

UKIDSS Data Access Policy

Proposal for UKIRT Board 2006 May 22nd

UKIDSS Survey Heads

Andy Lawrence (CPI), Steve Warren (CSS), Richard Jameson (LAS), Nigel Hambly (GCS), Phil Lucas (GPS), Alastair Edge (DXS), Omar Almaini (UDS), Paul Hewett (Data Processing WG), Ike Irwin (Data Processing WG), Simon Dye (CSV).

Draft 2006 May 13th

(1) Introduction.

There have been a number of discussions recently about data access policy for UKIDSS. These have involved individuals in UKIDSS, the VDFS management team, JAC, and the UKIRT Board. It was proposed that UKIDSS should propose a view on these issues, for discussion and hopefully endorsement by the UKIRT Board.

The only formal statement on data access policy is that on the JAC web pages at <http://www.jach.hawaii.edu/UKIRT/surveys/UKIDSSdatapolicies.html>

The wording is reproduced in Appendix A. There are also some words on the WSA web pages. These are reproduced in Appendix B. The consortium meeting of November 11th 2005 held a full discussion of data policy, but the specific recent issues were debated by the survey heads in a telecon held on May 4th. (Each survey head consulted the working group members before this telecon). An agreed position reached, which is described in this paper.

(2) Issue-A : who sets data access policy

Both at the November consortium meeting and during the recent debate on flat file access (see Issue-C) there has been considerable confusion over who has the right to determine data policy issues. The issue here is strictly about UKIDSS data, not WFCAM data overall.

The proposal from UKIDSS is as follows :

(i) Formal authority on all UKIRT data policies rests with the UKIRT Board, on behalf of PPARC. This statement includes UKIDSS data.

(ii) In the case of UKIDSS, the Board will expect the UKIDSS consortium to propose a policy that they can choose to endorse - or dispute, in which case the consortium should propose a revised policy.

(iii) Although UKIDSS proposes and the Board disposes, the JAC, the VDFS project, and the VISTA Data Users Committee (VDUC) are also stakeholders, and so must be consulted, and we must be assured that they are content with policy.

(iv) For major issues, the UKIDSS Working Group (WG) members should be consulted. For day-to-day issues, the consortium can reasonably expect that the Survey Heads deal with things on their behalf. Some rare issues would require agreement by a full open consortium meeting.

(3) Issue-B : guidance on fair use

The data policy statement on the web includes statements about sharing UKIDSS data with external collaborators, and publication of papers based on analysis of UKIDSS data. There have been some

queries and concerns about what this means in practice, mostly based around individuals who do have access rights to UKIDSS, but who want to write follow-up proposals with collaborators who are from outside ESO - for example, selecting a sample of Brown Dwarf candidates from the EDR, and proposing this source list to a US telescope, with US collaborators; or ditto following up a list of EDR sources cross-matched with faint radio sources.

Section IV of the existing data policy is already quite clear - you should not give other people access to the actual UKIDSS data, but once you have done a "scientific analysis" with the data you are free to do whatever you wish. Running a WSA query, eg on colours, to produce a candidate list, is such a "scientific analysis", so scientists should be free to do whatever they wish with such a list. (In principle, you could publish such a list; and we can't constrain what people do with a list in a table in a paper.) However, there is a grey moral area. One might imagine US folk approaching a UK postdoc and saying "hey, why don't you run a query like this and this.. then we can put a Keck proposal in together to get the spectra". This could be called "access by proxy".

The UKIDSS proposal is as follows :

(i) The strict limits on access are already clear.

(ii) However, access by proxy should be very clearly discouraged. The wording of the data policy should be revised to reflect this. The proposed wording is given at Appendix C.

(iii) Many potential problems could be prevented by providing some concrete examples. These can also be included in a revised version of the data policy, and also provided on the UKIDSS and WSA web pages.

A draft revised data policy prepared by Andy Adamson is given at Appendix C. More colourfully worded examples for possible inclusion on UKIDSS and WSA web pages are given at Appendix D.

