



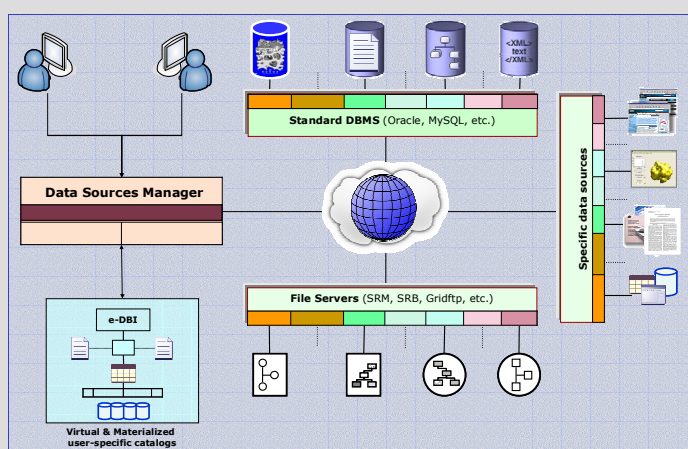
e-DBI: e-science Database Integration

A. Benabdelkader, V. Guevara
Science Park 107, 1098 XG, The Netherlands

Introduction

The main goal of the e-DBI tool is to allow the scientists to seamlessly **integrate several of multi-format data sources**. This tool, in essence, facilitates the navigation and exploration of potential scientific data repositories for data integration. In any scientific data integration scenario, the scientists need to perform several activities and tasks to gather and collect all the information from the data sources. With e-DBI, however, these tasks are performed in a **single-access point**, while the integration is carried out by **defining a virtual database** based on the collected data sources. The e-DBI tool allows the scientists to create the virtual database using their DBMS of their choice and it also **enables customizations** for the location and format of result sets coming from the data sources, in a flexible manner.

High-level view of the integration framework



Data integration procedure

The e-Science Database Integrator (e-DBI) aims at providing a data access interface more suitable to scientists. In order to define a virtual (integrated) database, a scientist needs to consider the following steps.

1. Define a Virtual Database VDB, using a RDBMS
2. Select the needed information (tables) from the different data sources, where it is possible to:
 - ✓ Filter the data
 - ✓ Rename table name and attributes
 - ✓ Reformat the data (apply some conversion if required)
3. Transfer the data into the new VDB, by copying the information
4. Enhance the virtual database VDB by:
 - ✓ Setting new constraints
 - ✓ Merging or fusing the data
 - ✓ Applying additional reformatting, etc.
5. Update or refresh the VDB
 - ✓ Check any time availability and completeness at the sources
 - ✓ Decide whether to perform an update or overwrite
 - ✓ Remove the data from the VDB, etc.

The current implementation of e-DBI supports the following data sources:

- ✓ Oracle, Sybase, MySQL, XML, Excel Spreadsheets, etc.

Data Integration Operation

