FITS and DRWG BoF

Flexible Image Transport System: Data Representation Working Group





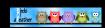
introductory talk

FITS WG

DRWG

10) conclusion

Contributions and discussion





DRWG

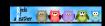
BoF session outline – 2

introductory talk

FITS WG

- 3) work done so far
- WG reorganization
- work to do

- 10) conclusion
- Contributions and discussion





introductory talk

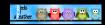
FITS WG

DRWG

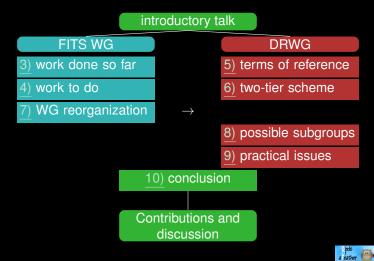
- 3) work done so far
- work to do
- WG reorganization

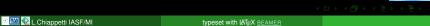
10) conclusion

Contributions and discussion







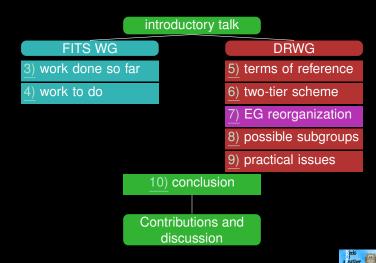


















FITS – recent updates to the standard – 3.a.

A version 4.0 draft of the FITS standard has been released in July 2016 via http://fits.gsfc.nasa.gov/fits_standard.html.

All items in the draft have been formally voted by the IAU FITS WG and therefore are already in effect since their announcement on FITSBITS.

The draft is currently undergoing a mere language editing and will be definitively released asap.







FITS – recent updates to the standard – 3.b

Refer to FITS 4.0 doc (in particular App. H.3) for details on changes/additions:

- chapter on time WCS (Rots et al. 2015 already incorporated in 2014)
- chapter on tiled image and table compression (incorporated conventions)
- incorporated CONTINUE convention for long-string valued kwds
- CHECKSUM and DATASUM data integrity keywords
- column limit keywords (TLMINn, TLMAXn ...)
- incorporated usage convention for preallocating scratch space in a FITS header
- mention of the Green Bank convention (App. L)
- reserved INHERIT keyword to support relevant convention (App. K)





- reorganization of FITS WG membership
- procedural vote on FITS 4.0 document
- reorganization of FITS voting rules
- resuming actions already initiated or mentioned
- anything else



- long keyword names (55-74 char)
- extended character set (a-z _-\$.: blank?)

- a MIME extension



- long keyword names (55-74 char)
- extended character set (a-z _-\$.: blank?)

- a MIME extension



- long keyword names (55-74 char)
- extended character set (a-z _-\$.: blank?)
- (incorp.) spatial region convention
- (incorp.) various WCS conventions (SIP, TNX, TPV, ZPX)
- (incorp.) ESO HIERARCH

- a MIME extension



- long keyword names (55-74 char)
- extended character set (a-z _-\$.: blank?)
- (incorp.) spatial region convention
- (incorp.) various WCS conventions (SIP, TNX, TPV, ZPX)
- (incorp.) ESO HIERARCH
- (incorp.) Hierarchical grouping
- (incorp.) FOREIGN extension

- a MIME extension



initiated by old task force/s (→Jan 2015) or mentioned otherwise (by me)

- long keyword names (55-74 char)
- extended character set (a-z _-\$.: blank?)
- (incorp.) spatial region convention
- (incorp.) various WCS conventions (SIP, TNX, TPV, ZPX)
- (incorp.) ESO HIERARCH
- (incorp.) Hierarchical grouping
- (incorp.) FOREIGN extension
- (incorp.) substring array in BINTABLEs
- (incorp.) various complex project-specific conventions
- dedicated METADATA extension
- support to Unicode strings
- an INDEX HDU for MEFs (FAR = FITS ARchive)
- a MIME extension



initiated by old task force/s (\rightarrow Jan 2015) or mentioned otherwise (by me)