(4) Issue-C : fast flat file access

The data processing chain has four stages. (i) The CASU pipeline produces processed images and per-frame catalogues as flat files, which are then sent to Edinburgh. (ii) Various curation tasks are run to ingest the data into the database structure. (Not as easy as it sounds !) (iii) A QC process is run which filters out bad data. (iv) Data is stacked and merged and re-pipelined to produce the deep survey data products. Currently we do not release data to the public until all these processes are complete; our current plan assumes a series of staged releases with well defined volumes and processing histories. The next release is DR1 on July 14th. The default assumption is that releases are made roughly every six months after that.

Richard McMahon has proposed that on a much faster timescale, we should make the flat files available. The idea would be to grow this part of the public WSA quasi-continuously, with the minimum possible delay between observations and availability. Currently the flat files are available through the WSA, but only those that are included in the staged releases. So in the short term, Richard's request is effectively for open access to the DR1 pipeline products from stage (i) before we have finished stages (ii) (iii) and (iv). This would involve releasing some data that would later fail the QC filter.

The arguments **for** fast flat file access are (a) adding scientific benefit, because some users can do science sooner with the flat files, and follow up observations sooner; (b) improving PR, by making data available rapidly and showing the VDFS system to be quick and efficient; and (c) clarifying the

CASU workplan. The arguments **against** fast flat file access are (a) creating extra work - setting this up and running it would add approximately 0.2FTE to VDFS; and (b) damaging PR by exposing sub-standard data, and by giving a perceived advantage to those on the inside track who are better able to take advantage of the flat files.

The survey heads consulted working group members, and found fairly uniformly that people felt that fast flat file access would be good in principle, but that they didn't feel strongly about it. A minority, including some survey heads, were distinctly worried about the "dodgy data" and "inside track" perception worries. Jim Emerson expressed a strong feeling that VDFS could not at least in the near future afford the necessary effort. To satisfy the desire for quick follow-up, some suggested that we should move towards 3-month releases, but this wasn't yet felt to be practical.

The eventual UKIDSS proposal was as follows :

- (a) Fast flat file access is in general seen as desirable, but not a high priority
- (b) No action to be taken before DR1
- (c) After DR1, UKIDSS would in principle like fast flat file access as a goal
- (d) Whether this is implemented is then an VDFS decision, balancing effort against priorities
- (e) We will revisit the possibility of faster staged releases in due course.

Appendix A : Current public policy from JAC web pages

(as at <http://www.jach.hawaii.edu/UKIRT/surveys/UKIDSSdatapolicies.html>)

UKIDSS SURVEY DATA POLICIES

I. Data, Transport & Storage

Raw data will be dispatched to CASU for pipeline processing within approximately one week of being taken.

1. UKIRT will retain a local tape copy of all raw data.
2. JAC staff and (registered) visitors at JAC will obtain WFCAM data for scientific research from the WFCAM science archive at WFAU.
3. A copy of the raw data will be made available to the ESO archive by CASU.
4. UKIRT will maintain a transients pipeline dedicated to producing results on moving and transient sources which require followup on a timescale shorter than will be possible through the science archive. Results from this pipeline will be made immediately world-public.

II. Access

1. Access to processed UKIDSS data (catalogues and calibrated images) will be through the WFCAM science archive (at WFAU, Edinburgh).
2. Raw data, where necessary, will be available from the Cambridge Astronomical Survey Unit (CASU) and via the ESO archive as appropriate.
3. In both cases, access will be granted to Registered Users only.
4. Processed data products via WFAU and raw data via CASU will be made available to all Registered Users immediately and simultaneously. Except where necessary to verify the quality of data and pipeline reduction, Individuals responsible for pre-processing and releasing the data will not undertake science analysis in advance of release to the science archive.
5. Access to any data downloaded from WSA, or any local copies of data files that result from the pipeline process, are subject to the same proprietary period and access rules defined elsewhere in the document.
6. The first public access (for the unregistered world community) to processed data through WFAU and raw data through CASU will occur 18 months after the data are taken.

