
Open Science Grid

Frank Würthwein
UCSD

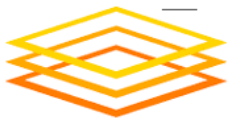


Overview

- ⇒ OSG in a nutshell
- ⇒ “Architecture”
- ⇒ “Sociology”
- ⇒ Present Utilization
- ⇒ Roadmap for new functionality

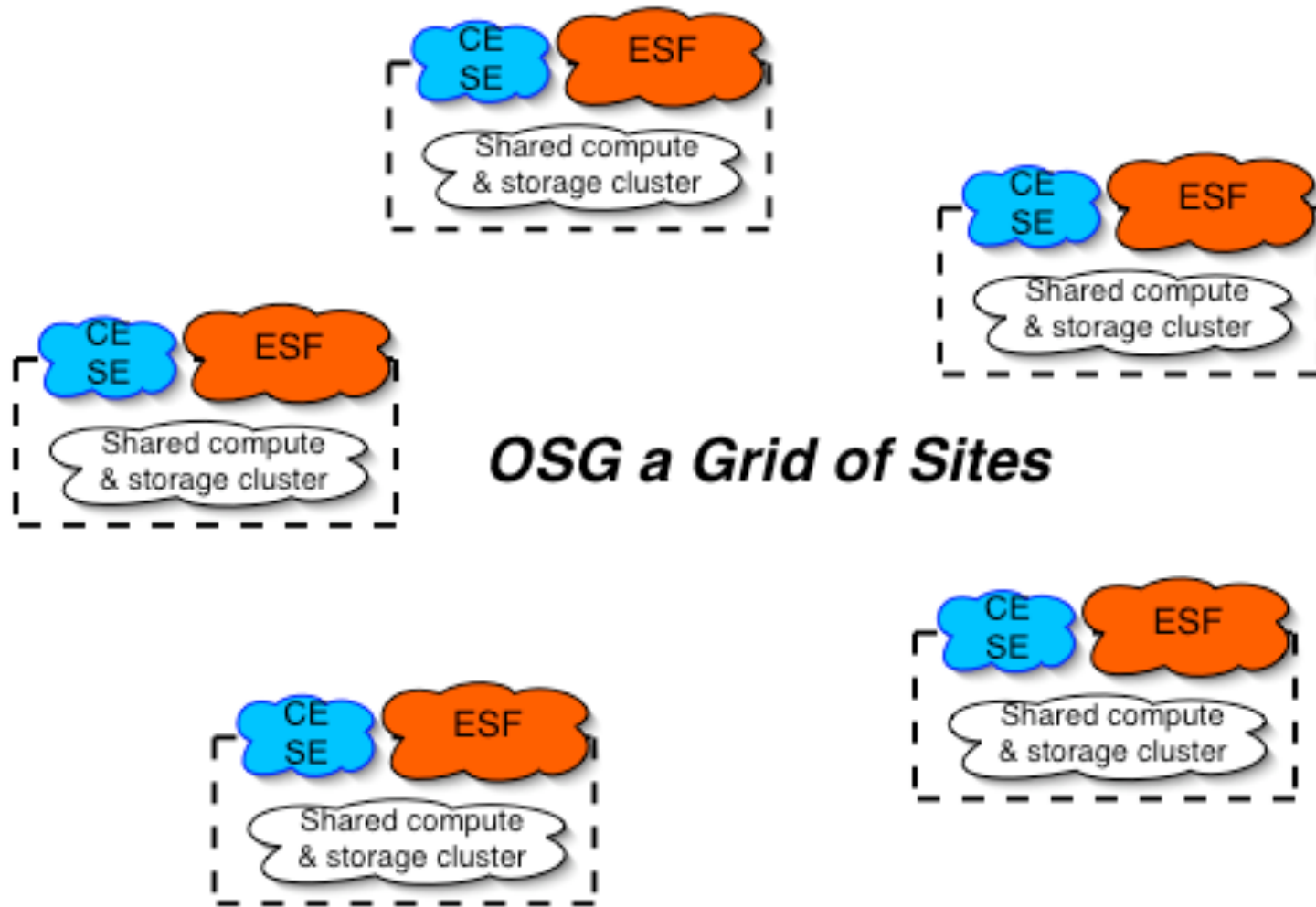
OSG in a nutshell

- ⇒ High Throughput Computing
 - Opportunistic scavenging on cheap hardware.
 - Owner controlled policies.
- ⇒ “Linux rules”: mostly RHEL3 on Intel/AMD
- ⇒ Heterogeneous Middleware stack
 - Minimal site requirements & optional services
 - Production grid allows coexistence of multiple OSG releases.
- ⇒ “open consortium”
 - Stakeholder projects & OSG project to provide cohesion and sustainability.
- ⇒ Grid of sites
 - Compute & storage (mostly) on private Gb/s LANs.
 - Some sites with (multiple) 10Gb/s WAN uplink.



“Architecture”

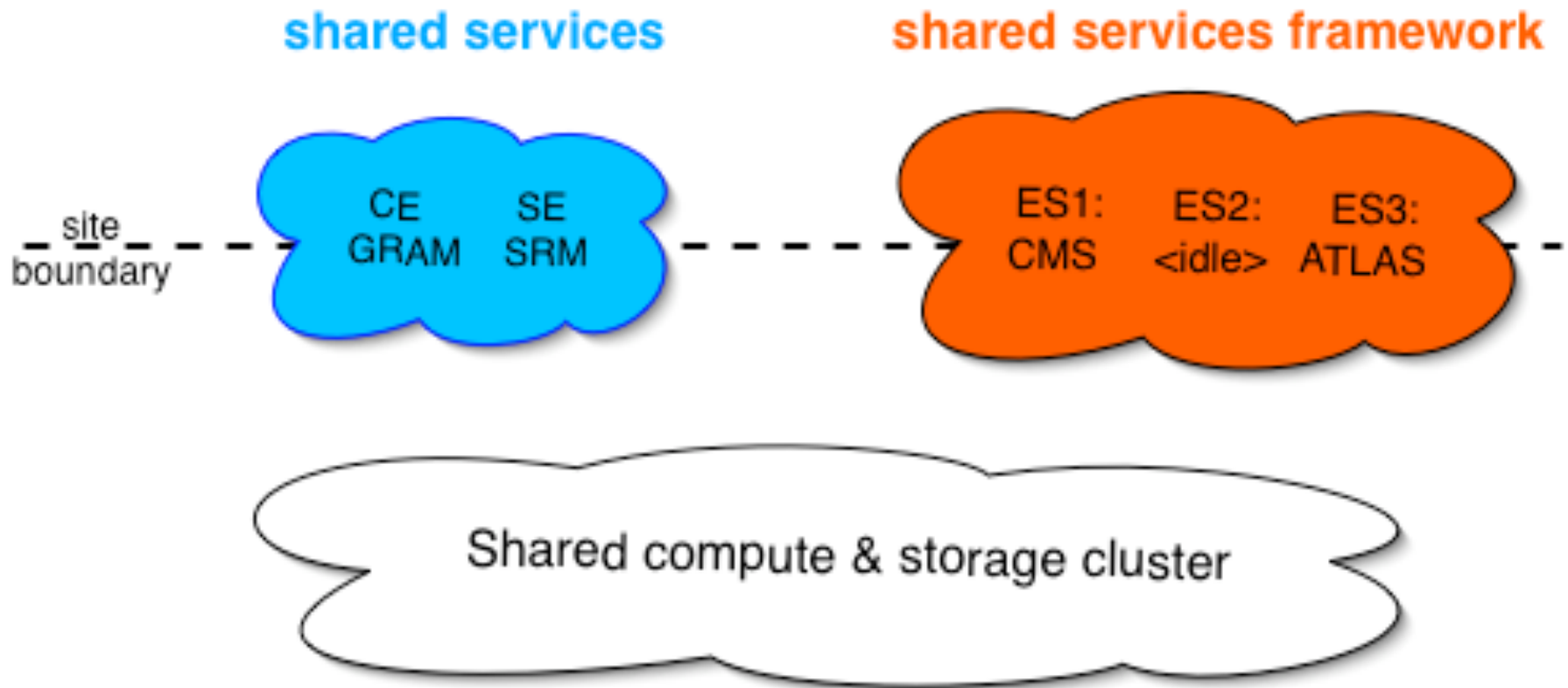




Today: ~50 sites, ~18,000 batch slots, ~500TB, up to 10Gb/s
 Vision: O(1e5) CPUs, O(1e5)TB, O(1e1-2)Gb/s in 5 years

OSG Site

(simplified snapshot of a typical OSG site in 2008)



Shared Services

⇒ CE:

- ⇒ Now: (modified) pre-WS GRAM
- ⇒ End of 2006: GT4 GRAM

⇒ SE:

- ⇒ Now: SRM ...
- ⇒ ... but with legacy support for GT4 gridftp & “Classic SE”

⇒ Authz:

- ⇒ VOMS & PRIMA & GUMS et al.

⇒ Monitoring

- ⇒ Now: one big mess
(GLUE schema 1.2 & ML & MIS & gridCat & ...)
- ⇒ End of 2006: well, one hopes for the best ...



Hardware Infrastructure

⇒ In principal:

⇒ Anything goes as long as there's truth in advertising.

⇒ In practice:

⇒ Intel/AMD.

⇒ RHEL 3 and its variants.

⇒ Gb/s LANs, up to multiple 10Gb/s WAN

⇒ Many (but not all) private/public network arrangements.

⇒ Lot's of cheap IDE disks



Two Infrastructure Details

Authz Model

Storage



- ⇒ Grid3, the pre-cursor to OSG, used group accounts, where entire VOs were mapped.
 - ⇒ Did not meet the security requirements of many sites, because it did not allow sites to easily distinguish the activities of users.
- ⇒ Goal was to enable finer grained authorization.
 - ⇒ Create multi-user environment in which traditional UID based security audits are possible if desired by site.
 - ⇒ “dynamic”, static, or group accounts according to site security policy.
 - ⇒ Move from host based to site based authz
 - ⇒ Authz = VO-allowed & !site-vetoed
 - ⇒ Distinguish user activities based on proxy cert with attributes attached.
 - ⇒ Utilize the capabilities of EDG developed Virtual Organization Management System (VOMS) to ...
 - ⇒ ... make authz decisions based on attribute information.
 - ⇒ One human can have different roles across multiple VOs, or within one VO.



Envisioned Use Cases

- ⇒ Enable support for priority in batch systems based on VO activities.
 - ⇒ One person may submit as either themselves, or as cms mc production, and receive different priority in batch system accordingly.
 - ⇒ One user who maintains a service (e.g. cms soft install) may get redirected to special batch slots for service maintenance.
- ⇒ Support write-authorization for sub-groups or individuals of VOs in storage systems, or application areas.
 - ⇒ One person installs cms application software on all OSG sites that all others have only read but not write access to.
- ⇒ Enable quotas (disk and/or CPU) for individuals or sub-groups based on published VO policy.
- ⇒ Allow data transfer requests from all users, and prioritize them based on role of the user.

OSG AuthZ Approach

- ⇒ VO defines Roles and associated privileges by specifying expected functionality.
 - ⇒ E.g. *cmssoft* may install software in area that is read-only by all *cmsuser* jobs running on site/campus.
 - ⇒ E.g. *cmsphedex* may have special access to SRM/dCache system.
- ⇒ Site maps VO scope identities to local scope identities.
 - ⇒ Site wide management of mapping.
 - ⇒ Service level granularity of mapping.
- ⇒ Site enforces VO privilege policies within local scope identities.
- ⇒ Authorization = (VO-allowed) && !(Site-vetoed)

Example:

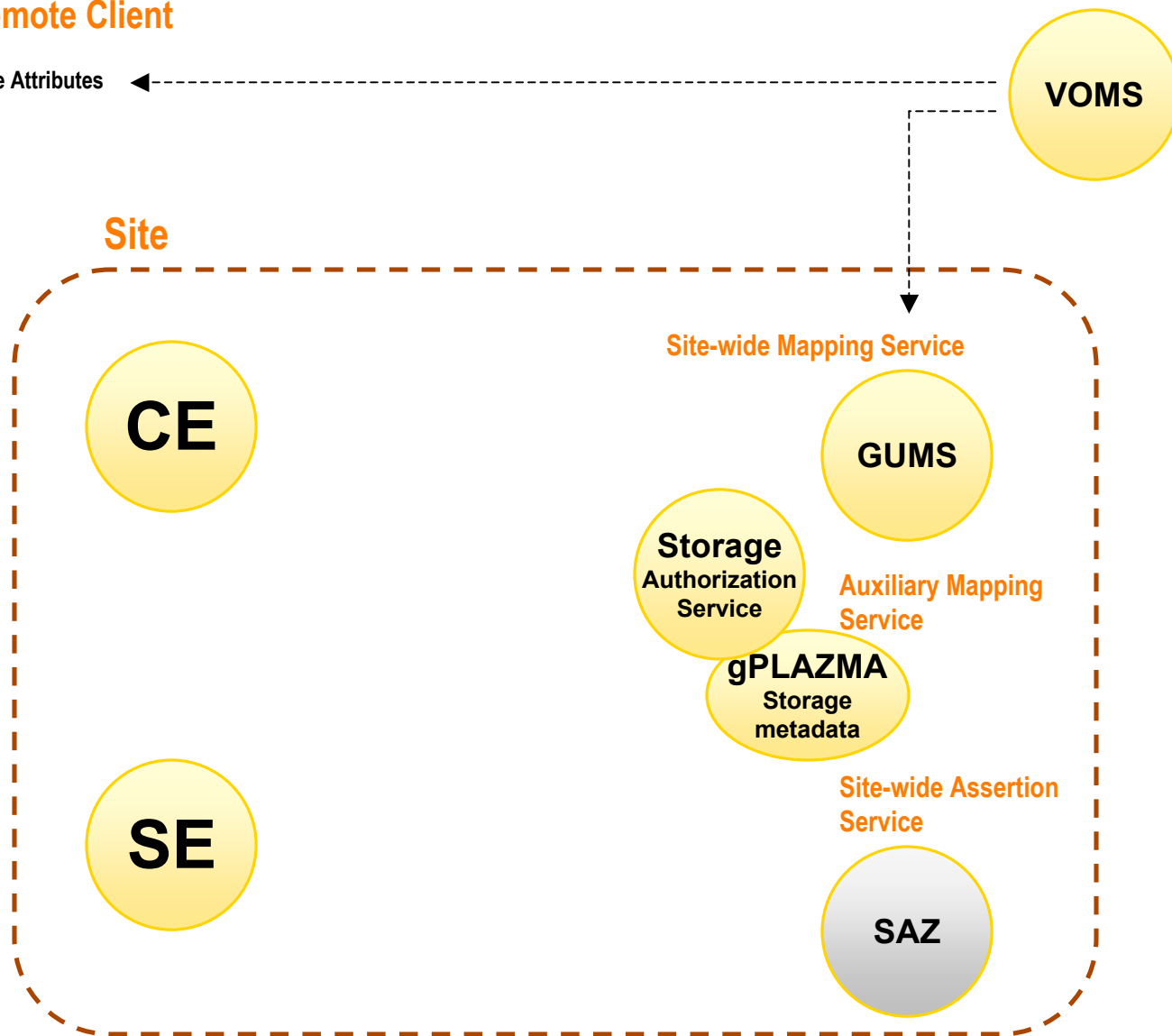
End-to-end Authz
for CE & SE





Local or Remote Client

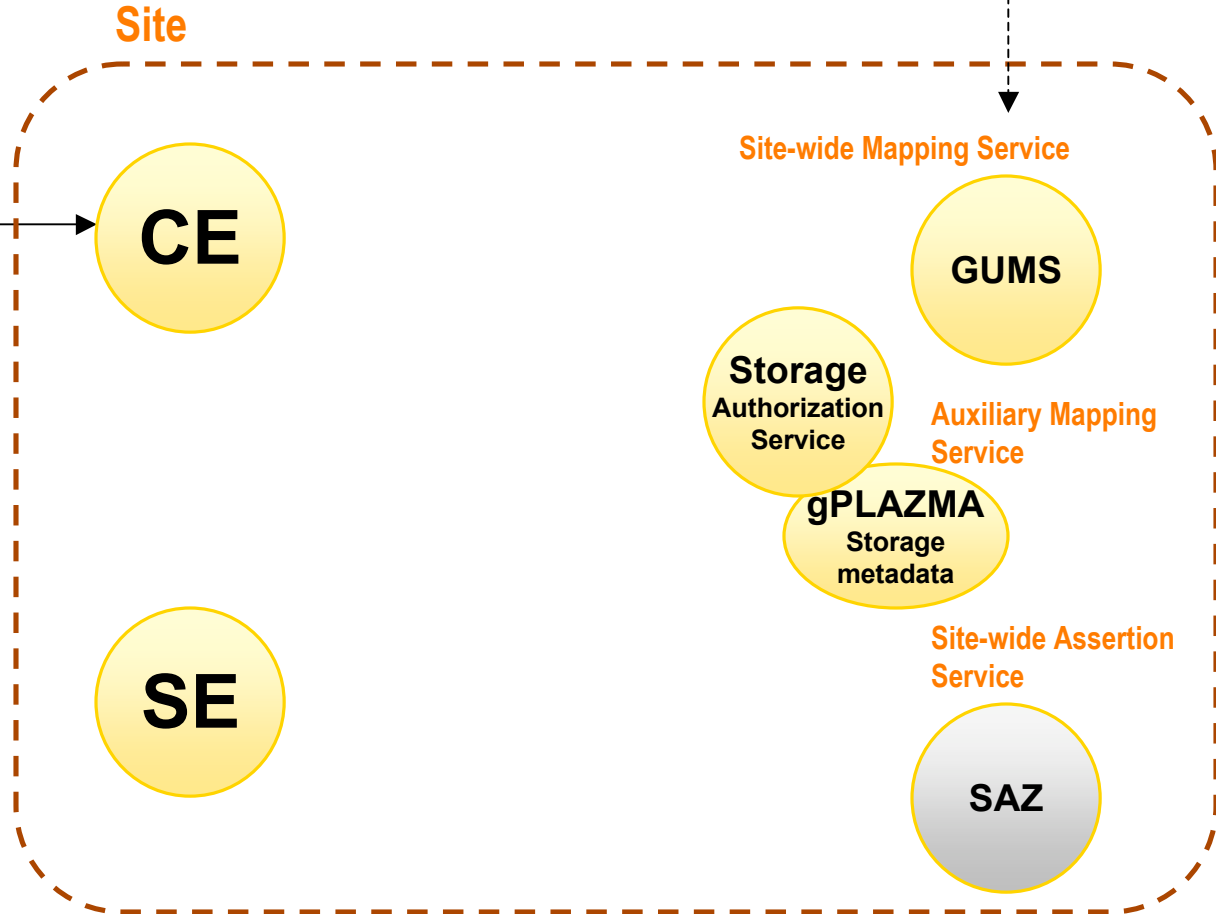
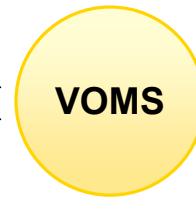
Proxy with VO
Membership | Role Attributes





