

OGSA-DAI

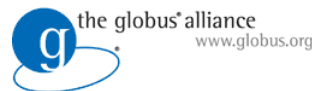
Sharing data resources to enable efficient collaboration

Mike Jackson

OGSA-DAI project team, EPCC, The University of Edinburgh

Databases in the Grid – tools, middleware and applications

OGF-23



Web: www.omii.ac.uk

Email: info@omii.ac.uk

Overview

- OGSA-DAI's vision
- What OGSA-DAI does – workflows
- OGSA-DQP
- OGSA-DAI and WS-DAI
- Pros and cons
- Summary

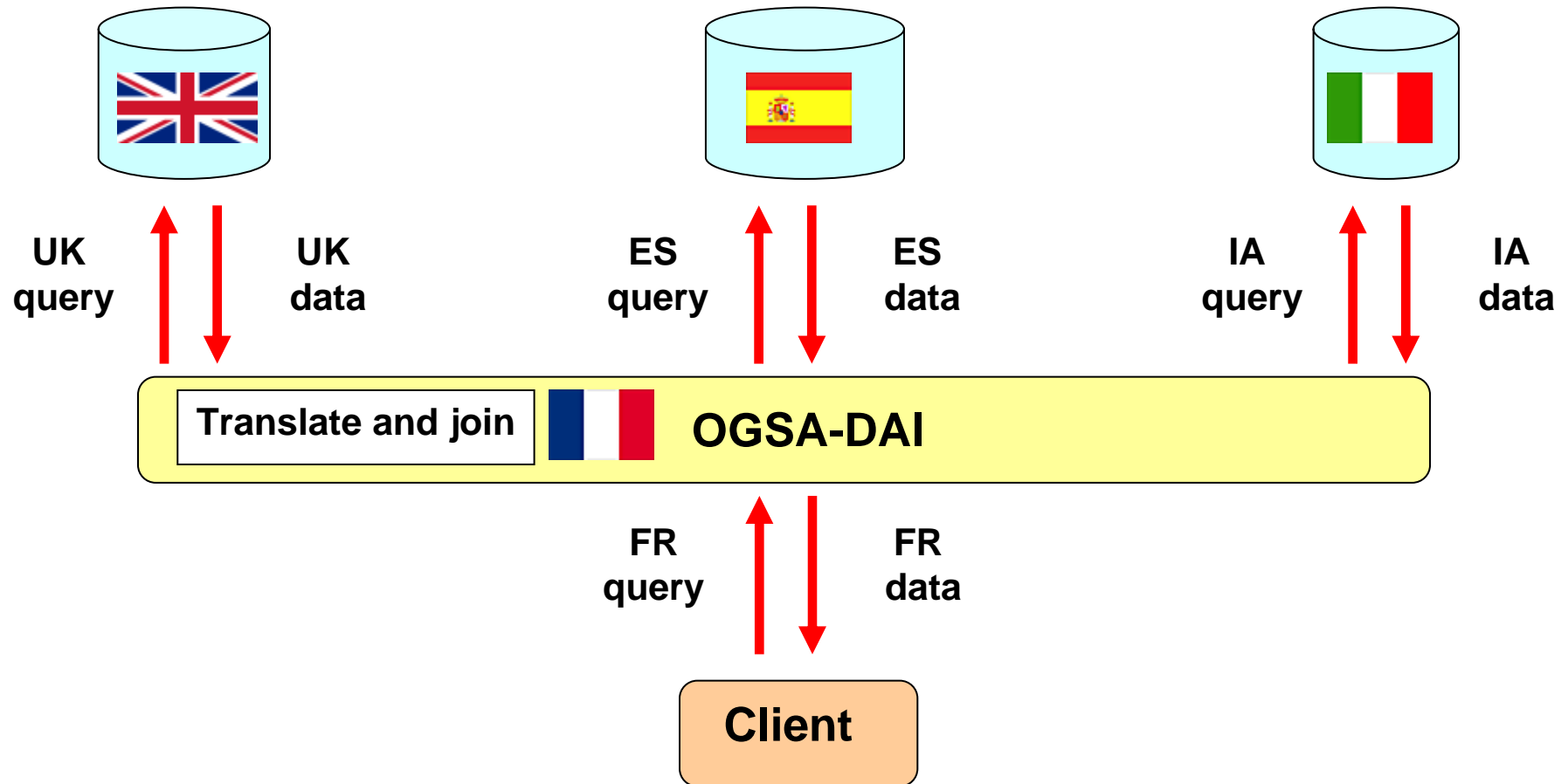
OGSA-DAI's vision

- Sharing data resources to enable **collaboration**
- Data **access**
 - Structured data in distributed heterogeneous data resources
- Data **transformation**
 - e.g. expose data in schema X to users as data in schema Y
- Data **integration**
 - e.g. expose multiple databases to users as a single virtual database
- Data **delivery**
 - To where it's needed by the most appropriate means
 - e.g. web service, e-mail, HTTP, FTP, GridFTP

