Support services for desktop grids and service grids by the EDGeS project

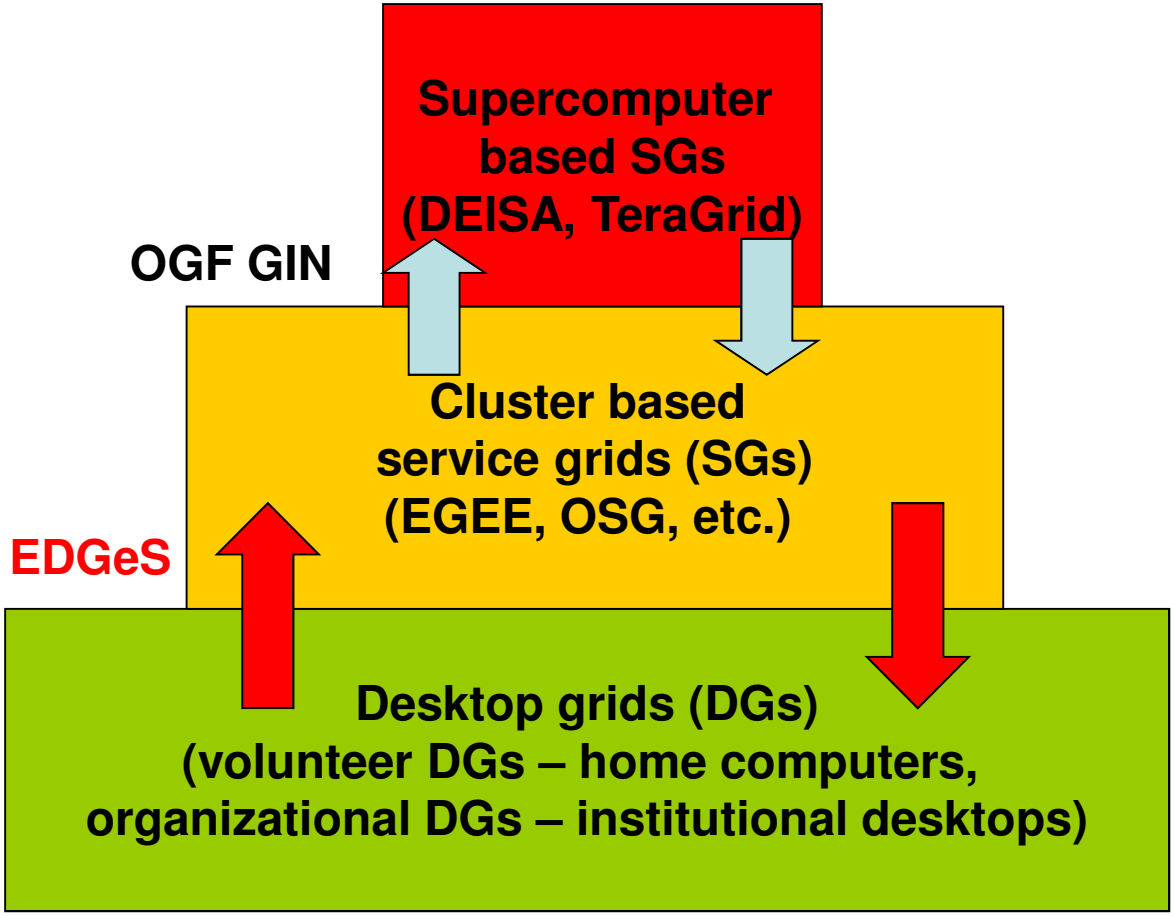
Tamas Kiss – University of Westminster

Leader of the EDGeS Application Support Service

E-mail: kisst@wmin.ac.uk

The EDGeS project receives Community research funding

The problem we address



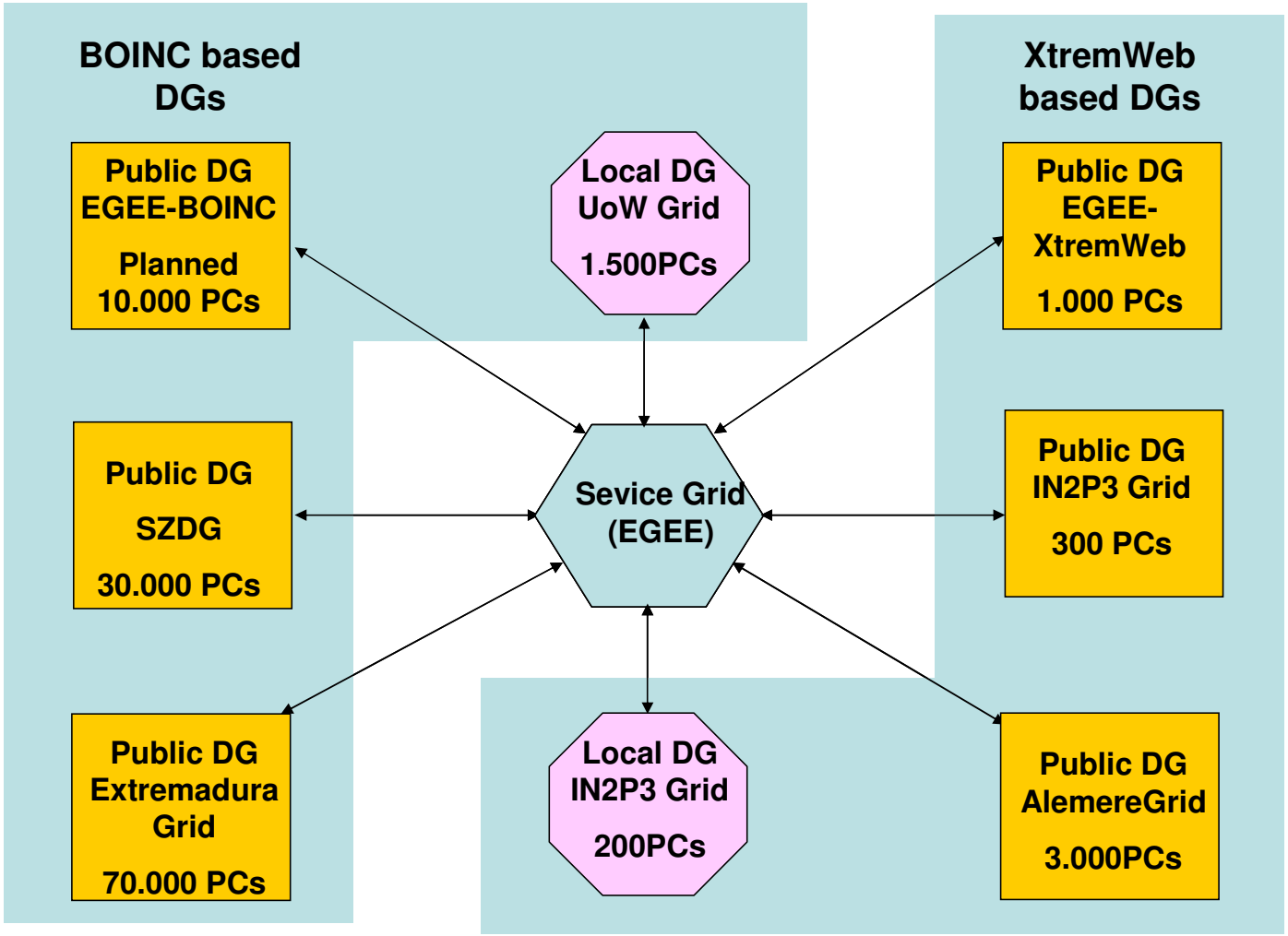
- Very expensive,
- small number of sites and large number of CPUs
- MPI appls
- Moderately expensive,
- moderate number of sites and CPUs
- any appls
- Inexpensive,
- very large number of CPUs (~100K – 1M)
- Bag of task appls

What is EDGeS?

EDGeS: Enabling Desktop Grids for e-Science

1. **EDGeS is a 2-year FP7 project started on the 1st of January 2008**
   - Research on integrating (bridging) SGs and DGs
   - Creates infrastructure
   - Provides services
2. **EDGeS is an infrastructure**
   - EDGeS VO interconnected with various DGs
   - Application repository
3. **EDGeS is a service for**
   - Application porting and validation

EDGeS - Connecting SG and DG infrastructures



BOINC: Berkeley Open Infrastructure for Network Computing

XtremWeb: Open Source DG platform developed by INRIA and IN2P3

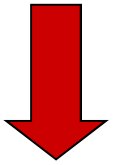
The EDGeS Application Support Service

- **Aims and objectives:**
  - facilitate the porting of applications to the EDGeS infrastructure by targeting
    - user communities already associated with EGEE or DGs
    - new user communities
- **Activities:**
  - develop a generic methodology for application porting
  - identify user communities that require the power of the EDGeS infrastructure
  - provide a service in order to aid the migration to and running of applications on the EDGeS infrastructure

Why do we need a methodology?