Local or Remote Client

Proxy with VO
Membership | Role Attributes



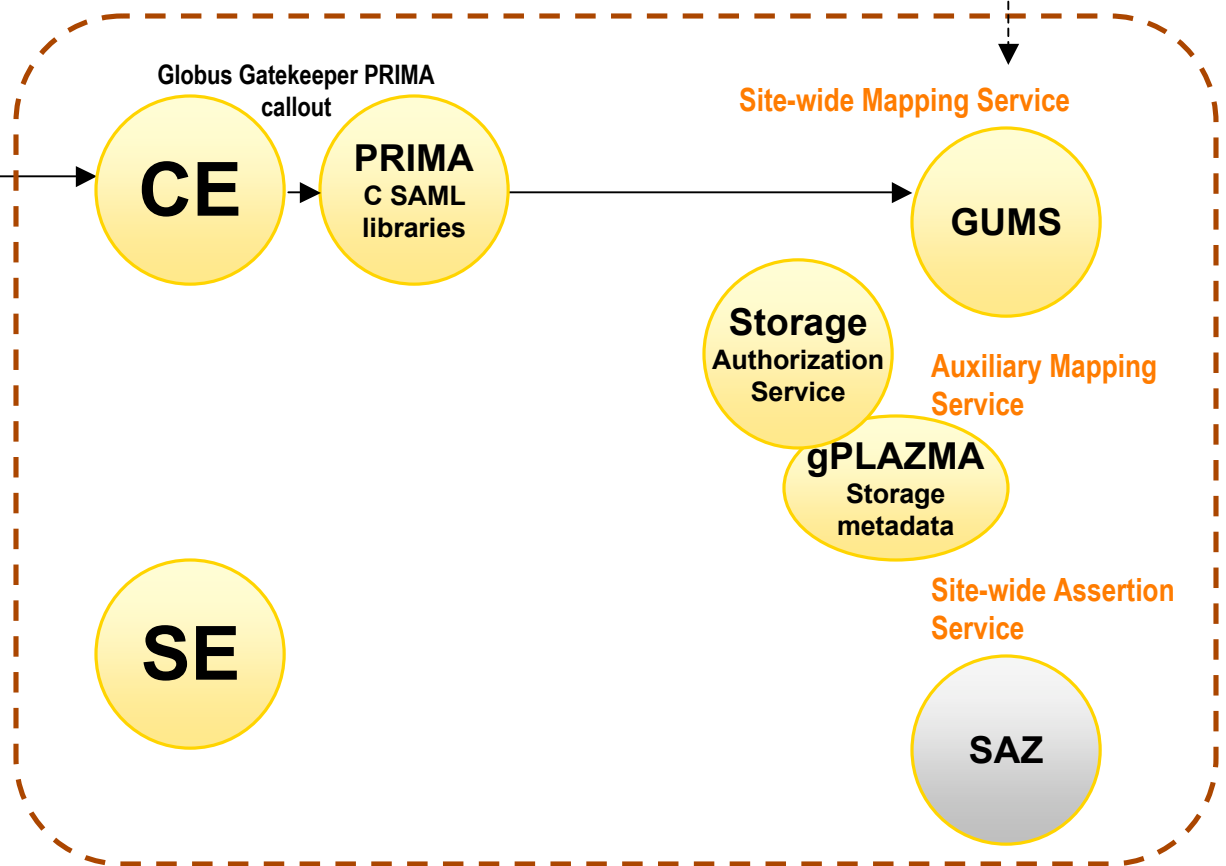


Local or Remote Client

Proxy with VO
Membership | Role Attributes



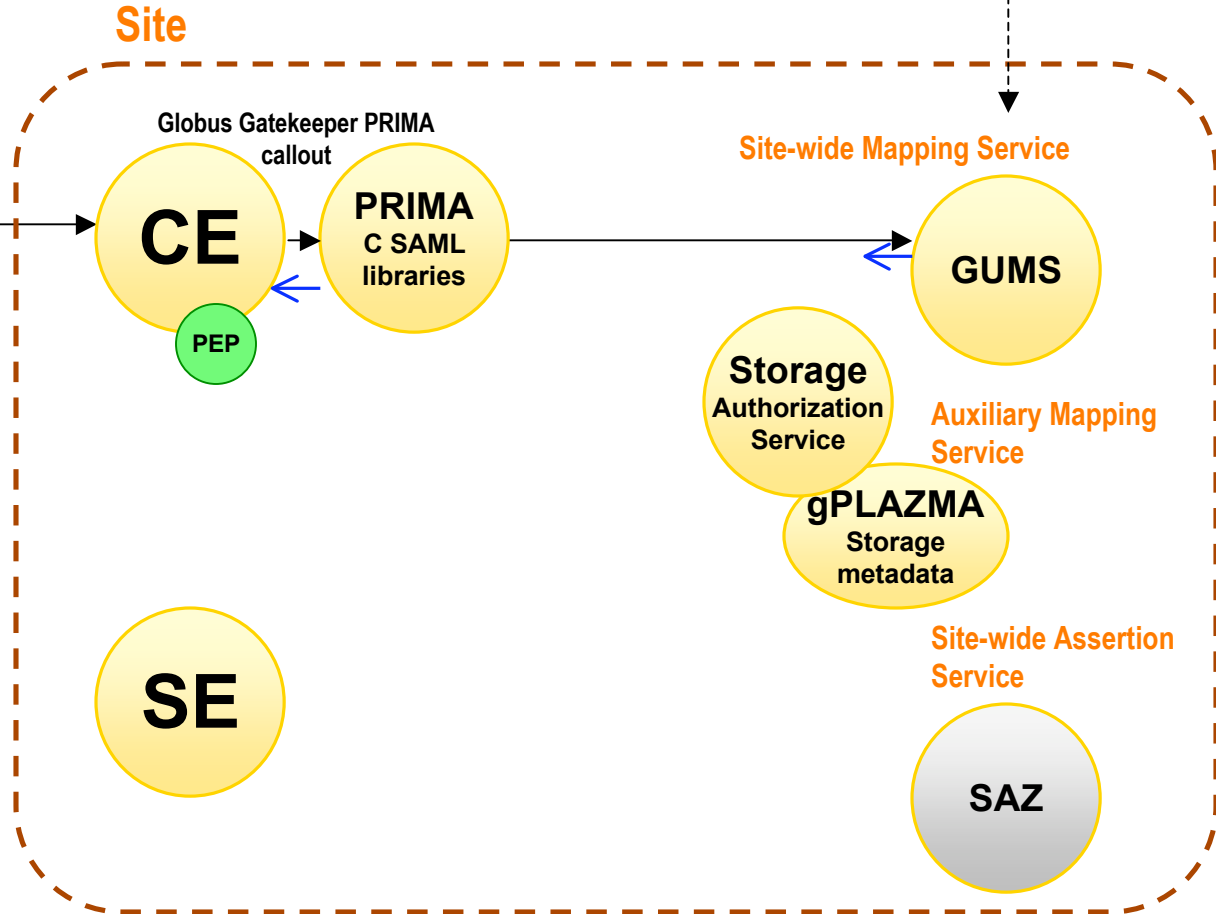
Site





Local or Remote Client

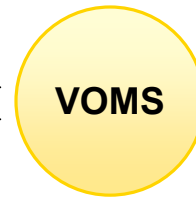
Proxy with VO
Membership | Role Attributes



