



Polska Infrastruktura
Informatycznego Wspomagania Nauki
w Europejskiej Przestrzeni Badawczej

Polska infrastruktura informatycznego wspomagania nauki w europejskiej przestrzeni badawczej – PL-GRID

Mariusz Sterzel

ACK Cyfronet AGH

m.sterzel@cyfronet.pl





Plan prezentacji

- ◆ **Co to jest grid?**
- ◆ **Enabling Grids for E-scienceE**
- ◆ **Europejska Inicjatywa Gridowa (EGI)**
- ◆ **Czym jest projekt PL-Grid**
 - ◆ idea i struktura PL-Grid
 - ◆ dostępne i planowane zasoby
- ◆ **Wykorzystanie gridu w Cyfronecie**
 - ◆ dostępne oprogramowanie
 - ◆ Wirtualne Laboratorium
 - ◆ rozwijane projekty
- ◆ **Jak zostać użytkownikiem**
- ◆ **Sposób pozyskania zasobów**



Czym jest/nie jest grid?

◆ Grid nie jest:

- ◆ nową generacją internetu
- ◆ nowym systemem operacyjnym
- ◆ tylko:
 - sposobem na wykorzystanie wolnych mocy obliczeniowych
 - nowym modelem obliczeń równoległych
 - nowym typem sieci P2P



Czym jest/nie jest grid?

- ◆ **Definicja – Vaidy Sunderam**

infrastruktura umożliwiająca współdzielenie, wybór czy gromadzenie geograficznie rozproszonych zasobów (komputerów, oprogramowania, (baz) danych czy ludzi)

- ◆ **czyli uwspólnionymi (wirtualnie) zasobami**

- ◆ zależnymi od ich dostępności, pojemności czy kosztów
- ◆ przeznaczonymi do rozwiązywania skomplikowanych problemów
- ◆ zgrupowanymi w ramach wirtualnych organizacji



Wg Iana Fostera grid to

- ♦ infrastruktura pozwalająca na współpracę i udostępnianie zasobów dla aplikacji naukowych
- ♦ serwis wykorzystujący otwarte standardy ogólne protokoły i interfejsy
- ♦ zintegrowane zarządzanie rozproszonych serwisów i technologii
- ♦ ‘niespokojna’ technologia umożliwiająca wirtualne i kolektywne tworzenie rozproszonego świata
- ♦ społeczność i technologia Open Source
- ♦ slogan marketingowy
- ♦ wszystkie powyższe razem... ;-)



[Download Folding@home](#)



EGEE

eGEE
Enabling Grids
for E-science

Archeology
Astronomy
Astrophysics
Civil Protection
Comp. Chemistry
Earth Sciences
Finance
Fusion
Geophysics
High Energy Physics
Life Sciences
Multimedia
Material Sciences
...

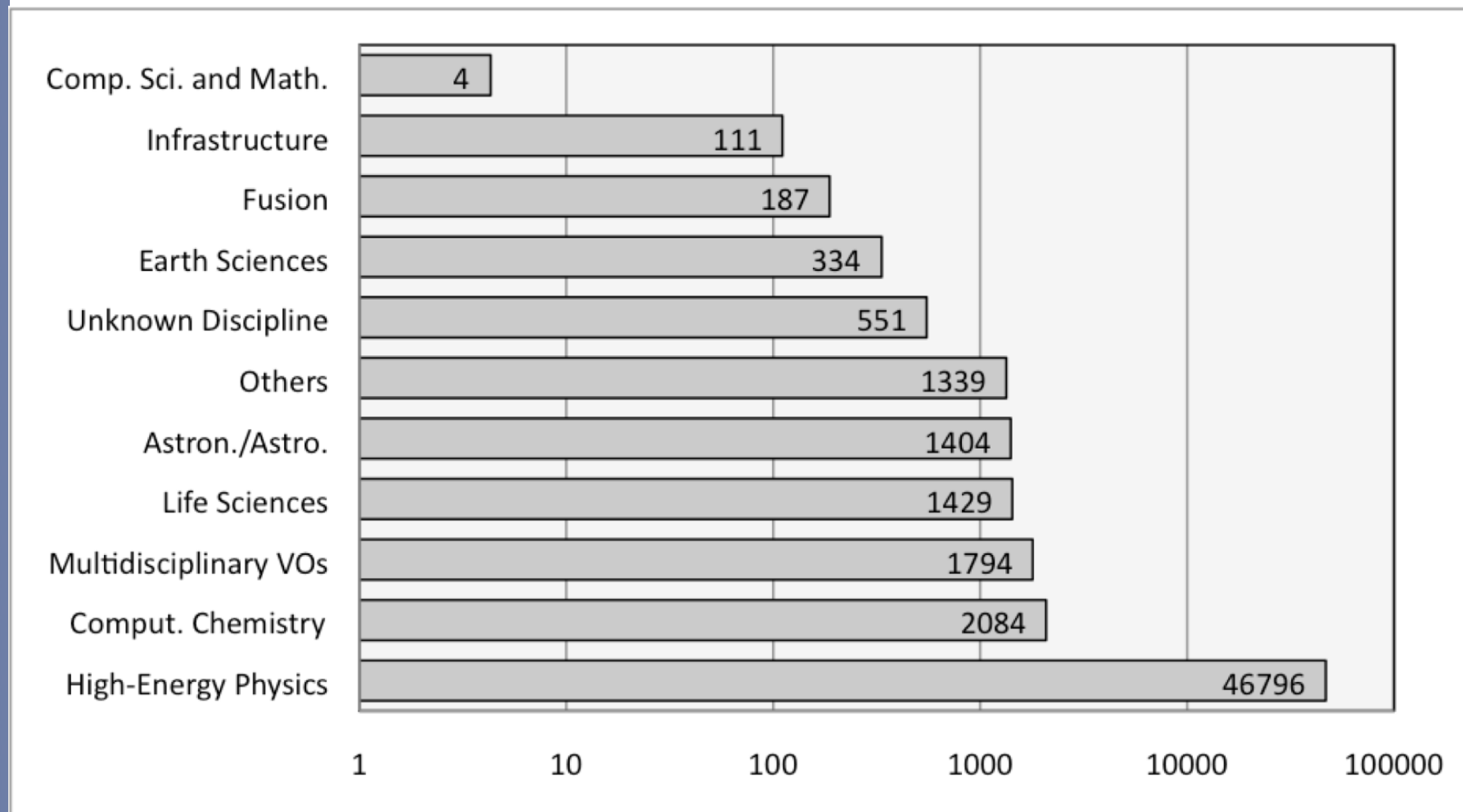
Scheduled = 21549
Running = 25368

>250 sites
48 countries
>100,000 CPUs
>30 PetaBytes
>10,000 users
>150 VOs
>150,000 jobs/day

