



Globalization of data access on computing infrastructures

Ł. Dutka, T. Lichoń, R. Słota, K. Zemek, M. Wrzeszcz, R.G. Słota, and J. Kitowski

ACC Cyfronet AGH
Department of Computer Science, AGH -UST

CGW'14
Kraków, Poland, October 27-29, 2014

Agenda



- Scientific data access – current state
- User requirements
- Existing solutions
- onedata
 - system overview
 - challenges
 - current state
- Conclusions



Scientific data access – current state



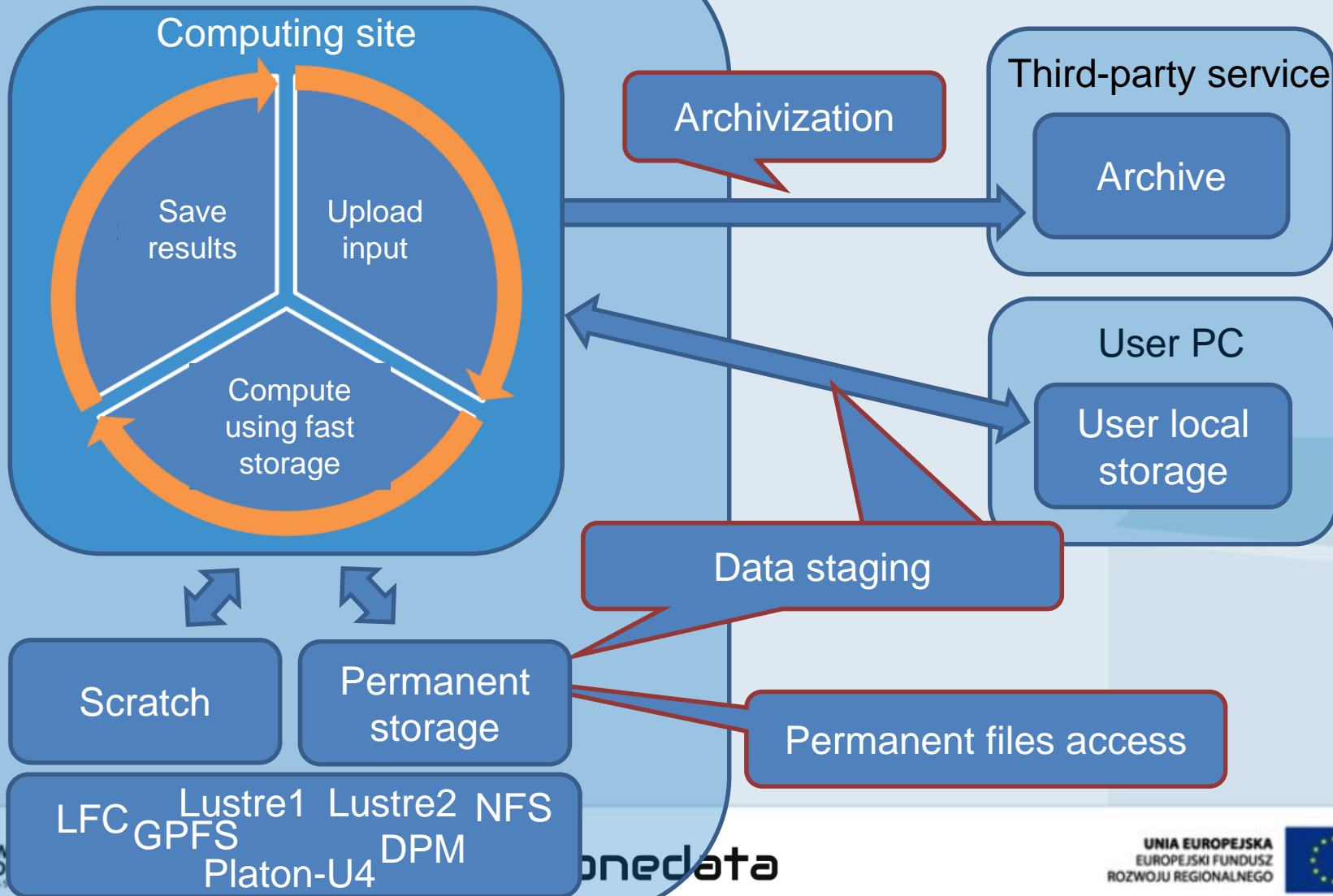
- Variety of entities with separate access endpoints
- High heterogeneity of storage with different access interfaces
 - Need for storage with various characteristics
 - Use of legacy storage resources
 - Local sites make autonomous decisions about their storage hardware and software



