

The Information Resources in Arcetri Astrophysics Observatory: Between Metadata and Semantic Web

Roberto Baglioni

*Dipartimento di Astronomia e Scienza dello Spazio,
Università di Firenze, Largo E. Fermi n.2, Firenze, Italia*
baglioni@arcetri.astro.it

Antonella Gasperini

*INAF-Osservatorio Astrofisico di Arcetri, Largo E.Fermi n. 5,
Firenze, Italia*
gasperini@arcetri.astro.it

Abstract. It is becoming apparent that libraries are going to play a key role in the new W3C's (World Wide Web Consortium) paradigm for the semantic web. For this reason, the Arcetri library is investigating methods for publishing different kinds of electronic documents on the net and a way of enriching them with semantic metadata. For the first phase, we are focusing on the library catalogue; and, in a second phase, we will consider bibliographies, preprints, technical reports, web pages, archives of astronomical data, and photographic and historical archives.

1. The project

The library of the Arcetri Astrophysical Observatory is investigating methods for publishing different kinds of electronic documents on the net and a way of enriching them with semantic metadata.

We have realized that RDF (Resource Description Framework) from W3C and the DCMS (Dublin Core Metadata Set) are, for the moment, the two main building blocks of the semantic web. Furthermore, RDF seems to be a good method for representing IFLA (International Federation of Library Associations and Institutions) FRBR (Functional Requirements for Bibliographic Records) information.

2. The steps of the project

For the first phase, we are focusing on the library catalogue; and in a second phase, we will consider bibliographies, preprints, technical reports, web pages, archives of astronomical data, and photographic and historical archives.

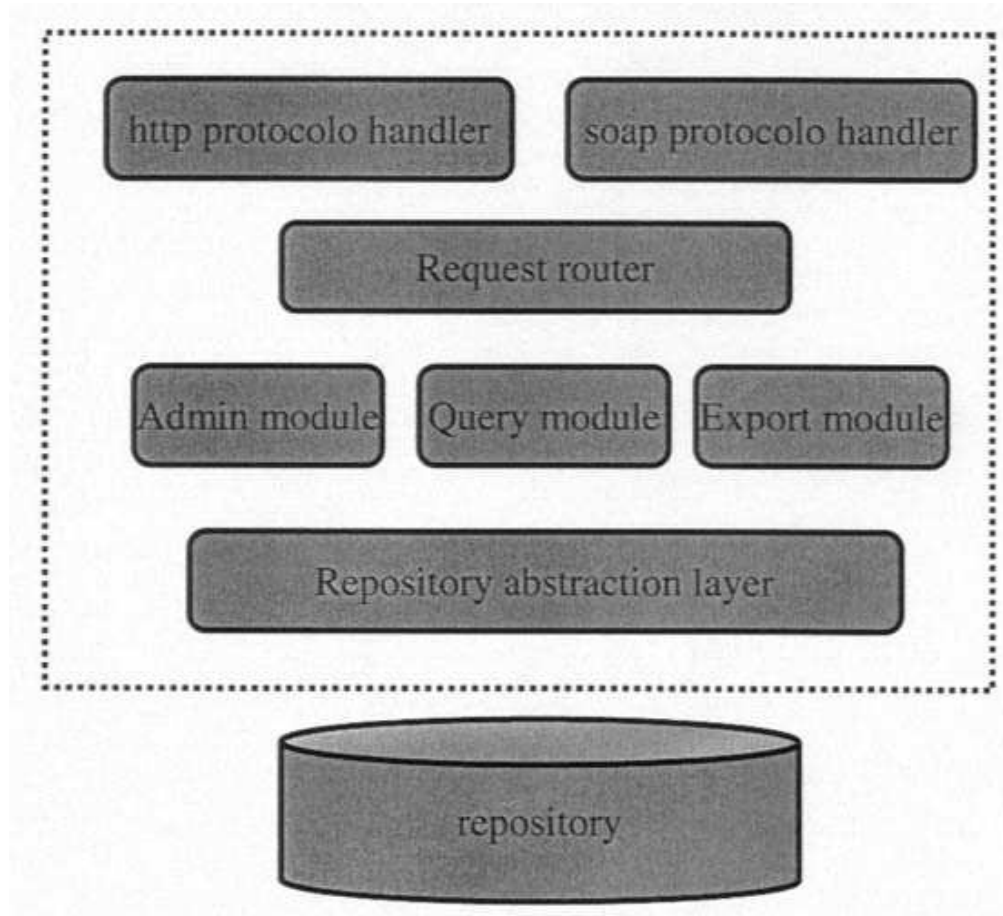


Figure 1. Architecture of Sesame

In order to build sophisticated relationships, all the documents contained on the Arcetri's web site can be filled with metadata. RDF superimposes a structure for enhancing the way we represent knowledge. This structure makes it easier for the machines to process statements. Each RDF statement is expressed in triplicate, and follows predefined patterns.

Actually, for better processing, RDF is usually expressed using the XML syntax.

For example:

```
<?xml version = '1.0' ?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:DC="http://purl.org/dc/elements/1.1/">
<rdf:Description
rdf:about="http://www.something">
```

```

<DC:creator>
Joe Smith
</DC:creator>
</rdf:Description>
</rdf:RDF>

```

3. Tools: RQL and Sesame

The first thing we need is a tool for storing, querying and modifying a model that will represent a set of RDF triples. Therefore, a DBMS (DataBase Management System), with an RDF front end, is the natural choice. RQL (RDF Query Language) is a specific language for RDF queries. The queries in RQL are translated in SQL (Structured Query Language) by the front end. The name of the front end is Sesame.

Sesame is an architecture for efficient storage and expressive querying of large quantities of RDF metadata. Sesame is developed by AIdministratoR Nederland b.v.

<http://www.aidadministrator.nl>

Sesame can handle requests from the net. Three modules are available for adding, querying and exporting RDF models. RQL is a query language specific for RDF and RDF schema.

```

select author,title
from{resource} DC:creator {author}
{resource} DC:title {title}
where title like "*stellar*"

```

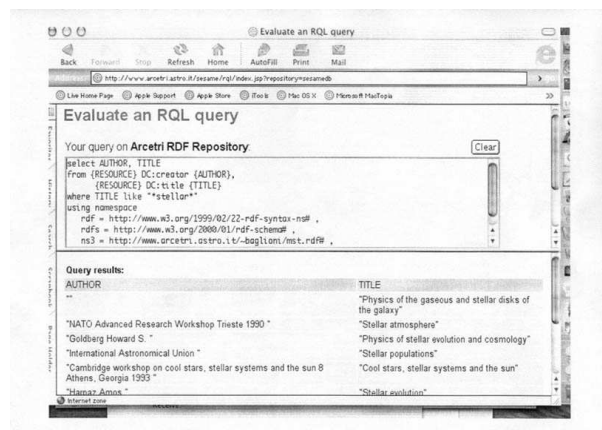


Figure 2. An RQL query