

**IDGF-SP**

**International Desktop Grid Federation - Support Project**

---

**Integration of European Grid Infrastructure with BOINC -  
latest steps and applications**

Peter Kacsuk and Jozsef Kovacs

MTA SZTAKI – LPDS

{kacsuk, smith}@sztaki.hu

IDGF-SP is to be supported by the FP7 Capacities Programme under contract nr RI-312297 .

# Why BOINC is not taken by the Grid community?

---

- David told:
  - The grid community ignored BOINC
  - The take of BOINC is not as wide spread as he has expected
- Indeed BOINC is very matured, proved to be useful in large grand challenge projects with even million volunteers
- In fact, this is the only technology that really enables volunteer computing
- So why isn't it as popular as we would expect?

# Why BOINC is not taken by the Grid community?



- The reasons:
  - To port an application to BOINC requires significant effort
  - To run a BOINC project you must become a BOINC expert
  - Originally BOINC was not designed to accept many different kind of submitted jobs (and this is what grids supported)
  - The response time of a volunteer grid is not as good as the one of a cluster grid (see the tail problem)
- Conclusions:
  - BOINC was designed to create long (even for years) running BOINC projects for a small number of grand challenge applications – and it is excellent for this

# Goal of our work

---

- Learning the lessons from the previous analysis the goal of our work was to enable the **mass usage of BOINC**:
  - To extend BOINC (without exchanging it) in order to enable
    - its dynamic, on-demand use (even only for days)
    - For a very large number of communities (even for individual researchers)
    - Without requiring any BOINC expertise
      - in porting applications
      - in set up and operate BOINC systems

# Goal of our work

---

- To enable the use of BOINC as an **HTC infrastructure** (like Condor, like many grids)
- To use BOINC as a **volunteer co-infrastructure**:
  - To enable the **extension of existing grid** and **supercomputer** infrastructures with BOINC to enable the **collection of cheap resources**
    - parameter sweep jobs should be **transparently transferred** to the BOINC co-infrastructure
    - where their execution is **much cheaper** due to the help of volunteers
- To enable the use of BOINC systems **via science gateways**

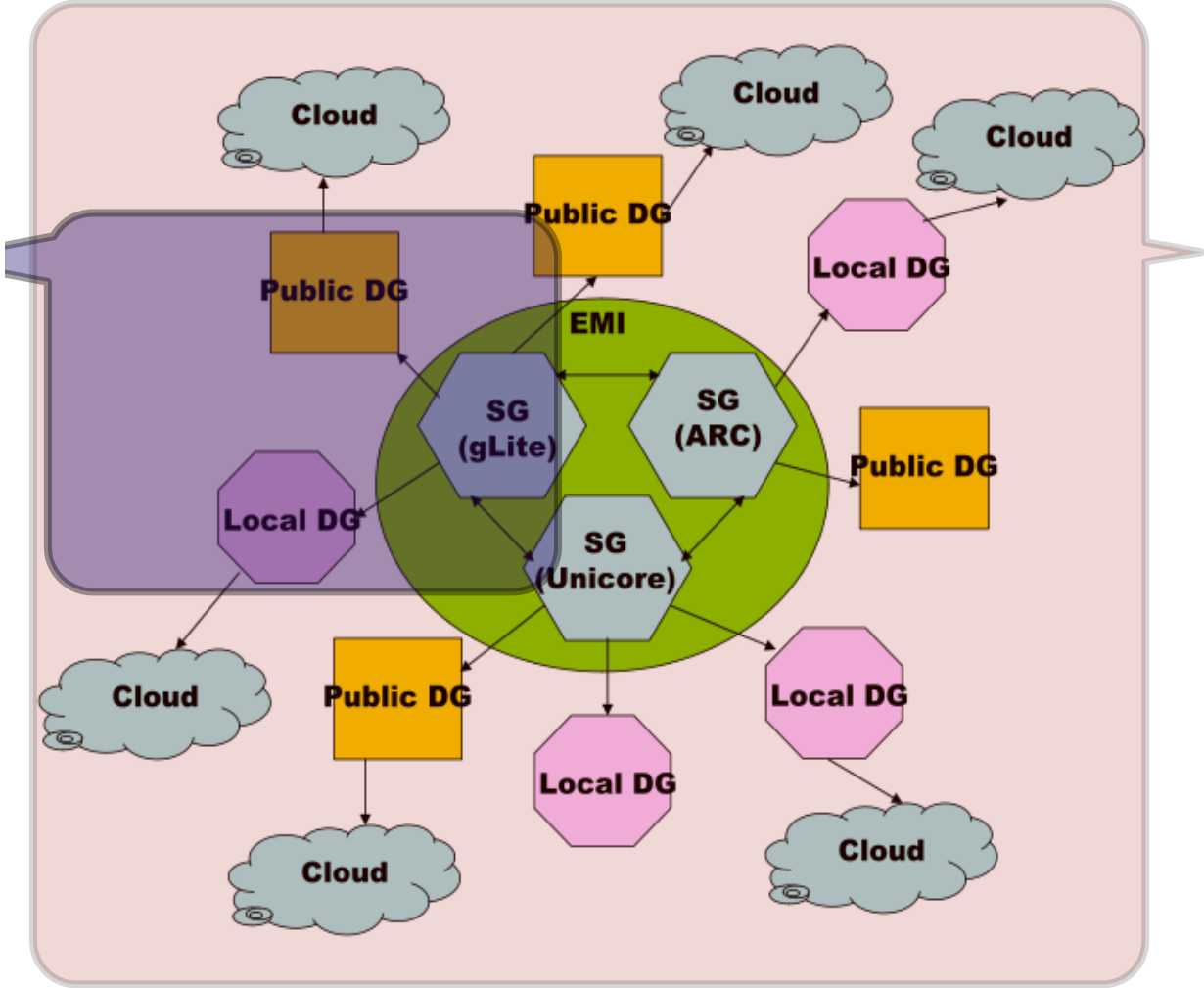
# Implementation of these goals

- **A series of EU projects** have been funded and successfully run
- Two technical projects (coordinator P. Kacsuk):
  - EDGeS: 2008-2010
  - EDGI: 2010-2012
- Two support action projects (coordinator R. Lovas):
  - DEGISCO: 2010-2012
  - IDGF-SP: 2012-2014
- **Result of these projects:** BOINC has been taken by many user communities for supporting the regular every day work of scientists and this approach getting more and more popular in Europe and elsewhere