21:14:04 UTC

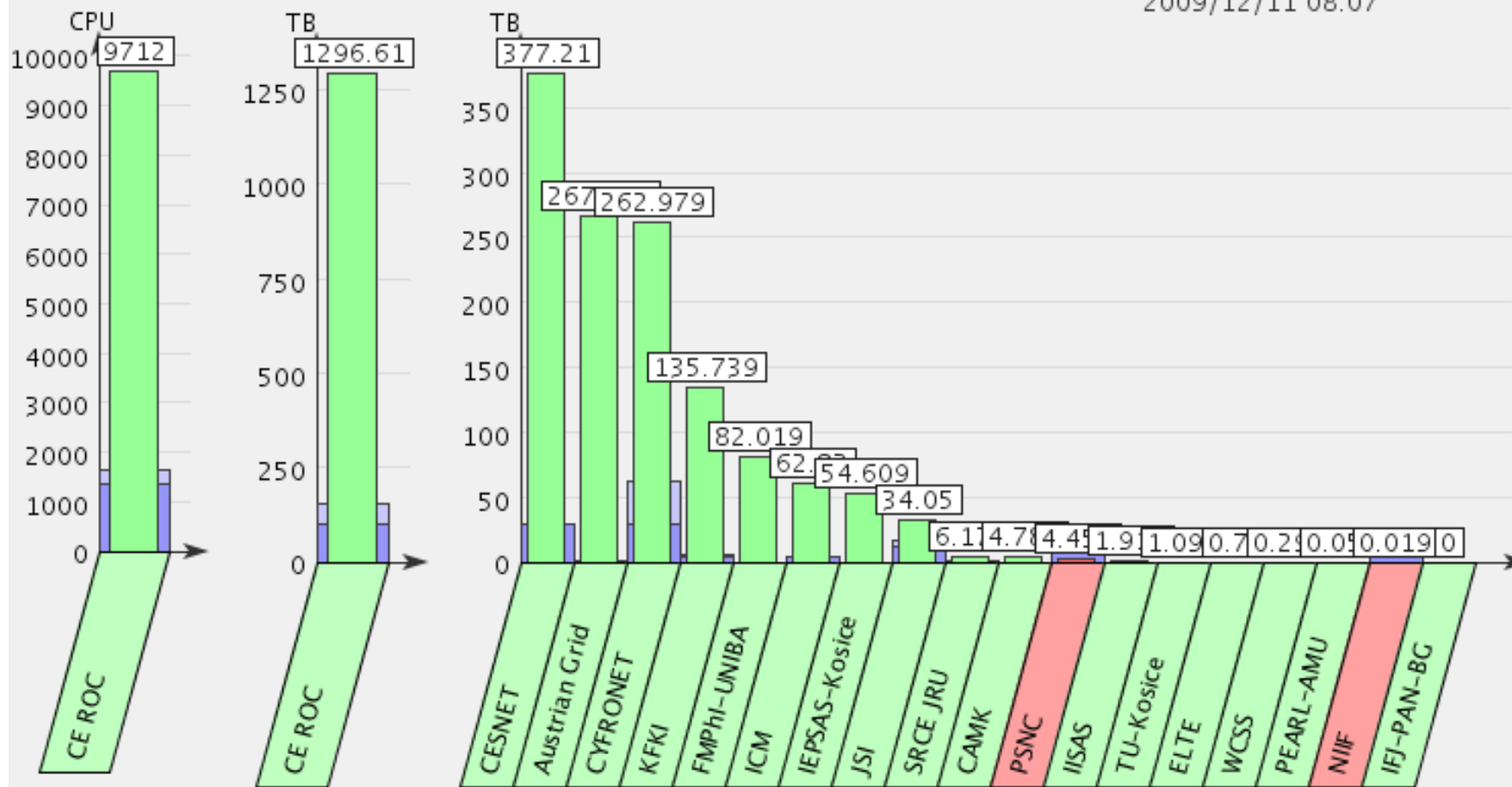


EGEE – wykorzystanie zasobów



Gridowe zasoby dyskowe w rejonie CE

2009/12/11 08:07



■ - Avg CPUs above current commitment ■ - PM 12 EGEE2 (March 2007) commitment
■ - Avg CPUs below current commitment ■ - PM 24 EGEE2 (March 2008) commitment
■ - Avg TBs above current commitment ■ - PM 12 EGEE2 (March 2007) commitment
■ - Avg TBs below current commitment ■ - PM 24 EGEE2 (March 2008) commitment



LHC Grid

- ◆ **Budowa środowiska obliczeniowego dla eksperymentów LHC**
- ◆ **Obszary rozwoju**
 - ◆ Budowa infrastruktury złożonej z klastrów komputerów PC
 - ◆ Rozwój oprogramowania Gridu zwanego middleware (gLite)
 - ◆ Budowa systemów zarządzania i pomocy dla użytkowników
- ◆ **Projekt przyjęty przez CERN Council we wrześniu 2001**
 - ◆ 2001-2007: Instalacja systemu i doprowadzenie do działania
 - ◆ 2005: Rozszerzenie do systemu światowego WLCG opartego o sieci Grid: EGEE, OSG, NorduGrid
 - ◆ 2007: odbiór wstępnej fazy serwisu
 - ◆ 2008/2009: produkcja symulacji dla eksperymentów oraz testy przed pierwszym uruchomieniem akceleratora LHC
- ◆ **Jesień 2009: powtórny start LHC i pierwsze dane**



Europejska Inicjatywa Gridowa

A zaczęło się tak:

**“...for Grids we would like to see
the move towards
long-term sustainable initiatives
less dependent upon
EU-funded project cycles”**

