



Enabling Grids for E-science

Въведение в Грид системите и в проекта EGEE

*Димитър Тодоров, ИПОИ – БАН
dttod@acad.bg*

*Използвани са материали на
Mike Mineter, Training Outreach and Education
National e-Science Centre, UK
mjm@nesc.ac.uk*

www.eu-egee.org

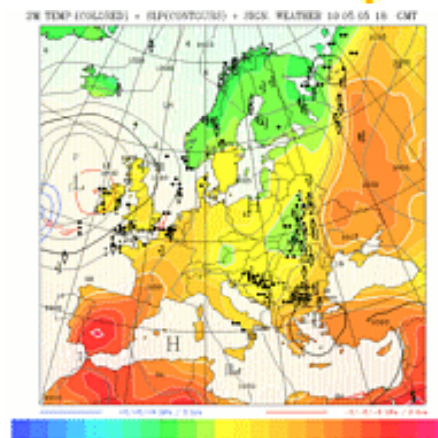
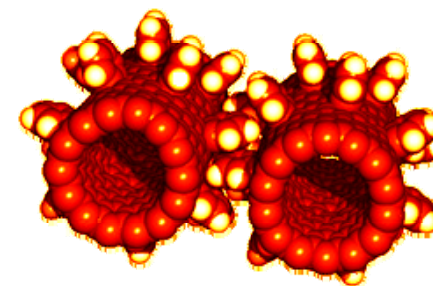


- **Въведение в**
 - е-изследванията и е-науките
 - Грид изчисления
 - е-инфраструктурата
- **Някои примери**
- **Грид концепции**
- **Повече за проекта EGEE**
- **Общо за мидълуера и услугите в EGEE Грид**

- Някои от важните научни области, при изследването на които често се надвишават **ресурсите на научните общности**, са:
 - Фундаменталните свойства на материята – нанотехнологиите
 - Генетика
 - Изменението на климата
 - Медицинска диагностика
- Изследванията в тези области са предимно върху цифрови данни, които са с **огромни и увеличаващи се обеми**
- **Обработката на тези данни** изисква все по-голяма изчислителна мощ

Пример: експерименталните науки използват все по-сложни сензори, които

- Генерират огромни количества от данни
- Достъпни са на учените по целия свят
- Стимулират международни сътрудничества



- Споделянето на ресурси (данни, инструменти, изчисления, човешки опит ...) през интернет води до възможност да се постигне сътрудничество в научната дейност, което се характеризира с
 - Размиване на границите между организациите
 - Много интензивни изчисления
 - Много големи обеми от данни
 - Много големи мащаби - многостранно и многообразно
- Първите примери са в точните науки - терминът **e-Наука (e-Science)** представлява провеждането на научна дейност с висока интензивност на изчисленията, която се изпълнява в разпределена мрежова изчислителна среда и която често е свързана с огромни информационни масиви
- Отнасянето на технологиите на **e-Науката** към нови научни области (социални науки, изкуства, хуманитарни науки и т.н.) доведе до възникване на по-универсалния термин **e-Изследвания (e-Research)**

**Сътрудничество
във “виртуалните
изчисления”**

Общи данни, компютри, софтуер - стават възможни чрез Грид структури:

- национални,
- регионални,
- международни: EGEE грид

**Импровизирани
сътрудничества**

Ел. Поща, Обмен на файлове, ssh достъп за изпълнение на програми - стават възможни чрез компютърните мрежи:

- национални,
- регионални и
- международни: GEANT

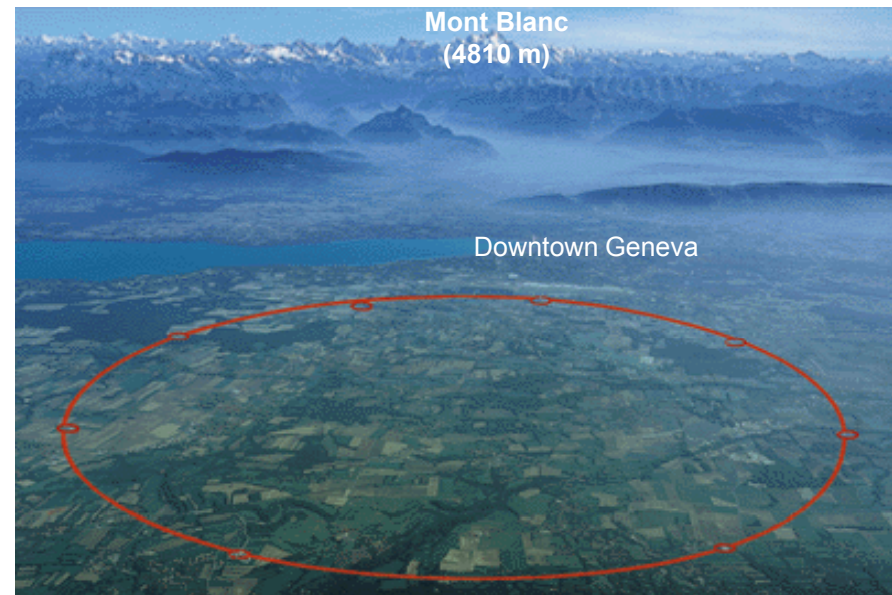
Хора с общи цели

- **Компютърни мрежи + Грид**
 - Мрежите свързват изчислителните ресурси
 - Грид предоставя възможност за реализиране на “виртуални изчисления” – споделено ползване на географски разпределени ресурси между различни административни единици
 - “административни единици”: институти; страни, където е разположен ресурса; процеси за управление на системите,...
- **+ Функциониране, поддръжка, обучение ...**
- **+ Изчислителни центрове, електронни архиви...**

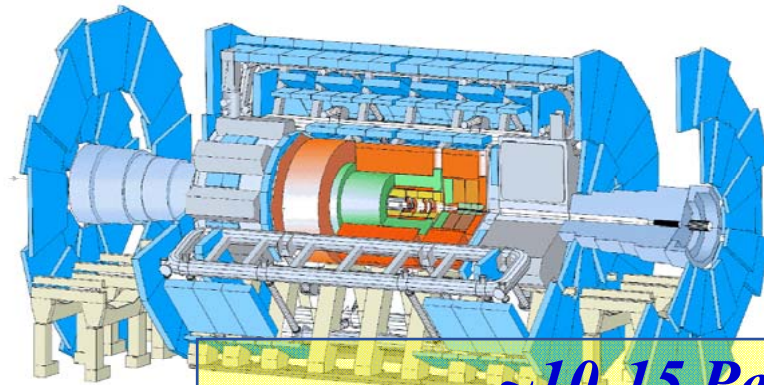
Някои области, в които има натрупан опит и положителни примери за използване на възможностите за **е-наука** са:

- Физика на елементарните частици
- Биомедицински приложения
- Приложения за Земята
- Археология

- Големи количества данни
- Мащабно сътрудничество за целия свят – скъпи експерименти
- Ресурсите за изчисления и управление на данните са разпределени в целия свят, като се притежават и управляват от различни организации
- Големият ускорител в (Large Hadron Collider-LHC) CERN, Женева, Швейцария:
 - Един от най-мощните инструменти създаден от хората за изследване на материята



ATLAS



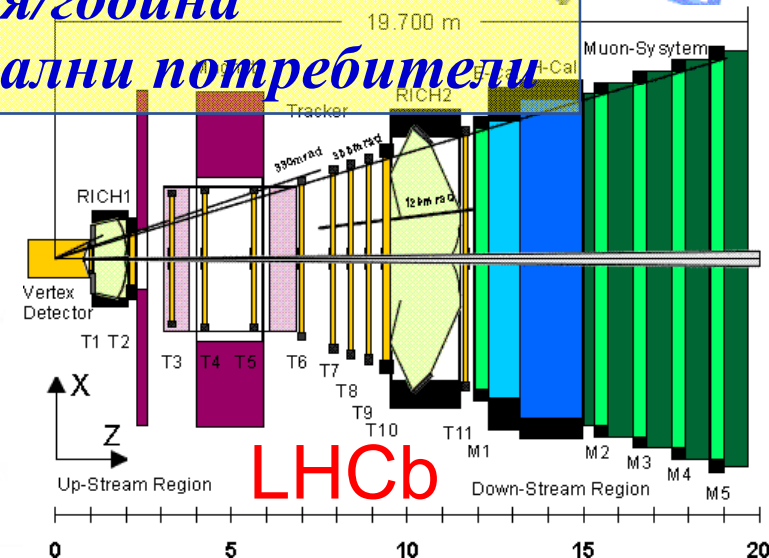
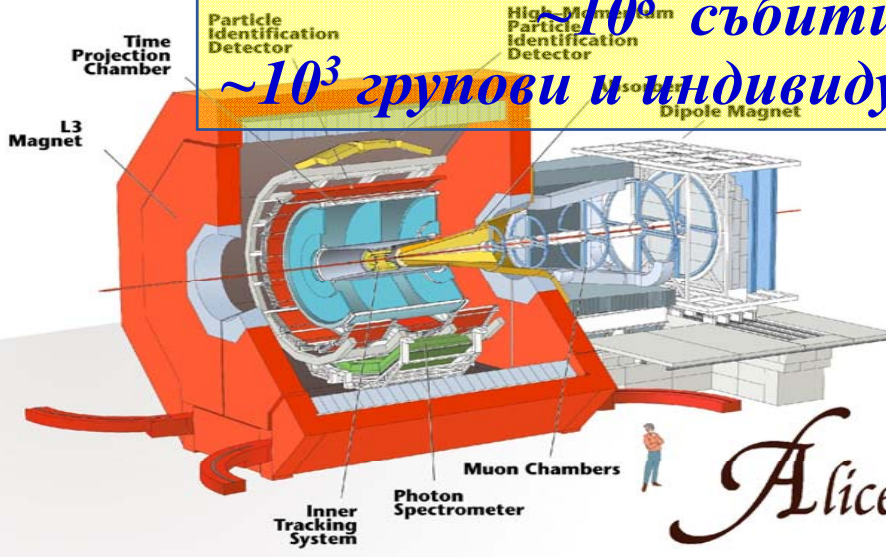
CMS



