

# XML и реляционные СУБД



РОССИЙСКИЕ  
ИНТЕРНЕТ  
ТЕХНОЛОГИИ

**Хранение и работа с XML  
в реляционных системах  
управления базами данных**

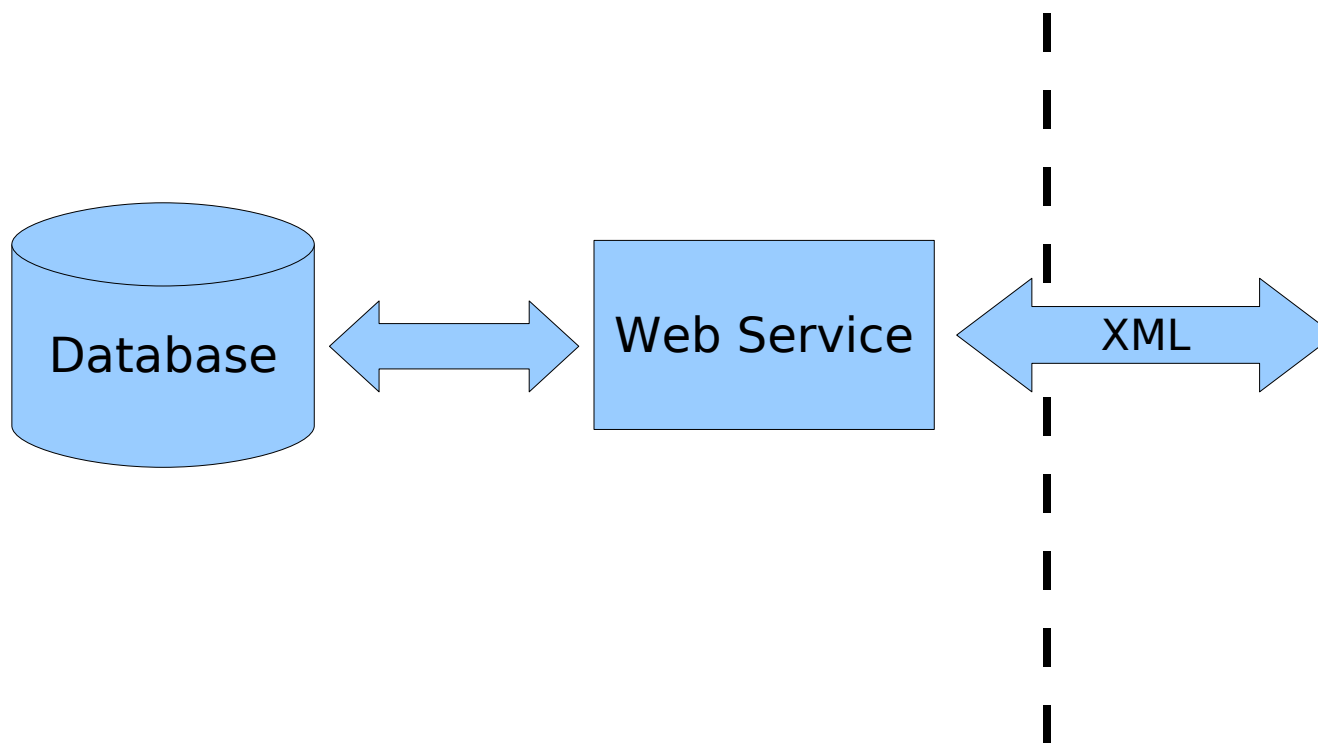
**Самохвалов Николай**  
Компания «Постгресмен»  
<http://postgresmen.ru>  
[rit2007@samokhvalov.com](mailto:rit2007@samokhvalov.com)

# XML и СУБД

## Зачем?

# XML и СУБД

- XML для общения с «внешним миром»