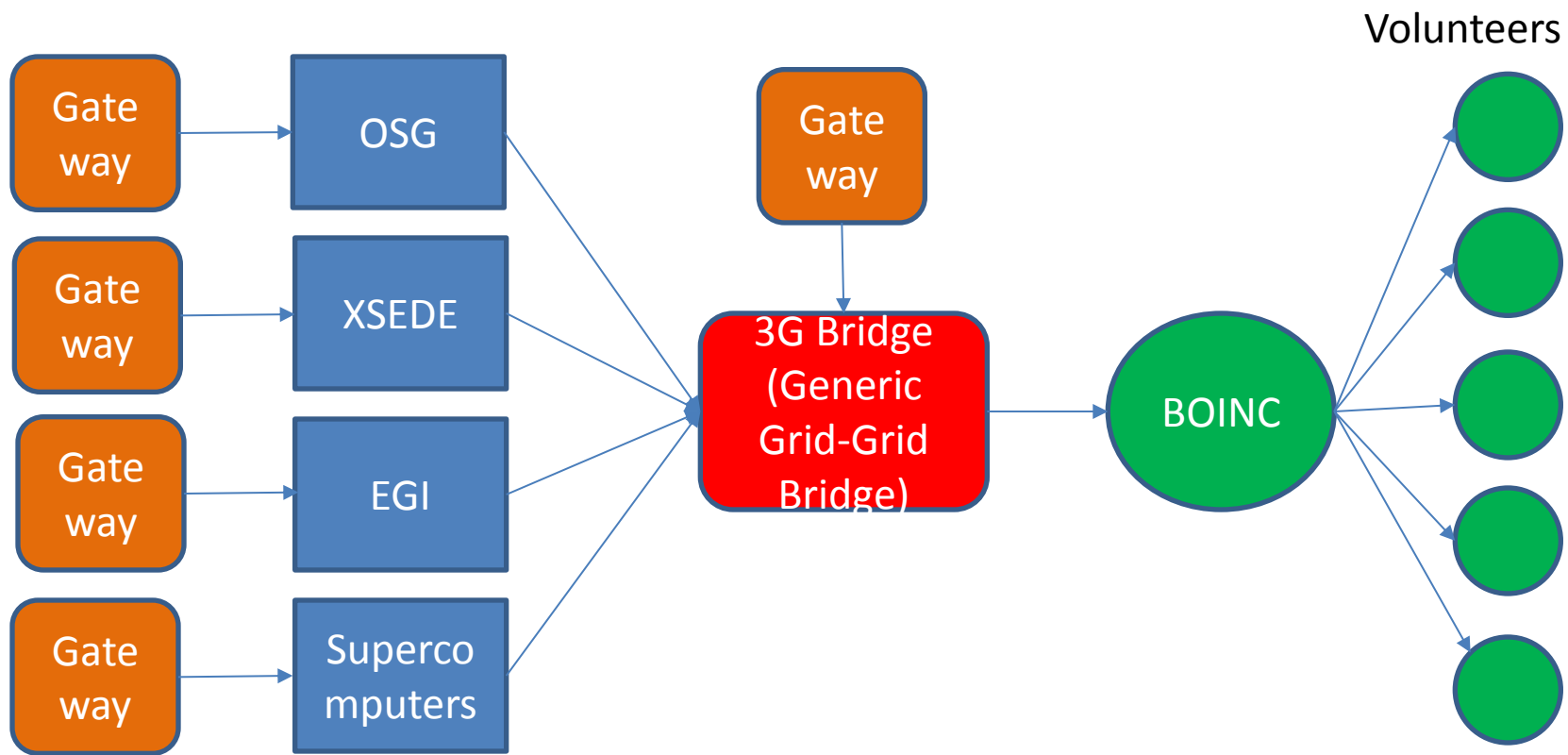
# Plans for EDGeS and EDGI

EDGeS scope  
for compute intensive applications  
for gLite



EDGI scope  
 for both compute and data intensive applications  
 for gLite, ARC, Unicore  
 Extend Desktop Grids with Clouds  
 Use virtualization

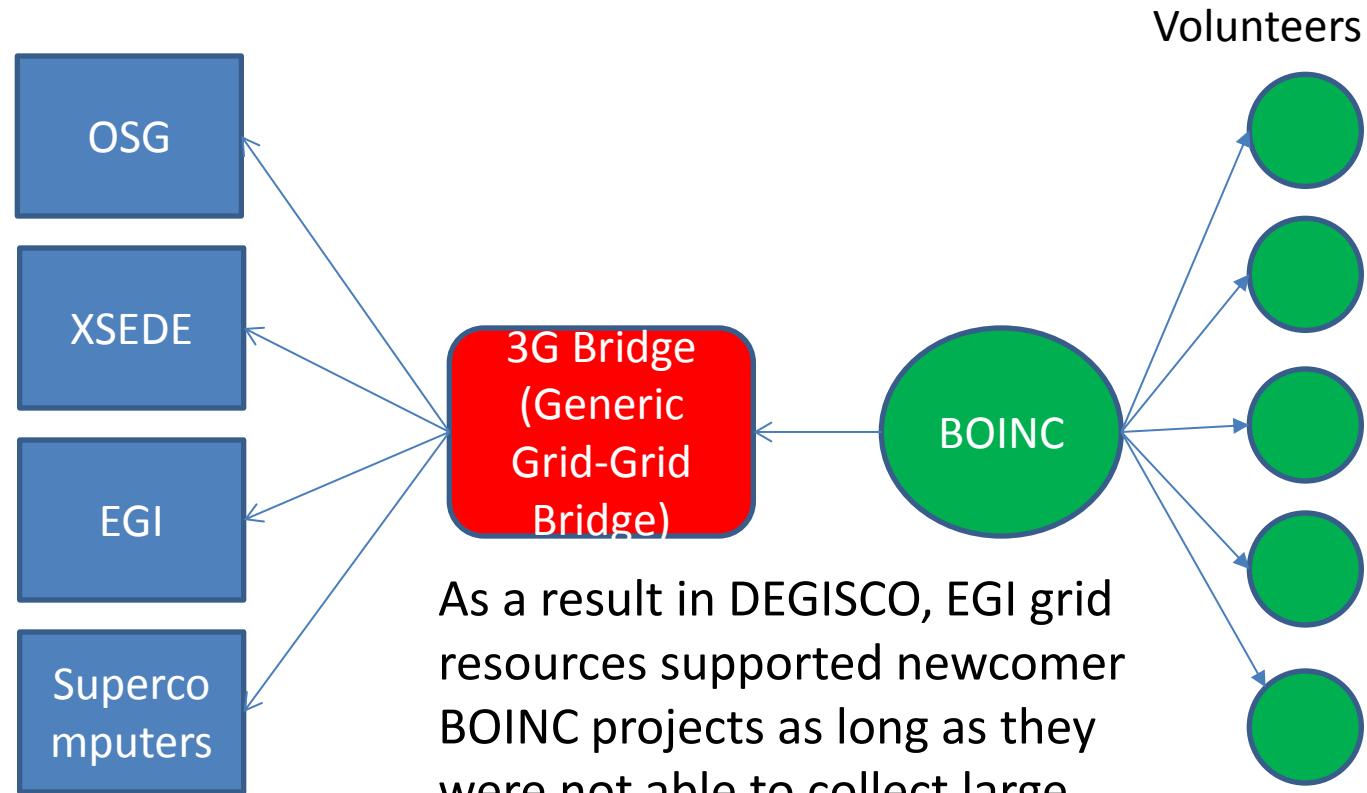
# Establishing the generic volunteer co-infrastructure concept and technical solution