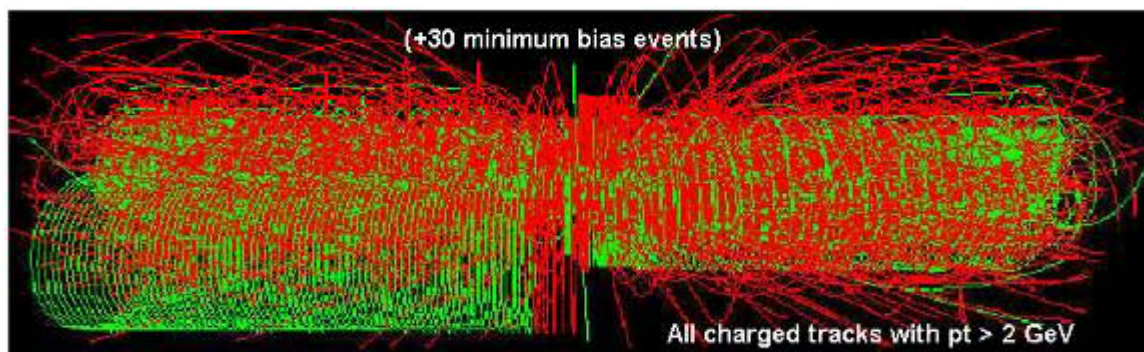
~10-15 PetaBytes /година

~10⁸ събития/година

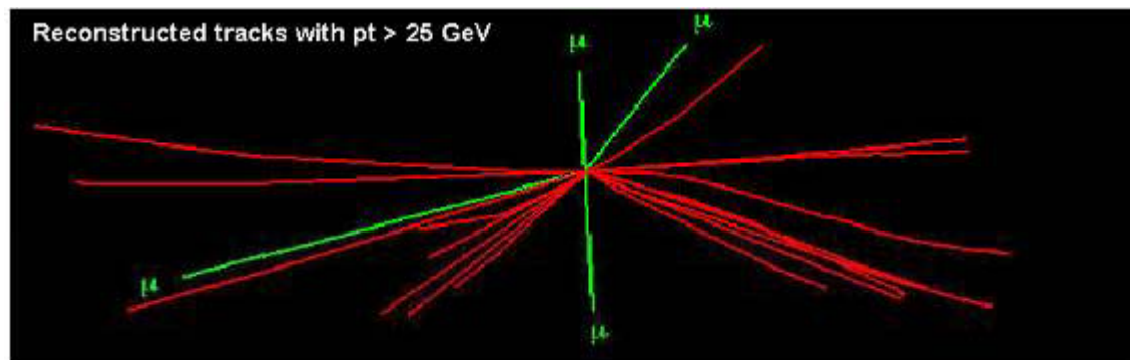
~10³ групи и индивидуални потребители



Започвайки от
това събитие



Търсейки този
“подпис”



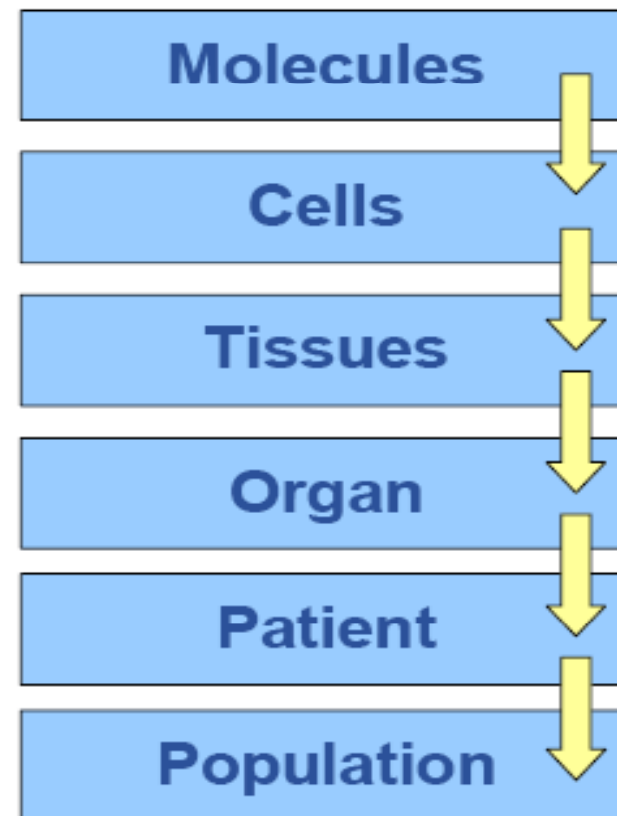
→ Избирателност: 1 на 10^{13}

(Като търсене на игла в 20 милиона купи сено)

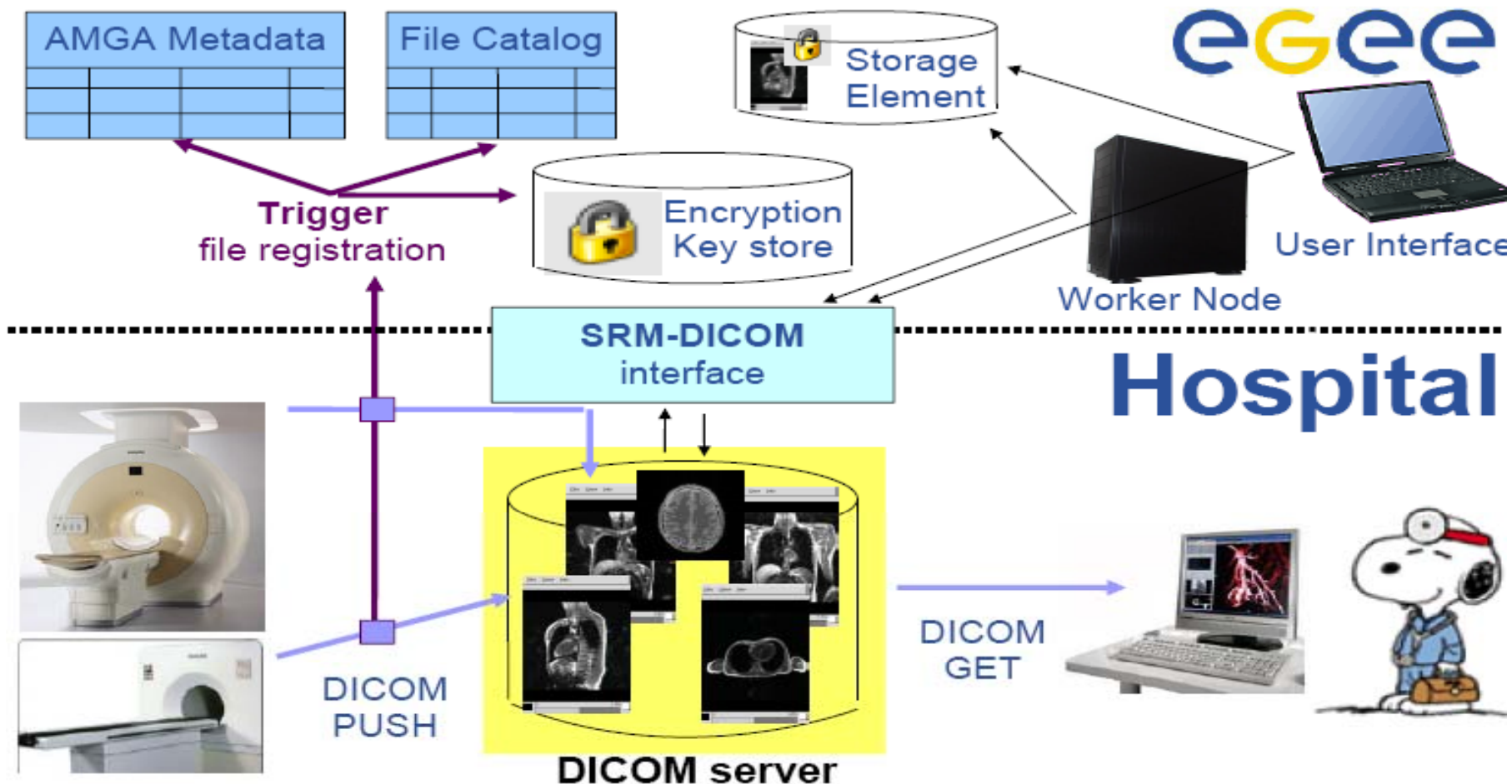
- **Bioinformatics**
 - Genomics
 - Proteomics
 - Phylogeny...

