Bimonthly newsletter of XML4Pharma,
Katzelbachweg 18, AT-8052 Thal, Austria
Web : www.XML4Pharma.com
Mail: Info@XML4Pharma.com
October 2012

XML4Pharma is a CDISC Registered Solutions Provider

Editorial

It is a long time ago since the last issue of this newsletter came out (february 2012). Since then, I bought a house, partially renovated it, and moved into it. At the university (University of Applied Sciences FH JOANNEUM, Graz, Austria), where I was in my first year as a professor in medical informatics, I did set up several new courses, and besides several student projects (of which I will tell something in later in the newsletter), I supervised 4 bachelor and 5 master thesises.
During the summer vacation I did some good mountaineering together with my son (we climbed Austria's highest mountain over a difficult route) and worked further on getting the new house in order, especially the garden which needed a lot of care.
But now summer is over and a new academic year has started. Although I still need to develop new courses, there is less stress, life has come to normality, and I do have time again to write a new issue of the newsletter.
A lot has happened in the last six months in the world of standards for clinical research, and this newsletter reports on the major new things.

CDISC publishes draft of v.2.0 of the define.xml standard

At the beginning of september, CDISC published the draft "for public comment" of the define.xml v.2.0 standard. You can find it here.
Although the comment period has ended (deadline was 1st of october) I would invite all those that are interested to read the draft specification and have a look at the examples.
The major new features are:

- Define.xml 2.0 is based on ODM 1.3.1, the latest version of the ODM standard, not anymore on the old ODM 1.2, which I consider as being outdated
- Define.xml introduces the concept of slices (subsets of a dataset that typically include a subset of the dataset rows) through the "where clauses". This e.g. allows to define that VSPOS="sitting" is only allowed in the case that VSTESTCD="SYSBP" or "DIABP".
- This was not only necessary to be able to better describe relationships between variable values (ValueLists did not provide a complete solution for this) for SDTM submissions, but also to allow for **full support for ADaM**.. A few screenshots are given at the end of this newsletter.

We have submitted a good number of remarks (mostly suggestions for improvements) and will also test the new (draft) standard on implementability in the next months.

FDA starts thinking about SDTM, ADaM and SEND submissions in XML

The FDA is organizing public meetings to discuss new (technical) solutions for study exchange standards. The full FDA announcement can be found here. The first public meeting will be held on November 5th.

Essentially, the FDA is trying to make a choice between CDISC ODM and HL7-v3 as the basis for the new exchange standard, and is asking participants to provide advantages and disadvantages of both these standards as the basis.

Fortunately, CDISC this times fully supports the idea to use the ODM standard as the basis for future SDTM, ADaM and SEND submission data submissions to the FDA, and rejects the idea of developing new HL7-v3 messages for this purpose.

Therefore, several CDISC teams are currently sending their comments to the FDA, and several CDISC representatives as well as CDISC stakeholders (pharma companies) will attend the meeting. If you would like to attend the meeting at the FDA in order to support CDISC, you can find the details about how you can attend on the CDISC website.

In our SDTM-ETL™ software, we do already use a prototype of such an ODM-based format for keeping the SDTM data. Only in the very last step this is then "downgraded" into SAS XPT.
Also several other vendors are using such an approach. This also means that such a new standard could be developed very rapidly (we estimate 6 months) as almost everything is already there and well tested and used. The only thing that needs to be done is an alignment of the current versions of the different vendors.

CDISC Interchange Baltimore

CDISC is organizing another International Interchange in Baltimore in the week of October 22nd (well, this week).
Very special and new is that there are almost no presentations at this Interchange: the only real presentation I could find in the programm is given by Wayne Kubick, the Chief Technical Officer of CDISC. All other sessions consist of panel discussions! I was pretty surprised when I first inspected the programm. Has this been a deliberate choice of the programming committee? Or have there been an insufficient number of abstracts send in?

Anyway, the Interchange concentrates on FDA issues and on therapeutic standards, so it is very SDTM-oriented. Therefore, the conference is not so very interesting for me (although we will have to do with therapeutic standards in the future in our SDTM consultancy work). Also given the costs, the long flight, other duties (at the university), and the bureaucratic hurdles (EU citizens like me again need a visa for the US, and at arrival we often still need to queue up for an hour or more for immigration - US citizens can enter Europe without visa and pass through immigration in just a few minutes) I decided not to attend the Interchange, even if I would love to hear Wayne Kubick's presentation.

However, there is also the upcoming European CDISC Interchange in April in Germany which I will surely attend. So if you want to meet me in person, ensure that you attend the European Interchange.

XML-Schema 1.1 is W3C recommendation

Since April this year, XML-Schema v.1.1 is a "W3C recommendation", which means it is an approved W3C specification.

