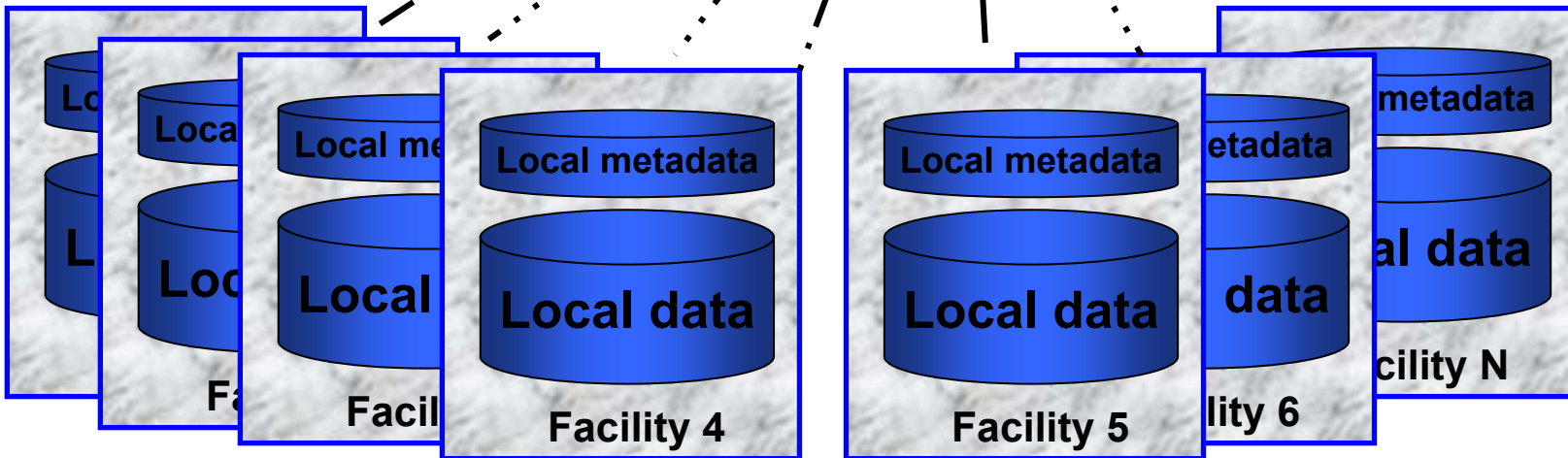
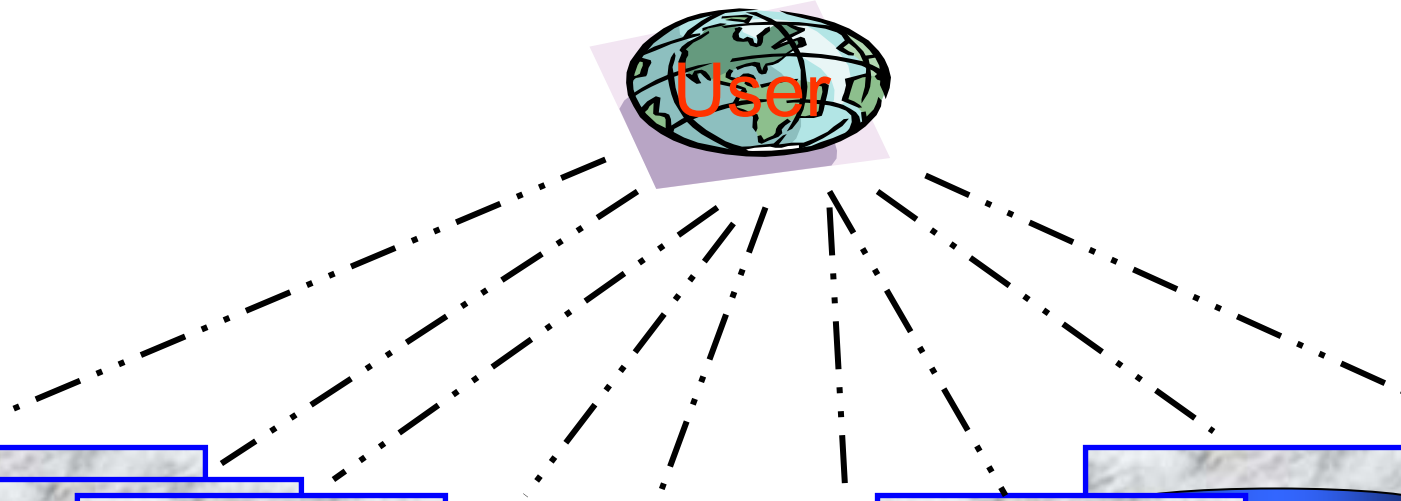
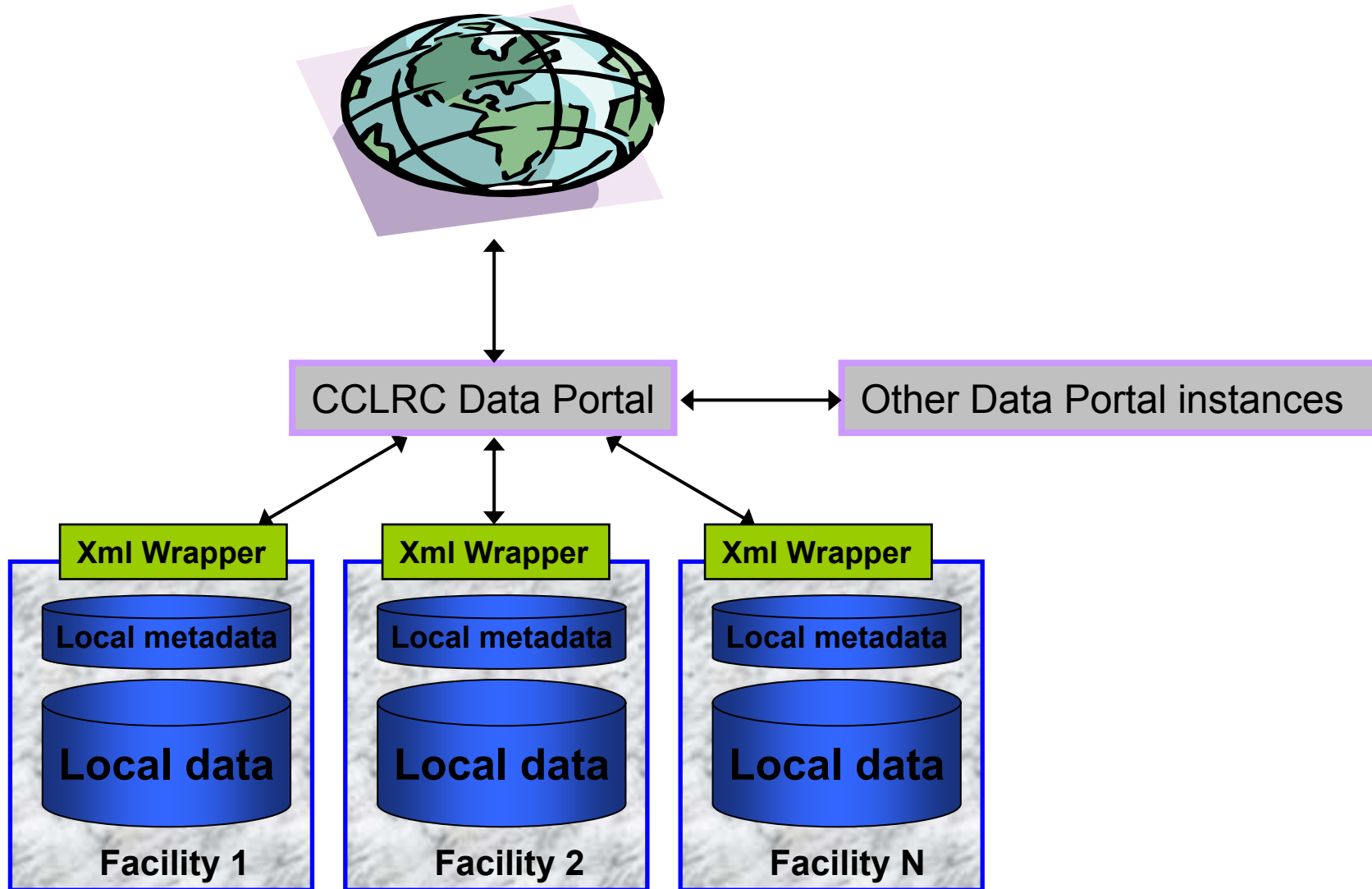


