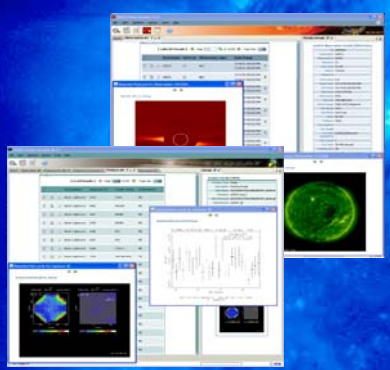
ESA New Generation Science Archives (II) : Data Layer

ABSI: Archives Building Systems Infrastructure:

- Common Scientific Archives Architecture.
- Set of tools and libraries.
- State of the art technology.
- Three tier architecture.
- Quick building and deployment.
- Easy maintenance.



Data Layer Requirements

- Access to metadata and data products
- Flexible granularity of data with millions of searchable elements
- Short response times
- Deliver an efficient and reliable access using open source technology:
  - Postgresql (database)
  - Spring (application framework)
  - Hibernate (persistence framework)
- Our software is built on top of these software solutions to provide additional functionality, through the following modules:
  - Query Manager Module → Metadata
  - Data Distribution Module → Data Products

Query Handler

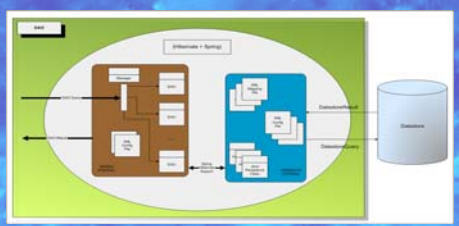
- Point of entry for all requests of the Query Manager
- Requests processing steps:
  - Call Query Builder to build a Query Object (if needed)
  - Call the corresponding DAO
  - Process the result with the Result Handler (if needed)
  - Give back the results

Query Manager Module → Metadata Access

- Handles all requests related with metadata information using state of the art software, such as:
  - Connection Pooling
  - Statement Pooling
  - Distributed Query & Objects Cache
- Composed of five different modules:
  - Query Handler
  - Query Builder
  - Geometry Handler
  - Result Handler
  - Data Access Objects (DAOs)

DAO (Data Access Object) Layer

- The DAO layer encapsulates access to stored metadata
- Hibernate mapping relational tool is used for several reasons:
  - Maps database tables to java classes
  - Provides data query and retrieval facilities
  - Manages connection pooling
  - Manages query and object cache
  - Offers portability between different databases



Geometry Handler

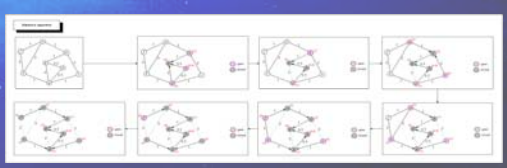
- Receives an geometrical condition
- This condition is replaced with spherical data types, functions and operators provided by the Postgresql module PgSphere

Result Handler

- Receives an object or a set of objects from the DAO layer, and "casts" them to the corresponding project dependent object
- Objects are recognized as simple (String, Integer,…) or complex elements of the project.

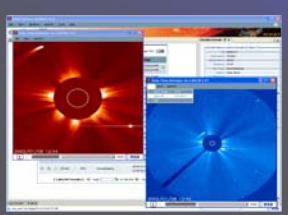
Query Builder

- Receives a QueryBean object and returns an equivalent object which can be understood by the DAO Layers
  - If there are any geometrical conditions, the Geometry Handler is called and its result added to the final result.
- The correspondence between the QueryBean entities and the database tables is achieved by the use of xml mapping.
- After the target entities are known, the Dijkstra algorithm is used to "join" and extract the requested attributes. The Dijkstra algorithm, conceived by Edsger Dijkstra in 1959, was chosen because:
  - It is a graph search algorithm that solves the shortest path problem for a graph, outputting a shortest path tree, which is often used in routing solutions.
  - Can be reapplied to our database tables and their relationships.



Data Distribution Module → Data Access

- Handles all requests related with repository files.
- Data can be requested in different granularity levels, dependant upon the project.
- Once the module finds the data requested, a HTTP URL will be provided to the user for easy data access. This URL is internally redirected to an FTP server.
- Offers synchronous and asynchronous data retrieval
- Synchronous (Direct Download)
  - Data available in FTP holding area, accessible from HTTP (redirected)
  - Authentication required
- Asynchronous (Shopping Basket)
  - Email notification when ready
  - Data available in FTP holding area, accessible from HTTP (redirected)
  - Authentication required
- Retrieval of postcards, when available, is treated as a synchronous request.
- Animation sequences can be performed by date order retrieval of postcards



Use Case: Soho Science Archive

- 3 levels of granularity in metadata display & data retrieval
- 1 884 997 observations in the database
- 765 356 coronal images in the database
- 941 424 full disk size images in the database
- 3 591 777 files in the data repository

… and growing !

SOHO Science Archive
http://soho.esac.esa.int/data/archive/index_ssa.html
EXOSAT Science Archive (coming soon)
http://www.sciops.esa.int/index.php?project=SAT

04-08 Oct. 2009
ADASS XIX
Sapporo

European Space Astronomy Centre (ESAC) - Villafranca del Castillo, Madrid (Spain)

I.Leon, C. Arviset, D.Baines, I. Barbarisi, J.Castellanos, N.Cheek, H.Costa, N. Fajersztejn, M.Fernandez, J. Gonzalez, A. Laruelo, I.Ortiz, P. Osuna, J. Salgado, A. Stebe, D. Tapiador