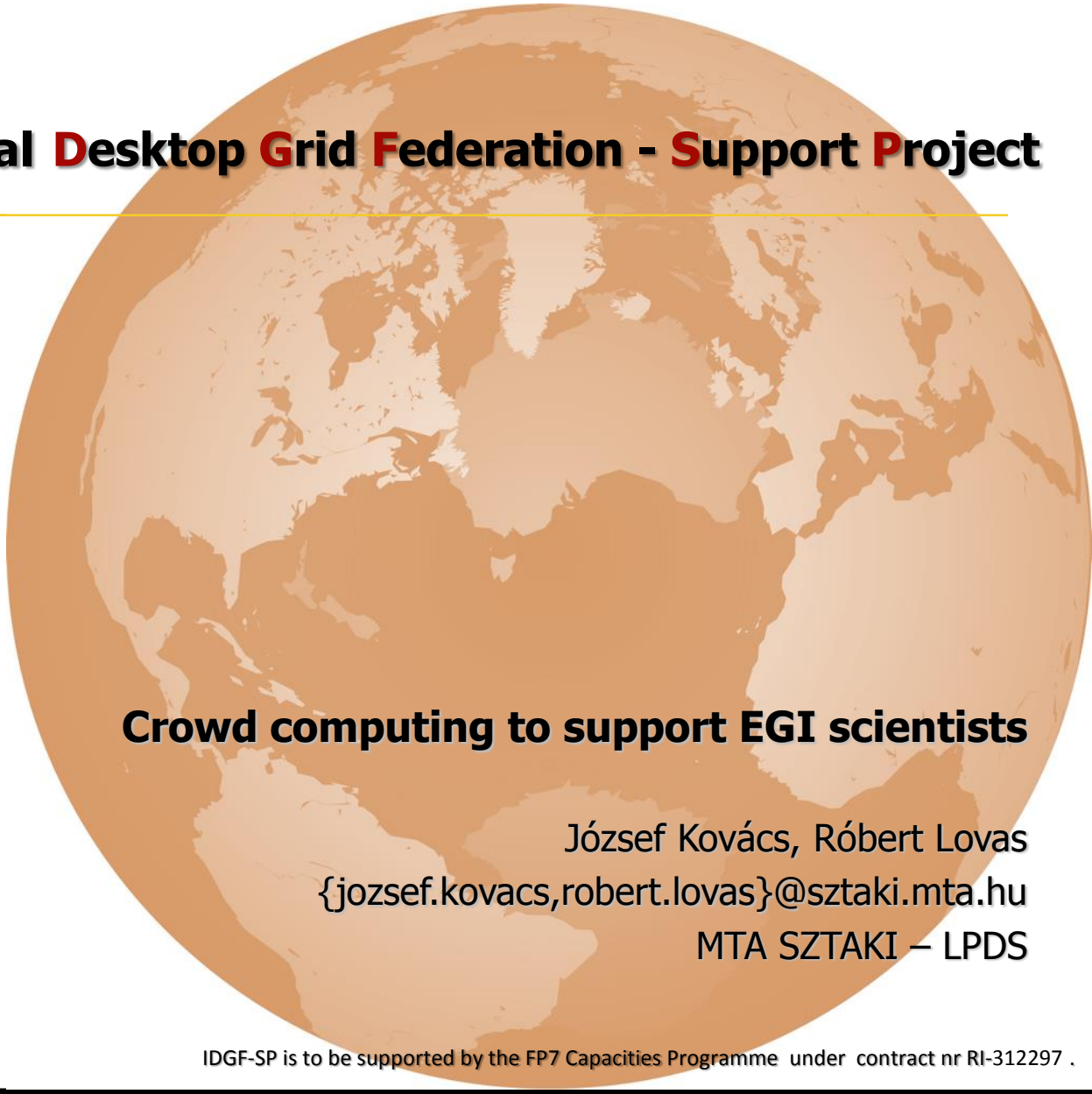


IDGF-SP

International Desktop Grid Federation - Support Project

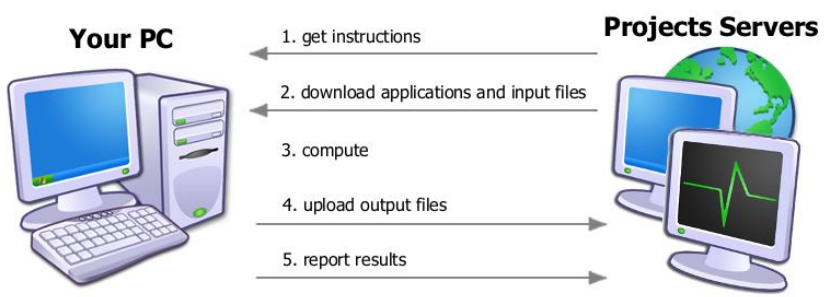
Crowd computing to support EGI scientists

