Bimonthly newsletter of XML4Pharma,
Schlossbergstrasse 20, DE-78224 Singen, Germany
Phone: +49 7731 975044
Web : www.XML4Pharma.com
Mail: Info@XML4Pharma.com
January 2010

XML4Pharma is a CDISC Registered Solutions Provider

January edition 2010

We wish all our readers a happy new year, and all the best for 2010.
The start of a new year is traditionally the moment to make some good intentions for the year to come. So do we.

One of our intentions is to publish at least 6 newsletters this year. We hope you like that.

Just like the previous years, we will continue working as a CDISC volunteer and contribute considerably to the development of new CDISC standards or of new versions of them. In that sense, one of my personal intentions is to develop a Schematron for the ODM standard, this to further enable to enforce the rules from the ODM specification that cannot (yet) be expressed in an XML-Schema.

CDISC publishes ADaM 2.1 and AdaM-IG 1.0

Only a few weeks ago, CDISC has published version 2.1 of the "Analysis Data Model" standard. It also published the first "Implementation Guide" for that model. Both can be found here.

Especially the fact that an "Implementation Guide" is now available will make it much easier for implementors to start working with the model, and to design and develop software for generating ADaM datasets.

As also the define.xml standard will be updated in future to provide the metadata for the ADaM datasets (now still in SAS Transport 5), also integration between SDTM and ADaM will make significant progress.

SDTM-ETL software to provide support for ADaM

We are currently working very hard on a new version of our popular SDTM-ETL software.
The new version will now also provide support for defining and generating subject-related ADaM datasets, and other user-specific datasets.

Basis of the software remains that clinical data (in ODM format) can be mapped to SDTM/ADaM datasets in a very user-friendy way, usually by drag-and-drop and smart wizards.

The beta version of the software is currently being presented to several customers, and is expected to be publicly released in March.

CDISC publishes document "XML Schema Validation for Define.xml"

As is well known, a number of departments at the FDA have still very little XML expertise. So it was no surprise when one of them reported that they have major difficulties with the validation of define.xml files. Especially, as a first step, they wanted to be able to validate incoming define.xml files against the define.xml XML-Schema, this though this is only one of the aspects of validation against the standard.

Therefore, the "CDISC XML Technology Team" (of which we are an active member) decided to write a "white paper" in order to help the FDA and other users to validate define.xml files against the XML-Schema.

The "white paper" describes not only best practices for validating define.xml files but also challenges and pitfalls. It also provides a list of commercially available tools for validation of XML files against the XML-Schema and whether they comply for use with the define.xml standard.

The white paper states: "*XML schema validation is a first step towards verifying that a define.xml instance matches the published standard. Since the schema cannot represent all the rules and requirements documented in the specification, additional checks should be executed to ensure compliance with the complete specification*".

Therefore it also lists a number of software tools that validate define.xml files beyond the XML-Schema, including our "Define.xml Checker", which is currently used by many companies that generate define.xml for submissions to the FDA.

ODM 1.3.1

The specification and XML-Schemas for the minor update ODM 1.3.1 have been published on the CDISC website for public review late november. The review period was until December 18th. The ODM team is currently processing the comments and will publish the final specification and XML-Schemas in the next few weeks.

This update corrects some small errors in the previously published specification and XML-Schema, and relaxes some constraints, e.g. to even better support extensions such as the upcoming Trial Design extension.

A number of clarifications and notes about "best practices" have also been added. The content of the latter may become part of the specification itself in future (as new constraints).

As usual, the ODM 1.3.1 is 100% downwards compatible: every valid ODM 1.3 file is also a valid ODM 1.3.1 file. This will however probably not be 100% the case however when we move to ODM 1.4 in future.

FDA changes direction about future transport format for SDTM submissions

This is the big news.

Before reading further, you might read Dave Iberson-Hurst's blog first, as this is the major source of the information.

