THE 6DF GALAXY SURVEY: PROGRESS AND DATA RELEASE 1

Will Saunders (AAO), Heath Jones (RSAA), Mike A. Read (IfA), Lachlan A. Campbell (RSAA), Matthew Colless, Malcolm Hartley, Paul Cass, Dionne James, Ken Russell, Kristin Fiegert, Fred G. Watson (AAO), Thomas Mauch (Sydney), John Dawe (RSAA) and the 6dFGS Science Advisory Group

INTRODUCTION

Wide-scale redshift surveys such as the 2dF Galaxy Redshift Survey and Sloan Digital Sky Survey have made significant advances in our understanding of the matter and structure of the wider universe. However, their optical selection inevitably biases them in favour of currently star-forming galaxies. They also have fields of view too small to allow full sky coverage, limiting their utility for dynamical and cosmographic studies, and the median distances of the galaxies observed are too great to allow useful peculiar velocities to be obtained. The 6dF Galaxy Survey (6dFGS) is a dual redshift/peculiar velocity survey of 150,000 galaxies that endeavours to overcome the limitations of the 2dFGRS and SDSS surveys in these areas.

To obtain a true three-dimensional distribution of galaxies it is necessary to determine the peculiar velocity as well as the redshift. We intend to measure peculiar velocities for a 15000 galaxy subsample, in order to directly probe the mass distribution of the universe and matter densities, large-scale galaxy biasing, and the amplitude of mass fluctuations in the fabric of the local universe. The primary sample of 114000 galaxies is K_s-selected from the 2MASS survey (Jarrett et al. 2000), and the magnitudes used in the selection are effectively total magnitudes. We selected all sources from the 2MASS Extended Source Catalog, version 3 (2MASS XSC; Jarrett et al. 2000), with $|b| \geq 10^o$ and $\delta < 0^o$, and with $K_{cor} < 12.75$ (Nov 2002 newsletter; Jones et al. 2004). We also include 16 additional samples at lower priority for very little observational overhead (details in AAO Newsletter 101, November 2000). Literature redshifts have been incorporated into the redshift catalogue, from ZCAT (Huchra et al. 1999) and the 2dF Galaxy Redshift Survey (Colless et al. 2000). Roughly half the sample is early type.

The project is now much more than half completed, with 881 fields observed by the end of 2003. At an expected rate of 400–450 fields/year, we are on target to meet the survey cutoff date of mid-2005, beyond which UK Schmidt funding is not currently secured.

Field Placement and Tiling Algorithm

The survey area is 17046 deg^2, and the 1360 6dF fields (5.7^o-diameter) contain a mean of 124 sources per field and cover the sky twice over. An adaptive tiling algorithm was employed to distribute the fields ("tiles") across the sky to maximise uniformity and completeness, described in full in Campbell et al. (2004). In order to achieve consistent completeness, and to counter the tendency to oversample galaxy clusters, it was necessary to inversely weight each galaxy by the local galaxy surface density. This resulted in a very small penalty in overall completeness except in the heart of the Shapley supercluster, where galaxy densities are orders of magnitude higher than elsewhere, and we added 10 tiles by hand in this region.

The original 2002 tiling of the 6dFGS catalogue (version A) was revised in February 2003 (version D), due to the higher-than-expected rate at which fibres were broken and temporarily lost from service, and a major revision in the primary sample itself from IPAC. Tests of the two-point correlation function were made on the sample selected through the final tiling allocation, to check for systematic effects. Mock catalogues were generated, using the correlation function observed by the 2dF Galaxy Redshift Survey (Hawkins et al. 2003), and tiled as the real data. The resulting 2-point correlation function was determined and compared with the original. This revealed an undersampling on scales up to ~1 h Mpc^{-1}, clearly the result of the fibre button proximity limit. No bias was seen on larger scales.

Theoretical tiling completenesses of around 95% were achievable for all except the lowest priority samples, and variations in uniformity were confined to < 5%. However, fibre breakages have meant that 6dFGS has consistently run with many fewer fibres than anticipated, hitting the completeness of the lower priority samples in particular. Since we have a fixed timeline for the survey (mid 2005) and a fixed number of fields to observe, there is little we can do about this.