**Viviane Reding, Commissioner, European Commission,
EGEE'06 Conference,
September 25, 2006**

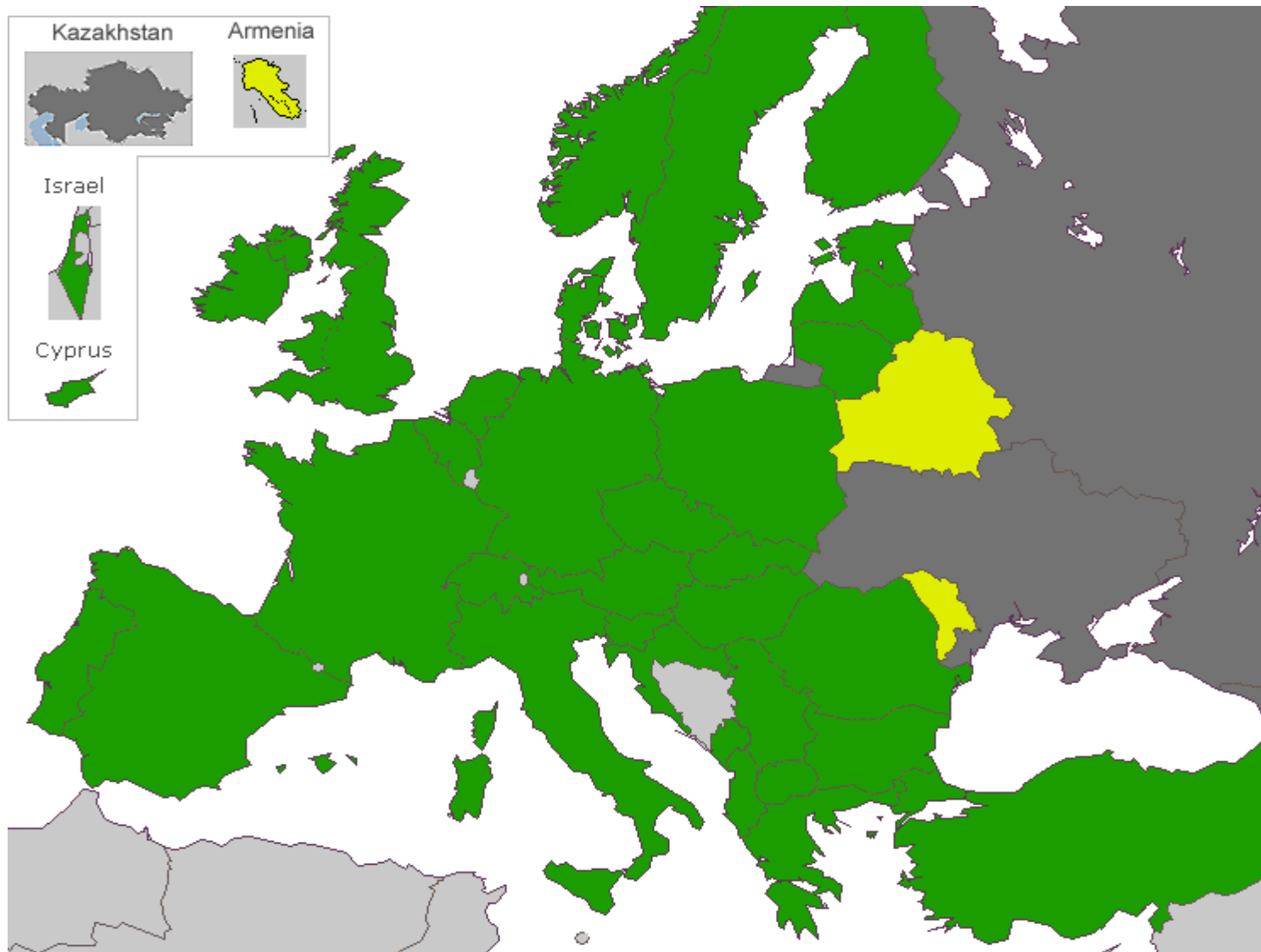
◆ **Cel:**

- ◆ Utworzenie trwałej e-infrastruktury gridowej w Europie
- ◆ Bazującej na Narodowych Inicjatywach Gridowych (NGI)

◆ **Każda NGI powinna:**

- ◆ reprezentować wszystkie zainteresowane strony w danym kraju
- ◆ zmobilizować się i uzyskać fundusze na działanie ze środków własnych danego kraju
- ◆ zbudować i zarządzać narodową e-infrastrukturą gridową
- ◆ wspierać społeczności naukowe
- ◆ brać udział w działaniach międzynarodowych, wносить wkład i jednocześnie wdrażać międzynarodowe standardy

Kraje zrzeszone w EGI





Idea PL-Grid

- ◆ **Otwarta ogólnopolska infrastruktura gridowa wspierająca uprawianie nauki w sposób umożliwiający integrację danych doświadczalnych i wyników zaawansowanych symulacji komputerowych**
 - ◆ badania mogą być prowadzone przez geograficznie rozproszone zespoły
- ◆ **Dostarczanie polskiej społeczności naukowej usług informatycznych opartych na gridowych klastrach komputerowych, służących e-Nauce w różnych dziedzinach**
 - ◆ infrastruktura musi być kompatybilna i interoperabilna z gridem europejskim i światowym
- ◆ **System skalowalny, pozwalający na dołączenie lokalnych klastrów komputerowych uczelni, instytutów badawczych czy „platform technologicznych”**
 - ◆ może być wykorzystany przez administrację państwową, zespoły zarządzania kryzysowego czy jednostki przemysłowe

Skład konsorcjum

- ◆ Akademickie Centrum Komputerowe CYFRONET AGH – koordynator
- ◆ Poznańskie Centrum Superkomputerowo-Sieciowe (PCSS)
- ◆ Wrocławskie Centrum Sieciowo – Superkomputerowe (WCSS)
- ◆ Centrum Informatyczne Trójmiejskiej Akademickiej Sieci Komputerowej (TASK)
- ◆ Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego (ICM)





Oferta PL-Grid

- ◆ **Docelowy wzrost udostępnianej mocy obliczeniowej do ok. 215 TFLOPs (5000 procesorów 4-rdzeniowych) oraz pamięci dyskowej o ok. 2500 TB**
- ◆ **Zestaw zaawansowanych narzędzi do organizacji eksperymentów obliczeniowych, który pomożemy dostosować do indywidualnych potrzeb**
- ◆ **Pomoc w zrozumieniu zagadnień związanych z uruchamianiem aplikacji naukowych na rozległych zasobach obliczeniowych**
- ◆ **Wsparcie technologiczne i informatyczne przy projektowaniu własnych aplikacji naukowych i ich wdrażaniu na infrastrukturze PL-Grid**



Oprogramowanie

- ♦ Jaka aplikacja może być przystosowana do użytkowania w PL-Grid?

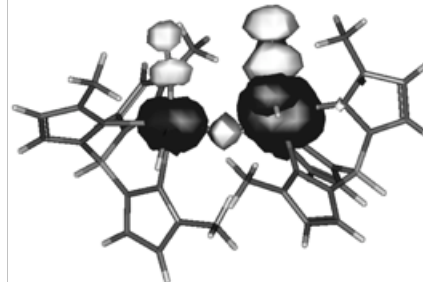
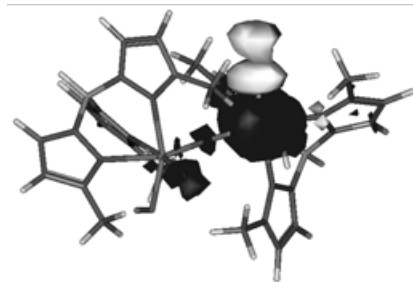
Każda!
(niektóre po lekkich modyfikacjach)

- ♦ Programy sekwencyjne – bez zmian
- ♦ Aplikacje równoległe (MPI) – bez zmian
- ♦ Programy interaktywne – wymagana zmiana w komunikacji z użytkownikiem
- ♦ Operacje na wielkich zbiorach danych – zmiana dostępu do danych

Możliwość jednoczesnego uruchamiania kilkudziesięciu/
kilkuset i więcej instancji danego programu

Programy QM

- ◆ Programy do szerokiego spektrum obliczeń kwantowochemicznych:
 - ◆ Gaussian
 - prosty w obsłudze, stosowany powszechnie
 - zaimplementowana wielość funkcji
 - ◆ Turbomole
 - bardzo dobrze zaimplementowane metody symulacji rozpuszczalnika
 - szybkie metody do opisu stanów wzbudzonych
 - możliwość przeprowadzania łączonych obliczeń QM/MM (zastosowanie m.in. QPOT)

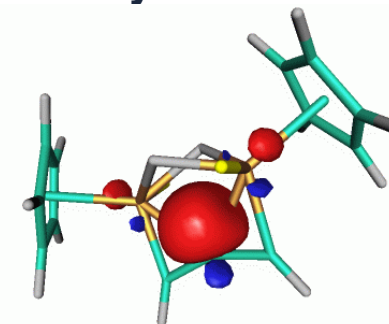


Programy QM

Programy do szerokiego spektrum obliczeń kwantowochemicznych:

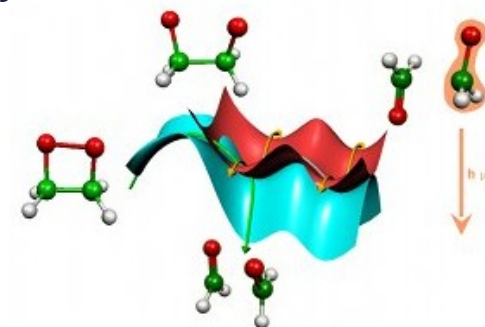
◆ ADF

- ◆ szybki (z racji obliczeń w bazach slaterowskich)
- ◆ szeroko stosowany w przypadku metali przejściowych
- ◆ obliczenia własności magnetycznych



