WHOLE EARTH TELESCOPE HEADQUARTERS STANDARD DATA FORMATS

R. L. Riddle and S. D. Kawaler

Dept. of Physics and Astronomy, Iowa State University, Ames, IA 50011, U.S.A.

January 20, 2003

Abstract. As the WET moves to CCD systems, we move away from the uniformity of the standard WET photometer into an arena where each system can be radically different. There are many possible CCD photometry systems that can fulfil the requirements of a WET instrument, but each of these will have their own unique native data format. During XCov22, it became readily apparent that the WET requires a defined data format for all CCD data that arrives at HQ. This paper describes the proposed format for the next generation of WET data; the final version will be the default format for XQED, the new photometry package discussed elsewhere in these proceedings.

Key words: techniques: photometric - methods: data analysis

1. INTRODUCTION

For most of the history of the WET project, the standard WET photometer has been the instrument of choice for observers around the world. Though there are many individually constructed PMT photometers in use at WET sites, each mimics the design and software of the WET three channel PMT system (Kleinman 1996). The output data from each system is in the venerable Quilt 9 format (or can be reformatted to Quilt 9), so that all of the data sent to HQ during past WET campaigns from the PMT instruments has been in the same format.

With the addition of CCD instruments to the WET network, this is no longer the case. The first CCD data added to the WET network were reduced in a somewhat ad hoc fashion, because there were one or, perhaps, two CCDs included in the campaign, and all the CCD frames were rereduced appropriately after the campaign (as

should always be the case). Because of the many advantages of the CCD system over the PMT system, the number of WET participants using CCDs will continue to increase in the future. Therefore, we need to examine their effect on the operations of the WET and WET HQ.

XCov22 was held in May 2002, and twenty one observatories participated (plus the Hubble Space Telescope). Seventeen of these sites were regular WET sites, and four were trial sites. Of the seventeen regular sites, seven contributed only CCD data to the campaign (South Africa contributed some CCD data in addition to the PMT data); three of the four trial sites were CCD systems, and HST used STIS, a CCD based spectrograph. This means that half of the participating WET sites during XCov22 collected their data using CCDs! The largest fraction before this was four of fourteen during XCov19 in June 2002, one of which was again HST and STIS; during XCov21 in April 2001, there were five CCD sites out of twenty, some of which did not send data to HQ for analysis (as planned). In the earlier campaigns, the CCD site data was easily converted to the WET format (i.e. differential magnitudes) for inclusion into the first look analysis, because of the small number of CCD sites (one or two) and existing conversion algorithms. We were not nearly as fortunate during XCov22, as each of the ten CCD sites submitted data in unique formats that required their own custom programs to convert to WET format.

2. THE DATA FORMAT "DISASTER" OF XCov22

Prior to XCov22, as it became obvious that the number of CCD sites that would participate would be large, the ADWO asked for all sites to send in a sample of their data in the format that they would be submitting to HQ during the campaign. Some sites promptly sent in their sample, while others sent their sample as their first campaign data set. While three were submitted in (or converted to) Quilt 9 format, the remaining seven CCD data formats required individual algorithms to convert them to a WET-friendly output ready for further HQ analysis.

After looking at the data formats, it became apparent that even the "well known" (i.e. previously used) data formats were different than used in earlier campaigns, sometimes in subtle ways. In any case, the HQ software for format conversion required extensive modifications to read the new formats. The software had received enough modifications in the past that it was, in some places, unclear what some program statements really meant (as happens in environments where many people contribute to a code). Therefore, we had to rework the entire format conversion code in order to guarantee the output differential magnitudes would be consistent for all the sites.

After converting the first few sets of data submitted, further problems developed. Some sites submitted data that did not match their initial test submissions. Some of the times were in UT, which was easy to discern, but some were in JD, or HJD, or BJED, which at first glance are very similar. Most sites did tell us beforehand what time format they would use, but one site switched from HJD to JD during the run, causing more confusion about the times. Even more troublesome were the "magnitudes"; some magnitudes were sky corrected, some weren't, some were inverse magnitudes, and so on. In one case, two sites claimed that they were sending the same data format, but the data formats were not the same upon examination. One site was gracious enough to send us three separate formats during the campaign and allowed us to choose the format to use! The HQ staff resorted to checking CCD data against concurrent PMT data, and attempting to diagnose timing errors and inverted magnitudes from that comparison. Data header and log information was, many times, not available, which increased the difficulty in sorting things out.

