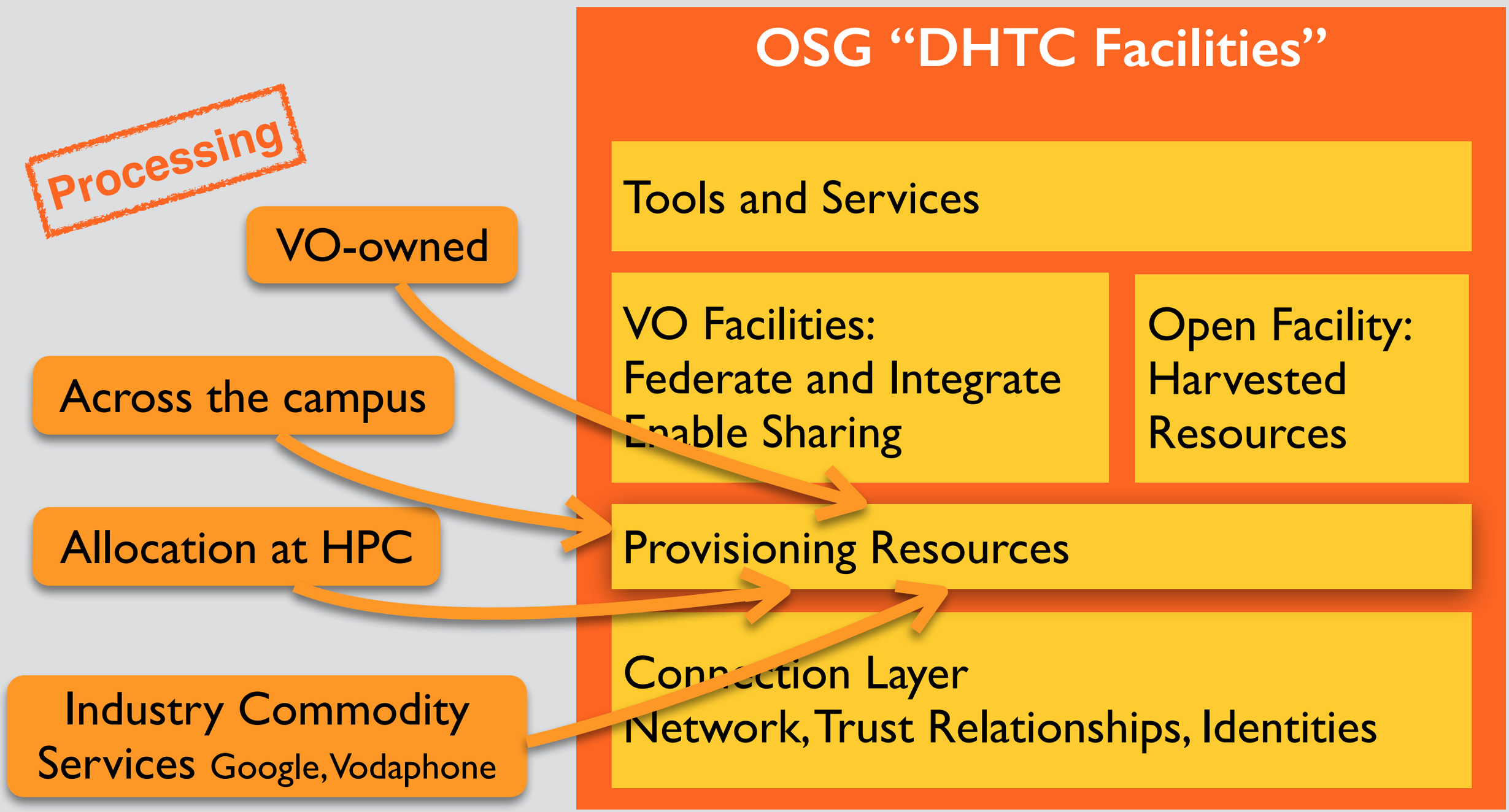
Processing

VO-owned

Across the campus

Allocation at HPC

Industry Commodity Services Google, Vodaphone



- Focus on *Resource Provisioning*
  - Statically federated resources need to be integrated with dynamically allocated resources causing new challenges for resource planning, acquisition, provisioning

200M?

