



Modifications to Final Reprocessed FOC Headers before Ingest

D. Fraquelli (CSC/STScI), L. Gardner (STScI), A. Alexov (STScI), Daniel Durand (CADC)

February 12, 2014

Abstract

The report describes the changes made to the FITS header keywords and the data files before the final reprocessed FOC data were ingested into the archive.

1. Introduction

In 2006 the FOC dataset, which consists of 6838 CAL datasets, containing 62,006 FITS files¹, was reprocessed using the best available reference files. Both science and non-science (i.e., DARKs, etc.) exposures were reprocessed to produce as homogeneous a dataset as possible. See FOC ISR-99 (Kamp et al., 2006) for details of the reprocessing.

After reprocessing, the dataset was copied from CADC to ST-ECF and STScI. At STScI, the dataset was placed on disk for direct download, but was not ingested into the archive, nor was the catalog updated. Archive users who searched and downloaded FOC data from STScI receive the original processed data, not the reprocessed version.

To remedy the situation, we investigated the reprocessed dataset, determined what was needed to allow ingest and cataloging of the data, made the changes and then ingested and cataloged the data. What was done to the keywords and data is detailed below.

2. Inspection and Correction of the FOC Dataset

¹ The FITS file total includes 6643 PDQ and 3230 OCX files. These are PODPS Data Quality and Observer Comment files produced manually by operations personnel.

The reprocessed dataset consists of a set of FITS files for each exposure. We ran a FITS verifier, *fverify*, on these files to ensure they were valid FITS. A number of errors were received, most of which fell into one of a number of types. Typical *fverify* errors are given in the Appendix.

- 1) File extensions: .c0f, .c1f, .q0f, .d0f, .cbf, .cuf, .cgf, .shf, .ulf
FITS ASCII Table HDU has reserved Image WCS header keys which are not allowed in non-Image HDUs.
- 2) File extensions: .dgr, .pdq, .shf
Data fill area invalid.
- 3) File extensions: .dgr, .pdq, .ocx
FITS ASCII Table HDU contains non-ASCII-text characters.
- 4) File extensions: .c0f, .c1f, .q0f, .d0f, .cbf, .cuf, .cgf, .shf, .ulf
FITS ASCII Table HDU TDISP11 value 'L1' is not allowed.
- 5) File extensions: .c0f, .c1f, .q0f, .d0f, .cbf, .cuf, .cgf, .shf, .trl, .ulf
EQUINOX value = J2000 is not a floating point number.
- 6) File extensions: .c0f, .c1f, .q0f, .d0f, .cbf
DATE-OBS date string has illegal format.
- 7) File extensions: .drg, .pdq, .ocx
FITS Warning (not error) about column names containing "-" characters.

The following changes were made to bring the data into compliance with the current FITS standard. Following the order of the *fverify* issues above, our solutions to resolve each issue were:

- 1) For those files with a FITS BINARY/ASCII Table extension (HDU1) which contained a copy of the WCS header keywords, along with a few additional non-WCS keywords, we removed all Image-related WCS keywords. Note that the WCS keywords were also in the primary (Image) header, so they were redundant in Table extension. The following keywords were removed from the Table HDU: DATAMAX, DATAMIN, CTYPE1, CTYPE2, CRPIX1, CRPIX2, CRVAL1, CRVAL2, CD1_2, CD2_1, CD2_2, CD1_1.
- 2) The "Data fill area invalid" FITS error went away when we opened, fixed, and closed all the headers in all the files. The process of fixing the headers also fixed the FITS block sizes when the headers were written to disk, hence fixing this problem.
- 3) All non-ASCII-text characters were converted to spaces. We found three such characters throughout the log HDU extensions: unix bells, tabs and nulls. None of these are allowed in ASCII FITS Tables.
- 4) In some of the FITS ASCII Table HDUs, the header keyword TDISP11 value had to be reset from 'L1' to 'A1' to be FITS-compliant. The display format was for the column MIR_REVR in logical format; the logical display format (L1) is deprecated in FITS and was changed to character type format (A1).

5) The type and value of the EQUINOX keyword was changed from string to floating point, and from 'J2000' to 2000 (float), to bring it into compliance with the current FITS standard.

