



Porting applications for a combined EGEE/Desktop Grid platform in the framework of the EDGeS infrastructure

Tamas Kiss - University of Westminster
Leader of EDGeS SA3 – Application Support Service
kisst@wmin.ac.uk



The EDGeS project receives Community research funding





The challenge

Service Grids (e.g. EGEE):

- Guaranteed service based on clusters **but**
- Relatively small number of processors (1K -50K)
- Many users but “misuse” of resources
 - Many applications are parameter sweep applications that better fit to desktop grids
 - Takes away the cluster resources from MPI type applications

Desktop (volunteer) Grids:

- No guaranteed service **but**
- Large number of processors (100K – 1M)
- Few users: 1-3 projects per desktop grid



What is EDGeS?

Enabling Desktop Grids for e-Science

EDGeS:

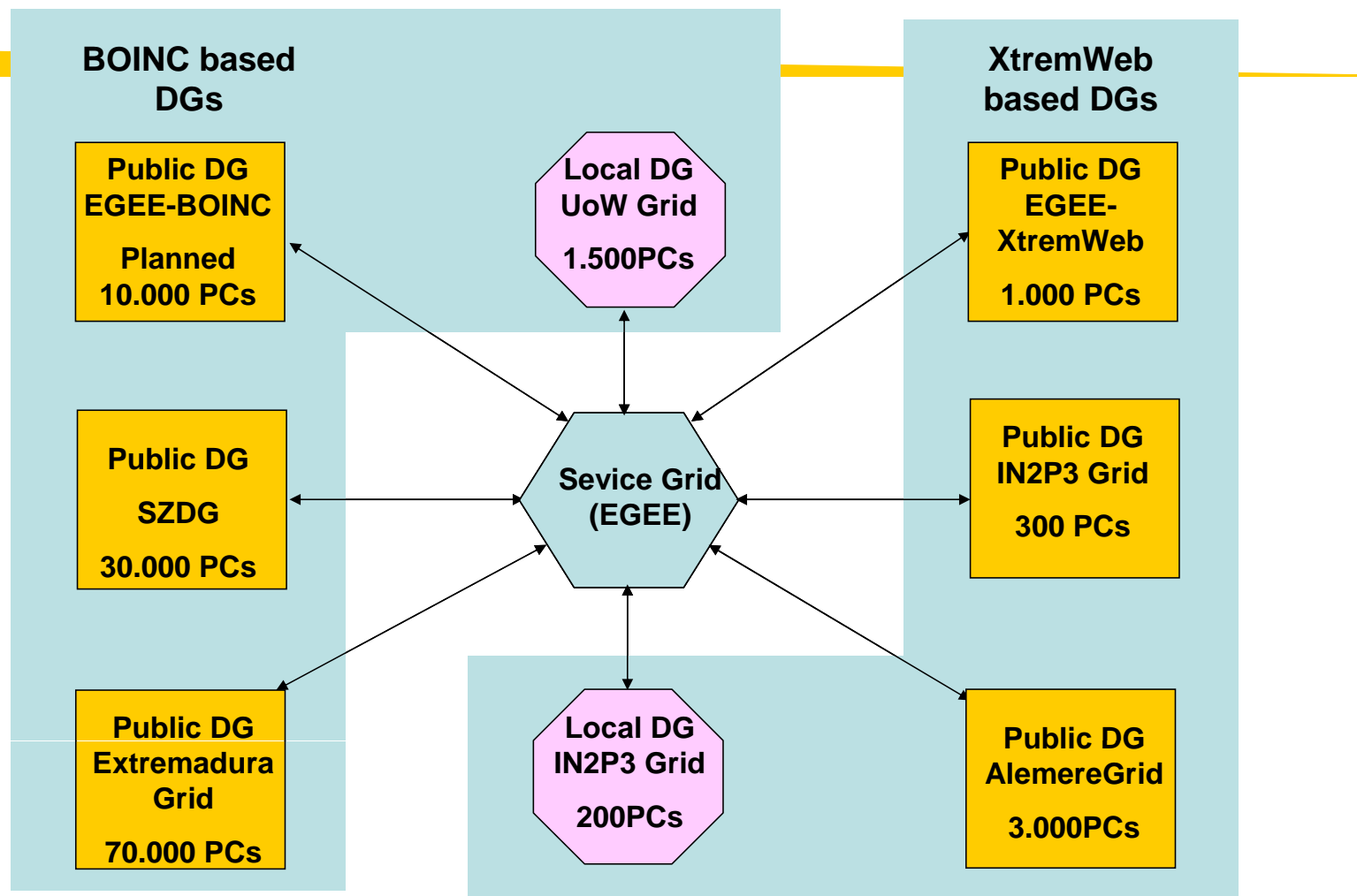
- FP7 project started on the 1st January 2008

Goals of the project:

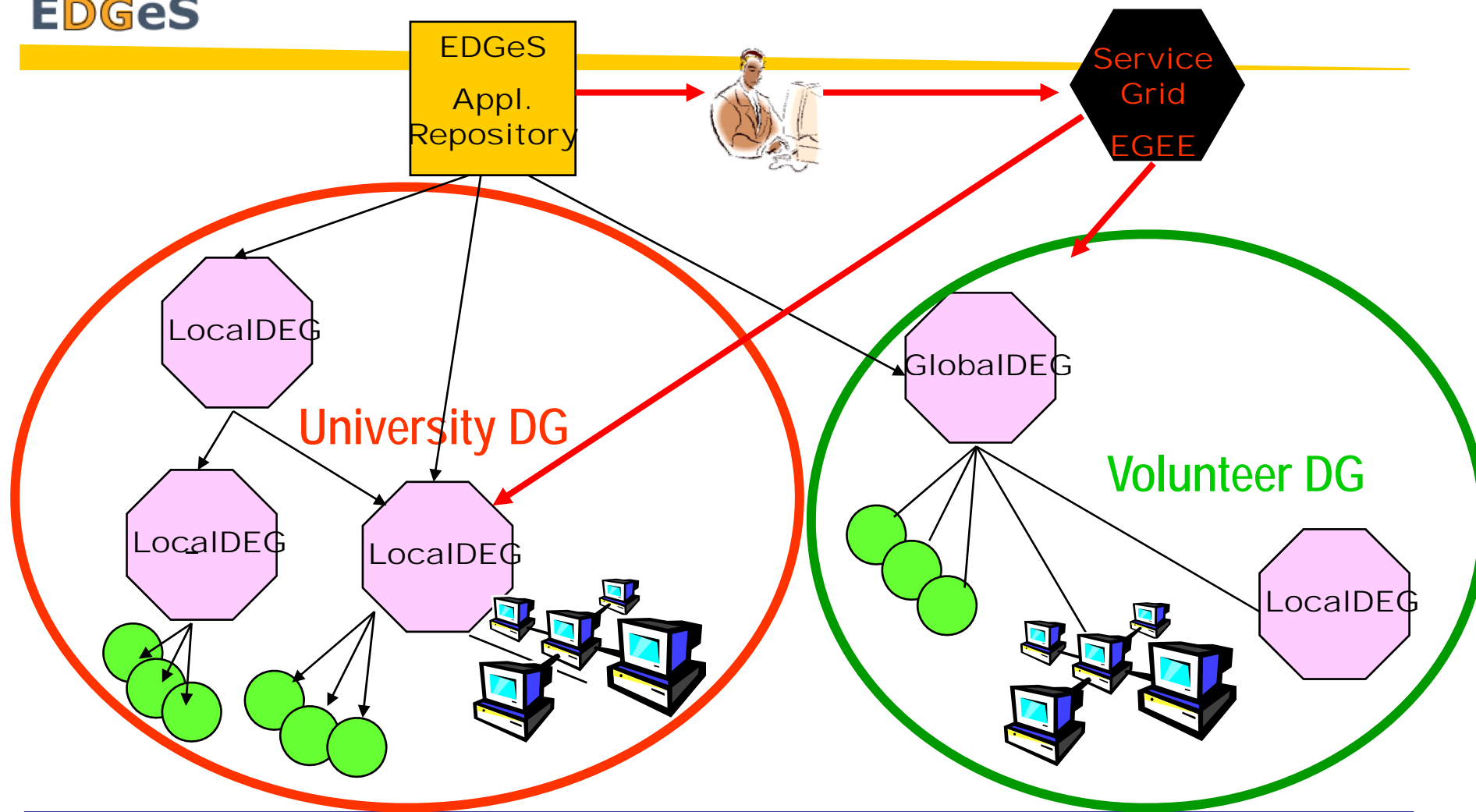
- Integrate Service Grids (EGEE) and Desktop Grids (BOINC and XtremWeb) to attract new scientific communities that needs very large number of computing resources (100K-1M processors)
- Extend the group of applications to be run on Desktop Grids
- Enable the seamless usage of the integrated SG-DG infrastructure, i.e. the user does not need to be aware if the application is executed in SG or DG
- Provide a Grid application development environment
- Offer application repository and broker for the seamless execution of applications in the SG-DG system
- **Provide application support to port applications to the EDGeS infrastructure**



