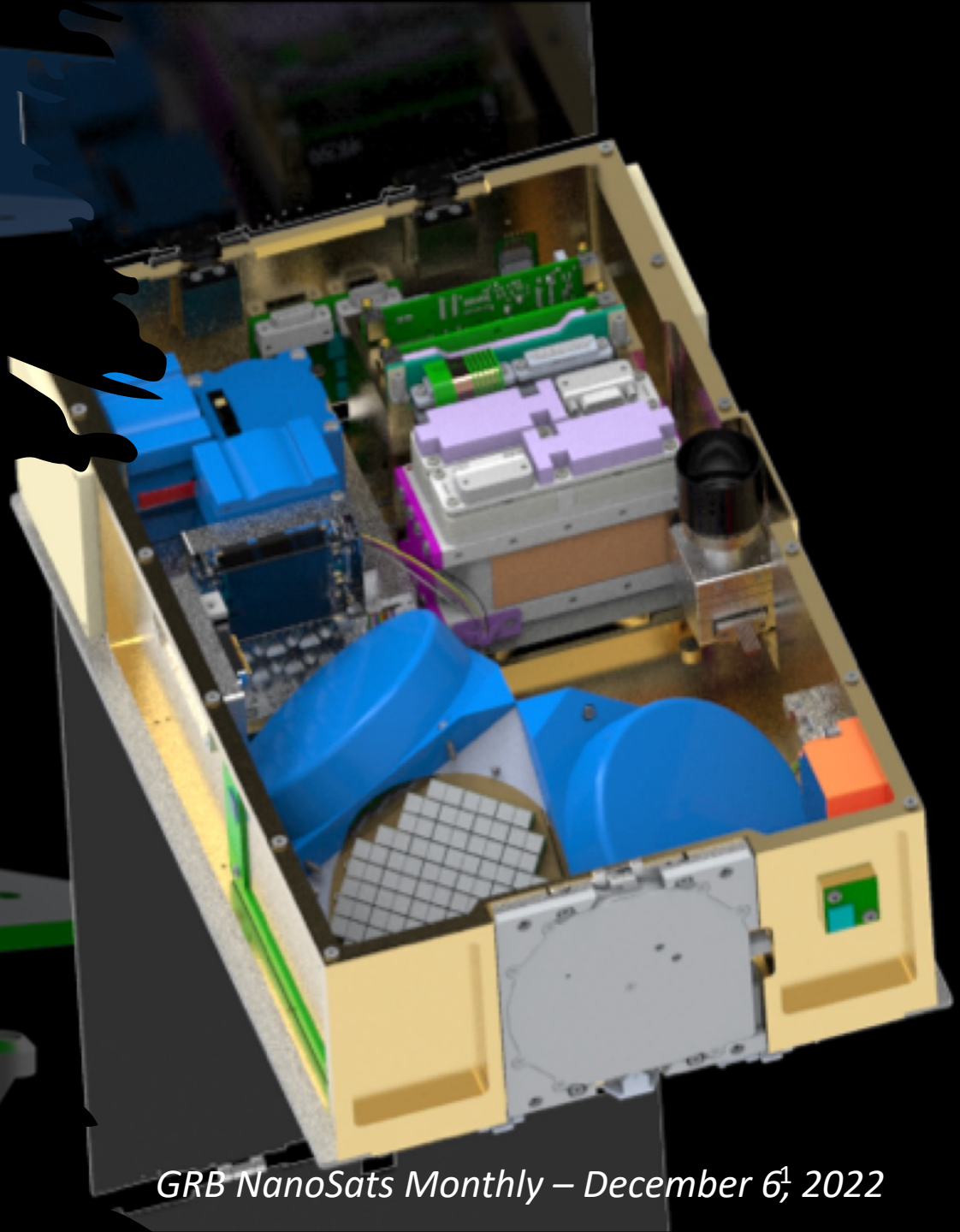
Data Formats and Archive

(BurstCube as an example for wide-field gamma-ray GRB NanoSats, and still in development)

Judy Racusin (NASA/GSFC)
on behalf of the
BurstCube Team



BurstCube Team

PI: Jeremy Perkins (NASA/GSFC)

Website: https://asd.gsfc.nasa.gov/burstcube/

Clemson
Dieter Hartmann

CRESST/GSFC
Israel Martinez

CUA/GSFC
Alessandro Bruno
Andrew Inglis
Teresa Tatoli

Delaware
Isabella Brewer

GWU
Sylvain Guiriec
Alyson Joens
Grant Mitchell
Pi Nuessle

ORAU/NPP
Ava Myers
Dan Violette

NASA/GSFC
Regina Caputo
Brad Cenko
Georgia de Nolfo
Carolyn Kierans
John Krizmanic
Julie McEnery
Judy Racusin
Lucia Tian

NASA/MSFC
Michelle Hui
Daniel Kocevski
Colleen Wilson-Hodge

LSU
Eric Burns

NRL
Lee Mitchell
Jacob R. Smith

UAH
Michael Briggs
Boyan Hristov

UCD
Lorraine Hanlon
Sheila McBreen
David Murphy
Alexey Uliyanov
Sarah Walsh

UMD
Peter Shawhan

USRA
Adam Goldstein
Oliver Roberts

Wisconsin
Sean Griffin

Action Item from Brno Panel Discussion

Action 11 - Data formats and sharing (where (HEASARC/Zenodo) - keep formats as common as possible