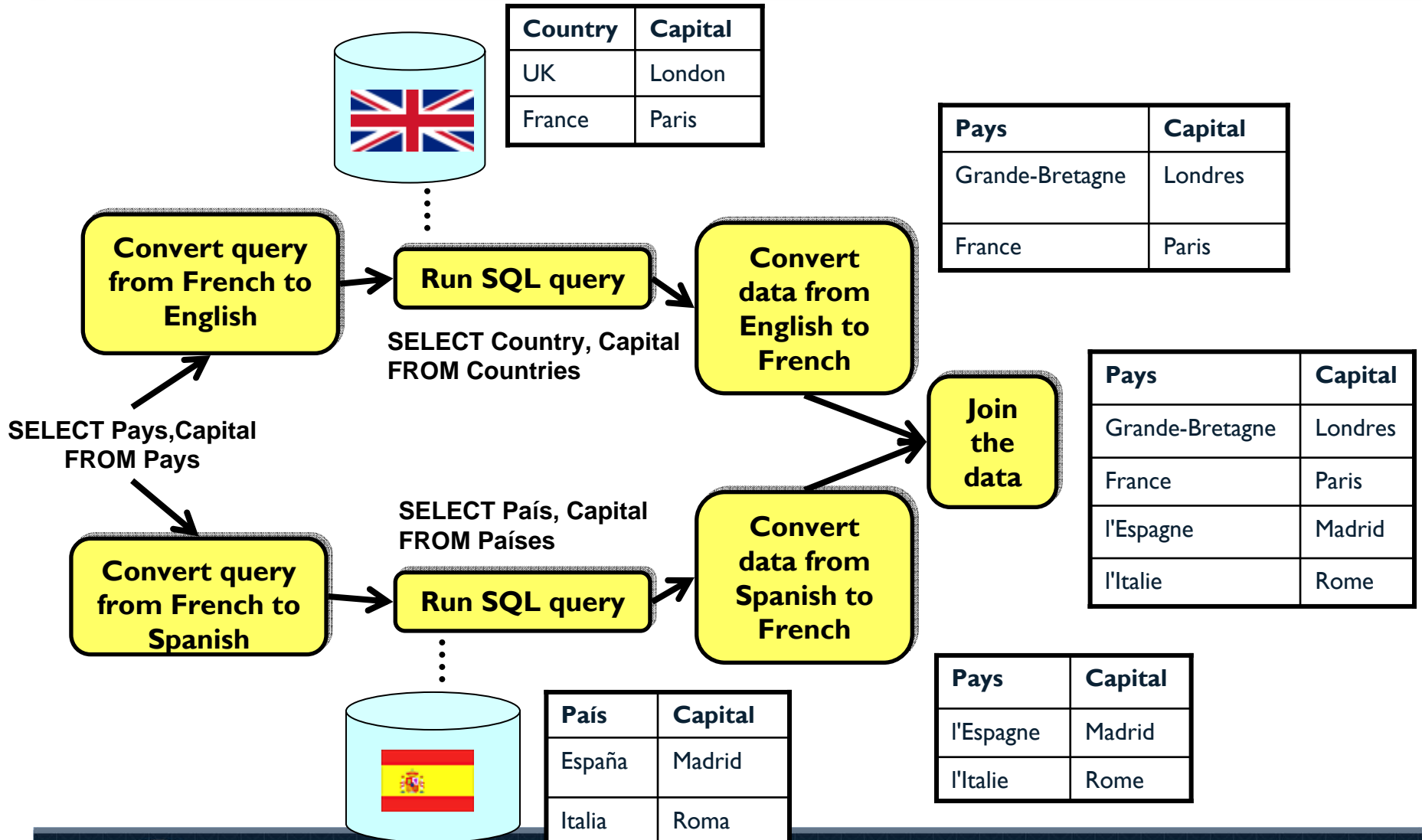
What OGSA-DAI does

- OGSA-DAI executes **workflows**
 - Equivalent to programs or scripts
- Workflows contain **activities**
 - Well-defined functional units
 - Data goes in, something is done, data comes out
 - Equivalent to programming language methods
- Workflows are submitted by clients
 - To an OGSA-DAI web service

