

edikt::BinX

www.edikt.org/binx

Binary XML

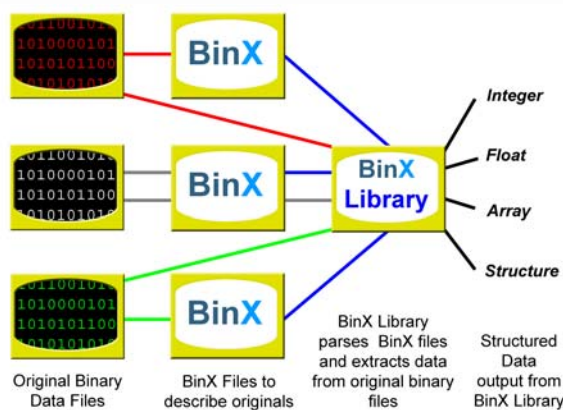
edikt

The datasets for many scientific users are stored in very large binary files, often one or more large arrays or tables. It is possible to represent such data in XML, but:

- in XML, the data file would be around 2-4 times larger than the binary representation, taking longer to write, transport etc.
- the proposed representation for a multidimensional array in XML is effectively a tree of lists - unhelpful for scientific users who often want to extract slices of data which cut across the tree.

However, there is enormous value, and interest, in representing the *metadata* associated with the data in XML. Metadata will typically describe such things as how the data was produced (parameters, algorithms used etc), when and by whom. It would be very useful if that metadata also contained a description of the *structure* and *representation* of the data itself. BinX is a software tool that addresses this need.

The BinX Library is being developed by edikt to allow the reading and writing of BinX XML schema files and the associated binary data files (see figure 1, below).



The library has functionality to:

- browse - read data from binary files;
- extract - select partial data from the dataset;
- transform - use XSL to transform the structure of binary data;
- create - create binary data and the associated BinX XML schema descriptions.

Case study:

BinX in the Virtual Observatory

To manage their increasing data volumes, astronomers are creating a global Virtual Observatory (VO), with goals to federate the world's astronomical data, provide resources for its exploitation and convenient access for all astronomers.

The interoperability of data formats is a key problem in the VO, greatly aided by XML and the astronomical standard VOTable schema.

However, XML is too verbose, and astronomers have a lot of data! A large amount of data is also stored in FITS, a binary format for storing images, and a problem arises: astronomers want their data in VOTable some of the time and FITS the rest. The answer is BinX.

BinX allows conversion between FITS and VOTable! edikt has conducted successful experiments in converting FITS to VOTable, and is currently working on the harder problem of VOTable to FITS. Thanks to edikt, BinX will find its way into the VO infrastructure, enabling scientists to make new discoveries in multi-wavelength astronomy.

BinX