A Roadmap for Space-Time Discovery in the VO Registry

Version 1.0

IVOA Note 2018-01-15

Working group
: Registry

This version
: http://www.ivoa.net/documents/regstcnote/20180115

Latest version
: http://www.ivoa.net/documents/regstcnote

Previous versions
: This is the first public release

Author(s)
: Demleitner, M.

Editor(s)
: Demleitner, M.

Version Control
: Revision 4730, 2018-01-31 11:36:04 +0100 (Wed, 31 Jan 2018)
: https://volute.g-vo.org/svn/trunk/projects/registry/regstcnote/regstcnote.tex

Abstract

In VODataService 1.1, the footprint of resources in space, time, and the electromagnetic spectrum (the "coverage") is defined using STC-X. Takeup of this mechanism has been low, and searchable registry interfaces have, by and large, not made use of it. With MOCs, the Virtual Observatory now has a technology that allows efficient handling of spatial footprints. This note lays out a way to develop the Registry ecosystem to accomodate them and to replace STC-X in the representation of resource coverage on the time and spectral axes as well.

Status of this document

This is an IVOA Note expressing suggestions from and opinions of the authors. It is intended to share best practices, possible approaches, or other perspectives on interoperability with the Virtual Observatory. It should not be referenced or otherwise interpreted as a standard specification.

A list of current IVOA Recommendations and other technical documents can be found at http://www.ivoa.net/documents/.

Contents

1 Introduction

Many important use cases for the Registry in the Virtual Observatory (VO) involve matching a region of interest in space, time, and spectrum with the regions covered by the registered resources. Examples include:

- Where can I find data in the near infrared for sources within two degrees of the Galactic center?
- Where can I find photometry on the southern sky for the first half of the 20^{th} century?

- Where can I find X-ray spectra for objects in the Large Magellanic Cloud?

To enable such use cases in the VO Registry, searchable registries must contain coverage ("STC", space-time coordinates) information. In VODataService 1.1 (Plante and Stébé et al., 2010), the intent was to represent such information mainly in STC-X *ResourceProfile* elements. As of January 2018, only about 1000 resources out of the 18000 resources in the Registry contain a *ResourceProfile*, and of those, the overwhelming majority declare essentially trivial metadata (full sky without explicit limits on time and wavelength).

Standard searchable registries have, so far, not actually exposed this information; in particular, RegTAP (Demleitner and Harrison et al., 2014) has dropped an attempt to add tables for STC coverage early on in the standardisation effort, mainly over disputes about how spatial coverage should be modeled. Prototype implementations of these early efforts existed, but were largely not used. The fact that coverage information was available for less than 10% (for the space axes, much less for the other axes) of the VO's resources certainly contributed to the perceived lack of usefulness.

On the side of registries trying to process VODataService 1.1 coverage information, dealing with the STC-X embedded within the resource records was made difficult by the large feature set of STC-X, where coverages could be provided in a myriad of reference frames and shapes that needed to be unified to standard systems before they could meaningfully be searched. Additionally, STC-X itself never reached the Recommendation status, so supporting coverage in the Registry meant implementing against a non-standard.

Meanwhile, Fernique and Boch et al. (2014) have introduced MOCs, a powerful technology for representing spatial coverage information. It is successfully employed in several custom services, as well as, for instance, through the Aladin 10 (Fernique, 2017) VO client software. Many resources have started to give MOC-based information on their spatial coverage by employing the *footprint* element in VODataService. By harvesting the URLs transmitted in this way in addition to in-record STC-X, spatial coverage is now available for 13503 records out of 19463 in the Registry as of 2018-01-15 (78 resources have temporal, 66 spectral coverage, mostly from STC-X).

The solution of giving a footprint URL, however, is still unsatisfactory because

1. No temporal and only rough spectral coverage is available.

2. Searchable registries have to harvest the MOCs in a separate step, which means many extra requests in addition OAI-PMH-based harvesting; in practical experience, a significant number of these extra requests fail for a wide variety of reasons. This further complicates the operation of searchable registries.