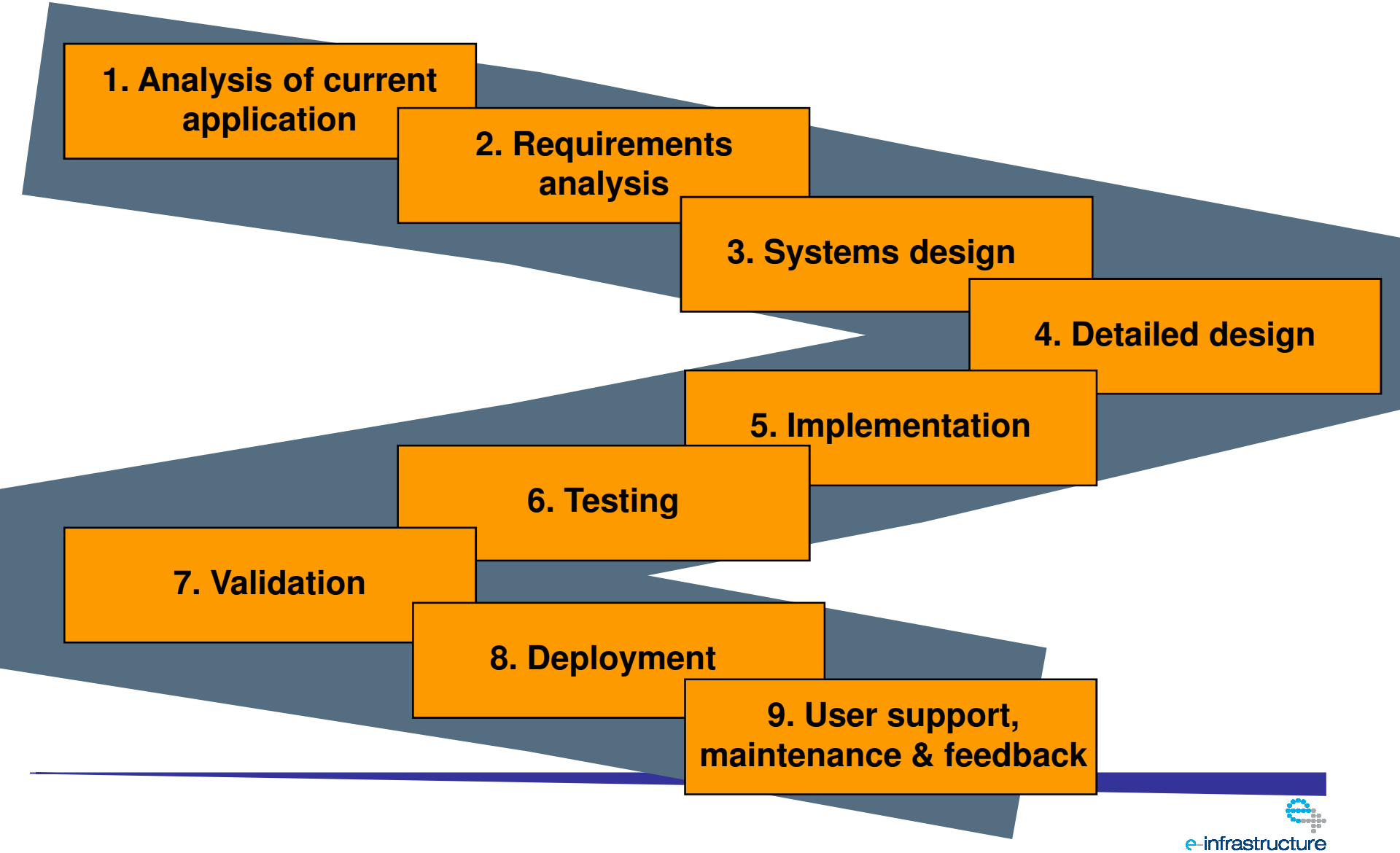
Motivations

- Grid application development is very often ad-hoc
  - Developers do not follow any methodology
  - Poorly documented systems
  - User expectations not fulfilled
- Systems design and development methodologies are too generic
  - Special focus is required when porting/developing an application to a SG/DG platform



EADM: EDGeS Application Development Methodology

EADM - at a glance



The EDGeS user community

- **Academic user community:**
  - 18 identified communities
    - BioInfoGrid, e-Health, Fusion, Computational Chemists, etc.
    - 4 new academic user partners
  - Establishing the EDGeS user forum
- **Industrial user community**
  - 5 identified companies
  - 3 new industrial user partners
  - Establishing the EDGeS industry forum

