



# Efektywne wykorzystanie zasobów PLGrid w chemii obliczeniowej

Klemens Noga

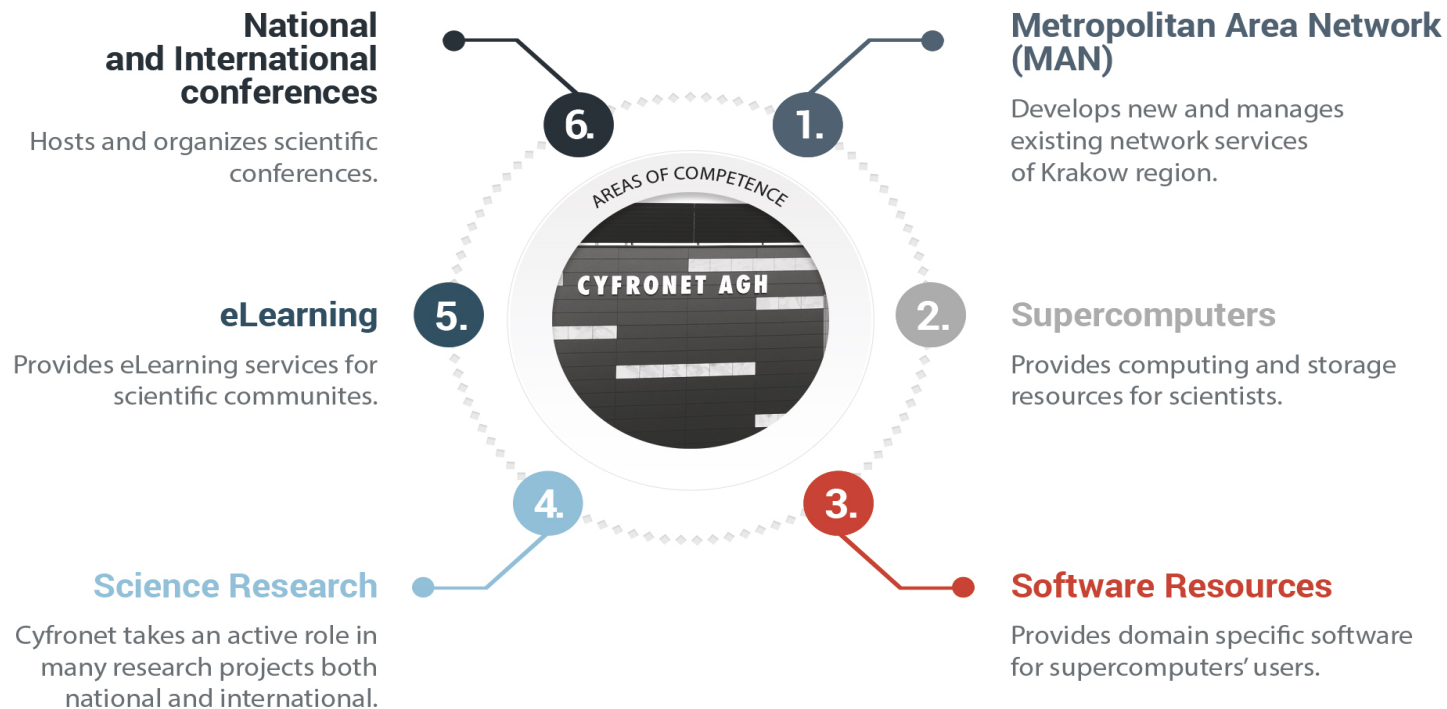
ACK Cyfronet AGH

Wydział Chemii Uniwersytetu Jagiellońskiego, Kraków 24 XI 2016

ACK Cyfronet AGH

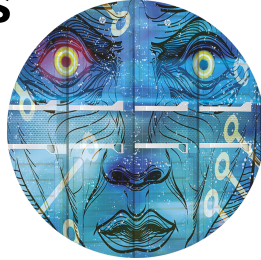


- The biggest Polish Academic **Computer Centre**
  - **Nearly 45 years of experience** in IT provision
  - Centre of excellence in **HPC** and **Grid Computing**
  - Home for **Prometheus** and **Zeus supercomputers**
- Legal status: an **autonomous** within AGH University of Science and Technology
- Staff: >150 , ca. 60 in R&D
- Leader of **PLGrid**: Polish Grid and Cloud Infrastructure for Science
- NGI Coordination in **EGI e-Infrastructure**



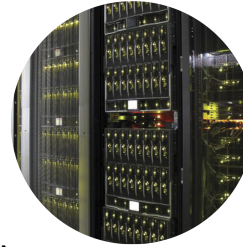
## Prometheus

- 2.4 PFLOPS
- 53 568 cores
- 1<sup>st</sup> HPC system in Poland (77<sup>th</sup> on Top 500, 38<sup>th</sup> in 2015)



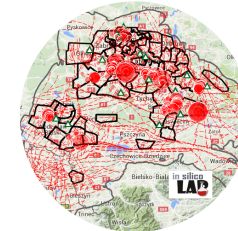
## Zeus

- 374 TFLOPS
- 25 468 cores
- 1<sup>st</sup> HPC system in Poland (from 2009 to 2015, highest rank on Top500 – 81<sup>st</sup> in 2011)



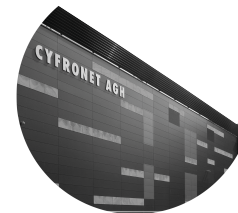
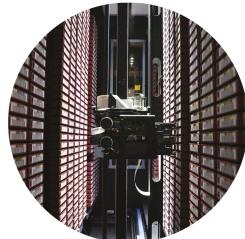
## Computing portals and frameworks

- OneData
- PLG-Data
- DataNet
- Rimrock
- InSilicoLab



## Storage

- 48 PB
- hierarchical data management



## Data Centres

- 3 independent data centres
- dedicated backbone links

## Computational Cloud

- based on OpenStack

## Research & Development

- distributed computing environments
- computing acceleration
- machine learning
- software development & optimization

