ESO - EUROPEAN SOUTHERN OBSERVATORY

EUROPEAN SOUTHERN OBSERVATORY
Organisation Européenne pour des Recherches Astronomiques dans l'Hémisphère Austral
Europäische Organisation für astronomische Forschung in der südlichen Hemisphäre

Virtual Observatory Systems Dept.

ESO/ST-ECF REGISTRY REQUIREMENTS for the Science Archive Facility

VOS-RRD-ESO-100

Issue 1.0/0
2005.02.24
12 pages

Prepared:	Marco C. Leoni		
	Name	Date	Signature
Approved:	Approver's Name		
	Name	Date	Signature
Released:	Releaser's Name		
	Name	Date	Signature

CHANGE RECORD

Issue	Date	Affected Paragraphs(s)	Reason/Initiation/Remarks

TABLE OF CONTENTS

1. Introduction

1.1 Purpose

This document lists and describes the requirements for the resource repository (registry) of the ESO/ECF Science Archive Facility.

This entails metadata requirements in support of the SAF user interface, the ability to resolve queries stemming from IVOA protocols and services, as well as a harvesting interface enabling Virtual Observatory registries to mirror and propagate public ESO/ECF resource metadata.

1.2 Scope

The scope of the ESO/ECF Registry is:

- register ESO/ECF observational data (i.e. no simulated/modelled data yet);
- register existing resources;
- support characterization and ingestion of external data collections based on ESO/ECF observations;
- implement IVOA local publishing registry (harvestable & searchable);
- support the Science Archive (Archive Level 3, see [8] for more details).

1.3 Reference Documents

In the following, only references to the latest standards are provided. The design document will specify the version actually used.

[1] *Resource Metadata for the Virtual Observatory*, http://www.ivoa.net/Documents/latest/RM.html
[2] *IVOA Identifiers*, http://www.ivoa.net/Documents/latest/IDs.html
[3] *VOTable Format Specification*, http://www.ivoa.net/Documents/latest/VOT.html
[4] *UCD (Unified Content Descriptor)*, http://www.ivoa.net/Documents/latest/UCD.html
[5] *STC (Space-Time Coverage metadata definition)*, http://www.ivoa.net/Documents/latest/STC.html
[6] *Key words for use in RFCs to Indicate Requirement Levels (IETF RFC 2119)*, http://www.ietf.org/rfc/rfc2119.txt
[7] *Specification of Physical Units within OGIP FITS files*, http://legacy.gsfc.nasa.gov/docs/heasarc/ofwg/docs/general/ogip_93_001/ogip_93_001.html
[8] *SAF User Requirements Document*, http://eurovodev.hq.eso.org/internal/Vof/ESOScienceArchiveURD/eso_urd2.pdf
[9] *Date and Time Formats*, http://www.w3.org/TR/NOTE-datetime
[10] *Representation of world coordinates in FITS*, http://arxiv.org/abs/astro-ph/0207407

2. Conventions

Following sections list all the conventions used in this document.

2.1 Notation & Typographic Conventions

Each requirement is given a short description and a number of the form **RR-nn** where *nn* is a running number in the document.

2.2 Acronyms

List of acronyms used in the document.