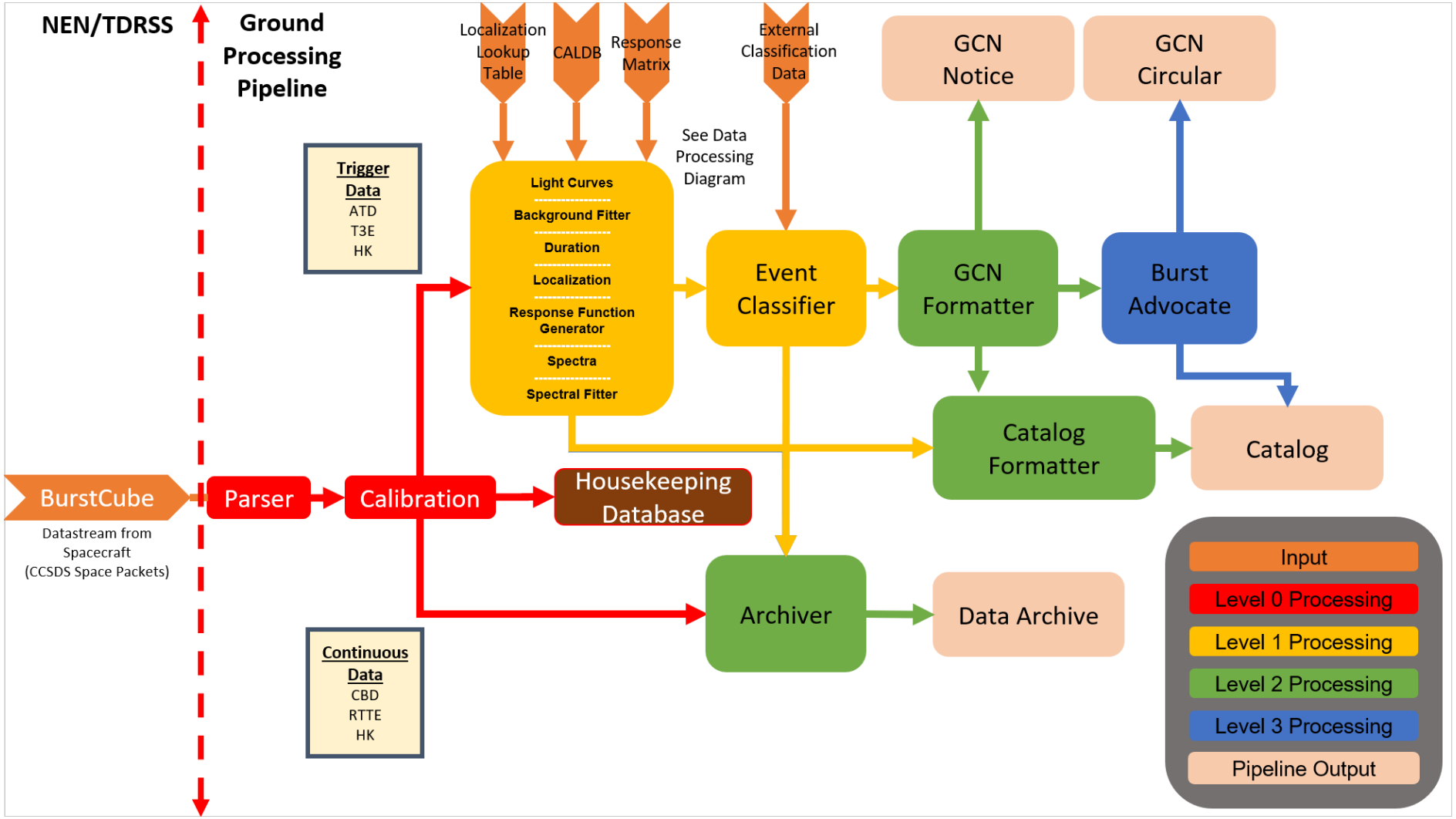
Why do we care about BurstCube-specific formats and archive?

- Maybe this will prevent your team from reinventing the wheel
- Uniformity in file formats will make it easier to reuse software and pipelines
- It'll be easier to combine datasets from different instruments for science analyses and joint localizations or sub-threshold searches