- long keyword names (55-74 char)
- extended character set (a-z _-\$.: blank?)
- (incorp.) spatial region convention
- (incorp.) various WCS conventions (SIP, TNX, TPV, ZPX)
- (incorp.) ESO HIERARCH
- (incorp.) Hierarchical grouping
- (incorp.) FOREIGN extension
- (incorp.) substring array in BINTABLEs
- (incorp.) various complex project-specific conventions
- dedicated METADATA extension
- support to Unicode strings
- an INDEX HDU for MEFs (FAR = FITS ARchive)
- a MIME extension
- an AtEnd convention (transmit files of unknown length)



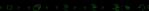


DRWG – terms of reference – 5.a

from original Letter of Intent for IAU Commission B2 "Data and Documentation" by Bob Hanisch

of the new WG. Subcommittees focused on other data representation standards can also be created







DRWG – terms of reference – 5.b

from original Letter of Intent for IAU Commission B2 "Data and Documentation" by Bob Hanisch

2.5 A WG for Data Representations (an expanded WG FITS) will help ensure and maintain the fluent interoperability of telescope data that has not only made multi-wavelength astronomical research commonplace, but has also made astronomy's data management practices the envy of many other disciplines. The WG FITS has been the custodian and advocate for the Flexible Image Transport System format since it was formally endorsed by the IAU in 1982; however, the more recent data landscape has broadened substantially, and many new facilities are exploring alternatives to the FITS standard in order to manage their issues of data scale and complexity. The WG FITS will therefore be expanded, and renamed Data Representations. It is vital to manage a careful and minimally disruptive transition from FITS to more modern and capable data representations, and in order to assure that continuity we plan to retain the FITS governance structure through a FITS Subcommittee that is part of the new WG. Subcommittees focused on other data representation standards can also be created as needed.



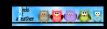


DRWG – terms of reference – 5.c

from original Letter of Intent for IAU Commission B2 "Data and Documentation" by Bob Hanisch

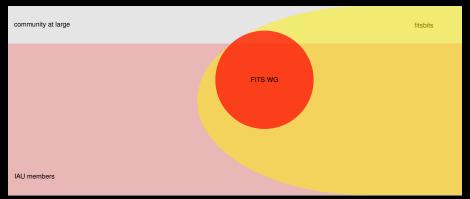
2.5 A WG for Data Representations (an expanded WG FITS) will help ensure and maintain the fluent interoperability of telescope data that has not only made multi-wavelength astronomical research commonplace, but has also made astronomy's data management practices the envy of many other disciplines. The WG FITS has been the custodian and advocate for the Flexible Image Transport System format since it was formally endorsed by the IAU in 1982; however, the more recent data landscape has broadened substantially, and many new facilities are exploring alternatives to the FITS standard in order to manage their issues of data scale and complexity. The WG FITS will therefore be expanded, and renamed Data Representations. It is vital to manage a careful and minimally disruptive transition from FITS to more modern and capable data representations, and in order to assure that continuity we plan to retain the FITS governance structure through a FITS Subcommittee that is part of the new WG. Subcommittees focused on other data representation standards can also be created as needed.

3. Associates The remit of this Commission encompasses the understanding of astronomical data in many forms, and requires both scientific and technical expertise. Commission 5 routinely relied upon Associates, particularly in the WG Libraries and WG FITS, and we expect that need to continue.



DRWG – two-tier scheme – 6.a

Outer tier: forum for IAU members interested in the matter Inner tier: Special Expert Groups (SEGs)



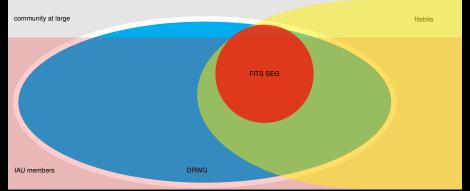
mimicked on FITS WG vs FITSBITS: community contributes, WG decides



DRWG – two-tier scheme – 6.b

Outer tier: forum for IAU members interested in the matter

Inner tier: Special Expert Groups (SEGs)