# Infrastruktura **PL**Grid



## Zasoby obliczeniowe

- 5+ PTFLOPS
- 130 000+ rdzeni



## Oprogramowanie naukowe

- 600+ aplikacji, narzędzi, bibliotek
- [apps.plgrid.pl](https://apps.plgrid.pl)



## Zasoby dyskowe

- 60+ PB
- archiwa
- kopie zapasowe
- rozproszony dostęp
- szybkie zasoby na pliki tymczasowe



## Narzędzia pracy zespołowej

- zarządzanie projektami (JIRA)
- kontrola wersji (Git)
- telekonferencje (Adobe Connect)



## Chmura obliczeniowa

- PaaS wykorzystująca OpenStack



Akademickie Centrum Komputerowe  
CYFRONET AGH w Krakowie  
(koordynator)



Centrum Informatyczne Trójmiejskiej  
Akademickiej Sieci Komputerowej



Poznańskie Centrum  
Superkomputerowo  
Sieciowe w Poznaniu



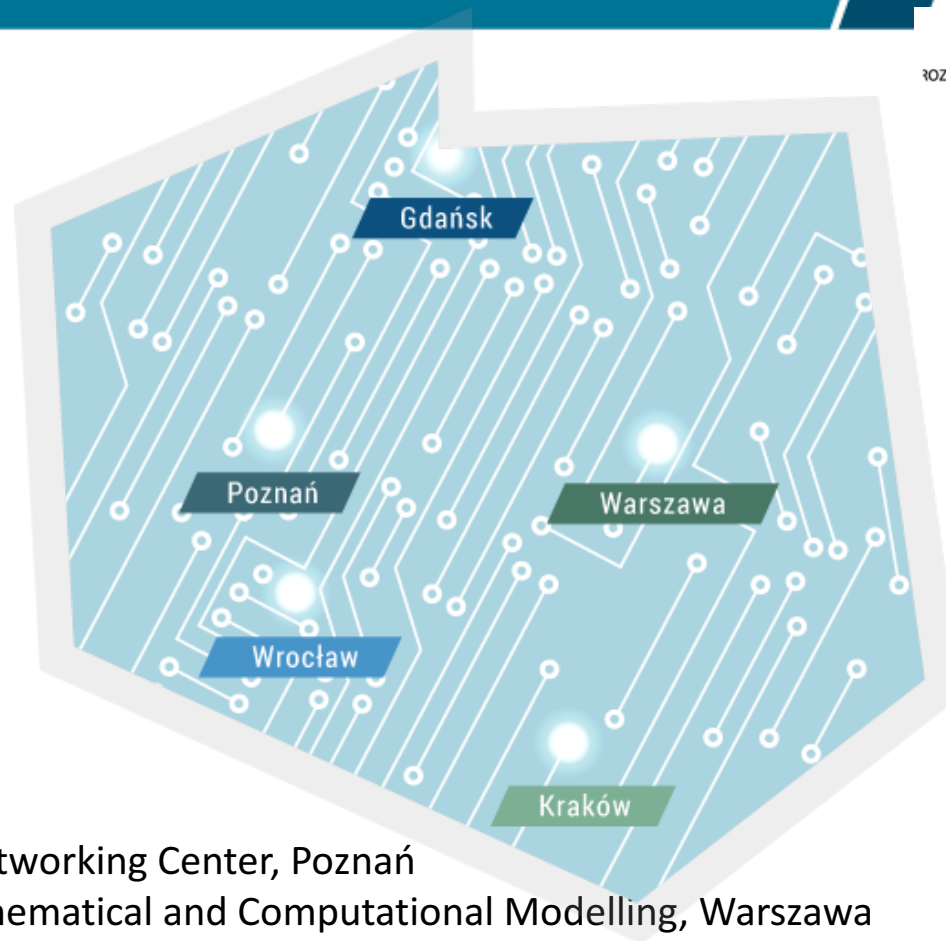
Interdyscyplinarne Centrum  
Modelowania  
Matematycznego i Komputerowego  
w Warszawie