József Kovács, Róbert Lovas
{jozsef.kovacs,robert.lovas}@sztaki.mta.hu
MTA SZTAKI – LPDS

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

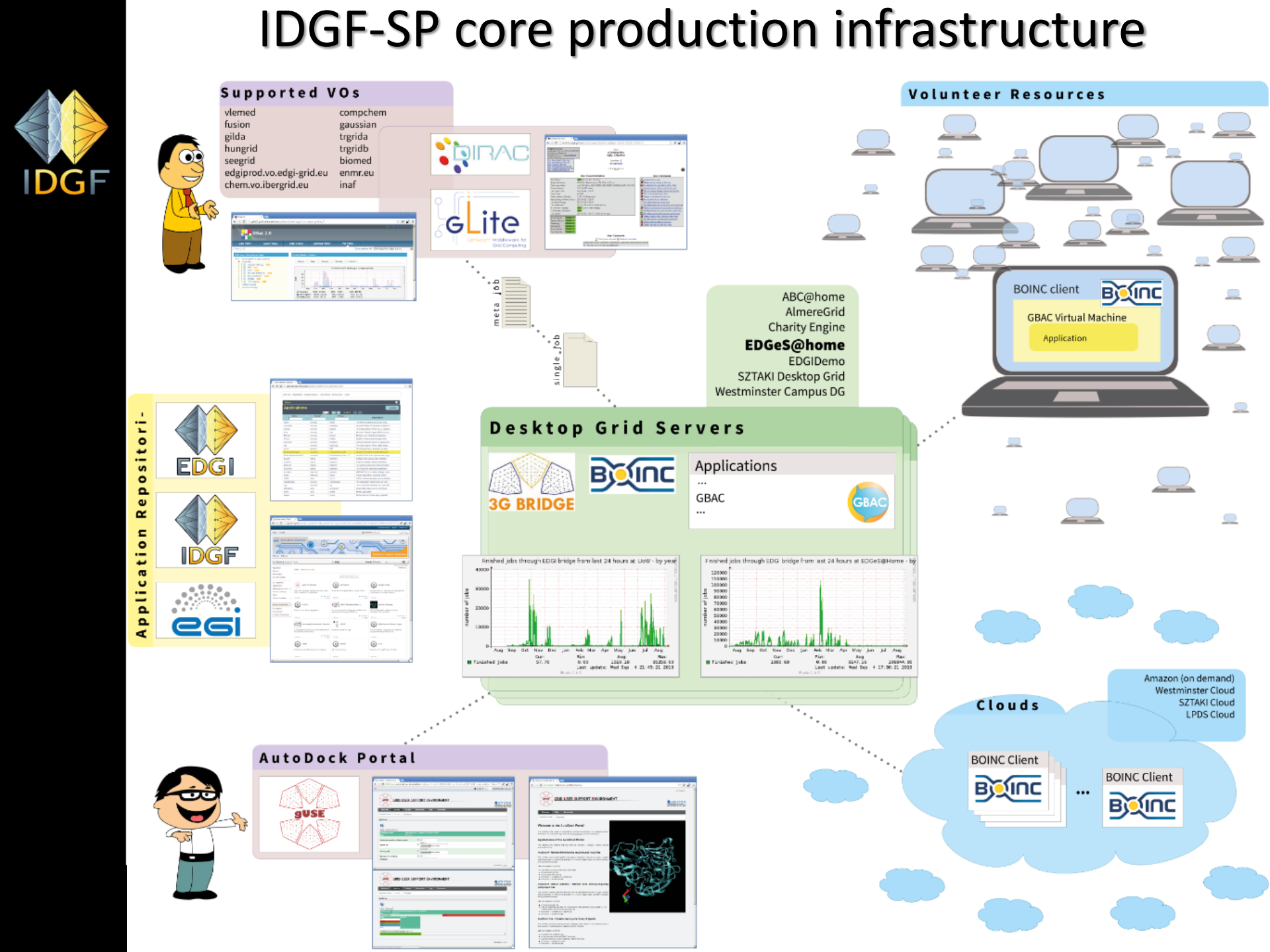
Volunteer and Crowd Computing

- **Volunteer computing** is an arrangement in which people (**volunteers**) provide computing resources to **projects**, which use the resources to do distributed computing and/or storage.
- The most widespread middleware to implent volunteer computing is **BOINC**.
- More than 1 billion PCs -> tablets, smart phones, game consoles