SURVEY IMPLEMENTATION

Observational Technique

Field acquisition with 6dF is carried out using four guide-fibre bundles, each of seven 100μm (6.7 arcsec) diameter fibres, giving a compact configuration ~20 arcsec in diameter. The use of acquisition fibres of the same diameter as the science fibres is sub-optimal, but in practice we achieve good alignment, particularly as we choose guide stars near the edge of the field. Guide stars are selected from the Tycho-2 catalogue (Hoeg et al. 2000), and have magnitudes typically in the range 8 < V < 11.

Each field is observed for 1 hr in V and 0.5 hr in R, although these times are increased in poor observing conditions. Calibration observations consist of quartz lamp flat fields and wavelength calibration lamp spectra (HgCd + Ne in R and HgCd + He in V). The spliced spectra have S/N ~ 10 $pixel^{-1}$, yielding > 90% redshift completeness. Three to five survey fields are observed on a clear night, depending on season. With 75% of the UKST time assigned to 6dFGS, and an average clear fraction of 60%, we observe about 400 fields per year. The observational strategy is to divide the sky into three declination strips. Initially, the survey has concentrated on the $\delta = -30^{o}$ declination strip (actually $-42^{o} < \delta < -23^{o}$); the equatorial strip ($-23^{o} < \delta < 0^{o}$) will be done next, and then finally the polar cap ($\delta < -42^{o}$).

Though observations started in June 2001, the data suffered from various problems with the instrument and the source data, and the 2001 data is not included in this data release. Initial observations were carried out at mid-latitudes for observational convenience, with the actual band corresponding to one of the Additional Target samples. Other Additional Target areas were included where programs had been assigned separate telescope time and the observations could be folded into 6dFGS.

Reduction of Spectra

The reduction of the spectra uses a modified version of the 2DFDR package developed for the 2dF Galaxy Redshift Survey. Unlike 2dF data, tramline fitting is done completely automatically, using the known gaps in the fibres to uniquely identify the spectra with their fibre number. We don't correct for scattered light unless there is a known problem, such as oil-contamination in the dewar. The extracted spectra for each field are combined, usually weighted by S/N to cope with variable conditions. Satisfactory observations usually have a S/N per pixel ≥ 10 in V and R. All data are crudely fluxed using 6dF observations of the standard stars Feige 110 and EG274; fibre to fibre transfer variations are corrected for by the flat-fielding and we assume a constant average spectral transfer function for each plate. The resulting R and V spectra for each source are then spliced together, using the overlapping region to match their relative scaling. We also rebin the lower dispersion R data in order to avoid a discontinuity in dispersion at the join.

Redshift Measurement

Accurate redshift measurement is a fundamental component of both the *z* and *v*-surveys. We started with the semi-automated redshifting RUNZ software used by the 2dF Galaxy Redshift Survey (Colless et al. 2001), and kindly provided by Will Sutherland. Extensive