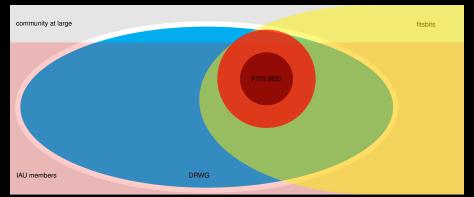
FITS WG becomes first SEG



DRWG – two-tier scheme – 6.c

Outer tier: forum for IAU members interested in the matter

Inner tier: Special Expert Groups (SEGs)



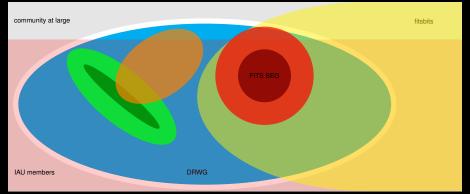
institutional vs individual members?



DRWG – two-tier scheme – 6.d

Outer tier: forum for IAU members interested in the matter

Inner tier: Special Expert Groups (SEGs)



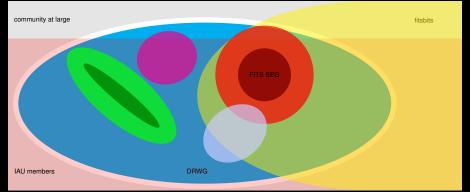
add other SEGs: community DRWG community contributes, SEG decides



DRWG – two-tier scheme – 6.e

Outer tier: forum for IAU members interested in the matter

Inner tier: Special Expert Groups (SEGs)



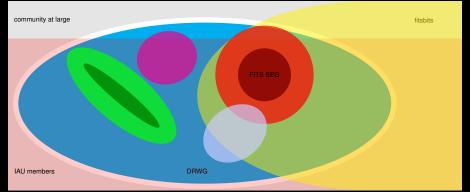
overlapping or distinct, permanent or short-term ...



DRWG – two-tier scheme – 6.e

Outer tier: forum for IAU members interested in the matter

Inner tier: Special Expert Groups (SEGs)



overlapping or distinct, permanent or short-term ...





FITS in DRWG – reorganization – 7.a







FITS in DRWG – reorganization – 7.b

- some members are stepping down or have retired replacements and new members (CDS StScI LSST China India ...) mantain total number within 25
- possible change in voting rules: institutional vs individual membership e.g. quorum of 50% of all members and 3/4 of institutional members institutional members must vote institutional NO delays by 3 months







FITS in DRWG – reorganization – 7.c

- some members are stepping down or have retired replacements and new members (CDS StScI LSST China India ...)
 mantain total number within 25
- possible change in voting rules: institutional vs individual membership current double quorum (3/4 members shall vote; 3/4 votes yes) a single NO delays taking into effect by 3 months e.g. quorum of 50% of all members and 3/4 of institutional members institutional members must vote institutional NO delays by 3 months







FITS in DRWG – reorganization – 7.d

- some members are stepping down or have retired replacements and new members (CDS StScI LSST China India ...)
 mantain total number within 25
- possible change in voting rules: institutional vs individual membership current double quorum (3/4 members shall vote; 3/4 votes yes) a single NO delays taking into effect by 3 months e.g. quorum of 50% of all members and 3/4 of institutional members institutional members must vote institutional NO delays by 3 months







DRWG – possible SEGs – 8.a

- Curation SEGs are long-term. Max 24 members. Member rotation?
- Topical SEGs have a limited duration (specific charter). Require significant time effort and competence.
- DRWG Executive made by SEG chairs/vice-chairs? Incompatibility DRWG chair vs SEG chair? Ex-officio members or observers (B2 chair, VO) ?

For discussion ...



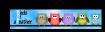




DRWG – possible SEGs – 8.b

- Curation SEGs are long-term. Time commitment limited (except occasional task forces). Max 24 members. Member rotation?
- Topical SEGs have a limited duration (specific charter). Require significant time effort and competence.
- DRWG Executive made by SEG chairs/vice-chairs? Incompatibility DRWG chair vs SEG chair? Ex-officio members or observers (B2 chair, VO) ?

