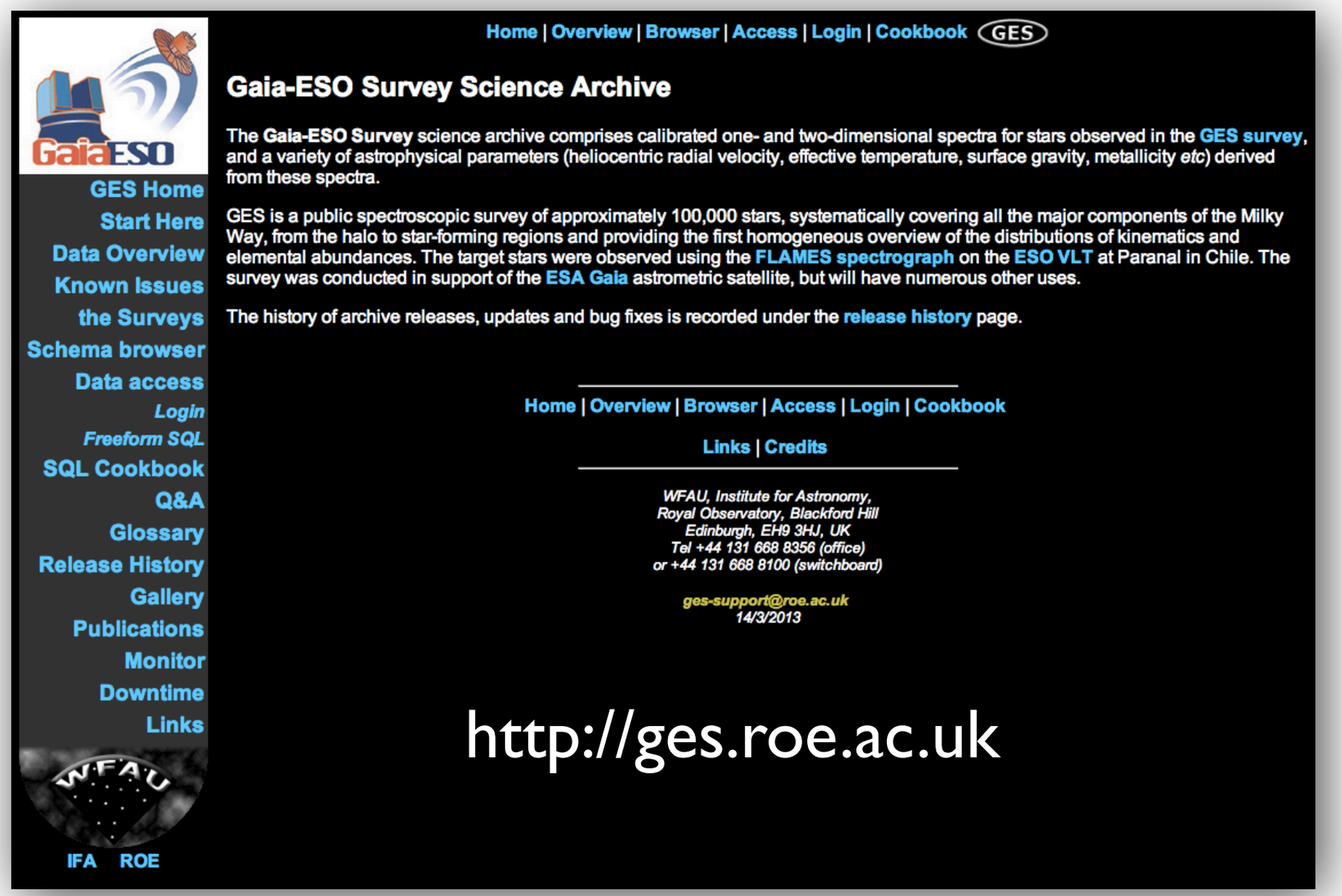
The GES Science Archive

Ross Collins

Nigel Hambly, Clive Davenhall, Mike Read, Eckhard Sutorius
WFAU, Institute for Astronomy, Royal Observatory Edinburgh



http://ges.roe.ac.uk

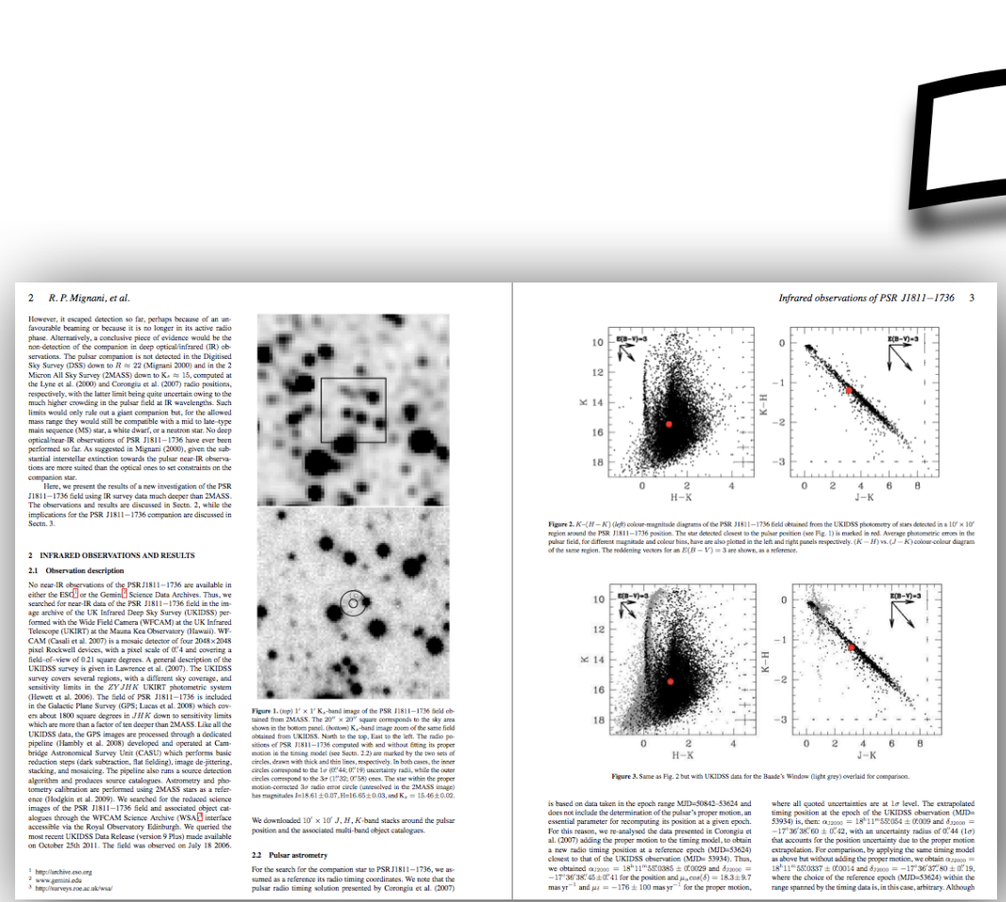
WFAU's Role

- Long-term collaboration between Cambridge's Astronomical Survey Unit (CASU) and WFAU led to the creation of the WFCAM & VISTA Science Archives (the VISTA Data Flow System or VDFS)

- Adopt similar model for GES:
  - WFAU mirrors data stored at Cambridge
  - Creates a rich, queryable database for easy data mining
  - Releases static, referencable databases with a web interface

Release databases

- Static release databases are there to be referenced (no more data reconstruction by measuring figures in papers)



2 *R. P. Mignani, et al.*

However, it escaped detection so far, perhaps because of an unfavourable beaming or because it is no longer in its active radio phase. Alternatively, a conclusive piece of evidence would be the non-detection of the companion in deep optical/infrared (IR) observations. The pulsar companion is not detected in the Digitised Sky Survey (DSS) down to $R \approx 22$ (Mignani 2000) and in the 2 Micron All Sky Survey (2MASS) down to $K_s \approx 15$, computed at the Lyne et al. (2000) and Corongiu et al. (2007) radio positions, respectively, with the latter limit being quite uncertain owing to the much higher crowding in the pulsar field at IR wavelengths. Such limits would only rule out a giant companion but, for the allowed mass range they would still be compatible with a mid to late-type main sequence (MS) star, a white dwarf, or a neutron star. No deep optical/near-IR observations of PSR J1811−1736 have ever been performed so far. As suggested in Mignani (2000), given the substantial interstellar extinction towards the pulsar near-IR observations are more suited than the optical ones to set constraints on the companion star.

Here, we present the results of a new investigation of the PSR J1811−1736 field using IR survey data much deeper than 2MASS. The observations and results are discussed in Sects. 2, while the implications for the PSR J1811−1736 companion are discussed in Sects. 3.