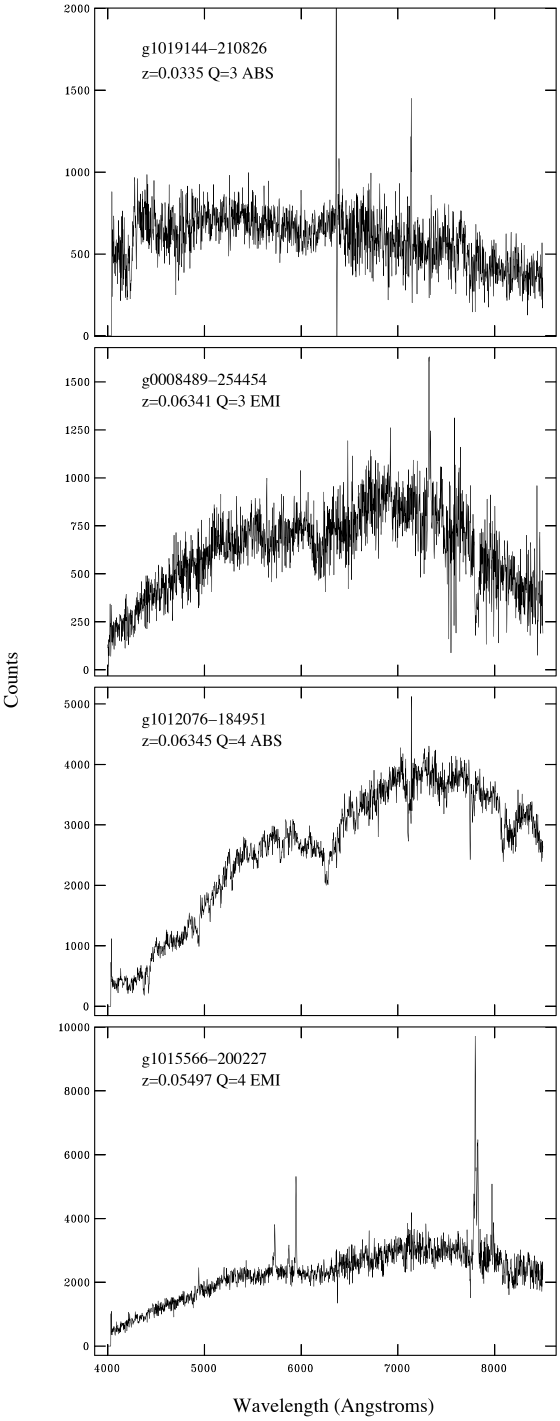

Figure 1. Examples of galaxy spectra from the 6dFGS exhibiting a range of types and redshift quality *Q*.

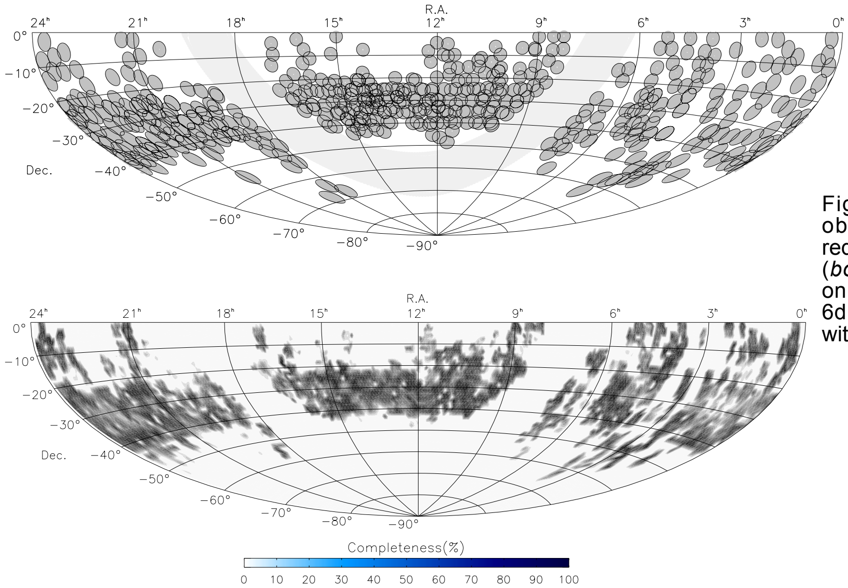

Figure 2. (*top*) Location of the observed fields contributing redshifts to the First Data Release. (*bottom*) Redshift completeness on the sky, combining the 44072 6dF First Data Release redshifts with the 24854 ZCAT sources.

modifications were made in order to accept 6dF data. The version used for 2dF determined the redshift separately for absorption and emission features, which improves the reliability but reduces the accuracy of the measurements. Since the line identification of the higher S/N and higher dispersion 6dF spectra was usually not in doubt, we decided to use both emission and absorption features to determine the cross-correlation redshifts; and in general the cross-correlation redshift is used in preference to the emission-line redshift.