EDGeS Application Areas

community	number of applications	academic	non-academic
Bioscience and healthcare	**7**	**5**	**2**
engineering	**5**	**5**	
physics	**4**	**4**	
business	**1**	**1**	
audio & image processing	**4**	**3**	**1**
TOTAL	**21**	**18**	**3**

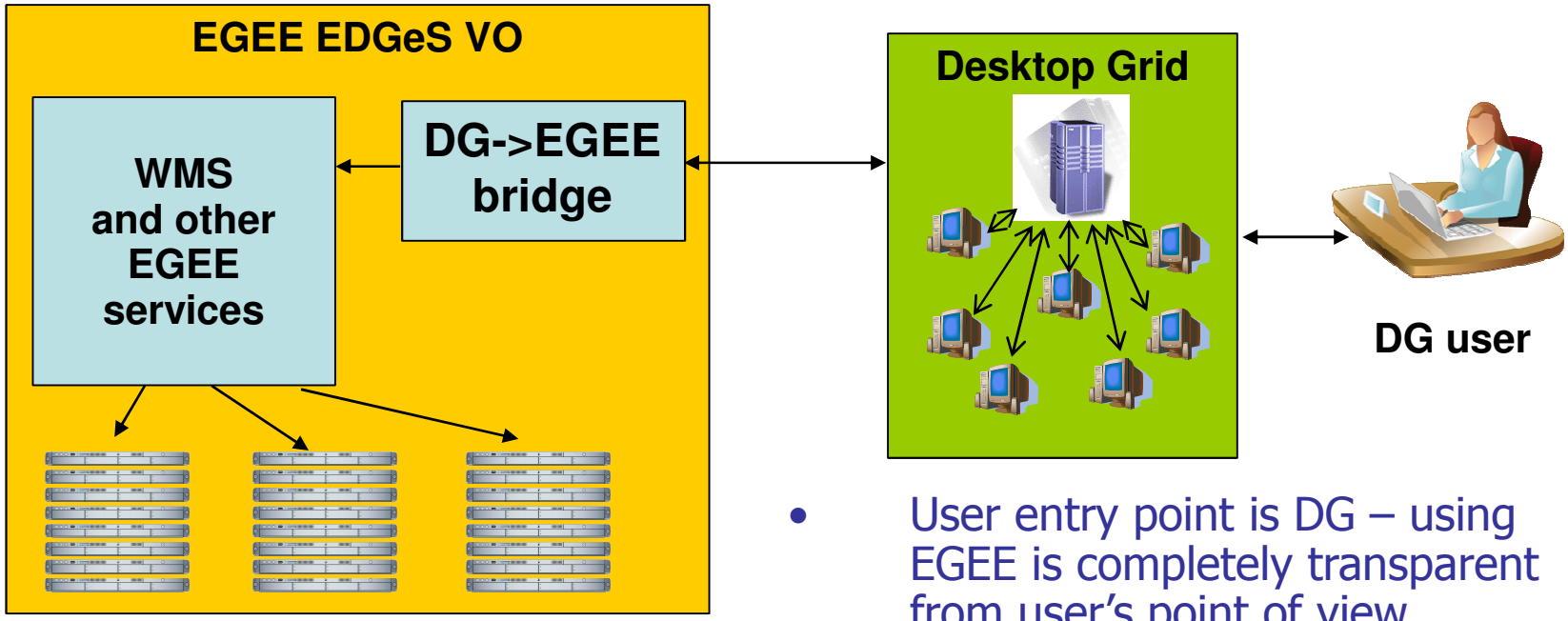
What is an EDGeS application?

EDGeS application

- **An EDGeS application is capable using both Service (EGEE) and Desktop Grid resources**
- **EDGeS application scenarios:**