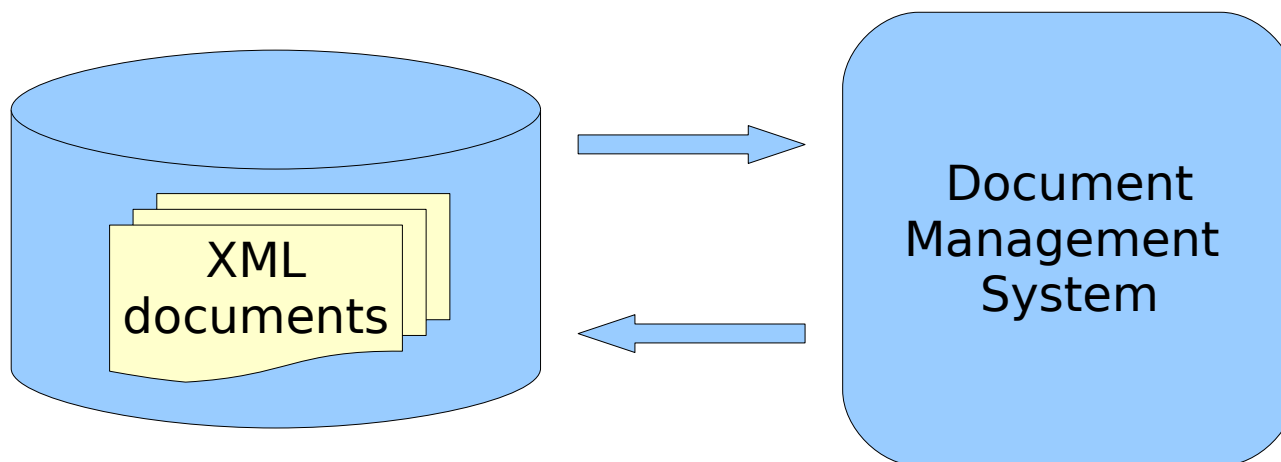
# Зачем хранить XML в РСУБД?



- Системы документооборота
- Полуструктурированные данные
- Архивы документов

# Пример 1: система документооборота

Цель: работа с документами в оригинальном формате



# Пример 2: разнородный каталог

Цель: работа с объектами с различными наборами СВОЙСТВ

**Авто** [+ Подать объявления в этот раздел](#)

<a href="#">Кредиты и ссуды</a> (22)	<a href="#">Комиссионное оформление и страхование</a> (5)
<a href="#">Легковые автомобили</a> (35679) <ul style="list-style-type: none"><li><a href="#">Новые автомобили</a></li><li><a href="#">Автомобили с пробегом</a></li></ul>	<a href="#">Прицепы, автодома и фургоны</a> (146)
<a href="#">Малый коммерческий транспорт</a> (2320)	<a href="#">Автобусы</a> (231)
<a href="#">Средние и тяжелые грузовики</a> (2039)	<a href="#">Спецтехника</a> (1426)
<a href="#">Грузовые прицепы</a> (508)	<a href="#">Мотоциклы и мопеды</a> (1505)
<a href="#">Водный транспорт</a> (365)	<a href="#">Снегоходы</a> (37)
<a href="#">Автозапчасти и принадлежности</a> (10507) <ul style="list-style-type: none"><li><a href="#">Для легковых автомобилей</a></li><li><a href="#">Для коммерческих автомобилей</a></li><li><a href="#">Для мотоциклов и мопедов</a></li><li><a href="#">Диски, шины, колеса</a></li><li><a href="#">Для спецтехники и прицепов</a></li><li><a href="#">Авто на запчасти</a></li><li><a href="#">Средства для ухода</a></li><li><a href="#">Другое</a></li></ul>	<a href="#">Автосервис и услуги</a> (780) <ul style="list-style-type: none"><li><a href="#">Сервис и ремонт</a></li><li><a href="#">Экспертиза и оценка</a></li><li><a href="#">Юридическая помощь</a></li><li><a href="#">Автошколы</a></li><li><a href="#">Другое</a></li></ul>
<a href="#">Другое</a> (400)	