Each redshift measurement is checked visually for errors introduced by spurious spectral features such as fibre interference patterns or poor sky subtraction. Occasionally, the redshift is redetermined manually, but in the majority of cases, the automated redshift value is accepted without change. The final redshift value is assigned a quality, Q, between 1 and 5 where Q=3,4,5 for redshifts included in the final catalogue. Q=4 represents a reliable redshift while Q=3 is assigned to probable redshifts; Q=2 is reserved for tentative redshift values and Q=1 for spectra yielding no value. Q=5 signifies a reliable redshift and very high signal-to-noise spectrum, although was used rarely for the 6dFGS. Figure 1 shows a few examples of galaxy spectra across the range of redshift quality, for both emission and absorption-line spectra.

The vast majority (76%) of the 6dFGS redshifts have Q=4 from spectra with a median signal-noise ratio of 9.4 pixel^{-1}. There are only 24 Q=5 sources. For Q=3 redshifts the median signal-to-noise ratio drops to 5.3 pixel^{-1}, indicating minimum of range of redshift-yielding spectra. The median signal-to-noise ratio for Q=2 redshifts (6.3 pixel^{-1}) is slightly higher than that for Q=3 (5.3 pixel^{-1}). This is due to the significant number of Galactic sources such as stars, the ISM, and planetary nebulae, which produce high signal-to-noise spectra, but are assigned Q=2 on account of their zero redshift.

FIRST DATA RELEASE

Statistics and plots

Between January 2002 and July 2003, the 6dF Galaxy Survey obtained spectra for 52355 sources from which 44186 redshifts were derived. Of these, 44072 were reliable redshifts with quality $Q \geq 3$ and $cz \geq 600$ km s^{-1}. Of the 113988 galaxies in the original sample, 31367 had existing literature redshifts: 24854 from the ZCAT compilation (Huchra et al. 1999) and 6513 from the 2dF Galaxy Survey (Colless et al. 2000).

Data from 524 fields have contributed to the first data release. As shown in Figure 2 (*top*), the majority of these occupy the central declination strip between $-42^{\circ} < \delta < 23^{\circ}$. Figure 2 (*bottom*) shows the corresponding distribution of redshift completeness on the sky for the K-band sample. The *redshift completeness*, R, is that fraction of galaxies in the parent catalogue of 171000 with acceptable ($Q \geq 3$) redshifts in a given area of sky, from whatever source. Of the first ~41000 sources observed with 6dF, around 3% were stars, 1% were PNe/ISM and 11% failed to yield a redshift.

The *field completeness* is the ratio of acceptable 6dF extragalactic redshifts in a given field to initial sources. Figure 3 shows the distribution of field completeness from the first 523 fields. This demonstrates that the redshift success rate of 6dF is very high, with both the median and mean completeness around 90%. Moreover, 95% of the fields are 65% complete or better, meaning

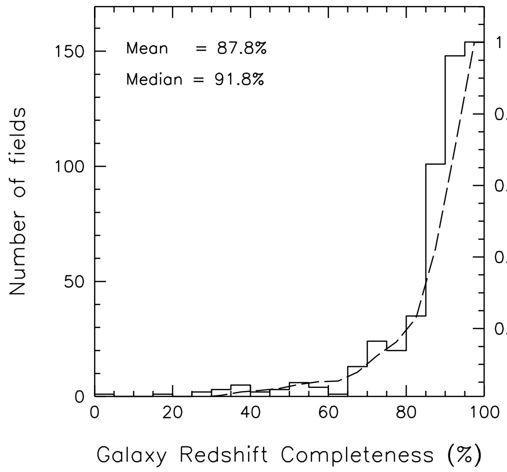

Figure 3. Galaxy redshift completeness by field, where completeness is the number of 6dF redshifts over the total 6dF redshifts and failures. The dashed line indicates the cumulative fraction as indicated at right.

that relatively few need to be recycled for repeat observation. The apparent discrepancy between the high *field* completeness values of Figure 3 and the lower *redshift* completeness in Figure 2 (*bottom*) is due to the high degree of overlap in the 6dFGS field allocation. Figure 2 (*top*) gives the impression of a single layer of tiles while in fact most parts of the sky need to be tiled two or more times over.