◆ Molcas

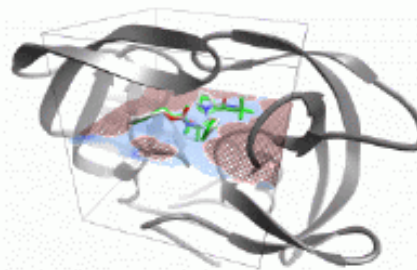
- ◆ bardzo dobra implementacja metod CASSCF oraz CASPT2:
- ◆ opis kompleksów metali przejściowych
- ◆ opis przejść pomiędzy różnymi stanami
- ◆ elektronowymi w czasie reakcji chemicznych



Programy wykorzystywane w biologii

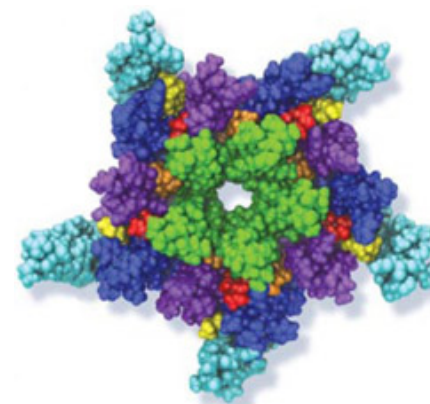
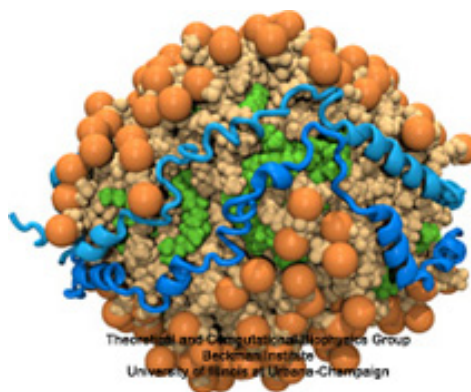
◆ AutoDock

- ◆ program wykorzystywany do przewidywania miejsc dokowania się ligandów do białek



◆ NAMD

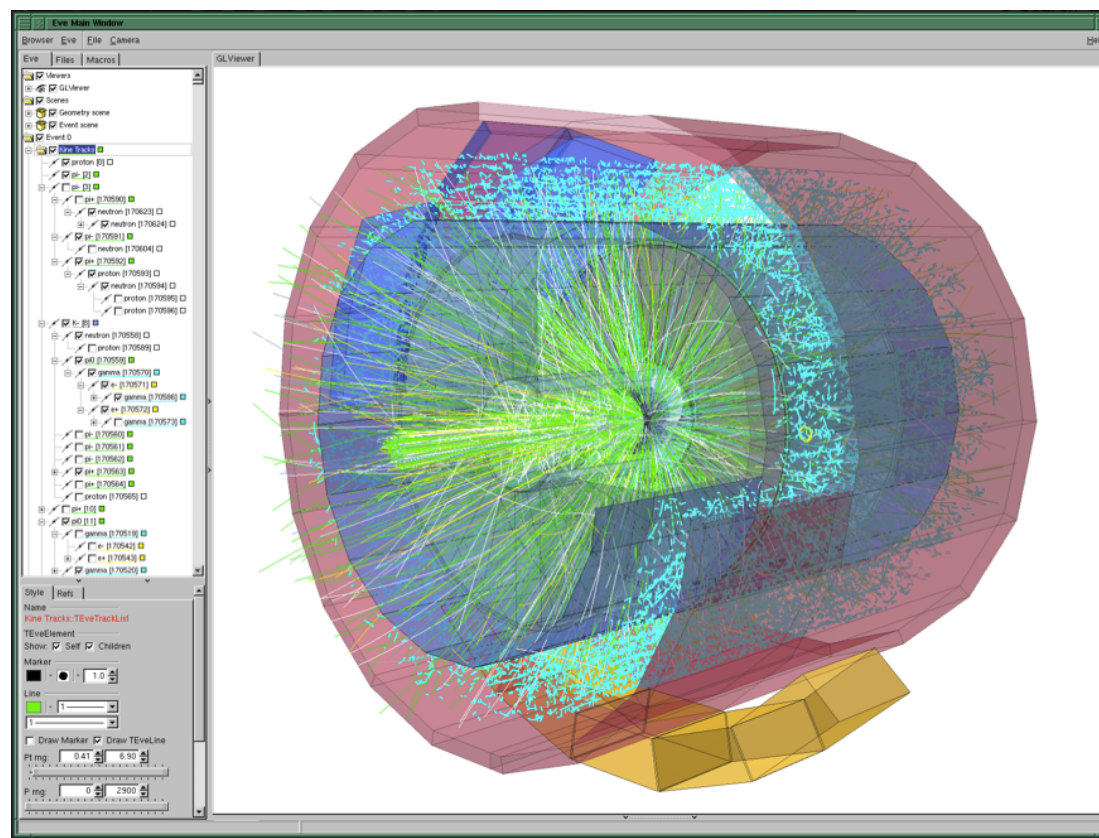
- ◆ program do przeprowadzania symulacji MM/MD dużych oraz bardzo dużych systemów biologicznych
- ◆ przeznaczony szczególnie do obliczeń równoległych na klastrach/gridach



Programy wykorzystywane w fizyce

◆ ROOT

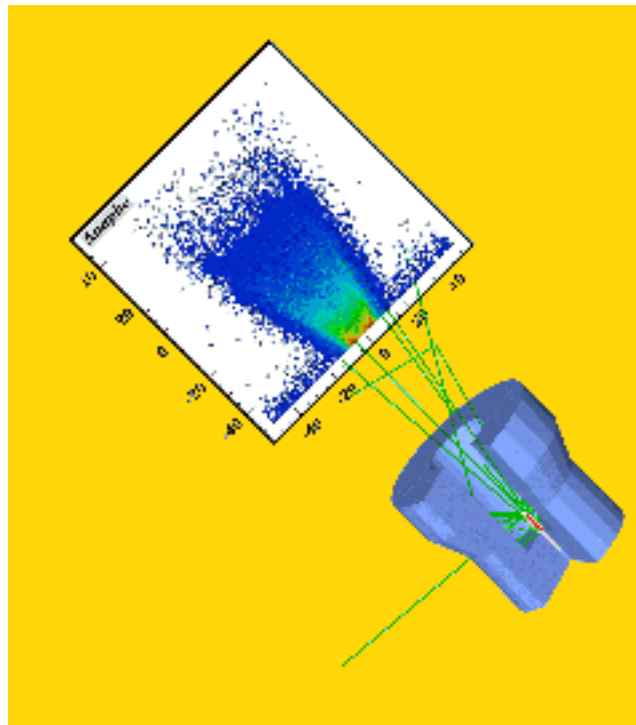
- ◆ Zaawansowany pakiet wykorzystywany do analizy wielkiej ilości danych doświadczalnych



Programy wykorzystywane w fizyce

◆ GEANT

- ◆ Narzędzie do modelowania przechodzenia cząstek przez materię. Wykorzystywany jest przez fizyków wysokich energii jak również w badaniach kosmosu a nawet medycynie



Oprogramowanie...