IRR.ru, раздел «Авто»

## Пример 2: разнородный каталог

item	
obj_id	INT8
item_section_id	INT8
item_vendor_id	INT8
item_model_id	INT8
item_year	INT2
item_price	NUMERIC(30,6)
<b>item_prop1</b>	<b>INT4</b>
<b>item_prop2</b>	<b>INT4</b>
<b>item_prop3</b>	<b>INT4</b>
<b>item_prop4</b>	<b>INT4</b>
...	
<b>item_prop21</b>	<b>TEXT</b>
<b>item_prop22</b>	<b>TEXT</b>
<b>item_prop23</b>	<b>TEXT</b>
...	
<b>item_prop41</b>	<b>BOOLEAN</b>
...	



item	
obj_id	INT8
item_section_id	INT8
item_vendor_id	INT8
item_model_id	INT8
item_year	INT2
item_price	NUMERIC(30,6)
<b>item_props</b>	<b>hstore</b>

hstore – тип данных в PostgreSQL,  
множество пар «ключ – значение» (аналог хэшей  
с поддержкой индексов (GiST))

## Пример 3: журнал действий

Цель: гибкость схемы (возможность добавить новые свойства)

action	
action_id	SERIAL
action_type_id	INT4
action_status_id	INT4
action_person_id	INT4
<b>action_data</b>	<b>hstore</b>

# SQL/XML (ISO/IEC 9075-14:2006)

- тип данных XML (XQuery 1.0 and XPath 2.0 Data Model)
- функции публикации (XMLDOCUMENT, XMLELEMENT, XMLATTRIBUTES, XMLFOREST, XMLAGG, XMLCONCAT, XMLPI, XMLCOMMENT)
- поддержка XQuery (операция XMLQUERY)

# Тип данных XML

doc		
PK	id	INTEGER
	data	XML
	created	TIMESTAMP

```

- <rss version="2.0">
- <rss version="2.0">
- <channel>
  <title>MoiKrug.ru - Вакансии (RSS)</title>
  - <description>
    MoiKrug.ru - Последние вакансии в первой социальной сети
  </description>
  <link>http://moikrug.ru/?from=rss</link>
  <language>ru-ru</language>
  - <image>
    <link>http://moikrug.ru/?from=rss</link>
    <url>http://moikrug.ru/images/logo.gif</url>
  - <title>
    MoiKrug.ru - найти хороших людей по рекомендациям
  </title>
  </image>
  - <item>
    <title>Менеджер по персоналу ($1500)</title>
    <link>http://moikrug.ru/vacancies/736200561/?from=rss</link>
    - <description>
      Менеджер по персоналу ($1500)<br /> HFLabs<br />
      персоналу.<br /> <br /> Обязанности:<br /> * кадровое
      поддержание корпоративной культуры компании.<br /> за
      года в роли HR в западной компании или компании в
      /> * английский язык на уровне чтения резюме;<br />
      зарплата 1000-1700 у.е. плюс бонусы 300-500 у.е. за
      "HR".
    </description>
    <dc:creator>Холодова Елена</dc:creator>
    <pubDate>Fri, 24 Nov 2006 18:08:36 GMT</pubDate>
  </item>
</channel>
</rss>
  
```

```

CREATE TABLE doc (
  id INTEGER PRIMARY KEY,
  data XML,
  created TIMESTAMP WITH TIMEZONE
);
  
```

# Производительность

- Выборка всего XML-значения по условию XPath

```
SELECT xmlcol  
FROM tbl1  
WHERE (xpath('/a/@b', xmlcol))[1] = '123';
```

```
SELECT xmlcol  
FROM tbl1  
WHERE (xpath('/a[@c="s"]/@b', xmlcol))[1] = '123';
```

