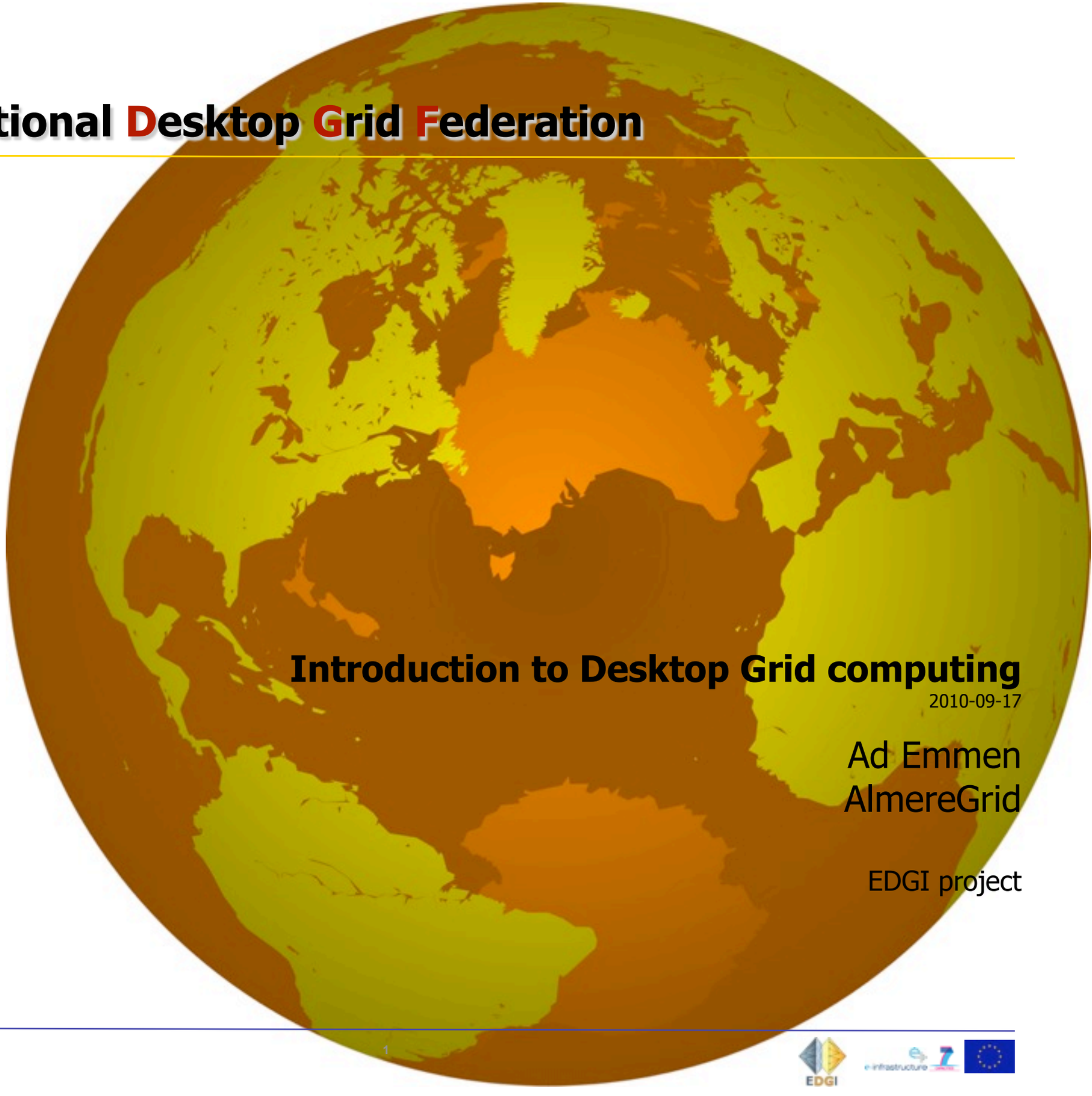
IDGF
International Desktop Grid Federation



Introduction to Desktop Grid computing

2010-09-17

Ad Emmen
AlmereGrid

EDGI project

EC Vice President Kroes

EC Vice President Kroes

- “On average, a desktop computer remains idle for around 60-85% of the time. Networks like EGI distribute computing tasks involving large amounts of data among the processing capacity of many thousands of separate desktop computers, putting their idle processor cycles to productive use. EGI-InSPIRE will give European researchers access to the aggregated processing power of 200,000 desk-top computers hosted by more than 300 centres around the world. The Commission is contributing €25 million over four years to the €73 million project. Other funding is provided from national sources such as National Grid Initiatives (NGI). “

EC Vice President Kroes

- "On average, a desktop computer remains i[illegible]ound 60-85% of the time. Networks like EG[illegible]uting tasks involving large amounts of [illegible] processing capacity of man[illegible]rate desktop computers, p[illegible]ssor cycles to productive use. F[illegible] European researcher[illegible]gated processing power of 200 0[illegible]rs hosted by more than 300 [illegible]orld. The Commission is contributing €[illegible] four years to the €73 million project. Other fun[illegible] provided from national sources such as National Grid Initiatives (NGI). "

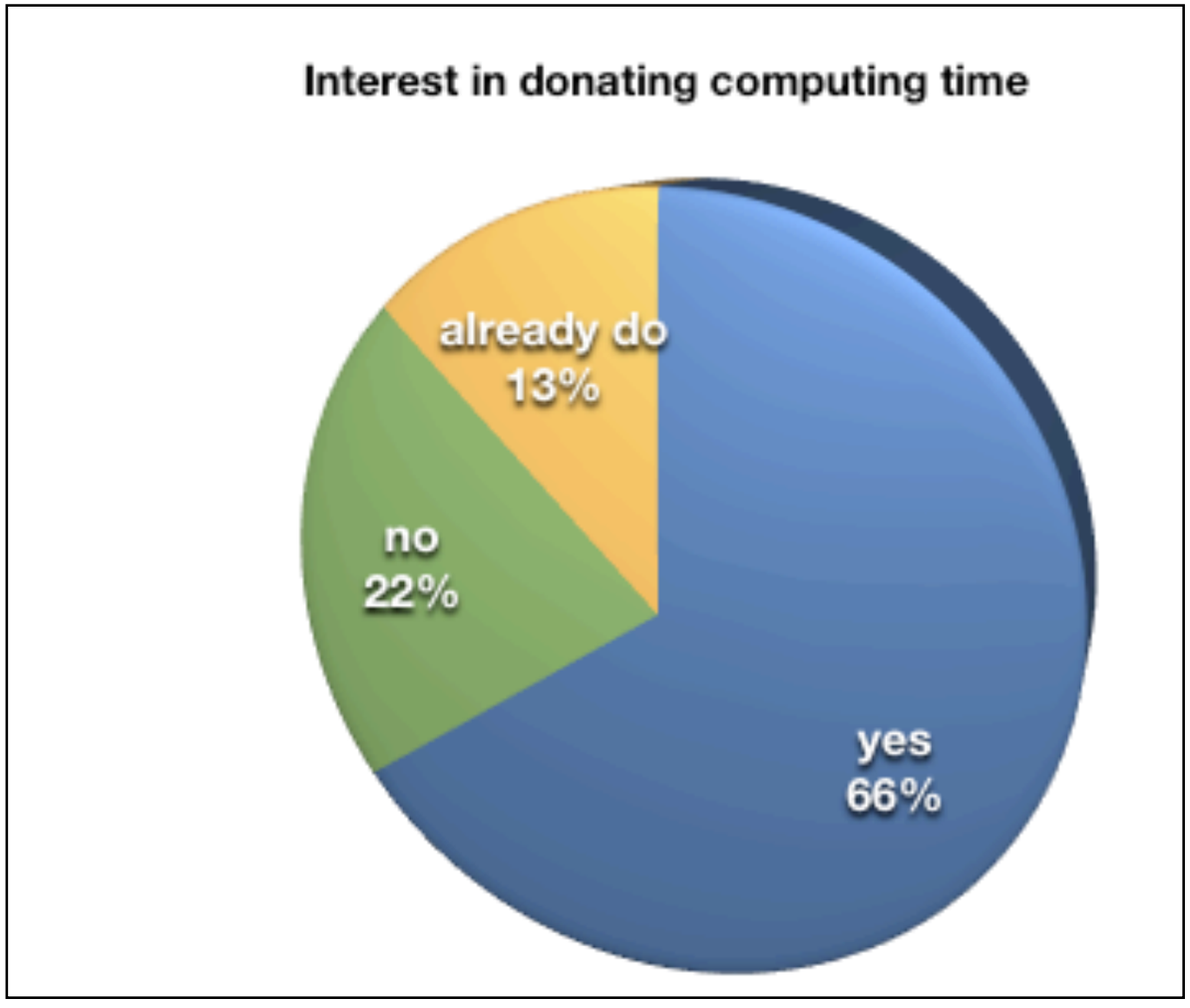
Not entirely True - Actually it will be through EDGI/DEGISCO/IDGF She gives about 3 million to these projects

But how much capacity?

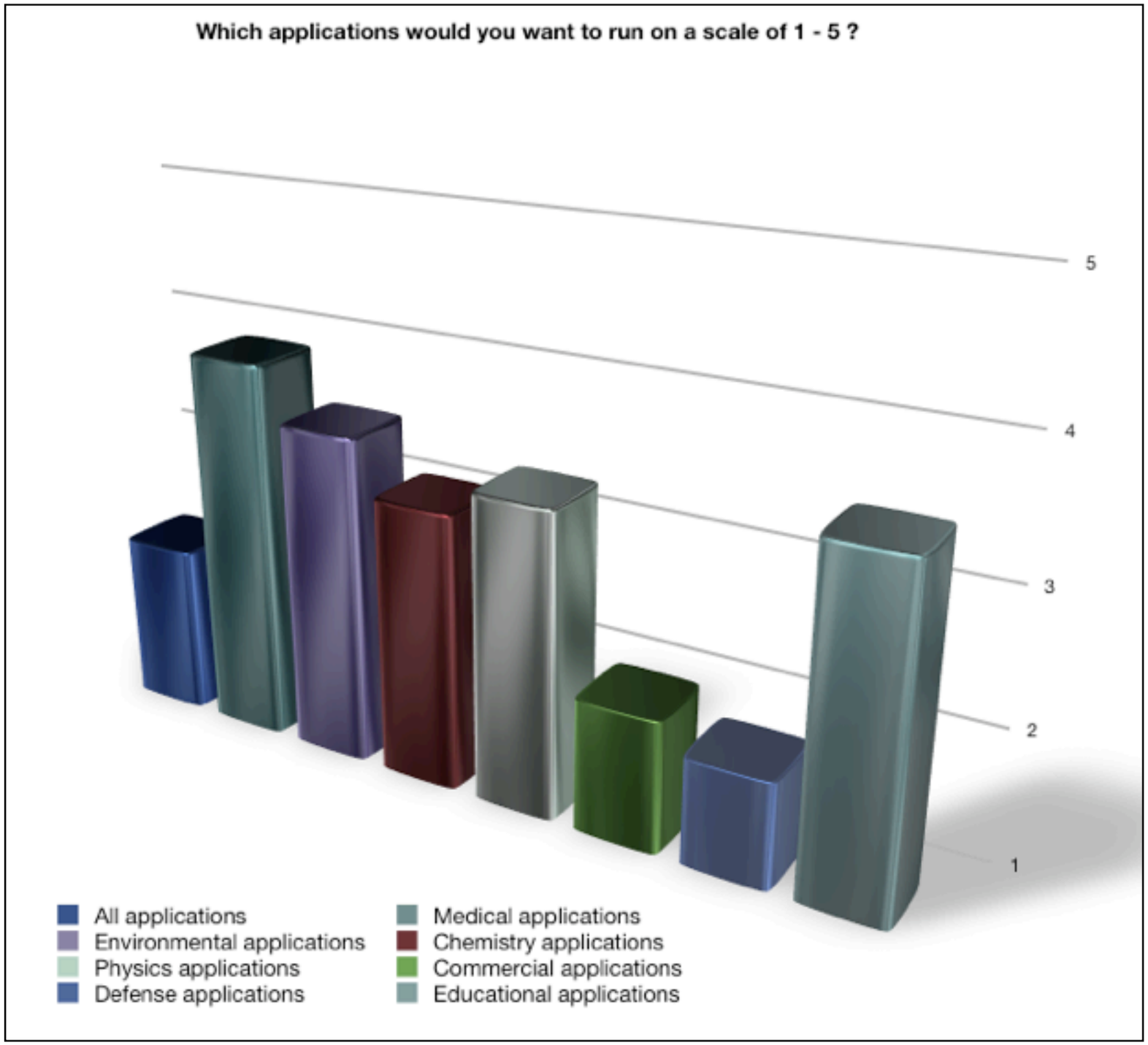
- OK, people want to donate computing time to science. But what is the capacity of a Desktop grid?
- Can be considerable: tens of thousands or more computers
- And also you can consider installing a local Desktop Grid at your institute university
- So enough capacity available for your scientific project or your scientific users
- But how and when to use it?

Are people interested in donating computing time?

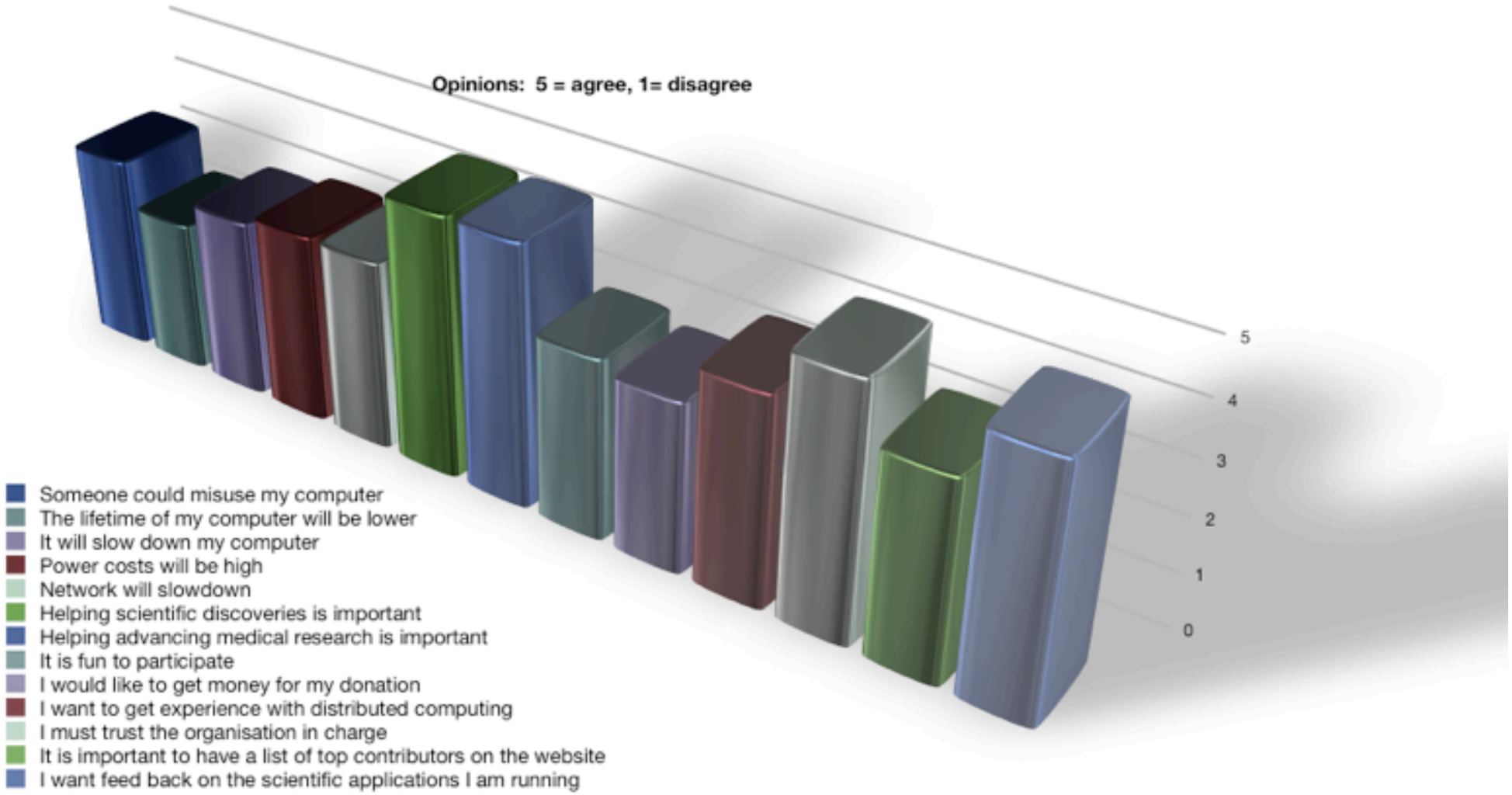
- Short answer: Yes, they are
- Of course, there is more needed to actually let them participate



Which applications?