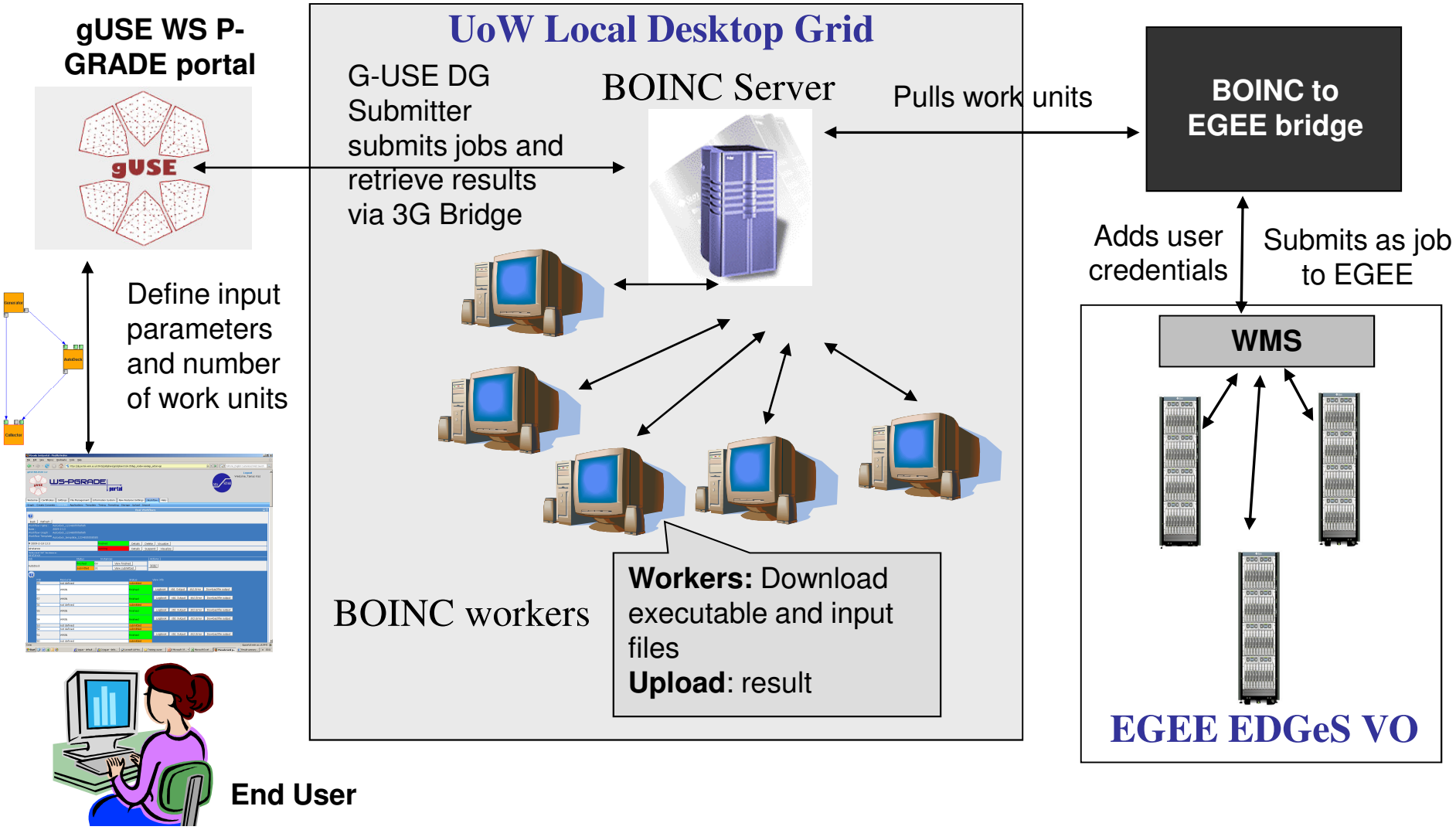
1. Application runs on DG and uses EGEE resources via the DG to EGEE bridge
2. Application submitted to EGEE and uses DG resources via the EGEE to DG bridge
3. Application uses both SG and DG resources via an external scheduling and job submission system (not through the EDGeS bridges)