Each of the seven sites performed their own CCD frame reductions, and submitted the extracted magnitudes and times of observations to HQ; it was only possible to convert the data to differential magnitudes and fit a polynomial to remove low order artifacts. Thus, the data HQ received from these seven sites did not preserve the uniformity of the data reduction process and the output product. This non-uniformity made handling the CCD data for XCov22 a particularly painful process.

The events of XCov22 created an impossible situation at HQ. If it were to happen again, the confusion about the incoming data could (and probably would) affect the efficiency of HQ operations, as well as negatively influencing the quality of the data reduction process during the campaign.

3. THE NEW WET DATA FORMAT

Given the trauma of the XCov22 HQ, it is obvious that the future CCD data would have to arrive at HQ in a standard format. This section describes our proposed format for WET CCD data submission.

Table 1. The Quilt 9 data file format. The left column is an actual Quilt 9 observation file from XCov21, and the right gives an explanation of each line of the file.

File element	Explanation
Quilt 9 data	Software identification key
15 Apr 01	Date at start of observations
3:16:50	Time at start of observations
RUN asm-0083	Observation run name
Drive BC	Photometer information
PG1336-018	Object observed
F1 5.000 0.000	Integration time in filter 1
F2 0.000 0.000	Integration time in filter 2
F3 0.000 0.000	Integration time in filter 3
F4 0.000 0.000	Integration time in filter 4
Mot 1F1 F1 F1	Photometer filter information
FC 1 5393	Number of data points
# Who: Anjum Mukadam	Observer
# Where: McDonald	Observatory
# Telscp: 36-in	Telescope
# Progrm: Quilt9.15	Photometer software
# Photom: P3M	Photometer
	This line intentionally left blank
16146 10021 16169	Counts in each channel; the rest
16060 10055 15878	of the data is in this format
15888 10054 16104	

3.1. Why not just use the Quilt 9 format?

With standard instrumentation, it's not surprising that in the past the data submitted to WET HQ was in a uniform format, or easily convertible to this format, the Quilt 9 data format. This is shown in Table 1 (this is an example from an XCov21 dataset taken at McDonald Observatory by Anjum Mukadam). For a PMT system, this works well, since it is possible to read out the PMT consistently in each integration cycle (presuming the computer clock remains accurate). All of the necessary elements are included in the header, and this is read natively by the reduction software (QED). However, this is ill suited to the output of most CCD systems that have sent data to HQ.

The two largest problems with adapting the Quilt 9 data format to CCD data are that it does not include more than three channels, and there is no provision for the inclusion of CCD timing effects. The first problem is easily solved, by adding more columns to the format; the drawback to this is that QED can not read more than three channels, so this requires a new data reduction program (such a program, XQED, is described elsewhere in these proceedings). The second is more problematic, as the Quilt 9 format assumes that every measurement is taken on equal time intervals (i.e. an even cadence). The CCD data cadence is determined by two elements: the integration time, and the readout time. These combine to define the cycle time for the image. If the computer clock is stable, and the control software timing is stable, then the cycle time can always be the same for every image. However, form most systems this is not the case, so each image may have slightly different timings; the entire dataset may suffer from nightly computer clock drifts as well. The Quilt 9 format cannot indicate such departures from uniform time intervals. In addition, the CCD photometry extraction algorithms may register the time of observation as the start of integration, end of integration, center of the integration, center of the cycle, or many other possibilities. It is appropriate to use the midtime of the integration for the time of observation, which, with all these possible timing issues, may not be an exactly evenly spaced time sequence.