6) The format of the DATE-OBS keyword was changed to comply with the FITS standard. The comment regarding the DATE-OBS format was similarly updated to 'yyyy-mm-dd'.

7) We also cleaned up FITS Warning messages by deleting the "-" character from the column name: (PDQ-FILE → PDQFILE; DRG-FILE → DRGFILE; OCX-FILE → OCXFILE)

Additionally, investigation showed the WCS were corrupt in x2fa0502t in most extensions. Inspection of exposures taken in the same orbit for the same program gave us confidence that the x2fa0502t.q0f.fits file contained valid WCS for this observation. The WCS were fixed based on the q0f value; a COMMENT block describing this change was added to the primary header (see Appendix for COMMENT block text). The HDU1 Table extension in these files had to be similarly changed so that the WCS Table values matched the keyword changes we made. The table contained the WCS keywords as column, row pairs. We updated the HDU1 Table values for those erroneous WCS keywords as in the header (see Appendix, Example #4, for COMMENT block text).

The FITSDATE header keyword was updated to reflect the processing date of the files.

We used python and shell scripts to make the changes listed above. Fverify was re-run after we fixed and cleaned up the data, to ensure no FITS errors remained.

We confirmed that the FITS files were backwards compatible in GEIS format by running the IRAF task 'strfits' on a number of test cases, to be sure no errors occurred. Files were randomly visually inspected in DS9 and FV. The HEASARC FITS diff tool (fdiff) was run on a number of test cases in order to be sure that the only differences present were those which were expected; we diff'ed the ECF versions vs our newly processed versions.

3. Keyword and Catalog Value Differences

The values of the FITS header keywords were compared to the existing values in the database (archive catalog). For the majority of the keywords, values matched between the headers and the database, or the differences were at the level of precision of the value or were a formatting difference. For those cases with a significant difference between the header and the database, an investigation was conducted to determine which should be used.

In the table below we list 29 keywords where a difference was seen between the FITS files and the database, and the number of datasets in which each difference was seen.

Header keyword	Number of datasets with differences	Comment
BACCORR	1	Header keyword deemed correct
BACHFILE	1285	Header keyword deemed correct
COMPTAB	6836	Set to N2A13493M_TMC.FITS, but see text
CRVAL1	3169	Header keyword deemed correct
CRVAL2	3091	Header keyword deemed correct
DATAMAX	5149	Header keyword deemed correct
DATAMIN	733	Header keyword deemed correct
EXPEND	136	Header keyword deemed correct
EXPSTART	136	Header keyword deemed correct
GEOCORR	117	Header keyword deemed correct
GEOHFILE	2406	Header keyword deemed correct
GRAPHTAB	6836	Set to M6E1130CM_TM.G.FITS, but see text
ITFCORR	1	Header keyword deemed correct
ITFHFILE	1277	Header keyword deemed correct
MTFLAG	4997	Pass through from scheduling
ORIENTAT	3856	Header keyword deemed correct
PARENTID	193	Pass through from scheduling
PHOTBW	5341	Header keyword deemed correct
PHOTPLAM	5259	Header keyword deemed correct
PR_INV_L	115	Pass through from scheduling
PXLCORR	20	Header keyword deemed correct
SDEHFILE	5159	Header keyword deemed correct
TARDESCR	2893	See text for resolution
TARGCAT	2494	See text for resolution
TARGNAME	4	See text for resolution
UNICORR	48	Header keyword deemed correct
UNIHFILE	596	Header keyword deemed correct
UNITAB	5716	Header keyword deemed correct
UTC0	273	Header keyword deemed correct

Those keywords whose values were used in calibration (all *CORR, *TAB and *FILE keywords) and/or computed by the processing (CRVAL1, CRVAL2, DATAMIN, DATAMAX) were deemed correct in the headers, with two exceptions. The COMPTAB and GRAPHTAB values in the headers were logical pointers to a file on disk rather than the name of the reference file. As only one GRAPHTAB and one COMPTAB file were ever produced for the FOC, we updated the headers of all but 4 datasets as follows.

