VizieR repository

Lessons learnt and issues

Repository implementation

- Uses the VTOAI OAI-PMH2 Perl Data Provider library (H. Suleman, V. Tech)
  - Slightly modified to handle namespaces (vs:...) in elements
- Answers OAI-PMH requests
  - verbs: Identify, ListSets, ListRecords, GetRecord
  - from, until
  - metadataPrefix (ivo_vor), identifier, resumptionToken (for long lists)

Repository implementation

- http://vizier.u-strasbg.fr/viz-bin/registry/vizier/oai.pl

- Latest version uses VOResource-v0.10

  - http://vizier.u-strasbg.fr/viz-bin/registry/vizier/oai_test.pl (in test... will move to oai.pl)

- Only a repository (not full registry)

- Dynamical content/answer

  - Resources are built on-the-fly from the contents of the VizieR DB

  - Uses VizieR metadata to fill resource elements

VizieR structure
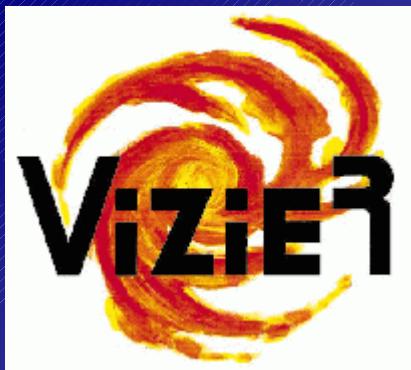
Catalogue 1:
- Table1
- Table2

Catalogue 2:
- Table1
- Table2
- Table3

Catalogue 3:   **Obsoleted by Cat34**
- Table

Clone2

+ Standardized metadata

Metadata extraction

METAcat	VizieR META catalogue	catid=14810095
	Catalogues (one catalogue = set of tables); PK=catid	

Column	Value	Explain
catid	14810095	Identification of catalog
kslot	63	Slot number in Kohonen map [0 .. 99]
status	0	0x80=Obsolete, 0x40=noQbox, 0x10=inPrep, 0=OK
name	J/ApJ/481/95	Catalog designation (CDS)
title	**Radio identification of EGRET sources (Mattox+ 1997)**	Short title of catalog
bibcode	1997ApJ...481...95M	Reference of the catalogue
authors	**Mattox J.R., Schachter J., Molnar L., Hartman R.C., Patnaik A.R.**	Authors of catalog (comma-separated)
explain		Long title of catalog
catype	0	Catalog type: 1=Model 2=Observation 4=Compilation, 5=Critical Comp., 6=General Comp.

Metadata extraction

Column	Value		Explain
	VizieR META catalogue		catid=14810095, tabid=1
METAtab	Contains details of all tables; PK=(catid, tabid)		
name	J/ApJ/481/95/table1		logical table name in catalog
dbname	**c14810095t1**		name of table in db or script
dbaid	**2**		Data-base identifier FK(METAdba)
catid	**14810095**		catalogue identification FK(METAcat)
tabid	**1**		identifier of table within catid
notid	**0**		METAnot_Note related to the Table
flags	**0**		flags -- not yet used
equinox		yr	Equinox of RA and DE coordinates
epoch		yr	Epoch of coordinates (need pm)
cooframe	**2**		Frame of Coordinates FK4/G/S/Ecl/FK5
status	**0**		0x80=Obsolete, 0x40=noQbox, 0x10=inPrep
explain	**Potential radio identifications of sources in the second EGRET catalog**		Caption of this table
comment	**The column 2EG provides a direct link to 2EG data**		Free-text comments
records	**154**		Number of records (tuples) in table
loadate	**1998-09-19T16:25:08**		Date/time table loaded (sec/2000)
release	1998-09-19T16:25:16		Date/time table Released (sec/2000)
filename	**/ftp/cats/J/ApJ/481/95/table1.dat**		original file (at CDS only)

Metadata extraction

METAcol	Column Description: PK=(catid,tabid,colid)												

To get all details for a row, just click on the row number in the leftmost `Full' column.