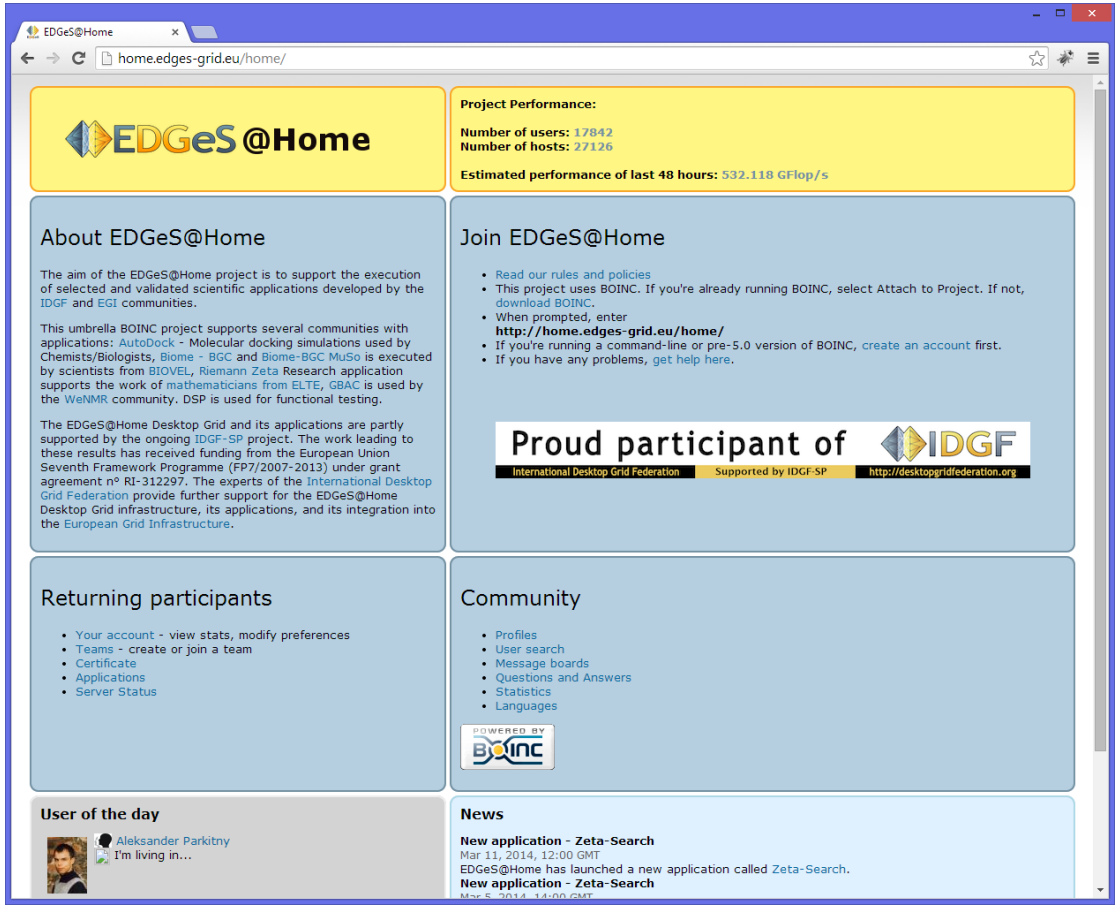


- **Crowd computing**: volunteer, but focusing on utilising **human intellectual capabilities** (e.g. recognition)

IDGF-SP core production infrastructure



Applications at EDGeS@home for EGI scientists



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

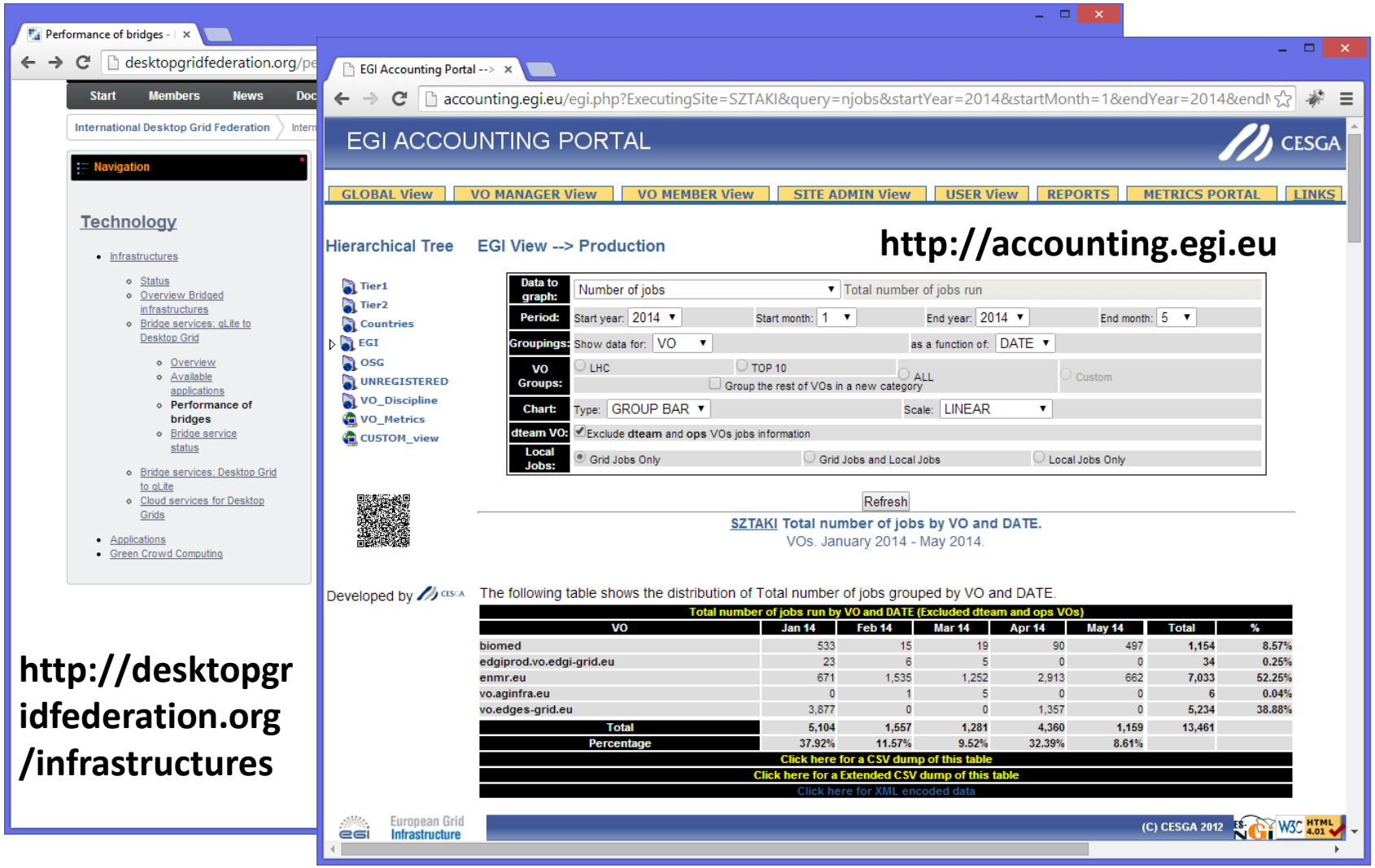
Promoting Desktop Grids Virtual Team in EGI