2 INFRARED OBSERVATIONS AND RESULTS

2.1 Observation description

No near-IR observations of the PSRJ1811−1736 are available in either the ESO[1] or the Gemini[2] Science Data Archives. Thus, we searched for near-IR data of the PSR J1811−1736 field in the image archive of the UK Infrared Deep Sky Survey (UKIDSS) performed with the Wide Field Camera (WFCAM) at the UK Infrared Telescope (UKIRT) at the Mauna Kea Observatory (Hawaii). WFCAM (Casali et al. 2007) is a mosaic detector of four 2048×2048 pixel Rockwell devices, with a pixel scale of 0.″4 and covering a field-of-view of 0.21 square degrees. A general description of the UKIDSS survey is given in Lawrence et al. (2007). The UKIDSS survey covers several regions, with a different sky coverage, and sensitivity limits in the *ZYJHK* UKIRT photometric system (Hewett et al. 2006). The field of PSR J1811−1736 is included in the Galactic Plane Survey (GPS; Lucas et al. 2008) which covers about 1800 square degrees in *JHK* down to sensitivity limits which are more than a factor of ten deeper than 2MASS. Like all the UKIDSS data, the GPS images are processed through a dedicated pipeline (Hambly et al. 2008) developed and operated at Cambridge Astronomical Survey Unit (CASU) which performs basic reduction steps (dark subtraction, flat fielding), image de-jittering, stacking, and mosaicing. The pipeline also runs a source detection algorithm and produces source catalogues. Astrometry and photometry calibration are performed using 2MASS stars as a reference (Hodgkin et al. 2009). We searched for the reduced science images of the PSR J1811−1736 field and associated object catalogues through the WFCAM Science Archive (WSA)[3] interface accessible via the Royal Observatory Edinburgh. We queried the most recent UKIDSS Data Release (version 9 Plus) made available on October 25th 2011. The field was observed on July 18 2006.

[1] http://archive.eso.org
[2] www.gemini.edu
[3] http://surveys.roe.ac.uk/wsa/

Figure 1. (*top*) $1' \times 1'$ K_s-band image of the PSR J1811−1736 field obtained from 2MASS. The $20'' \times 20''$ square corresponds to the sky area shown in the bottom panel. (*bottom*) K_s-band image zoom of the same field obtained from UKIDSS. North to the top, East to the left. The radio positions of PSR J1811−1736 computed with and without fitting its proper motion in the timing model (see Sects. 2.2) are marked by the two sets of circles, drawn with thick and thin lines, respectively. In both cases, the inner circles correspond to the 1σ (0.″44; 0.″19) uncertainty radii, while the outer circles correspond to the 3σ (1.″32; 0.″58) ones. The star within the proper motion-corrected 3σ radio error circle (unresolved in the 2MASS image) has magnitudes J=18.61 ±0.07, H=16.65 ±0.03, and K_s = 15.46 ±0.02.

We downloaded $10' \times 10'$ *J*, *H*, *K*-band stacks around the pulsar position and the associated multi-band object catalogues.

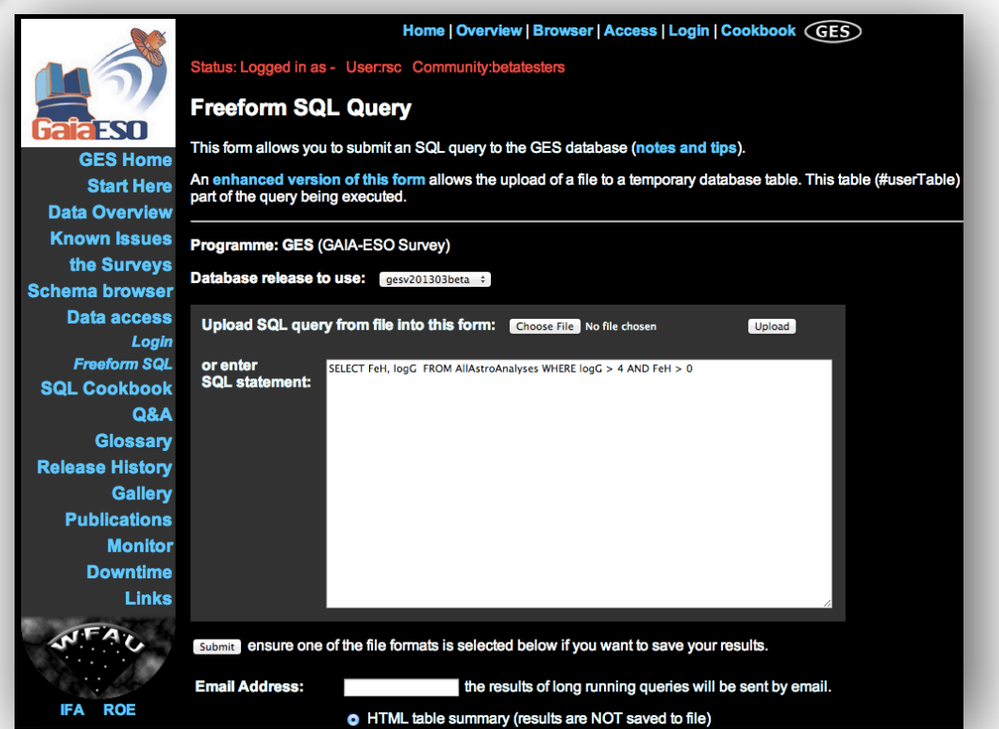
2.2 Pulsar astrometry

For the search for the companion star to PSRJ1811−1736, we assumed as a reference its radio timing coordinates. We note that the pulsar radio timing solution presented by Corongiu et al. (2007)

Figure 2. *K*–(*H*−*K*) (*left*) colour-magnitude diagrams of the PSR J1811−1736 field obtained from the UKIDSS photometry of stars detected in a $10' \times 10'$ region around the PSR J1811−1736 position. The star detected closest to the pulsar position (see Fig. 1) is marked in red. Average photometric errors in the pulsar field, for different magnitude and colour bins, have are also plotted in the left and right panels respectively. (*K* − *H*) vs. (*J* − *K*) colour-colour diagram of the same region. The reddening vectors for an $E(B-V) = 3$ are shown, as a reference.

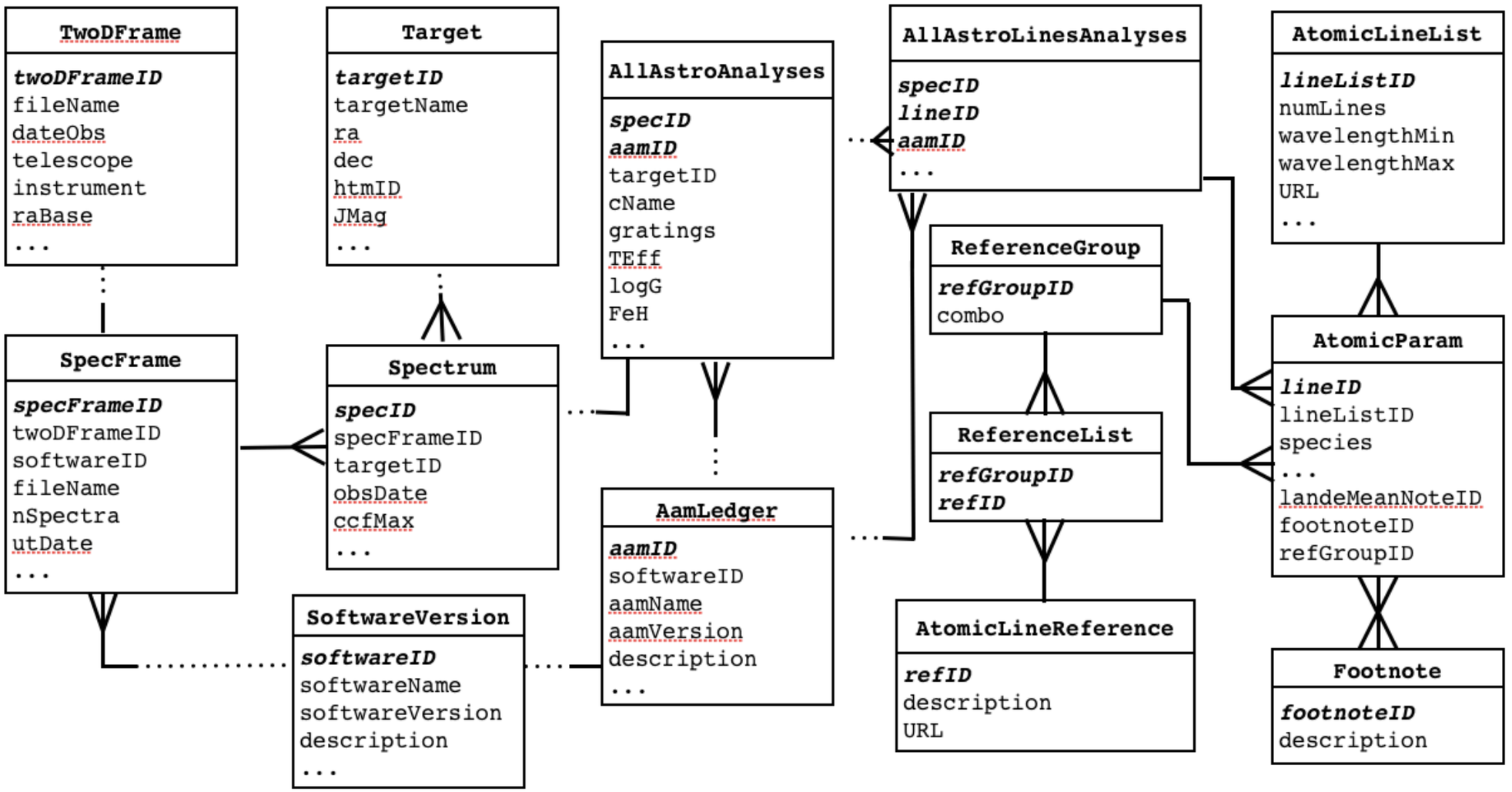
Figure 3. Same as Fig. 2 but with UKIDSS data for the Baade's Window (light grey) overlaid for comparison.