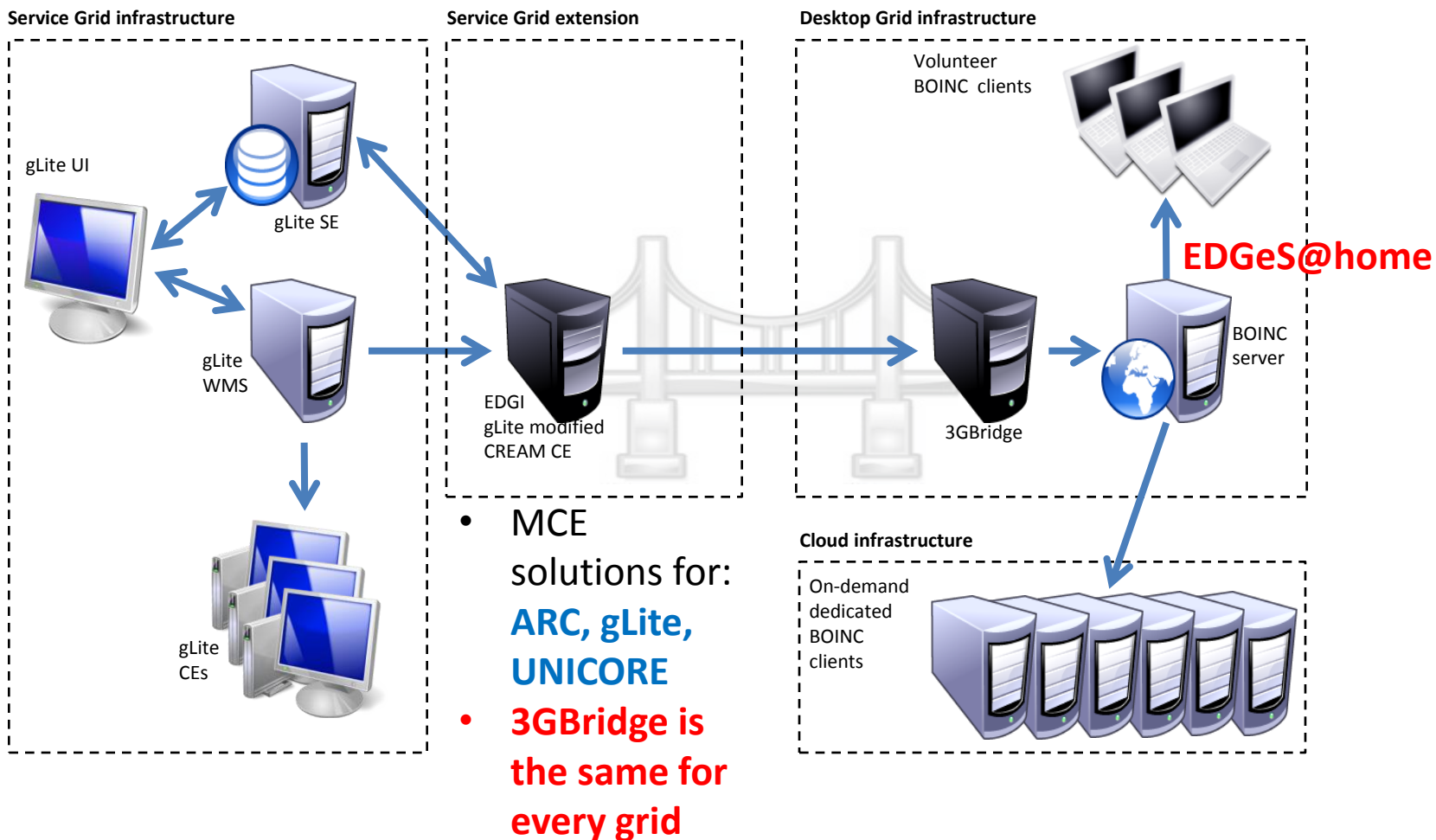
# Establishing the generic volunteer co-infrastructure concept and technical solution

The co-infrastructure concept works in the other direction, too.



As a result in DEGISCO, EGI grid resources supported newcomer BOINC projects as long as they were not able to collect large number of volunteer resources.

