Wide Field Astronomy Unit (WFAU)

Virtual Observatory
Data Access service
Target use cases:

- JOIN queries combining data in from catalogs within ROE
- JOIN queries combining data from external TAP services
- Storage for user data, including uploaded data and query results
Virtual Observatory
Data Access service

Implemented using

- ADQL parser from CDS

- SQL Server cluster at ROE

- OGSA-DAI DQP service from EPCC
Virtual Observatory
Data Access service

ADQL library from CDS

- ADQL syntax checking
- Data structure validation
- SQL dialect translation

What is ADQL?
ADQL is an SQL-like language which includes astronomical facilities to query a database. This language has been defined by the IVOA in the Recommendation of 30 Oct 2006 (Version 2.0) and is mainly used in the Table Access Protocol (TAP).

Why this library?
In order to help Java developers to parse, manipulate and translate ADQL queries quickly and with as few lines of code as possible.

Functionnalites:
- Parse: read ADQL queries in text and transform them into a Java object (actually, a syntactic tree).
- Manipulate: the generated object can be manipulated so than modifying the original query.
- Translate: an interface and some implementations lets translating SQL into other languages like SQL.

How to use it?
- Getting started: to start with this library.
- Documentation: to have more details about all provided functionalities.
- JavaDoc: Java documentation of all available classes.
- Link: What's new?: Last modifications of the library.

This library is free: you can redistribute it and/or modify it under the terms of the GNU Lesser General Public License.

Extensible architecture makes it easy to customize.

D.Morris
Institute for Astronomy,
Edinburgh University
May 2013
SQL Server cluster at ROE

- Supports cross catalog queries within the system
- All the tables are accessible from a single namespace

```
SELECT
catalog1.schema1.table1.column1,
catalog2.schema2.table2.column2,
FROM
catalog1.dbo.table1
JOIN
catalog2.dbo.table2
ON
 ....
WHERE
 ....
```
Virtual Observatory
Data Access service

OGSA-DAI service from EPCC

- Middleware services enabling federation of heterogeneous data resources.
- Used in a wide range of applications including medical research, geographical information systems, meteorology, transport, computer-aided design, engineering and astronomy.
Virtual Observatory
Data Access service

Distributed Query Processor (DQP)

- Splits a JOIN query into separate sub-queries for each resource
- Combines the results from each sub-query to recreate the original JOIN

```
SELECT resource1.table1.column1,
     resource2.table2.column2,
FROM resource1.table1
JOIN resource2.table2
....
```

```
SELECT table1.column1
FROM table1
WHERE ....
```

```
SELECT table2.column2
FROM table2
WHERE ....
```
Virtual Observatory
Data Access service

IVOA TAP services

- Provides a common abstraction for databases
- Hides the implementation details
- Services from different data providers behave the same way

OGSA-DAI astronomy extensions

- Developed for ROE by EPCC
- Enabling OGSA-DAI to use IVOA data services
Virtual Observatory
Data Access service

OGSA-DAI TAP Factory

- Prototype federated TAP service
- Developed for ROE by EPCC

- Using OGSA-DAI DQP to combine data from external TAP services

- Using AstroGrid DSA to provide TAP interface

SELECT
  service1.table1.column1,
  service2.table2.column2,
FROM
  service1.table1
JOIN
  service2.table2

D.Morris
Institute for Astronomy,
Edinburgh University
May 2013
Virtual Observatory
Data Access service

New data access service

- New metadata service to create a virtual 'data space'
- Combining local and remote data
- Local JOINS executed within SQL Server
- Remote JOINS processed by DQP

```
SELECT
    resource1.table1.column1,
    resource2.table2.column2,
FROM
    resource1.table1
JOIN
    resource2.table2
....
```
Virtual Observatory
Data Access service

New data access service

- User data appears within the same virtual space
- Results from ADQL queries automatically stored in users space
- Available for query in combination with local catalogs and remote TAP services

```
SELECT resource.table1.column1, userdata.table2.column2,
FROM resource.table1
JOIN userdata.table2
```
Virtual Observatory
Data Access service

Target use cases:

- JOIN queries combining data from catalogs within ROE - processed locally within SQL Server

- JOIN queries combining data from external TAP services - processed using OGSA-DAI DQP

- Storage for user data, including uploaded data and query results - stored as database tables, available to be used in queries