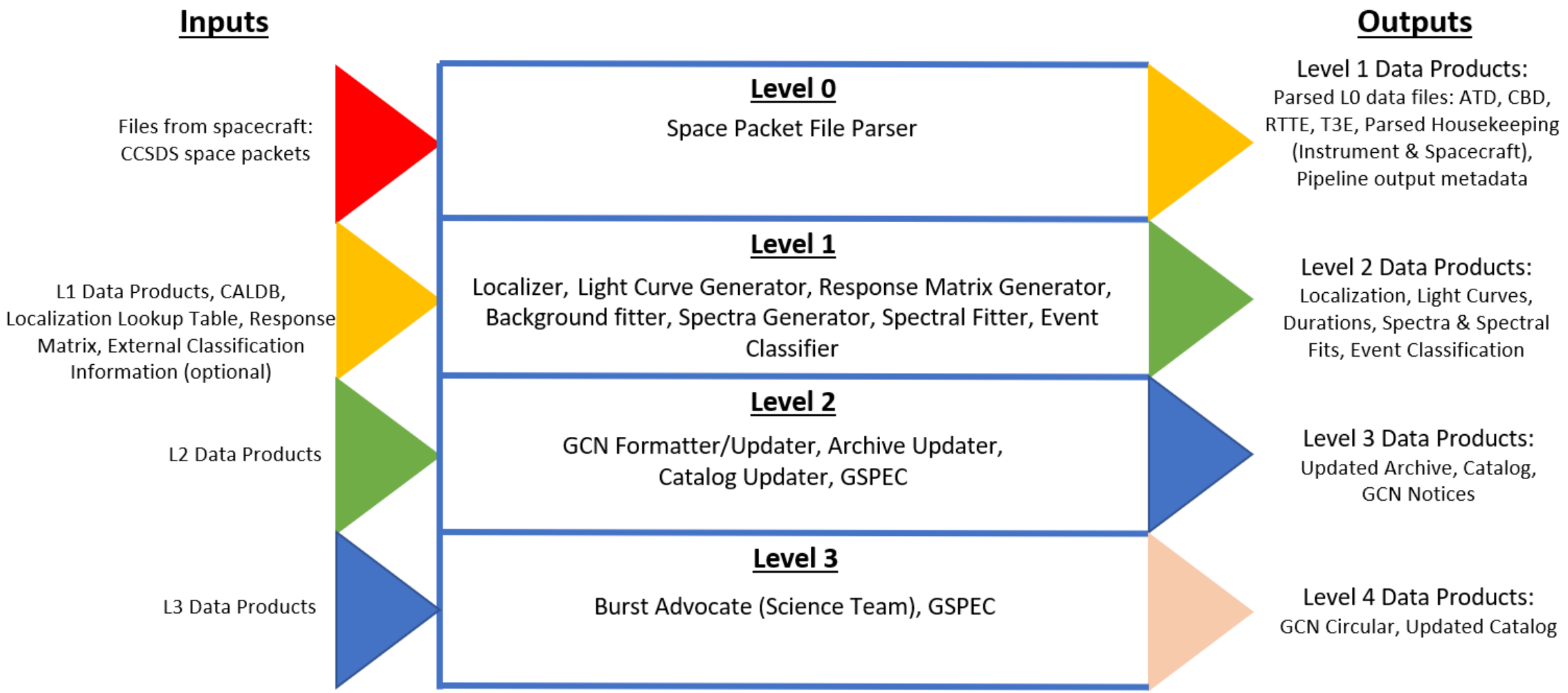
BurstCube-Specific Limitations/Restrictions

- HEASARC compliance (designed for pointed X-ray instruments)
  - FITS standards
  - CALDB
  - FTOOLs compatibility
  - https://heasarc.gsfc.nasa.gov/docs/heasarc/heasarc_req.html
- All data public immediately upon processing (except for most spacecraft housekeeping)
- Fast processing for generating GCN Notices
- Bandwidth
  - Direct to Earth (DTE): s-band, once daily pass, 2 Mbit/s
  - TDRSS (space network): s-band, as needed, 1 kbit/s