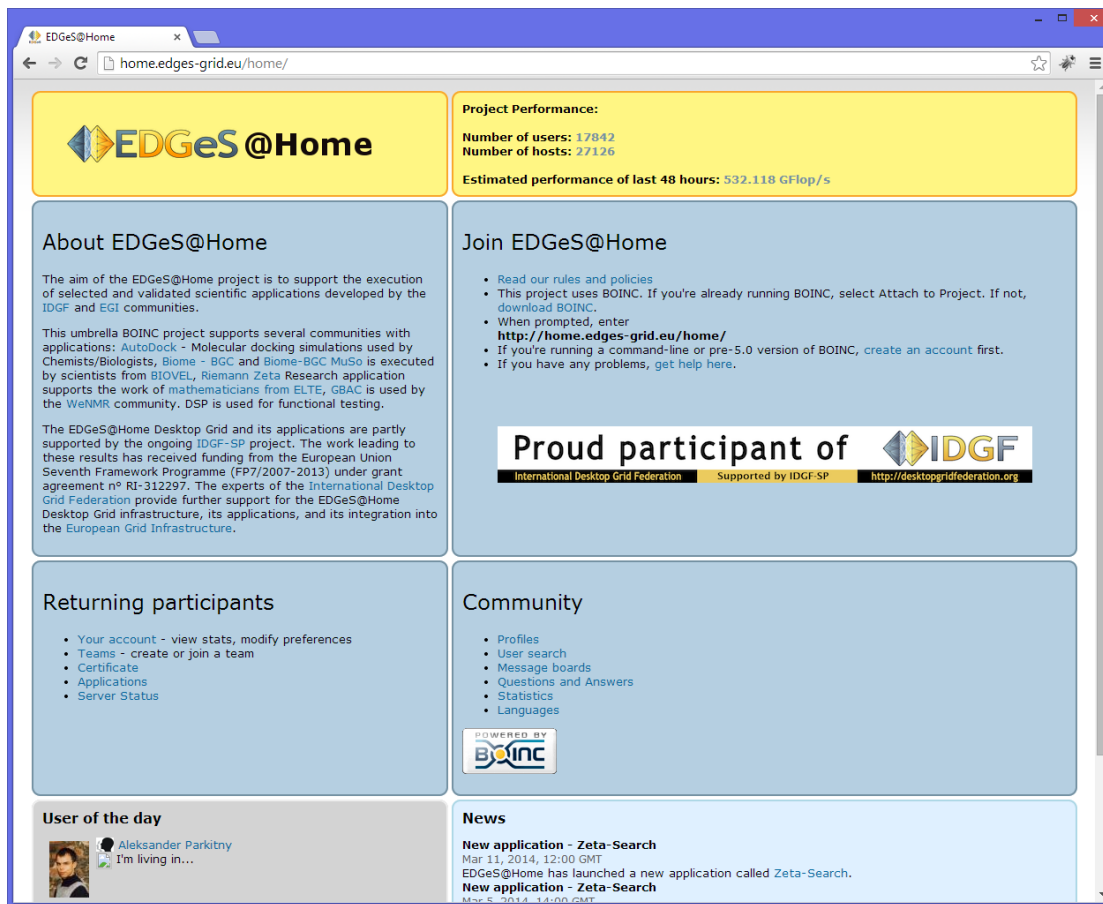
# Concrete solution for using BOINC as co-infrastructure of EGI grids



# EDGI solutions for show-stopping problems

- GBAC (Generic BOINC Application Client):
  - To avoid application porting
  - To run any applications on any type of client machines
  - By using virtualbox based virtualization
- Dedicated cloud resources as clients to solve the tail problem
- Result: EDGeS@home is now actively used by EGI user communities as
  - co-infrastructure or
  - HTC infrastructure

# Applications at EDGeS@home for EGI scientists



**EDGeS@Home**

**Project Performance:**  
 Number of users: 17842  
 Number of hosts: 27126  
 Estimated performance of last 48 hours: 532.118 GFlop/s

**About EDGeS@Home**  
 The aim of the EDGeS@Home project is to support the execution of selected and validated scientific applications developed by the IDGF and EGI communities.  
 This umbrella BOINC project supports several communities with applications: *AutoDock* - Molecular docking simulations used by Chemists/Biologists, *Biome - BGC* and *Biome-BGC MuSo* is executed by scientists from BIOVEL, *Riemann Zeta Research* application supports the work of mathematicians from ELTE, *GBAC* is used by the WeNMR community. *DSP* is used for functional testing.  
 The EDGeS@Home Desktop Grid and its applications are partly supported by the ongoing IDGF-SP project. The work leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° RI-312297. The experts of the International Desktop Grid Federation provide further support for the EDGeS@Home Desktop Grid infrastructure, its applications, and its integration into the European Grid Infrastructure.

**Join EDGeS@Home**

- [Read our rules and policies](#)
- This project uses BOINC. If you're already running BOINC, select Attach to Project. If not, download BOINC.
- When prompted, enter <http://home.edges-grid.eu/home/>
- If you're running a command-line or pre-5.0 version of BOINC, [create an account](#) first.
- If you have any problems, [get help](#) here.

**Proud participant of**   
 International Desktop Grid Federation | Supported by IDGF-SP | <http://desktopgridfederation.org>


**Returning participants**

- [Your account](#) - view stats, modify preferences
- [Teams](#) - create or join a team
- [Certificates](#)
- [Applications](#)
- [Server Status](#)

**Community**

- [Profiles](#)
- [User search](#)
- [Message boards](#)
- [Questions and Answers](#)
- [Statistics](#)
- [Languages](#)