- **Medical imaging**
 - Medical imaging
 - Computer Aided Diagnosis
 - Therapy planning
 - Simulation...

- **Life sciences**
 - Drug discovery
 - Epidemiology
 - ...



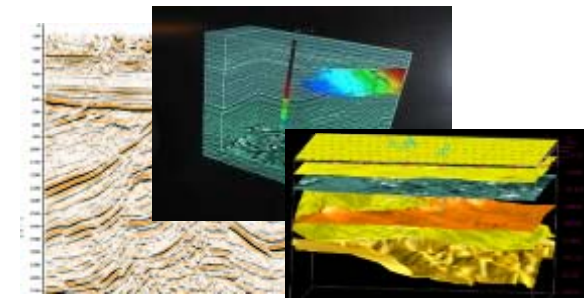
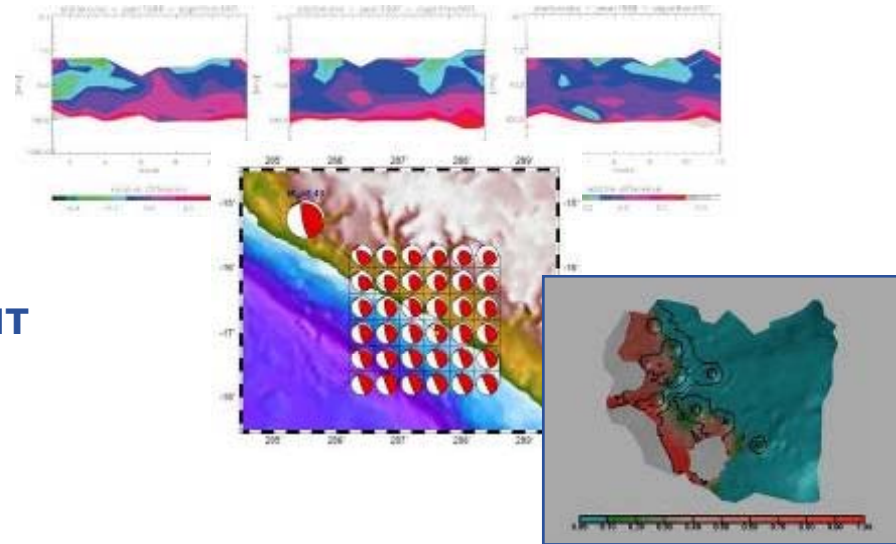
*Биомедицинската общност и Грид, EGEE потребителски форум,
1 март 2006, I. Magnin*