Wrocławskie Centrum  
Sieciowo - Superkomputerowe we  
Wrocławiu



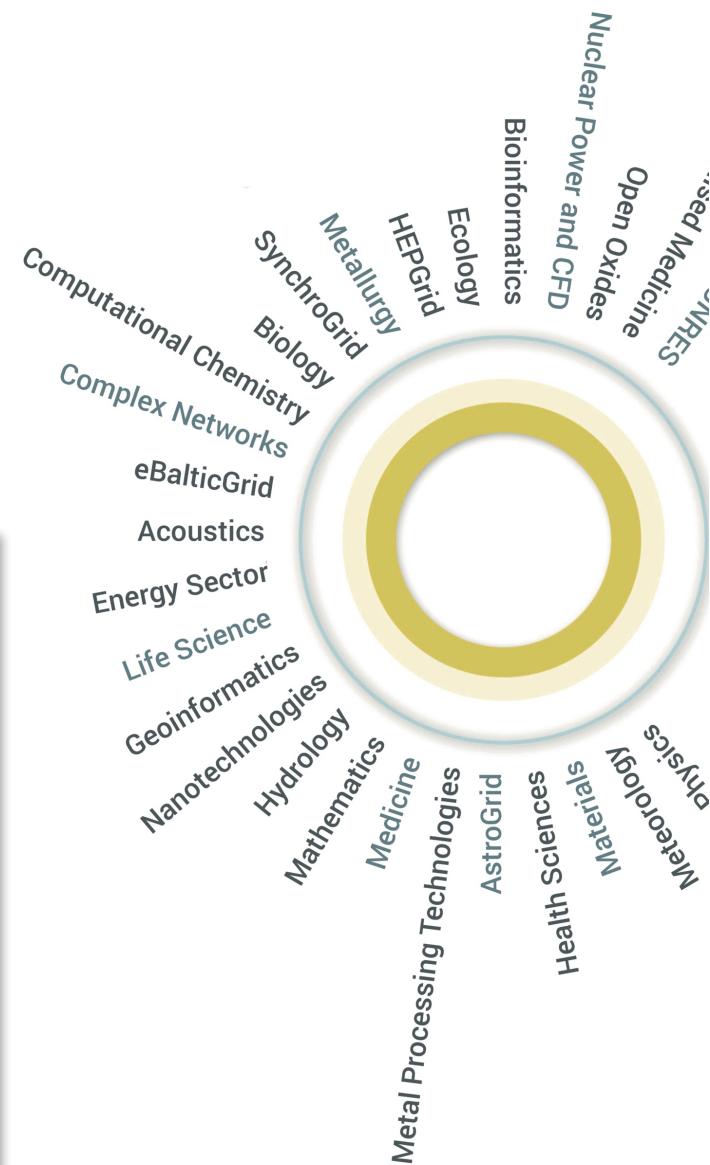
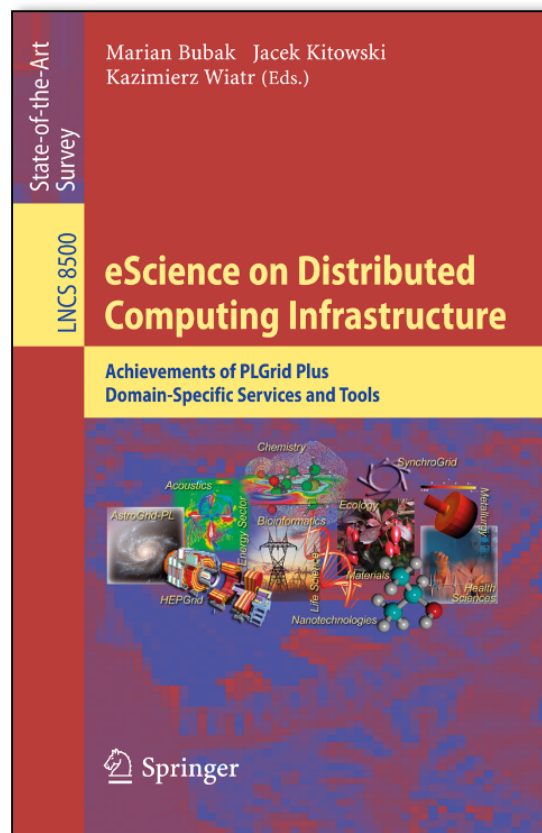
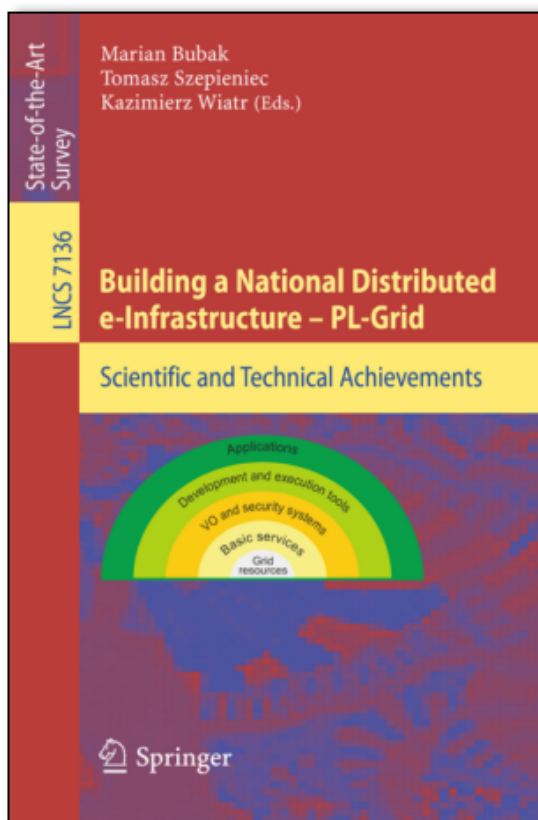
- Projects:
  - PL-Grid
  - PLGrid Plus
  - PLGrid NG
  - PLGrid Core
- PLGrid Consortium
  - Coordinator: ACC Cyfronet AGH
  - Partners:
    - Poznan Supercomputing and Networking Center, Poznań
    - Interdisciplinary Centre for Mathematical and Computational Modelling, Warszawa
    - Wrocław Centre for Networking and Supercomputing, Wrocław
    - Tricity Academic Computer Centre, Gdańsk





## Projekty PLGrid Plus oraz PLGrid NG

- współpraca pomiędzy ekspertami dziedzinowymi oraz IT
- rozwiązania dla 27 domen nauki zintegrowane z infrastrukturą PLGrid



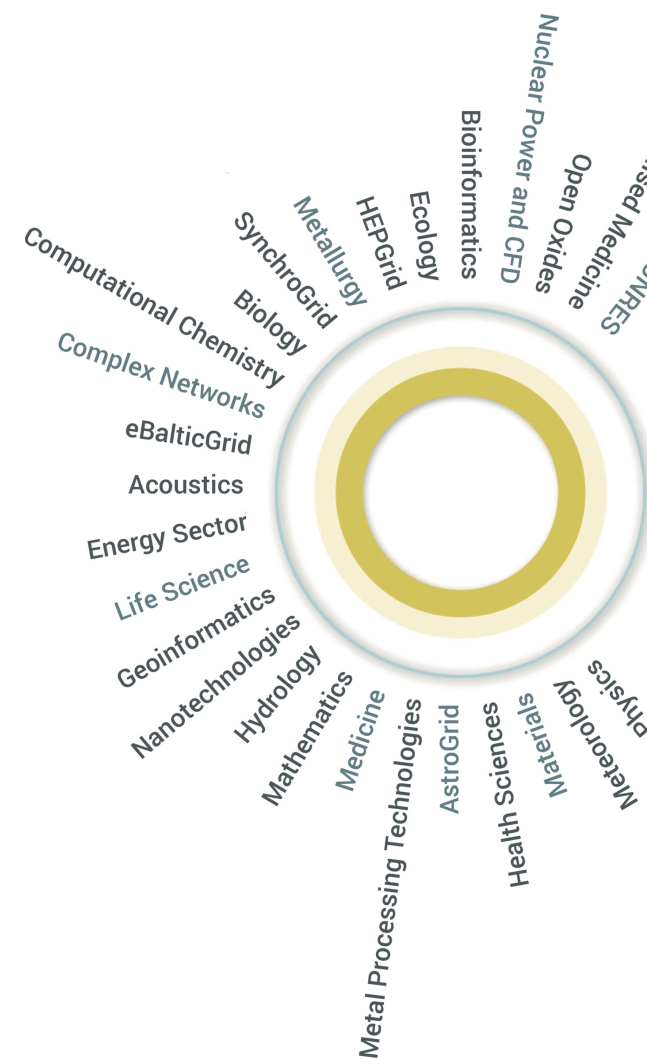
W infrastrukturze liczy blisko **7000 użytkowników**.

