

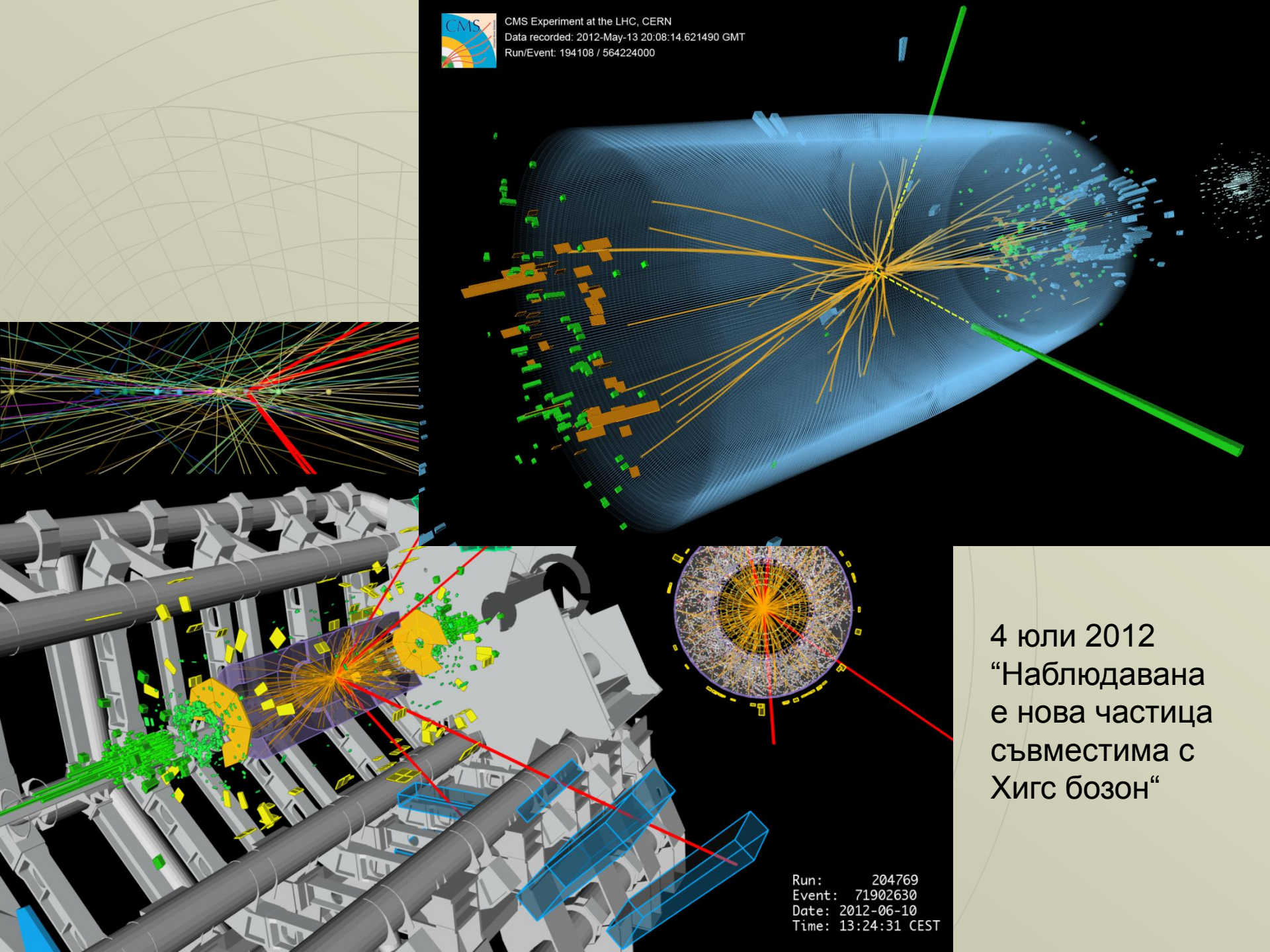


Въведение в GRID

Преслав Константинов



CMS Experiment at the LHC, CERN
Data recorded: 2012-May-13 20:08:14.621490 GMT
Run/Event: 194108 / 564224000



4 юли 2012
“Наблюдавана
е нова частица
съвместима с
Хигс бозон“

Run: 204769
Event: 71902630
Date: 2012-06-10
Time: 13:24:31 CEST

Global Effort → Global Success

Results today only possible due to
extraordinary performance of
accelerators – experiments – Grid computing