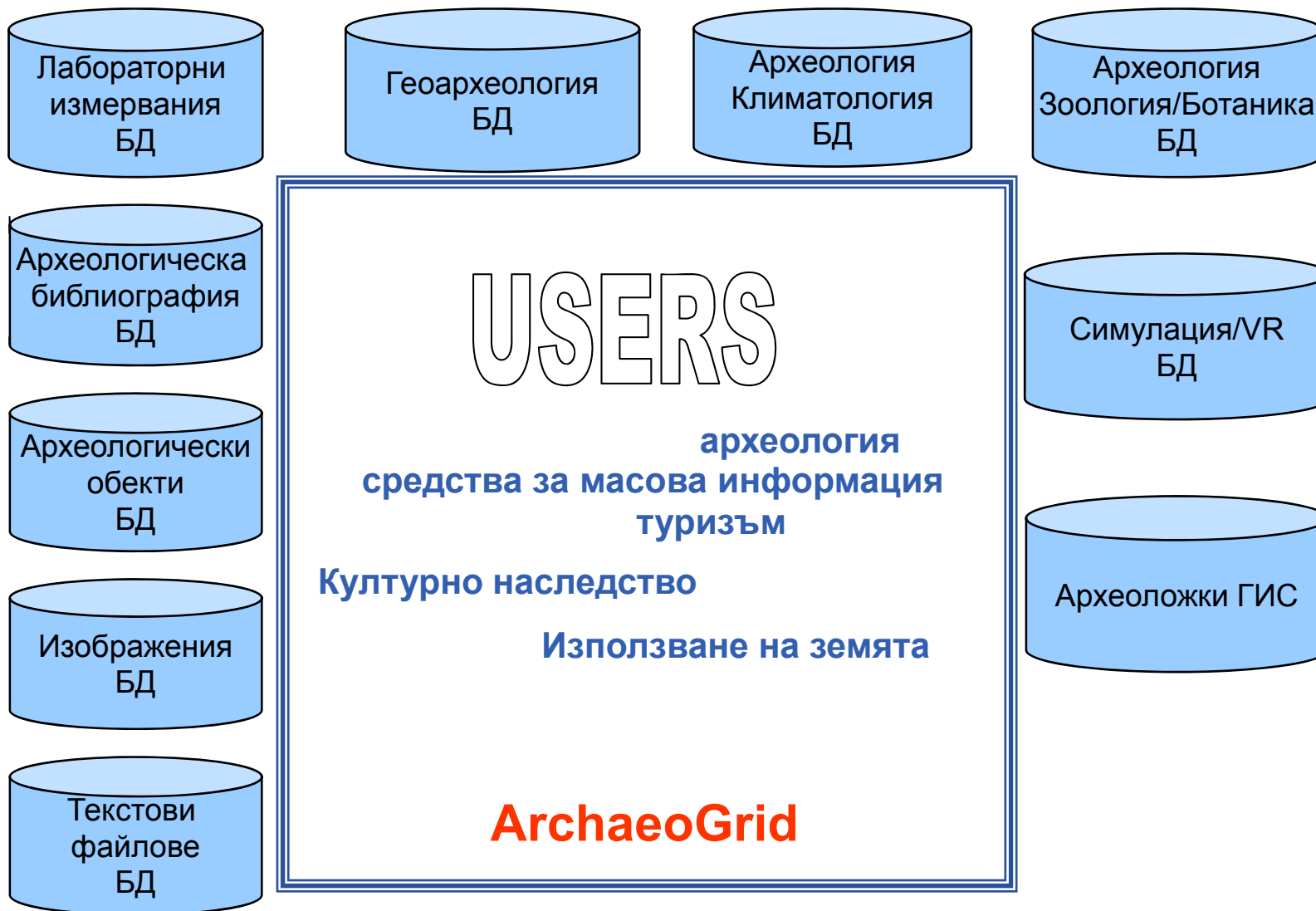


Биомедицинската общност и Грид, EGEE потребителски форум, 1 март 2006, I. Magnin

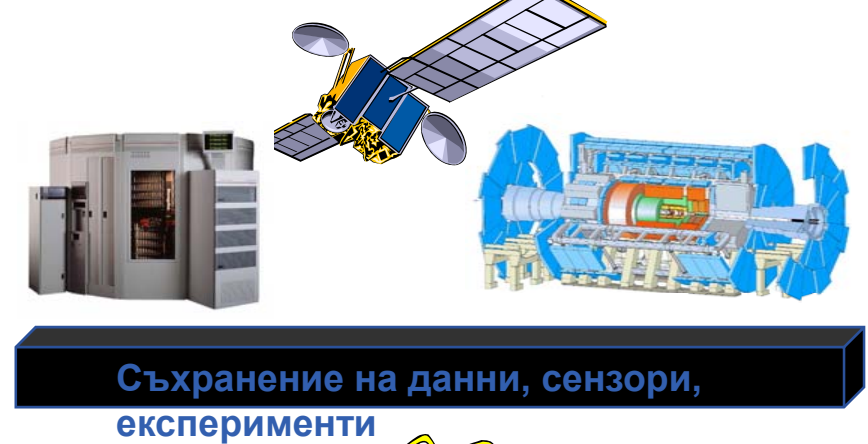
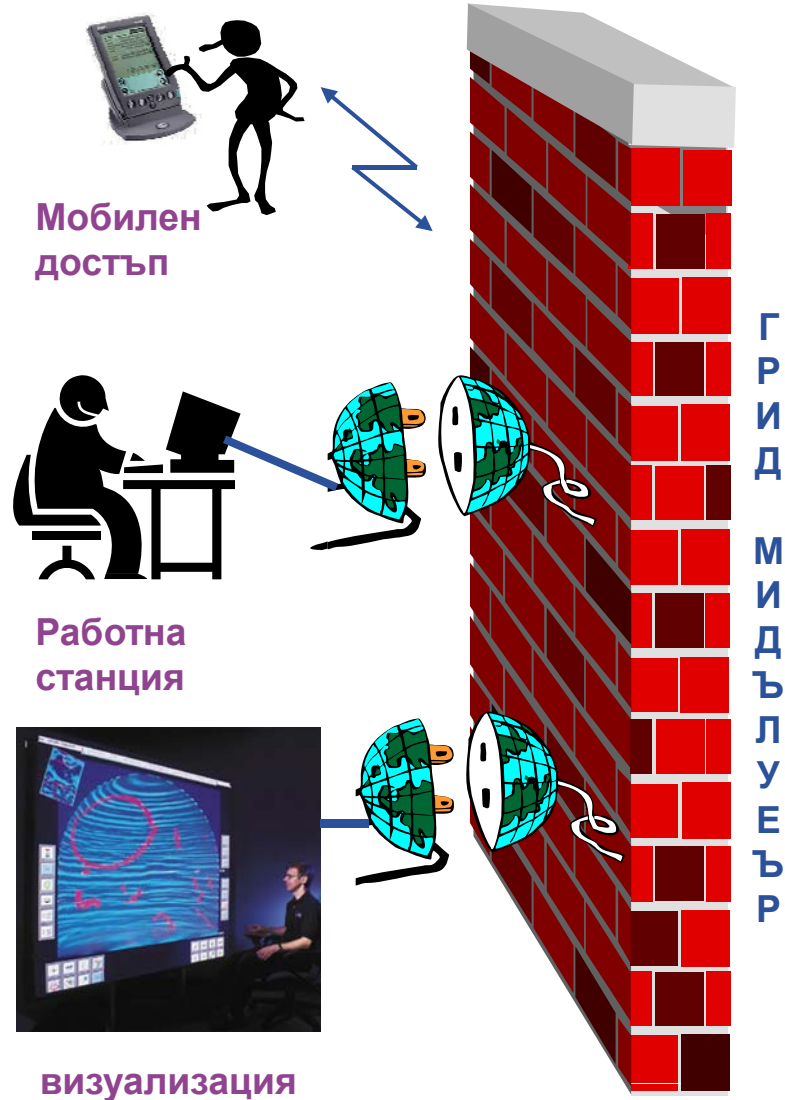
- **Инициативата WISDOM (Wide In Silico Docking On Malaria –**
(<http://cordis.europa.eu/ictresults/index.cfm/section/news/tpl/article/ID/80155/BrowsingType/Features>)
 - Демонстрира възможностите за ползване на изчислителни ресурси с Грид организация за откриване на нови лекарства.
 - Целта е чрез скрининг по определен критерий на данните за голямо множество съединения да се открият кандидати за лекарство, които впоследствие да бъдат допълнително анализирани по традиционни методи.
 - Две различни молекулярни приложения (Autodock и FlexX)
- **Големи числа**
 - Около 46 милиона ligands docked за 6 седмици
 - 1ТВ генерирани данни
 - До 1000 компютъра в 15 страни използвани едновременно за около 80 CPU години
- **Резултатите от експеримента водят до неговото продължаване за търсене на лекарства за “птичи грип”**
(http://wisdom.eu-egEE.fr/avianflu/press/NR_avianfluEGEE.pdf)

- **Наблюдения на Земята от сателит**
 - За озоновия слой
- **Физика на Земята**
 - Откриване на механизмите на големи земетресения
- **Хидрология**
 - Управление на водните ресурси
- например в средиземноморието (SWIMED)
- **Геология**
 - Геоклъстър: R&D инициатива на Compagnie Générale de Géophysique



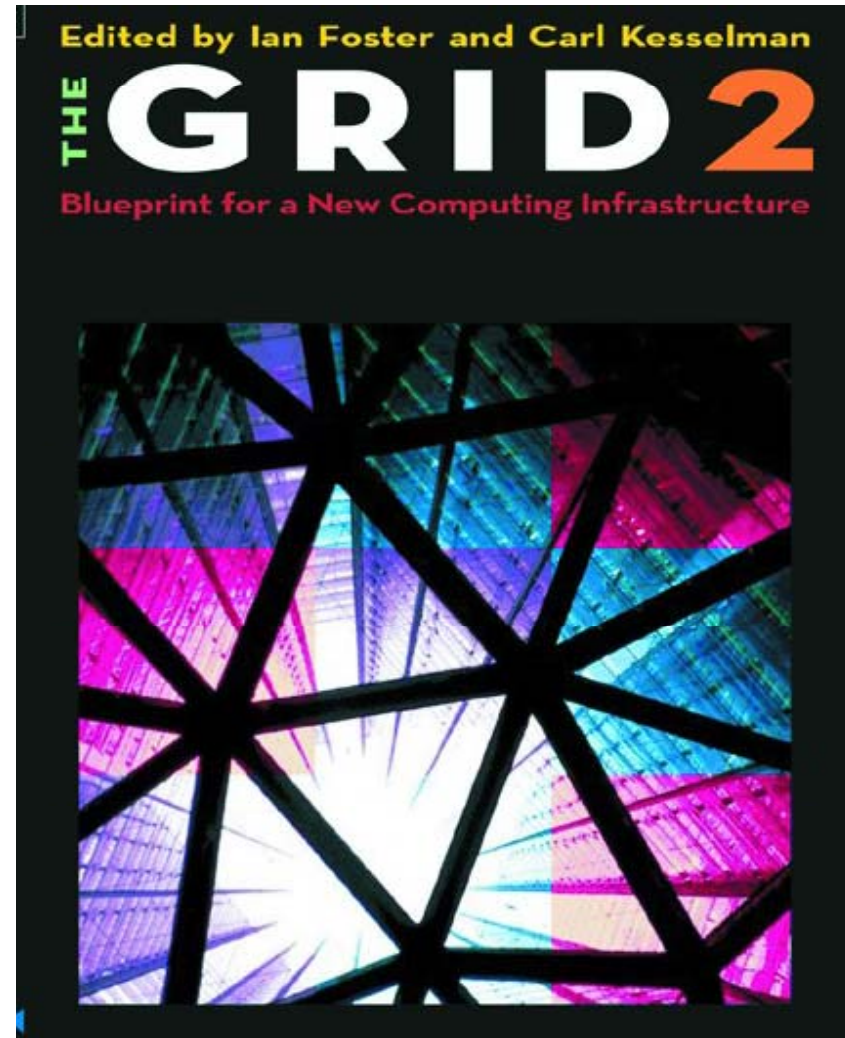


P.G.Pelfer, EGEE User Forum, March 1-3, 2006



- Грид визията е “виртуални изчисления” (+ информационни услуги за локализиране на изчисленията и на ресурсите памет)
 - Сравнение: Web: “виртуални документи” (+ услуги за тяхното търсене и локализиране)

- **ОБОСНОВКА:** необходимост от сътрудничество в споделяне на ресурси и опит за разширяване възможностите на
 - Изследванията
 - Търговия – инженерство, ...
 - Обществени услуги – здраве, околна среда,...



- Става възможен цялостен подход
- Предизвикателство към въображението
- Ефект > Σ частите



