Metadata Synchronization Protocol for a Decentralized Network of Data Providers

Łukasz Opioła | Łukasz Dutka | Renata G. Słota | Jacek Kitowski

ACK CYFRONET AGH, Kraków, Poland

AGH University of Science and Technology, Kraków, Poland

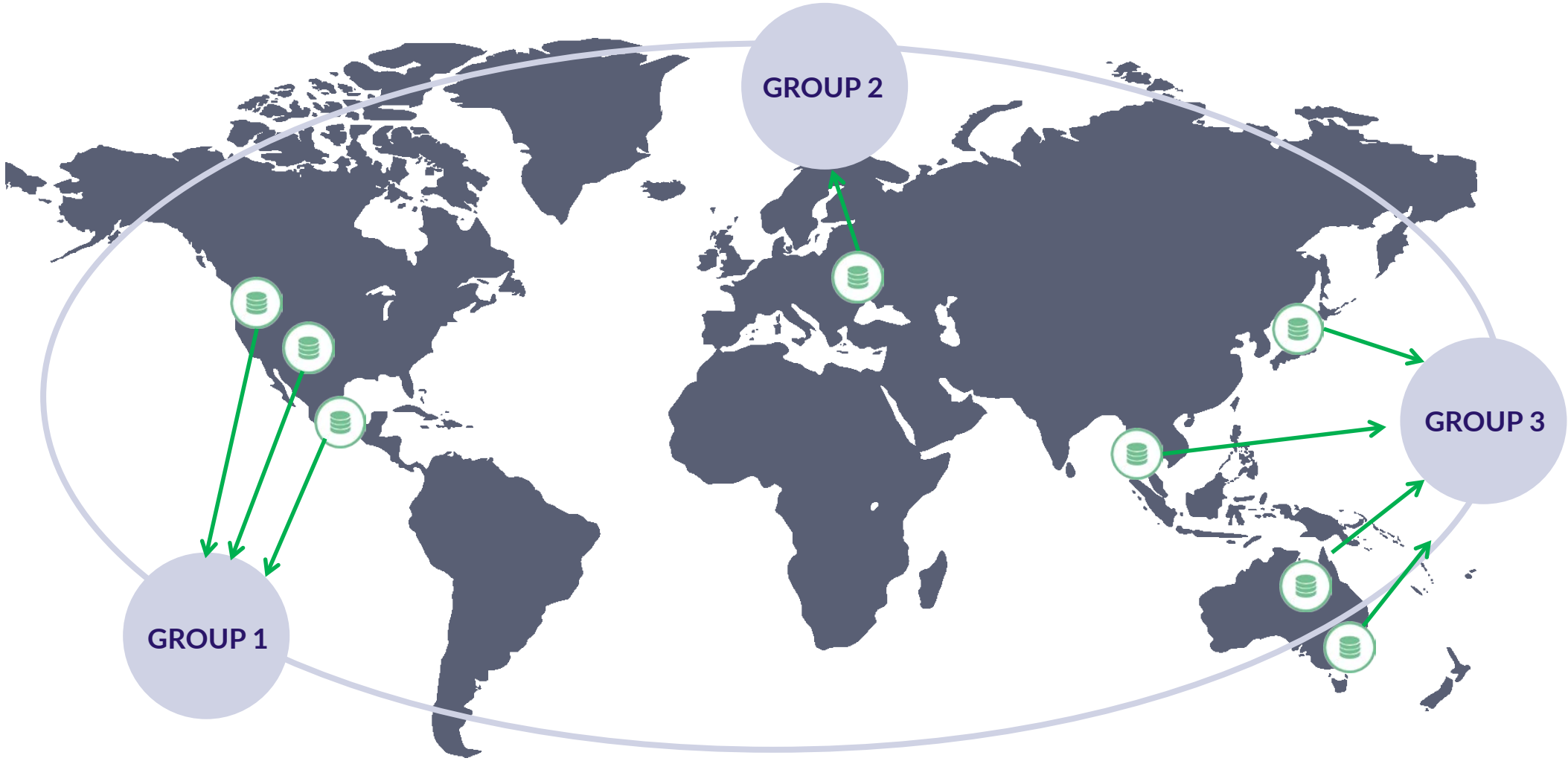
Faculty of Computer Science, Electronics and Telecommunications

Department of Computer Science

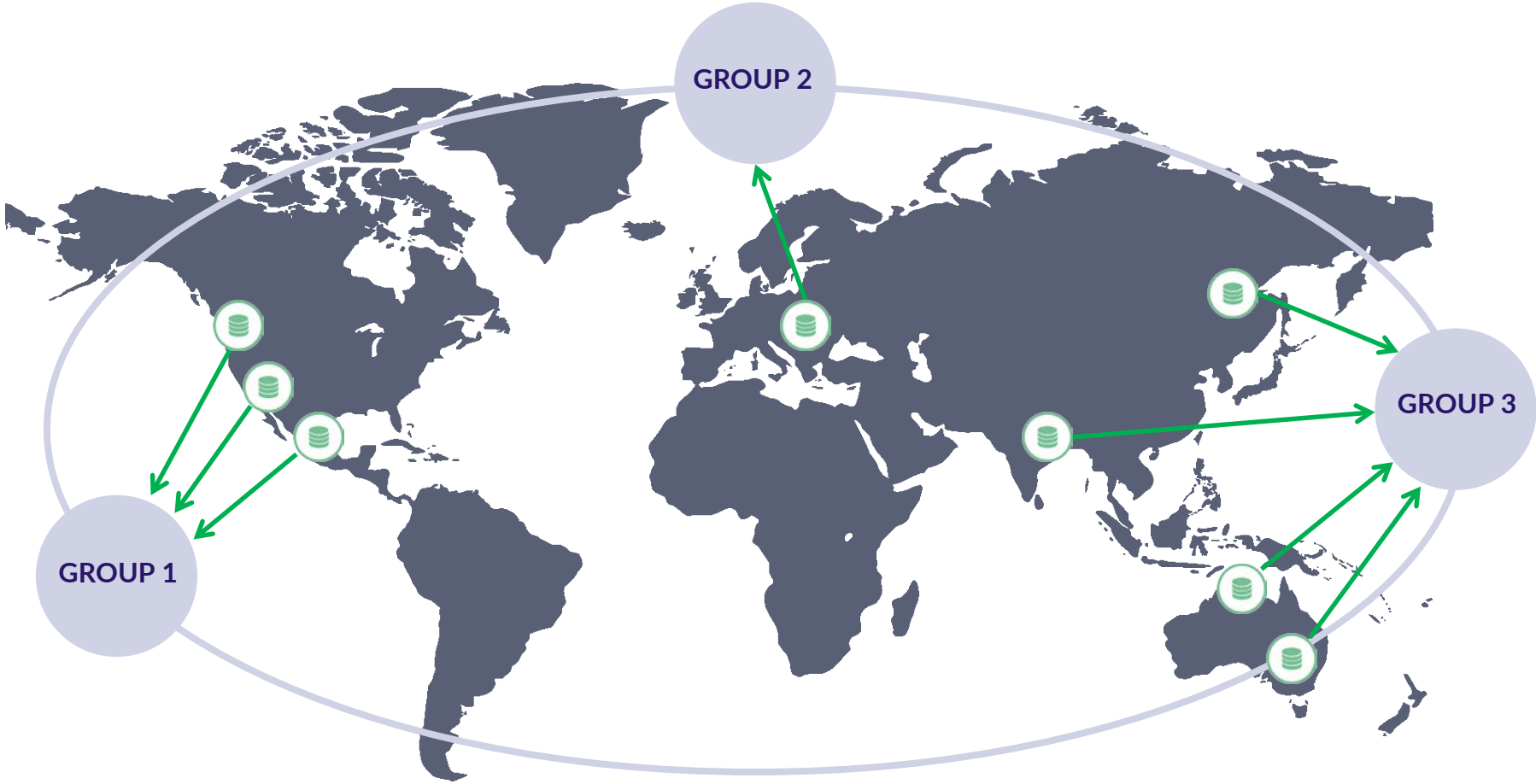
AGENDA

1. Global data access for modern science
2. Challenges of global data access
3. Metadata synchronization protocol - requirements
4. **Proposed concept of metadata synchronization protocol**
5. Conclusions