Scientific data access – current state (PL-Grid)



PL-Grid distributed infrastructure



User requirements



- Location transparency
- High performance transfer
- Usability & ease of access
- Decentralization

Data Globalization



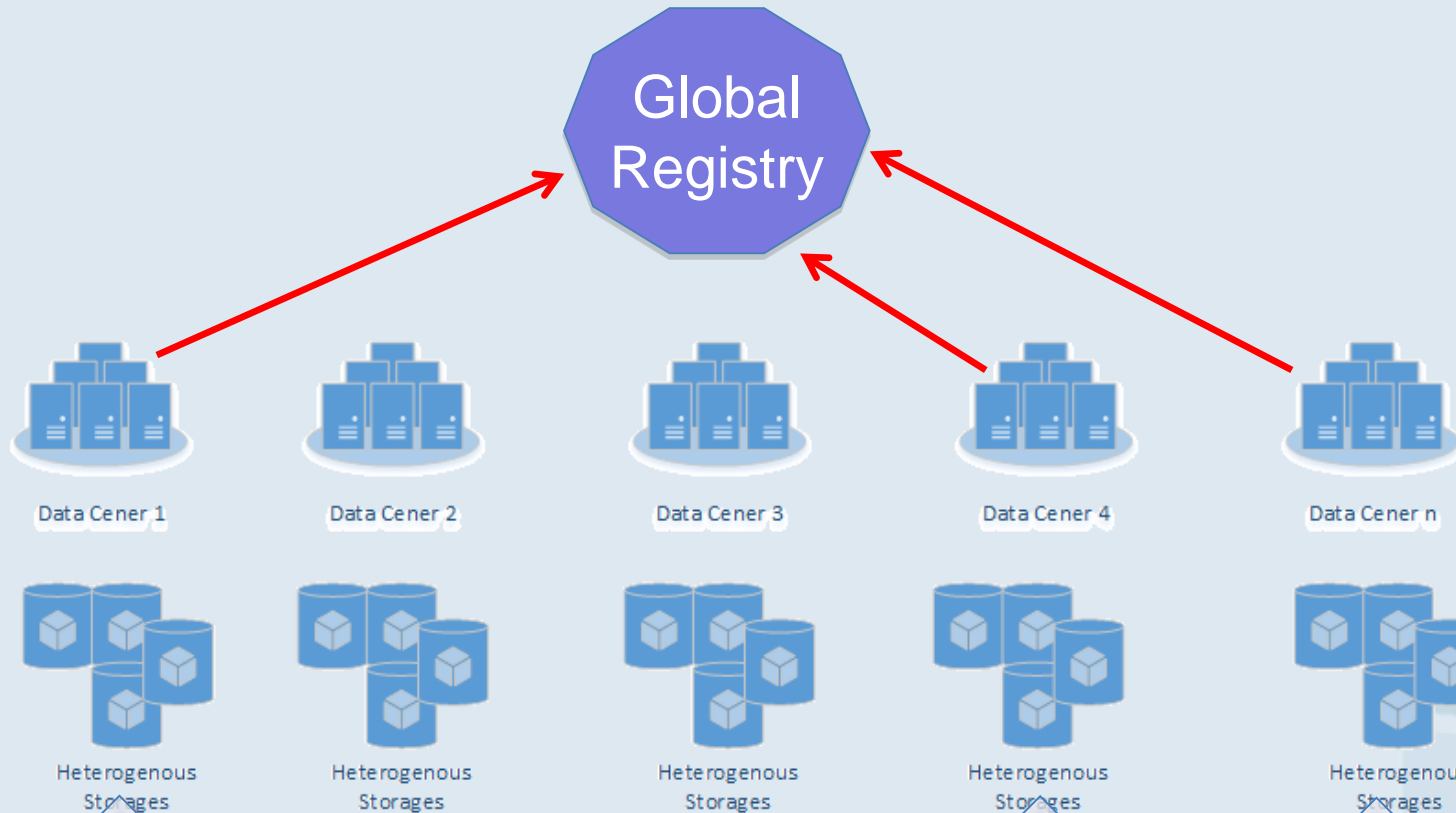
Existing solutions



- **Location-transparent filesystems**
 - GlusterFS, PanFS, Scality RING
 - Centrally managed
- **Federated data-management systems**
 - iRODS, Parrot
 - No location transparency
- **Data transfer services**
 - Globus Connect
 - No location transparency