Sharing distributed heterogeneous resources with OGSA-DAI



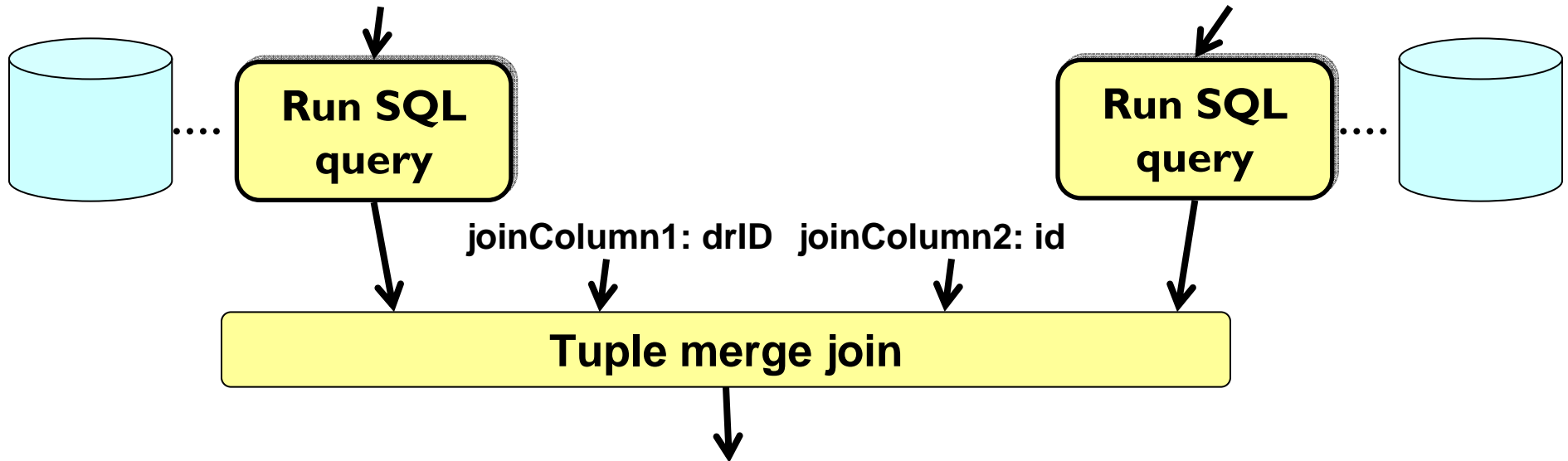
Data-centric workflows



Example – a join activity for VOTES

**SELECT drID, name, id FROM
patients ORDER by drID**

**SELECT id, drName FROM doctors
ORDER by id**

