Astronomy and Computing 7–8 (2014) 101–107

Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom

Full length article

The virtual observatory registry

M. Demleitner[a,*], G. Greene[b], P. Le Sidaner[c], R.L. Plante[d]

[a] Universität Heidelberg, Astronomisches Rechen-Institut, Mönchhofstraße 12-14, 69120 Heidelberg, Germany
[b] Space Telescope Science Institute, 3700 San Martin Dr, Baltimore, MD 21218, USA
[c] VOParis/Observatoire de Paris, 61 Av de l'Observatoire, 74014 Paris, France
[d] National Center for Supercomputing Applications, University of Illinois, 1205 W. Clark St., Urbana, IL 61821, USA

ARTICLE INFO

Article history:
Received 30 April 2014
Received in revised form
7 July 2014
Accepted 7 July 2014
Available online 19 July 2014

Keywords:
Virtual observatory
Registry
Standards

ABSTRACT

In the Virtual Observatory (VO), the Registry provides the mechanism with which users and applications discover and select resources – typically, data and services – that are relevant for a particular scientific problem. Even though the VO adopted technologies in particular from the bibliographic community where available, building the Registry system involved a major standardisation effort, involving about a dozen interdependent standard texts. This paper discusses the server-side aspects of the standards and their application, as regards the functional components (registries), the resource records in both format and content, the exchange of resource records between registries (harvesting), as well as the creation and management of the identifiers used in the system based on the notion of authorities. Registry record authors, registry operators or even advanced users thus receive a big picture serving as a guideline through the body of relevant standard texts. To complete this picture, we also mention common usage patterns and open issues as appropriate.

1. Introduction

The Virtual Observatory (VO) is a distributed system—by design, there is no central node either running services, delivering data, or even just a single link list-style directory. In order to still maintain the appearance of a single, integrated information system, users and clients must have a means of discovering metadata of VO-compliant resources (in the sense discussed in Section 3). This means is provided by the VO Registry.[1]

Following the VO philosophy, the VO Registry is not a single, central system but rather a network of several types of services, some of which host and publish metadata collections, while others provide capabilities for querying such collections. All follow standard protocols for exchanging information between them and between them and client software.

The VO Registry is governed by a fairly large set of standards; one of the goals of this paper is to review this body of text and discuss how each standard fits into the architecture. Anticipating some terms that will be explained later, let us collect and arrange the relevant standards already in the introduction.[2] Where the standards have short names in common use in the VO community, we introduce these here and refer to the standards by their mnemonic names in the following.

- *IVOA Identifiers* (Plante et al., 2007) lays out how resources and resource records in the VO are referenced.
- *Resource Metadata for the Virtual Observatory* (*RM* for short; Hanisch, 2007) specifies what entities need descriptions in the VO and what pieces of metadata these should contain to satisfy the VO's use cases.
- *VOResource* (Plante et al., 2008) lays out the basics of encoding resource metadata information as specified in *RM* in XML and defines the basic types. When we talk about *VOResource* in the following, we usually mean not only (Plante et al., 2008) but also the registry extensions introduced next.
- Several *Registry extensions* apply the building blocks from *VOResource* to more specialised types of services or interfaces. All of these combine a definition of the metadata as well as its XML serialisation.

* Corresponding author. Tel.: +49 6221541837.
E-mail address: msdemlei@ari.uni-heidelberg.de (M. Demleitner).

1 Written in upper case in the following, the term "Registry" refers to the entire system, as opposed to the lower-case "registry", which denotes a concrete service.

2 For an even bigger picture of the VO and its components, see Arviset and Gaudet (2010).

http://dx.doi.org/10.1016/j.ascom.2014.07.001
2213-1337/

  – *VODataService* (Plante et al., 2010) defines extra metadata to describe data collections and services exposing them; in particular, this concerns table and column metadata as well as metadata on service parameters.
  – *SimpleDALRegExt* (Plante et al., 2012) defines what extra metadata applies to services implementing several "simple" protocols of the VO's Data Access Layer (DAL).
  – *TAPRegExt* (Demleitner et al., 2012) defines what extra metadata applies to services implementing the Table Access Protocol TAP.
  – *StandardsRegExt* (Harrison et al., 2012) contains resource types for standard texts and thus defines how standards can be referenced, e.g., when declaring protocol support.
- *Registry Interfaces* (Benson et al., 2009) specifies how registries exchange the XML records defined in *VOResource* and extensions. It also contains a Registry extension for the services implementing Registry services themselves. Furthermore, its current version also defines two APIs for registry clients; in a forthcoming version, these APIs will be dropped.
- *Registry Interfaces* re-uses the non-VO *OAI-PMH* (Various, 2002) standard. This Protocol for Metadata Harvesting defined by the Open Archives Initiative governs the interactions of the registries among themselves. Its use by the VO is subject to several idiosyncrasies laid out in *Registry Interfaces*.
- *RegTAP* (Demleitner et al., 2014) defines how registry users can query the Registry's data content using IVOA's Table Access Protocol. An alternative, parameter-based API is currently being designed. We defer the discussion of the client APIs to a forthcoming article.