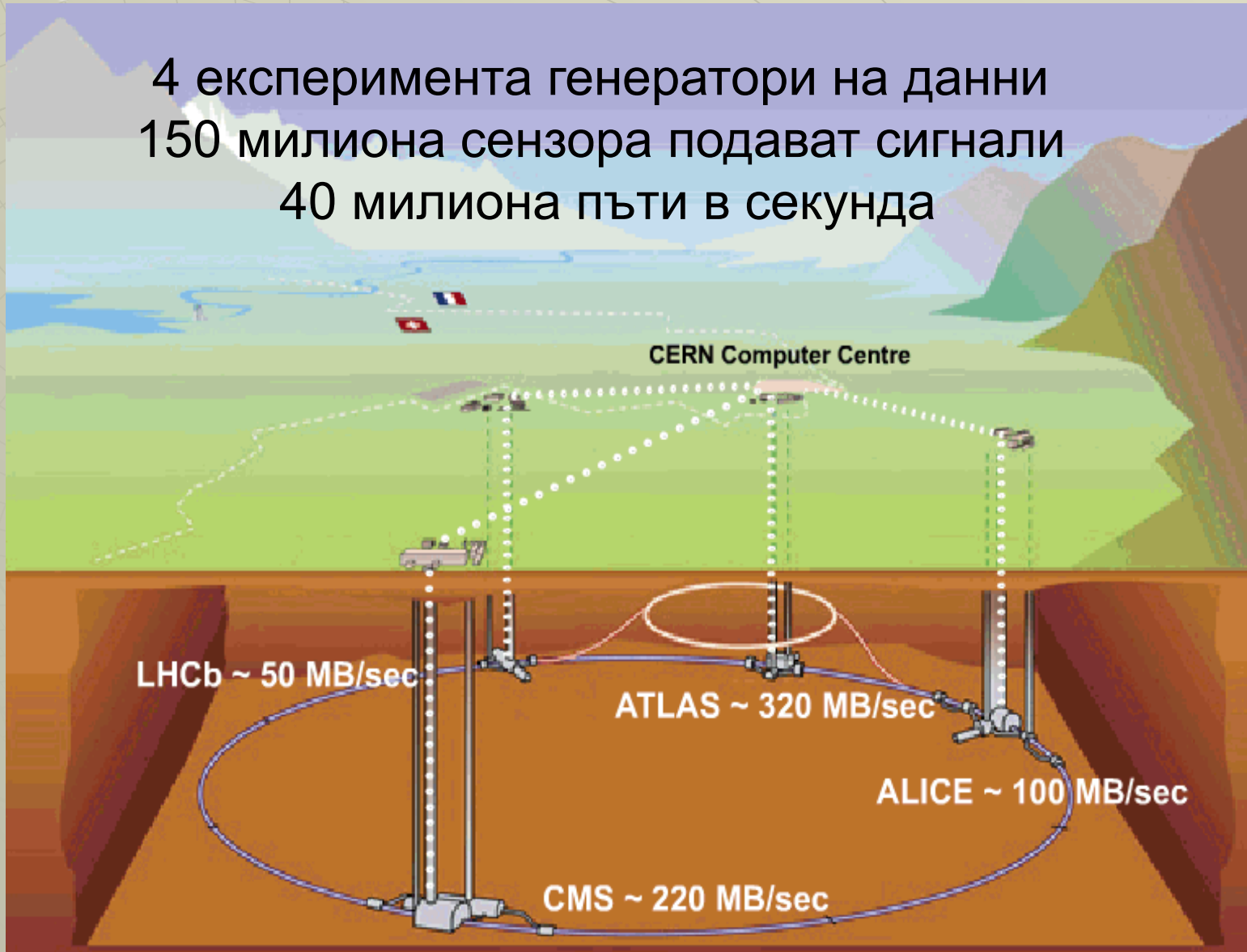
Observation of a new particle consistent with
a Higgs Boson (but which one...?)

Historic Milestone but only the beginning

Global Implications for the future

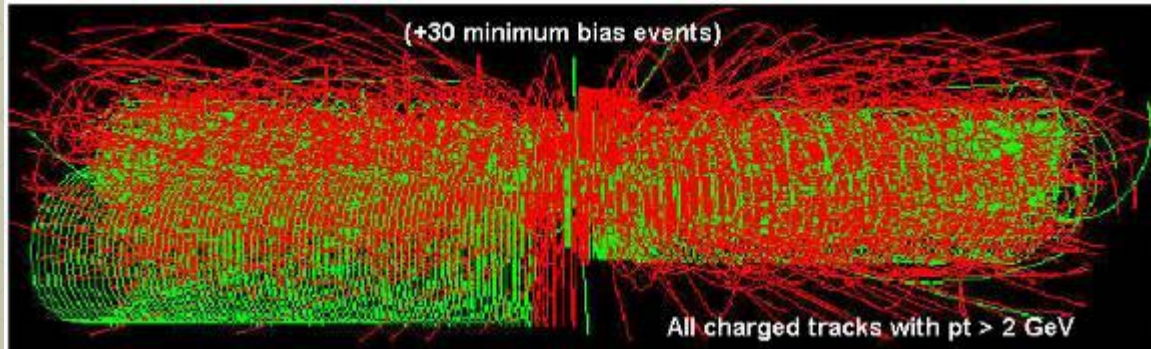
Търсене на Хигс бозона

4 експеримента генератори на данни
150 милиона сензора подават сигнали
40 милиона пъти в секунда