- Выборка произвольной части XML-значения

```
SELECT xpath('/a/@b', xmlcol)  
FROM tbl1  
WHERE id = 111;
```

- Модификация части XML-значения

# Oracle 10g

- Тип XMLType. Варианты хранения:
  - object-relational storage
  - CLOB
- XML Schemas
- SQL/XML (включая ф-ции публикации, XMLQUERY, XMLVALIDATE, XMLTABLE)
- XQuery
- XSLT
- DOM
- Node-level updates
- Индексы:
  - обычные B-деревья в случае object-relational storage
  - XMLIndex для обоих типов хранения (индексация путей, значений, функциональные индексы)
  - индексы для полнотекстового поиска

# MS SQL 2005

- Тип XML. Хранение: BLOB.
- XML Schemas
- SQL/XML publishing functions
- XQuery
- XSLT
- Node-level updates
- Индексы:
  - «первичные» XML-индексы (ORDPATH)
  - «вторичные» XML-индексы
    - PATH
    - VALUE
    - PROPERTY
- Web services

# IBM DB2 version 9

- Тип XML. Собственное хранилище (native storage: pureXML)
- XML Schemas (с поддержкой эволюции)
- SQL/XML (включая ф-ции публикации, XMLQUERY, XMLVALIDATE, XMLTABL
- XQuery
- XSLT
- Node-level updates
- Индексы:
  - XML value indexes
  - XML full text indexes
- Annotated XML Schemas для декомпозиции (XML->rel)

# MySQL 5.1

```
mysql> SET @xml = '<a>111<b:c>222<d>333</d><e:f>444</e:f></b:c></a>';  
Query OK, 0 rows affected (0.00 sec)
```

```
mysql> SELECT ExtractValue(@xml, '//e:f');
```

```
+-----+  
| ExtractValue(@xml, '//e:f') |  
+-----+  
| 444 |  
+-----+
```

```
1 row in set (0.00 sec)
```

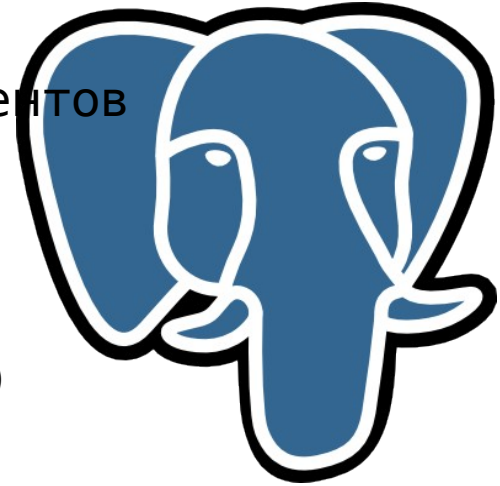
```
mysql> SELECT UpdateXML(@xml, '//b:c', '<g:h>555</g:h>');
```

```
+-----+  
| UpdateXML(@xml, '//b:c', '<g:h>555</g:h>') |  
+-----+  
| <a>111<g:h>555</g:h></a> |  
+-----+
```

```
1 row in set (0.00 sec)
```

# PostgreSQL 8.3

- Тип XML (over VARCHAR), поддержка XML-фрагментов
- SQL/XML
  - SQL/XML publishing functions
  - XMLPARSE, XMLSERIALIZE
- XSLT (via contrib)
- XPath с поддержкой XML Namespaces (xpath())
- Индексы:
  - функциональные индексы над xpath()
  - XLABEL (GiST based; в разработке)
- Альтернативные функции отображения rel->XML
  - table\_to\_xml(), query\_to\_xml(), cursor\_to\_xml()
  - table\_to\_xmlschema(), ...
  - table\_to\_xml\_and\_xmlschema(), ...
  - database\_to\_xml(), database\_to\_xmlschema(), ...

