The Neurophysiology Data Translation Format (NDF) Specification – V1.2.1

Document Ref:	The Neurophysiology Data Translation Format (NDF) Specification – V1.2.1
Author:	Bojian Liang, Jennifer Simonotto, Alastair Knowles, Martyn Fletcher
Date:	September 5, 2010
Pages:	42
Abstract:	The purpose of the Neurophysiology Data Translation Format (NDF) is to provide a means of sharing neurophysiology experimental data and derived data between services and tools developed within the CARMEN project (www.carmen.org.uk). This document specifies the NDF. The specification supports the types of data that are currently used by members of the CARMEN consortium and provides a capability to support future data types. It is capable of accommodating external data file formats as well as metadata such as user defined experimental descriptions and the history (provenance) of derived data.

Contents

1 Status of this document

This is V1.2.1 of the **N**europhysiology **D**ata Translation **F**ormat (NDF) Specification. This is the version that reflects all changes on the XML header during the NDF API implementation.

2 Introduction

CARMEN aims to create an environment for sharing neurophysiology experiment data and algorithms using distributed computing technology [Foster, Kesselman, 2004]. For this purpose, a variety of user contributed applications will be published, either from "wrapped" legacy implementations or from new implementations. CARMEN will need to cater for a wide range of incoming data types (from various acquisition systems, models, etc.). Analysis applications will also be executed on raw experimental data, model data and on derived data. There is a need for data translation to allow these applications to access a single, stable data format.

The NDF specifies a uniform file and format structure for data sharing and inter-application[1] data communication. In the first instance this will be implemented and provided within the CARMEN system. An NDF dataset consists of a configuration file which manages a variable number of host data files. The configuration file uses an XML format. It contains metadata and references to the associated host data files. Users and applications can embed additional information in the configuration file, for example to record the history of data processing activity, or to associate the dataset with other datasets. A client application can extract the metadata directly from the configuration file remotely without the need to look into the associated binary data file.

The NDF also specifies a set of the most commonly used experimental data entities as "NDF internal data types". These may be equally applicable to other signal or image processing applications, and include continuous time series, fixed/arbitrary length time series segments and spike times (e.g. events). NDF may be extended arbitrarily to accommodate other data entities. The applications developed within the CARMEN project will provide support for the NDF specified data types. Further, third party format data files can be attached to an NDF dataset as an "external format". This provides backward compatibility for non-CARMEN developed applications as well as allowing CARMEN applications to, at least, detect the format of the data even if they cannot provide full support for it. Examples of third party data include images and image series, and proprietary and bespoke time series data formats.

The MAT data format [Mat-File, 2008] is used as the main data format to encode the host data files, providing a generic and open framework for encoding numerical entities and arrays. To optimise storage space and network load, it is recommended that original digitized data input sets in integer type are stored as such, rather than converting them to floating point data types.

All the data files comprising one set of NDF data are, by default, placed in the same

[1] For brevity the term application is used throughout to indicate tools, services, software packages and applications.

directory as the appropriate configuration file, but if desired links to data in other locations can be specified using URIs in the configuration file. The NDF provides mechanism to store large data set to multiple data files as well as split a single channel to separated data files. As such, the NDF supports data size either a single channel or multiple channel that larger than 2GBs. Supported by the NDF APIs, splitting data to multiple files is transparent to users.

In this document, Section-3 specifies the NDF data types and their file formats. Section-4 specifies the format of the XML configuration files.

3 The NDF Data Types and the host data file format

For data storage purposes, an NDF data set may contain any type of data, managed by an NDF configuration file. However, the use of data types that are defined outside the CARMEN system (i.e. *external data formats*) may not have full support from the CARMEN applications. A set of "*CARMEN internal data types*" are defined for data sharing and inter-application communication within the CARMEN system. An *internal data type* is the data type that can be read/write by the NDF API without need the help of a third party package. This section will briefly look at the internal data types, support for external data, and annotation files defined as part of an NDF dataset for use in the CARMEN system.

3.1 The NDF internal data types

Technical Report [BL,JS, 2008] identified the data entities that will be acquired and created by user groups within the CARMEN project. In addition to time series data, these include images and image series. As image handling is broadly standardized, the NDF uses existing standard formats for image data. Images are therefore represented as third party data, specified by internal data type *ImageData*. Time series data are specified by the following types: continuous analog data (*TimeSeriesData*), fixed length waveform data segments (*SegmentData*) and neural event data (*NeuralEvent*). A further data type *ExperimentEventData* is defined for annotation/marker event data.

In addition, users can specify custom data types and associate instances of them with NDF datasets. Custom types may only be supported by a subset of the services within CARMEN. It is anticipated that custom data types may be fully incorporated by the NDF specification as they become prevalent. Table 3.1 lists the CARMEN internal data types and their host data file formats.

Table 1: NDF Internal Data Types

Data Type Name	Descriptions	Host data file format
ImageData	Image and image sequence data	Industrial standard
TimeSeriesData	One session waveform data	MATLAB Mat file
SegmentData	Waveform data segment	MATLAB Mat file
NeuralEventData	Such as spike time data	MATLAB Mat file
ExperimentalEventData	Event marker or annotation data	XML file, MATLAB Mat file
GenericMatrix	Reserved for internal use	MATLAB Mat file

CARMEN does not define new file formats for the storage of the raw source data, these are stored in their original proprietary or bespoke formats.

1. **ImageData** files are stored in the appropriate industry standard formats. When associated with an NDF dataset, metadata are extracted and copied into the configuration file. Image data are therefore fully defined by the host data file.

2. **TimeSeriesData** is the basic data type for storage of multichannel signal data. Data can be of any numeric data types defined by the associated MAT file. Metadata is stored in the associated configuration file. The host MAT data file may

or may not contain full metadata. This data type is therefore fully defined by the configuration file together with the host data files. Data element used within the host MAT file is a n-by-1 *numeric array* and stored as one channel per array. Channel names are mapped to a valid MatLab variable names as the matrix names (labels of the channel in a MAT file). Both the channel names and their MatLab Labels are stored in the configuration file.

3. **SegmentData** is the type used to store time-series data segments of one or more channels. For fixed length segment, data elements used within the host MAT file are in a 2-by-1 or 3-by-1 *cell array* with numeric matrix sub-elements. The first cell (a n-by-1 numeric matrix) represents time offset. The second cell (a m-by-n numeric matrix) represents n segments each with length m. The optional third cell (n-by-1 uint8 type matrix) represents the sorted ID. Data of the first two cells can be represented using any numeric data type. Other metadata are defined in the configuration file. For variable length segment, data elements are in a 3-by-1 or 4-by-1 cell array. The first cell (n-by-1 numeric matrix) represents index offset. The second cell (n-by-1 uint32 type matrix) represents end position of the n segments. The third cell is a L-by-1unmeric matrix with the n segments saved one after the other in sequence, where L is the number of data points of all the segments. The optional fourth cell (n-by-1 uint8 type matrix) represents the sorted ID. Figure-1 shows the layout of a variable length segment data set in a 4-by-1 cell. The data set with four segments, each with length 3, 5, 4 and 6 respectively. The second cell shows the end point of each segment as 3, 8, 12 and 18.

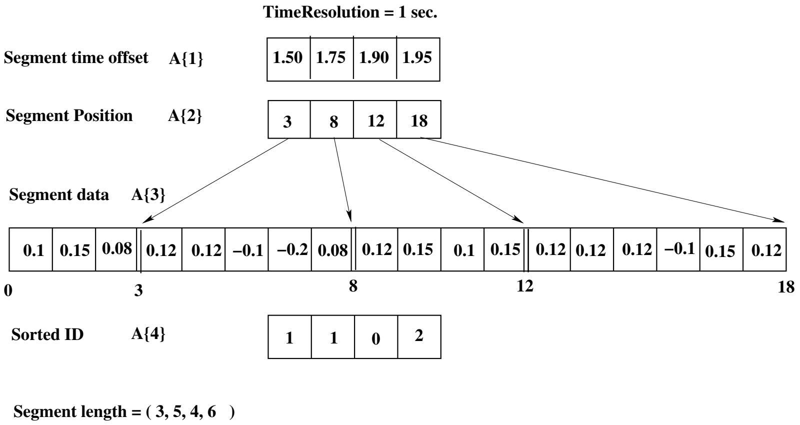

Figure 1: Variable Length Segment Data Layout

4. **NeuralEventData** is the data type used to store event data, where only time instances are of concern, such as spike time data. Neural event data can be encoded in any permissible numerical data type defined within the host MAT file. Metadata is stored in the associated configuration file. Data elements used within the host MAT file is a n-by-1 *numeric array* and stored as one channel per array. Array names (MatLab Labels) correspond to channel names. Both the channel names and their MatLab Labels are stored in the configuration file. The time unit of a neural event data channel must be in seconds.

5. **ExperimentalEventData** is the data type used to store “time stamp – annotation” or “time stamp – event values” data pair sequences. In the first instance, each pair identifies a time instance and attaches a short text comment or description of the event. The host file for this type is in XML format represented as *NDF Annotation Data*. In the second instance, the time stamp and event values are stored in a 2-by-1 cell array of a MAT file represented as *Binary Experimental Event Data*. For NDF annotation data, event data pairs are not necessarily ordered by time. Two data pairs that define the start and end time instances of a given event can be grouped together by the XML tab *interval*. Time stamps used in the annotation data must be consistent with the time course in the raw data. Details of the XML schema used by the host file can be found in Section-3.3. A binary experimental event data channel is in a 2-by-1 cell array. The two cells are both n-by-1 numeric matrices. The first cell is the time offset of the events and the second one is the event values. In both cases, unit of the time stamps must be in seconds.