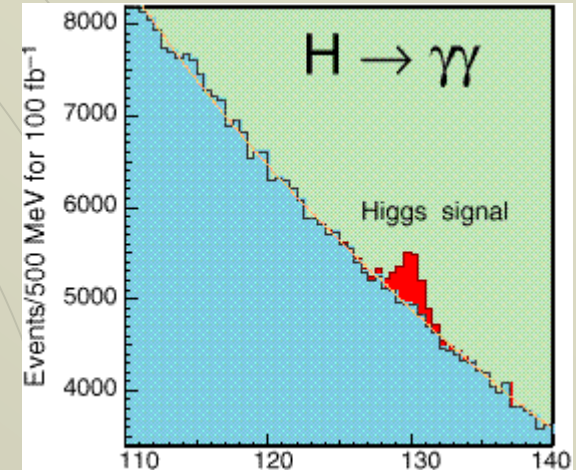
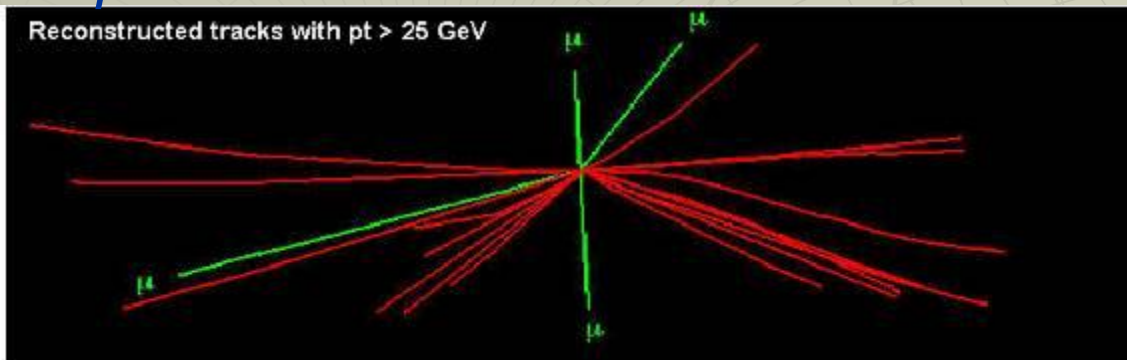


Търсене на Хигс бозона

Започвайки от това ...



търсите това



Селективност: 1 в 10^{13}

**Като да търсиш 1 човек
в хиляда пъти
населението на Земята!**

**Или игла в 20 милиона
купи сено!**

Данни и още данни ...

Експериментите на LHC произвеждат **15 Петабайта (милиона Гигабайта)** нови данни всяка година (около 20 милиона CD!)

1 PB 13.3 години HD видео

1.5 PB всички снимки във Фейсбук

350000 PB твърдите дискове, произведени през 2011

За 2010 и 2011 в ЦЕРН са записани около **38 PB**

През 2012 се очакват още **30 PB**

Към това трябва да се прибавят и обработените и симулираните данни



Обработка на данните

Анализът на данните на LHC изисква изчислителна мощ равна на ~200 000 съвременни процесорни ядра.