III. Data Sharing and Publication

1. Registered Users will not export proprietary data or data products to persons other than Registered Users. Nor will they make the data public in digital form. UKIRT reserves the right to revoke registration from anyone found to have broken this fundamental rule.
2. Scientific collaborations that make use of UKIDSS data are encouraged. However, the export of UKIDSS proprietary data to external members of a scientific collaboration is permitted only upon the explicit authorization from the UKIDSS Consortium Principal Investigator (CPI : A.Lawrence).
3. Scientific results from analyses of UKIDSS data are the property of the individuals responsible for the analyses and may be published in any form at any time in the usual way.
4. All users of UKIDSS data should include an acknowledgement in all publications as follows: *This work is based in part on data obtained as part of the UKIRT Infrared Deep Sky Survey.*

IV. Registration

Initial registration will be open to the following:

1. Professional astronomers, students and postdoctoral staff currently engaged in astronomical research at Universities and Research Institutes in ESO-member countries.
2. Current members of the JAC scientific staff.
3. UKIDSS members from the Japanese astronomical community.

In the medium term, these users will automatically be registered via organisation-based community servers within the AstroGrid infrastructure.

Change of status (e.g. a move out of the ESO countries) will normally result in UKIDSS data access being discontinued.

It is understood that relatively sparse registration deadlines may prove problematic; an "urgency" procedure will be in place whereby direct contact can be made with UKIRT in case of need. Examples of acceptable cases might be (i) a new recruit requiring access to the data outwith the normal registration cycle and (ii) a request for access from a user who would not normally require it (e.g. an X-ray astronomer wanting to gain an additional point in a SED).

V. Appeals.

Requests in the following areas will be assessed by the Head of UKIRT Operations and the UKIDSS Survey scientist. Their decision will be final.

1. For access by a potential user not fitting the three Registration criteria above.
2. For continued access following a change of status or location.
3. For the right to copy data to a non-Registered User.

Some of the above policies are based on their equivalents in the CFHT Legacy Survey.

Appendix B : Wording on WSA web pages

(as at <http://surveys.roe.ac.uk/wsa/policy.html>)

Data policy

The WFCAM Science Archive contains data from *all* pipeline processed WFCAM programmes, both survey (i.e. the UKIDSS) and non-survey (open time PI) programmes. In general, these data have proprietary periods attached to them: UKIDSS data are for use by consortium members and astronomers in ESO member countries for a period of 18 months after observation; data obtained for open time programmes are for use by the PI and named co-Is only for a period of 12 months after observations. Once these proprietary periods have expired, the data become available to all.

If you wish to access proprietary data and are entitled to do so, then you must be registered with an appropriate user community. This enables you to log in to the archive server and gain full access to the proprietary data associated with that community. Click here for [further details concerning data access policies for UKIDSS](#).

Details of proprietary data are always available in the archive, even if the data themselves are protected.

Appendix C : Proposed revised public policy page

I. Data transport and storage

1. Data will be dispatched to CASU for pipeline processing within approximately one week of being taken.
2. UKIRT will retain a local tape copy of all raw data.
3. JAC staff and (registered) visitors at JAC will obtain WFCAM data for scientific research from the WFCAM science archive at WFAU.
4. A copy of the raw data will be made available to the ESO archive by CASU.
5. UKIRT will maintain a transients pipeline dedicated to producing results on moving and transient sources which require followup on a timescale shorter than will be possible through the science archive. Results from this pipeline will be made immediately world-public.

II Access

1. Access to processed UKIDSS data (catalogues and calibrated images) will be through the WFCAM science archive (at WFAU, Edinburgh).
2. Raw data, where necessary, will be available from the Cambridge Astronomical Survey Unit (CASU) and via the ESO archive as appropriate.
3. In both cases, access will be granted to Registered Users only.
4. Processed data products via WFAU and raw data via CASU will be made available to all Registered Users immediately and simultaneously. Except where necessary to verify the quality of data and pipeline reduction, Individuals responsible for pre-processing and releasing the data will not undertake science analysis in advance of release to the science archive.
5. Access to any data downloaded from WSA, or any local copies of data files that result from the pipeline process, are subject to the same proprietary period and access rules defined elsewhere in the document.
6. The first public access (for the unregistered world community) to processed data through WFAU and raw data through CASU will occur 18 months after the data are taken.