The distribution of 6dFGS redshifts exhibits the classic shape for magnitude-limited surveys of this kind (Figure 4). The median survey redshift, <cz> = 16008 km s^{-1} (|z| = 0.055), is less than half that of the 2dFGRS or SDSS surveys.

Figure 5 shows the radial distribution of galaxies across the southern sky, projected across the full range of southerly declinations (δ = 0 to –90°). Projecting in this way has the drawback of taking truly separate 3D space structures and blending them on the 2D page. However, as most of the redshifts to date come from the central declination strip the projection effect is limited for now. Other variations in galaxy density seen in Figure 5 are due to the incomplete coverage of observed fields and the projection of the Galactic Plane. No 6dFGS galaxies lie within galactic latitude |b| ≤ 10° as is evident in the two sparsely populated sectors at ~ 8 hr and ~17 hr right ascension. The 6dFGS is also clumpier than optically-selected redshift surveys such as 2dFGRS and SDSS. This is because the near-infrared selection is biased towards early-type galaxies, which cluster more strongly than spirals.

The 6dFGS provides the largest sample of near-infrared selected galaxies to determine the fraction of mass in the present-day universe existing in the form of stars. Jones et al. (2004) have derived the *J* and K_s-band luminosity functions from the first 51000 redshifts of the 6dF Galaxy Survey, combining data from both before and after the First Data Release. Using the near-infrared luminosity functions and stellar population synthesis models, the galaxy stellar-mass function for the local universe can be estimated. When this is integrated over the full range of galaxy masses, the total mass of the present-day universe in stars can be expressed in units of the critical density.

6dFGS Online Database

Data from the 6dF Galaxy Survey are publicly accessible through an online database at http://www-wfau.roe.ac.uk/6dFGS/, and maintained by the Wide Field Astronomy Unit of the Institute for Astronomy, University of Edinburgh. An early data release of 17K redshifts was made in December 2002, along with the opening of the web site and tools for catalogue access. This paper marks the First Data Release of 44K redshifts measured between January 2002 and July 2003. The design of the database is similar to that used for the 2dF Galaxy Redshift Survey in that parameterised data are stored in a relational database. Each TARGET object is also represented by a multi-extension FITS file which holds thumbnail images of the object and the spectra. The database is accessed/queried using Structured Query Language (SQL). A combined 6dF-literature redshift catalogue is provided in a separate single master catalogue.

The 6dFGS database is housed under Microsoft's relational database software, *SQL Server 2000*. The data are organised in several tables (Table 1). The master

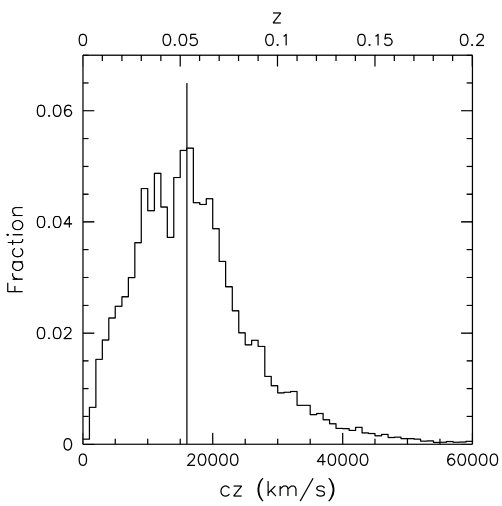

Figure 4. Distribution of redshifts for the 44072 6dF First Data Release galaxies with redshift quality *Q* ≥ 3 and *cz* > 600 km s^{-1}. The mean redshift for the survey (<cz> = 16008 km s^{-1}) is indicated with a vertical solid line.

