Grid and Cloud Operations Interoperability – An overview

Alberto Di Meglio (CERN, EMI)
Morris Riedel (FZJ, EMI)
Shahbaz Memon (FZJ, EMI)
Cal Loomis (CNRS/LAL, StratusLab)
Davide Salomoni (INFN, WNoDeS)

EUROPEAN MIDDLEWARE INITIATIVE

StratusLab

WNoDeS
Grid / Cloud Integration Framework

Content

- Definitions
- Aiming at the right target
- Implementations and implementers
- Conclusions

definitions

Definition of the grid

- "A grid is a system that coordinates resources that are not subject to centralized control (within different control domains) using standard, open, general-purpose protocols and interfaces to deliver nontrivial qualities of service" [1]

[1] Ian Foster, What is the Grid? A Three Point Checklist, Argonne National Laboratory & University of Chicago , July 20, 2002

Definition of the cloud

- "Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [2]

[2] National Institute of Standards and Technology (NIST) - Computer Security Division - Computer Security Resource Center

Grid and Cloud Relationship

- From the previous basic-principle definitions, the major difference seems to be a difference in focus:
  - grid is a model for federating resources to provide a set of computing and data services
  - cloud is a resource and services provisioning model
- However, there is no hard incompatibility between the definitions

Grid and Cloud Comparison

	grid	cloud
Users own resources	No	No
Resources are in a known physical place	No	No
Resources are used and discarded when not needed anymore	Yes	Yes
API-based access	Yes	Yes
Federation or coordination of resources across different domains	Yes	No (not yet?)
Computational/storage abstraction	Job/file	Machine/application/ service/disk
Fast VM provisioning/de-provisioning (using APIs)	No	Yes
Interactive	No	Yes
Customized environment	No	Yes
Cost model	Investment+ maintenance	Pay-per-use

Interoperability and Integration

- **Interoperability**: the property by which two or more systems can work together.
  - Finding ways by which grid can make use of cloud without too much change in either of them
- **Integration**: the method or procedure by which a larger system is composed from smaller elements
  - Creating a new generation infrastructure using elements of grid and cloud and open standards