3. Hence, the validity of a resource record with a footprint effectively depends on two artefacts: the resource record itself and the MOC referenced.

4. The updating architecture based on OAI-PMH is bypassed, leading to significant extra logic in maintaining up-to-date coverage information.

In this note we propose a modification of the VODataService schema (in section 2) and an extension to RegTAP (in section 3) that will facilitate VO-wide, standards-based dissemination and interrogation of STC coverage. We conclude with a collection of points that should be discussed during further standardisation, and a roadmap for the next steps in section 4.

2 Extending the VODataService Schema

2.1 Axes Modeled as Floating Point Intervals

We propose to model time and spectral coverage as a union of simple intervals over the real numbers. To avoid burdening the searchable registries with error-prone transformations – that often enough are actually impossible without information not conveyed within the registry records –, we fix reference systems and units.

Specifically, times are assumed to be given in Barycentric Dynamical Time (TDB) at the solar system barycenter. They must be specified as Modified Julian Dates. For now, we do not forsee discovery use cases that require a temporal resolution significantly below one hour. Hence, resource record authors are encouraged to pad their actual temporal coverage such that differences in time scales (of the order of seconds) or reference position (of the order of minutes between ground-based observatories and the barycenter) do not matter.

Following common VO practice, spectral limits must be given in meters of vacuum wavelength for the solar system barycenter. Again, discovery use cases on a level where the difference between reference frames of ground-based observatories versus the solar system barycenter matters are not forseen, and resource record authors are advised to pad their intervals on that level.

2.2 Spatial Coverage

Spatial coverage is conveyed as a MOC. To enable easy embedding into resource records written in VOResource (i.e., XML), we represent MOCs in the ASCII convention proposed in the MOC specification (Fernique and Boch et al., 2014). This representation is not normative so far, but it is in fairly wide use, and its proper standardisation should be a minor matter as part of the work proposed here.

The MOCs are to be understood in ICRS. Possible extensions to non-celestial coordinates are discussed in Sect. 4.1.

An important characteristic of MOCs is the order of the smallest scale (the “MOC resolution”). Higher orders yield more faithful representations of the actual coverage, but also lead to a possibly significant increase of the size of the serialised MOC. We suggest a “typical resolution” of the Registry of about a degree. If the resources in the Registry were uniformly distributed, that resolution would mean a selectivity of roughly 1:50000 (since there are roughly 50000 different patches with a diameter of a degree on the sky), which seems sufficient.

This translates into the recommendation to have MOCs within resource records at orders between 5 and 7. In a standard, this should be given as a recommendation, not as a requirement. Resources that really only cover a few compact sources could and probably should use higher resolutions.

At least within the 1.x series of VODataService, we additionally have to keep the FOOTPRINT element. It has been used to convey URLs of coverage MOCs in FITS serialisation for a while now, and some searchable Registries (for instance the Heidelberg RegTAP service) already harvest and interpret such footprints.

Although, as argued above, we do not believe footprint is a well-scaling and general solution to the problem at hand, it is perfectly conceivable that sepcialised applications or services want to harvest high-resolution MOCs. Hence, continued support for the FOOTPRINT element might be desirable even in the presence of a COVERAGE element in VOResource.

To distinguish MOC-based footprints from those giving footprints in other formats – there have footprint services returning STC-S strings, for instance –, services supplying high-resolution MOCs on an extra endpoint should add an ***ivo-id*** attribute valued

`ivo://ivoa.net/std/moc`

to their footprint elements.

On the timeframe of VODataService 2.0, we suggest use regular capabilities for the declaration footprint service metadata rather than the specialised FOOTPRINT element.