is based on data taken in the epoch range MJD=50842–53624 and does not include the determination of the pulsar's proper motion, an essential parameter for recomputing its position at a given epoch. For this reason, we re-analysed the data presented in Corongiu et al. (2007) adding the proper motion to the timing model, to obtain a new radio timing position at a reference epoch (MJD=53624) closest to that of the UKIDSS observation (MJD= 53934). Thus, we obtained $\alpha_{J2000} = 18^h11^m55\overset{s}{.}0385 \pm 0\overset{s}{.}0029$ and $\delta_{J2000} = -17°36'38''45 \pm 0.''41$ for the position and $\mu_\alpha \cos(\delta) = 18.3 \pm 9.7$ mas yr^{-1} and $\mu_\delta = -176 \pm 100$ mas yr^{-1} for the proper motion, where all quoted uncertainties are at 1σ level. The extrapolated timing position at the epoch of the UKIDSS observation (MJD= 53934) is, then: $\alpha_{J2000} = 18^h11^m55\overset{s}{.}0384 \pm 0\overset{s}{.}0039$ and $\delta_{J2000} = -17°36'38''60 \pm 0.''42$, with an uncertainty radius of $0.''44$ (1σ) that accounts for the position uncertainty due to the proper motion extrapolation. For comparison, by applying the same timing model as above but without adding the proper motion, we obtain $\alpha_{J2000} = 18^h11^m55\overset{s}{.}0337 \pm 0\overset{s}{.}0014$ and $\delta_{J2000} = -17°36'37''80 \pm 0.''19$, where the choice of the reference epoch (MJD=53624) within the range spanned by the timing data is, in this case, arbitrary. Although

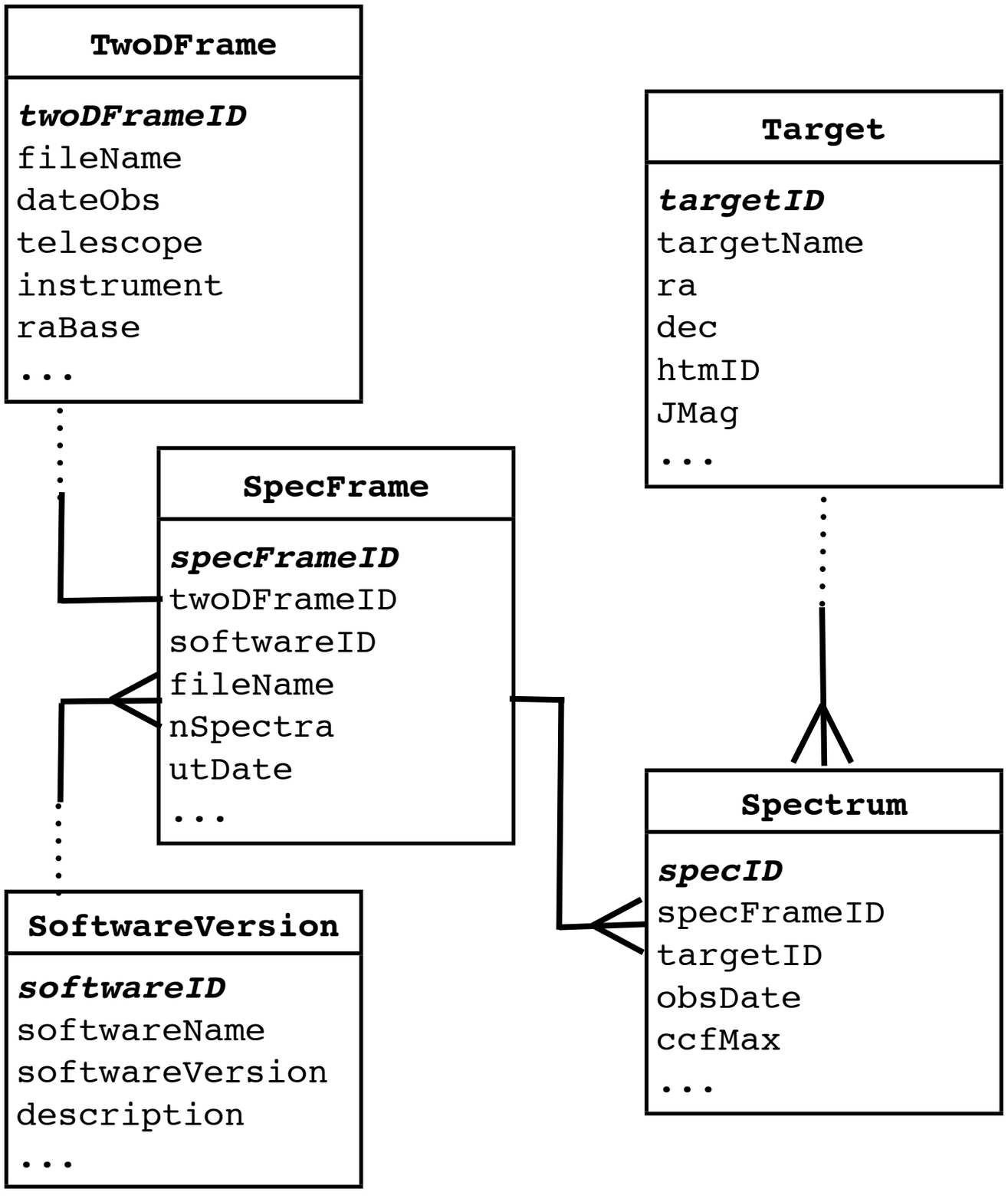


GES Data Model

- The key to a good database is a good data model
- The key to a good data model is to fully understand how all the data are related



