VISTA DATA FLOW SYSTEM (VDFS)

for VISTA & WFCAM data

Science Archive User Interface Document

author

M.A. Read (WFAU Edinburgh)
Project assistant
number
VDF-WFA-VSA-008
issue
Issue 1.0
date
19 September 2006
co-authors
N.C. Hambly

	VDF-WFA-VSA-008,	Issue 1.0.	, 19 September	2006
--	------------------	------------	----------------	------

2

Contents

1	SC	OPE	4
2	INT	TRODUCTION	4
3	VD	FS WEBSITES	4
	3.1	Construction	5
	3.2	Content summary	5
4	$\mathbf{U}\mathbf{S}^{1}$	ER AUTHENTICATION	5
5	PIX	KEL DATA	6
	5.1	Small image extraction - pixel form 1 (PF1)	7
	5.2	Batch small image extraction - pixel form 2 (PF2) $\dots \dots \dots \dots \dots \dots$	7
	5.3	Large image extraction - pixel form 3 (PF3)	8
	5.4	Stacked image generation - pixel form 4 (PF4)	8
	5.5	Browsable access to pixel data	8
	5.6	Other access to pixel data	8
	5.7	Demonstrations of pixel access	9
6 OBJECT CATALOGUE DATA			
	6.1	Radial Search - Object Form 1 (OF1) $\dots \dots \dots \dots \dots \dots \dots \dots \dots$	9
	6.2	Rectangular Search - Object Form 2 (OF2)	10
	6.3	SQL query - Object Form 3 (OF3)	10
	6.4	Cross-matching a catalogue - Object Form 4 (OF4)	10
	6.5	Browsable access to catalogue data	10
	6.6	Other access to catalogue data	10
	6.7	Demonstrations of catalogue access	11
7	VII	RTUAL OBSERVATORY CONSIDERATIONS	11
8	но	USEKEEPING	11
9	US	ER DOCUMENTATION	11
	9.1	Schema Browser	12
10	AP	PENDIX	13
11	AC	RONYMS & ABBREVIATIONS	19
12	AP	PLICABLE DOCUMENTS	20

VDF-WFA-VSA-008, Issue 1.0, 19 September 2006	3
13 CHANGE RECORD	20
14 NOTIFICATION LIST	21

1 SCOPE

This document describes the interface and underlying software that will sit between the data held in the science archives of the VISTA Data Flow System (VDFS) and the end user. In this context a user might be an astronomer from an ESO member country, a non-ESO astronomer or to a lesser extent a member of the public.

The VDFS science archives are the WFCAM Science Archive (WSA) and the VISTA Science Archive (VSA). The Science Requirements Analysis Document (AD01) lays out the conditions for user access to these archives and drives the development of the user-interface. From a user-interface point of view, many of the requirements and methods of implementation are similar or indeed the same for the WSA and VSA. This document will take a joint approach to the archives highlighting any differences as they occur. VDFS design policy was to develop the WSA first and some examples of existing functionality are included.

2 INTRODUCTION

The WSA and VSA websites will be the primary point of contact for users and Section 3 gives a summary of their construction and content.

If users wish to access data still under a proprietary period they will first need to be authenticated in order to verify their access rights to the data. User authentication is outlined in Section 4.

Users of the archives will be mainly interested in access to the stored pixel data (FITS images) and/or the generated object catalogues. The archives will store pixel data in flat files with the meta-data loaded into a Relational DataBase Management System (RDBMS). Object catalogue data are read from FITS catalogue files during the curation procedures and stored directly in the RDBMS (though the FITS catalogue files will also be archived). During curation any enhanced object catalogue products are also produced and loaded into the RDBMS e.g. merged source and neighbour tables.

The way the data are held in the archive and the different software required to process user requests lead us to consider these two types of access in separate sections.

Section 5 discusses the user interface to the pixel data splitting it up into the different ways in which users will be able to extract image data and the underlying software required.

The user interface providing access to the object catalogue data is similarly broken down and the subject of Section 6.

Ensuring the archive interface conforms to emerging Virtual Observatory standards and housekeeping issues are briefly discussed in Sections 7 and 8.

Section 9 lists the documentation that will be provided to assist the user.