GLOBAL DATA ACCESS FOR MODERN SCIENCE

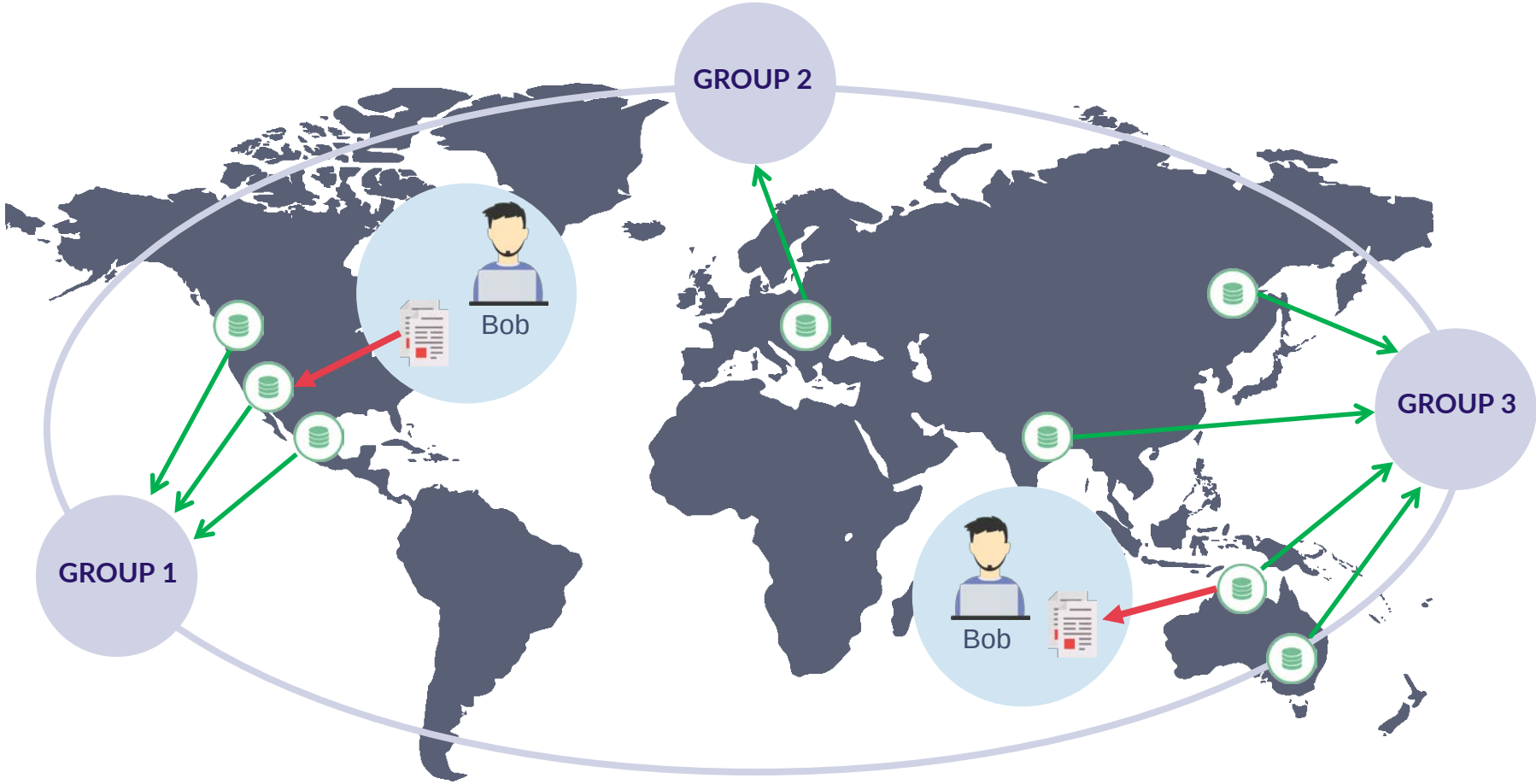


CHALLENGES OF GLOBAL DATA ACCESS



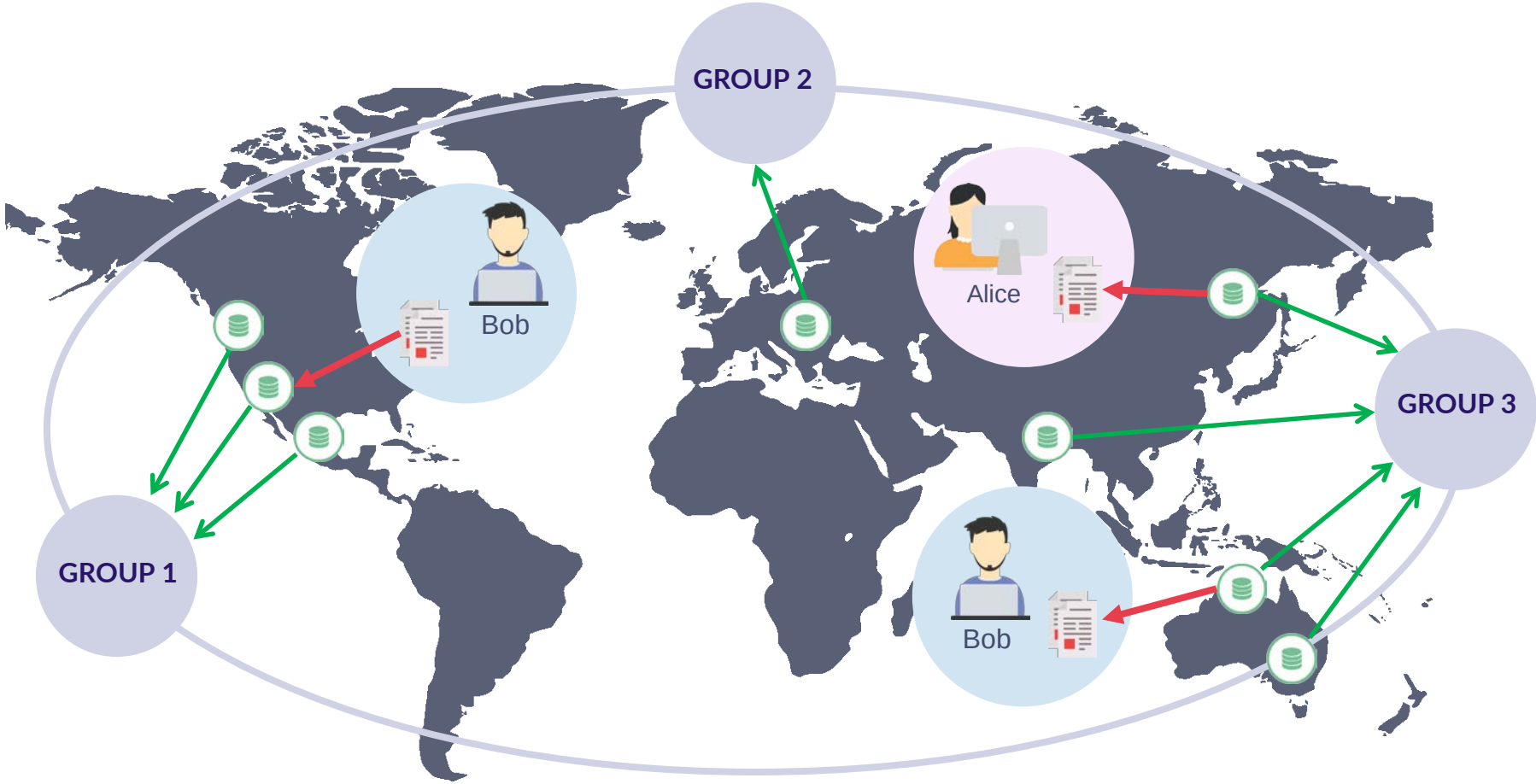
CHALLENGES OF GLOBAL DATA ACCESS

- Transparent data access



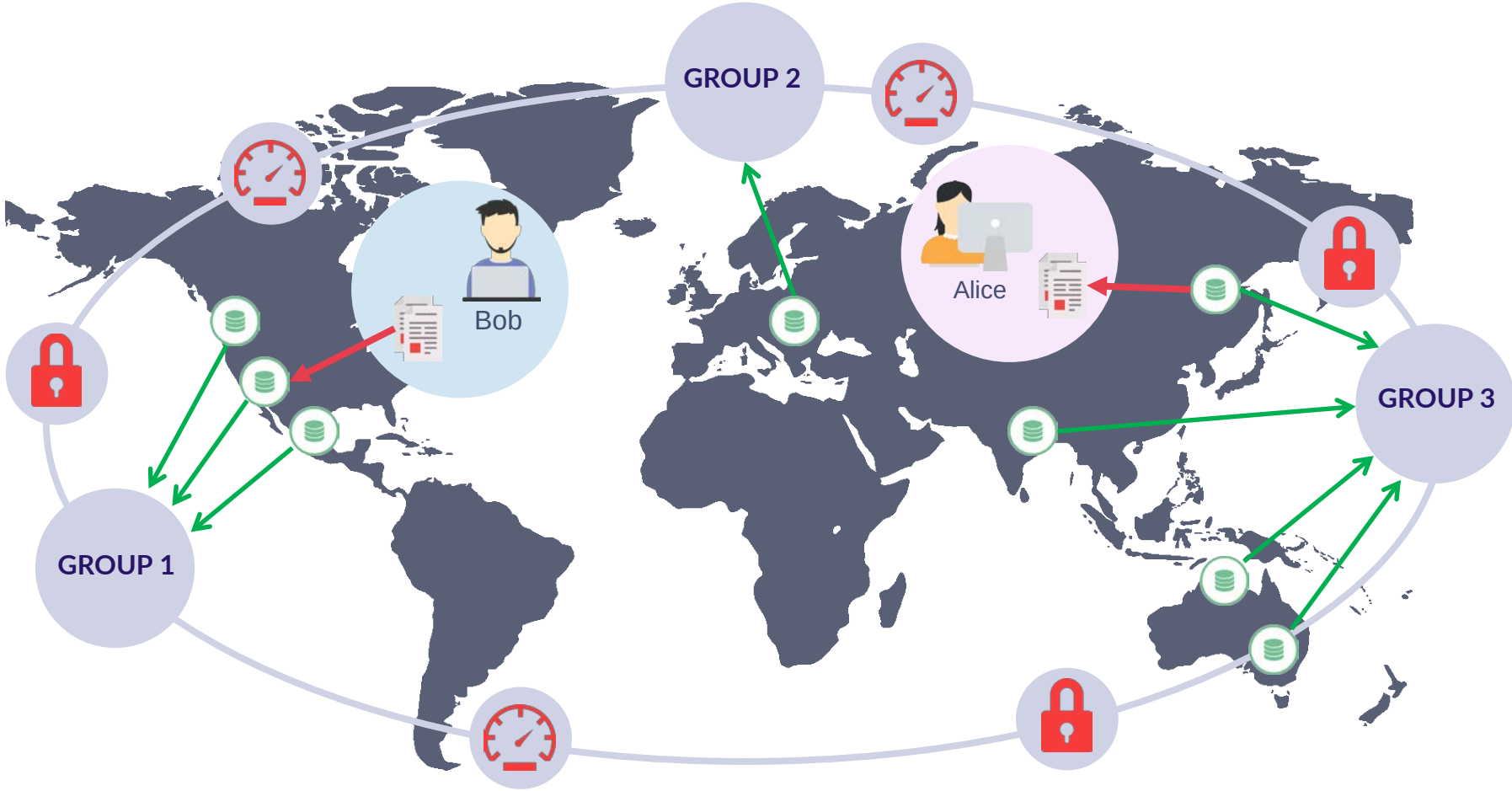
CHALLENGES OF GLOBAL DATA ACCESS

- Transparent data access
- Cross-border collaboration



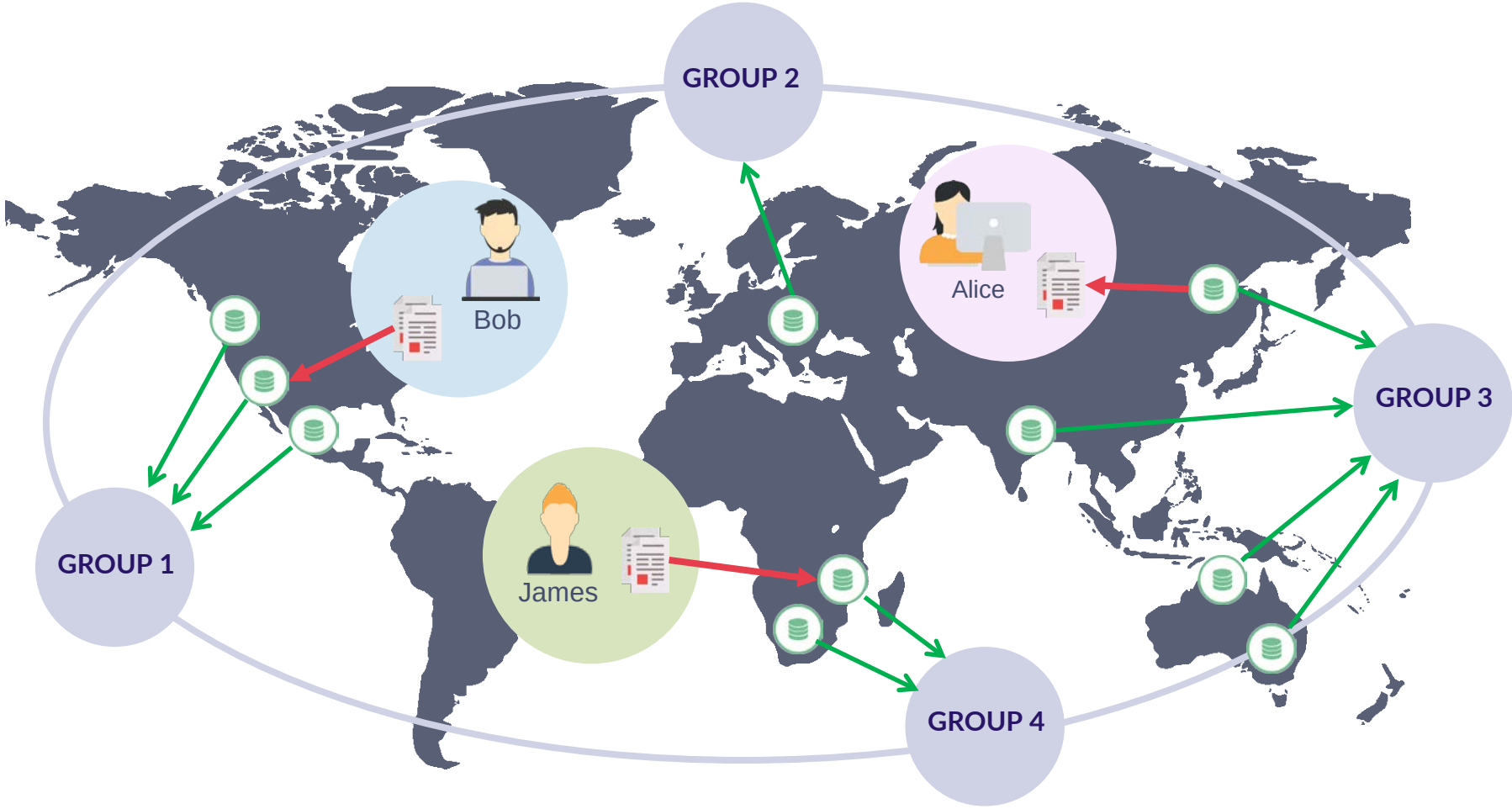
