



*International
Virtual
Observatory
Alliance*

Educational Resources in the Virtual Observatory

Version 1.0

IVOA Working Draft 2023-03-29

Working Group
Registry

This version
<https://www.ivoa.net/documents/DocRegExt/20230329>

Latest version
<https://www.ivoa.net/documents/DocRegExt>

Previous versions
Version 1.1

Author(s)
Demleitner, M., Molinaro, M., Iafrate, G., Heigl, H.

Editor(s)
Demleitner, M.

Abstract

Following the spirit of the Virtual Observatory (VO), the production and provision of material to teach VO usage, both for introducing VO workflows and as an element of a larger astronomy curriculum, should be a joint, global, and distributed effort. In order to ensure global discoverability of the results of that effort, this standard defines an extension to the VOResource metadata scheme suitable for educational texts, and it describes how RegTAP can be used to then discover educational material registered using that extension.

Status of this document

This is an IVOA Working Draft for review by IVOA members and other interested parties. It is a draft document and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use IVOA Working Drafts as reference materials or to cite them as other than “work in progress”.

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

Contents

1 Introduction	2
1.1 Previous Work	3
1.2 Use Cases	3
1.3 Role within the VO Architecture	4
2 The DocRegExt Schema	5
3 Locating Texts using RegTAP	8
A An Example DocRegExt Record	8
B A versioned repository for tutorials	9
C Changes	10
C.1 Changes from the Note	10
References	10

1 Introduction

The Virtual Observatory is a complex tool, introducing numerous technologies not immediately familiar to the average astronomer. In enabling university level students and active researchers to fully exploit the VO’s capabilities, course materials and worked-out use-cases have been found an efficient means of developing the necessary skills both in interactive course situations like “VO Days” and beyond. Efficient ways for interested users to locate such material as well as for VO operators to curate it are highly desirable.

At the same time, advances in technology and communications are creating new and exciting opportunities for teachers to bring astronomy into their classrooms. As the VO makes science-grade data publicly available and classroom sets of (suitably) networked computing devices are now standard in schools, exciting projects come within reach of teachers. To realise these potentials,

it is necessary to disseminate educational material to help teachers preparing classes. Such material – including documented step-by-step tutorials and use cases explaining how to perform basic astrophysical research using VO tools and resources – exists in various formats and has been translated in different languages.

Hence, both training professionals for VO exploitation and the provision of educational material to school teachers and the greater public require some way of handling metadata on the available material. Since the VO has no central body that could curate such a metadata collection but instead uses a decentralised registry for the analogous problem of service discovery, it appears natural to employ the VO Registry for the management of text resource metadata as well.

1.1 Previous Work

In 2013, some providers of educational material began to address the problem of registering educational texts with the goal of improving their findability.

The VO at that time already had a registry extension for standards, which of course are also text-like: StandardsRegExt (Harrison and Burke et al., 2012). This extension, however, focuses on metadata important for standards – e.g., vocabularies and status – that is not pertinent for educational material. Conversely, it is not concerned with document language (which can safely be assumed to be English for standards), and it disregards the issue of locating formatted and source versions, which for educational material is important.

Therefore, a new registry extension was designed, provisionally called DocRegExt. With perhaps a dozen records of that type in the Registry, in 2016 a web page giving a near real-time display of the available material, was created, VO Text Treasures, or VOTT¹ for short. This extension was published as part of the IVOA note “Educational Resources in the Virtual Observatory” (Demleitner and Molinaro et al., 2018).

Since then, the number of registered resources of the type *doc:Document* had risen to 42 as of March 2023. It hence seems advisable to give the registry extension a proper normative basis, which is what this standard provides.

1.2 Use Cases

The design of DocRegExt has been guided by the following discovery cases:

- Is there a tutorial covering discovering intermediate mass black holes? (Standard VOResource is sufficient).
- Is there a tutorial covering working with X-Ray data? (Standard VOResource is sufficient).

¹<https://dc.g-vo.org/VOTT>

- Is there a tutorial dealing with planets suitable for school use? (Standard VOResource is sufficient).
- Is there a tutorial dealing with planets suitable for school use in Italian? (That requires the declaration of the document language).
- What are the subjects of maintained (in the sense of: probably working in the VO as found by the students) tutorials? (The active flag of standard VOResource is unsuitable here since even outdated resources will still be accessible; instead, this standard proposes to include the date the authors last tried a tutorial using standard VOResource mechanisms).
- Are there tutorials using redshifts? (This is addressed by allowing table metadata in DocRegExt).
- Where can I find an editable version of tutorial ivo://auth/tut1? (This is solved by interfaces with a *role* of *source*).
- Are there translations of tutorial ivo://auth/tut2? (This is solved by modelling different translations as capabilities of the same resource).
- Is there material using service ivo://auth/svc1? (Declaring standard VOResource relationships covers this use case).
- Is there material about something visible tonight? (When tutorials declare their *coverage*, this is possible – but few do that at this point).
- I found this VO tutorial somewhere on the net (“on a mirror”). Is it the latest version? If not, where can I find an update? (Unless the title of the text changed, standard VOResource should suffice).