- **General project info**
  - **Start (planned) :** November 2013
  - **Duration :** 6 months **+ 3 months extension**
  - **Leader:** Robert Lovas
  - **URL:** http://go.egi.eu/vtdg

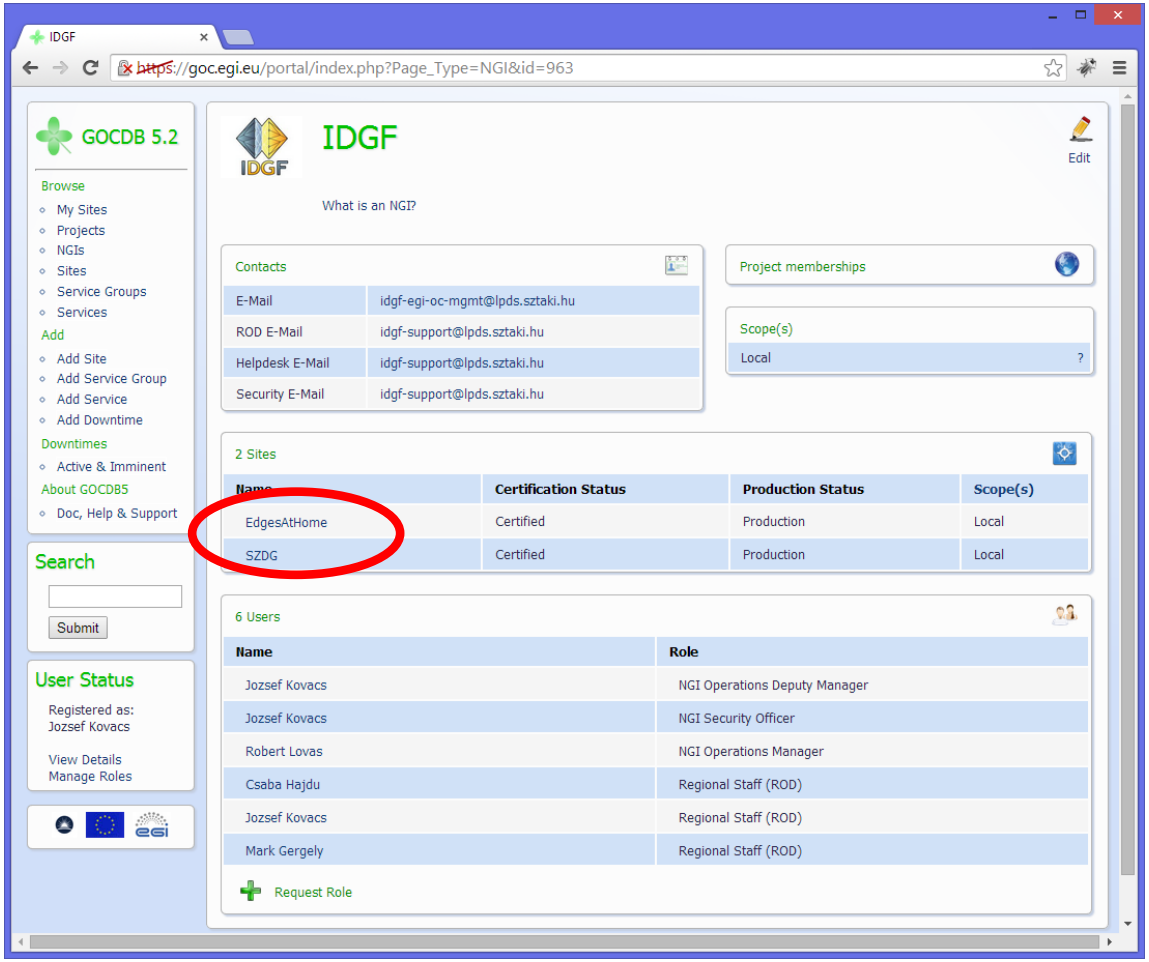
- *"EGI-InSPIRE is ideally placed to **join** together the new Distributed Computing Infrastructures (DCIs) such as clouds, supercomputing networks and **desktop grids**, for the benefit of user communities within the European Research Area."* --- EGI-InSPIRE Project fact sheet

