VO compliant CAM products

R. Lorente
April 2006

1 Purpose of the project

The existing CAM pipeline products have been evaluated in order to check their compliance in the VO framework. In view of this assessment, we have designed and produced a set of new CAM files to be selected as the CAM VO default product as the VO constraints evolve.

2 Current VO constraints

The current VO constraints come from the protocol to query and to get images, the SIMPLE IMAGE ACCESS PROTOCOL or SIAP:

- SIAP output is one single file per observation.
- in the case of a FITS file, it must contain a 2-dimensional image as the Primary Header Unit (PHU). Extensions are permitted but are not fully supported by the interface (see *Simple Image Access Specification Version 1.0* at http://www.ivoa.net/Documents/WD/SIA/sia-20040524.html).

Apart from that, the tools currently available for displaying images in VO, i.e. Aladin, can only display an image if it is in the first extension of the file. If the image is multidimensional, it only displays the first plane of the first extension. This is equivalent to the second SIAP constraint.

3 CAM pipeline products assessment

We will concentrate only on the CAM products that contain images (flux per pixel), i.e. CMOS, CMAP and CSP. There are other byproducts offered at the end of the pipeline that contain more information, like the spectrum of the source in the CVF case (CSSP), or the list of detected sources (CPSL) and their photometry, but we assume that the VO user wants to see an image coming from ISOCAM.

- The file CMOS, or CMAP in the case a mosaic is not built, contains all the information that a user could need for interpreting the data: date and time, coordinates and matrix for sky projection, wavelength and bandwidth, data and errors. Therefore they represent the most complete product to do science in the VO framework. Their main disadvantage is that the images are part of the binary extension which contains the rest of the information, and there is not any image extension as such. Nor the main standard tools to display images, **fv**, **ds9** and **Aladin**, neither SIAP can interpret the CMOS or CMAP contents as an image.

- CSP files are the CAM Browse products. These are the only products that contain an IMAGE extension with the data. They can be displayed by the standard tools, included VO tools. This is the reason why they were selected until now as the CAM VO default products. Their main disadvantages are:
  - they are not scientifically validated.
  - they do not offer all the information needed to interpret the data. As an example, the wavelength appears as a single value in the header. Therefore in multi-wavelength observations there is no information about the wavelength of each image. Moreover, there is no information about the errors associated to each pixel of the image.
  - in some cases there is more than one browse product per observation.
  - in other cases the CSP files are three-dimensional images.
  - finally, the CSP header is not even completely standard, since what it seems to be the CD matrix, is the PC transformation matrix, leading to wrong displayed coordinates in some interpreters (see Calabretta & Greisen, 2002).

4 Requirements for the new CAM Products

Since the VO is continuously evolving it is difficult to assess with confidence the need of a new CAM product, and the pros and cons of each existing product that will persist in the VO framework.

Even guessing that the constraints would relax in the future, we do not foreseen that an special tool will be developed to read a CMOS or CMAP as an image and to be displayed as such. This means that at least the conversion of the CMOS or CMAP files into standard images, plus some essential information, justifies the task described in this document.

As we have seen in Section 2, the main VO constraints for image observations are related to:

- the number of extensions allowed in file
- the dimensions allowed in the image extension
- the number of files allowed per observation

On top of all this, the CAM default VO product should be as much realistically representative of the observation as possible, even if this complicates the structure of the file. As an example, a CAM observation with different pointings and different wavelengths on each of them, i.e. a multi-extension and multi-layer FITS file, is far from being currently VO compliant. Therefore the first issue is to produce a 2D single extension FITS file representative of this kind of observations and already queriable from VO. On the other hand an 'ideal' CAM VO product should contain all the data taken during such an observation together with all the information needed to do science, i.e. an image-like CMOS or CMAP file, what we could consider the *self-contained CAM VO product.* But from now to the point when this kind of file would be allowed in VO, intermediate products should be available from VO at any time and, simultaneously, as close as possible to the *self-contained CAM VO product.*

Taking all this points into account, we suggest to design more than one VO oriented CAM product that could adapt in different ways to the VO evolution. Concerning simplicity, we could design a complex and complete FITS file with as many extensions as needed, or to split it into a set of simpler FITS files. With regard to the VO constraints, we can design products fully compatible with the current VO, an others free from technical constraints. Therefore, the characteristics of the files to be designed are summarized in these two points: 'Are they technically VO compliant?' and 'Do they represent the observation in a self-consistent way?' Each of the produced files should have an affirmative answer to the first question at some stage of the VO evolution, and simultaneously should be as close as possible to the assertion of the second question.

