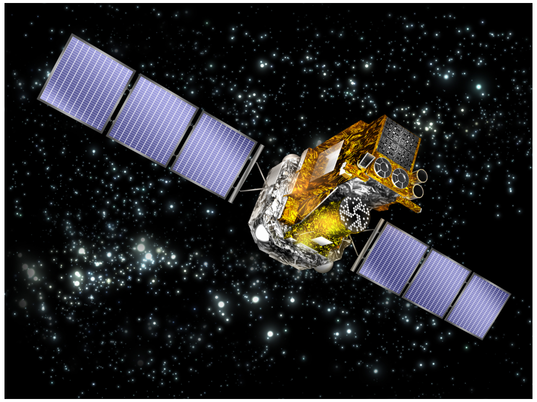
DL3 for high-energy telescopes

Experience from space-based missions

Jürgen Knödlseder

Topics covered

- FITS, HEASARC, OGIP and all that
- What others did (and what we can learn from them)
  - COMPTEL
  - INTEGRAL/SPI
  - Fermi/LAT


FITS : A flexible Image Transport System

- FITS was invented in the early 80ies for the interchange of astronomical images on magnetic tapes
  (Wells, Greisen, Harten, 1981, A&AS, 44, 363)

1. Introduction.- With the advent of the WSRT and the VLA in radio astronomy, the increased use of CCD arrays and other digital techniques in optical astronomy, and the development of satellites for astronomical observations at other frequencies, the number of images in digital form has increased enormously. In order to compare images made at different observatories and in order to continue the processing of the images, astronomers often wish to transport their data to their home sites or to other observatories. This interchange of data has traditionally been hampered by the fact that each installation has generated its own software system for image processing tailored to its own computer facilities, which differ enormously. Almost every installation has developed at least one unique data format and produced a large quantity of software based on the use of that internal format. Given this situation, the adoption of a single format for use in all installations would be prohibitively expensive and would lead, in general, to less efficient computing within the individual systems. However, a feasible course of action is the adoption of a unique interchange tape format to be used for transferring digital imagery between cooperating institutions. Each institution wishing to exchange imagery would then need to write only two programs - one to translate the transfer format into the internal format used by that institution and one to perform the reverse translation.

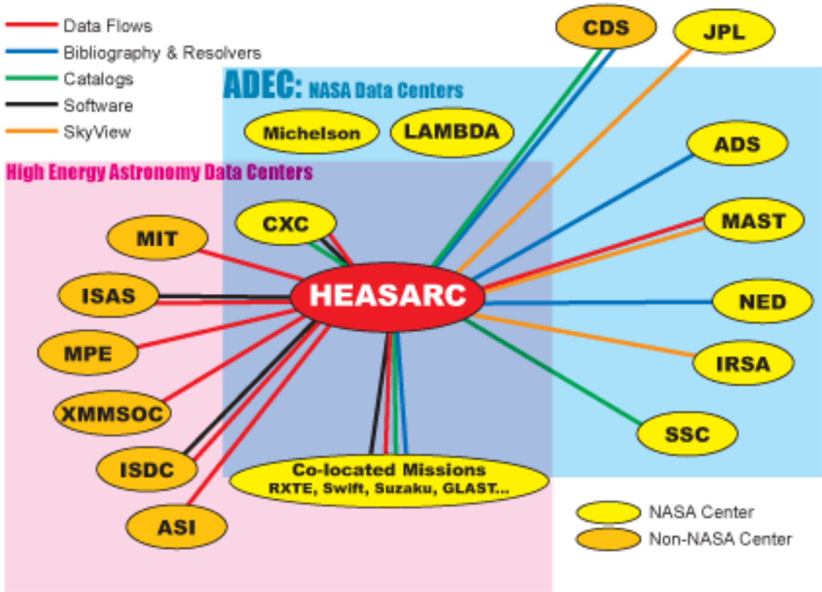
Evolutions

- 1988: ASCII table extension
- 1989: floating point numbers
- 1994: image & binary table extension
- 2002: World Coordinate Systems
- 2005: 64-bit integers
- 2008: standard v3.0

Developments overseen since 1988 by IAU FITS Working Group

HEASARC : NASA's archive for high-energy space observatories

- Established in 1990
- Motivation was to support of some imminent missions using a multi-mission approach to achieve cost savings by reusing software and archive resources
  - ROSAT (German, U.S. and U.K. X-ray satellite; 1990 – 1999)
  - CGRO (U.S. 17 tons gamma-ray satellite; 1991 – 2000)
  - ASCA (Japanese / U.S. X-ray satellite; 1993 – 2000)
  - RXTE (X-ray timing satellite; 1995 – 2010)
- Today provides data for 32 HE missions and 25 CMB experiments



HEASARC : (C)FITSIO and FTOOLS

- FITSIO
  - developed since 1992 as machine-independent Fortran-77 subroutine interface for reading and writing FITS files
  - rewritten in C in the late 90ies (cfitsio)
  - major contributions by INTEGRAL, XMM and Chandra data centres
  - current release 3.38 (2/2016)
- FTOOLS
  - developed since 1992 as a generic set of software utilities to manipulate FITS files
  - initial version for processing data of the ASCA satellite
  - current release 6.18 (2/2016) supports 13 missions

OGIP : HEASARC's FITS working group

http://heasarc.gsfc.nasa.gov/docs/heasarc/ofwg/ofwg_recomm.html

- **Event Lists Definition and File Format**. (also available in **postscript** and **latex** format).
  A list of recommended columns and keywords in an Event List has also been published in **Legacy 6, 36 (1995)**.