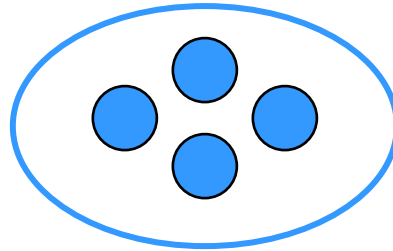
The Data Portal

Glen Drinkwater

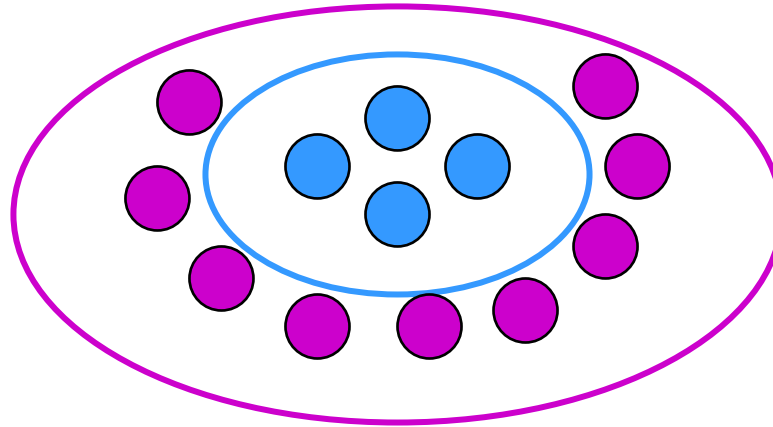
- Council for the Central Laboratory of the Research Councils' has many large scale facilities.
- E.g. ISIS, a pulsed neutron source and Diamond, the UK's new synchrotron light source.
- Data Portal
 - Making existing scientific data resources accessible through a single interface.
 - Acting as a broker between scientists, facilities and data.
- Benefits
 - Repetition of experiments can be avoided.
 - Collaborations can be built by identifying that someone else is working in a similar area.
 - Data about a related material can be found and used to aid a new analysis.
 - Data can be reanalysed when better analysis tools becomes available