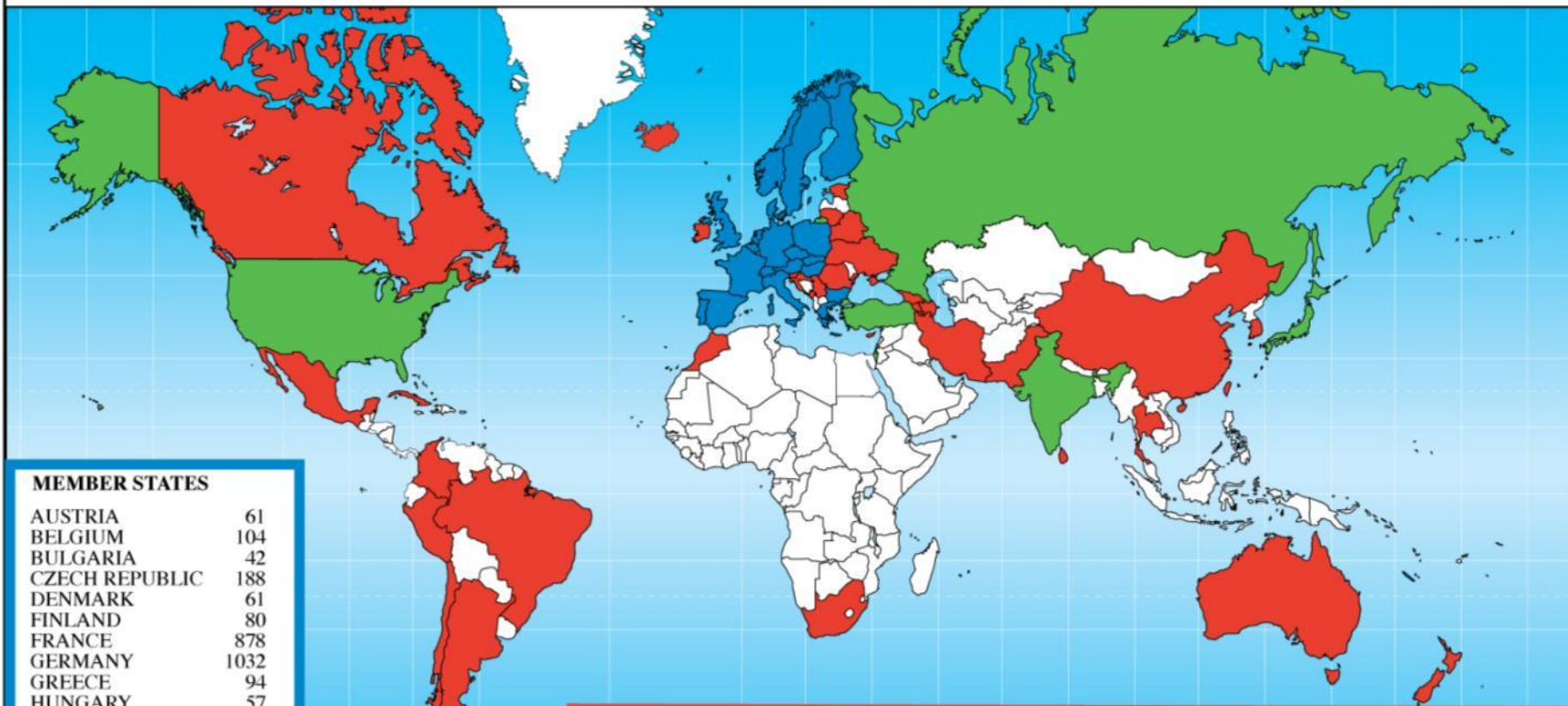
Данните са организирани като събития

Събитията са независими и програмите ги обработват едно по едно. Може да се прави паралелно.



Голяма научна общност в целия свят

Distribution of All CERN Users by Nation of Institute on 6 January 2009



MEMBER STATES

AUSTRIA	61
BELGIUM	104
BULGARIA	42
CZECH REPUBLIC	188
DENMARK	61
FINLAND	80
FRANCE	878
GERMANY	1032
GREECE	94
HUNGARY	57
ITALY	1483
NETHERLANDS	175
NORWAY	78
POLAND	174
PORTUGAL	111
SLOVAKIA	49
SPAIN	286
SWEDEN	73
SWITZERLAND	330
UNITED KINGDOM	715

6071

OBSERVER STATES

INDIA	89
ISRAEL	59
JAPAN	200
RUSSIA	883
TURKEY	52
USA	1485

2768

OTHER STATES

ARGENTINA	10	CUBA	3	MONTENEGRO	1	SRI LANKA	1
ARMENIA	15	CYPRUS	6	MOROCCO	5	TAIWAN	42
AUSTRALIA	14	ESTONIA	11	NEW ZEALAND	6	THAILAND	1
AZERBAIJAN	1	GEORGIA	11	PAKISTAN	24	UKRAINE	18
BELARUS	19	ICELAND	1	PERU	1		
BRAZIL	73	IRAN	12	ROMANIA	49		
CANADA	136	IRELAND	12	SERBIA	17		
CHILE	4	KOREA	51	SLOVENIA	16		
CHINA	64	LITHUANIA	5	SOUTH AFRICA	8		
COLOMBIA	11	MEXICO	28				
CROATIA	20						

