SDF: Stanford Data Format

Zak Sahul, Oscar Carrillo, Jed Black
Stanford University

SDF Philosophy

- Allow storage of arbitrary data types
  - signed and unsigned integers
  - arbitrary size representations (1, 2, 4, etc bytes)
  - floating point numbers
  - complex numbers
  - ascii
- Record events
  - as flags, or text with a timestamp
  - simultaneous events

SDF Philosophy

- Arbitrary signal characteristics
  - each signal at different sampling frequency
  - signal range
  - any number of signals
- Store *any* and *all* ancilliary information
  - calibration parameters
  - filter settings
  - patient identification
- Easily extensible
  - need to store new information

SDF Overview
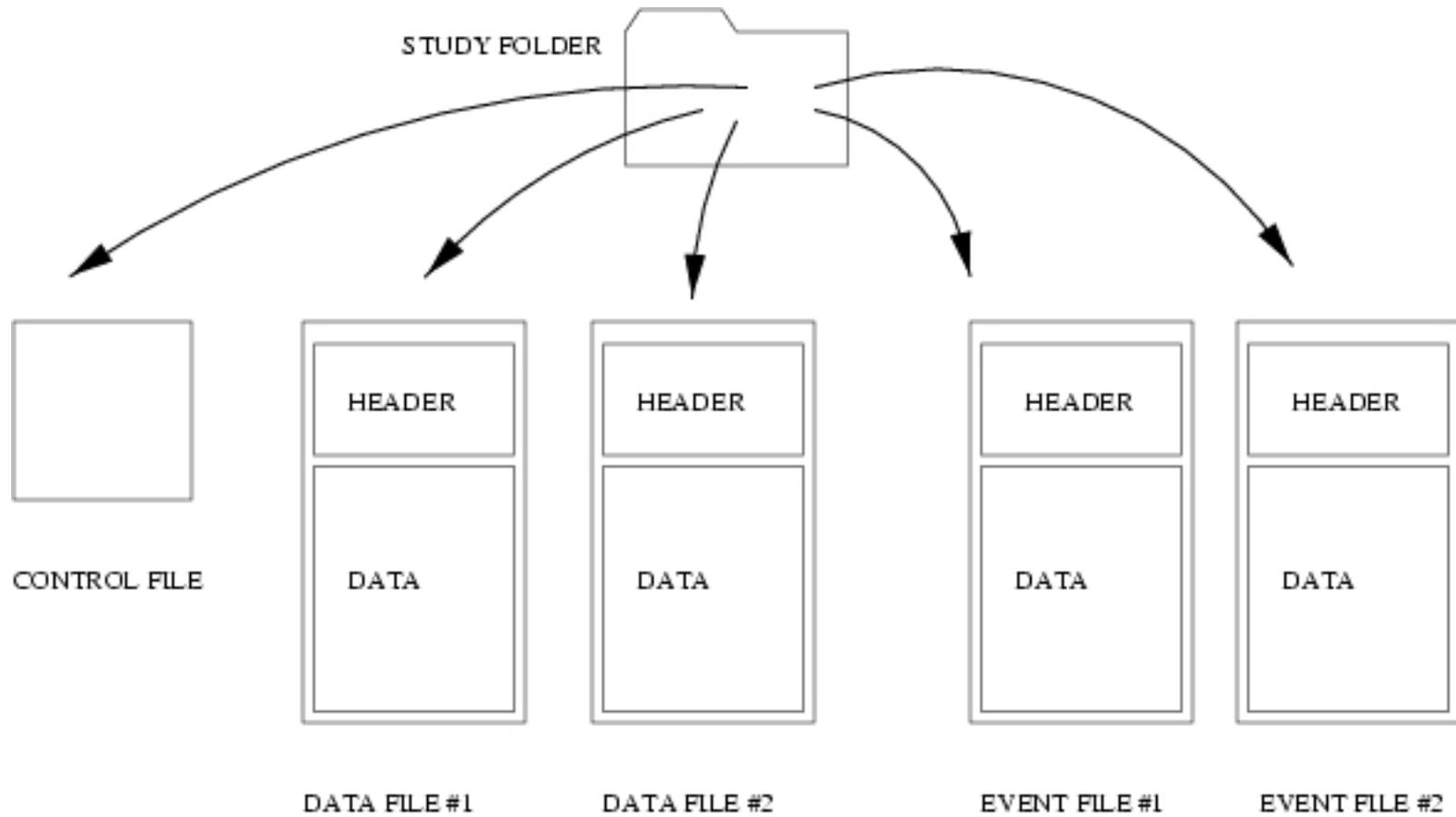
SDF

- Each study in a folder
- "Control.txt" file contains information about the study and about the other files *in ASCII text form.*
- Each channel data in a separate file
  - Each file contains header and data sections
- Each type of event in a separate file
  - Each file contains header and data sections
- Header section contains information on what is in the data section.
- Header section *in ASCII text* human readable form.

control.txt

- human readable text file

- contains lines of keywords with attribute name and value pairs:

  - patient -id "John Doe" -age 32

- Similar to tcl/tk, attribute names start with a "-".

- Minimal but not necessary set of keywords defined for PSGs

- An application may define new keywords or attribute name-value pairs.

- unfamiliar keywords or name-value pairs should be ignored.

control.txt keywords

- patient
  - -name
  - -id
- recording
  - -id
- datafile
  - -name
  - -label
- eventfile
  - -name
  - -label

Example

patient -name "John Doe" -age 42 -number 12345

recording -id "Sleep Study"

datafile -name ch0.x -label EEG

datafile -name ch1.x -label EEG

datafile -name ekg.x -label EKG

eventfile -name artifact.dat -label "Artifact"

eventfile -name stage.dat -label "Sleep Stage"

videofile -name video.mpeg -label "PLMS_Video"

Data Files

- Header and data sections with header in ascii.

- Similar to control.txt, there are keywords and zero or more attribute name and value pairs.

- Keywords and attributes are used to describe the data and how to read it.

- A minimal but not necessary set of keywords are defined for PSGs.

- Unrecognized keywords and attributes should be ignored by applications.

Data Files

- Default values should be assumed for keywords and attributes that are not present in the data - so theoretically no application should die because of what the application considers as incomplete header.

- Raw data starts after the keyword "data" with no attributes

Data File Minimal Keyword Set

- patient
  - -id
  - -name

- recording
  - -id

- data
  - -ascii   true or false
  - -binary   true or false
  - -bytes   byte length per data point
  - -type   char, int, float,

Data File Minimal Keyword Set

- data
  - -samplingrate   in Hz
  - -calibration   calibration method
  - -storage   "big-endian or little endian"

Example data file

patient -name "John Doe" -id 1234 -age 42

recording -site "Sleep Disorders Clinic"

data -ascii true

data -type "integer" -calibration insignal

data -sampling rate 100.0

data

515

790

1160

...

Event files

- Header and event data section

- All event data should be ASCII as well

- Data section consists of columns.

- First column represents a measure of time - clock time, time elapsed since start of recording, epoch number, sample number etc.

- Rest of the columns represent the event that happened the time specified in the first column.

- The header documents how the events are stored.

Event File Minimal Keyword Set

- patient
  - -name   -id
- recording
  - -id
- event
  - -label   name of event
  - -basis   time, sample or epoch #
  - -basisunits   units of the basis (eg. seconds)
  - -variable   variable name representing data

Example

patient -name "John Doe" -age 42

recording -site "Sleep Clinic"

event -label "Calibration" -datafile ch0.x

event -basis "time" -basisunits "seconds"

event -variable "slope" -type "float" -units "uV/t"

event -variable "offset" -type "float" -units "uV"

data

0.0	100.0	5.0
3600.0	110.1	11.8
10800.0	100.3	6.3

SDF Advantages

- ASCII header - easy to understand and decipher data files

- Minimal keyword and attribute pairs allows exchange of data

- Other keywords can be defined by individual institutions for their own benefit.

- All types of data supported. Event data stored efficiently.

- Each study is a collection of smaller files

SDF Advantages

- Can access each channel's data independently and quickly - faster throughput of analysis programs and one can easily blind certain channels for RCTs.

- One can encrypt certain attribute values after data has been recorded.

- Applications not encumbered with epoch duration, sampling rate, number of signals etc imposed by data format.

SDF Disadvantages

- A collection of files can lead to data loss if not careful in cataloging.

- Carriage-return, line-feed in unix versus DOS has been an irritant.

- Combining ASCII and binary in one file not aesthetically pleasing.