For discussion ...







DRWG – possible SEGs – 8.c

- Curation SEGs are long-term. Time commitment limited (except occasional task forces). Max 24 members. Member rotation?
- Topical SEGs have a limited duration (specific charter). Require significant time effort and competence. Max 12 members.
- DRWG Executive made by SEG chairs/vice-chairs ? Incompatibility DRWG chair vs SEG chair? Ex-officio members or observers (B2 chair, VO) ?

For later discussion ...





DRWG – possible SEGs – 8.d

- SEG 0 : FITS SEG (curation)
- SEG n : curation SEG for astronomical usage of data format xxxx (proposals?)
- design, prototype, formulate proposal
- define an IAU standard?
- SEG 3 : Event Stream SEG formalize protocols, VO cooperation, define an IAU standard

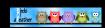
For discussion ...



DRWG – possible SEGs – 8.e

- SEG 0 : FITS SEG (curation)
- SEG n: curation SEG for astronomical usage of data format xxxx (proposals?)
- SEG 1 : Next Generation data format SEG design, prototype, formulate proposal
- SEG 2 : Data Provenance SEG define an IAU standard?
- SEG 3 : Event Stream SEG formalize protocols, VO cooperation, define an IAU standard

For discussion...



DRWG – practical issues – 9.a

Needed operational tools

- a DRWG website
- a DRWG mailing list
- SEG tools (each on its own)

- Find a permanent hosting institution? volunteers?
- Use (ad interim?) external tools (pbworks wiki?)

For discussion ...or contact me (lucio@lambrate.inaf.it)



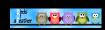
DRWG – practical issues – 9.b

Needed operational tools

- a DRWG website
- a DRWG mailing list
- SEG tools (each on its own) Current FITS WG tools (website @ HEASARC, lists FITSBITS and IAUFWG @ NRAO) will continue

- Find a permanent hosting institution? volunteers?
- Use (ad interim?) external tools (pbworks wiki?)

For discussion ...or contact me (lucio@lambrate.inaf.it)





DRWG – practical issues – 9.c

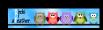
Needed operational tools

- a DRWG website
- a DRWG mailing list
- SEG tools (each on its own) Current FITS WG tools (website @ HEASARC, lists FITSBITS and IAUFWG @ NRAO) will continue

I have the capability to mantain an Apache website and a mailman mailing list, but may not have the time (and ... in the long run we will be all ... retired).

- Find a permanent hosting institution? volunteers?
- Use (ad interim?) external tools (pbworks wiki?)

For discussion ...or contact me (lucio@lambrate.inaf.it)





DRWG – practical issues – 9.d

Needed operational tools

- a DRWG website
- a DRWG mailing list
- SEG tools (each on its own) Current FITS WG tools (website @ HEASARC, lists FITSBITS and IAUFWG @ NRAO) will continue

I have the capability to mantain an Apache website and a mailman mailing list, but may not have the time (and ... in the long run we will be all ... retired).

- Find a permanent hosting institution? volunteers?
- Are there IAU-provided facilities?
- Use (ad interim?) external tools (pbworks wiki?)

For discussion ...or contact me (lucio@lambrate.inaf.it)



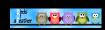




DRWG – final remarks – 10

I solicit contributions and open discussion on

- FITS related ideas
- Other formats related ideas
- DRWG member candidatures
- Proposals for DRWG SEGs
- Setting up the DRWG (tools)





... the word to you

SIC TRANSIT GLORIA MVNDI

...the word to you

SIC TRANSIT GLORIA

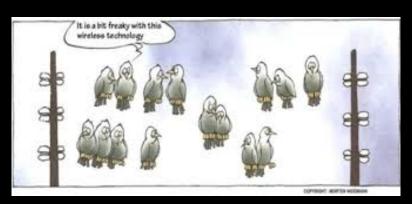
... the word to you

SIC TRANSIT

...the word to you

SIC

... the word to you



It is a bit freaky with this wireless technology

A voice from the past – 12.a

XAS format (circa 1996)