Spectra Observations Data Model



TwoDFrame
raw observation image metadata

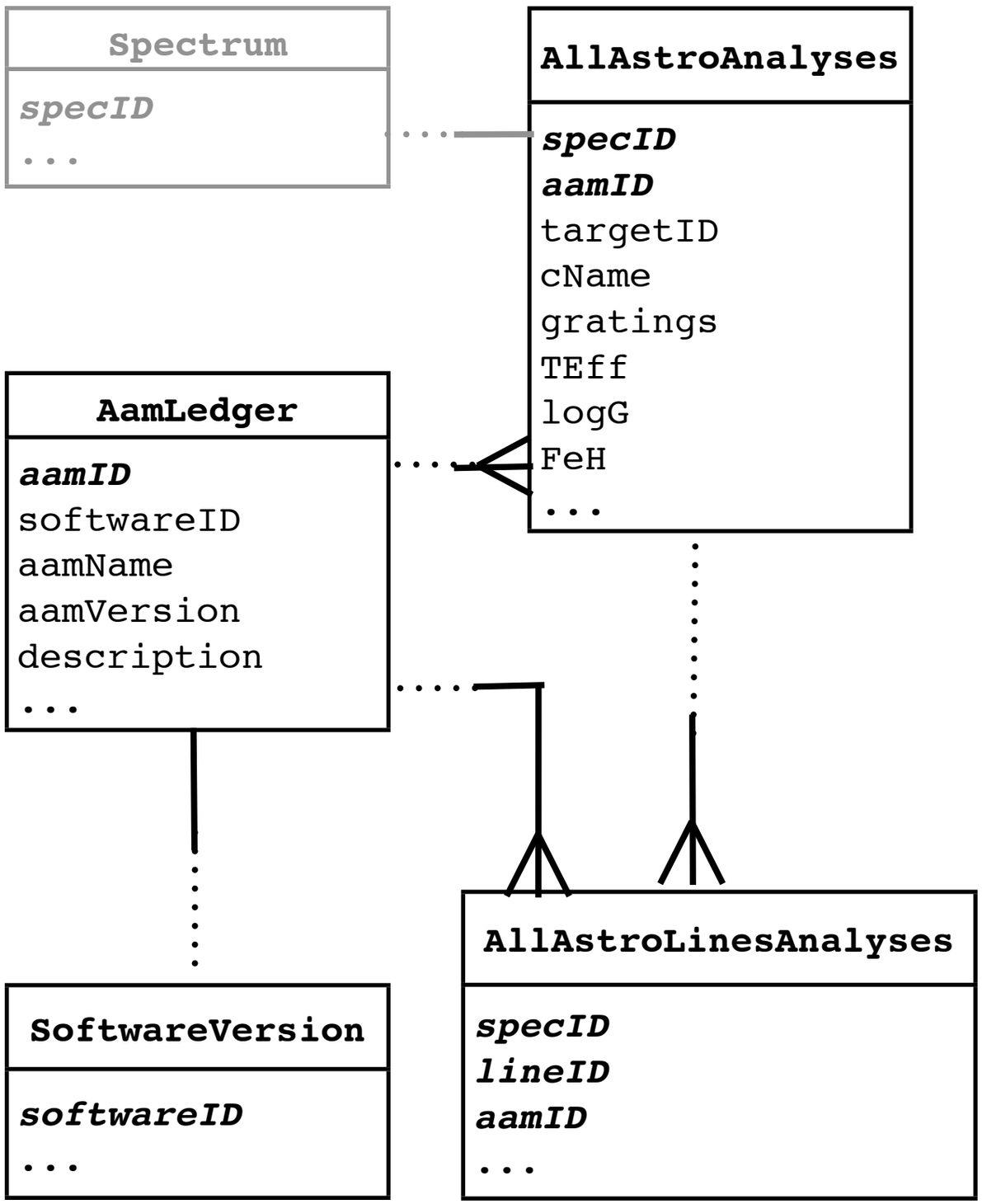
SpecFrame
processed image metadata

Spectrum
metadata of each spectrum extracted from image

Target
Confusion in observation/ analysis data between Targets/Objects:

Here we use Target in sense of the object for which a spectrum was taken, but in the data it is often used to mean the observation field target. May change this nomenclature.

Spectra Analyses Data Model



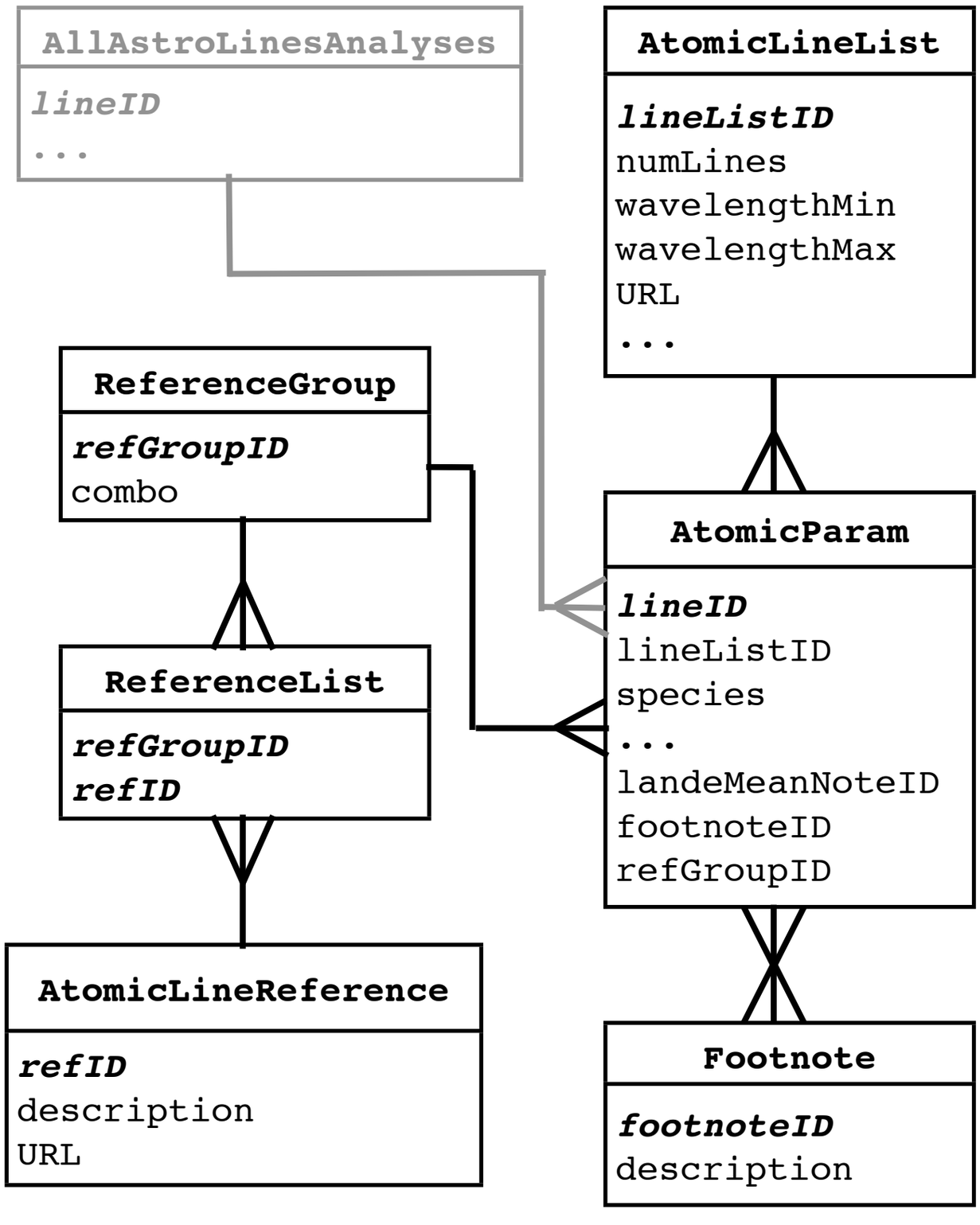
Best parameters and abundances for each spectrum in a single table, one per WG & instrument combination:

- **GiraffeAstroAnalysisWG10**
- **UvesAstroAnalysisWG11**
- **GiraffeAstroAnalysisWG12**
- **UvesAstroAnalysisWG12**
- **AstroAnalysisWG13**

Combined view of all best parameters and abundances:

- **AllAstroAnalyses**

Atomic Line List Data Model



AtomicLineList
Description of complete list

AtomicParam
Individual atomic line data

Footnote
Footnotes associated with line data

ReferenceGroup
References used for line data

AtomicLineReference
Individual bibliographical references

ReferenceList
Links reference group to individual references

Data Homogenisation

- We define the database *schema* from our data model
  - Includes well described columns with consistent
    - names
    - data types
    - units
    - default values
- This makes it easy to query data from all of the WGs together in the **`AllAstroAnalyses`** table for science verification etc.
- Unfortunately, the data we have received to date from WGs has been highly inconsistent and is poorly described

Plea to Working Groups

- Please send us the analysis data in self-describing [multi-extension] FITS table format
  - with unique column names
  - enforces consistent data types
  - encourages descriptions and units with the data (not in separate files) as well as clearly defined default values
  - quick & easy to program a computer to parse

- Even within a single WG we've had best parameters and best abundances variously split between files or in a single file (and with different column names for the same data in each case)

Plea to Working Groups

- We received a multitude of different data formats from WGs:
  - FITS, tab-separated, comma-separated
  - Some had delimited strings, others not
  - Some files had column headers, others didn't.
  - Those with column headers sometimes had non-unique names and inconsistent column separations:

```
TARGET NAME    OBJECT    CNAME    Teff dTeff Num eTeff enum logg dlogg Num elogg
```

  - Invalid values for data type (or inconsistent type) e.g. these "floats" even when proper defaults defined: `9.99.0, ---`
- *Bottom-line:* Not difficult for the human eye to parse and understand, but takes more time and effort to program a computer to parse all data variations accurately

Non-absolute Values

- Some data received so far has been a mixture of absolute values and some maximum values

- A mixture of floating point numbers and numbers like `<2.109` may be OK in a text file, but databases (and FITS files) aren't happy with this

- Suggestion: column of floating point values with an additional column to flag whether the value is absolute or a maximum (or minimum), e.g. `0` is an absolute value, `1` is a maximum value

Default Values

- We define a single default value for each data type in the GES Science Archive database (defaults defined by individual WGs in their data are converted to our uniform default values):

bigint	-99999999	float/real	-9.999995e+08
int	-99999999	datetime	9999-12-31
smallint	-9999	varchar(<4)	---
tinyint	0	varchar(>3)	NONE

