Domain-oriented services and resources of Polish Infrastructure for Supporting Computational Science in the European Research Space – PLGrid Plus

Data Storage Solution Using PL-Grid UNICORE Infrastructure

P.Bała, K.Benedyczak, M.Borcz, R.Kluszczyński, G.Marczak, M.Stolarek

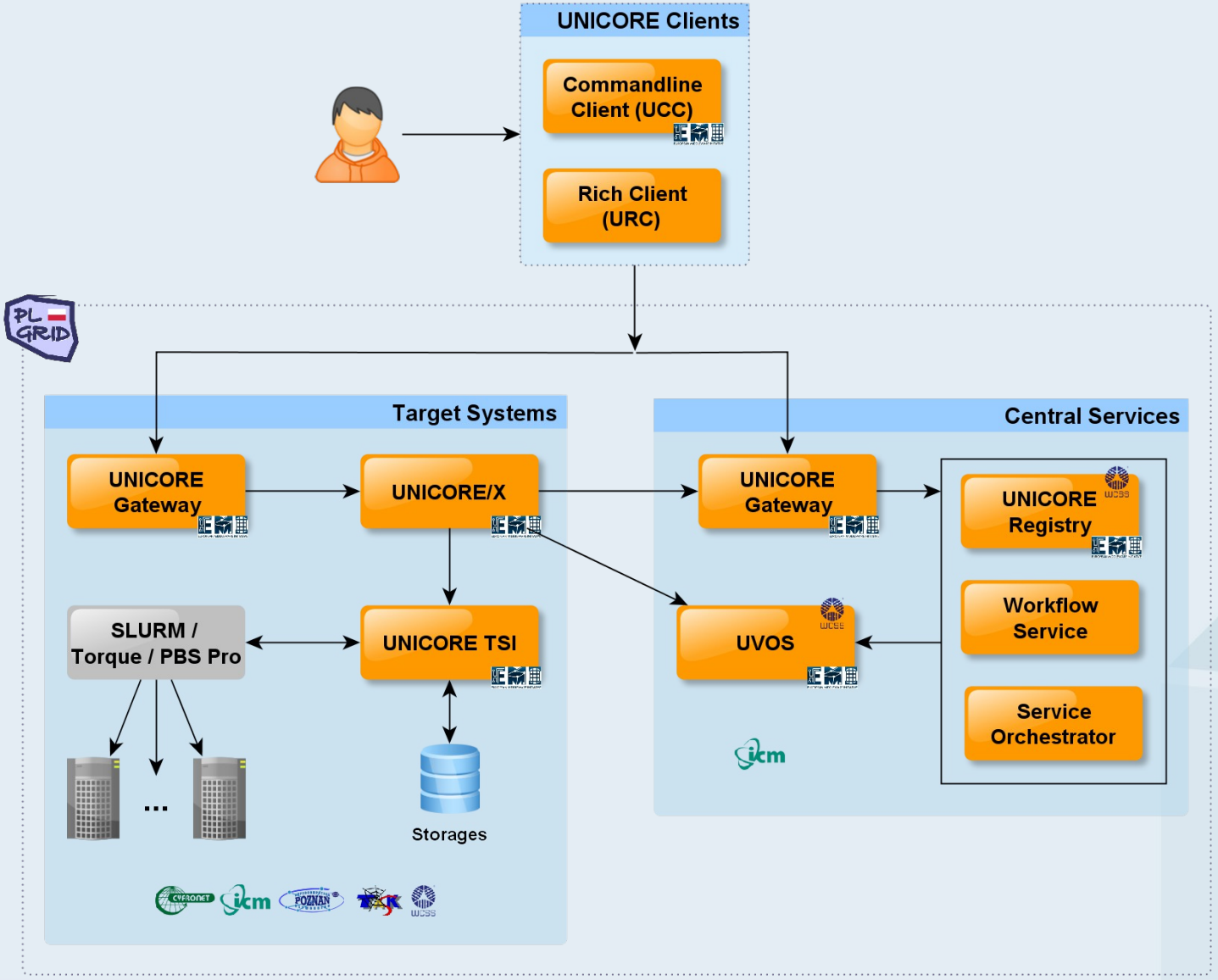
ICM, University of Warsaw

UNICORE Summit 2013,
Leipzig, Germany, June 18th

PL-Grid

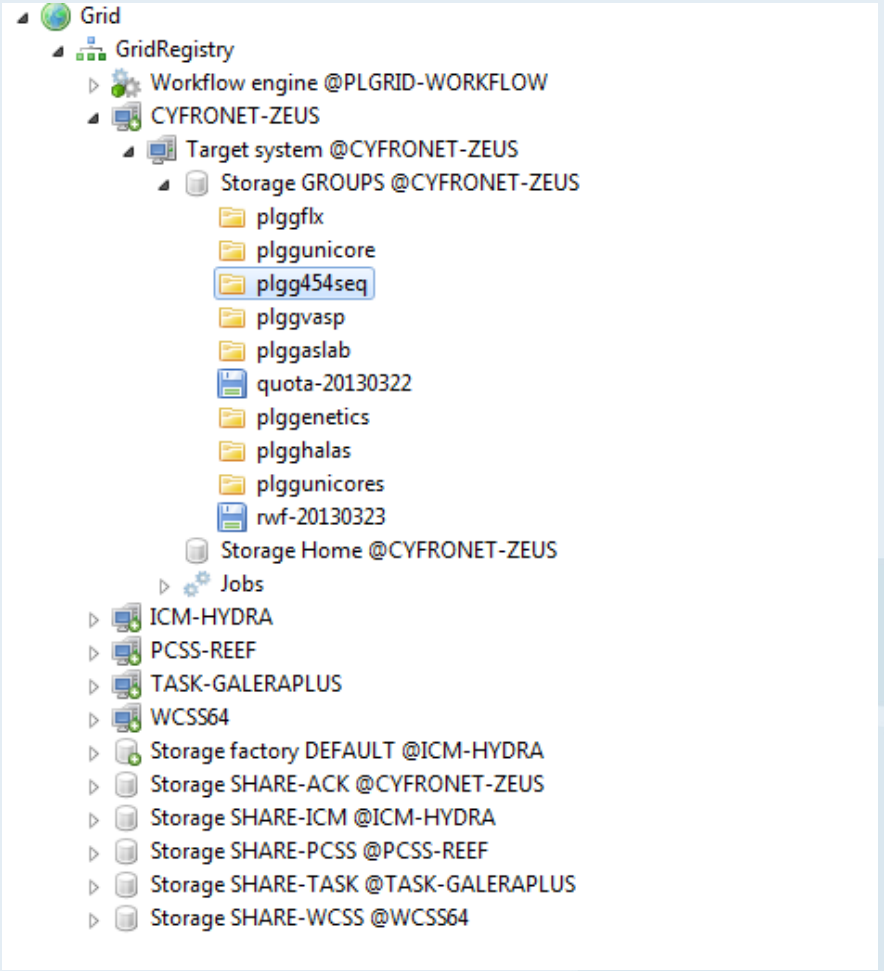
- National Grid Initiative
- Partners:
  - Polish supercomputer centers
  - CYFRONET, ICM, PCSS, WCSS, TASK
- UNICORE
  - UNICORE Operating Center located in ICM
  - Installed in all partners centers
    - UNICORE Gateway, UNICORE/X, TSI
  - Central services
    - UNICORE Registry, UVOS, Workflow System (ICM)
    - Backup copy of Registry and UVOS (WCSS)