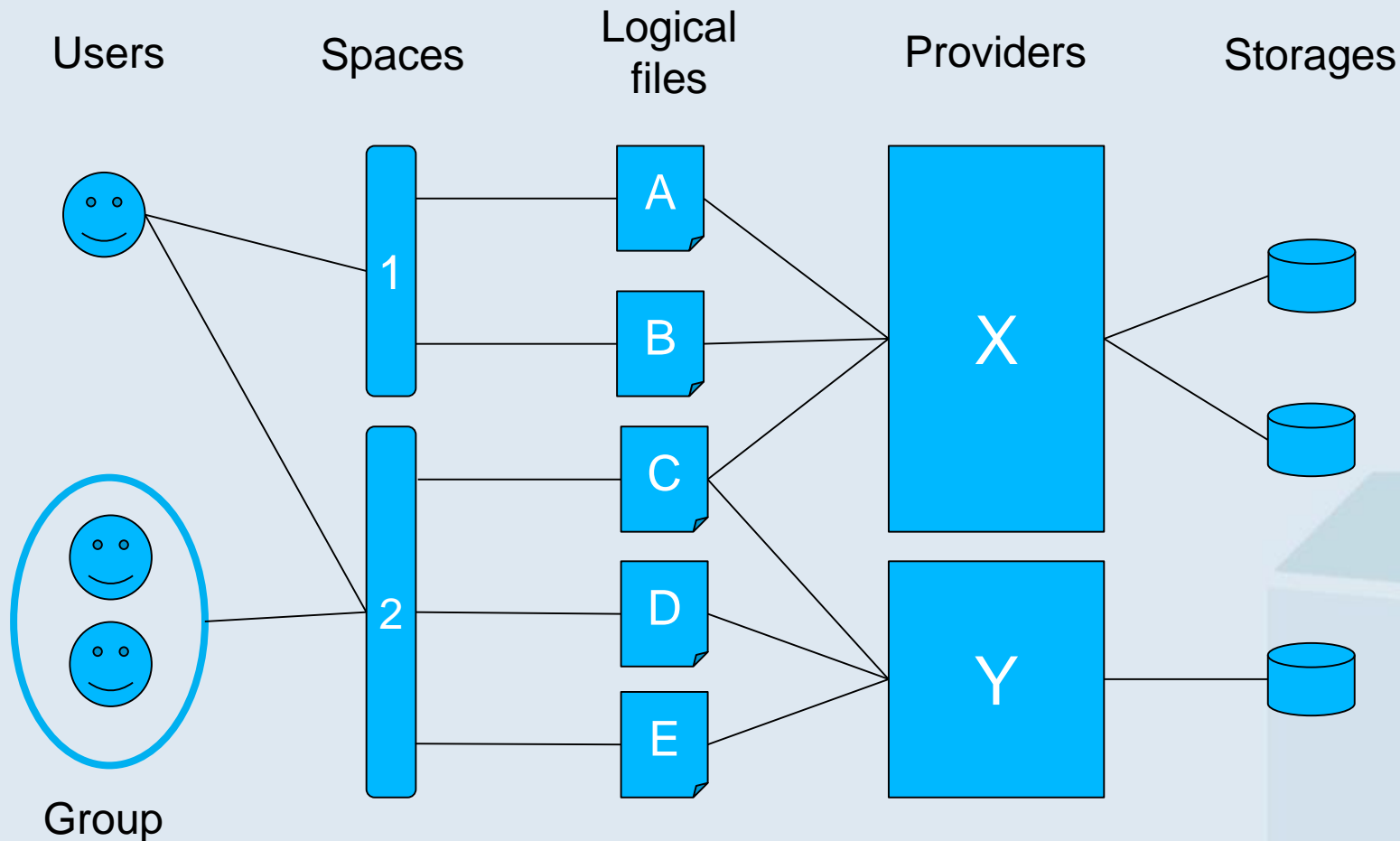
onedata – system overview



onedata globally unified space for User1 provided by many DCs



onedata – challenges: simplicity of files organization



onedata – challenges: data center cooperation