3 VDFS WEBSITES

Initial access to information on the WSA and VSA and to the data is via the websites maintained by WFAU. The WSA website is at

http://surveys.roe.ac.uk/wsa

3.1 Construction

Apache web servers running under Linux and offering up CGI host the VDFS websites. Tomcat is installed on the servers providing Java servlet functionality. All the sites' pages and forms are HTML 4.01 compliant. This ensures they work with a wide range of browsers on common operating systems. Specifically they have been tested using Microsoft Internet Explorer under Windows and Netscape variants running on Unix/Linux. Cascading Style Sheets (CSS) will also be used to build the website.

Experience with the WSA has shown that it has been necessary to make use of Javascript in some of the web forms to enhance the functionality.

3.2 Content summary

The VDFS websites will consist of

- introductory pages
- information on the various surveys, their progress and what is held in the archives
- web forms enabling the user to access the data and pages describing other access points (see Sections 5 and 6)
- a schema browser, used to explore the structure and content of the databases in the archives (see Section 9)
- documentation and help pages (see Section 9)
- links to relevant resources (e.g. WFCAM, UKIDSS, VIRCAM etc).

4 USER AUTHENTICATION

Data held in the WSA and VSA are subject to a proprietary period during which time they should only be accessible to authorised users. To facilitate this, proprietary data are to be held in protected databases until their proprietary period has expired at which point the database will become world-readable.

Users wishing to access proprietary data through the user-interface will need to be registered with the archives. For UKIDSS data held in the WSA the list of registered users is maintained by a group of community contacts. These contacts administer their community (adding users, changing passwords etc) though a password protected interface (Java servlet) provided by WFAU. The list of registered non-survey users who wish to access their open-time data is maintained by WFAU staff. The lists of registered users are held in a database in the archive.

When accessing data under its proprietary period users must first login using a web form. The attempt to login is authenticated against the lists of users in the database. Successful logins are stored in a browser session. The session login status is then used by the access points described below to determine which database(s) the user is allowed to access. The underlying database connections are also based on login status. These connections will fail if a user attempts un-authorised access to a proprietary database. A user's login status is displayed at the top of all the web forms accessing the archive.

5 PIXEL DATA

As detailed in document AD02 the WSA and VSA will store the pixel data in multi-extension FITS files in a flat file system. Tables in the SQL databases will hold the meta data extracted from each image and will also record the path/filename of the corresponding FITS file. The SQL database resides on a separate machine (Windows PC) that is networked to the web server. The FITS files are on disks also visible to the web server.

Different ways for accessing the pixel data are described below with a schematic representation given in Fig 1.

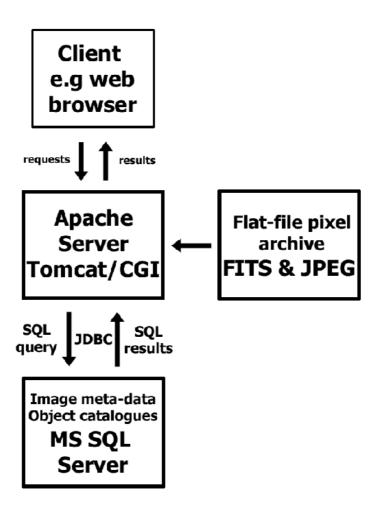


Figure 1: Schematic overview of data access

The primary access route for users will be via web forms on the WSA website. In general terms the forms will action a Java servlet that queries the relevant database pulling out any matching results. The resulting parameters (filenames, extension numbers etc) are passed to CGI scripts to perform any necessary image manipulation. Speed and efficiency is more important in the low level manipulation of the binary data and this is carried out using C code.

Some specific examples of pixel access methods are give below:

5.1 Small image extraction - pixel form 1 (PF1)

This form will enable users to extract an image cut-out from a SINGLE extension of a stored FITS file. The maximum size extracted will therefore be limited by the area covered by the extension accessed.