Scenario 1 – DG to EGEE via bridge

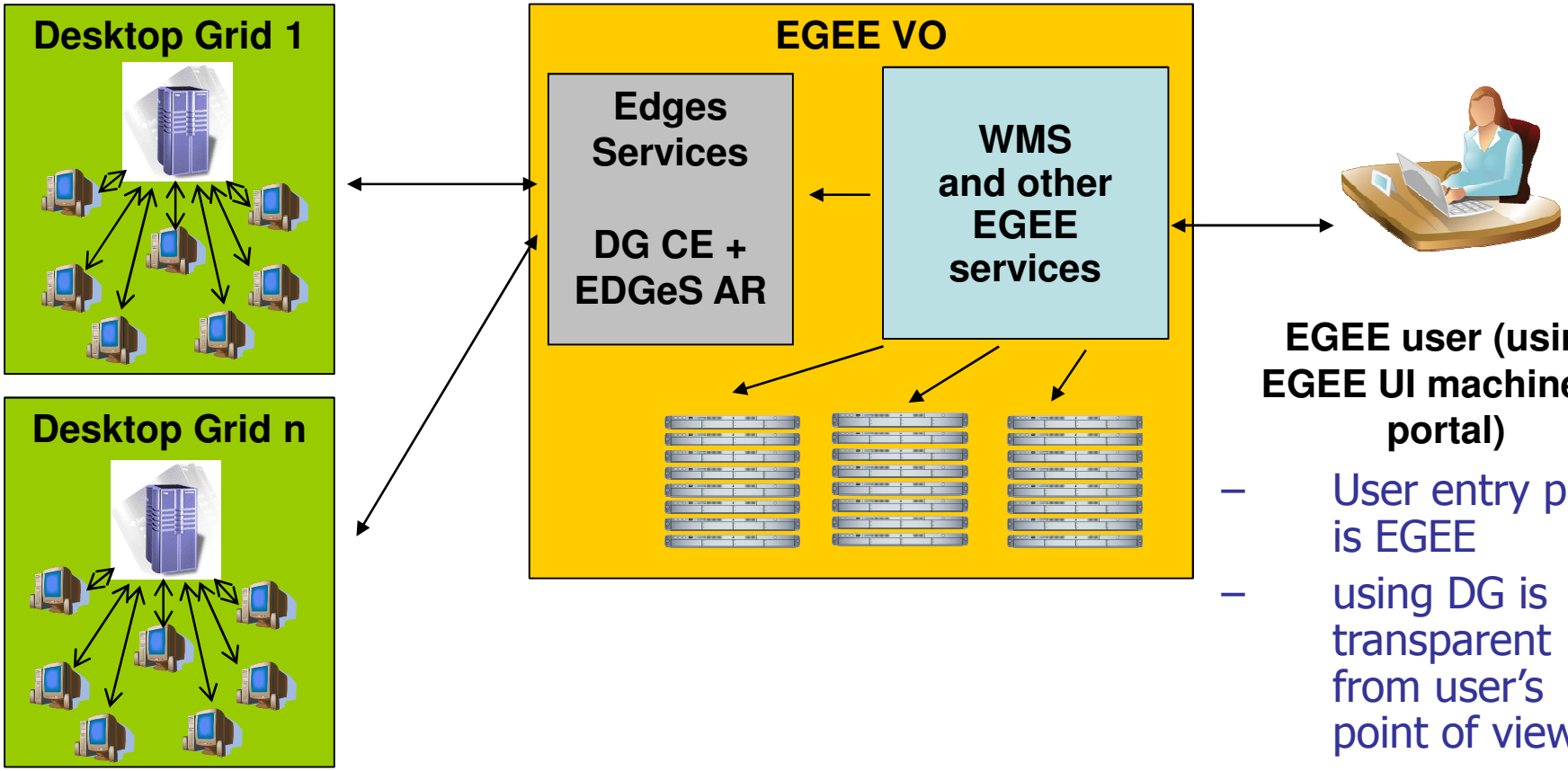


- User entry point is DG – using EGEE is completely transparent from user's point of view

Protein Molecule Simulation on a BOINC based Desktop Grid



Scenario 2 – EGEE to DG via bridge



EGEE user (using EGEE UI machine or portal)

- User entry point is EGEE
- using DG is transparent from user's point of view

Scenario 2 – EGEE to DG via bridge

Porting to EDGeS:

1. Develop **DG version** of application
2. Test **DG version** and write **test report**
   - Use the EDGeS Development DG
   - Use your own development DG
   - Test report template downloadable from EDGeS website
3. Have your application **validated**
   - Validation is done by EDGeS Validation Team
   - Outcome: validation document
   - Validation includes tests via the EDGeS EGEE->DG test bridge
4. Publish your application in **EDGeS Application Repository**
   - EDGeS Services (AR and EDGeS CE) should be deployed by your EGEE VO as pre-requisite
   - DG Admins can download your application from EDGeS AR if they are ready to support it

Which EGEE applications are suitable for EDGeS?