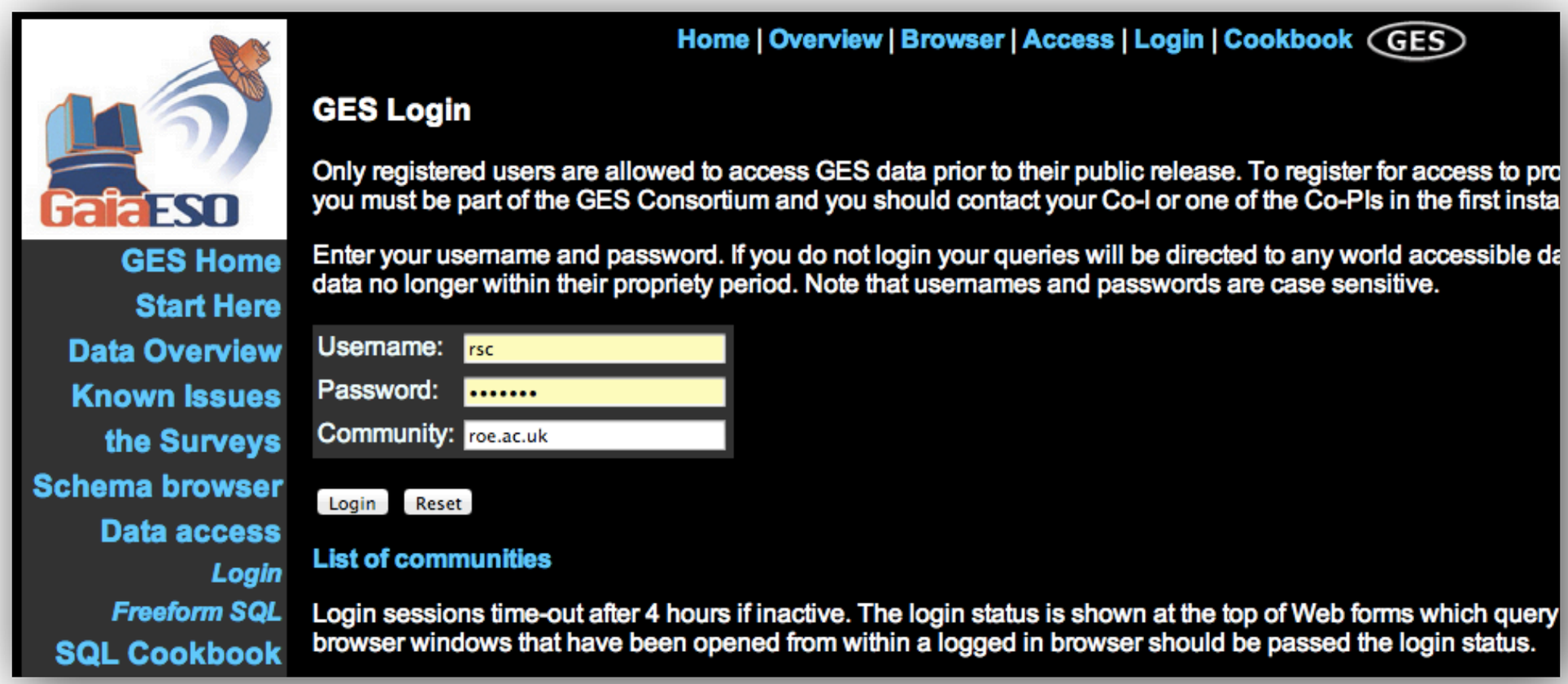
- This makes it easy to automatically always exclude default values when you select a range to query
- Most tables contain a row of default values to make joins between tables easier when value being joined is default

Quality Control & Versioning

- All earlier data remains in the main database
- As new processing versions of earlier observations / analyses arrive, the earlier ones are marked deprecated with the `deprecated` column of all relevant tables
- Where deprecated column exists: ensure to select on `deprecated=0`
- Can exclude earlier data versions from subsequent static release databases
- `deprecated` column can be used to mark data of poor quality with specific code values
- This allows for either clean selections (`deprecated=0`) or more specific selections that include certain quality issues

Using the Archive

- Access restricted via login accounts
- Currently only available to beta testers, will be available to the GES community shortly
- Remember to always login first



Login Accounts

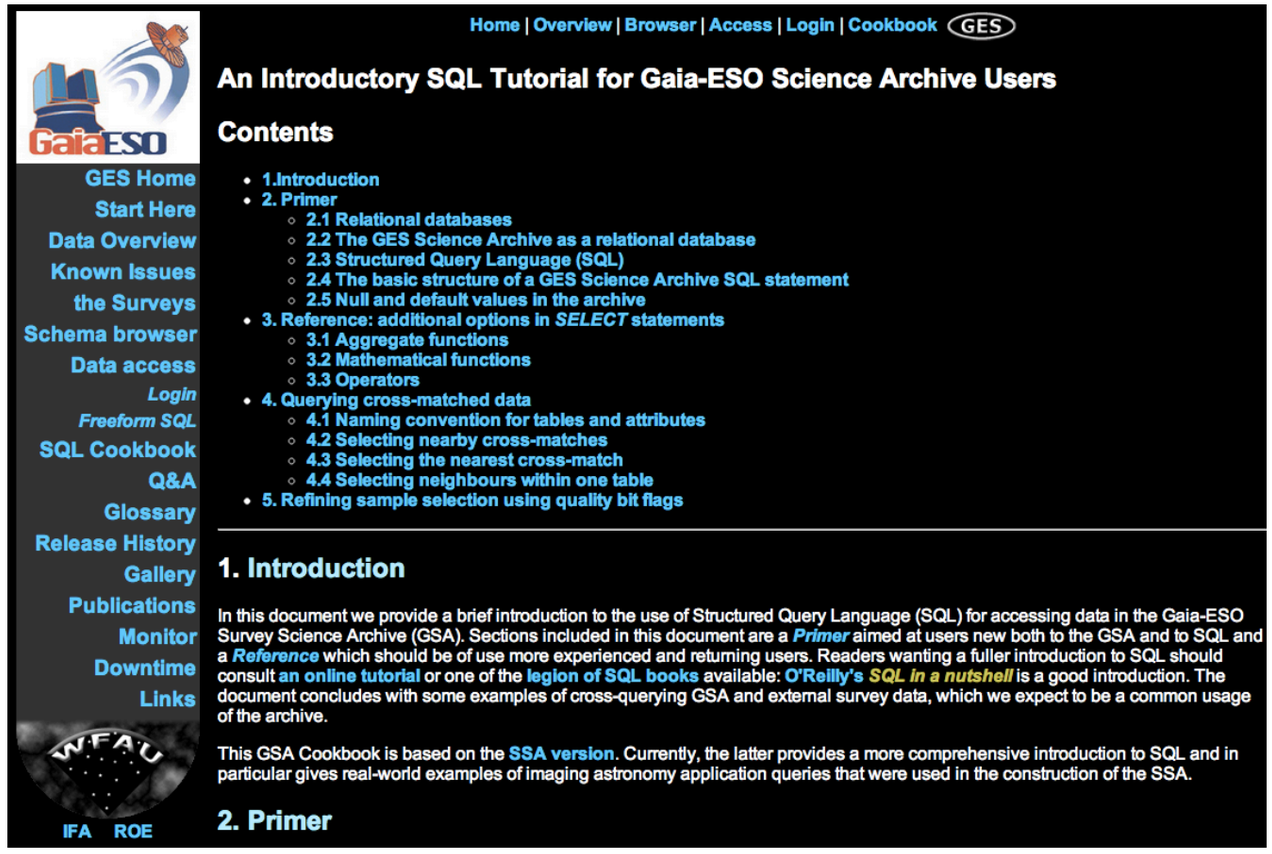
- User account administration delegated to a "community contact" for each of the communities listed on our website
    - They arrange accounts for everyone in their institution that requires one
- Database access restricted by community (currently only betatesters have an accessible database), normally all communities will be able to access the same set of databases

Example communities:

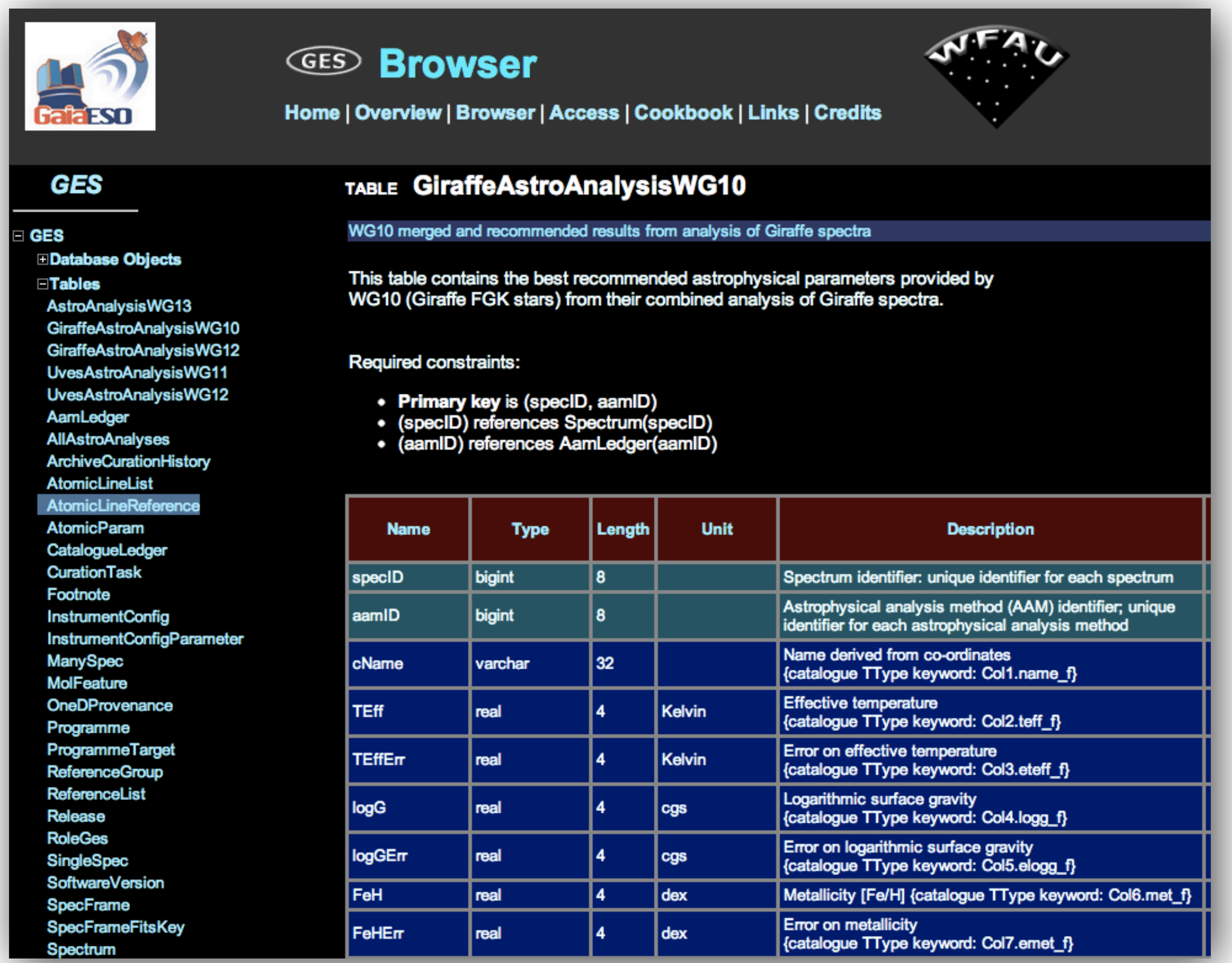
`betatesters`	community for initial archive testing
`cam.ac.uk`	IoA, Cambridge, UK
`eso.org`	ESO
`oca.eu`	Observatoire de la Cote d'Azur, Nice, France

SQL Cookbook

- Interface is only freeform SQL at present: providing direct access to the database
- Will provide some quick access forms later if helpful
- We give a tutorial in the use of SQL for beginners on the website with some example queries to play with



Schema Browser



Name	Type	Length	Unit	Description
specID	bigint	8		Spectrum identifier: unique identifier for each spectrum
aamID	bigint	8		Astrophysical analysis method (AAM) identifier; unique identifier for each astrophysical analysis method
cName	varchar	32		Name derived from co-ordinates {catalogue TType keyword: Col1.name_f}
TEff	real	4	Kelvin	Effective temperature {catalogue TType keyword: Col2.teff_f}
TEffErr	real	4	Kelvin	Error on effective temperature {catalogue TType keyword: Col3.eteff_f}
logG	real	4	cgs	Logarithmic surface gravity {catalogue TType keyword: Col4.logg_f}
logGErr	real	4	cgs	Error on logarithmic surface gravity {catalogue TType keyword: Col5.elogg_f}
FeH	real	4	dex	Metallicity [Fe/H] {catalogue TType keyword: Col6.met_f}
FeHErr	real	4	dex	Error on metallicity {catalogue TType keyword: Col7.emet_f}

Query Interface

One good use of the archive in these early days is science verification tests.

Here we query the **`AllAstroAnalyses`** table in the archive to produce a plot of T_Eff vs. log G

Select to save as FITS file for option to view results directly in TOPCAT

Query Results

GES Database - SQL Query Results

Data file generating queries can take a bit longer to execute as they write to a file ALL rows returned by the query.

A web link to your generated output file will appear at the bottom of this page.

Connecting to gesv201303beta database
QUERY STARTED: Sat Apr 06 17:15:31 BST 2013 [1 active, 29 total]

Please keep this browser window open and wait for your results or further information to appear below...

timeout: 3600

Connected to database

Submitted query: SELECT TEff, logG FROM AllAstroAnalyses WHERE TEff > 0 AND logG > 0

••• OK

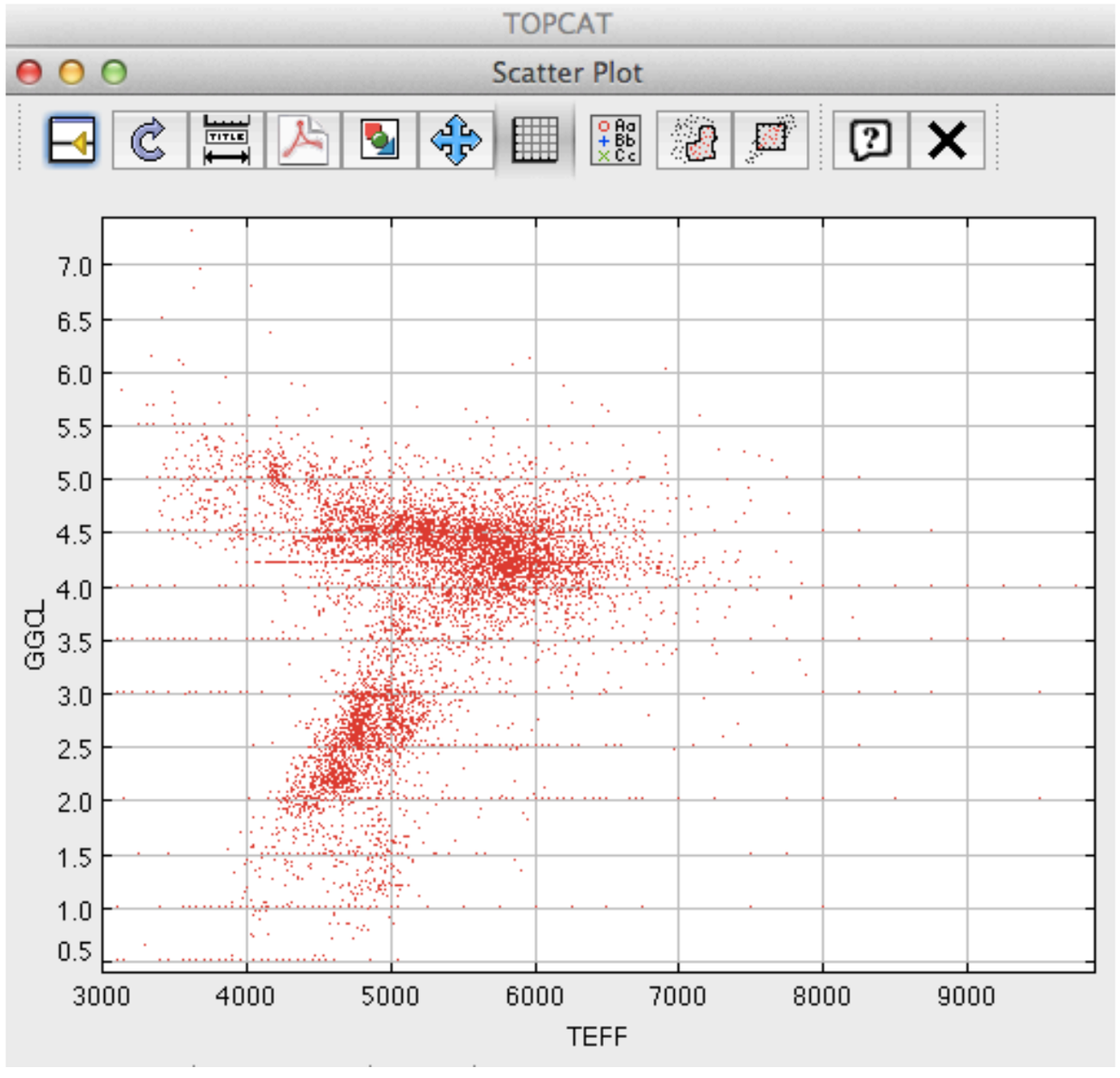
	TEff	logG
1	+4528.000000	+1.900000
2	+4470.000000	+2.530000
3	+4538.000000	+2.420000
26	+4111.000000	+1.830000
27	+4280.000000	+1.890000
28	+4138.000000	+1.550000
29	+4398.000000	+2.100000
30	+4360.000000	+2.100000

(Query returned 10260 result rows, only the first 30 rows are shown in the displayed table.)

