NAME

streamarchive - StreamArchive file format

DESCRIPTION

StreamArchive typed archives are a series of **keyword** and **value** records that are similar to content of the POSIX.1-2001 extended headers called **TAR (PAX) HEADERs**, based on a proposal from **Sun Microsystems** from 1997.

A new file always begins with the **path** keyword and after the mandatory **size** keyword, file content may follow. Each file record is terminated by a **status** keyword.

An archive begins with an **archtype=StreamArchive** record and ends with a **status=EOF** record.

The archive meta data do not add non-printable characters. If the file names in the archive are only made from ASCII characters and if the archive only contains files with ASCII content, the whole archive contains only ASCII content.

HEADER FORMAT

The header records use the following format:

"%d %s=%s\n", <*length*>, <*keyword*>, <*value*>

Each record starts with a a decimal length field. The length includes the total size of a record including the length field itself and the trailing new line.

The *keyword* may not include an equal sign. All keywords beginning with upper case letters are reserved for local extensions.

If the value field is of zero length, it deletes any header field of the same name that is in effect from the same extended header or from a previous global header.

Null characters do not delimit any value. The data used for *value* is only limited by its implicit length.

HEADER KEYWORDS

All numerical values are represented as decimal strings. All texts are represented as UTF-8 or an unspecified binary format (see **hdrcharset** keyword) that is expected to be understood by the receiving system:

atime

The time from **st_atime** in sub second granularity. A nanosecond granularity is currently supported.

charset

The name of the character set used to encode the data in the following file(s).

The following values are supported for **charset**:

ISO-IR 646 1990	ISO/IEC 646:1990
ISO-IR 8859 1 1998	ISO/IEC 8859-1:1998
ISO-IR 8859 2 1998	ISO/IEC 8859-2:1998
ISO-IR 8859 3 1998	ISO/IEC 8859-3:1998
ISO-IR 8859 4 1998	ISO/IEC 8859-4:1998
ISO-IR 8859 5 1998	ISO/IEC 8859-5:1998
ISO-IR 8859 6 1998	ISO/IEC 8859-6:1998
ISO-IR 8859 7 1998	ISO/IEC 8859-7:1998
ISO-IR 8859 8 1998	ISO/IEC 8859-8:1998
ISO-IR 8859 9 1998	ISO/IEC 8859-9:1998
ISO-IR 8859 10 1998	ISO/IEC 8859-10:1998
ISO-IR 8859 11 1998	ISO/IEC 8859-11:1998
ISO-IR 8859 12 1998	ISO/IEC 8859-12:1998
ISO-IR 8859 13 1998	ISO/IEC 8859-13:1998
ISO-IR 8859 14 1998	ISO/IEC 8859-14:1998
ISO-IR 8859 15 1998	ISO/IEC 8859-15:1998
ISO-IR 10646 2000	ISO/IEC 10646:2000

ISO-IR 10646 2000 UTF-8 ISO/IEC 10646, UTF-8 encoding

BINARY None

comment
Any number of characters that should be treated as comment. The comment is ignored.

ctime
The time from **st_ctime** in sub second granularity. A nanosecond granularity is currently supported.

dev The device id from **st_dev** of the file as decimal number.

The value is a signed int. An implementation should be able to handle at least 64 bit values. Note that the value is signed because POSIX does not specify more than the type should be an int.

devmajor
The device major number of the file if it is a character or block special file. The argument is a decimal number.

The value is a signed int. An implementation should be able to handle at least 64 bit values. Note that the value is signed because POSIX does not specify more than the type should be an int.

devminor
The device minor number of the file if it is a character or block special file. The argument is a decimal number.

The value is a signed int. An implementation should be able to handle at least 64 bit values. Note that the value is signed because POSIX does not specify more than the type should be an int.

filetype
A textual version of the real file type of the file. The following names are used:

unallocated An unknown file type that may be a result of a **unlink**(2) operation. This should never happen.

regular A regular file.

contiguous A contiguous file. On operating systems or file systems that don’t support this file type, it is handled like a regular file.

symlink	A symbolic link to any file type.
directory	A directory.
character special	A character special file.
block special	A block special file.
fifo	A named pipe.
socket	A UNIX domain socket.
mpx character special	A multiplexed character special file.
mpx block special	A multiplexed block special file.
XENIX nsem	A XENIX named semaphore.
XENIX nshd	XENIX shared data.
door	A Solaris door.
eventcount	A UNOS event count.
whiteout	A BSD whiteout directory entry.
sparse	A sparse regular file.
volheader	A volume header.
unknown/bad	Any other unknown file type. This should never happen.

arfiletype

The following additional file types are used in **arfiletype**:

hardlink

A hard link to any file type.

fsdevmajor

The device major number of the file (from **st_dev**) as a decimal number.

The value is a signed int. An implementation should be able to handle at least 64 bit values. Note that the value is signed because POSIX does not specify more than the type should be an int.

fsdevminor
The device minor number of the file (from **st_dev**). as a decimal number.

The value is a signed int. An implementation should be able to handle at least 64 bit values. Note that the value is signed because POSIX does not specify more than the type should be an int.

gid The group ID of the group that owns the file. The argument is a decimal number.

gname
The group name of the following file(s) coded in UTF-8 or (if the **hdrcharset** keyword is present) coded to fit the charset value.

hdrcharset
The name of the character set used to encode the data for the **gname**, **linkpath**, **path** and **uname** fields in the POSIX.1-2001 extended header records.

The following values are supported for **hdrcharset**:

ISO-IR 10646 2000 UTF-8	ISO/IEC 10646, UTF-8 encoding
BINARY	None

ino The inode number from **st_ino** of the file as decimal number.

The value is an unsigned int. An implementation should be able to handle at least 64 bit unsigned values.

linkpath
The name of the *linkpath* coded in UTF-8 or (if the **hdrcharset** keyword is present) coded to fit the charset value.

mtime
The time from **st_mtime** in sub second granularity. A nanosecond granularity is currently supported.

nlink

The link count of the file as decimal number.

The value is an unsigned int. An implementation should be able to handle at least 32 bit unsigned values.

path

The name of the *path* coded in UTF-8 or (if the **hdrcharset** keyword is present) coded to fit the charset value.

size

The size of the file as decimal number. The **size** keyword may not refer to the real file size but is related to the size if the file in the archive.

status

The **status** keyword appears after file data and is used to signal whether the last file has been transferred correctly. The first **status** keyword that appears after file data, has a number as parameter. If this number is equal to **0**, then the file data has been successfully transferred into the archive. If this number is non-zero, it is the **errno** from the creating system.

In addition, each archive is terminated by a **status** keyword with the argument **EOF** to singal the end of the archive.

uid The uid ID of the group that owns the file. The argument is a decimal number.

uname

The user name of the following file(s) coded in UTF-8 or (if the **hdrcharset** keyword is present) coded to fit the charset value.

VENDOR.keyword

Any keyword that starts with a vendor name in capital letters is reserved for vendor specific extensions by the standard.

SEE ALSO

star(5).

BUGS

None currently known.

Mail bugs and suggestions to **schilytools@mlists.in-berlin.de** or open a ticket at **https://codeberg.org/schilytools/schilytools/issues**.

The mailing list archive may be found at:

https://mlists.in-berlin.de/mailman/listinfo/schilytools-mlists.in-berlin.de.

AUTHORS

Joerg Schilling and the schilytools project authors.