The EDGeS infrastructure

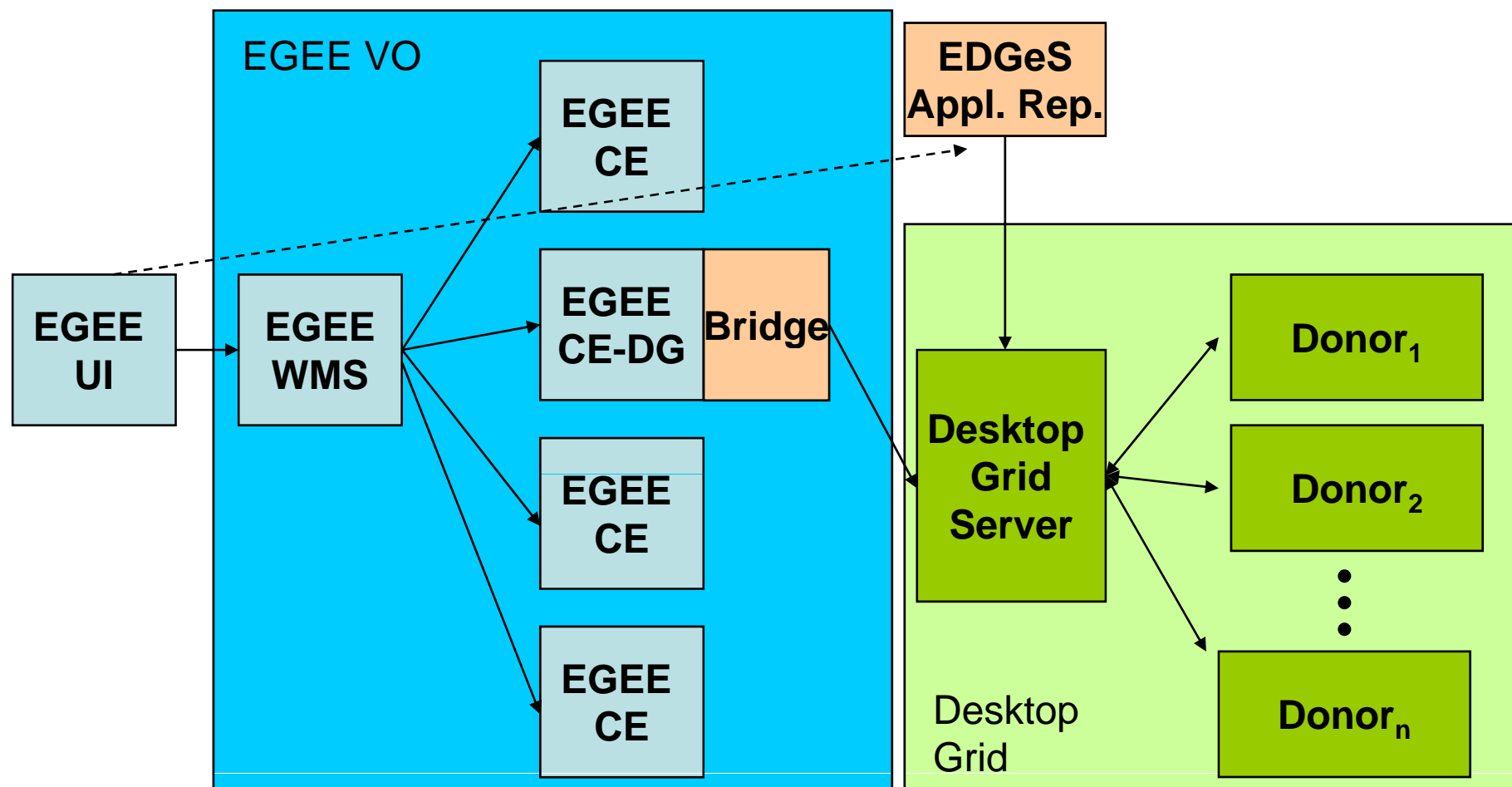


A typical user scenario



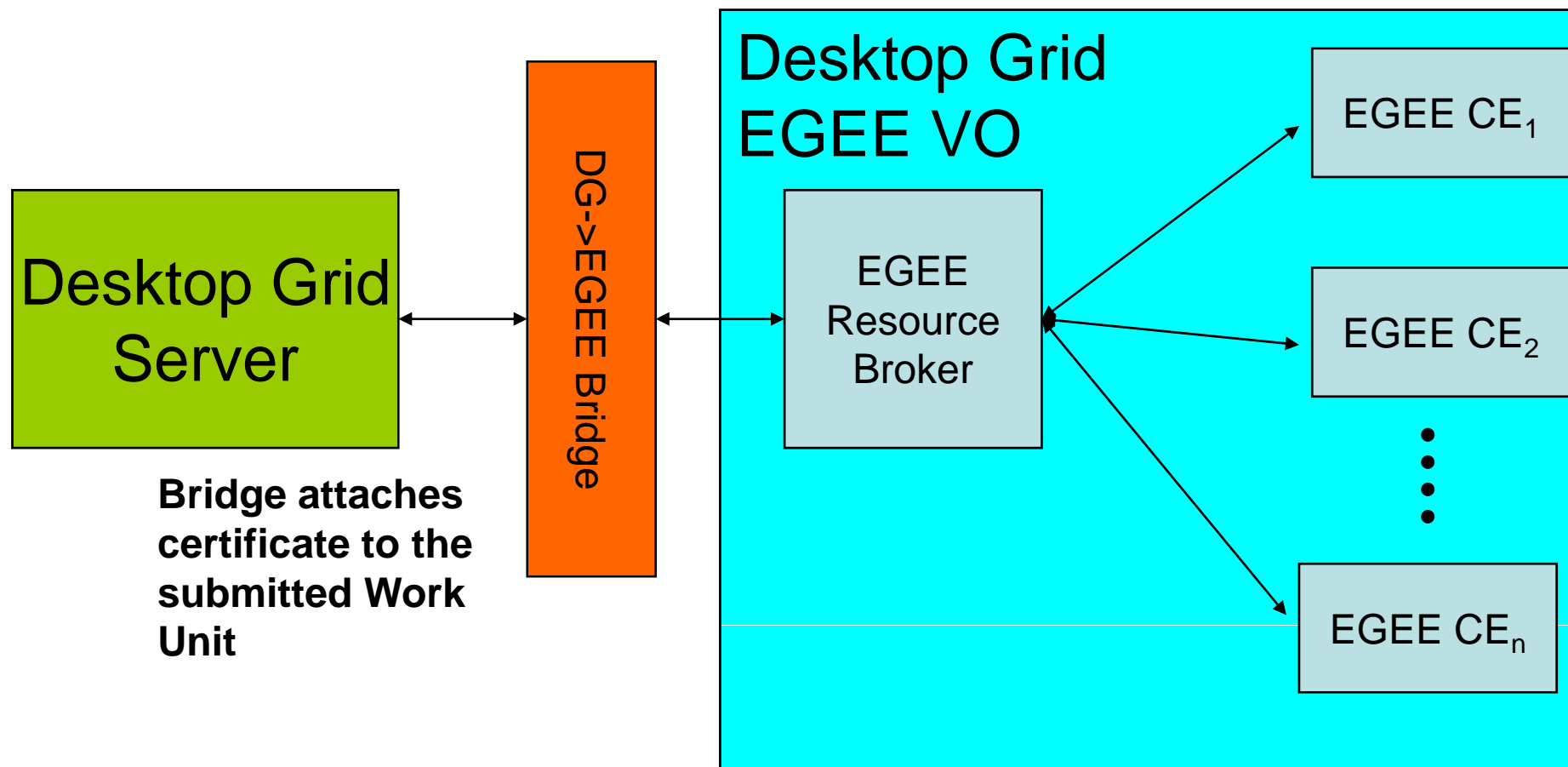


High level architecture of the EGEE to DG Bridge





High level architecture of the DG to EGEE Bridge





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:**
 - Identify user communities that require the power of the EDGeS infrastructure
 - develop a generic methodology for application porting
 - provide a service in order to aid the migration and running of their applications on the EDGeS infrastructure