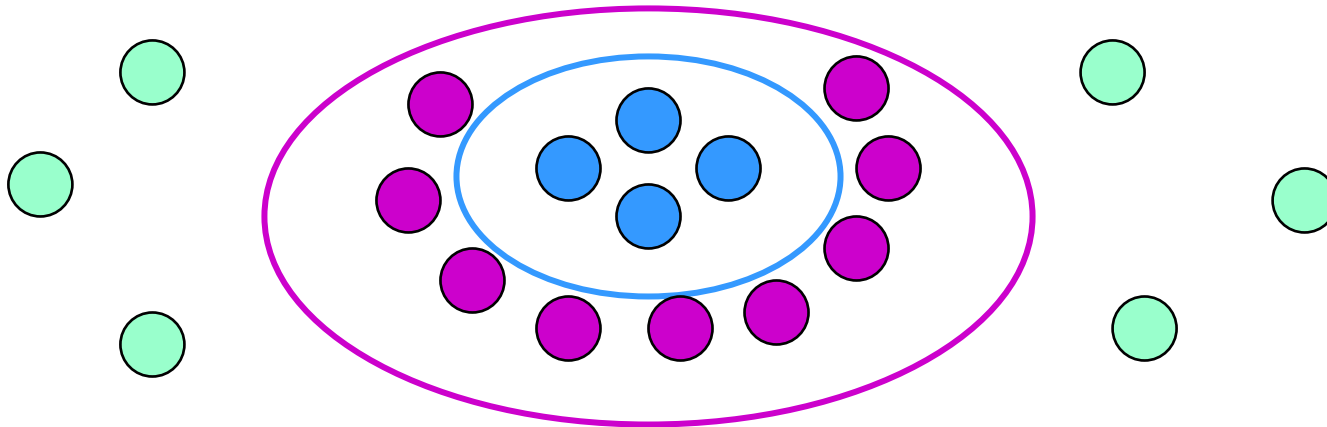




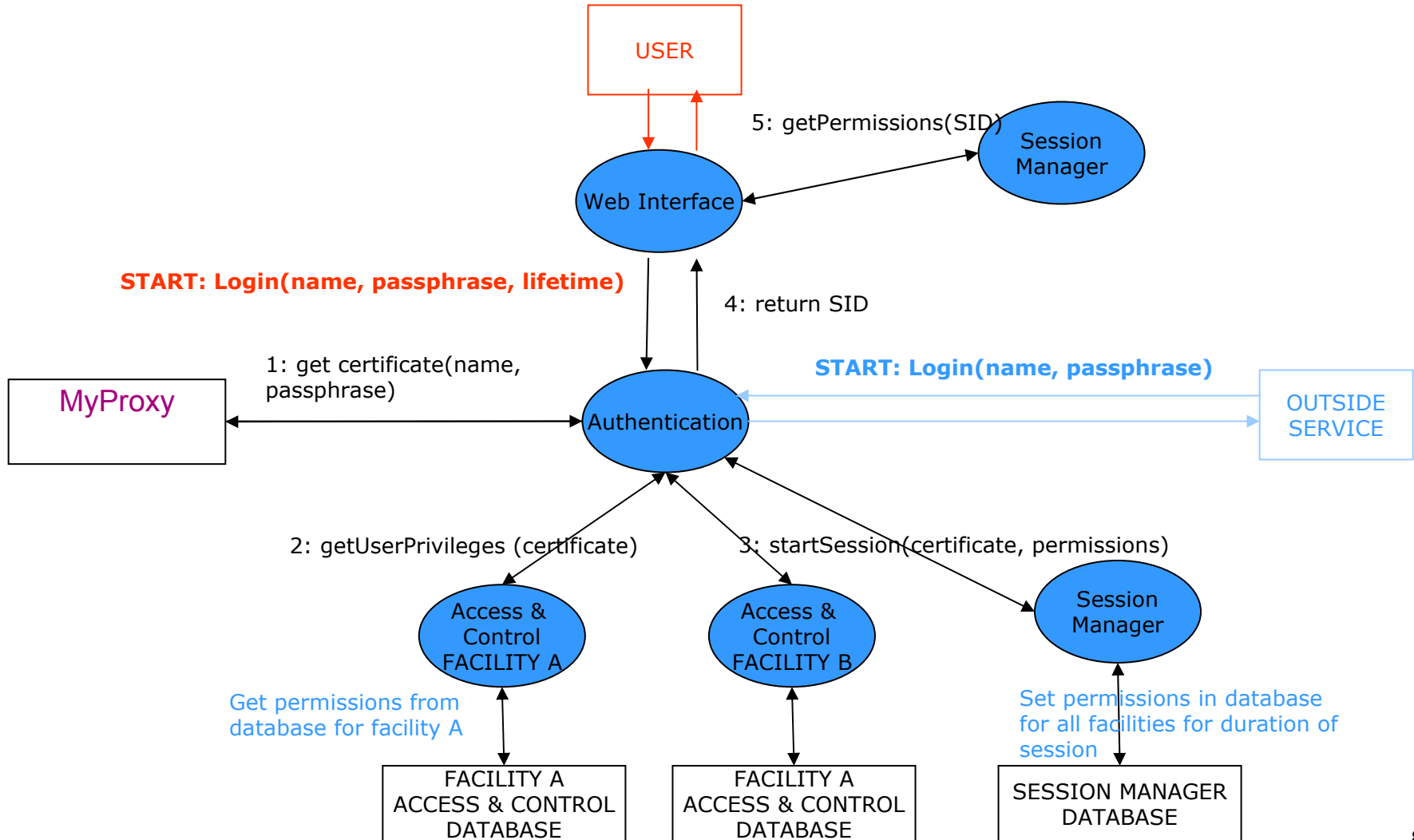
- Web Interface, Query and Reply, Lookup and Help.
- Important function grouped into modules, each modules with a web service interface, interface description in WSDL and communicate via SOAP.