UNICORE Infrastructure

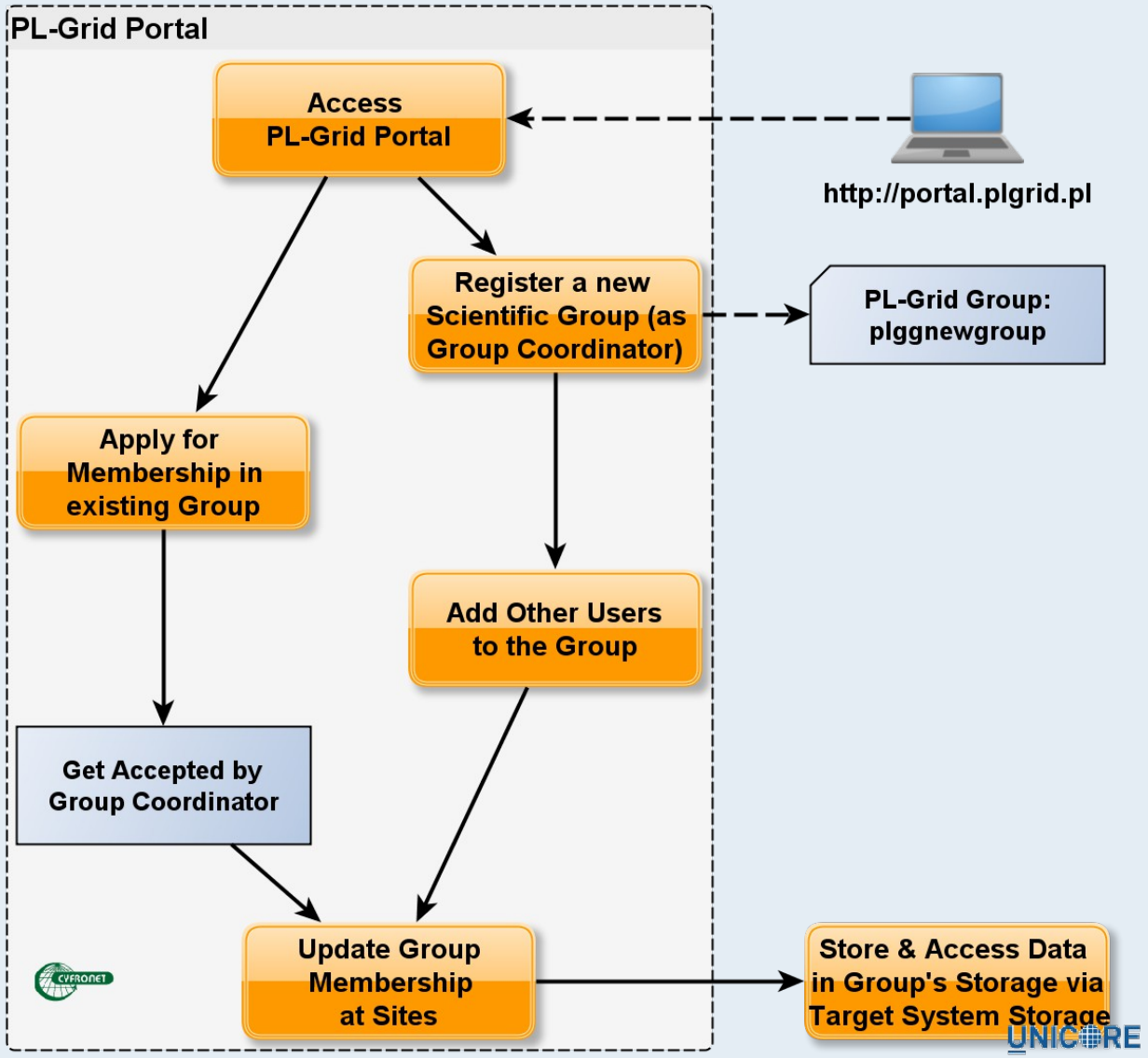


PL-Grid Storages

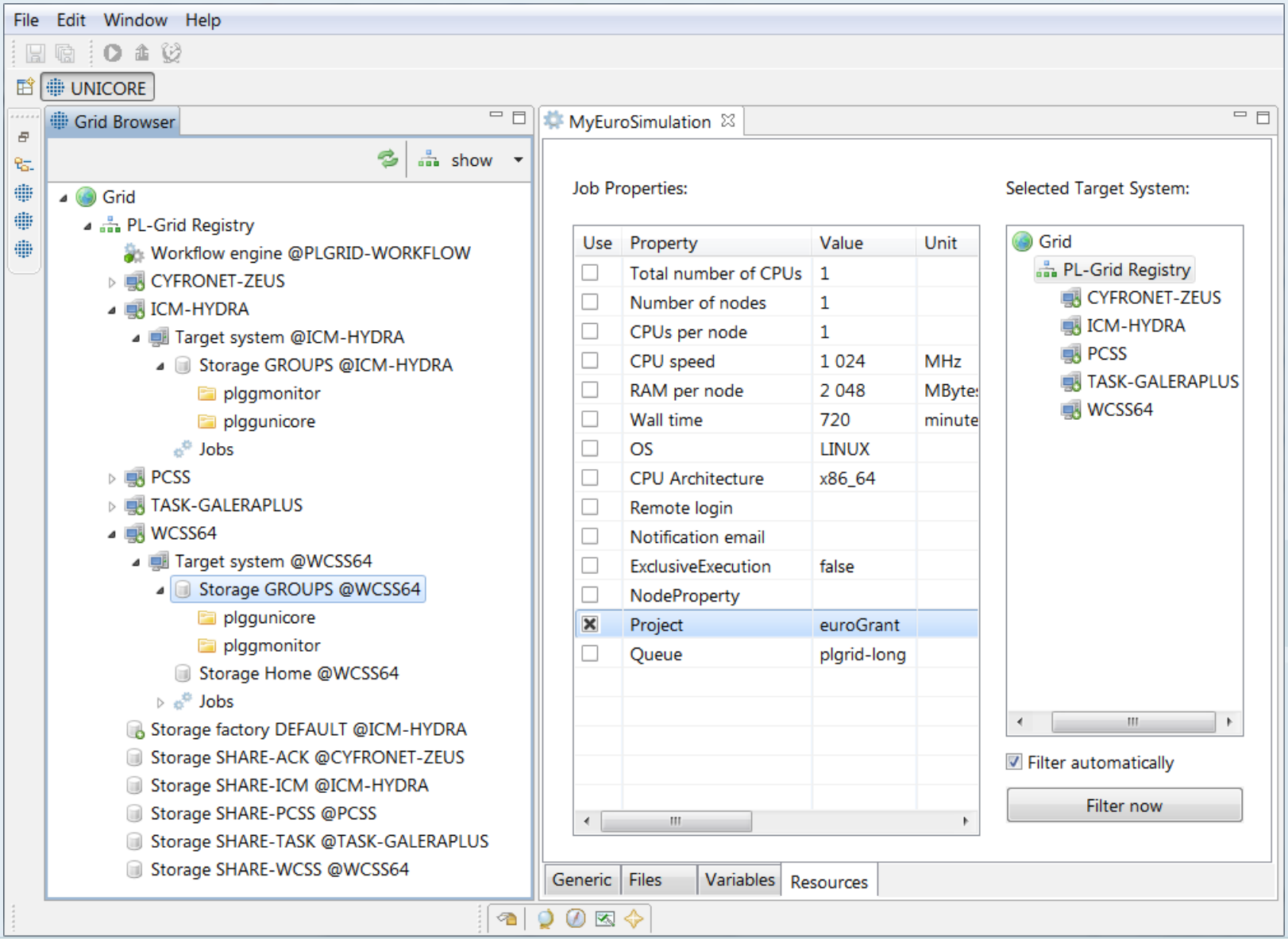
- UNICORE at PLG have 3 types of storages:
  - Home storage
  - Groups storage
  - Global storage
- PL-Grid offers support for scientific groups
  - Separate documents folder, forum and wiki portlets in PL-Grid portal
  - Storage accessible by all group members (via Target System Storage)
- Users can form a scientific groups:
  - Every group has a coordinator
  - Grants are assigned per scientific groups (max. walltime, storage size)