Efekt prac prowadzonych z pomocą Infrastruktury jest ponad **2500 publikacji**.

Znakomita większość publikacji jest znajduje się w czasopismach z **Listy Filadelfijskiej** oraz jest wysoko punktowana przez **MNiSW**

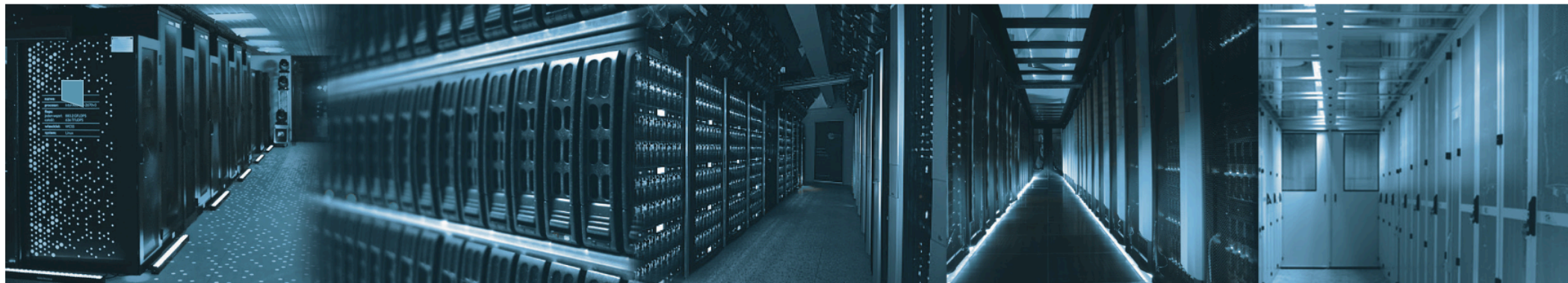
Z infrastruktury korzystają:

- Pracownicy naukowci i dydaktyczni
- Doktoranci
- Magistranci w ramach prac dyplomowych
- Studenci
- Inne osoby uczestniczące w badaniach naukowych (w tym obcokrajowcy)



## Superkomputery z Polski

- 77 – Prometheus (ACC Cyfronet AGH) – 2.4 Pflops (PLGrid)
- 172 – Eagle/Orzeł (PSNC) – 1.37 PFLOPS (PLGrid)
- 176 – Tryton (CI TASK) – 1.41 PFLOPS (PLGrid)
- 223 – Okeanos (ICM) – 1.08 PFLOPS
- 384 – Bem (WNSC) – 0.86 PFLOPS (PLGrid)





- HP Apollo 8000
- 2,35 PFLOPS (38 na Top500 XI 2015, **77** na Top500 XI 2017)
- 2232 węzłów, 2x Intel Haswell E5-2680v3 2,5 GHz, 12 rdzeni
  - w tym 72 węzły z 2 x Tesla K40 XL
- 53 568 rdzeni, 13824 na wyspie
- 279 TB DDR4 RAM (128 GB na węzeł)
- PUE ~1.06, 816 kW mocy całkowitej
- 20 szaf typu rack
- CentOS 7
- system gotowy do rozbudowy w trakcie produkcyjnego działania





- efektywne chłodzenie wodne
- wydajność energetyczna (~1,06 PUE)

## Dostępne rodzaje procesorów

Intel Xeon 4-, 6-, 12-, 14-rdzeniowe (do 28 rdzeni na węźle)

AMD Opteron 6-, 12-, 16-rdzeniowe (do 64 rdzeni na węźle)

## Akceleratory obliczeń

karty GPGPU NVidia Tesla (do 8 kart na węźle)

Intel Xeon Phi (do 2 kart na węźle)

## Różne konfiguracje węzłów obliczeniowych

od 8 do 64 rdzeni na węzeł

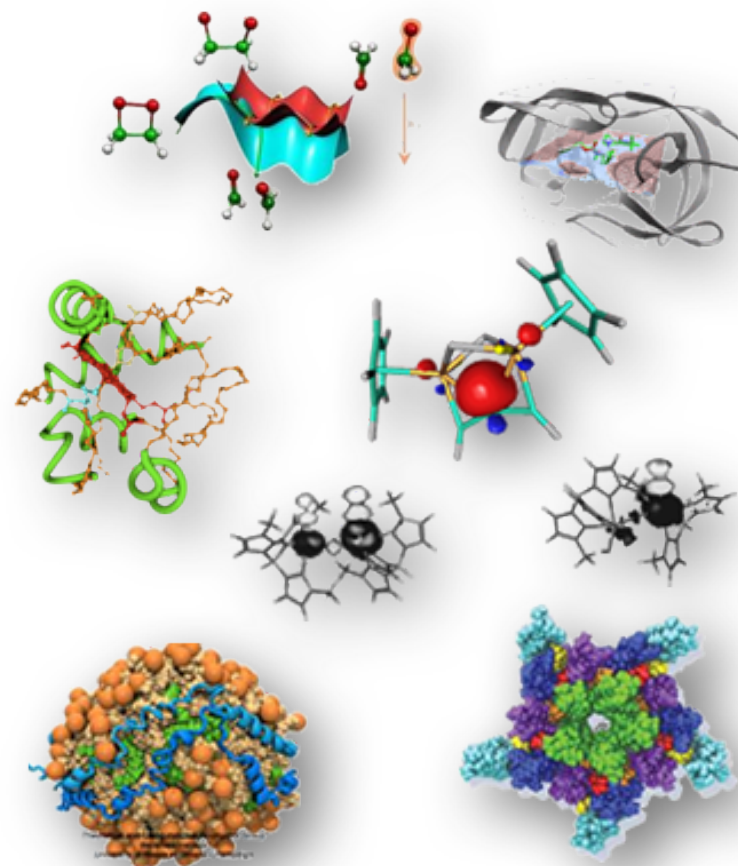
do 512 GB RAM na węzeł

vSMP (Intel Xeon) – do 6 TB RAM oraz 768 rdzeni

Pełna konfiguracja na stronie:

[www.plgrid.pl/oferta/zasoby\\_obliczeniowe/opis\\_zasobow/HPC](http://www.plgrid.pl/oferta/zasoby_obliczeniowe/opis_zasobow/HPC)

- **Chemia i Biologia:** ADF, AMBER, CFOUR, Dalton, GAMESS, Gaussian, Molcas, Molpro, MOPAC, NWChem, TURBOMOLE, Jaguar, LAMMPS, AutoDock/AutoGrid, BLAST, Clustal, Siesta, Quantum Espresso, VASP (na licencjach własnych użytkowników), CP2K, CPMD, Gromacs, NAMD
- **FEM:** Abaqus, Ansys/FLUENT, Nastran, Marc, OpenFOAM
- **Nanotechnologia:** ABINIT, Quantum Espresso, NAMD, CP2K, CPMD, Crystal, CASTEP
- **Interdyscyplinarne:** Mathematica, MATLAB, Simulink
- **Kompilatory i narzędzia:** Intel, PGI, GNU, MKL, CUDA, MPI, OpenMP, OpenCL,, Python, R, Ruby, Java, Alinea, ROOT, WRF, GA, HDF, NAG
- **Bazy danych**
- **Możliwość instalacji dowolnego oprogramowania**





- PLGrid to różne sposoby dostępu do zasobów
  - lokalne systemy kolejkowe na klastrach (CLI oraz GUI)
  - oprogramowania pośredniczące (UNICORE, QosCosGrid, gLite)
  - maszyny wirtualne
  - portale sieciowe (InSilicoLab, GridSpace2, PLG-Data, DataNet,...)
- Zarządzanie dostępem do usług poprzez Portal Użytkownika PLGrid (<https://portal.plgrid.pl>)
- Katalog aplikacji i usług (<https://apps.plgrid.pl/>)
  - usługi ogólne - <https://docs.plgrid.pl/uslugi>
  - usługi dedykowane wybranym dziedzinom nauki - [https://docs.plgrid.pl/uslugi\\_dziedzinowe](https://docs.plgrid.pl/uslugi_dziedzinowe)

**Adobe Connect** - to cenione na rynku rozwiązanie do prowadzenia telekonferencji do 50 uczestników jednocześnie.

telekonferencje

**Confluence** - to przestrzeń typu "wiki" do agregowania wiedzy.

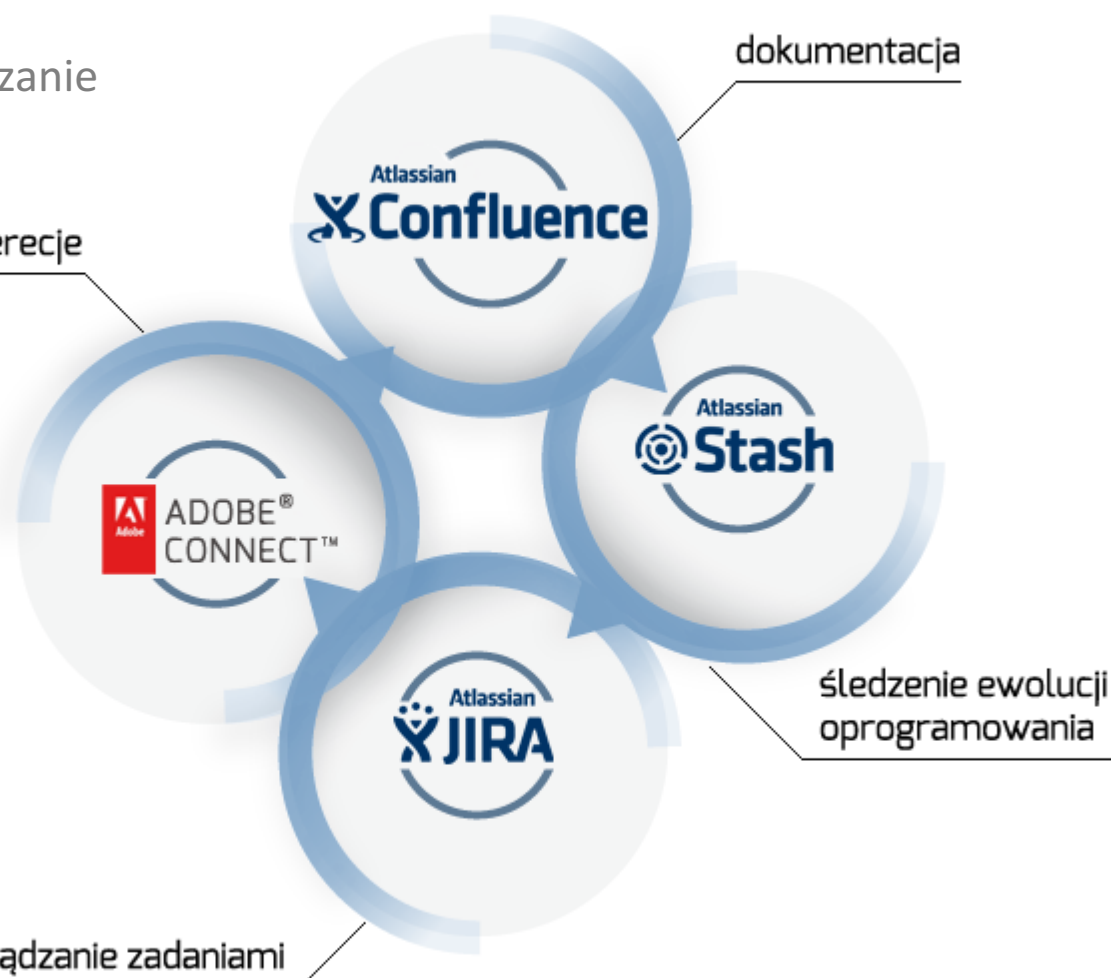
dokumentacja

**Jira** - to narzędzie do planowania pracy i śledzenia jej efektów.

**Stash** - oprogramowanie do śledzenia zmian w kodzie źródłowym

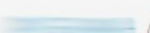
śledzenie ewolucji oprogramowania













zarządzanie zadaniami



<https://data.plgrid.pl/>

- Dostęp do danych przez przeglądarkę internetową

ZAWARTOŚĆ FOLDERU: PEOPLE /  / BADANIA / LUTY 2014 / [Dodaj pliki](#) [Nowy folder](#)