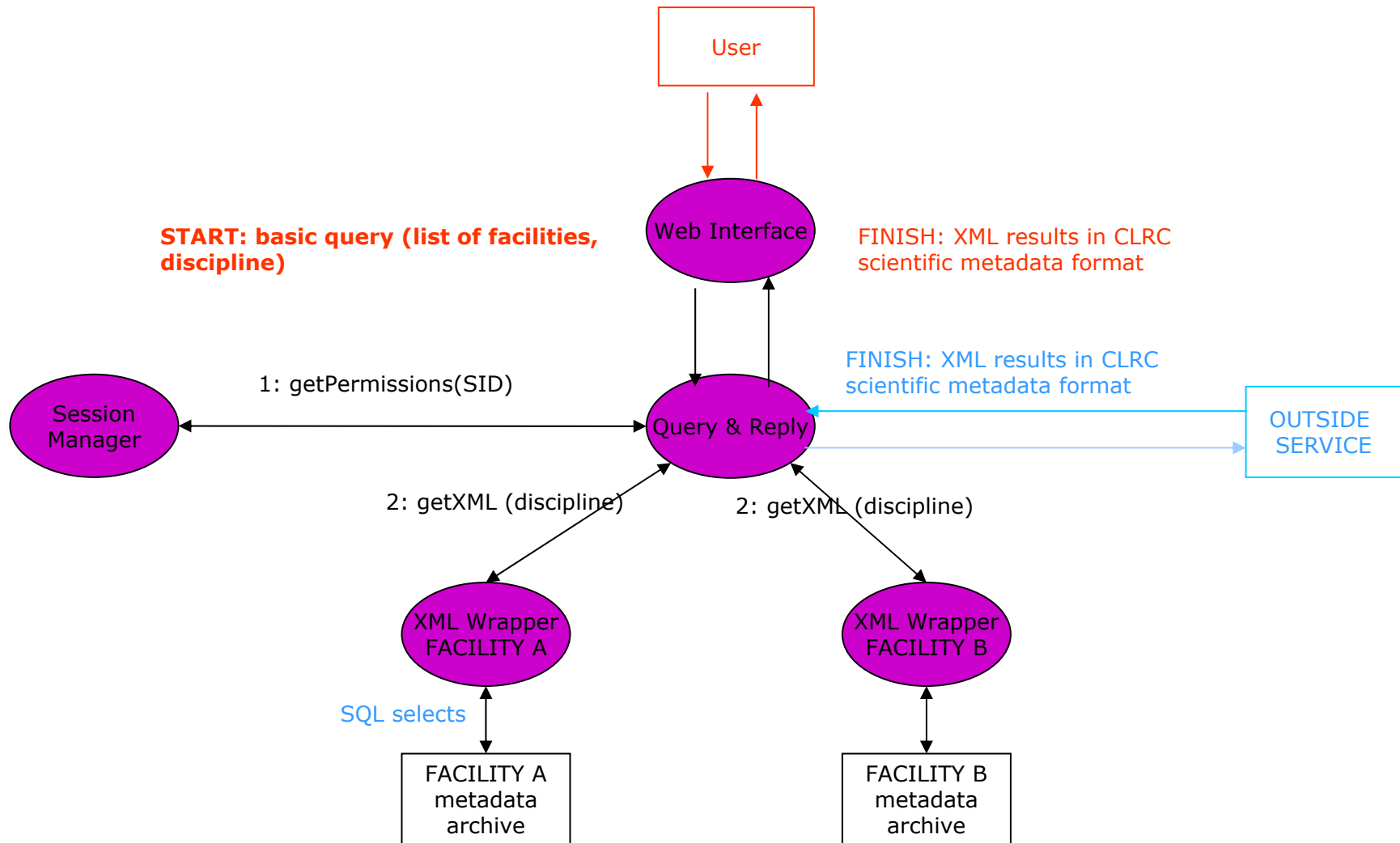


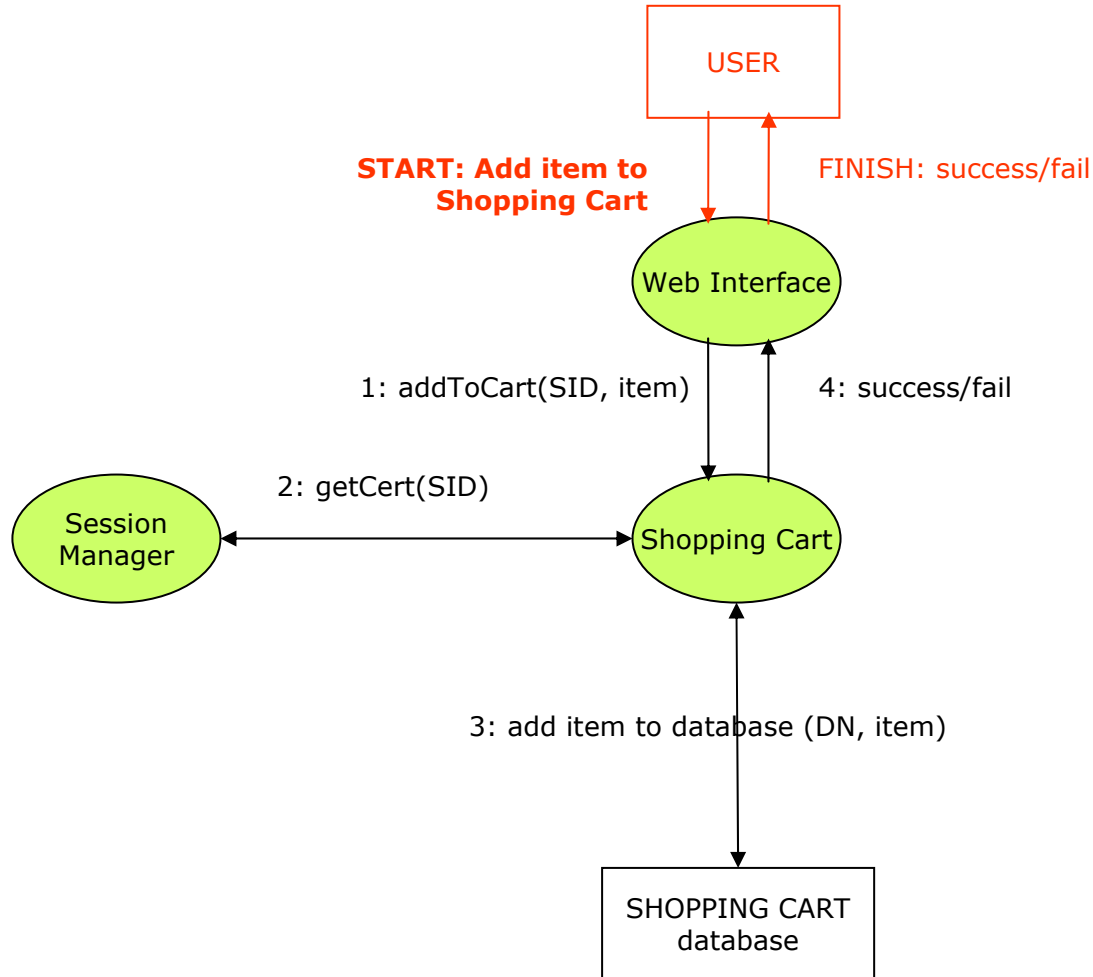
- Access Control, Authentication and Authorisation, Data Transfer, Shopping Cart, User Administration, Facility Administration and Accounting.
- Functions grouped into modules, each with web services interface, interface description in WSDL, communication via soap. These Modules could be shared or exchanged with their own implementation.