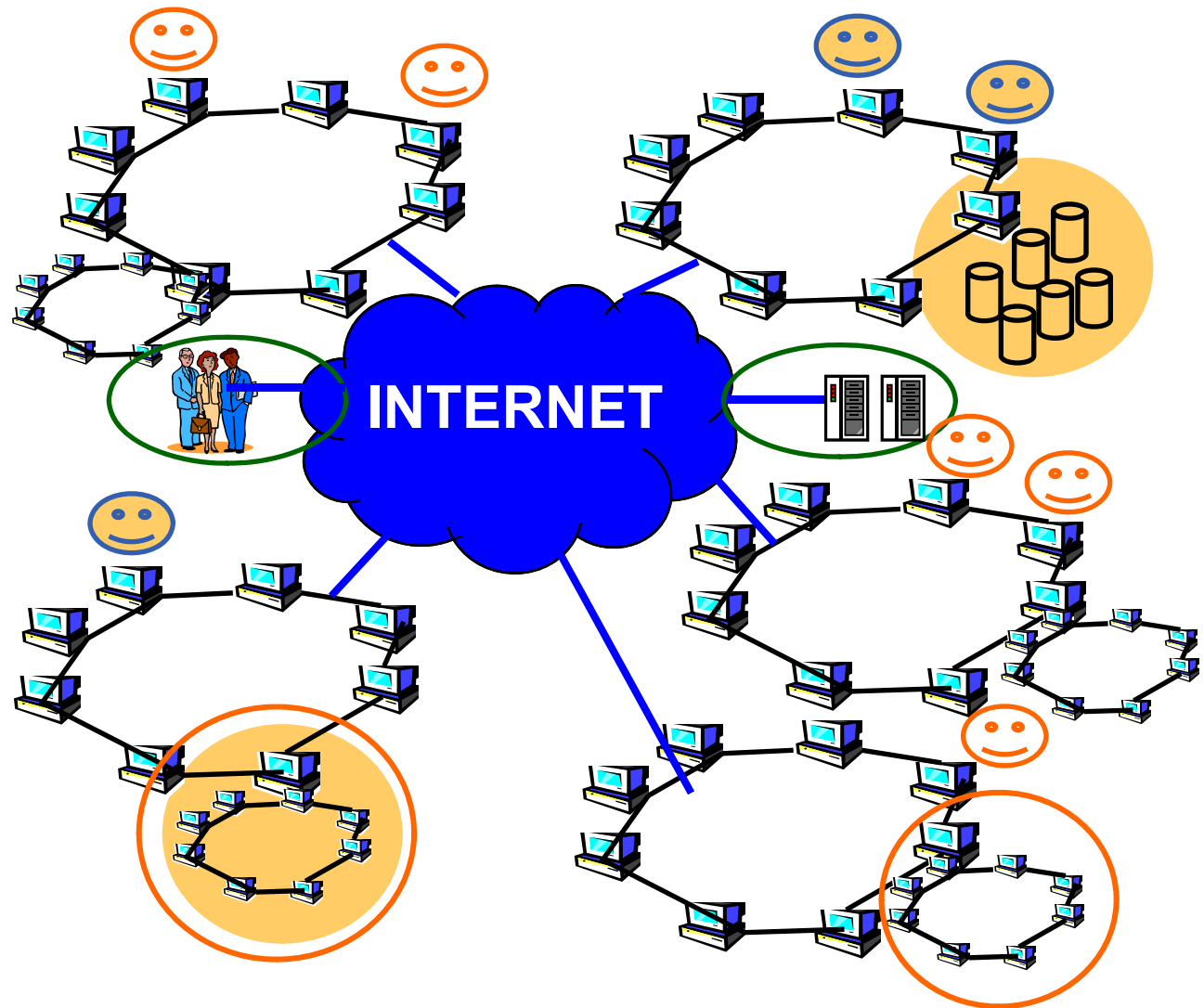
Диаграма на базата на слайд на Ян Фостер

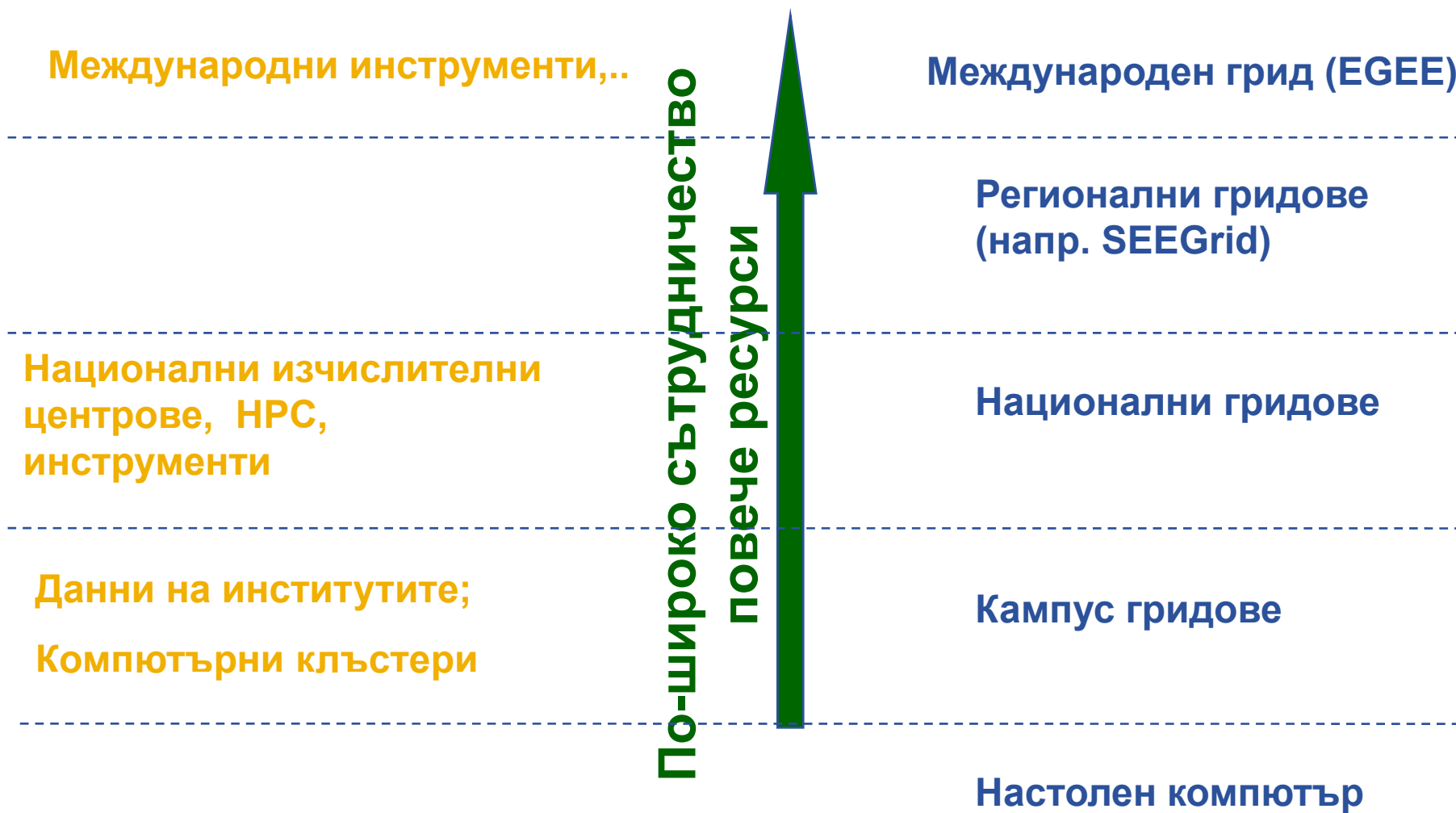
- **Гъвкаво и опростено общо ползване на ресурсите, което дава предимства при сътрудничество**
 - Преминават границите на административните единици
 - Абстракциите прикриват детайлите
 - Полза от съобразяване с Грид процедурите
 - Услуги по експлоатацията (хора и софтуер)

- **По-рационално използване на ресурсите**
 - Всеки споделя ресурси според осреднените си изисквания и има възможност в определени моменти при необходимост да ползва повече
 - Споделените общи ресурси имат минимален престой
 - Всеки има полза от
 - *Маскиране на хетерогенността*
 - *Мащаба*

- **Какво е Виртуална Организация?**
 - Хора от различни организации, които желаят да си сътрудничат и да споделят ресурси извън техните организационни граници
 - Например: общи проекти, научно сътрудничество ...
- **Всеки Грид е инфраструктура, която позволява една или повече “виртуални организации” да имат общ достъп до ресурсите**
- **Всеки ресурс в грид е на разположение чрез абстракция, която маскира хетерогенността и местоположението**
 - Различни изчислителни платформи
 - Различни ресурси данни
 - Географски разпределени
- **Ресурсите се притежават от ВО, които ги ползват на базата на проведени преговори**

- **Виртуални организации се договарят със сайтове за достъп до ресурси**
- **Грид мидълуер се изпълнява на всеки общ ресурс, като осигурява**
 - Услуги за управление на данни
 - Изчислителни услуги
 - Единен унифициран вход
- **Разпределените услуги (хората и мидълуер) правят работата на грида**







Взаимодействие на компютърните науки с приложенията на научната общност!