- ◆ **Kompilatory**
 - ◆ Intel
 - ◆ Portland
- ◆ **Biblioteki numeryczne**
 - ◆ MKL
 - ◆ ACML
 - ◆ GotoBLAS
- ◆ **Obliczenia równoległe**
 - ◆ MPICH
 - ◆ OpenMPI
 - ◆ Intel MPI



Wasze programy

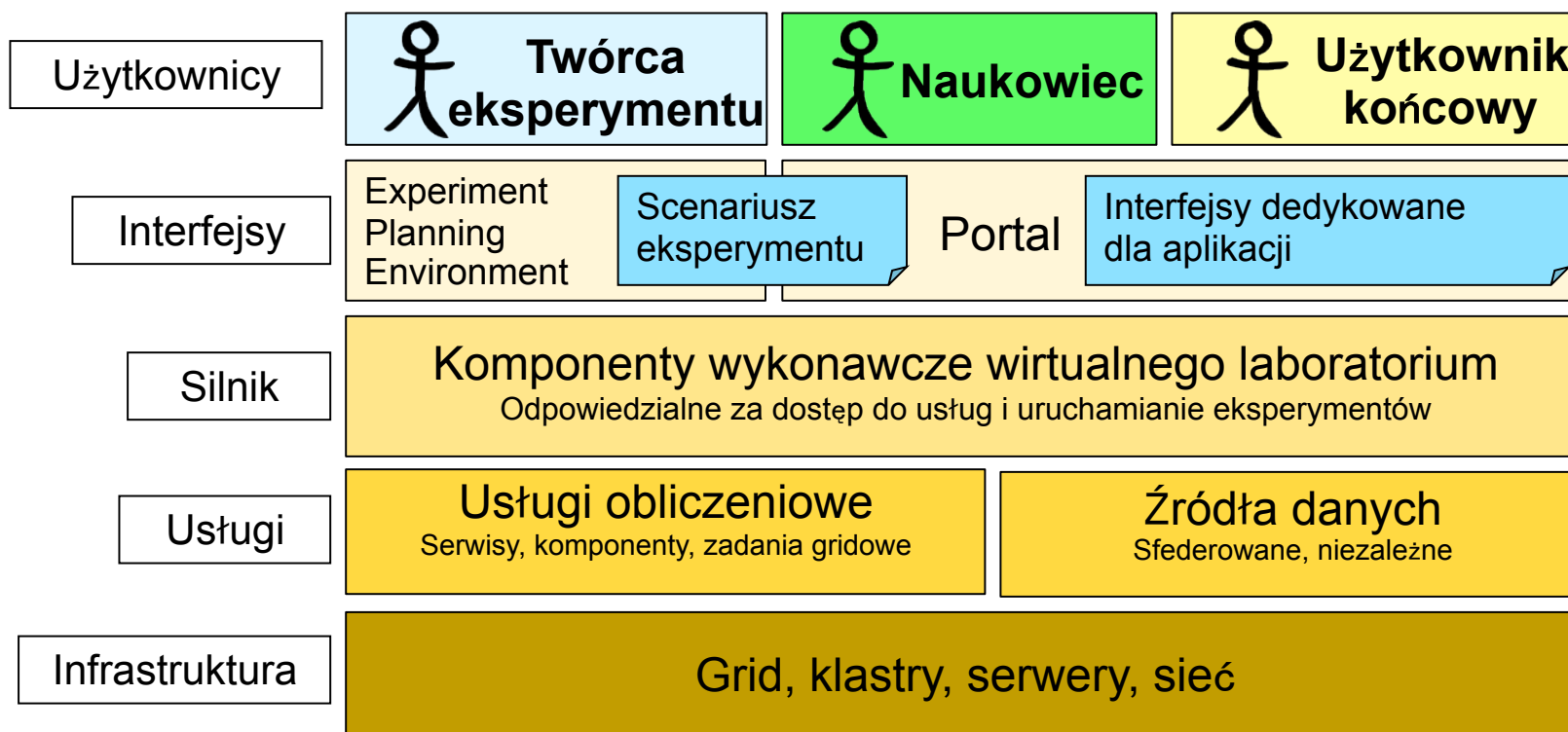
- ◆ **Grid umożliwia wydajną pracę praktycznie wszystkich aplikacji UNIXowych**
 - ◆ programów komercyjnych i darmowych
 - ◆ programów autorskich napisanych w ogólnodostępnych językach programowania tj. C/C++, FORTRAN, Python, Ruby etc.

Czekamy na propozycje potrzebnych Wam programów!



Wirtualne Laboratorium GridSpace

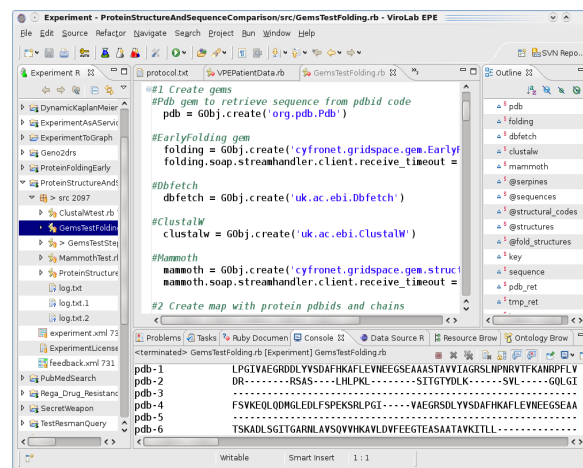
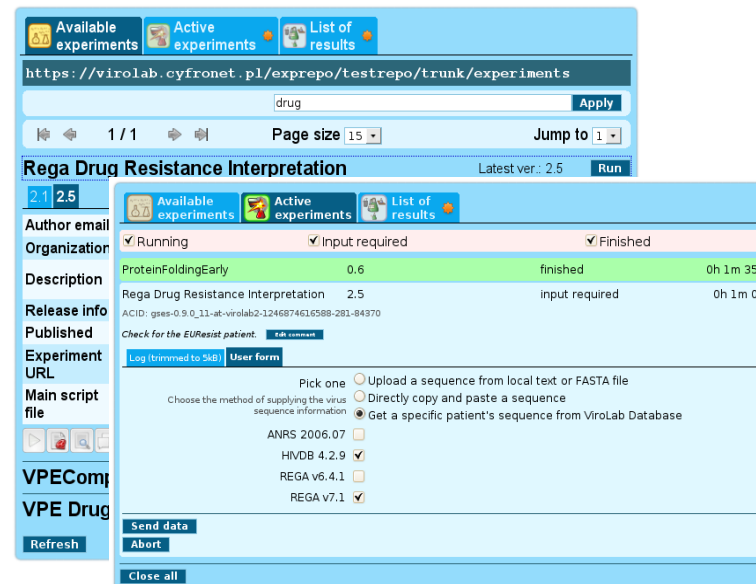
- ◆ Wirtualne laboratorium GridSpace umożliwia łączenie rozproszonych zasobów obliczeniowych oraz źródeł danych.
- ◆ Dostarcza użytkownikom wygodnych, wysokopoziomowych narzędzi do planowania i wykonywania eksperymentów *in-silico*.