An important additional use case is enabling attractive, browsable lists of registred educational material. In the operation of GAVO’s VOTT service mentioned above it was found that one requirement resulting from this use case is direct access to formatted material in order to enable thumbnail generation.

On the use cases of locating editable forms of such texts – which has been found to be necessary fairly regularly – we note in passing that representing source-product relationships is in principle in the domain of provenance and thus not in the Registry’s main scope. However, in the case discussed here the relation is so simple and its representation so useful that we propose to include it in a DocRegExt.

1.3 Role within the VO Architecture

Fig. 1 shows the role this document plays within the IVOA architecture (Dowler and Evans et al., 2021). It is closely related to the following other standards:

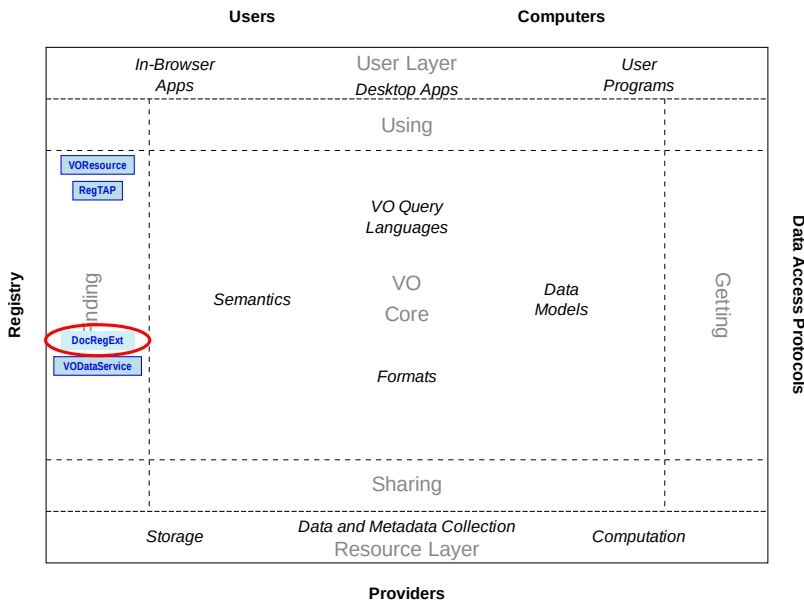


Figure 1: Architecture diagram for this document

- It extends the VO’s basic resource metadata schema *VOResource* (Demleitner and Plante et al., 2021). In particular, we take certain liberties with *VOResource*’s notion of “capability” by re-using it for the different translations of a document.
- We give discovery recipes in terms of the Registry access protocol *RegTAP* (Demleitner and Harrison et al., 2019).
- The base type of the *Document* resource type defined here is *vs:Catalog-Resource* from *VODataService* (Demleitner and Plante et al., 2021). While that type is really intended to describe astronomical *data* (rather than material using such data), pieces of metadata like the tableset or the coverage have clear applications within the current context. With the moderate extension of the notion of a capability mentioned above, the match in the concepts is surprisingly smooth.

2 The DocRegExt Schema

To satisfy the requirements derived above, we have designed a registry extension with two definitions. To avoid unnecessary incompatibilities when migrating to a proper IVOA standard, we use the namespace URI

<http://www.ivoa.net/xml/DocRegExt/v1/>

for DocRegExt even while the schema cannot actually be retrieved from there. The canonical schema location until the extension is endorsed by the IVOA is <http://docs.g-vo.org/xml/DocRegExt/v1/>.

The canonical schema prefix for DocRegExt is `doc`.

To let authors define comprehensive metadata, the schema re-uses the `vs:CatalogResource` type from VODataService 1.2 (Demleitner and Plante et al., 2021) to construct the `doc:Document` resource type. We import the schema by major version, such that further enhancements within the VODataService version 1 series will apply to DocRegExt records, too.

While the schema does not limit what kinds of capabilities a `doc:Document` record has – it is conceivable that custom services for use in a particular tutorial are communicated in this way –, access to actual files is enabled using `doc:Edition`-typed capabilities. It may be argued that this use of VOResource capabilities stretches their semantics a bit. We argue, however, that these documents can well be understood as parameterless service endpoints. Using capabilities furthermore allows a complete representation of the metadata in RegTAP without any extra tables (cf. sect. 3).