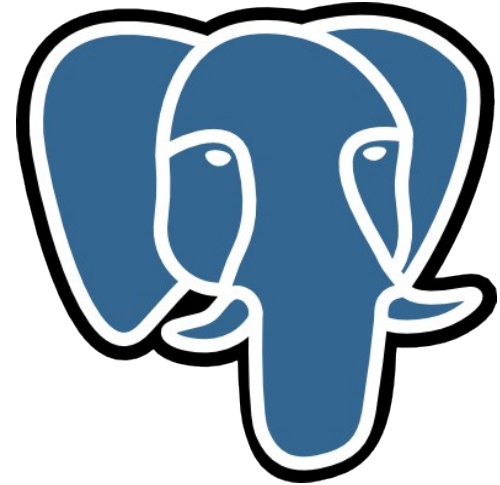


# PostgreSQL 8.3

```
xmltest=# SELECT XMLELEMENT(  
           NAME "person",  
           XMLATTRIBUTES(  
               person_name || ' ' || person_lastname  
                   AS "name",  
               age(person_birthday) AGE "age"  
           )  
       )  
FROM person;
```

xmlelement

```
-----  
<person name="Ivan Sidorov" age="29 years 6 mons 6 days" />  
<person name="Petr Ivanov" age="22 years 8 mons 14 days" />
```



# PostgreSQL 8.3

```
xmltest=# CREATE TABLE table1(
        id INTEGER PRIMARY KEY,
        created TIMESTAMP NOT NULL
        DEFAULT CURRENT_TIMESTAMP,
        xdata XML
    );
CREATE TABLE

xmltest=# INSERT INTO
        table1(id, xdata)
        VALUES(
            1,
            '<dept xmlns:smpl="http://example.com" smpl:did="DPT011-IT">
                <name>IT</name>
                <persons>
                    <person smpl:pid="111">
                        <name>John Smith</name>
                        <age>24</age>
                    </person>
                    <person smpl:pid="112">
                        <name>Michael Black</name>
                        <age>28</age>
                    </person>
                </persons>
            </dept>'
        );
INSERT 0 1
```



# PostgreSQL 8.3

```
xmltest=# SELECT xpath('//person/name/text()', xdata)
FROM table1;
```

        xpath

```
-----
{"John Smith","Michael Black"}
(1 row)
```

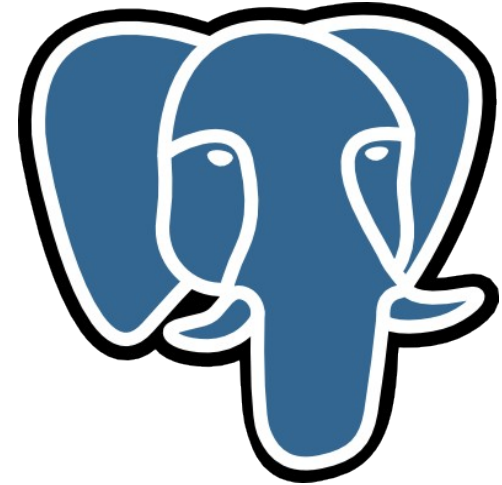
```
xmltest=# SELECT xpath('//person/@simpl:pid', xdata,
                  ARRAY[ARRAY['simpl', 'http://example.com']])
FROM table1;
```

        xpath

```
-----
{111,112}
(1 row)
```

```
xmltest=# CREATE INDEX i_table1_xdata ON table1 USING btree(
          xpath('//person/@name', xdata)
```