CHALLENGES OF GLOBAL DATA ACCESS

- Transparent data access
- Cross-border collaboration
- Security & efficiency



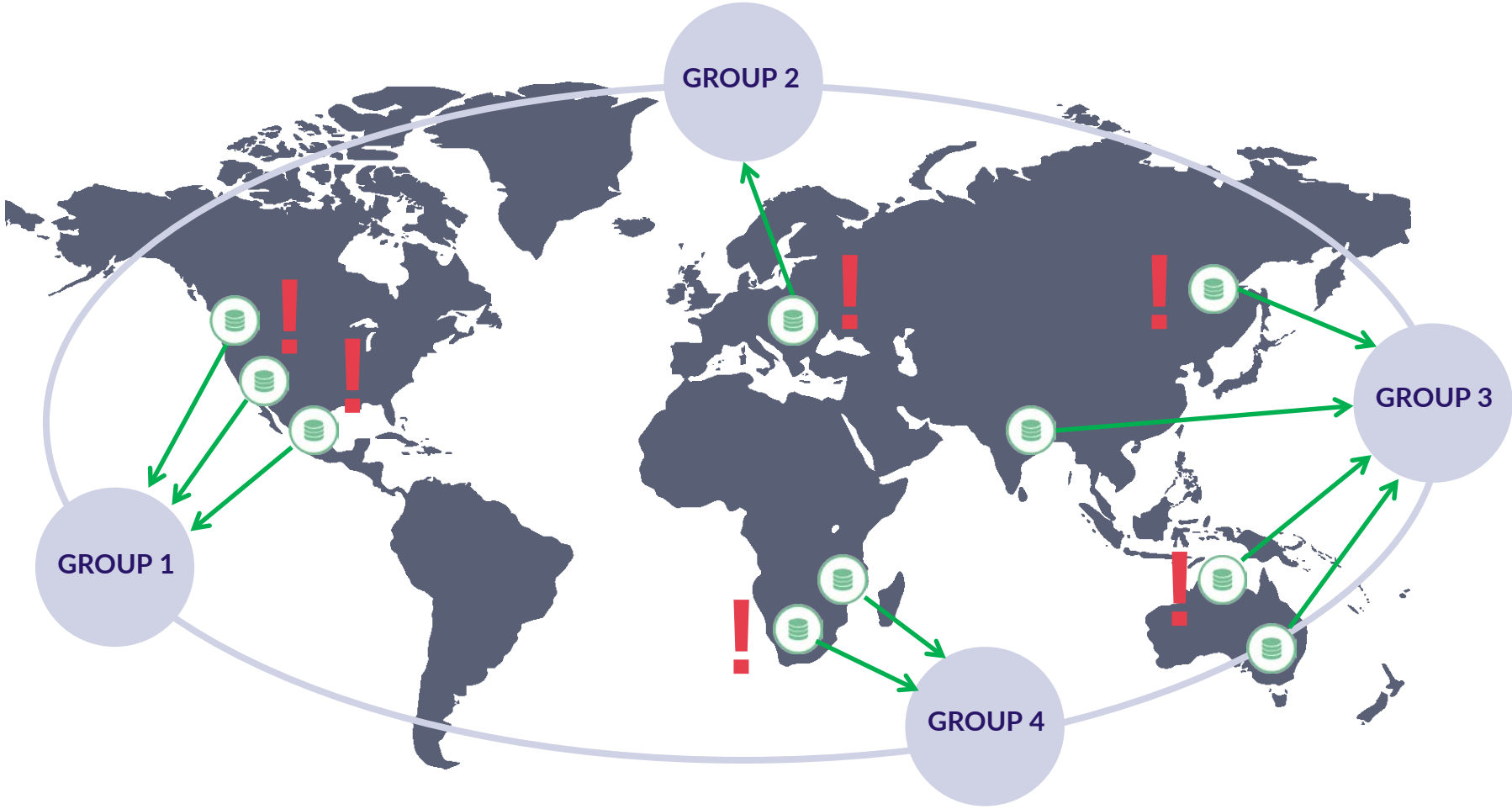
CHALLENGES OF GLOBAL DATA ACCESS

- Transparent data access
- Cross-border collaboration
- Security & efficiency
- Openness of the network



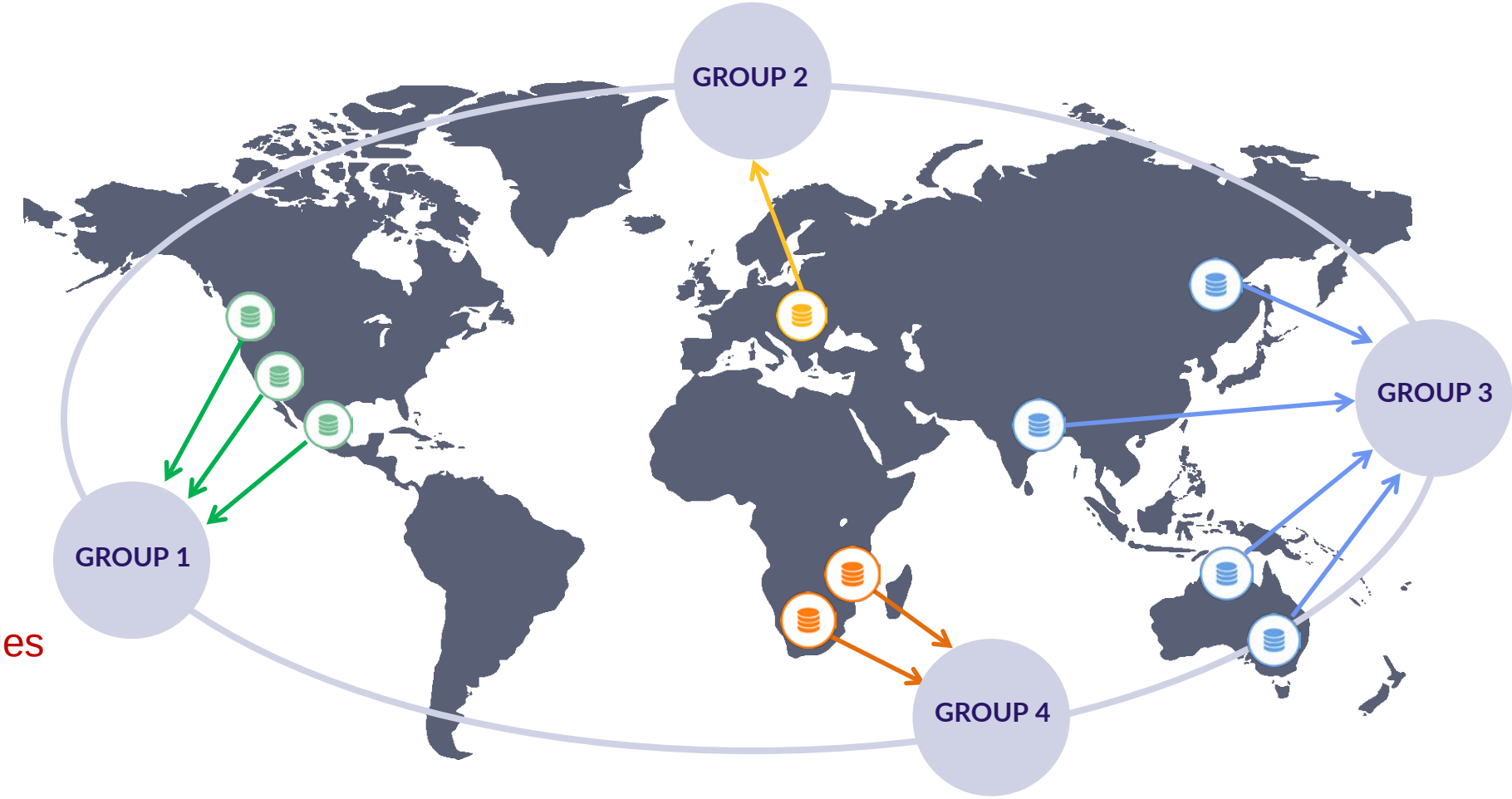
CHALLENGES OF GLOBAL DATA ACCESS

