EUROPEAN SOUTHERN OBSERVATORY

European Organization for Astronomical Research in the Southern Hemisphere
Organisation Europeenne pour des Recherches Astronomiques dans l'Hemisphere Austral
Europaische Organisation fur astronomische Forschung in der sudlichen Hemisphare

Internal Memorandum

From: Alain Smette, Ricardo Schmutzer, Paul Eglitis, Adam Dobrzycki

To: Andreas Kaufer, Fernando Comeron, Michael Sterzik, Ueli Weilenmann

Cc: Ricardo Schmutzer, Paul Eglitis, Adam Dobrzycki, Pedro Baksai, Ismo Kastinen, Ferdinando Patat, Mark Neeser, Pascal Ballester, Michael Hilker, Monika Petr-Gotzens, Fernando Selman, Martino Romaniello, Reinhard Hanuschik, Antoine Merand, Francesca Primas

Date: October 28, 2010

Subject: **Instrument Name Identifier**

On October 20, 2010, a phone meeting was held to summarize the issues regarding instrument identifier name. Its goal is to avoid the recurrence of problems associated with the instrument name which lead: (a) to the XSHOOTER instrument name change, as the identifiers for the front and back ends were different and, (b) to large amount of time spent on considering the possibility to change the prefix of the VIRCAM filenames from VCAM to VIRCAM (see PPRS-034105). The timing of this meeting, about 6 months after the XSHOOTER name change was meant to make sure that all possible problems associated with filenames were identified (cf. PPRS-035000).

This memorandum summarizes (1) the limitations to the filename length caused by various systems and (2) the current naming conventions. It also (3) suggests 4 recommendations to avoid repetition of the above problems. Finally, (4) it summarizes the status of the coming instruments, pointing out that a decision is urgently needed in the case of OmegaCAM.

Participants to the meeting were: Alain Smette, Instrument Operations Team coordinator; Ricardo Schmutzer, Paranal Software Group; Paul Eglitis, Archive Department head; Adam Dobrzycki, ESO Data Interface Control Scientist.

1. Limitations to the filename length.

a. Limitations caused by the detector controllers.

i. Names created by FIERA are limited to 32 characters. SPR VLTSW20100148, following up on PPRS-035000 proposes a workaround in BOSS to fix the filename in the OLDB callback, This workaround works only for filenames up to 37 chars. long (32 + ".fits").

ii. Names created by IRACE and NGC are limited to 256 characters.

b. FITS limitations.

Length of file names (FITS or not) that are to be stored as values of keywords in the headers of FITS files are limited to 68 characters. This unbreakable limit comes from the fact that from 80 available characters in a card, 8 characters are used for the keyword name, 2 characters are used for '<SPACE> =' in columns

9-10 of the card, and 2 characters are used for the two delimiting quotes.

Filenames stored in a hierarchical keyword have an even shorter length.

The following 3 keywords are used for files produced by an instrument: (i) ARCFILE: archive file name, (ii) ORIGFILE: file name on instrument workstation, (iii) CDBFILE: pipeline output file name. Of those, only ARCFILE is "safe" from having problems with the 68-character limit, because of well defined convention that puts the names below the limit. The other do not: ORIGFILE can in principle be asked to store whatever IRACE and NGC produce - both with 256 character limit. Similarly, pipeline developers could *a priori* create CDBFILE filenames of arbitrary lengths, however, see point (d) below.

Once Phase 3 is implemented, additional keywords may be added: developers must be aware of this limitation.

c. Limitations originating from databases.

i. Most serious limitation comes from the ngas database. Apparently for historical reasons, the length of the file identifier entry (file_id) is 64 characters. For FITS files, the identifier is the file name without ".fits" (i.e. the actual limit is 69 characters). For all other file types file_id is the entire file name (note that this is the *archive* file name, i.e. for FITS files this is equivalent to ARCFILE).

ii. Second strongest limitation is on the full name length of pipeline products in the QC products database: 80 characters. But since this information is put into CDBFILE keyword in the FITS file, the FITS 68 character limit is strictest. This should ensure that this DB limitation will not come to play.

iii. For completeness, other - less likely to matter - database limitations are:

1. Identifier in observations database (ARCFILE minus ".fits"): 75 chars.
2. Full file name in ngas database (file name plus directory structure inside ngas disk): 255 chars

iv. Modification of the ngas and observations databases would be difficult. Both are large databases, replicated between Chilean sites and Garching. Thus any fix would have strong (albeit one-time) impact on observatory operations. Any changes to the ngas and observations databases - both operations critical DBs - would have to undergo full blown integration testing prior to implementation, competing for SED resources with lots of other projects, which would almost certainly resulting in a rather long implementation timescale.

d. Other.

