Use Case (astrovirtel)

Date: 8-April-2003
Version: DRAFT 0.1
Contact: Alberto.Micol@eso.org

Table of Contents

Use Case Background

The Use Case here described is loosely based on the cycle 2 ASTROVIRTEL program by R. De Grijs (IoA, Cambridge). The case has been simplified to ease the description; parts not relevant to the VO Registry have been omitted.

Notation:

While describing the use case, some Registry functionalities and/or requirements are highlighted by the string ">> RGST_REQ"

Science Domain:

Star cluster luminosity functions of nearby galaxies, and their dependence on the environment and evolutionary state of the host galaxy.

Scientific Work Plan:

The Astronomer wants to perform a statistical study of the luminosity functions of star clusters (CLF) in nearby galaxies. To study the dependency of the CLF on the environment and age, galaxies are grouped into different bins.
To simplify this use case, let's define the bins in this naive way: elliptical, spiral, irregular, each bin divided in two categories: with or without nuclear activity, for a total of 6 bins.

For statistical significance, about a dozen of galaxies should be studied in each galaxy-type bin.

In order to study the luminosity functions of star clusters in nearby galaxies, the Astronomer needs to measure star cluster magnitudes (U, B, V, R, I, J, H) on all available archived images of nearby galaxies closer than 50Mpc, provided the resolution, field of view and exposure times are adequate.

The observations must be deep enough to probe the cluster luminosity functions down to at least 1.5 mag below the turnover magnitude of Mv ~ -7.4 found for old globular CLFs.

To resolve spurious detections, observations in at least two passbands are required.
Good quality (photometric conditions), well-calibrated observations are required.

Data:

This use case is not focusing on the exploitation of a known-by-the-user data center. At the contrary the Registry is used to identify those registered services that could provide relevant pixel or catalogue data.

Actors and system components

The primary actor is the Astronomer who uses the Registry to select (first) and query (afterwards) the services (both archives and catalogues) that potentially provide the type of data s/he is interested in.

Use Case Detailed Description

Astronomer requirements and workflow scenario

A1) the Astronomer is planning to use the Registry to find services that will help him/her to compile a list of nearby galaxies. Such list should include those galaxies for which the embedded stellar clusters can be observed by existing groundbase or spacebased instruments (constraint on field of view, resolution, telescope area).
Such list of galaxies does not need to be complete, but it needs to be big enough to increase the chances of finding enough multicolour observations in the data archives.

The list should contain the following attributes:

- Galaxy Identifier
- Galaxy Type (Elliptical, Spiral, etc.)
- Distance
- Position (RA, DEC, J2000)

The attribute "Position" is in the list not because scientifically relevant, but because it might be necessary for later querying the data archives, in case the archive services do not provide the "name resolver" capability.

Note: Which ones are the ancillary attributes (not directly required by the user) needed by subsequent services ?

Since the position is only used to query the data archives, the astrometric accuracy does not need to be particularly good; a precision better than, let's say, half of the field of view of the queried instruments, e.g. 1 arcmin, would be consider adequate for this project.

A2) the Astronomer is planning to use the Registry to identify which archives/telescopes/instruments can provide data of the required type (resolution, field of view, spectral coverage, data type [calibrated or calibrate-able], photometric conditions, etc.). In this particular case:

A2.a) UBVRIJH fits images in at least two bands
A2.b) Adequate Field of view (larger than 2 arcmin)
A2.c) Adequate Resolution
First order: Pixel scale < 0.6 arcsec
Second order (adaptive): Pixel scale < 90ly/Galaxy Distance
(the same instrument might be ok for a closeby galaxy, but could not have enough resolving power for more distant galaxies)

A2.d) Adequate limiting magnitude (Absolute Mag(V) >= -5.9)
A2.e) Seeing < 1.0
A2.f) Photometric conditions (for groundbased observations)

Also, the images must be calibrated, or should be possible to calibrate them:

A2.g) calibrated products or, otherwise, raw images with all necessary ancillary files for user reduction and calibration.

A3) the Astronomer is planning to use the Registry to identify the services that will help him/her to query the selected archives (A2) for each of the galaxies selected in (A1)

> **Note**: While each archive will provide its own service to perform this task, there should be an integrated VO functionality to perform this on the Astronomer behalf.
> This is not discussed any further since is not involving the Registry.
> It is food for thoughts for a VO Portal that should be able to integrate the various services.

A4) the necessary requests are submitted to the relevant archives.

As for A3.

After A4, the Astronomer will happily analyse the retrieved data.

Registry requirements and possible scenarios

R1(A1).- Selecting services that could help compiling the desired sample

The Registry is queried to identify services of a certain type.

R1a.- services of type: CATALOG_BROWSER
that is, catalog search engine, e.g., Vizier or ADC

R1b.- services that have to do with a particular class of objects: GALAXY

R1c.- services that return a particular class of data: OBJECT_CATALOG

>> RGST_REQ: Thesaurus of all service types, object classes, data classes

Among those, the Registry must identify those services whose interfaces provide access (input & output parameters) to some project-specific attributes.

R1d.- services that can return as output parameters:

- galaxy name or id
- galaxy position (precision better than 1arcmin)
- galaxy type (MORPH_TYPE)

R1e.- services that can be constrained by the distance

In general, a service is not ready to accept constraints for all its output parameters.
This requirement could be avoided, but, for his/her convenience, the Astronomer is going to request only services which are able to accept a constraint on the distance; this will limit the number of returned records (once the service is queried), and will eliminate the need to go through the sample and remove the entries with distance > 50 Mpc.

This requirement has some non-trivial implications, and requires some attention also from the UCD working group. Please refer to the SPECIAL REQUIREMENTS section below (SR5 Explicit vs Implicit Constraints, and SR6 UCD Equivalence).

R1f.- Gather information on the number of records offered by the services previously identified (R1a-R1e)

The user would need to know which of the services previously identified best fits his/her own user requirements, typically in terms of Data Quality (, precision of returned measurements, sample), or simply because it is easier to use.

For this particular use case, the astrometry is not an issue, hence there is no need to qualify on the coordinate precision. Instead, what is relevant here is the fact that the constructed sample should be big enough to maximise the likelihood of finding observations in the data archives.
Hence, the Astronomer prefers to avoid services that do not cover large number of galaxies.

R2(A2).- archives/telescopes/instruments able to provide required data

Once the sample is built, the Astronomer must identify which archives/telescopes/instruments provide the necessary data.

The Registry is used to identify services of a certain type.

R2a.- services of type: PIXEL_ARCHIVE
R2b.- services that can return a particular class of data: IMAGE

>> RGST_REQ: As for requirement A1, there must be a standard list of service types and a standard list of data classes

R2c.- Cover some (at least 2) of the UBVRIJHK passbands (see A2.a)
R2d.- Provide a Field of View FOV > 2 arcmin (see A2.b)
R2e.- Offer a resolution better than 0.6 arcsec (see A2.c)

Standard description of archives, telescopes, instruments, detectors, filters, gratings, instrument modes, is required. This is up to the Data Model WG.

>> RGST_REQ: Discuss the Registry-Data Model interface for requirements like R2c, R2d, and R2e with the Data Model WG

R2f.- Offer calibrated images (with associated meta information, i.e. not jpeg) or, otherwise, raw files along with the necessary ancillary information (reference files, etc.) to calibrate them.

> *The word "image" is too generic. Let's introduce the word "scimage" for science image, which signifies that the image is accompanied by its metadata (read: header), as opposed to a jpeg image.*

R2f might be seen as a subclass of R2b; in that case R2b could be re-written:

R2b*.- can return a particular class of data:

> data_class = SCIMAGE/CALIBRATED or SCIMAGE/CALIBRATE-ABLE

>> RGST_REQ: Standard list of data classes and subclasses

Other Astronomer's requirements have to do with Limiting magnitude (A2.d), seeing (A2.e), and photometric conditions (A2.f). These constraints are not to be used by the Registry, unless these parameters are stored in the metadata associated with the services. Typically it will be the data provider service that will deal with such complex and detailed aspects.