6. **GenericMatrix** is a semi-defined data type. it provides a means of creating application specific data types during CARMEN application development. Generic matrix can be of any numeric data type supported by the MAT format. Instances are defined and used within CARMEN. Instances of *GenericMatrix* are signed by the identifier *applicationID* in the configuration file, which defines a group of applications that can read the host data file format and ’understand’ the data read.

3.2 NDF external data support

The NDF supports any third party data formats including image data files and the other vendor created data formats such as Multi-Channel System files (mcd). Such files are stored as part of the data set (in host data files) and the configuration file contains an entry in *UserDefinedData* to describe the data.

3.3 The Annotation data file format

The CARMEN internal data type *ExperimentalEventData* supports data annotation. The host annotation file contains event markers (time stamps), paired with descriptions of the events that took place. The major advantage of separating annotation data from raw data is that annotations can be applied to secondary data derived from raw data without any additional changes. Host files of *ExperimentalEventData* annotation data are encoded in XML, which can be generated programmatically and/or modified manually.

The host data file consists of one *description* element with multiple *eventNote* and/or *interval* elements which define the time instances at which events occurred.

- The *version* element provides the version number of the annotation file.
- The *description* element provides general information in text format about the annotation that the data represents.

- The *timeMarker* element defines the type of markers. When *timeMarker* is true, the time instances of the marked events are the time in seconds. Otherwise, the value of *timeMarker* represents the number of the frame of the image sequence.
- The *timeResolution* element defines the time resolution of the time offset at each time stamp in seconds. This value overrides the default parameter defined in the data configuration file – see Section-4.4.5. When the *timeMarker* element is set as false, this element is not meaningful.
- The *groupInfo* element provides additional description for each group of event data. The info text of in a *group* sub-element is applied to events that are with the same group id.
- The *eventNote* element contains four attributes and one text message.

  1. The *timeOffset* attribute represents the value of the time instance at which the event occurs. The unit of this attribute is defined by element *timeMarker*. This is a required attribute.
  2. The *group_id* attribute is optional and is reserved for identifying different event groups. This allows applications to only process events of specific types from the host data file.
  3. The *attachedFile* element allows an external data file (such as an audio recording of the sound of the event) to be attached to the event. This attribute is optional.
  4. The *application* attribute can only be specified when *attachedFile* is also defined. It defines a recommended application for opening the file defined by the *attachedFile* attribute. If *application* is not specified, the associated file will be opened by the system default application for the file type. *application* is therefore optional.

  An *eventNote* element can be empty, or can contain a text message to describe the event.

- The *interval* element contains two *eventNote* elements that specify the start and end time instances of the event in question. An optional *group_id* attribute can be applied to identifying events of different groups.

3.3.1 The Annotation data file XML elements example

This section explains the structure of the annotation data file in XML.

The *version* element declares the version number of the annotation file:

```
<version>1.0</version>
```

The *description* element contains a description of the annotation data:

```
<description>
    Description of the annotation data.
</description>
```

The *timeMarker* defines attribute *timeOffset* is time in seconds. Otherwise, *timeOffset* is the point index where the event happened.

```
<timeMarker>true</timeMarker>
```

The *timeResolution* element defines the time resolution of the marked time instances. This parameter should the same as the one defined in the NDF configuration file, or otherwise, overrides the one defined from the configuration file. For example, the following element defines a time resolution as 1 μs (10^{-6} second).

```
<timeResolution>0.000001</timeResolution>
```

If the *timeOffset* attribute is set as decimal value 12345678.99 and *timeMarker* is true, it represents that the time offset from the beginning of the data set is $12345678.99 \times 0.000001 = 12.34567899$ seconds. If *timeMarker* is false, the integer part (floor) of *timeOffset* is the index of the data point from the raw data where the event happened.

The following *groupInfo* element provides description to the event data, i.e. group 1 is the event data for video record and group 4 for audio data.

```
<groupInfo>
   <group id="01">Video record</group>
   <group id="04">Audio record</group>
</groupInfo>
```

The *interval* elements contains two *eventNote* elements that define the start and end time instances of the event:

```
<interval group_id="04">
  <eventNote timeOffset="1237888.23"
             attachedFile="sound1.wmv"
             application="realplayer">
    Text message for the event start.
  </eventNote>
  <eventNote timeOffset="18958585.23">
    Text message for the event end.
  </eventNote>
</interval>
```

An *eventNote* element must contain a *timeOffset* attribute. When *timeMarker* is set to false, *timeOffset* is an integer representing a frame number within an image series. As described in the previous section, the other three attributes (*attachedFile*, *application*, and *group_id*) are optional. *eventNote* can be an empty element, or can contain a descriptive text string:

```
<eventNote timeOffset="123456.789" group_id="04"
```

```
        attachedFile="sound1.wmv"
        application="realplayer">
   Pulse detected by the acquisition system
</eventNote>
```

These are also valid elements:

```
<eventNote timeOffset="123456.789"
        attachedFile="sound1.wmv" >
   Pulse detected by the acquisition system
</eventNote>

<eventNote timeOffset="123458.000" >
   Pulse detected by the acquisition system
</eventNote>

<eventNote timeOffset="123490.222" />
```

The last element is empty, illustrating that the *eventNote* marker can be used to specify the time instance of a predefined (or known by convention) event. The *eventNote* elements need not necessarily be ordered by time.

The following example demonstrates a complete annotation file combining the above components:

```
<?xml version="1.0" encoding="utf-8"?>
<NDTF_Annotation
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:schemaLocation="http://www.carmen.org.uk
                                   ndtfAnnotation.msd"
      xmlns="http://www.carmen.org.uk" >
  <description>
    A new annotation file.
  </description>
  <timeMarker>true</timeMarker>
  <timeResolution>0.000001</timeResolution>

  <interval group_id="04">
    <eventNote timeOffset="1237888.230"
               attachedFile="sound1.wmv"
               application="realplayer">
      Text message for the event start.
    </eventNote>
    <eventNote timeOffset="18958585.232">
      Text message for the event end.
    </eventNote>
  </interval>

  <eventNote timeOffset="123456.789" group_id="04"
```

```
        attachedFile="sound2.wmv"
        application="realplayer">
   Pulse detected by the acquisition system
  </eventNote>

  <eventNote timeOffset="123456.222"
                      attachedFile="sound1.wmv" >
       Pulse detected by the acquisition system
   </eventNote>

   <eventNote timeOffset="123458.000" >
      Pulse detected by the acquisition system
   </eventNote>
</NDTF_Annotation >
```

4 The NDF Configuration file

One set of NDF data consists of a configuration file (XML) and an arbitrary number of host data files. The configuration file contains metadata associated with the host data files. The configuration file can be considered as a separated header, providing a common means for applications to interpret the host data file(s). Within a data processing chain, the output of one application may become the input for another. The XML configuration file, which can be created and modified programmatically, provides a means to pass information about transformations that have been applied to data along the data processing chain. Only the most recent transformation need be described at any point in the chain, as earlier transformations can be inherited from the data associated with previous steps.

4.1 Structure of the configuration file

A NDF configuration file consists of one *Version* element, one *NdtfDataID* element and three sections: *GeneralInfo*, *DataSet*, and *History*. Each section is enclosed within XML tags. The *Version* element records the version of the NDF specification that has been used to structure the configuration file. The *NdtfDataID* is the identifier of the NDF data set. The *GeneralInfo* section describes the general information about the NDF data; the *DataSet* section contains metadata required for interpreting the included host data file(s), and the *History* section is used to record the data processing history. Figure-2 shows the tree diagram of a configuration file.

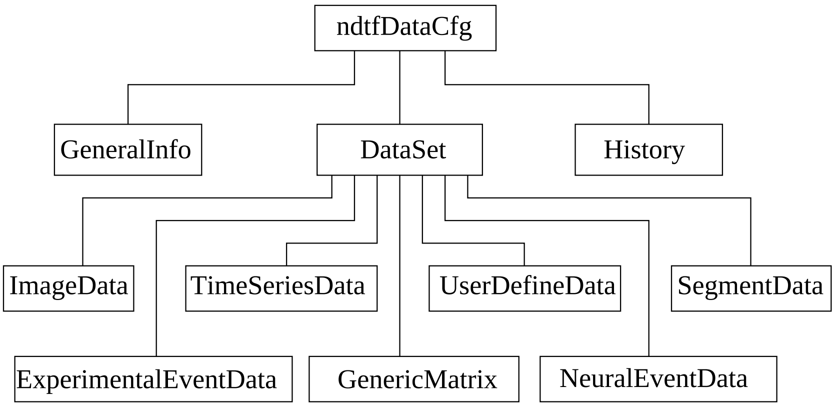

Figure 2: Configuration file tree diagram – summary

The following sub-sections describe the information that should be recorded inside the *GeneralInfo*, *DataSet*, and *History* sections.

4.2 General description element

This section provides general information about the NDF dataset. It contains seven sub-elements, some of which are optional. The sub-elements need not conform to any particular ordering.