- **Spectral analysis** FITS file formats for Xspec-compatible PHA files. See also the **Addendum**.
- **Definition of RMF and ARF Spectral calibration file formats**. See also the **Addendum**.
- **XSPEC Table Models** File format. (also available in **postscript** and **latex** format).
- **WMAP FITS file format in spectral files**

- **Timing analysis data file formats** for High Energy Astrophysics Data.

- **Effective Areas FITS file format**
- **Radial PSF datasets FITS file format**
- **2D (image) PSF datasets FITS file format**
- **Vignetting Functions FITS file format**

- **Strings for specifying Physical Units within FITS files**
- **Standard Strings for Mission, Instrument, Filter, Detector & Grating Names for OGIP FITS files**
- **Standard strings for Detector Operating and Observing Modes**. (Also available in **postscript** and **latex** format).
- **Table Indexing** - A Proposal for creating indices on FITS tables (**Postscript**)

- **Calbration Database**: documentation on installation and use of the CALDB.

Many of these documents are relevant to the discussions at this meeting

OGIP : HEASARC's FITS working group

- Example: event list definition
  - multi-mission column names
    - http://heasarc.gsfc.nasa.gov/docs/heasarc/ofwg/docs/events/ogip_94_003/ogip_94_003.html
    - spatial: DETX, DETY, RA, DEC, L, B
    - spectral: ENERGY
    - temporal: TIME
  - recommended keywords
    - http://heasarc.gsfc.nasa.gov/docs/journal/fits6.html
    - General keywords (instrument configuration, observed object)
      - TELESCOPE / INSTRUME / DETNAM / FILTER
      - OBS_MODE = 'POINTING' / 'SLEW' / 'SCAN'
    - Time keywords
      - [DATE-OBS, DATE-END], [TIME-OBS, TIME-END] (covered time interval)
      - MJDREF (reference time)
      - DEADC = LIVETIME / ONTIME (average deadtime correction)
    - Spatial keywords
      - TCTYPx, TCRPXx, TCRVLx, TCDLTx (WCS for event-wise pixel to sky projection)
    - Spectral keywords
      - E_MIN, E_MAX, EUNIT (covered energy range)

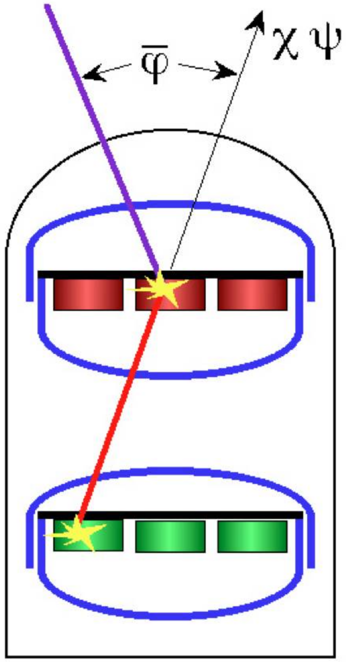
OGIP : HEASARC's FITS working group

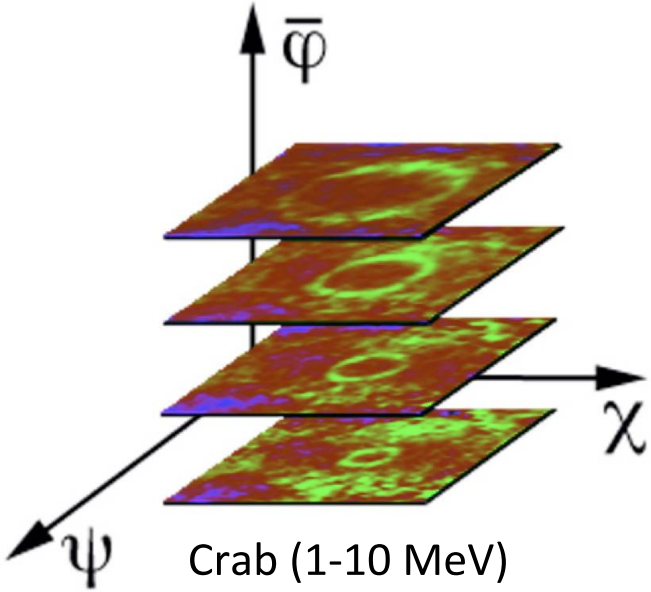
http://heasarc.gsfc.nasa.gov/docs/heasarc/ofwg/ofwg_recomm.html

The following 15 specific recommendations were formally approved by the **HEASARC FITS Working Group** from 1993 through 1995.