aiming at the right target

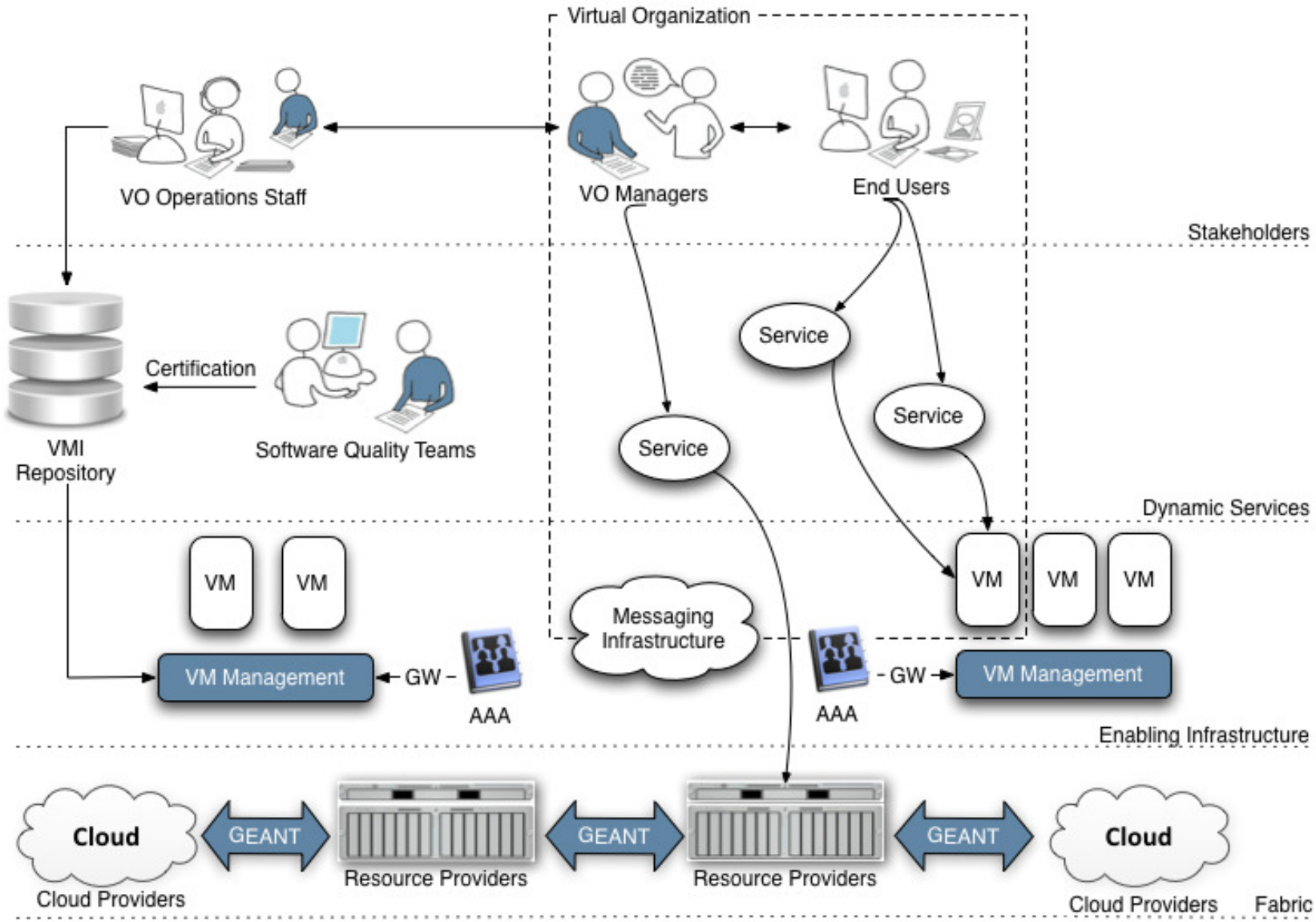
Main Use Case

- We refer to the specific case of interoperability and integration between grid and cloud in the context of the Distributed Computing Infrastructures (DCI) represented by EGI