In the remainder of this paper, we will first delineate the Registry's role in the VO and outline its scope (Section 2), before establishing some basic notions on the relation between resources and their descriptions as the VO treats it in Section 3. Having thus introduced the concept of a resource record, in Section 4 we proceed to discuss how registries maintain collections of them. Section 5 explains the process of transmission and dissemination of the records and the separation of responsibilities in this process, as well as a common implementation error that has long plagued the Registry. The VO's way to generate globally unique identifiers as required by the harvesting protocol is then considered in Section 6.

With the basic architecture described, we proceed to discuss the current Registry content in Section 7, in particular as regards what resource records are contained. This provides some insight into the data model underlying the Registry. For the most relevant case where the resources described are services, special care must be taken in the description of "capabilities", i.e., facilities that operate on a client's behalf. We give an overview of these capabilities in Section 8. Finally, we briefly touch the issue of the validation of services and their descriptions in Section 9.

2. Scope

The Registry's role in the VO primarily is resource discovery. Hence, it must collect data sufficient to answer requests at least of the following types (or their combination):

- Resources of type X (as in: image service, database service, etc.),
- Resources on topic X (defined through keywords or via a full text search in the resource descriptions),
- Resources with physics X (defined through waveband, observables, queriable phenomena, etc.),
- Resources by author(s) X,
- Resources suitable for use X,
- Resources with spatial or temporal coverage X.

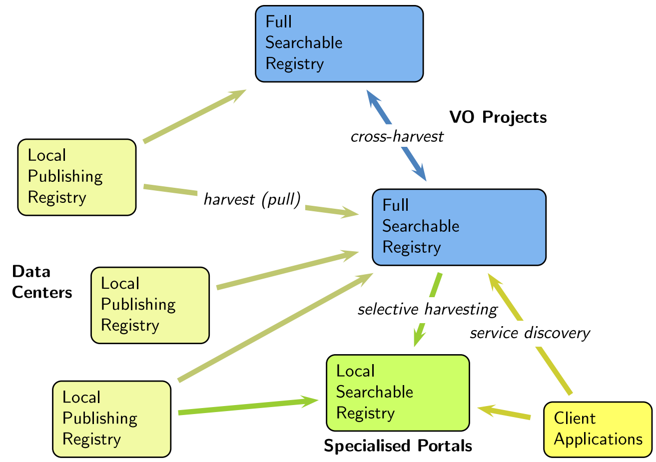

Fig. 1. A sketch of the registry system in the Virtual Observatory as laid out by Plante and Greene (2008): searchable registries harvest from publishing registries operated by the data providers. Users and client applications can then discover VO resources through queries to a searchable registry, either a full searchable registry that contains everything known to the VO, or a specialised one focused on a particular subset.

Once a resource record has been located by any of these constraints, it provides sufficient information at least to let users

- Assess suitability of the resource for purpose X,
- Access the resource,
- Identify who to credit for results obtained using the resource,
- Contact technical support for the resource.

The VO Registry is also used to monitor the health and functionality of the VO. The registries themselves are routinely validated and curated to ensure consistency with IVOA standards, which uncovers errors in the metadata supplied by the service operators. Even more importantly, services within the Registry are validated to comply to the standards they claim to implement, and registry records, where necessary, contain test input parameters suitable for exercising a service.

The Registry as such is *not* a mechanism of data preservation, and it does not provide persistent identifiers. The identifiers within the VO Registry, the IVORNs, are simple URIs with a scheme of `ivo`, an authority part as discussed in Section 6, and a local part governed by some reasonable restrictions on which characters are allowed to occur.

They can be resolved to resource records and, if applicable, access URLs by searchable registry and thus, not unlike DOIs (ISO Technical Committee 46, 2012), introduce a level of indirection between a service identifier and its access URLs. However, the indirection in the Registry mainly is a side effect of the requirement to provide rich, structured metadata for the services.

Unlike with DOIs, an operator is free at any time to discard identifiers, and the current VO infrastructure would stop resolving it on a short timescale. The conceptual reason why IVORNs as such are not suitable as persistent identifiers is that, as laid out in Section 3, they are in the first place identifiers of the resource records. Although the VO Registry could be exploited as a basis for (external) data preservation services and persistent identifiers for resources–Accomazzi (2011) reports on one such effort –, it does not in itself provide such facilities.

3. Resources and resource records

The Virtual Observatory can be seen as a collection of *resources*. Hanisch (2007) defines a VO resource as a "VO element that can be described in terms of who curates or maintains it and which can be given a name and a unique identifier". He goes on to name sky