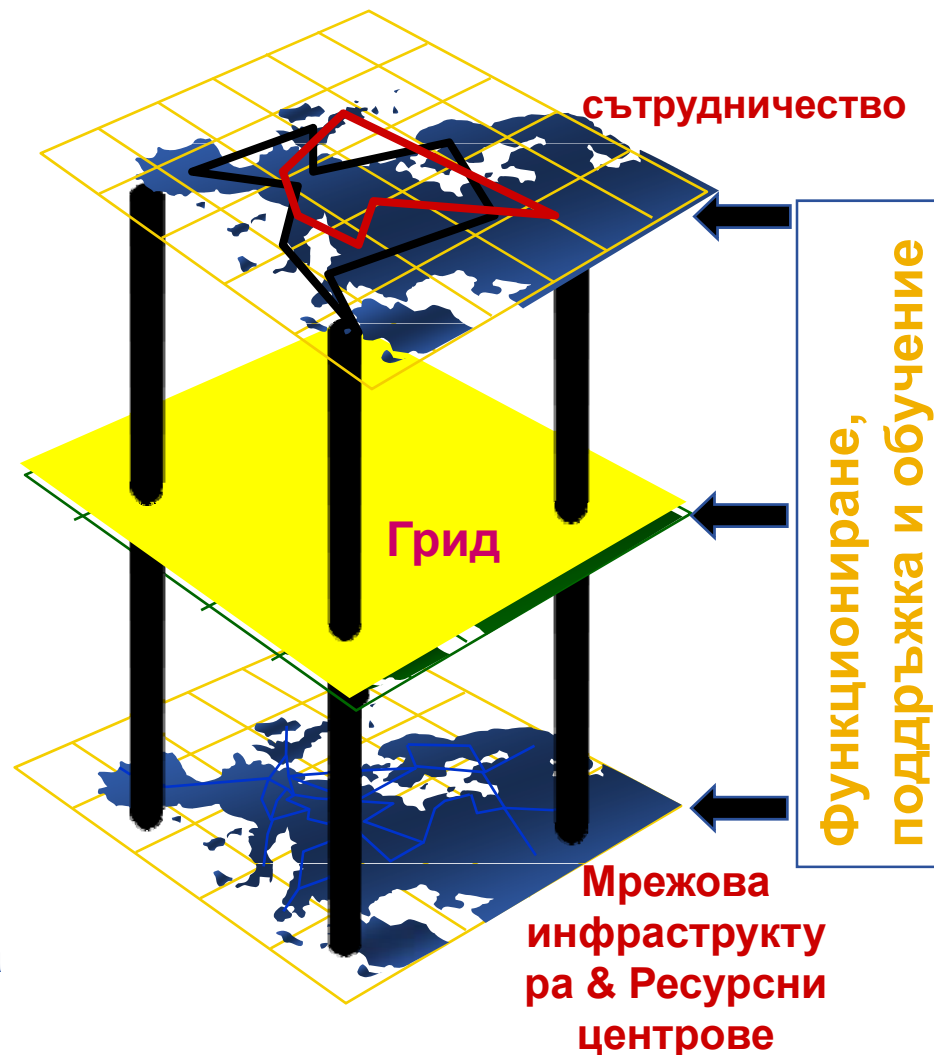
Специфични разработки на VO:

- Портали
- Виртуални изследователски среди
- Семантика, онтология
- Работен поток
- Регистри за услугите на VO

Гридовете в производствен режим осигуряват тези услуги.

- **Организациите, предоставящи ресурсите (компютри, бази данни,...), трябва да управляват рисковете от доверяване на потребители, които те не познават**
 - Доверие във ВО
 - ВО трябва да вярва на своите членове
- **Нужди на потребителите по отношение на сигурността**
 - Единствен вход - да бъде възможно с едно влизане на потребителя в изчислителната среда, неговата идентичност да се предаде на останалите ресурси в е-инфраструктурата
 - Да се довери на собствениците на ресурсите, които използва
- **Мидълуеърът на грида трябва да осигурява:**
 - *Идентифициране на потребителите - кой иска да ползва ресурса*
 - *Ауторизация – да предоставя на потребителя права за ползване на ресурсите и да следи за тяхното спазване*
 - *Сигурност - да намали уязвимостта от външни въздействия*
 - *Да се знае кой, кога и какво прави и е правил*
- **Мидълуерът, осигуряващ сигурността на Грид инфраструктурата е в основата на (повечето) промишлени гридове**

- **Грид: виртуални изчисления независимо от границите на административните единици**
 - Ресурсите имат обща идентификация
 - Общо управление на правата на потребителите
 - Достъпът до ресурсите е абстрактен
- **Необходимост от Грид**
 - Сътрудничество в изследванията,
 - Сътрудничество в техническото обслужване, техническата реализация, общите услуги ...
 - По-добро оползотворяване на ресурсите



- **Грид** са компютърни ресурси, които се предоставят при поискване през Интернет посредством стандартизирани процедури и при гарантирана висока надеждност.
 - Акцент е наличието на глобален унифициран достъп
 - Потребителят получава компютърни ресурси имащи желани от него възможности (хардуер и софтуер), които ползва за ограничено време.
- При **Cloud Computing** се акцентира върху получаването на компютърни услуги, които се предоставят при поискване през Интернет посредством стандартизирани процедури и при гарантирана висока надеждност.
 - Потребителят не се интересува от компютърния ресурс зад тези услуги.
 - В частност компютърната услуга може да е под формата на компютърна среда, която потребителят организира и ползва по желан от него начин
- Може да се приеме, че **Cloud Computing** е еволюция на концепцията за ползване на обществени компютърни услуги, в която междинен етап са **Грид**.
- От гледна точка на техническата реализация **Cloud Computing** е потребителско ниво на достъп до разпределени компютърни ресурси, което е:
 - над нивото на услугите предоставяни от компютърните ресурси, които могат да са и с **Грид** организация;
 - фокусирано върху това какво потребителя иска да прави, а не върху начина по който то се обезпечава.
- **Разлики от гледна точка на организацията на ползване**
 - **Cloud Computing** най-често не решава въпросите по взаимодействието на различни собственици на ресурсите т.к. е с един собственик
 - **Грид** акцентира върху взаимодействието и организирането на различните потребители (Виртуални организации)

- Проектът EGEE (Enabling Grids for E-Science - <http://www.eu-egee.org/>) протича в три части, всяка с продължителност от 2 години
 - EGEE-1 от 2004 до 2006
 - EGEE-2 от 2006 до 2008 и
 - EGEE-3 от 2008 до 2010 - това е последния етап
- Следва EGI (European Grid Initiative - <http://web.eu-egi.eu/>) , където ударението се измества
 - от европейско към национално финансиране
 - от развитие и внедряване на Грид инфраструктурата към нейната надеждна експлоатация
- Цел на проекта EGEE - да развие, инсталира и експлоатира **единна, надеждна и широкомащабна Грид услуга**, която
 - Да е свързана с националните, регионални и международни Грид инициативи
 - Да развие и поддържа мидълуер, който да е в състояние да осигури устойчиви и надеждни Грид услуги
 - Да привлече нови потребители от научните институти и индустрията и да осигури тяхното обучение

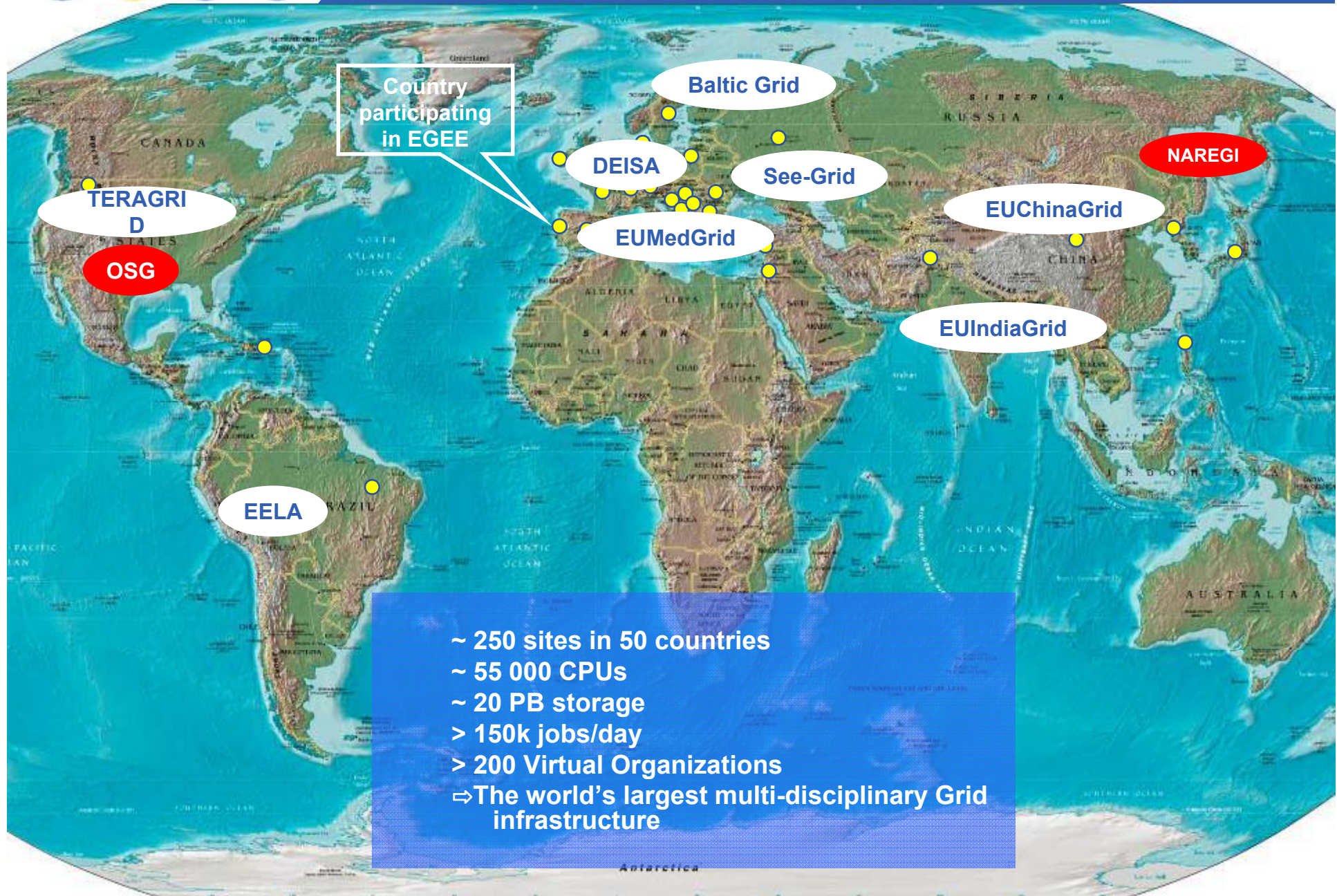
- **Инфраструктура**

- Брой на сайтовете: 267
- Брой на страните: 54
- Брой на CPU (на разположение в 24 часов режим): ~ 114,000
- Обем на достъпната масова памет: ~ 20 PB disk + tape