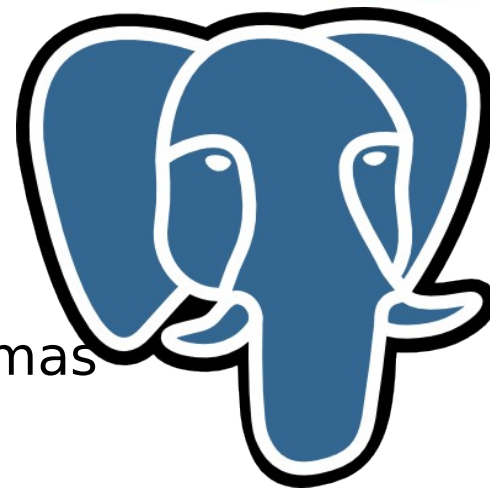
```
);
CREATE INDEX
```



# PostgreSQL 8.3

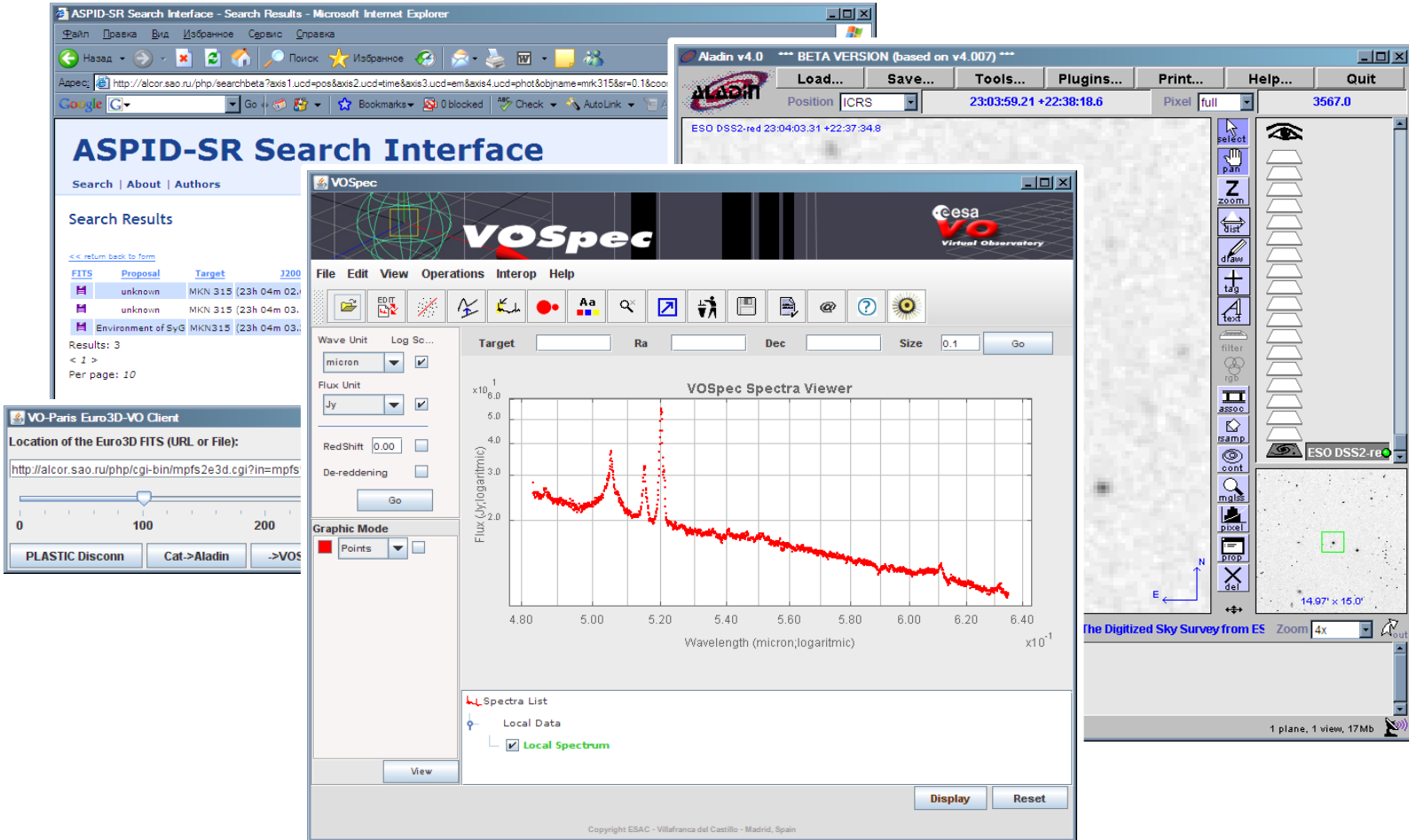
Планы по развитию:

- XLABEL
- XML Schemas catalogue
- полнотекстовый поиск по XML-данным
- decomposition with annotated XML Schemas
- XQuery (XMLQUERY)



# Пример использования

Хранение метаданных в наблюдательных архивах  
International Virtual Observatory



ASPID-SR Search Interface - Search Results - Microsoft Internet Explorer

ASPID-SR Search Interface

Search Results

FITS	Proposal	Target	J2000
unknown	MKN 315	(23h 04m 02s	
unknown	MKN 315	(23h 04m 03s	
Environment of SyG	MKN315	(23h 04m 03s	

Results: 3  
< 1 >  
Per page: 10

VO-Paris Euro3D-VO Client

Location of the Euro3D FITS (URL or File):  
<http://alcor.sao.ru/php/cgi-bin/mpfs2e3d.cgi?in=mpfs>

PLASTIC Disconn Cat->Aladin ->VOS

Aladin v4.0 BETA VERSION (based on v4.007)

Position |CRS| 23:03:59.21 +22:38:18.6 Pixel full 3567.0

ESO DSS2-red 23:04:03.31 +22:37:34.8

VOSpec

VOSpec Spectra Viewer

Wave Unit Log Sc...  
micron

Flux Unit  
Jy

RedShift 0.00

De-reddening

Graphic Mode  
 Points

Target \_\_\_\_\_ Ra \_\_\_\_\_ Dec \_\_\_\_\_ Size 0.1 Go

Flux (Jy,logarithmic) x10<sup>0</sup>

Wavelength (micron,logarithmic) x10<sup>-1</sup>

Spectra List

- Local Data
  - Local Spectrum

Display Reset

Copyright ESAC - Villafranca del Castillo - Madrid, Spain

The Digitized Sky Survey from ES Zoom 4x

1 plane, 1 view, 17Mb

**Спасибо за  
внимание!**

# Ссылки, литература

- **SQL:200x, 2006, Part 14: XML-Related Specifications (SQL/XML)**, ISO/IEC JTC CD 9075-14:200x(E). International Standard, ISO, ANSI.
- **R. Bourret. XML Database Products.** <http://www.rpbouret.com/xml/XMLDatabases>
- **M. Nicola and B. van der Linden. Native XML Support in DB2 Universal Data**  
In VLDB, pages 1164 1174, Trondheim, Norway, 2005.
- **P. O Neil, E. O Neil, S. Pal, I. Cseri, G. Schaller, and N. Westbury. ORDPATHS  
insert-friendly XML node labels.** // In SIGMOD 04: Proceedings of the 2004 ACM  
international conference on Management of data, pages 903 908, New York, NY, USA  
ACM Press.
- **Бартунов О. С. Что такое PostgreSQL?** // Корпоративные базы данных-2005. (C  
M.: 2005. [http://www.citforum.ru/database/postgres/what\\_is/](http://www.citforum.ru/database/postgres/what_is/)
- **N. Samokhvalov. XML Support in PostgreSQL.** // Proceedings of the SYRCoDIS-2  
Moscow, Russia, 2007.
- **I. Zolotukhin, N. Samokhvalov, F. Bonnarel, and I. Chilingarian. Comprehensive  
Metadata Query Interface for Heterogeneous Data Archives Based on Open  
PostgreSQL ORDBMS.** // In *Astronomical Data Analysis Software & Systems XVI*, p  
Tucson, Arizona, USA, 2006.