

OGSA-DAI

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

OGF-23



Web: www.omii.ac.uk

Email: info@omii.ac.uk

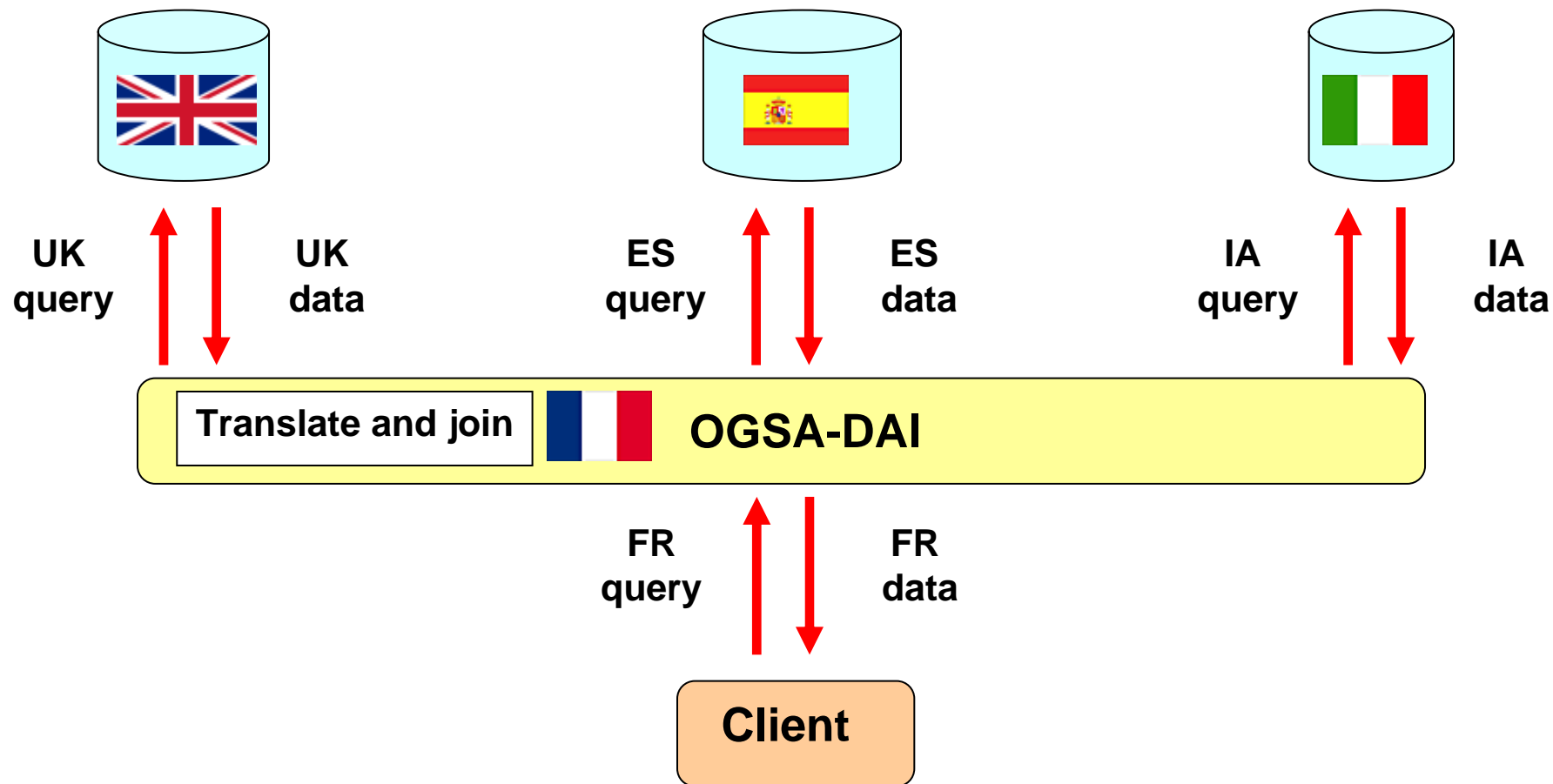
OGSA-DAI's vision

- 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

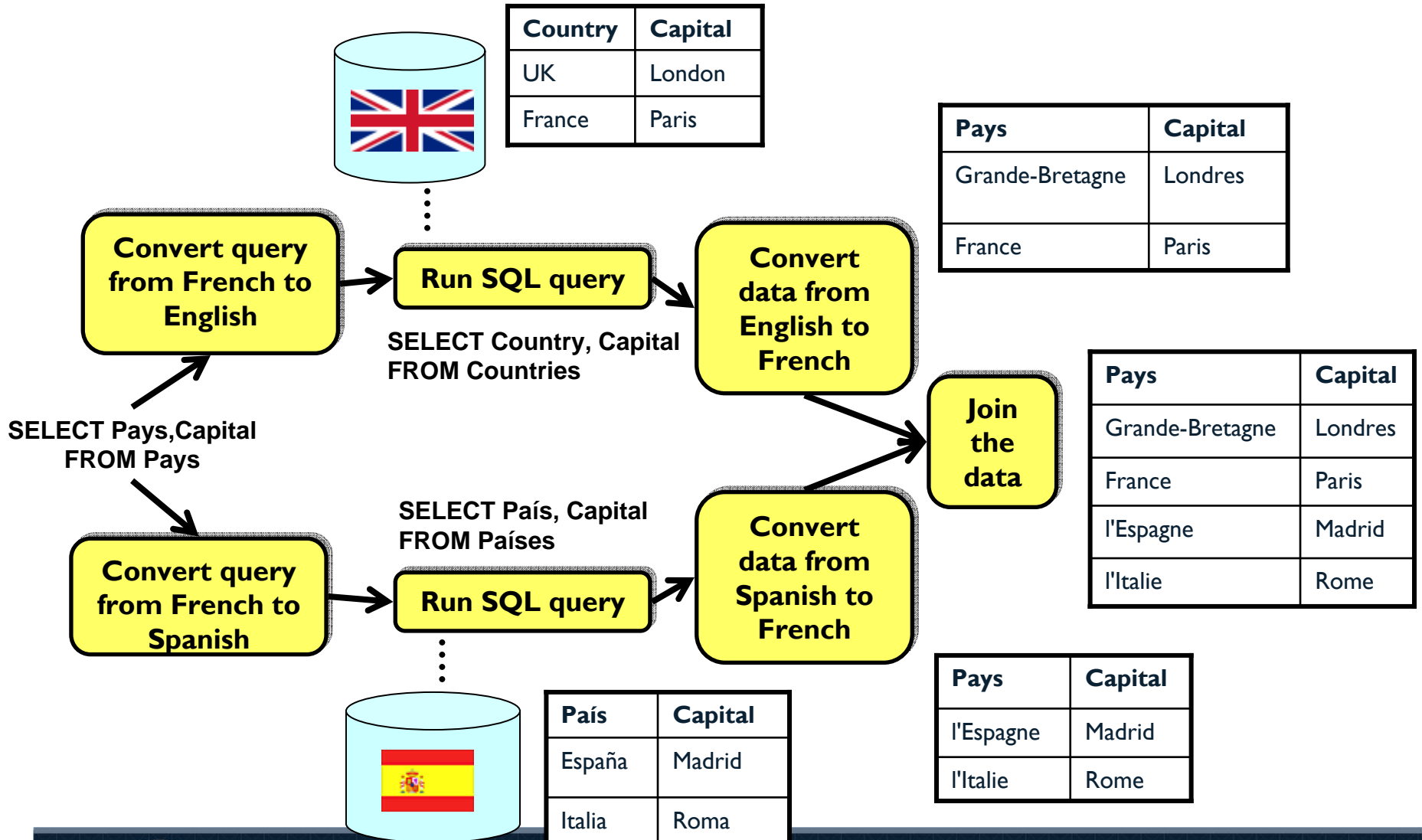
What OGSA-DAI does

- 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



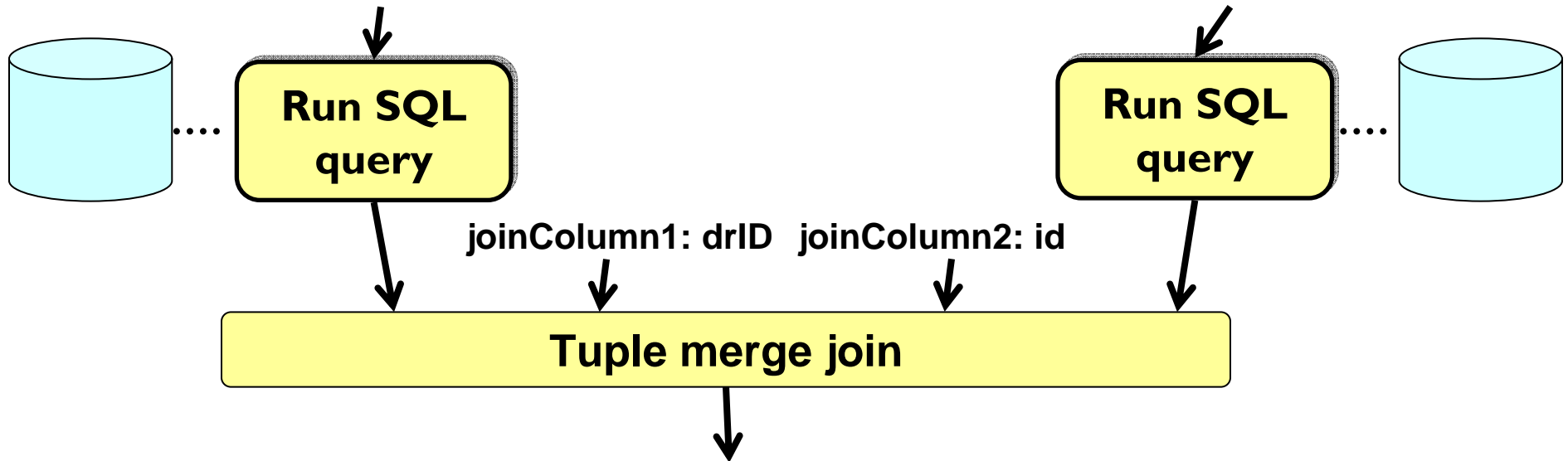
Data-centric workflows



Example – a join activity for VOTES

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

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

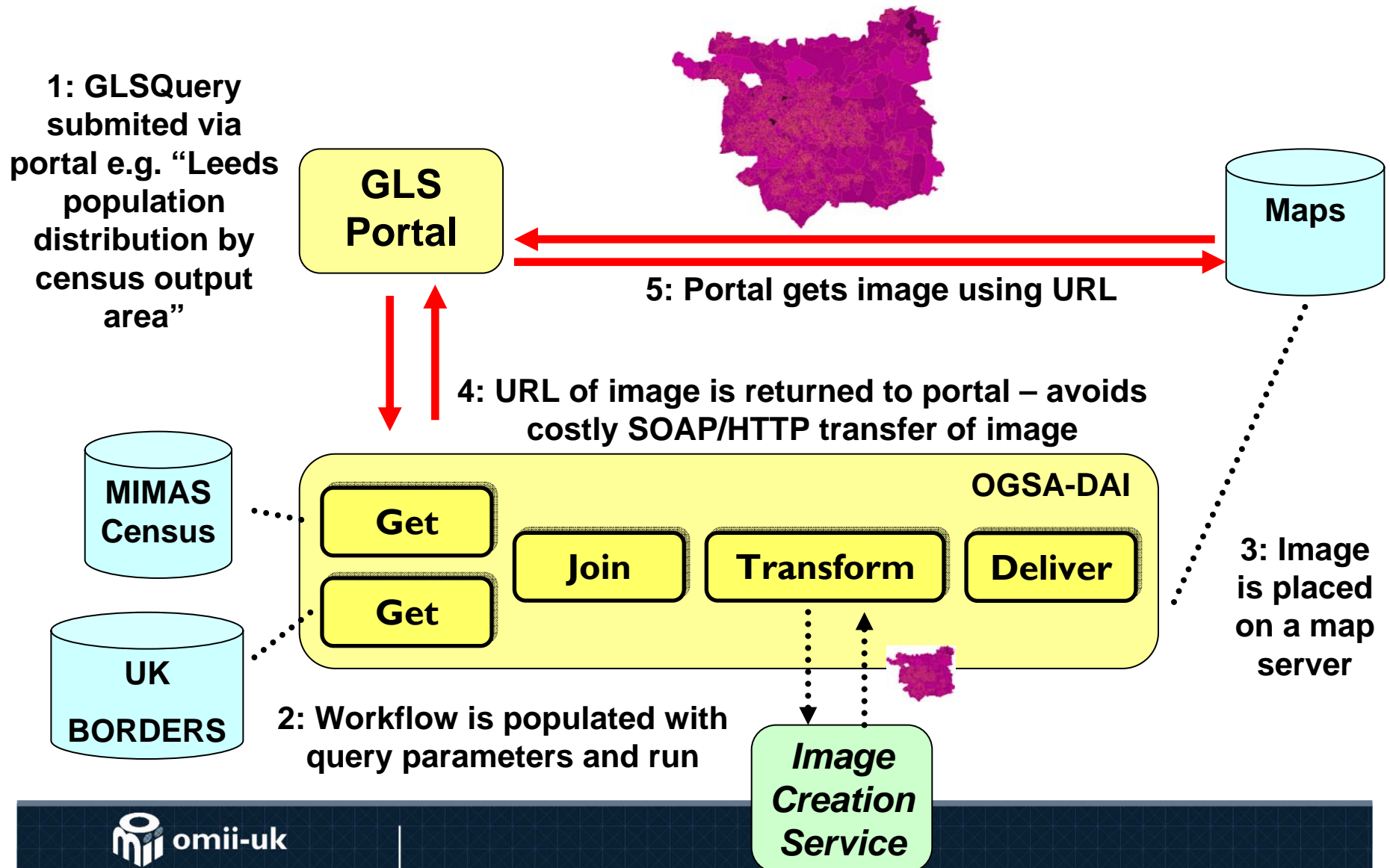


- Equivalent to

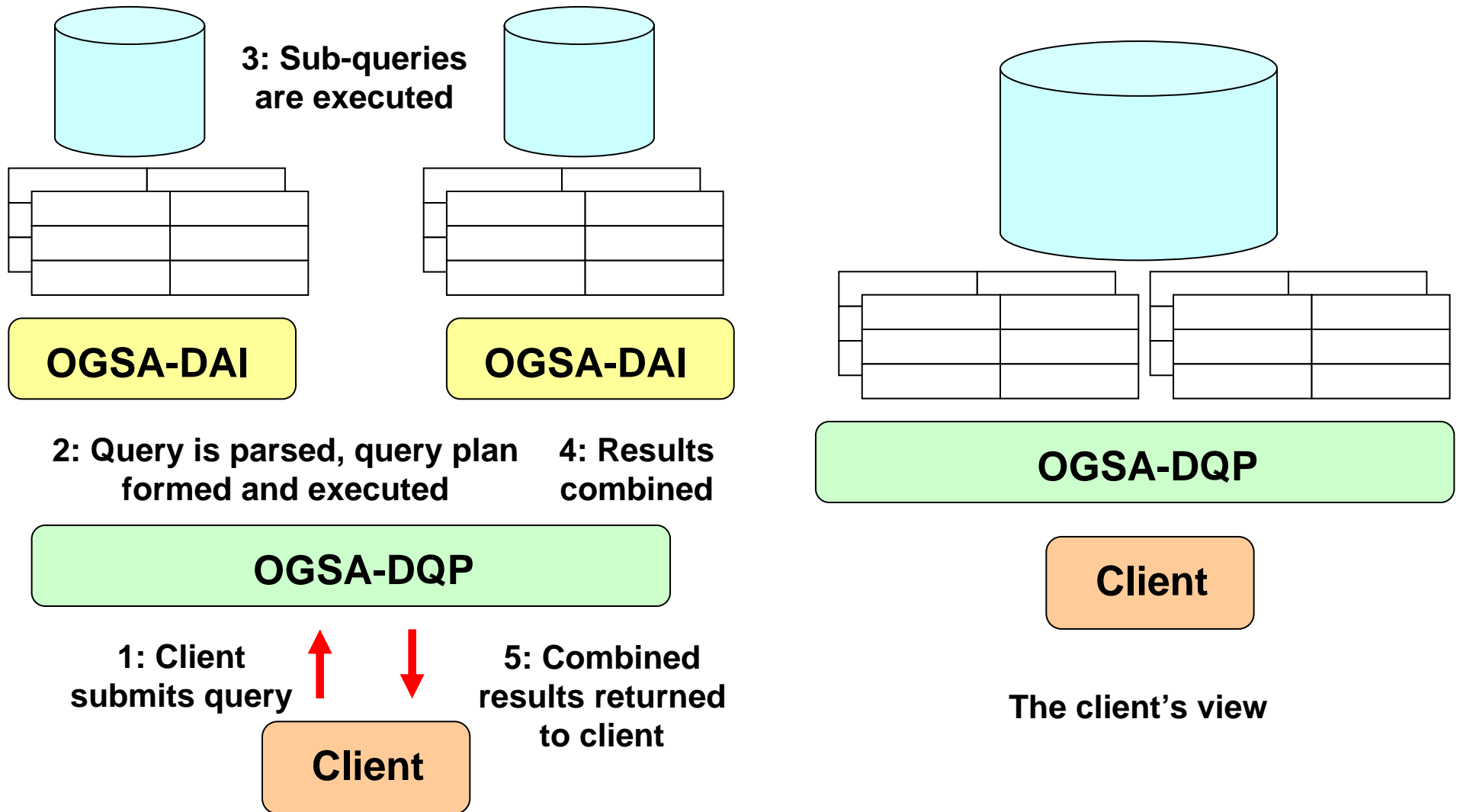
**SELECT id, name, drID, drName FROM patients, doctors WHERE
patients.drID = doctors.id;**

