JRA3: New mining and Integration Tools

Contact person : <u>J.Tennyson</u>

**Objectives :** 

This JRA will develop extensions to the baseline infrastructure. Key objectives are the design of advanced data mining tools and the design of cross-matching and cross-federating tools, providing sophisticated integrated science services aimed at maximising the scientific utility to the end user community of the VAMDC services.

WP Lead: UCL(3)

Description of work :

Through the activities of JRA1 and JRA2, the AM resources will be searchable and will provide information in a standardised way. The following step is to build the query protocols that will access those published AM data and then to design software that will handle and process those data.

Tasks:

## Task1: Registry Queries (lead by CNRS(1) with (12))

We will need to use protocols to query the registries at a fine level of granularity. Indeed we don't wish to only find resources having implemented a type of service such as SSAP or TAP, but rather be able to select resources according to their content through key words. The purpose of Task 1 is to implement those protocols.

Task 2: Tools for Manipulation of Data (lead by KOLN(7) with (1))

Our queries will return data organised according to schemes defined in JRA1. Those schemes will be quite complex because they will reproduce all the scientific concept attached to the data. Therefore the handling of the XML files will be complex and will require specific tools. For now we identify too main generic tools: one performing cross-matching of data and one performing cross-federation of data. These tools are particularly difficult because they require to compare the content of many fields in the schema. Those generic tools will be made available for download in SA1 to the end users and developers. Support to adapt those tools to specific applications will be provided in SA2. We plan to provide libraries to allow users to develop their own applications.

## Task 3: VAMDC advanced data mining services (lead by UCL(3))

With the deployment of a vast range of high value data services through the standard VAMDC infrastructure, this task will investigate optimal strategies to best mine these AM data resources to both advance the creation of new AM fundamental data, and by providing stream lined automated access to appropriate AM data targeted at specific user groups (for the astronomy community benefiting from the availability of high energy data from satellites such as Swift, XMM, Chandra, who require specific atomic data for high excitation species of elements such as iron). This task would investigate the provision of application services wrapping complex work flows combining AM data access, manipulation, and integration into user processing chains – e.g. in solar physics, astro-biology/ chemistry and so forth.

Deliverables : <u>JRA3: New mining and Integration Tools</u>

Milestones : WP8