Full	name	catid	tabid	colid	notid	famid	type	dbtype	length	flags	morid	fmt	unit	dbunit
1	recno	14810095	1	0	0	1243	0	0	8	0	0	%8d		
2	2EG	14810095	1	1	0	179	4	24	10	3	1	%-10s		
3	n_2EG	14810095	1	2	1	313	4	24	1	3	0	%-1s		
4	GLON	14810095	1	3	0	1172	301	0	8	23	0	%6.2f	deg	.01deg
5	GLAT	14810095	1	4	0	1171	1	0	8	23	0	%6.2f	deg	.01deg
6	n_GLAT	14810095	1	5	2	313	4	24	1	0	0	%-1s		
7	VI	14810095	1	6	3	1393	1	0	8	C0	0	%5.2f		.01
8	ID	14810095	1	7	0	156	4	24	8	3	0	%-8s		
9	u_ID	14810095	1	8	0	102	4	24	1	0	0	%-1s		
10	RadioID	14810095	1	9	4	166	4	24	12	3	0	%-12s		
11	n_RadioID	14810095	1	10	5	313	4	24	1	0	0	%-1s		
12	Names	14810095	1	11	0	156	4	24	12	0	0	%-12s		
13	n_Names	14810095	1	12	6	313	4	24	1	0	0	%-1s		
14	S5GHz	14810095	1	13	0	513	0	0	8	83	0	%5d	mJy	mJy
15	SI	14810095	1	14	0	1326	1	0	8	C3	0	%4.1f		.1
16	r0	14810095	1	15	0	134	1	0	8	80	0	%6.1f	arcmin	.1arcmin
17	eta	14810095	1	16	0	1351	1	0	8	80	0	%9.6f		.000001
18	r	14810095	1	17	0	1110	1	0	8	C0	0	%5.1f	arcmin	.1arcmin
19	r95	14810095	1	18	0	134	1	0	8	C0	0	%5.1f	arcmin	.1arcmin
20	Contour	14810095	1	19	0	1199	1	1	8	80	0	%7.3f	%	%
21	LR	14810095	1	20	0	1346	1	1	8	80	0	%9.2e		
22	p(id/r)	14810095	1	21	0	1351	1	1	8	80	0	%9.2e		

Metadata extraction

J/ApJ/481/95   Radio identification of EGRET sources (Mattox+ 1997)
==
The identification of EGRET sources with flat-spectrum radio sources.
        Mattox J.R., Schachter J., Molnar L., Hartman R.C., Patnaik A.R.
        <Astrophys. J. 481, 95 (1997)>
        =1997ApJ...481...95M   (SIMBAD/NED BibCode)
==
ADC_Keywords: Active gal. nuclei ; Radio sources ; Gamma rays ;
              Cross identifications
Mission_Name: CGRO

Keywords: galaxies: active - gamma rays: observations - quasars: general

Abstract:
    We present a method to assess the reliability of the identification of
    EGRET sources with extragalactic radio sources. We verify that EGRET
    is detecting the blazar class of active galactic nuclei (AGNs).
    However many published identifications are found to be questionable.
    We provide a table of 42 blazars that we expect to be robust
    identifications of EGRET sources. This includes one previously
    unidentified EGRET source, the lensed AGN PKS 1830-210, near the
    direction of the Galactic center. We provide the best available
    positions for 16 more radio sources that are also potential
    identifications for previously unidentified EGRET sources. All high
    Galactic latitude EGRET sources (|b|>3) that demonstrate significant
    variability can be identified with flat-spectrum radio sources. This
    suggests that EGRET is not detecting any type of AGN other than
    blazars. This identification method has been used to establish with
    99.998% confidence that the peak gamma-ray flux of a blazar is
    correlated with its average 5GHz radio flux. An even better
    correlation is seen between gamma-ray flux and the 2.29GHz flux
    density measured with VLBI at the base of the radio jet. Also, using
    high-confidence identifications, we find that the radio sources
    identified with EGRET sources have greater correlated VLBI flux
    densities than the parent population of flat radio spectrum sources.

