

## 2024-11-16 Registry session notes

### Christophe Arviset on statistics per country of registered protocols and services

- \* note that it's quantity not quality being measured
- \* 29% from Managed Authority outside IVOA
- \* E.g., registries (publishing) in Belgium and Poland, neither in IVOA
- \* Remember it is also a function of how the resource is registered, whether one service for many datasets or many for many.
- \* Same issue for HiPS, some questions over numbers.
- \* So there are countries publishing data but not members (Czechia and Belgium), and members who don't publish many resources (India, but remember caveats above that it could be one serving many).
- \* Question: for one to many vs many to many, is there a recommendation.
- \* Markus: Register datasets not services. Good arguments for collective services like ObsCore. But if more of this info is in the Registry, that would be ideal. And move from specifying and searching on service type to product type.
- \* Comparing publishing registry stats? Difficult.

### Robert Nikkuta on publishing registry at NOIRLab

- \* Followup of May 2023 Bologna when we asked what to do, now showing what we have done.
- \* Motivated by knowledge drain and evolution of development technology (languages, ci/cd, tooling)
- \* Got good support from IVOA
- \* Python code (OAI PMH part), FastAPI, Terraform infrastructure code, CI on Gitlab. Also CI to automate XML creation.
- \* Buy in with dev staff and management was crucial (and sometimes difficult)
- \* <https://gitlab.com/nsf-noirlab/csdc/vo-services> has our code and templates for yours
- \* Internal service that asks "what services do you have", takes response, converts to XML, triggers MR, human approval, then deployment
- \* First iteration focused on Astro Data Lab services
- \* Lots still to do
- \* MRs welcome, help available.
  
- \* Theresa: Excited! Huge challenge at ST is moving our OAI interface to Python. She's been sketching out FastAPI, messing with Terraform. But lack of resources of course, will def look at this.
- \* Joshua: Wrote a package to consume records and generate XML (because he hates XML) and would like help testing the round-tripping
- \* Q: we have a TAP and cone, used in a viewer tool. Not registered. What/how should we do this?
- \* Robert: explains
- \* Q Gilles: other formats enabled? Robert says no reason we couldn't add that given time

\* Renaud: OAI PMH library from something existing? No, was redeveloped, not that hard.

## Markus Demleitner on new features of VOResource 1.2

- \* Reminder: general base type for different things in the Registry.
- \* Maintenance release mostly. altIdentifier, DOIs, etc.
- \* How to add DOIs on our resources? Nobody wants to use IVO IDs. Or ORCIDs. To link to external resources.
- \* Could add to vr:ResourceName? Then applies to contributor, facility, instrument, managingOrg, name, publisher. So let's add it there.
- \* Attribute or element? Prefer attribute for ... reasons.
- \* If you need multiple altIdentifier, repeat
- \* creator/name also vr:ResourceName. But not used so deprecate.
- \* DOIs as sources: vr:Source has @format attribute. It can have bespoke values, recommended to be bibcode, but you can also give DOI
- \* Format? <http://doi.org/foo.bar> or doi:foo.bar ? Currently favor latter.
- \* content/subject keywords used to be freeform, but let's use UAT. Now a SHOULD. "Even Elsevier is using them."
- \* SPDX License URIs? rights/@rightsURI to be machine readable, so you can look for "only CC0 data". SPDX is the spec the open source community is using, so let's use that.
- \* Need reference usage, so please try it out and let us know during RFC.
- \* Q: somebody said three days ago DOI released a formal doc explaining how to specify them.
- \* Ann: value URIs? Reduces ambiguity, human error. Markus said discussed previously in an endorsed note. Notetaker TJ didn't follow.
- \* Gilles: add RORID to list of things that can be specified? Ann: if the string should take a machine to somewhere else, you need the URI, but if it's for humans, use the human readable format.
- \* Joshua: SPDX identifier, reference, or detailed URL? M: look at what Redhat uses.
- \* Sebastien: what do registry folks think of UAT terms usability, using them to make suggestions to users. M: yeah, surprisingly easy to do.
- \* James: is there a way to say this is proprietary for 18months then public? M: no. No discovery use case for that. Ann: I have one in theory. Interaction of public dates with proposal deadlines.
- \* Gilles: possibility for users to put something in the Registry, tag with some kind of creation level? M: there's already enough crap, so let's not let scientists add to it.

## Markus on how to write a resource people will find

- \* Giving my usual lecture, I wanted to find something and the same string worked in some clients and not others. Why?
- \* Look at RegTAP data model, there are multiple places such keywords might appear, for different search use cases.
- \* Because different clients search different parts of the data model. (Aside: TopCat uses Glots, which is only still around because some providers not yet providing tablesets.)

\* Take away? Make sure your acronym and its expansion both appear in the resource title and in the table description (for TAP).

\* Would be much more useful to search by physics by UTCs, TAT, and coverage. So please add these!

\* Currently this doesn't give you many, but it gives you the higher quality services.

\* Homework: catalog vs catalogue. Need synonym dictionaries, which Postgres at least can use. Who would maintain one? I have a few pairs.

\* Manon: add wild cards instead of synonyms? M: prefer synonyms.

\* Robert: spec require english? Otherwise synonyms gets harder. M: title, description etc. must be in English. Otherwise this cannot work. However, in DocRegExt dealing with tutorials, allows localized non-English fields. So this wouldn't work there. R: maybe someday AI/ML will solve this. M: not holding my breath.

\* Gilles: not a fan of synonyms. Difficult for users to query the registry, so users usually use a client that constructs the SQL for them. Why not use something like Elastic Search that the big search engines use? This is what ADS does. So catalog/ue is solved by resolved. M: we've tried, but we need more, e.g. relations (?). Pierre IS was doing that at one point.

\* G: following ADS with table\_name:ucd:string ? M: meh.

\* Ann: the RegTAP DM is trivial compared to planetary. NoSQL databases are really good at fishing expeditions like searching many places for text searches. M: Can you tell us more about that next interop? Ann: yes.

Renaud: next steps, roadmap

\* productTypeServed status and client updates.

\* TableReg NOTE to register obscure tables edited by Markus

\* DocRegExt edu resources, working draft to PR?

\* Update discovering data collections note for new CatalogResource.