Opinions about Grid computing



	Someone could misuse my computer	The lifetime of my computer will be lower	It will slow down my computer	Power costs will be high	Network will slowdown	Helping scientific discoveries is important	Helping advancing medical research is important	It is fun to participate	I would like to get money for my donation	I want to get experience with distributed computing	I must trust the organisation in charge	It is important to have a list of top contributors on the website	I want feed back on the scientific applications I am running
Average	3,6	2,8	3,3	3,5	3,5	4,8	4,8	3,3	3,1	3,8	4,6	3,6	4,5
Standard deviation	1,5	1,6	1,5	1,5	1,6	0,5	0,5	1,3	1,4	1,4	0,9	1,5	0,9

Survey - Conclusions

- Overall: *there is interest in Desktop Grid computing in Europe.*
- However, that people are willing to change their current practice and say that they want to participate in Grid efforts does not mean that they are actually going to do that.
- Need to *generate trust in the organisation that manages the Grid.*
- People want to *donate computing time for scientific applications,* especially medical applications. They *do not like to donate computing time to commercial or defense applications.*
- People want *feedback on the application they are running.*
- No clear technical barriers perceived by the respondents: so this does not need much attention.
- Overall the respondents were rather positive about donating computing time for a Grid or about running applications on a Grid.

A Grid definition

- Connecting a large number of resources and make them available for an application
- Resources belong to different owners
- Typically for Desktop Grids: it is the Desktop Grid computer that initiates all communications: there is no central control over the desktop computer
- Most computers in a Desktop Grid are desktop computers, but they can also be notebooks, or servers, or mobile phones

What is Grid

- Using physical resources belonging to different organisations for a single application
- Main problem: crossing (administrative) boundaries between the organisations

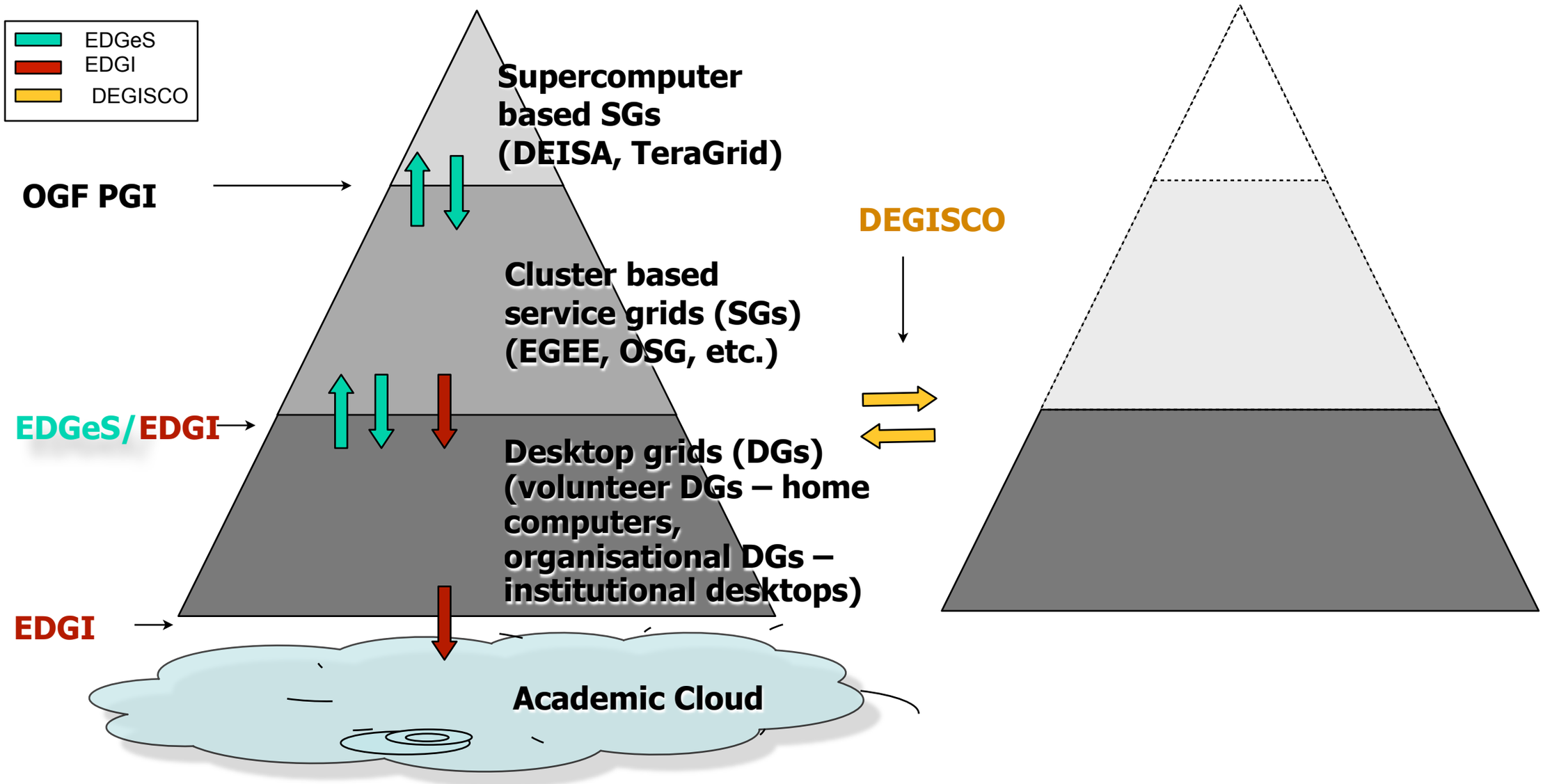
Resources that can be shared

- Processors - compute Grid
- Storage - storage Grid
- Data bases, repositories - data Grid
- Sensors - sensor Grid

Different owners of resources

- Citizens' and SMEs' PC's: Desktop Grid
  - Thousands or even millions of owners
- Company divisions: Enterprise Grid
  - A dozen or so organisations within one enterprize
- Universities and research institutes - eScience Grid
  - Up to a few hundred (in the case of EGI, a pan-European Grid)

Where Desktop Grids fit in



A Grid is an infrastructure

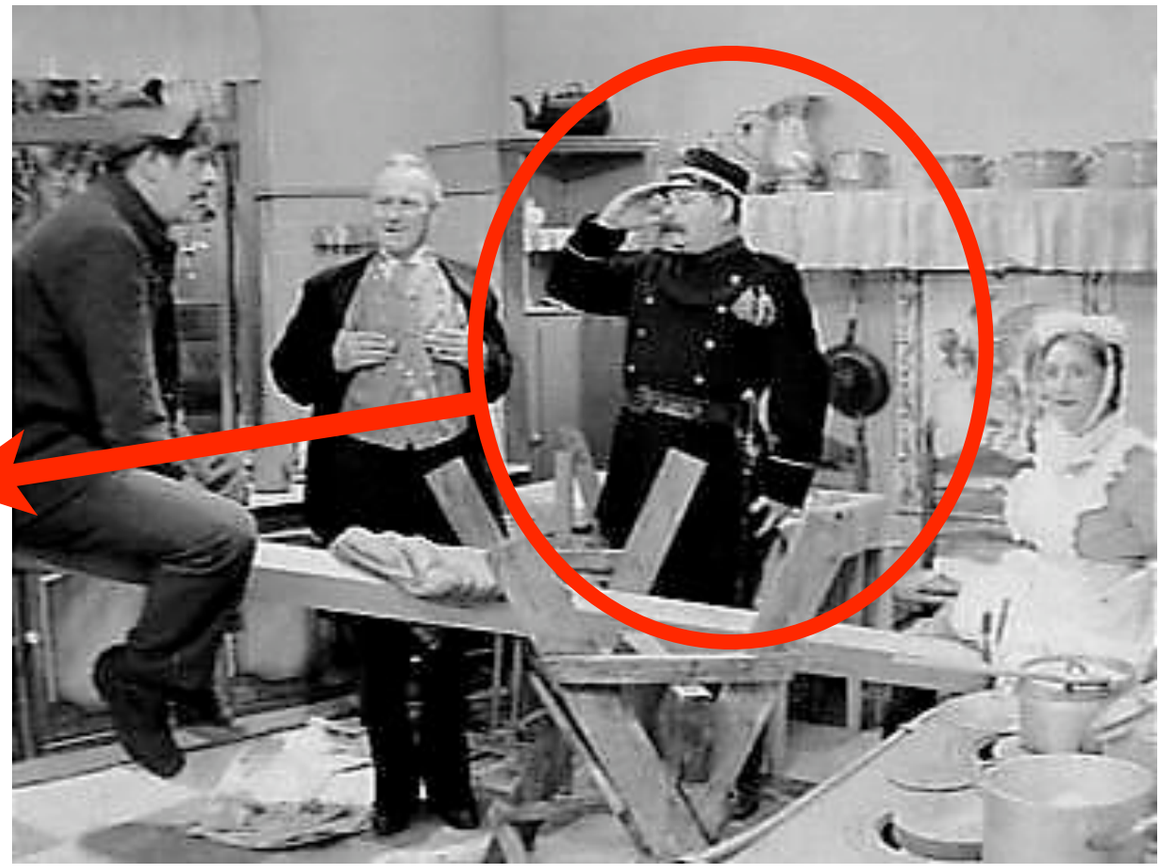
- Recipe:
  - Take resources
  - Connect them through a fast network
  - Add Grid middleware
  - And your Grid is ready
  - *But you have to port your applications to the Grid*

One resource is not a Grid

- One police man in a village
- One resource

One resource is not a Grid

- One police man in a village
- One resource



Coordination between many resources

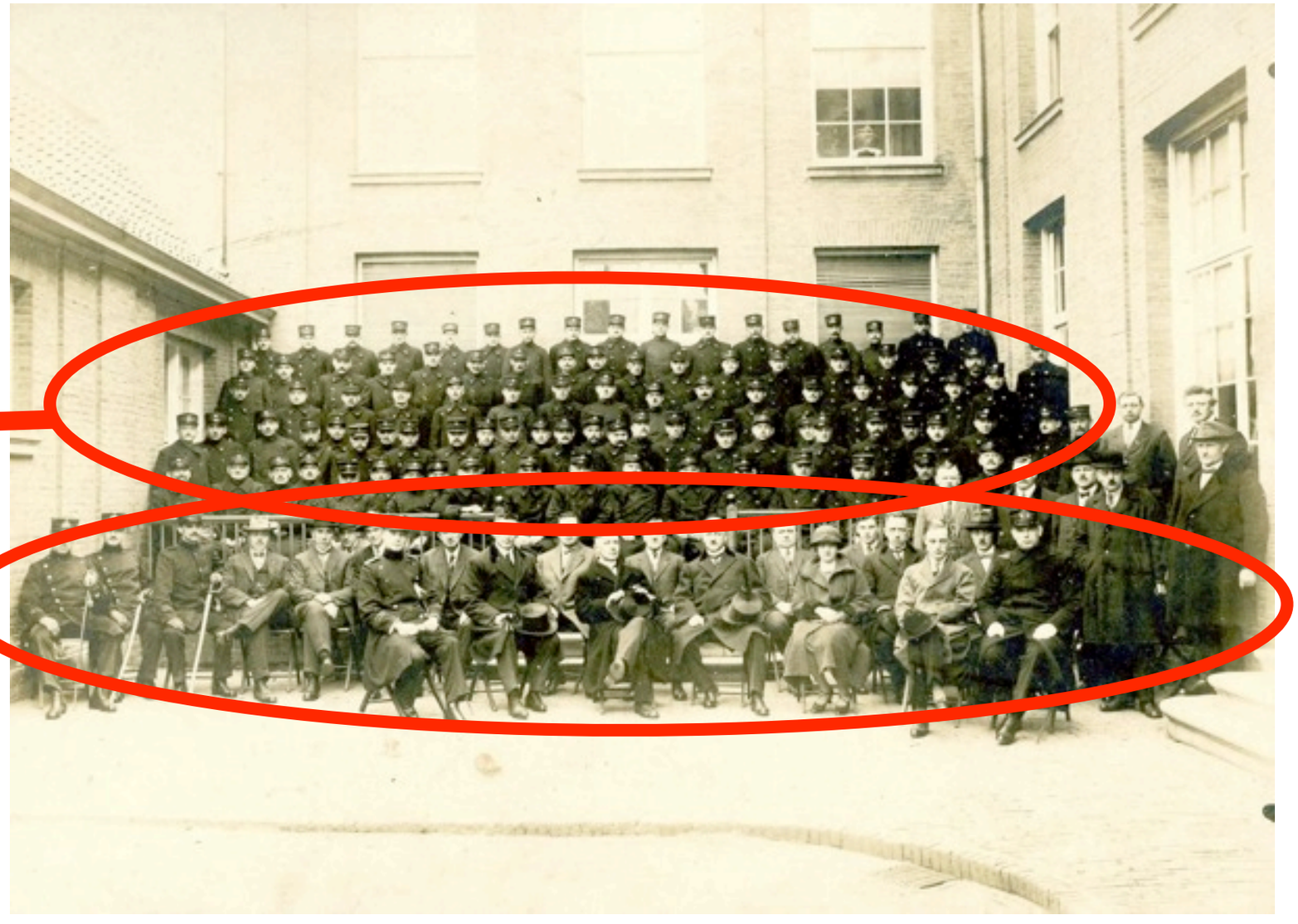
- One police corps
- Many resources
- Coordination and management