- GRAPHTAB = 'M6E1130CM_TM.G.FITS'

- COMPTAB= 'N2A13493M_TMC.FITS'

For 4 datasets, X19P5101T, X3IU0806R, X4560201R and X4KO2701R, GRAPHTAB and COMPTAB were left blank because there was no best SDE reference file for these exposures. The SDE reference file contains the spectrographic detector efficiency correction.

UTC0 reports the base value of the spacecraft clock. The value is corrected from time to time, but is not changing constantly. It is used in the calculations for all other times reported in the FITS headers, including EXPSTART and EXPEND. All cases where the database and the headers were different are confined to the time interval December 17, 1996 to April 7, 1997. We could find no problem reports addressing this issue. Because the other times are in agreement and the final processing would have the full set of UTC0 updates available during the processing, we elected to update the database with the UTC0, EXPSTART and EXPEND values from the headers.

ORIENTAT holds the position angle of the aperture's y-axis with respect to North measured through East. It is calculated from two angles: the position angle of the +y axis of the aperture with respect to the V3axis and the position angle of the V3 axis of HST with respect to North measured through East. For the aperture, the position angle of the +y axis was taken from the Project Data Base (PDB). In OPUS (aka the pre-archive or production pipeline) the value resides in the si_aperture table in the opus database. The position angle of the V3 axis is taken from the Mission Schedule. During the time the FOC was active a number of problem reports involving the value of ORIENTAT were filed, suggesting some differences between the database and headers are expected.

Using the PDB (via the opus database table si_aperture), and using the position angle of the V3 axis w. r. t. N from the dadsops database, we calculated ORIENTAT for a number of cases, both for the F/96 and the F/48 cameras. In almost all cases, the values calculated for ORIENTAT agree with the headers. The remaining cases appear to be exposures taken during Earth occultation (i.e., not on guide stars). Based on these results, we updated the catalog with the ORIENTAT values from the headers.

A number of header keywords held "pass through" values from the proposal or Planning and Scheduling. Where there were differences, the value was taken from the database and placed in the FITS header. These keywords are listed here, although not all appear in the above table.

ACCPDATE
MTFLAG
PARALLAX
PARENTID
PR_INV_L
PSTPTIME
PSTRTIME
REDSHIFT
SCAN_ANG
SCAN_LEN

MAG_R
MAG_V
MU_RA

The FITSDATE header keyword was updated in all the files which contained this keyword; the value was updated to “9-DEC-2013” to reflect the fact that the FITS files were updated on this date.

The TARDESCR, TARDESCR2, TARGCAT, TARGCAT2, and TARGNAME keywords hold the target description, target categories, target keywords and target name as provided in the proposal. During the first ten or so years of the mission, instructions to the proposers on how to provide this information and the code that packaged the information into keywords that were passed to OPUS changed a number of times. A major effort was undertaken in 2005 to update the archive catalog (database) to bring the target descriptors, target categories and target keywords into agreement with the, by then, approved method of specifying the information. Rather than repeat this work, we elected to update TARGNAME in the header with the value from the database. We removed TARDESCR, TARDESCR2, TARGCAT and TARGCAT2 as keywords and placed them in a series of comments in the header. Both the values from the header and the values from the database are contained in the comments added to the shf file of each dataset. An example is show here.

```
COMMENT * ----- STScI -----  
COMMENT *  
COMMENT * Header keywords TARGCAT, TARGCAT2, TARDESCR, TARDESC2 have  
COMMENT * been removed from the shf headers; their values and database values  
COMMENT * have been preserved in this comment block (created July 2013).  
COMMENT *  
COMMENT * Header (old) keyword value TARGCAT:  
COMMENT * 'GALAXY'  
COMMENT * Header (old) keyword value TARGCAT2:  
COMMENT * 'CLUSTER OF GALAXIES'  
COMMENT * Header (old) keyword value TARDESCR:  
COMMENT * 'GALAXY;ELLIPTICAL;;CLUSTER OF GALAXIES;HIGH REDSHIFT CLUSTER;GRAVITA'  
COMMENT * Header (old) keyword value TARDESC2:  
COMMENT * 'TIONAL LENS'  
COMMENT *  
COMMENT * Database (new) keyword value TARGCAT:  
COMMENT * GALAXY  
COMMENT * Database (new) keyword value TARGCAT2:  
COMMENT * CLUSTER OF GALAXIES  
COMMENT * Database (new) keyword value TARDESCR:  
COMMENT * GALAXY;ELLIPTICAL;;CLUSTER OF GALAXIES;HIGH REDSHIFT CLUSTER;GRAVITA  
COMMENT * Database (new) keyword value TARDESC2:  
COMMENT * TIONAL LENS
```