- The main use case is therefore the provision of accessible, efficient and scalable computing and data services to scientific researchers, including both existing and new users

The DCI Projects

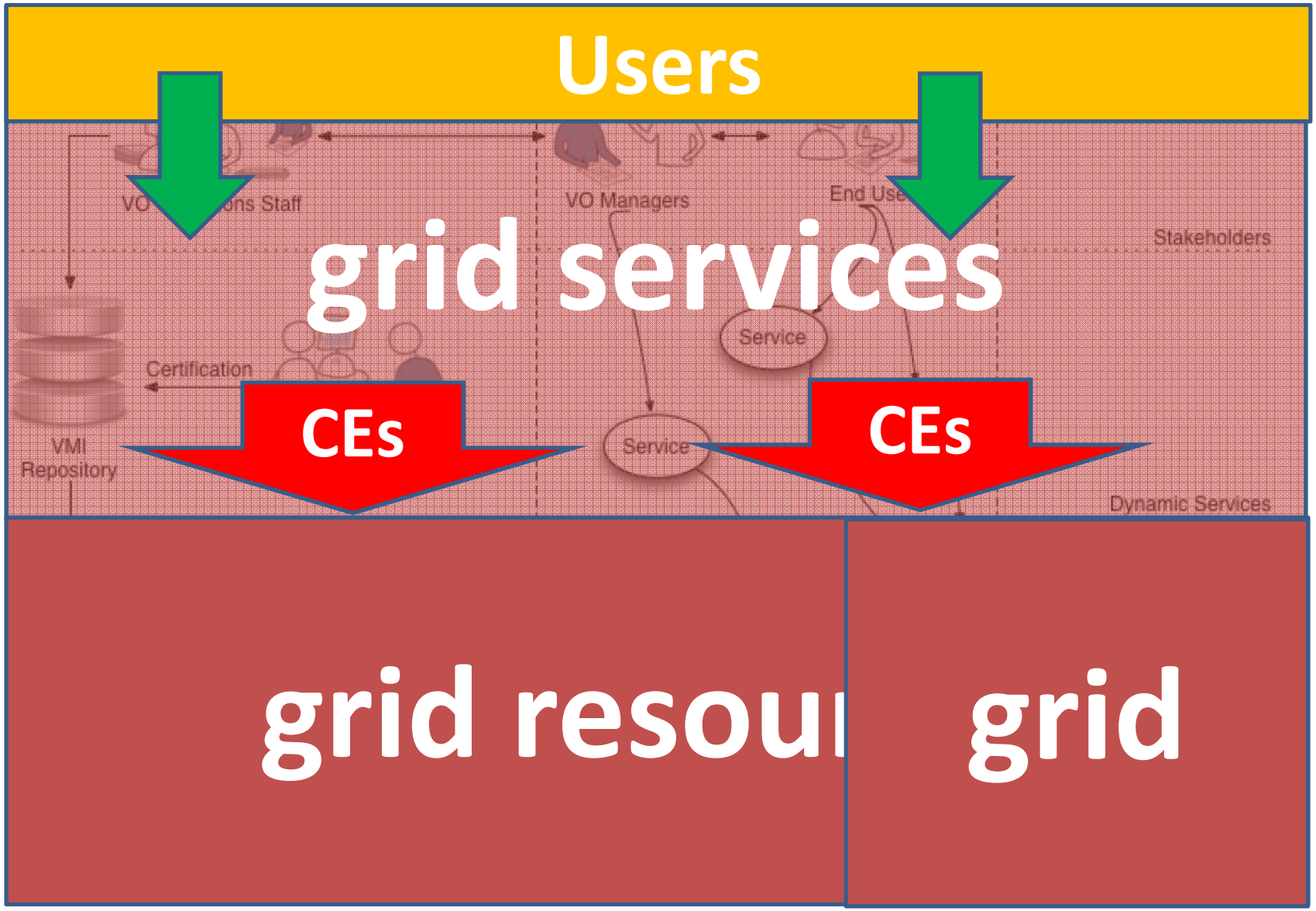
The DCI Roadmap: https://documents.egi.eu/document/172