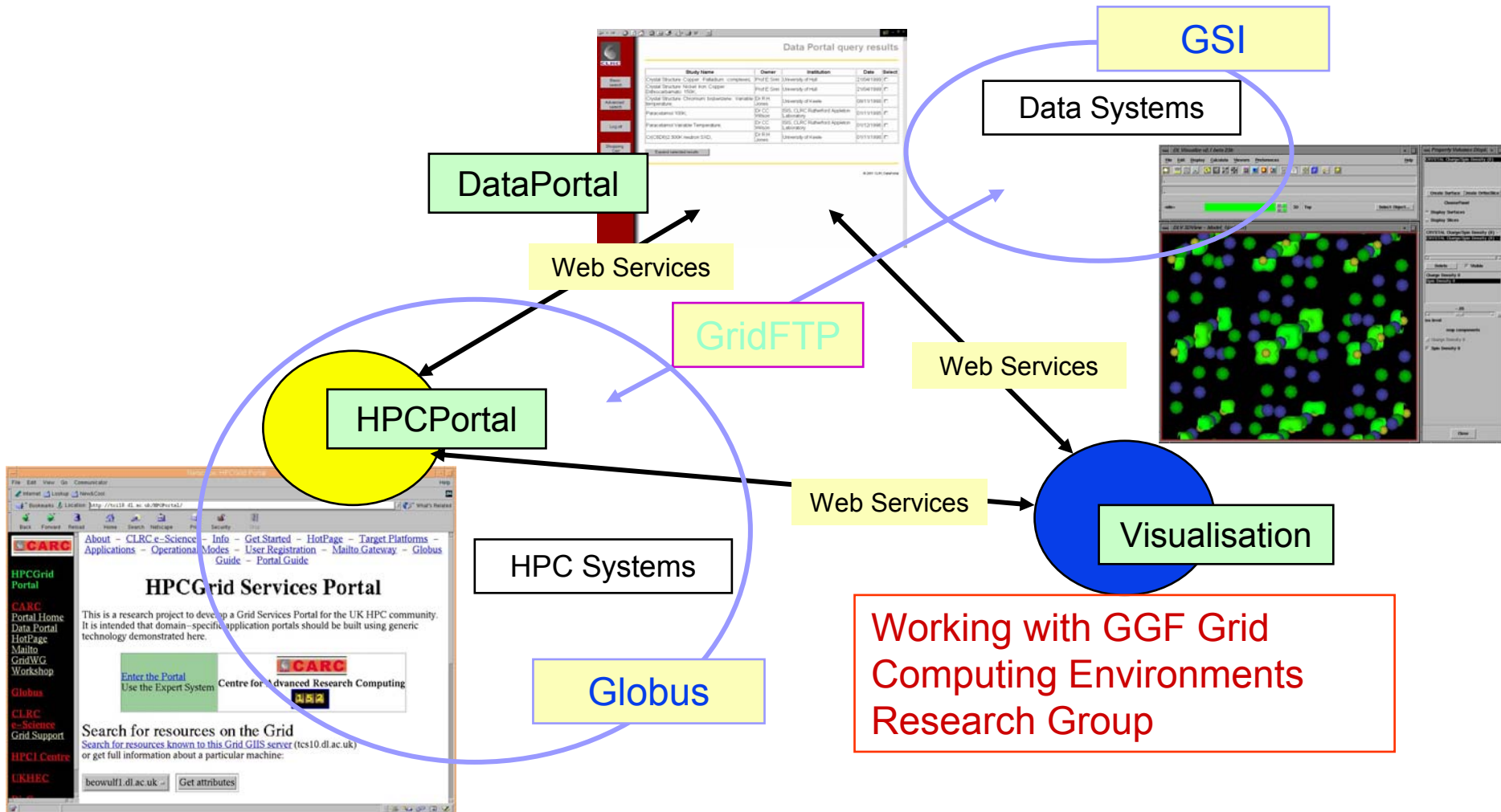
- XML-Wrapper, RasDaMan Insertion and Extraction, Specialised Search, HPCPortal, Visualisation Portal, SRB, other DataPortal Instances .
- Other services that are linked with the DataPortal, but are not integral part of it. Registered with the Portal and accessible via web services interface.