PL-Grid Scientific Groups



Groups and Grants in URC



Home and Global Storages

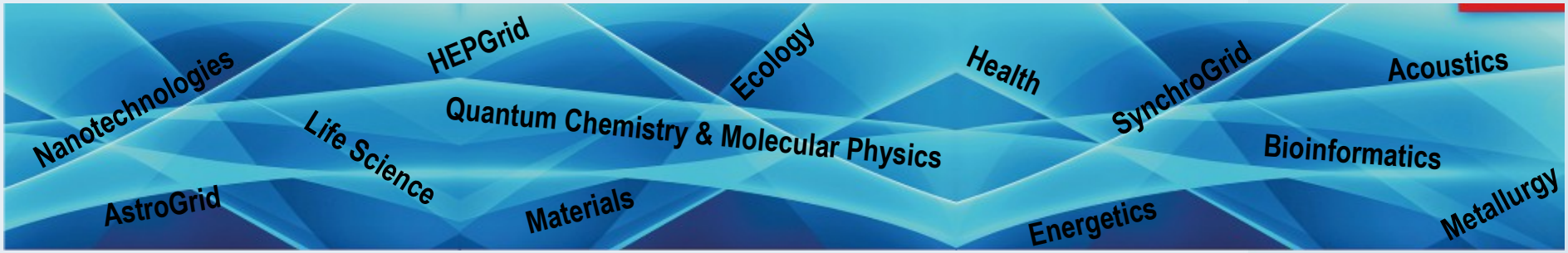
- Group directories are said to be reliable for users
- UNICORE/X provides access to Home storage
  - Accessible as Target System Storage
  - Available also during computations
  - This is usually the same physical device which is available as a UI's home directory for gLite users
  - This is also officially reliable storage (backup)
- At every site there is also configured Global Storage
  - Additional storage for UNICORE users
  - Easy UNICORE access via Registry
  - The goal of this storage is similar to group directories
  - Located mostly on fast filesystem
- ICM provides also Storage Factory (i.e. workflow processing)

PL-Grid Plus

- The main aim is to increase potential of the Polish Science by providing the necessary IT services for research teams in Poland, in line with European solutions
- Five PL-Grid Consortium Partners:

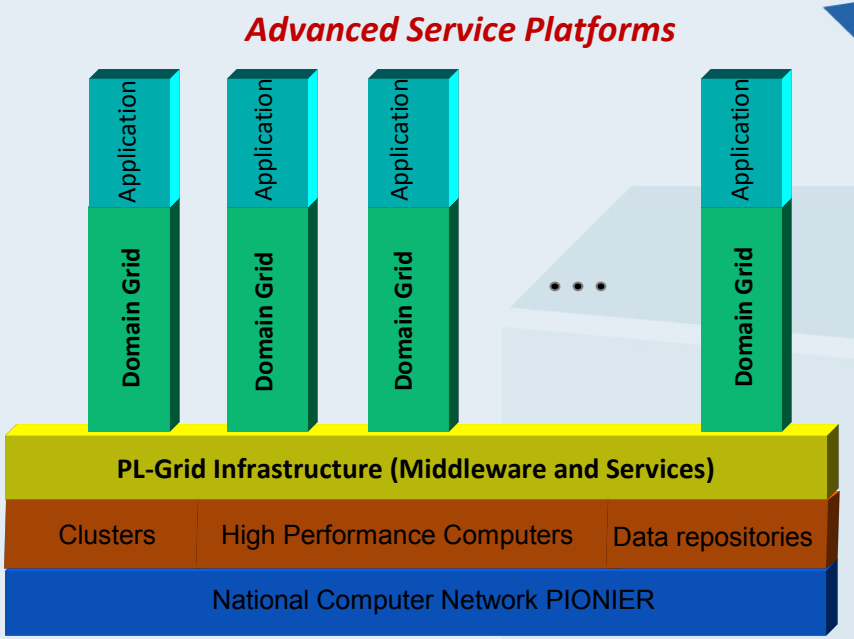
CYFRONET, ICM, POZNAN, CI TASK, WCSS

- Project Coordinator: Academic Computer Centre CYFRONET AGH
- The Project most important task is preparation of specific computing environments – so called *Domain Grids* – i.e. solutions, services and extended infrastructure, tailored to the needs of different groups of scientists
- 3 domain grids have almost ready platforms, others are still testing solutions and approaches (most of them will use one of available middlewares)



PL-Grid Plus Tasks

- Design and deploy *Domain Grids* - solutions for scientific-domain related services, tools and software packages for 13 identified scientific domains
- Extensions of resources available in the PL-Grid infrastructure by ca. 500 TFlops of computing power and ca. 4.4 PBytes of storage capacity
- Design and start-up of support for new domains grids
- Deployment of Quality of Service system for users by introducing SLA agreement
- Deployment of new infrastructure services
- Deployment of Cloud infrastructure for users
- Broad consultancy, traning and dissemination offer