696

Компютърният център на ЦЕРН

10000 инсталирани компютъра

65000 процесорни ядра

62 PB дисково пространство

65 PB капацитет на лентовите устройства

3.5 MW електрическо захранване

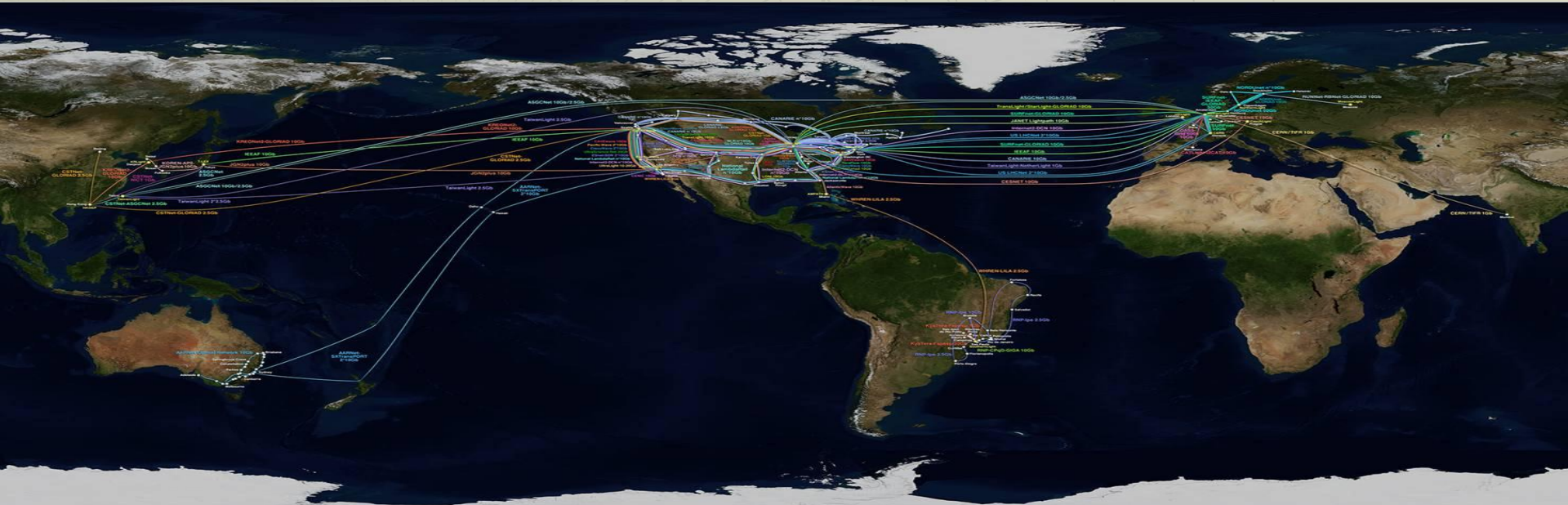
Не стига!



Решението

Проблем: ЦЕРН може да осигури само част от компютърните ресурси за работата с всички тези данни.

Решението: да се използва ГРИД, за да се свържат изчислителните центрове на физиката на елементарните частиците и да се обединят ресурсите по целия свят.



Какво е GRID (ГРИД)?

Има различни определения

- ◆ **Грид** е координирано *споделяне* на ресурси и решаване на задачи в динамични *многоинституционални* виртуални организации.

От гледна точка на потребителя

- ◆ **Грид** е инфраструктура, която осигурява лесен достъп до компютърна мощ и място за съхранение на данни, без значение къде се намират и чии са.

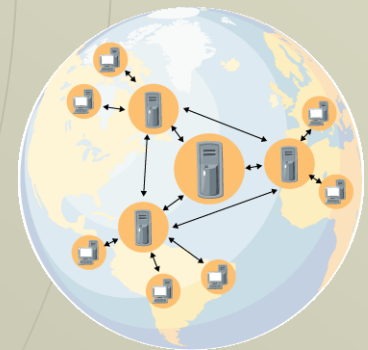
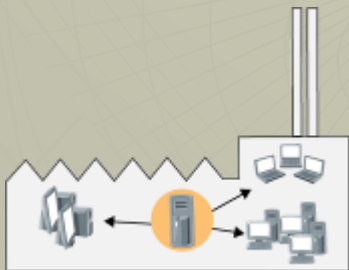
Няма централизиран контрол на ресурсите и потребителите

Различни ГРИД-ове за различни цели

Думата Grid се използва като събирателно за различни типове разпределени изчисления, например в рамките на една компания или ползването на много домашни компютри.

Няма унифициран Grid (както има само един web) **има различни ГРИД-ове** за различни приложения.

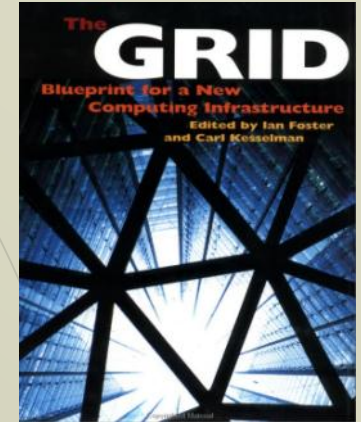
OpenGridForum работи за създаването на стандарти



Произход на названието Grid

Терминът Grid добива популярност от книгата на **Ян Фостър** и **Карл Кеселман** (“The Grid: blueprint for a new computing infrastructure”, 1997г.).

Името Grid (мрежа, решетка, скара) е избрано по аналогия с **електрическата мрежа**: “включваме се” към компютърната мрежа, без да се интересуваме откъде идва ресурса, както използваме тостер.



Какви са предизвикателствата?

Трябва да **споделя данни** между хиляди учени с различни интереси.

Трябва да **свързва цели компютърни центрове**, не само отделни компютри.

Трябва да гарантира, че **данните са достъпни навсякъде, по всяко време**.

Трябва да се развива въпреки това да остава **надежден**.

Трябва да се справя с **различните политики на управление на различните центрове**.

Трябва да гарантира **сигурност** при липса на централизиран контрол.

Да улеснява комуникацията между потребителите и доставчиците.