- **Проекта**

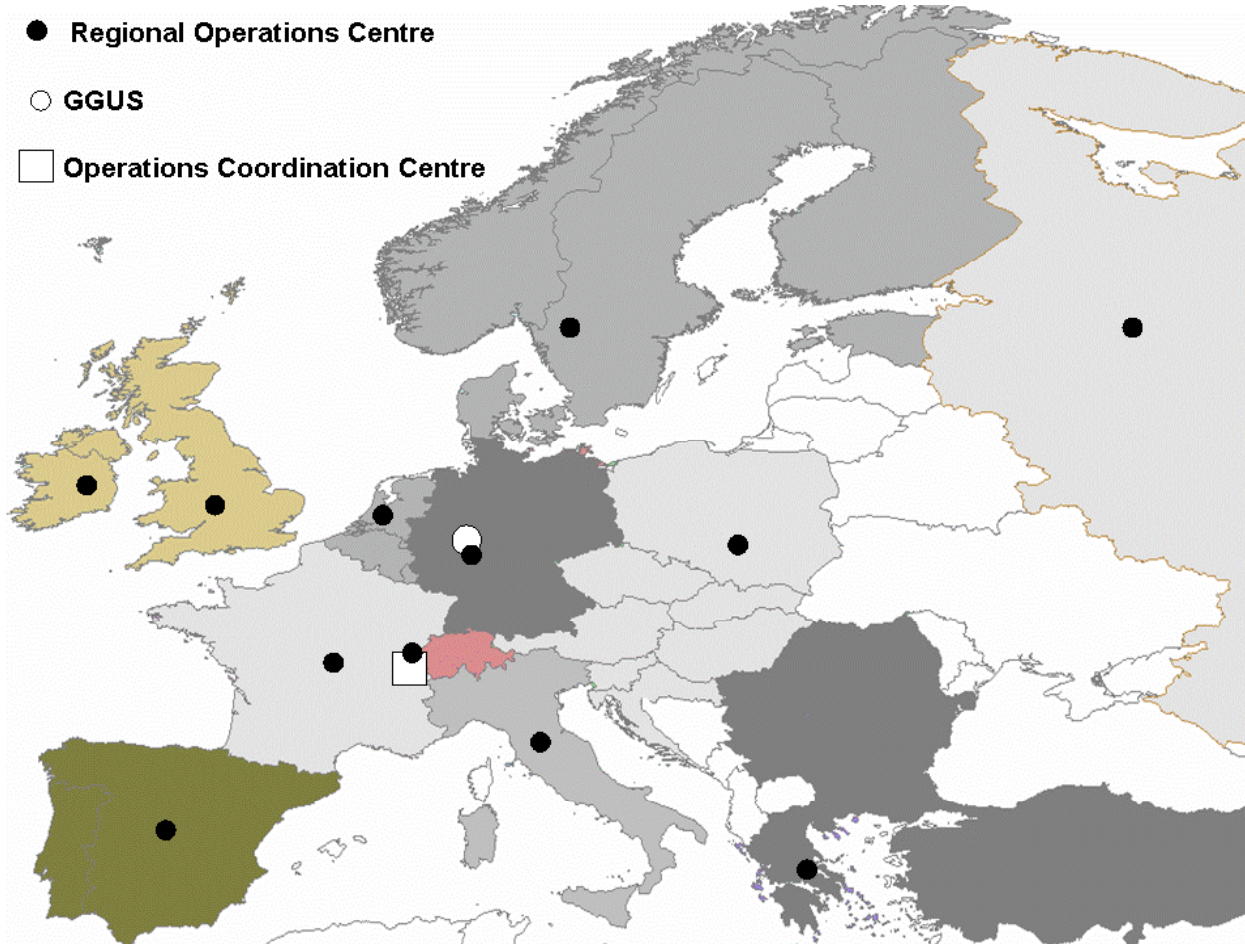
- Продължителност: 24 месеца
- Финансов принос на ЕК: 32,000,000 euro
- Общ бюджет: около 47,150,000 euro
- Цена на компютърните ресурси, предоставени от участниците в проекта - около 50,000,000 euro
- Общо изразходван труд: 9,010 човекомесеца, от които над 4,500 ч.м. са принос на участниците с тяхно финансиране (<http://project.eu-egee.org/>)

- Брой на Виртуалните организации, ползващи EGEE инфраструктурата: > 200
- Брой на регистрираните Виртуални организации: >152
- Брой на регистрираните потребители: > 16000
- Брой на хората, имащи полза от EGEE инфраструктурата : ~20000
- Брой на задачите (jobs): >150k jobs/day
- Брой на приложните области ползващи EGEE инфраструктурата: > 15
 - Archeology
 - Astronomy & Astrophysics
 - Civil Protection
 - Computational Chemistry
 - Computational Fluid Dynamics
 - Computer Science/Tools
 - Condensed Matter Physics
 - Earth Sciences
 - Finance (through the Industry Task Force)
 - Fusion
 - Geophysics
 - High-Energy Physics
 - Life Sciences
 - Multimedia
 - Material Sciences



~ 250 sites in 50 countries
~ 55 000 CPUs
~ 20 PB storage
> 150k jobs/day
> 200 Virtual Organizations
⇒ The world's largest multi-disciplinary Grid infrastructure

- Regional Operations Centre
- GGUS
- Operations Coordination Centre



- **Operations Coordination Centre (OCC)**
 - management, oversight of all operational and support activities
- **Regional Operations Centres (ROC)**
 - providing the core of the support infrastructure, each supporting a number of resource centres within its region
 - Grid Operator on Duty
- **Resource centres**
 - providing resources (computing, storage, network, etc.);
- **Grid User Support (GGUS)**
 - At FZK, coordination and management of user support, single point of contact for users

Грид услуги

Как мидълуера на EGEE Грид осигурява услугите, които подържат сътрудничеството между множество разнообразни Виртуални Организации и осигуряват ползването на техните ресурси?

- **Когато ползвате ПК или работна станция:**
 - Влизате с потребителско име и парола (“Authentication”)
 - Ползвате предоставените ви права (“Authorisation”)
 - Изпълнявате задания
 - Работите с файлове: създавате ги, четене/запис, списъци директории
- **Компонентите са локално свързани с шина**
- **Операционна система**
- **Един администр. домейн**
- **При използване на Грид**
 - Влизате с цифрови акредитиви (“Authentication”)
 - Ползвате предоставените ви права (“Authorisation”)
 - Изпълнявате задания
 - Работите с файлове: създавате ги, четене/запис, списъци директории
- **Услугите са свързани с Интернет**
- **Мидълуер**
- **Много администр. домейни**



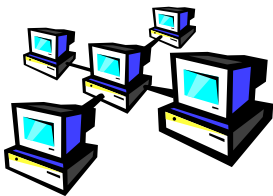
Потребителски интерфейс (UI): Машината през която потребителят влиза в Грида



Ресурсен брокер (RB): Съчетава нуждите на потребителя с моментното състояние на ресурсите на Грида



