

# Data Model View

J.L. Contreras  
UCM-GAE Madrid

# Outline

- Introduction
- Low Level Data
- Proposal

# Introduction

We are trying to document things  
Look for more information at redmine:

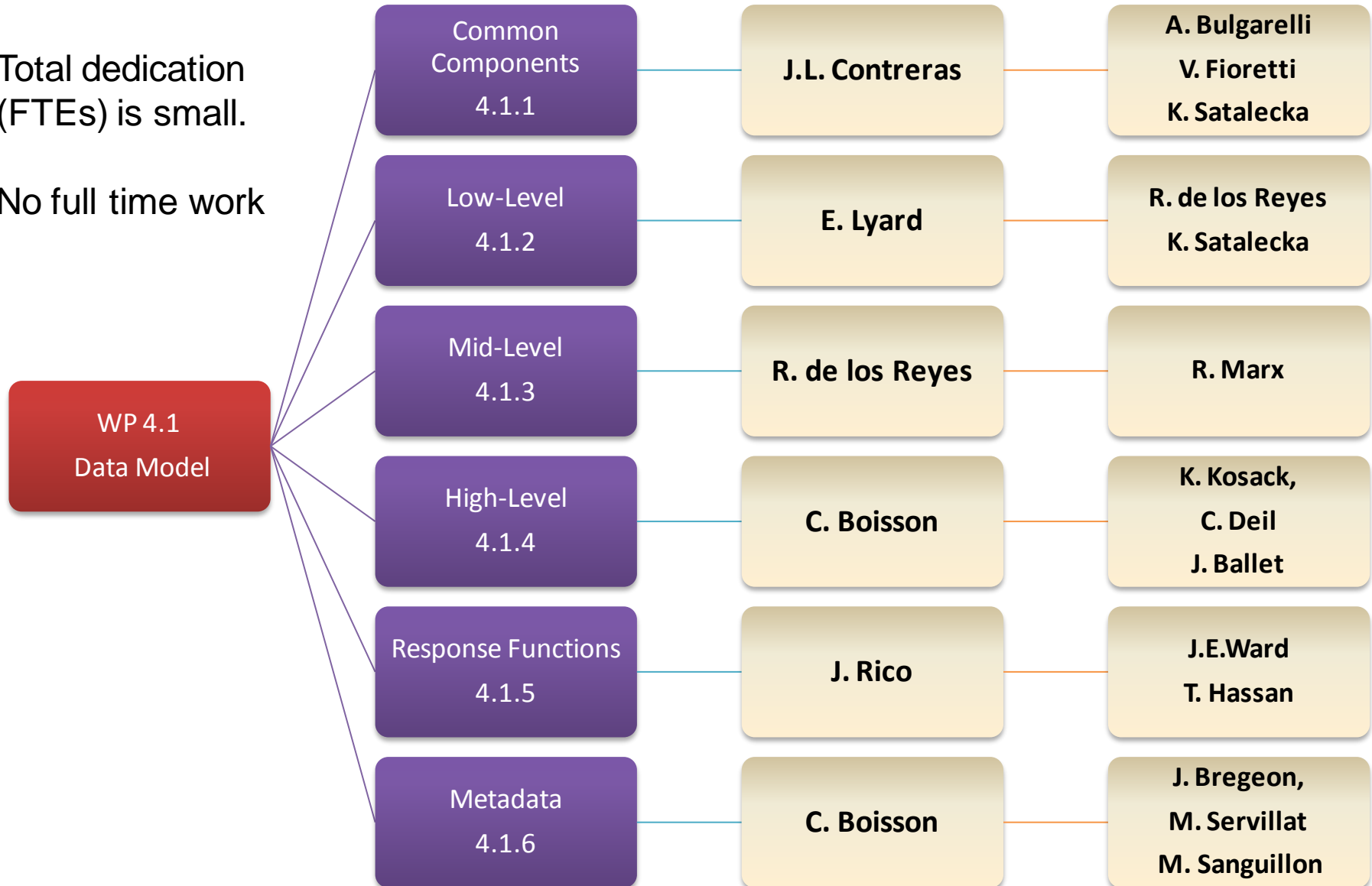
<https://forge.in2p3.fr/projects/model/wiki/DL0>

# Data Model PBS

Where are people working

Total dedication  
(FTEs) is small.