EADM

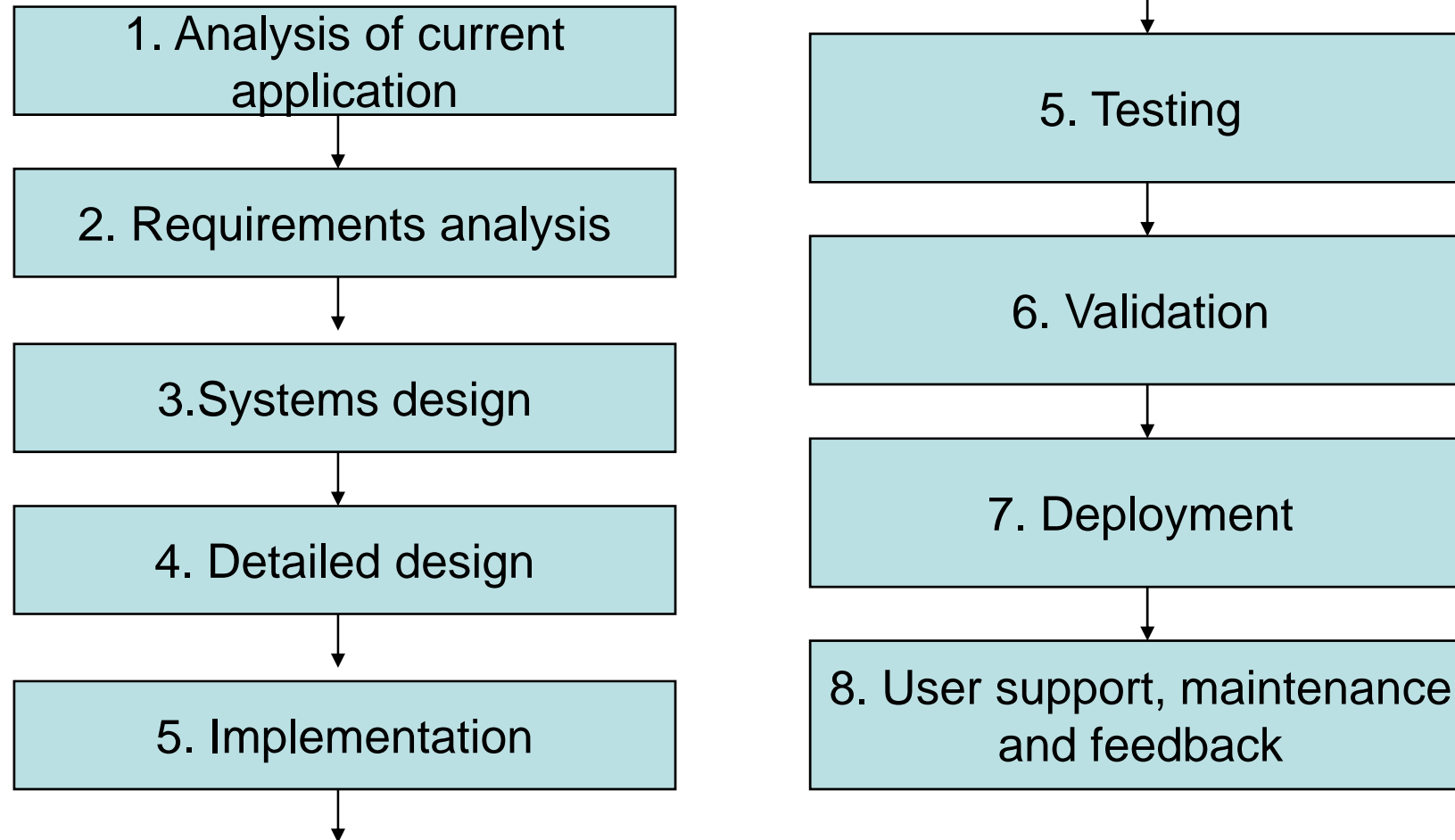
EDGeS Application Development Methodology

- **Aims and objectives:**
 - a generic methodology that addresses the problem of application porting and defines how the recommended software tools, developed by EDGeS, can aid this process.
- **EADM – an iterative approach**
 - EADM identifies well defined stages that have a suggested logical order. However, the overall process is in most cases is non-linear allowing revisiting and revising the results of previous phases at any point.