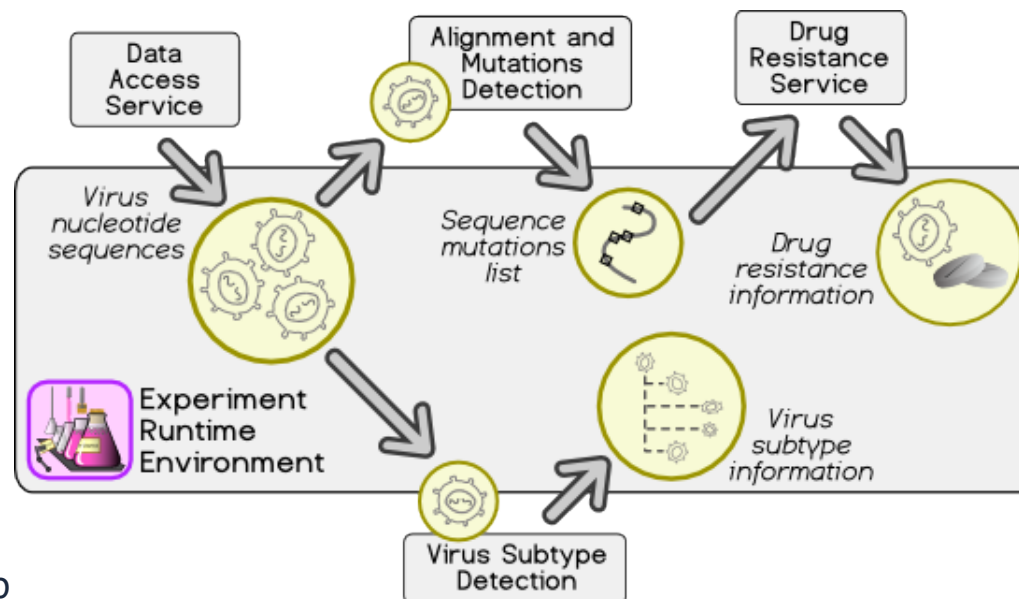
Zastosowania wirtualnego laboratorium

- ◆ Przetwarzanie dużych ilości danych
- ◆ Łączenie wielu programów, narzędzi i usług do pobierania, przetwarzania danych oraz obliczeń – dostępne poprzez jednolity interfejs
- ◆ Uruchamianie aplikacji w różnych technologiach (lokalne, zdalne, klastry, grid, komponenty, serwisy)
- ◆ Planowanie eksperymentów; projektowanie, tworzenie i modyfikowanie ich przebiegu na podstawie wyników
- ◆ Łatwe testowanie zmian i rozmaitych algorytmów
- ◆ Zarządzanie rezultatami
- ◆ Tworzenie interfejsów WWW



Przykład eksperymentu - ViroLab

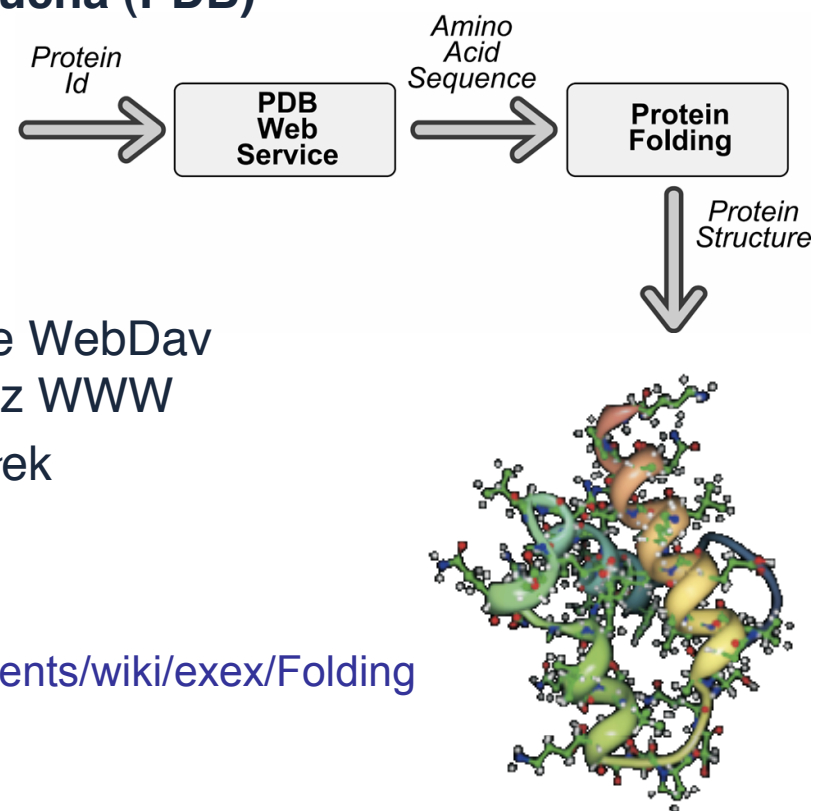
- ◆ **Dane wejściowe z laboratorium**
 - ◆ badanie pacjenta
 - ◆ sekwencje genetyczne wirusa HIV umieszczone w bazie danych
- ◆ **Eksperyment *in-silico* w wirtualnym laboratorium**
 - ◆ sekwencje nukleotydowe wirusa HIV pobrane z bazy danych
 - ◆ dopasowanie sekwencji i/lub określenie ich typu za pomocą właściwych algorytmów
 - ◆ oporność wirusa jest wyznaczana poprzez wywołanie usługi obliczeniowej (DRS)



<http://virolab.cyfronet.pl>
<http://www.virolab.org>

Przykład eksperymentu – zwiżanie białka

- ◆ **Cel:**
 - ◆ demonstracja zastosowania wirtualnego laboratorium dla aplikacji bioinformatycznych
- ◆ **Wejście: identyfikator białka i łańcucha (PDB)**
- ◆ **Rezultat: struktura 3D**
- ◆ **Wykorzystane usługi:**
 - ◆ Serwis Protein Data Bank (PDB)
 - ◆ Fałdowanie metodą Early-Stage
 - ◆ Zapisanie rezultatów na serwerze WebDav celem zapewnienia dostępu przez WWW
 - ◆ Powtarzanie w pętli dla wielu białek



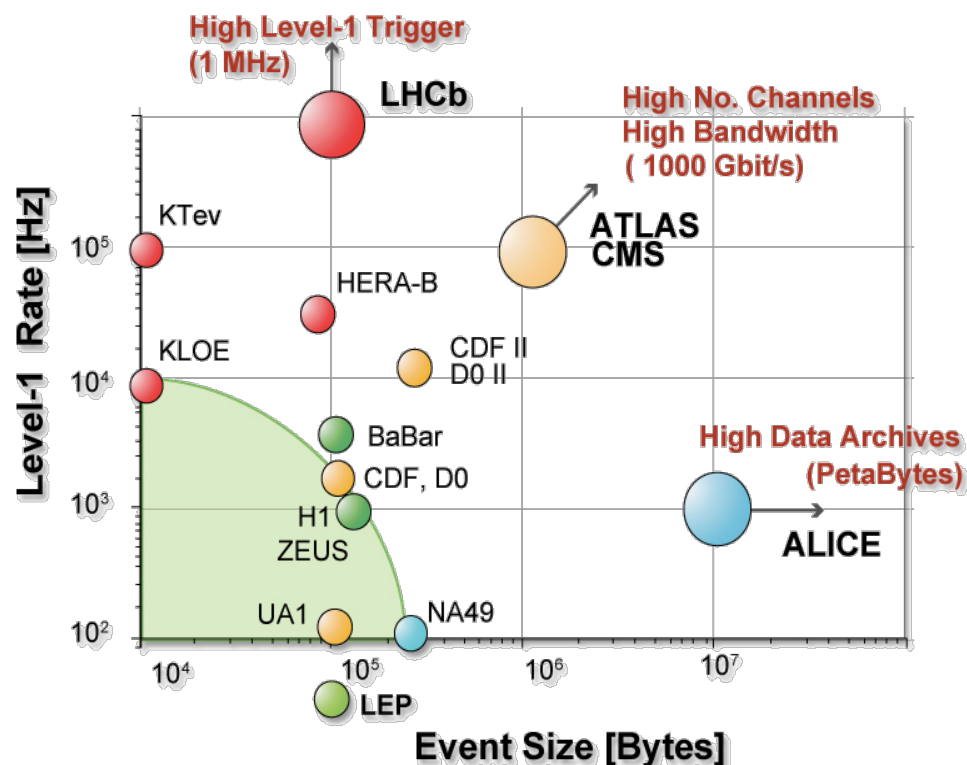
<http://virolab.cyfronet.pl/trac/exampleExperiments/wiki/exex/Folding>

