Registry Session

Bologna, Tuesday May 09 2023 @14:00 CEST: Room 216 and online (Session 7)

Schedule: https://wiki.ivoa.net/twiki/bin/view/IVOA/InterOpMay2023Registry

Hendrik Heinl - Time Series in the VO -- how not to be seen
* Monty Python
* ObsCore worked to find light curves for TESS but not if you don't know what ObsCore is. So you naively search the Registry for TESS and find 141 tables. Add the keyword 'lightcurve' and get six from Vizier not providing the LCs themselves. Nor 'timeseries'. You have to know to specify datamodel='obscore', far frm ideal, and then can specify in a TAP query for timeseries.
* Repeat on TopCat. Not quite the same results, depends whether you search table descriptions or service descriptions. But still hard.
* Same for Gaia if you don't already know where to look or go to Google.
ESA site found on Google nicely explains Gaia LCs. Googling for TESS gets you MAST but not CADC (which has a good LC service.)
* Proposal: dataproduct_type goes in the Registry!

Markus Demleitner - Confessions of a Registry Janitor
MD runs a registry and has to keep it clean of 'lint' or 'turds'. Today talking 'lint', things that aren't pretty and make things not work well.
* Coverage is the number one thing needed for blind discovery. (Spatial and spectral.)
* Author names: last vs first, multiple word names, etc. Do use last comma first. Don't use citations or acknowledgements in the name fields.
* Contact info: avoid personal addresses but something that will persist at the institution.
* Subjects: Use the UAT vocabulary for machine-readable subjects!   (Unified Astronomy Thesaurus).
* Resource descriptions: should be useful to non-specialists and full-text searches. Minimum: What, From where, What for.
* Relationships: i.e., if there's a TAP service that serves this dataset, the entry should have a ServedBy relationship pointing to it. Don't use "related-to" anymore. Check relationship_type vocab.
* Tablesets: They are optional but should be mandatory. Tells you what the tables, columns, UCDs etc. in the Registry.
* Column descriptions: make useful for non-specialists. Let's work on column footnotes in VOTables.
* UCDs: use them!! and 1+ versions (check with astropy or TopCat)
* Long version of talk at http://blog.g-vo.org/registry-a-janitor-speaks-out.html
* Q: Vizier puts references
* Q: Footnotes? A: see after session.

Markus Demleitner - VOResource-1.2 - new features
* Most basic metadata element in the VO.
* UATs: 1.1 says terms should be from UAT, but that's a problem because terms are identified by long numeric URIs that we dn't want in our subject fields. Last July clarified it in Endorsed Note 2022-07-22 says what to do: "53-persei-stars" etc.
* Please update your subjects from UAT
* URI updates: referenceURI now required to start with 'http'. And rights now from SPDX (machine-readable identifier for data license, SPDX is a license registry)
* altIdentifier: More of them needed. Relationships to VO resources are easy but what about to non-VO resources with a DOI? E.g., <relatedResource>astropy</relatedResource>, where's the DOI? Add as "relatedResource altIdentifier="doi:...">astropy</> .
* BUT: VOResource 1.1 has had altId for creator and contact (for ORCIDs). Let's deprecate it since nobody has used them?
* DOIs: 1.2 will allow "doi" as an attribute to add to content/source in addition to bibcode. (ADS trying to move away from bibcodes.)
* Other? Protest? Submitting the PR after this Interop.
* Q: UAT not helpful for planetary. A: Fix it! Q: Fair, but will take a while. What should we do in the meantime? A: Propose to the UAT, predict what it's going to be, and it'll likely work. Better than nothing.
* Q: We don't use SPDX license list, but from Creative Commons URI. Not going to change that for VO. SPDX isn't the authority. A: Dang. Email registry email list and we'll discuss.
 * Q: In re relationship_type, adding doi, thanks. But you mentioned cites... <notetaker didn't understand>
 * Q: In re UAT terms in subjects. Is anybody machine-reading these in practice yet? A: In principle. Important elsewhere.
 * Q: Would this solve Hendrik's issue with timeseries discoverability? A: No. That doesn't go in subject.

Renaud Savalle - StandardRegExt -1.1
* Extension of VOResource XML to register the standards documents themselves.
* Tags: Standard, ServiceStandard, and StandardKeyEnumeration
* Look at ivoa.net/std/hips. It has keys with names hipslist-1.0 and hips-1.0. IRSA declares its service as standardID="ivo://ivoa.net/std/hips#hipslist-1.0". Nice. But in October MD identified four issues.
* 1. How to register Endorsed Notes? Add statuses "en" and "pen" to EndorsedVersion alongside "rec", "pr", etc.
* 2. Which standards to register? Only those that make use of identifiers "to" or "into" (the hash # above is an "into").
* 3. Who writes StandardsRegExt records? Editors of the standards should when they submit the first Working Draft. See ivoatex doc section 3.10.
* 4. What to do with StandardKeyEnumeration? Deprecate it in favor of Controled Vocabularies. Never was used, and CVs now better, maintained by Semantics WG, updatable, machine-readable. Keep those already existing.

* 5. Who writes the standard keys? See Section 4 of the proposed update.
Basically when you write your standard, you write how the keys should be
updated.
* Next step: see github, going for WG after this Interop, go read and contribute.
* Q: Do we remove the IWD (internal working draft)? A: It would make sense
since it's like an Endorsed Note. ...

All - General Discussion
Renaud's suggested topics:
1.Deprecate Registry Interface 1.0 client search functionality
Mark: don't think anybody is using them. No objections to dropping it.
Markus: look at how TopCat uses Registry. Or PyVO. Don't want another
backend. Wrap the relational backend like PyVO does instead of maintaining
another one.
TomD: The STScI registry does have a Solr-based web interface:
https://vao.stsci.edu/keyword-search/
Mark: Switch them off until the next Interop and revisit.

2.How to bring RegTAP query implementations up-to-date (MOC usage in ADQL
geometric queries)
E.g., want this query to work in all RegTAPs "... WHERE 1=contains(MOC('6/2623
2666-2667'), coverage)".
GAVO: it works with PostgreSQL+pgsphere; long term support?
ESAC: VOLTT fork? Plan to upgrade to RegTAP 1.2 in the next year. Not sure
what needed yet, but planned.
NAVO: Won't commit, but curious to explore more MOC usage. Not currently
using pgsphere, so it may have to wait. Haven't yet committed resources to it. We
lobby for it quarterly.

3-4.Curation and validation of the Registry
Meeting earlier this month.

Thanks to Markus for the harvest trigger service. Will discuss more next Interop's
Registry session.

Q: How to find out how best to validate, latest validator versions?
Ops: We talked about this but it changes often. Registry not the best place for
that. Web page. Human readable =>
https://wiki.ivoa.net/twiki/bin/view/IVOA/IvoaValidatorsSummary