BurstCube Pipeline

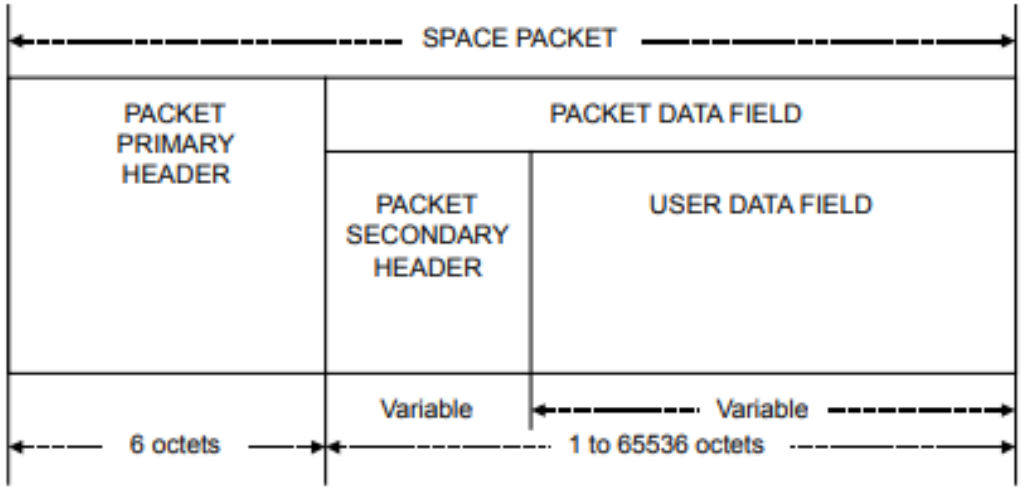


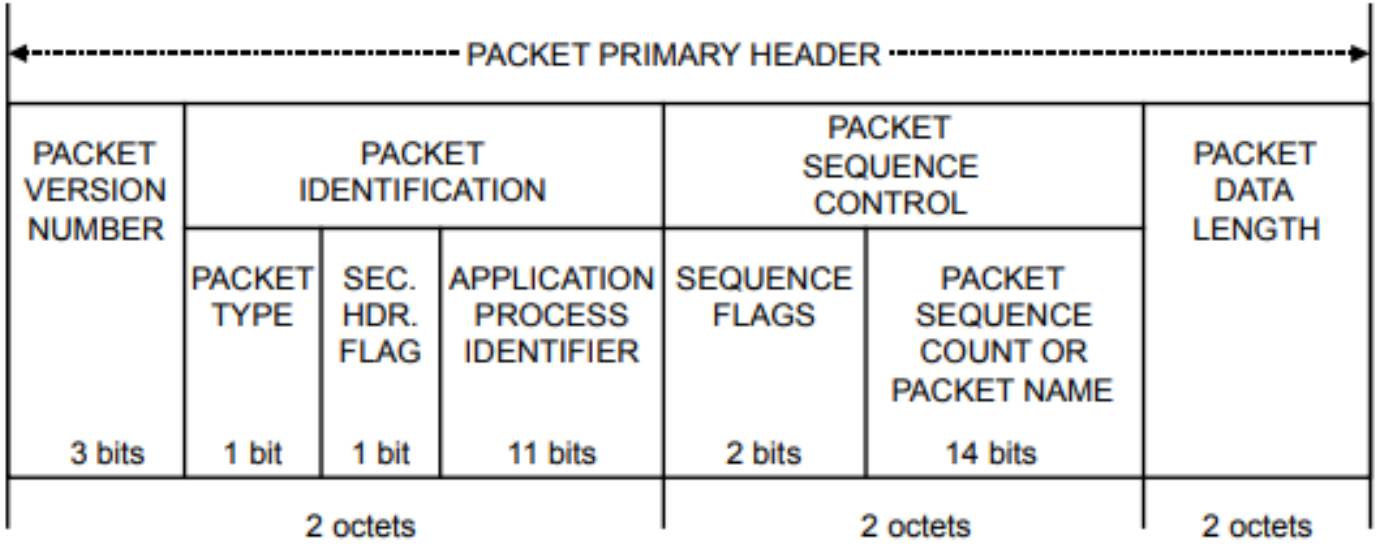
BurstCube Data Products



BurstCube Data Products (Level 0)

- Raw Space packets -> FITS
- Packet handling
  - packet sorting
  - gaps





https://public.ccsds.org/Pubs/133x0b2e1.pdf

BurstCube Data Products (Level 1)

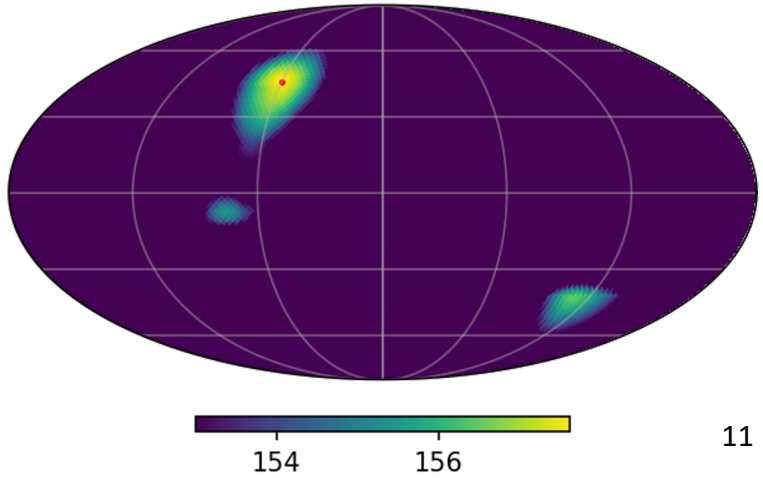
Datatype	Latency	Type	Energy Channels	Time Resolution	Time Coverage
Alert Trigger Data (ATD)	≲15 min	Trigger	16	50 ms to 2 s	-60 to +60 s
Triggered Time Tagged Event (T^3E)	≲1 day	Trigger	1024	10 µs	-30 to 100 s
Requested TTE (RTTE)	≲1 day	Requested	1024	10 µs	requested
Continuous Binned Data (CBD)	≲1 day	Continuous (excluding SAA)	16	256 ms	Continuous (excluding SAA)
Instrument Housekeeping (IHK)	≲1 day	housekeeping	N/A	1 s, 60 s	continuous
Orbit/Attitude Spacecraft Files	≲1 day	housekeeping	N/A	60 s	continuous

Requested Time Tagged Event Data (RTTE)

- Onboard buffer holds ~48 hours of TTE data
  - Passes insufficient to downlink all
  - Can downlink <1 hour/day
- Onboard file system setup such that we can downlink upon request
  - community request via a webform
  - requests generated via GCN Notices on detections by other instruments
- Prioritization based upon likelihood of subthreshold counterparts and astrophysical interestingness
- Window of data around time of interest will be downlinked at daily pass on a best effort basis
- Same format at T^3E

BurstCube Data Products (Level 2)

- Triggers Only
- Light Curves (FITS and images)
  - 1 file per detector
    - 50-100, 100-300, 300-1000 keV at 64 ms (TTE) and 1 s resolution (ATD + TTE)
  - 1 file with summed detectors
    - 50-100, 100-300, 300-1000 keV at 64 ms (TTE) and 1 s resolution (ATD + TTE)
- Spectra
  - source PHA2 files
  - response functions
  - background spectra
- Duration
  - Includes automated background estimation
  - Probability that a GRB is short or long
- Spectral Modeling
  - PL, CPL, Band, SBPL fits
- Localization
  - 2D image with WCS coordinates
  - Healpix probability map
  - image of localization map
- Trigger Classification
  - GCN Notice
  - Catalog



BurstCube Data Products (Level 3)

- GCN Notices
  - Standard parameters
    - trigger details (time, trig algorithm, timescale, etc.)
    - localization (RA/Dec circular approximation radius, HEALPIX map)
    - classification (GRB, solar flare, TGF, particles, galactic transient)
  - Automated analysis parameters
    - fluxes, fluences
    - spectral fits
- Trigger Catalog
  - summary of details in notices and more characterization (spectral fits, duration, etc.)
  - based upon Fermi-GBM trigger/GRB catalogs

BurstCube Data Products (Level 4)

- GCN Circulars
  - descriptions of triggers based upon data in notices
- Updated Catalog
  - From science team improved background and spectral fits

TITLE: GCN CIRCULAR
NUMBER: 40234
SUBJECT: GRB 231206A: BurstCube Detection
DATE: Judith Racusin at NASA/GSFC (judith.racusin@nasa.gov)

J. L. Racusin (NASA/GSFC) reports on behalf of the BurstCube Team:

At 22:09:04.72 UT on 6 December 2023, BurstCube triggered and located GRB 231206A (trigger 193492323/20231206923).