- Transparent data access
- Cross-border collaboration
- Security & efficiency
- Openness of the network
- Autonomy (lack of trust)



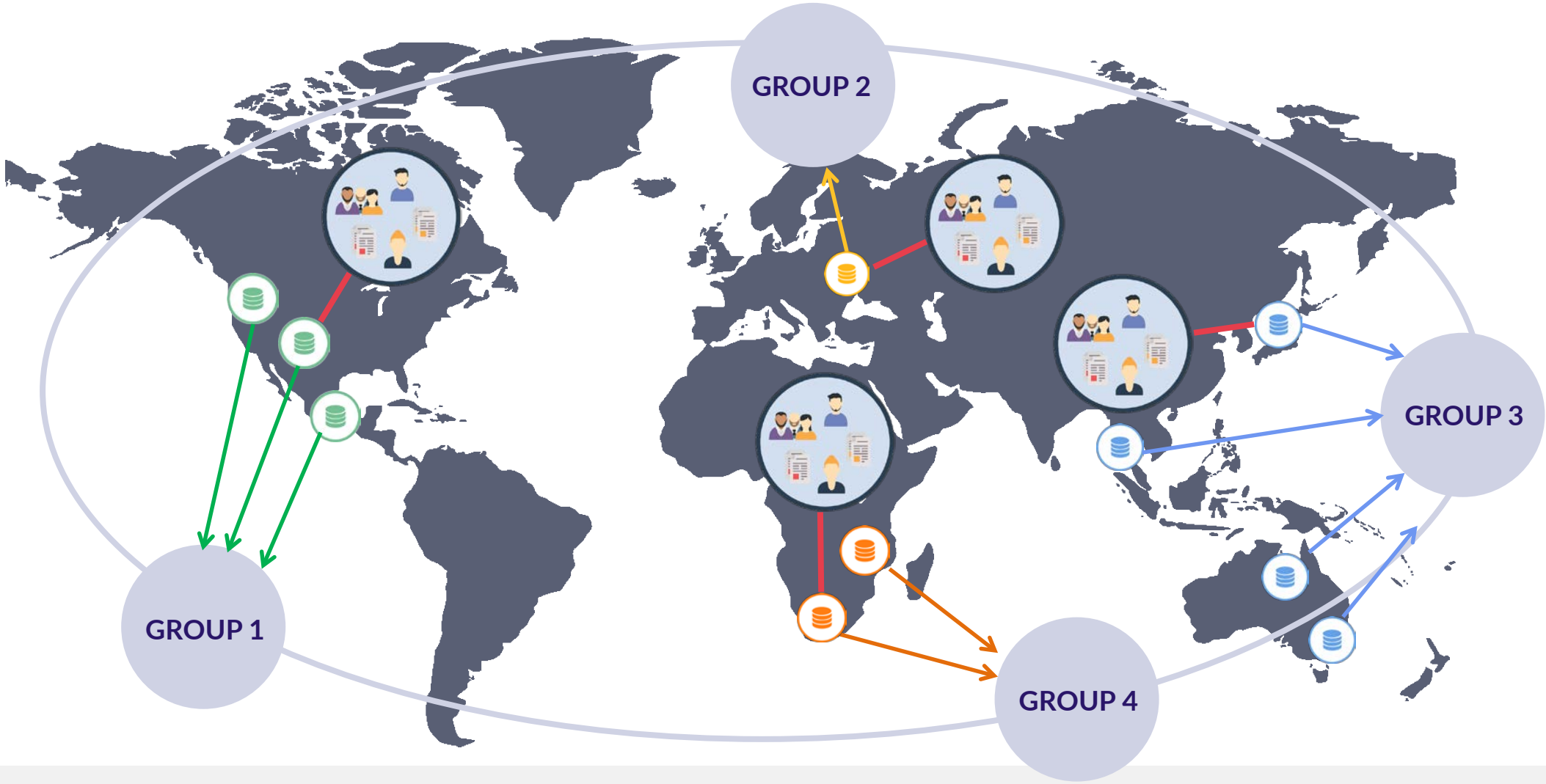
CHALLENGES OF GLOBAL DATA ACCESS

- Transparent data access
- Cross-border collaboration
- Security & efficiency
- Openness of the network
- Autonomy (lack of trust)
- Reflecting existing hierarchies