File Summary:

        FileName   Lrecl   Records   Explanations

    x ReadMe   80   .   This file
    x table1.dat   141   154   Potential radio identifications of sources in
                                      the second EGRET catalog
    x table2.dat   139   60   Potential radio identifications of EGRET sources

References:
    Fichtel et al., The first EGRET catalog   1994ApJ....94..551F
==
(End)   James Marcout, Patricia Bauer [CDS]   07-Oct-1997

XML generation

```xml
- <OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
    <responseDate>2004-09-24T16:35:35Z</responseDate>
  - <request verb="GetRecord" metadataPrefix="ivo_vor" identifier="ivo://CDS/VizieR/J/ApJ/481/95/table1">
      http://vizier.u-strasbg.fr/cgi-bin/registry/vizier/oai_test.pl
    </request>
  - <GetRecord>
    - <record>
      - <header>
          <identifier>ivo://CDS/VizieR/J/ApJ/481/95/table1</identifier>
          <datestamp>1998-09-19T17:25:16Z</datestamp>
          <setSpec>J.ApJ</setSpec>
        </header>
      - <metadata>
        - <resource xsi:type="vs:TabularSkyService" xsi:schemaLocation="http://www.ivoa.net/xml/VOResource/v0.10
            http://www.ivoa.net/xml/VOResource/VOResource-v0.10.xsd http://www.ivoa.net/xml/VODataService/v0.5
            http://www.ivoa.net/xml/VODataService/VODataService-v0.5.xsd http://www.ivoa.net/xml/ConeSearch/v0.3
            http://www.ivoa.net/xml/ConeSearch/ConeSearch-v0.3.xsd">
          - <title>
              Radio identification of EGRET sources (Mattox+ 1997) - Potential radio identifications of sources in the second EGRET catalog
            </title>
            <shortName>J/ApJ/481/95/tab</shortName>
            <identifier>ivo://CDS/VizieR/J/ApJ/481/95/table1</identifier>
          - <curation>
              <publisher ivo-id="ivo://CDS/VizieR">VizieR</publisher>
            - <creator>
              - <name>
                  Mattox J.R., Schachter J., Molnar L., Hartman R.C., Patnaik A.R.
                </name>
              </creator>
              <contributor>James Marcout, Patricia Bauer [CDS]</contributor>
              <date role="creation">1998-09-19T17:25:16Z</date>
              <version>07-Oct-1997</version>
            - <contact>
                <name>CDS support team</name>
              - <address>
                  CDS, Observatoire de Strasbourg, 11 rue de l'Universite, F-67000 Strasbourg, France
                </address>
                <email>question@simbad.u-strasbg.fr</email>
              </contact>
            </curation>
          - <content>
              <subject>AGN</subject>
              <subject>Stars</subject>
            - <description>
                We present a method to assess the reliability of the identification of EGRET sources with extragalactic radio sources. We verify that EGRET is detecting the
                blazar class of active galactic nuclei (AGNs). However many published identifications are found to be questionable. We provide a table of 42 blazars that we
                expect to be robust identifications of EGRET sources. This includes one previously unidentified EGRET source, the lensed AGN PKS 1830-210, near the
                direction of the Galactic center. We provide the best available positions for 16 more radio sources that are also potential identifications for previously
```

Dynamical repository

- The VizieR repository currently has no memory:
  - :o) always up-to-date
  - :o) easy to maintain (especially useful for HUGE collections like VizieR)
  - :o(not the OAI spirit (OAI is book-oriented: once a resource has existed, it must exist forever, even as 'deleted')

- Need hard-coded XML version ?

- What date should be used ?
  - Metadata change => created=? modified=?

Issues

- Resources granularity
  - Currently, VizieR description at Table level only
  - Why not Catalogues ?
  - Create resources for Tables AND Catalogues ?
  - Make use of OAI Sets (4000 sets?)
  - How to handle hierarchy in a 'flat' registry?

- Handling mirrors (no, not again?)
  - Dupplicate resources?
    - 10,000 resources * 9 mirrors = 90,000 !
  - Dupplicate interfaces (and accessURL)?

Issues

- Currently, all VizieR tables are presented as <resource xsi:type="vs:TabularSkyService">
  - (needed to describe the table columns)
- For each table, there are different interfaces
  - WebBrowser (for humans)
  - ParamHTTP (retrieve VOTable)
- If the table contains positions, I can make a ConeSearch on this table
  - new <resource xsi:type=...>?
  - new interface?
  - single resource VizieRConeSearch ?
    - thousands possible table-id as parameters!

Issues

- How to describe a WS in an interface (methods?)

- Relations between resources with <relatedResource>

  - Standard 'deprecatedBy' ?

  - Point to ivo-id and not Name?