

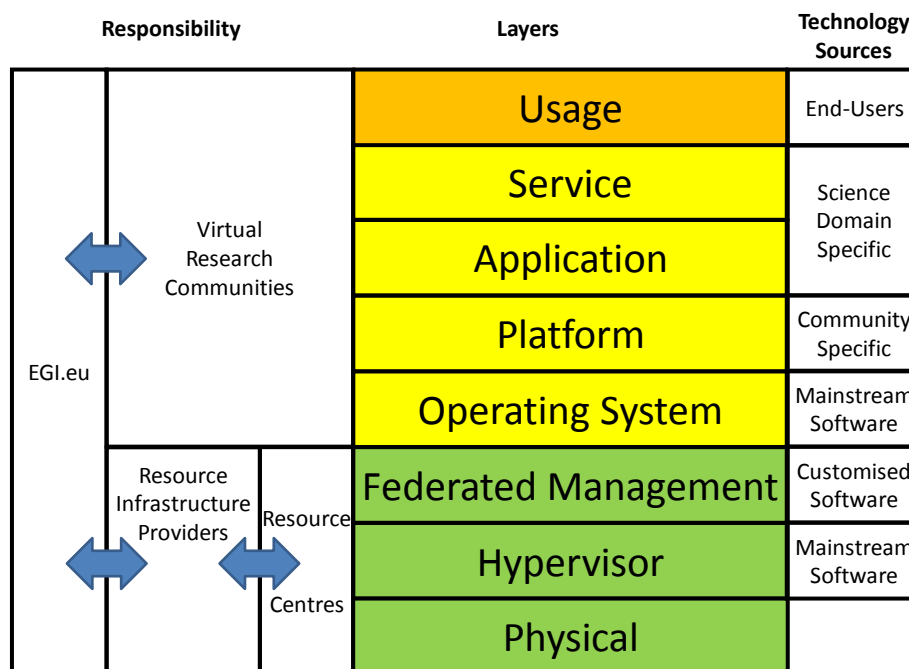
2 Technology

EGI relies on external technology providers to provide the software components needed to deliver the software environments used by its end-users. Some aspects of this software environment are under the direct responsibility of EGI as the infrastructure provider working on behalf of the resource infrastructure providers and their resource centres. Other elements of the software environment are of direct interest of the virtual research communities and their end-users.

EGI's technical architecture is aligned along the infrastructure, platform and software layers adopted by the cloud paradigm. Although EGI does not currently operate a cloud based infrastructure it is moving in that direction, and currently it does provide an operational infrastructure that supports a single platform for virtual research communities to integrate their applications (software) for the use by their end-users.

2.1 Architecture

As EGI moves towards an infrastructure based upon virtualised resources the dynamics between service development, service integration and service operation will change. Alongside this change the services that are a responsibility for the infrastructure provider and the services that are the responsibility of the user-communities or the end-user will also evolve. The diagram below identifies a number of layers in the technology stack and identifies the direct responsibility for each of those layers and the possible sources of technology that make up those layers. Clearly sustainable sources of technology are ones that build on mainstream software used across whole sectors (i.e. beyond EGI). Technology that is specific to a community or a science domain becomes the responsibility of that community, though EGI.eu will help facilitate that interaction.



2.1.1 Infrastructure

For EGI.eu the priority will be to provide a secure integrated e-infrastructure that enables an on-line European Research Area through the deployment of a diverse set of infrastructure services to meet the needs of different virtual research communities. As a community we need to ensure that the infrastructure offered by each resource centre through the NGI can be federated into a uniform and integrated set of core services. These services are oriented around the needs of the infrastructure and include:

- Federated identities through X.509 certificates for those who need to access infrastructure services to just manage the infrastructure
- Wide-area access control to infrastructure management services deployed at a resource centre (e.g. VOMS and Argus)
- Scalable distributed information service to enable the discovery resources and capabilities (e.g. BDII)
- EGI-wide messaging infrastructure (e.g., ActiveMQ network) inter-connecting infrastructure management services (e.g. Accounting, Monitoring, etc.)
- Accounting (e.g. capturing low-level usage records and feeding these into regional and aggregated infrastructure wide repositories)
- Monitoring (e.g. Nagios with probes for the infrastructure at a resource centre)
- Virtual Machine Image replication (e.g. GridFTP & Globus on-line)
- Virtual Machine Management environments (e.g. through integration of hypervisors and packages such as OpenNebula, OpenStack, etc.)
- Operation of these services (helpdesk, management, escalation, coordination, portals, etc.)