Files created by the pipelines add up to 11 characters to the ARCFILE filename: the prefix 'r.', the suffix '_tpl_<iiii>' . A possible additional character is the 't' in 'tfits' to indicate that the file is a binary FITS table, although this convention is now dropped by the pipelines.

2. Current naming conventions for archival files.

The current naming convention is not fully standardized:

a. ORIGFILE names are based on INS.CON.ID. Additional characters correspond to

instrument modes, day of the year (DDD), identifier (_NNNN), which usually add 5 characters but sometimes 10 characters (in some cases the identifier is repeated). Note that several instrument modes can be selected at the same time: for example, in the case of XSHOOTER, the corresponding string is composed of 3 substrings separated by '_': (1) 'SLT or IFU', (2) 'AFC', 'OBJ', 'STD', (3) 'UVB', 'VIS' or 'NIR'.

b. ARCFILE for pipeline products and External Data Products.

i. ARCFILE names are based on <INS-PREFIX>. Additional characters correspond to the restricted ISO 8601 format for the time tag, 28 characters, incl. the '.fits'. Section 10.1 of the Data Interface Control Document GEN-SPE-ESO-19400-0794 Issue 4, 8 April 2008 states that the <INS-PREFIX> should have a length of between 4 and 10 characters.

ii. This <INS-PREFIX> is set by the OLAS_ID (online archiver system identifier) environment variable, which so far followed a convention to limit it to the first 5 characters of the configuration file keyword INS.CON.ID. (The difference between INS.CON.ID and the FITS INSTRUME keyword is the possible presence of 'improper' characters in the latter.) The reason for this 5 characters limit seems to be related to an old limitation of a previous database. ***However, this convention is only valid for instruments that obey the instrument common software and is not strictly followed*** (cf. VIRCAM).

iii. CDBFILE names follow a convention defined by the pipeline developers; see 1.d above.

3. Recommendations.

a. The configuration keyword INS.CON.ID is the unique source for naming an instrument that follows the instrument common software. The name of the instrument in INS.CON.ID should have a length of 4 to 10 characters and only contents either a letter or a number; no other characters are permitted. This name should be used for all operations, front- and back- ends, in other words, from the name used in the Call for Proposals, ESOFORM, p2pp ,to names of templates, of files etc…

b. Consequently, the OLAS_ID environment variable must be identical to the INS.CON.ID configuration file keyword for instruments for which commissioning has not yet started at the time of approval of this recommendation. Such a statement should be part of the Data Interface Control Document GEN-SPE-ESO-19400-0794. For new instrument for which commissioning data have been obtained, the previous convention holds, i.e. the first 5 letters of INS.CON.ID.

c. The document VLT-MAN-ESO-17240-2240 Common Software for Templates User Manual should be updated similarly.

d. The instrument name, as it appears in INS.CON.ID, and its uniqueness should be part of the Preliminary Design Review for new instruments. Its length should take into account the limitations described in Sec. 1 above and the naming convention described in Sec. 2.

4. Status of coming instruments.

The status of the coming instruments is as follow:

a. OmegaCAM:

i. All the back end SW is based on the value of INS.CON.ID, set to OCAM, as can be seen in the file ocmcfg/config/ocmcfgINS.cfg (ocmcfg version 1.111, Jun 18, 2008).

i) All templates have names starting with OCAM.

ii) All filenames (ORIGFILE and ARCFILE) have names starting with OCAM. Only Intermediate Level Test files have been produced.

iii) QC1 log files have names starting with QC1_OCAM.

ii. As far as we can tell, the name used in the front end SW has not yet been defined. For consistency, we have the choice of either use the name OCAM, which would have the minimal manpower impact, or change all the back end to OMEGACAM to insure consistency with the publicized name. ***A decision on the name to be used in the front end is urgently needed.***

b. PACMAN:

i. The value of INS.CON.ID is set to PACMAN, as can be seen in the file pamcfg/config/pamcfgINS.cfg (pamcfg version 2.12, Sep 28, 2010).

i) All templates have name starting with PACMAN.

ii) The OLAS_ID is set to PACMA, so that filenames have ORIGFILE starting with PACMAN and ARCFILE starting with PACMA.

iii) Commissioning data have already been ingested in the archive.

ii. There is no reason to justify any change. However, consistency must be insured at the front end.