Coordination between many resources

- One police corps
- Many resources
- Coordination and management

Coordination between many resources

- One police corps
- Many resources
- Coordination and management



Coordination between many resources

- One police corps
- Many resources
- Coordination and management

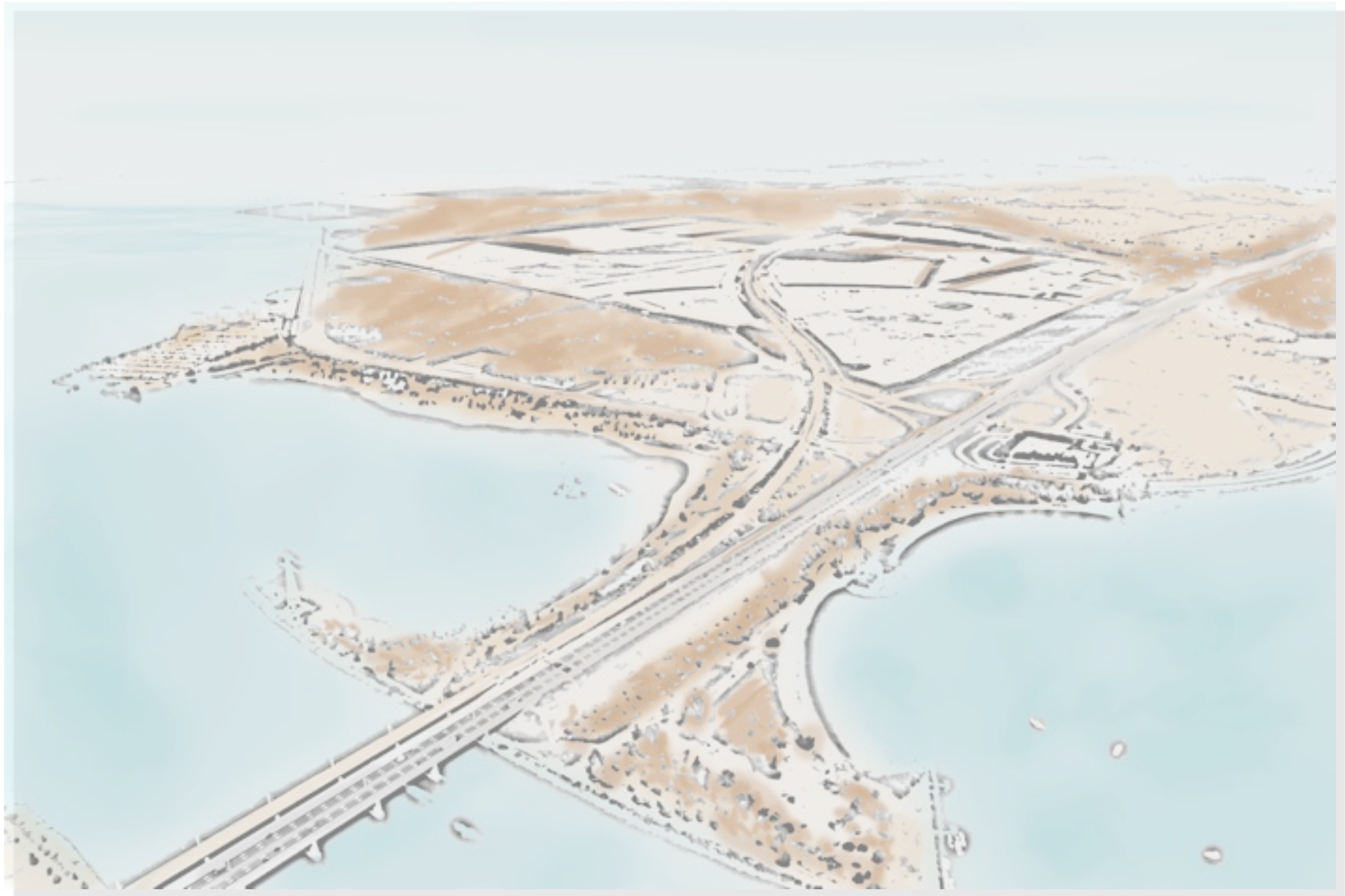
Almere - a New Town

Almere - a New Town

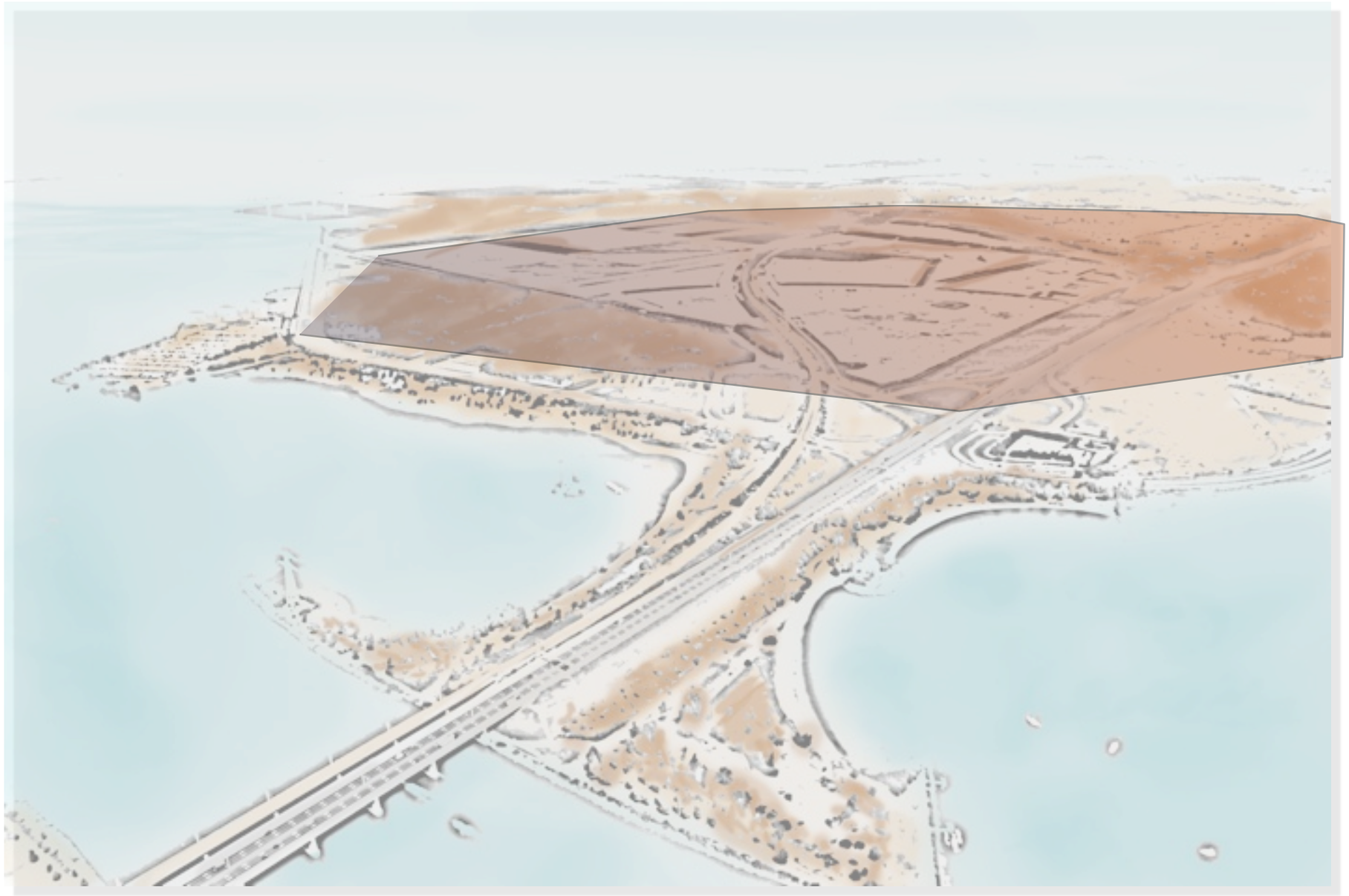
- Almere is a new town: build on the bottom of the former *Zuiderzee* sea
- Fastest growing town in The Netherlands
- From 0 to 190.000 inhabitants in about 25 years
- Perfect breeding ground for AlmereGrid