- *Description* contains a short textual description of the data. This is a required element.
- *Laboratory* is a text string which can be used to record a name or identifier for the laboratory where the data was generated. This is an optional element.
- *Investigator* is a text string of the name of the investigator. It is an optional element.
- *SpecimenID* is a text string which can be used to record a name or identifier for the biological specimen that was used for the recording (e.g. animal or tissue section). This is an optional element.
- *CreateDate* is a text string in CCYY-MM-DD format describing the date when the dataset was generated. This is a required element.
- *CreateTime* is a text string in hh:mm:ss format describing the time when the dataset was generated. This is an optional element.
- *RecordID* is an optional element reserved for the storage of the data ID.

An sample *GeneralInfo* section is as follows:

```
<GeneralInfo>
    <Description>
         This is a NDF configuration file
       GeneralInfo section example.
    </Description>
    <Laboratory>CARMEN Lab</Laboratory>
    <Investigator>Smith Carmen</Investigator>
    <SpecimenID>anyname</SpecimenID>
    <CreateDate>2008-01-01</CreateDate>
    <CreateTime>12:38:38</CreateTime>
    <RecordID>abc123</RecordID>
</GeneralInfo>
```

4.3 DataSet element

The *DataSet* element is the core of the configuration file. It provides the information that applications require to interpret the dataset. The host data files contained in the dataset are defined as sub-elements of *DataSet*. Table-3.1 contains the six CARMEN internal data types. Within the *DataSet* element, each may be defined as a sub-element. *DataSet* can contain multiple sub-elements representing the same or different data types. For example, multiple *ExperimentalEventData* instances (e.g. annotations) may be created to describe different interpretations of the same data, or *ImageData* and *TimeSeriesData* instances may be combined to register concurrent recording modalities (e.g. imaging and multi-electrode array). Sub-elements within the *DataSet* section have an arbitrary order. Figure-2 shows the seven data types as the children of section *DataSet*.

In addition to the internal data types, an external data type *UserDefinedData* may be defined. This allows third party data format and other user defined extensions to be added, providing backward compatibility with proprietary or legacy software tools. The configuration file provides seamless access to these different representations based on the applications used to read the data. Users may arbitrarily define their own sub-elements within *UserDefinedData*.

While *UserDefinedData* may be extended by the user, two predefined methods are provided to access user defined datasets. One method defines an *applicationID* attribute, which can be used to specify an application or group of applications that can be used to access the data. The other defines a *recommendedApp* attribute, which can be used to specify a software package that is capable of accessing the data. This might be used to specify that a particular third-party software package provides support for the data format. If both *applicationID* and *recommendedApp* attributes are specified, *recommendedApp* is used by default.

Both *GenericMatrix* and *UserDefinedData* are semi-defined. Only the applications that defines (or is capable to use) the data can fully understand what the data is. However, *GenericMatrix* is the NDF internal data type and the NDF API directly supports data I/O from a *GenericMatrix* data record whilst the latter can only be read and written by specified applications.

4.3.1 The *DataInfo* element

The *DataInfo* element may be defined within several internal data types. Its contents and constraints are modified to make it specific. It contains metadata specific to neurophysiology experiment data such as sampling rate, filters and other equipment settings. Table-4.3.1 contains the full set of sub-elements for *DataInfo* and describes their respective constraints for different data types. The sub-elements of *DataInfo* are defined as follows:

1. *AcquisitionEquipment* is a text string allowing a name or other identifier (e.g. model number, device identifier) for the data acquisition system to be specified. It is enclosed by a pair of XML tags:

   ```
   <AcquisitionEquipment>
      Name of the acquisition equipment
   </AcquisitionEquipment>
   ```

2. *EquipmentSettings* is a short text string to describe the data acquisition system settings (only applicable if equipment is defined).

   ```
   <EquipmentSettings>
   ```

Table 2: The *Datainfo* sub-element and usages

Sub-elements of *DataInfo*	**NDF DataTypes**			
	ImageData	*TimeSeriesData*	*SegmentData*	*NeuralEventData*
AcquisitionEquipment	Optional	Optional	Optional	Optional
EquipmentSettings	Optional	Optional	Optional	Optional
StartDatetime	Required	Required	Required	Required
EndDateTime	Optional	Optional	Optional	Optional
NumberOfChannels	N/A	Required	Required	Required
ItemCount	N/A	Required	Required	Required
SamplingRate	N/A	Required	Required	Required
TransducerType	N/A	Optional	Optional	Optional
ADCSettings	N/A	Optional	Optional	Optional
LowPassFilter	N/A	Optional	Optional	Optional
HighPassFilter	N/A	Optional	Optional	Optional
ChannelLabels	N/A	Required	Required	Required
Trigger	N/A	N/A	Required	N/A

```
   Message for equipment settings
</EquipmentSettings>
```

3. *StartDateTime* defines the time instance of the start of data recording. This sub-element consists of two attributes: the data-time string in CCYY-MM-DDThh:mm:ss format, and the decimal part of the second. The latter is stored as a floating point number, allowing different resolutions to be used.

```
<StartDateTime DateTime="2008-01-28T18:03:28"
                              decimalSeconds="0.000031"/>
```

4. *EndDateTime* defines the time instance of the end of the recording in the same format as *StartDateTime*.

```
<EndDateTime DateTime="2008-01-28T18:04:28"
             decimalSeconds="0.000031"/>
```

5. *NumberOfChannels* is the number of channels in the recording, e.g.

```
 <NumberOfChannels>64</NumberOfChannels>.
```

6. *ItemCount* is the number of items in the recording. For NDF data types *NeuralEventData*, *SegmentData* and *ExperimentalEventData*, element *ItemCount* is a string in CSV format, each field represents the number of items in a relevant channel. For *TimeSeriesData*, all channels defined in one element should have the same number of items.
For *TimeSeriesData* or other data types with only one channel defined in the section, the XML element *ItemCount* looks like:

```
 <ItemCount>25000</ItemCount>.
```

For other data types with multiple channels defined in the section, it looks like:

```
 <ItemCount>25001,10000,3456</ItemCount>.
```

7. *SamplingRate* defines the data sampling rate in Hz.

   ```
   <SamplingRate>25000</SamplingRate>
   ```

8. *TransducerType* defines the type of the signal transducer. It is a text string and specified as follows:

   ```
   <TransducerType>active electrode</TransducerType>
   ```

9. *ADCSettings* describes the Analog-to-Digital conversion. Attribute *precision* is the precision of the Analog-to-Digital convertor, e.g. 12-bit. Attribute *zeroOffset* is the offset of the real zero level. *resolution* defines the quantization step. The unit of both *zeroOffset* and *resolution* is given by attribute *unit*. For example, given an Analog-to-Digital convertor of 12-bit precision, ADC resolution 0.00001221V and zero offset of -0.020V, the maximum input signal level will be $0.00001221 \times (2^{12} - 1) - 0.020 = 0.030$(V) and the minimum input signal level will be -0.020(V). If *ADCSettings* is not applicable, e.g. for a non-digitized signal, it may be omitted.

   ```
   <ADCSettings precision="12" zeroOffset="0.020"
                 resolution="0.0000121" unit="v"/>
   ```

10. *LowPassFilter* and *HighPassFilter* defines the settings for any low-pass and / or high-pass filters used. The attributes are self explanatory. *cutoffFrequency* is specified in Hz.

    ```
    <LowPassFilter cutoffFreqency="60000"
                 filterType="Chebyshev" order="10"/>
    <HighPassFilter cutoffFreqency="5.10"
                 filterType="Butterworth" order="10"/>
    ```

11. *ChannelLabels* defines a list of channel names in CSV (Comma Separated Value) format. The values in this list must correspond with the location of the data channels defined with the MAT file data element list defined on page 20 (*StructInfo*).

    ```
    <ChannelLabels>ch11, ch12, ch15, ch86</ChannelLabels>
    ```

12. *Trigger* defines the trigger threshold and time span of the triggered data segments (where applicable). Attribute *triggerType* defines the type of the trigger. *triggerType*=0, no trigger is used. *triggerType*=1, a signal level trigger is used. In this case, the unit of the threshold must match the unit of the signal values defined for the channel(s). *triggerType*=2, a slope based trigger is used. *triggerType*=3, a shape similarity trigger is used. Left-span and right-span are the signal time span on the left and right sides of the trigger instance in seconds.

    ```
    <Trigger triggerType="1" threshold="0.0003"
             leftSpan="0.0020" rightSpan="0.0025"/>
    ```

    In this example, the segment starts at $T - 0.0020$ and $T + 0.0025$ seconds, where T is the trigged time instance.

4.4 Data type definitions

4.4.1 ImageData

An *ImageData* section containing all the possible sub-elements looks like this:

```
<ImageData filename="imagefilename"  memberID="0">
    <Location>
      http://carmen/data/exampleData.mov
    </Location>
    <DataInfo>
       <AcquisitionEquipment>
         Name of the acquisition equipment
            </AcquisitionEquipment>
       <EquipmentSettings>
           Camera AGC turned off, Gamma=0.75
             </EquipmentSettings>
       <StartDateTime DateTime="2008-01-28T18:03:28"
                      decimalSeconds="0.000031"/>
       <EndDateTime DateTime="2008-01-28T19:03:28"
                    decimalSeconds="0.000031"/>
    </DataInfo>
    <Dimension width="100" height="288"/>
    <FrameInfo frameCnt="268" frameRate="15"/>
    <ExtraParams>
        <Params  name="pressure" value="10.1"  unit="pa"/>
        <Params  name="depth"    value="1000.01" />
    </ExtraParams>
    <RecommendedApp>imageviewer</RecommendedApp>
</ImageData>
```

- The *filename* attribute specifies the host data file name. By default, the host data file is in the same directory as the configuration file. Depending on the application, the host data file may be downloaded from the remote site defined by the *Location* parameter. Optional attribute *memberID* specifies a ID number to combine multiple data channels to one group. This is useful, for example, if raw data and processed data are stored in the same file, the member ID allows user to identify what the data is, even they use the same variable name.

- The *Location* sub-element defines the URI of the host data file wherever applicable. This need not be specified if the host data files are stored in the same directory as the configuration file.

- The *DataInfo* sub-elements are as defined in Section-4.3.1.

- The *Dimension* sub-element defines the dimensions of the image data (i.e. frame dimensions) in pixels.

- *FrameInfo* defines the total number of frames in the host file, and the frame rate (frames per second) if the host file is an image sequence.

- *ExtraParams* is optional. It defines a set of additional parameters that applied to the data set. Each parameter consists of attributes *name*, *value* and optional attribute *unit*. The number of parameters defined within the element is arbitrary.
- *RecommendedApp* is an optional sub-element. It provides the name of a recommend application to open the host data file.

4.4.2 TimeSeriesData

The *TimeSeriesData* sub-element defines one section of contiguous time-series recording. Both single and multiple channel recordings may be defined. When multiple channels of data are defined in one element, all the parameters except *ChannelLabels* and *MatElementLabels* must be the same. A section tnat defines multiple channels is called a compact section.

```
<TimeSeriesData filename="multipleChannelDataFilename"
                unit="mv" memberID="0">
  <Location>http://carmen/data/exampleData.mat</Location>
  <DataInfo>
     <AcquisitionEquipment>
        Name of the acquisition equipment
     </AcquisitionEquipment>
     <EquipmentSettings>
       Message for equipment settings
     </EquipmentSettings>
     <StartDateTime DateTime="2008-01-28T18:03:28"
                    decimalSeconds="0.000031"/>
     <EndDateTime DateTime="2008-01-28T19:03:28"
                  decimalSeconds="0.000031"/>
     <NumberOfChannels>4</NumberOfChannels>
     <ItemCount>25000</ItemCount>
     <SamplingRate>20000</SamplingRate>
     <TransducerType>active electrode</TransducerType>
     <ADCSettings precision="12" zeroOffset="0.0388"
                   resolution="0.000015" unit="mv"/>
     <LowPassFilter cutoffFreqency="60000"
                  filterType="Chebyshev" order="10"/>
     <HighPassFilter cutoffFreqency="5"
                  filterType="Butterworth" order="10"/>
     <ChannelLabels>
          ch 11, ch 12, ch 15, ch 86
     </ChannelLabels>
     <PositionList>1,1;1,2;1,3;1,4</PositionList>
  </DataInfo>
  <StructInfo>
     <MatElementLabels timeOffset="0.0000345">
        ch_11, ch_12, ch_15, ch_86
     </MatElementLabels>
     <ChildrenFiles elementID="0">
```

```
         <File startIndex="0" itemCount="1024"
               startTime="0.001" endTime="0.00279"
                                  filename="abc1.mat" >
             https://www.carmen.org/datastorage/0001185
        </File>
        <File startIndex="1024" itemCount="1024"
              startTime="0.0028" endTime="0.00459"
                                 filename="abc2.mat" >
            https://www.carmen.org/datastorage/0001186
        </File>
     </ChildrenFiles>
     <ChildrenFiles elementID="1">
         <File startIndex="0" itemCount="1024"
               startTime="0.001" endTime="0.00279"
                                 filename="xyz1.mat" >
            https://www.carmen.org/datastorage/0001195
         </File>
         <File startIndex="1024" itemCount="1024"
               startTime="0.0028" endTime="0.00459"
                                  filename="xyz2.mat" >
             https://www.carmen.org/datastorage/0001196
          </File>
    </ChildrenFiles>
  </StructInfo>
  <ExtraInfo>
      <InfoText  elementID="1">
          Plain text message provide extra information
          such as sample type, position, etc that are
          not defined from the DataInfo element for data
          element 1.
       </InfoText>
       <InfoText  elementID="2">
           Plain text message provide extra information
           such as sample type, position, etc that are not
           defined from the DataInfo element for data
           element 2.
       </InfoText>
   </ExtraInfo>
   <ExtraParams>
      <Params  name="pressure" value="10.1" unit="pa"/>
      <Params  name="depth"    value="1000.01" />
   </ExtraParams>
   <RecommendedApp>Signal Data Explorer</RecommendedApp>
</TimeSeriesData>
```

In addition to the (*filename*, *memberID*, *Location*, *RecommendedApp* and *DataInfo*) parameters defined already, the following parameters may be set within the *TimeSeriesData* section:

- Attribute *unit* defines the unit of the signal value. This is a required attribute, however, the attribute *ADCSettings::unit* will override this setting if both are defined.

- *StructInfo* provides information about the data location within the host data file. The sub-element *MatElementLabels* defines a list of MAT elements in which the relevant channels of data defined in the section *DataInfo::ChannelLabels* are stored. This mapping allows arbitrary names to be used for NDF data channel lables whilst valid MatLab variable names are used to save data to a MAT file. Each element in the list is separated by a comma. This list must correspond with the channel labels list defined in *DataInfo*. The attribute *timeOffset* within element *StructInfo* defines the time offset between the observed start of the recording and the first recorded data point. The NDF data set supports split data files, i.e. multiple host data files can be used to store one channel of data. This allows NDF data set to save data channel with item count larger than 2GB in a 32-bit system. If the split data files are in the local directory, the NDF API will automatically manage the data I/O for multiple data files. However, if the multiple data files are in the remote site, additional information must be applied in order for the client to know what data files are required to download. Each *ChildrenFiles* sub-section defines a set of data files that is belong to one data channel. *elementID* is a zero-based index of the channel name in list *ChannelLabel* that the *ChildrenFiles* applied. Each *File* sub-element defines a URI for a remote child data file. Attribute *startIndex* is the index of first data point of the child file referencing to the whole data channel. *itemCount* is the number of items in the child file. *startTime* and *endTime* define the time instance of the first data and last data points in the child file. These parameters can be used to fast locate the required child data file by either the index or time interval without need to look into all the binary child data files. *filename* is the filename of the child file.

- The *ExtraInfo* sub-element provides additional information about the data, in plain text. Attribute *elementID* is a zero-based index defined the channel that the info text applied. If the text applies to all channels defined in the *ChannelLabels* list , elementID should be set to -1.

- The *ExtraParams* sub-element provides additional parameter set for the data. Each sub-element within *ExtraParams* is with the same meaning as that in data type *ImageData*.

4.4.3 SegmentData

The *SegmentData* section defines the data structure of multiple channel, multiple segment time series data such as triggered MEA (Multi Electrode Array)data. Data is stored as a cell array within a MAT data file:

```
<SegmentData
  filename="multipleChannelTriggeredDataFilename"
  unit="mv" memberID="0"
  memberID="0"
  fixedLength="true">
```

```
<Location>
   http://carmen/data/exampleData.mat
</Location>
<DataInfo>
  <AcquisitionEquipment>
     Name of the acquisition equipment
  </AcquisitionEquipment>
  <EquipmentSettings>
     Message for equipment settings
  </EquipmentSettings>
  <StartDateTime DateTime="2008-01-28T18:03:28"
                 decimalSeconds="0.000031"/>
  <EndDateTime DateTime="2008-01-28T19:03:28"
              decimalSeconds="0.000031"/>
  <NumberOfChannels>4</NumberOfChannels>
  <ItemCount>2048,2048,1000,680</ItemCount>
  <SamplingRate>20000</SamplingRate>
  <TransducerType>active electrode</TransducerType>
  <ADCSettings precision="12" zeroOffset="0.0388"
                 resolution="0.000015" unit="mv"/>
  <LowPassFilter cutoffFreqency="60000"
                 filterType="Chebyshev" order="10"/>
  <HighPassFilter cutoffFreqency="5"
               filterType="Butterworth" order="10"/>
  <Trigger triggerType="1" threshold="0.0003"
             leftSpan="0.0025"  rightSpan="0.0025"/>
<ChannelLabels>ch11, ch12, ch15, ch86</ChannelLabels>
<PositionList>1,1;1,2;1,3;1,4</PositionList>
</DataInfo>
<StructInfo>
    <MatElementLabels timeOffset="0.0000345">
        Elm1, Elm2, Elm3, Elm4
    </MatElementLabels>
    <ChildrenFiles elementID="0">
        <File startIndex="0" itemCount="1024"
              startTime="0.001" endTime="0.00279"
                                  filename="abc1.mat">
            https://www.carmen.org/datastorage/0001185
        </File>
        <File startIndex="1024" itemCount="1024"
              startTime="0.0028" endTime="0.00459"
                                 filename="abc2.mat" >
            https://www.carmen.org/datastorage/0001186
        </File>
    </ChildrenFiles>
    <ChildrenFiles elementID="1">
        <File startIndex="0" itemCount="1024"
              startTime="0.001" endTime="0.00279"
                                filename="xyz1.mat" >
            https://www.carmen.org/datastorage/0001195
            </File>
            <File startIndex="1024" itemCount="1024"
                startTime="0.0028" endTime="0.00459"
                                   filename="xyz2.mat">
                https://www.carmen.org/datastorage/0001196
            </File>
        </ChildrenFiles>
    </StructInfo>
    <ExtraInfo>
        <InfoText sortedID="5" elementID="1">
            Plain text message provide extra information
            such as sample type, position, etc that are
            not defined from the DataInfo element for data
            element 1.
        </InfoText>
        <InfoText sortedID="8" elementID="2">
            Plain text message provide extra information
            such as sample type, position, etc that are
            not defined from the DataInfo element for data
            element 2.
        </InfoText>
    </ExtraInfo>
    <ExtraParams>
        <Params  name="pressure" value="10.1" />
        <Params  name="depth"    value="1000.01" />
    </ExtraParams>
    <RecommendedApp>Signal Data Explorer</RecommendedApp>
</SegmentData>
```

All parameters defined within section *SegmentData* have the same specification as *TimeSeriesData*, with the exception of attribute *fixedLength*, sub-elements *DataInfo::Trigger* and *DataInfo::ItemCount*. Optional attribute *fixedLength* provides information of the nature of the semgment length. *fixedLength="true"* indicates the data segment in the channel are with fixed length. With *fixedLength="false"*, the data segments in the channel are with variable length. If attribute *fixedLength* is not defined, it implies that a fixed length segment is presented. Details of XML sub-element *DataInfo::Trigger* can be found in Section-4.3.1. Instead of using a single number to identify the data item count as in *TimeSeriesData* where all channels have the same length, element *DataInfo::ItemCount* for segment data is a list of numbers in CSV format represented the number of segments in each channel. This list must corresponds to the channel name list *DataInfo::ChannelLabels*.

4.4.4 NeuralEventData

The *NeuralEventData* section defines the data structure of neural event data such as spike times. The values of this data type represent the time instances of events:

```
<NeuralEventData filename="spikeTimeDataFilename"
                 timeResolution="0.00001" memberID="0">
  <Location>http://carmen/data/exampleData.mat</Location>
  <DataInfo>
        <AcquisitionEquipment>
            Name of the acquisition equipment
        </AcquisitionEquipment>
        <EquipmentSettings>
            Message for equipment settings
        </EquipmentSettings>
        <StartDateTime DateTime="2008-01-28T18:03:28"
                       decimalSeconds="0.000031"/>
        <EndDateTime DateTime="2008-01-28T19:03:28"
                     decimalSeconds="0.000031"/>
        <NumberOfChannels>4</NumberOfChannels>
        <ItemCount>2048,2048,1000, 680</ItemCount>
        <SamplingRate>20000</SamplingRate>
        <TransducerType>active electrode</TransducerType>
        <ADCSettings precision="12" zeroOffset="0.0388"
                      resolution="0.000015" unit="mv"/>
        <LowPassFilter cutoffFreqency="60000"
                      filterType="Chebyshev" order="10"/>
        <HighPassFilter cutoffFreqency="5"
                    filterType="Butterworth" order="10"/>
        <ChannelLabels>
            ch11, ch12, ch15, ch86
        </ChannelLabels>
        <PositionList>1,1;1,2;1,3;1,4</PositionList>
  </DataInfo>
  <StructInfo>
         <MatElementLabels timeOffset="0.0000345">
             Elm1, Elm2, Elm3, Elm4
         </MatElementLabels>
         <ChildrenFiles elementID="0">
             <File startIndex="0" itemCount="1024"
                   startTime="0.001" endTime="0.00279"
                                       filename="abc1.mat">
                 https://www.carmen.org/datastorage/0001185
             </File>
             <File startIndex="1024" itemCount="1024"
                   startTime="0.0028" endTime="0.00459"
                                      filename="abc2.mat" >
                 https://www.carmen.org/datastorage/0001186
             </File>
         </ChildrenFiles>
         <ChildrenFiles elementID="1">
             <File startIndex="0" itemCount="1024"
                   startTime="0.001" endTime="0.00279"
                                    filename="xyz1.mat" >
                 https://www.carmen.org/datastorage/0001195
             </File>
             <File startIndex="1024" itemCount="1024"
                   startTime="0.0028" endTime="0.00459"
                                       filename="xyz2.mat">
                 https://www.carmen.org/datastorage/0001196
             </File>
         </ChildrenFiles>
  </StructInfo>
  <ExtraInfo>
       <InfoText sortedID="5" elementID="1">
           Plain text message provide extra information
           such as sample type, position, etc that are
           not defined from the DataInfo element for data
           element 1.
       </InfoText>
       <InfoText sortedID="8" elementID="2">
           Plain text message provide extra information
           such as sample type, position, etc that are
           not defined from the DataInfo element for data
           element 2.
       </InfoText>
  </ExtraInfo>
  <ExtraParams>
       <Params  name="pressure" value="10.1" />
       <Params  name="depth"    value="1000.01" />
  </ExtraParams>
  <RecommendedApp>Signal Data Explorer</RecommendedApp>
</NeuralEventData>
```

The parameters defined within *neuralEventData* have the same specifications as those in *TimeSeriesData*, with the exception of *timeResolution*, which is specific to this data type. *timeResolution* defines the time resolution in seconds of the neural event time instance values. For example, given a time resolution of 0.000005 seconds and an event time value of 100000, the real time value will be $100000 \times 0.000005 = 0.5$ seconds. *timeResolution* may or may be the same as the reciprocal of parameter *SamplingRate*.

4.4.5 ExperimentalEventData

As described previously, *ExperimentalEventData* with *recordType*=”Text” (annotations) are fully defined by their host data files (XML), generated programmatically or manually. The *ExperimentalEventData* element contains cursory information required for the interpretation of these data and links out to their respective file locations:

```
<ExperimentalEventData
       filename="AnnotationFilename"
       timeResolution="0.00001"
       memberID="0">
       <Location>
          http://carmen/data/datafilename
       </Location>
</ExperimentalEventData>
```

The *timeResolution* attribute defines the time resolution (in seconds) of the time offset markers defined in the host data file. This parameter may be overridden by the same parameter defined within the annotation file. The format of the annotation file is described in Section-3.3.

ExperimentalEventData with *recordType*=”Binary” indicates that the event data is stored in a MAT file in binary format. An example of *ExperimentalEventData* in binary type looks like this:

```
<ExperimentalEventData filename="BinaryExpEventDataFile.mat"
         recordType="Binary"
         timeResolution="0.00001"
         memberID="1">
    <Location>http://www.carmen.org.uk</Location>
    <BinaryEventData>
        <StartDateTime dateTime="2008-01-28T18:03:28"
                             decimalSeconds="0.000031"/>
        <NumberOfChannels>4</NumberOfChannels>
        <ItemCount>1180,1023,1000,680</ItemCount>
        <ChannelLabels>
            ch11, ch12, ch15,ch86
        </ChannelLabels>
        <MatElementLabels>
            Elm1, Elm2, Elm3,Elm4
        </MatElementLabels>
        <ChildrenFiles elementID="0">
            <File startIndex="0" itemCount="1024"
                  startTime="0.001" endTime="0.00279"
                                     filename="cde1.mat" >
                https://www.carmen.org/datastorage/0001185
            </File>
            <File startIndex="1024" itemCount="2048"
                  startTime="0.0028" endTime="0.00459"
                                     filename="cde2.mat" >
                https://www.carmen.org/datastorage/0001186
            </File>
        </ChildrenFiles>
    </BinaryEventData>
    <ExtraInfo>
        <InfoText  elementID="0">
            Plain text message provide extra information for
             experimental event for element 1.
        </InfoText>
        <InfoText  elementID="3">
            Plain text message provide extra information for
            experimental event for element 4.
        </InfoText>
     </ExtraInfo>
</ExperimentalEventData>
```

With attribute *recordType*=”Binary”, addition XML element *BinaryEventData* must attached. An optional element *ExtraInfo* can also be used to provide additional information about the data. Sub-elements of *BinaryData* are defined as follows:

1. *StartDataTime* defines the time instance of the start of data recording. This sub-element consists of two attributes: the data-time string in CCYY-MM-DDThh:mm:ss format, and the decimal part of the second. The latter is stored as a floating point number, allowing different resolutions to be used. Time stamps of the event instances are relevant values that start from date and time defined by this sub-element.
2. *NumberOfChannels* is the number of channels in the recording.
3. *ItemCount* is a string in CSV (Comma Separated Value) format and each field represents the number of items in a relevant channels.
4. *ChannelLabels* defines a list of channel names in CSV format. The values in this list must correspond with the location of the data channels defined with the MAT file data element list *MatElementLabels*.
5. *MatElementLabels* is a list with valid MatLab variable names that map to the *ChannelLabels* list.

Element *ExtraInfo* is defined as the same as that in *TimeSeriesData*.

4.4.6 GenericMatrix

This data type is defined as an extensible data type, for the purpose of CARMEN application development. The syntax of the data type is as follows:

```
<GenericMatrix filename="InternalDataFilename"
                     applicationID="appID"
                     unit="mA"
                     memberID="0">
     <Location>
         http://carmen/data/exampleData.mat
     </Location>
     <DataName>FFTSpectrum</DataName>
     <MatLabel>FFTSpectrum_MatLabel</MatLabel>
     <AppDefinedDataInfo>
        <DefineYourTagsHere>
            Parameters defined here describe
            the data structure.
        </DefineYourTagsHere>
```

```
    </AppDefinedDataInfo>
</GenericMatrix>
```

The *filename* attribute defines the host data file name. Attribute *applicationID* is the data type identifier, which in effect defines a group of applications that support the data format. A *GenericMatric* together with the *applicationID* defines a new application specific data type. The method for storing and referencing these applications is a system level consideration that is out of the scope of this document. The optional attribute *unit* defines the unit of the data values contained in the bespoke data, where this is applicable. Optional attribute *membarID* is used to integrate multiple data into group. If all *GenericMatrix* data is belong to the same group or it is not necessary to distinguish the group of data this attribute can be ignored. Element *DataName* and *MatLabel* define the name mapping between the data names and MAT file variable labels. *GenericMatrix* only supports single matrix per section. The key sub-element of this data type is *AppDefinedDataInfo*. Applications may define any number of XML elements and attributes within this element, to arbitrarily specify their dataset. For example, the *DefineYourTagsHere* element in the above example is a user defined element, which has been added to the *GenericMatrix* definition to create a custom type. This data type has two purposes. First, there may be some application specific datasets such as intermediate data that need to be shared within groups of applications but are not already defined as internal data types. By making use of *GenericMatrix* it will be possible to reuse these data types, making algorithm development and deployment more straightforward. Second, it provides a migration path for new CARMEN internal data types to be added to the NDF specification over time.

4.4.7 UserDefinedData

A *UserDefinedData* may be treated as the combination of *TimeSeriesData* and the *GenericMatrix* data type. It is used to reference a third party data file or application specified data that can't be represented by the NDF internal data type. The syntax of a *UserDefinedData* element is as follows,

```
<UserDefinedData filename="UserDefinedDataFilename"
                           applicationID="appID"
                           unit="mv"
                           memberID="0">
   <Location>http://carmen/data/dataname</Location>
   <DataInfo>
      <AcquisitionEquipment>
              Name of the acquisition equipment
      </AcquisitionEquipment>
      <EquipmentSettings>
         Message for equipment settings
      </EquipmentSettings>
      <StartDateTime DateTime="2008-01-28T18:03:28"
                     decimalSeconds="0.000031"/>
      <EndDateTime DateTime="2008-01-28T19:03:28"
                   decimalSeconds="0.000031"/>
      <NumberOfChannels>4</NumberOfChannels>
      <SamplingRate>20000</SamplingRate>
      <TransducerType>active electrode</TransducerType>
      <ADCSettings precision="12" zeroOffset="0.0388"
                    resolution="0.000015" unit="mv"/>
      <LowPassFilter cutoffFreqency="60000"
                     filterType="Chebyshev" order="10"/>
      <HighPassFilter cutoffFreqency="5"
                   filterType="Butterworth" order="10"/>
      <ChannelLabels>
          ch11, ch12, ch15, ch86
      </ChannelLabels>
      <PositionList>1,1;1,2;1,3;1,4</PositionList>
    </DataInfo>
    <ExtraInfo>
        <InfoText sourceID="2" elementID="1">
              Plain text message provide extra information
            such as sample type, position, etc that are not
            defined from the DataInfo element for data
            element 1.
        </InfoText>
        <InfoText sourceID="5" elementID="2">
              Plain text message provide extra information
            such as sample type, position, etc that are not
            defined from the DataInfo element for data
            element 2.
        </InfoText>
    </ExtraInfo>
    <ExtraParams>
        <Params  name="pressure" value="10.1" />
        <Params  name="depth"    value="1000.01" />
    </ExtraParams>
    <RecommendedApp>preferred app</RecommendedApp>
    <UserInfo>
       <DefineYourTagsHere>
             Define your data with your new tags.
       </DefineYourTagsHere>
     </UserInfo>
</UserDefinedData>
```

Because this is a user defined data type, the sub-elements *DataInfo* and *ExtraInfo* are all optional (in contrast to the typical *TimeSeriesData* element, where they are mandatory). The *UserInfo* element is similar to the *AppDefinedData* sub-element in *GenericMatrix*. Users can define an arbitrary range of XML tags here to specify the custom elements of their data type. This approach allows both CARMEN and third party applications to understand the data. For *UserDefinedData* the *unit* attribute is also optional.

4.5 History element

The *History* element provides a mechanism to record the data processing history. Each record is stored within a pair of *Processor* XML tags. It is recommended that all CARMEN applications copy this section from the input NDF data set to the output configuration file of the output data set and append their history to the end of the section as a new *Processor* sub-section. In this way, all datasets will be traceable within CARMEN.

```
<History>
   <Processor>
     <ProcessingDateTime
          StartDateTime="2008-01-28T18:03:28"
          EndDateTime="2008-01-28T18:03:30" />
     <CommandLine>
"spikedector.exe -f mydata.csd -a 5 -b 10 > this_data"
     </CommandLine>
     <ProcessingSettings>
        If it is not a command line application, use
        free form text message to describe the data
        processing parameters, application name,
        input filename and output filename, etc.
     </ProcessingSettings>
  </Processor>
</History>
```

The children of element *processor* are defined as follows,

- *ProcessingDateTime* defines the date/time instance when the process was initiated. The optional attribute *EndDateTime* can be used to record the date/time instance when the process concluded. This element is in plain text.
- *CommandLine* records the execution call(s) that invoked the process in plain text format whenever applicable. Command line recording is the recommended method for storing processing records. This makes it possible to read the processing history and apply the same process to other datasets automatically. More than one command line call may be recorded. Each is enclosed in quotes.
- *ProcessingSettings* is used when a command line record is not applicable. This sub-element allows the recording of a piece of script or a short text message to describe the data processing settings or description of the output data.

5 Summary

The NDF is a wrapped data format consisting of one configuration file and one or more host data files. The NDF specification does not define new formats for the host data files. Instead, established third party data formats that are open and provide mature tooling for data access (such as the MAT data format and XML) are used. The wrapped dataset allows applications in the CARMEN system to understand data and supports backward compatibility with third party tooling. The wrapped dataset contains an XML configuration file consisting of three sections. In addition to the *GeneralInfo* section, the *DataSet* section supports six predefined internal data types and one user defined data type which is extensible. The *History* section also provides a mechanism to retain knowledge of data processing steps along the processing chain. The NDF configuration file provides a rich set of information about the experiments. Being in XML format, it is possible to quickly extract and display the contents of an NDF configuration file of interest within a web page such as the CARMEN portal.

Appendices A-C provide XML templates for the annotation and configuration files, which can be used as examples to create new datasets. For more details of the NDF configuration file definitions, the XML schema files can be found on the CARMEN web site.

6 Technical Information

6.1 Authors

Bojian Liang
Department of Computing Science, University of York,
York, YO10 5DD United Kingdom
bojian@cs.york.ac.uk

Jennifer Simonotto
School of Computing Science, Newcastle University,
Newcastle upon Tyne, NE1 7RU United Kingdom
Jennifer.Simonotto@newcastle.ac.uk

6.2 Copyright Notice and License

References

[BL,JS, 2008] Liang, B. and Simonotto, J. (2008). Technical Report: Data Format in CARMEN V2.0 *http://metagenome.ncl.ac.uk/subversion/carmen/trunk/Documents*.

[Mat-File, 2008] The MathWorks, Inc.(2008). Mat-File Format 7.6 *http://www.mathworks.com/access/helpdesk/help/pdf_doc/matlab/matfile_format.pdf*.

[Foster, Kesselman, 2004] Foster, I. and Kesselman, C. (2004). The Grid: Blueprint for a new computing infrastructure (2nd ed.). Morgan Kaufmann Publishers, ISBN: 1-55860-475-8.

Appendices

A Annotation file XML file example

```
<?xml version="1.0" encoding="utf-8"?>
<NDTF_Annotation
       xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
       xsi:schemaLocation="http://www.carmen.org.uk ndtfAnnotation.xsd"
       xmlns="http://www.carmen.org.uk" >
  <version>1.0</version>
  <description>
    A new testing file
  </description>
  <timeMarker>true</timeMarker>
  <timeResolution>0.000001</timeResolution>
  <groupInfo>
     <group id="01">Vedio record</group>
     <group id="04">Audio record</group>
  </groupInfo>

  <eventNote timeOffset="1233445.23"
             attachedFile="sound1.wmv">event 1</eventNote>
  <eventNote timeOffset="1233445.23">event 1</eventNote>
  <eventNote timeOffset="1233445.23">event 1</eventNote>

  <interval group_id="01">
    <eventNote timeOffset="1237888.23"
               attachedFile="sound1.wmv"
               application="realplayer">
      Setup data
    </eventNote>
    <eventNote timeOffset="18958585.23">Stopped</eventNote>
  </interval>

  <interval group_id="04">
    <eventNote timeOffset="1237888.23"
               attachedFile="sound2.wmv"
               application="realplayer">
      Setup data
    </eventNote>
    <eventNote timeOffset="18958585.23"
               attachedFile="sound3.wmv"
               application="realplayer">
      Stopped
    </eventNote>
  </interval>

  <eventNote timeOffset="1233445.23">event 1</eventNote>
  <eventNote timeOffset="1233445.23">event 1</eventNote>
  <eventNote timeOffset="1233445.23">event 1</eventNote>
  <interval group_id="02">
    <eventNote timeOffset="1237888.23">
      Setup data
    </eventNote>
    <eventNote timeOffset="18958585.23">
      Stopped
    </eventNote>
  </interval>
</NDTF_Annotation>
```

B A simple NDTF data configuration file example

```
<?xml version="1.0" encoding="utf-8"?>
<ndtfDataCfg
     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
     xsi:schemaLocation="http://www.carmen.org.uk ndtfDataCfg.xsd"
     xmlns="http://www.carmen.org.uk">

  <Version>1.0</Version>
  <NdtfDataID>E06A305A-31AA-44B6-8B1A-178F4A748208</NdtfDataID>
  <GeneralInfo>
      <Description>Experiment 1 </Description>
      <Laboratory>CARMEN Lab</Laboratory>
      <Investigator>John Smith</Investigator>
      <SpecimenID>anyname</SpecimenID>
      <CreateDate>2008-01-01</CreateDate>
      <CreateTime>12:38:38</CreateTime>
      <RecordID>abe123</RecordID>
  </GeneralInfo>

  <DataSet>
     <NeuralEventData filename="spikeTimeData.mat"
                                timeResolution="0.00001"
                                memberID="0">
          <DataInfo>
              <StartDateTime dateTime="2008-01-01T18:03:28"
                             decimalSeconds="0.000031"/>
              <NumberOfChannels>4</NumberOfChannels>
              <ItemCount>2048,2048,1000, 680</ItemCount>
              <SamplingRate>20000</SamplingRate>
              <ChannelLabels>ch11, ch12, ch15, ch86</ChannelLabels>
          </DataInfo>
          <StructInfo>
              <MatElementLabels timeOffset="0.0000345">
                   ch11, ch12, ch15, ch86
              </MatElementLabels>
          </StructInfo>
          <ExtraInfo>
             <InfoText  sourceID="2" sortedID="5" elementID="-1">
                   Plain text message provide extra information such as
                sample type, position, etc that are not defined from the
                DataInfo element for data element 1.
              </InfoText>
          </ExtraInfo>
     </NeuralEventData>
  </DataSet>

  <History>
      <Processor>
          <ProcessingDateTime StartDateTime="2008-04-01T18:03:28"/>
          <CommandLine>
                spikedector.exe -f mydata.mcd -a 5 -b 10
          </CommandLine>
      </Processor>
  </History>
</ndtfDataCfg>
```