Wymagania LHC

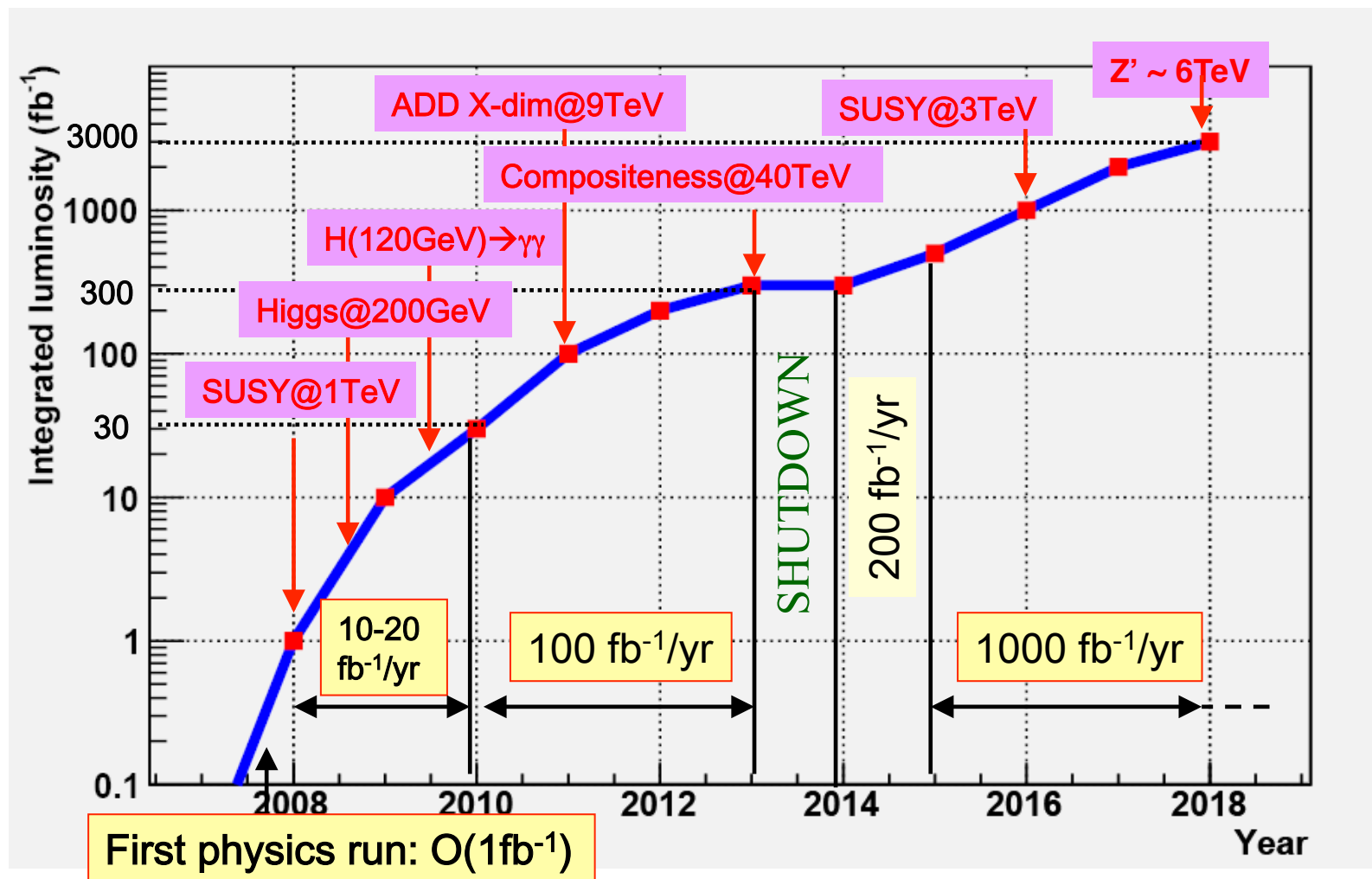
Wysoka częstość zderzeń w akceleratorze LHC i duże rozmiary przypadku rejestrowanego przez detektor stawiają ekstremalne wymagania systemom obliczeń i zapisu danych

Energia	7 TeV
Światłość	$10^{34} \text{ cm}^{-2} \text{ s}^{-1}$
Liczba zderzeń na przecięcie wiązek	≈ 23
Wysoka częstość przecięć	40 MHz
Rozmiar przypadku	$\approx 1.6 \text{ MB}$

CPU	100,000 PC
Pamięć masowa	20,000 TB/rok



Przyszłość LHC





Projekt nEDM

- ◆ Projekt mający na celu bardzo dokładne pomiary momentu dipolowego neutronu o rozdzielczości kilku 10^{-28} e.cm
- ◆ Ponad 12 partnerów
- ◆ Strategia: „ultra-zimne” neutrony przechowywane w próżni w temperaturze pokojowej
- ◆ Obecnie realizacja fazy I i II projektu – test komponentów i budowa aparatu pomiarowego
- ◆ Organizacja wirtualna nEDM umożliwia współpracę i wymianę danych pomiędzy partnerami
- ◆ Wymagane zasoby dyskowe – ok. 20 TB



Jak zostać użytkownikiem

- ◆ Każda osoba związana z polską nauką może korzystać z zasobów PL-Grid
- ◆ Aby zostać użytkownikiem należy zarejestrować się oraz uzyskać odpowiedni certyfikat
- ◆ Procedura rejestracji oraz dalsze informacje dostępne są na stronie:

<http://www.plgrid.pl>

- ◆ Z chęcią odpowiemy na pytania:

m.sterzel@cyfronet.pl



www.plgrid.pl



Polish Infrastructure
for Supporting Computational Science
in the European Research Space



[Home](#) [Wiki](#) [Intranet](#)

[Zaloguj](#)

Start

Aktualności

PL-Grid na Seminarium Wydziałowym AGH



skali.

Zapraszamy na seminarium poświęcone ofercie projektu PL-Grid, dla użytkowników zainteresowanych obliczeniami dużej

[Więcej...](#)

Kalendarz PL-Grid na 2010



Kalendarz ścienny projektu PL-Grid na rok 2010 został przygotowany do pobrania w formacie A3.