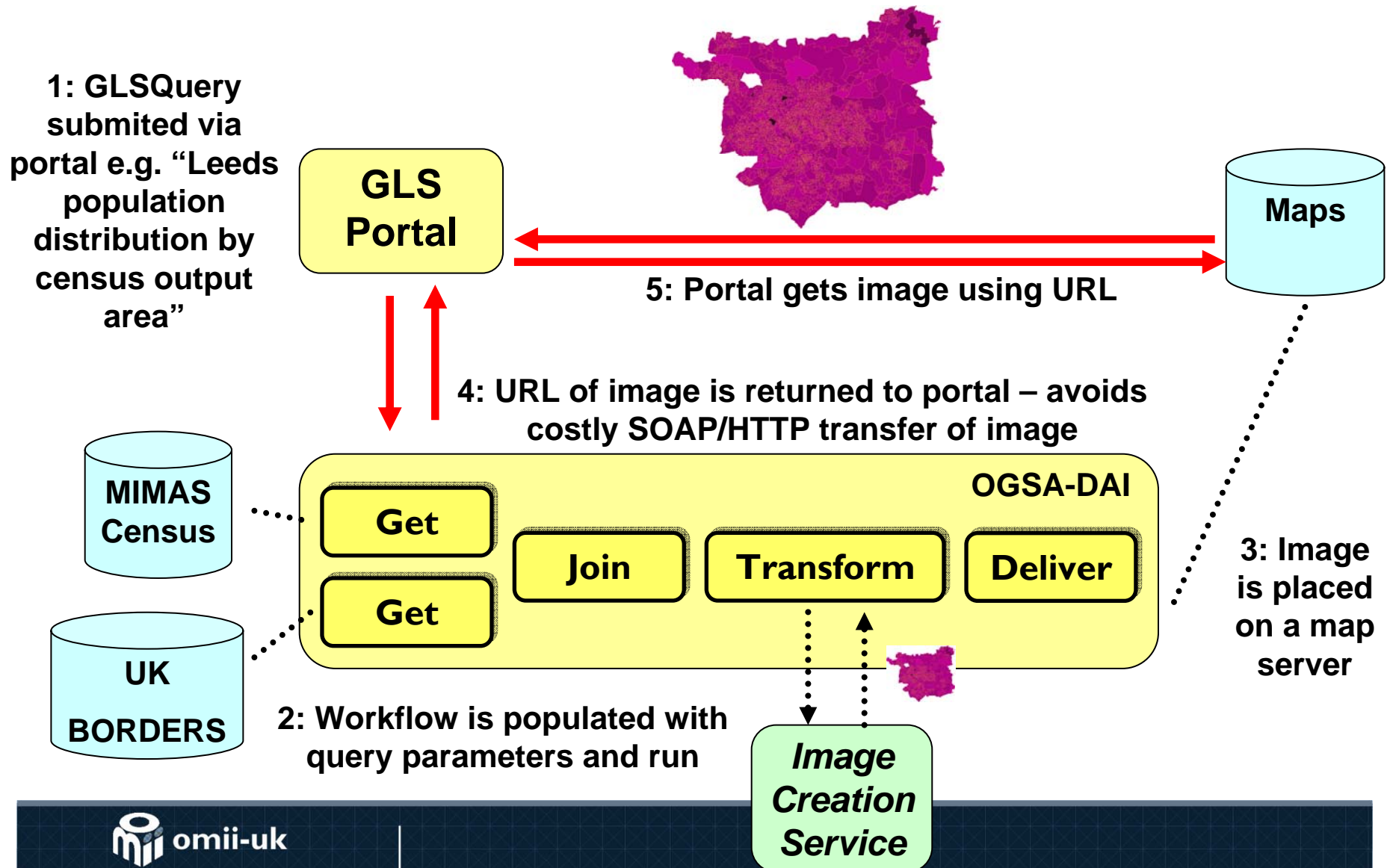


- Equivalent to

**SELECT id, name, drID, drName FROM patients, doctors WHERE
patients.drID = doctors.id;**

- patients and doctors are in two different databases

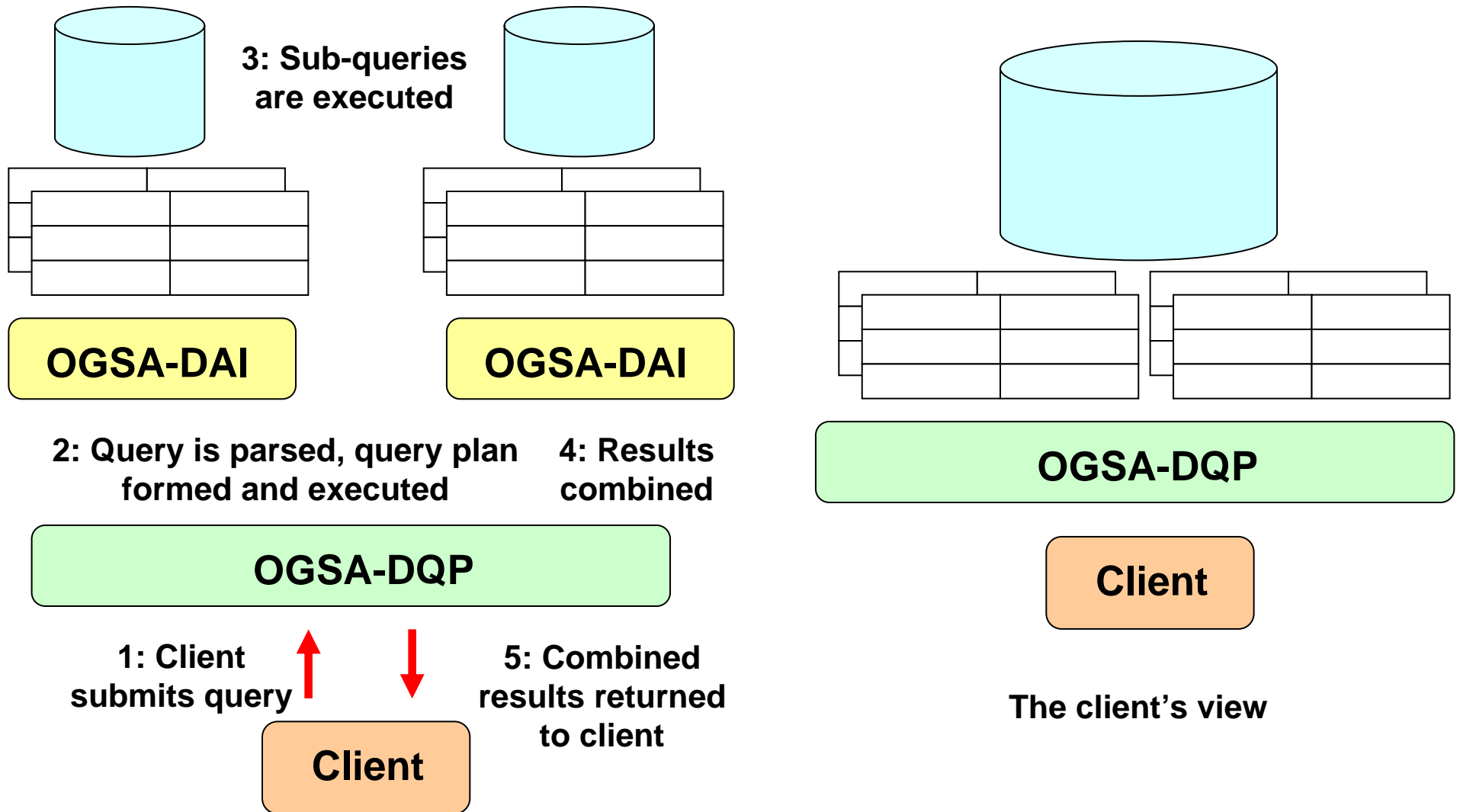
Example – SEE-GEO – geo-linking service portal