For science and beyond

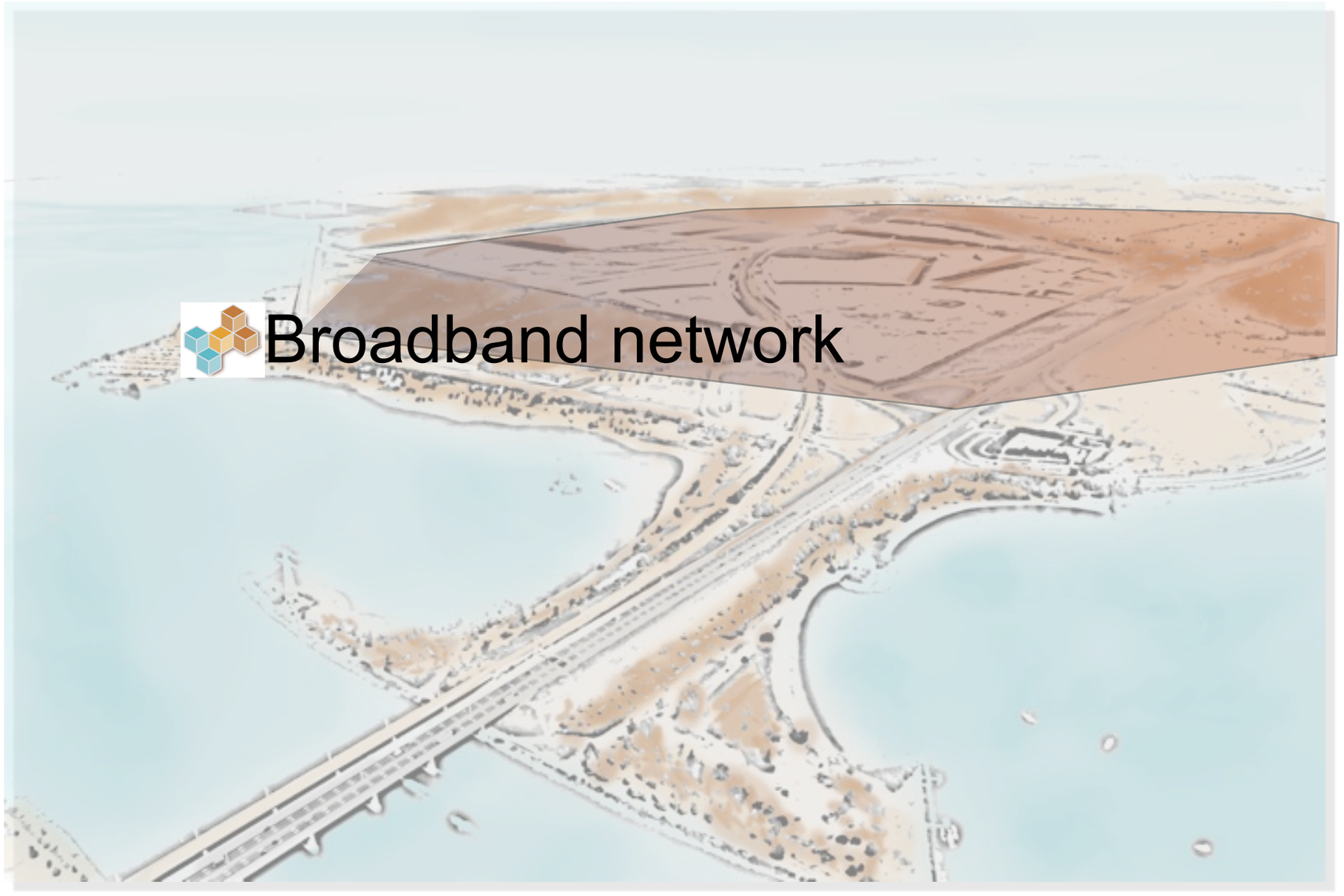
For science and beyond



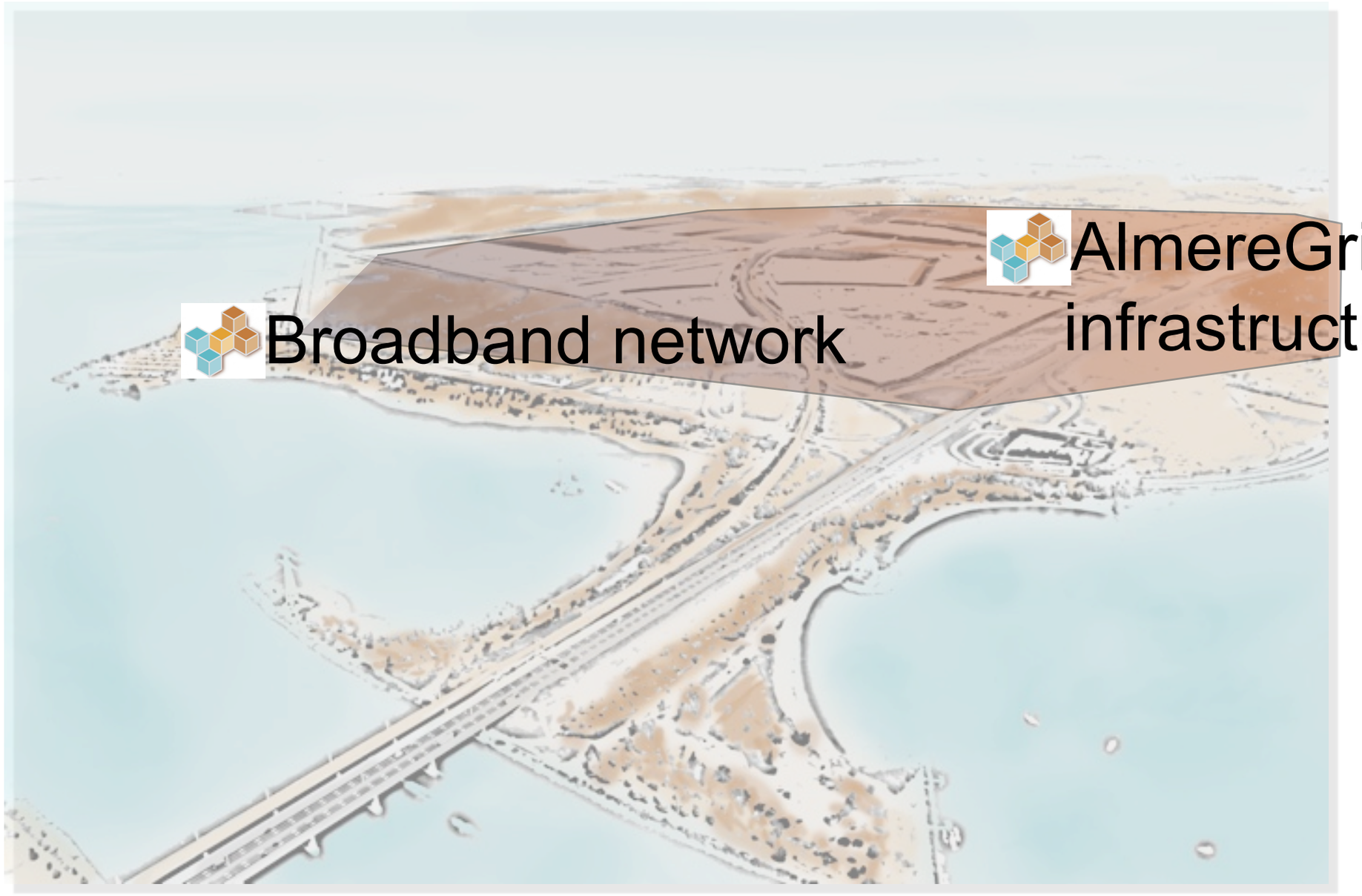
For science and beyond



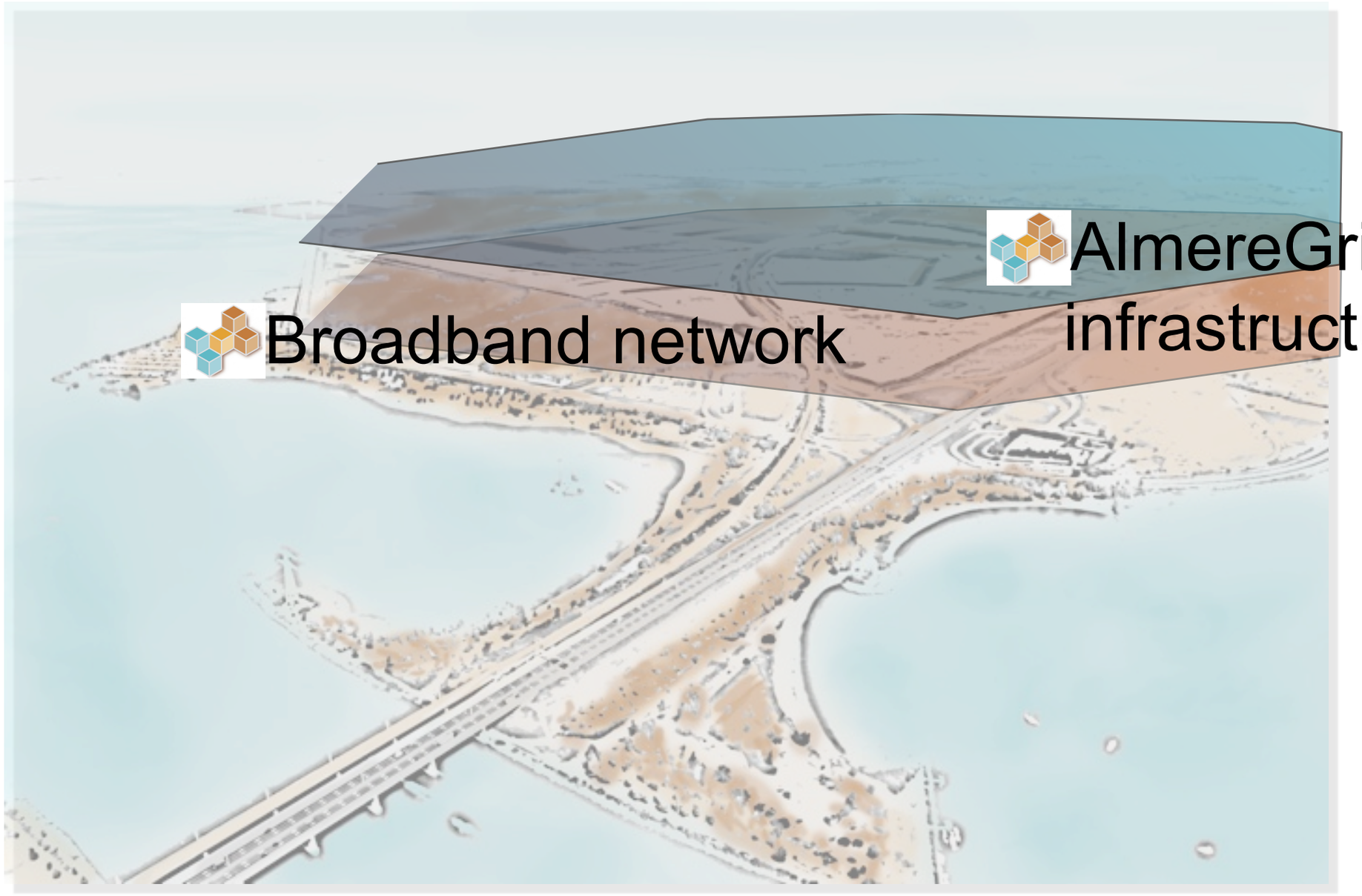
For science and beyond



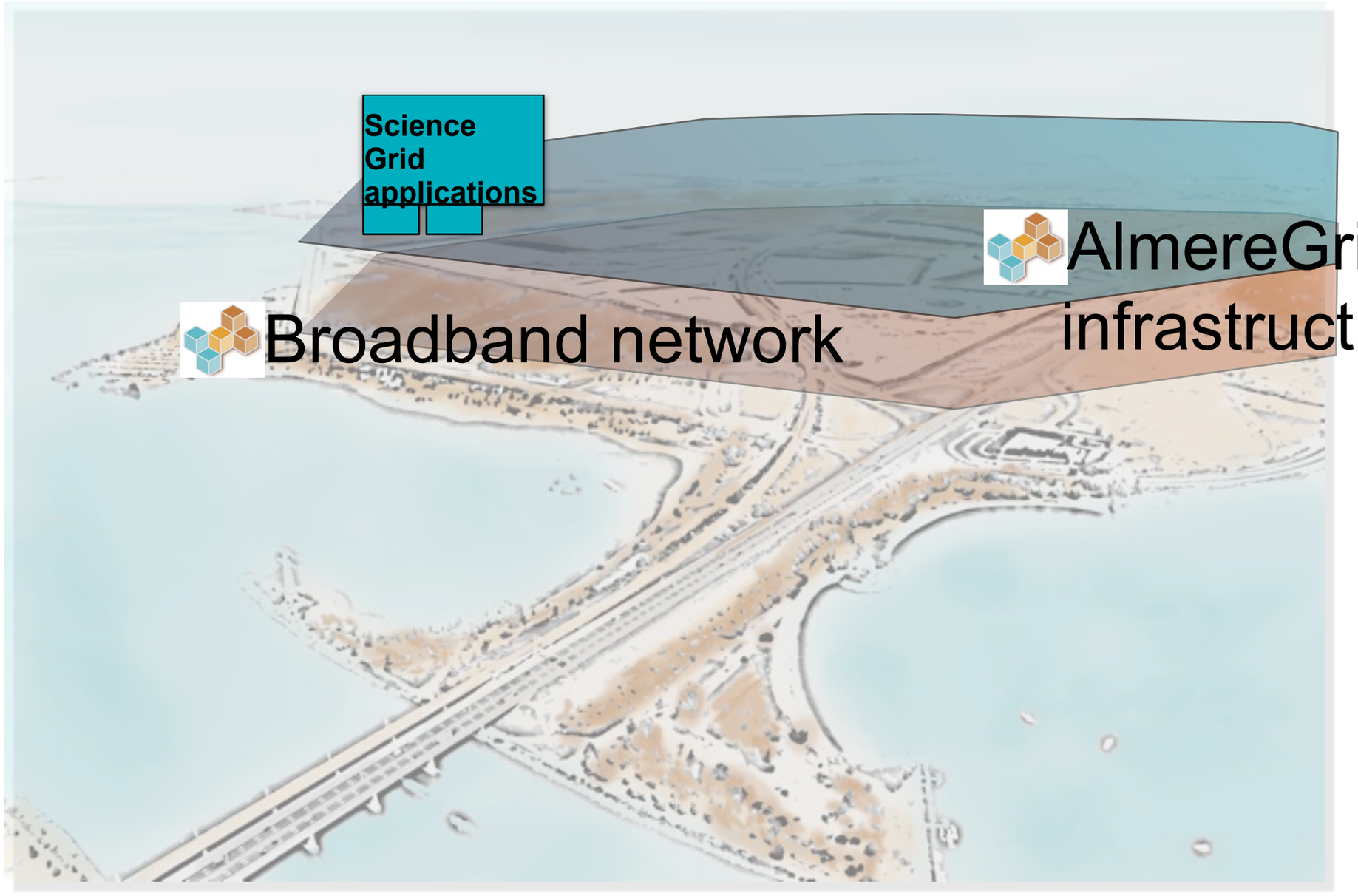
For science and beyond



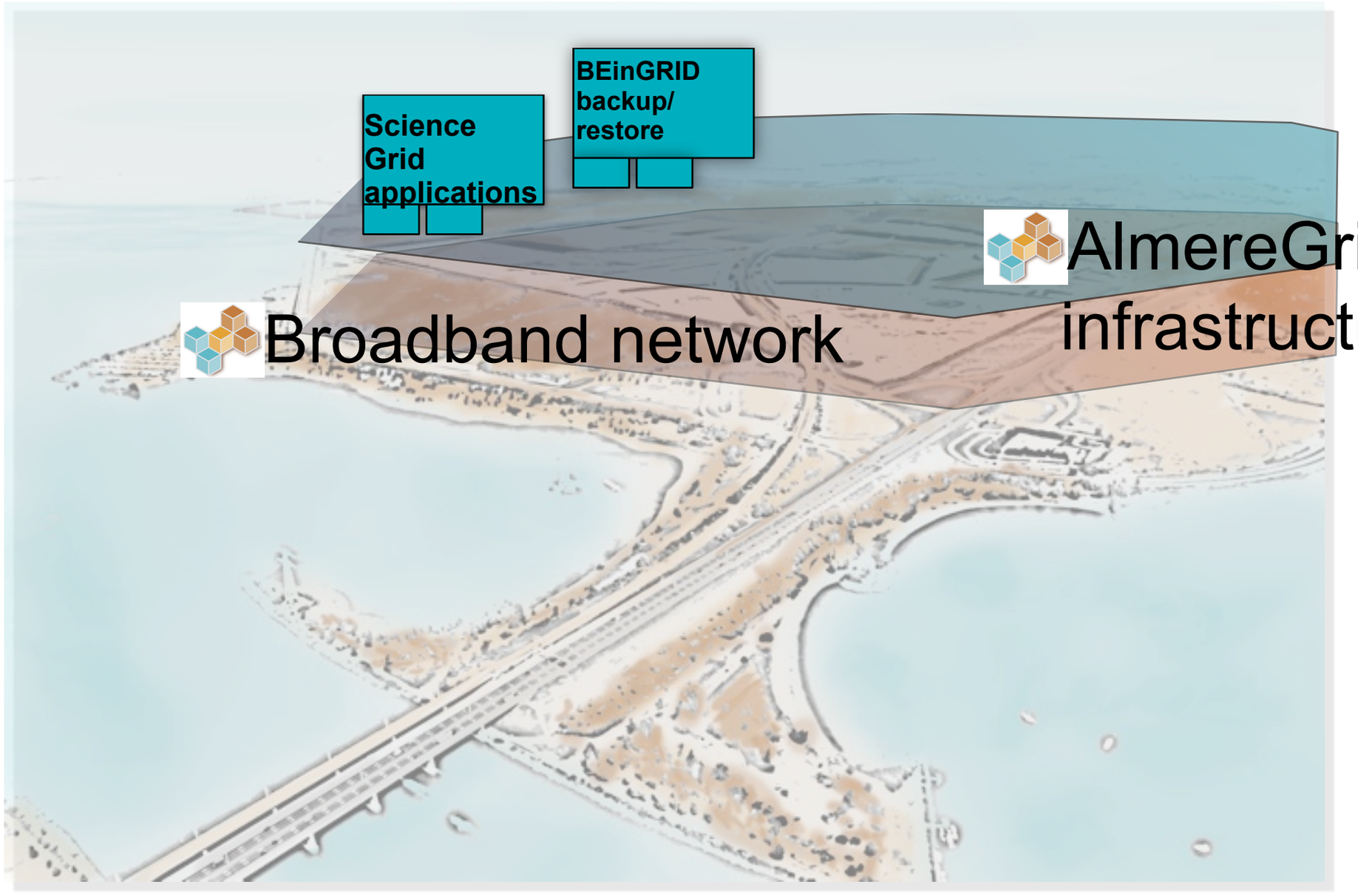
For science and beyond



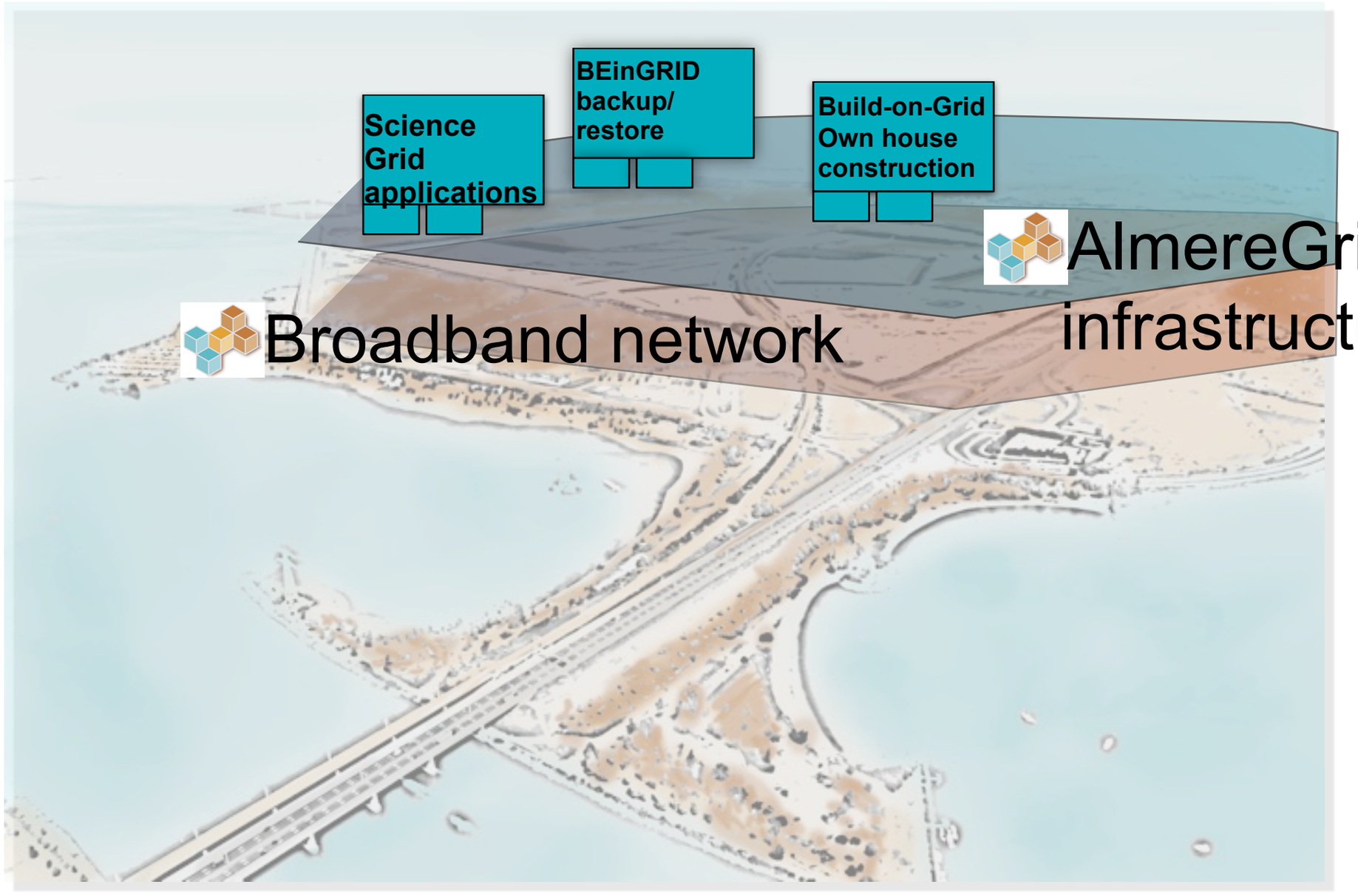
For science and beyond



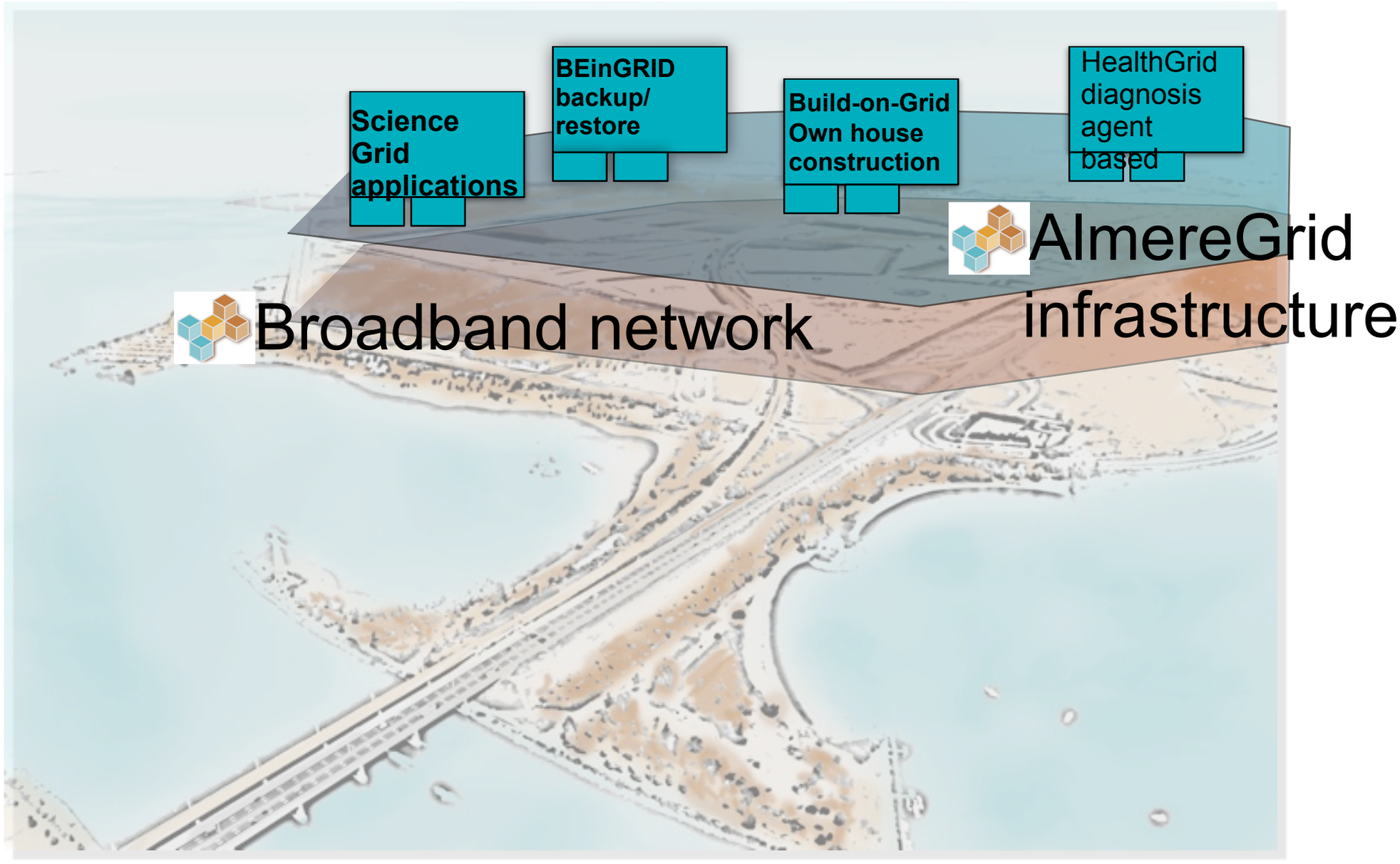
For science and beyond



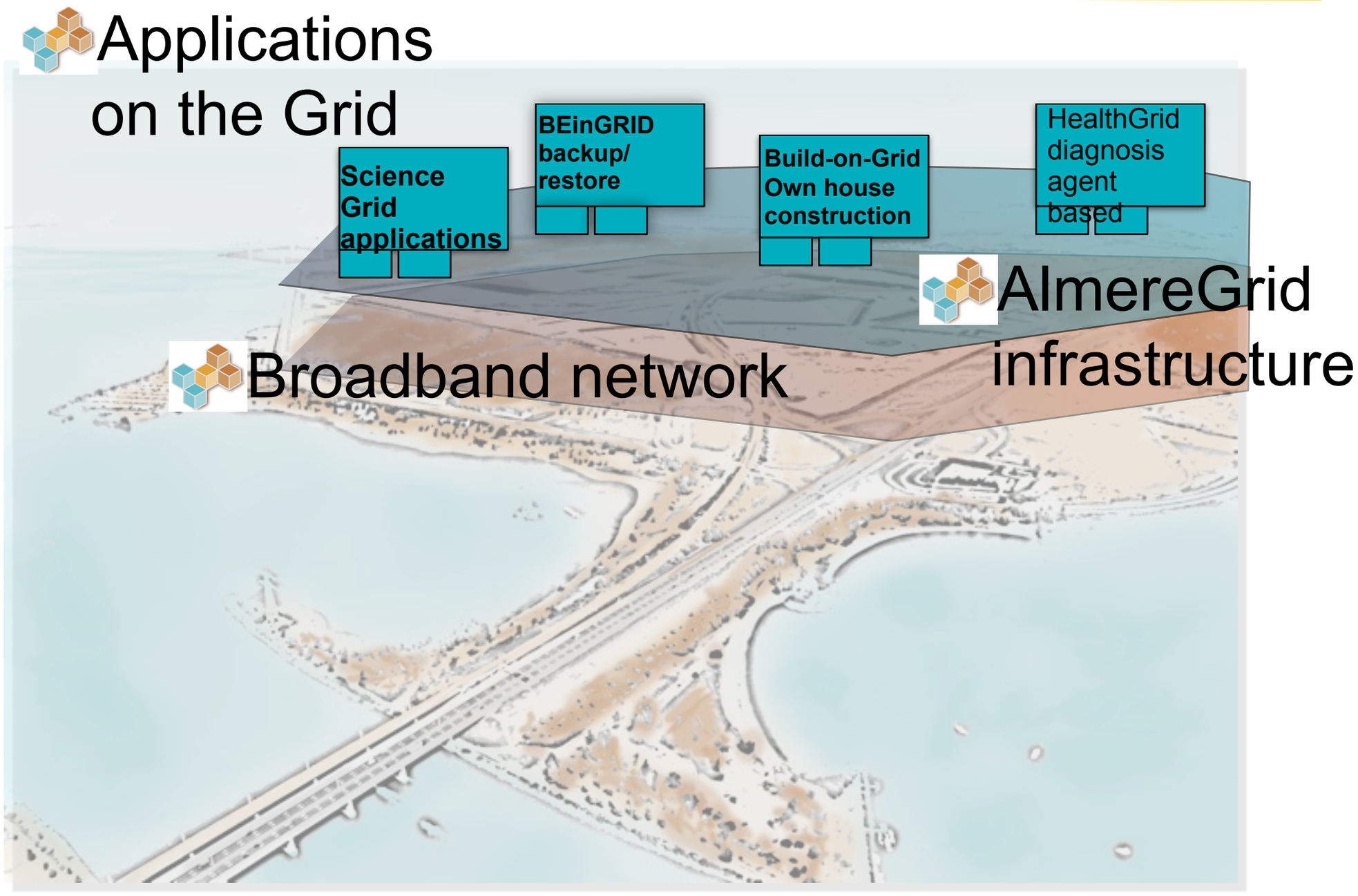
For science and beyond



For science and beyond



For science and beyond



For science and beyond

Applications on the Grid

e-Education using Grid

Strengthen knowledge infra-structure of the city

Science Grid applications

BEinGRID backup/ restore

Build-on-Grid Own house construction

HealthGrid diagnosis agent based

Broadband network

AlmereGrid infrastructure

Desktop Grids - introduction

- Agenda
  - Basic techniques
  - Open source examples
    - XtremWeb
    - BOINC
  - Commercial examples
    - LSF Desktop
    - Datasynapse (Tibco)
    - Parabon

Basic structure

- Resource computers
  - The resource computers (also called workers, clients,..) execute the work
- User system or interface
  - Interface for the Grid user to submit his/her application, monitor the progress, and retrieve results
- Grid server
  - Handles the requests from the resource computers to get new work or submit the results
  - Handles the requests from the users to execute jobs

Basic desktop Grid

Basic desktop Grid

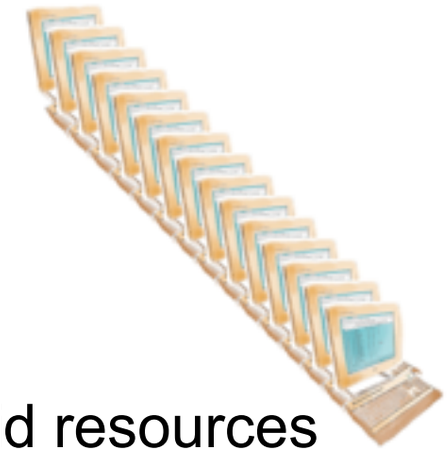
Companies

Grid organisation

Universities & research institutes

Residential areas

Basic desktop Grid



Grid resources

Grid resources

Grid server

Grid users

Volunteer desktop Grids

- Uses a pull model: it is always the resource computer that asks for work and initiates all communication with the server
- There is an application repository: users can only run validated trusted applications
- The resources, however, are untrusted and volatile: you are not sure they deliver the correct answer

Local Deskop Grids

- Organisations (universities, companies) can also run an internal Desktop Grid.
- Can be composed of trusted resources
- Use spare capacity
- Can also include servers and clusters