Boxes and menus on the form will allow the user to specify the celestial coordinates of the extraction (decimal degrees or sexagesimal), the coordinate system (J2000, B1950, Galactic), the size of the area to extract (x,y in arcmin), the waveband, and type of image to use (normal, stack, interleave etc)

When submitted this form will action a Java servlet on the server. A summary of the tasks the script will perform is given below:

- parse the inputted parameters and check they are valid. If problems are detected output message to browser and exit;
- convert inputted coordinates/system to RA and Dec (J2000 decimal degrees).
- check RA and DEC are valid i.e. $0 \le RA \le 360$ and $-90 \le DEC \le +90$, if not output message to browser and exit.
- construct the SQL query string needed to select the relevant image parameters of any matching archived FITS image from the relevant database table
- execute SQL query via JDBC
- parse the results using the AST library and WCS values to determine the pixel coordinates of the requested position.
- output table containing HTTP links, that lists matching images.
- these links action CGI scripts that use C-code to generate and display the jpegs and produce FITS images of the cut-out region.

Small image extractions are completed in real time (under 30 seconds).

5.2 Batch small image extraction - pixel form 2 (PF2)

Uploading a file of coordinates to this multi-part form provides the user with a batch mode front end to the functionality offered in Section 5.1. Users are asked to supply the path of the upload file, size of extraction, waveband, and a valid email address.

A limit needs to be placed on the total amount of pixel data that can be extracted. For the WSA this is currently set to a total area of 500 sq. arcmin.

The Java servlet actioned by this form carries out similar steps to those described above in Section 5.1 as it loops through the input file. Output is written to a temporary results file and if a given image can not be extracted the script will skip to the next one. After initial checking of the input parameters and execution of the SQL query, the actual extractions are run in the background with a message being returned to the browser informing the user that an email will be sent to them when the script has finished.

The email sent on completion provides the user with a URL where the results can be viewed an downloaded. These results include a tar saveset of the extracted FITS files and a PDF file of the jpeg images.

For the WSA in order to narrow down the number of matching images for a given coordinate the user is also be asked which survey or programme to use. The underlying SQL query is then joined with the relevant mergeLog table which ensures the optimum science frames are returned.

5.3 Large image extraction - pixel form 3 (PF3)

This form will enable users to construct an image from multiple detector frames covering an area of sky up to 1.0 degree across. However the underlying mosaicing software will be able to generate arbitrarily large areas.

The inputted parameters will be similar to PF1 (Section 5.1) but with the addition of a text box to enter the scrunch factor (pixel binning) of the returned image and a box to enter the user's email address.

This form will again action a servlet. The main difference in the processing steps outlined in Section 5.1 is that the initial SQL query constructed and sent to the image database table will return the path/filenames of all images held that overlap with the area of sky requested. This list of files together with the size and binning requested for the output image will in turn be passed to a local copy of the CASU mosaicing code that will combine the files together into a single image. As a large amount of pixel re-sampling can be involved the main work will be done in the background with the user being notified by email on completion.

5.4 Stacked image generation - pixel form 4 (PF4)

After submitting user inputted values for position and waveband this form will return a list of matching images as part of another web form. Users will then select which of the images should be stacked and supply an email address.

On submission this second form will action a CGI script that inputs the selected images to a local copy of the CASU stacking tool. Again the intensive processing is done as a background task with an email notification being sent on completion.

5.5 Browsable access to pixel data

The WSA website has (and the VSA website will have) static tables and charts displaying the contents of the archive. These pages are generated periodically usually coinciding with data releases. In addition a web form will allow users to browse and list the contents of the archive. Options on the form such as filter, date range, observation type etc. allow the listing to be refined.

Once again the form actions a Java servlet that performs the required SQL query on the relevant database. The servlet returns the lists as HTML tables. These tables contain links allowing the user to view the library jpeg images of the multiframe extensions and download the FITS image file and any associated FITS catalogue.

5.6 Other access to pixel data

The functionality offered in Section 5.1 can be made compatible with the GAIA (based on SkyCAT) and Aladin tools. The underlying servlet will be very similar, (or indeed possibly the same but with different options being executed). The main difference being that the generated FITS image is piped directly back to the querying tool. Essentially this method just exposes a queryable URL that could be accessed directly from the command line (e.g. using wget) allowing users to create their own batch extractions.

The methods outlined in this section will only access data classed as world readable.