- **Motivation**
  - Despite the on-going work and the benefits of the integrated Desktop Grid DCI resources available for EGI users, only **a part of the EGI community is aware** of the latest Desktop Grid related achievements, and a fraction of the EGI users and infrastructure operators take their advantages in everyday practice.
  - In order to fill this gap; a part of the targeted objectives are technical ones but there is stronger focus on the human aspects and **promotion** – training, networking and support activities.

Monitoring and accounting

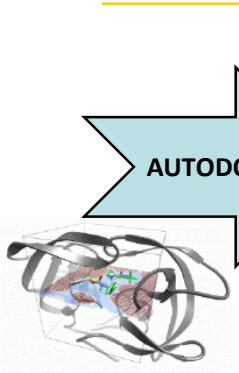


IDGF (Regional) Operation Center



- IDGF OC has been established
- IDGF OC to collect Desktop Grid resources for EGI
- Sites are being set-up
- One site represent a DG server
- EDGeS@home & SZDG

Overview of various utilisation of EDGeS@home volunteer resources



AUTODOCK

gUSE portal

BIOVEL

DIRAC, WeNMR

gLite grid

gLite CREAM mCE

cr2.edgi-grid.eu:8443/cream-pbs-homeboinc

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

EDGeS@home

ELTE Univ.

3GBridge

GBAC
BBGC
MUSO
AUTODOCK
ZETA

METAJOB PLUGIN

BOINC project

WORK UNITS

PANNON Univ.

