



CADC Resources in the NVO Registry

Registered Resources

The Canadian Astronomy Data Centre (CADDC) has added resources to the NVO Registry using the NCSA Registry Portal. This report outlines the decisions and strategy we followed in organising our resources, followed by a few comments on the the experience.

Authority

The CADDC Authority resource defines the Authority ID under which all other resources are declared. The identifier is `ivo://cadc.nrc.ca`.

Organisation

The CADDC is registered as a organisation. The group leader is the registered contact person.

DataCollection

We have registered resources of type DataCollection for four archives: CFHT, GSA, HST, JCMT. The WFPC2 Associations project is also registered as a DataCollection. In each case, the archive scientist is registered as the contact person. The goal of creating these resources is to store descriptive metadata for the content here and to not duplicate this information in the description of services (current and future).

The level of granularity we have used here is adopted to keep the number of resources to a minimum. The downside is that a considerable portion of the useful metadata cannot be supplied because it varies across instruments within the archive. The alternative would be to register one data collection per archive-instrument combination, and in some cases even subdivide these by instrument mode/configuration when such factors change the registry metadata. It was deemed that this fine level of granularity would create too many resource declarations that were only marginally more useful. A middle ground might be to subdivide archive-level data collections into types (imaging and spectroscopy, for example) since some of the simple VO service interfaces make such a distinction.

Project

We have also registered Canadian Virtual Observatory (CVO) project as a resource of

type Project. On the surface, this appears slightly odd since the CVO currently presents several web applications with documentation. As a project, CVO is an experiment in the architecture of interacting VO services and a high-level data model that ties several levels of services together. The current web applications are registered separately (see below).

BrowserBasedService

For each data collection, we have registered the main science web form as a `BrowserBasedService`. These services are linked to the data collection using two optional components: `Relationship` (service-for) and `RelatedResource` (the appropriate data collection). The archive scientist is the declared contact person for the science web form.

The CVO web applications are registered; the project technical lead is the registered contact person.

In theory, one would like to add and edit all the characterisation information to the data collection and have a fairly spartan service description. The registry portal at NCSA allows you to create the service resource using the data collection as a template, thus populating the characterisation metadata, but does not appear to automatically keep optional metadata synchronised. Also, it is feasible for users of the registry to follow the relationship from a service resource to the data collection resource to gather this extra information, but it is unknown if this is standard or recommended practice at this time.

SIAService

SIAP services for three of our data collections (HST, CFHT, JCMT) have recently been added. The first two are registered as cutout services and the last one as a pointed observation archive (simply because the images are quite small). We have declared the service-for relationship to connect these services to the underlying data collection resource. As mentioned above for the browser based services, we would prefer to keep content metadata in the data collection resource. However, in the case of SIAP services, only a portion of the collection is accessible via this service interface (at most, only the imaging data – in practice, only the imaging data which is sufficiently processed and characterised) so it is inevitable that service metadata will differ from that of the underlying data collection; generally, it should be possible to make the service metadata more complete.

Comments

The resource metadata includes optional things like spatial, spectral, and temporal coverage and sampling. With our approach, we leave the instrument blank and try to enter "typical" values, which becomes almost meaningless and useless when the

archive contains imaging and spectra and, as is the case in the real world, the lineup of instruments in operation change over time. Thus, the natural granularity of registry resources appears to be more suited to the instrument level. However, this means we would have several logical data collections per archive (say 20 collections instead of the current 4-5). As we add services (browser, SIAP, SSAP, SkyNode) to each, we will reach a rather unmanageable 50-100 resource entries fairly quickly.

Notes

The NVO Registry contains ConeSearch and SIAP resources for CNOC1 which show up in some queries. While these are served by CADC, they are not under the `ivo://cadc.nrc.ca` authority ID. Some action needs to be taken (remove them or put them under the standard authority ID and support them).

Experiences

1. I used the NCSA registry portal to create authority, organisation, data collection, project, and service resources. This site was the easiest to use to create sensible resource entries. The ability to create a new resource using an existing one as a template (by having it selected before clicking the "Add Resource" button) saved a lot of tedious typing. Given the single web-page view, it is hard to maintain a degree of consistency without a good memory or having planned ahead and opened two browser windows/tabs. (Suggestion: Add a link in the "Resource List" page so the user can view a resource; with a link, this can be in a separate window at the user's discretion. Alternatively, use links instead of buttons to submit the form and then the user can easily edit in a new window and keep the list for viewing other resources at the same time.)
2. It took about one day to create, edit, and make the 15 resources somewhat consistent. In future, it should take 15 minutes to add a new service resource for an existing data collection resource. Adding a data collection will take longer since it should contain a more detailed description.
3. Normally, searching for CADC in the searchable registry web page at STScI yields most of the records we have added. Also, a small bug in the NCSA portal caused a "CADC Registry" to be created. This was clarified by Ray Plante and Ramon Williamson very quickly and the resource was deleted. For some reason, the CVO project resource (type Project) never appears when searching for CVO, but the two browser-based service resources do appear.
4. Soon after making some changes at the NCSA registry portal, searches for CADC and CVO at the STScI registry search page show no results under the `ivo://cadc.nrc.ca` authority ID (except for the deleted registry resource). This could be due to the way harvesting works (old records removed, new ones added later), but it is disconcerting that making small changes to a resource description causes them to disappear for some time. (Details: This occurred at approximately 9:30am

PST on Feb 4, 2005. I made changes at about 9 and until ~10am I could find nothing. Then at about 10:10am, searching for CADC started to show results again.)

4. None of the resources appear when searching for CADC or CVO in the Caltech registry. I assumed that the three sites harvest from each other. I have not contacted Ray Plante or Matthew Graham about this yet.
5. For the registered SIAP services, the editing page on the NCSA portal specifies some fields as mandatory. However, I successfully left some blank when I didn't know the value or it was not applicable (maximum query response size, for example).