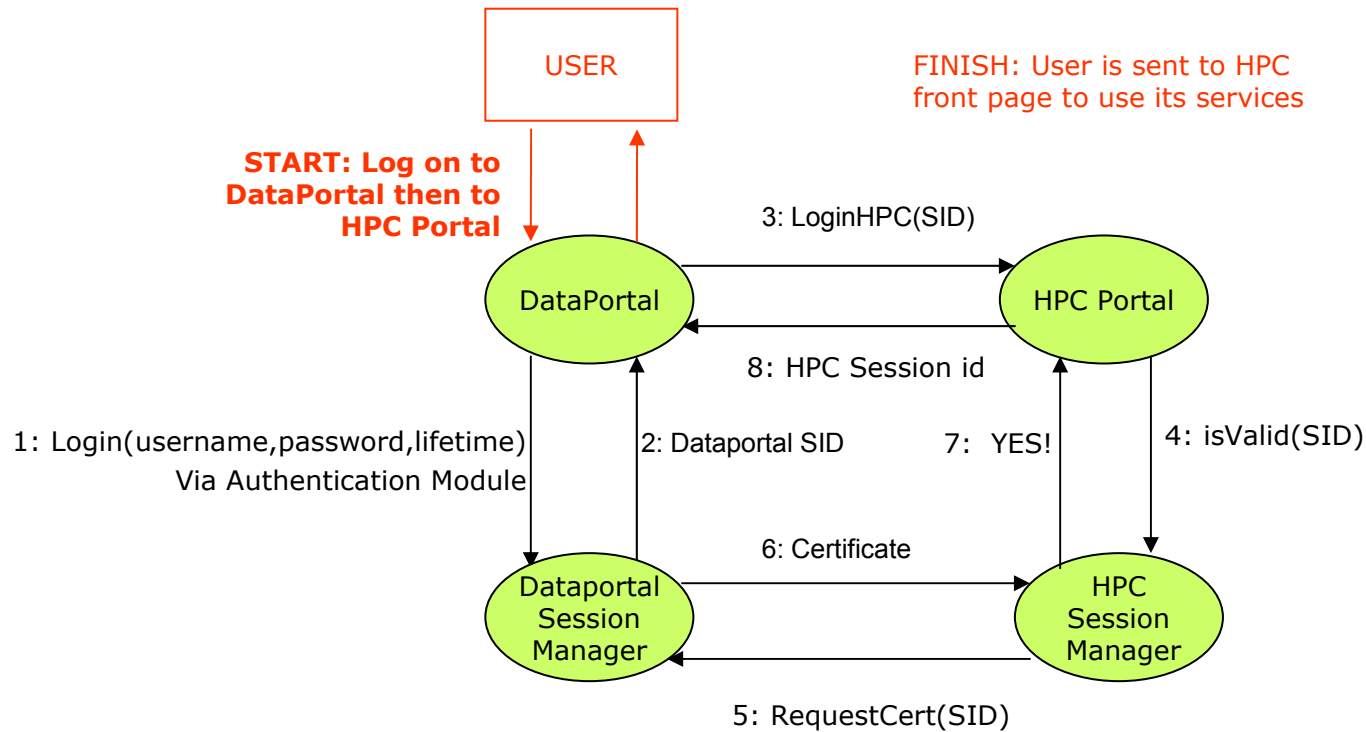




- Another e-Science Centre project to develop a Web portal to search for resources and submit HPC applications to a computational Grid.
- Uses Globus toolkit v2.2
- Functionalities include:
 - Resource Management: GRAM.
 - Information Services: MDS.
 - Data Management: GridFTP and GASS.
 - All use GSI security protocol as the connection layer.