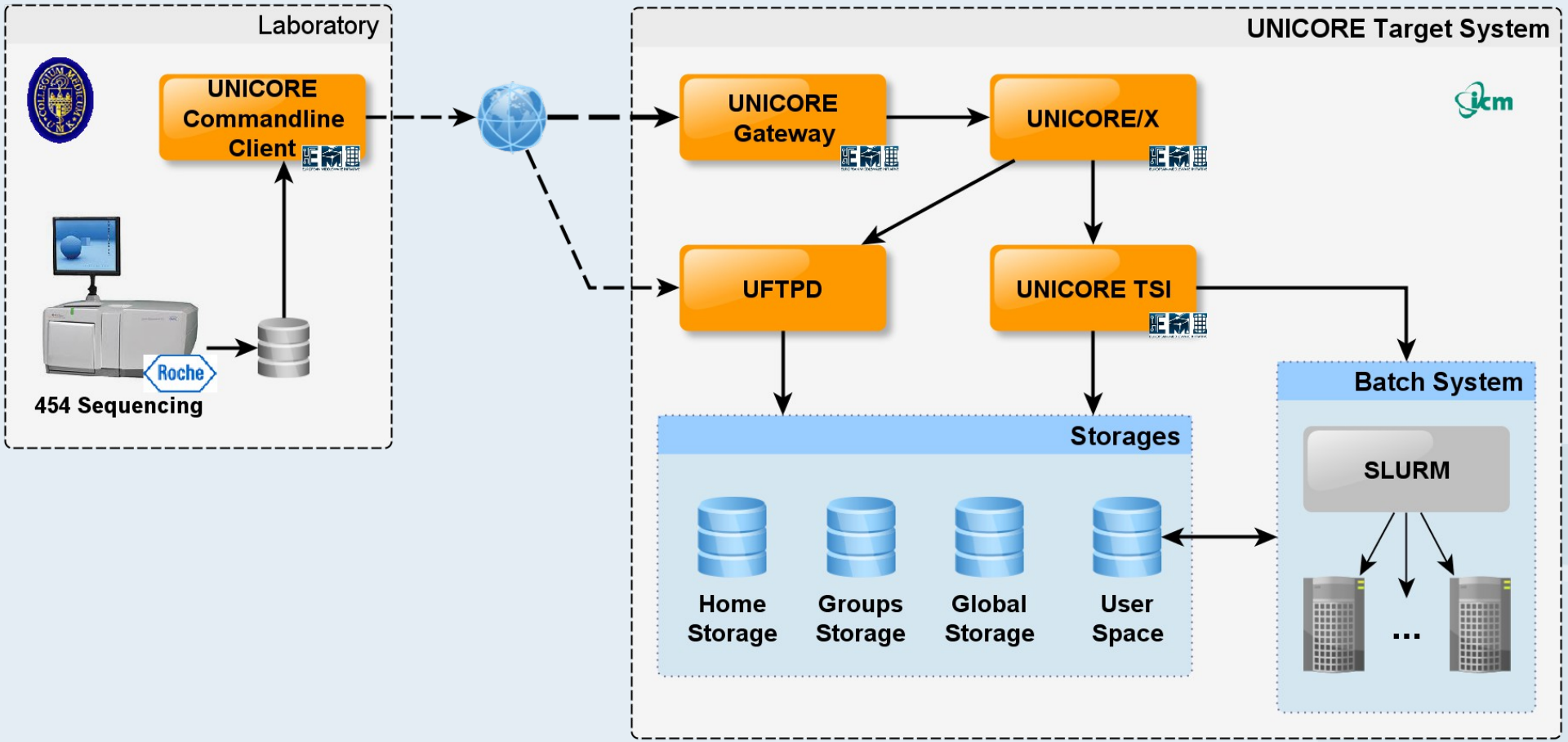
Bioinformatics Domain Grid

- First almost ready domain-specific solution prepared at ICM
- Main tasks so far:
  - Bring FLX application included in Roche Instrument software on the Grid
  - Design and prepare data analysis workflow
  - Prepare safe storage solution for experiments data
  - Design handling and storing data from Roche instruments
- Observations:
  - There is an interest in solutions for storing data
  - Analysis is usually run by user which prepared workflow – there is a need for easier way of handling worklows (sharing, running)
  - There are questions about handling metadata – this still needs some tests

Experiments

- Partners:
  - Department of Molecular and Forensic Genetics, Institute of Forensic Medicine, Ludwik Rydygier Collegium Medicum, Nicolaus Copernicus University
  - The Postgraduate School of Molecular Medicine, Medical University of Warsaw
  - ICM, University of Warsaw
  - Medical University of Warsaw
- GS FLX Instrument (Roche Diagnostics) and its software
  - Analysis of patients' mtDNA having colorectal cancer (mutation identification)
- Illumina HiSeq 1500 Instrument
  - Every run is expensive and takes up to several days
  - Safe and secure data storage needed