No full time work



# DL0 (Remember: “First data to be stored permanently”)

- **Crucial for:**

1. Long term development of CTA  
Save time, space, last for 30 year
2. Short term work  
Needed for prototypes of cameras, pipelines, data distribution...

→ **Pressure to have a definitive model soon**

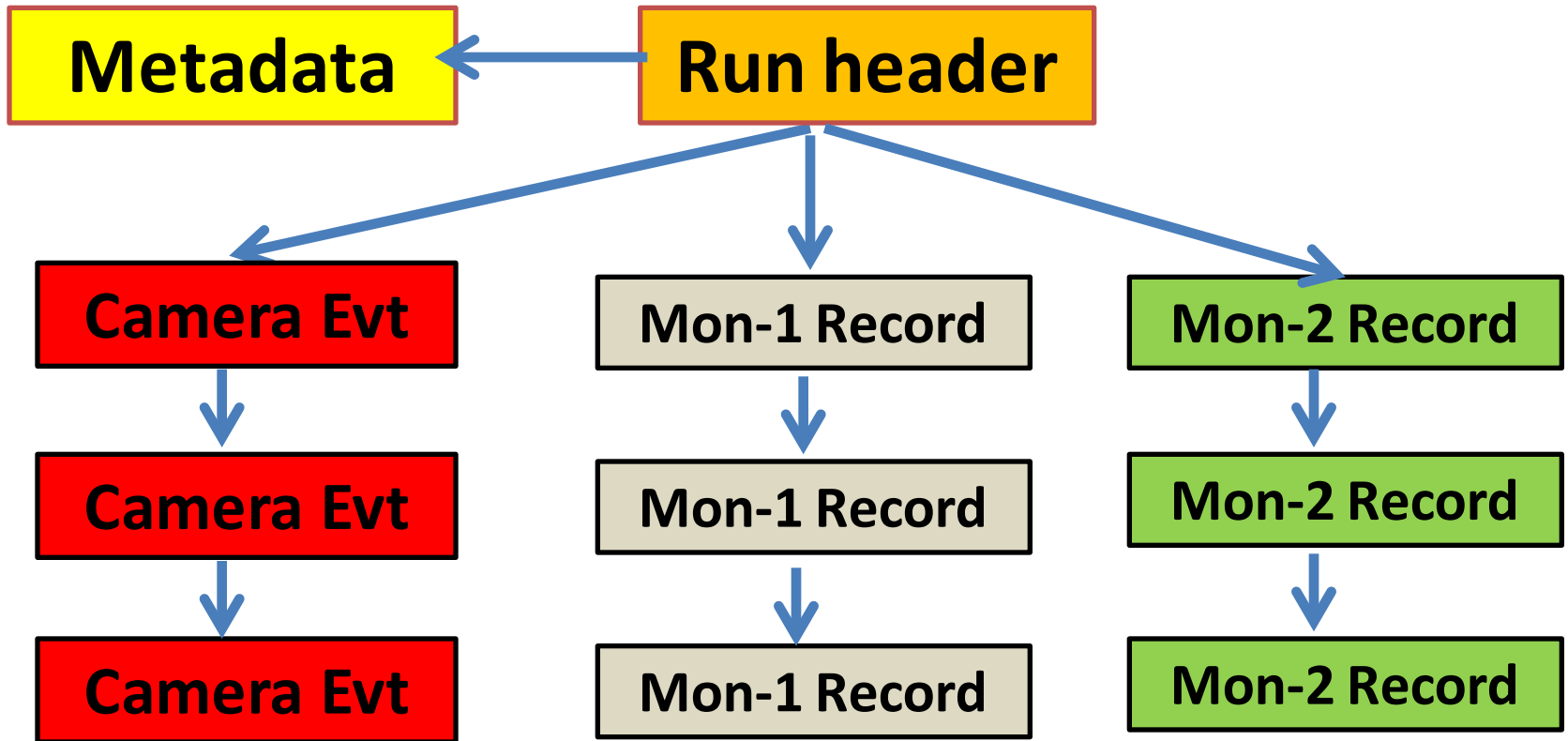
- **Difficult:**

1. Being the front line with hardware has to absorb camera differences
2. We still do not know many details (e.g: calibration, data reduction)
3. The astronomical community has many doubts  
(Astronomy & Computing special issue “The future of Astronomical Data Formats” in prep )