2.3 A VOResource Example

With the proposed VODataService schema (version tag 1.11), a full coverage element could look like this:

```
<coverage>
  <spatial>
```

```
    4/2068
    5/8263,8268-8269
    6/33045-33047,33049,33051,33069,33080-33081,
    33083,33104-33106,33112,33124-33126,33128-33130
  </spatial>
  <temporal>51845.1 52262.2</temporal>
  <spectral>3e-07 1.1e-06</spectral>
  <footprint ivo-id="ivo://ivoa.net/std/moc"
    >http://dc.zah.uni-heidelberg.de/cdfspect/q/ssa/coverage</footprint>
  <waveband>Optical</waveband>
</coverage>
```

2.4 Schema Changes

The changes to the VODataService schema necessary to implement the proposed scheme are limited to the VS:COVERAGE type. For a rough, first-order approximation to the new schema one can add the following elements after the current STCResourceProfile-typed child:

```
        <xs:element name="spatial" type="xs:string"
              minOccurs="0">
           <xs:annotation>
              <xs:documentation>
                 An ASCII-serialised MOC defining the spatial coverage
                 of the resource.
              </xs:documentation>
              <xs:documentation>
                 The MOC is to be understood in the ICRS reference frame.
                 Resources should give the coverage at least to order 6
                 (a resolution of about one degree). The order should be
                 chosen so as to keep the resulting MOC smaller than a few
                 dozens of kB.  If desired, a more precise MOC can be provided
                 on a dedicated endpoint declared in the footprint element.
              </xs:documentation>
           </xs:annotation>
         </xs:element>

<!-- TODO: make a FloatPair or FloatInterval or whatever type -->
         <xs:element name="temporal" type="xs:string" minOccurs="0"
              maxOccurs="unbounded">
           <xs:annotation>
              <xs:documentation>
                 A pair of lower, upper limits of a time interval
                 for which the resource offers data.
              </xs:documentation>
              <xs:documentation>
                 This is written as for VOTable tabledata (i.e.,
                 white-separated C-style floating point literals).
                 The limits must be given as MJD. While they
                 are not intended to be precise, they are to be understood
                 in TDB for the solar system barycenter. The total coverage
                 of the resource is the union of all such intervals.
```

```
      </xs:documentation>
    </xs:annotation>
  </xs:element>

  <xs:element name="spectral" type="xs:string" minOccurs="0"
      maxOccurs="unbounded">
    <xs:annotation>
      <xs:documentation>
        A pair of lower, upper limits of a spectral interval
        for which the resource offers data.
      </xs:documentation>
      <xs:documentation>
        This is written as for VOTable tabledata (i.e.,
        white-separated C-style floating point literals).
        The limits must be given in meters of vacuum wavelength,
        e.g., 655e-9 658e-9. While the limits are not intended
        to be precise, they are to be understood for the
        solar system barycenter.
      </xs:documentation>
    </xs:annotation>
  </xs:element>
```

While we probably do not want to include schema validation for ASCII MOCs, the floating-point intervals can easily be validated using XML schema.

A draft VODataService 1.2 schema (version attribute on the root element: 1.11) that includes these elements is part of this note and can be obtained from the IVOA document repository[1].

3 STC coverage in RegTAP

The mapping of the proposed VOResource extension to a relational model is fairly straightforward. Given the forseeable query patterns, any denormalisation does not seem warranted. The interval-valued axes thus map to tables that can have multiple rows per ivoid.

Since essentially arbitrary geometric operations between the spatial coverages and other ADQL-defined geometries are allowed, explicit support for MOCs or at least rather similar spatial representations is required in the database. Beta-level support for representing MOCs is available in recent releases of pgsphere (Chilingarian and Bartunov et al., 2004).

3.1 Relational Mapping

These considerations yield the following three tables:

[1] http://www.ivoa.net/documents/regstcnote/20180115/VODataService-v1.2.xsd

Table rr.stc_temporal All columns in this table should be indexed. The `ivoid` column should be an explicit foreign key into `rr.resource`. It is recommended to query this table using the `ivo_interval_overlaps` user defined function.

Column names, utypes, ADQL types, and descriptions for the `rr.stc_temporal` table

ivoid `xpath:/identifier`	char[*]	The parent resource.
time_start `stc:AstroCoordArea.TimeInterval.StartTime`	float	Lower limit of a time interval covered by the resource.
time_end `stc:AstroCoordArea.TimeInterval.EndTime`	float	Upper limit of a time interval covered by the resource.