Collectively, funding will need to be secured to provide these core technologies and to operate them.

2.1.2 Platform

A platform needs to be deployed on top of the federated virtualised infrastructure either directly by the virtual research community or by experts acting on their behalf. This is the integration of through the resource infrastructure provider network (and their resource centres operating Infrastructure as a Service) that are coordinated by EGI.eu its network of to provide a federated Infrastructure as a Service. A virtual research community will need to use the provided infrastructure services (either directly or through experts) to deploy and operate a software environment (i.e. a Platform as a Service) needed by their end-users.

The platform provides application developers within a virtual research community with a 'programming model' (a distributed service infrastructure is included in this definition) to facilitate access through the platform to the available resources to meet their end-users' needs. Different platforms will be needed by different virtual research communities. This means end-users, or experts acting on their behalf either within their own community or elsewhere, will need to take responsibility to develop, integrate and deploy the software (in encapsulated in virtual machine appliances) onto the European-wide infrastructure. This means sourcing technology directly from within their own community, external technology providers, or other sources and assembling into virtual machine images that meet their specific needs. An existing virtual machine image could be used by the virtual research community as a basis for creating their own. Prior to the deployment of

a virtual machine image a resource centre may require the endorsement of the image by a recognised body. The resulting endorsed virtual machine images could be published and made available for others either directly, or to form the basis of their own virtual machine images.

The expertise for undertaking these different tasks already exists within the EGI community or could be developed so as to provide a number of service sources:

- Within the European production infrastructure (e.g. EGI.eu)
 - To promote re-use of software and virtual machine images between different communities a 'marketplace' (e.g. by building on the work of the current Application database and Software Repository) containing software providers and their technologies could be maintained. This will enable different communities to contribute and use each other's software and services. Social networks could be used to assess individual contributions.
 - From this software collection a set of virtual machine images could be generated and deployed to provide a basic environment for virtual research communities with basic common needs.
 - Endorsement of community prepared virtual machine images that comply with agreed best practice and policy can be provided as a service to virtual research communities.
 - The minimal policy relating to the endorsement of virtual machine images for use within the production infrastructure will be established.
 - This distributed environment needed to monitor the deployment of the virtual machine images and to ensure their availability and reliability to the consuming virtual research community will be established.
- Within individual research groups (e.g. WeNMR)
 - An individual research group can draw upon the software in the software marketplace to build their own virtual machine images (or use images contributed from the community) that can be customised with their own software services and applications.
 - If the expertise does not exist in house, training material will demonstrate best practice in building and operating virtual machines in a distributed environment. Some resource centres may require that the virtual machine images be endorsed by a recognised authority (e.g. the local NGI, EGI.eu, or recognised individual).
 - The correct operation of the deployed environment will be the responsibility of the research group. A dashboard framework may be provided centrally to support this activity.
- Within virtual research communities (e.g. WLCG@CERN)
 - A virtual research community will generally have greater contributed resources to draw on to develop their own community specific software and to provide an environment that meets their specific needs. Externally maintained software may form part of this environment.
 - The virtual research community will generally have the expertise to build their own virtual machines and have sufficient expertise to endorse their own virtual machine images through recognised trusted individuals.