METADATA SYNCHRONIZATION PROTOCOL

METADATA SYNCHRONIZATION



METADATA SYNCHRONIZATION - REQUIREMENTS

- Decentralized (retaining autonomy)
- Globally consistent
- Scalable – hundreds of providers
- Secure – decentralized AAI*
- Location discovery mechanisms

** AAI – authentication and authorization infrastructure*

Proposed solution

Metadata synchronization protocol

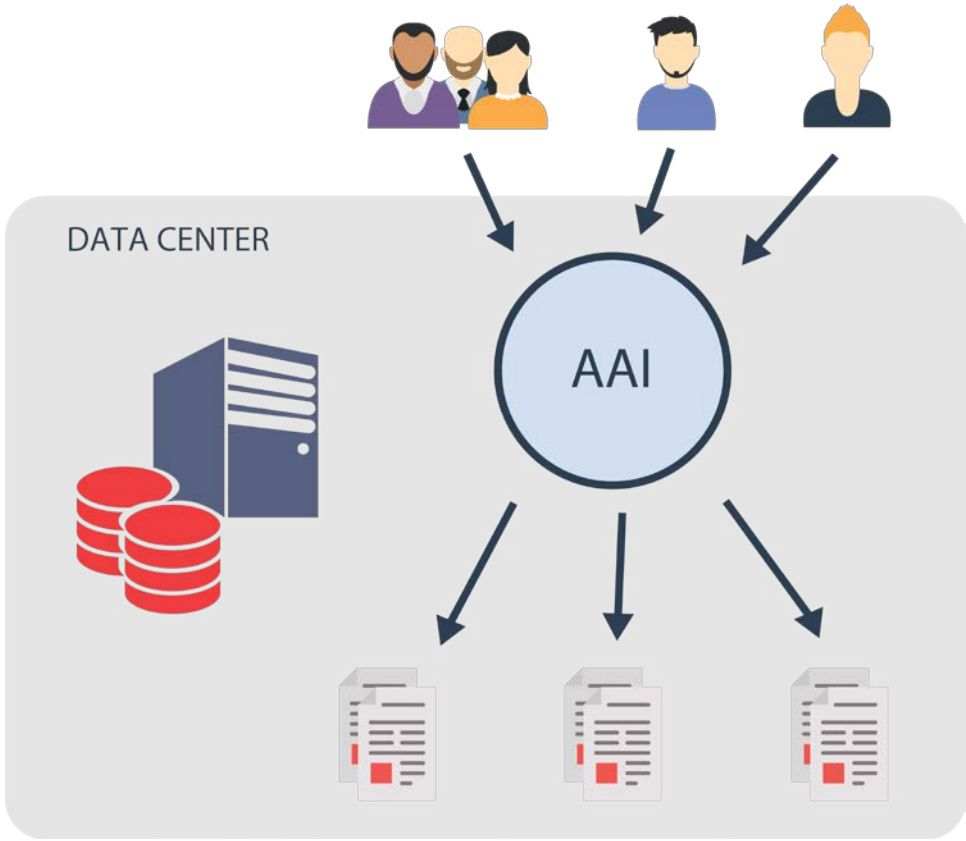
based on

hybrid, multi-tier architecture

PROPOSED SOLUTION – PROVIDER SCOPE

Provider service – basic building block

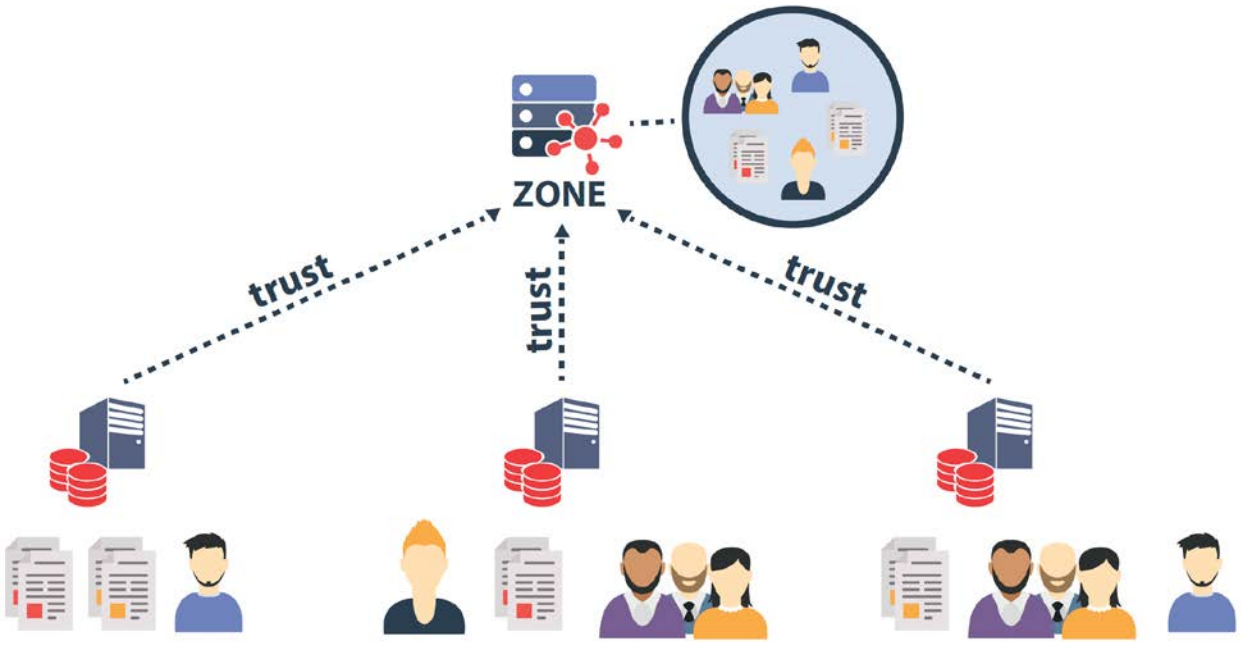
- Gathers knowledge using metadata sync
- Manages data on underlying storages



AAI – Authentication and Authorization Infrastructure

PROPOSED SOLUTION – ZONE SCOPE

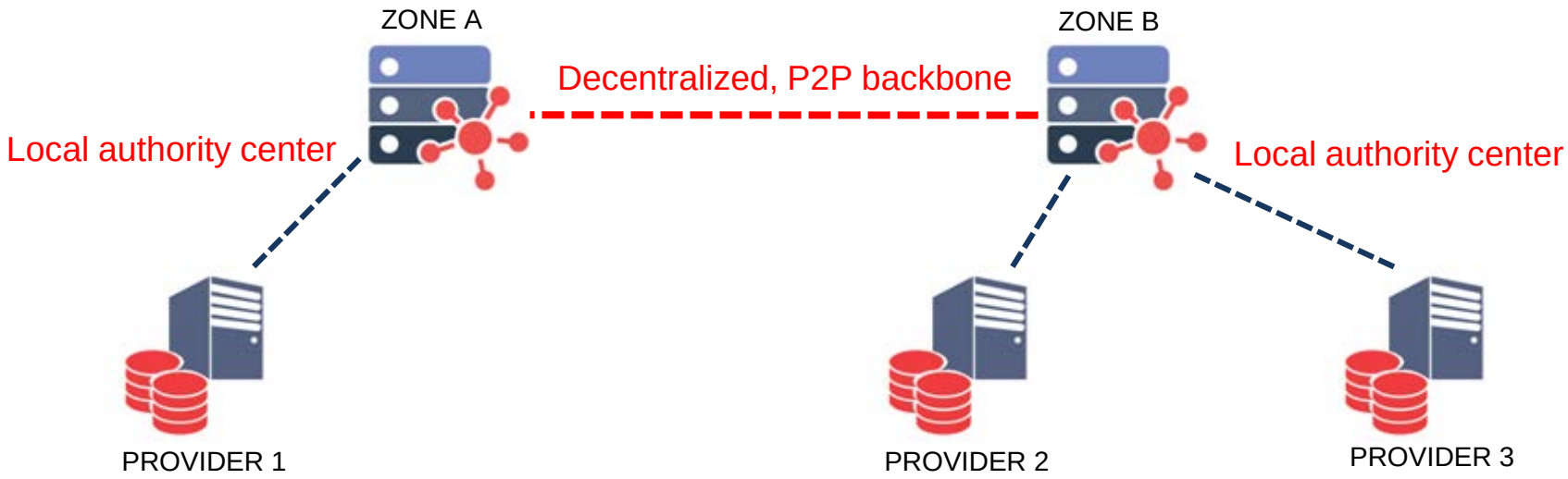
Zone service – oversees a group of providers