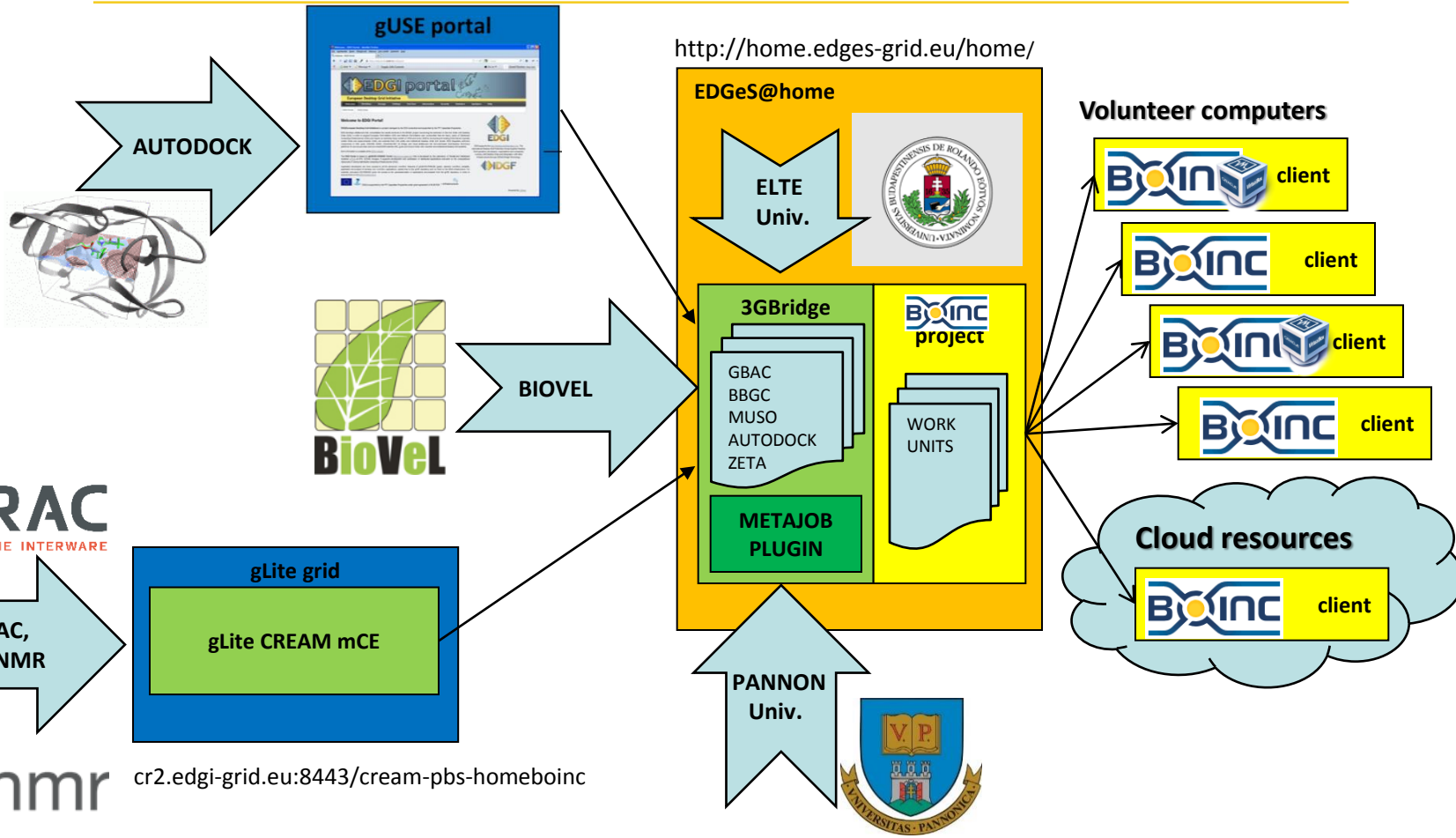
**POWERED BY** 

**User of the day**  
 **Aleksander Parkitny**  
 I'm living in...

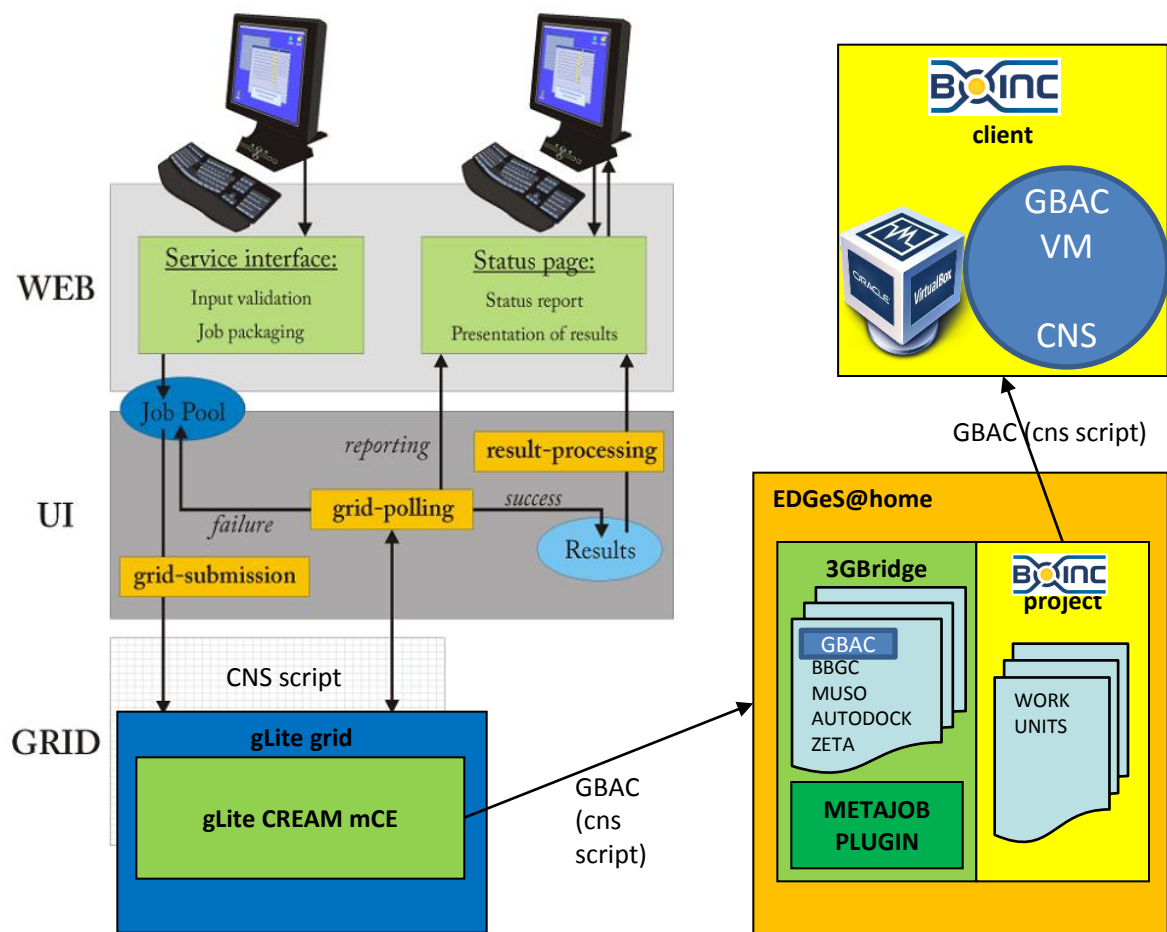
**News**  
**New application - Zeta-Search**  
 Mar 11, 2014, 12:00 GMT  
 EDGeS@Home has launched a new application called Zeta-Search.  
**New application - Zeta-Search**  
 Mar 5, 2014, 14:00 GMT

- **CNS [GBAC] (WeNMR)**
- **DIRAC [GBAC] (DIRAC)**
- **BBGC/MUSO (BIOVEL)**
- **Autodock (publicly available)**
- **Zeta-search (ELTE, Hungary)**
- **LinAlgOpt [GBAC] (Pannon, Hungary)**

# Overview of job submission alternatives for EDGeS@home



# Use case 1: submission through gLite (WeNMR/CNS, [DIRAC])



- Suggested for those who
  - want to skip **porting application** for Desktop Grid
  - prefer **gLite interface** instead of learning 3GBridge API
  - submit jobs in the **range of hundreds**

# Use case 2: submission through the gUSE portal (e.g. Autodock portal)

The screenshot displays the 'Import' page of the gUSE portal. The page header includes the gUSE logo and the text 'GRID USER SUPPORT ENVIRONMENT'. The user is logged in as 'József Kovács'. The page shows a table of application selectables with the following data:

Name	Notes	Exported by
<input checked="" type="radio"/> PublicAutoDock423	AutoDock 4.2.3 Application	10195
<input type="radio"/> PublicAutoDock423_noautogrid	AutoDock 4.2.3 Application without AutoGrid	10195
<input type="radio"/> Public_AutoDockVina112	AutoDock Vina 1.1.2 Application	10195

Below the table, there are input fields for renaming options:

- New Graph name:
- New Template name:
- New concrete Workflow name:

A 'Message:' field is also present at the bottom.

- Suggested for those who
  - prefer **customised web interface and/or workflows** with the easily customisable gUSE portal
  - require **higher abstraction** than jobs
  - submit jobs in the **range of (tens of) thousands**

# Use case 3: direct 3GBridge submission (BIOVEL)



**@BioVeL**

**@SZTAKI**

SZTAKI DESKTOP GRID

3G-BRIDGE

APP BIOME-BGC

Ecos-SERVER

- o INTERACTION PAGES
- o WEB-SERVICES

Biome-BGC Projects database

User login

Username:

Password:

Request new password

The Biome-BGC Projects Database HOME

The main aims of this Biome-BGC Projects database homepage are

- o to support developments of ecosystem modelling web services and workflows under BioVeL project (www.bioveel.eu); related scientific workflows can be found at myExperiment site.
- o to help to manage and run Biome-BGC models embedded in Taverna workflow environment or on
- o to support developments of enhanced new Biome-BGC model versions

Supported by:

MTA ÖKOLÓGIAI KUTATÓKÖZPONT

IDGF

Workflows

This page shows workflows that you can execute on BioVeL Portal

A workflow is the automation of series of data analysis (steps) to process data (from small to very large-scale), be that from one's own research and/or from existing sources.

Public

Register to upload your own workflows and to search for workflows on my experiment

Title	Author(s)	Uploaded	Updated	Operations
BioVeL Generic ENM workflow with interaction This workflow takes as input a file containing species occurrence points to cre...	Alan R. Williams, Renato De Giovanni, Vera Hernandez & Robert Kulawik	2013/07/25 08:16	2013/10/22 15:46	Run Download
[BETA] Data Refinement Workflow v14 This is a beta version of the Data Refinement Workflow and should be used only ...	Cherian Matthew, Vera Hernandez	2013/09/23 13:43	2013/09/23 16:23	Run Download
Biodem workflow with interaction This workflow takes as input a file containing species occurrence points to cre...	Renato De Giovanni, Alan R. Williams, Robert Kulawik	2013/10/08 07:59	2013/10/08 08:09	Run Download
Biome-BGC ESI Regulation, test & demo version 1.0 a bi-biome-BGC-cbo is a process-based biogeochemical model that can be used to ...	Dóra KRASSER, Péter ITTZEZ, Ferenc HORVATH	2013/08/12 10:29	2013/08/15 05:18	Run Download
BioVeL ESW - ENM Statistical Workflow with raster difference computation The ENM Statistical Workflow (ESW) allows the computation of the extent and int...	Sarah Boufad & Robert Kulawik	2013/09/18 16:44	2013/09/18 16:44	Run Download
blast_align_and_Tree This workflow accepts a protein sequence as input. This sequence is compared to ...	Katy Wolstencroft	2013/01/30 12:21	2013/09/06 09:30	Run Download
blast_and_Interprocan This workflow performs an NCBI blast at the EBI. It accepts a protein sequence ...	Katy Wolstencroft	2013/01/30 12:49	2013/12/04 11:11	Run Download
clustal_phylogeny This workflow accepts a ClustalW protein sequence alignment and produces a phyl...	Katy Wolstencroft	2013/01/30 13:17	2013/06/11 15:12	Run Download

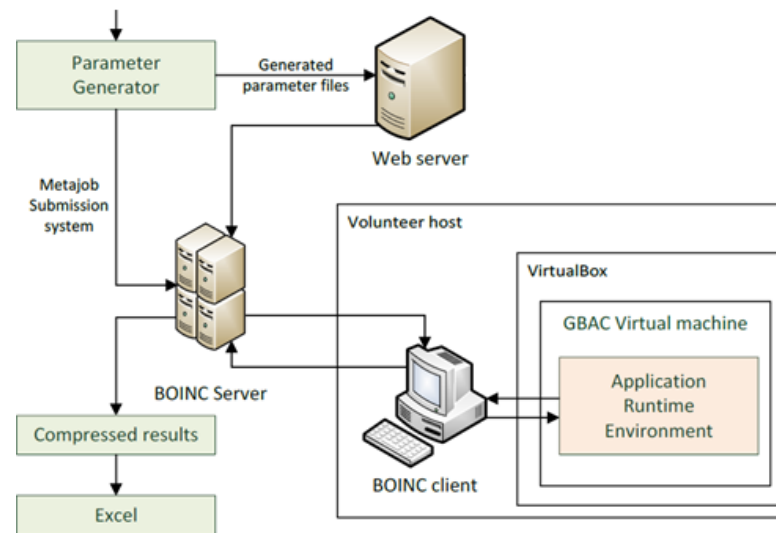
- Suggested for those who
  - prefer low-level interface integration to the desktop grid server
  - prefer hiding the single job submission with an **own portal**
  - submits jobs in the range of (tens of) thousands



# Use case 4: 3GBridge with Metajob (Pannon University, linear programming)



- **Parameter-sweep Metajob file** describes thousands of jobs
- **Submitted in one step**
- Can be combined with **GBAC**
- Batch level **directives** (e.g. stop at 90% for Monte Carlo type)
- **Results are downloadable in one step**



**LinAlgOpt** at E@H: trying to find optimal values for system solvers of Linear Programming problems by doing parameter sweep of a large number of run time parameters.