The DCI Blueprint



implementations and implementers

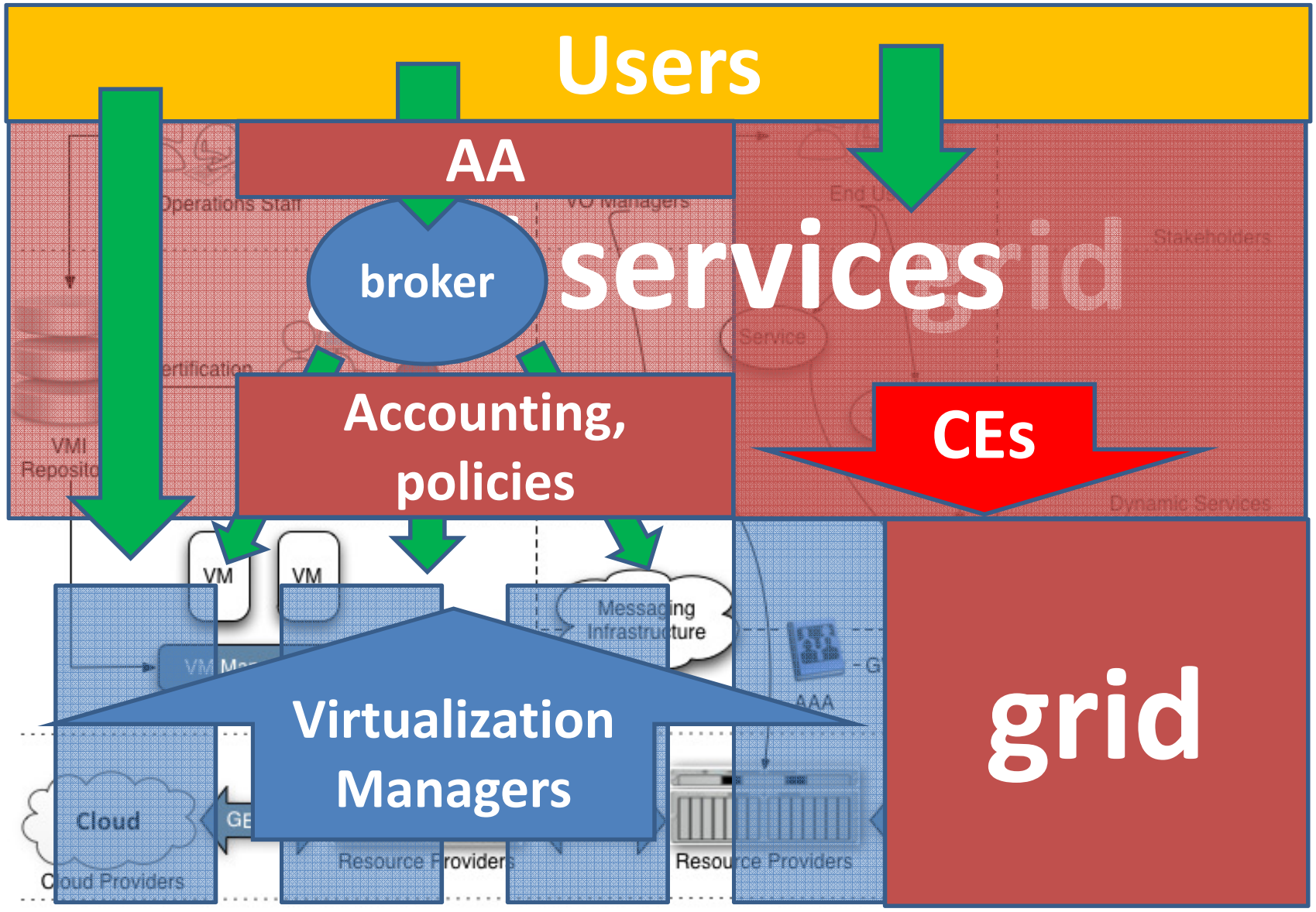
Model 1: Grid with private clouds



Model 1: Grid with private clouds

- Simplest to implement
- Very little modifications on the grid services (adapt CEs to see VM managers as an LRMS)
- Transparent for users
- Minimal impact for site admins
- Standardization role: VM managers APIs (for the CE support)
- Virtualized WN and storage space (non-persistent)