5 CAM products VO compliance oriented

The new products will be inserted in IDA. Each observation will have associated a set of files with different naming convection and one among them should be chosen as default VO product according to the VO dynamics.

5.1 Description of the different files produced for each observation

Since a large variety of files will be produced per observation, it is important to describe them in detail in order to be able to know which is the exact file to be chosen depending on the VO contraints. Table 1 summarises which files are associated to which kind of observations and describes their internal structure.

NAMING	OBS	INTERNAL STRUCTURE
CSLI_tdt	all	single file, single 2-D image ext
CCUB_tdt	single conf multi-layer	single file, single 3-D image ext
CICx_tdt	multi-conf	multiple files, single n-D image ext
CIMA_tdt	all	single file, multi-ext (2-D image at PHU, + data)
CSCP_tdt	all	single file, multi-ext, + data

Table 1: Summary of the different files produced. Column 1: naming convention of the FITS files. Column 2: Observations that have associated a given product: all observations, single, multi-configuration or multi-layer observations. Column 3: Number of files per observation with the given name, and its internal structure (number of extensions + number of dimensions).

The files that have been produced are:

- CSLI_tdt.fits: *single file, single 2-D image extension.* The filename stands for *CAM SLIce.* This file is already SIAP compliant. Only if the observation consists of a single image, it is fully represented by this file. The 2D image, or *slice* contained in this file is:
  - the 2-D image of the observation if this is composed by a single image, single configuration.
  - the brightest 2-D image, if the observation is composed by several images with the same configuration. The *brightness* of an image has been defined as the one of the 10×10 central pixels.
  - the 2-D image corresponding to the most used configuration, if the observation is composed by several configurations and if the most used one is a 2-D image.
  - the brightest 2-D image of the most used configuration, if the observation is composed by several configurations that in turn are composed by several images.
- CCUB_tdt.fits: *single file, single 3-D image extension.* The filename stands for *CAM CUBe.* This option is not yet SIAP compliant, since, even if it is a single file, it consists of a 3-D image extension (*cube*). Only if the observation can be expressed by a single configuration (see Appendix), it is fully represented by this file. In the remaining cases it contains the most used configuration, if in turn this configuration consists of more than one 2-D image.
- CICx_tdt.fits (x=a,b,c,...): *set of files with single 3-D or 2-D image extension.* The filename stands for *CAM Individual Configuration.* This set exists only for observations with more than one configuration (see Appendix) and in this case it is more complete than the files described above. The set has as many files as configurations in the observation, each one with an image extension, 2-D if there is a single image with this configuration, and 3-D if there is a multi-layer image with this configuration. The different files are distinguished by the fourth letter in the filename (x=a,b,c... until reaching the number of configurations). This set would be suitable only if several files per observation are allowed. The only requirement met by this set is that they are all single extension images. It is an intermediate product only valid at an hypothetical case where several files are allowed but their internal structure is restricted to be a single extension file.
- CIMA_tdt.fits: *single file, multi extension.* The filename stands for *CAM IMAge.* This file is far from being VO compliant and it should be suitable

only if at some VO stage, complex files are allowed but a 2-D image in the PHU is still required. Therefore the first extension is kept to be a 2-D image as the one in CSLI_tdt.fits, followed by a binary table with all the data contained in the CMAP or CMOS binary table. After it, a set of m extensions which contain n_i-D images, being m the number of configurations in the observation, and n_i the dimension (2-D or 3-D) of the i configuration.

- CSCP_tdt.fits: *single file, multi extension*. The filename stands for *CAM Self Contained Product*. This file is equivalent to the previous one, but it does not contain the first 2-D image extension. No restriction has been put in its contents, only the obvious requirement of having the data as n-D images, which, as explained, will remain to be the main disadvantage of all the pipeline CAM products. It contains the main header of the CMAP product (CMOS when available), followed by a binary table with all the contents of the CMAP or CMOS binary table and finally the set of m extensions as in CIMA_tdt.fits.