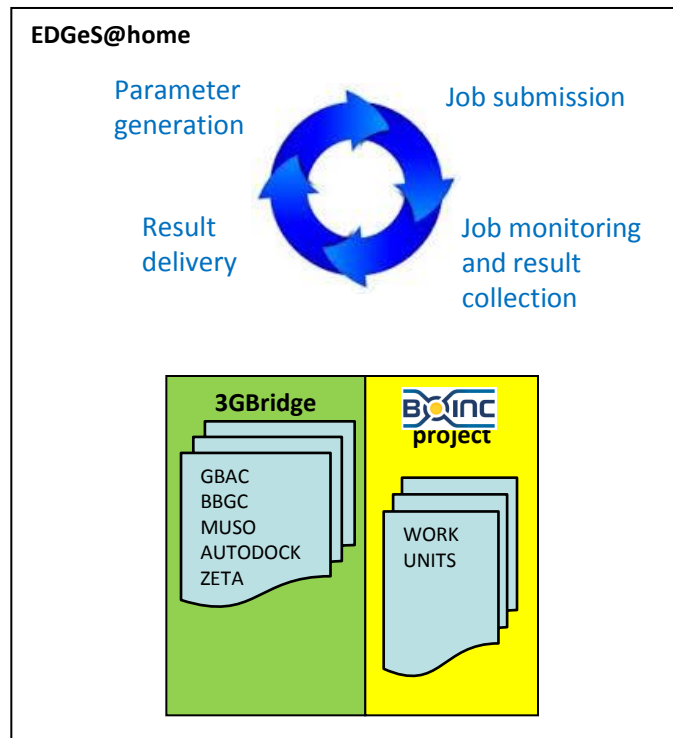
- **Suggested for those who**
  - prefer **low-level CLI interface** for job submission
  - prefer **batch submission** instead of individual job handling
  - submits jobs in the **range of tens of thousands**

More details: <http://doc.desktopgrid.hu>

# Use case 5: Automatic workunit generation and submission (ELTE)



- A huge **parameter space** is defined
- **Parameter generator** is executed as part of the infrastructure
- Jobs are automatically generated and **submitted**
- Results are **collected** (preprocessed if needed) and **sent** to the application owner



**Zeta-search** at E@H and SZDG: locating many values where  $Z(t)$  (Riemann-Siegel formula) is large in order to get a better understanding of the behavior of the distribution of primes, scans the numbers towards infinite

- **Suggested for those who**
  - have easily programmable parameter sweep application **running for years**
  - generate jobs in the **range of millions**



# Acceptance of the co-infrastructure concept by EGI: establishing IDGF (Regional) Operation Center



The screenshot shows the IDGF portal interface. The main content area displays the IDGF logo and a section titled "What is an NGI?". Below this, there are several data tables:

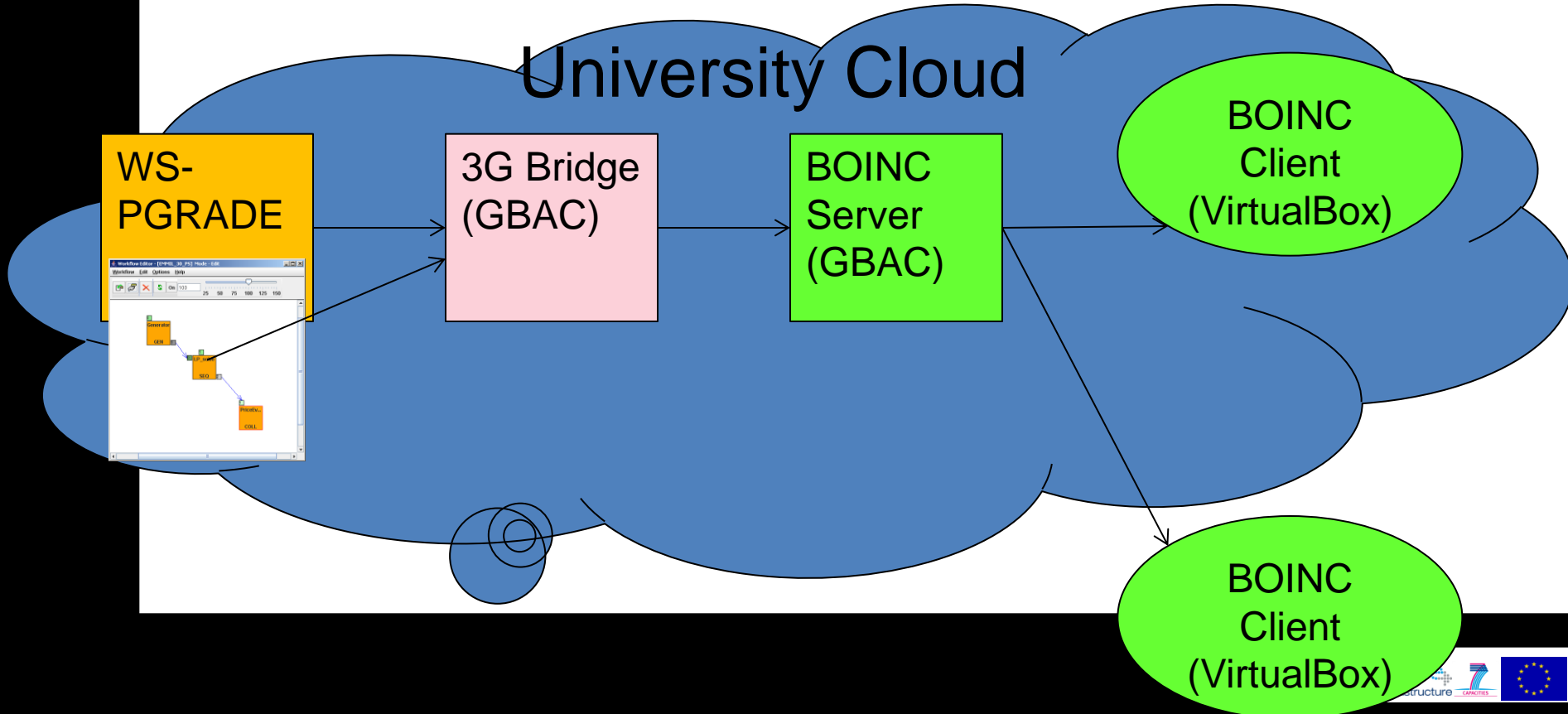
- Contacts:** A table with columns for E-Mail and IDGF support email addresses.
- Project memberships:** A section with a "Scope(s)" dropdown menu set to "Local".
- 2 Sites:** A table with columns for Name, Certification Status, Production Status, and Scope(s). The "EdgesAtHome" site is circled in red.
- 6 Users:** A table with columns for Name and Role, listing various staff members and their roles.