5.7 Demonstrations of pixel access

The single and batch image cut-out access methods (PF1 and PF2) have been implemented in the WSA

Access to SuperCOSMOS Sky Surveys (SSS) data has been made available from within GAIA (under Data-Servers > Browse Catalog Directories... and open/expand SuperCOSMOS catalogues) and Aladin (under load > image servers > others).

The queryable URL functionality described in Section 5.6 has been supplied to the 6dF observers who routinely use it to generate SSS finders for checking target objects (3000-10000 extractions per week).

Screenshots of some of the access methods described above are provided in the Appendix (Section 10).

6 OBJECT CATALOGUE DATA

Object catalogue data are stored in SQL Server database tables on a Windows machine networked to the web server. The basic recipe for access is described below:

- Java servlet running under Tomcat on web server parses and checks parameters supplied by user
- servlet carries out any necessary coordinate manipulation/conversion
- servlet constructs SQL query based on inputted parameters and required function.
- servlet connects to SQL server using JDBC and submits query
- SQL server performs query and returns results
- servlet reads results and parses data into desired output format
- servlet outputs results to user either as html, links to files or directly as stream.

Note that the intensive part of any query is carried out by SQL Server. Much of the differences in functionality of the access methods described below lie in the construction of the SQL query, the code for executing the query is largely generic and re-usable.

The amount of data returned by a query and written to file needs to be restricted. For the WSA this is currently set at 15 million cells (e.g. a million rows with 15 attributes).

Again the primary access method is via web forms. A user's login status is used to ensure they are authorised to query the requested database.

6.1 Radial Search - Object Form 1 (OF1)

This form enables a user to extract objects within a specified distance of a supplied position (RA/DEC, Galactic) or object name. Other options passed from the form to the servlet are: which survey to search, which parameters to extract (all, subset or user specified), any additional constraints (used to form the SQL WHERE clause), the format of the output data (HTML, delimited ascii file, binary FITS table, VOTable).

The actioned servlet first converts the inputted coordinates or resolves the inputted name (using SIMBAD) to an RA and DEC in J2000 decimal degrees. An SQL query is then formed that efficiently searches through the required table using the indexes on RA and Dec.

The servlet submits the query as a separate thread, with the main thread keeping the browser connection active and checking that it hasn't been stopped by the user.

On completion the query thread parses and formats the returned rows, writes the data to file if requested and prints any HTML table output and links to files to the browser window.

6.2 Rectangular Search - Object Form 2 (OF2)

Similar to the radial search but this time an SQL function is used that extracts objects bounded by limits in RA and DEC (or the coordinate system requested).

6.3 SQL query - Object Form 3 (OF3)

There are two versions of this form. The first uses drop down menus and text boxes to guide the user in building an SQL query for execution by the servlet. Users are able to choose/input values to construct the SELECT, FROM and WHERE clauses of SQL. An option enables the querying of a second table joined with the primary table. The secondary table can be one of the non-WFCAM based tables held in the archive e.g. 2MASS or SDSS. Joins are made via the neighbours table for a given combination.

The second version is based around a text box into which users with knowledge of SQL and the contents of the WSA or VSA (which will be documented) can directly input their SQL query.

Output options for both versions will be as OF1. Results of long-running queries are sent by email to users.

6.4 Cross-matching a catalogue - Object Form 4 (OF4)

This multipart form offers the user the ability to upload a list of coordinates and match them against tables held in the database.

Users specify the pairing radius and whether they want just the nearest object or all matching objects extracted.

Unlike the previous object catalogue forms several SQL statements are executed by the servlet. The first of these creates a temporary database table, that table is then populated with the contents of the upload file. Finally the requested database table and temporary table are paired. The temporary table is automatically dropped when the JDBC connection is closed.

6.5 Browsable access to catalogue data

As detailed in Section 5.5 the archive listing form provides a browsable way to reach links to the object catalogues which are stored as FITS binary tables (generated from and associated with a given FITS image file by the CASU standard source extraction tool).

6.6 Other access to catalogue data

The functionality offered in Section 6.1 will be made compatible with the GAIA and Aladin tools. Some of the options available on the web form will be hard-wired to sensible defaults for this implementation (e.g. the parameters returned) as the interface, especially with GAIA, is not fully configurable. As previously mention in Section 5.6 the underlying queryable URL can be made accessible from the command line using wget.