Volunteer computers

client

client

client

client

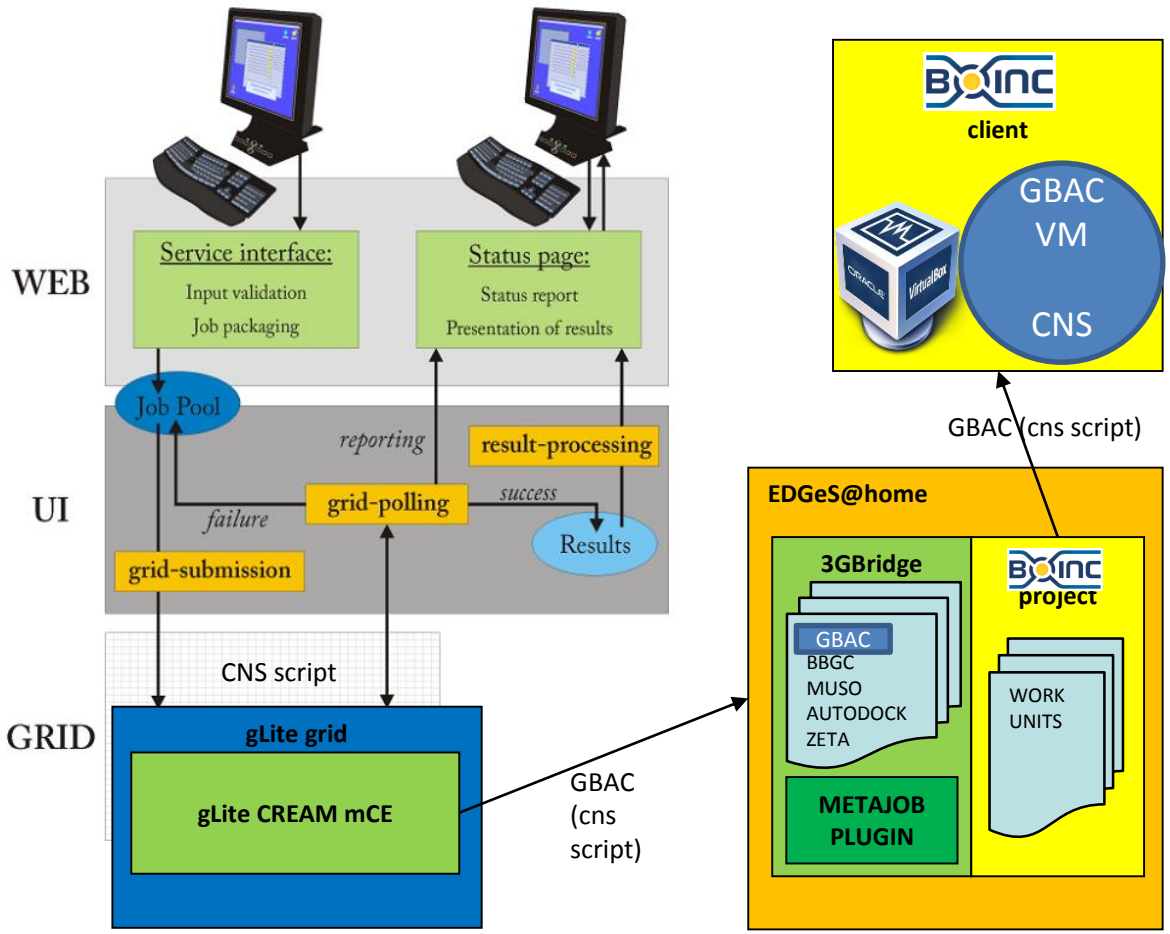
Cloud resources

client

FP7 WeNMR

- WeNMR (https://www.wenmr.eu/) is a **Virtual Research Community** supported by EGI, the largest one within the **life science area**. WeNMR aims at bringing together complementary research teams in the structural biology and life science area into a virtual research community at a worldwide level and provide them with a platform integrating and streamlining the computational approaches necessary for NMR and SAXS data analysis and structural modeling.
- WeNMR has a **portal and workflow system** that links and integrates into a variety of computing infrastructures. An important WeNMR application has been ported to Desktop Grids. Also **EDGeS@home** infrastructure has been integrated into WeNMR, it is in production phase.
- The coordinator of WeNMR project has joint the IDGF initiated **"Promoting Desktop Grid" Virtual Team** in the framework of EGI-InSPIRE project.

WeNMR HADDOCK portal: modeling of biomolecular complexes



Haddock portal is sending 1 out of 10 jobs to **EDGeS@home** desktop Grid through the modified computing element (Bridge).

Utilisation of volunteer resources with **minimal effort**: no porting, no windows version development

FP7 BioVEL

- BioVeL is a virtual e-laboratory that supports research on **biodiversity** issues using large amounts of data from cross-disciplinary sources. BioVeL offers the possibility to use computerised "workflows" (series of data analysis steps) to process data, be that from one's own research and/or from existing sources.
- BioVeL applications have been ported to Desktop Grids. BioVEL announced collaboration with IDGF-SP partner SZTAKI and support for EDGeS@HOME. IDGF became one of the "**Friends of BioVEL**".
- BioVeL : http://www.biovel.eu/
- Portal: http://ecos.okologia.mta.hu/bbgcdb/