As it has been described CIMA_tdt.fits and CSCP_tdt.fits contain the binary extension with the data exactly as they appear in the CMOS or CMAP files. Even if this bynary extension has redundant information it has been decided to keep it to avoid puzzling the users already familiar with CAM pipeline products. The presence of this binary extension in CIMA_tdt.fits and CSCP_tdt.fits means that they contain all the information that was decided to be stored in the main CAM pipeline products, together with as much images as the observation requires. Therefore this two products have to be consider as the *self-consistent CAM VO products*.

5.2 Files associated to the observations and rules for file selection

These are the files that can be found depending on the complexity of the observation:

- CSLI_tdt.fits exists for all the observations
- CCUB_tdt.fits exists only for the observations with multi-layer images: observations with a single configuration composed by a multi-layer image, and observations with multiple configurations where the most used one is a multi-layer image.
- CICx_tdt.fits files exist only if the observation contains more than one configuration.

- CIMA_tdt.fits and CSCP_tdt.fits exist for all the observations. Their number of extensions depends on the observation.

The choice of the ad-hoc file should be made following Table 2 and the enclosed diagram. Since not all the files exist for all the observations, the optimized option is always the one that is closer to the most complete one.

VO CONSTRAINTS	single extension		various extensions
	single file	various files	
2-D image	CSLI_tdt.fits		CIMA_tdt.fits
2/3-D image	CCUB_tdt.fits CSLI_tdt.fits	CICx_tdt.fits CCUB_tdt.fits CSLI_tdt.fits	CSCP_tdt.fits

Table 2: Checklist to choose the correct file according to the requirements and to the available files per observation. When different options are present (more than one file per box) a file should be chosen if and only if the previous one does not exist for an observation

As an example, let's take the most difficult situation: various files per observation can be accessed from VO with no restriction on the image dimensions. In this case the m CICx_tdt.fits files should be taken. But these files exist only for multi-configuration observations. Therefore for the single-configuration observations the CCUB_tdt.fits should be taken if available, CSLI_tdt.fits otherwise. This is because we have to choose always the most complete existing product.

As it can be seen from Table 2 the option where various files per observation are accessible from VO but which contents is limited to 2-D images is not covered by the current design for obvious reasons. In this case the CSLI_tdt.fits file should be chosen.

6 Summary

New CAM products have been generated with the aim of having a set of files per observation that at the same time that they could fulfill the VO constraints in each of its stages, they keep as much information and data as possible. From the file already SIAP compliant, CSLI_tdt.fits, to the *self-consistent CAM VO products*, CSCP_tdt.fits and CIMA_tdt.fits, intermediate files have also been produced. Some of them cover different VO constraints relaxation stages, that could be beside the point. In all these intermediate cases, the header of the file warns the user about its incompleteness.

CSCP_tdt.fits and CIMA_tdt.fits are, at best, the most suitable CAM products to be accessible from VO. They contain all the information kept in the main CAM pipeline products, CMAP or CMOS, together with an image-like reformatting of the data stored in as many extensions as needed.

The VO compliant CAM products have been produced for CAM01, CAM03 and CAM05 observations with available CMAP or CMOS: 6474 CAM01, 374 CAM03 and 62 CAM05. CAM04 observations will have a different VO compliant product that will not be based on CMAP or CMOS contents and that will be inserted in IDA.

Table 1 explains the structure of the different files. Diagram 1 and Table 2 guides to the suitable file to be chosen according to the requirements.

7 Bibliography

- Calabretta, M. R., Greisen, E. W., 2002, A & A 395, 1077

8 Appendix: What is an image configuration?

The observations can be classified according to the number of configurations it contains. A *configuration* is the set of parameters that must be part of an image fits header. Therefore a change in any of these parameters should mean that there is a change of configuration, or in another words, a new image extension, since the same header cannot be shared if the parameters are different.

The set of parameters that are part of the configuration are:

- CRVALn
- CROTA2
- CRPIXn
- CDELTn
- NAXISn

An important problem arises when defining what it is the meaning of "same CRVALn" or "same CROTA2". Often there is not a real change of coordinates but a small difference due to drifts during the observation, mainly in the CROTA2 values. The limit of "same coordinates" has been set to around 0.5 arcseconds: a drift in coordinates bigger than this is consider to be a change in configuration.