Desktop Grid applications

- There is in general no communication between the resources
- Hence you can only run *pleasantly parallel* programmes
- In volunteer computing the bandwidth can be rather limited and asymmetrical
- But there can be large numbers of resources

XtremWeb$^{HEP\text{-}E}$

- Written in Java
- Can be used for all type of Grids, from simple local to advanced volunteer Desktop Grids
- Versions:
  - XtremWeb by High Energy Physics (XtremWeb^{HEP}) is developed by IN2P3. *http://xtremweb-hep.org/*
  - It is based on XtremWeb 1.8.0. by INRIA *http://xtremweb.net*
  - There is a special extensively tested EDGeS distribution of XtremWeb^{HEP} called short: XtremWeb$^{HEP\text{-}E}$ *http://edges-grid.eu*

Structure of XtremWeb$^{HEP\text{-}E}$

- Three components:
  - Server
  - Worker
  - Client
- Server
  - Sends work to a Worker
  - Receives the results
  - Admininstration of all components

Structure XtremWeb$^{HEP\text{-}E}$ -2

- Worker
  - Requests works
  - Executes work
  - Sends results back to the server
- Client
  - Interface to the server
  - Execute all commands
  - Command line or GUI
  - Submit work, monitor work. Retrieve results from the server

XtremWeb$^{HEP\text{-}E}$ installation

- From sources
  - *Installation guide:* http://www.edges-grid.eu/c/document_library/get_file?p_l_id=29093&folderId=23974&name=DLFE-1023.pdf
  - There is also an experimental installer for Linux
- Documentation of XtremWeb is not good: look for assistance
  - *Discussion group:* http://www.edges-grid.eu/web/edges/10/message_boards/category/15353
- Will be moved to: http://desktopgridfederation.org

XtremWeb$^{HEP\text{-}E}$ applications

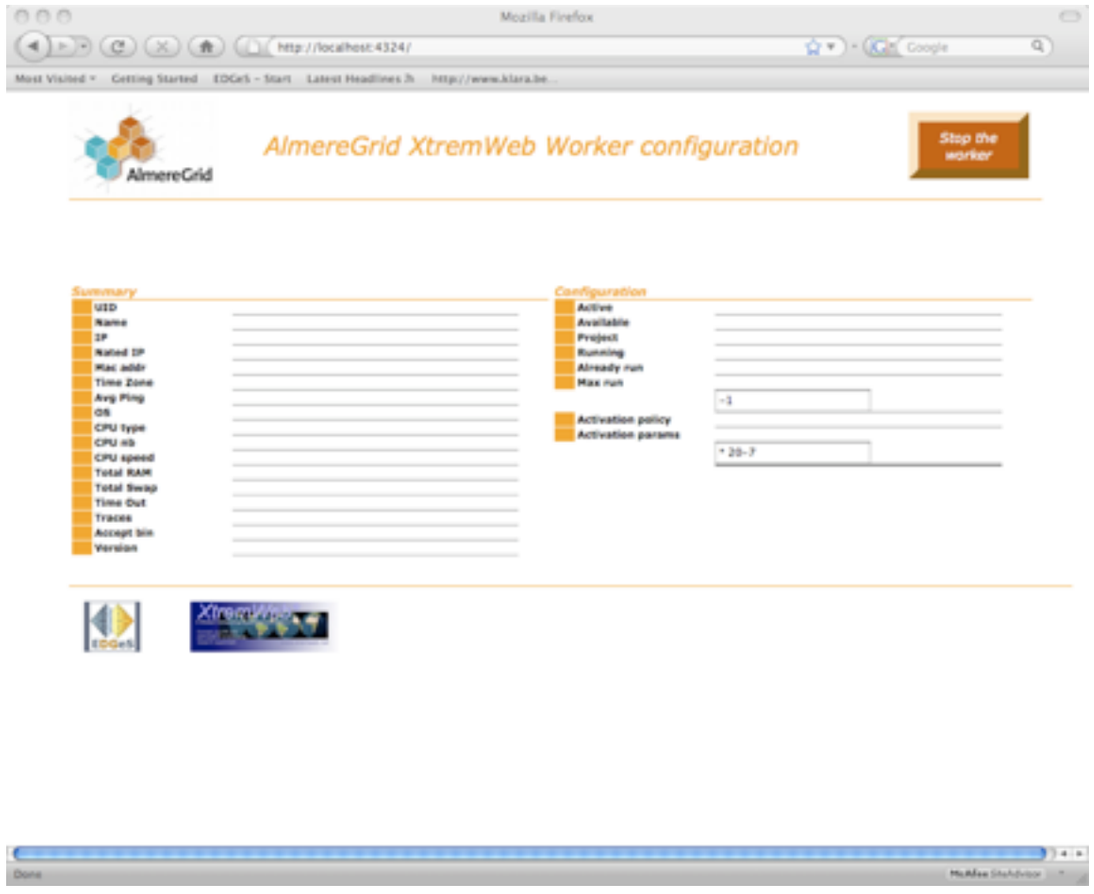
- XtremWeb resource computers run any binary static executable for the resource's Operating System
- Applications run inside and under control of the XtremWeb worker application
- Versions for Windows, MacOSX, Linux

XtremWeb workers - resources

- On the worker: far less sophisticated than the BOINC client
- Also overview
  info on grid server
  website

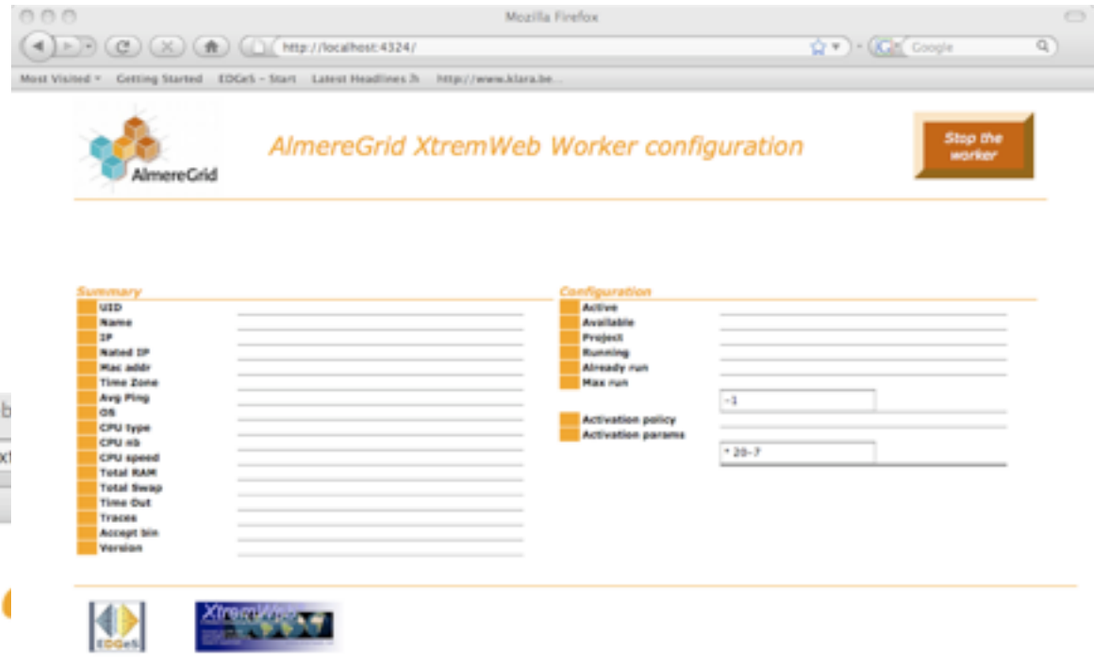
XtremWeb workers - resources

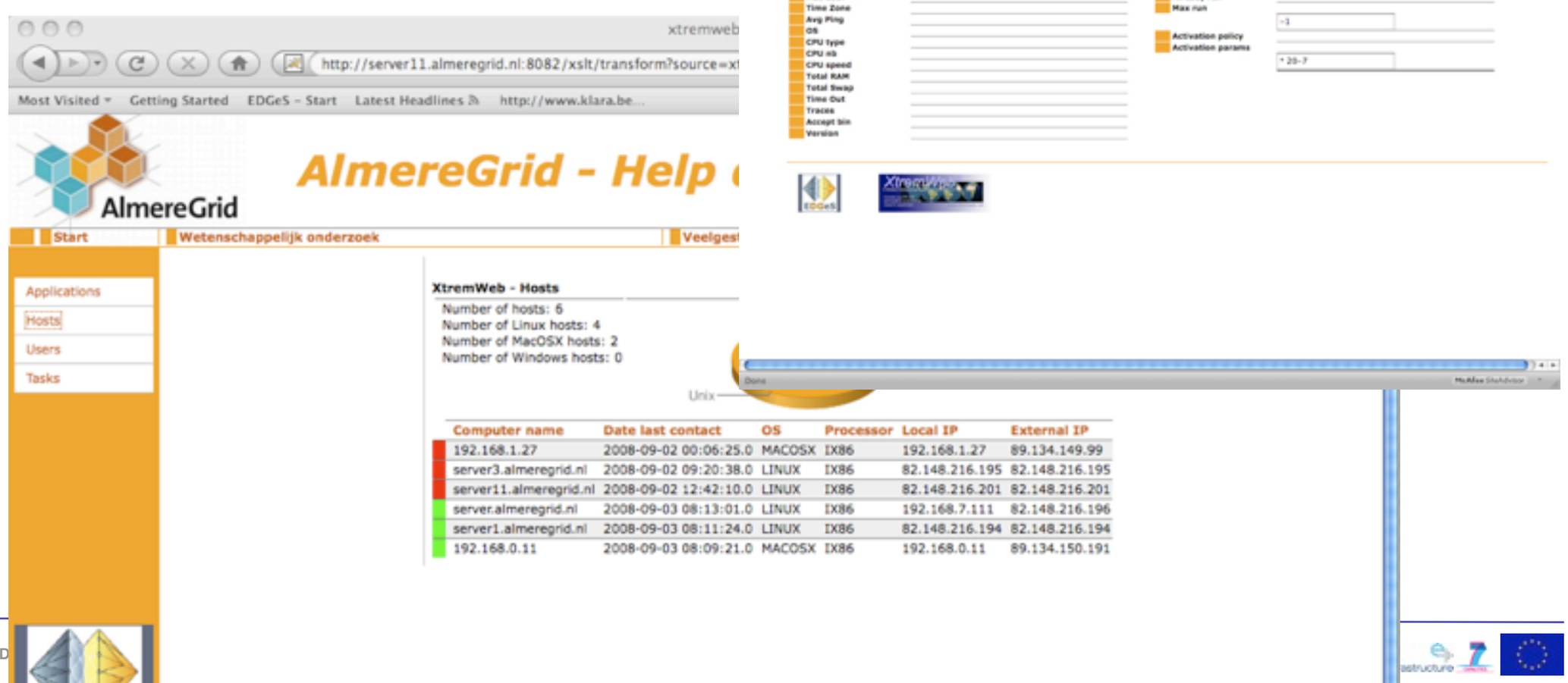
- On the worker: far less sophisticated than the BOINC client
- Also overview info on grid server website