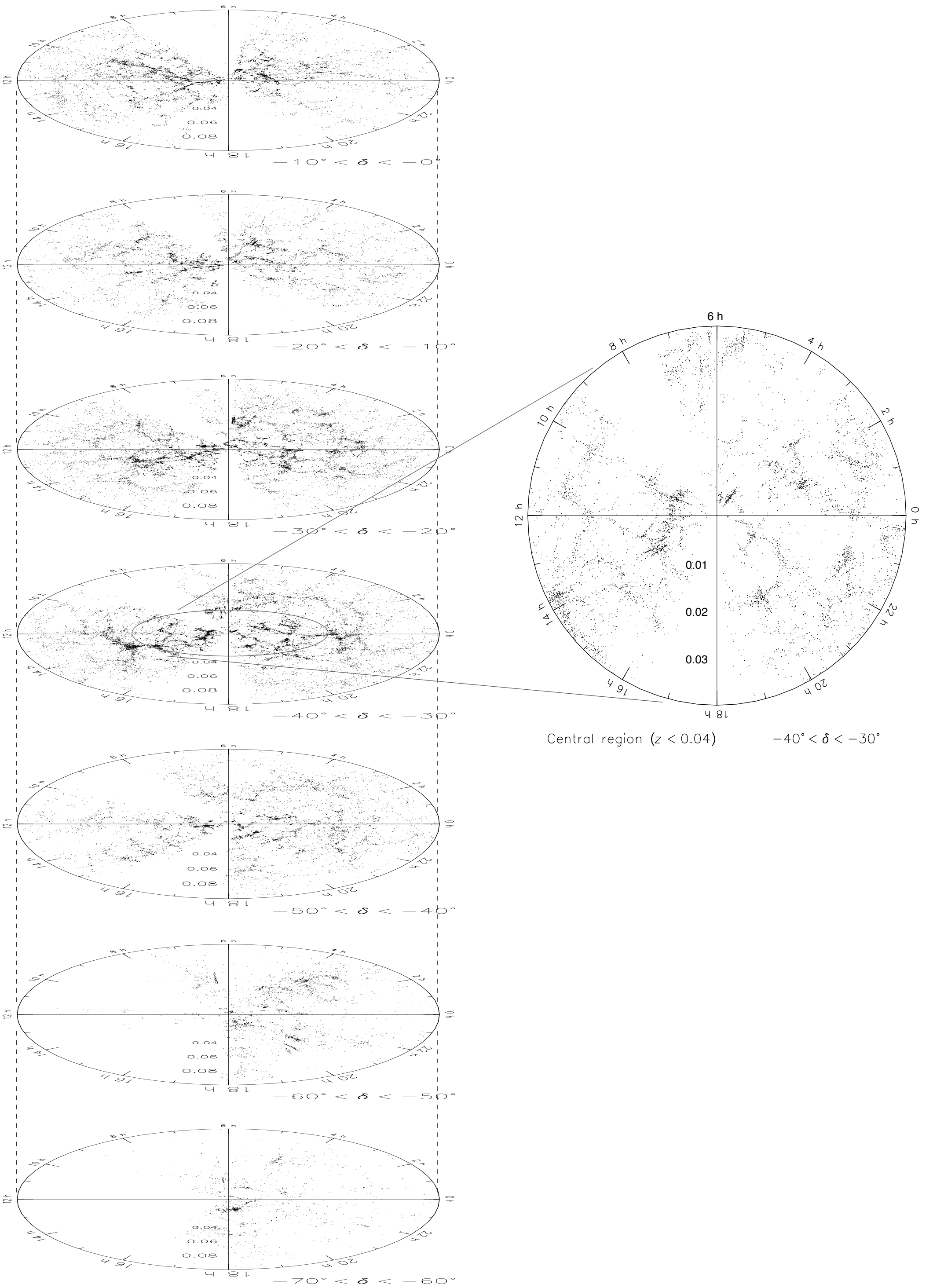

Figure 5. Spatial redshift distribution combining the 44072 6dF First Data Release redshifts with the 24854 of ZCAT. The redshift slice projects through all southerly declinations, $\delta < 0^o$. The sparse sampling around 8 and 17 hr is due to non-coverage in the Galactic Plane. Variations elsewhere in the sky are due to different sky regions having different observational completenesses at the time of the First Data Release.