C Full NDTF data configuration file template

```
<?xml version="1.0" encoding="utf-8"?>
<ndtfDataCfg
     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
     xsi:schemaLocation="http://www.carmen.org.uk ndtfDataCfg.xsd"
     xmlns="http://www.carmen.org.uk">
  <Version>1.0</Version>
  <NdtfDataID>E06A305A-31AA-44B6-8B1A-178F4A748208</NdtfDataID>
  <GeneralInfo>
      <Description>Experiment 1 </Description>
      <Laboratory>CARMEN Lab</Laboratory>
      <Investigator>True Man</Investigator>
      <SpecimenID>anyname</SpecimenID>
      <CreateDate>2008-01-01</CreateDate>
      <CreateTime>12:38:38</CreateTime>
      <RecordID>abe123</RecordID>
  </GeneralInfo>

  <DataSet>
      <ImageData filename="imagefilename" memberID="0">
          <Location>Location URI</Location>
          <DataInfo>
              <AcquisitionEquipment>
                  Name of the acquisition equipment
              </AcquisitionEquipment>
              <EquipmentSettings>
                  Camera AGC turn off, Gamma=0.75
              </EquipmentSettings>
              <StartDateTime dateTime="2008-01-28T18:03:28"
                                 decimalSeconds="0.000031"/>
              <EndDateTime dateTime="2008-01-28T19:03:28"
                                 decimalSeconds="0.000031"/>
            </DataInfo>
          <Dimension width="100" height="288"/>
          <FrameInfo frameCnt="268" frameRate="15"/>
          <ExtraParams>
              <Params  name="pressure" value="10.1" unit="pa"/>
              <Params  name="depth"    value="1000.01" />
          </ExtraParams>
          <RecommendedApp>imageviewer</RecommendedApp>
      </ImageData>

      <TimeSeriesData filename="multipleChannelDataFilename"
                                     unit="mv" memberID="0">
          <Location>http://www.carmen.org.uk</Location>
          <DataInfo>
              <AcquisitionEquipment>
                  Name of the acquisition equipment
              </AcquisitionEquipment>
              <EquipmentSettings>
                  Message for equipment settings
              </EquipmentSettings>
              <StartDateTime dateTime="2008-01-28T18:03:28"
                                  decimalSeconds="0.000031"/>
              <EndDateTime dateTime="2008-01-28T19:03:28"
                                  decimalSeconds="0.000031"/>
              <NumberOfChannels>4</NumberOfChannels>
              <ItemCount>2048</ItemCount>
              <SamplingRate>20000</SamplingRate>
              <TransducerType>active electrode</TransducerType>
              <ADCSettings precision="12" zeroOffset="0.0388"
```

```
                      resolution="0.000015" unit="mv"/>
         <LowPassFilter cutoffFrequency="60000"
                        filterType="Chebyshev" order="10"/>
         <HighPassFilter cutoffFrequency="5"
                        filterType="Butterworth" order="10"/>
         <ChannelLabels>ch11, ch12, ch15, ch86 </ChannelLabels>
         <PositionList>1,1;1,2;1,3;1,4</PositionList>
     </DataInfo>
     <StructInfo>
         <MatElementLabels timeOffset="0.0000345">
              Elm1, Elm2, Elm3, Elm4
         </MatElementLabels>
         <ChildrenFiles elementID="0">
              <File startIndex="0" itemCount="1024"
                        startTime="0.001" endTime="0.00279"
                                     filename="abc1.mat" >
                    https://www.carmen.org/datastorage/0001185
              </File>
              <File startIndex="1024" itemCount="1024"
                        startTime="0.0028" endTime="0.00459"
                                     filename="abc2.mat" >
                    https://www.carmen.org/datastorage/0001186
              </File>
         </ChildrenFiles>
         <ChildrenFiles elementID="1">
              <File startIndex="0" itemCount="1024"
                        startTime="0.001" endTime="0.00279"
                                        filename="xyz1.mat" >
                    https://www.carmen.org/datastorage/0001195
              </File>
              <File startIndex="1024" itemCount="1024"
                        startTime="0.0028" endTime="0.00459"
                                       filename="xyz2.mat" >
                    https://www.carmen.org/datastorage/0001196
              </File>
         </ChildrenFiles>
     </StructInfo>
     <ExtraInfo>
         <InfoText  elementID="1">
            Plain text message provide extra information such as
         sample type, position, etc that are not defined from the
         DataInfo element for data element 1.
         </InfoText>
         <InfoText  elementID="2">
            Plain text message provide extra information such as
         sample type, position, etc that are not defined from the
         DataInfo element for data element 2.%
         </InfoText>
     </ExtraInfo>
     <ExtraParams>
         <Params  name="pressure" value="10.1" unit="pa"/>
         <Params  name="depth"    value="1000.01" />
     </ExtraParams>
     <RecommendedApp>Signal Data Explorer</RecommendedApp>
 </TimeSeriesData>

<SegmentData filename="multipleChannelTriggeredDataFilename"
                   unit="mv" memberID="0", fixedLength="true">
     <Location>http://www.carmen.org.uk</Location>
     <DataInfo>
         <AcquisitionEquipment>
              Name of the acquisition equipment
```

```
        </AcquisitionEquipment>
        <EquipmentSettings>
            Message for equipment settings
        </EquipmentSettings>
        <StartDateTime dateTime="2008-01-28T18:03:28"
                           decimalSeconds="0.000031"/>
        <EndDateTime dateTime="2008-01-28T19:03:28"
                           decimalSeconds="0.000031"/>
        <NumberOfChannels>4</NumberOfChannels>
        <ItemCount>2048,2048,1000, 680</ItemCount>
        <SamplingRate>20000</SamplingRate>
        <TransducerType>active electrode</TransducerType>
        <ADCSettings precision="12" zeroOffset="0.0388"
                        resolution="0.000015" unit="mv"/>
        <LowPassFilter cutoffFrequency="60000"
                      filterType="Chebyshev" order="10"/>
        <HighPassFilter cutoffFrequency="5"
                    filterType="Butterworth" order="10"/>
        <Trigger triggerType="1" threshold="0.0003"
                   leftSpan="0.0025" rightSpan="0.0025"/>
        <ChannelLabels>ch11, ch12, ch15, ch86</ChannelLabels>
        <PositionList>1,1;1,2;1,3;1,4</PositionList>
    </DataInfo>
    <StructInfo>
        <MatElementLabels timeOffset="0.0000345">
             Elm1, Elm2, Elm3, Elm4
        </MatElementLabels>
        <ChildrenFiles elementID="0">
             <File startIndex="0" itemCount="1024"
                       startTime="0.001" endTime="0.00279"
                                       filename="abc1.mat">
                   https://www.carmen.org/datastorage/0001185
             </File>
             <File startIndex="1024" itemCount="1024"
                      startTime="0.0028" endTime="0.00459"
                                       filename="abc2.mat" >
                   https://www.carmen.org/datastorage/0001186
             </File>
        </ChildrenFiles>
        <ChildrenFiles elementID="1">
             <File startIndex="0" itemCount="1024"
                        startTime="0.001" endTime="0.00279"
                                       filename="xyz1.mat" >
                   https://www.carmen.org/datastorage/0001195
             </File>
             <File startIndex="1024" itemCount="1024"
                         startTime="0.0028" endTime="0.00459"
                                          filename="xyz2.mat">
                   https://www.carmen.org/datastorage/0001196
             </File>
        </ChildrenFiles>
    </StructInfo>
    <ExtraInfo>
        <InfoText sortedID="5" elementID="1">
           Plain text message provide extra information such as
        sample type, position, etc that are not defined from the
        DataInfo element for data element 1.
        </InfoText>
        <InfoText sortedID="8" elementID="2">
           Plain text message provide extra information such as
        sample type, position, etc that are not defined from the
        DataInfo element for data element 2.
        </InfoText>
    </ExtraInfo>
    <ExtraParams>
        <Params  name="pressure" value="10.1" />
        <Params  name="depth"    value="1000.01" />
    </ExtraParams>
  <RecommendedApp>Signal Data Explorer</RecommendedApp>
</SegmentData>

<NeuralEventData filename="spikeTimeDataFilename"
           timeResolution="0.00001" memberID="0">
    <Location>http://www.carmen.org.uk</Location>
    <DataInfo>
        <AcquisitionEquipment>
            Name of the acquisition equipment
        </AcquisitionEquipment>
        <EquipmentSettings>
            Message for equipment settings
        </EquipmentSettings>
        <StartDateTime dateTime="2008-01-28T18:03:28"
                           decimalSeconds="0.000031"/>
        <EndDateTime dateTime="2008-01-28T19:03:28"
                           decimalSeconds="0.000031"/>
        <NumberOfChannels>4</NumberOfChannels>
        <ItemCount>2048,2048,1000, 680</ItemCount>
        <SamplingRate>20000</SamplingRate>
        <TransducerType>active electrode</TransducerType>
        <ADCSettings precision="12" zeroOffset="0.0388"
                        resolution="0.000015" unit="mv"/>
        <LowPassFilter cutoffFrequency="60000"
                      filterType="Chebyshev" order="10"/>
        <HighPassFilter cutoffFrequency="5"
                    filterType="Butterworth" order="10"/>
        <ChannelLabels>ch11, ch12, ch15, ch86</ChannelLabels>
        <PositionList>1,1;1,2;1,3;1,4</PositionList>
    </DataInfo>
    <StructInfo>
        <MatElementLabels timeOffset="0.0000345">
             Elm1, Elm2, Elm3, Elm4
        </MatElementLabels>
        <ChildrenFiles elementID="0">
             <File startIndex="0" itemCount="1024"
                      startTime="0.001" endTime="0.00279"
                                    filename="abc1.mat" >
                   https://www.carmen.org/datastorage/0001185
             </File>
             <File startIndex="1024" itemCount="1024"
                        startTime="0.0028" endTime="0.00459"
                                       filename="abc2.mat" >
                   https://www.carmen.org/datastorage/0001186
             </File>
        </ChildrenFiles>
        <ChildrenFiles elementID="1">
             <File startIndex="0" itemCount="1024"
                       startTime="0.001" endTime="0.00279"
                                      filename="xyz1.mat" >
                   https://www.carmen.org/datastorage/0001195
             </File>
             <File startIndex="1024" itemCount="1024"
                       startTime="0.0028" endTime="0.00459"
                                        filename="xyz2.mat" >
                   https://www.carmen.org/datastorage/0001196
             </File>
        </ChildrenFiles>
    </StructInfo>
    <ExtraInfo>
        <InfoText  sourceID="2" sortedID="5" elementID="1">
           Plain text message provide extra information such as
        sample type, position, etc that are not defined from the
        DataInfo element for data element 1.
        </InfoText>
        <InfoText  sourceID="6" sortedID="8" elementID="2">
           Plain text message provide extra information such as
        sample type, position, etc that are not defined from the
        DataInfo element for data element 2.
        </InfoText>
    </ExtraInfo>
    <ExtraParams>
        <Params  name="pressure" value="10.1" />
        <Params  name="depth"    value="1000.01" />
    </ExtraParams>
  <RecommendedApp>Signal Data Explorer</RecommendedApp>
</NeuralEventData>


<ExperimentalEventData filename="AnnotationFilename"
                       recordType="Text"
                       timeResolution="0.00001"
                       memberID="0">
     <Location>http://www.carmen.org.uk</Location>
</ExperimentalEventData>

<ExperimentalEventData filename="BinaryExpEventDataFile.mat"
                       recordType="Binary"
                       timeResolution="0.00001"
                       memberID="1">
     <Location>http://www.carmen.org.uk</Location>
     <BinaryEventData>
        <StartDateTime dateTime="2008-01-28T18:03:28"
                                decimalSeconds="0.000031"/>
        <NumberOfChannels>4</NumberOfChannels>
        <ItemCount>1180,1023,1000, 680</ItemCount>
        <ChannelLabels>ch11, ch12, ch15, ch86</ChannelLabels>
        <MatElementLabels>Elm1, Elm2, Elm3, Elm4</MatElementLabels>
        <ChildrenFiles elementID="0">
             <File startIndex="0" itemCount="1024"
                       startTime="0.001" endTime="0.00279"
                                      filename="cde1.mat" >
                   https://www.carmen.org/datastorage/0001185
             </File>
             <File startIndex="1024" itemCount="2048"
                       startTime="0.0028" endTime="0.00459"
                                        filename="cde2.mat" >
                   https://www.carmen.org/datastorage/0001186
             </File>
        </ChildrenFiles>
     </BinaryEventData>
     <ExtraInfo>
         <InfoText  elementID="0">
            Plain text message provide extra information for
            experiment event for element 1.
         </InfoText>
         <InfoText  elementID="3">
            Plain text message provide extra information for
            experiement event for element 4.
         </InfoText>
      </ExtraInfo>
</ExperimentalEventData>

<GenericMatrix filename="InternalDataFilename"
               applicationID="1EEE48EE-9A46-4CCD-8D54-9227188AEC08"
               unit="mA"
               memberID="0">
     <Location>http://www.carmen.org.uk</Location>
     <DataName>FFTSpectrum</DataName>
     <MatLabel>FFTSpectrum_MatLabel</MatLabel>
     <AppDefinedDataInfo>
         <DefineYourTagsHere>
             Place parameters within the application defined tags
         </DefineYourTagsHere>
     </AppDefinedDataInfo>
</GenericMatrix>

<UserDefinedData filename="UserDefinedDataFilename"
                 applicationID="AC953D7C-362C-4C28-B050-81C3F62CE12E"
                 unit="mv"
                 memberID="0">
     <Location>http://www.carmen.org.uk</Location>
     <DataInfo>
        <AcquisitionEquipment>
             Name of the acquisition equipment
        </AcquisitionEquipment>
        <EquipmentSettings>
             Message for equipment settings
        </EquipmentSettings>
        <StartDateTime dateTime="2008-01-28T18:03:28"
                              decimalSeconds="0.000031"/>
        <EndDateTime dateTime="2008-01-28T19:03:28"
                              decimalSeconds="0.000031"/>
        <NumberOfChannels>128</NumberOfChannels>
        <SamplingRate>20000</SamplingRate>
        <TransducerType>active electrode</TransducerType>
        <ADCSettings precision="12" zeroOffset="0.0388"
                        resolution="0.000015" unit="mv"/>
        <LowPassFilter cutoffFrequency="60000"
                      filterType="Chebyshev" order="10"/>
        <HighPassFilter cutoffFrequency="5"
                    filterType="Butterworth" order="10"/>
        <ChannelLabels>ch11, ch12, ch15, ch86</ChannelLabels>
        <PositionList>1,1;1,2;1,3;1,4</PositionList>
     </DataInfo>
    <ExtraInfo>
        <InfoText sourceID="2" elementID="1">
           Plain text message provide extra information such as
        sample type, position, etc that are not defined from the
        DataInfo element for data element 1.
        </InfoText>
        <InfoText sourceID="5" elementID="2">
           Plain text message provide extra information such as
        sample type, position, etc that are not defined from the
        DataInfo element for data element 2.
        </InfoText>
    </ExtraInfo>
    <ExtraParams>
        <Params  name="pressure" value="10.1" />
        <Params  name="depth"    value="1000.01" />
```

```
          </ExtraParams>
           <RecommendedApp>preferred app</RecommendedApp>
           <UserInfo>
               <DefineYourTagsHere>
                   Defined your new tags and parameters here.
               </DefineYourTagsHere>
           </UserInfo>
      </UserDefinedData>
  </DataSet>

  <History>
      <Processor>
          <ProcessingDateTime StartDateTime="2008-01-28T18:03:28"
                              EndDateTime="2008-01-28T18:03:30"/>
          <CommandLine>
               spikedector.exe -f mydata.mcd -a 5 -b 10
          </CommandLine>
          <ProcessingSettings>
              If it is not a command line application, use free
              form text message to describe the data processing
              parameters, application name,input filename and
              output filename, description, etc.
          </ProcessingSettings>
      </Processor>
  </History>

</ndtfDataCfg>
```