- Gathers and serves metadata
- Trusted authority & mediator
- AuthN & AuthZ center
- Macaroon based tokens
- Reflects existing hierarchies

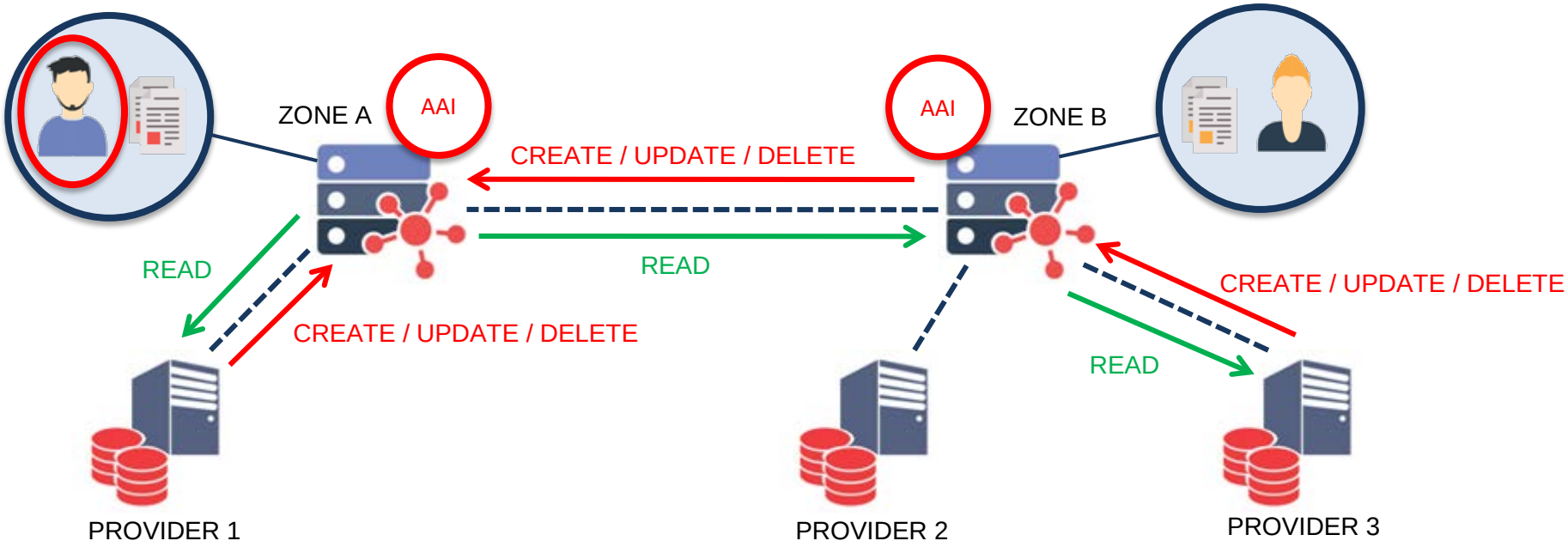
PROPOSED SOLUTION – GLOBAL SCOPE

Cooperation between *Zones* – hybrid architecture



PROPOSED SOLUTION – METADATA SYNC

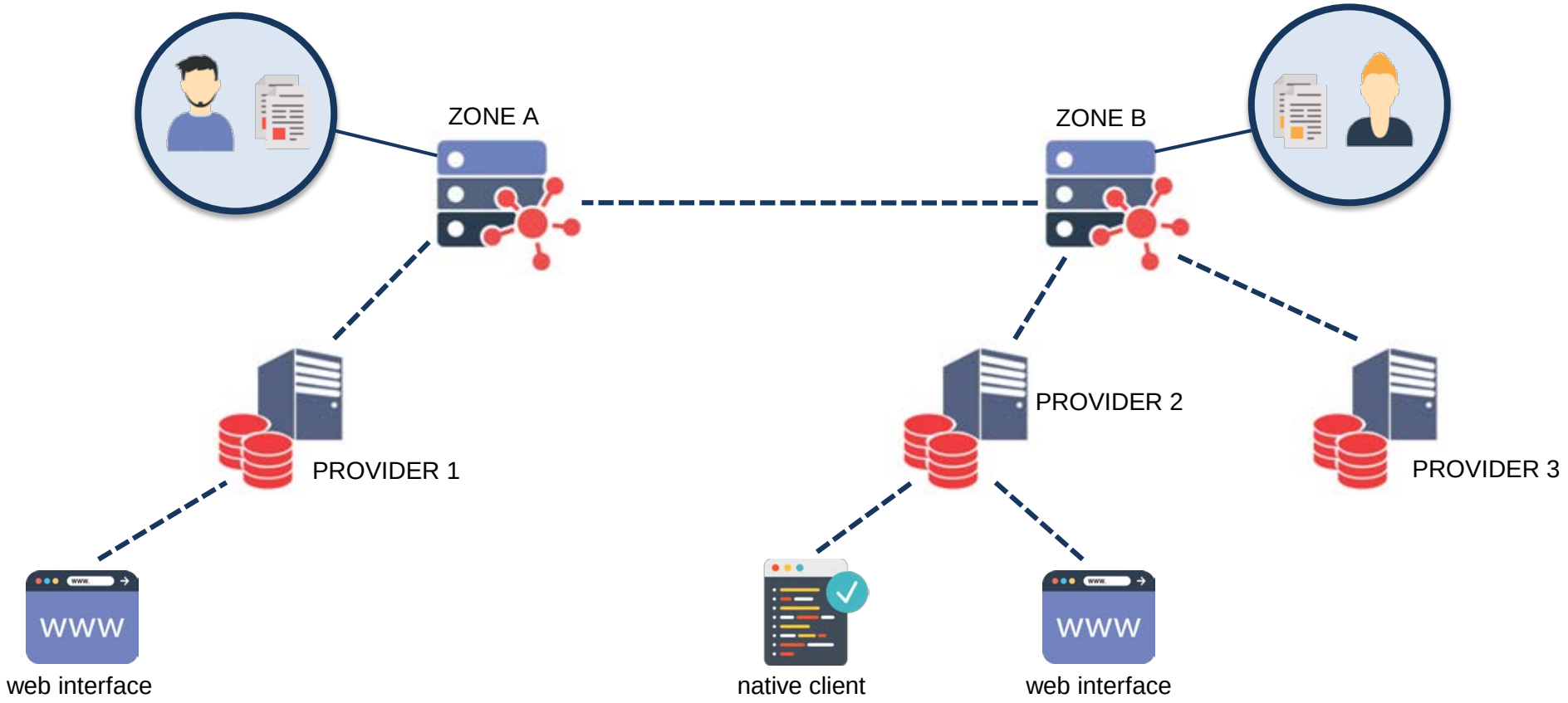
Metadata handled by the *Zone* of origin – local **AuthN & AuthZ** center



- No concurrent modification conflicts
- Metadata sync overheads evenly distributed

