A Quick Guide To XCEDE 1

This is a brief description of **XCEDE 1** (**X**ML-Based **C**linical **E**xperiment **D**ata **E**xchange schema). The schema provides a hierarchical means for storage and sharing of biomedical data, including images, and analysis annotations. Typical sources and targets for **XCEDE** data transfer include Web services, databases, data browsers, and data analysis tools.

The quick guide by no means replaces the extensive documentation available on the schema. For more information about **XCEDE** please visit:

http://www.xcede.org/
http://www.nbirn.net/tools/xcede/index.shtm

Schema Overview

The **XCEDE** schema provides data on a project, studies within that project, subjects involved in them, visits of all the subjects, as well as a detailed description of every visit. Its root element is <projectlevel>. Whenever a small portion of the schema is needed, a <spliceFrom> element is used to incorporate data from other levels. As for the root element, the following data is provided:

<projectlevel>

<ID>: a unique string identifier.
<description>, <funding>: free-form text description of the project and the funding sources respectively.
<provenance>, <statistics>, <annotations>: common to each level and will be explained at a later time.
<subject>: spliced in from <subjectlevel>.

<subjectlevel>

The <subjectlevel> data does not vary between visits. It is spliced in from the <projectlevel> and includes:

<ID>: a unique string identifier for the subject.
<birthDate>, <deathDate>: the dates of the subject's birth and death in the ISO 8601 form.
<name> : the subject's full name.
<sex>: one of *male, female* or *other*
<species>: the subject's species described with its common name, latin name, strain and a free-form text description (<commonName>, <latinName>, <strain>, <description>).
<extendedDescriptor>: contains name, value, description and annotation fields. It may appear more than once.
<provenance>, <statistics>, <annotations>
<visit>: spliced in from <visitlevel>.

<visitlevel>

<visitlevel> is a top-level (root) element, which encapsulates data for a subject visit, such as assessments performed before the visit and the sequence of elements detailing each study. The children of the <visitlevel> element are:

<project>, <subject>: elements spliced in from the corresponding higher-level elements.
<ID>: a unique string identifier.
<location>: the location for this visit.
<subjectVar>: the assessments gathered during a subject scan. It has the following children

- **<age>:** floating point number which indicates the age of the number at the time of the visit.
- **<assessment>:** assessments that appear more than once (<TaxonomyClass>, <name>, <description>, <assessmentValue>).
- **<annotation>**

<study>: spliced in from <studyLevel>.
<provenance>, <statistics>, <annotations>

<studylevel>

<studylevel> is a top-level (root) element which encapsulates data for a group of scans. The children of the <studylevel> element are:

<project>, <subject>, <visit>: elements spliced in from the corresponding higher-level elements.
<series>: spliced in from <serieslevel>.
<provenance>, <statistics>, <annotations>

<serieslevel>

<serieslevel> is a top-level element and its purpose is to accommodate data for series of images that were acquired from one medical scanning device during a study. Its children are:

<project>, <subject>, <visit>, <study>: elements spliced in from the corresponding higher-level elements.
<scanner>: contains details about the scanning device (<manufacturer>, <model>, <additional Equipment> and <annotation>).
<expProtocol>: the events and conditions that occurred during the scan, for example, the events that were presented to the subject (<ID>, <taxonomicClass>, <name>, <conditionDescriptor>, <annotation>).
<acqProtocol>: the scanner acquisition parameters that were used in the scan (<ID>, <taxonomicClass>, <name>, <acqParam>, <annotation>).
<datarec>
<provenance>, <statistic>, <annotation>

<datarec>

<datarec> is a data interface describing how to read/interpret encapsulated files. This element can

appear more than once, but must have a unique type/ subtype pair. Its immediate children are:

<rasorigin>: If the datarec represents an image, it provides the location in patient coordinates of the first data element (pixel, voxel).
<dimension>: stores information about one of the N dimensions in the data record. Each dimension has the following children, <size>, <origin>, <spacing>, <gap>, <datapoints>, <direction>, <units>, <type>.
<byteorder>: describes whether the individual data elements are stored with the most-significant byte first or least-significant-byte first.
<element type>: describes the type of individual data elements in the data record.
<datarec fragment>: The external data pointed to by this data record are represented as a list of "fragments", that are defined as streams of data contiguously stored in the same file. Data stored in multiple files necessitates multiple fragments. It consists of <filename>, <fileoffset> and <filerecordsize>.
<description>: is optional free-form text describing the data record.
<provenance>, <statistic>, <annotation>

Common children of top-level elements

<provenance>: describes a pipeline of processing steps that was used in order to generate some subset of the data in an element. Every <provenance> element has an <ID>, while the info of every processing step is given in a <processStep> child.
<statistic>: stores basic statistical information, which was calculated from the data of an element. Its immediate children are <sourceData>, <process>, <provenanceID>, <description>, <annotation>. It can also be extended with the addition of more specialized children.
<annotation>: includes the annotation data for its parent element. It has three children: <text>, <annotator> and <timeStamp>.

Usage and Support Tools

With a definite shift towards data centric applications, XML databases are gaining popularity. Using **XCEDE** format for storing biomedical data makes moving to such databases intuitive and easy.

Next step in the evolution of distributed computing, extensively used in biomedical applications, is to move it over the internet. The fundamental building blocks for this are XML Web services, clearly applications using **XCEDE** format can easily adapt to this change.

BXH XCEDE Tools

This is a suit of tools designed to read, write and manipulate XML descriptors using BHX or **XCEDE** schemas. These descriptors wrap the original image data files and provide a consistent and portable interface to image and other meta data extracted from images. The tools suit can be downloaded from
http://www.nbirn.net/tools/bxh_tools/index.shtm

SPM XML Toolbox

This is a MATLAB add on toolbox available for SUN, LINUX and Microsoft Windows operating systems. The toolbox supports both SPM99 and SPM2 statistical structures and has been used to capture PET and fMRI analysis results and the associated analysis model specifications. It has an intuitive GUI interface for configuring results capturing and output formatting. The toolbox can be downloaded from
http://www.nbirn.net/tools/xcede_spm/index.shtm