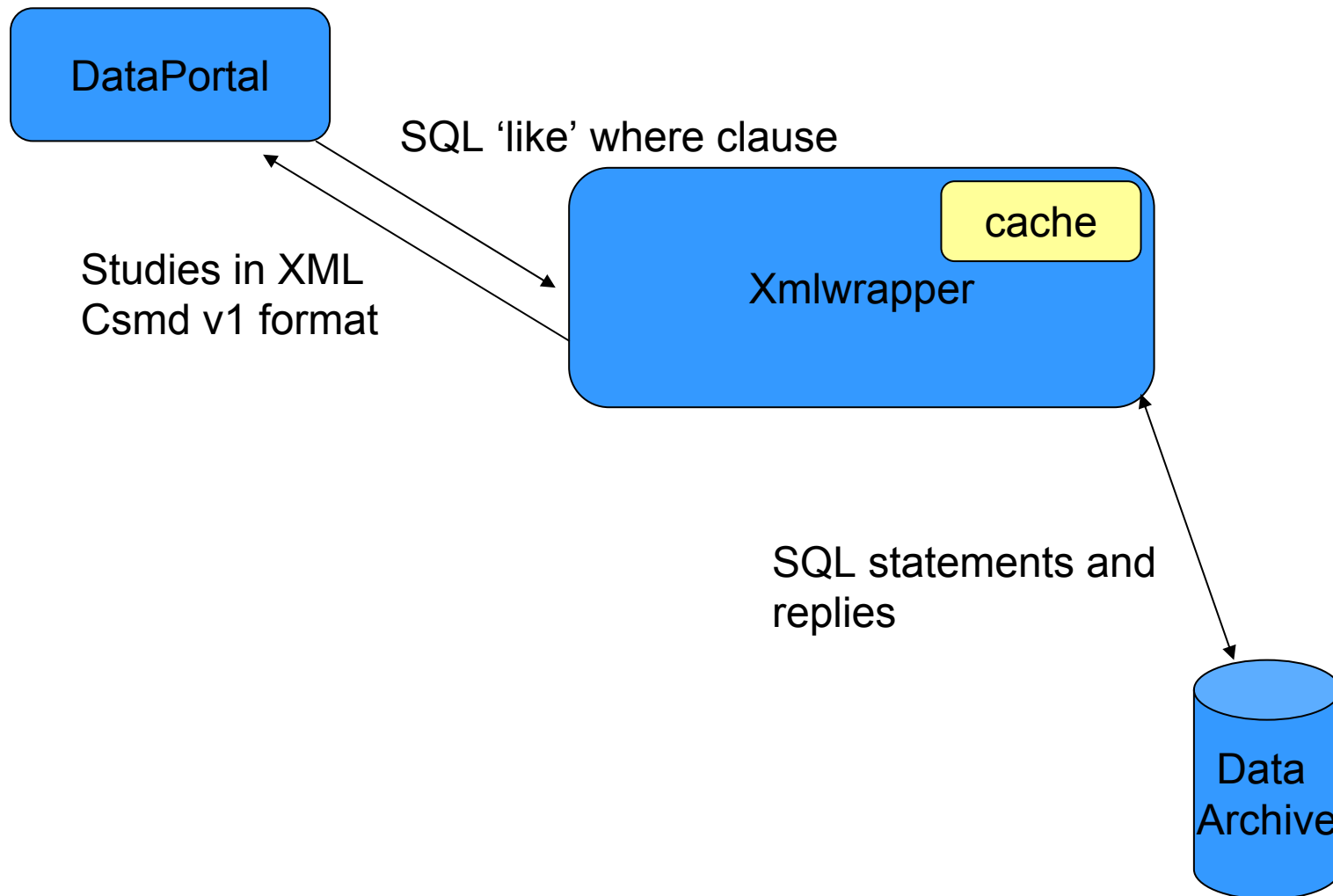


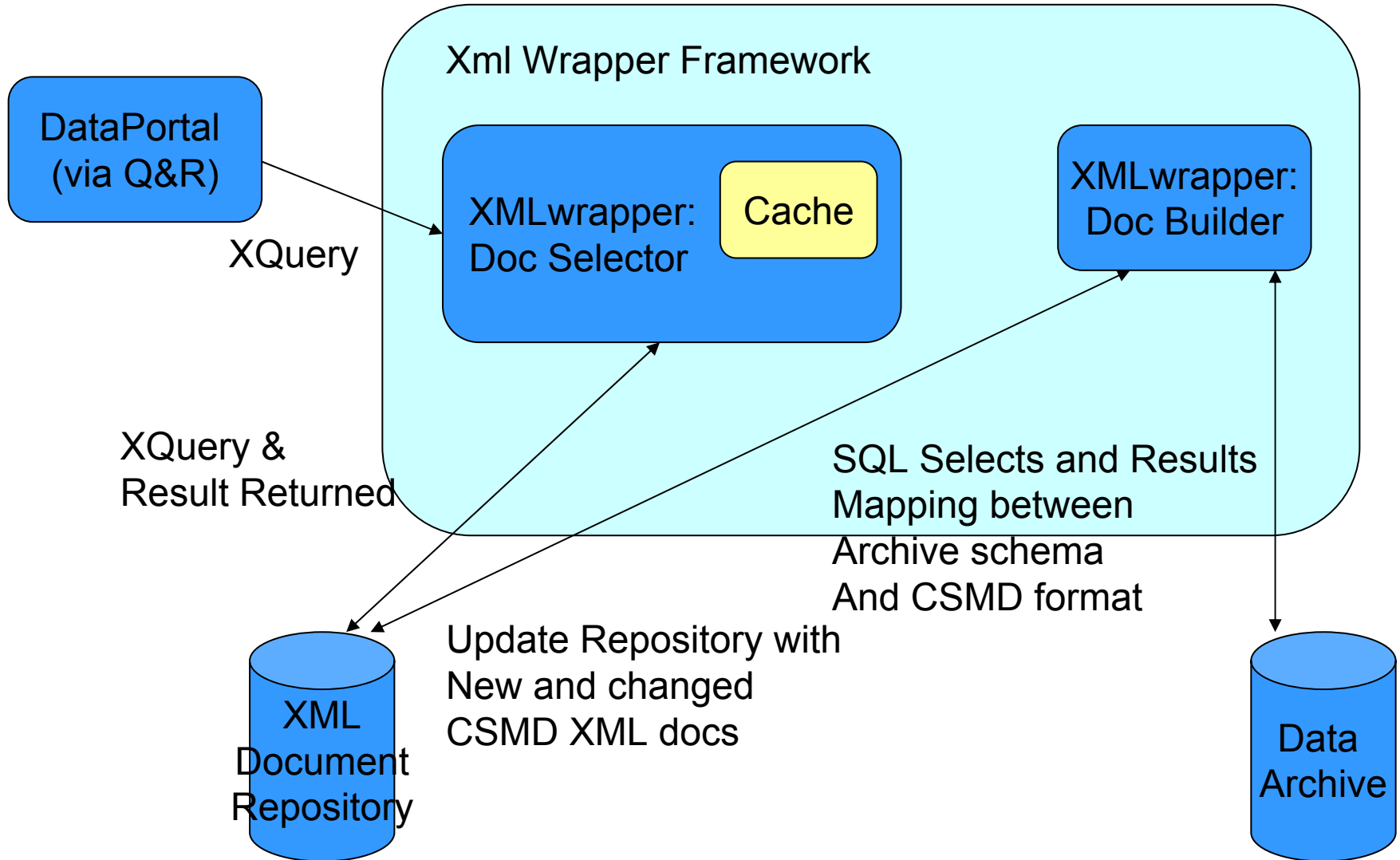
- How do you have signal sign on?
- Both HPC and DataPortal have their own Session Managers which rely on Globus Proxy Certificates.
- Integrated session managers communicate over SSL using mutual authentication between the web servers.
- Allows certificates to be transferred between portals allowing single sign on.
- The certificate can then be used for GSI authentication.



- User logs on to Data Portal and searches for data.
- The data found is added to the persistent shopping cart.
- The user could then transfer the data to another machine using GSI FTP, either using the Data Portal or the HPC Portal.
- Using single sign on, the user could then go directly to the HPC, and use a remote job run on the data that they have just transferred and using GSIFTP, transfer the results back to their machine for analysis.

- Converting to a new Xml Wrapper architecture
- Problems:
 - DTD limits us to one investigation per study.
 - Any new wrappers should be written using the new standard.
 - Only allows one investigation per study (as using the old format).
 - XSLT time consuming to write and must operate on the whole study.
 - Need to transfer the whole studys – this can be large.
 - Data Archive goes down we cannot access study even though it is cached.





- Xml Wrapper still operates if archive is down.
- Concurrency much better supported as selecting from the archive is essentially one thread.
- Only relevant data sent between xml wrapper and Data Portal – also XQuery can include formatting rules which can potentially do away with the need to XSLT.

Questions?