

# IVOA Registry WG Running Meeting

2024-05-06 13:00 UTC

<https://cnrs.zoom.us/j/91455321058?pwd=ZHc3K3loOUpTWUJlOQzh6Y08wUXdDZz09>

## Participants

- RS Renaud Savalle PADC
- MD Markus Demleitner GAVO
- GL Gilles Landais CDS
- MM Manon Marchand CDS
- SD Sebastien Derriere CDS
- SN Sara Nieto ESAC
- TJ Tess Jaffe HEASARC
- Gregory Mantelet CDS
- Theresa Dower STSci
- Pierre Fernique CDS

## Topic: Linking capabilities with tablesets

Email discussion thread:

<http://mail.ivoa.net/pipermail/registry/2024-March/005572.html>

RS introduces the matter by describing VOResource record `ivo://CDS.VizieR/I/355` of type `CatalogService` from Vizier Pub Reg using the Oxygen XML editor. This records features capabilities and a tableset

GL presents the need to link a capability to a table and proposes a new `<servees>` element of `<capability>` referencing a table in the `<tableset>` to link a Simple Cone Search services to its underlying table

MD the Discovering Data Collection note has solves a similar issue for TAP with the "auxiliary". Discovery use case: GAIA + lightcurve: first find capabilities, then tables. See also PR <https://github.com/astropy/pyvo/pull/505> Another case: Resources from Gaia that SERVES Time Series ? How to know which table has time series in the tableset ? Attaching product type to table does not seem right. Interest is data discoverability for user: how to search, and what would the query would look like.

GL:

MD: product type should not (significantly) change between parts of one

VOResource to provide good user experience.

GL: Adding utype on table is possible, could we use that ?

MD: Yes. We could also attach coverage to tables but it will make discovery harder and its queries ugly. In our current architecture, a resource is a collection of things held together by the same metadata.

GL: Could we put capabilities into tables ?

MD: Such a solution would be a significant change to Registry data model, severe enough that we would have to change essentially all of our discovery interfaces. To keep within what we have, a Gaia data release needs to be represented in multiple resource records: gaia\_source vs the light curves, or the republished 2MASS.

MM: How to discover what tables can be joined together ?

MD: Look at the served-by relationship of the TAP service used to compute the join (but note we don't do that kind of discovery yet).

SD: SCS services have a 1-1 mapping with tables. One reason why have several tables (with unrelated data product type) in the same VOResource. is to factorize common metadata (curation, content kw) vs repeat it. If we have different VOResources per product type, searching for Gaia DR3 would return lots (50?) VOResources.

GL: Example of how to setup the VOResource correctly in our case ?

MD: Most VOR already have common metadata. Repetition is not necessary an issue! However, ranking resources matched during discovery, which would increase visibility of the Registry is one, but we don't have this yet. From a data discovery viewpoint, one should first discover data then the ways a given client would access access it: such discovery pattern is what we need!

GL my proposal would allow that

MD: Show me how TOPCAT / PyVO would work with your proposal

GL We are afraid of breaking the Vizier to Registry architecture, if we add a tableset in a separate resource, with the same metadata, including the DOI of the CATALOG which is for the WHOLE collection ! Our Reg is harvested by EUDAT B2FIND then EOSC: they will find the same DOI linked to several resources !

MD: You can use the DOI of the publication or the DOI of the catalog which you (CDS) register

GL: Danger is that EOSC could expose several entries for the same DOI and then merge them with other sources (ADS, DataCite etc). We don't know their merging algorithm, but they merge things.

MD: They should not merge such different information !

GL: We are not sure we can do that today... also not many discovery clients follow the relationships !?

MD: There is still the assumption 1 resource = 1 data collection with 1 primary table. Tables that can stand by itself (roughly: don't have a foreign key) SHOULD be in different resources.

MM: when should a legitimate join should done between tables which have different data products ?

MD: JOINABILITY is not the criteria for being in the same resource, the criteria should be "can I use them together", or are there describing different things ?

SD: Consider a single Vizier catalog with several tables which are related. Why would people need to explore several resources to find them ? Ex: a catalog having its data split into 2 tables, one for the northern hemisphere and one for the

southern one.

MD: It's a legal gray zone for Vizier, but I'd totally advocate making a single cone search for north and south; it's almost certainly what users would want.

SD: Our proposal helps the link dataset <-> capability, so why is it different ?

MD: Think of the API for using it first. We want to keep our APIs simple.

GL: What would be the impact of using small VOResources on current discovery clients ?

MD: None in current ones

GL: If we would go for smaller VOResources now they would be exceptions which would cause maintenance problems for us. So if we decide in the long term to split our resources , it will be for all Vizier catalogs. Sara do you have a solution at EASC ?

SN: In our resources we have several capabilities

GL A "global" Cone Search like TAP could help. Should we wait for a new version of SCS ? MD, do you prefer the SCS or the TAP global approach ?

MD: TAP ! SCS is the cause of the whole problem ! Evolving CS could be more realistic solution rather than to turn around discovery using your <serve> proposal. Fixing CS by allowing it to use several tables, or a (multiple) selection of tables, is less work than restructuring the Registry data model.

GL Nice but it does not solve "this table has a Time Series data". UCDs and kws are not sufficient for that.

MD: DataProductType belongs to the VOResource level and is on track to be added there. A vocabulary exists already <http://www.ivoa.net/rdf/product-type>. The current VODataService draft (<https://github.com/ivoa-std/VODataService>) features it (look for "productTypeServed") => Let's plan a talk at the upcoming Registry Session in Sydney with slides provided by ML about that (TJ might present it)

GL: also needed a title in <capability>...

MM: It already has a description.

MD: We could add a label too...

RS: Thanks to all participants and thanks to the speakers. To summarize, the one-to-one vs one-to-many mapping between Vizier catalogs and the Registry VOResources are not ideal but this design choice won't likely evolve in the short term at CDS. However it is useful for others and new publishers to know that smaller resources ease data discovery. In the meantime, CDS will explore the impact of their <serve> proposal on the discovery APIs to see how it impacts usability. See you all at Sydney Interop (I will attend remotely).