Research

Updates on our Workflows

Achievements for ecosystem modeling workflows

By Ferenc Horváth, Zoltán Barcza, Péter Ittzés and Dóra Krasser

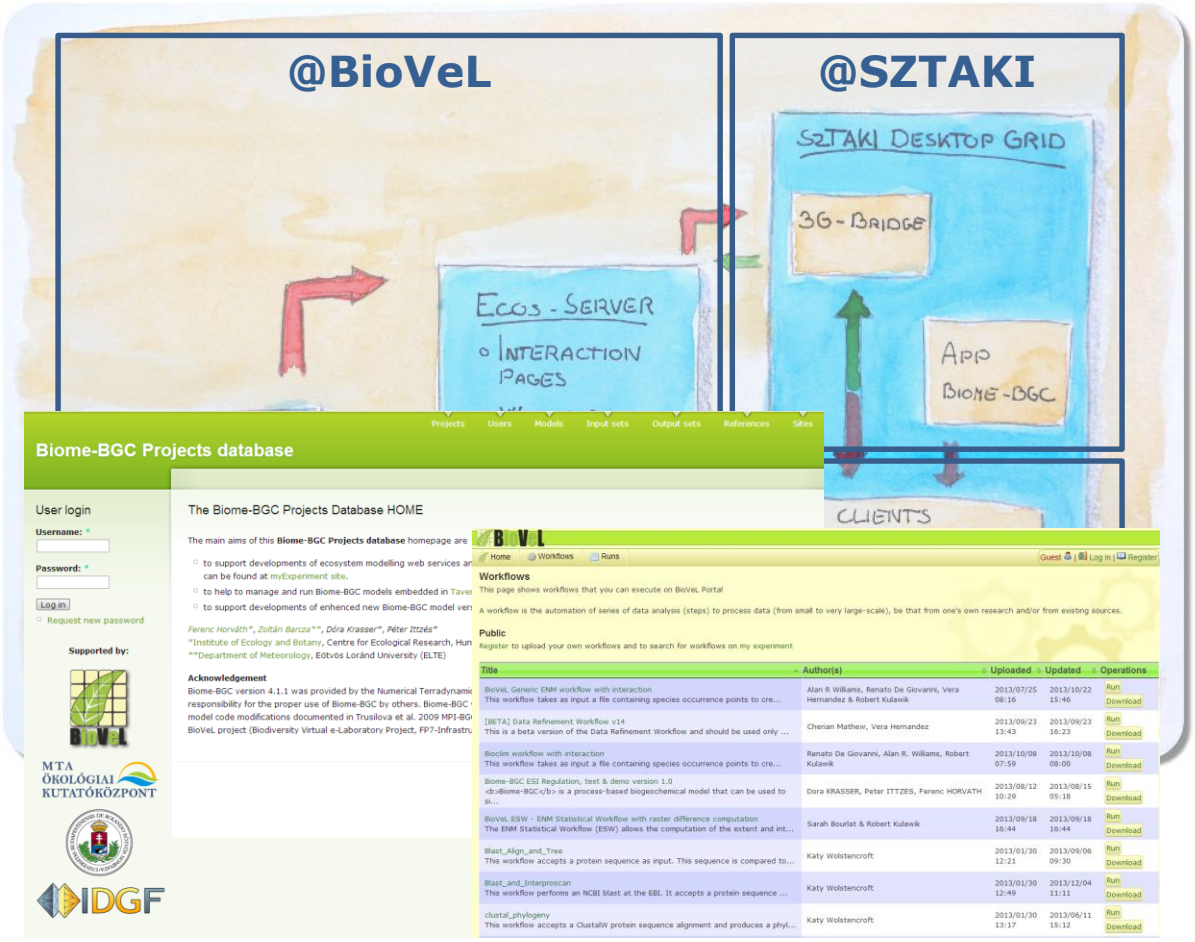
The Biome-BGC team has achieved considerable scientific and technological breakthroughs recently.

The cooperation with OpenNESS has resulted in a broad range of quantitative ecosystem service indicators (ESI) based on ecosystem simulation and the first Biome-BGC ESI workflow on the BioVeL Portal. Using these indicators (such as annual wood production; yearly maximum of total above ground biomass of grasslands —as biomass provisioning ESI; energy absorption by evapotranspiration —as micro and regional scale climate regulation ESI; damping of ecosystem daily water outflow —as hydrological cycle and water flow maintenance ESI, etc.) in real ecosystem simulations, different climate or land use change scenarios can be evaluated. These new developments were demonstrated at INTECOL 2013.

We are also proud to report that the full integration of Biome-BGC based web services and BOINC desktop grid technology was reached, and the 'Biome-BGC Ecosystem Modelling' application is running on the public EDGeS@home at beta level (see figure). Although single run simulations can be executed on the ECOS server, Monte Carlo experiment based sensitivity analysis, data-model harmonization and spatial extension workflows are computationally very demanding.

We also had a very fruitful collaboration with the team of the SZTAKI PERL and we welcome them as "Friends of BioVeL". (see page 4)

BIOVEL: ecosystem modeling on E@H



- Ecosystem modelling: **BBGC** and **BBGC-Muso**
- Both apps are driven by the **Biovel Web-Service** component
- Jobs are generated by **Taverna workflows**
- Taverna workflows are started by users through **BIOVEL portal**

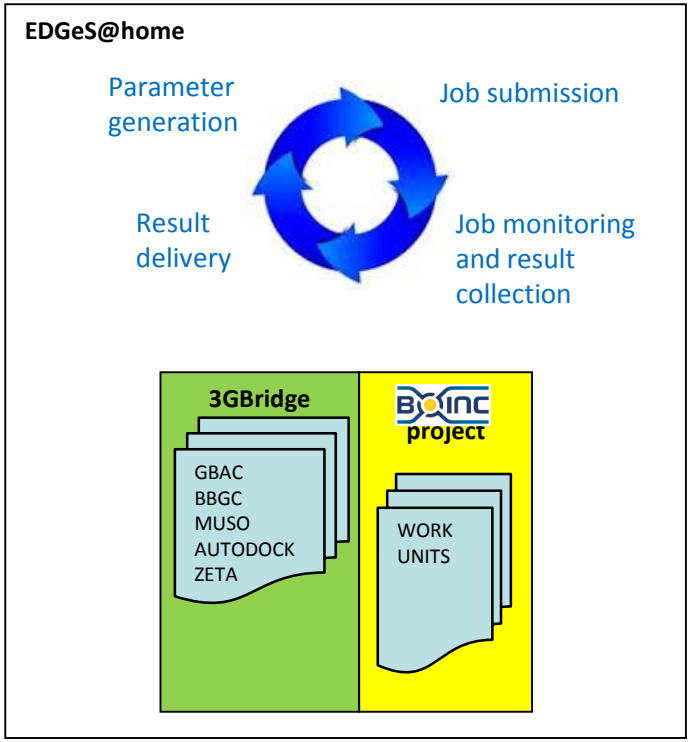
Zeta-search application by ELTE

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.