- R1: Recommends **using underscores rather than hyphens** in FITS keywords and column names (**text version**)
- R3: **RA and Dec keywords** for giving the position of the observed object, the telescope pointing, and the spacecraft axes (**text version**)
- R4: The **CATIDn FITS keywords** to contain source catalog identifications (**text version**)
- R5: **Specifying physical units in FITS files.** This recommendation is contained in OGIP memo ogip_93_001 (**postscript**, **latex**). A slightly earlier version of this memo was published in **Legacy** 4, 57 (1994).
- R6: On using the **TLMINn, TLMAXn, TDMINn and TDMAXn keywords** to record the the minimum and maximum values within columns of FITS tables (**text version**)
- R7: The **CREATOR keyword** to denote the name and version of the software program that generated that FITS file/extension (**text version**)
- R2/R8: The **HDUCLASn/HDUVERS keywords** to provide a hierarchial classification scheme for the various types of FITS extensions (**text version**). Also available is a **list of all HDUCLASn/HDUVERS values** currently defined by the HFWG.
- R9: **Quality flags** with a value of zero should be reserved to indicate GOOD (not rejected) quality data. Non-zero quality flag values may be used to indicate varying degrees of 'badness'. (**text version**)
- R10: Keywords to define **channel & energy boundaries** (**text version**)
- R11: Keywords and definitions relating to **exposure times** (**text version**)
- R12: Standard strings to denote the **mission, instrument and filters** within X- and gamma-ray astronomy. This recommendation is contained in OGIP memo ogip_93_013 (**postscript**, **latex**).
- R13: **Long string** keyword convention (**text version**)
- R14: **TSORTKEY** - a keyword for specifying the sort order of a FITS Table (**text version**)
- R15: Recommendations for **naming columns** in a FITS table (**text**, **latex**, **postscript**)

COMPTEL





Crab (1-10 MeV)

- Principle
  - measurement of 2 interaction locations defines apex χ,ψ of a cone
  - measurement of 2 energy deposits allows computation of Compton scattering angle φ

COMPTEL

- Context (pre-FITSIO era)
  - event files, telescope data, etc. in custom binary format
  - event and response cubes in FITS format (custom I/O library)
  - database driven F77-based analysis system (COMPASS; 4 sites, automatic synchronisation of database between all sites)
  - data converted to standard FITS format for legacy HEASARC archive at the end of the mission (not used by COMPASS), but lacks response functions (not really exploitable)
- Some lessons
  - conversion from custom binary to FITS format is possible
  - database driven analysis system is extremely constraining and blocks outside users

COMPTEL – response factorisation

$$R(\boldsymbol{p}', E', t' | \boldsymbol{p}, E, t) = \frac{DRX(\boldsymbol{p})}{T} \times DRG(\chi, \psi, \bar{\varphi}) \times IAQ(\chi, \psi, \bar{\varphi} | \boldsymbol{p}, E),$$