■ **Metadata synchronization**

Providers supporting a common space need to keep files' metadata in sync

■ **On-demand file transfer**

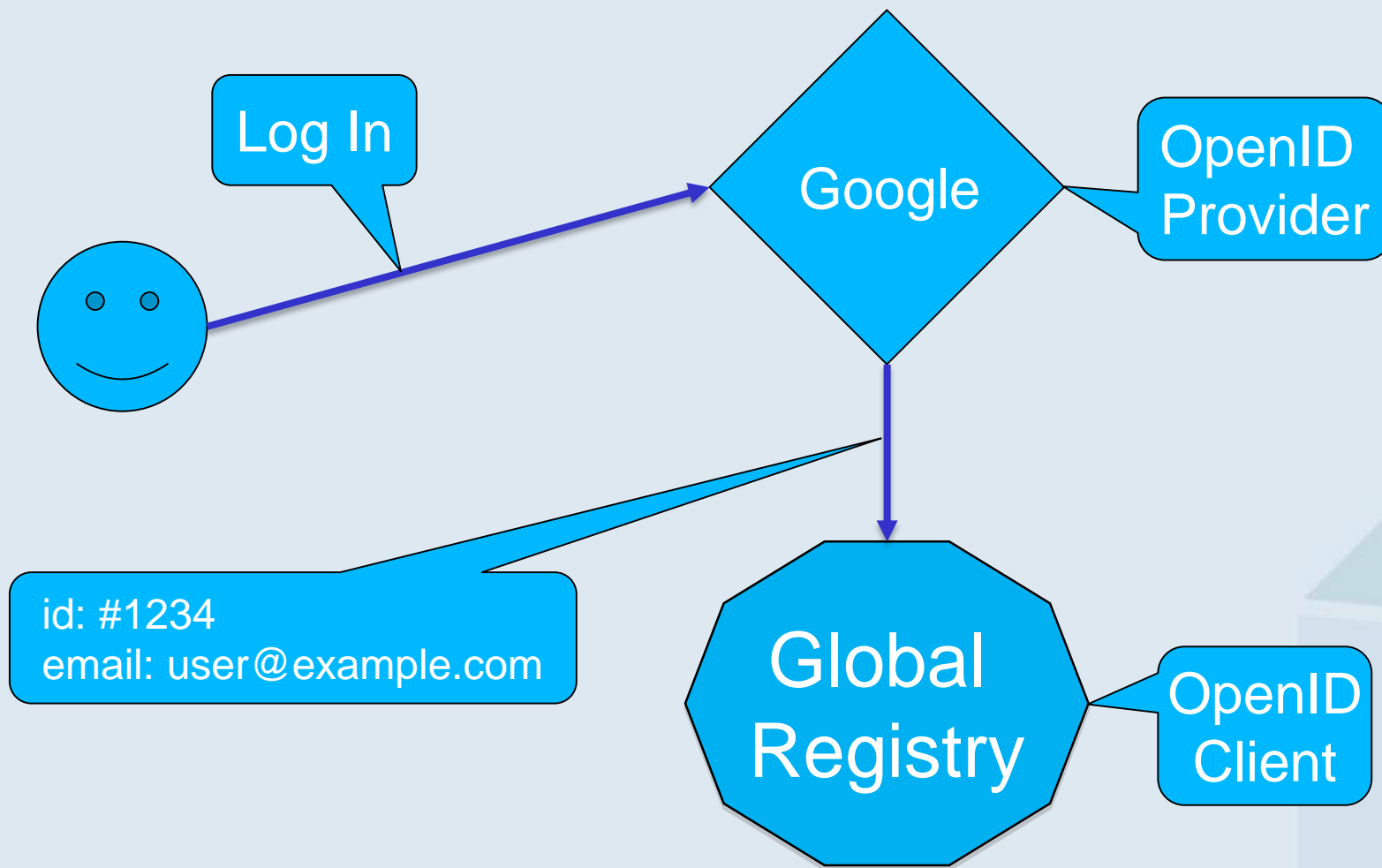
Create a file while connected to one provider, read the file from the other

