

An XML-based Data Access Interface for Image Analysis and Visualization Software

Syam Gadde, Charles R. Michelich, James Voyvodic
Duke-UNC Brain Imaging and Analysis Center (BIAC)
Duke University, Durham, NC

Constructing a robust image analysis pipeline from several, often disparate, software elements is not a trivial process. Many valuable software tools have arbitrary data format requirements. Combining such software tools in analysis requires data conversion steps that may slow down processing and introduce errors.

We have defined a general data access interface in the form of a structured text file, written in XML (Extensible Markup Language). This XML file provides instructions on how to extract image data from files in any uncompressed data format, without explicit conversion of data. This interface supports data of arbitrary dimensionality, data type, and byte order, and is applicable to most MR image formats, physiological and stimulus data. The XML file is placed alongside the data it "encapsulates", and points to the raw data files by reference.

The XML markup language provides several advantages over simpler plain text formats. Because XML is merely structured text, it is easily parsed and used as a common access interface for image dimensions, subject information, acquisition parameters, or transformations between image and world coordinate systems. The XML structure itself is extensible using XML namespaces, facilitating the integration of additional metadata modules.

We have created subroutine libraries to provide software support for the XML file format for C, C++, and MATLAB. These libraries employ XML tools based on platform- and language-independent standards such as XPath and DOM. Incorporating these library tools into existing analysis programs allows those programs to read images of many types (e.g. DICOM, GE Signa 5.x, Analyze 7.5/SPM, MINC, etc.) via the XML file, and makes them forward-compatible with future data formats that can be encapsulated within this XML specification. Conversely, any image files with an associated XML file, regardless of the underlying format, are automatically readable by all XML-aware applications that support this interface. The XML interface allows existing software to take advantage of the additional information and data formats the metaheader supports. For example, our analysis tools that formerly required users to manually specify orientation, position, and image dimensions now obtain this information from the XML files.

We also use this approach to support legacy software. Using an XML-based metaheader as the common intermediary greatly simplifies our conversion software; conversions between N formats require only 2N software components, rather than the N^2 required by a direct format-to-format approach. Moreover, the XML file maintains the image metadata we might otherwise lose when converting from a richly annotated data format (e.g. DICOM) to one that is sparsely annotated (e.g. Analyze 7.5).

This general framework, implemented as the BIAC XML Header (BXH), has been in wide use at the Duke-UNC Brain Imaging and Analysis Center since late 2002. It has also been adopted as a component of the XML schema developed by the Biomedical Informatics Research Network (BIRN).