III Data Sharing & Publication

1. Registered Users will not export proprietary data or data products to persons other than Registered Users. Nor will they make the data public in digital form. UKIRT reserves the right to revoke registration from anyone found to have broken this fundamental rule.
2. Scientific results from analyses of UKIDSS data are the property of the individuals responsible for the analyses and may be published in any form at any time in the usual way.
3. All users of UKIDSS data should include an acknowledgement in all publications as follows: *This work is based in part on data obtained as part of the UKIRT Infrared Deep Sky Survey.*
4. Scientific collaborations that make use of UKIDSS data are encouraged. However, the export of UKIDSS proprietary data to external members of a scientific collaboration is permitted only under certain circumstances, as laid out in section IV. In case of uncertainty, users should seek explicit authorization from the UKIDSS Consortium Principal Investigator (CPI : A.Lawrence).

IV Collaborative Followup

1. Within the ESO community, UKIDSS data may be freely shared and used to generate telescope time proposals without restriction.
2. For collaborative followup involving non-ESO scientists, the following guideline should be adhered to: the followup project should result from analysis of UKIDSS data **initiated and carried out by ESO scientist(s)**, and ESO scientists should have **substantial involvement in all aspects of the followup project**.
3. Some illustrative examples follow:

Acceptable

- ESO scientist selects a colour-based subsample and decides to write a proposal for spectroscopy using a non-ESO facility. Non-ESO colleagues are invited on board to increase the chances of getting time.

- ESO scientist receives contact from non-ESO source offering to collaborate on the basis of an unique sample of gamma-ray sources. ESO scientist takes this list and crossmatches with the UKIDSS survey, and the combined list is used to generate a paper and/or collaborative time proposal.

Probably not acceptable (See section VI - Appeals)

- ESO scientist receives contact from non-ESO source asking for a catalogue to be generated from the UKIDSS survey to be followed up on Keck. Whether this is in fact acceptable depends on the degree of involvement of ESO scientists in the project.

Not acceptable

- ESO scientist receives contact from non-ESO source offering to collaborate on the basis of an unique sample of gamma-ray sources. ESO scientist sends list of all UKIDSS sources in the relevant region of the sky and cross-matching is done by the non-ESO scientist.

V Registration

Initial registration will be open to the following:

1. Professional astronomers, students and postdoctoral staff currently engaged in astronomical research at Universities and Research Institutes in ESO-member countries.
2. Current members of the JAC scientific staff.
3. UKIDSS members from the Japanese astronomical community.

In the medium term, these users will automatically be registered via organisation-based community servers within the AstroGrid infrastructure.

Change of status (e.g. a move out of the ESO countries) will normally result in UKIDSS data access being discontinued.

VI Appeals

Requests in the following areas will be assessed by the UKIDSS PI, UKIDSS Survey scientist and Associate Director UKIRT. Their decision will be final.

1. For access by a potential user not fitting the three Registration criteria above.
2. For continued access following a change of status or location.
3. For the right to copy data to a non-Registered User.

Some of the above policies are based on their equivalents in the CFHT Legacy Survey.

Appendix D : Proposed "fair use" examples for UKIDSS and WSA web pages

(1) You select a colour based subsample and decide to write a Keck proposal for spectra. You invite US colleagues on board to increase your chances of getting time. THIS IS FINE.

(2) Somebody phones you from the US and says "why don't pick you a sample along such and such lines and we will put in a Keck proposal together". THIS IS SEEN AS "ACCESS BY PROXY" AND IS STRONGLY DISCOURAGED.

(3) somebody phones you from the US and says "I have a sample of gamma-ray sources that nobody else has.. how about it ?" You say "send me the list". You crossmatch the gamma-ray list and send back the cross-matched sources. THIS IS FINE.

(4) Somebody phones you from the US and says "I have a sample of gamma-ray sources in the blah field that nobody else has.. how about it ?" You say "I will send you the UKIDSS sources in that area of sky, so you can cross-match". THIS DEFINITELY CONTRAVENES THE DATA POLICY.