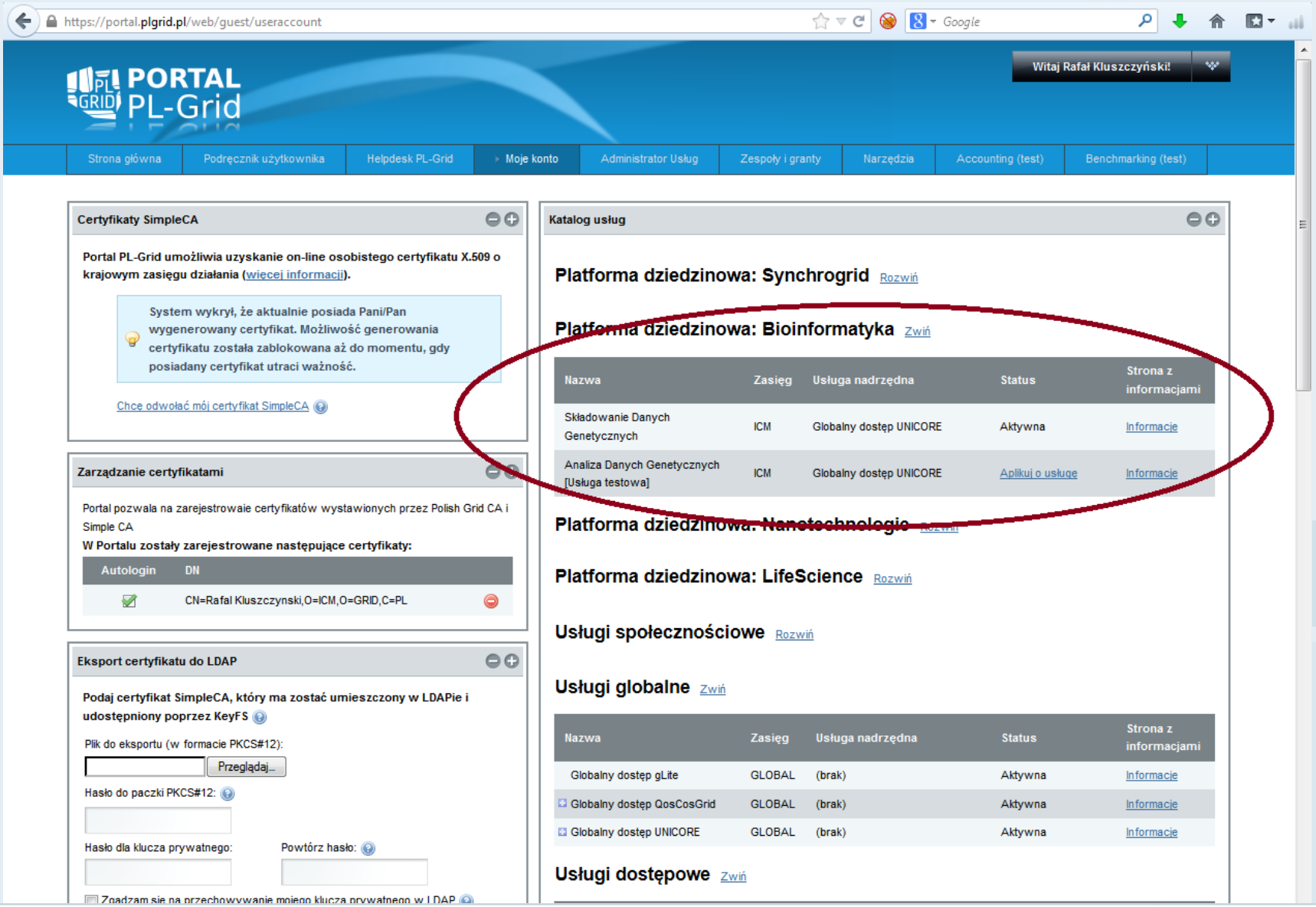
Genetic Data Analysis – Experiment



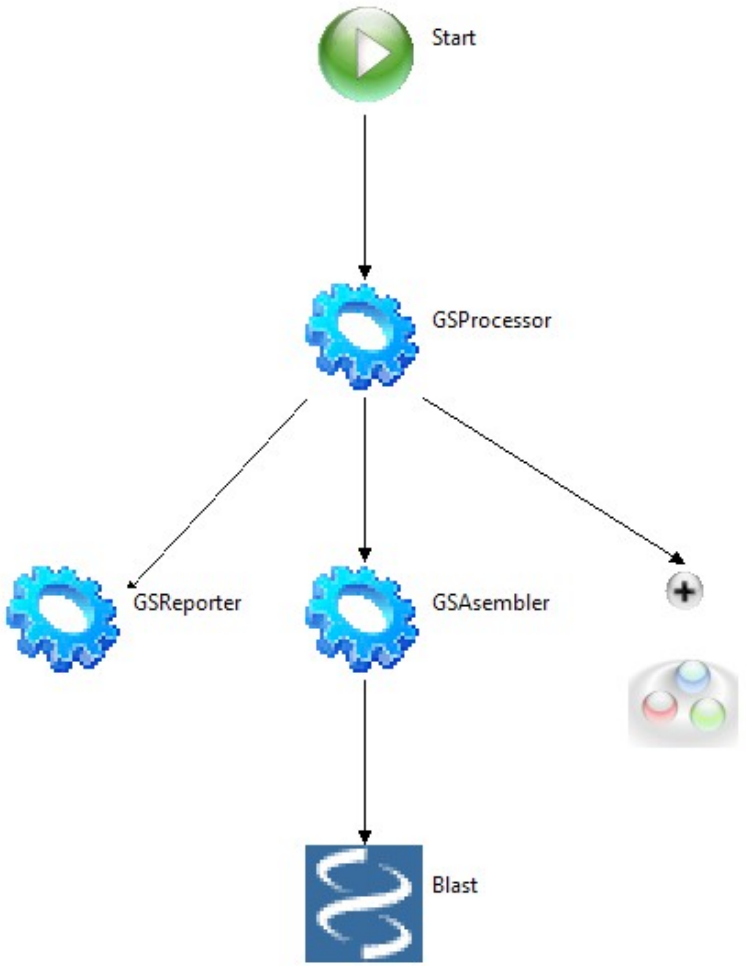
Genetic Data Analysis – Getting Access

- Every domain service has to be activated by users via PL-Grid Portal
  - Domain services have their own documentation
  - Portal may trigger additional actions like assigning to particular group or activate other dependency services
- Genetic Data Analysis
  - Needs to have active: UNICORE Global Access
  - Assigns to group ***plggflx***
    - Restrict access to software
    - Group directory contains workflow template
- In order to use, user needs to:
  - Register in PL-Grid portal
  - Activate access of domain service: Genetic Data Analysis
  - Download URC, get workflow template and run analysis

Getting Domain Service Access



Genetic Data Analysis – Workflow



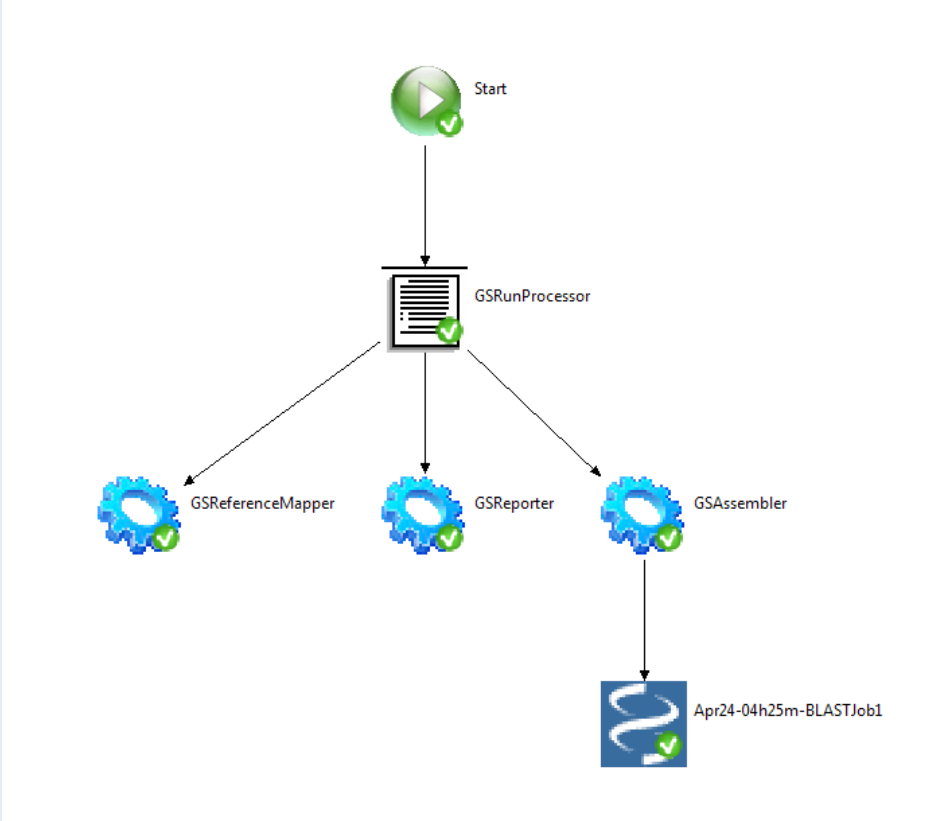
Genetic Data Analysis – Speed up

- The most time consuming part of analysis pipeline is GS Run Processor
- Table below show significant speedup for 64 cores comparing to single core run
- Different execution time for different processors
  - Internal Cache Size