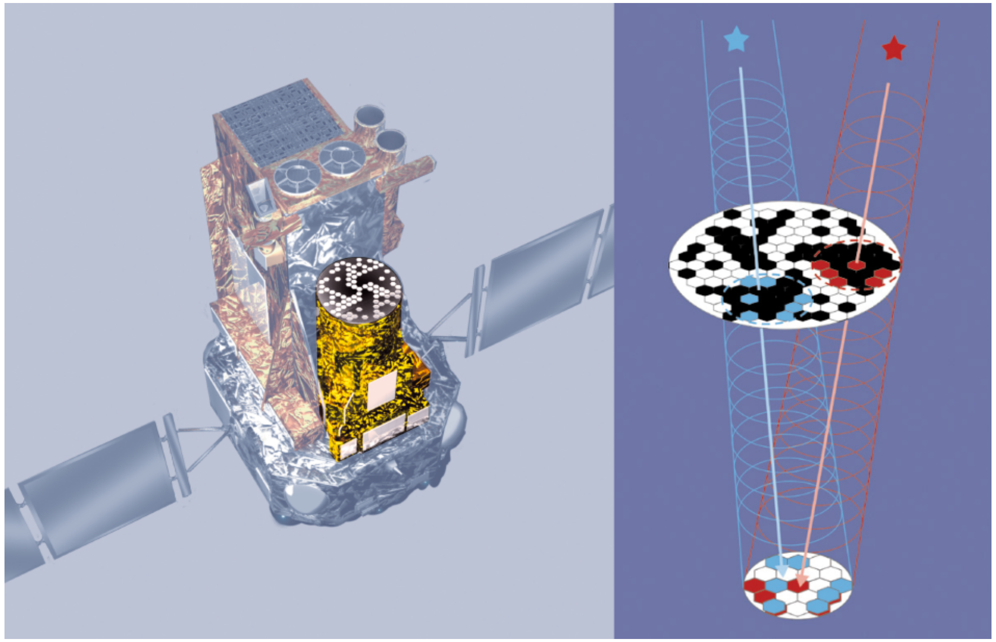
Exposure
- geometric D1 surface

Geometry effects
- D1 shadow on D2
- Earth horizon cuts

Compton kinematics
- Efficiency
- PSF
- Energy resolution

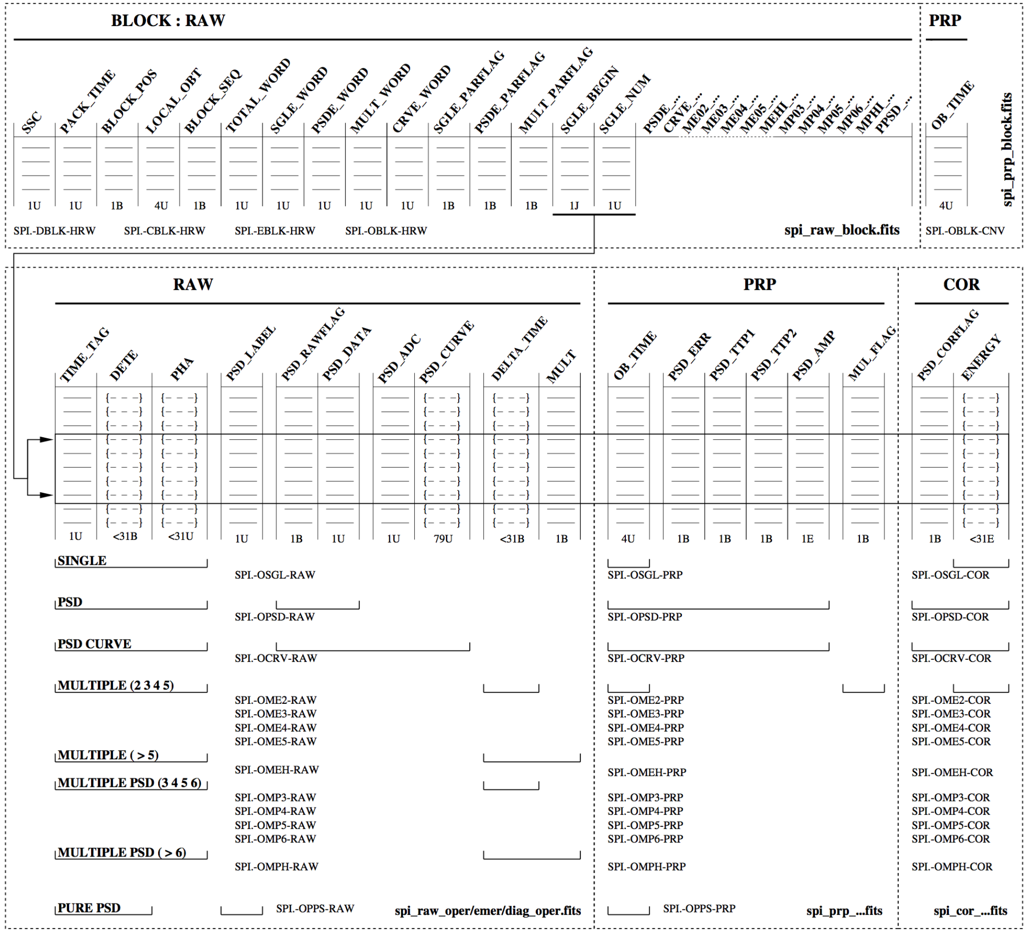
- Clever response factorisation …
  – simplifies data analysis (e.g. adding observations)
  – speeds up computations (DRX and DRG pre-computation, energy independent)

