XML4PharmaNewsletter

Bimonthly newsletter of XML4Pharma, a subsidiary of Computer Chemistry Consultancy
Feldstrasse 20, CH-8488 Turbenthal, Switzerland
Phone: +41 52 3851745 Mobile : +41 52 3851745
Mail: XML4Pharma@CompChemCons.com
April 2003

XML4Pharma at the Pharma IT Summit and at IT Pharma Solutions 2003

XML4Pharma recently gave a very successful course at the Pharma IT Summit conference in New York. This workshop was titled "XML: The Language of Success for Pharmaceutical IT". Over 15 high-level managers from pharmaceutical and biotech companies, CROs and software companies active in the pharma industry attended this course. They learned about XML, DTDs, XML-Signature (the new vendor-neutral standard for electronic signatures) and about web services.

IBC Life Sciences also recently contracted us to give a similar course at the IT Pharma Solutions congress ("What you need to know about XML in Pharma") in Brussels on June 17th. Full information about this workshop can be found at:
http://www.ibc-lifesci.com/itsolutions/

CDISC soon to release XML-Schema for Lab standard.

CDISC (the Clinical Data Interchange Standards Consortium) will now soon release the XML-Schema for its new Lab standard, of which the specification was released recently. The Lab standard will be used for exchanging laboratory information of clinical trials between labs, CROs and clinical trials sponsors.

XML-Schemas will eventually replace most Document Type Definitions (DTDs), as the former enables better control and checking of input in XML files. For example, a DTD cannot impose a specific format for a date, whereas this is very easy in an XML-Schema.
XML-Schema is a rather new technology, and surely as difficult to 'read' as Document Type Definitions. However, it is important for IT-people in pharma companies to be able to 'read' XML-Schema, to be able to give

appropriate IT-support to the Clinical Data departments.

Therefore, XML4Pharma organizes special courses on XML technologies at pharma companies (on location at the company), where a large part of the course is devoted to understanding DTDs and XML-Schema. For more information, see:

http://www.XML4Pharma.com

An XML-Schema is also currently being developed by CDISC for the ODM, which currently uses a DTD.

XML4Pharma develops CDISC-XML-Checker

We are currently developing an XML-Checker for CDISC, the Clinical Data Interchange Standards Consortium. This tool will enable to check whether a CDISC XML file (ODM, Lab) conforms to the DTD or XML-Schema. If non-conformity is observed, the tool gives the exact location of the error, and gives suggestions (in understandable language for non-XML specialists) for correcting the error. This will enable CROs and sponsors to check on the validity of a CDISC XML file, before and after transport between a laboratory, CRO and sponsor. The tool will be made available for free to CDISC members in a foreseeable future.

Web Services

Web services (or applications talking to applications over the intra- or internet) are already well established in e-commerce and B2B transactions. This rather new technology is now also booming in bioinformatics, where researchers want to be able to perform large-scale calculations where some of the calculation methods are only available on remote computers.

Also on an intranet, web services can be very useful, not only in bioinformatics, but also e.g. in virtual high throughput screening. Another use is to enable new applications to use methods of already existing applications, therefore avoiding replication of software development work. Currently we are developing web services and wrappers for a number of existing applications (even written in Fortran !) of one of our customers, so that these applications become available as web services.
If you or your company is also interested in establishing web services on the intra-, extra- or internet, just call or mail us.

Presenting: Generalized Analytical Markup Language (GAML)

Generalized Analytical Markup Language is a still relative unknown member of new standards in pharma R&D. GAML has been created to allow long-time storage of analytical data, independent of operating system or computer system (can your current computers still read the analytical data created 10 years ago on a no-more-existent analytical instrument ?), and to allow the exchange of analytical data between computers of different types and with different software. Major considerations during building GAML were compliance with GPC, GERM (Good Electronic Record Management) and FDA 21 CFR Part 11.
Those now who think that GAML is yet-another-propriety, unfexible analytical standard, are completely wrong. First of all, the standard is open and its use if free of any license issues. Secondly, it is a very flexible standard, allowing every instrument manufacturer and chemical or pharma company to implement it.

The GAML website (**www.GAML.org**) lists a nice set of examples coming from companies like Bruker, Micromass, Perkin Elmer, Thermo LabSystems, and some others.