Prawa	Rozmiar	Data modyfikacji	Nazwa	Typ	
-rw-r--r-- 	11.6 MB	Jan 27 22:24	 GSM1019938_NBM6-P.CEL	Plik	Na pewno? <a href="#">Tak</a> <a href="#">Nie</a>
<input type="text" value="rw-r-----"/>  	3.5 MB	Jan 27 22:31	 GSM933008_US91903683 Dec08_1_1.txt	Plik	<a href="#">Usun</a>
-rw-r--r-- 	5.11 MB	Feb 5 16:25	<input type="text" value="GSM980568_MB2009031253.CEL"/>	 	<a href="#">Usun</a>
drwxr-xr-x 	2 KB	Feb 8 14:47	 aktualne_pliki	Folder	<a href="#">Usun</a>
-r--r--r-- 	128 KB	25 21:02	 all-new-3class-tr	Plik	<a href="#">Usun</a>

<https://submit.plgrid.pl/>



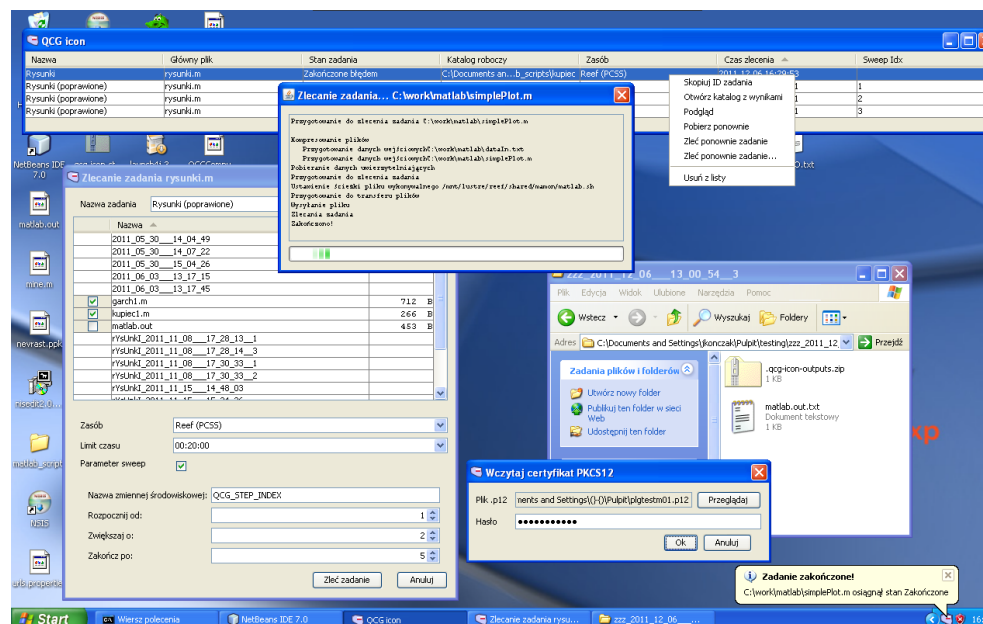
- Rimrock – Robust Remote Process and Job Controller
- Ułatwia pracę z zadaniami obliczeniowymi na zdalnych zasobach
- Umożliwia pracę z zadaniami wsadowymi oraz uruchomienie interaktywnej aplikacji
  - dostarcza prostego interfejsu RESTowego do wysyłania i odbioru danych oraz kontroli zadań



**QCG-NOW**  
MOC HPC NA TWOIM BIURKU

<http://www.qoscosgrid.org/qcg-now/>

- Lekka aplikacja graficzna do zarządzania zadaniami gridowymi
- Szerokie portfolio wspieranych aplikacji z chemii obliczeniowej
  - Gaussian
  - GAMESS
  - Molpro
  - Dalton
  - Turbomole
  - ADF
  - CRYSTAL09
- integracja z GaussView










<http://insilicolab.grid.cyfronet.pl/>

- Środowisko pracy z systemem zintegrowanych narzędzi, które:
  - wspomagają zarządzanie złożonymi obliczeniami
  - automatyzują powtarzalne cykle obliczeń
  - umożliwiają w wygodny sposób zarządzanie procesem obliczeń
  - ułatwiają zarządzanie rozproszonymi danymi eksperymentu
  - umożliwiają wspólną analizę rezultatów wielu równoległych obliczeń
  - ułatwiają współpracę pomiędzy badaczami pracującymi nad wspólnymi projektami
  - nie rozpraszają użytkowników wykorzystywaną technologią - bez forsowania zmiany sposobu myślenia naukowców






Log in... You are logged in as **anonymousUser** 

**Your Experiments**

Menu    

- ✓ Benzene...ground  
state...RHF/STO-2G

Welcome  



**in silico LAB** 

**Welcome to InSilicoLab Portal**

To take advantage of the full portal functionality (including job submission and LFC directory management), it is required that you have a valid proxy configured. You can configure it now - using the left panel, or do it later at any time - by clicking the "Configure proxy..." button in the workspace.

If you choose not to configure your proxy now, you can still access the portal and work with limited functionality. For this purpose, choose one of the options right panel below.