The on-ground calculated location, using the BurstCube trigger data is:

RA/Dec = 234.3, -47.3 (J2000)
with a circular equivalent statistical error of 10.4 deg (radius, 90% confidence).

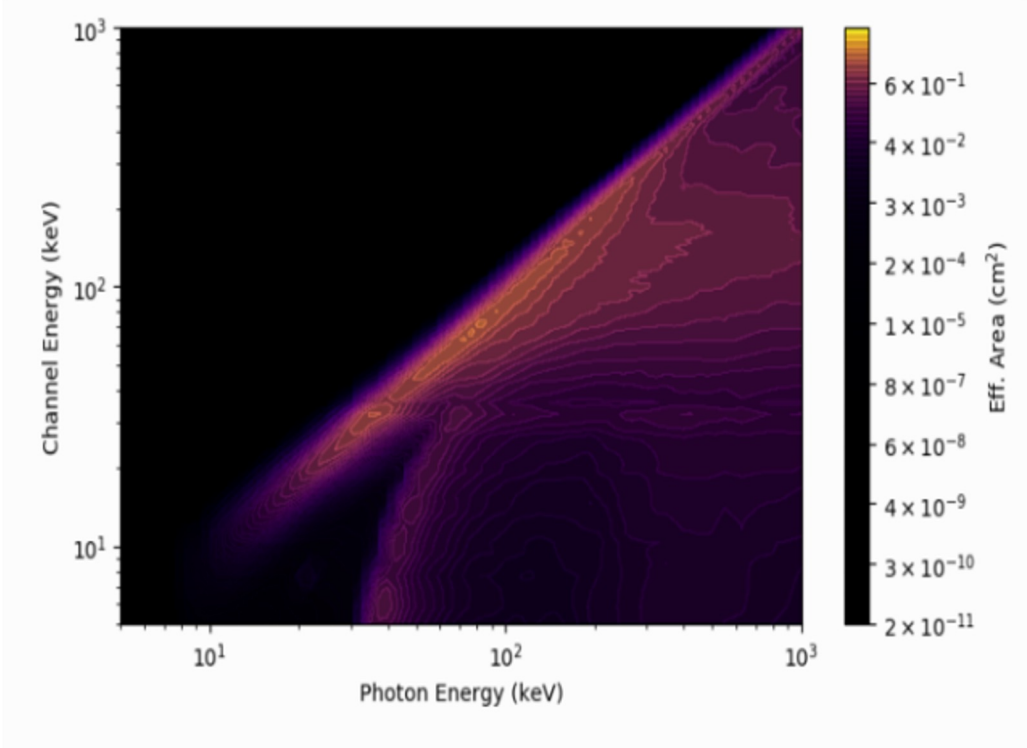
The skymap and HEALPix FITS file can be found here:
https://heasarc.gsfc.nasa.gov/FTP/burstcube/data/transient/2023/bn231206923/quicklook/bc231206923csa_atdloc.fits

The light curve can be found here:
https://heasarc.gsfc.nasa.gov/FTP/burstcube/data/transient/2023/bn231206923/quicklook/bc231206923csa_atdlc.fits

Totally made-up example based on GBM

Calibration Database (CALDB)

- Alignment file
  - describes detector orientations and relative coordinate systems
- SAA polygon definition
- Energy bounds
  - 16 energy channels (ATD, CBD)
  - 1024 energy channels (TTE)
- Full set of Response functions
  - 3072 per detector (NSIDE=16)
- Not actually used in analyses, but for reference