Again the methods outlined in this section will only access data classed as world readable.

6.7 Demonstrations of catalogue access

The core functionality described in Sections 6.1, 6.3 and 6.4 is implemented in the WSA under:

http://surveys.roe.ac.uk:8080/wsa/region_form.jsp

http://surveys.roe.ac.uk:8080/wsa/menu_form.jsp

http://surveys.roe.ac.uk:8080/wsa/SQL_form.jsp

http://surveys.roe.ac.uk:8080/wsa/crossID_form.jsp

respectively.

Cross-matching an uploaded file of 1000 records with the UKIDSS Large Area Survey takes approximately 30 seconds.

Screenshots of some of the examples described above are provided in the Appendix (Section 10).

7 VIRTUAL OBSERVATORY CONSIDERATIONS

This topic will be covered in detail AD03. However it is worth noting here that VOTable format is one of the output options for the object catalogue access.

8 HOUSEKEEPING

Temporary output files generated by user requests are written to a publicly (HTTP) accessible area of the file system. A cron job run daily and deletes any of these files more than 48 hours old.

Log files and database tables of user access are archived and used to monitor and generate statistics on archive usage (hits, queries, data volume served). In the two months following the WSA DR1 release some 2800 SQL queries were made via OF2 and nearly 0.5 billion rows of data were returned. In the same period just over 900 archived FITS files were downloaded by users via the archive listing form.

9 USER DOCUMENTATION

The WSA website has and VSA websites will have extensive online documentation to help the user including:

- Information specific to each access method
- Links to explanations of options/input on web forms
- ullet Example usage and queries (e.g use cases "how do I extract ..." and filled in forms)
- Detailed description of database contents e.g. tables, parameters (type, range), functions, constraints
- SQL tutorial
- FAQ
- E-mail contact for support and feedback

9.1 Schema Browser

The database contents of the archives are documented in detail in the Schema Browser section of the website. Users can navigate through the database design viewing documentation on schemas, tables, views, columns and functions. For a given attribute or column the following information is available: name, type, length, unit, description, default value and Unified Content Descriptor (UCD). Where necessary more detailed information on a given attribute is provided via a link to the glossary section of the Schema Browser.

10 APPENDIX

This section shows screenshots of some of the ways users can access the archives held at WFAU.



Figure 2: Screenshot of the WSA (PF1) pixel web form.

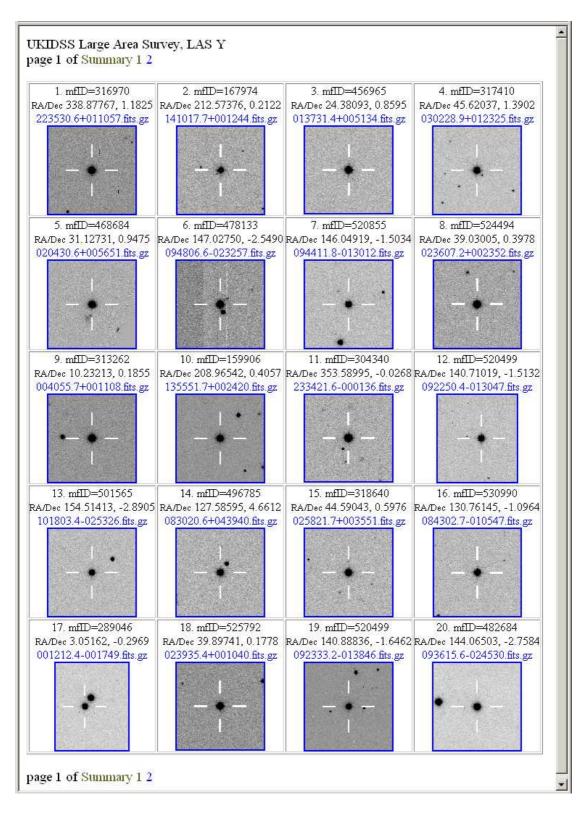


Figure 3: Screenshot of results from the WSA batch pixel web form (PF2).