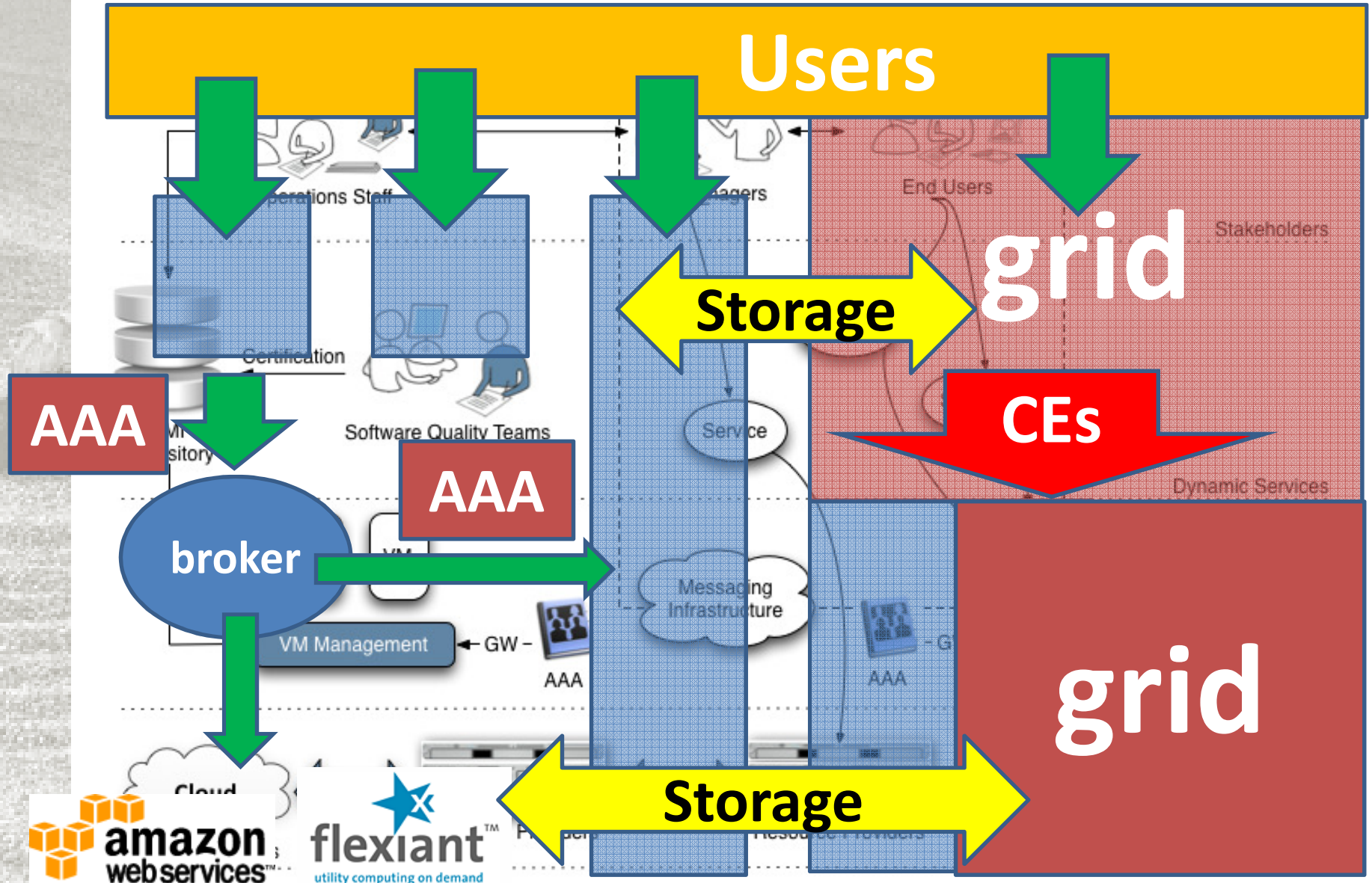
Model 2: Grid and cloud access



Model 2: Grid and cloud access

- Infrastructure part simple to implement
- Access model requires extension of AA policies and technologies, accounting and brokering mechanisms
- Very little modifications on the grid services (adapt CEs to see VM managers as an LRMS)
- Grid access transparent for users, new cloud access
- Minimal impact for site admins
- Standardization role: VM APIs, AA, storage interfaces, SRM, CDMI
- Virtualized WN, custom execution environments, persistent/shared storage space via mounted cloud file systems