Download Results File , your results in a gzipped FITS file (Contains **10260 rows**, 35.1 KB)

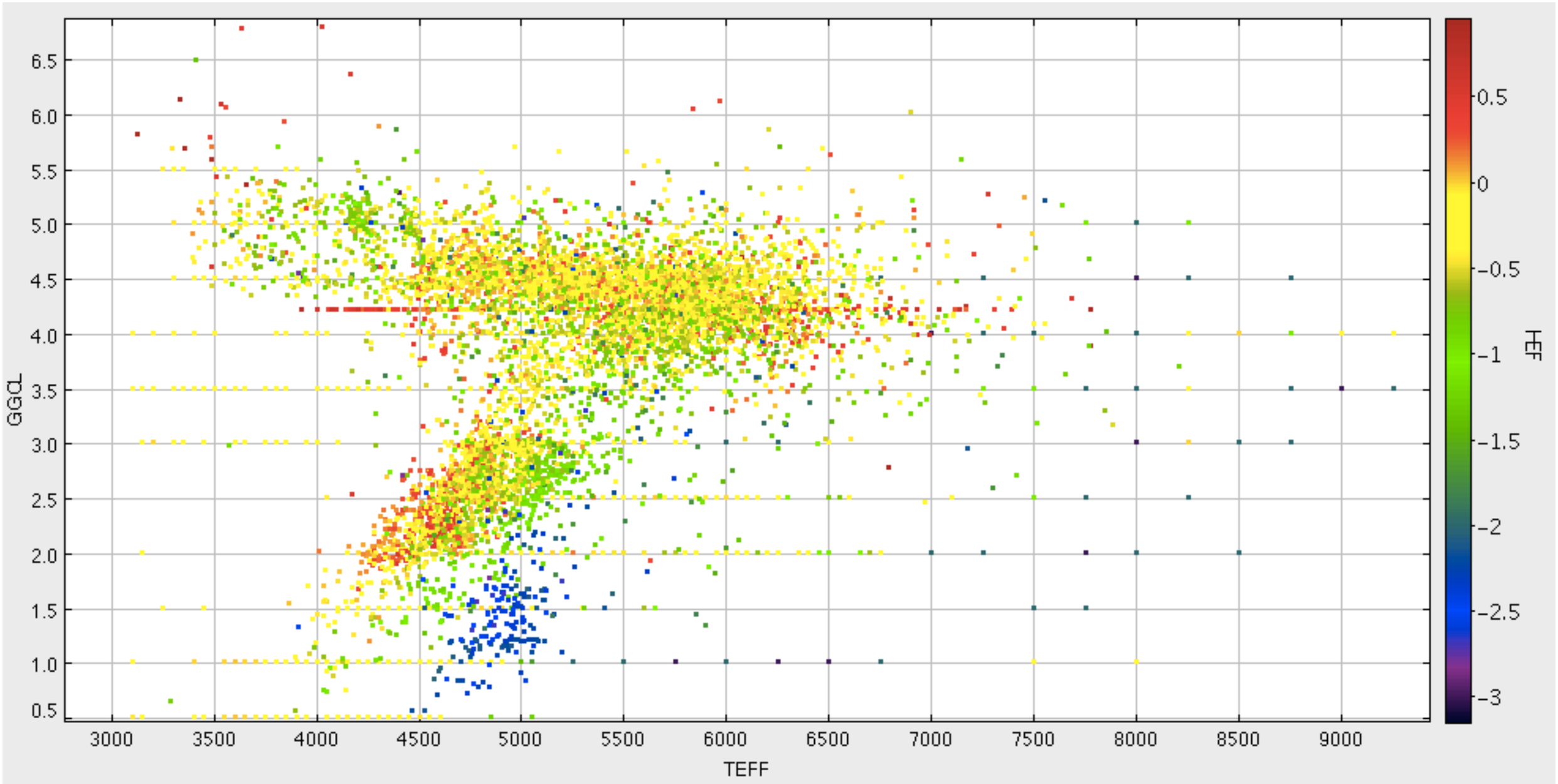
Launch file in Topcat (requires Java 1.5 and Java Web Start, approx 12Mb download for Topcat application)

QUERY FINISHED: Sat Apr 06 17:15:34 BST 2013



Query Results

- Additional columns, e.g. FeH, can be included in the results



Spectrum Browser

- Any queries that contain both **`Spectrum.`**`nSpec` and **`SpecFrame.`**`fileName` columns in the results (*hint*: join on both **`Spectrum.`**`specID` & **`SpecFrame.`**`specFrameID`) offer a link to a preview of the spectrum

GES Database - SQL Query Results

Data file generating queries can take a bit longer to execute as they write to a file ALL rows returned by the query.

A web link to your generated output file will appear at the bottom of this page.

Connecting to GES database
QUERY STARTED: Sat Apr 06 11:06:25 BST 2013 [1 active, 27 total]

Please keep this browser window open and wait for your results or further information to appear below...

timeout: 3600

Connected to database

Submitted query: SELECT TOP 10 * FROM Spectrum S, SpecFrame F WHERE S.specFrameID=F.specFrameID

••• OK

The sLink column can be used to display the spectrum.

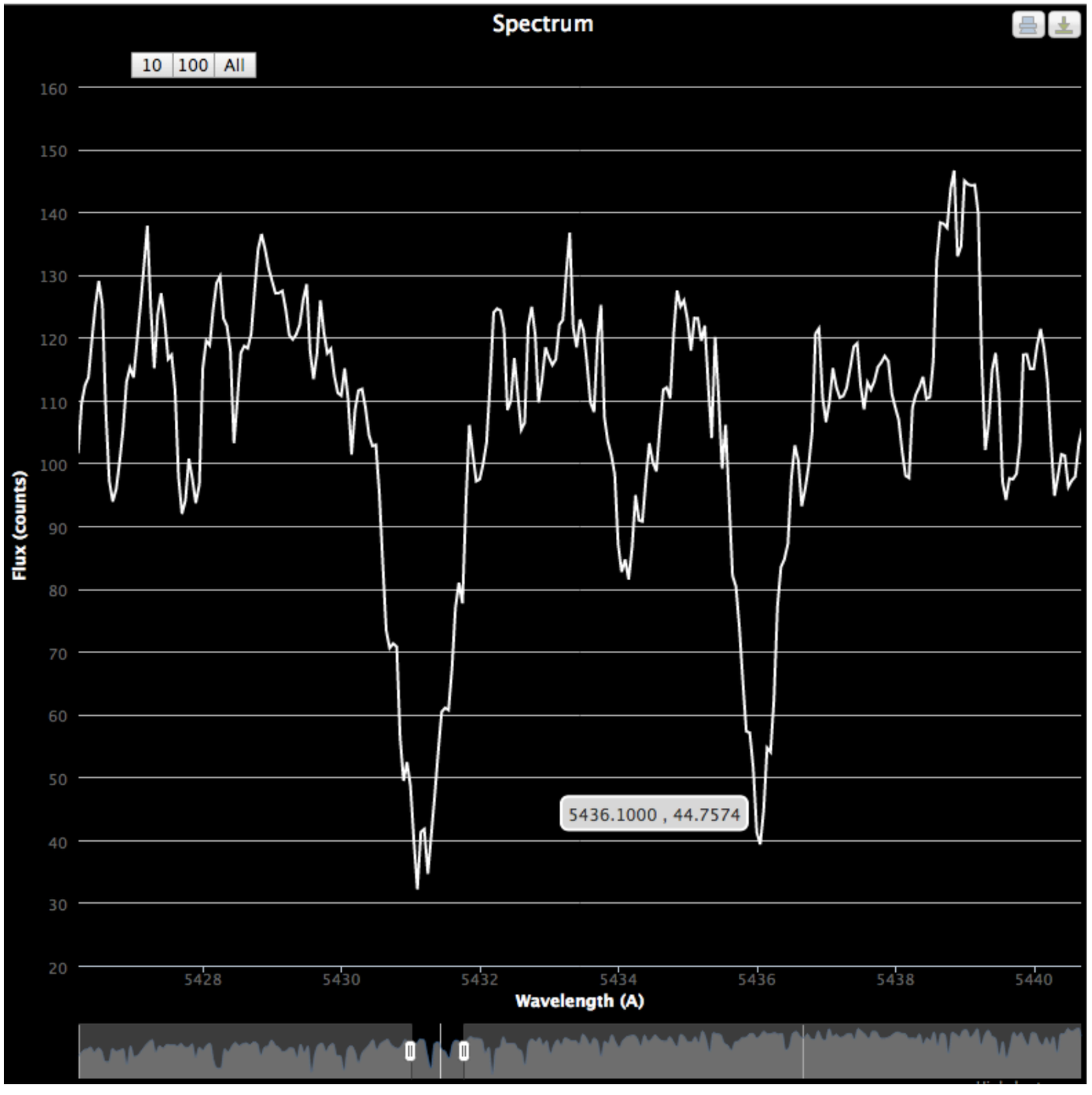
	sLink	specID	specFrameID	nSpec	nSpecOld	targetID	targetName	object	ra	dec	htmID	cx	cy
1	view	4294967331	1	35	40	-99999999	NONE	Sr_0_218	+44.1216000	-0.3806583	8933772630748	+0.7178481	+0.6961681
2	view	4294967332	1	36	42	-99999999	NONE	Se_0_122	+44.0726125	-0.4338944	8933788168714	+0.7184383	+0.6955495
3	view	4294967333	1	37	43	-99999999	NONE	Se_0_214	+44.0908958	-0.3679306	8933771514337	+0.7182221	+0.6957843
4	view	4294967334	1	38	44	-99999999	NONE	Sb_0_213	+44.0921000	-0.3785417	8933771528574	+0.7182066	+0.6957986
5	view	4294967335	1	39	45	-99999999	NONE	Sb_0_192	+44.1145875	-0.5064306	8933796841061	+0.7179211	+0.6960684
6	view	4294967336	1	40	46	-99999999	NONE	Se_0_96	+44.0052500	-0.5038417	8933745836685	+0.7192483	+0.6946974
7	view	4294967337	1	41	47	-99999999	NONE	Se_0_210	+44.1569250	-0.4778472	8933798674646	+0.7174096	+0.6966017
8	view	4294967338	1	42	48	-99999999	NONE	Sb_0_97	+44.0161792	-0.5082417	8933745280413	+0.7191153	+0.6948341
9	view	4294967339	1	43	49	-99999999	NONE	sky_134	+44.0524542	-0.4418833	8933788294689	+0.7186822	+0.6952960
10	view	4294967340	1	44	50	-99999999	NONE	Sb_0_209	+44.1301292	-0.4361361	8933790593192	+0.7177395	+0.6962702

(Query returned 10 result rows, all rows are shown in the displayed table.)