**Actions**

-  **Configure your proxy**
-  **Create a new experiment:**

<http://insilicolab.grid.cyfronet.pl/>



## ■ Chemia kwantowa oraz biochemia

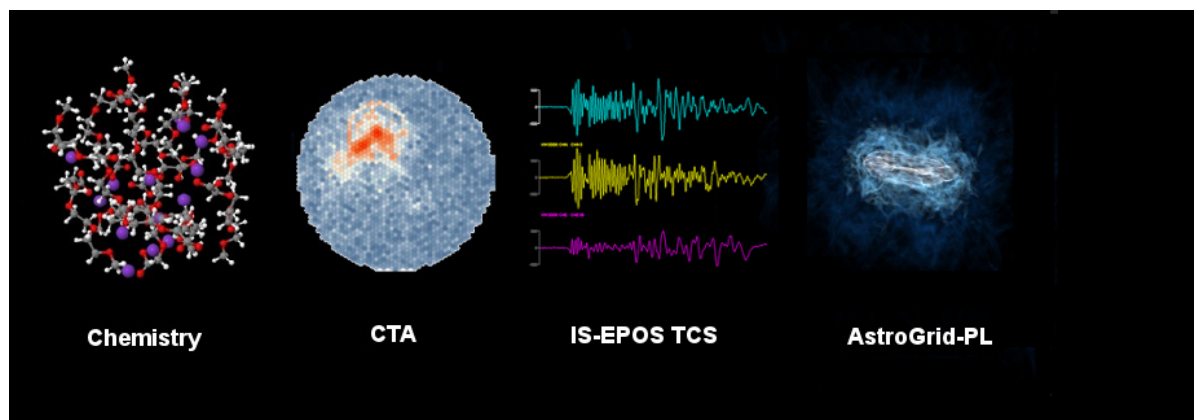
- obliczenia pakietami Gaussian, GAMESS, Turbomole
- Trajectory Sculptor - półautomatyczne przycinanie dużych układów molekularnych (np. wyników symulacji MD)
- możliwość łączenia obu typów eksperymentów

## ■ Astrofizyka

- obliczenia hydrodynamiczne metodami objętości skończonej
- obliczenia dla konsorcjum Cherenkov Telescope Array (CTA)

## ■ Geofizyka

- badania sejsmiczności indukowanej (IS-EPOS)



<https://statystyka.plgrid.pl/>

- Zintegrowana platforma usług dedykowanych statystycznej analizie danych w wielodziedzinowych zagadnieniach naukowych
  - ułatwia wybór testu statystycznego dla analizy badanego zagadnienia
  - udostępnia otwarte repozytorium kilkudziesięciu popularnych testów statystycznych
  - oraz zestaw aplikacji do analizy szeregów czasowych.

Statystyczna Analiza Danych ▾ Kategorie ▾ Dodaj test Dashboard Tomasz Piątek ▾

[Lista testów](#) / [Jednoczynnikowa analiza wariancji \(ANOVA\)](#) / [domyślna](#) / [Uruchomienie testu](#)

## Jednoczynnikowa analiza wariancji (ANOVA)

Wyniki **Dane** Raport

wpisz lub wklej z Excela dane, oddzielając liczby spacjami.

wpisz lub wklej z Excela podział na grupy

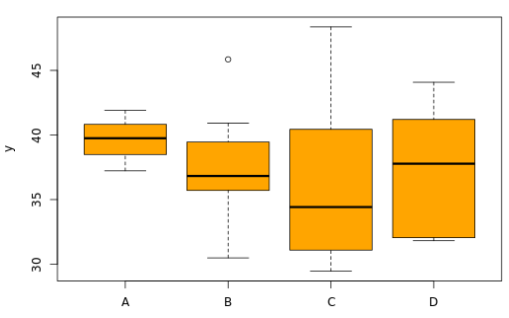
załaduj plik Excela

No file selected.

Opis osi y

Zmienna 1: kolor:



Porównanie grup przy pomocy jednoczynnikowej analizy wariancji:  $p=0.831$  wartość statystyki  $F=0.291$  na 3 i 36 stopniach swobody

### Wyjaśnienia

Program do porównywania grup przy pomocy analizy wariancji (ANOVA). Należy podać kolumnę danych z wartościami numerycznymi, oraz, w drugim polu, kolumnę z podziałem na grupy (do której grupy dana wartość należy).

Dane można wprowadzać na trzy sposoby:

- wpisując dane bezpośrednio do okienek po lewej stronie aplikacji
- w Excelu lub OpenOffice podświetlamy kolumnę (kolumny nie powinny mieć nazw), Ctrl+c i po kliknięciu okienka po lewej stronie, Ctrl+v
- wybierając przy pomocy przycisku po lewej stronie plik Excela, który w pierwszym arkuszu, pierwszej kolumnie ma wartości liczbowe a w drugiej podział na grupy. Kolumny nie powinny mieć nazw.



# Dostęp do Infrastruktury



- Rejestracja elektroniczna przez Portal Użytkownika PLGrid (<https://portal.plgrid.pl>)
- Weryfikacja użytkownika na podstawie nr OPI i danych w bazie Ludzie Nauki (<http://www.nauka-polska.pl/>)



Po rejestracji użytkownik uzyskuje możliwość:

- Zarządzanie dostępem do usług i wielu aplikacji
- Monitoring zużycia zasobów
- Zarządzanie grantami obliczeniowymi
- Zarządzanie certyfikatami gridowymi
- Dostęp do wszystkich zasobów przy użyciu **jednego konta i jednego hasła** (lub certyfikatu gridowego)

## REJESTRACJA

1

Wypełnij gotowy formularz rejestracji i wyślij go do weryfikacji. Następnie kliknij w link aktywacyjny, który otrzymałeś w mailu. Konto zostanie aktywowane!

## AFILIACJE

2

Uzupełnij informacje na temat instytucji naukowej, do której przynależysz. Afiliacja polskiej jednostki naukowej umożliwi dostęp do pełnej oferty Infrastruktury PLGrid.

## ZAŁÓŻ ZESPÓŁ NAUKOWY

3

Jest to jednostka, za pomocą której możliwe jest negocjowanie zasobów obliczeniowych. Odzwierciedla cały zespół badawczy lub użytkownika pracującego samodzielnie.

## UTWÓRZ GRANT

4

Grant stanowi umowę między centrum komputerowym a użytkownikiem. Określany jest w niej rodzaj usługi, jej parametry oraz gwarancje otrzymania.