I have been awaiting this new version for a long time, and am now very happy that it has obtained the "final"status. The reason for this is that XML-Schema 1.0 has many limitations, and some parts of it (e.g. the "redefine" mechanism) has been differently interpreted by different implementors.

So recently, I bought the book "Definitive XML-Schema - second edition" written by Priscilla Walmsley. I can recommend this book to anyone interested in XML-Schema, be it v.1.0 or 1.1 (both are described).

XML-Schema 1.1 has many advantages over 1.0. For example, it allows to add complicated rules in the form of "assertions", a feature that has been borrowed from "Schematron", which is another language for validating instance documents against a standard. Furthermore, the "alternative" element allows to define an XML-element being of different types, depending on the value of one or more attributes. This means that the structure of an element (i.e. which child elements are allowed, how many of them and in which order), can be made dependend on the context of the element as being defined by one of its attributes.

Another interesting feature of XML-Schema 1.1 is that attributes can be declared "inheritable" meaning that the attribute value is then inherited by the child elements. This is important, as in the past, attribute inheritance was sometimes assumed but it could not be declared in the XML-Schema whether this is the case or not.

An important new feature (from the CDISC perspective) is that the "redefine" element has been deprecated, and is being replaced by the "override" element. All ODM-extensions such define.xml and SDM-XML are based on "redefine, but some tools vendors like Altova (XMLSpy) had a different interpretation of the functioning of the "redefine" element, so that these tools do not support ODM-extensions like define.xml. This has led to confusion, which has led the CDISC XML Technology team to write a white paper concerning Schema validation for define.xml.

Now the new "override" element has been described unambiguously in XML-Schema 1.1, so that different interpretations are no longer possible.

Until now, there are not many XML editors available that fully support XML-Schema 1.1, but there are already some very good Java libraries available such as Saxon and Xerces. Those editors that have implemented one of these librariers however can already work with XML-Schema 1.1.

As XML-Schema will allow to much better describe the rules for XML-based CDISC standards such as ODM and define.xml, I have proposed a student's project at our university to transform the existing XML-schemas for ODM (1.3.1) and define.xml (1.1) to schemas according to the new W3C XML-Schema 1.1 standard.

HL7 certification received

Electronic Health Record systems (EHRs) often use HL7-CDA-based documents like the CCD (Continuity of Care Document) to exchange information between systems and persons.
Also the national EHR system ELGA that is currently being developed in Austria (where I live) is based on HL7 standards. The further development and deployment of the ELGA system was just given "green light" two weeks ago by the government, so that we expect an acceleration in the development of systems for working with EHRs in Austria.

As also all the profiles of IHE that have to do with clinical research (and in which I have been involved) use HL7-v3 standards, I believed that the time was right to start studying again, so last spring I decided to sign up for the HL7 certification class that was organised by HL7 Austria. The course was an e-learning course with weekly online tests and exams. It was very intense - I had to study for 4-8 hours per week to pass the exams, but it was also great fun, and I am now proud to announce that I am now an "**HL7 certified person**".

This certification will in future further help me to make contributions to the development of methods and systems that allow to better integrate clinical research with systems used in healthcare.

It will also allow me to further improve my teaching at the university where HL7 standards is an important part of the 3-semester course "Medical Informatics".

Some of our students projects

Although this has nothing to do with XML4Pharma directly, I would like to tell the interested reader about a few of our students projects (some of them leading to a bachelor or masters thesis) that have to do with CDISC standards.

First, one of our bachelor students is currently developing a "schematron" for ODM 1.3.1. A schematron is an XML-technology to describe complicated rules for XML-based standards such as CDISC-ODM and HL7-CDA. Typically, they are used as an addition to XML-schema, as the latter is very good in describing the hierarchy of the XML elements and simple restrictions on attributes and element contents, but not in describing more complicated rules such as dependencies between attributes.This is where Schematron has its strength.

Another project that is currently running is a prototype of a system for preselecting eligible patients using a repository of EHRs in the form of HL7-CDA documents. The system reads an ODM-SDM file containing eligibility criteria in a machine-readable form, and then presents these criteria to the user. The user can then choose which criteria to apply. The system then checks every EHR file (CDA format) and selects those for which the information passes the eligibility criteria that were selected. It then displays the list of selected CDA files, which can then be inspected by the user.

A third project is about SDTM and define.xml. Define.xml essentially contains the information in order to automatically create a database. The SDTM database tables essentially contain the information to populate this database. The FDA tried to develop such a system in the past but failed. In my personal opinion, the FDA made the error that they tried to skip the "database" step and directly wanted to populate a datawarehouse.

Three of our bachelor students have now developed a system to automatically create an SDTM database from the define.xml, and then populate it with the information from the SAS-XPT datasets. They will hand over their results to me at the end of this month, and make a public presentation in November.
We will then probably present the results as a poster on the next european interchange in Germany in spring.