■ **Global Registry as a mediator**

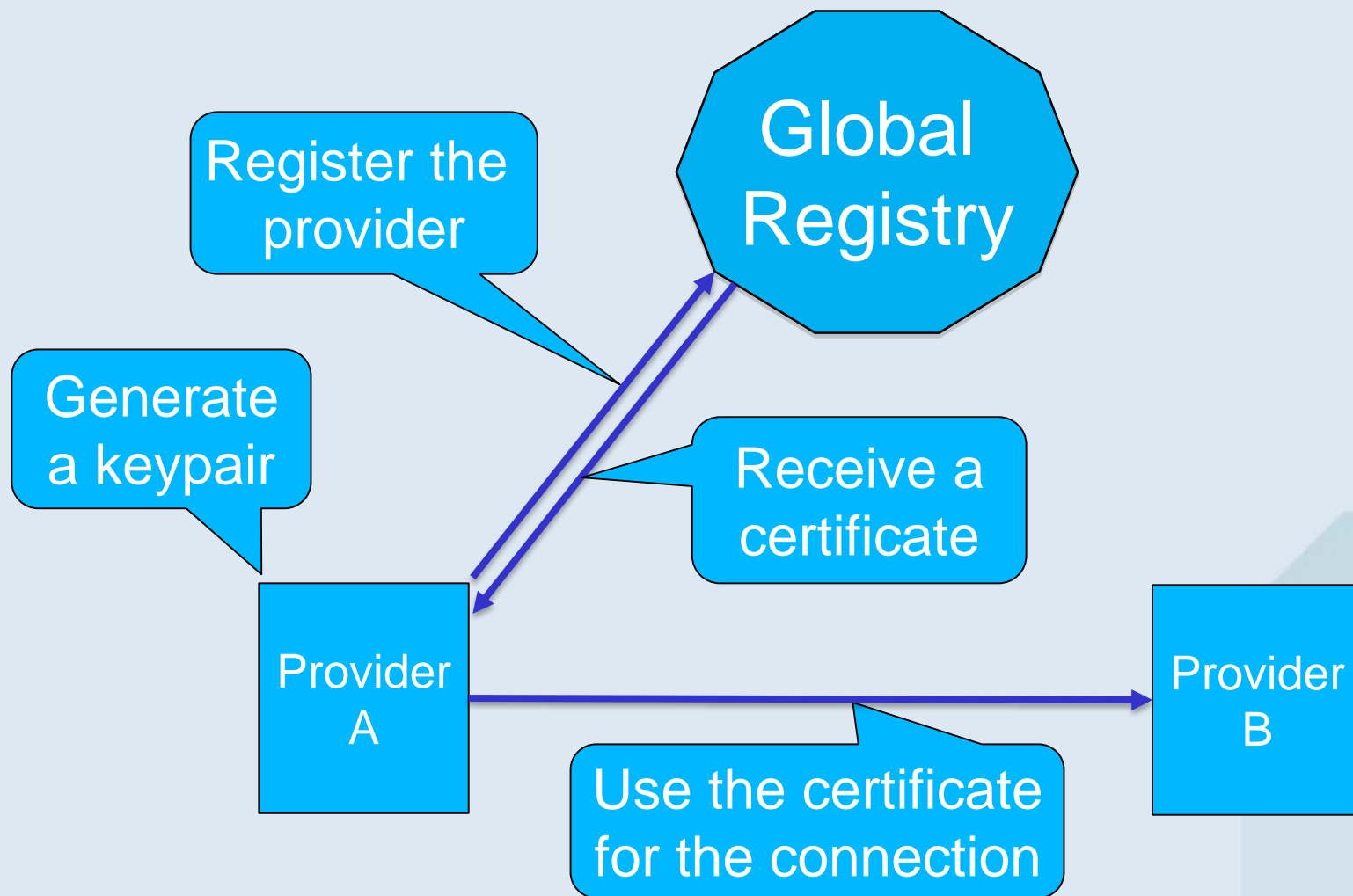
Global registry oversees synchronization and transfer processes