- Dedicated operations staff within the community will ensure (on behalf of the community) that their domain specific services are operating effectively. A dashboard framework may be provided centrally to support this activity.
- Within national grid infrastructures (e.g. IGI, UK NGS)
 - An NGI may on behalf of its own users develop and deploy its own software environment.
 - The environment may be deployed directly onto the NGI-owned local hardware or encapsulated within virtual machine images on the NGI-owned resources. Endorsement of these virtual machine images may be required depending on the policy of the resource centres within the NGI. This endorsement service by the NGI could be offered to research groups or virtual research communities – both those within the NGI and beyond.
 - NGI operations staff will ensure that this environment is available for its own users. A dashboard framework may be provided centrally to support this activity.

A number of platforms, based around software components emerging from the DCI, related projects, main stream software projects or commercial providers can be envisaged to meet the needs of different virtual research communities. Some of these are illustrated below:

| Platform | DCI Software Components | Related Software Components |
|-------------------|---|-----------------------------|
| MPI | UNICORE GridFTP | QosCosGrid |
| Intensive HTC | CREAM, FTS, WMS, BDII, Argus | GANGA, DIANE |
| Low Volume HTC | CREAM, GridFTP GridWay | |
| Pilot HTC | CREAM, FTS, WMS, BDII, Argus, | Pilot Frameworks |
| VENUS-C | VENUS-C VM images ready for direct deployment into the infrastructure | |
| Basic Environment | CE, SE & WN | |

2.1.3 Software

On top of the provided platform environment end-users may need further software to complete their research activities. This software may come in the form of services or in the form of applications run on the provided platform. The support of this software, beyond the provision of stable well-defined lower-level infrastructure services to support this work is seen as outside the scope and responsibility of EGI.

2.2 Sustainability

Different sustainability strategies are proposed for the maintenance and incremental development of the different areas of the EGI Technology Architecture – infrastructure, platform and software. The optimal route to sustaining new innovative functionality is through specific national or European funding mechanisms. These can be summarised as:

- Infrastructure: The maintenance and introduction of incremental functionality into the software used for the infrastructure will be funded by the EGI.eu participants.
 - Action Now (TP): Assess effort required for each component.
 - Action Now (TP): Assess cost for the technology provider to deliver this effort.
 - Action Now (EGI.eu & TP): Evaluate definitive current components that would comprise an EGI federated virtualised infrastructure.
 - Initial List: VOMS, BDII, Argus, Nagios & additional monitoring components, Apel & further accounting components, OpenNebula/StratusLab, ActiveMQ, existing Dashboard services.
- Platform: While each platform is customised to the particular needs of each virtual research community the software components used to build that environment maybe shared between platforms. As these components are used directly by end-users, responsibility for the maintenance and development of these components falls to the consuming virtual research community. EGI.eu will continue to facilitate the connection between virtual research communities and technology providers through the bi-annual forums and dedicated workshops for virtual research communities to present their requirements to the community’s technology providers. In exceptional circumstances where there is a software component common across many communities EGI.eu *may* consider its support.
 - Action Later (EGI.eu): Define how EGI.eu may help facilitate connections between technology providers and users.
 - Action Now (EGI.eu – UCB): How ready are the virtual research communities to take up this responsibility – both technically and financially?
- Software: Much of the software in this area is specific to a single virtual research community. Some user-communities have produced software that has been adopted by other communities and have or are beginning to build a support community around those software components that can sustain themselves (e.g. GANGA, DIANE).

In summary the sources of sustainability for the different technology areas:

| Area | Maintenance (including incremental minor new functionality) | Innovation (major new functionality) |
|-----------------------|---|---|
| Infrastructure | EGI.eu has contracts with key technology providers for just the required functionality funded by its participants | Projects (National, EC & EGI.eu funds) |
| Platform | EGI.eu <i>may</i> have contracts with key technology providers for <i>some</i> commonly used components funded by its participants. Virtual Research Communities will need to develop sustainability strategies for their own software components for their platforms funded by national or European research bodies. Virtual Research Communities can work collaboratively with technology providers to contribute to commonly used services funded by national or European research bodies. | Projects (National & EC funds) |
| Software | Virtual Research Communities develop their own software funded by national or European research | Projects (National & EC funds) |

| | | |
|--|--|--|
| | bodies. Individual Research Groups develop their own software funded by national or European research bodies. | |
|--|--|--|