

edikt

Effective management of the flood of data becoming accessible on Grids throughout the world will only be possible if we can provide it with meaningful interpretations. To do this, **edikt** believes that linkage between computer and applications scientists is vital.

edikt, based at the National e-Science Centre in Edinburgh, is a software engineering group aiming to build these links, producing robust, scalable and usable data management tools that enable new research areas in e-Science.

edikt's goals are to:

- transfer existing research into use in challenging scientific applications, thereby spinning off new research projects as theories are tested;
- work with companies to exploit existing industrial technologies and extend functionality;
- encourage commercialisation of novel techniques developed.

edikt is funded initially by the Scottish Higher Education Funding Council under their largest Research Development Grant to date. This underpinning funding allows **edikt** to develop generic spin-off technologies that may have commercial applications in Scotland and beyond in areas such as drug discovery, financial analysis and agricultural development.

For this reason, **edikt** has a dedicated commercialisation manager who can push out the knowledge, expertise and software of **edikt** to industry and business.

Realms of inquiry

Initially, **edikt** is addressing e-Science problems in the areas of

- astronomy – looking at data interoperability and portability;
- particle physics – looking at data replication technologies and metadata;
- bioinformatics – looking at data integration, interoperability and accessibility;
- virtual organisations – investigating the general technologies needed to integrate individual organisations into a single virtual entity.

From these areas – and others – **edikt** draws together common requirements, distils common problem areas and engineers generic solutions based on cutting edge computer science.

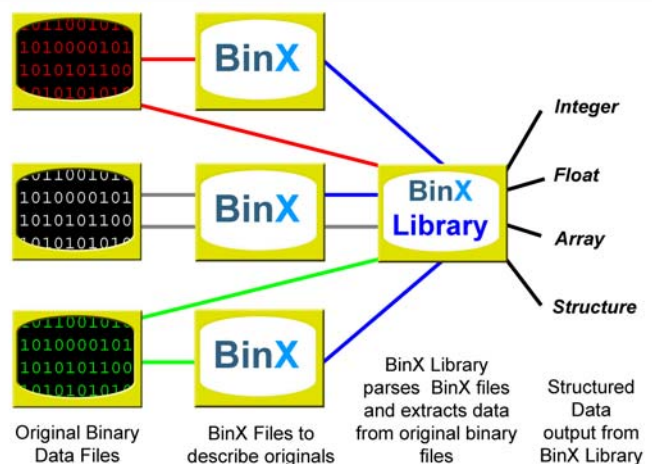
edikt::BinX

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 an XML Schema 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).



edikt:: BinX

Continued

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.

(**edikt::Eldas**, *figure 2 below*)

edikt::Eldas

edikt's Enterprise-Level Data Access Services (Eldas) use Java 2 Enterprise Edition (J2EE) and Enterprise Java Beans (EJB) technologies to enable Grid database access and integration.

Based on the specifications produced by the Data Access and Integration Services (DAIS) working group at the Global Grid Forum (GGF), Eldas delivers data access services for existing XML and relational databases within the Open Grid Services Architecture (OGSA) framework and within an enterprise Java architecture.

Eldas is enterprise middleware, facilitating the access and integration of data from separate sources via the Grid.

Eldas implements a subset of the DAIS specifications, focusing on providing scalable, secure and high-performance data access within the EJB framework. By adopting a modular design, Eldas allows details specific to particular databases or data sources to be released into in the Data Access Component (DAC) (*see figure 2, left*).

Multiple databases types can be connected to Eldas, via the DAC, making Eldas easily scalable and extensible.

edikt will release the generic Eldas software as a core product, but will also use it to underpin its work in data integration for e-Science.

More information

For more information about **edikt**, please see our website, www.edikt.org