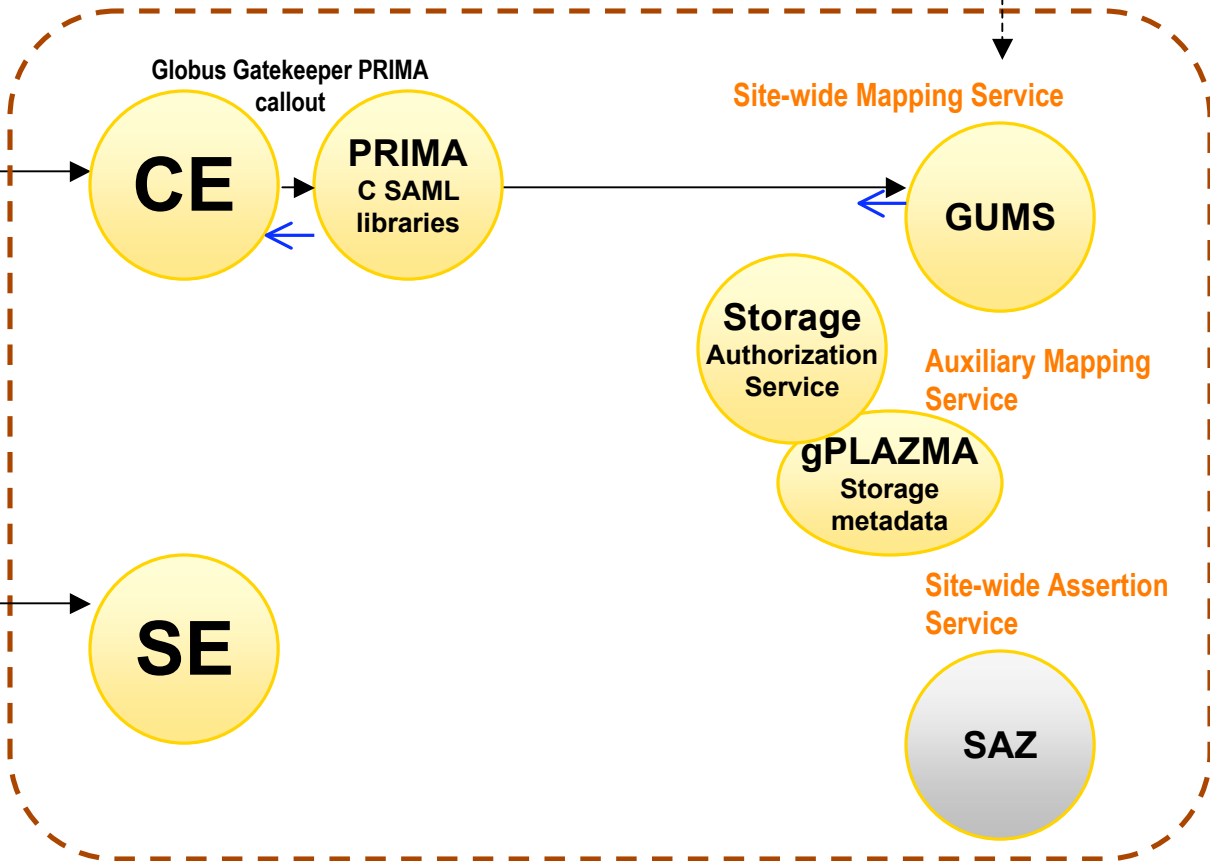


Local or Remote Client

Proxy with VO
Membership | Role Attributes



Site



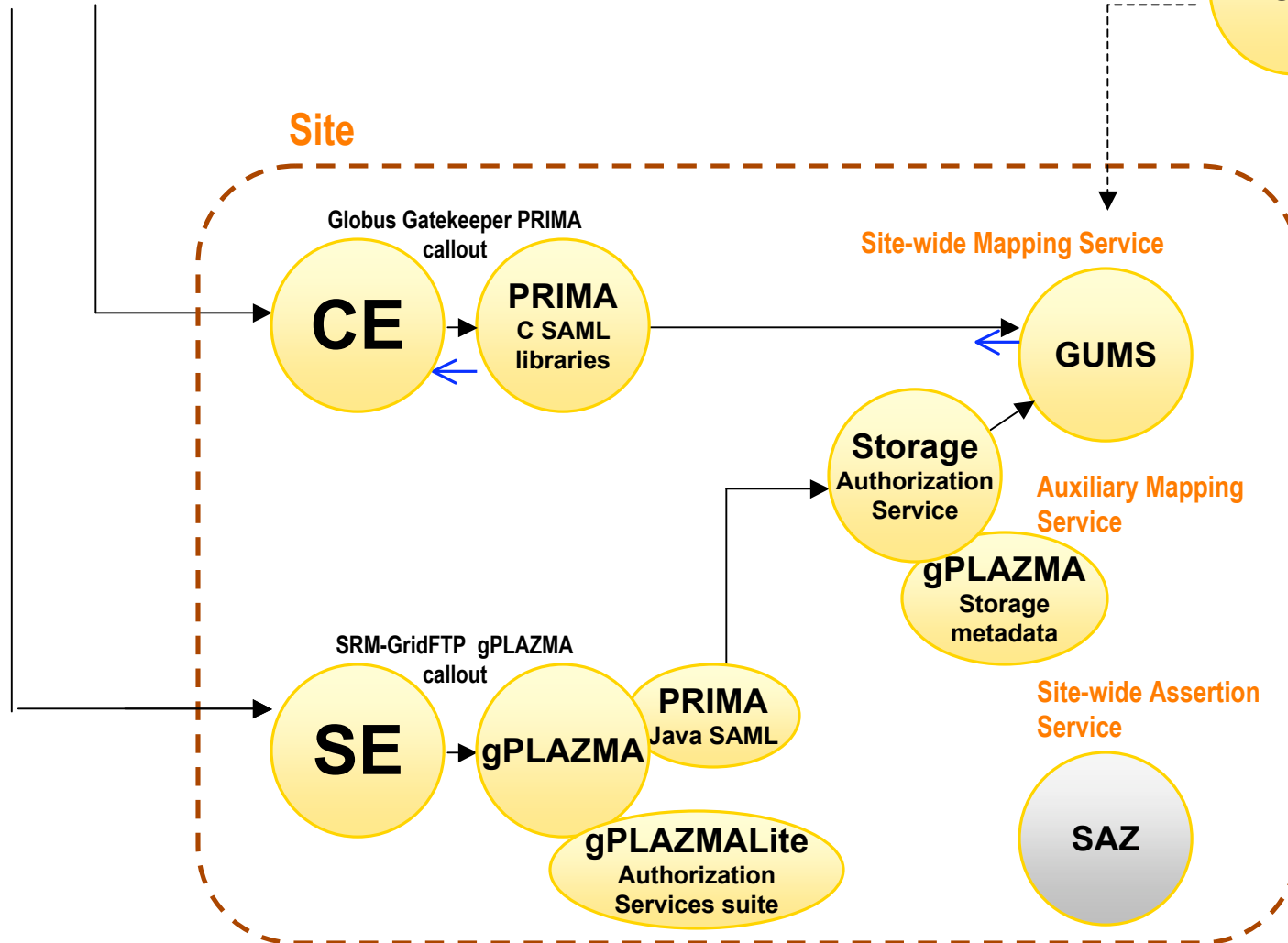


Local or Remote Client

Proxy with VO
Membership | Role Attributes



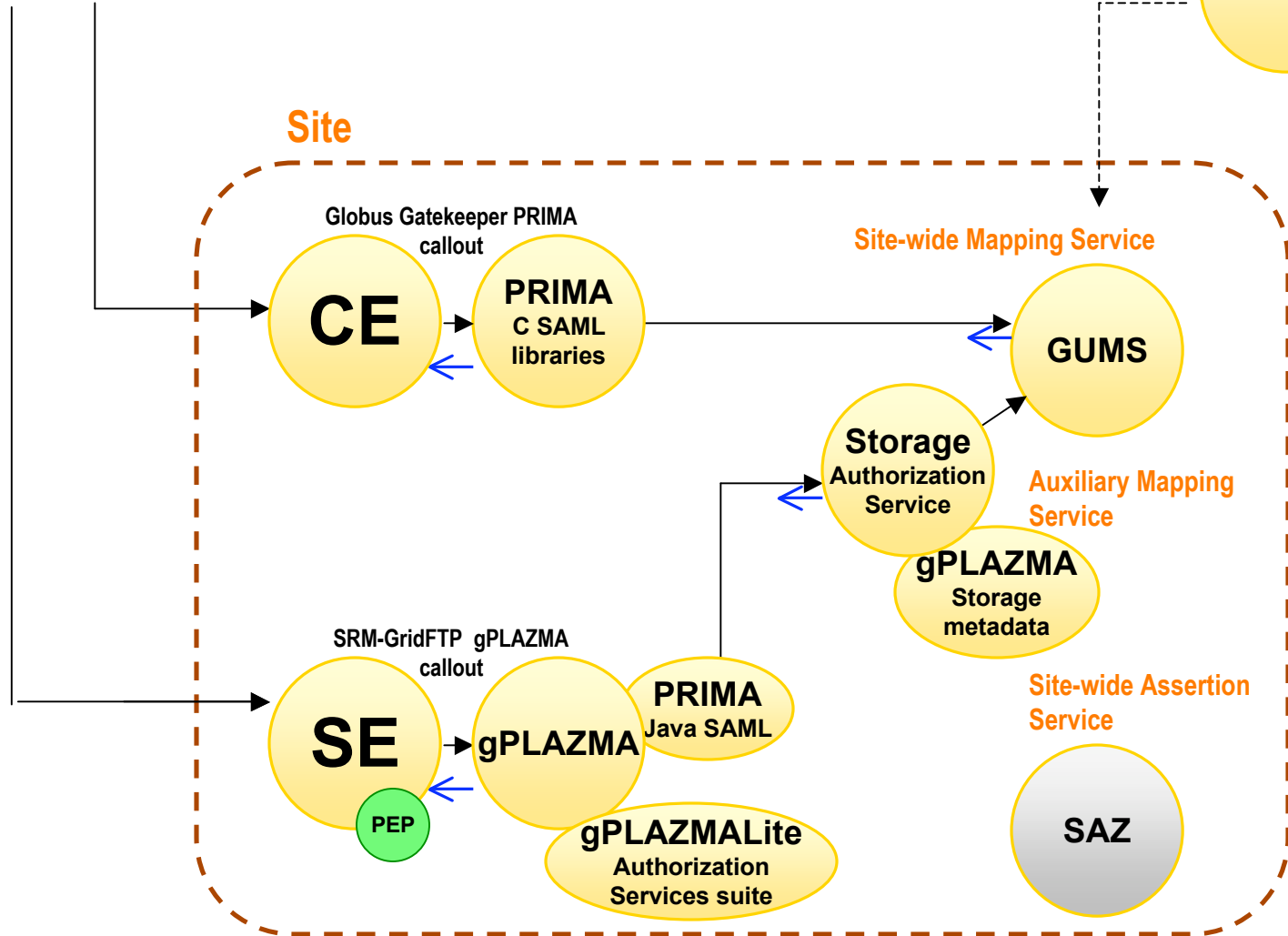
Site





Local or Remote Client

Proxy with VO
Membership | Role Attributes



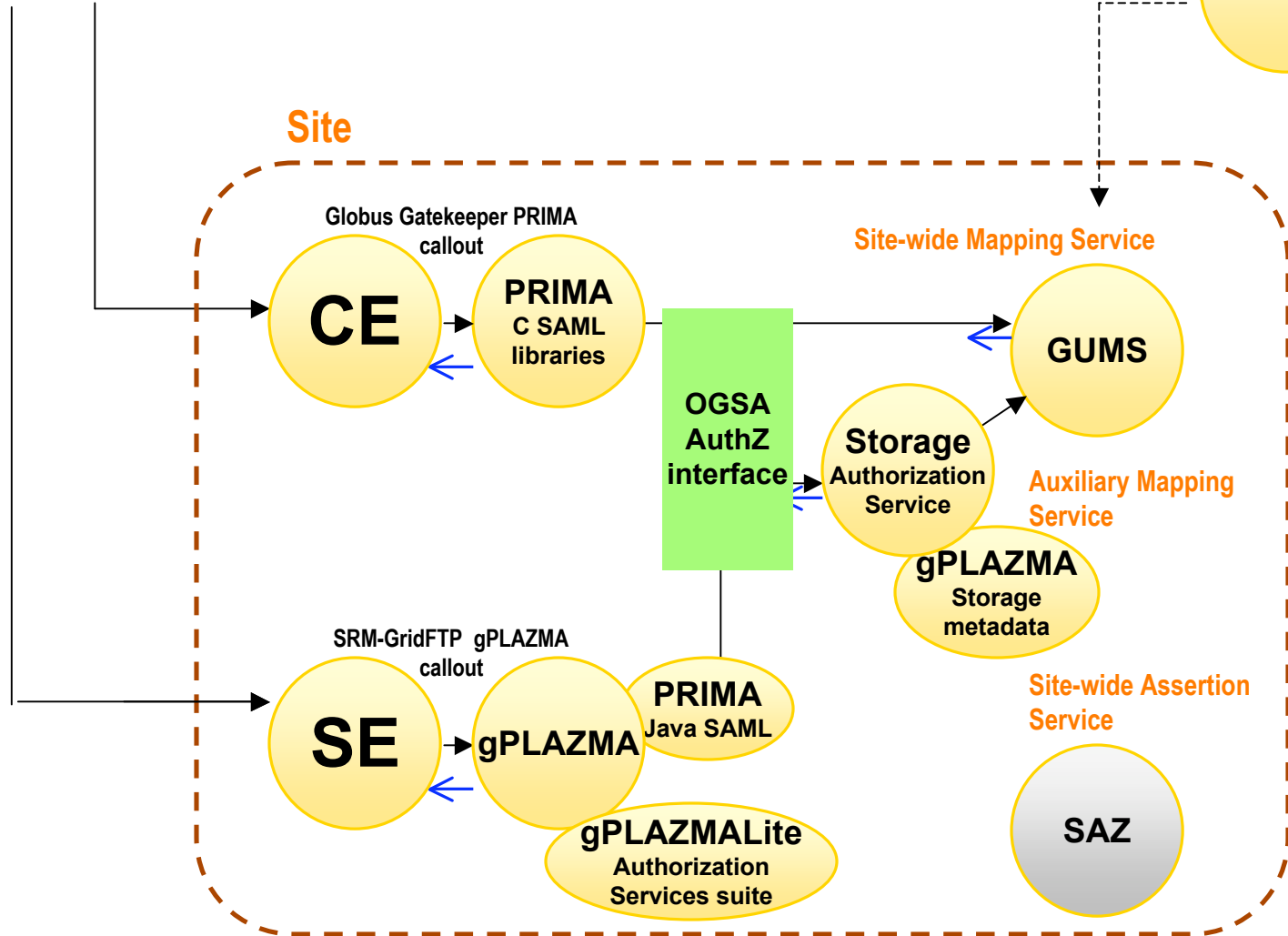


Local or Remote Client

Proxy with VO
Membership | Role Attributes



Site



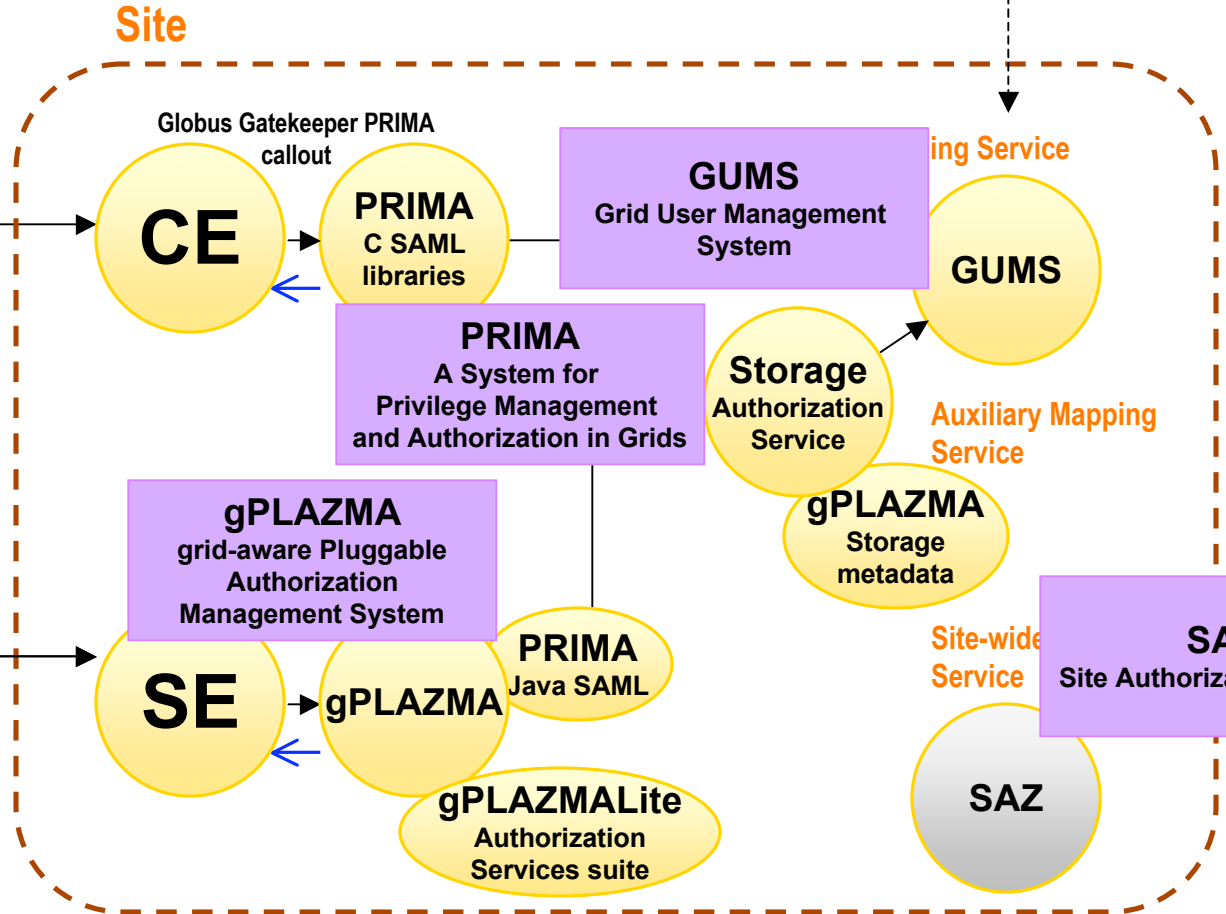


Local or Remote Client

Proxy with VO
Membership | Role Attributes

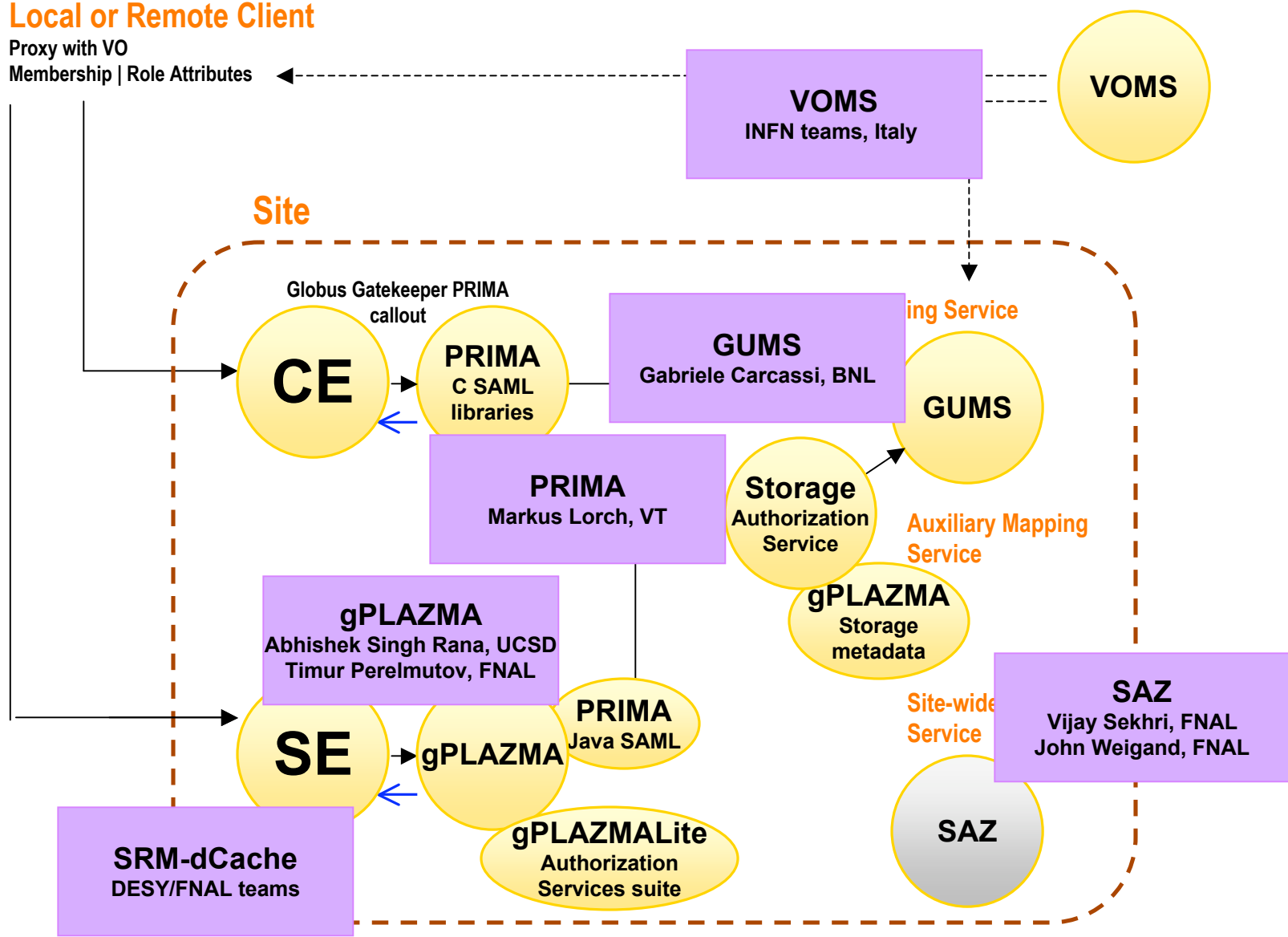
VOMS
Virtual Organization
Membership Service

VOMS



Local or Remote Client

Proxy with VO
Membership | Role Attributes



Note:

OSG Authz approach extends beyond traditional Authz.

Generic Attribute Authorization Framework!

Different Services may use different extended attributes!



Storage

No global file system.

All storage is local to site.

Managed WAN data movement.



Open Science Grid



Disk areas in some detail:

- ⇒ Shared filesystem as applications area at site.
 - Read only from compute cluster.
 - Role based installation via GRAM.
- ⇒ Batch slot specific local work space.
 - No persistency beyond batch slot lease.
 - Not shared across batch slots.
 - Read & write access (of course).
- ⇒ SRM controlled data area.
 - Job related stage in/out.
 - “persistent” data store beyond job boundaries.
 - SRM v1.1 today.
 - SRM v2 expected in next major release (summer 2006).

SRM/dCache in a nutshell

⇒ Goals:

- Virtualize large amounts of commodity disk.
- Provide fail-over & load balancing.

⇒ Strategy:

- Separate physical & logical namespace.
- Separate file request from file open.
 - One SRM manages many data servers for various protocols.

⇒ WAN “upload”:

- One SRM interface manages many gftp servers.
- “Lambda station” to schedule λ 's.