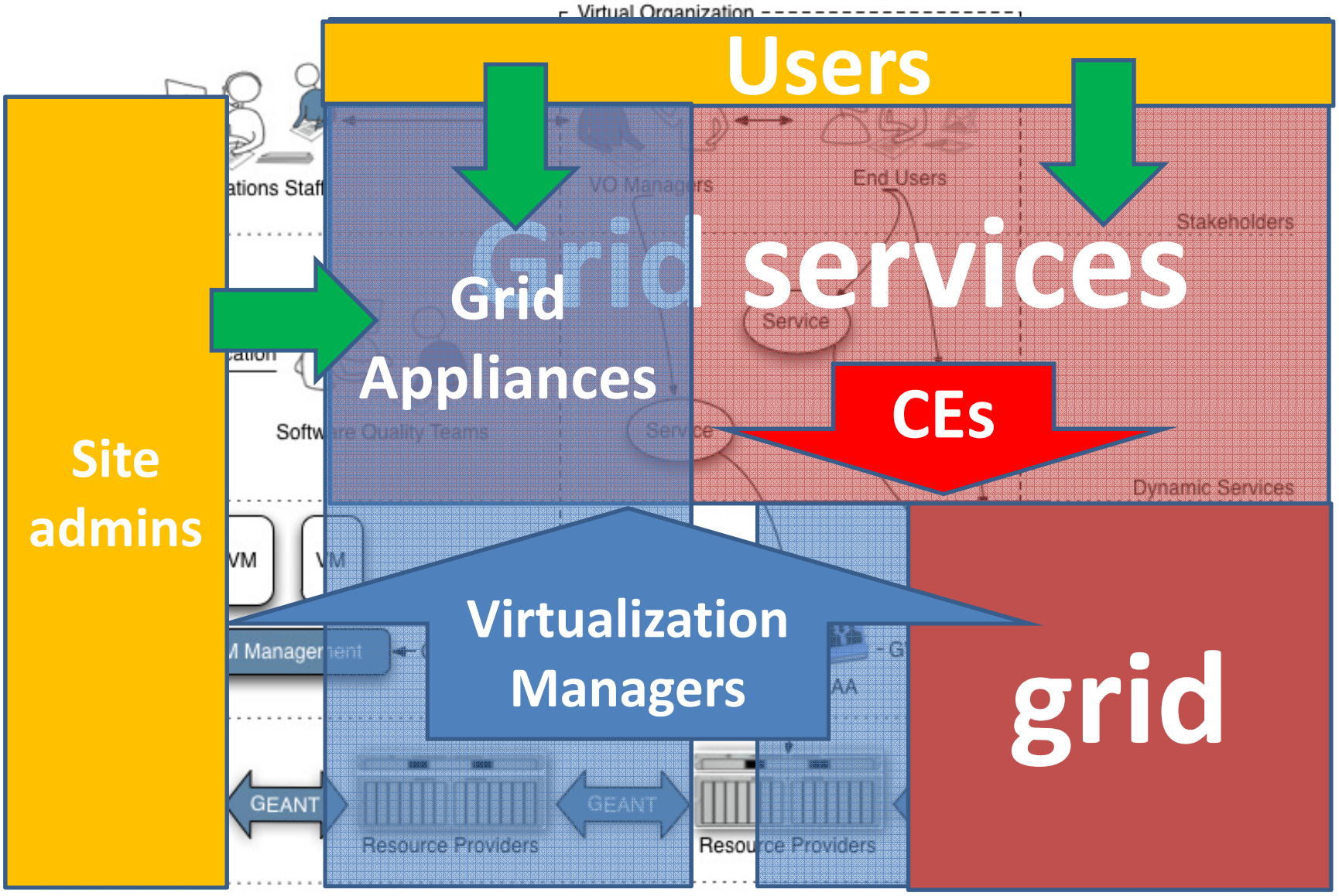
Model 3: Grid and hybrid cloud access



Model 3: Grid and hybrid cloud access

- Infrastructure part more complex to implement
- Access model requires extension of AA policies and technologies, accounting and brokering mechanisms
- Very little modifications on the grid services (adapt CEs to see VM managers as an LRMS)
- Grid access transparent for users, new cloud access
- More effort for site admins in setting up and managing the private/public cloud, peak time management possible
- Standardization role: VM APIs, AA, storage interfaces, SRM, CDMI,
- Virtualized WN, custom execution environments, persistent/shared storage space via mounted cloud file systems, commercial cloud storage
- Elasticity
- Complex integrated accounting and billing system

Model 4: Virtual grid services



Model 4: Virtual grid services

- Infrastructure and service part more complex to implement
- No special additional AAA requirements
- Potentially non trivial modifications to grid services to run as dynamically instantiated appliances
- Grid access transparent for users
- Grid potentially easier to set up for site admins
- Standardization role: no special requirements for grid/cloud interop, but pressure on grid standardization for service discovery and registration, dynamic configuration, info systems
- Potentially any service can be virtualized

Operational Challenges

- Deployment models
  - Impact on users, site admins, required technology, many combination or variations of the previous models can be envisaged, which ones make sense?
- Monitoring and accounting
  - Not necessarily a new technical problem
  - More a policy or standardization topic
  - Use Nagios, APEL, standard messaging technology