Table rr.stc_spectral All columns in this table should be indexed. The `ivoid` column should be an explicit foreign key into `rr.resource`. It is recommended to query this table using the `ivo_interval_overlaps` user defined function.

Column names, utypes, ADQL types, and descriptions for the `rr.stc_spectral` table

ivoid `xpath:/identifier`	char[*]	The parent resource.
wavelength_start `stc:AstroCoordArea.SpectralInterval.LoLimit`	float	Lower limit of a vacuum wavelength interval covered by the resource (for the solar system barycenter).
wavelength_end `stc:AstroCoordArea.SpectralInterval.HiLimit`	float	Upper limit of a vacuum wavelength interval covered by the resource (for the solar system barycenter).

Table rr.stc_spatial All columns in this table should be indexed. The `coverage` is given as having a type `string` in the following table, but it is to be understood as a geometry suitable for the common ADQL geometry operators and operands (CONTAINS, INTERSECTS, CIRCLE, POLYGON, etc). Details on the serialisation of this column on output should be given by a future version of the IVOA DALI standard (**?**).

The column `ref_system_name` is present for extensability and must always be NULL (meaning ICRS BARYCENTER coordinates). Non-ICRS coverages are not yet supported. Still, clients must contrain any query against coverage with `ref_system_name IS NULL` to avoid false positives later. Non-ICRS systems will be denoted by strings drawn from a controlled vocabulary that is yet to be defined.

Column names, utypes, ADQL types, and descriptions for the `rr.stc_spatial` *table*		
ivoid `xpath:/identifier`	char[*]	The parent resource.
coverage	char[*]	A geometry representing the area a resource contains data for; this should be tight at least with a resolution of degrees.
ref_system_name	char[*]	The reference frame coverage is written in. This is currently reserved and fixed to NULL. Clients should always add a constraint to NULL for this to avoid matching non-celestial services later.

3.2 A user-defined function

In order to facilitate robust query patterns, searchable registries that host these tables must also provide a user defined function

$$\texttt{ivo\_interval\_overlaps} \quad (l_1 \text{ NUMERIC}, h_1 \text{ NUMERIC}, \\ l_2 \text{ NUMERIC}, h_2 \text{ NUMERIC})$$

that returns 1 if the intervals $[l_1 : h_1]$ and $[l_2 : h_2]$ overlap and 0 otherwise. This is defined to be equivalent to the expression

$$h_1 \geq l_2 \text{ and } h_2 \geq l_1$$

with the conventional boolean mapping of 1 as True and 0 as False.

3.3 Relationship to RegTAP

This extension could become part of RegTAP 1.2; this would in effect mean making it mandatory for all RegTAP searchable registries. In the interest of maximal interoperability and least surprise to clients this would probably be the most desirable way forward, although it would mean requiring rather extensive support for geometric operations in the backend database of all RegTAP services.

An alternative to globally requiring the three tables would be to make them optional; the presence of the `TAP_SCHEMA` (and tablesets) would let clients determine support for the coverage tables (and potentially discovery supporting searchable registries).

A last possibility would be to define the three tables in a separate standard and define an extra data model, to be used independently in discovery.

3.4 Sample Queries

Here are RegTAP queries fulfilling the use cases mentioned in the introduction using the three proposed tables. We only select ivoids in the following examples – the actual discovery of interface URLs or other pieces of resource metadata proceeds as normal in RegTAP.