The resource-level `REFERENCEURL` in `doc:Document` records should be some sort of landing page with an abstract of the text and links to the full texts and perhaps the document source(s). When using the versioned repository (sect. B), this could be the top-level README file within the VCS. For simple documents, it is acceptable to use the English-language document itself as `referenceURL`; documents only available in non-English should provide a landing page with an English-language abstract, though.

The `FACILITY` and `INSTRUMENT` items should only be set if the text in question actually exploits particular properties of the concrete instrument. A `TABLESET` can be given for the central table-like structures a text deals with and facilitates discovery by physics via the UCDs given in the tableset.

Document-typed resource records should define relations to other general resources (e.g., applications or services) they use. The IVOA vocabulary relationship_type² defines what terms are legal in the declaration of relationships. Document records should preferably use *Cites* and in particular declare relationships to tools. If these are not registered, use the name of their binary as the name of the related resource; this will very typically be lowercase-only. Tutorials specifically introducing one or more services should define *IsSupplementTo* relationships to these services.

Document-typed resource records should contain a `DATE` element with a `ROLE` of *Inspected*. This should correspond to the last time that the current editor has made sure the tutorial or use case works as expected. See the `date_role` vocabulary³ for details.

Each `doc:Edition`-typed capability should correspond to a translation of the document. It is recommended to list the English-language version first if it exists.

²https://www.ivoa.net/rdf/voresource/relationship_type

³http://www.ivoa.net/rdf/voresource/date_role

Remove this when we upload the schema to the IVOA XML repo.

The following description of the *doc:Edition* capability is generated from the schema file.

doc:Edition Type Schema Documentation

An “edition” (typically: translation) of the document.

Although for a while, multiple editions of the document in one language may be given (corresponding perhaps to two “major” versions), in general, only the latest version of the document per language should be present.

At least one `vr:WebBrowser`-typed interface with `role="rendered"` must be present. The access URL of the interface points to a rendered version of the edition (preferably in PDF, but HTML is acceptable, too).

Editors are strongly encouraged to also provide an interface with `role="source"`, the `accessURL` of which should point to an editable version of the document, a version controlled repository, or the like.

doc:Edition Type Schema Definition

```
<xs:complexType name="Edition" >
  <xs:complexContent >
    <xs:extension base="vr:Capability" >
      <xs:sequence >
        <xs:element name="languageCode" type="xs:token" minOccurs="1"
          maxOccurs="1" />
        <xs:element name="locTitle" type="xs:token" minOccurs="0"
          maxOccurs="1" />
      </xs:sequence>
    </xs:extension>
  </xs:complexContent>
</xs:complexType>
```

doc:Edition Extension Metadata Elements

Element *languageCode*

Type string: *xs:token*

Meaning The language this document is (mainly) written in, as an BCP 47 language code.

Occurrence required

Comment The country codes must be given in all lowercase. This results in strings like `en`, `de`, or `or`, where regions might actually matter, `es-es` or `es-mx`.

This language is also the language for `locTitle`, irrespective of that element’s `xml:lang` setting.

Element *locTitle*

Type string: *xs:token*

Meaning The translated document’s title in the language specified by the language sibling.

Occurrence optional

Comment It is recommended to set this element’s `xml:lang` element to the same value as the language sibling. The reason `xml:lang` is not used in the first place is that with an extra element, enforcing that the content language is given is more straightforward.

3 Locating Texts using RegTAP

In the relational registry (Demleitner and Harrison et al., 2014), DocRegExt is straightforwardly represented in the standard VOResource tables. In particular, to find all access URLs for documents together with their English-language titles, one would write:

```
SELECT res_title, access_url FROM
  rr.resource
  NATURAL JOIN rr.interface
WHERE
  res_type='doc:document'
  and intf_role='rendered'
```

The *languageCode* and *locTitle* elements from the *doc:Edition* capability extension are mapped into *res_details* with the following *detail_xpaths*:

- */capability/languageCode* – the document language as an BCP 47 (Phillips and Davis, 2009) language code.
- */capability/locTitle* – the title in the national language.