OGSA-DQP

- OGSA-DQP utilises OGSA-DAI
 - Developed by Universities of Manchester and Newcastle
 - Rewritten for OGSA-DAI 3.0 by EPCC as part of NextGrid
- Distributed query processing
 - Multiple tables on multiple databases are exposed to clients as multiple tables in one “virtual database”
 - Databases can be exposed
 - EITHER within one OGSA-DAI server
 - OR via multiple remote OGSA-DAI servers
- Clients are unaware of the multiple databases

OGSA-DQP – what it does and the client's view



OGSA-DAI and WS-DAI

- OGF WS-DAI specifications
 - Web services for data manipulation
 - WS-DAIR and WS-DAIX
- OGSA-DAI product web services
 - execute(workflow)
- WS-DAI specification web services
 - GetDocuments(“bangles”, “madness”, “jam”)
 - XPathExecute(“/bands/pop/@name=‘bangles’”)
 - GetItems(1,3)
- OGSA-DAI WS-DAIX 1.0
- Façade – makes OGSA-DAI more usable

OGSA-DAI pros

- **Efficient**
 - Workflows encapsulate many operations within a single web service invocation
- **Data providers retain control of data**
- **Security layer**
 - Federation, data resource, table, document, file, row, line
- **Data transformation and federation layer**
- **Lightweight clients**

OGSA-DAI pros

- **Extensible**
 - Data resources, activities, security, presentation layers
- **Enabling framework**
 - Save development time – focus on your application
- **100% Java open source freeware**
 - Runs under Windows, UNIX, Linux
- **Compliant with**
 - Globus Toolkit 4.0.x
 - Apache Axis/Tomcat
 - OMII 3.4.0
 - UNICORE – by OMII-Europe
 - VOMS – by OMII-Europe



OGSA-DAI cons

- Another “thing” between you and the data
- Performance implications
 - Is access speed paramount?
 - Are the benefits on the previous slides more important?
- OGSA-DAI is not a silver bullet
 - Do the strengths outweigh the weaknesses for your specific requirements?

Conclusion

- **OGSA-DAI**
 - Sharing data resources to enable collaboration
 - Workflows to access, transform, filter, integrate and deliver data
 - Distributed query processing
 - Facades can make workflows more usable
 - Highly customisable enabling environment
- **More information**
 - “Accessing and integrating structured data using OGSA-DAI 3.0 and links with WS-DAI” – Friday, 09:00, Sarria
 - <http://www.ogsadai.org.uk>

Releases roadmap

- OGSA-DAI 3.0 extension packs – out now
 - Inter-OGSA-DAI data delivery activities
 - Dynamic data resource plug-in deployment
 - Servlet
 - VOTES join activity and document-based client
- OGSA-DAI WS-DAIX – out now
- OMII-Europe OGSA-DAI 3.0 components – out now
 - Plug-ins for VOMS authorization
 - OGSA-DAI 3.0 UNICORE presentation layer
- OGSA-DQP 3.2 – May 2008 in conjunction with NextGrid
- SQL views – June 2008 in conjunction with BEinGRID
- OGSA-DAI 3.1 – September 2008
 - Integration of extension packs
 - Bug fixes and new features
- OGSA-DAI WS-DAIR – September 2008