INTEGRAL/SPI



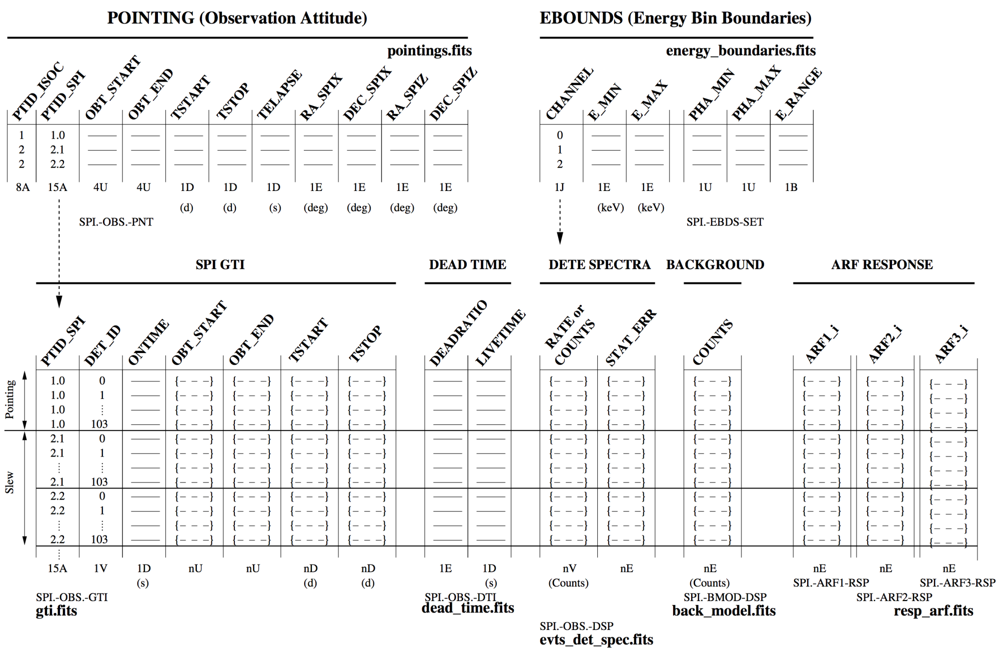
- Principle
  - sources lead to characteristic shadows on the detector
  - telescope dithering turns spatial information into count rate variations on individual detectors

INTEGRAL/SPI – event lists



- Context
  - Event data were initially split over several files (no-copy rule)
  - Later the event data format was changed to a single file
  - Created some headache for code developers
- Lessons learned
  - Better do it right in the first place
  - Be primarily concerned about the end user

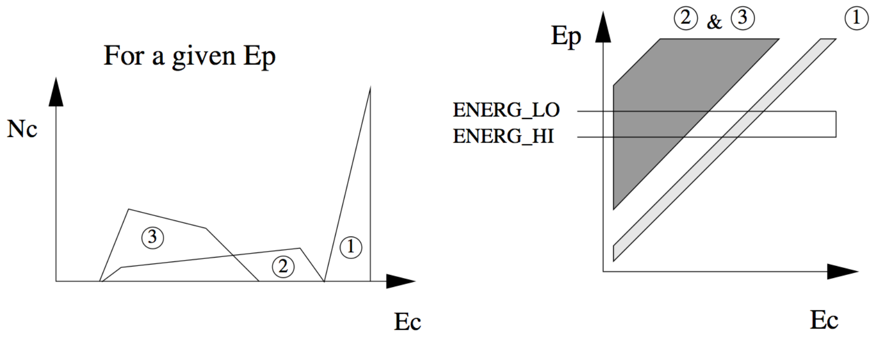
INTEGRAL/SPI – binned data



- Related software
  - data access library dal3spi little used by software
  - if you think about writing a data access library, ask first the software developers if they actually need one (otherwise you waste a lot of money)

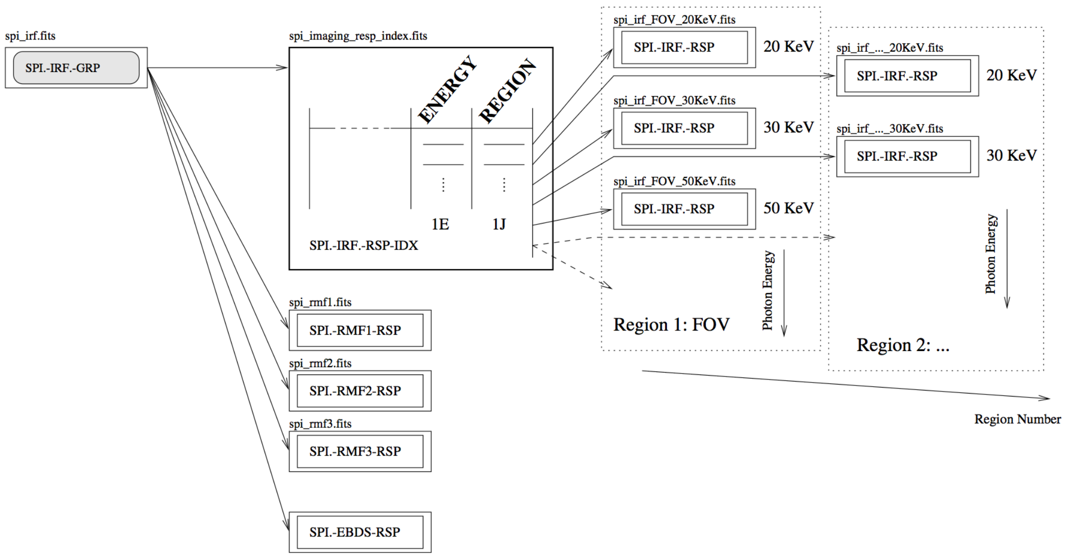
- Context
  - Data are binned per pointing and detector in a flat binary table
  - Table columns spread over 5+ files
  - File information stored in observation group (FITS database)
- Lessons learned
  - Ties data closely to software
  - Sometimes difficult to move data (filenames in observation group)
  - Sometimes needs manual editing of grouping tables