XtremWeb workers - resources

- On the worker: far less sophisticated than the BOINC client
- Also overview info on grid server website





XtremWeb$^{HEP\text{-}E}$ tasks & bridge

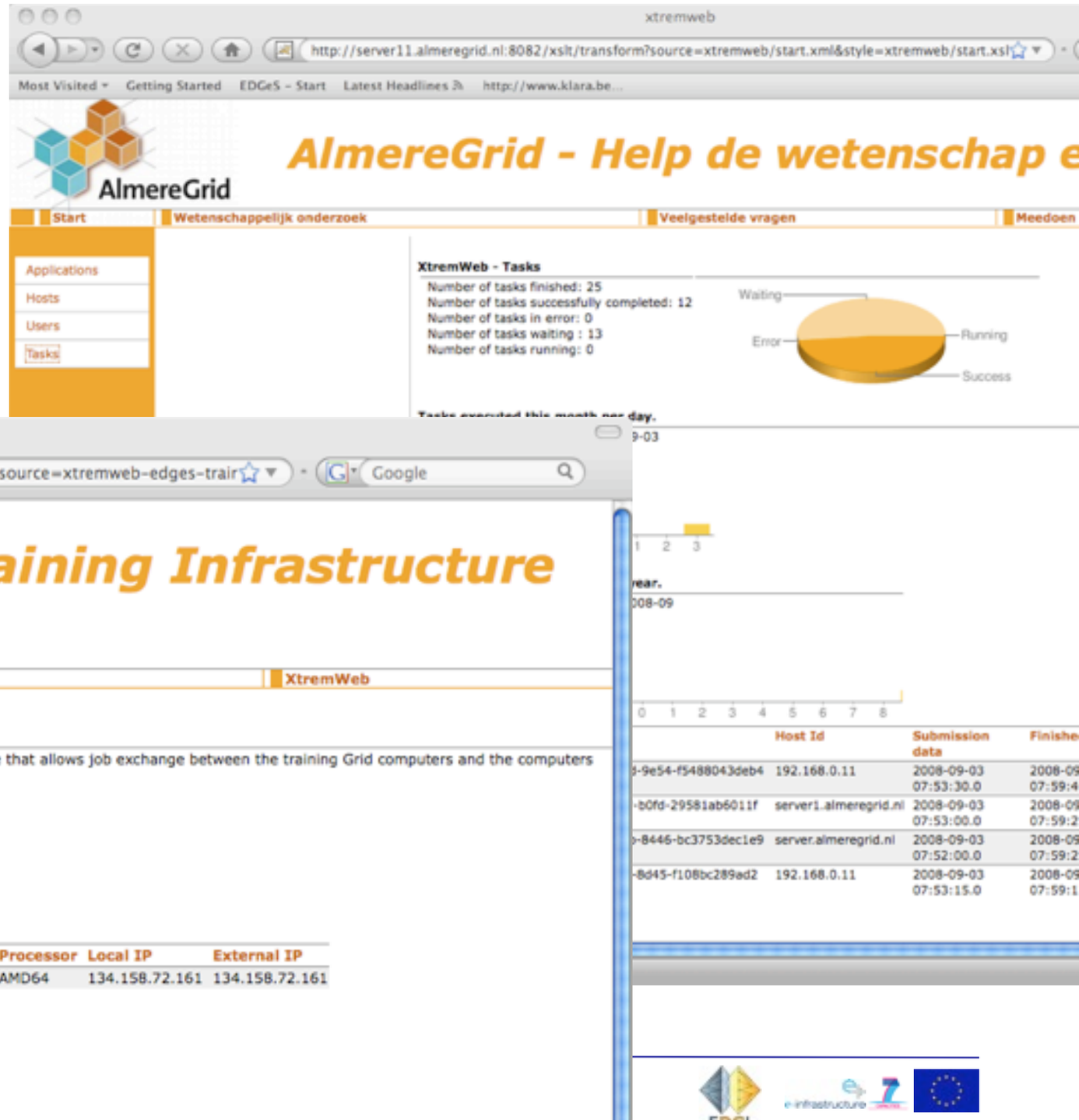
- In XtremWeb jobs are called tasks
- There is a bridge
  possible to
  Service Grids

XtremWeb$^{HEP\text{-}E}$ tasks & bridge

- In XtremWeb jobs are called tasks
- There is a bridge possible to Service Grids

XtremWeb$^{HEP\text{-}E}$ tasks & bridge

- In XtremWeb jobs are called tasks
- There is a bridge possible to Service Grids



BOINC

- Middleware designed for volunteer Desktop grid computing
- *http://boinc.berkeley.edu/*
- Used for the largest Grid in the world: Seti@Home
- More than 4 million computers world wide have BOINC installed
- Many other projects use BOINC

Some other Desktop Grids

- World Community Grid - IBM (http://www.worldcommunitygrid.org/) *900.000 PCs*
- Leiden Classical Grid - Education on Grid (http://boinc.gorlaeus.net/) *45.000 PCs*
- SZTAKI - Hungarian Grid (http://www.desktopgrid.hu/) *80.000 PCs*
- AlmereGrid (http://almeregrid.nl) 3.000 PCs
- PS3GRID (Based on Playstations)
- (http://ps3grid.net/)

BOINC Structure

- BOINC is organised around projects: a project is typically one application
- Resource computers (BOINC clients) can connect to one or more projects
- A Grid server can run more projects: but most only have one or perhaps a few
- User interface is integrated with the project

BOINC installation

- Installing a BOINC Grid is not *click and play*
- Collection of C++ programmes, PHP scripts and other sources, with many dependencies on operating system, the database, and other tools
- Easiest way: get one of the distributions. For instance *the SZTAKI BOINC distribution.* This installs relatively fast if you start with a Debian 5.0 system

BOINC Client on resource

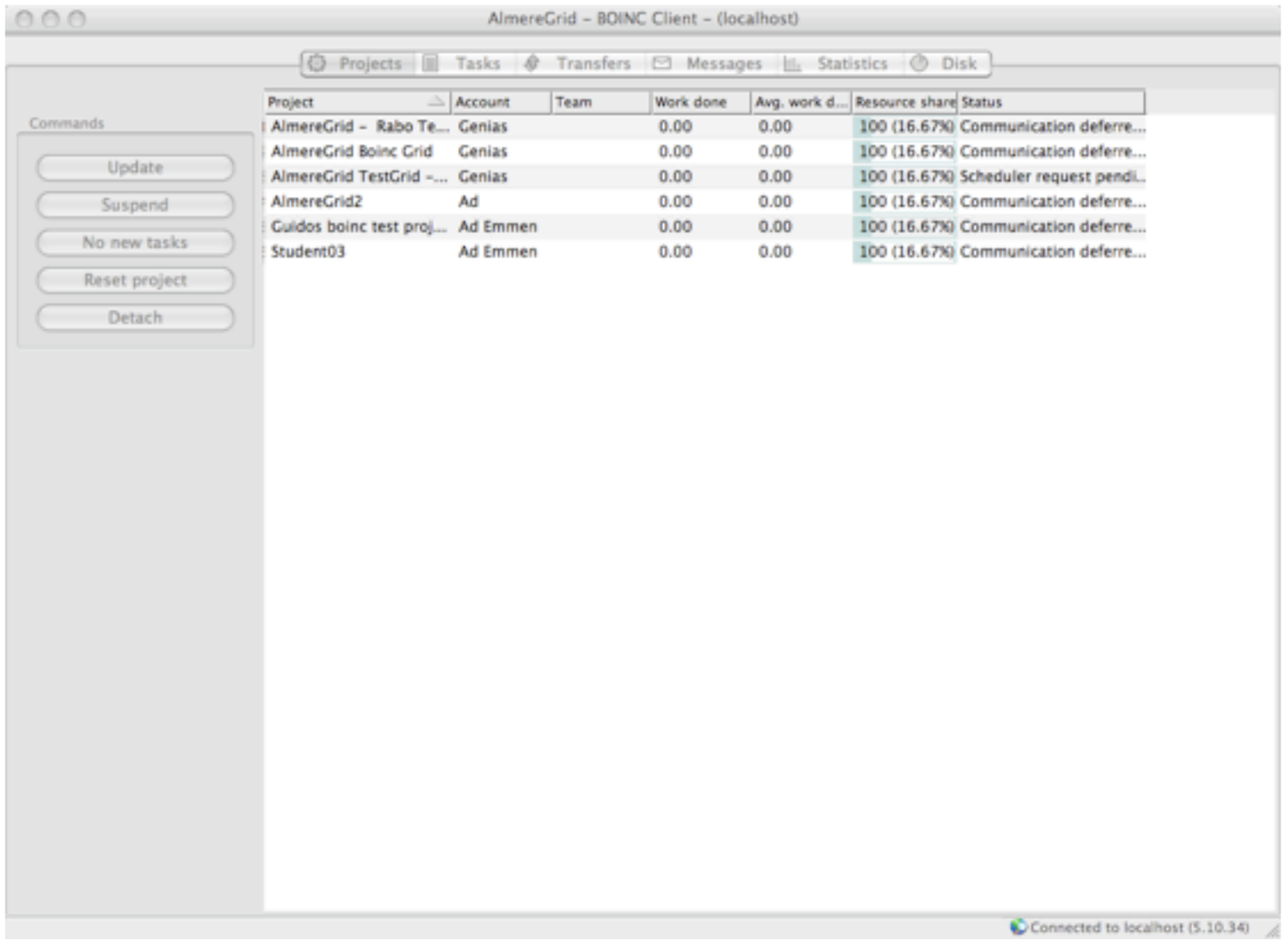
- Customisable
- Easy view and complex view with a lot of details on what is happening

BOINC Client on resource

- Customisable
- Easy view and complex view with a lot of details on what is happening

BOINC Client on resource

- Customisable
- Easy view and complex view with a lot of details on what is happening



BOINC applications

- BOINC applications have to be compiled with calls to a special BOINC API.
- There are some wrappers for executables worka fin for simple applications (one executable, a few input/output files
- More extensive API's exist for instance for master worker-applications: SZTAKI DC-API

Computing nomads

- BOINC has a rudimentary credit system. You get credits for each application that is run on your system.
- Some people are collecting credits to get on top of the ranking of a project (for fun or competition)
- This leads to computing nomads that go from project to project trying to get on top
- With BOINC you can also form *teams*. So you can alos try to get your team on top
- However, for many SME's and private persons the credit system is a barrier to participate: they do not want to compete, just donate computing time to science

Commercial desktop Grids

- Desktop Grids are a big commercial success
- Many large companies (pharmaceutical, engineering, banks,..) have installed big Desktop Grids
- But they do not tell much about it: it is a competitive advantage and proven technology
- Several companies provide Desktop Grid middleware; although some now call in differently because they do not want to be associated with the Globus and gLite based Grids that by many are seen as failed in the commercial area

Datasynapse

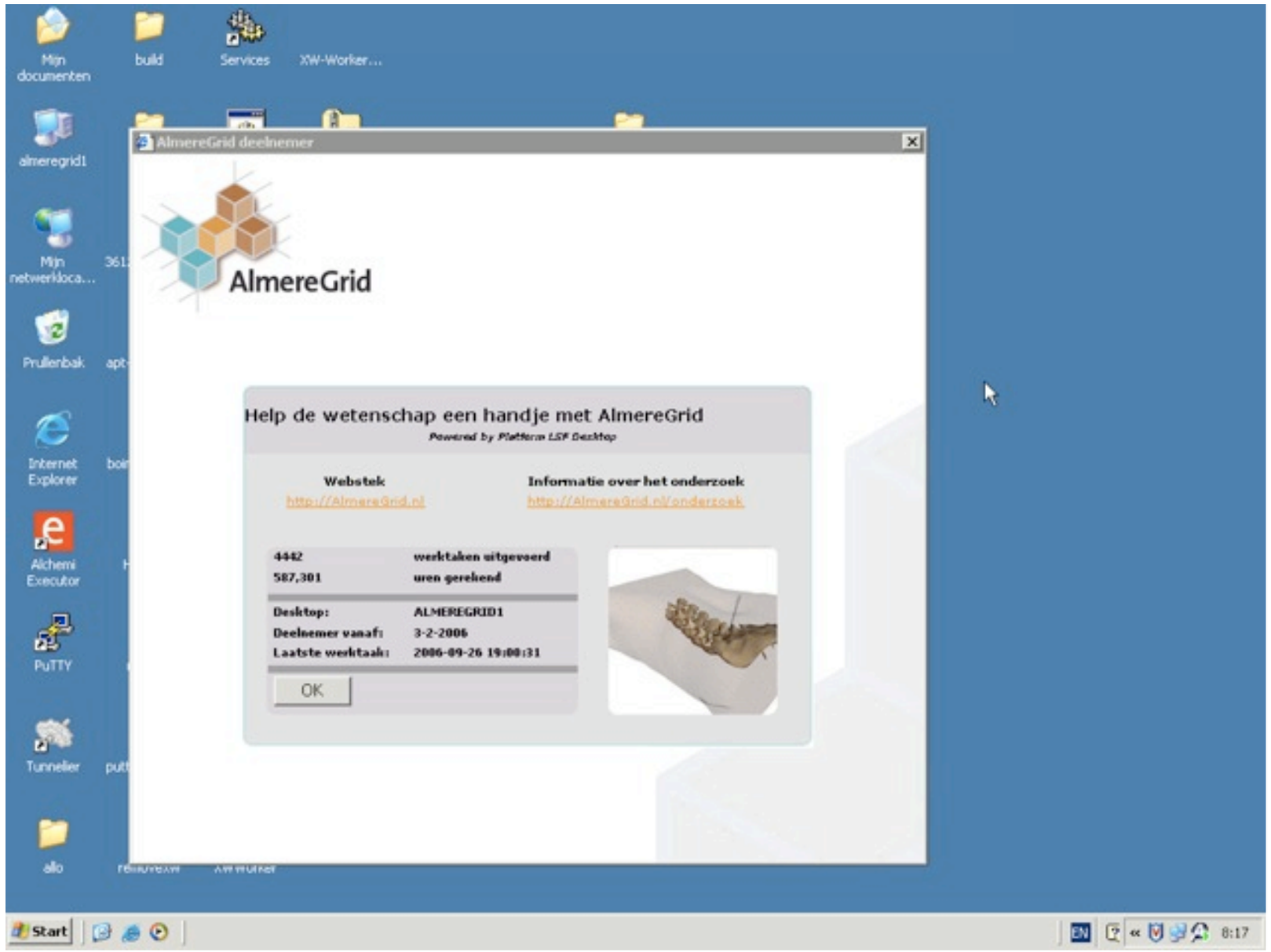
- More than 1 million computers run Datasynape software in the world.
- Making it, probably, the most successful Grid middleware after BOINC
- Acquired by Tibco (Cloud provider) last year for US$28 million

Platform LSF Desktop

- Platform computing is involved in job management systems for already many years
- Platform LSF is mainly focused at managing systems with many processors and (large) clusters. This works in push mode
- Addition: LSF Desktop for Windows machines. This works in pull mode and is used to connect Desktops to an LSF cluster