“Sociology”



Driven by LHC Physics

- ⇒ Computing Challenge:
 - 20PB of data in 2008 served across 30PB disk distributed across 100 sites worldwide to be analyzed by 100MSpecInt2000 of CPU.
 - ⇒ Many orders of magnitude increased physics reach.
 - x7 increase in beam energy \Rightarrow x150 increase in top Xsection.
 - x10 increase in instantaneous luminosity.
 - Read & write access (of course).
 - ⇒ At least three orders of magnitude increase in reach for new physics.
 - ⇒ Not just any 3 orders of magnitude, but expect threshold effect.
 - ⇒ Many people expect revolutionary discoveries in year 1 of data taking.
- ⇒ *The stakes for computing have never been this high in HEP!*

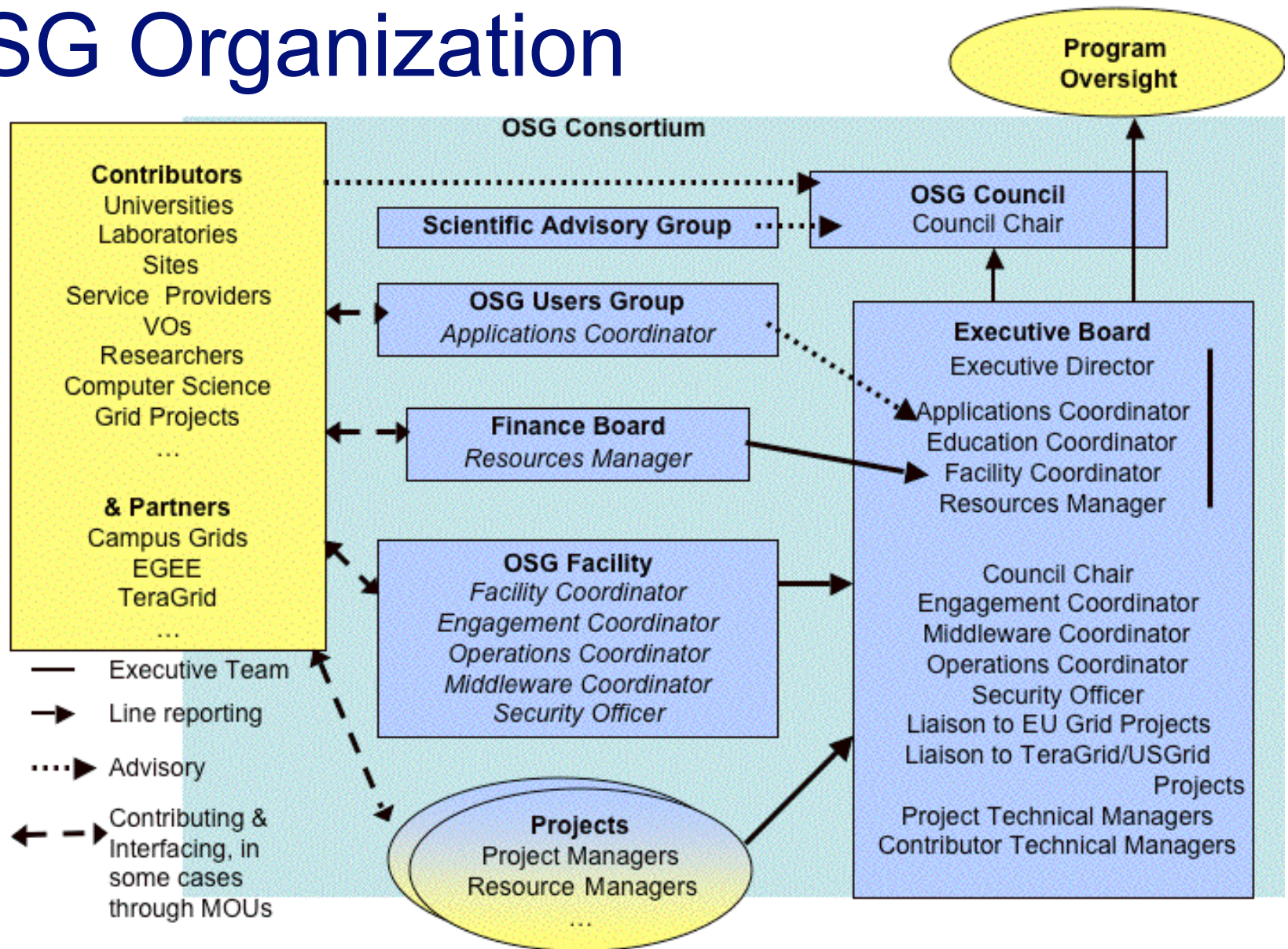


OSG Organization

Mix of Consortium & Project



OSG Organization



OSG organization (explained)

- ⇒ OSG Consortium
 - Stakeholder organization with representative governance by OSG council.
- ⇒ OSG project
 - (To be) funded project to provide cohesion & sustainability
 - OSG Facility
 - “Keep the OSG running”
 - “Engagement of new communities”
 - OSG Applications Group
 - “keep existing user communities happy”
 - Work with middleware groups on extensions of software stack
 - Education & Outreach

OSG Management

Executive Director:	Ruth Pordes
Facility Coordinator:	Miron Livny
Application Coordinators:	Torre Wenaus & fkw
Resource Managers:	P. Avery & A. Lazzarini
Education Coordinator:	Mike Wilde
Council Chair:	Bill Kramer

OSG Management (continued)

Engagement Coord.:	Alan Blatecky
Middleware Coord.:	Alain Roy
Ops Coordinator:	Leigh Grundhoefer
Security Officer:	Don Petravick
Liaison to EGEE:	John Huth
Liaison to Teragrid:	Mark Green

The Grid “Scalability Challenge”

- ⇒ Minimize entry threshold for resource owners
 - Minimize software stack.
 - Minimize support load.
- ⇒ Minimize entry threshold for users
 - Feature rich software stack.
 - Excellent user support.

Resolve contradiction via “thick” Virtual Organization layer of services between users and the grid.

Me -- My friends -- The grid

Me: thin user layer

My friends:
VO services
VO infrastructure
VO admins

The Grid: anonymous sites & admins

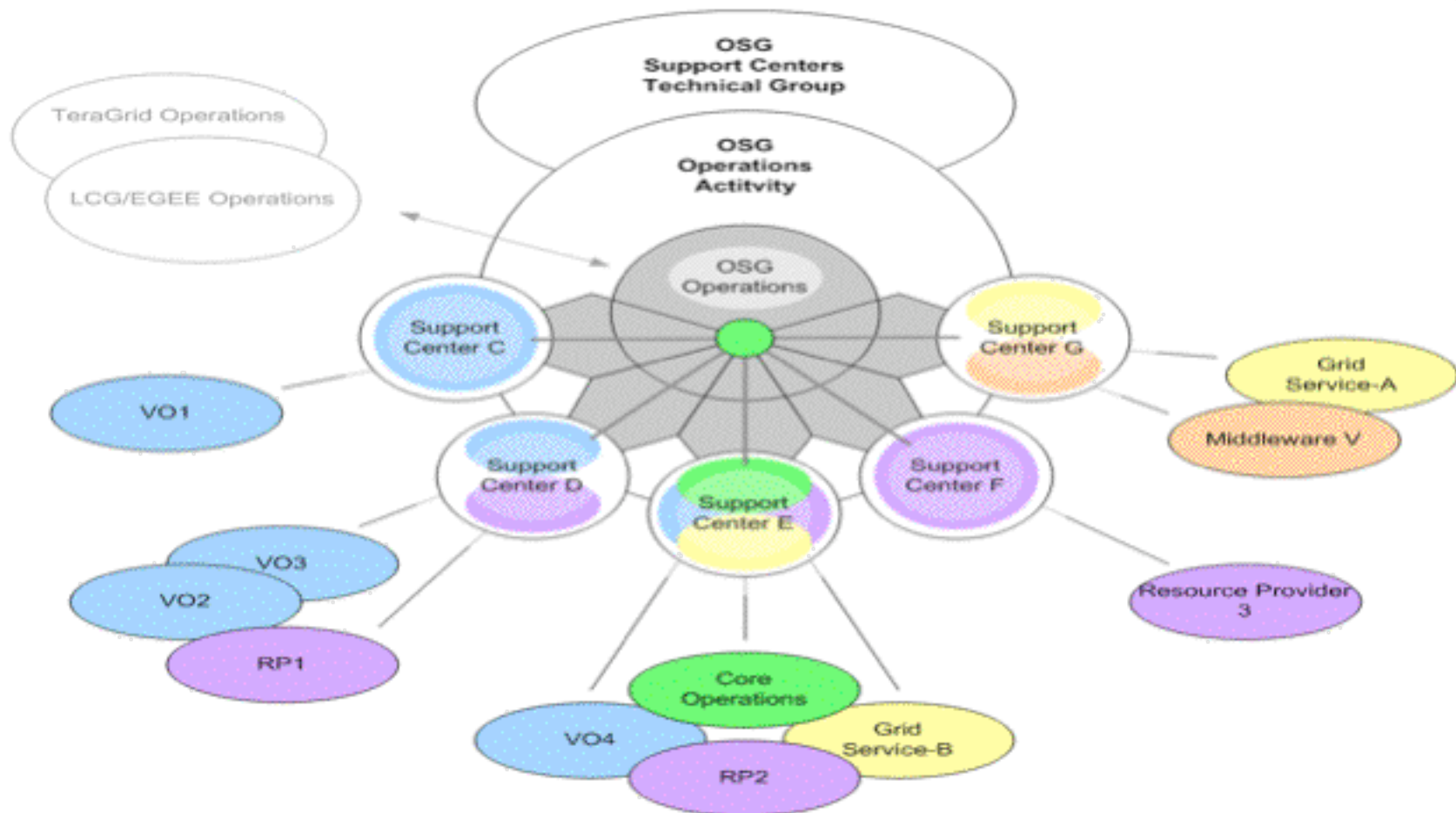
Me & My friends
are domain science
specific.

Common to all.



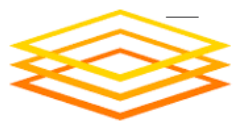
OSG Operations Model

Physical View



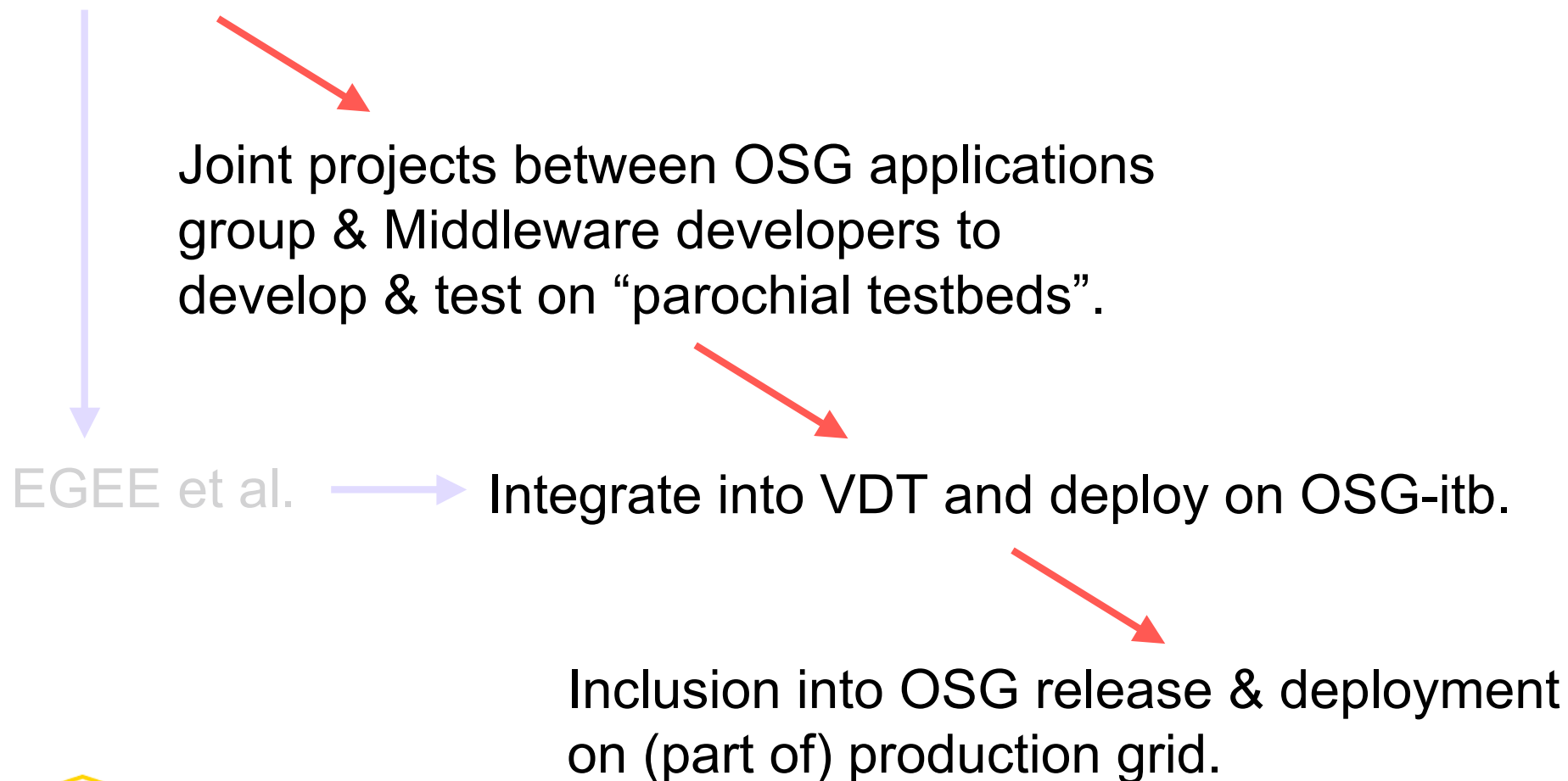
User Management

- ⇒ User registers with VO and is added to VOMS of VO.
 - VO responsible for registration of VO with OSG GOC.
 - VO responsible for users to sign AUP.
 - VO responsible for VOMS operations.
 - VOMS shared for ops on both EGEE & OSG by some VOs.
 - Default OSG VO exists for new communities.
- ⇒ Sites decide which VOs to support (striving for default admit)
 - Site populates GUMS from VOMSeS of all VOs
 - Site chooses uid policy for each VO & role
 - Dynamic vs static vs group accounts
- ⇒ User uses whatever services the VO provides in support of users
 - VO may hide grid behind portal
- ⇒ Any and all support is responsibility of VO
 - Helping its users
 - Responding to complains from grid sites about its users.



Middleware lifecycle

Domain science requirements.



Status of Utilization



Principle versus Practice

- ⇒ 53 Compute Elements registered.
- ⇒ More than 18,000 batch slots registered.
 - ... but only 10% of it used via grid interfaces that are monitored.
 - ⇒ Large fraction of local use rather than grid use.
 - ⇒ **Policy & Metrics challenged.**
 - ⇒ *Not all registered slots are available to grid users.*
 - ⇒ *Not all available slots are available to every grid user.*
 - ⇒ *Not all slots used are monitored.*

OSG by numbers

- ⇒ 53 Compute Elements
- ⇒ 9 Storage Elements
(8 SRM/dCache & 1 SRM/DRM)
- ⇒ 23 active Virtual Organizations

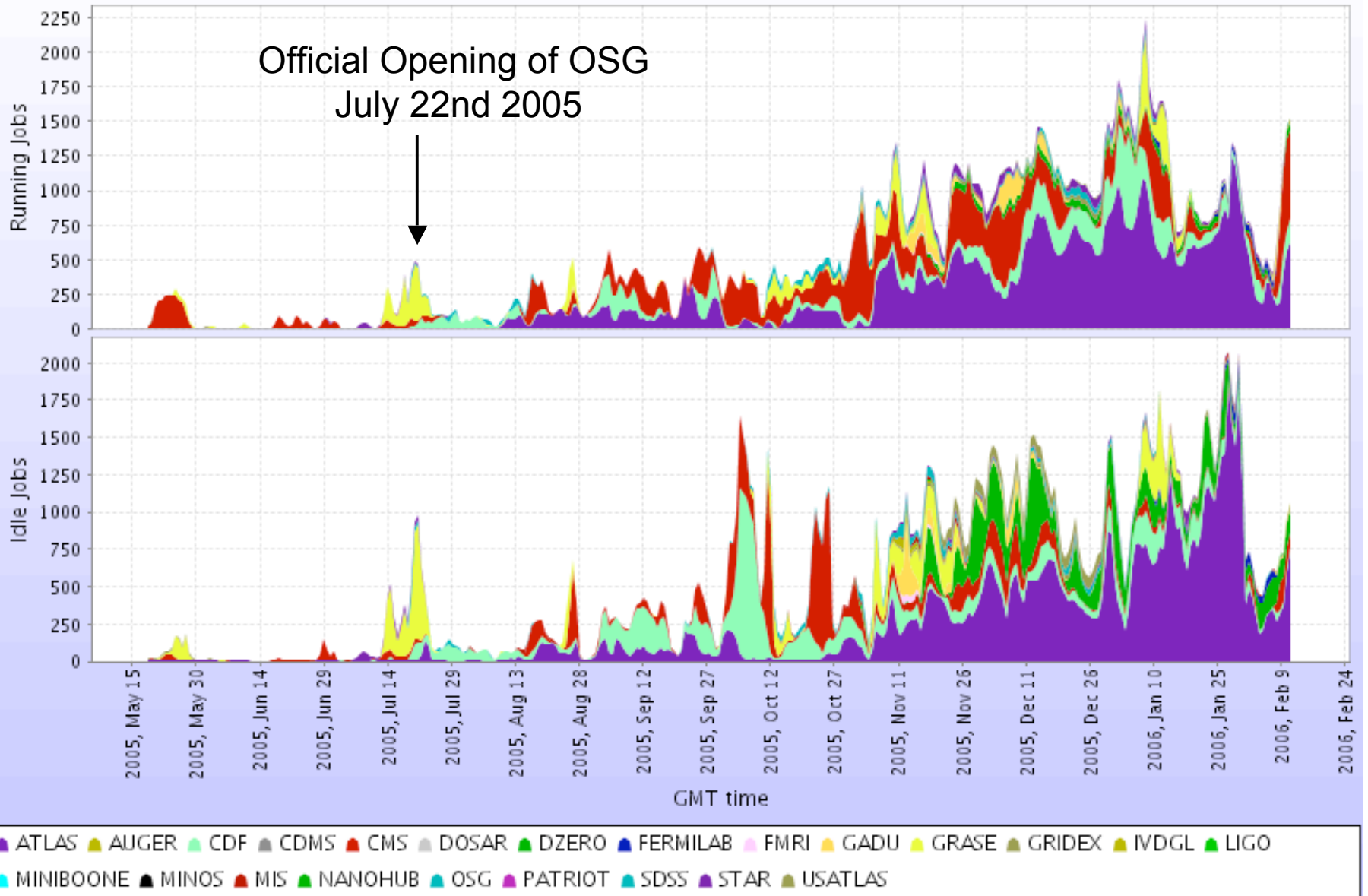
4 VOs with >750 jobs max.

