Linked Data in VOTables?

GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

Markus Demleitner

msdemlei@ari.uni-heidelberg.de

e-inf-astro

- What is Linked Data?
- Reminder: RDF
- Reminder: RDFa in TAP examples
- Trivial RDFa lite in VOTable metadata
- Less Trivial RDFa lite in VOTable metadata
- RDFa lite in VOTable TABLEDATA
- Should we go there?

What is Linked Data?

In current practice: Just another buzzword.

Potentially: Rich, machine-readable metadata for otherwise "normal" data files using embedded annotation understandable to *standard* (non-VO) tools…

…linking data to metadata items, complex entities, and possibly other data using RDF.

Remind me: RDF

RDF, the Resource Description Framework, represents information about "resources" in triples of

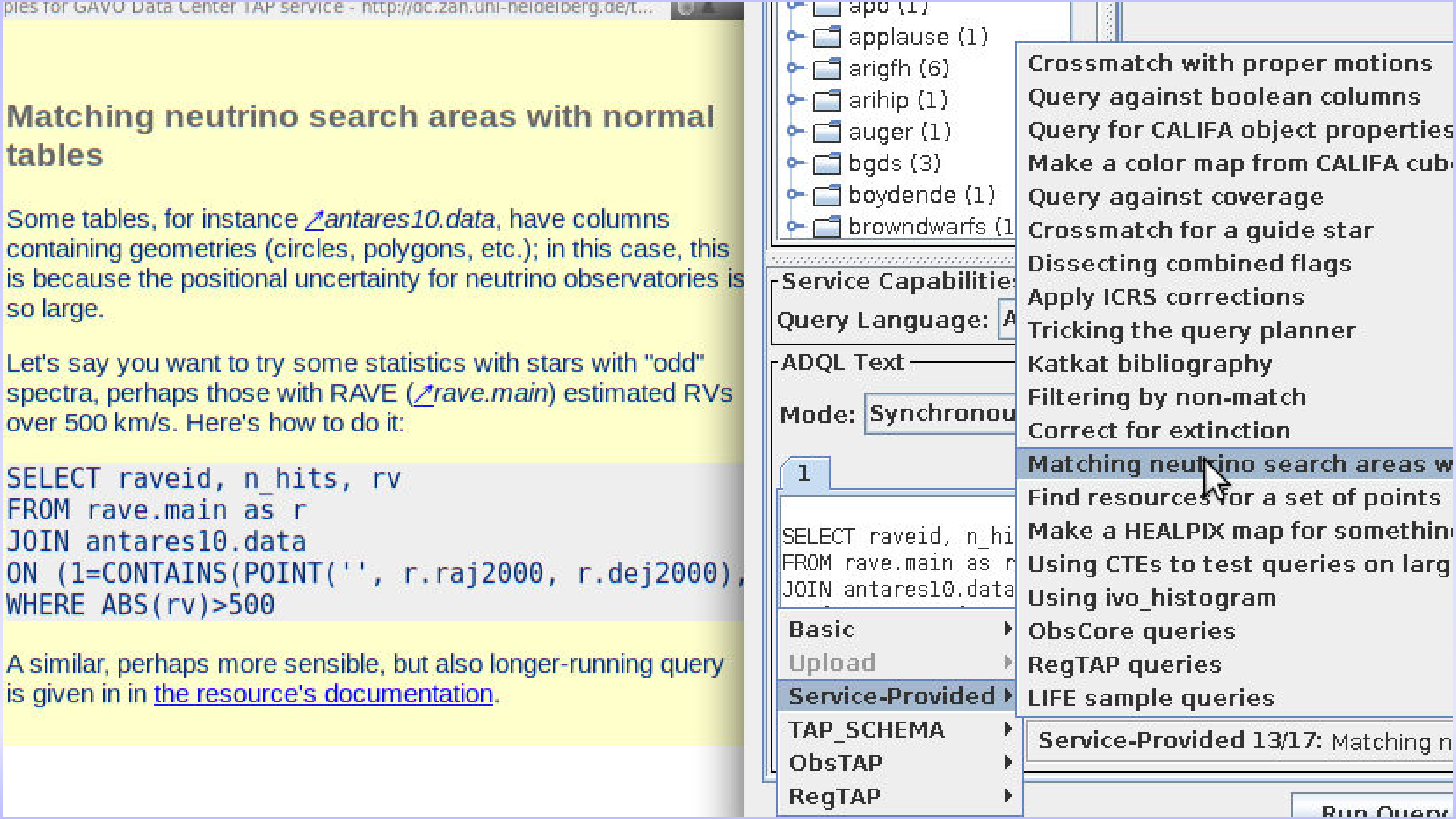
$$\langle \mathit{Subject}, \mathit{Predicate}, \mathit{Object}\rangle.$$

All of these usually are URIs. Think of datalink: a datalink row could be (partially) represented in the RDF triple:

⟨ivo://org.gavo.dc/ ?plts/data/1960Sep_PLS/04573_1.fits,
http://ivoa.net/rdf/datalink/core#documentation,
http://dc.g-vo.org/plts/q/dl/static/marked/4573.jpg⟩

Remind me: RDFa

There are many ways to write such triples. A rather nice one in RDFa, which we are already using in DALI examples:



ples for GAVO Data Center TAP service - http://dc.zah.uni-heidelberg.de/t...

Matching neutrino search areas with normal tables

Some tables, for instance ↗antares10.data, have columns containing geometries (circles, polygons, etc.); in this case, this is because the positional uncertainty for neutrino observatories is so large.