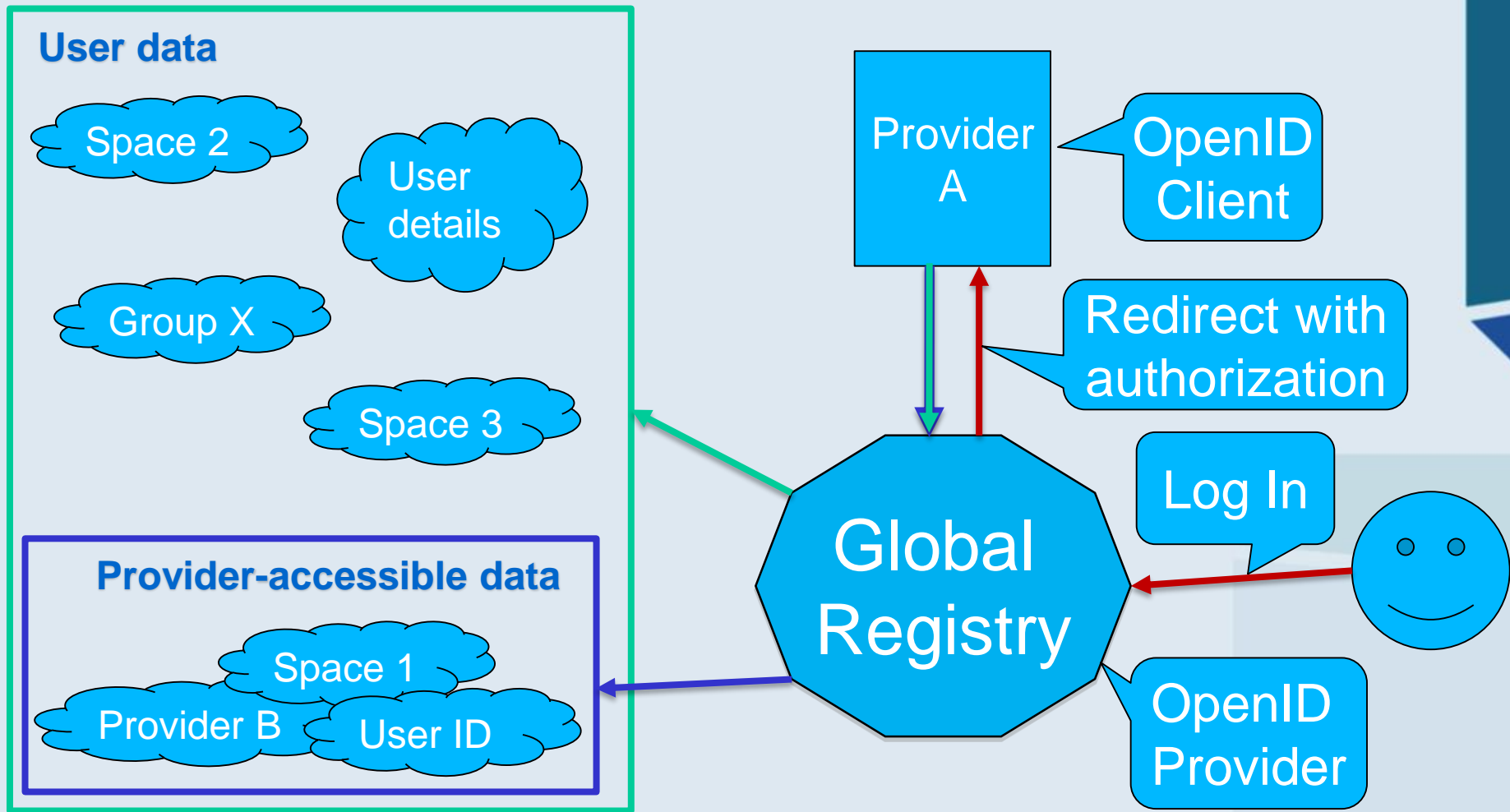
onedata – challenges: security – OpenID Client



onedata – challenges: security – Certification Authority



onedata – challenges: security – OpenID Provider



onedata – current state



■ Release 2.0 deployment started

- Unified cross-datacenter access from anywhere
- Support for file sharing and publishing
- System monitoring for administrators
- Multi-provider space support
- CDMI support

■ Plans for the near future

- Support for grants and accounting
- Metadata support
- Intelligent, automatic data management
- Management GUI for administrators
- File locks



Conclusions



- onedata is a result of research projects:
 - PLGrid PLUS (v1.5 deployed at ACK Cyfronet)
 - PLGrid CORE (development of v2.0)
- Current solutions are not sufficient for scientific environment
- Meets requirements of users, administrators & developers
 - Sought after by multinational federation - EGI





Thank you