Download Results File , your results in a gzipped FITS file (Contains **10 rows**, 7 KB)

Launch file in Topcat (requires Java 1.5 and Java Web Start, approx 12Mb download for Topcat application)

Spectrum Browser



FITS Spectra Download

- Any queries that contain the **`SpecFrame.`**`fileName` column in the results (*hint*: join on **`SpecFrame.`**`specFrameID`) offer a link to download the VO-compliant FITS file of the spectra that can be viewed in SPLAT etc.

GES Database - SQL Query Results

Data file generating queries can take a bit longer to execute as they write to a file ALL rows returned by the query.

A web link to your generated output file will appear at the bottom of this page.

Connecting to gesv201303beta database
QUERY STARTED: Mon Apr 08 12:53:58 BST 2013 [1 active, 9 total]

Please keep this browser window open and wait for your results or further information to appear below...

timeout: 3600

Connected to database

Submitted query: SELECT TOP 10 fileName, nSpec FROM Spectrum S, SpecFrame F WHERE S.specFrameID=F.specFrameID

••• OK

The sLink column can be used to display the spectrum.

The getFLink column can be used to download the referenced FITS file.

	sLink	getFLink	fileName	nSpec
1	view	download	/disk40/ges/ingest/fits/tests/May2012/C20111231_H548.8_MW_0255555-002846_1.fit	35
2	view	download	/disk40/ges/ingest/fits/tests/May2012/C20111231_H548.8_MW_0255555-002846_1.fit	36
3	view	download	/disk40/ges/ingest/fits/tests/May2012/C20111231_H548.8_MW_0255555-002846_1.fit	37
4	view	download	/disk40/ges/ingest/fits/tests/May2012/C20111231_H548.8_MW_0255555-002846_1.fit	38
5	view	download	/disk40/ges/ingest/fits/tests/May2012/C20111231_H548.8_MW_0255555-002846_1.fit	39
6	view	download	/disk40/ges/ingest/fits/tests/May2012/C20111231_H548.8_MW_0255555-002846_1.fit	40
7	view	download	/disk40/ges/ingest/fits/tests/May2012/C20111231_H548.8_MW_0255555-002846_1.fit	41
8	view	download	/disk40/ges/ingest/fits/tests/May2012/C20111231_H548.8_MW_0255555-002846_1.fit	42
9	view	download	/disk40/ges/ingest/fits/tests/May2012/C20111231_H548.8_MW_0255555-002846_1.fit	43
10	view	download	/disk40/ges/ingest/fits/tests/May2012/C20111231_H548.8_MW_0255555-002846_1.fit	44

(Query returned 10 result rows, all rows are shown in the displayed table.)

Download Results File , your results in a gzipped FITS file (Contains **10 rows**, 628 bytes)

Launch file in Topcat (requires Java 1.5 and Java Web Start, approx 12Mb download for Topcat application)

Cross-matches

- Cross-match targets with other survey catalogues also hosted/mirrored at WFAU:
    - VISTA PS - ESO infrared photometry
    - VST PS - ESO visible photometry
    - 2MASS - all sky infrared photometry
    - SSS - all sky visible photometry
    - SDSS - Sloan Digital Sky Survey
    - *And many more...*
- Data model allows joins with these catalogues enabling a single query to retrieve for any targets in the GES archive the observation measurements from all surveys

What's next?

- Tidying up the data schema
  - from beta testing feedback

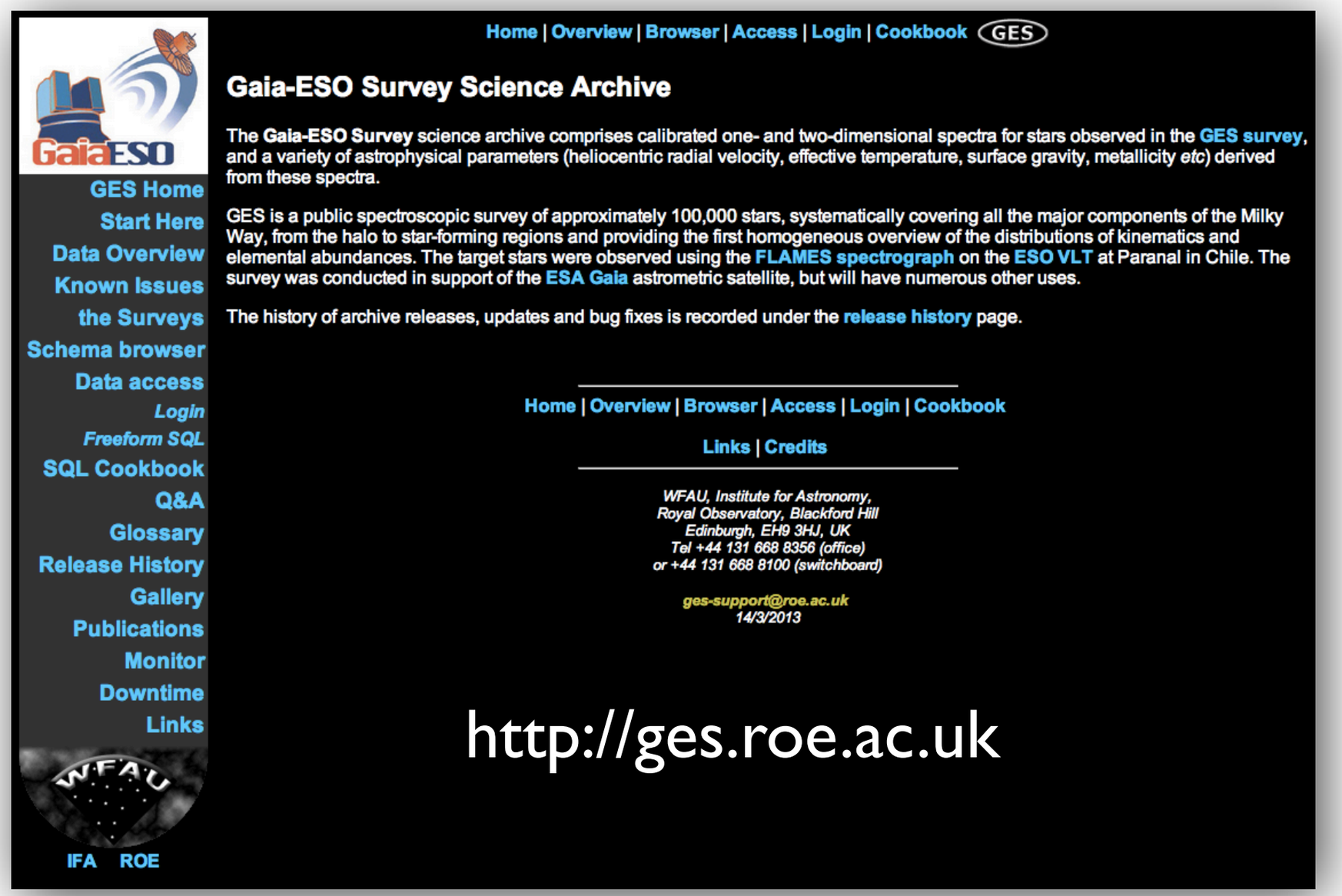
- Implementing all links in the data model
  - key missing links are Spectrum & Target table data
    - need latest processed spectra from Cambridge

- Produce a consortium-wide release database
- Eventually this data will be publicly accessible (inc. VO) and *referencable:* can include SQL queries used against specific data releases in your papers for reproducibility

The GES Science Archive

Ross Collins

Nigel Hambly, Clive Davenhall, Mike Read, Eckhard Sutorius
WFAU, Institute for Astronomy, Royal Observatory Edinburgh



http://ges.roe.ac.uk