- **Different Levels:**

1. Logical format
2. Content
3. File format

# EVT0/CAL0 Logical Format



**Single Telescope Files: Merging at DL1:**  
**Motivated by a use case: data flow in Calibration**

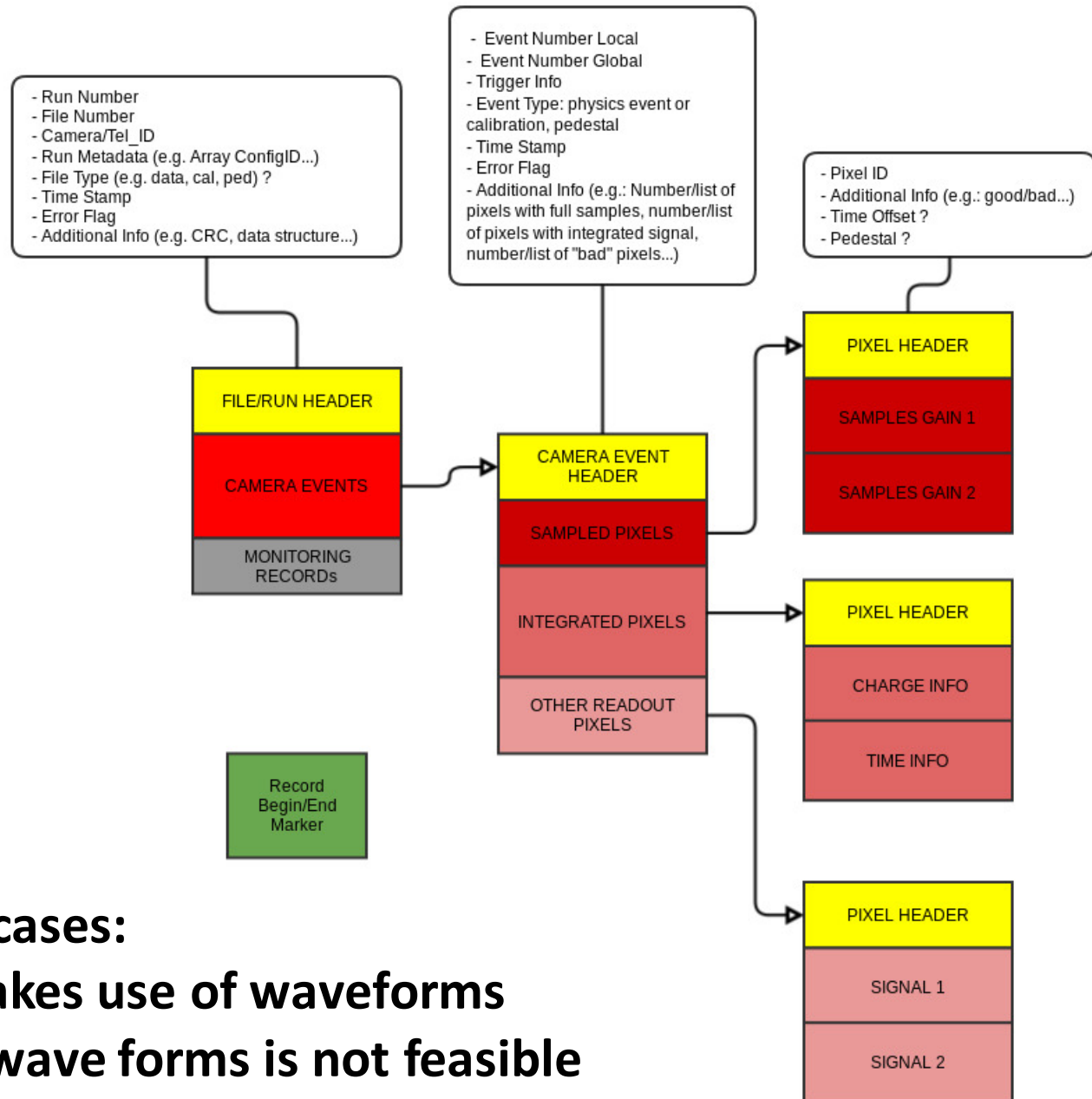
# EVT0 Camera

## Format/content

This can be a possible scheme of the content, extremely general, to be discussed with the cameras

On top of it we can impose an event format and a file format. Chosen from the options being discussed

Part of it already coded in protobuf model



**Motivated by 2 use cases:**

- **Present analysis makes use of waveforms**
- **Data storage of all wave forms is not feasible**

# End-of-May Heidelberg meeting Cameras-ACTL

Data Model assisted as observer

Data Level	Description
Front End Boards	Data produced by the FEB
Camera Server	Data produced by the CS
DAQ	Data produced by DAQ
DL0	Lowest level of data stored ...
DL1	Calibrated Data. Telescope wise
.....	.....

Defined Data levels below DL0

Agreed that DAQ will use google protobuf <sup>(1)</sup> and cameras will check if they can use it

IF it is approved

Started to define ICDS with cameras

Cameras will look at ASTRI ICD document as an example

(1) Protobuf is a method developed by google to define data and serialize them to be sent over the network or to files

[https://en.wikipedia.org/wiki/Protocol\\_Buffers](https://en.wikipedia.org/wiki/Protocol_Buffers)



# What to do for the short term work ?

Pipelines need a DL0 prototype format (Data)

Cameras Need FEB, CS, format (xST)

DAQ also needs a format (ACTL)