PROPOSED SOLUTION – METADATA SYNC

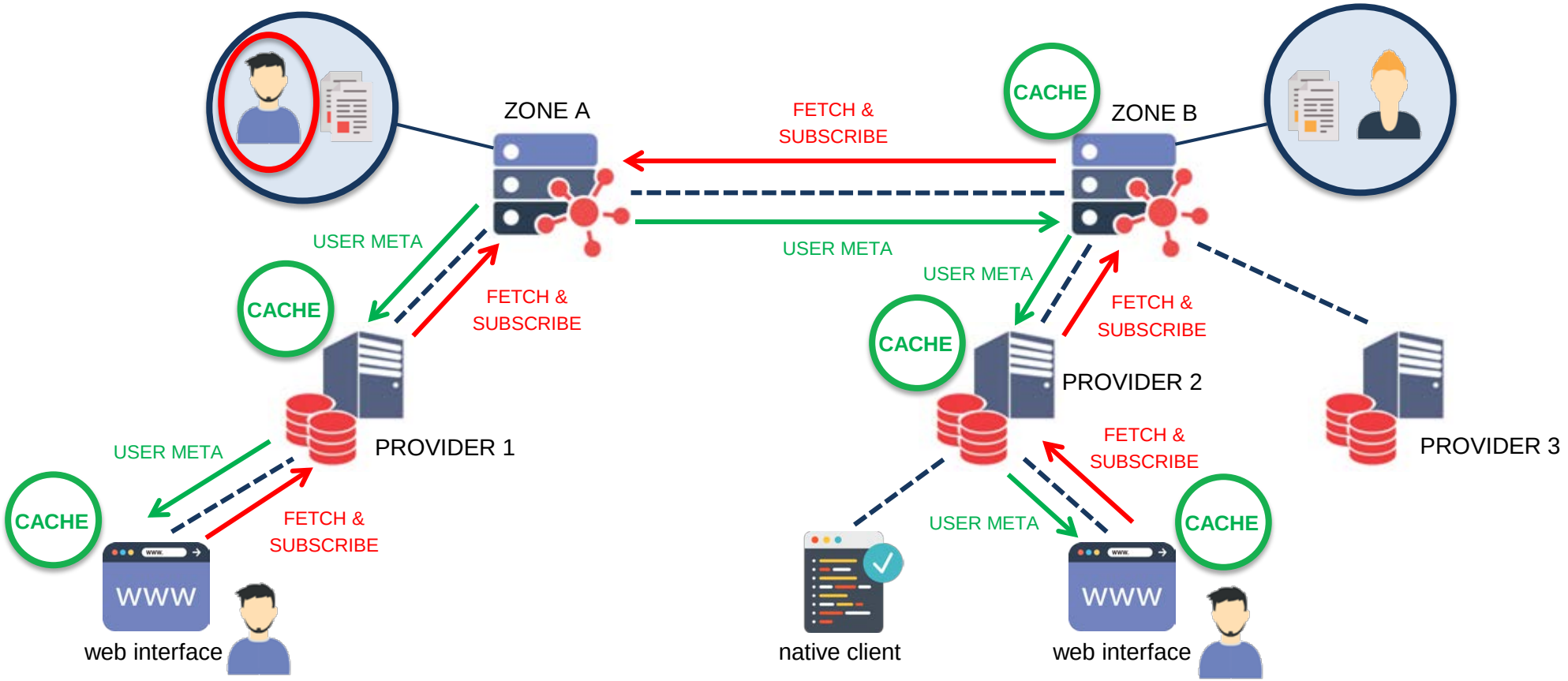
Multi-tier client-server architecture



PROPOSED SOLUTION – METADATA SYNC

Publish / subscribe and multi-tier caching

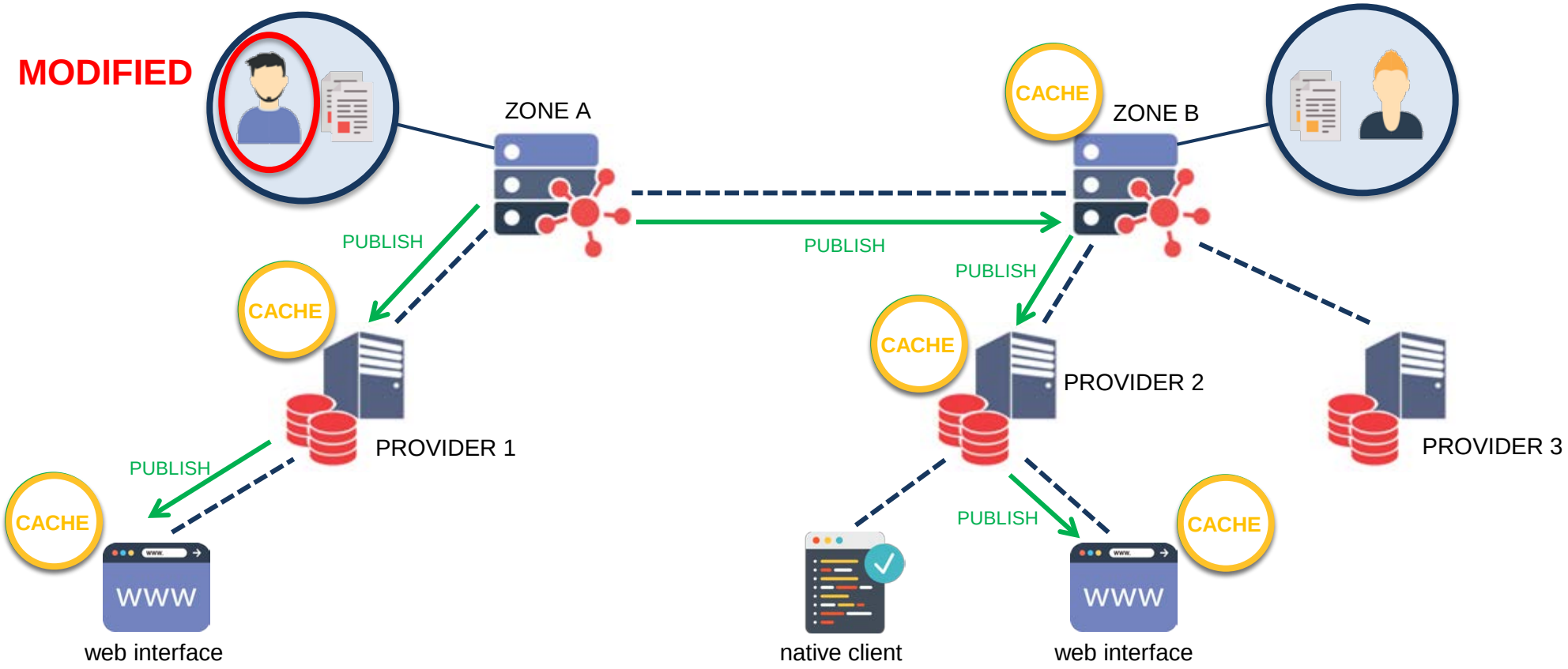
Vast performance improvement



PROPOSED SOLUTION – METADATA SYNC

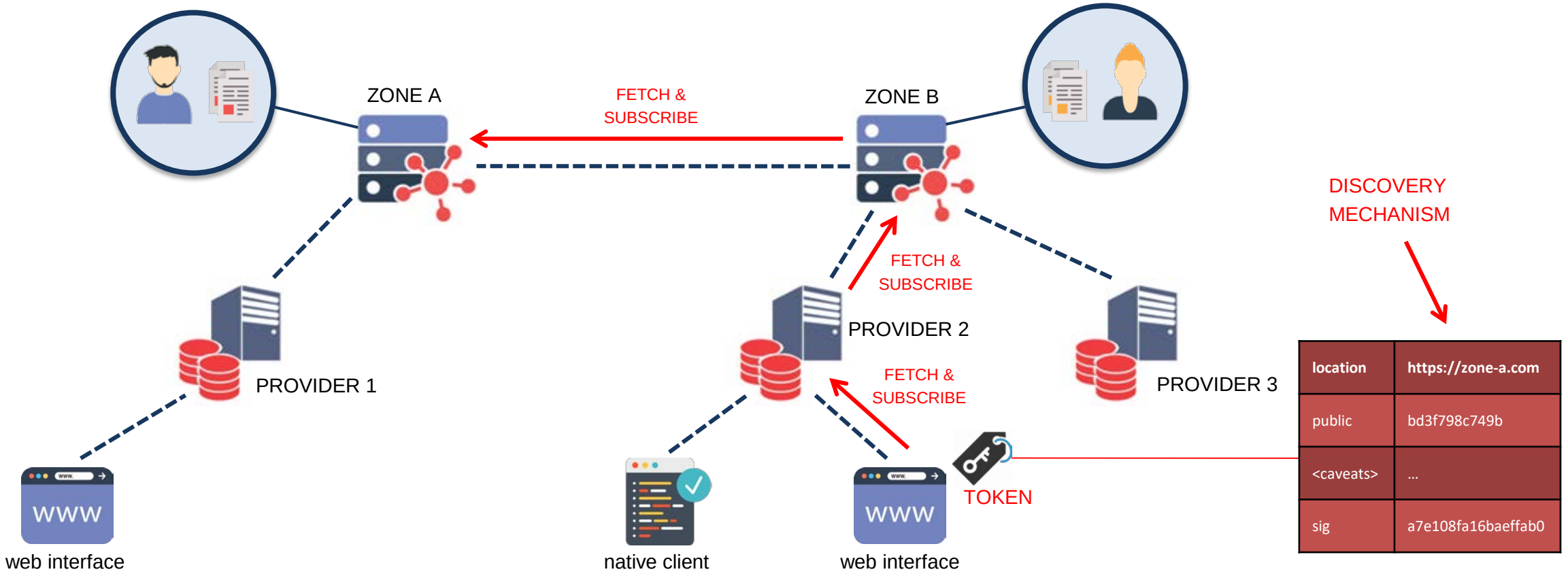
Publish / subscribe and multi-tier caching