1. Where can I find data in the near infrared for sources withing two degrees of the Galactic center?

```
SELECT ivoid
FROM stc_spatial
NATURAL JOIN stc_spectral
WHERE
        1=INTERSECTS(
                CIRCLE('', 266.5 -29., 2),
                coverage)
        AND ref_system_name IS NULL
        AND ivo_interval_overlaps(
                wavelength_start, wavelength_end,
                4e-6, 8e-7)
```

2. Where can I find photometry on the southern sky for the first half of the 20th century?

```
SELECT ivoid
FROM rr.stc_spatial
NATURAL JOIN rr.stc_temporal
NATURAL JOIN rr.table_columns
WHERE
        1=INTERSECTS(
                CIRCLE('', 0, -90, 90),
                coverage)
        AND ref_system_name IS NULL
        AND ivo_interval_overlaps(
                time_start, time_end,
                15000, 34000)
        AND ucd LIKE 'phot.mag%'
```

3. Where can I find X-ray data for the center of the Large Magellanic Cloud?

```
SELECT ivoid
FROM rr.stc_spatial
NATURAL JOIN rr.stc_temporal
WHERE
        1=CONTAINS(
                POINT('', 80.9, -69.8),
                coverage)
        AND ref_system_name IS NULL
        AND ivo_interval_overlaps(
                wavelength_start, wavelength_end,
                1e-8, 1e-11)
```

Note that as of early 2018, the prototype at http://dc.g-vo.org/tap cannot yet intersect circles and MOCs. Hence, the first two queries do not yet work on that service as they are given here; use the pattern from query three for experiments.

4 A Roadmap

4.1 Questions to be Discussed

Should we include further axes? The STC-X *ResourceProfile* element allows the definition of coverage in redshift space. We could not identify clear discovery cases that would require distance information, although "resources on nearby stars" or "resources on distant galaxy clusters" might count. However, to satisfy the first one, redshifts will not work as a proxy for distance, whereas using physical distances on a cosmological scale involves prescribing models. It is not clear whether a model for distance information in astronomy exists that enables sufficiently many discovery scenarios and still admits simple query patterns.

Other reference systems? We currently require all spatial coordinates to be in the ICRS. This obviously is not enough whenever objects move fast against the ICRS, as for instance for solar system objects and, in particular, surface features and the like. To enable future extensions to these domains, a column `ref_system_name` must currently always be filled with `NULL` on the service side, and clients must always constrain coverage queries with a `ref_system_name IS NULL` condition.

Is this enough to cover forseeable and plausible use cases? Should we write `'ICRS'` rather than `NULL` already, and then perhaps already define some system names we already have resources for? Given it will be present in almost all STC queries, should we have a less verbose name than `ref_system_name`?

Non-electromagnetic coverage? With the advent of neutrino and gravitational wave observatories, the location in the electromagnetic spectrum does not necessarily characterise observations any more even for conventional celestial observations. The diversity even increases when observations of charged primary particles are included. Should the VO Registry enable Metadata discovery for such resources? If so, how? We remark that moving to energy instead of wavelength in the declaration of spectral coverage could already go a long way here.

MOC representation in VOTables This draft suggests MOCs should be serialised to the (currently non-standard) ASCII form when they appear in VOResource documents. Below, we propose to follow this practice when packing MOCs into VOTable cells. The current standard representation, on the other hand, is essentially an array of integers with some metadata not actually necessary for decoding. It would be almost trivial to port this convention to VOTables. Doing so might yield some space savings when using binary coding as well as more straightforward parsing at the expense of generally less compact representations in TABLEDATA and less human readability.

4.2 Next Steps

The scheme proposed here is implemented by DaCHS (Demleitner and Neves et al., 2014) on the side of resource record production and by the Heidelberg RegTAP mirror[2] on the side of searchable registries. As of January 2018, no client uses the new STC tables, although Aladin 10's registry interface (Fernique, 2017) has global resource discovery based on spatial constraints using a scheme fairly similar to `rr.stc_spatial`.

To refine this proposal and to update the related standards, we believe the following steps need to be taken:

Software support for MOCs in pgsphere Current betas of pgsphere know how to match points against MOCs. To avoid surprising behaviour, MOCs should behave in ADQL like POLYGONs and CIRCLEs, in particular in INTERSECTS and OVERLAPS, ideally also in AREA, CENTROID and DISTANCE (though the latter are perhaps not quite as central for RegTAP). This is currently being worked on in Heidelberg.

