

e-VLBI File-Naming Conventions

As agreed by Ed Himwich, Yasuhiro Koyama, Cormac Reynolds, Alan Whitney
29 Nov 2004

Goals

As e-VLBI continues to develop, it is essential that procedures be adopted to ensure reliable identification of e-VLBI raw-data files. One important aspect is an agreement on file-naming conventions.

The goals of the file-naming convention are as follows:

- Identify the experiment, station and scan name
- Identify the file format (VSI, K5, Mark 5A, Mark 5B, PC-EVN, etc)
- Provide enough information in the filename so that the data file can be transformed from one file format to another, as necessary, with no additional information. For example, if the data do not have embedded time codes, the start time of the data must be specified.

Filename Format

The filename format for a file containing data from a single scan is

`<exp name>_<station code>_<scan name>[_<data start time>_<aux info1>_<aux info2>...].<file type>`

where

- `<exp name>` - experiment name; max 6 chars (consistent with current limit)
- `<station code>` - standard 2-character station code, followed optionally by a numeric suffix (example: 'ef2') for the case where a scan spans a multi-file set all spanning the same time period of data.
- `<scan name>` - assigned scan name (derived from VEX file or other source); max 16 chars
- `<data start time>` - (optional) start time of data in file; required if data start time is not unambiguously embedded in the data itself. Format may be either 1) VEX time format or 2) undelimited time of form 'yyyydddhhmmss' (13 digits), 'dddhhmmss' (9 digits), 'yyyyddd' (7 digits), or 'hhmmss' (6 digits). Fractional seconds should be specified as necessary. The `<data start time field>` is mandatory when a single scan is broken into a time series of files with the same `<scan name>`.
- `<aux info>` - (optional) auxiliary information field(s) in format 'cc=ppp' where 'cc' is a standardized 2-char identifier for information and 'ppp' is the information value in some specified standardized format (example: 'bm=0x0000ffff' specifies the VSI 'bit mask' used in collecting the data)
- `<file type>` - identifies high-level data format within file (for example: 'vsi', 'm5a', 'evn', etc. for VSI, Mark5A and PC-EVN formats, respectively)

Example filename: 'gre53_ef_scan035_154d12h43m10s.vsi'

Rules

1. No embedded white space.
2. No underscore or '=' characters are allowed in parameter fields.
3. <Aux info> fields may be in any order and must be self-identifying with a 2-character ID.
4. If <data start time> is not present, the first <aux info> field may be delimited from <scan name> by a single underscore.
5. Lower-case is *preferred* for all fields; this preserves selecting file names by filename wildcards
6. Maximum filename length – 64 chars

Example File Names

Time is fully encoded in data:

'gre53_ef_scan035.vsi'

Year and day not encoded in data:

'gre53_ef_scan035_2004y154d.vsi'

'gre53_ef_scan035_2004154.vsi'

Scan occupies four parallel files:

'gre53_ef1_scan035.k5'

'gre53_ef2_scan035.k5'

'gre53_ef3_scan035.k5'

'gre53_ef4_scan035.k5'

Scan broken into four time segments starting at 10-second intervals:

'gre53_ef_scan035_154d12h43m10s.vsi'

'gre53_ef_scan035_154d12h43m20s.vsi'

'gre53_ef_scan035_154d12h43m30s.vsi'

'gre53_ef_scan035_154d12h43m40s.vsi'

File Types

The following <file type> suffixes are suggested:

'evn' – PC-EVN

'k5' – K5

'm5a' – Mark 5A

'm5b' – Mark 5B

'vsi' – VSI

Others may be added as necessary.

Aux Info Codes and Parameters

The following general aux info codes and parameter formats are suggested:

Name	2-char code	Parameter value
Bit mask (VSI)	bm	VSI hex bit mask (example - '0x0000ffff')
BSIR (VSI)	sr	VSI Bit-stream information rate (sample rate) in MHz (example - '32')

Additional aux info codes may be agreed on by the community.

Further Suggestions

It may be useful to consider naming other various types of non-data files in a similar manner so that all filenames have a consistent format. For example, an experiment-wide schedule file might be named

<exp name>.skd

A station-specific log file might be named

<exp name>_<station code>.log

A scan-specific auxiliary file might be named

<exp name>_<station code>_<scan name>.aux

Though there is no hurry to adopt these immediately, but it seems logical to move in that direction as resources allow.