- not FITS but mappable to FITS
- except for smaller calibration files (plain ASCII)
- defined as working format as opposed to transport format
- X-ray astronomy data mode.
- two physical binary formats (BINTABLE was just being standardized at the time
- issue at the time: portability (VAX/Unix, big/little endian), efficiency
- binary format used native representation (for work and export)
- unidirectional utility to import and localize in place
- files used native record length (image or table row
- 28-byte miniheader (magic number, record length, number of data and header records
- 1 or exceptionally more minih records followed by data area then ...
- metadata header (actually trailer) at end
- header arranged in binary keywords with binary type, length byte, name, value
- kwd names limited to 8 char
- numeric (I J E D) kwds could be array-valued (potentially up to 255 bytes)
- string kwds limited to 68 bytes for FITS compatibility
- hence all other arrays limited to same max length
- mandatory kwds: BITPIX, NAXIS1 and NAXIS2, TFIELDS, TFORMn, TTYPEn

typeset with LATEX BEAMER

A voice from the past - 12.b

- not FITS but mappable to FITS
- except for smaller calibration files (plain ASCII)
- defined as working format as opposed to transport format
- X-ray astronomy data model: images, resp. matrices, spectra, time profiles, photon lists, gen. histo
- two physical binary formats (image and tabular) (BINTABLE was just being standardized at the time)
- issue at the time: portability (VAX/Unix, big/little endian), efficiency
- binary format used native representation (for work and export)
- unidirectional utility to import and localize in place
- files used native record length (image or table row
- 28-byte miniheader (magic number, record length, number of data and header records
- 1 or exceptionally more minih records followed by data area then ...
- metadata header (actually trailer) at end
- header arranged in binary keywords with binary type, length byte, name, value
- kwd names limited to 8 char
- numeric (I J E D) kwds could be array-valued (potentially up to 255 bytes)
 - string kwds limited to 68 bytes for FITS compatibility
- hence all other arrays limited to same max length
- mandatory kwds: BITPIX, NAXIS1 and NAXIS2, TFIELDS, TFORMn, TTYPEn
- ev. unnamed columns (record length padded to 4n bytes for Fortran direct access)

A voice from the past – 12.c

- not FITS but mappable to FITS
- except for smaller calibration files (plain ASCII)
- defined as working format as opposed to transport format
- X-ray astronomy data model: images, resp. matrices, spectra, time profiles, photon lists, gen. histo
- two physical binary formats (image and tabular) (BINTABLE was just being standardized at the time)
- issue at the time: portability (VAX/Unix, big/little endian), efficienc
- binary format used native representation (for work and export)
- unidirectional utility to import and localize in place
- files used native record length (image or table row
- 28-byte miniheader (magic number, record length, number of data and header records)
- 1 or exceptionally more minih records followed by data area then ...
- metadata header (actually trailer) at end
- header arranged in binary keywords with binary type, length byte, name, value
- kwd names limited to 8 char
- numeric (I J E D) kwds could be array-valued (potentially up to 255 bytes)
 - string kwds limited to 68 bytes for FITS compatibility
- hence all other arrays limited to same max length
- mandatory kwds: BITPIX, NAXIS1 and NAXIS2, TFIELDS, TFORMn, TTYPEn
- ev. unnamed columns (record length padded to 4n bytes for Fortran direct_access)

A voice from the past – 12.d

- not FITS but mappable to FITS
- except for smaller calibration files (plain ASCII)
- defined as working format as opposed to transport format
- X-ray astronomy data model: images, resp. matrices, spectra, time profiles, photon lists, gen. histo
- two physical binary formats (image and tabular) (BINTABLE was just being standardized at the time)
- issue at the time: portability (VAX/Unix, big/little endian), efficiency
- binary format used native representation (for work and export)
- unidirectional utility to import and localize in place
- files used native record length (image or table row)
- 28-byte miniheader (magic number, record length, number of data and header records)
- 1 or exceptionally more minih records followed by data area then . . .
- metadata header (actually trailer) at end
- header arranged in binary keywords with binary type, length byte, name, value
 - kwd names limited to 8 char
- numeric (I J E D) kwds could be array-valued (potentially up to 255 bytes)
 - string kwds limited to 68 bytes for FITS compatibility
- hence all other arrays limited to same max length
- mandatory kwds: BITPIX, NAXIS1 and NAXIS2, TFIELDS, TFORMn, TTYPEn
- ev. unnamed columns (record length padded to 4n bytes for Fortran direct access)