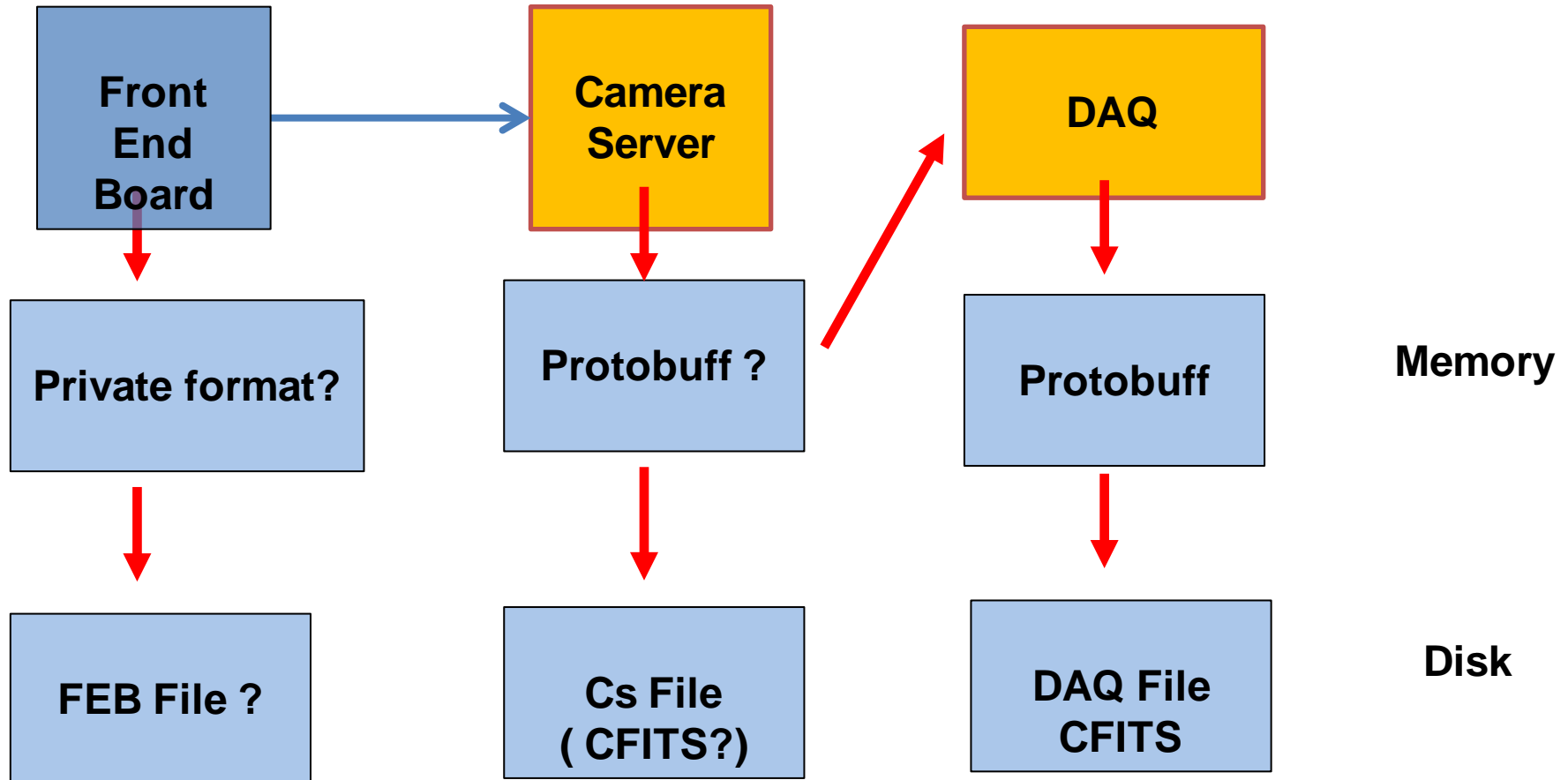


Use temporary structures “scaffolds”

**To be discussed !**

- Pipelines has to use generated MC for input → working basic python interface now available (Jean Jacquemier)
- DAQ will use protobuf and has provided libraries to w/r in CFITS (next slide)
- Cameras will try to use protobuf with content as common as possible following present scheme. Use CFITS for files ?

# From FEB to DAQ (Proposal EL)



# DL0 File Format

- So far 3 file formats have been proposed:

- ACTIO (Konrad) → Adapted from H.E.S.S.
- CFITs (Etienne) → Adapted from FACT
- Streaming (Andrea) → Adapted from satellites

All of them are possible but need to be tested and compared.  
And we might have more proposals ...(HDF5 ?)