~~3~~ ~~30~~M hours!

developing work plans

- In order to develop the concrete road maps/work plans I propose to break down our goals and approaches into the following three focus areas, each of which presents a clear value proposition:

  - 1. Provide value to VOs and resource owners by enabling DHTC and resource sharing for their set of applications

  - 2. Provide opportunities for the "long tail of science" (Joe the biologist)

  - 3. Promote DHTC for science

- straw-man goals/deliverables for these three focus areas, that should inform the development of roadmaps and work plans for the OSG project areas -- to be extended and prioritized today and tomorrow

Open Science Grid

1. provide value to VOs and resource owners

- ride the wave of the LHC computing upgrades
  - ★ validate operation at 2X scale
  - ★ enable sharing and opportunist use
- support access to dynamically provisioned resources
  - ★ XD, DOE HPC, commercial and community clouds
- extend to new VOs: inclusion of DOE Intensity Frontier program
- ...

Open Science Grid

2. provide opportunities for the long tail of science

- provide seamless access/usability for researchers based on OSG connect
  - mature and extend the OSG connect services
  - target particular campus opportunities: e.g. 3/6 ACI-REF campuses?
  - target partner communities: LHC, others?
- provide the Open Facility (opportunistic cycles)
  - increase throughput, manageability, accountability
- extend the supported use/services/science applications
  - data: support of workloads that require data on input, or produce data on output
    - Brian's goal of supporting large collections of jobs that require up to 10GB input data each.
    - Xrootd based caching at a subset of sites? Target the 10GB to 1TB crowd?
    - Ian's proposal of caching/dropbox-like data provisioning
  - data movement service based on Globus Online
  - portal-based workflows like Galaxy etc
  - add a Panda service for non-Atlas Panda users
  - support for special processing needs
    - support meta-scheduling across multiple clusters that each have co-processor cards?
    - large RAM support, long runtime support, etc. i.e. apps that are slightly out of our main sweet spot
- …

Open Science Grid

3. promote use of DHTC for science

- Focus Groups, provide technical leadership/guidance
  - LHC, IF, others?
- Blueprint activity
- Campus Infrastructure Communities, workshops etc
- evolve partnership XD/XSEDE
- extending the OSG eco system to the HPC community
  - technology, communication, lowering barriers?
- ...

Open Science Grid

work plans to reach these goals

- These goals and deliverables should inform us what work is needed to develop the OSG platforms. Here's a collection of talking points we should address in the different areas. Most of these will require targeted work in more than one OSG area:
  - operations/production: process for consolidation of services? operations model for OSG direct and OSG connect services?
  - addressing scale: x2 resources, extending OSG direct use, extending OSG connect
  - provisioning of dynamic resources
    - XSEDE allocations, clouds
    - support for VM-based workflows
  - security infrastructure, certificate provisioning (user, host)
    - phase out the PKI based IDM, enable use of campus identities
    - bring down the number of certs: users, VOs, site, piggy-back on CIlogon
    - develop traceability and trust relationship models

- ★ extend Open Facility based on harvesting opportunistic cycles
  - ✦ technical deliverables to support this
- ★ develop the software platform and software factory
  - ✦ improve usability
  - ✦ where do we go with the OSG client, low-impact CE and CE-less resource access
- ★ data: support for static input data, including software provisioning, database acces etc
  - ✦ OASIS, squids
  - ✦ support use case where every job in a workflow needs the same 10GB of the input.
- ★ network monitoring and performance metric
- ★ …

other points

- Organization of OSG project
  - ★ how are we doing, any changes, clarifications needed?
- outgoing activities, dissemination, lowering the barrier of entry
- partnerships:
  - ★ XD/XSEDE, ACI-REF, WLCG, EGI?
- developing community, CIC, next campus workshop
- communications, web pages, newsletter

plan for retreat

- start with a initial session going through the high-level goals and talking points, followed with discussion sessions for each of the OSG project areas, led by the relevant AC, with the goal to develop a roadmap for that area. ACs should come prepared to outline a straw man for their area and the discussion should focus on fleshing it out.