[Więcej...](#)

Witamy na stronach projektu PL-Grid

W ramach projektu PL-Grid powstaje otwarta, ogólnopolska infrastruktura gridowa, która będzie wspierać badania naukowe w sposób umożliwiający integrację danych doświadczalnych i wyników zaawansowanych symulacji komputerowych, prowadzonych przez geograficznie rozproszone zespoły.

[więcej o projekcie](#)

[kontakt](#)

[nasza oferta](#)

[helpdesk](#)
zgłoś problem

[wypróbuj grid](#)
zostań użytkownikiem

[szkolenia](#)
zgłoś swój udział

Dlaczego warto korzystać z infrastruktury PL-Grid?

Udostępniamy nowoczesny sprzęt. Uzyskasz dostęp do zasobów o dużej mocy – docelowo 215 Tflops, które znacznie przyspieszą Twoje obliczenia.

Prezentacje i demonstracje projektu PL-Grid

Rozwój Polskiej Infrastruktury Gridowej w ramach Projektu PL-Grid (prezentacja)



Polska Infrastruktura
Informacyjnego Wspomagania Nauki



Wydarzenia

Seminarium poświęcone ofercie projektu PL-Grid
Kraków, Polska,
2009-12-11 - 2009-12-11

OGF28 - 28th Open Grid Forum
Monachium, Niemcy,
2010-03-15 - 2010-03-19

5th EGEE User Forum
Uppsala, Szwecja,
2010-04-12 - 2010-04-15

[Poprzednie wydarzenia...](#)

[Nadchodzące wydarzenia...](#)

[Nowe broszury](#)



Procedura rejestracji

Wysyłamy email na plgrid-rejestracja@rt.grid.cyfronet.pl zawierający:

1. Nazwisko:
2. Imię:
3. Login:
4. Instytucja naukowa:
5. Dziedzina nauki:
6. Tematyka badań:
7. Adres do rekordu w bazie „Ludzie Nauki” OPI:
8. Jestem / Nie jestem pracownikiem projektu PL-Grid
9. Numer telefonu komórkowego:

Oświadczam, że wyrażam zgodę na przetwarzanie moich danych osobowych zgodnie z ustawą z dnia 29 sierpnia 1997r. o ochronie danych osobowych (Dz. U. 1997 r. Nr 133, poz. 883). Administratorem danych jest ACK CYFRONET AGH.

Zobowiązuję się do przestrzegania
[Regulaminu Użytkownika Infrastruktury PL-Grid.](#)

Rejestracja cd.

- ◆ **W razie braku jakichś danych użytkownik będzie poproszony o ich weryfikację:**
 - ◆ Zwykle baza „Ludzie Nauki” w OPI nie zawiera adresu e-mail. Procedura uaktualnienia bazy OPI przez użytkownika zajmuje kilka dni
- ◆ **Gdy wszystkie dane są kompletne**
 - ◆ użytkownik otrzymuje e-mailem dane logowania, linki do dokumentacji itd.
 - ◆ hasło – e-mailem lub w SMS
- ◆ Cała procedura zajmuje zwykle ok. tygodnia



Rejestracja – wkrótce

Application

plgrid.pl https://team.plgrid.pl/rejestracja/ pierśi

mBank

nedm kozela – Szukaj w Goo... spotkaniaco:071209 – PL-Grid spotkaniaco:071209 – PL-Grid Application

Portal Użytkowników Infrastruktury PL-Grid

Login

Hasło

Zaloguj

Zapomniałem hasła
Zarejestruj się

Prawa autorskie Projektu PL-Grid

Find: co może Next Previous Highlight all Match case

Done sterzel ABP Z

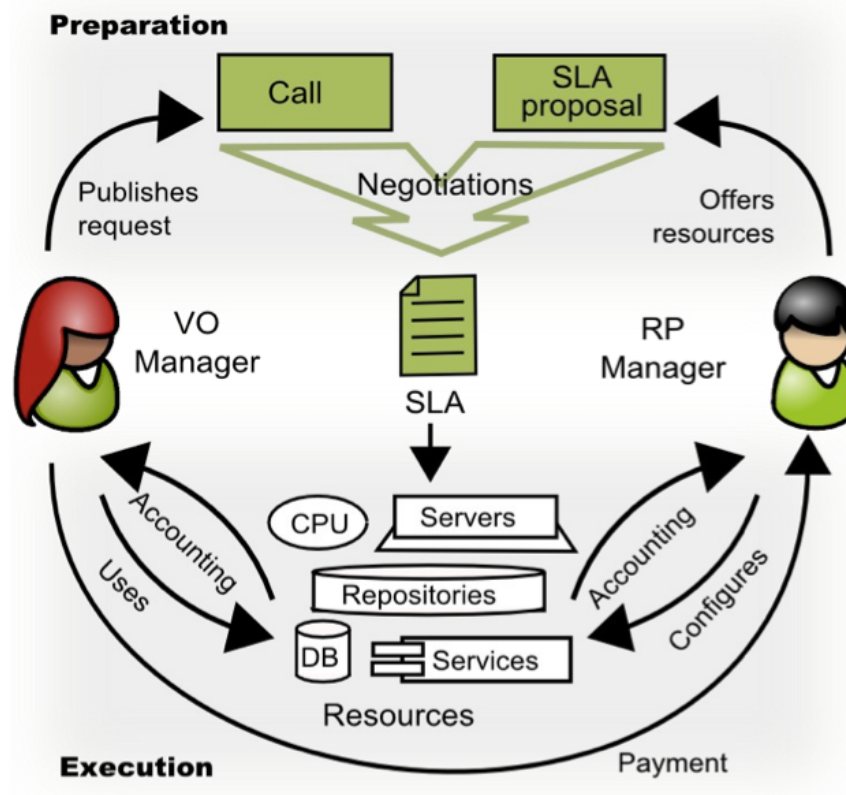


Granty obliczeniowe

- ◆ **Cel:**
 - ◆ dostarczanie informacji o aktualnym wykorzystaniu zasobów przez grupy naukowe
 - CPU
 - zasoby dyskowe
 - ◆ gwarantowane zasoby obliczeniowe dla grup badawczych
 - ◆ nagradzanie najbardziej aktywnych (naukowo) użytkowników
- ◆ **Przyznawane zwykle na okres 1 roku**
 - ◆ uwzględniają zarówno obliczenia gridowe jak i lokalne
- ◆ **Odnowienie grantu możliwe jedynie w przypadku rozliczenia poprzedniego**

Bazaar

- ◆ Narzędzie do dynamicznego występowania o zasoby obliczeniowe i ich przydzielania
- ◆ zintegrowane z systemem przydziału grantów i rozliczania



Manage call

Here you can change your call properties and requirements. If you want your call to be visible to another people, set field 'Call status' to PUBLISHED.
Bold fields are mandatory.

Your call

ELEMENTARY INFORMATION

Call ID: 449

Call name:

Responsible person: Malgorzata Tomanek

VO Name: 

Computation start date: 

Computation closing date:  ☐ Unspecified date

Estimated number of CPU/cores:

Estimated storage space [in GB]:

Call status: 

Call open from:

Call open to:

Description:

April, 2010						
Today						
wk	Sun	Mon	Tue	Wed	Thu	Fri Sat
12					1	2 3
13	4	5	6	7	8	9 10
14	11	12	13	14	15	16 17
15	18	19	20	21	22	23 24
16	25	26	27	28	29	30
Tue, Apr 13						



Pytania?