See also talks by Israel Martinez

/burstcube

/csa

/ bcf /…

/cpf /…

/index file1 file2

BurstCube Archive Structure

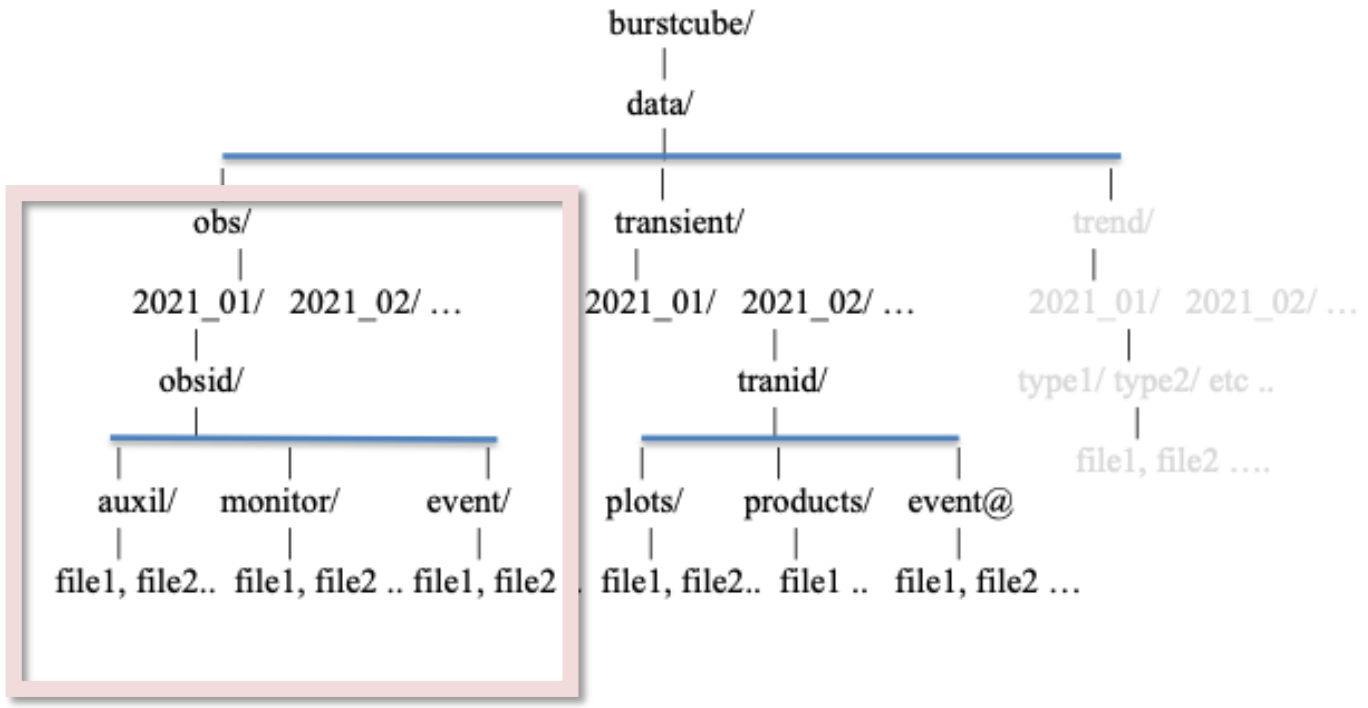
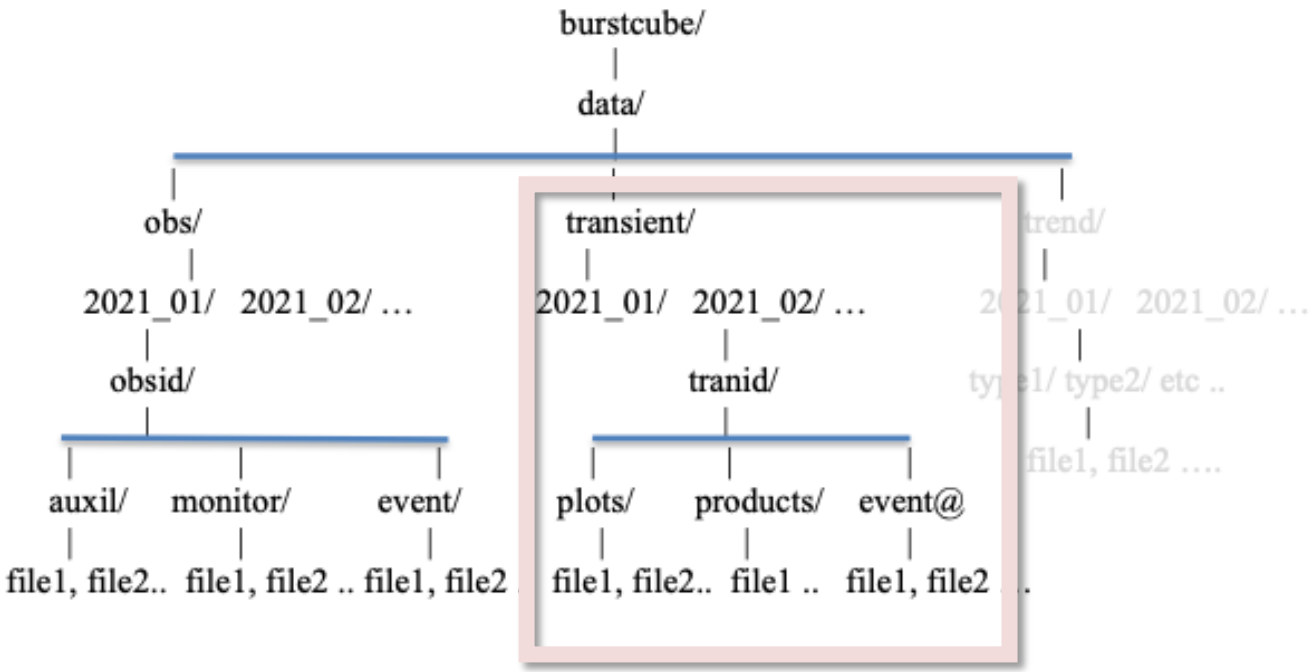