Processor type	Cache size	Interconnect	Number of cores	Time (hours)
Intel Xeon CPU @ 3.60GHz	1024 kB	none	1	70.0
AMD Interlagos Opteron Processor 6272 @ 2.10GHz	2048 kB	Gigabit	64 (1 x 64 cores)	6.5
AMD Opteron Processor 6174 @ 2.20GHz	512 kB	Gigabit	64 (8 x 8 cores) 96 (8 x 12 cores)	4.5 4.5
Intel Xeon CPU, X5660 @ 2.80GHz	12288 kB	Infiniband	64 (8 x 8 cores) 96 (8 x 12 cores)	2.5 2.5

UFTP Data Transfer

- Single sequencing experiment generates ca. 30 GB of data
- Organized in 834 files of the size about 33 MB each
- In order to gain reasonable upload time – UFTP protocol is a must
- Example of input data size for workflow atomic jobs:
  - GSRunProcessor: 27 GB (PIF files)
  - GSReferenceMapper: 3.2 GB (SFF files)
  - GSReporter: 2.5 GB (CWF files)
  - GSAssembler: 3.2 GB (SFF files)
  - BLAST: result may have up to 6 GB (FASTA file)



UFTP Transfer Statistics

- As for now UFTP was mostly used to automate files upload
  - Almost ready to use it also during workflow's stage-(in/out)s
- Experiments upload statistics (last months)
  - Only receiving transfers over 30 MB in size
  - About 1,600 of uploads (mostly 33MB PIF files), most of them used 4 streams
  - Average upload transfer: 10.7 MB/s
    - Files of size about 33MB – 10.6 MB/s
    - Files with size over 10GB – 11.1 MB/s
  - Avarage time of one file: 3.04 s
- **iperf** shows 20MB bandwith between endpoints
  - Still needs to check where it is reduced

Data Storage Requirements

- Roche instruments generate genetic data for every experiment
  - Every experiment costs so those are expensive data sets
- Usually storage capacity included with machine does not exceed 3 or 4 experiments
- Requirements for genetic data:
  - It should allow for automatic and fast upload
  - Access to data during computations should be possible
  - Transfer and access should be secure
  - Only partners should have access to data (confidentiality)
  - Safe and realiable backup policy

External Storage for Genetic Data

- *Genetic Data Storage* domain service has been initially prepared for storing data generated by Roche instrument
  - Cost of every experiment and large data size results in the need of storing them using external stable storages in secure and safe way
- As an eternal backup service PLATON U4 has been used
  - Ensures safety and replication of data
  - Security provided by X509 certificates
  - Data access using FUSE (sshfs)
- Experiment data sets
  - Stored as a single archive per experiment
  - Encrypted using OpenSSL to provide requested confidentiality