- **EDGeS applications should run on both EGEE and DGs**
  - EGEE supports a much wider scale of applications than DGs
  - We should assure that the application runs on DGs

- **Requirements towards an EGEE application to be executable on DGs:**
  - **Parallelization:**
    - Only **master/worker** or **parameter sweep** parallelisation
      - **No MPI** or internal communication between worker nodes
      - Nodes can only use the results of other nodes **through the server**
  - **Data handling:**
    - small or medium-sized (max. 100 MB per worker) inputs and outputs
    - No shared data storage
    - No confidential data (sent down to potentially un-trusted worker)

Application examples – Scenario 2

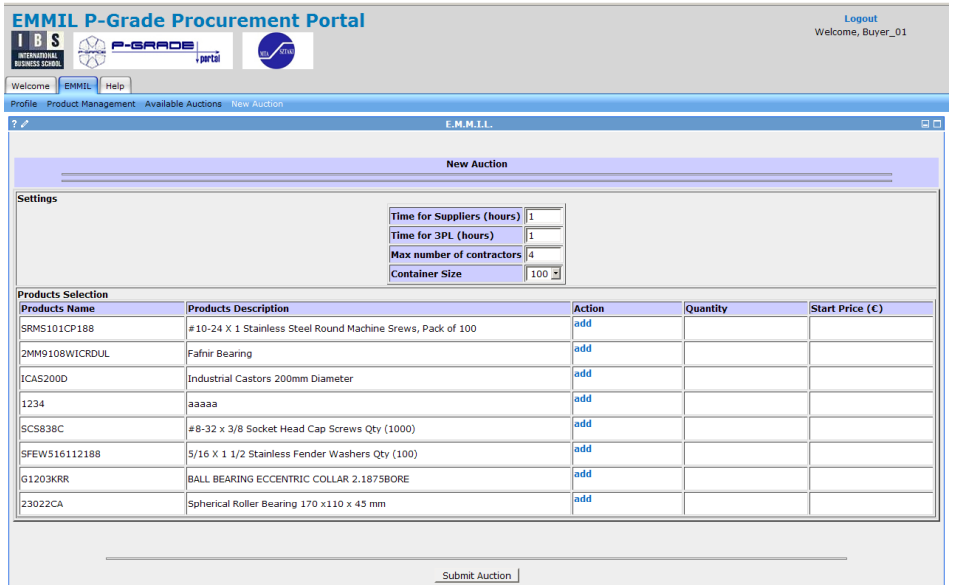
EDGeS application examples:

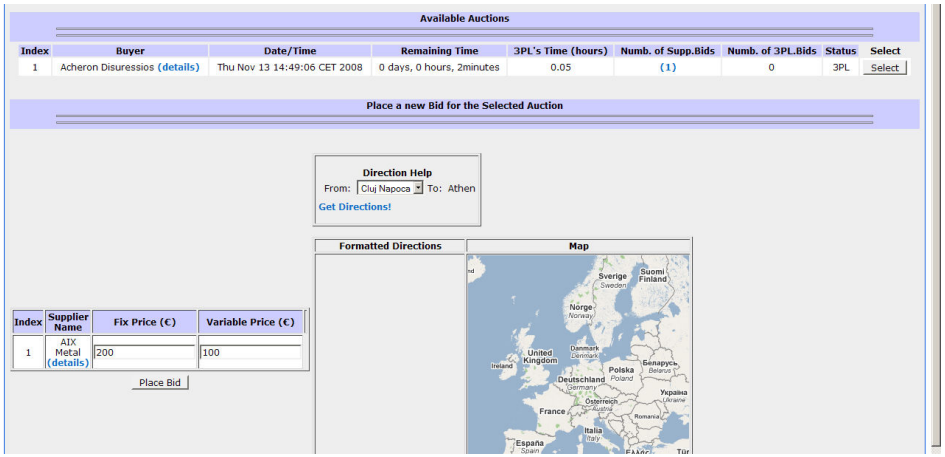
- **EMMIL E-marketplace application** – run on SEE-Grid already
- **VisIVO -** Visualisation Interface to the Virtual Observatory – EGEE application by Astrophysical Observatory of Catania, currently ported to DG by UoW
- **3D video rendering –** application ported to EGEE (however, DG version already existed)

Scenario 2 – an example

EMMIL – E-Marketplace Model Integrated with Logistics

- three sided negotiation between buyers, sellers and third party logistics providers
- Aims to optimise the total coast
- Ported to EGEE as a parameter sweep application in the SEE Grid project
- Large matrix solving – matrix can be separated and solution parallelised
- Ported to EDGeS using GenWrapper