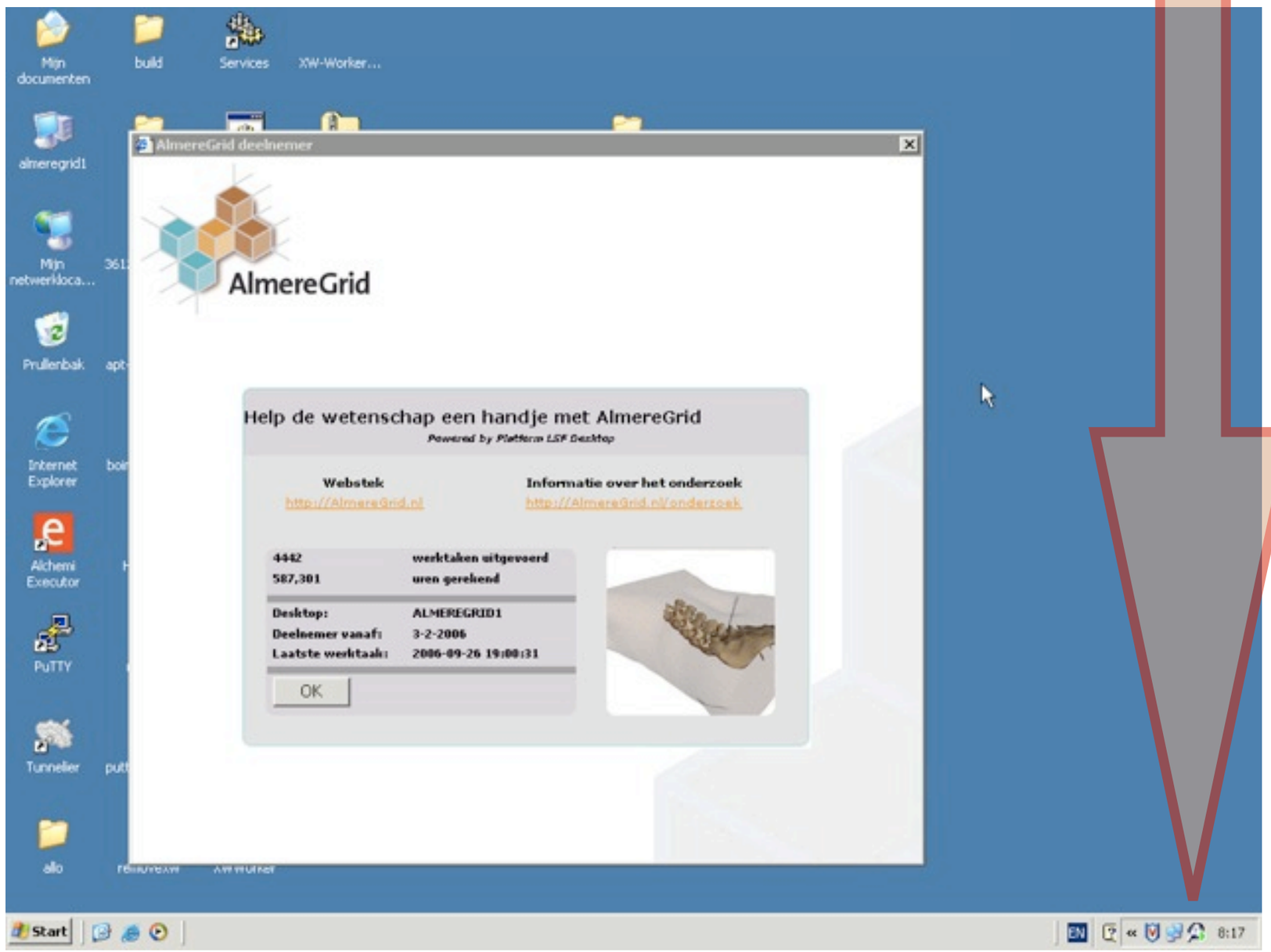
LSF Desktop client

- Tray shows activity
- Customisable



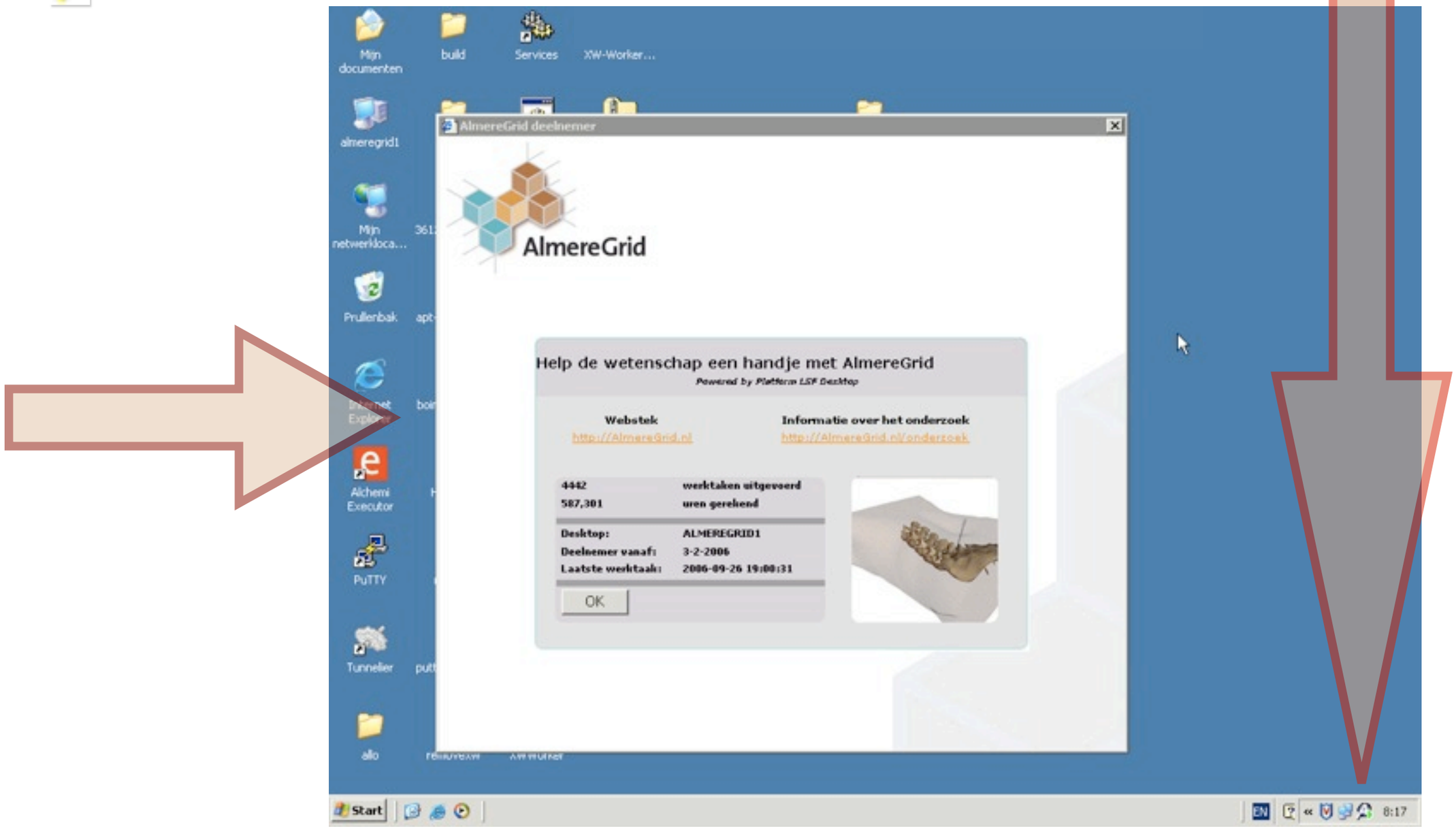
LSF Desktop client

- Tray shows activity
- Customisable



LSF Desktop client

- Tray shows activity
- Customisable



Parabon

- Frontier Grid platform
- Works as a traditional Desktop Grid
- Fancy, easy to use user GUI

And there are more...

- Fura: Grid systems, Spain
- ProActive: ActiveEon, France
- Digipede, USA,
- …

Desktop Grid QA

- Q. A Desktop Grid consists of otherwise unused computing sources. So it is for free?
  - A. It is much cheaper than any other solution, but of course if you use a computer it uses power and produces heat
- Q. Are Desktop Grids Green?
  - A. Yes: less computers have to be built.
- Q. So we do not need supercomputers anymore?
  - A. No. Although there are many applications that can run on a Desktop Grid, some require fast communication between the processors. For those supercomputers are more appropriate
- Q. It took me a long time as a user to get used to Glite within EGEE. Do I have to learn a new Grid technology
  - A. No. The EDGeS project provides a seamless bridge between EGEE and Desktop Grids

End of part I

EDGeS project background

- *Service Grids* have typically order 10.000 trusted processors; *Desktop Grids* have order 100.000 untrusted processors.
- Largest Service Grid: EGEE: 150.000 processor cores
- *Largest Desktop Grid*: SETI@Home: 2.100.000 hosts registered
- Service Grids have *untrusted application*s but trusted resource providers; Desktop Grids have *trusted applications*, but untrusted resource providers
- http://edges-grid.eu

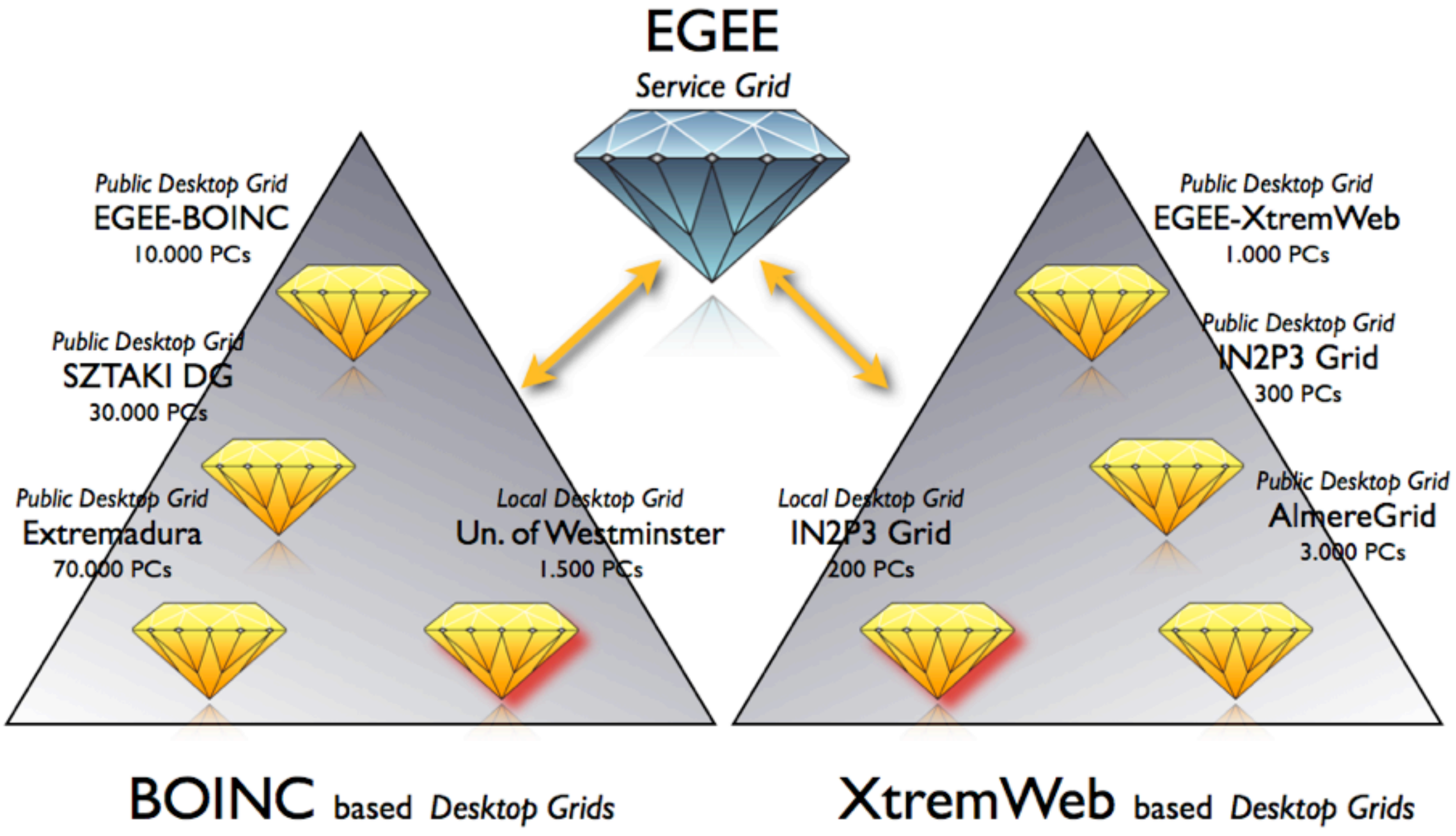
EDGeS -
Enabling Desktop Grids for e-Science

- Connect Service Grids (EGEE, BIG-Grid,…) with Desktop Grids (SZTAKI Desktop Grid, AlmereGrid, Extramadura Grid…, to provide even more resources to scientists
- Provide a Bridge between these types of Grids for automatic job sharing
- Provide an Application Development Methodology to port applications to the Grid; Port a number of (new) applications to the Grid
- Organise users and industry fora
- Two year, EU funded project (started 1-1-2008, ends 31-3-2010)
- Infrastucture continues and is supported by follow-on projects

EDGeS Partners

	Partner	*Participant short name*	*Country*
1	**Computer and Automation Research Institute of the Hungarian Academy of Sciences**	**MTA SZTAKI**	**Hungary**
2	Centro de Investigaciones Energéticas Medio Ambientales y Tecnológicas	CIEMAT	**Spain**
3	Foundation for the Development of Science and Technology in Extremadura	Fundecyt	**Spain**
4	The French National Institute for Research in Computer Science and Control	INRIA	**France**
5	University of Westminster	UoW	**UK**
6	Cardiff University	CU	**UK**
7	Faculty of Sciences and Technology of the University of Coimbra	FCTUC	**Portugal**
8	Stichting AlmereGrid	AlmereGrid	**The Netherlands**
9	Centre National de la Recherche Scientifique - Institut National de Physique Nucleaire et de Physique des Hautes Energies	IN2P3	**France**