EADM Stages

Earlier stages can be revisited at any time





EADM

EDGeS Application Development Methodology

- 1. Analysis of current application:** describe the currently existing application in detail (problem domain, computing platform, parallelism, data access, functionalities etc.)
- 2. Requirements analysis:** requirements towards the ported application concerning efficiency of execution and data access are analysed from a user perspective
- 3. Systems design:** answer the major questions of systems design principles concerning the ported application (what will be the target platform for execution and how it will be accessed, what level/type of parallelism will be utilised, what data access mechanisms will be applied, etc.)



EADM

EDGeS Application Development Methodology

4. **Detailed design:** Technical Design Specification - forms the basis of the implementation and details all the necessary technologies and solutions that will be utilised
5. **Implementation:** implement the system based on the detailed design specification
6. **Testing:** both functional and performance testing
7. **Validation:** only validated applications can run on DG resources
8. **Deployment:** application published in the EDGeS application repository – select and run application with custom parameters
9. **User support, maintenance and feedback**



Application Examples

Video Stream Analysis in a Grid Environment (ViSAGE)

Correlation Systems Limited - Israel

- **Video Analysis:** the use of advanced algorithms to process video data.
- **ViSAGE:** processing video streams using a Grid of computational nodes
- The demand for computation power is very difficult for prediction
 - the ratio between average demand and peak demand is very high.
 - Not possible (economically) to have computation power for peak period
 - Cost must be competitive

Source: Erel Rosemberg – Correlation Systems



Visage in EDGES mode

Visage processes Image pairs over the BOINC grid and paints movement in yellow.

Insert video source:

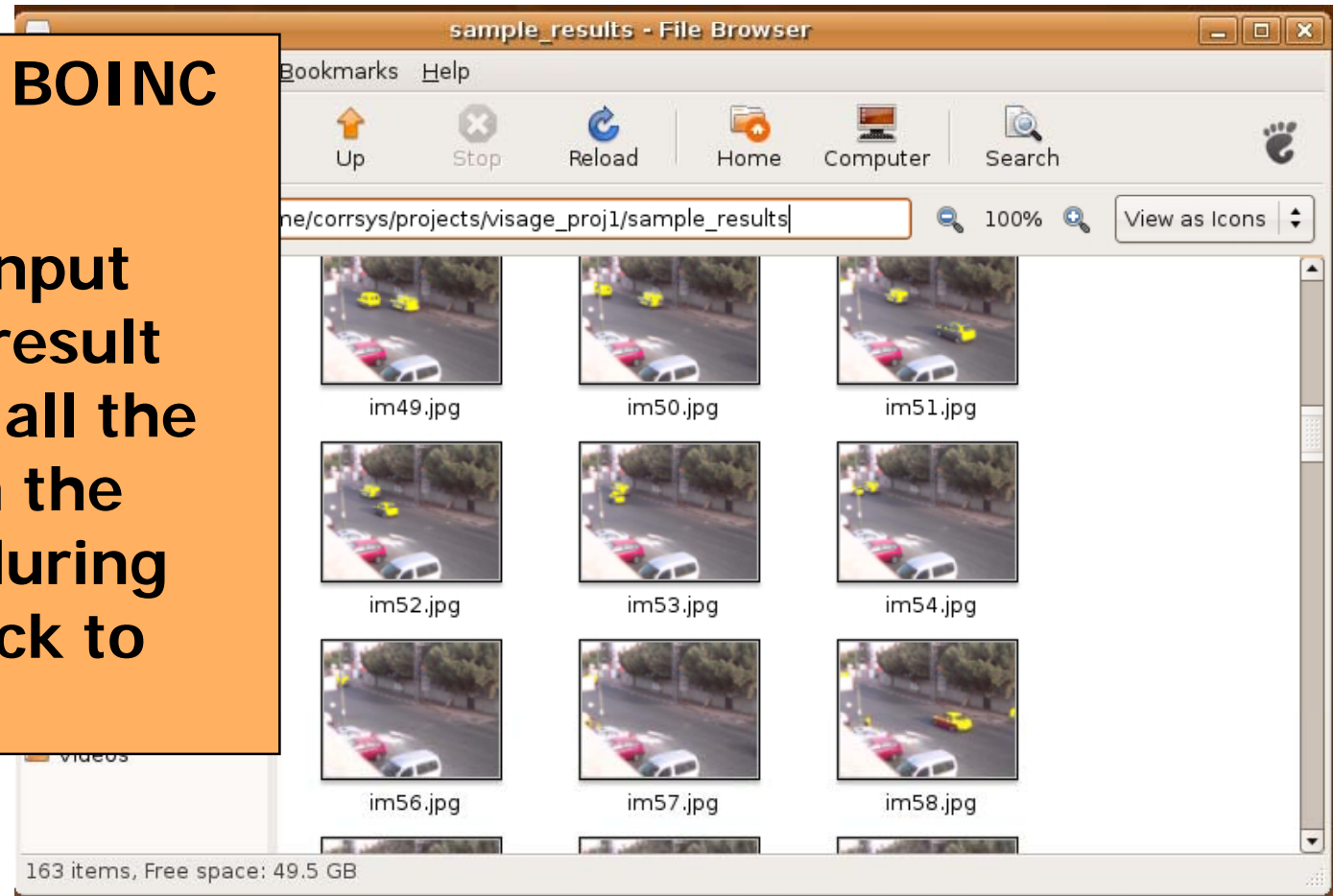
Insert Url of Server:

Video options: forward, backward, pause, play..

