

W0165

NeXus With Binary XML - Efficient Support of Binary Data in the XML Implementation of the NeXus File Format. Georgi Darakev, Vassil Litchev, Kostadin Z. Mitev, Herbert J. Bernstein, Dept. of Mathematics and Computer Science, Dowling College, Oakdale, NY 11769 USA.

There are two formats for the management of synchrotron image data under serious consideration to facilitate moving raw image data among beam lines and user home institutions: imgCIF and NeXuS. imgCIF provides a very efficient representation of large binary image data in its native binary form. NeXuS is equally efficient in internal storage and in transfer as an image of its internal HDF4/5 format, but to date it lacks an agreed efficient external representation. Use of SOAP serialization of the internal objects has been proposed, but that imposes a serious performance penalty in working with XML, since XML does not directly support a binary representation. For the smaller volumes of data seen in collecting neutron data, this has not been a serious problem. For high-data-rate synchrotron data collection, efficiency is more important. We propose a new, reasonably efficient format with only a 1 part in 15 overhead for encoding as printable characters, and have augmented the NeXuS and CBFlib APIs to support this format. These augmented APIs support encoding and decoding of binary data thereby producing valid UTF-16 XML documents, or if desired, mixed UTF-8/UTF-16 XML documents by use of non-character codepoints as control flags (extending the UTF-16 BOM convention). The availability of these APIs should contribute to achieve seamless transitions among imgCIF, NeXuS, imgML (an XML presentation of imgCIF) and SOAP serializations of NeXuS data and internal NeXuS file formats.