EDGeS - Grid infrastructure



Typical scenario for EGEE users

EGEE

Service Grid

User

EDGeS

Application Repository

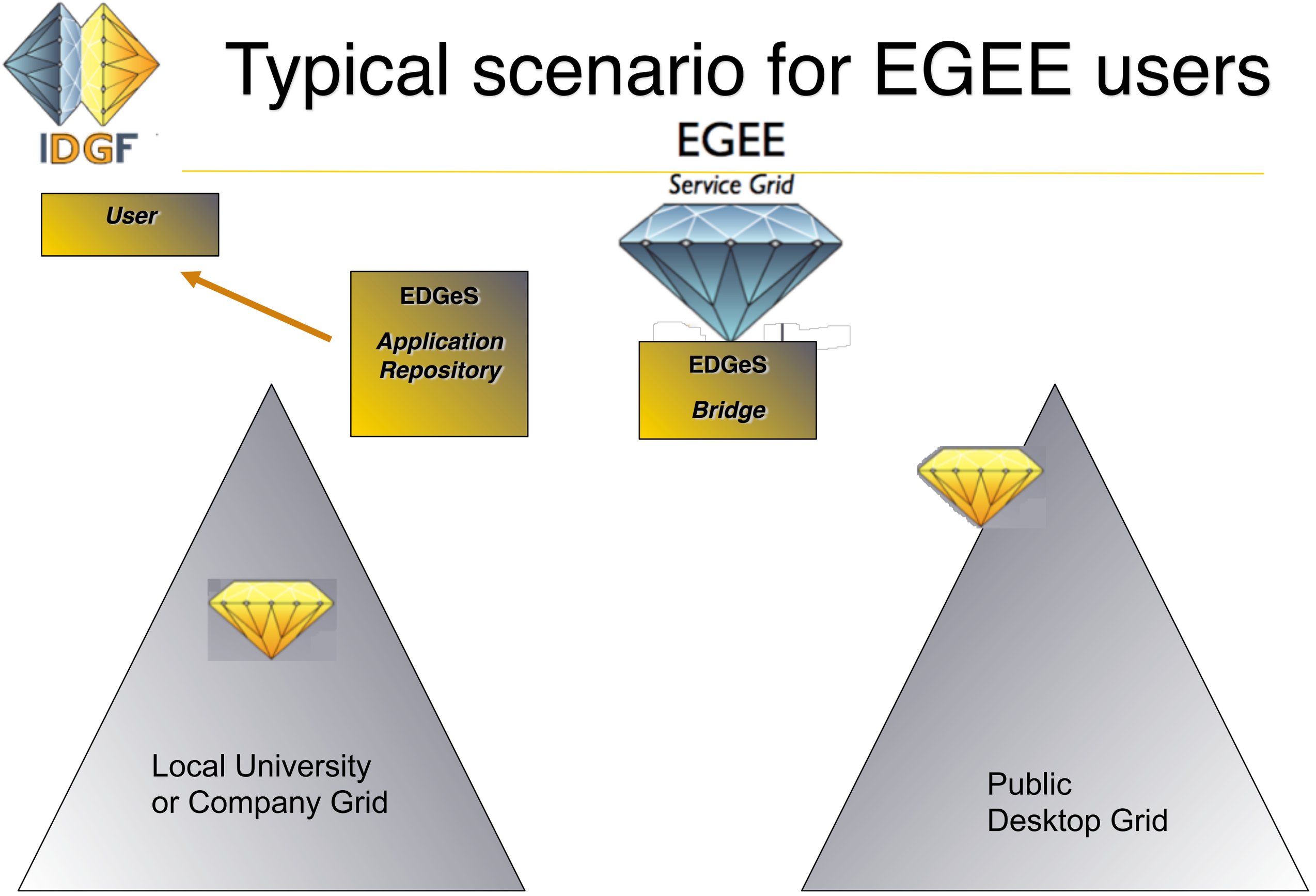
EDGeS

Bridge

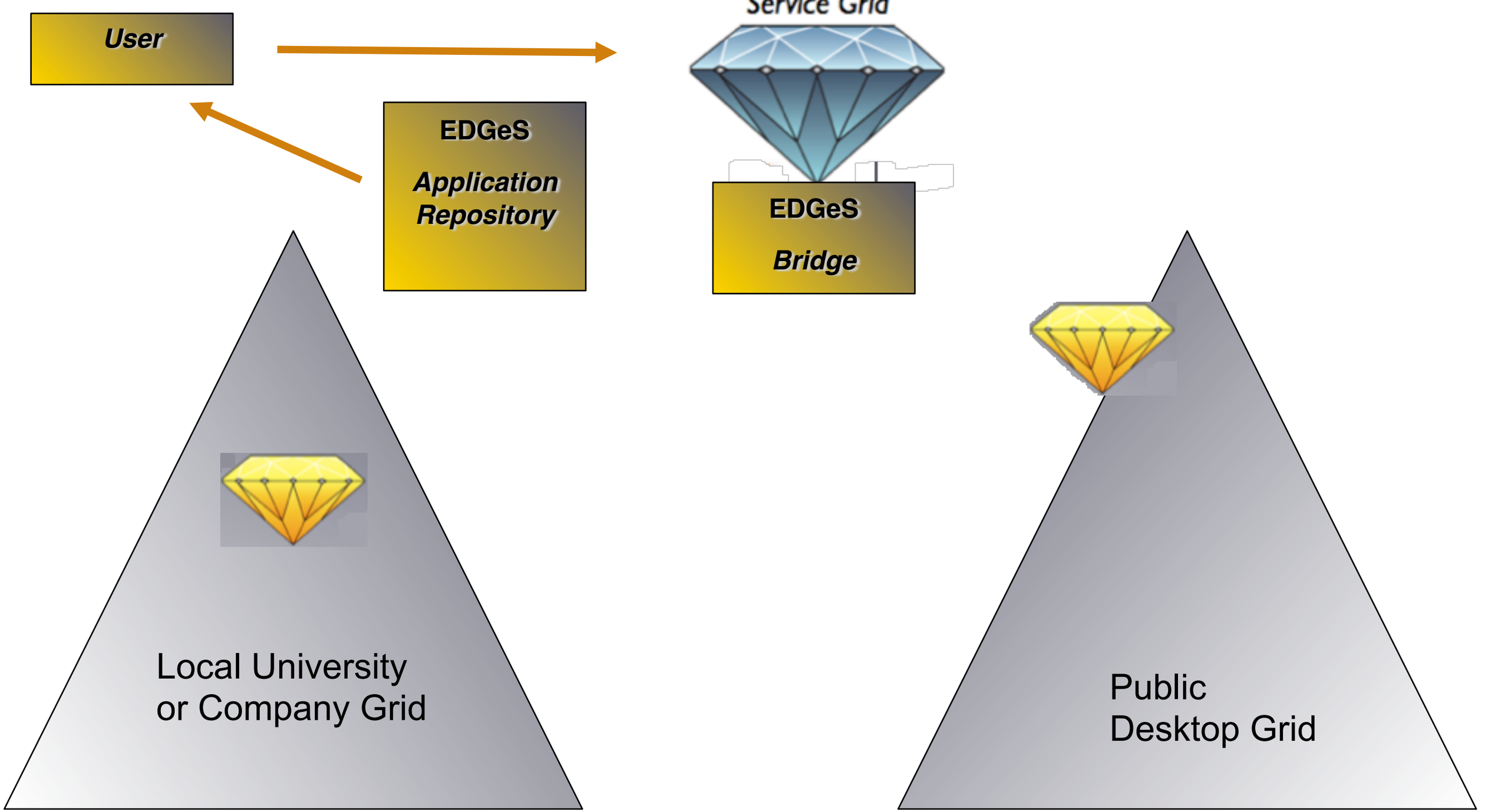
Local University
or Company Grid

Public
Desktop Grid

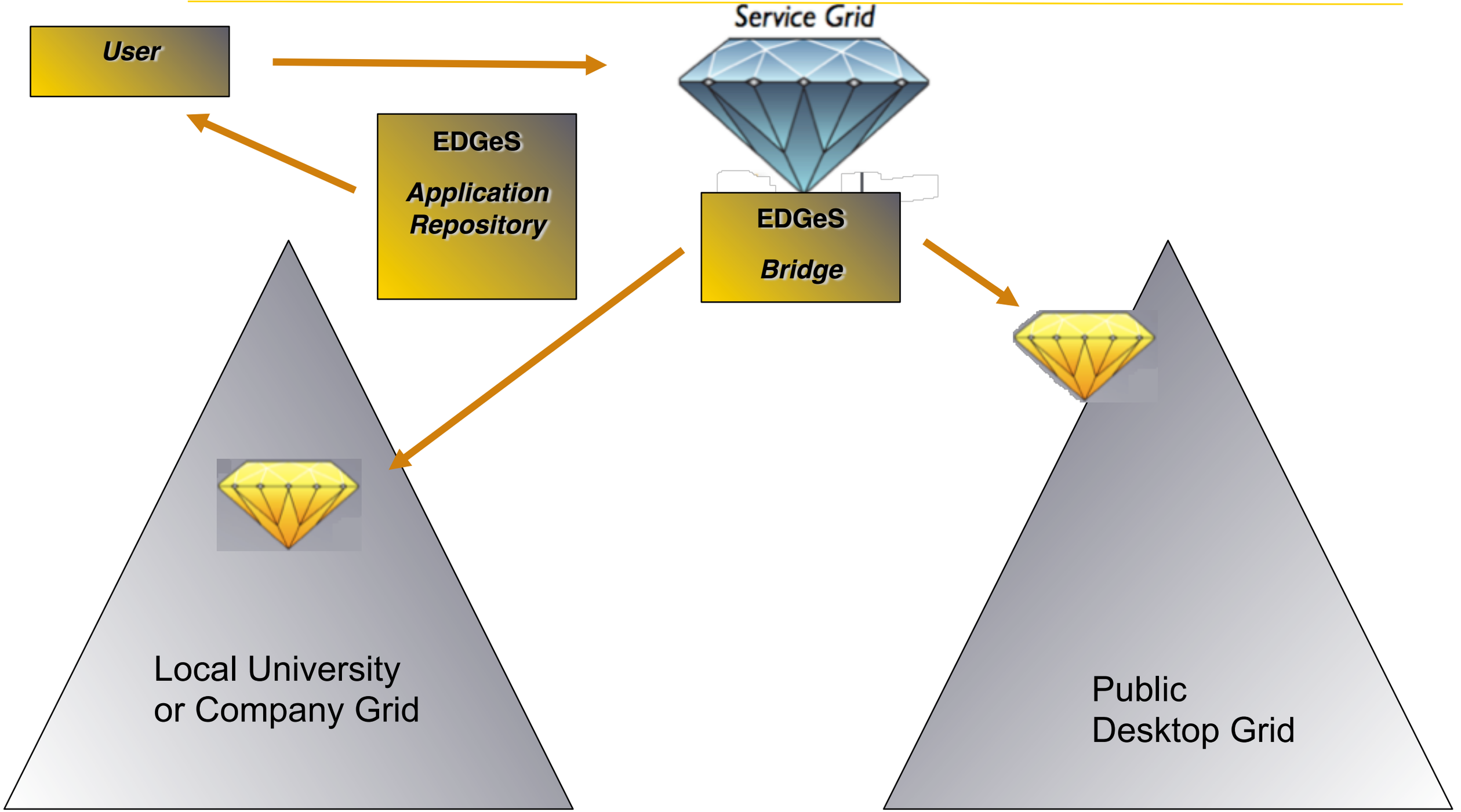
Typical scenario for EGEE users



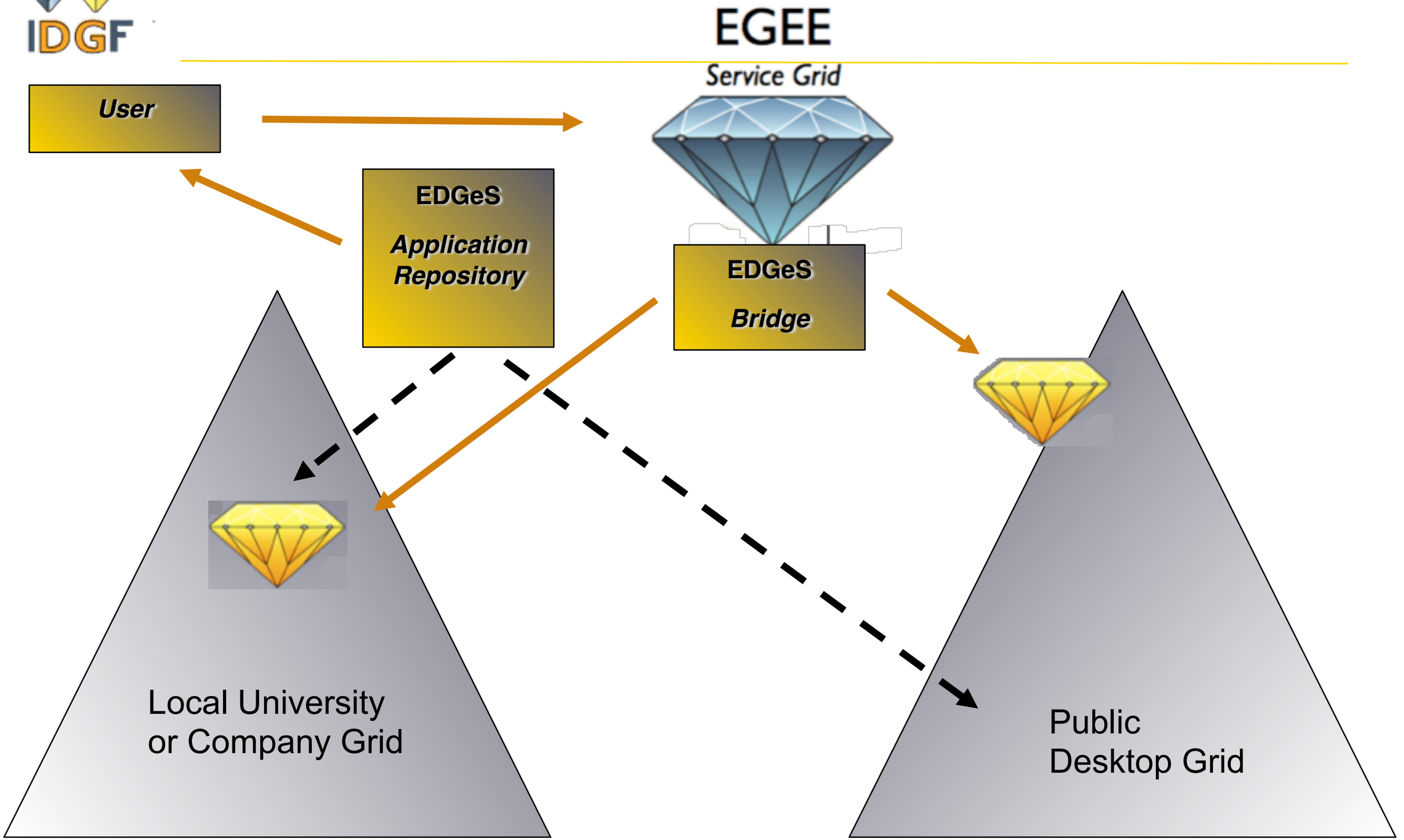
Typical scenario for EGEE users



Typical scenario for EGEE users



Typical scenario for EGEE users



For EGEE users wanting to run on Desktop Grids

- Bag-of-tasks type of EGEE applications are selected and ported to BOINC and XtremWeb
- Porting is a joint work of EDGeS and the application developer
- No certificate support in BOINC: all applications are validated by the EDGeS team
- Validated applications are placed in the EDGeS Application Repository
- Desktop Grids connected by EDGeS can register applications from the EDGeS Application Repository
- EGEE users can take applications from the Application Repository and submit via the EGEE->Desktop Grid bridges

For Desktop Grid users wanting to run jobs on EGEE

- EDGeS did create a new Virtual Organisation in EGEE (VO) called as *EDGeS VO*
- Sites in the EDGeS VO support the applications of the EDGeS connected Desktop Grids that are available in the *Application Repository*
- Jobs of the connected Desktop Grids can be seamlessly executed in the EDGeS VO

EADM - EDGeS Application Development Methodology

1. Analysis of current application

EADM - EDGeS Application Development Methodology

1. Analysis of current application
2. Requirements analysis
3. Systems design
4. Detailed design
5. Implementation
6. Testing
7. Validation
8. Deployment
9. User support, maintenance & feedback

Applications ported to the EDGeS Grid

Application	Organisation	Runs on Desktop Grid	Runs on EGEE (EDGes VO)
Video Stream Analysis in a Grid Environment (VISAGE)	Correlation Systems Ltd - Israel	√	√
Digital Alias-free Signal Processing	U. of Westminster	√	√
Protein Molecule Simulation using Autodock	U. of Westminster	√	
Patient Readmission Application	U. of Westminster	√	

Applications ported to the EDGeS Grid

Application	**Organisation**
Ultrasound computer tomography	Forschungszentrum Karlsruhe
Digital Alias-free Signal Processing	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

Illustration

Illustration

- EGEE IIII's *first* Memorandum of Understanding was signed with EDGeS
- Goals of the MoU:
  - Jointly select EGEE applications that should be migrated to EDGeS
  - Collaboration in porting these applications to EDGeS
  - EGEE support for the establishment and operation of the EDGeS VO
  - Organizing joint events (summer school, user forum meetings)
- Used as basis for a press release
- Mentioned in flyers targeted towards potential users/grid owners
- In presentations (like this one ☺)

Standardisation needs of EDGeS

- The EDGeS Bridges are there because of lack of standards
- Standards needed:
  - Authentication, authorization of users and applications. (To work in insecure and potentially hostile environments). Current schemes used in EGEE and the like are complicated.
  - Widely adopted jobs description language with not too many “profiles” *OGF JSDL* is a a good start. *OGF BES* could be usefull too, but not implemented for Desktop Grids yet.
  - Application repository. With possibilities to have approved unremovable trusted applications, There is *OGF ASC*. But that is not widely used.
  - We are also looking at OGF standards such as GLUE 2.0 (information on a Grid) and AUTZ (authentication on the Grid).
- EDGes participates in PGI: Production Grid Infrastructures standardisation working group

Desktop Grid User & Industry Forum

- The User Forum is intended for people who have applications that run, or could run on a Desktop Grid or a combined Desktop/Service Grid.
- The Industry Forum is intended for people from companies that are interested in setting up a local Desktop Grid (and port/run applications)
- The Fora provide:
  - Two meetings a year to meet face-to-face with the Desktop Grid developers and other Desktop Grid users
  - A directory with Desktop Grid resources. Grid news feed
  - Guidance with putting applications on the Grid. A Desktop Grid application development environment (2009)
  - Potential access to a European wide Grid with potentially millions of computer nodes

Desktop Grid User & Industry Forum

- The User Forum is intended for people who have applications that run, or could run on a Desktop Grid or a combined Desktop/Service Grid.
- The Industry Forum is intended for people from companies that are interested in setting up a local Desktop Grid (and port/running applications)
- The Fora provide:
  - Two meetings a year to meet face-to-face with the Desktop Grid developers and other Desktop Grid users
  - A directory with Desktop Grid resources. Grid news feed
  - Guidance with putting applications on the Grid. A Desktop Grid application development environment (2009)
  - Potential access to a European wide Grid with potentially millions of computer nodes

Will become: Global Desktop Grid Federation

http://desktopgridfederation.eu

Globe adapted from http://upload.wikimedia.org/wikipedia/commons/f/fa/Globe.svg

http://
desktopgridfederation.eu

Globe adapted from http://upload.wikimedia.org/wikipedia/commons/f/fa/Globe.svg