R3(A3) (query archives) and R4(A4) (submit requests) are not treated here (see A3 and A4)

Special Requirements

Requirements **R1a, R1b, R1c, R2a, R2b** and **R2f/R2b*** requires the availability of various standard lists:

- SR1 Service Types (e.g., catalog_browser, pixel_archive, etc.)
- SR2 Data Classes (scimage, spectrum-1d, event list, radio map, catalog, etc.)
- SR3 Data SubClasses (calibrated, raw, mosaic, preview, binned, etc.)
- SR4 Object Classes and SubClasses (galaxy, star, etc.; elliptical, novae, etc.)

Requirement **R1e** ("Services that can be constrained by the distance") highlights at least two problems:

SR5(R1e): Explicit vs Implicit Constraints

Example: there might be a service known (in its metadata) to offer a catalogue of galaxies all within 50Mpc, but not having explicitly the "galaxy distance" in the list of constrainable parameters. In such case, it is in the meta information that the Registry should place a constraint (implicit).

>> RGST_REQ: The Registry should be able to handle both explicit and implicit constraints; the ability of mapping an explicit constraint to an implicit one is the challenge.

SR6(R1e): UCD Equivalence

There are various possible UCDs that can be used for the "galaxy distance" attribute, the most obvious being:

POS_GAL_GC (Galactocentric Distance) in Mpc
PHOT-DIST-MOD (Distance Modulus)
REDSHIFT_GC (Galactocentric Redshift)
REDSHIFT_PHOT (Photometric Redshift)

NOTE: ASTROVIRTEL used the LEDA service to build the sample of galaxies by posing a constraint on the distance modulus.

The Registry needs to know that all those UCDs are, at least for this project, equivalent.

Comment: Some other project might want to reject the photometric redshift from the list of possible UCDs, but most likely the rejection should come from the error associated with the measurements and not by the type of experiment that led to a given measure.

>> RGST_REQ: Make sure UCD WG tackles the UCD Equivalence problem.

>> RGST_REQ: The ability to convert between equivalent UCDs would be required to support constraints on UCD values.

APPENDIX

Depicting Registry functionality in sort-of-SQL/Scripting Language

F1(R1).- Identifying services to compile a list of nearby galaxies

```
Select
     service_name, service_description, service_rowcount
from
     Registry
where
     service_category   = "CATALOG_BROSWER"
 and data_class         = "OBJECT_CATALOG"
 and subject            = "galaxy"          (IVOA Thesaurus?)
 and querable_parameter = "galaxy distance"(mapped to list of UCDs?)
 and output_parameter   = "galaxy name"

save into thisSession.catalogServices
```

At this point two scenarios are possible:

(1) The optimal service has been identified using R1a to R1f: the Astronomer will need to learn how to interact with it (e.g., how to formulate the query, and how to handle the results).

(2) There is not a single optimal service; instead, various services are to be queried, and results have to be merged to compile the required sample of galaxies.

F2(R2).-Identifying archives/telescopes/instruments able to provide the required data

```
select REG.serviceID, REG.DM.instrument,
       REG.DM.observing_mode, REG.service_interface
from Registry
where
     service_category = "PIXEL_ARCHIVE"
 and wavelength in (U,B,V,R,I,J,H,K)
 and data_class ="SCIMAGE/CALIBRATED"
 and limiting_magnitude > 23    => OPTIONAL (only if in service metadata)
 and pixel_scale   < 600  (mas) => OPTIONAL (")
 and field_of_view > 2 (arcmin) => OPTIONAL (")

Save into thisSession.archiveServices
```

Once the relevant archives have been identified, the Astronomer or, better, a VO portal should provide the functionality here described.

F3(R3).- Querying all selected Archives

Results collected (list of galaxies) are sent to all archives identified in thisSession.archiveService

WHO ?
Registry?
Astronomer?
VO Portal?

```
Select
     observation_ID, archive_ID
From
     thisSession.archiveServices
Where
     thisSession.archiveServices.objName =
thisSession.catalogServices.objName

Save into thisSession.datasetList
```

F4(R4).- Submitting Requests

```
Foreach ( dataset, archive_id ) in thisSession.datasetList

   submit_request( avo_user_id, dataset, archive_id )
```

Submission of archive requests implies solving the following problems:
- user registration/authentication
- the archiveService must provide a submitRequest service

These aspects should be tackled by the Data Access Layer WG.