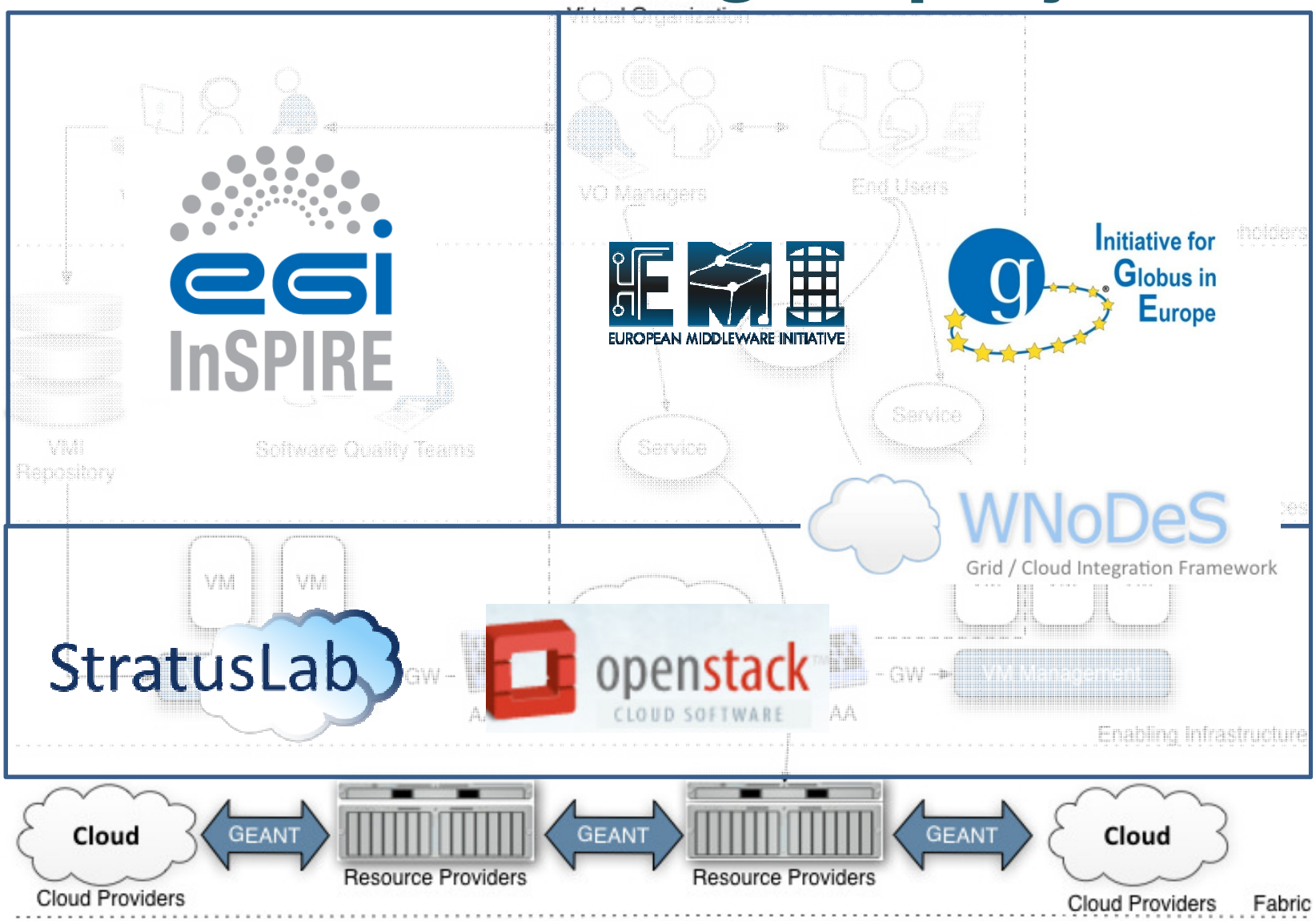
Operational Challenges

- Security
  - Different requirements, but existing VO based AA policies/technology have to be extended to the cloud
  - Reuse of existing services, e.g. VOMS, Argus, Gridsite delegation libraries
  - Support for VOMS certificates, proxies, roles, attributes is already being added to cloud VMs, work on cloud security and delegation done by Cloudsec in OGF

Operational Challenges

- Data management and access
  - Data in grid SEs accessed using SRM clients, posix or webdav interfaces from virtualized clients (AAI?)
  - Cloud storage space not a problem
  - Access of data stored on cloud storage from grid?
    - New standard interfaces (CDMI), existing interfaces (SRM)? On both sides?
  - Cost of data movements from/to public clouds
  - Data protection, international regulations

Relevant technologies/projects



Conclusions

- Clear definitions of cloud and of the use cases is important
- There are a number of possible models, some of them are technically feasible today, but production infrastructures cannot be disrupted, need to go step by step
- Cloud principles and technologies can be very beneficial to users and admins of grid, but several grid services are mature and relatively standard, should be extended and reused

Thank you

EMI is partially funded by the European Commission under Grant Agreement INFSO-RI-261611

StratusLab is co-funded by the European Community' under Grant Agreement INFSO-RI-261552