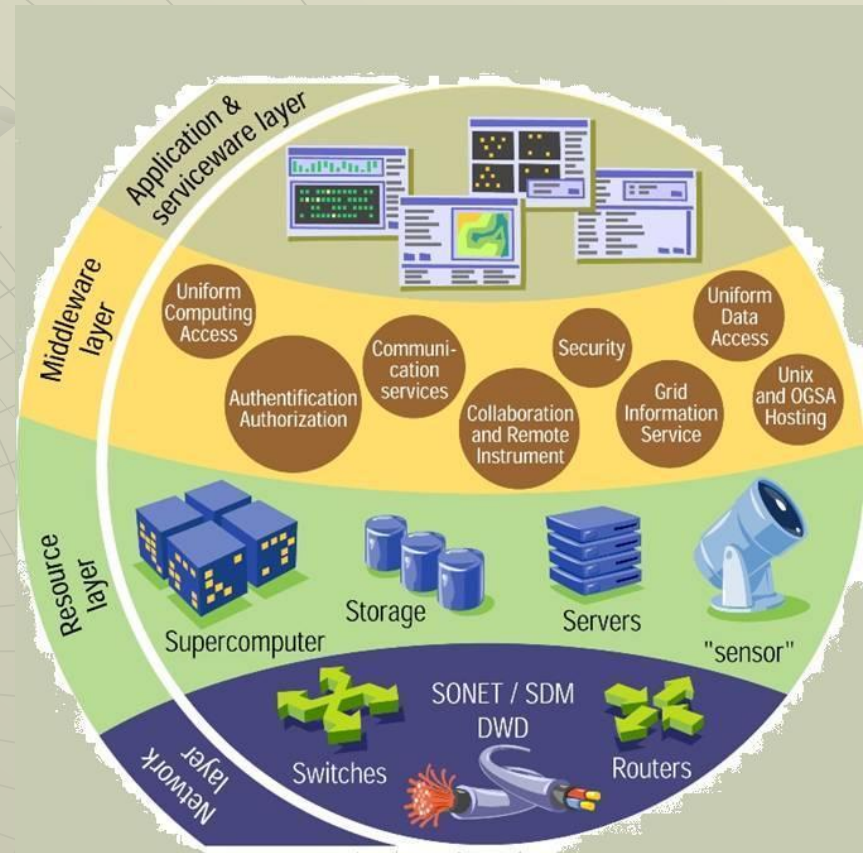
Не е само технически проблем.

Как работи?

Grid разчита на сложен софтуер - **междинен софтуер (мидълуеър)**, който осигурява лесна комуникация между различните компютри в различни части на света.

Информационната система на Grid **не само намира данните**, необходими на учения, но **също програмите за обработката им** и изчислителната мощ за тази обработка.

Разпределя изчислителната задача до **мястото, където има свободни ресурси** и връща резултата обратно.



Предимства

Дава възможност да се изпълняват мащабни компютърни изчисления, обхващащи хиляди компютри за широка гама от приложения.

Прозрачен достъп до разпределени ресурси, позволяващ работа от личния компютър, или дори чрез мобилен телефон.

По-ефективно и лесно сътрудничество между отдалечени групи, както в науката, така и в бизнеса.

Области на приложение

Астрономия и астрофизика

Молекулярна биология

Медицина

Климатология

Геофизика

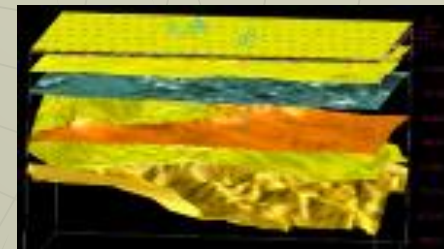
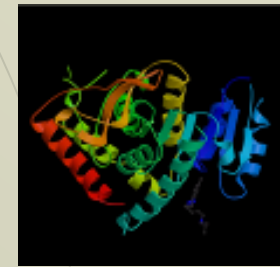
Гражданска защита

Термоядрен синтез

Материалознание

Постоянно се прибавят нови

Използва се рутинно и ежедневно



WLCG

WLCG Worldwide LHC Computing GRID разпределена компютърна инфраструктура, осигуряваща съхраняването и обработката на информацията от експериментите на LHC.

Управлява се, поддържа се и се използва в сътрудничество между експериментите и участващите компютърни центрове от целия свят.

Федерация от 3 ГРИДа:

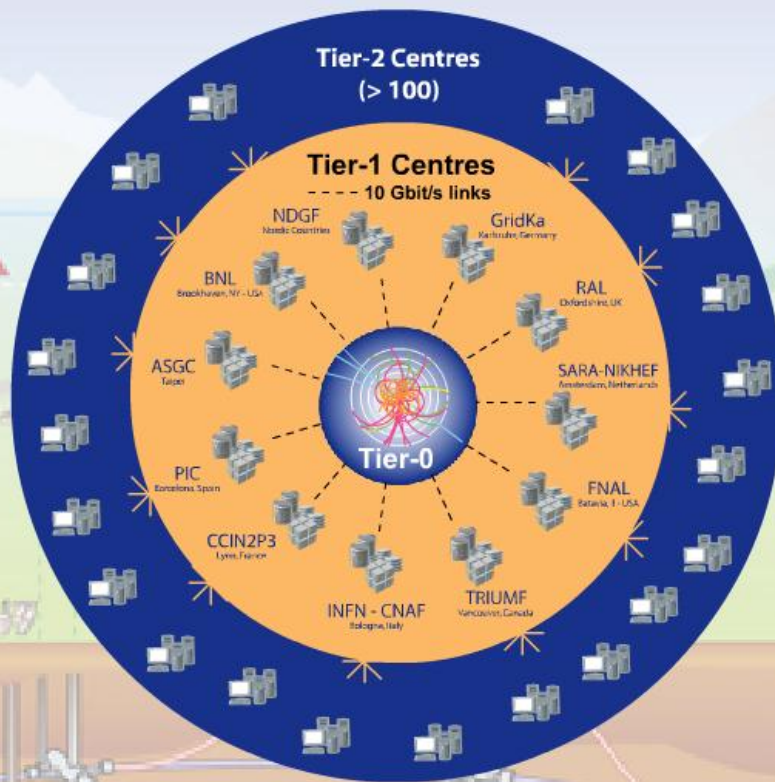
EGI — European Grid Infrastructure

OSG — Open Science Grid

NDGF — Nordic Data Grid Facility



Структура на WLCG



Ниво 0 (ЦЕРН): 15%

- Запис на данните
- Начална реконструкция
- Разпределение на данните

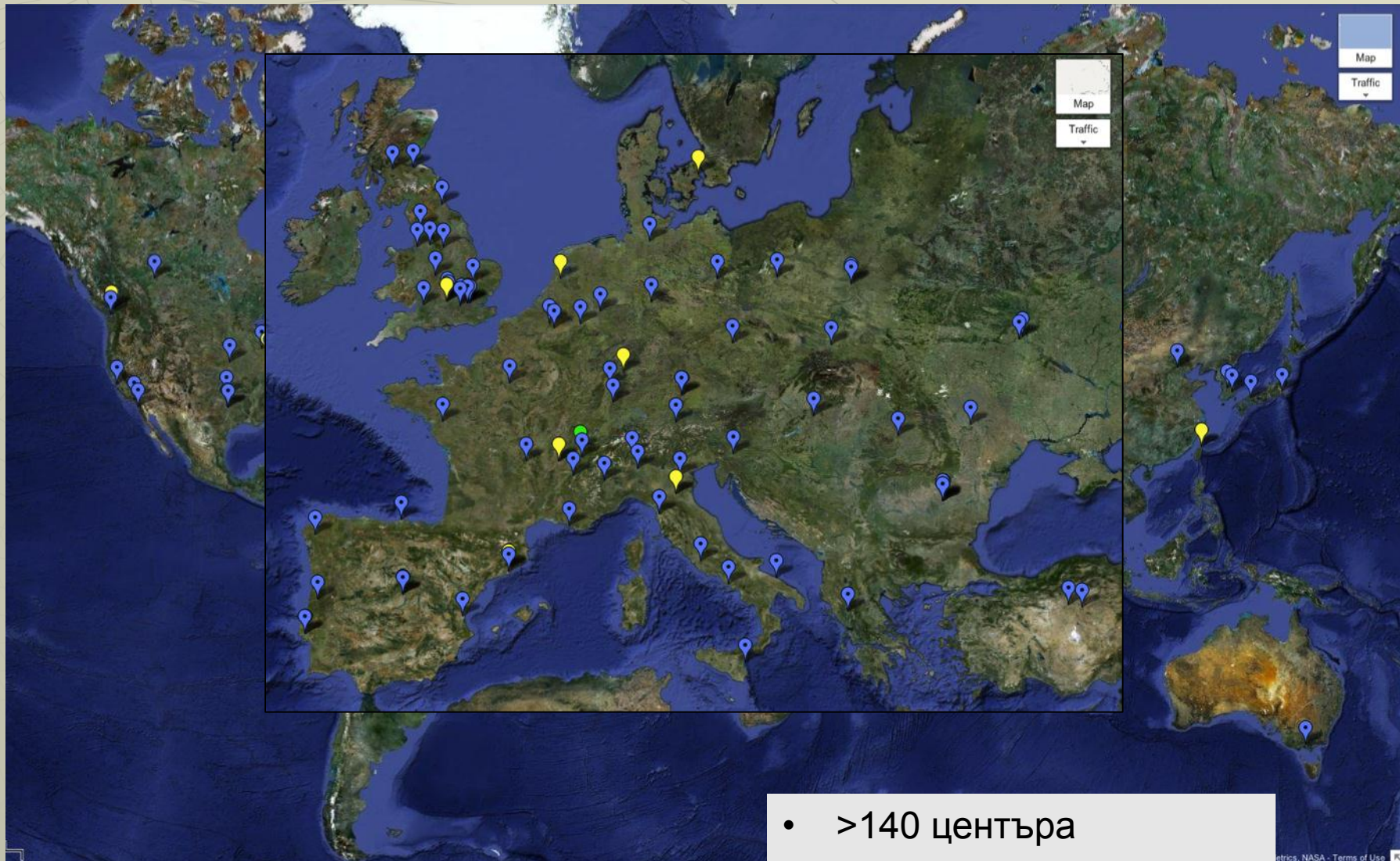
Ниво 1 (11 центъра): 40%

- Постоянно съхранение
- Преработка
- Анализ
- Съвързани с 10 Gb/s
- Лентови библиотеки