Table 3.4	
a)1st data set : daily observation data	
obs/	
Auxiliary filename	**Description**
auxil/	
bcYYMMDDcsa.hk	HK for the parameters of the detectors
bcYYMMDD.hk	HK for spacecraft including attitude and orbit
bcYYMMDD.cat	Catalog file
bcYYMMDD_joblog.html	Log file of the processing
Science Data	**Description**
monitor/	
bcYYMMDDcs0_Xcbd.fits	Binned 16 chan spectra NNN s detector CS0
bcYYMMDDcs1_Xcbd.fits	Binned 16 chan spectra NNN s detector CS1
bcYYMMDDcs2_Xcbd.fits	Binned 16 chan spectra NNN s detector CS2
bcYYMMDDcs3_Xcbd.fits	Binned 16 chan spectra NNN s detector CS3
	*NNN is one of the selectable integration time
events/	
bcYYMMDDcs0_ atdYYMMDDXXX.fits	Alert trigger cs0 YYMMDDXXX
bcYYMMDDcs1_ atdYYMMDDXXX.fits	Alert trigger cs1 YYMMDDXXX
bcYYMMDDcs2_ atdYYMMDDXXX.fits	Alert trigger cs2 YYMMDDXXX
bcYYMMDDcs3_ atdYYMMDDXXX.fits	Alert trigger cs3 tYYMMDDXXX
bcYYMMDDcs0_ tteYYMMDDXXX_uf.evt	T3E/RTTE det cs0 trig YYMMDDXXX
bcYYMMDDcs1_ tteYYMMDDXXX_uf.evt	T3E/RTTE det cs1 trig YYMMDDXXX
bcYYMMDDcs2_ tteYYMMDDXXX_uf.evt	T3E/RTTE det cs2 trig YYMMDDXXX
bcYYMMDDcs3_tteYYMMDDXXX_uf.evt	T3E/RTTE det cs3 trig YYMMDDXXX

BurstCube Archive Structure



b)2nd data set : transient data	
transient/	
Science Data	**Description**
products/	
bcYYMMDDXXXcs0_ttelc	Lighcurves several extensions per cs0
bcYYMMDDXXXcs1_tte.lc	Lighcurves several extensions per cs1
bcYYMMDDXXXcs2_tte.lc	Lighcurves several extensions per cs2
bcYYMMDDXXXcs3_tte.lc	Lighcurves several extensions per cs3
bcYYMMDDXXXcsa_tte.lc	Lighcurves several extensions per all csa
bcYYMMDDXXXcs0_tte.pha	PHA2 Spectra per cs0
bcYYMMDDXXXcs1_tte.pha	PHA2 Spectra per cs1
bcYYMMDDXXXcs2_tte.pha	PHA2 Spectra per cs2
bcYYMMDDXXXcs3_tte.pha	PHA2 spectra per cs3
bcYYMMDDXXXcs0_bg.pha	Background spectrum for cs0
bcYYMMDDXXXcs1_bg.pha	Background spectrum for cs1
bcYYMMDDXXXcs2_bg.pha	Background spectrum for cs2
bcYYMMDDXXXcs3_bg.pha	Background spectrum for cs3
bcYYMMDDXXXcs0.rsp	Response per cs0
bcYYMMDDXXXcs1.rsp	Response per cs1
bcYYMMDDXXXcs2.rsp	Response per cs2
bcYYMMDDXXXcs3.rsp	Response per cs3
bcYYMMDDXXXcsa.fits	Localization for the trigger YYMMDDXXX: 2d image primary plus keywords for trigger information, healpix representation in 1st extension
bcYYMMDDXXX.cat	Catalog file
bcYYMMDDXXX_joblog.html	Log file of the processing
plot/	
bcYYMMDDXXXcs0_atdlc.gif [png]	Plot lightcurve cs0
bcYYMMDDXXXcs1_atdlc. gif [png]	Plot lightcurve cs1
bcYYMMDDXXXcs2_atdlc. gif [png]	Plot lightcurve cs2
bcYYMMDDXXXcs3_atdlc. gif [png]	Plot lightcurve cs3
bcYYMMDDXXXcs0_atdpha. gif [png]	Plot spectra cs0
bcYYMMDDXXXcs1_atdpha. gif [png]	Plot spectra cs1
bcYYMMDDXXXcs2_atdpha. gif [png]	Plot spectra cs2
bcYYMMDDXXXcs3_atdpha. gif [png]	Plot spectra cs3
bcYYMMDDXXXcs0_ttelc.gif [png]	Plot lightcurve cs0
bcYYMMDDXXXcs1_ttelc. gif [png]	Plot lightcurve cs1
bcYYMMDDXXXcs2_ttelc. gif [png]	Plot lightcurve cs2
bcYYMMDDXXXcs3_ttelc. gif [png]	Plot lightcurve cs3
bcYYMMDDXXXcs0_ttepha. gif [png]	Plot spectra cs0
bcYYMMDDXXXcs1_ttepha. gif [png]	Plot spectra cs1
bcYYMMDDXXXcs2_ttepha. gif [png]	Plot spectra cs2
bcYYMMDDXXXcs3_ttepha. gif [png]	Plot spectra cs3
bcYYMMDDXXXcsa. gif [png]	Plot image of the localization

What are practical archive solutions for other missions?