target list used to configure 6dFGS observations is represented by the TARGET table. Spectral observations are stored in the SPECTRA table. The input catalogues that were merged to make up the master target list are also held in individual tables (TWOMASS, SUPERCOS etc.). The TARGET table forms the hub of the database. Every table is interlinked via the parameters targetid and targetname. These parameters are unique in the master TARGET table but are not necessarily unique in the other tables, (e.g. SPECTRA) as objects can and have been observed more than once. The SPECTRA table holds all the observational and redshift related data. Parameters are recorded for both the V and R frames (with a lot of the values being the same for both frames), and redshift information is derived from the combined VR frame. The TWOMASS table contains the *K*, *J* and *H*-selected samples originating from the 2MASS extended source catalogue. The *K*-selected sample represents the primary 6dFGS input catalogue.

Initially every FITS file, representing each target (targetname.fits), holds thumbnail images of the target. As data are ingested into the database the reduced spectra are stored as additional FITS image extensions. Table 1 summarises the content within each FITS extension. The first 5 extensions contain the thumbnail images and each has a built-in World Coordinate System (WCS). The optical *B* and *R* images come from SuperCOSMOS scans of blue B_j and red R_f survey plates. The 2MASS *J*, *H* and *K* images were extracted from datacubes supplied by IPAC. Note that although some objects in TARGET do not have 2MASS images, the corresponding extensions still exist in the FITS file but contain small placeholder images. The remaining extensions contain the spectra. Each 6dFGS observation will usually result in a further 3 extensions, the V band spectrum, the R band spectrum and the combined/spliced VR spectrum.

Table 1. Contents of each extension in the database FITS files

FITS Extension	Contents
1st	SuperCOSMOS B_j image (1 x 1 arcmin)
2nd	SuperCOSMOS R_f image (1 x 1 arcmin)
3rd	2MASS *J* image (variable size)
4th	2MASS *H* image (variable size)
5th	2MASS *K* image (variable size)
6th	V-spectrum extension
7th	R-spectrum extension
8th	combined VR-spectrum extension
*n*th	additional V, R, and VR data

The V and R extensions are images with 3 rows. The 1st row is the observed reduced SPECTRUM, the 2nd row is the associated variance and the 3rd row stores the SKY spectrum as recorded for each data frame. Wavelength information is provided in the header keywords CRVAL1, CDELT1 and CRPIX1, such that:

Wavelength (A) = CRVAL1 – CRPIX1 - (pixel number) x DELT1 .

Additional WCS keywords are also included to ensure the wavelength information is displayed correctly when using image browsers such as Starlink's GAIA or SAOimage DS9.

The VR extension also has an additional 4th row that represents the WAVELENGTH axis, which has a continuous dispersion, achieved through the continuation of the V dispersion into the R half from rescrunching.

Access to the database is through two different Hypertext Mark-up Language (HTML) entry forms. Both parse the user input and submit an SQL request to the database. For users unfamiliar with SQL, the menu driven form provides guidance in constructing a query. The SQL query box form allows users more comfortable with SQL access to the full range of SQL commands and syntax. Both forms allow the user to select different types of output (HTML, comma separated value (CSV) or a TAR save-set of FITS files).

There are online examples of different queries using either the menu or SQL form at http://www-wfau.roe.ac.uk/6dFGS/examples.html. More information about the database is available directly from the 6dFGS database website.

References

Burkey, D. & Taylor, A., 2004, MNRAS, submitted
Campbell, L. A. et al., 2004, MNRAS, in press
Colless, M. et al., (2dFGRS team), 2001, MNRAS, 328, 1039
Colless, M. et al., 2001, MNRAS, 321, 277
Hawkins, E. et al., (2dFGRS team), 2003, MNRAS, 346, 78
Hoeg, E. et al., 2000, A&A, 355, L27
Jarrett, T.-H. et al., 2000, AJ 120, 298
Jones, D. H. et al., 2004, ApJ, submitted.
Jones, D. H. et al., 2004, in prep.
Metropolis, N. et al., 1953, J. Chem. Phys., 21
Saunders W. et al., 2001, AAO Newsletter, 97, 14
Saunders W. et al., 2002, AAO Newsletter, 101, 10