ADQL	**A**stronomical **D**ata **Q**uery **L**anguage
ADU	**A**stronomical **D**etector **U**nit
DB	**D**ata **B**ase
FITS	**F**lexible **I**mage **T**ransport **S**ystem
FOV	**F**ield **O**f **V**iew
GSC-II	**G**uide **S**tar **C**atalogue, version 2.0
GUI	**G**raphical **U**ser **I**nterface
ICRS	**I**nternational **C**elestial **R**eference **S**ystem
IFU	**I**ntegrated **F**ield **U**nit
I/F	**I**nter**F**ace
IVOA	**I**nternational **V**irtual **O**bservatory **A**lliance
OAI	**O**pen **A**rchive **I**nitiative
OGIP	**O**ffice of **G**uest **I**nvestigator **P**rograms
PI	**P**rincipal **I**nvestigator
SAF	**S**cience **A**rchive **F**acility
SIA	**S**imple **I**mage **A**ccess
SOA	**S**ervice **O**riented **A**rchitecture
SSA	**S**imple **S**pectrum **A**ccess
SQL	**S**tructured **Q**uery **L**anguage
UCD	**U**nified **C**ontent **D**escriptor
URI	**U**niform **R**esource **I**dentifier
URL	**U**niform **R**esource **L**ocator
USNO	**U**S **N**aval **O**bservatory
UType	**U**nique **Type**
VO	**V**irtual **O**bservatory
VOResource	**V**irtual **O**bservatory **R**esource
VORegistry	**V**irtual **O**bservatory **R**egistry
VO	**V**irtual **O**bservatory
VOTable	**V**irtual **O**bservatory **T**able
WCS	**W**orld **C**oordinate **S**ystem
ZP	**Z**ero **P**oint

2.3 Terminology

The words "MUST", "SHALL", "SHOULD", "MAY", "RECOMMENDED", and "OPTIONAL" (in upper or lower case) used in this document are to be interpreted as described in IETF standard, RFC 2119 [6].

3. Overview

3.1 Registry in general

A general definition of what a register is can be the following: "an official written record of names or events or transactions" (http://dictionary.com).
In this case, we are talking about an entity that is storing a list of addresses (like yellow-pages) and providing some services based on those registered items.

The concept of registry is more complex than simple yellow-pages, but most of this complexity is hidden from the end user, e.g. accessing it through a web portal or a GUI interface.

Moreover, a register is the central point for an SOA: it provides the ability of dynamically discover and use services (as well as to manage them) is based upon a registry.

It should be noted, that this registry is not the same as a *VO Registry* (see 3.2), but a superset.

Each registered item is called "resource".

There will be two kinds of resources: *data* and *services*.

3.2 VO Registry

More specifically, in the VO the definition of a registry can be the following: "A registry is a dynamic database of metadata describing a set of Internet-available resources. The registry is used to identify and locate resources satisfying user-specified criteria, and to direct more detailed information requests to the relevant services."
A VO registry is a collection of resources – or better to say VO resources [1].

The ESO/ECF registry collection of resources can contain more metadata then the VO registry one, but it has to be possible to map the former to the latter.

4. Resource discovery/search capabilities

Querying the registry is surely the main feature; for this scope:

RR-01 It MUST be possible to perform online search on the metadata associated with the resources.

RR-02 Metadata values MUST be typed, restricted, and SHALL avoid observatory specific terminology.

RR-03 A VO interface compliant to the IVOA registry MUST be provided[†].

RR-04 A harvesting interface MUST be provided, in order to get data published on a full IVOA registry.

RR-05 A query interface compliant to IVOA standards SHALL exist.

RR-06 The capability to access data using different protocols MUST be implemented: the priority is for IVOA data access standards.

[†] Definition of IVOA registry standards is still work-in-progress at the time of this writing.

5. Resource Unique Identification

In order to be registered, a resource has to have a unique identifier that can be used to retrieve or refer to it.

RR-07 Each resource MUST have a unique identifier, used to address it.

RR-08 The identifier MUST have a structure that automatically allows the mapping of this ID with one specified by the "IVOA Identifiers" standard [2], and with the internal ESO/ECF archive identifier.

6. Metadata

Each resource has to be properly characterized.
This meta-information should give a full description - from the scientific point of view - of what is in the resource, enabling the science-readiness of the served resource.

RR-09 A resource MUST have all the mandatory (RR-12) metadata information in order to be registered.

RR-10 The mandatory IVOA metadata [1] MUST be provided.

RR-11 Technical validation of registered resources SHOULD be provided.

RR-12 Units MUST be specified for each measured quantity (see [7] for a list of valid units).

RR-13 *UCD* [4] and *UType* [4] SHALL be generated for each value.

RR-14 Each measured quantity MUST have its associated error.