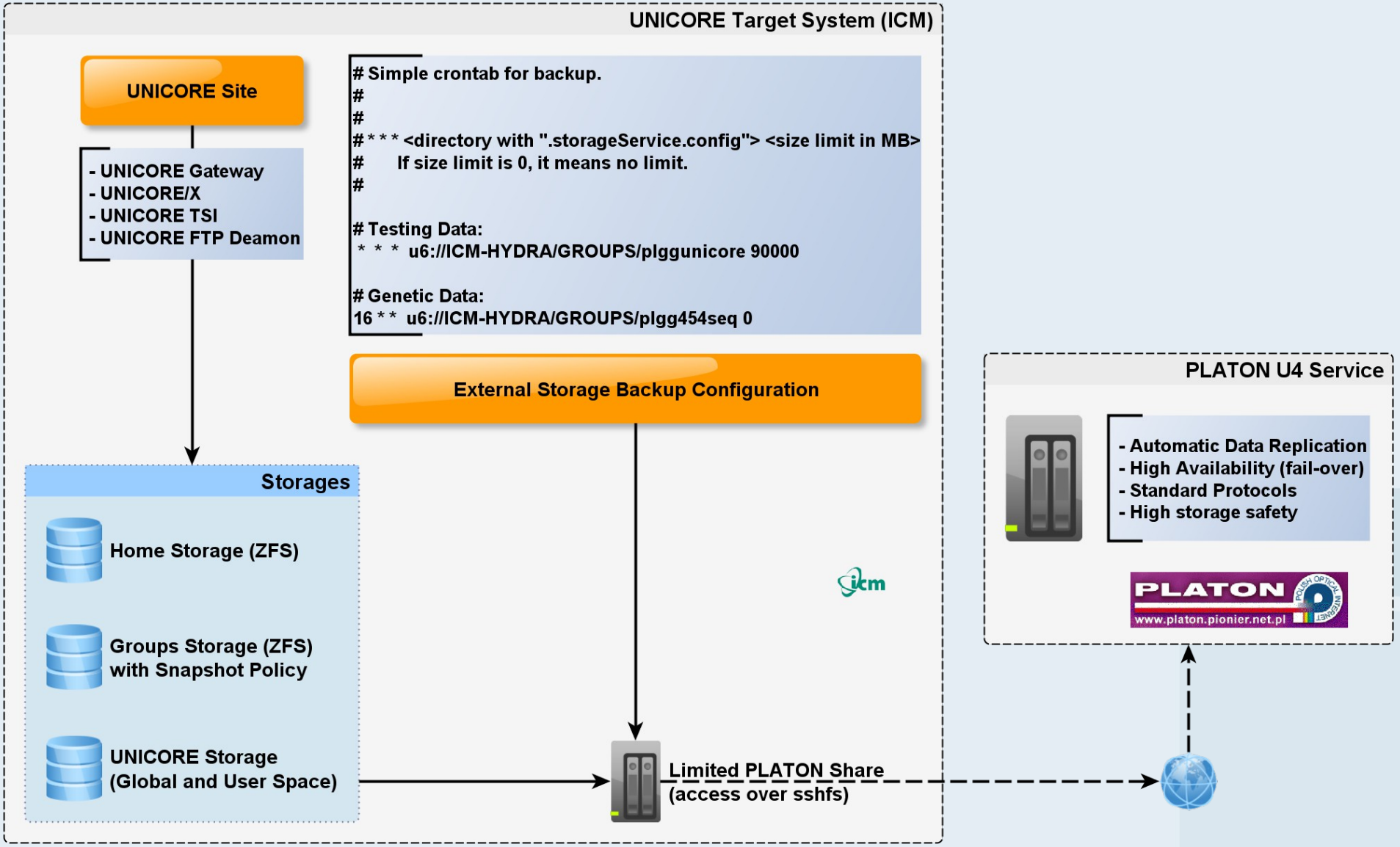
PLATON U4 Service

- Reliable data storage facility with easy and efficient access and replication
  - Automatic and transparent data and meta-data replication
  - Fail-over among multiple redundant components
  - Standard protocols including SFTP and WebDAV
  - High storage safety and security thanks to support for end-to-end data encryption, data integrity control, SSL-protected data transmission and security procedures employed at storage sites
- Deployed in the environment of Polish Optical Network PIONIER that offers 10Gbit channels for the data access and replication
- Main target users include the scientific and academic community in Poland, digital libraries, cultural heritage institutions and public administration
- http://storage.pionier.net.pl/en/

Genetic Data Storage – Features

- User has access to official PL-Grid storages via UNICORE
  - Home directories (standard Linux filesystem)
  - Groups directories (standard Linux filesystem)
- Every user may upload/download data with the help of UNICORE
  - Available protocols: BFT and multistreaming UFTP
  - Encryption should be enabled during the transfers
- Multi-level backup
  - Data stored on official PL-Grid storage (accessible by user at any time)
  - Storage snapshot run daily/weekly (may be accessible by user)
  - Backup to external storage (Platon U4) based on user request (access only by contacting with administrator)
- Data stored within external services are additionally encrypted

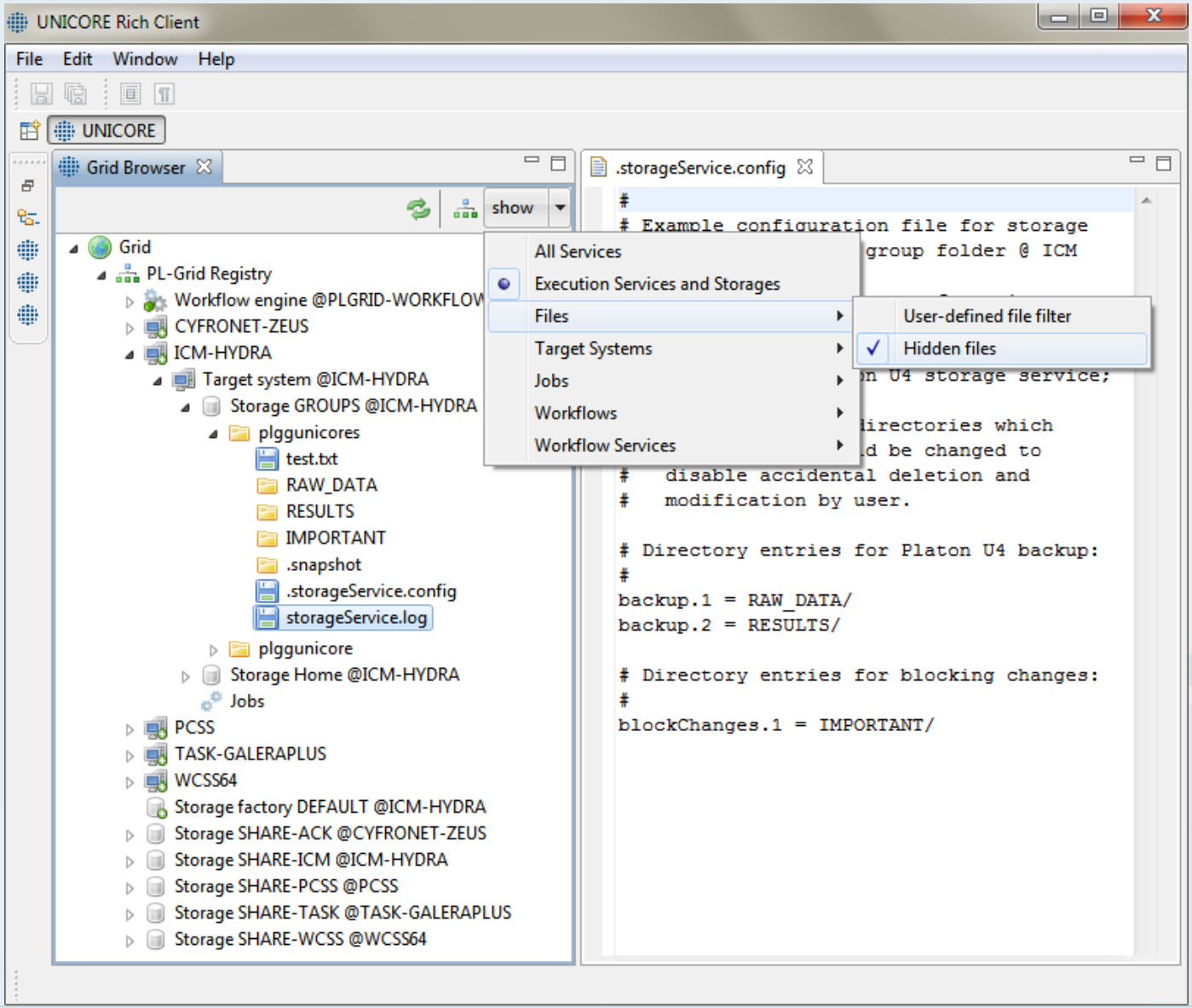
Handling Genetic Data Diagram



Genetic Data Storage – User PoV

- Registration in PL-Grid portal and activation of „Genetic Data Storage” domain service
- User requests for safe storage backup using HelpDesk
- Administrators verify the user and his purpose
- Determination of backup parameters:
  - Data localization (which type of storage: Home or GROUPS)
  - Backup frequency
  - Data size limit
- Administrators add entry to configuration and confirm that the service is ready
- User determines via simple configuration file which subdirecotires located at chosen storage will have external backup
- User verifies if backup was successful by checking log file located in his storage after every time the backup process was run