Timing issues aside, there are other limitations to the Quilt 9 header. All of the PMT information should be replaced with essential CCD information. The length of the information is limited by the QED software; for example, in addition to the essential information, it would be nice to see all the names of the observers instead of a collection of initials. Including CCD header information directly could be critical in some circumstances, and learning exactly how time was handled during the observations and extractions is absolutely necessary. Included comments could be helpful, in case the data log file was lost. The list of possible items to include in the header is long, and none of them are possible with the Quilt 9 format. A new header format increases flexibility and passes along the information required to manage timing issues with the CCD data.

3.2. Header items

The Quilt header is organized in a particular order, but that isn't ideal when looking at CCD information, since not all CCDs are created equal (and there is not a standard WET CCD system as of yet). We adopt a keyword based system, which preserves flexibility and keeps the information organized.

The initial list of keywords are listed in Table 2. These keywords convey the basic information that must be transmitted to HQ for data reduction. Most of the keywords here are related to telescope,

Keyword	Values	Description
XQEDDATA		All files start with this line
FILENAME	text	Name of the data run file
OBSERVER	text*	The people who gathered the data
HEADER.	text*	CCD header line
COMMENT	text*	Comment line
OBSERVATORY	text	The observatory
TELESCOPE	text	Telescope used at the observatory
LONGITUDE	HH:MM:SS.S	Telescope longitude W of Greenwich, Eng.
LATITUDE	DD:MM:SS.S	Telescope latitude
ALTITUDE	meters	Telescope altitude above sea level
OBJECT	text	The star that was observed
OBJCLASS	text	The class of the object observed
RA	HH:MM:SS.S	Right ascension of star at EPOCH
DEC	DD:MM:SS.S	Declination of star at EPOCH
EPOCH	YYYY.Y	Epoch of observations, either a standard
	START	(i.e. 2000.0), or the observation time at
		the start of the data run
PHOTOMETER	text	Photometer used to acquire data
PHOTTYPE	CCD, PMT	Type of photometer used
ACQSOFT	text	Data acquisition software
OBJKEY	text	Object keyword
OBSKEY	text	Observatory/telescope keyword
PHOTKEY	text	Photometer keyword

Table 2. Header keywords. Note that values marked with a star can occur on multiple lines in the header file.

object and instrument information. This is, of course, not a complete list, as there are many other possible keywords that could be added to the system (such as all kinds of CCD information). The most useful keyword might be the COMMENT, which would allow any pertinent information to be included within the extracted data file.

The last three header keywords in Table 2 are "shortcuts" to some of the other keywords in Table 2. A library of information for observatories, objects and photometers will be created for the new XQED software; each element of the library will be associated with a keyword. This keyword list will be published, and the keywords can be used within XQED to access the information when a data file is loaded.

3.3. Time formats

During XCov22, the CCD data submitted to HQ came in all kinds of time formats, and HQ had to determine what format was

Keyword Values Description DATE-OBS YYYY-MM-DD Date at start of first integration TIME-BEG HH:MM:SS.SSS Time at start of first integration HH:MM:SS.SSS TIME-END Time at start of last integration DRIFT seconds Computer clock drift through the night (compared to a stable clock). Time is drift from start to end of night. **EXP-TIME** Exposure (integration) time seconds READ-TIME seconds CCD readout time (average) CYC-TIME seconds Length of exposure and CCD readout time TIME-FMT The format of the submitted times. Values: **XQED-STD** Time in standard FITS type format (2002-05-12T06:00:00.000). Time is at the beginning of the integration. UT-ONLY Only the UT of observation, no date. Times at beginning of integration. START-END Start time of integration in first column, end time of integration in second column. Times are UT only, not FITS format. PMT-SYSTEM No time is included, as it is presumed that

Table 3. Header keywords related to time.

actually submitted in many cases. To avoid such confustion in the future, all times that are submitted to HQ will be in Universal Time (UTC) format only; no time conversion to JD, HJD, MJD, BJED, local time, or any other time will be accepted. Even with that limitation, it is still possible that the extracted times will have problems. For example, an extraction can return the beginning of the observation, or the center of the observation cycle (the time for integration and readout), the middle of the integration, the end of the integration, or possibly alternate times. Another complication is that the cycle in observing time may not be constant, or there may be clock drift through the night. Even the integration time itself might not be equal for every observation!

the PMT cycling is of even cadence.

To get around all these possible issues, we can expand the time header parameters to include several bits of information (remember, all times mentioned are UT), as shown in Table 3. The TIME-FMT keyword can include the information about what time is extracted, or how the times are encoded in the data; it is the key to understanding the values for the time that are transmitted to HQ. CCD photometry extraction routines should return the time that the exposure began,

or the start and stop times of the exposure (not the cycle), which is ideal for systems that do not have an even integration time cadence. Keywords can be developed to describe other timing situations, but this will be limited; otherwise, we will end up with twenty unique formats again.

3.4. Data formats

Table 4. Keywords for aperture information.

Keyword	Values	Description
EXTRSOFT NCHAN APTYPE	text integer ONESKY, MANYSKY	CCD extraction software Number of data channels submitted The type of aperture setup for the apertures for each object.
APSHAPE	ALLSKY "shape" ANNULAR	data, either one sky aperture or sky The kind of aperture used in the observations.
APSIZE APSCALE	pixels numerical list	The size of the apertures, in pixels A list of numbers showing the ratios of each aperture if they are not all the same size. They should be
		related to the target aperture.

PMT photometers convert the photons that enter the instrument aperture and strike a photomultiplier tube into an electronic signal. This signal is read out by the acquisition software and saved as an integer value; the PMT is then "cleared" and the measurement of a new signal is started. Similarly, CCDs "trap" photons that reach their pixels, convert them to charge, and read them out; the pixel values are saved in a file (usually FITS format), the CCD charge cleared and the next image started. The files created by PMTs are text files (with a header and one line of text per data cycle), easily transmittable by email. A CCD frame is produced for each data cycle; typically raw CCD frames are too large to allow for a night's worth of data to be transmitted to HQ. Therefore, it is necessary for observers to reduce their CCD data into a format that is easily sent to HQ.

As mentioned in Section 2, there are many different CCD extraction programs available. There is no uniform format to the data that results from photometry extraction with these programs. For HQ to receive uniform datasets that allow production of a combined ("first look") timeseries with understood noise properties, it is necessary to make the pre-HQ reductions as minimal as possible. The

Table 5. The Quilt 9 file from Table 1 is reformatted into the new WET format and displayed below.

```
# XQEDDATA
# ACQSOFT: Quilt 9.15
# DATE: 2001-04-15
# TIME-BEG: 3:16:50
# FILENAME: asm-0083
# OBJECT: PG 1336-018
\# EXP-TIME: 5.000
# OBSERVER: Anjum Mukadam
# OBSERVATORY: McDonald Observatory, Texas, USA
# TELESCOPE: 36 inch
# PHOTOMETER: P3Mudgie
# PHOTTYPE: PMT
# TIME-FMT: PMT-SYSTEM
# COMMENT: McDonalds...supersized photons!
16146
      10021
            16169
16060
      10055
            15878
15888
      10054
            16104
```

simplest data is the extracted counts from the CCD pixels, added up to make an "aperture"; this is analogous to the PMT aperture. There are many possible aperture arrangements, but there are three basic procedures for aperture photometry of a CCD frame:

- 1. A sky aperture for all the stellar apertures;
- 2. More than one separate sky aperture measurement;
- 3. A sky aperture for each separate stellar aperture.

For the first case, a series of (whatever shape) apertures are placed around each star (target plus comparison objects), while an extra (empty) aperture is used for a separate measurement of the sky (to be listed last in the aperture measurements). The second case is similar to the first case, except that there are more than one separate sky measurement channel in the data; the data header must note which is which. For the third case, each star has the sky near it measured at the same time; for example, using an annulus surrounding and excluding the star for the sky measurement. Each paired measurement must be listed in the order of (data sky), and each line of data must start with the target star. In any case, all of the apertures used should be the same size; if they are not, then the data values must be scaled and this must be noted.

To summarize, the new WET format for photometry measure-

ment will consist of only the pixel measurements from the CCD, summed over the aperture. No converted magnitudes will be accepted, and the pixel sums for each aperture should be simple sums (in other words, no PSF fitting or other advanced aperture extraction methods). Keywords for the various pieces of information that describe the data submission are summarized in Table 4.

4. A SAMPLE DATA FILE

Table 6. An imaginary CCD data file, using many of the keyword concepts developed.

```
# XQEDDATA
# DATE: 2004-02-29
# TIME-BEG: 12:32:43.000
                            / start of first exposure integration
# TIME-END: 21:52:12.000
                            / start of last exposure integration
# FILENAME: bob-0001.dat
# OBJECT: PG 1336-018
# OBJCLASS: sdB
# OBJKEY: pg1336
# EPOCH: 2000.0
# EXP-TIME: 10.000
# READ-TIME: 2.500
# CYC-TIME: 12.500
# DRIFT: 12
# OBSERVER: Britney Spears
# OBSERVER: Bob Dylan
# OBSERVATORY: Keck Observatory, Mauna Kea, HI
# TELESCOPE: Keck I, 10 meters
# OBSKEY: keck1
# PHOTOMETER: CCD prototype
# PHOTKEY: SP1
# PHOTTYPE: CCD
# NCHAN: 6
# APTYPE: ALLSKY
# APSHAPE: ANNULAR
# APSIZE: 10
# APSCALE: 1 0.85 1 0.85 1 0.85
# ACQSOFT: Spears 1.0b12
# EXTRSOFT: IRAF DAOPHOT
# TIME-FMT: XQED-STD
# COMMENT: Oops, I did it again.....
2001-04-15T03:16:50.000 12147 8490 13181 8654 11881 9002
2001-04-15T03:16:55.000 12199 8422 13177 8621 11923 8887
2001\text{-}04\text{-}15\text{T}03\text{:}17\text{:}00.000\ 12307\ 8448\ 13123\ 8633\ 11874\ 9028
```

With the new data format paradigm in mind, let's look at the Quilt 9 header described above if it was submitted in the new format. Table 2 shows the translated data file; this includes the basics that would be required to reduce the data. All header lines must start with a "#", to differentiate them from the data entries, and all the keywords are followed by a ":". The header data can be followed by a "/", as in a FITS header; everything after this character will be ignored. Header lines also must be no more than 80 characters in length, and keywords should be uppercase.

Table 6 shows a more complete example, this time of a fictitious CCD data set. The library keywords OBSKEY, OBJKEY and PHOTKEY have been used to shorten the example; note that there are still header lines with related information. Completeness is a good thing for a header file. In this example, the CCD aperture used a circle around the star, with a second outside for the sky measurement; the sky annuli were larger (in pixels) than the star aperture, so all were multiplied by 0.85 to scale them appropriately (as noted in the APSCALE keyword).

Including the time for every point is critical, especially when the times are not evenly cycled in a CCD system. It will be a requirement that all sites do this in the future when they submit their CCD data; the choice of time format depend on the circumstances of the instrument (mainly whether it has a variable integration or cycle time).

5. CONCLUSION

There should never be an ambiguity as to what the "raw" data that are submitted to the headquarters of a WET run represent. That, unfortunately, happened during XCov22; it should never happen again. Any CCD data that is submitted to HQ in the future will be required to arrive in the proper format. Quilt 9 format will remain acceptable for PMT data, but PMT observers are encouraged to rework their submissions into the new format, since it allows more information to be transmitted to HQ.

This is our proposed format for CCD data transmission to WET HQ during the XCov campaigns. Readers are encouraged to submit their own comments on and improvements to the format. We want it to be as useful to the community as possible, especially since this format will be the default for XQED (the planned replacement for QED) and be used far into the future.

REFERENCES

Kleinman S. J., Nather R. E., Phillips T. 1996, PASP, 108, 356 Nather R. E., Winget D. E., Clemens J. C., Hansen C. J., Hine B. P. 1990, ApJ, 361, 309