6.1 General and Specific Metadata

Following is a list of required metadata (bold face), with associated description.
The list is split into three parts: common, data specific, service specific.

RR-15 The registry MUST be able to manage all the following metadata.

6.1.1 Common Metadata

- **Title**: name given to the resource.
- **Identifier**: unambiguous reference to the resource.
- **Publisher**: entity responsible for making the resource available.
- **Subject**: a list of the topics, object/target types, or other descriptive keywords about the resource.
- **Description**: an account of the content of the resource.
- **ReferenceURL**: a URL pointing to additional information about the resource.
- **Type**: the nature or genre of the content of the resource.
- **Release Date**: start of validity of the resource.
- **Expiration Date**: end of validity of the resource.
- **Public Flag**: visibility of the resource itself to different user groups.
- **Searchable Flag**: visibility of the resource metadata to different user groups (only metadata can be searched and served).
- **Updated**: last date-time stamp the resource has been updated, (format `YYYY-MM-DDThh:mm:ssZ` [9]).
- **Tech Check**: closed list indicating the result of technical check about completeness of the resource and compliancy to the standards.

6.1.2 Data Metadata

This set of metadata is further split up according to resource types.

- **Proposal ID**: the identifier of the original proposal (string, list).

- **Target Name**.
- **Instrument**: (string, list). See definition in [1].
- **Detector Mode**: list of observation modes.
- **Optical Elements**: Filter, Grism, etc.
- **Format**: mime type to specify the format hence the type of resource.
- **Processing type**: description of SW/pipeline used for processing (e.g. MVM, IRAF, etc.), with version.
- **Comments**: free text description field.
- **Number of Frames**: number of simple objects used to create the result.
- Temporal info:
  - **Start Date Time**
  - **Stop Date Time**
  - **Total Exposure Time**
- **Type of Observation** (e.g. imaging, spectroscopy, etc.)
- **Technique of Observation** (e.g. IFU, Echelle, Long Slit, etc.)
- WCS:
  - **Reference Source for WCS** (e.g. GCC2, USNO)
  - **Spatial Accuracy**: dispersion astrometry solution, **$<\Delta\alpha>$**, **$<\Delta\delta>$**, **σ_α**, **σ_δ** (arcsec), **Number of Stars used**
  - **Spectral Accuracy**: dispersion wavelength solution, **σ_λ** (Å)
  - **CR Pixel, CR Val, CD Matrix**
  - **C-type** (see [10])
  - **Equinox**
- **Field Of View Size** (Sampling, Orientation, Coverage, and Footprint/finding charts can all be derived from WCS) or **Slit Geometry** and **Orientation**.
- **RA** and **DEC** of the target.
- Spatial info:
  - **Spatial Corners**: coordinates of the corners of the image or spectrum (decimal degrees)
  - **Number of Bins**: number of spatial bins – for each axis.
  - **Bin Size**: pixel size or bin size – for each axis (arcsec or Å).
  - **Seeing** (arcsec) + **Seeing Error**.
- Spectral info:
  - **Wavelength Min**
  - **Wavelength Max**
  - **Resolution Min, Resolution Max**
  - **Resolution Average**
- **Magnitude System**: e.g. Vega, AB.
- **Product Units** (of pixel values).
- **ZP** – and time scale: $mag = -2.5 \cdot \log(ADU) + ZP + 2.5Log(T)$ where *T=Effective Exposure Time*
  - **Effective Exposure Time**
  - **ZP Error**
- **Depth**[‡]: e.g. Limiting Magnitude in the specified Magnitude System, corresponding S/N
  - **Depth SN**
  - **Depth Aperture**
  - **Depth Error**

Imaging

- **S/N of the Spectrum** → Spectroscopy
- **Effective Gain**: in Electrons/ADU.
- **Scientific Validation Status**: closed list indicating the level of scientific validation (e.g. refereed journal/paper).
- **Paper Reference**: link to most prominent paper (e.g. *bibcode*[‡]).

[‡] Bibliographic code.