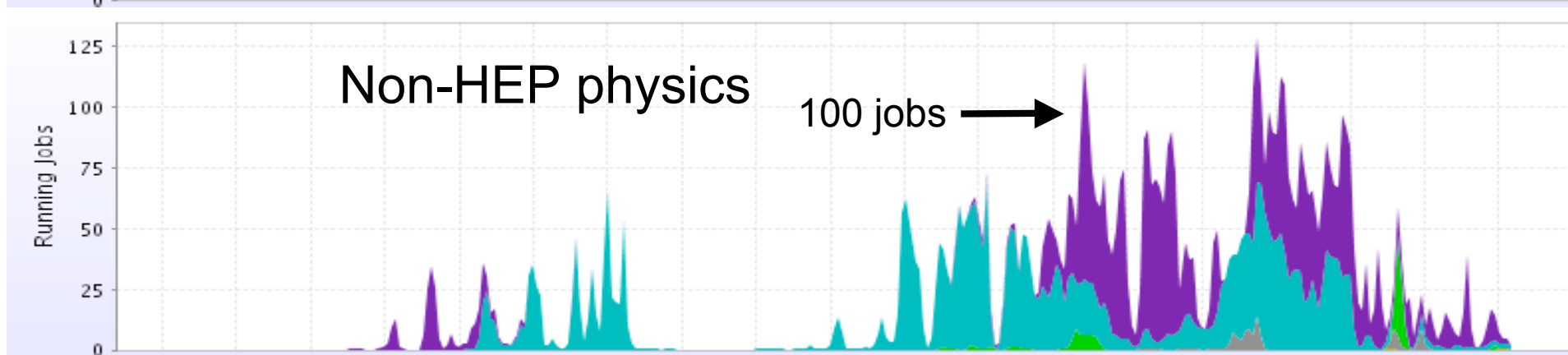
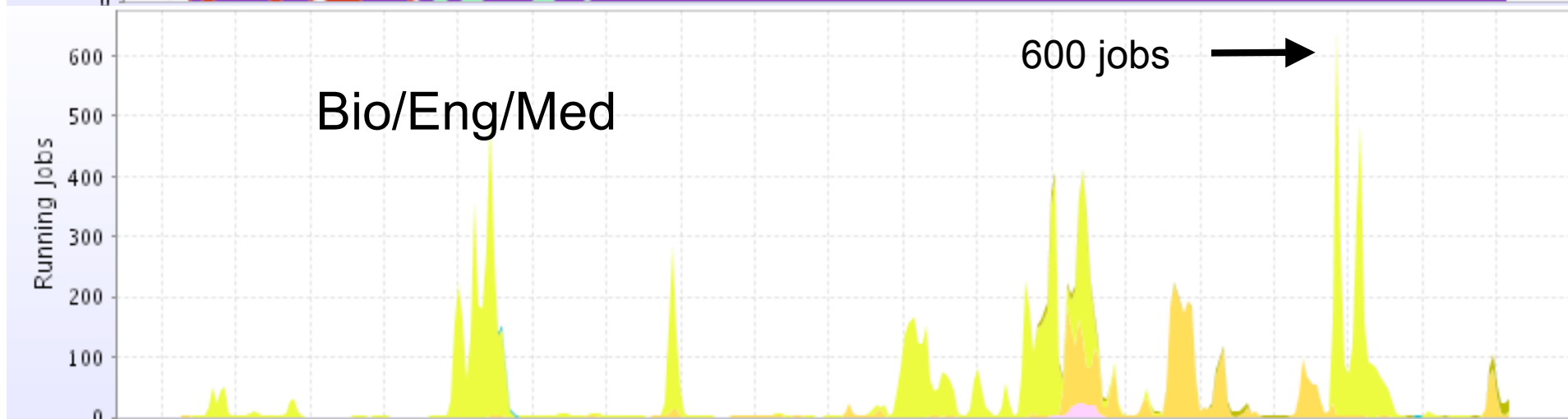
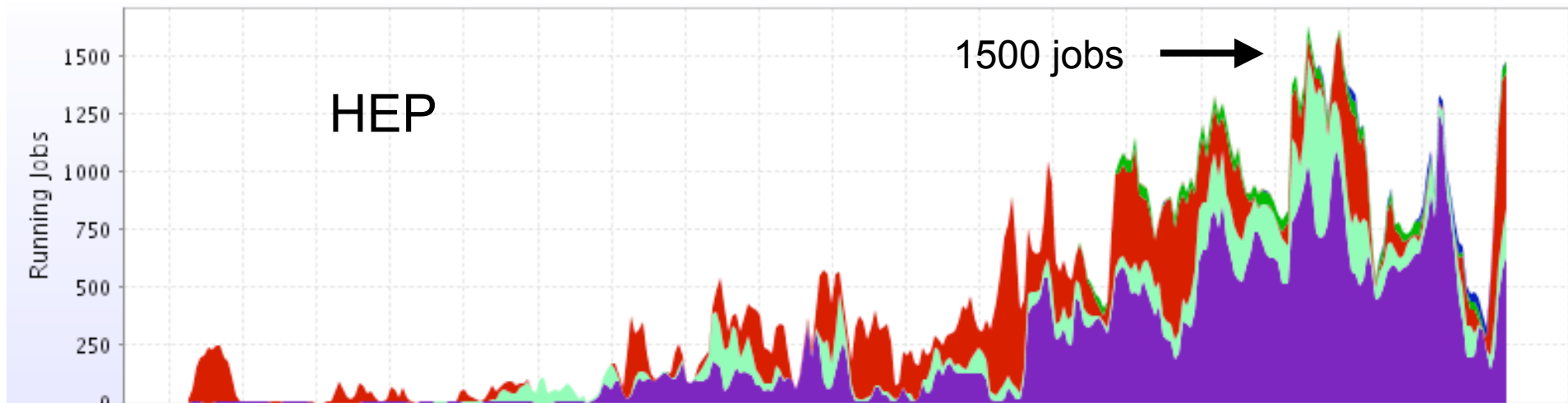
4 VOs with 100-750 max.

Farm	Running Jobs			
	Last value	Min	Avg	Max
ATLAS	618.6	0	232	2282
AUGER	0	0	0	5
CDF	171.9	0	74.22	752
CDMS	0	0	0.623	40
CMS	636.7	0	133.6	980
DOSAR	0	0	0.004	10
DZERO	56.69	0	17.57	77
FERMILAB	0	0	3.709	186
FMRI	0	0	0.459	46
GADU	0	0	10.71	301
GRASE	0.155	0	34.52	1045
GRIDEX	10.66	0	6.39	38
IVDGL	19.9	0	2.517	83
LIGO	0	0	0.681	82
MINIBOONE	0	0	0.001	14
MINOS	0	0	0	15
MIS	0.041	0	0.437	34
NANOHUB	0	0	0.002	1
OSG	0	0	0.095	14
PATRIOT	1	0	0.057	2
SDSS	3.421	0	11.59	244
STAR	0	0	11.92	114
USATLAS	0	0	2.166	16
Total	1519			



Total Jobs per VO





Roadmap



Extending the functionality (examples)

- ⇒ Storage Systems & data management
 - Widespread deployment of SRM v2, and beyond ...
- ⇒ Edge Services Framework
- ⇒ Advanced network services
- ⇒ Security enhancements
- ⇒ Advanced workflow and workload management
 - “late binding”
 - VDS enhancements

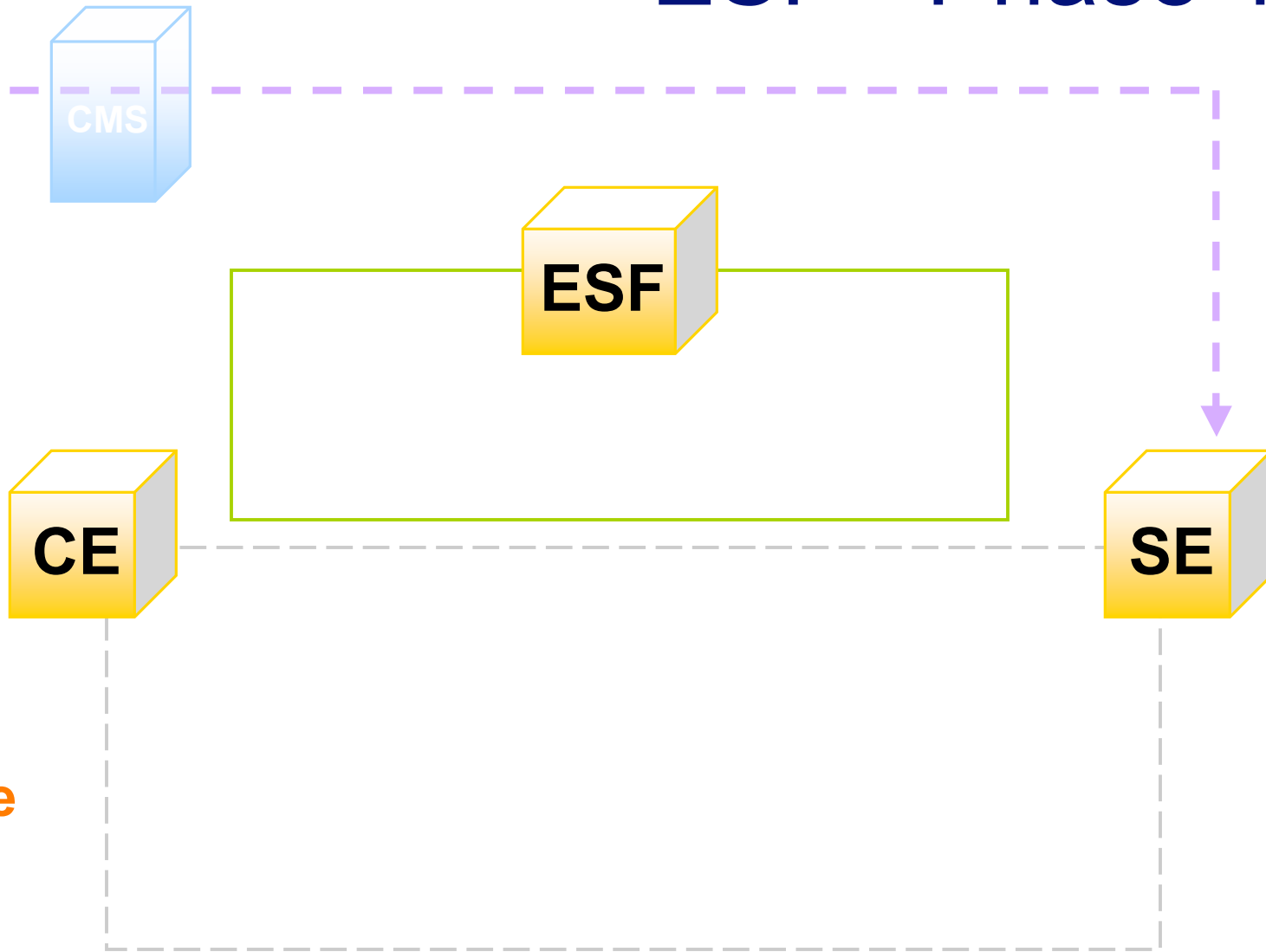
Can there be a shared Services Framework that makes site admins happy?

- ⇒ No login access to strangers.
- ⇒ Isolation of services.
 - ⇒ VOs can't affect each other.
 - ⇒ VOs receive a strictly controlled environment.
- ⇒ Encapsulation of services.
 - ⇒ Service instances can receive security review by site before they get installed.

Explore solution based on virtual machines.

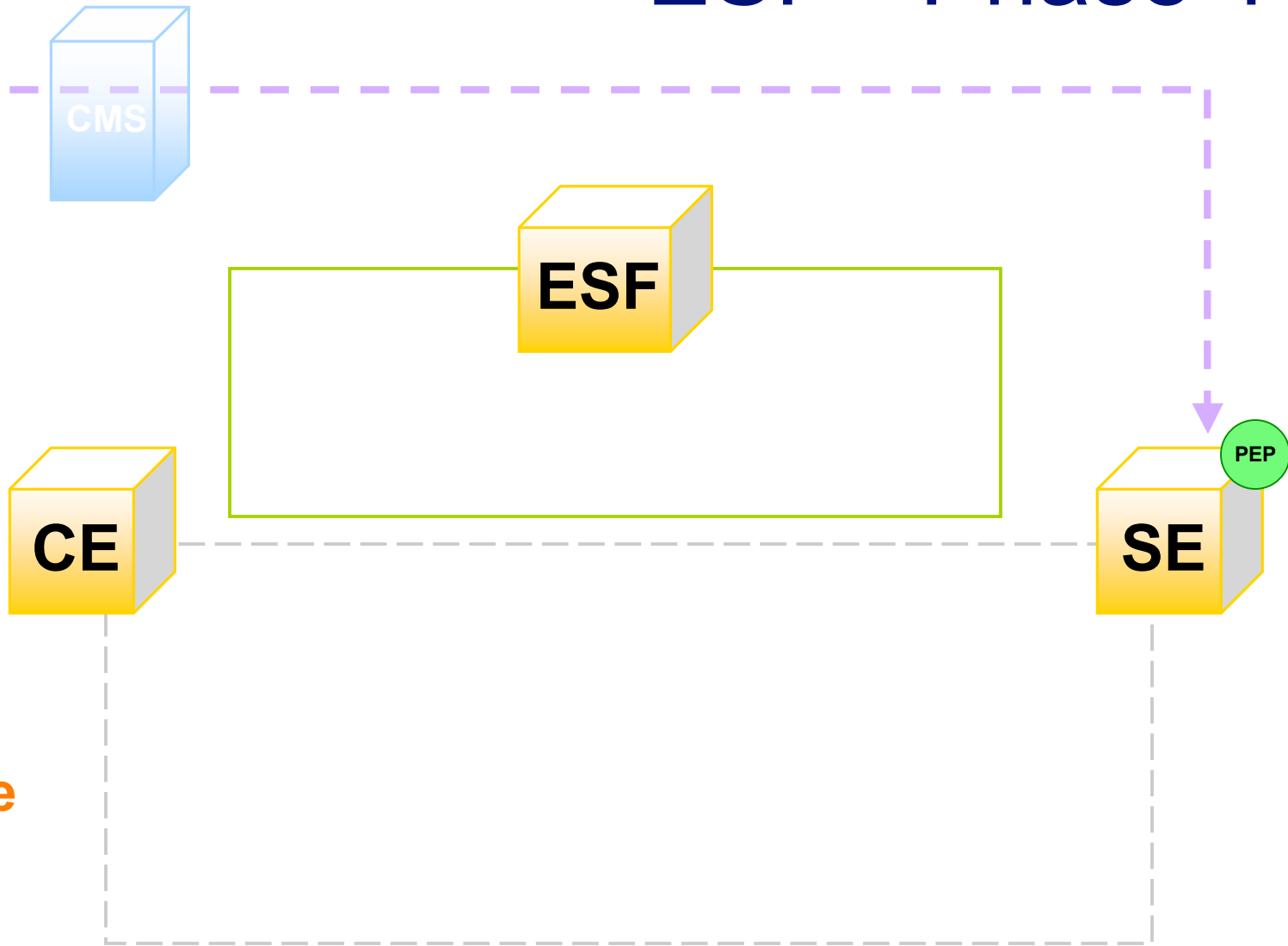
ESF - Phase 1

Role=VO Admin



ESF - Phase 1

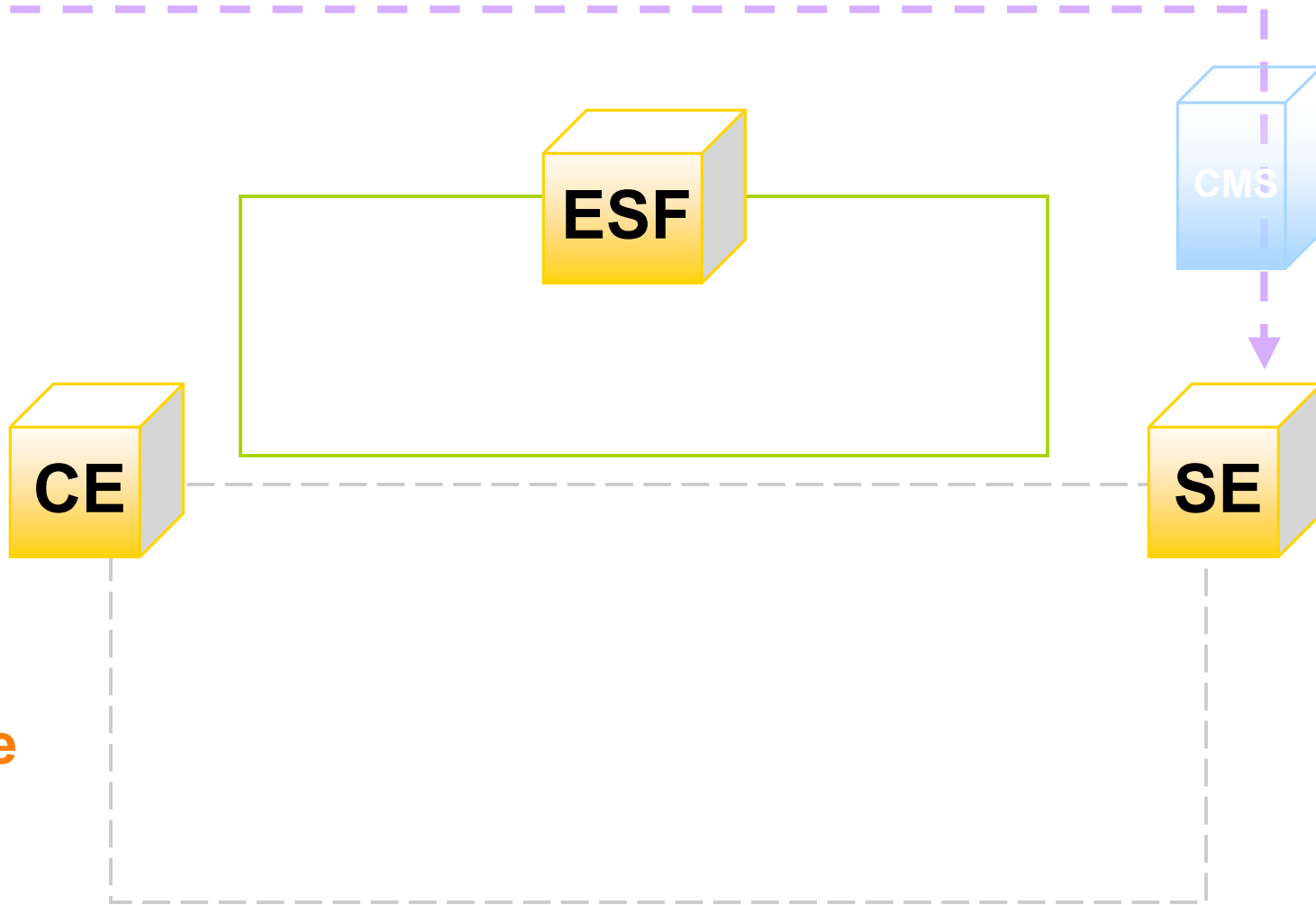
Role=VO Admin



Role=VO Admin

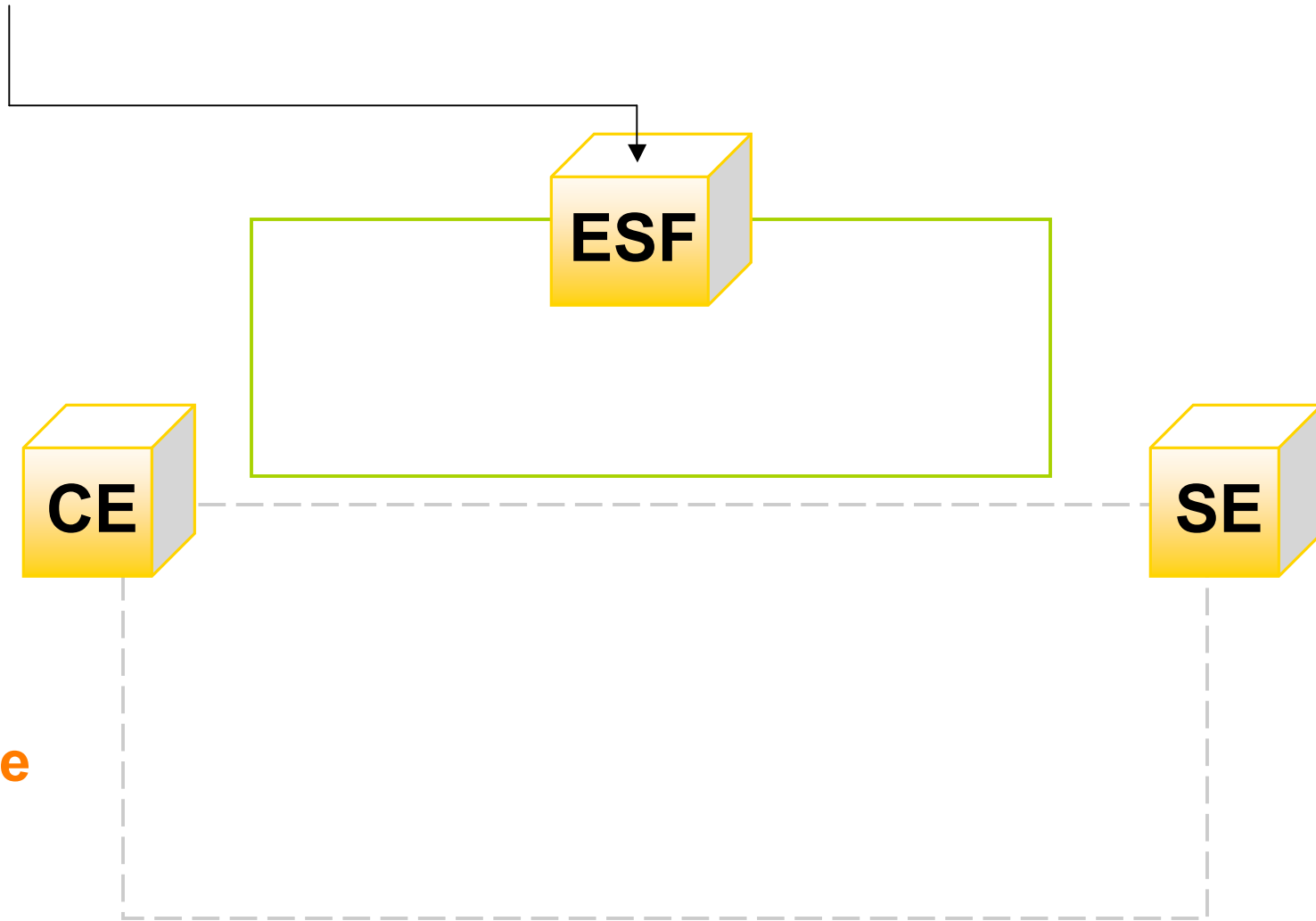


ESF - Phase 1



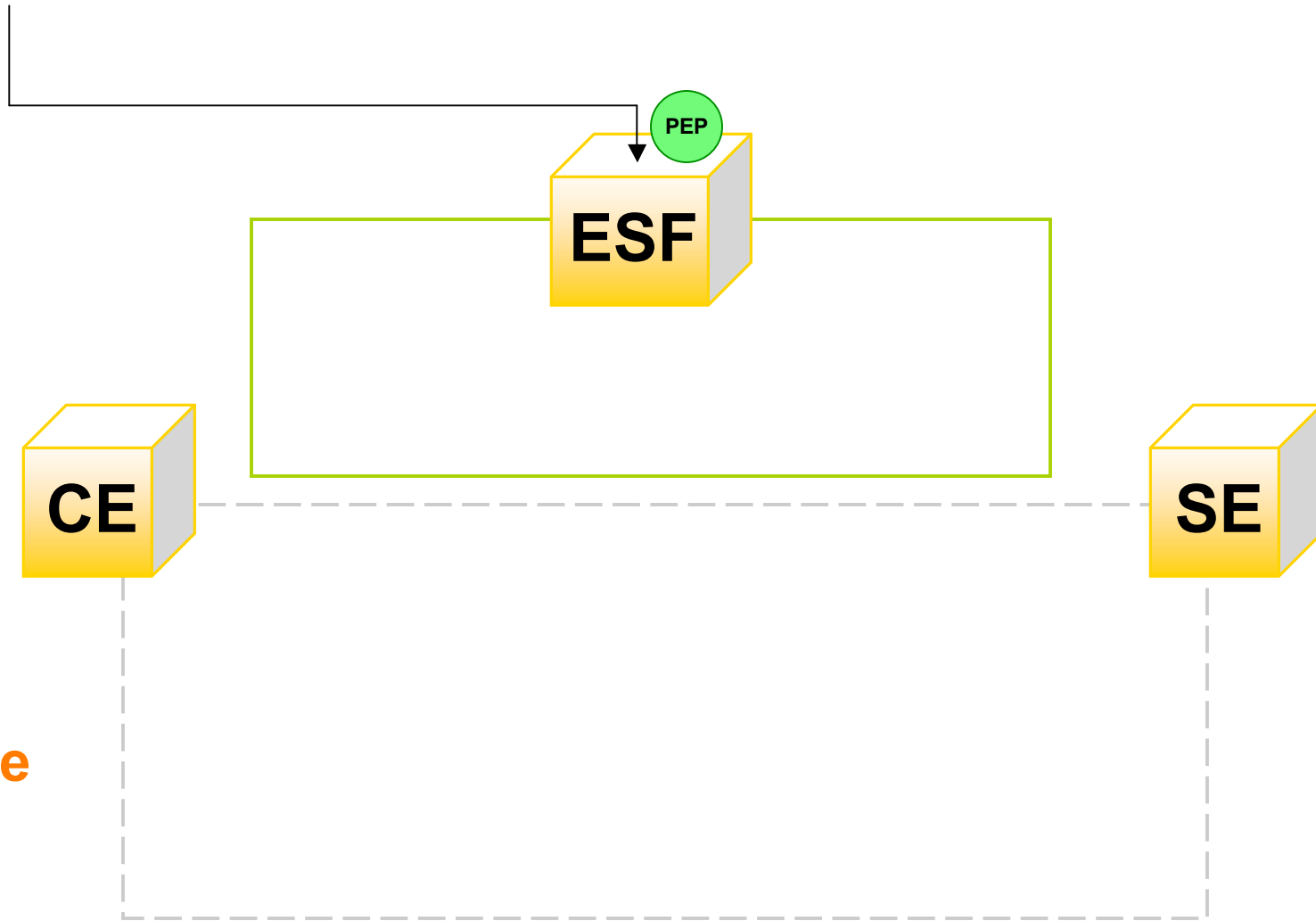
ESF - Phase 1

Role=VO Admin



ESF - Phase 1

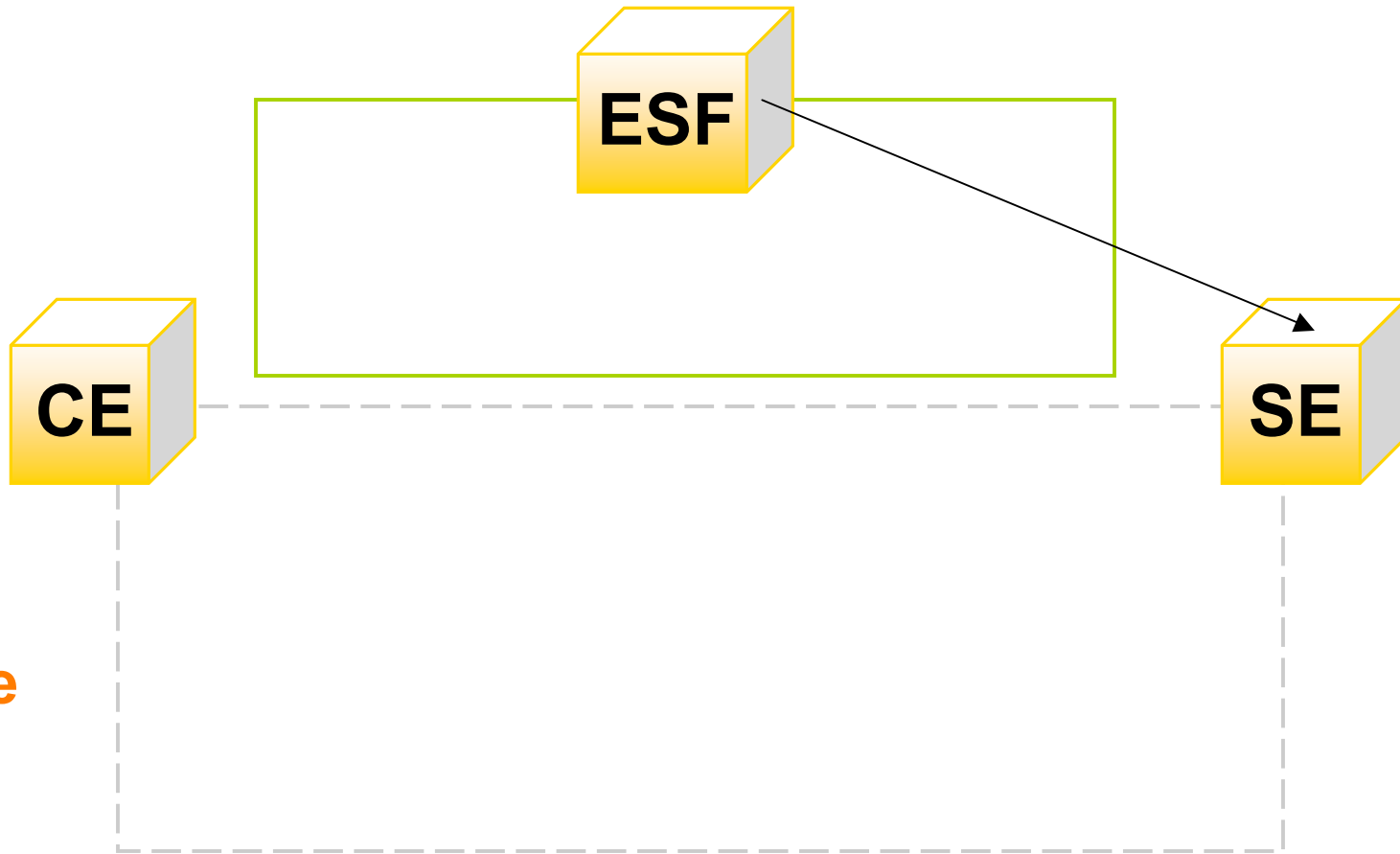
Role=VO Admin



Role=VO Admin



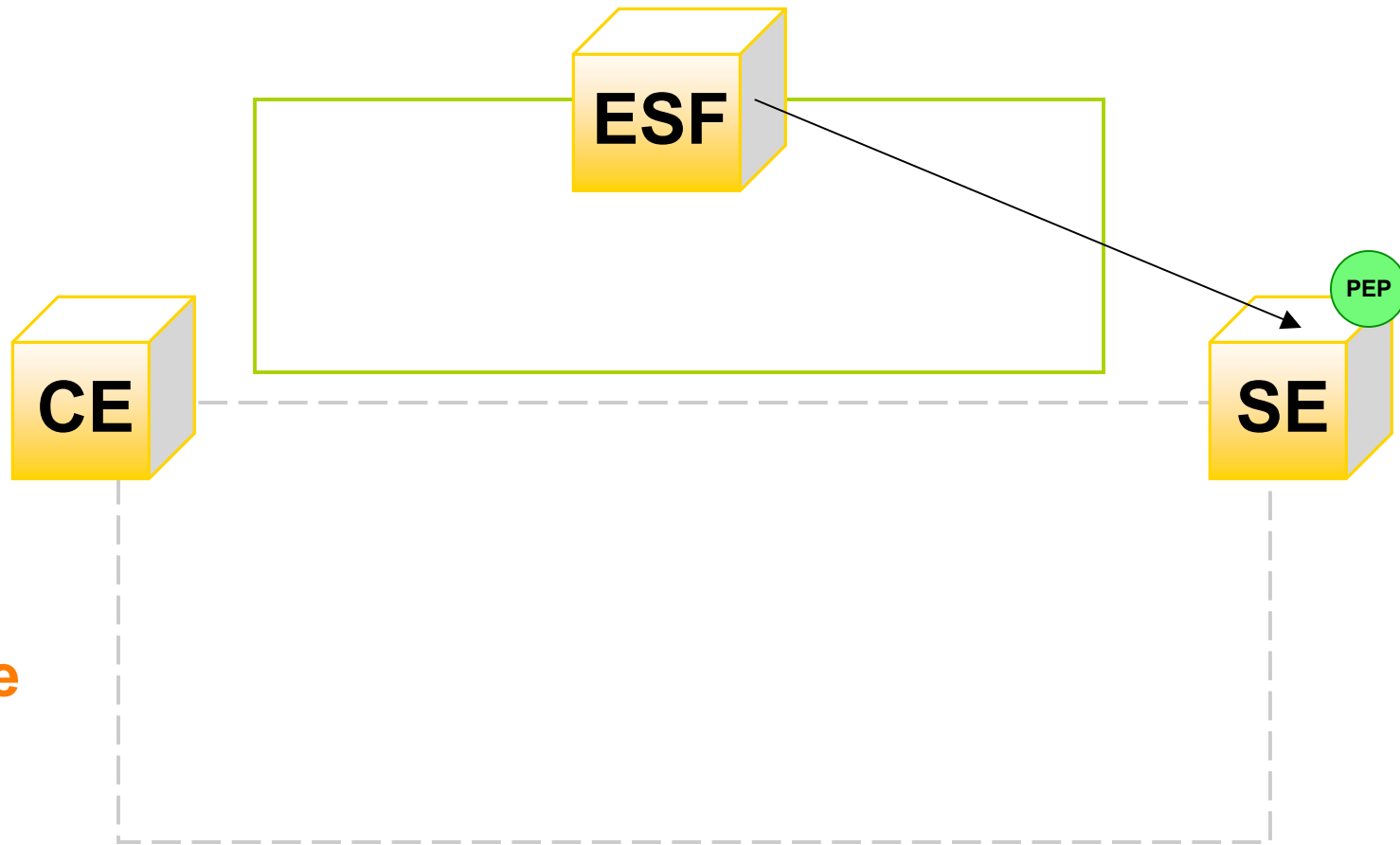
ESF - Phase 1



Role=VO Admin