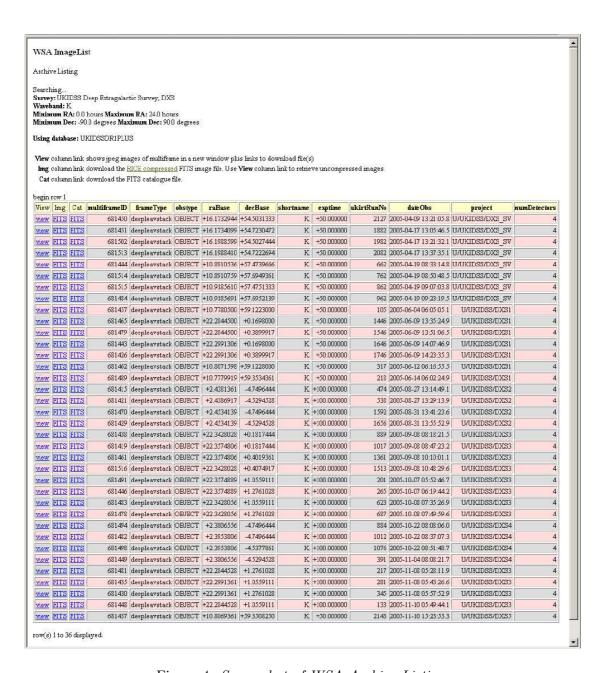


Figure 4: Screenshot of WSA Archive Listing.

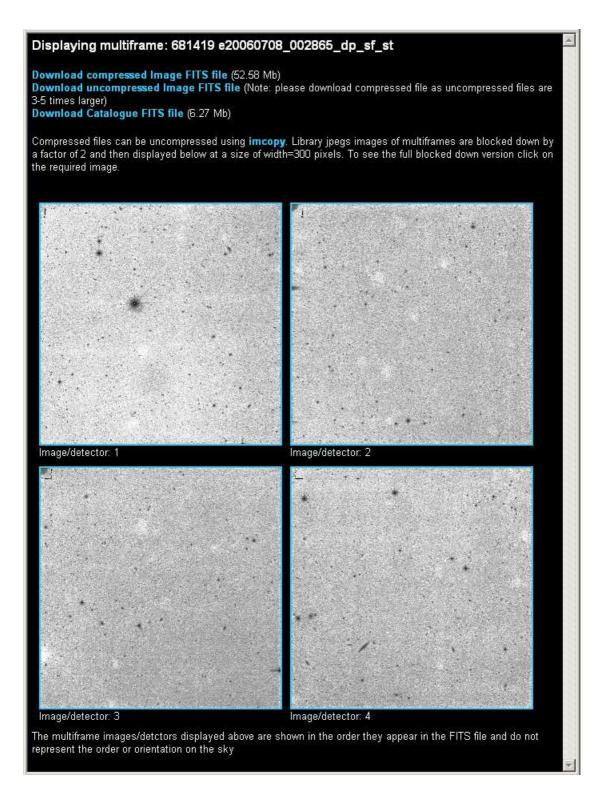


Figure 5: Screenshot displaying WSA multiframe from an archive listing.

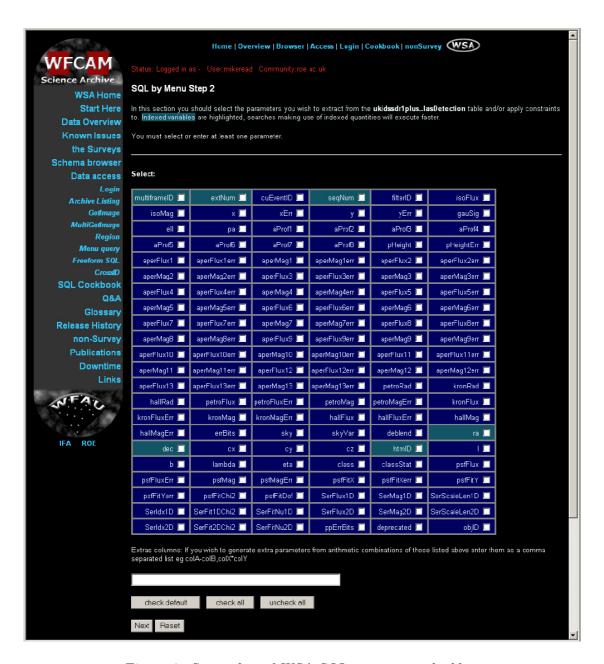


Figure 6: Screenshot of WSA SQL menu query builder.