Finally, we documented our work in the headers. The following FITS COMMENT block was added to ALL FOC files (note, shf file has an additional COMMENT block, which is located after the following summary block, see above):

```
COMMENT * ----- STScI -----
COMMENT *
COMMENT * Details on the final header updates can be found in the FOC ISR 100.
COMMENT * (1) Headers were fixed to be FITS-compliant.
COMMENT * (2) The following header keys were updated with values from the db:
COMMENT *     ACCPDATE, MAG_R, MAG_V, MTFLAG, MU_RA, PARALLAX, PARENTID,
COMMENT *     PR_INV_L, PSTPTIME, PSTRTIME, REDSHIFT, SCAN_ANG, SCAN_LEN,
COMMENT *     and TARGNAME
COMMENT * (3) Many reserved WCS keywords were removed from the FITS ASCII
COMMENT *     Table extension (HDU1) because they are only allowed in IMAGE
COMMENT *     extensions.
COMMENT * (4) In HDU1, when applicable, TDISP11 (MIR_REVR column) was
COMMENT *     changed from 'L1' to 'A1', since 'L' is no longer allowed.
COMMENT * (5) GRAPHTAB and COMPTAB were updated from logical pointers to
COMMENT *     the reference file names: GRAPHTAB = M6E1130CM_TMG.FITS and
COMMENT *     COMPTAB='N2A13493M_TMC.FITS
COMMENT *
COMMENT * Final updates performed at STScI in December 2013.
COMMENT * ----- STScI -----
```

4. Ingest Changes

The database was populated from datasets produced by the production pipeline, which produces a standard dataset as follows, where the pdq extension was manually produced.

x58h0107m_c0f.fits	x58h0107m_dgr.fits	x58h0107m_shf.fits
x58h0107m_c1f.fits	x58h0107m_pdq.fits	x58h0107m_trl.fits
x58h0107m_d0f.fits	x58h0107m_q0f.fits	x58h0107m_ulf.fits

The final reprocessed datasets were produced using a standalone version of calfoc, the FOC calibration software, which would often produce an alternate to the c0f and c1f files. These alternate extensions, cbf, cgh and cuh, indicate full processing was not done and are defined as follows.

cbf - Output from first set of processing steps (background, ITF, pixel splitting, and absolute DE) when no subsequent processing is done.

cgf - Geometrically corrected image if no flat-field correction is done.

cuf - Flat-fielded image if no geometric correction is done.

For example, for x19p5101t, a cbf file is produced instead of the standard cof/c1f pair.

x19p5101t_cbf.fits x19p5101t_dgr.fits x19p5101t_q0f.fits x19p5101t_trl.fits
x19p5101t_d0f.fits x19p5101t_pdq.fits x19p5101t_shf.fits x19p5101t_ulf.fits

When data are ingested, the database (archive catalog) is populated by placing the value of the headers keywords into specific tables. Because each dataset consists of a number of files, and some keywords may appear in the headers of more than one of these files, there is a defined order that indicates which file should be the source of the keyword value. The ingest software starts with the first file in the order listed below. If the keyword is found, that value is placed in the database. If the keyword is not found, the next file is opened, and so on. The order of the files is fixed and the value in the database comes from the first instance of the keyword in the ordering.

For pipeline processing the order was c0f, d0f, shf, ulf.

For the standalone processing the order was c0f, cbf, cgf, cuf, d0f, shf, ulf.

Acknowledgements

We thank Phil Hodge and Robert Jedrzejewski for very helpful discussion on the "non-ASCII" character issues, and Mike Swam for his insight on the use of UTC0 in the science processing pipeline. We thank Brad Whitmore and Lou Strolger for their review comments on the ISR.

Reference

Kamp, I., Durand, D. and Micol, A. 2006, Instrument Science Report FOC-99, "Final reprocessing of FOC data."

Appendix

The following two examples list typical fits verify errors as discussed in Section 2.

1) Example w.r.t. issues: EQUINOX non-floating point number, TDISP11 format not allowed, WCS Image keys not allowed in Table HDU, Data fill area invalid)

```
archopsvm1> fverify x19p0901t.c1f.fits
```

```
FVERIFY V4.0.0 (CFITSIO V3.250)
```

```
-----
```

HEASARC conventions are being checked.

File: x19p0901t.c1f.fits

2 Header-Data Units in this file.

===== HDU 1: Primary Array =====

*** Error: Keyword #115, DATE-OBS: (from CFITSIO error stack:)
input date string has illegal format (ffs2dt):
7/05/93

*** Error: Keyword #109, EQUINOX: value = J2000 is not a floating point
number. The value is entered as a string.

170 header keywords

32-bit floating point pixels, 2 axes (1024 x 1024),

===== HDU 2: ASCII Table =====

*** Error: Keyword #75, TDISP11: Format L1 cannot be used for TFORM "A1".
*** Error: Keyword #128, DATAMAX is not allowed in the Bin/ASCII table.
*** Error: Keyword #127, DATAMIN is not allowed in the Bin/ASCII table.
*** Error: Keyword #135, CTYPE1 is not allowed in the Bin/ASCII table.
*** Error: Keyword #136, CTYPE2 is not allowed in the Bin/ASCII table.
*** Error: Keyword #121, CRPIX1 is not allowed in the Bin/ASCII table.
*** Error: Keyword #122, CRPIX2 is not allowed in the Bin/ASCII table.
*** Error: Keyword #119, CRVAL1 is not allowed in the Bin/ASCII table.
*** Error: Keyword #120, CRVAL2 is not allowed in the Bin/ASCII table.
*** Error: Keyword #124, CD1_2 is not allowed in the Bin/ASCII table.
*** Error: Keyword #125, CD2_1 is not allowed in the Bin/ASCII table.
*** Error: Keyword #126, CD2_2 is not allowed in the Bin/ASCII table.
*** Error: checking data fill: Data fill area invalid
137 header keywords

X19P0901T_CVT.C1H.TAB (18 columns x 1 rows)

Col#	Name (Units)	Format
1	CRVAL1	D25.17
2	CRVAL2	D25.17
3	CRPIX1	E15.7
4	CRPIX2	E15.7
5	CD1_1	E15.7
6	CD1_2	E15.7

```

7 CD2_1      E15.7
8 CD2_2      E15.7
9 DATAMIN    E15.7
10 DATAMAX   E15.7
11 MIR_REVR (LOGICAL-) A1
12 ORIENTAT  E15.7
13 FILLCNT   I12
14 ERRCNT    I12
15 FPKTTIME  D25.17
16 LPKTTIME  D25.17
17 CTYPE1 (CHARACTER*8) A8
18 CTYPE2 (CHARACTER*8) A8

```

+++++ Error Summary +++++

HDU#	Name (version)	Type	Warnings	Errors
1	Primary Array	0	2	
2	X19P0901T_CVT.C1H.TAB	ASCII Table	0	13

**** Verification found 0 warning(s) and 15 error(s). ****

```

////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////

```

2) Example w.r.t. issue: non-ASCII-text characters

```
archdevvm1> fverify x0n00302t.trl.fits
```

```
FVERIFY V4.0.0 (CFITSIO V3.250)
```

```
-----
```

HEASARC conventions are being checked.

File: x0n00302t.trl.fits

2 Header-Data Units in this file.

```
===== HDU 1: Primary Array =====
```

9 header keywords

Null data array; NAXIS = 0

```
===== HDU 2: ASCII Table =====
```

*** Error: row 91 data contains non-ASCII-text characters.

*** Error: This ASCII table contains 5 non-ASCII-text characters

===== HDU 2: ASCII Table =====

*** Warning: Column #1: Name "DGR-FILE" contains character '-' other than letters, digits, and "_".

*** Error: checking data fill: Data fill area invalid

13 header keywords

X19P0101T.DGR (1 columns x 5 rows)

Col#	Name (Units)	Format
1	DGR-FILE	A132

+++++ Error Summary +++++

HDU#	Name (version)	Type	Warnings	Errors
1	Primary Array	0	0	
2	X19P0101T.DGR	ASCII Table	1	1

**** Verification found 1 warning(s) and 1 error(s). ****

//
//

4) Files x2fa0502t.c0f.fits, x2fa0502t.c1f.fits and x2fa0502t.d0f.fits had erroneous WCS header keys in HDU0 and table values in HDU1. Both were fixed and the following COMMENT block was added to both HDU0 and HDU1 in order describe the issue in both places (header keys and table values, respectively).

COMMENT * ----- STScI -----
COMMENT *
COMMENT * Files x2fa0502t.c0f.fits, x2fa0502t.c1f.fits and x2fa0502t.d0f.fits
COMMENT * had erroneous values within HDU1. These (WCS) header keyword values
COMMENT * were updated using a copy of the x2fa0502t.q0f.fits WCS header
COMMENT * values because the file appears to be correct out of the root set:
COMMENT *
COMMENT * CRVAL1 = 322.4935
COMMENT * CRVAL2 = 12.16683333333
COMMENT * CRPIX1 = 212.
COMMENT * CRPIX2 = 281.
COMMENT * CD1_1 = -1.522923E-6
COMMENT * CD1_2 = 3.682517E-6
COMMENT * CD2_1 = 3.682517E-6
COMMENT * CD2_2 = 1.522923E-6

COMMENT * ORIENTAT= 67.53223
COMMENT * FPKTTIME= 49622.2772777965
COMMENT * LPKTTIME= 49622.2790992664
COMMENT *
COMMENT * Previous values were overwritten; these values were:
COMMENT *
COMMENT * x2fa0502t.c0f.fits
COMMENT *
COMMENT * CRVAL1 = 216385.207354576
COMMENT * CRVAL2 = 4.0048880340425E-20
COMMENT * CRPIX1 = 3.390789E-20
COMMENT * CRPIX2 = 3.390789E-20
COMMENT * CD1_1 = -3.986111E-6
COMMENT * CD1_2 = 3.596501E-27
COMMENT * CD2_1 = 3.596501E-27
COMMENT * CD2_2 = 3.986111E-6
COMMENT * ORIENTAT= 5.169558E-20
COMMENT * FPKTTIME= 3.3907891472831E-20
COMMENT * LPKTTIME= 3.3907891472831E-20
COMMENT *
COMMENT * x2fa0502t.c1f.fits
COMMENT *
COMMENT * CRVAL1 = 216385.207354576
COMMENT * CRVAL2 = 4.0048880340425E-20
COMMENT * CRPIX1 = 3.390789E-20
COMMENT * CRPIX2 = 3.390789E-20
COMMENT * CD1_1 = -3.986111E-6
COMMENT * CD1_2 = 3.596501E-27
COMMENT * CD2_1 = 3.596501E-27
COMMENT * CD2_2 = 3.986111E-6
COMMENT * ORIENTAT= 5.169558E-20
COMMENT * FPKTTIME= 3.3907891472831E-20
COMMENT * LPKTTIME= 3.3907891472831E-20
COMMENT *
COMMENT * x2fa0502t.d0f.fits
COMMENT *
COMMENT * CRVAL1 = 216385.207354576
COMMENT * CRVAL2 = 4.0048880340425E-20
COMMENT * CRPIX1 = 3.390789E-20
COMMENT * CRPIX2 = 3.390789E-20
COMMENT * CD1_1 = 3.390789E-20
COMMENT * CD1_2 = 6.879854E11
COMMENT * CD2_1 = 1.722074E11
COMMENT * CD2_2 = 4.398567E21
COMMENT * ORIENTAT= 5.169558E-20
COMMENT * FPKTTIME= 3.3907891472831E-20

COMMENT * LPKTTIME= 3.3907891472831E-20
COMMENT *
COMMENT * Software used:
COMMENT * heasoft-6.10 - ftools - fv & fmodhead
COMMENT *
COMMENT * Generated: 30-DEC-2013 @STScI
COMMENT *