Eventual consistency

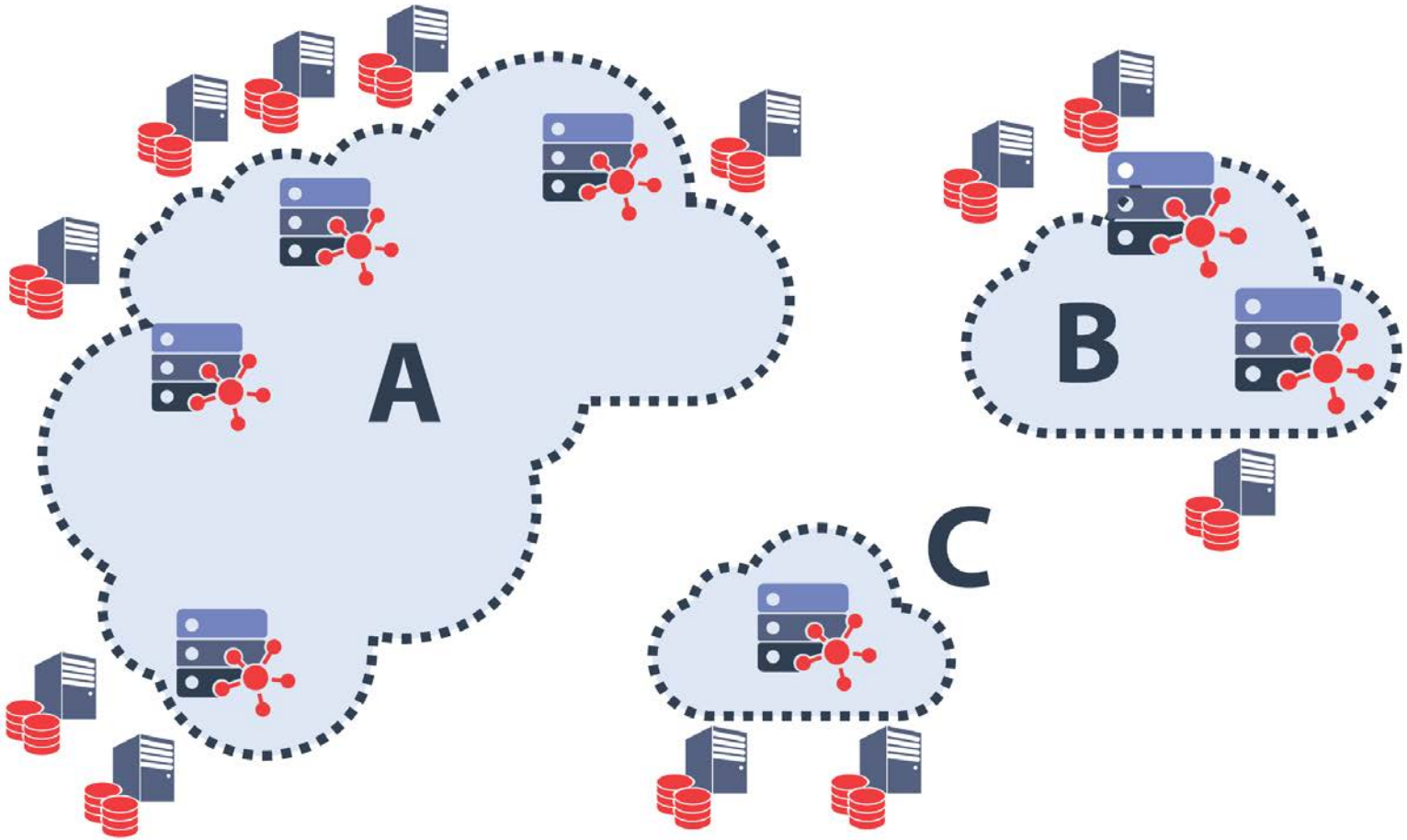


PROPOSED SOLUTION – AAI

Macaroon-based decentralized authorization



Proposed solution – zone mode

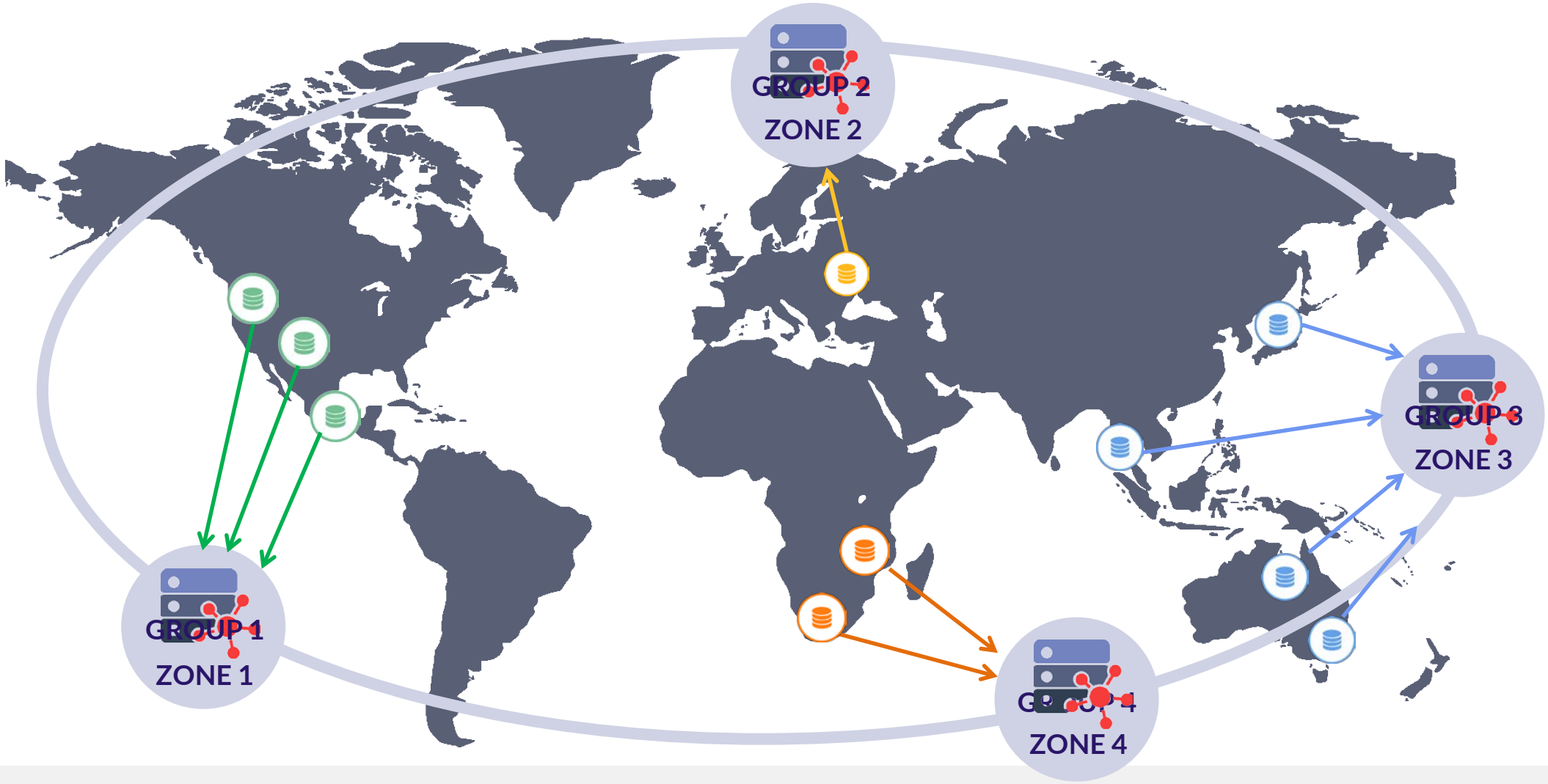


A) Open

B) Restricted

C) Isolated

METADATA SYNC PROTOCOL IN GLOBAL DATA ACCESS



CONCLUSIONS

- Global data access can be achieved by creating a decentralized network of data providers
- We propose a **metadata synchronization protocol** for such network, based on **hybdrid, multi-tier architecture** with **P2P backbone**
- Proposed concept is being implemented in **Onedata**, a distributed virtual file system

- ✔ Zone service acting as central authority and SP server
- ✔ Synchronization Protocol (SP) for single zone scope
- ✔ Macaroon based AAI for single zone scope
- ✔ Data provider service employing the SP
- 🚧 Cross Zone cooperation support
- 🚧 Sync protocol supporting global, cross zone scale

Onezone

Oneprovider

THANK YOU

https://onedata.org

The authors gratefully acknowledge the financial support of this work from AGH-UST