- patients and doctors are in two different databases

Example – SEE-GEO – geo-linking service portal



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



OGSA-DAI workflows pros

- **Efficient**
 - Workflows encapsulate many operations within a single web service invocation
- **Lightweight clients**
- **Opaque and transparent data federations**
- **Enabling framework**
 - Save development time – focus on your application
- **Fits into WSRF resource-centric view**

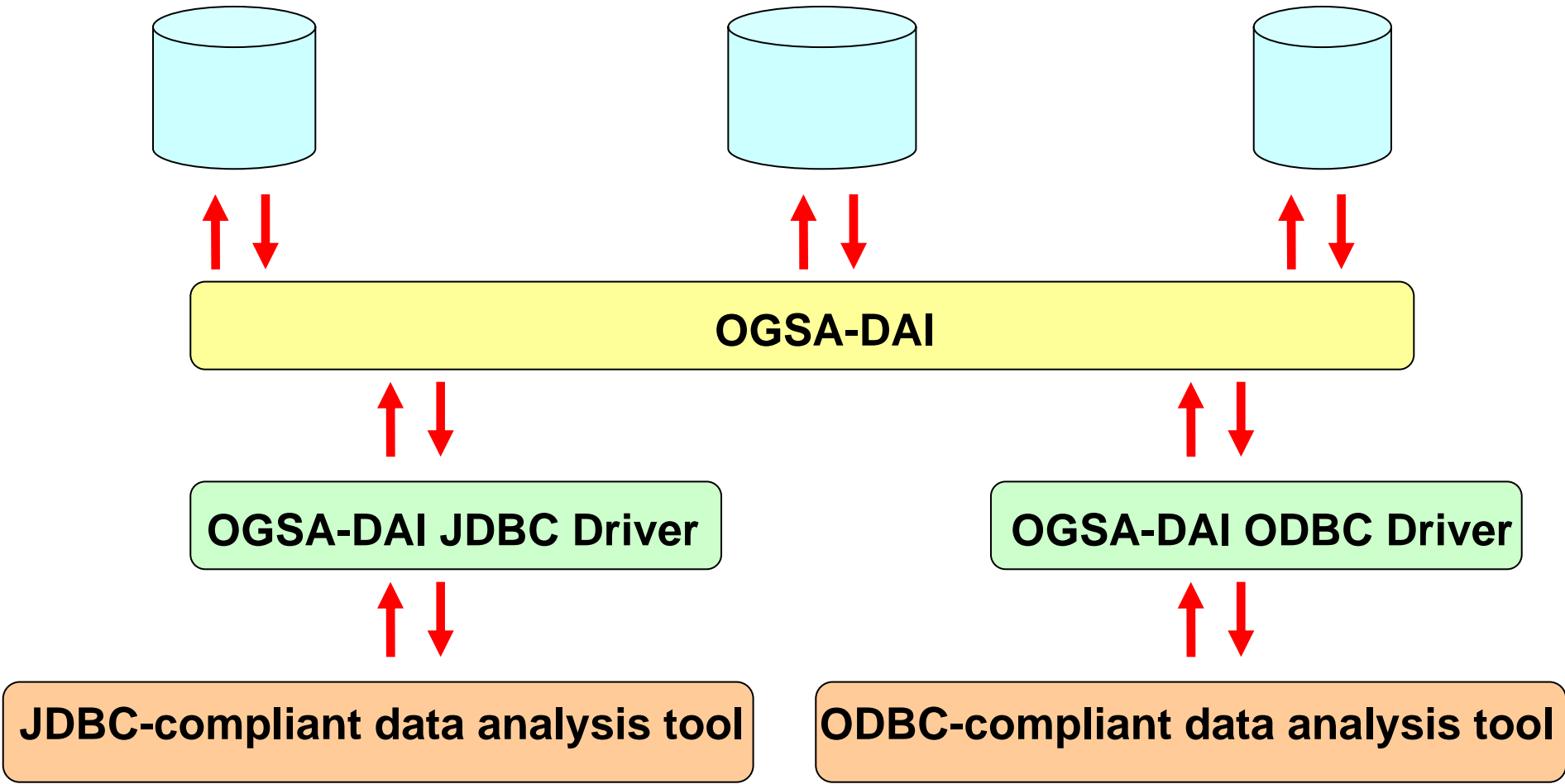
OGSA-DAI workflows cons

- Too expressive
 - Semantics must be inferred from names of activities
 - Need to query the server
 - Problems using OGSA-DAI services in workflows
 - e.g. Taverna
- A de-facto open standard only
- Incompatible with certain data analysis tools

Facades

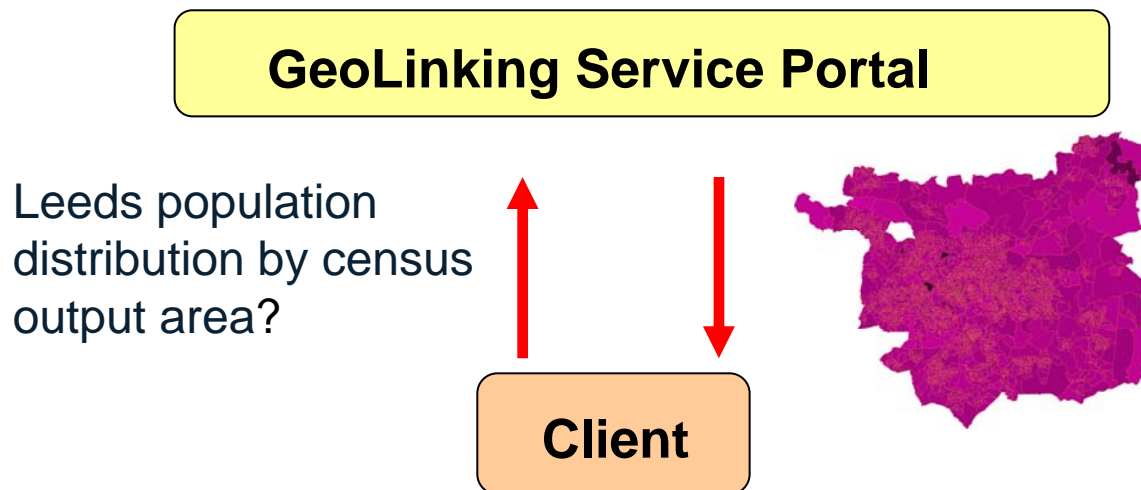
- Define facades on top of OGSA-DAI
- Why?
 - Provide interfaces with more tightly-defined semantics
 - Comply with standards
 - Exploit existing data analysis tools
- Continue to exploit the power of workflows under-the-hood
 - “Canned workflows”
 - Templates selected and populated, executed and parsed

JDBC and ODBC-compliant facades



SEE-GEO and GEOGrid – standards and facades

- SEE-GEO and GEOGrid's services
- OGC GeoLinking Services
 - Open Geospatial Consortium specifications
 - Specifications being hardened into standards



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
 - Implements WS-DAIX operations using workflows

OGSA-DAI and standards

- Pros
 - Access workflow power via well-defined interfaces
 - Facilitate inter-operability with other components
- Cons
 - Evolution of specifications into standards
 - Shifting sands – WS-DAI
 - Pulled rugs – OGSi
 - Development overhead
 - Relying on 3rd parties to implement standards e.g. GT and WSRF
 - Compliance versus core functionality
- Moral
 - Lightweight presentation layers
 - APIs and interfaces everywhere
 - Don't force but don't prohibit

Conclusion

- **OGSA-DAI**
 - Sharing data resources to enable collaboration
 - Workflows to access, transform, integrate or deliver
 - Facades and standards contribute to inter-operability
 - Don't force but don't prohibit
- Further information
 - <http://www.ogsadai.org.uk>