- HEASARC does host some non-NASA missions
  - Significant cost/effort in implementation
  - Community familiarity with interfaces
  - Easy access to data products and tools
- Common formats make that easier
  - compatibility with analysis tools even better
- Public data has lots of benefits, but requires community support, which is difficult for small missions
- Public alerts for transient missions are even more important
- Long-term archive adds to scientific output of missions

Data Analysis Tools

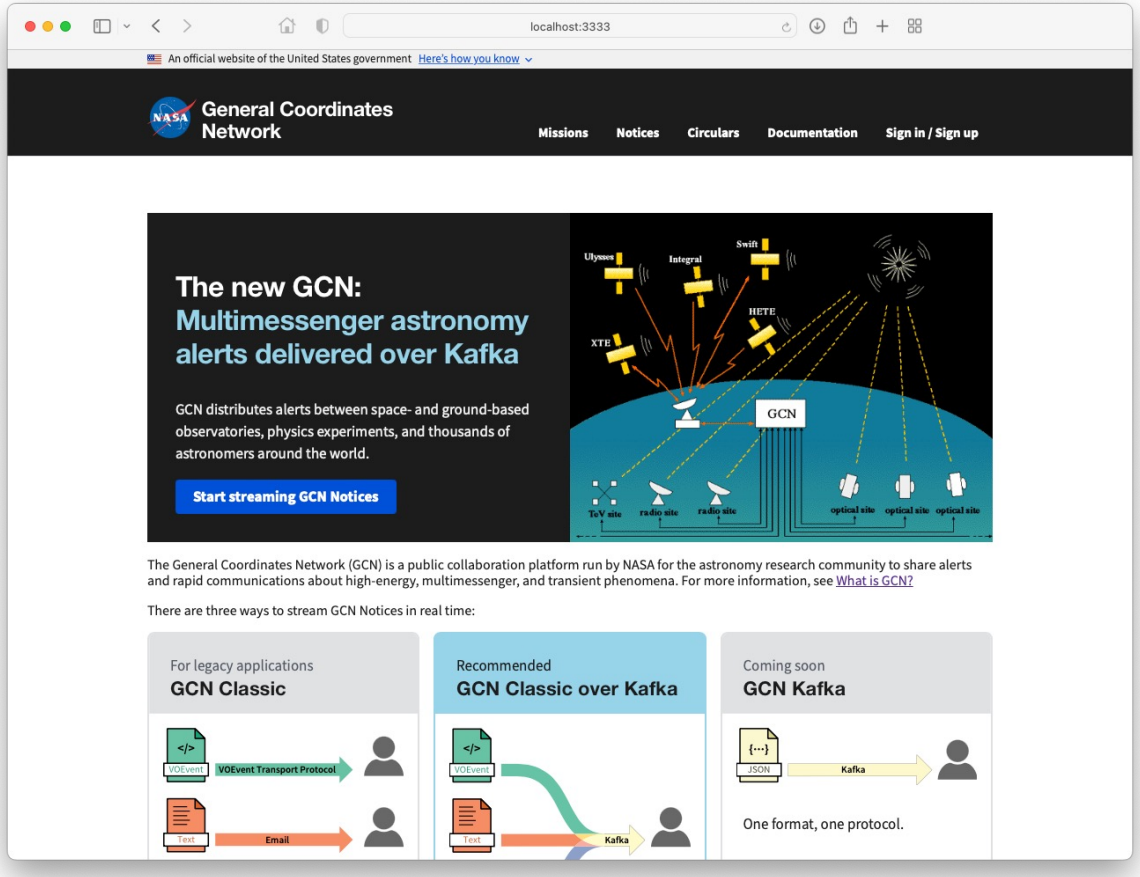
- GBM Data Tools
  - Fermi-GBM python analysis package
    - https://fermi.gsfc.nasa.gov/ssc/data/analysis/gbm/gbm_data_tools/gdt-docs/index.html
- bc-tools
  - BurstCube specific tools built upon GBM Data Tools
  - https://gitlab.com/burstcube/bc-tools/
  - Towards an effort to make detector-agnostic programming interface for count-based gamma-ray instruments

GCN Modernization on behalf of GCN Team:

- General Coordinates Network (https://gcn.nasa.gov)
- Phasing out legacy GCN system (GCN Classic) over next couple of years

Why switch to the new GCN?

	GCN Classic	GCN Classic over Kafka
Self-service	**NO.** Users need to contact administrator in order to make account and subscription changes	**YES.** Manage your own account and subscription settings through the web site
Open standards	**NO.** Notices are sent using three custom protocols	**YES.** Notices are sent using one standard protocol, Apache Kafka
Open source	**NO.** Custom software needed to receive notices	**YES.** Receive notices using open-source software
Highly available	**NO.** Notices are broadcast by a single server	**YES.** Notices are broadcast by a cluster of highly-available Kafka brokers in the cloud
Secure	**NO.** Notices are sent as plaintext	**YES.** Notices are protected with SSL/TLS



New GCN web portal

GCN Notices

- Producer Instruments
  - All new notice types will go thru the new GCN not legacy system (GCN Classic)
  - Unifying schema between instruments (as much as possible) - available soon
- Consumers – Switch now!
  - Self-managed GCN Notice subscriptions via email
  - Self-managed GCN Notices over Kafka (replaces socket & VOEvent transport protocol)
  - Still available in all 3 legacy formats (text, 160-byte binary, VOEvent)
- More Details
  - https://nasa-gcn.github.io/gcn-presentation/

GCN Circulars (Coming Soon)

- Modernization of GCN Circulars
  - Self-service subscription management
  - Self-service submitter registration using peer endorsement system
    - automatic onboarding of existing subscribers and submitters
  - Webform submission with Astro-flavored Markdown (email submission still accepted)
    - Correct handling of Unicode
    - Contextual parsing of links
- Migrate archive
- Smooth transition of subscribers
- Future enhancements
  - Automatically minted DOIs and BiBTeX entries
  - Link ORCIDs to GCN Circulars
  - Filtering by source type
  - Erratum