The standard allows storage of as well original raw data, plots and graphics, as textual results, with binary data stored as ASCII characters (Base64 encoded), so that there is a guarantee that these can still be read some 20 years from now.

```
- <trace name="TIC(steroids02)" technique="MS">
  - <coordinates label="Time (min)" units="MINUTES" linkid="MSTIME0">
      <values byteorder="INTEL" format="FLOAT32"
        numvalues="336">1uoYOzWlYjxg5dA8UIOXPfCnRj1TcXY9Wx2TPau
        9Sg+zBM1PiUGQT4uJE0+hxZZPi/dZD6Iz3A+4cF8PvVvhD4haYo+Tr
        qD4EVq4+WDm0Pg50uj46bcA+ZmbGPmt1zD6/WNI+dZPYPsl23j7
        PsI8AT9YOQQ/7jUHP3E9Cj/zRAO/iUEQPx8+Ez+OUBY/OEIZP+IzH
        6t8oPw50Kj9F/Ss/fIYtP8cELz8meDA/SgwyP22gMz9pSjU/jd42P3
        eEA/DElDP3pbRj/9Ykk/k19MPxVnTz+EeVI/BoFVP5x9WD8zels/yX
        aj+4ZWO/TmJwP+Recz96W3Y//WJ5P6dUfD89UX8/6SaBPzWlgj+
        POZvkz8I85Q/U3GWP57vlz/pbZk/IPeaP+F6nD8s+ZO/7nyfP8P1oD
        iYioP8ARqj+Blas/4AitP8kvrj8WIK8/dgWwP3TasD9eurE/SJqyPOZv
        6bY/xdm3P3TauD+F67k/5Re7PwllvD8Zvb0/x0u/P/7UwD81Xs1/b
```

An example of Base64 encoded binary data.

Instead of using a Document Type Definition (DTD), GAML uses the new W3C XML-Schema, as this allows a much more precise definition of which elements and attributes may occur, in which order and how many times, and in which format (for example for dates). A DTD does for example not allow to define that an element should contain Base64-encoded binary data, whereas this is very easily done in an XML-Schema.

The GAML website is worth visiting: it has a good number of application examples, an excellent presentation about what GAML is all about, and a good amount of documentation, including a white paper and the full XML-Schema.

For more information, see:

www.GAML.org

```
<?xml version="1.0" encoding="UTF-8" ?>
<!-- edited with XML Spy v4.4 U (http://www.xmlspy.com) by Ja
- <GAML version="1.00" name="nylon66">
  - <experiment name="FID">
      <collectdate>1994-07-05T20:26:00Z</collectdate>
      <parameter name="TITLE" label="TITLE" group="acqus">Para
        930901</parameter>
      <parameter name="JCAMPDX" label="JCAMPDX" group="acqu
      <parameter name="DATATYPE" label="DATATYPE" group="ac
        Values</parameter>
      <parameter name="ORIGIN" label="ORIGIN" group="acqus">
        GmbH</parameter>
      <parameter name="OWNER" label="OWNER" group="acqus">
      <parameter name="FileDate" label="FileDate" group="acqus":
      <parameter name="File" label="File"
        group="acqus">File: /u/data/general/nmr/mikek/18/acq
      <parameter name="AQ_mod" label="acquisition mode" group=
      <parameter name="AUNM" label="AUNM" group="acqus"><au
      <parameter name="BF1" label="BF1" group="acqus">500.13
      <parameter name="BF2" label="BF2" group="acqus">125.75
      <parameter name="BF3" label="BF3" group="acqus">500.13
      <parameter name="BF4" label="BF4" group="acqus">500.13
      <parameter name="BYTORDA" label="Byte order for binary d
      <parameter name="CNST0" label="CNST0" group="acqus">1
      <parameter name="CNST1" label="CNST1" group="acqus">1
      <parameter name="CNST2" label="CNST2" group="acqus">14
```

An example of Generalized Analytical Markup Language (an NMR experiment) from the GAML website.

The eCTD (v.3) in table form: new version available

Last year, we developed a table (25 pages) explaining the DTD of the eCTD in a very well readable table format. This table (available for free) was highly popular at regulatory and IT departments of many pharma companies. Since that time however, the eCTD specification has seen some changes. Therefore we have adapted the table to meet the version 3 specifications, which were published in November 2002 by the ICH.
The new explanation of the eCTD is as of now available. Just simply send us a short e-mail, and we will send you a copy of the table.