INTEGRAL/SPI – response data



(a) Illustration of the three kinds of RMFs for a given photon energy

(b) Schematic RMF matrix

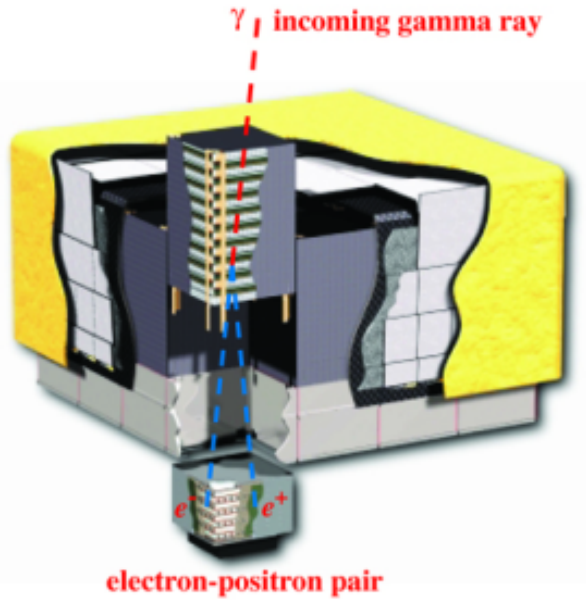


- Context
  - Energy response factorised into 3 components (photo peak, Compton scattering, backscattering from passive material)
  - Response file information stored in response group (FITS database)

Lessons learned

  - Think about clever factorisation of the response function (brute force may be prohibitive)

Fermi/LAT



- Principle
  - gamma rays pair convert in a tracker
  - resulting electron-positron pair is absorbed in a calorimeter

Fermi/LAT – event lists

	ENERGY	RA	DEC	L	B	THETA	PHI	ZENITH_ANGLE	EARTH_AZIMUTH_ANGLE
Select	E	E	E	E	E	E	E	E	E
All	MeV	deg	deg	deg	deg	deg	deg	deg	deg
Invert	Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify
1	4.943488E+02	8.080436E+01	2.707862E+01	1.788864E+02	-5.162598E+00	3.502382E+01	2.652596E+02	1.916358E+01	3.024701E+02
2	4.571115E+02	7.348530E+01	3.367534E+01	1.697908E+02	-6.341403E+00	3.500423E+01	2.491867E+02	2.820110E+01	3.098061E+02
3	6.667039E+02	8.358914E+01	9.268201E+00	1.955380E+02	-1.250010E+01	4.927800E+01	2.830118E+02	1.918320E+01	2.492156E+02
4	7.470766E+02	9.776036E+01	1.761573E+01	1.949593E+02	3.589495E+00	3.561794E+01	2.954099E+02	4.401269E+00	2.839326E+02
5	2.163904E+02	7.582805E+01	2.581853E+01	1.773278E+02	-9.526855E+00	4.205932E+01	2.605479E+02	2.869867E+01	2.952794E+02
6	5.391274E+02	7.852840E+01	1.237749E+01	1.901241E+02	-1.509467E+01	5.084543E+01	2.757419E+02	2.638160E+01	2.667675E+02
7	3.495143E+02	7.588198E+01	1.347919E+01	1.876893E+02	-1.663797E+01	5.161323E+01	2.722870E+02	2.881646E+01	2.700143E+02
8	3.725128E+02	7.716067E+01	2.606655E+01	1.778411E+02	-8.413099E+00	4.190646E+01	2.617491E+02	2.911594E+01	2.965721E+02
9	2.339170E+02	9.484727E+01	2.424977E+01	1.878038E+02	4.254071E+00	3.359674E+01	2.826212E+02	1.728199E+01	3.088091E+02
10	3.609297E+02	9.858973E+01	1.753086E+01	1.953976E+02	4.253620E+00	3.706775E+01	2.946416E+02	1.238954E+01	2.906511E+02
11	2.656009E+02	8.486671E+01	8.672123E+00	1.967170E+02	-1.171327E+01	5.259469E+01	2.875174E+02	2.676332E+01	2.632943E+02
12	7.181124E+02	8.688122E+01	2.104490E+01	1.869774E+02	-3.721150E+00	4.246061E+01	2.784423E+02	2.570693E+01	2.920928E+02
13	8.820988E+02	8.822605E+01	1.788360E+01	1.903517E+02	-4.240757E+00	4.393013E+01	2.831611E+02	2.409891E+01	2.854463E+02
14	3.238407E+03	7.911369E+01	3.161705E+01	1.743019E+02	-3.792160E+00	4.177920E+01	2.599167E+02	3.636702E+01	3.070055E+02