VODataService 1.2 The addition to the VODataService schema will necessitate a new version of the VODataService standard. Several other minor updates to VODataService need to be undertaken anyway, in particular in order to improve alignment with VOSI.

Standardising ASCII MOCs If MOCs are indeed to be represented in VODataService instance documents, it is highly desirable to standardise a representation beyond FITS[3]. As the current MOC specification already proposes a usable ASCII serialisation, this could happen in a very minor update to the MOC specification, but see the next paragraph.

[2] http://dc.g-vo.org/tap

[3] It would, of course, be possible to include base64-encoded FITS bytestreams in the XML documents. While this might be marginally acceptable in VOResource documents, it seems the overhead of encoding large amounts of whitespaces and null bytes is excessive when many MOCs are serialised into VOTables with a MOC-valued column.

Defining MOC representation in VOTable cells For visualisation and possibly more advanced filtering, it is highly desirable that clients can use the `coverage` column of `stc_spatial` in select clauses. This means that we need a representation of MOCs at least in VOTables. The ASCII serialisation proposed in the MOC specification would suffice, but for unambiguous identification of MOC literals, a VOTable *xtype* needs to be defined. This would probably require a new version of DALI. For standardisation convenience, this update to DALI could then also provide the normative text for the ASCII serialisation of the MOCs themselves.

This option needs to be weighed against the alternative as discussed in Sect. 4.1: Using arrays of NUNIQs analogous to the current FITS-based MOC serialisation.

Updating RegTAP As discussed in sect. 3.3, the new RegTAP tables would preferably come in an update to RegTAP.

A Changes from Previous Versions

No previous versions yet.

References

Chilingarian, I., Bartunov, O., Richter, J. and Sigaev, T. (2004), PostgreSQL: the suitable DBMS solution for astronomy and astrophysics, *in* F. Ochsenbein, M. G. Allen and D. Egret, eds, 'Astronomical Data Analysis Software and Systems (ADASS) XIII', Vol. 314 of *Astronomical Society of the Pacific Conference Series*, p. 225.
http://adsabs.harvard.edu/abs/2004ASPC..314..225C

Demleitner, M., Harrison, P., Molinaro, M., Greene, G., Dower, T. and Perdikeas, M. (2014), 'IVOA Registry Relational Schema Version 1.0', IVOA Recommendation 08 December 2014, arXiv:1510.02275.
http://adsabs.harvard.edu/abs/2014ivoa.spec.1208D

Demleitner, M., Neves, M. C., Rothmaier, F. and Wambsganss, J. (2014), 'Virtual observatory publishing with DaCHS', *Astronomy and Computing* **7**, 27–36, arXiv:1408.5733.
http://adsabs.harvard.edu/abs/2014A%26C.....7...27D

Fernique, P. (2017), 'All-VO registry in Aladin v10: the good, the bad…', Talk given at May 2017 IVOA Interoperability Conference, Shanghai.
http://wiki.ivoa.net/internal/IVOA/InterOpMay2017-Reg/AladinV10Registry-Fernique.pdf

Fernique, P., Boch, T., Donaldson, T., Durand, D., O'Mullane, W., Reinecke, M. and Taylor, M. (2014), 'MOC - HEALPix Multi-Order Coverage map Version 1.0', IVOA Recommendation 02 June 2014, arXiv:1505.02937. http://adsabs.harvard.edu/abs/2014ivoa.spec.0602F

Plante, R., Demleitner, M., Plante, R., Delago, J., Harrison, P. and Tody, D. (2017), 'Describing Simple Data Access Services Version 1.1', IVOA Recommendation 30 May 2017. http://adsabs.harvard.edu/abs/2017ivoa.spec.0530P

Plante, R., Stébé, A., Benson, K., Dowler, P., Graham, M., Greene, G., Harrison, P., Lemson, G., Linde, T. and Rixon, G. (2010), 'VODataService: a VOResource Schema Extension for Describing Collections, Services Version 1.1', IVOA Recommendation 02 December 2010, arXiv:1110.0516. http://adsabs.harvard.edu/abs/2010ivoa.spec.1202P