Result Images

Results from BOINC clients:

The project input folder holds result images from all the computers in the BOINC grid during transition back to Visage.





Application Examples

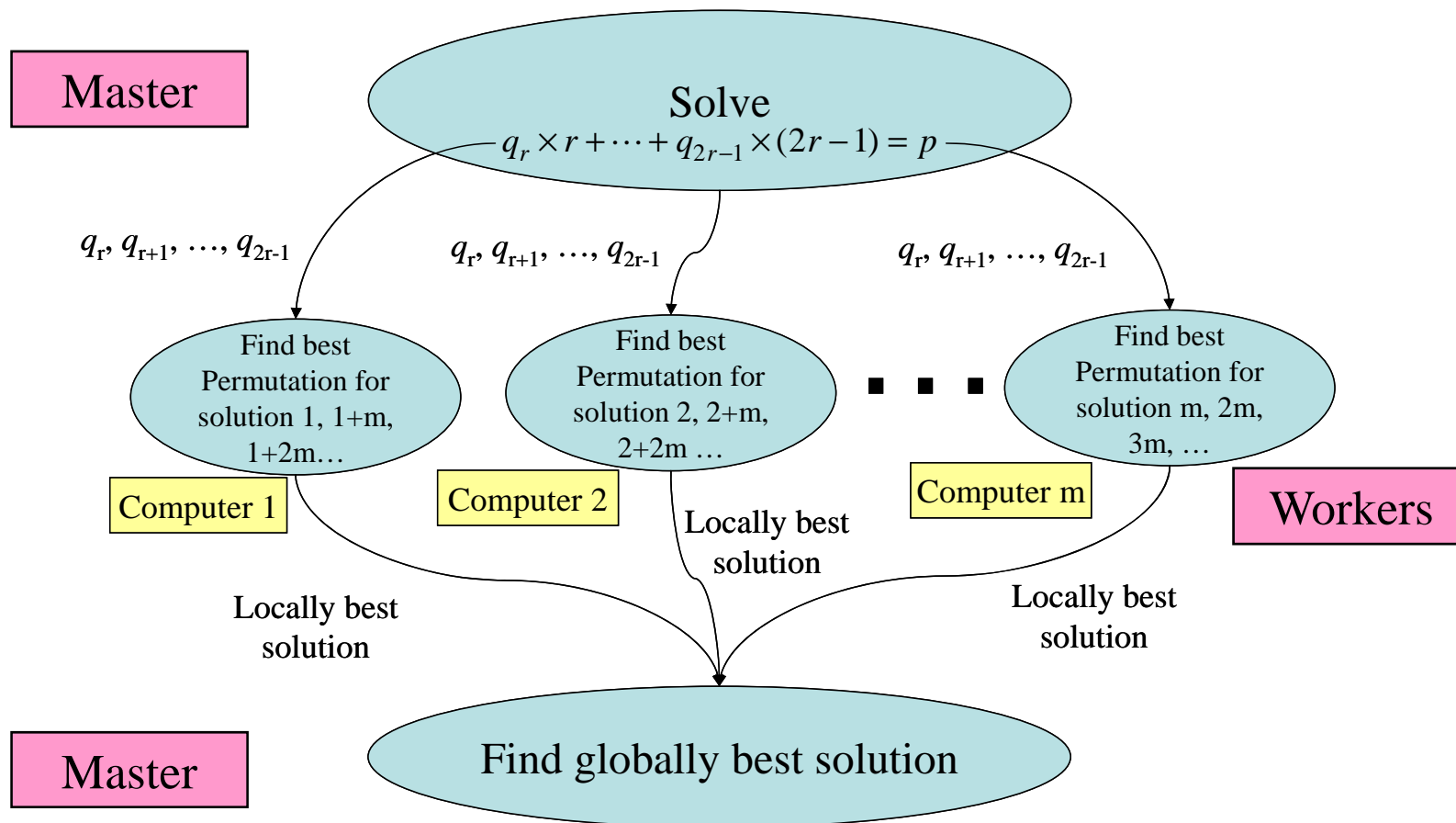
Digital Alias-free Signal Processing (DASP)

University of Westminster

- **The challenge:**
 - designing periodic non-uniform sampling sequences for digital alias free signal processing is rather computationally expensive
 - A linear equation has to be solved and a large number of solutions (in the magnitude of 10^{10} has to be compared).
 - Solutions are independent from each other, suitable for DG style of parallelisation
 - Both SG and DG based solutions are developed – ideal as pilot project to test principles

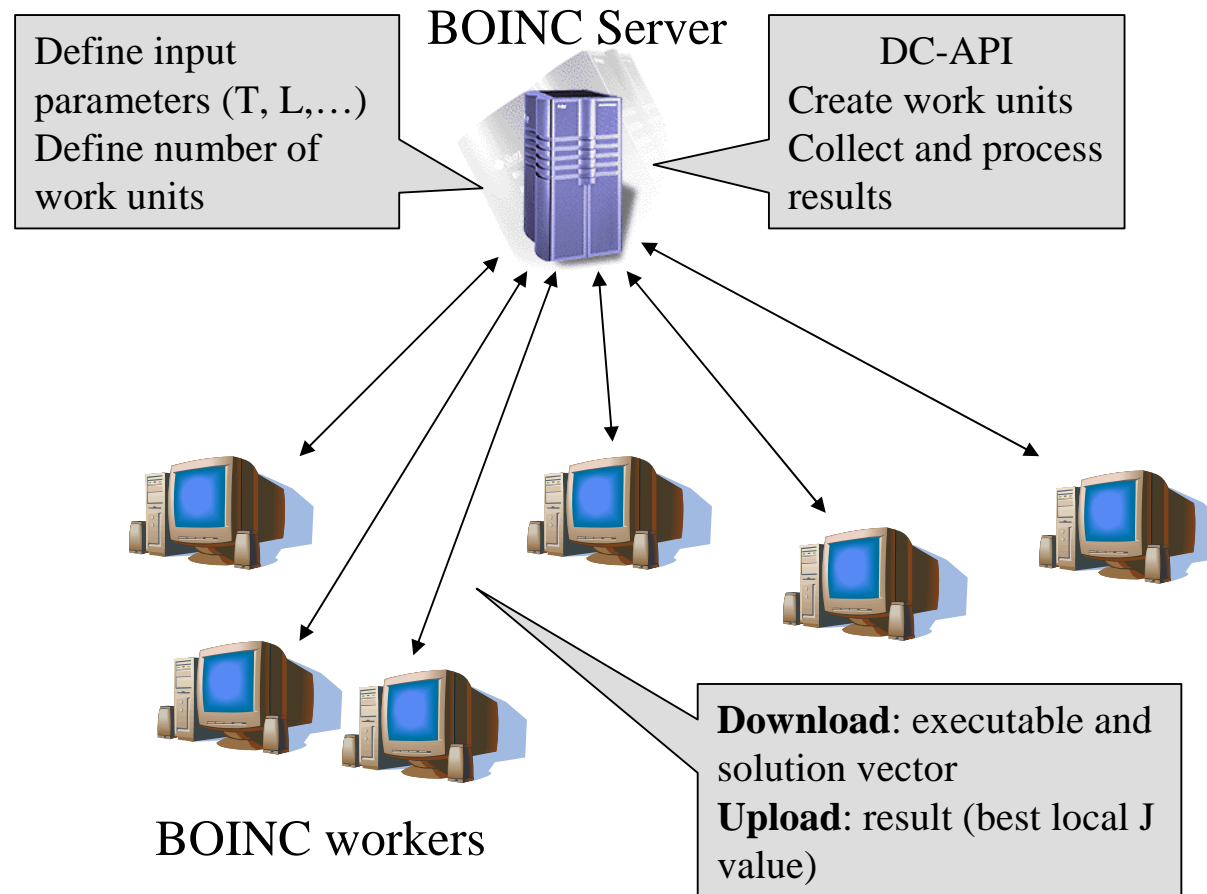
Application Examples

DASP – the parallelisation

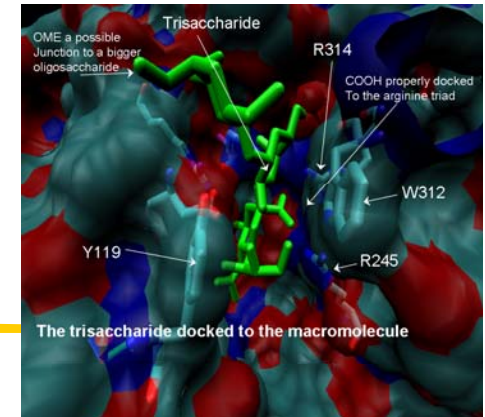


Application examples

DASP – the Desktop Grid Solution



Applications ported using EADM- Protein Molecule Simulation using Autodock

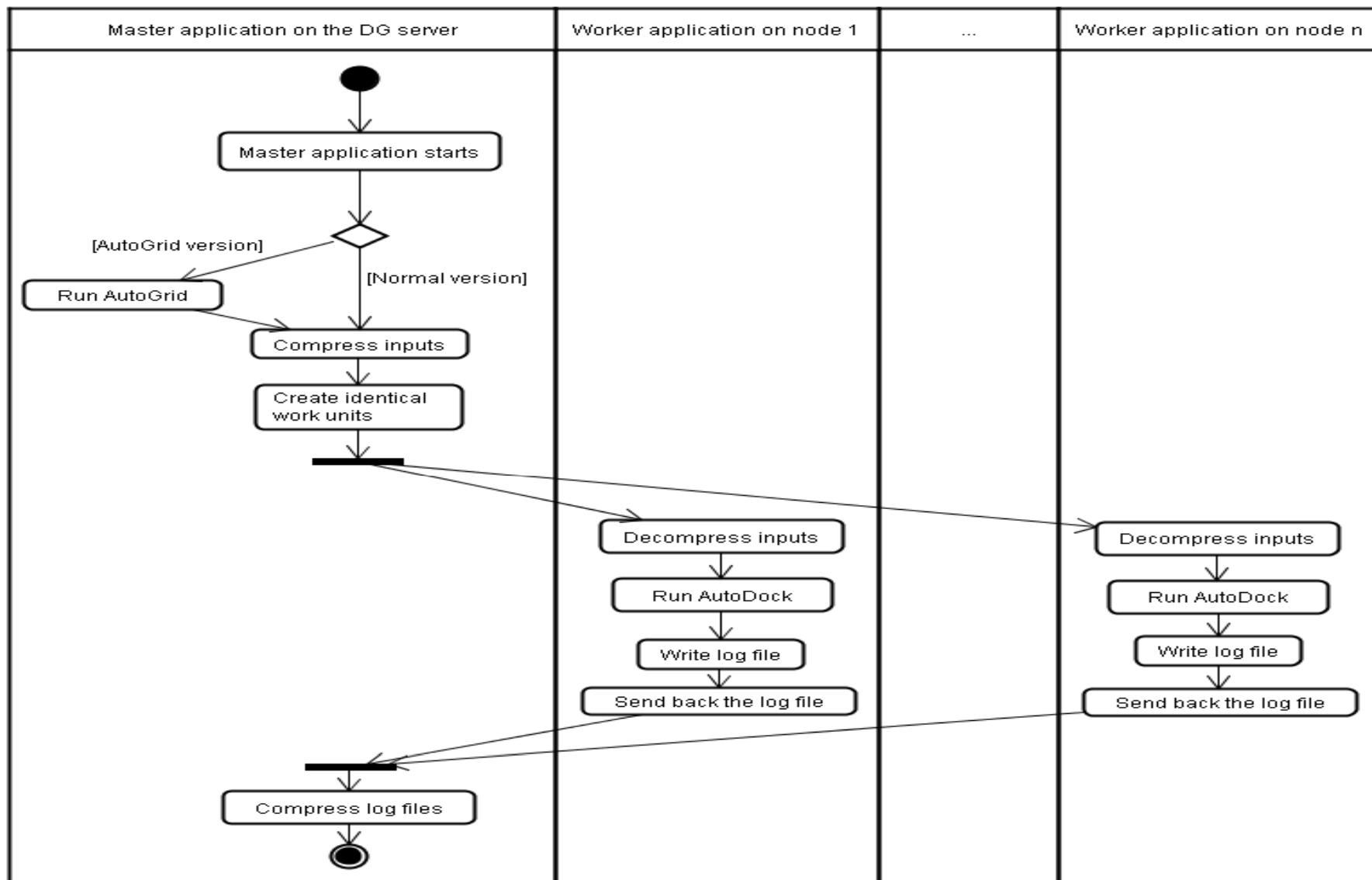


- **The challenge:**
 - *AutoDock*: a suite of automated docking tools; designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure
 - Undertake the simulations on a larger scale and in a much shorter timeframe
 - **Application components:**
 - *AutoDock* performs the docking of the ligand to a set of grids describing the target protein
 - *AutoGrid* pre-calculates these grids
 - To be run on the Westminster local desktop grid and on EGEE – being capable to utilise bridges to both directions
- **Users:**
 - Department of Molecular & Applied Biosciences at the UoW



Applications ported using EADM

Autodock – the parallelisation





Applications ported using EADM

Patient Readmission Application

University of Westminster

- **The challenge:**
 - A statistical model is developed in R, where individual hospitals' propensity for first readmission, second readmission, third (and so on) are considered to be measures of performance.
 - The national (England) hospital episodes statistics dataset comprises more than 5 million patient readmissions between Jan 1998 and Dec 2003.
 - Implementing a statistical model using complete population dataset could possibly take weeks to estimate the parameters.



Applications ported using EADM

Patient Readmission Application

- **The solution:**
 - To resolve the problem, we extract 1000 random samples from original data, where each random sample is likely to lead to differing hospital performance measures.
 - Each client gets 1000 random samples, and does the statistical computation separately.
 - The master application summarizes the received result files.



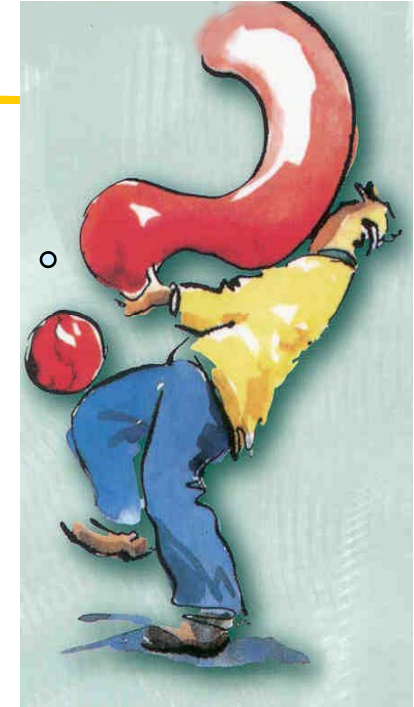
Some further applications currently being ported to EDGeS

- Ultrasound Computer Tomography - **Forschungszentrum Karlsruhe**
- EMMIL – E-marketplace optimization - **SZTAKI**
- Anti-Cancer Drug Research (CancerGrid) - **SZTAKI**
- Integrator of Stochastic Differential Equations in Plasmas - **BIFI**
- Distributed Audio Retrieval - **Cardiff University**
- Cellular Automata based Laser Dynamics - **University of Sevilla**
- Radio Network Design – **University of Extramadura**
- An X-ray diffraction spectrum analysis - **University of Extramadura**
- DNA Sequence Comparison and Pattern Discovery - **Erasmus Medical Center**
- PLINK - Analysis of genotype/phenotype data - **Atos Origin**
- 3D video rendering - **University of Westminster**



Thank you for your attention ...

Any
questions?



For more information please visit the EDGeS Website:

Join the EDGeS User & Industry Forum here!

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

Please contact us if you need support in porting your application to EDGeS!

Email: kisst@wmin.ac.uk