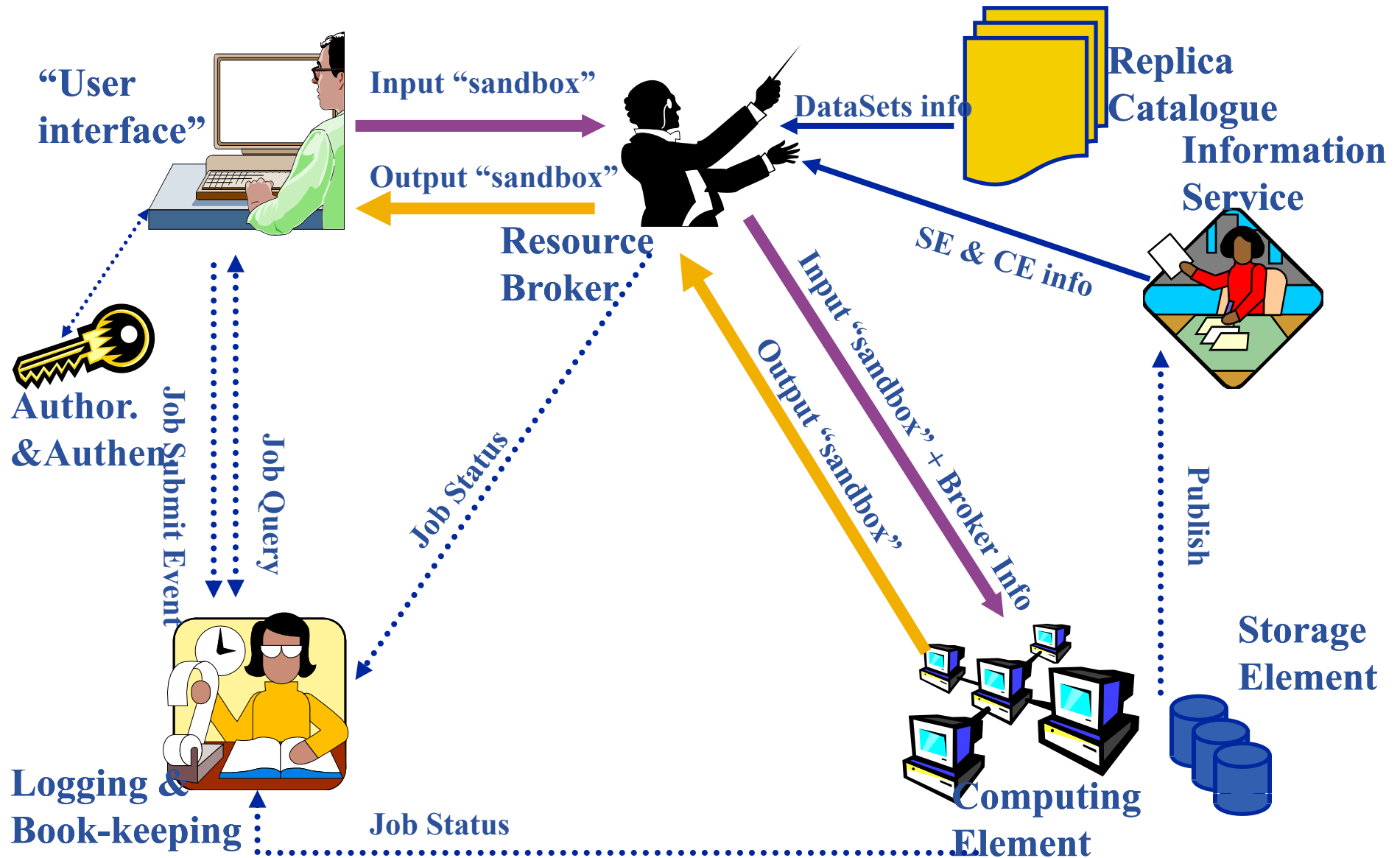
Информационна Система: Дава информация за характеристиките и състоянието на ресурсите на Грида (на CE и SE)



Изчислителен Елемент (CE): Създава и управлява опашката от задачи на потребителя към работните компютри (Workers) на сайта, където те се изпълняват

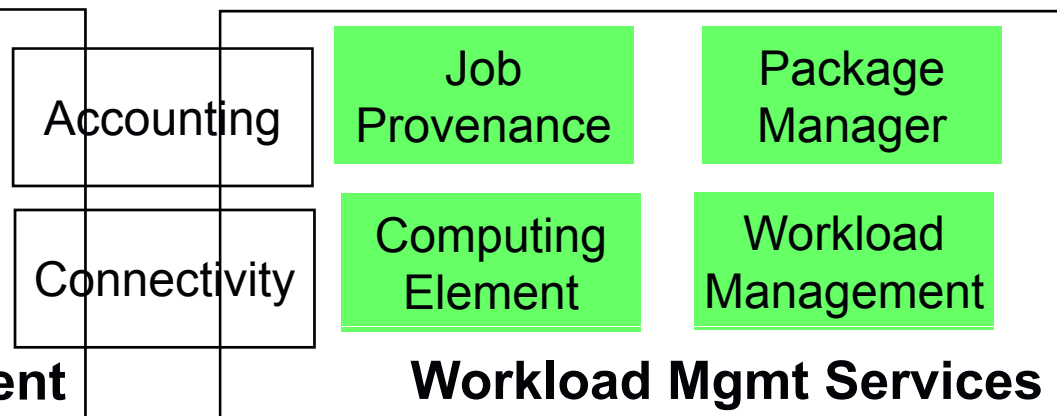
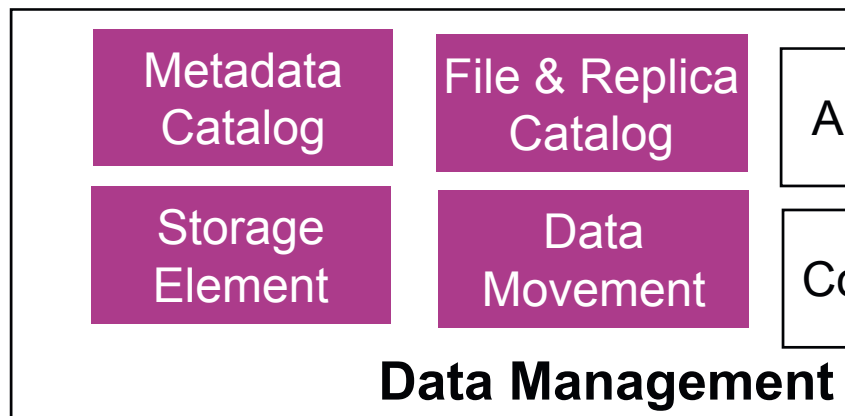
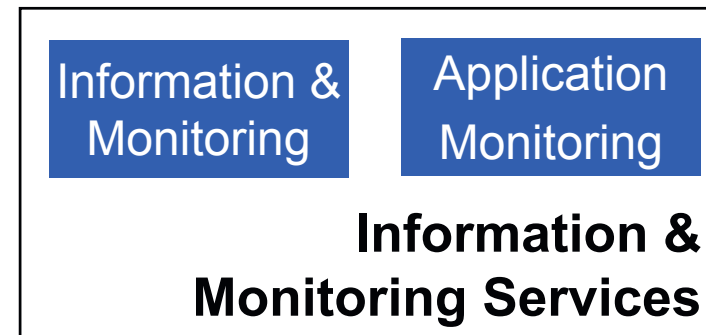
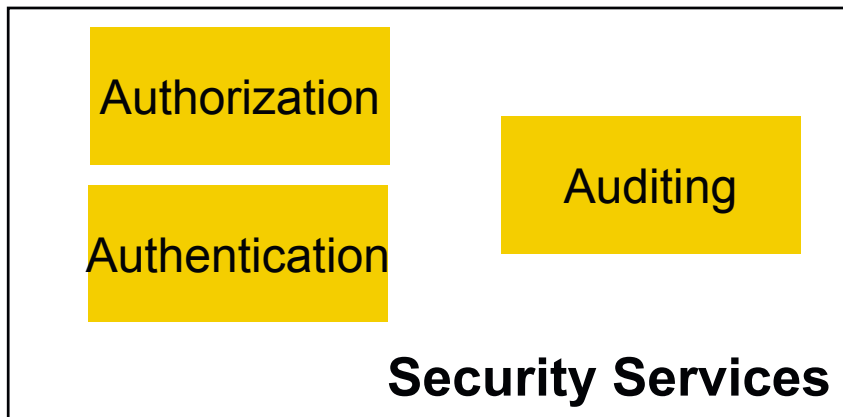
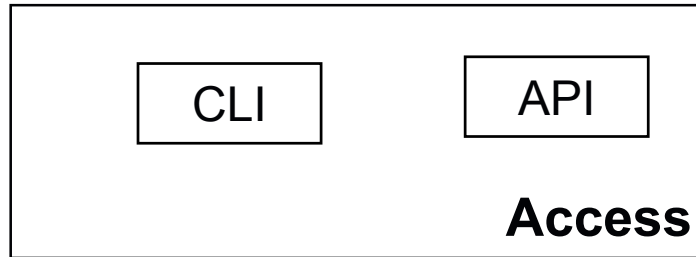


Елементи памет (SE): осигурява необходимата памет за потребителските файлове



Кой осигурява ресурсите?!

<u>Услуга</u>	<u>Осигуряващ</u>	<u>Забележка</u>
<u>Услуга за управление на достъпа</u>	потребител / институт / Виртуална Организация	Компютър с клиентски софтуер
<u>Ресурсен Брокер (RB)</u>	Виртуална Организация	
<u>Информационна система</u>	Грид – управление на операциите	Услугата се финансира от бюджета на проекта EGEE
<u>Изчислителен елемент (CE)</u>	Виртуална Организация	Променлив размер - ВО трябва да осигури ресурси, които да съответстват на нейните средни нужди за даден период
<u>Елементи памет (SE)</u>	Виртуална Организация	



- **EGEE експлоатира най-голямата Грид структура с много Виртуални организации в света**
- **Налице е Грид ниво в е-инфраструктурата, което е предназначено за научни изследвания, обществени услуги и индустрията**
- **Основни предимства на EGEE Грид**
 - Стабилност и висока надеждност, позволяващи дългосрочно планиране
 - Качество, съответстващо на изискванията за работа в непрекъснат производствен режим