Group's Backup Configuration



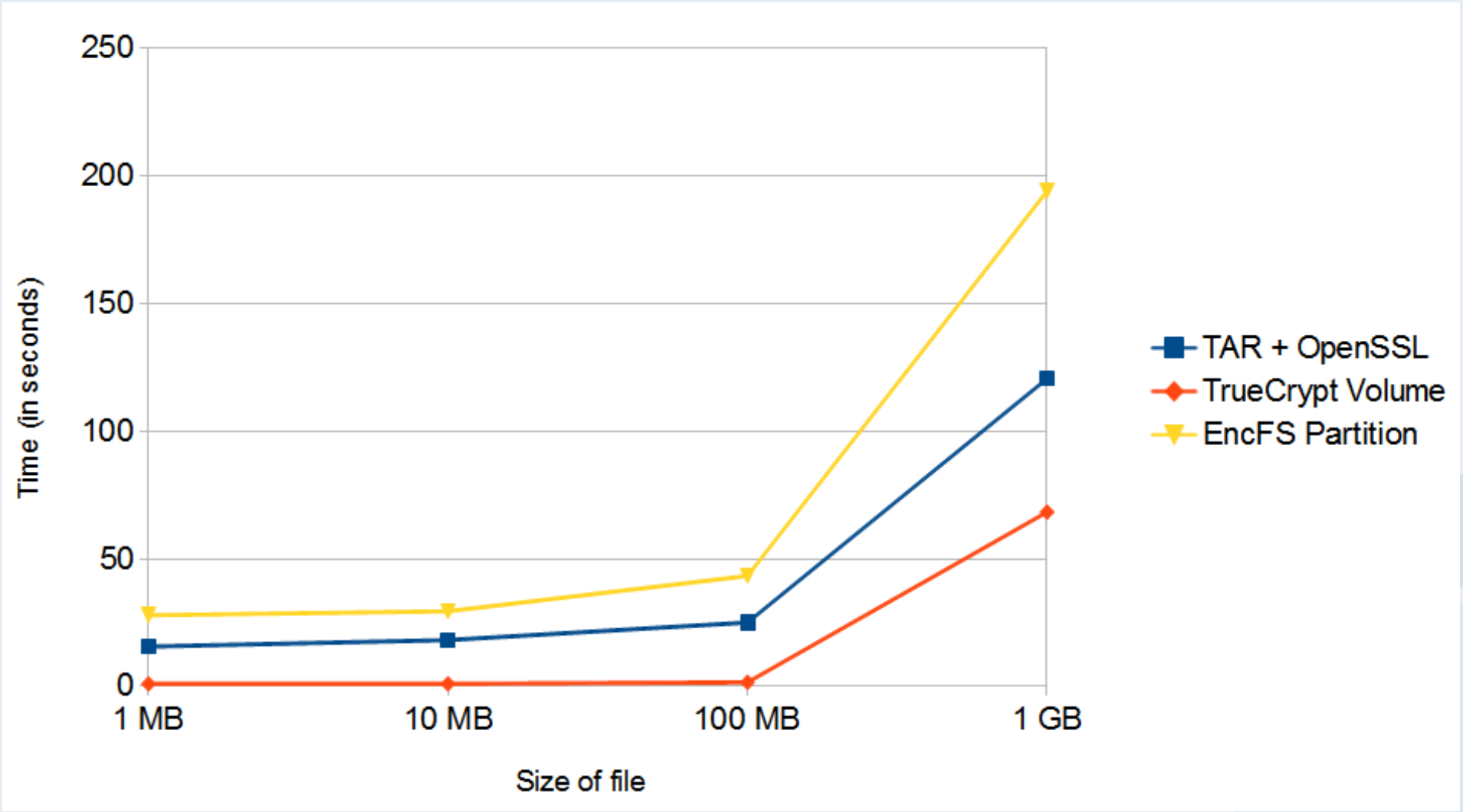
Backup Observations

- When using TAR+OpenSSL (enc)
  - Changing one file forces backup of whole subdirectory
  - Whole process should be faster (at least when there are changes of few files)
- Some latest results
  - 217 GB → about 6h 27'
  - 229 GB → about 10h 30'
  - 218 GB → about 6h 40'
  - 257 GB → about 8h 41'
- Slow process
  - Need to check network connection (find bootlenecks)
  - OpenSSL encryption reduces speed almost by half
- Check some other solutions: TrueCrypt / EncFS (over PLATON share)

Tests using PLATON U4 Storage

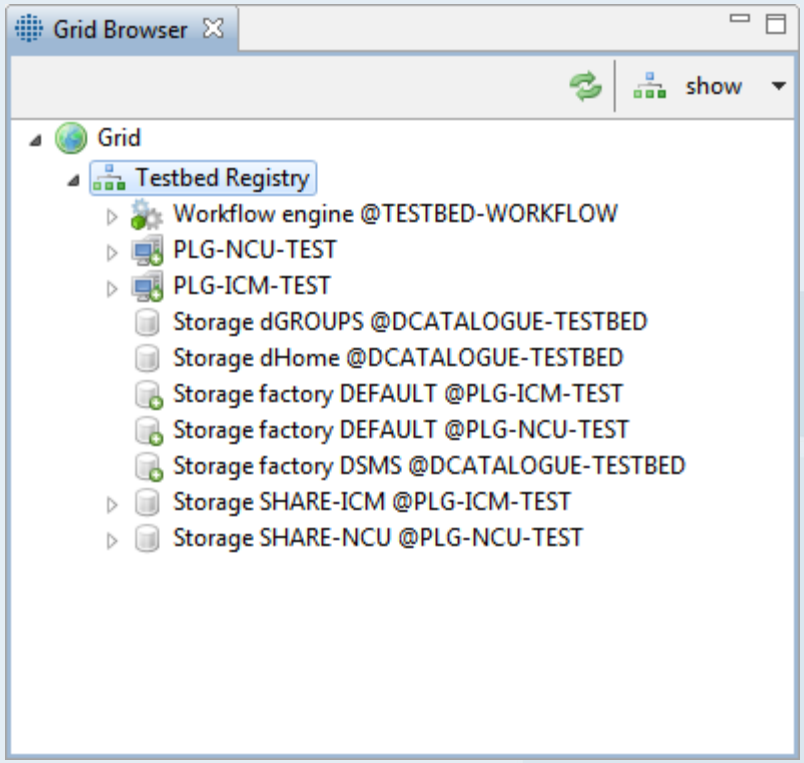
- TrueCrypt 7.1a
  - Created 2GB (normal) volume as Linux Ext4 file system
  - AES encryption algorithm, RIPEMD-160 has algorithm
  - Using password, no keyfile
- EncFS 1.7.4-1
  - Using standard configuration during setup
  - Filesystem cipher: "ssl/aes", version 3:0:2
  - Filename encoding: "nameio/block", version 3:0:1
  - Key size: 192 bits, block size: 1024 bytes
  - Each file contains 8 byte header with unique IV data.
  - Filenames encoded using IV chaining mode.
- TrueCrypt volume and EncFS raw data stored directly on PLATON U4 storage (sshfs)

Time Comparision of Single File



Summary

- Data problem is becoming more and more important
- Distributed Storage Managment Service (DSMS)
  - Still have some issues (many tests were made)
  - It works as a proxy to SMSes, it won't be fast
  - Allows easy access to Target System Storages
- Issues we will look at:
  - SMS metadata access
  - UFTP server-to-server transfers (also as a part of workflows)



Thank You for Attention

- Questions
- Remarks
- Ideas
- ?