## **We are developing a “DL0 Testbed” to compare formats:**

1. Using the same MC samples
2. Debugging and comparing performances

Look in <https://forge.in2p3.fr/projects/model/wiki/DL0> → DL0 testbed

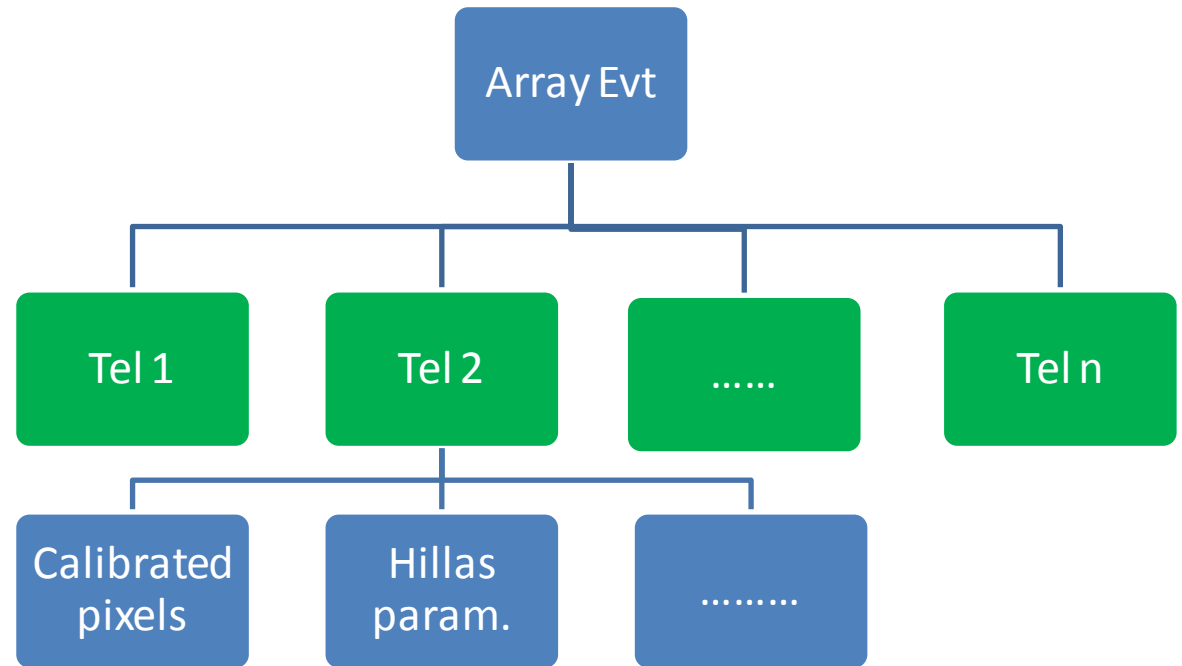
Other factors are important and will be weighted: User base, support, applicability to higher Data Levels.

My lecture of the articles published so far in  
“The future of Astronomical Data Formats”  
(+discussions at ADASS)

- The standard (FITS) has aged.
- Solutions:
  - Private formats. Observatory dependent
  - Upgraded FITS → To be defined
  - Format built on HDF5 → To be defined
- How urgent is to have a DL0 file format ?

# Higher Data Levels pose additional constraints

## Array Events EVT1



Can the file format used for DL0 be similar to the one of DL1 and DL2 (shower) ?

Remember that DL3 is already required to be FITS (uncompressed)

# Proposal: Converge first on Camera Events at DAQ Level

1. Tests of ACTL software. Package provided by Etienne.
2. Check that the right camera data information is available in the wiki
3. We ACTL/DATA provide a baseline file with the data model definition for camera events in protobuf language (updated version of <svn+ssh://svn.in2p3.fr/cta/ACTL/CamerasToACTL/trunk/DataModel/L0.proto>)
4. Cameras customize it to their needs
5. If not possible for cameras to produced protobuf format, DAQ will convert it from low level to camera-customized protobuf
6. We hold a teleconf in 3 weeks to check progress
7. Is it possible to find a slot in the Barcelona meeting ?

## Proposal: Additional points

1. Do we have the correct list or responsible persons ?
2. ICDs
3. Can cameras use CFITs for their prototyping data ?

# Thank you



# Data reduction

- Data reduction is closely linked to data format.
- The proposed event format should enable to keep pixels at different levels of processing.
- Still I suggest to delay the issue until we see how we progress on Data Format.

# Data Reduction for DLO

1. Data reduction takes place before shipping data to data center.
2. Must comply with Science requirements
3. Assuming 3% samples/97% integrated  
We need an additional 10 x total reduction
  1. 2 x from compression
  2. 5 x from other means
4. DR should be reproducible offline
5. Internal quality control

# Data Reduction Steps

1. Suppress some samples/gains ??
2. Identify pixels with signal
3. Choose pixels of interest (region/set)
4. Integrate rest of pixels
5. Suppress pixels
6. Suppress events
7. Compress (is lossy compression accepted?)

# Data Reduction Options

1. FEB
2. Camera Server
3. Real Time analysis
4. Delayed Online analysis

# Data Reduction Options

1. FEB
2. Camera Server
3. Real Time analysis
4. Delayed Online analysis