Application by the Eotvos Lorand University, Hungary

Zeta-search execution on EDGeS@home and SZDG

- 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 is an easily programmable parameter sweep application potentially **running for years**
- It generates jobs in the **range of millions**

http://riemann-siegel.com/

FP7 SOCIENTIZE + KOPI

- One of the main aims of the FP7 SOCIENTIZE project (http://www.socientize.eu) is to **promote the usage** of science infrastructures (dedicated and external resources), including professional and amateur scientists, and also sets up a network where infrastructure providers and researchers recruit volunteers from a **general public to perform science at home**.
- Some of the experiments from SOCIENTIZE are based on the *PyBossa* (pybossa.com) platform. PyBossa is an open source platform for crowd-sourcing online (volunteer) assistance **to perform tasks that require human cognition, knowledge or intelligence** (e.g. image classification, transcription, information location etc.)
- The **KOPI Online Plagiarism Search Portal** (http://kopi.sztaki.hu) is an open service that enables (among others) the **users to check for identical or similar contents between their own documents** and the files uploaded by other authors or available in Wikipedia, where KOPI uses the ***SZTAKI Desktop Grid*** to pre-process the immense database of Wikipedia.

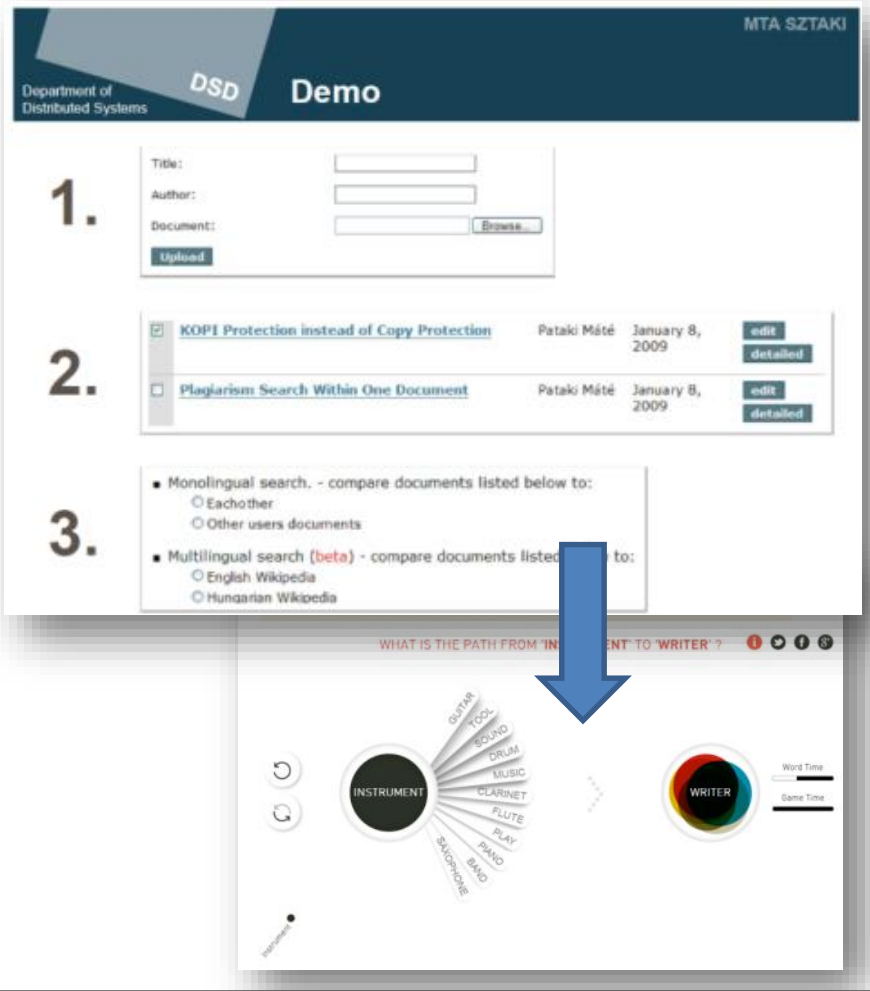
Conclusion

Volunteers (citizen scientist) can be classified in two ways:

- **Beneficiaries** (plagiarism service, improving teaching)
- **Active contributors** (uploaded documents, computer capacity, human intellectual effor)

Scientists

- are able to **harness the power of the general public**, and the
- citizens will get **more** information and **insight** view about the **science**



Thank you for the attention!

Project websites:

http://idgf-sp.eu

http://desktopgridfederation.eu

http://doc.desktopgrid.hu

József Kovács, Róbert Lovas
{jozsef.kovacs,robert.lovas}@sztaki.mta.hu

Acknowledgement:

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