	START	STOP	SC_POSITION	LAT_GEO	LON_GEO	RAD_GEO	RA_ZENITH	DEC_ZENITH	B_MCILWAIN	L_MCILWAIN
Select	D	D	3E	E	E	D	E	E	E	E
All	s	s	m	deg	deg	m	deg	deg	Gauss	Earth_Radii
Invert	Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify
1	2.395574174942E+08	2.395574466000E+08	Plot	1.846488E+01	-9.255068E+01	5.427307500666E+05	9.693066E+01	1.834127E+01	3.142248E+00	1.430840E+00
2	2.395574466000E+08	2.395574766000E+08	Plot	1.788266E+01	-9.084282E+01	5.426378661531E+05	9.876013E+01	1.776295E+01	3.067991E+00	1.424369E+00
3	2.395574766000E+08	2.395575066000E+08	Plot	1.726479E+01	-8.909514E+01	5.425587716882E+05	1.006332E+02	1.714921E+01	2.987788E+00	1.417619E+00
4	2.395575066000E+08	2.395575366000E+08	Plot	1.662967E+01	-8.736010E+01	5.424964489131E+05	1.024935E+02	1.651834E+01	2.903018E+00	1.409828E+00
5	2.395575366000E+08	2.395575666000E+08	Plot	1.597808E+01	-8.563747E+01	5.424500008359E+05	1.043415E+02	1.587111E+01	2.812711E+00	1.401637E+00
6	2.395575666000E+08	2.395575966000E+08	Plot	1.531082E+01	-8.392696E+01	5.424235896018E+05	1.061774E+02	1.520832E+01	2.718733E+00	1.393232E+00
7	2.395575966000E+08	2.395576266000E+08	Plot	1.462868E+01	-8.222823E+01	5.424150562293E+05	1.080014E+02	1.453074E+01	2.609803E+00	1.384421E+00
8	2.395576266000E+08	2.395576566000E+08	Plot	1.393244E+01	-8.054092E+01	5.424293467878E+05	1.098141E+02	1.383917E+01	2.522435E+00	1.374749E+00
9	2.395576566000E+08	2.395576866000E+08	Plot	1.322290E+01	-7.886463E+01	5.424621573937E+05	1.116157E+02	1.313437E+01	2.421814E+00	1.364964E+00
10	2.395576866000E+08	2.395577166000E+08	Plot	1.250084E+01	-7.719890E+01	5.425162502960E+05	1.134068E+02	1.241715E+01	2.319188E+00	1.354737E+00
11	2.395577166000E+08	2.395577466000E+08	Plot	1.176703E+01	-7.554327E+01	5.425920998159E+05	1.151878E+02	1.168826E+01	2.218510E+00	1.343705E+00
12	2.395577466000E+08	2.395577766000E+08	Plot	1.102227E+01	-7.389725E+01	5.426918217307E+05	1.169591E+02	1.094848E+01	2.117492E+00	1.332434E+00
13	2.395577766000E+08	2.395578066000E+08	Plot	1.026733E+01	-7.226027E+01	5.428122159571E+05	1.187214E+02	1.019859E+01	2.017626E+00	1.320609E+00
14	2.395578066000E+08	2.395578366000E+08	Plot	9.502963E+00	-7.063181E+01	5.429545860599E+05	1.204753E+02	9.439344E+00	1.919699E+00	1.308093E+00
15	2.395578366000E+08	2.395578666000E+08	Plot	8.729950E+00	-6.901128E+01	5.431221747281E+05	1.222211E+02	8.671507E+00	1.823539E+00	1.294656E+00
16	2.395578666000E+08	2.395578966000E+08	Plot	7.949041E+00	-6.739810E+01	5.433103981009E+05	1.239596E+02	7.895826E+00	1.731785E+00	1.281078E+00
17	2.395578966000E+08	2.395579266000E+08	Plot	7.161003E+00	-6.579165E+01	5.435237741766E+05	1.256914E+02	7.113063E+00	1.643070E+00	1.266206E+00