Let's say you want to try some statistics with stars with "odd" spectra, perhaps those with RAVE (↗rave.main) estimated RVs over 500 km/s. Here's how to do it:

```
SELECT raveid, n_hits, rv
FROM rave.main as r
JOIN antares10.data
ON (1=CONTAINS(POINT('', r.raj2000, r.dej2000),
WHERE ABS(rv)>500
```

A similar, perhaps more sensible, but also longer-running query is given in in the resource's documentation.

```
SELECT raveid, n_hi
FROM rave.main as r
JOIN antares10.data
```

RDFa Source Code

There’s actually not much magic behind this. Look at ⟨the source⟩:

```
<body vocab="http://www.ivoa.net/rdf/examples#">
  <div typeof="example" id="Matchingneutrinosearchareaswithnormaltables"
   resource="#Matchingneutrinosearchareaswithnormaltables">
    <h2 property="name">Matching neutrino search areas with
      normal tables</h2>
    <p>Some tables, ...</p>
    <pre class="dachs-ex-tapquery literal-block" property="query">
SELECT raveid, n_hits, rv
FROM rave.main as r
JOIN antares10.data
ON (1=CONTAINS(POINT(’’, r.raj2000, r.dej2000), origin_est))
WHERE ABS(rv)&gt;500
    </pre>
  </div>
</body>
```

For VOTable?

Minimal RDFa (“RDFa lite”) only needs the attributes vocab, typeof, property, resource, prefix, and possibly href and src.

If we allowed these for some (GROUP, INFO, PARAM) or all (TR, TD!) VOTable elements, we could make certain VOTable metadata readable for non-VO software.

Level 0: Atomic Dublin Core

Quite a bit of our VOTable metadata maps nicely to Dublin Core. Standard tools can interpret them when we write:

```
<VOTABLE...>
  <DESCRIPTION property="dcterms:description">
    This schema contains data re-published from the official...
  </DESCRIPTION>
  <RESOURCE type="results">
  <INFO property="dc:rights" name="copyright">
    If you use public Gaia DR3 data in a paper, please take note of
    ...
  </INFO>
```

Level 0: It Works!

```
$ get-triples.py withdc.py
@prefix dc: <http://purl.org/dc/terms/> .

<>
  dc:description """
    This schema contains data re-published from the official
    ...""" ;
  dc:rights """If you use public Gaia DR3 data in a paper, please...
    """ .
```

This means: We have a description and a legal notice on the current document (that's what <> means in turtle), and here's what they read.

Level 1: Typed Annotations

There's a ready-made vocabulary "description of a project" (doap) that states how, well, projects are described. We could declare the project that created the data like this:

```
<GROUP property="dc:creator"
  id="srcproj"
  resource="#srcproj"
  typeof="doap:Project"
  prefix="doap: http://usefulinc.com/ns/doap#">
        <INFO property="doap:name">DPAC consortium</INFO>
        <INFO property="doap:homepage"
          >https://www.cosmos.esa.int/web/gaia/dpac/consortium</INFO>
</GROUP>
```

Level 1: It Works!

Again, standard RDFa tooling can pull out structured information:

```
$ get-triples.py with-typed-ann.xml
@prefix dc: <http://purl.org/dc/terms/> .
@prefix doap: <http://usefulinc.com/ns/doap#> .

<> dc:creator <#srcproj> .

<#srcproj> a doap:Project ;
    doap:homepage "https://www.cosmos.esa.int/web/gaia/dpac/consortium" ;
    doap:name "DPAC consortium" .
```

Level 2: Datalink By The Book

Datalink already is RDF-enabled via semantics. We *could* expose the *data content* as RDF, too:

```
<FIELD arraysize="*" datatype="char" name="ID"/>
<FIELD arraysize="*" datatype="char" name="access_url"/>
<FIELD arraysize="*" datatype="char" name="semantics"/>

<TABLEDATA vocab="http://www.ivoa.net/rdf/datalink/core#"
  resource="ivo://org.gavo.dc/~?kapteyn/data/fits/POT015_000317.fits">
<TR>
  <TD>ivo://org.gavo.dc/~?kapteyn/data/fits/POT015_000317.fits</TD>
  <TD property="preview-image" >http://dc.g-vo.org/[...]/POT015_000317.jpg</TD
  <TD>#preview-image</TD>
</TR>
```

Level 2: It Works!

```
@prefix ns1: <http://www.ivoa.net/rdf/datalink/core#> .
@prefix ns2: <http://www.w3.org/ns/rdfa#> .

<> ns2:usesVocabulary ns1: .

<ivo://org.gavo.dc/~?kapteyn/data/fits/POT015_000317.fits>
  ns1:calibration
  "http://dc.g-vo.org/kapteyn/q/dl/static/wedges/POT015_000317w.fits" ;
  ns1:preview
  "http://dc.g-vo.org/kapteyn/q/dl/static/jpegs/thumb_POT015_000317.jpg" ;
  ns1:preview-image
  "http://dc.g-vo.org/kapteyn/q/dl/static/jpegs/POT015_000317.jpg" ;
  ns1:this
  "http://dc.g-vo.org/getproduct/kapteyn/data/fits/POT015_000317.fits" .
```

Should we?

- Is there sufficient external takeup of RDFa to make it worthwhile?
- Do we even want non-Astronomers to understand our VOTables?
- Would RDF tooling help us do our job? Note ⟨all the great vocabularies⟩ other people have already come up with…
- Major ouch factors: VOTable ID vs RDFa id, links in literals or non-href attributes, no data annotation in BINARY/FITS serialised tables.

…Opinions?