6.1.3 Service Metadata

Priority has to be given to IVOA standards [1].

- **InterfaceURL**: a URL pointing to a document that presents or describes the service interface.
- **BaseURL**: the base portion of a URL used to invoke a service with the expectation that an additional string must be appended for the service to execute properly. The syntax of the appended string is defined by the specific service.
- **HTTPResultsMIMEType**: the MIME type that is returned by a service.
- **StandardURI**: an identifier for a standard service. The syntax for identifiers is described in [2].
- **StandardURL**: a URL that points to a human-readable document that describes the standard upon which a service is based.
- **MaxSearchRadius**: service providers may choose to restrict the scope of searches done against their services, lest they be swamped with requests for millions or billions of results records.
- **MaxReturnRecords**: service providers may choose to restrict the number of records returned in order to avoid swamping the user with responses to an overly general query. If no value is provided, it is assumed that there is no restriction on the number of records returned.
- **Relationship**: a resource may be related to another resource in a way that is important to document, so that associated services or duplicate copies may easily be located.
    - *mirror-of* The resource is a mirror of another resource. Information gathered from the resources is indistinguishable.
    - *service-for* The resource is a service associated with a data collection.
    - *derived-from* The resource is a derivative of another resource, e.g., a subset selected for a particular scientific interest, or a reprocessed data collection.

- **RelationshipID**: the identifier of an associated resource. The relationship is described in the Relationship metadata element. The syntax for identifiers is described in [2].

7. Reliability

7.1 Robustness

In general, once operative, the registry will require a stable environment, from both H/W and S/W point of view.

RR-16 This means, the machine MUST be daily backed up.

RR-17 A second machine for failover MUST be available and kept in synch with the main one.

RR-18 The registry MUST scale with increasing usage.

7.2 Log Facility to Measure Usage and Monitoring

This feature - as well as the analysis feature – is essential to be able to better understand the system requirements, i.e. what the users want, and to tweak the system on the result of such an analysis.

RR-19 The capability to log access and usage of the system SHOULD be implemented.

RR-20 A mechanism for log analysis SHOULD be provided.

RR-21 An automatic self-test of the system itself SHALL be available.

8. Users

As in most of the systems, there will be different users interacting in various ways with the system itself.

This depends on the type of the user as well as on the kind of information required.
For this reason, different users will have distinct needs resulting in multiple set of requirements: an overlap exist for some functionality.
See diagram in Appendix A.

8.1 Type of Users

Three main categories of users have been identified.
In the future, it may be necessary to update the list, according to new needs.

8.1.1 Operators

Query, browse (maybe this can be done directly through the db interface), update, backup.
This is the only internal user, and the only one that can modify the repository.

8.1.2 External Users

This category includes generic users - ESO archive users, too.
They need to be able to query the registry as well as to submit new data products for the inclusion in the registry. They also need to update metadata of resources already included in the registry.

8.1.3 VO Registries

Other VO registries may harvest public resource metadata using IVOA standard protocols (e.g. currently agreed standard for harvesting is OAI).

9. Functional Integration

9.1 Required interfaces and their capabilities: for populating, searching, access privileges

According to the type of users (Sec. 5) there will be different interfaces offering the required functionalities.

9.1.1 Operator Interface

The main task of this I/F is to allow the administration of the registry. This includes the monitoring of the activities, and the creation of usage statistics (useful for further tuning of the system itself).
This I/F will also serve batch functionalities, e.g. for submitting a long list of new resources to the registry or for harvesting other registries.

RR-22 There MUST an operator interface for maintaining resource metadata.

9.1.2 Middleware

RR-23 An interface for programmatic access to the stored data MUST be provided.

9.1.3 VO Harvesting Interface

RR-24 An interface for harvesting stored resource metadata, using IVOA agreed standard, MUST be provided.

10. Documentation

RR-25 All systems SHALL be documented.

The “Policy & Procedures” document will describe what kinds of responsibilities the various involved parties (e.g. data providers) will have.

Appendix A

Diagram showing registry-users interactions: