

A FITS convention to support longer keyword names

W. Pence, L. Chiappetti, M. Currie, J. Mink, R. Seaman

Version 0.4
11 November 2014

1 Introduction

The FITS scientific data format was invented in 1979 and has greatly benefited astronomical research by providing a common data format that is used by virtually all scientists and institutions in the field. FITS continues to serve astronomers well and will likely still be used for decades to come. It is widely acknowledged, however, that some of the original design decisions made when the format was created, while reasonable at the time, are now unnecessarily restrictive and inhibit new uses of FITS files in modern applications.

One of the most frequently cited restrictions is the 8-character limit on the length of keyword names in FITS headers. This restriction¹ negatively affects FITS file designers and users in several significant ways:

1. **Reduces Clarity** - The 8-character limit leads to cryptic mnemonics names that are confusing and difficult to remember for both software developers and end users of the data files. Examples can be found in almost every FITS file, such as `UCH2CJTM`, instead of a more descriptive name like `TEC_COLD_JUNCTION_2_TEMP`, or `ROTRTRTG`, instead of perhaps `TARGET_ROTATION_RATE`.

This problem is exacerbated when dealing with arrays of indexed keywords or keywords associated with columns in a FITS table (e.g., `TUNITn`) where the root name is limited to even fewer characters to allow space for the numeric suffix. The World Coordinate System (WCS) keywords (as shown in Table 22 of Version 3 of the FITS Standard) are a good illustration of this problem. In several instances, the root name of the WCS keyword is limited to just 2 characters (e.g., `12PC104A`) and in the most extreme case, the root name is reduced to a single letter (e.g., the letter `V` in `7V104_9A`) because of the need to also encode the column number, up to 2 coordinate axis numbers, a parameter sequence number, and the alternate WCS version code letter, all within only 8 characters!

2. **Hinders Innovation** - The keyword length limitation hinders innovation in developing new conventions for representing the complex data products that are being generated by current and future astronomical instruments. FITS binary tables, for example, offer great flexibility in storing complex data structures, but the 8-character limit is a major obstacle when trying to invent keyword names that also convey relationships or associations to other data elements (e.g., by adding a unique and easily identifiable prefix to the names of the related set of keywords). To cite another WCS example, in the draft paper on representations of distortions within coordinate systems, it was necessary to invent an entirely new type of ‘record-valued’ keyword structure because it is impossible to represent the large number of possible distortion keywords in any coherent way with only 8-character names.

¹While the 8-character limit is restrictive by today’s standards, it was actually a significant increase over the 6-character limit on variable names in the Fortran-77 computer language which formed the basis of much of the FITS header record syntax.

- 3. **Hinders Data Exchange** - The keyword length restriction is also an obstacle when using FITS as an ‘interchange format’ for transporting data between different computer systems, which was the primary motivation for developing the FITS format in the first place. Other current scientific data formats allow meta-data parameters to have symbolic names longer than 8 characters, so representing them as FITS header keywords is problematic.

In order to help alleviate these problems, this paper describes a simple FITS convention for supporting keyword names up to 55 characters long. One important feature of this convention is that it conforms to the current FITS format requirements and thus requires no modifications to the FITS Standard document to be considered ‘legal’ FITS usage.

2 Free-format convention to support longer FITS keyword names

The 8-character limit on the length of keyword names ultimately comes from the fact that the ‘value indicator’ (the equals sign character followed by a space character, ‘=␣’, that separates the keyword name from the value field) is constrained to be in bytes 9 and 10 of the 80-byte keyword record. The obvious way to make room for longer keyword names, then, is to lift this fixed format restriction and allow the value indicator to be located anywhere (within reason) within the keyword record. Since there is a trade-off between the length of the keyword name and the number of characters remaining in the 80-byte record for the keyword value and comment fields, this convention limits the keyword names to at most 55 characters in length (with the value indicator in bytes 56 and 57), which leaves a minimum of 23 characters available for the value field. This is enough space to represent the longest 64-bit integer values (20 characters long) as well as floating-point numbers expressed in exponential notation with 16 decimal places of precision and 3-digit exponents (23 characters long) as shown below:

```
1234567890123456789012345678901234567890123456789012345678901234567890
KEY_NAME_AABBCCDDEEFFGGHHIIJJKKLLMMNNOOPPQQRRSSTTUUVVWW= -1.234567890123456E-123
```

Section 4.1 of the FITS Standard document defines how FITS header records should be interpreted. Of particular relevance here is sub-section 4.1.2.3:

In keyword records that contain the value indicator in bytes 9 and 10, the remaining bytes 11 through 80 of the record shall contain the value, if any, of the keyword, followed by optional comments. In keyword records without a value indicator, bytes 9 through 80 should be interpreted as commentary text, however, this does not preclude conventions that interpret the content of these bytes in other ways.

In other words, the default interpretation of keyword records that do not have the value indicator in bytes 9 and 10 is to treat the characters in bytes 1–8 as the keyword name, and all the remaining characters in bytes 9–80 as commentary text. However, the Standard explicitly allows other conventions, such as this one, to be established which interpret the keyword record differently. In fact, this convention is not the first to make use of this provision in the FITS Standard: The ESO HIERARCH convention, with keywords like
HIERARCH ESO INS OPTI-3 ID = 'ES0427 ' / Opt. ID
has been in widespread use since 1990, and is another example of a convention that ascribes an alternate interpretation to FITS header records that do not have a value indicator in bytes 9–10.

2.1 Allowed character set

The longer keyword names supported by this convention inherit all of the requirements that apply to normal 8-character keywords, such as the name must begin in byte 1 of the keyword record and the name cannot contain embedded space characters. The FITS Standard also requires that the 8-character keyword names only contain the Latin alphabetic characters ‘A’ through ‘Z’, the digits 0 through 9, and the underscore and hyphen. The current convention, however, relaxes this requirement for characters 9–55 of the name and allows the following additional set of characters: the 26 lowercase letters ‘a’–‘z’, and the special characters ‘+’ (plus sign), ‘\$’ (dollar sign), ‘.’ (period), and ‘@’ (“at” sign). If the lowercase letters are used, the keyword name should be treated as case-insensitive, so, for example, the names `VOLTAGE_MAX`, `VOLTAGE_Max`, and `VOLTAGE_max` all refer to the same keyword.

It should be emphasized again that the FITS Standard does not allow these additional characters within the first 8 characters of the keyword name.

2.2 Detailed requirements of this convention

Any FITS header that uses this long keyword name convention must contain a keyword named `FITSVERS` with a floating-point value greater than or equal to 2.0. The presence of this keyword serves as a flag to FITS reading and writing software that this long keyword name convention should be used when interpreting the header records within that HDU.² This keyword also helps guard against the unlikely event that a header record in an existing FITS file would be mistakenly interpreted as using this convention.

A FITS 80-byte keyword record should be interpreted under this convention if all the following conditions are met:

- The header contains the `HEADVERS` keyword with a floating-point value greater than or equal to 2.0.
- The header record contains a value indicator (‘=␣’) with the equals sign located somewhere between byte 10 and byte 56, inclusive, of the record.
- The bytes that precede the value indicator only contain characters that are allowed within the keyword name, as defined above. Note that 1 or more space characters may follow the keyword name, preceding the value indicator.

If all these conditions are met, then the remaining bytes on the keyword record following the value indicator should be interpreted as containing the value field followed by an optional comment field, according to the existing rules defined in the FITS Standard.

3 Impact of this convention

It is anticipated that use of this convention within new FITS files should have minimal impact on existing software packages. If legacy software does not support this convention, then according to the FITS Standard, it should simply interpret the first 8 characters of the keyword record as the keyword ‘name’, and because there is no value indicator in bytes 9–10, the rest of the keyword record should be interpreted as commentary text, similar to the way `COMMENT` or `HISTORY` keywords

²Other future FITS conventions that are developed may wish to use this same keyword for a similar purpose, but may place different requirements on the value.

are treated. Such legacy software, of course, will not be able to correctly interpret the intended value of this new type of keyword, but as long as FITS file designers are judicious in the use of this convention (i.e., at least initially, only use this convention for new mission-specific keywords that are unlikely to be required by legacy software packages) this is not likely to be a significant problem.

The cost of retrofitting existing software to fully support this convention is also anticipated to be relatively modest. The biggest impact will be on the standard FITS libraries that most applications software packages rely on to read and write FITS keywords. As an example, a trial version of the CFITSIO library (not yet publicly released) was upgraded to support this convention with about 2 programmer-days of effort.

Once a FITS library has been upgraded, the applications programs that are linked to that library will then inherit the ability to read and write longer keyword names. In some cases, however, it may also be necessary to make further small modifications to the application program itself, for example, to increase the size of string variables in order to store the longer keyword names in computer memory..

4 Summary

The 8-character keyword name limitation has a significant negative impact on current FITS users and is impeding the development of new FITS conventions. Removing this limitation provides significant benefits to future projects by allowing them to create clearer, more self-documenting keyword names at a relatively small cost to existing software systems.