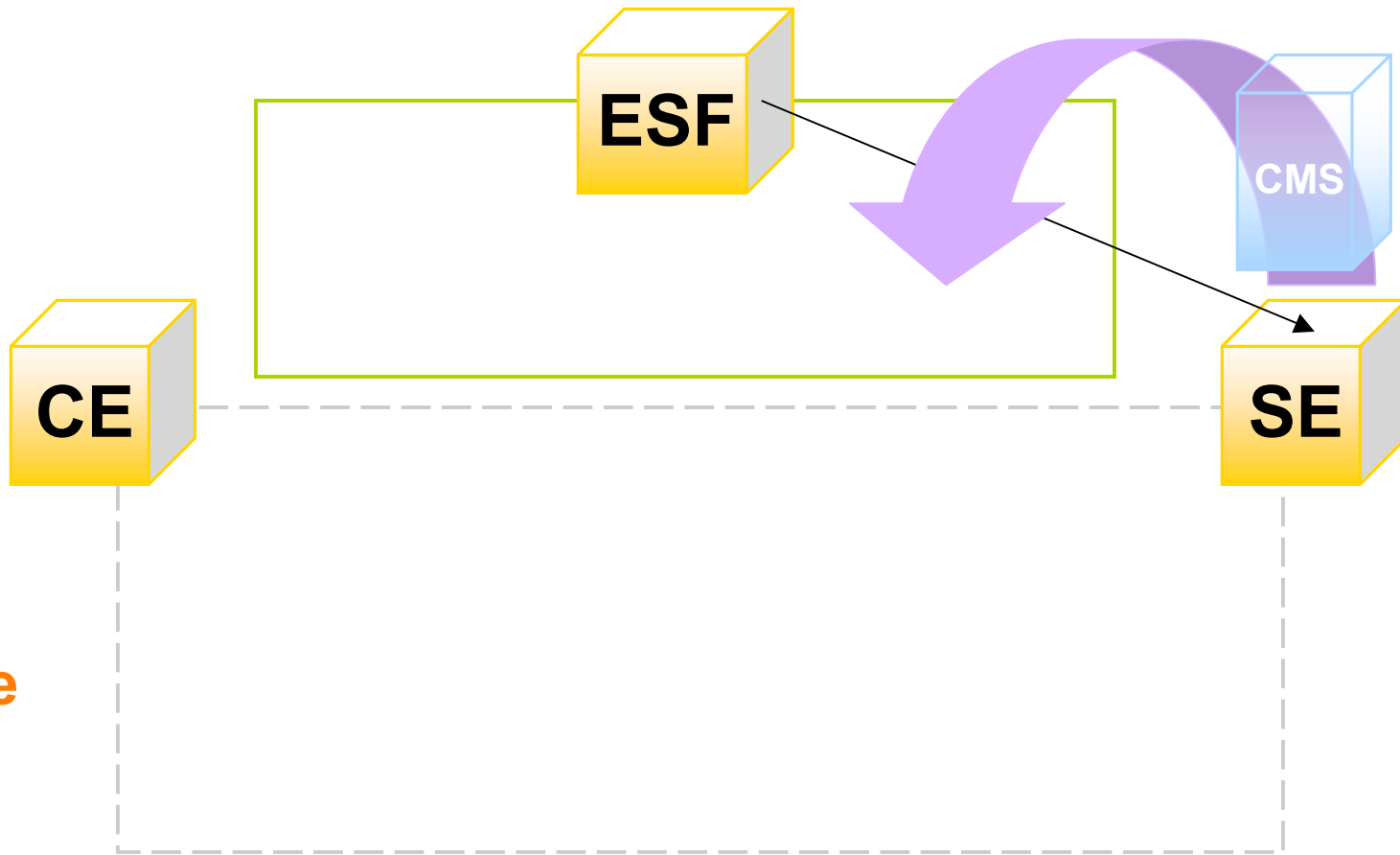
ESF - Phase 1



Role=VO Admin



ESF - Phase 1



Open Science Grid

2/21/2006

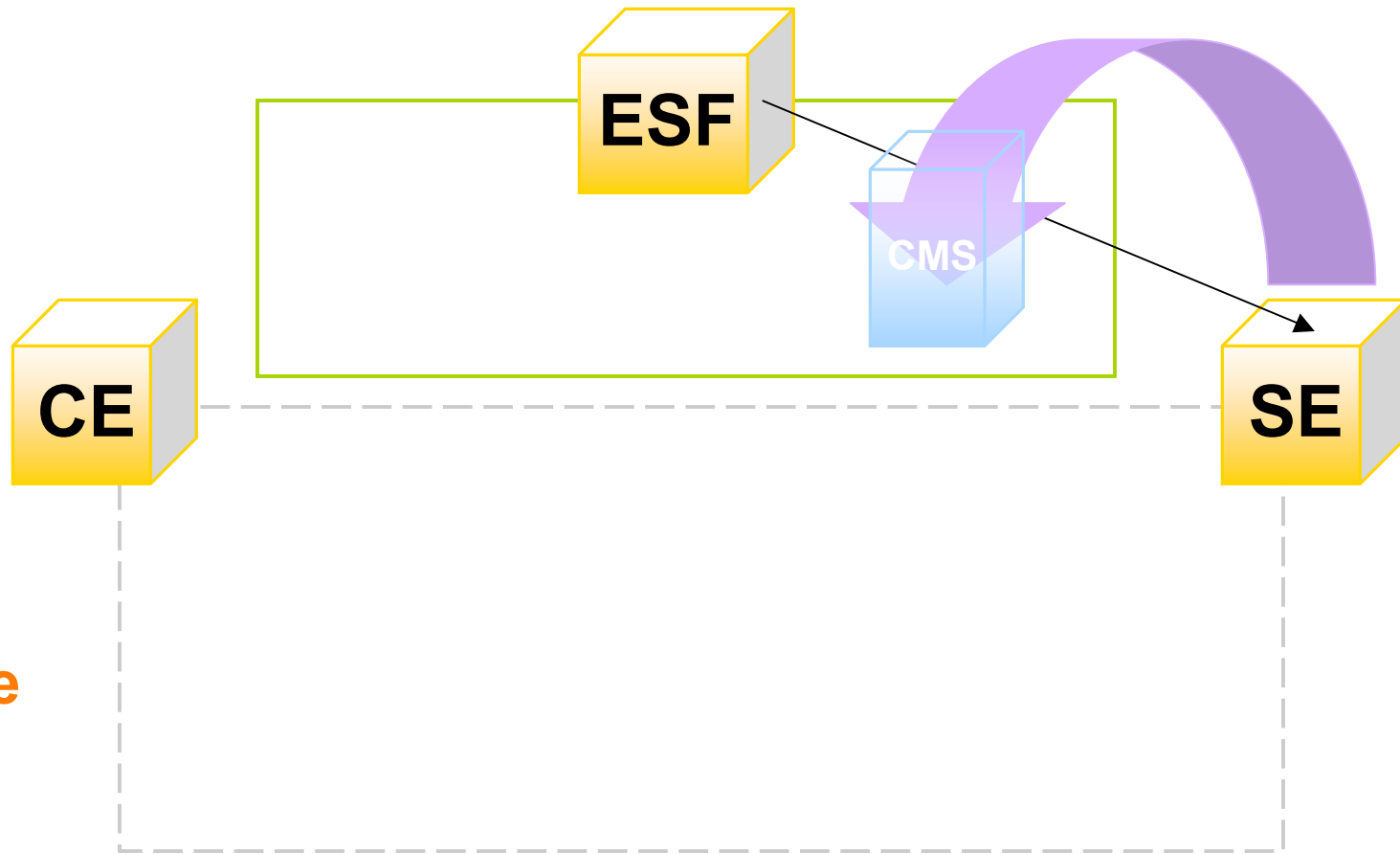
SDSC/CIP

55

Role=VO Admin



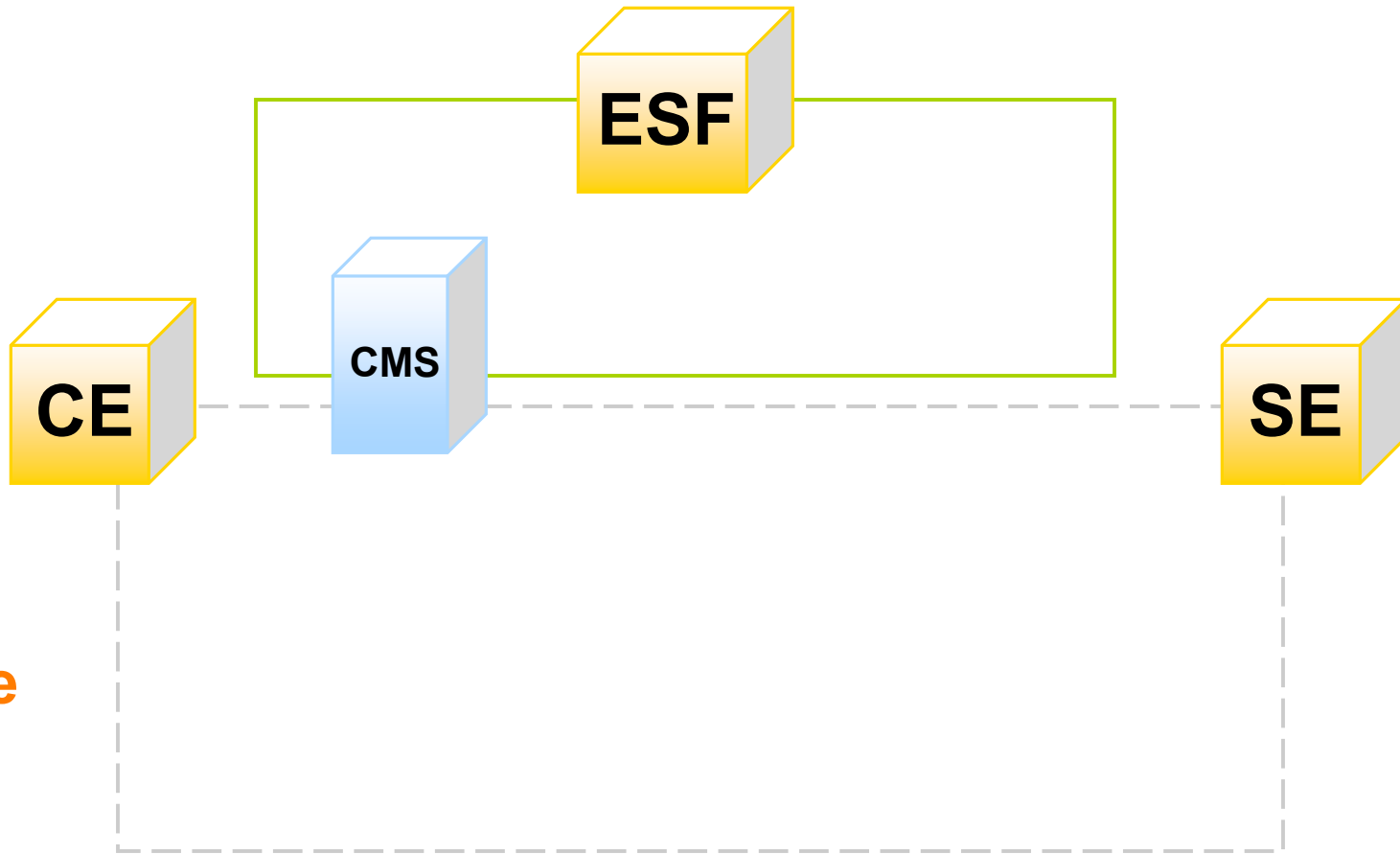
ESF - Phase 1



Role=VO Admin



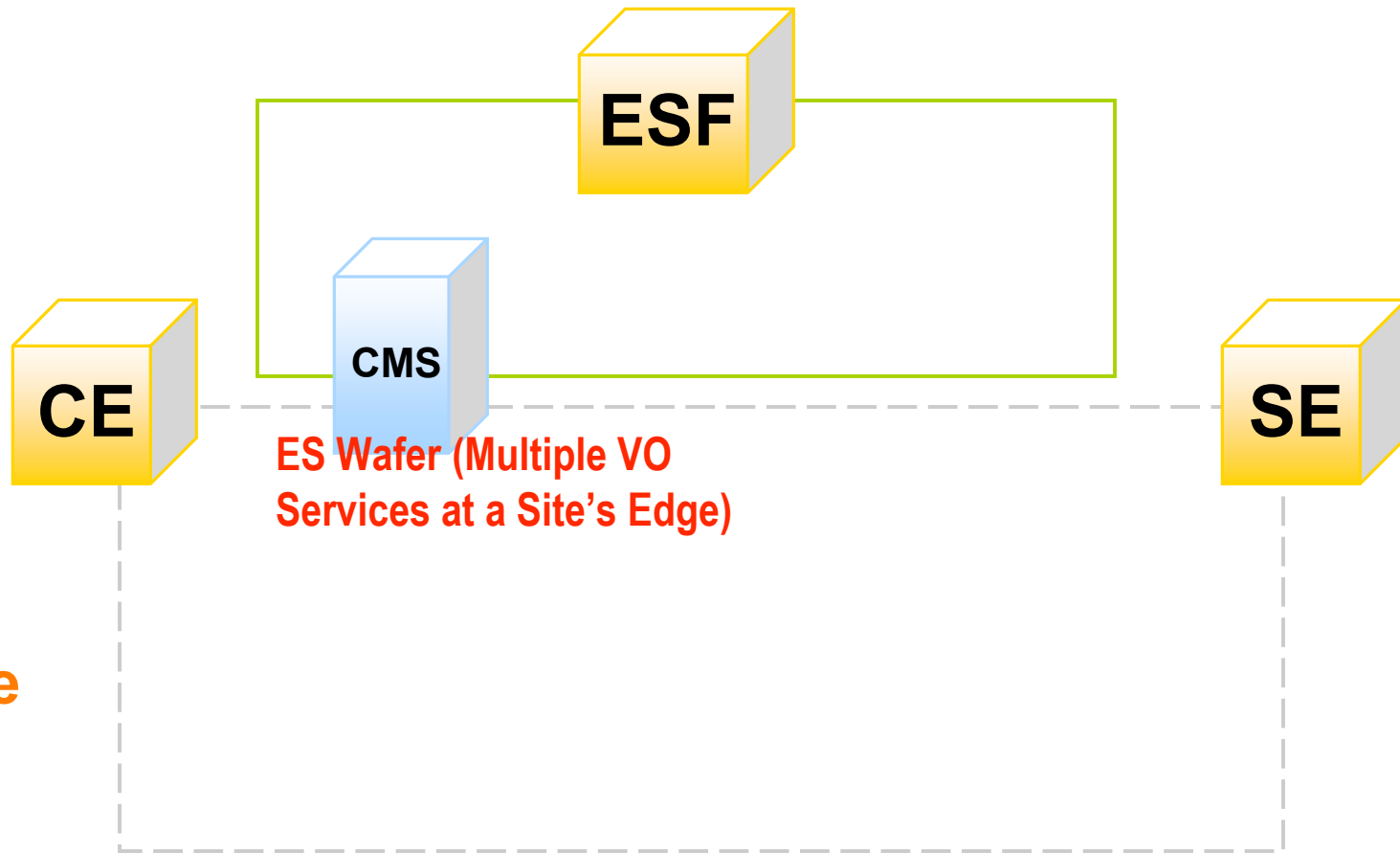
ESF - Phase 1



Role=VO Admin

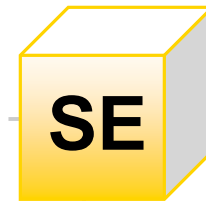
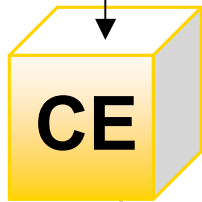
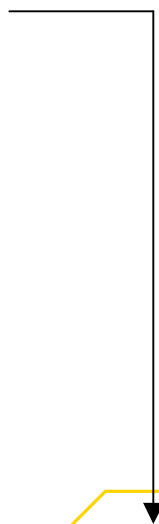