Ниво-2 (~130 центъра): 45%

- Симулация
- Анализ

ГРИД центрове в WLCG



Слой 0



Слой 1

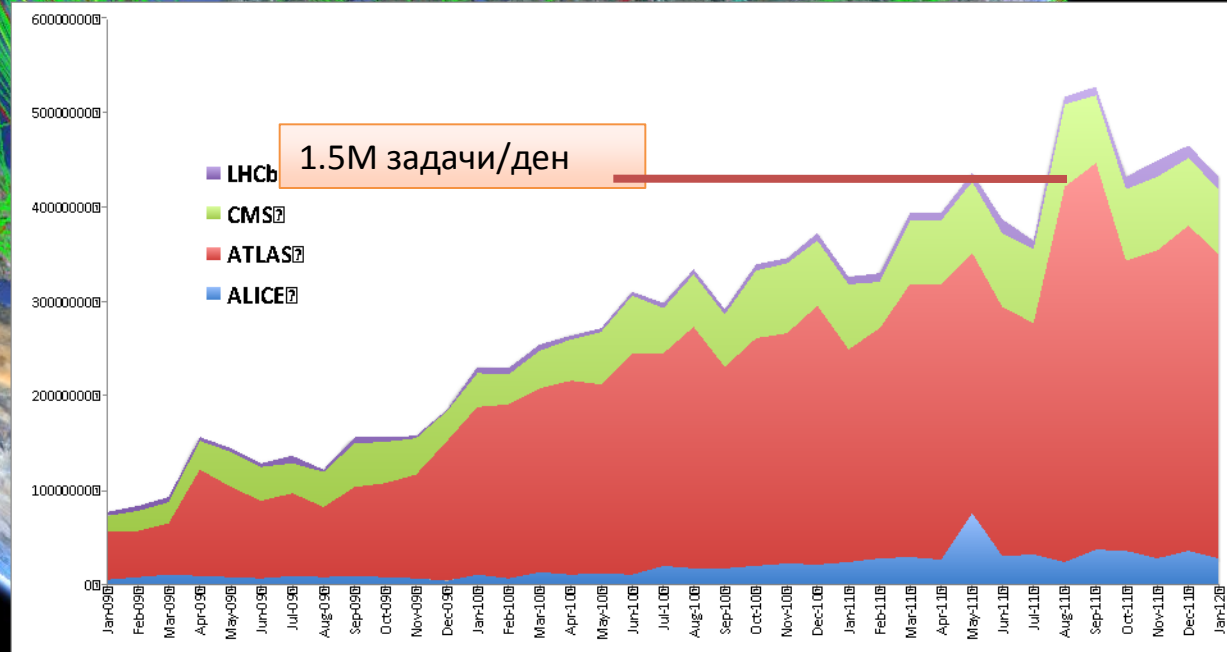


Слой 2

Използване на WLCG

9/7/2012 6:32:44 pm

Running jobs: 252955
Transfer rate: 17.05 GiB/sec

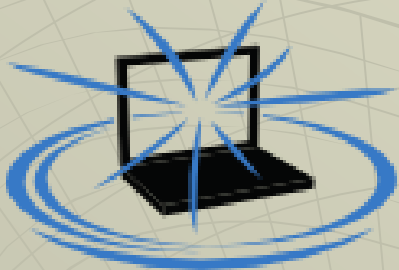


Включете се и вие

Как? Възможно ли е?

Като дарите свободното време на
на вашите компютри за LHC.

Лесно е. Инсталирате софтуера
и се включвате.



LHC@home

<http://lhcatome.web.cern.ch>

Test4theory: Симулиране на взаимодействията между частиците

Sixtrack: Моделиране на движението на протоните в ускорителя

Малък виртуален LHC вкъщи

Volunteer computing

Използва доброволно предоставено компютърно време на стотици хиляди компютри за решаване на научни задачи с голям обществен интерес.

Редовно поддържан уеб сайт за напредъка по проекта.

Обсъждане на научната страна на задачата във форум с участието и на учените и на доброволците.

Понякога любителите дават добри научни идеи и имат принос.

Участниците се състезават поотделно или в отбори.

Множество проекти. Ще ги познаете по @home.

SETI@Home: Търсене на извънземен разум

Climateprediction.net: Моделиране на климата

Folding@Home: Изследвания на заболявания като Алцхаймер

Einstein@Home: Търсене на пулсари

GridRepublic и World Community Grid: събират на едно място множество проекти



Благодаря за вниманието!