A voice from the past – 12.e

- not FITS but mappable to FITS
- except for smaller calibration files (plain ASCII)
- defined as working format as opposed to transport format
- X-ray astronomy data model: images, resp. matrices, spectra, time profiles, photon lists, gen. histo
- two physical binary formats (image and tabular) (BINTABLE was just being standardized at the time)
- issue at the time: portability (VAX/Unix, big/little endian), efficiency
- binary format used native representation (for work and export)
- unidirectional utility to import and localize in place
- files used native record length (image or table row)
- 28-byte miniheader (magic number, record length, number of data and header records)
- 1 or exceptionally more minih records followed by data area then . . .
- metadata header (actually trailer) at end
- header arranged in binary keywords with binary type, length byte, name, value
- kwd names limited to 8 char
- numeric (I J E D) kwds could be array-valued (potentially up to 255 bytes)
- string kwds limited to 68 bytes for FITS compatibility
- hence all other arrays limited to same max length
 - mandatory kwds: BITPIX, NAXIS1 and NAXIS2, TFIELDS, TFORMn, TTYPEn
 - ev. unnamed columns (record length padded to 4n bytes for Fortran direct access)

A voice from the past – 12.f

- not FITS but mappable to FITS
- except for smaller calibration files (plain ASCII)
- defined as working format as opposed to transport format
- X-ray astronomy data model: images, resp. matrices, spectra, time profiles, photon lists, gen. histo
- two physical binary formats (image and tabular) (BINTABLE was just being standardized at the time)
- issue at the time: portability (VAX/Unix, big/little endian), efficiency
- binary format used native representation (for work and export)
- unidirectional utility to import and localize in place
- files used native record length (image or table row)
- 28-byte miniheader (magic number, record length, number of data and header records)
- 1 or exceptionally more minih records followed by data area then
- metadata header (actually trailer) at end
- header arranged in binary keywords with binary type, length byte, name, value
- kwd names limited to 8 char
- numeric (I J E D) kwds could be array-valued (potentially up to 255 bytes)
- string kwds limited to 68 bytes for FITS compatibility
- hence all other arrays limited to same max length
 - mandatory kwds: BITPIX, NAXIS1 and NAXIS2, TFIELDS, TFORMn, TTYPEn
 - ev. unhamed columns (record length padded to 4h bytes for Fortian direct access)

A voice from the past -12.g

- not FITS but mappable to FITS
- except for smaller calibration files (plain ASCII)
- defined as working format as opposed to transport format
- X-ray astronomy data model: images, resp. matrices, spectra, time profiles, photon lists, gen. histo
- two physical binary formats (image and tabular) (BINTABLE was just being standardized at the time)
- issue at the time: portability (VAX/Unix, big/little endian), efficiency
- binary format used native representation (for work and export)
- unidirectional utility to import and localize in place
- files used native record length (image or table row)
- 28-byte miniheader (magic number, record length, number of data and header records)
- 1 or exceptionally more minih records followed by data area then ...
- metadata header (actually trailer) at end
- header arranged in binary keywords with binary type, length byte, name, value
- kwd names limited to 8 char
- numeric (I J E D) kwds could be array-valued (potentially up to 255 bytes)
 - string kwds limited to 68 bytes for FITS compatibility
- ٠ hence all other arrays limited to same max length
- ۰ mandatory kwds: BITPIX, NAXIS1 and NAXIS2, TFIELDS, TFORMn, TTYPEn
- ٠ ev. unnamed columns (record length padded to 4n bytes for Fortran direct access)