ESF - Phase 1



ESF - Phase 1

Role=VO User

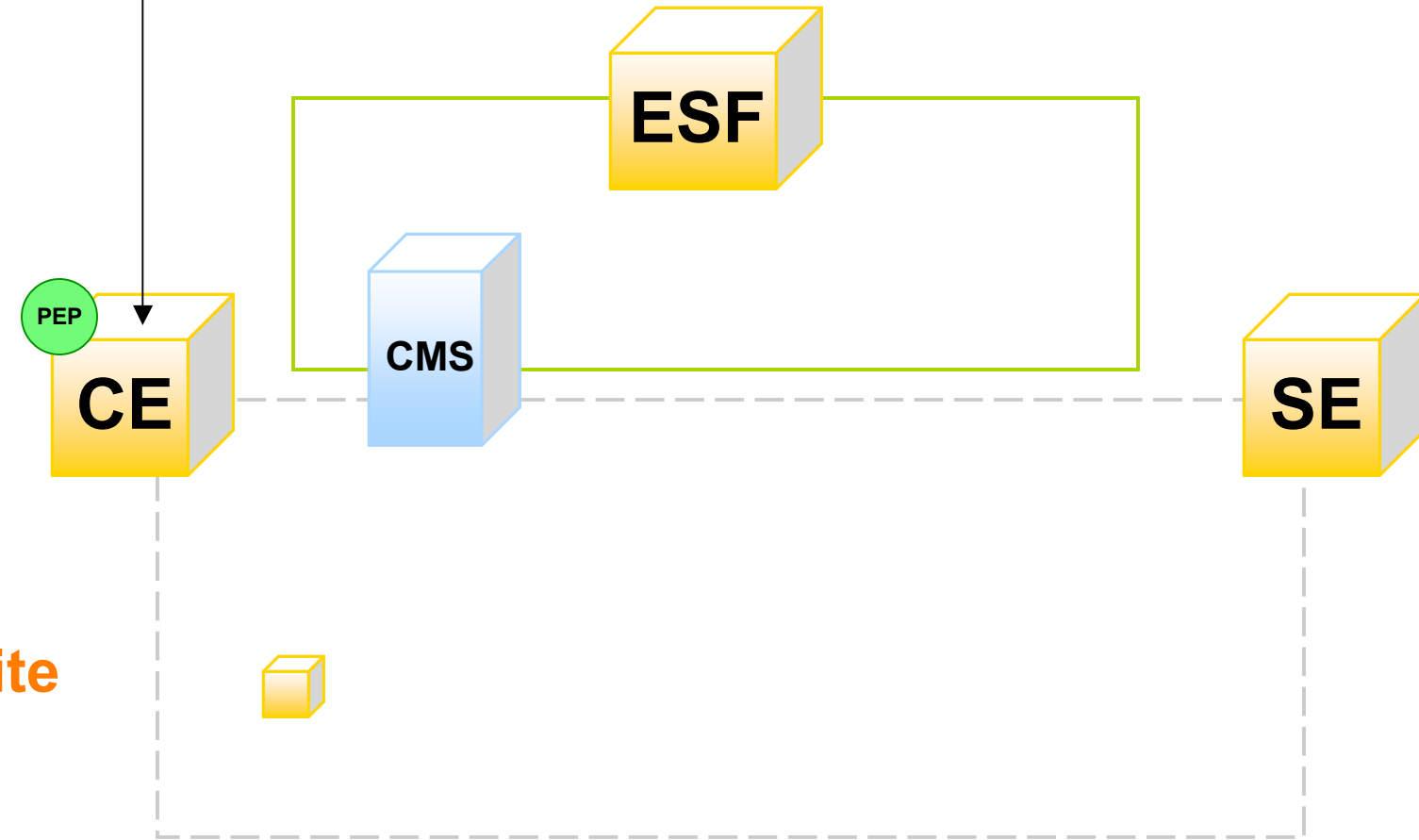


Site



ESF - Phase 1

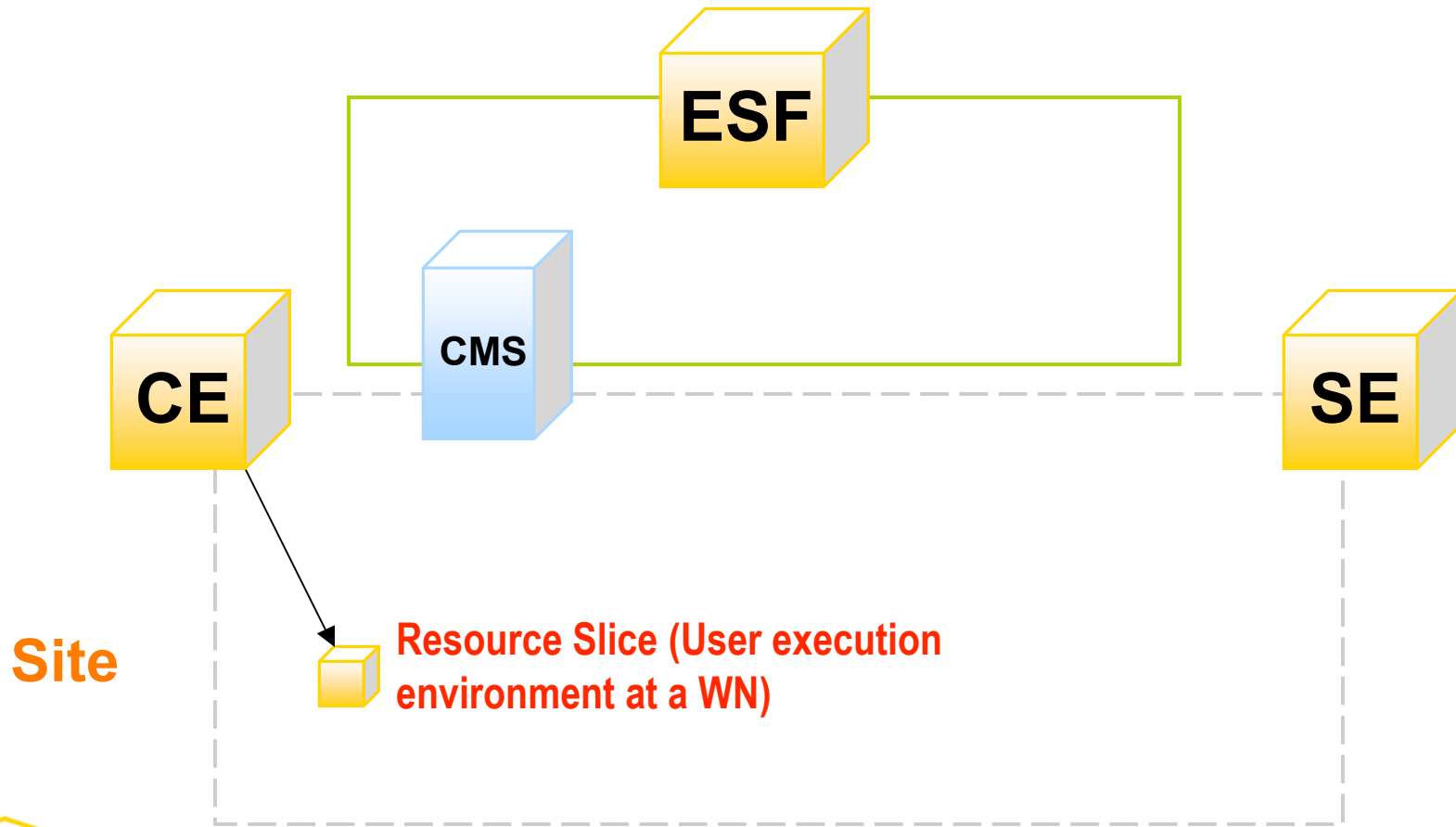
Role=VO User



Role=VO User



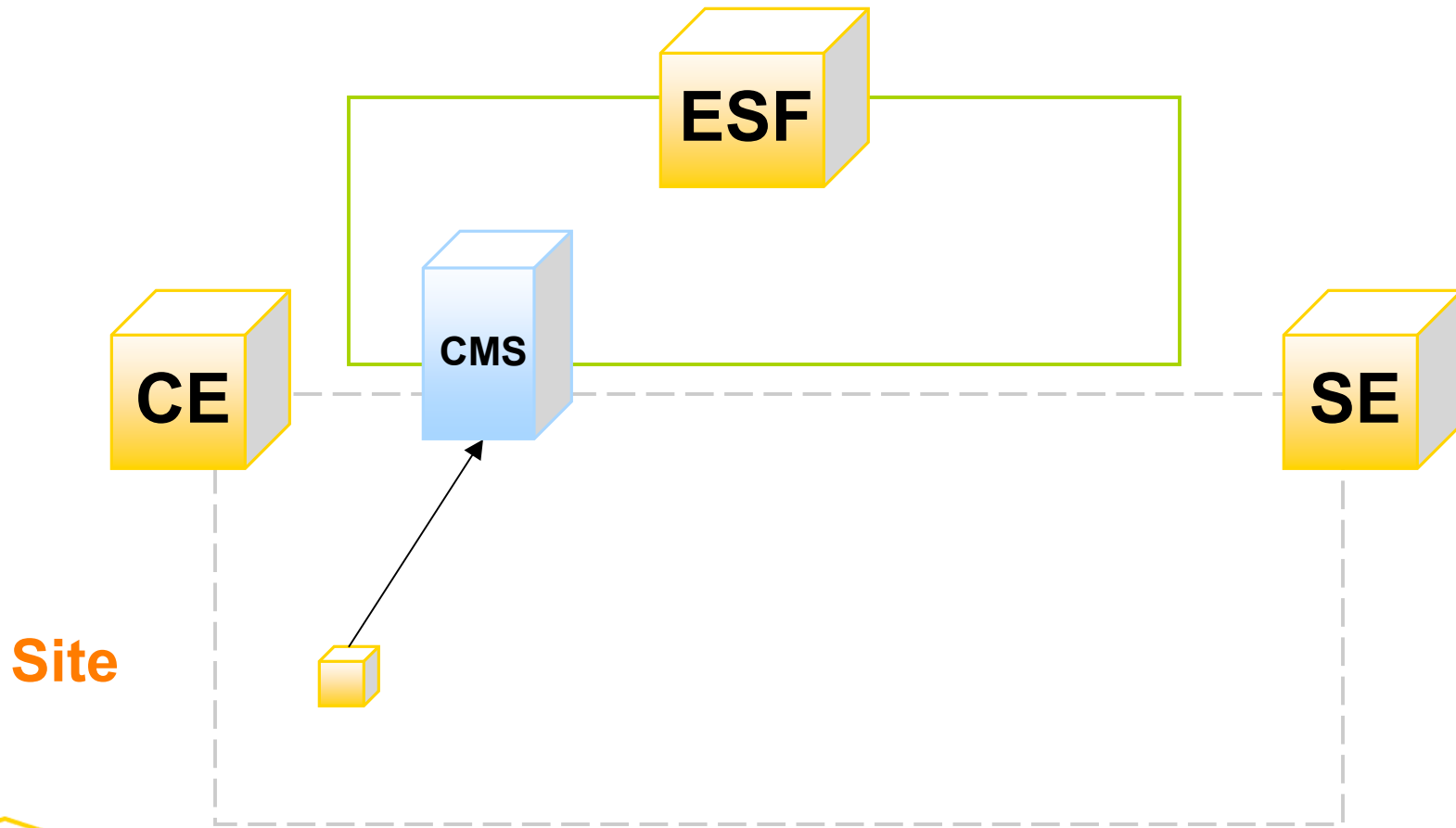
ESF - Phase 1



Role=VO User



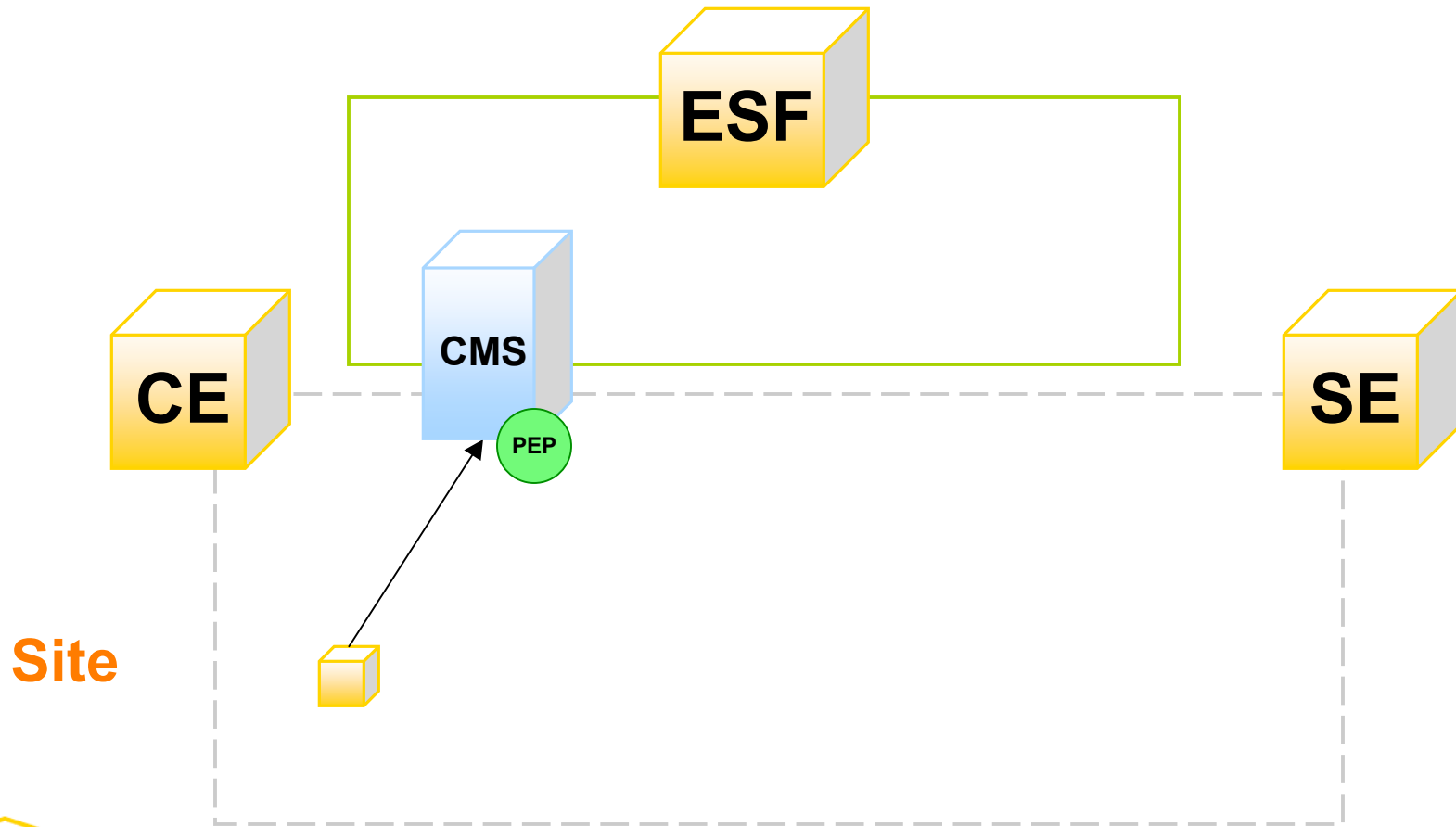
ESF - Phase 1



Role=VO User



ESF - Phase 1



Short term Roadmap



Release Schedule

	Planned	Actual
OSG 0.2	Spring 2005	July 2005
OSG 0.4.0	December 2005	January 2006
OSG 0.4.1	April 2005	
OSG 0.6.0	July 2006	

Dates here mean “ready for deployment”.
Actual deployment schedules are chosen by each site,
resulting in heterogeneous grid at all times.

Summary

- ⇒ OSG facility opened July 22nd 2005.
- ⇒ OSG facility is under steady use
 - ~20 VOs, ~1000-2000 jobs at all times
 - Mostly HEP but large Bio/Eng/Med occasionally
 - Moderate other physics (Astro/Nuclear)
- ⇒ OSG project
 - 5 year Proposal to DOE & NSF
 - Facility & Extensions & E&O
- ⇒ Aggressive release schedule for 2006
 - January 2006: 0.4.0
 - April 2006: 0.4.1
 - July 2006: 0.6.0