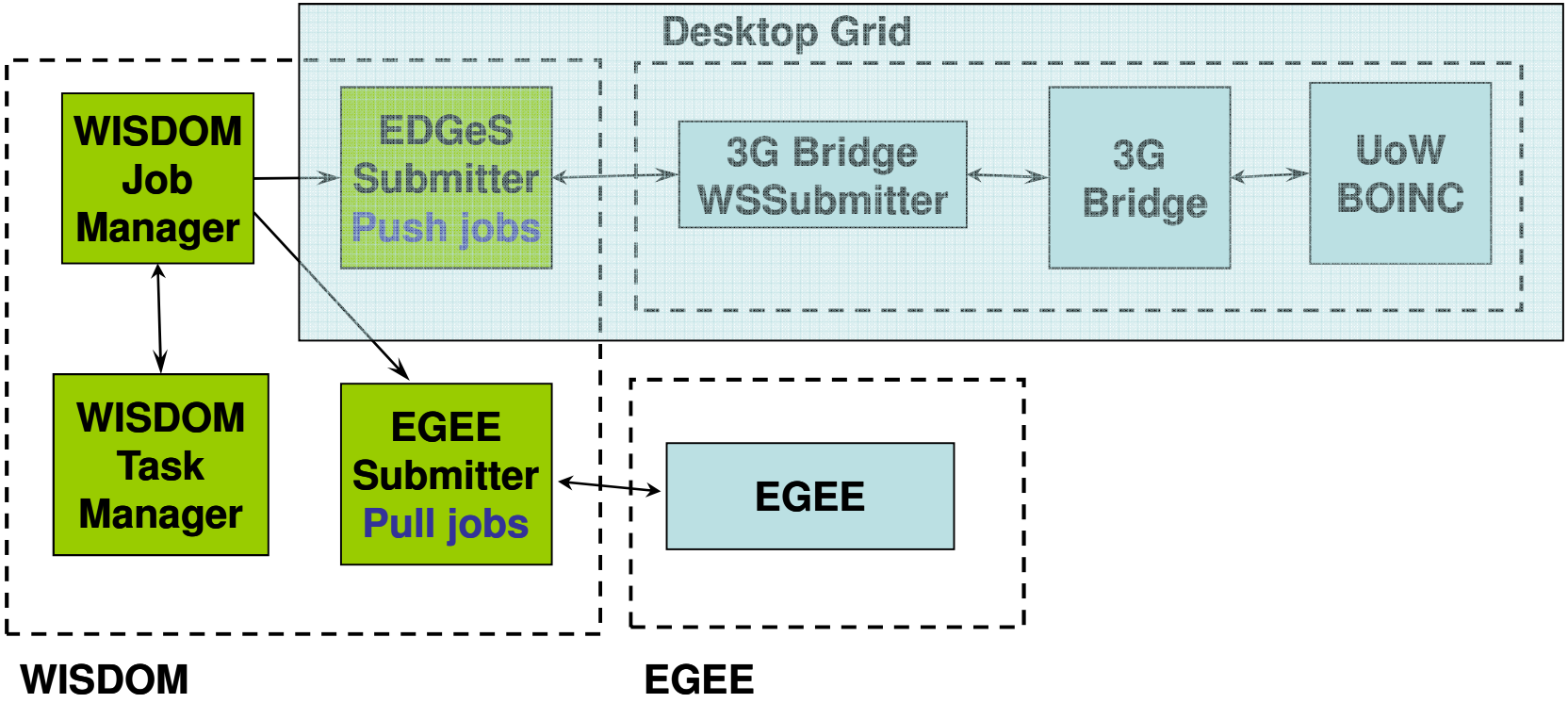


Scenario 3 – SG/DG resources but not through EDGeS bridges

- **Using external job submission and scheduling system to submit jobs to both EGEE and DG resources**
  - **G-USE portal supports this scenario by default**
    - E.g.:
      - CancerGrid application uses EGEE and DG resources via the g-USE portal (but not through the bridges - see presentation on Friday afternoon)
  - **Could be justified with specific user requirements**
    - E.g.:
      - WISDOM project uses only pull jobs on EGEE that are unsuitable to be bridged to DG
  - **Both solutions use EDGeS technology:** 3GBridge

Scenario 3 – e.g. DG submitter for WISDOM

- **WISDOM:** Meta middleware to submit pull (pilot) jobs to EGEE
- **The DG submitter:**
  - Submit push (direct) jobs to the DG when EGEE resources are overloaded



Thank you for your attention …

For more information please visit the EDGeS Website:

Join the EDGeS User & Industry Forum here!

http://www.edges-grid.eu/

Further EDGeS related sessions during EGEE'09:

EDGeS: Extending resources of EGEE VOs with desktop grid resources

Wednesday 16.30 and Thursday 16.30 – Demo Booth

The usage of EDGeS services for EGEE users

Thursday, 11.00 – 13.00 and 14.30 – 16.30, Saria