- Dla użytkowników PL-Gridu dostępne są dwa CA
  - Simple CA (<http://plgrid-sca.wcss.wroc.pl>)
  - Polish Grid CA (<https://plgrid-ca.pl>)
- Simple CA:
  - certyfikat uzyskuje się poprzez portal <https://portal.plgrid.pl/>
  - generowany na żądanie dla każdego użytkownika (automatycznie)
  - dostęp ograniczony do polskich zasobów
- PL-Grid CA
  - certyfikat uzyskuje się poprzez portal <https://plgrid-ca.pl>
  - użytkownik musi potwierdzić tożsamość w Urzędzie Rejestracji (RA)
  - umożliwia pracę na całości gridu w European Grid Initiative (EGI)

- Certyfikaty gridowe mogą być przechowywane w różnych formatach. Najbardziej rozpowszechnione to:
  - PKCS #12
    - certyfikat wraz z kluczem prywatnym znajduje się w jednym pliku binarnym zwykle o rozszerzeniu `.p12`
  - PEM
    - certyfikat stanowi para plików tekstowych:
      - klucz prywatny (zwykle `userkey.pem`)
      - plik certyfikatu (zwykle `usercert.pem`)
- Certyfikaty w formacie PEM używane są
  - większość oprogramowania pośredniczącego (często domyślny format)
- Certyfikaty w formacie PKCS #12 używane są przez
  - przeglądarki internetowe
  - większość oprogramowania pośredniczącego

## Recepta:

- Wygeneruj swój certyfikat gridowy
  - np. certyfikat z Simple CA na Portalu Użytkownika PLGrid
    - zakładka "Certyfikaty"
- Aplikuj o usługę "Globalny dostęp QosCosGrid"
  - zakładka "Usługi" i przycisk „Zarządzaj usługami” w Portalu PLGrid lub strona <https://apps.plgrid.pl/>
  - wyszukaj usługę "Globalny dostęp QosCosGrid"



## Recepta:

- Wygeneruj swój certyfikat gridowy
  - np. certyfikat z Simple CA na Portalu Użytkownika PLGrid
    - zakładka "Certyfikaty"
- Aplikuj o usługi "Globalny dostęp gLite" oraz "InSilicoLab for Chemistry"
  - zakładka "Usługi" i przycisk „Zarządzaj usługami” w Portalu PLGrid lub strona <https://apps.plgrid.pl/>
  - wyszukaj usługę "Globalny dostęp gLite" i "InSilicoLab for Chemistry"





## Szkolenia

- zdalne - <https://ekursy.plgrid.pl/>
- z trenerem - informacje na <http://www.cyfronet.krakow.pl/>

## Podręcznik Użytkownika

- Dostępny online:  
[https://docs.plgrid.pl/podrecznik\\_uzytkownika](https://docs.plgrid.pl/podrecznik_uzytkownika)

## Konsultacje z ekspertami

- Za pośrednictwem Helpdesk PLGrid:  
<https://helpdesk.plgrid.pl>  
[helpdesk@plgrid.pl](mailto:helpdesk@plgrid.pl)

## Forum użytkowników

- <https://zapytaj.plgrid.pl/>

## **Efektywne wykorzystanie Komputerów Dużej Mocy w obliczeniach naukowych**

- zasoby obliczeniowe ACK Cyfronet AGH/PLGrid
- przeprowadzanie obliczeń z wykorzystaniem systemów kolejkowych
- zarządzanie danymi na klastrach obliczeniowych

## **Warsztaty Software Carpentry**

- automatyzacja zadań przy wykorzystaniu CLI/Bash
- kontrola wersji oraz współdzielenie przy użyciu Git/GitHub
- podstawy programowania w języku Python/R
- zarządzanie danymi wykorzystując SQL
- <http://poland.software-carpentry.org/>

## **Wykorzystanie Pythona w obliczeniach naukowych**

- możliwości języka
- zastosowanie w obliczeniach chemicznych

## Programowanie kart graficznych prostsze niż myślisz – CUDA/OpenCL

- architektura GPGPU,
- podstawy CUDA C/OpenCL,
- czynniki warunkujące wydajność kodu
- zagadnienia zaawansowane: konteksty, użycie wielu kart, zastosowanie tekstur,
- biblioteka realizująca algebrę liniową na GPGPU: CUBLAS

## Programowanie równoległe (wątki, OpenMP)

- Wykorzystanie wątków od zrównoleglania obliczeń
- Możliwości biblioteki OpenMP
- czynniki wpływające na wydajność obliczeń równoległych

## Programowanie rozproszone (MPI)

- Wykorzystanie MPI do rozpraszania obliczeń
- Komunikacja między procesami
- Czynniki wpływające na wydajność kodów rozproszonych

## Narzędzia ułatwiające przeprowadzanie projektów badawczych.

- Stash: sposoby użycia systemu kontroli wersji Git oraz przechowywania i administracji repozytoriami oraz projektami,
- JIRA: zastosowanie przy planowaniu i przeprowadzaniu projektów badawczych oraz programistycznych,
- Confluence: zastosowanie do agregowania wiedzy oraz dokumentacji,
- Platforma telekonferencyjna Adobe Connect

<http://tinyurl.com/cyfronet-2017-chemia>

## ■ Dzień Otwarty ACK Cyfronet AGH: 27 XI 2017, Kraków AGH A0

- Prezentacja aktywności ACK Cyfronet AGH
- Projekty badawcze prowadzone przez użytkowników na KDM Cyfronetu
- Rozstrzygnięcie konkursu na najlepszą pracę doktorską zrealizowaną w oparciu o zasoby ACK Cyfronet AGH



## ■ Konferencja Użytkowników Komputerów Dużej Mocy KU KDM'18, 7-9 III 2018, Zakopane

- Wykorzystanie infrastruktury obliczeniowych w nauce (obliczenia wieloskalowe oraz symulacje obliczeń kwantowych)
- Prezentacje wyników naukowych użytkowników ACK Cyfronet AGH
- Szkolenie z zastosowania Pythona w analizie danych lub wprowadzenie do modelowania sztucznej inteligencji (AI)





Rejestracja: <https://portal.plgrid.pl>

[helpdesk@plgrid.pl](mailto:helpdesk@plgrid.pl)

<https://helpdesk.plgrid.pl>

+48 12 632 51 63