- IDGF OC has been established
- IDGF OC to collect Desktop Grid resources for EGI
- Each site represents a DG server
- So far two sites have been set-up
- EDGeS@home and SZDG

# Creating BOINC system with or without a job submission gateway on-demand in a cloud



- User can deploy by one click a BOINC infrastructure that is extended with a science gateway
- The BOINC infrastructure running in the cloud can be extended with ordinary home computers as in ordinary BOINC projects.
- Recommended to user communities having no BOINC expertise



# Summary: IDGF-SP core production infrastructure



## Supported VOs

vlemed  
 fusion  
 gilda  
 hungrid  
 seegrid  
 edgiprod.vo.edgi-grid.eu  
 chem.vo.ibergrid.eu

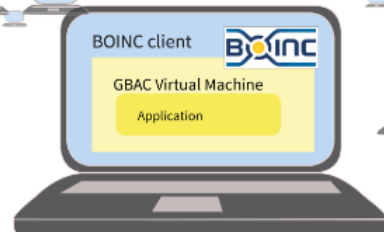
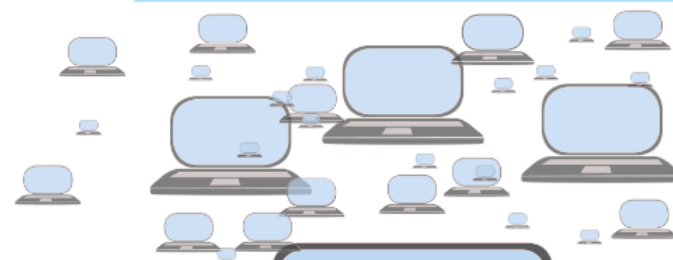
compchem  
 gaussian  
 trgrida  
 trgridb  
 biomed  
 enmr.eu  
 inaf



meta\_job

single\_job

## Volunteer Resources



ABC@home  
 AlmereGrid  
 Charity Engine  
**EDGEs@home**  
 EDGIDemo  
 SZTAKI Desktop Grid  
 Westminster Campus DG



### Desktop Grid Servers

Applications

- GBAC

Finished jobs through EDGI bridge from last 24 hours at UoW - by year

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
2007	0	0	0	0	0	0	0	0	0	0	0	0	0
2008	0	0	0	0	0	0	0	0	0	0	0	0	0
2009	0	0	0	0	0	0	0	0	0	0	0	0	0
2010	0	0	0	0	0	0	0	0	0	0	0	0	0
2011	0	0	0	0	0	0	0	0	0	0	0	0	0
2012	0	0	0	0	0	0	0	0	0	0	0	0	0
2013	0	0	0	0	0	0	0	0	0	0	0	0	0

Finished jobs through EDGI bridge from last 24 hours at EDGEs@Home - by year

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
2007	0	0	0	0	0	0	0	0	0	0	0	0	0
2008	0	0	0	0	0	0	0	0	0	0	0	0	0
2009	0	0	0	0	0	0	0	0	0	0	0	0	0
2010	0	0	0	0	0	0	0	0	0	0	0	0	0
2011	0	0	0	0	0	0	0	0	0	0	0	0	0
2012	0	0	0	0	0	0	0	0	0	0	0	0	0
2013	0	0	0	0	0	0	0	0	0	0	0	0	0



## AutoDock Portal

## Clouds

Amazon (on demand)  
 Westminster Cloud  
 SZTAKI Cloud  
 LPDS Cloud

# Summary

---

- The BOINC extension (called as SZTAKI Desktop grid) helped to make generally accept the co-infrastructure concept in Europe
- More and more user communities start to use the concept
- If this will generally be accepted then BOINC will be used by massive number of user communities as we planned

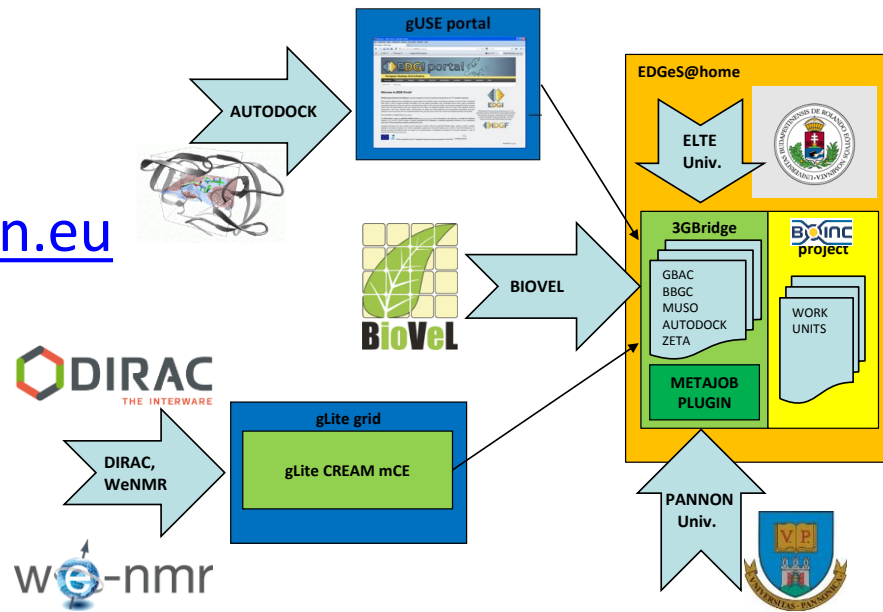
# Thank you for your attention!

Project websites:

<http://idgf-sp.eu>

<http://desktopgridfederation.eu>

<http://doc.desktopgrid.hu>



Acknowledgement:

- IDGF-SP EU support project (RI- 312297)