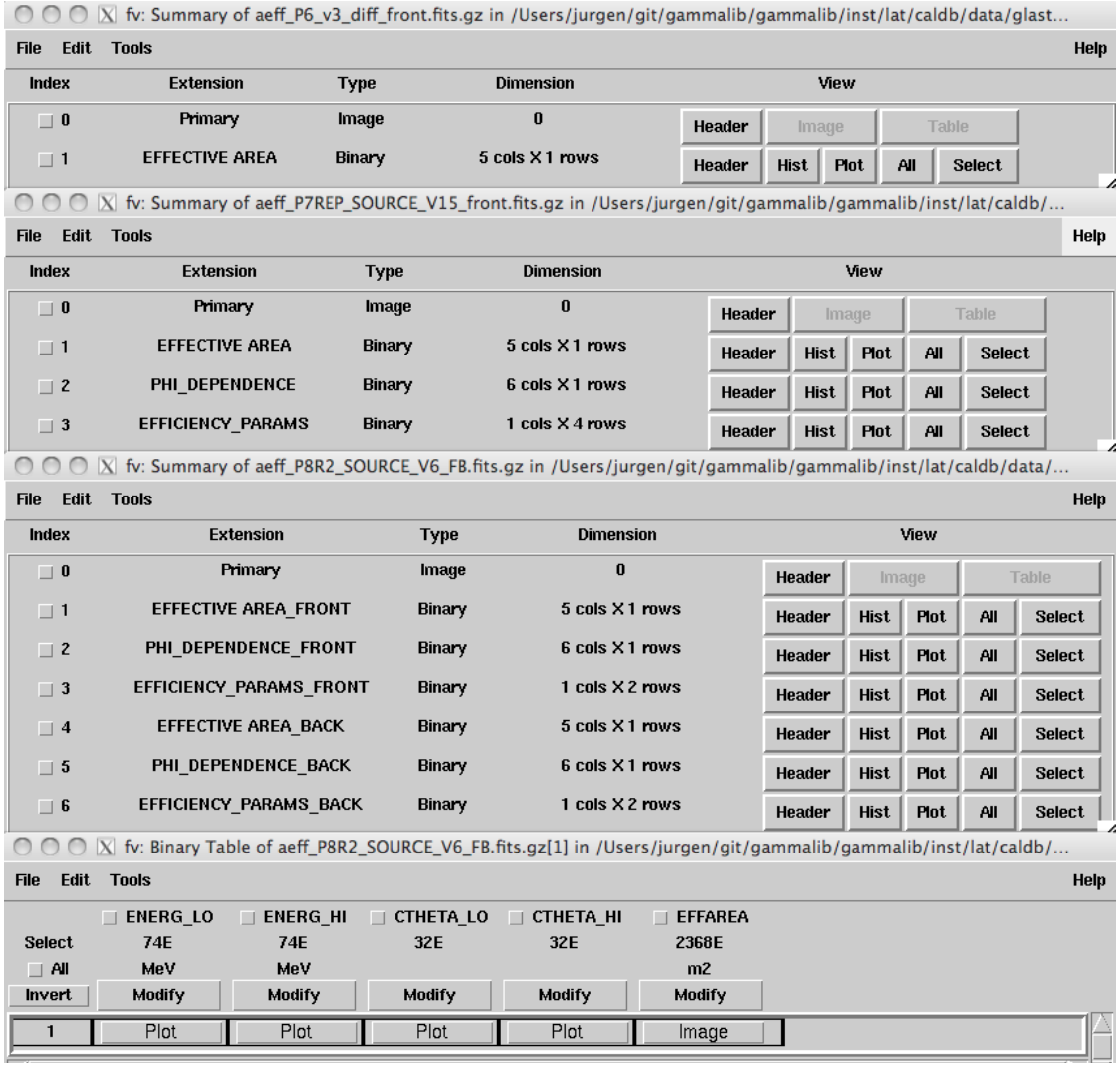
- Context
  - Relevant event data (ft1) and space craft data (ft2) in single files
  - Multiple files can be appended into a single large file
- Lessons learned
  - Simple files make file transport easy
  - Simple files allow for standalone data analysis (e.g. Fermi bubbles)

Fermi/LAT – response data

$$X^{\alpha}(\boldsymbol{p}, E) = f_1^{\alpha}(E) \int_{\theta} \tau(\boldsymbol{p}, \theta)\, A_{\text{eff}}^{\alpha}(E, \theta)\, \mathrm{d}\theta \; + f_2^{\alpha}(E) \int_{\theta} \tau_{\text{wgt}}(\boldsymbol{p}, \theta)\, A_{\text{eff}}^{\alpha}(E, \theta)\, \mathrm{d}\theta,$$

- Context
  - Time variability of response function encoded in a “livetime cube” τ (spacecraft scans the sky continuously)
  - Energy dependent detection efficiency variations due to ghost events
- Lessons learned
  - Factorisation speeds-up data analysis (“livetime cube” needs only to be computed once and is independent of energy)
  - Factorisation allows to store time independent response functions

Fermi/LAT – response data



- Context
  - Fermi/LAT response format evolved from pre-launch (Pass 6) to post-launch (Pass 7) and change in reconstruction algorithms (Pass 8)
- Lessons learned
  - Flexible response format is important to cope with unforeseen changes

Preliminary conclusions

- What others did well
  - Use FITS (standard, broad support, legacy)
  - Distribute only data that is needed for high-level analysis and make it compact (simplifies transport, eases usage)
  - Think about response factorisation (brute force is memory intensive and likely also slow)
  - Be flexible (you may need to adapt to real live (or user needs) once your observatory is online)
- What others did (in my opinion) not so well
  - Do not tie data too closely to software (limits data use, may create legacy problematic)
  - Avoid changing of data format after start of operations (software overhead)
  - Think twice before developing a data access library (may be waste of money)