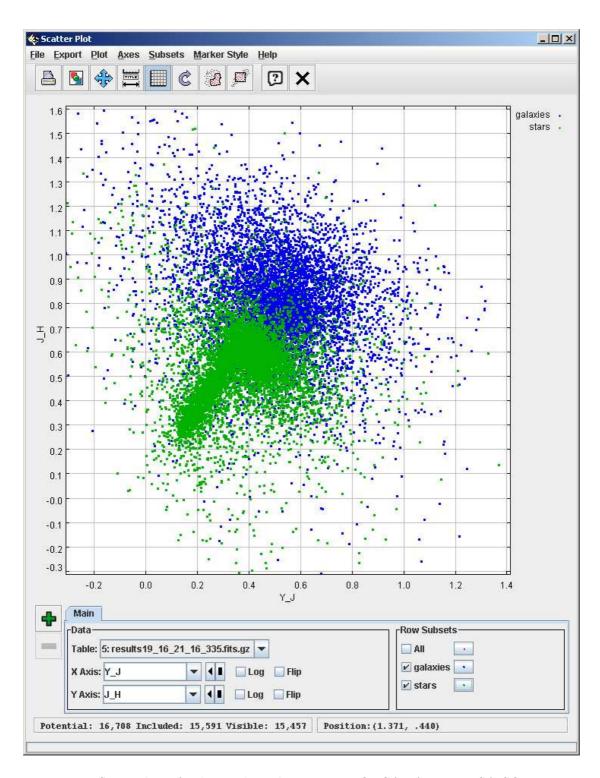


Figure 7: Screenshot of colour-colour diagram in TOPCAT from a WSA SQL query.



Figure 8: Screenshot of WSA Schema Browser.

11 ACRONYMS & ABBREVIATIONS

6dF: Six-degree field

6dFGS : Six-degree field Galaxy Survey ADnn : Applicable Document No nn CGI : Common Gateway Interface

CASU: Cambridge Astronomical Survey Unit

DBD : Database Driver DBI : Database Interface

DXS: Deep Extragalactic Survey JDBC: Java Database Connectivity

 $\begin{array}{l} LAS: Large\ Area\ Survey \\ LWP: LibWWW-Perl \end{array}$

HTML: HyperText Markup Language

HTTP: Hypertext Transfer Protocol SOAP: Simple Object Access Protocol SQL: Structured Query Language SSS: SuperCOSMOS Sky Surveys

SSA: SuperCOSMOS Science Archive UDS: Ultra Deep Survey

UKIDSS: UKIRT Infrared Deep Sky Survey

VIRCAM: VISTA InfrarRed Camera VISTA: Visible and Infrared Survey Telescope for Astronomy

VPO: VISTA Project Office

W3C : World-Wide Web Consortium

WFAU: Wide Field Astronomy Unit (Edinburgh)

 $\begin{aligned} & \text{WFCAM}: \text{Wide-Field Camera} \\ & \text{XML}: \text{eXtensible Markup Language} \end{aligned}$

12 APPLICABLE DOCUMENTS

AD01	Science Requirements Analysis Document	VDF-WFA-VSA-002
		Issue: 1.0 09/06
AD02	Database Design Document	VDF-WFA-VSA-007
		Issue 1.0 09/06
AD03	Virtual Observatory integration	VDF-WFA-VSA-010
		Issue 1.0 09/06

13 CHANGE RECORD

Issue	Date	Section(s) Affected	Description of Change/Change Request
			Reference/Remarks
1.0	06/09/06	All	New document based on VDF-WFS-WSA-008

14 NOTIFICATION LIST

The following people should be notified by email whenever a new version of this document has been issued:

WFAU: P Williams, N Hambly CASU: M Irwin, J Lewis

QMUL: J Emerson ATC: M. Stewart JAC: A. Adamson

UKIDSS: S. Warren, A. Lawrence