The downside of not defining an extra table for the documents is that the query patterns in RegTAP are somewhat clumsy. For instance, to list the English and Italian titles of all texts available in Italian, one has to carefully join two subqueries to *res_details*:

```
SELECT res_title, loctitle FROM
  rr.resource
  NATURAL JOIN (
    SELECT ivoid, loctitle FROM (
      SELECT ivoid, cap_index, detail_value as loctitle
      FROM rr.res_detail
      WHERE detail_xpath='/capability/locTitle') AS titles
    NATURAL JOIN (
      SELECT ivoid, cap_index
      FROM rr.res_detail
      WHERE
        detail_xpath='/capability/languageCode'
        AND detail_value LIKE 'it%') AS italiancaps
    ) as loctitles
WHERE
  res_type='doc:document'
```

A An Example DocRegExt Record

This document is accompanied with two example records showcasing most of DocRegExt's features. One example describes a tutorial for determining the

distance to the Crab nebula⁴ and shows a resource with multiple translations. Note that no interfaces with *source* roles are given for the German and Spanish versions. This is because the corresponding files were lost (see sect. B for a proposal designed to reduce such problems).

The second example describes a use case for a particular set of low-resolution spectra⁵ and is thus related to special services having particular tables. This is indicated using related resources and by including a tableset describing these tables.

B A versioned repository for tutorials

Registering text documents as VO resources allows searching for tutorials and similar material through standard registry interfaces, but keeping tutorials up to date, in their master form and also in their translated versions, is an obviously important management issue not really addressed by the Registry.

For tracking changes and versions, the standard tool is a version control system. Therefore, a versioned repository (using subversion as the version control system) has been set up at GAVO data centre⁶. It collects part of the already existing VO tutorials with the goal of preserving them and letting users update and translate them.

The repository has an internal structure designed to enable:

- different national languages (master language set to english)
- translation vs. master language updates
- licensing, in order to clarify how and whether a tutorial can be changed or re-used
- additional material used by tutorials
- access roles to let everyone access tutorials but prevent untrusted updates or additions to it

Details of this structure are discussed in a `README` file at the root of the repository⁷.

The repository is intended to work as an archive and a space for cooperative development of educational material for the VO. A migration to some git-based platform is under consideration.

Within the repository, there is also a relatively lightweight style file with several definitions useful for tutorials. Authors starting new tutorials are encouraged to consider using it. For a short introduction to it, see its `README`⁸.

⁴<https://www.ivoa.net/documents/DocRegExt/20230329/m1distance-example.xml>

⁵<https://www.ivoa.net/documents/DocRegExt/20230329/dfbs-example.xml>

⁶<http://svn.ari.uni-heidelberg.de/svn/edu/>

⁷<http://svn.ari.uni-heidelberg.de/svn/edu/trunk/README>

⁸<http://svn.ari.uni-heidelberg.de/svn/edu/trunk/latex-votut/README>

C Changes

C.1 Changes from the Note

- Following XML itself, we now ask for BCP 47 rather than RFC 3066 language codes.
- Now basing doc:Document on vs:CatalogResource.
- Removed the section on the “educational registry”.

References

- 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://doi.org/10.5479/ADS/bib/2014ivoa.spec.1208D>
- Demleitner, M., Harrison, P., Molinaro, M., Greene, G., Dower, T. and Perdikeas, M. (2019), ‘IVOA Registry Relational Schema Version 1.1’, IVOA Recommendation 11 October 2019.
<http://doi.org/10.5479/ADS/bib/2019ivoa.spec.1011D>
- Demleitner, M., Molinaro, M., Ramella, M., Iafrate, G. and Heintl, H. (2018), ‘Educational resources in the virtual observatory’, IVOA Note 11 April 2018.
<http://ivoa.net/documents/Notes/EDU/>
- Demleitner, M., Plante, R., Stébé, A., Benson, K., Dowler, P., Graham, M., Greene, G., Harrison, P., Lemson, G., Linde, T. and Rixon, G. (2021), ‘VO-DataService: A VOResource Schema Extension for Describing Collections, Services Version 1.2’, IVOA Recommendation 02 November 2021.
<http://doi.org/10.5479/ADS/bib/2021ivoa.spec.1102D>
- Dowler, P., Evans, J., Arviset, C., Gaudet, S. and Technical Coordination Group (2021), ‘IVOA Architecture Version 2.0’, IVOA Endorsed Note 01 November 2021.
<http://doi.org/10.5479/ADS/bib/2021ivoa.spec.1101D>
- Harrison, P., Burke, D., Plante, R., Rixon, G., Morris, D. and IVOA Registry Working Group (2012), ‘StandardsRegExt: a VOResource Schema Extension for Describing IVOA Standards Version 1.0’, IVOA Recommendation 08 May 2012, arXiv:1402.4745.
<http://doi.org/10.5479/ADS/bib/2012ivoa.spec.0508H>
- Phillips, A. and Davis, M. (2009), ‘Tags for identifying languages’, IETF BCP 47.
<https://tools.ietf.org/html/bcp47>