The configuration definition means that a change on any other parameter does not imply a different image, but a different plane of a 3-D image. As an example, the wavelength is not part of the configuration, since images with different wavelenths can share the same image header as far as all the parameters above do not change. In this case the keyword "WAVE" in the header will contain a string with as many wavelengths as needed and the corresponding extension will be a 3-D image, with one plane per wavelength.

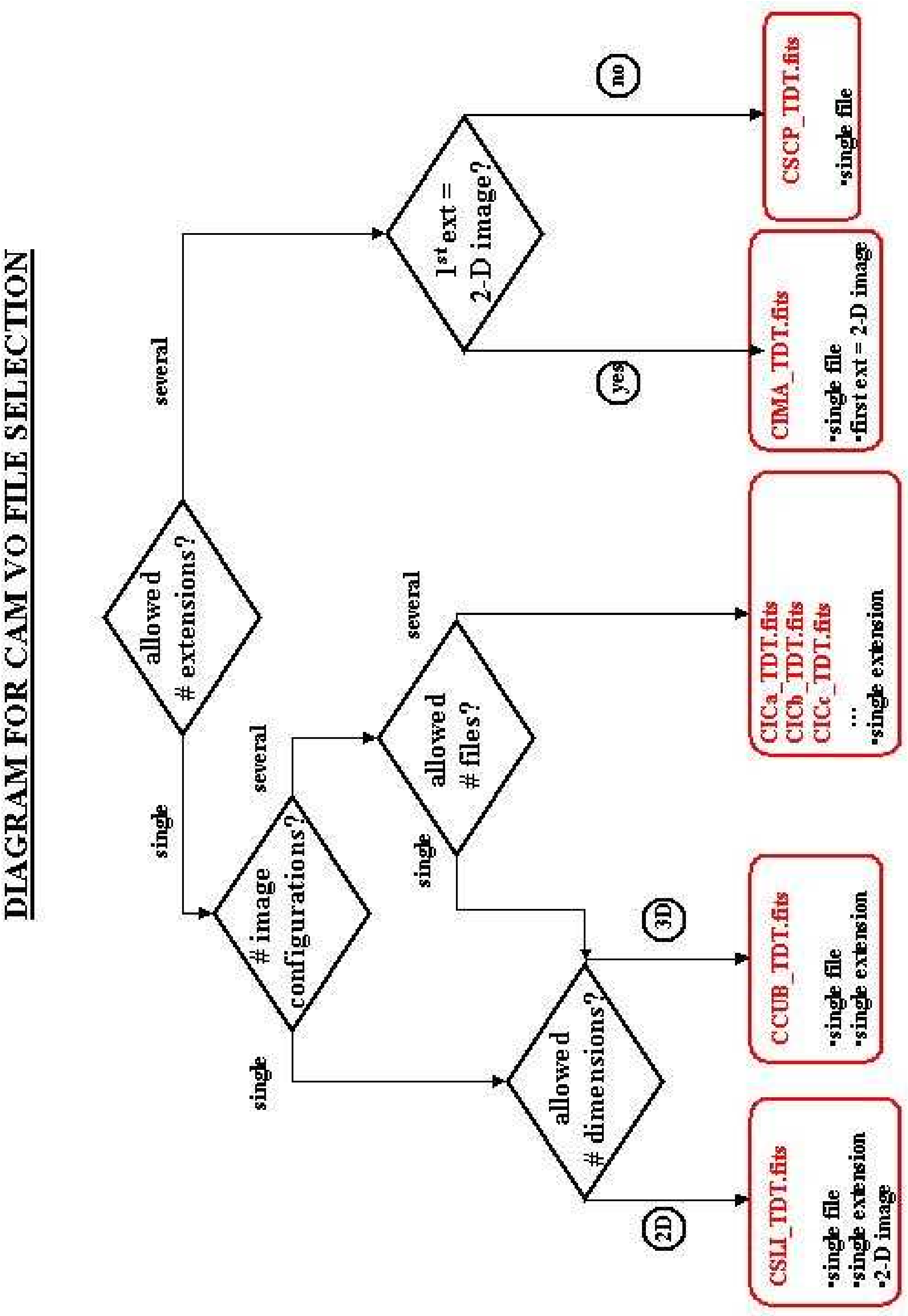

Figure 1: Diagram for the selection of the suitable file according to the VO constraints. Box in red contain the name of the corresponding file, together with a short description of their contents.