During an FDA session at the last US Interchange, where critical questions were asked about the development of HL7-messages for future SDTM submissions, it became apparent that there are at least differences of opinion between the "Office of the Commissioner" (OC), which are the "developers[1]" of the standards, and CBER and CDER, who are the "users" of the standards. Whereas the latter are still putting SDTM in place[2] (training, processes, new software: "*we are trying to create a modern review process*") the former are envisaging the use of HL7-XML technology for future submissions. The original plan was to have these HL7-messages in place in 2013.

From the FDA session (and Dave's blog) it now has become clear that the date of 2013 has been dropped[3]. Another target date has however not been given. It also became clear that the FDA is no longer envisaging an HL7-**message** for submission of the SDTM data (at least the subject-related domains), but will perform an investigation whether the HL7-**CDA document** (which is envisaged by HL7 for electronic health records) can port SDTM submission data.

Both are important pieces of information.
Firstly (the disappearing timeline): it highly probably means that SAS Transport 5 (in my eyes, a prehistorical format) will remain in use for many many years at the FDA (my estimate is 10 years). One may say this is bad news or this is good news. The good news is that it means stability – i.e. processes, software and systems that produce SDTM datasets in SAS Transport 5 format, must not be updated (at least not for the output format) in the next 5-10 years.
The bad news is that we are forced to accept that we have to live further with all these terrible limitations of SAS Transport 5, such as (the real list is long) 8 characters for the variable names, 40 for the labels, and 200 for text content.
You can choose yourself what is most important for you.

Secondly, the change in choice from an HL7-message to the HL7-CDA document. Although both are related and based on the same technology (the highly criticized HL7-XML), this is an important change of direction. The CDA ("Clinical Data Archictecture") specification is envisaged by HL7 to be used for the future persistent electronic health record (EHR). Exchange of information between

1 "Developer" is a great word here, as due to the lack of XML knowledge at the FDA, the develoment of the standards itself is outsourced by the FDA.
2 Remind that SDTM 1.1 as well as define.xml were published almost 5 years ago.
3 A few weeks later, J.Levine from the FDA Office of the Commissioner denied this however !

healthcare takers should then be done by extracting information from the CDA into exchange HL7-messages, such as the CCD ("Continuity of Care Document").
I agree that this may be confusing: does the FDA want the persistent EHR, or does it want a transport/exchange format, which by definition, ... should be an HL7-message?

Anyhow, it is clear that ... a lot is unclear. Essentially, the vision of the FDA (or better: of the Office of the Commissioner) is that SDTM submissions should be done in a format very similar to that envisaged for EHRs by HL7. I say "envisaged", as although HITSP ("hitspy") has recognized CCD as a format for exchange between EHR systems, others may get the same status. Arguments like "single source" come into play.

However, my personal opinion is that this is a wrong vision. First of all, CDA is US-specific. It is nearly not used outside the US. This nails down the argument of integration with EHRs and single source.
Secondly, we are talking about transport formats. It is not because you are using the same truck to transport oranges as to transport cows, that you can breed a cow that produces orange juice (see the previous newsletter).
Thirdly, it still needs to be proven that CDA can transport SDTM information – that is far from sure sofar. Fourth, as CDA (and HL7-XML in general) is enormously complex, I do not see how CDER and CBER will be able to implement it, seen the technical infrastructure they have and the lack of XML knowledge at the FDA. So a visionary idea may be found to be impossible to implement.

So, what to do?

As SDTM (as it is now with the SAS Transport file mechanism) will remain in place for many many years in my opinion, organizations that do not use SDTM to submit to the FDA (thinking the HL7-mechanism will soon come) should definitely make the switch to SDTM **now**. This has also clearly been stated in the excellent blog contribution of Becky Kush, which was published shortly after Dave's blog. It also includes the strong recommendation to use CDASH forms for data collection (CDASH forms in ODM format will be published soon).

Personally, I do not think that CBER and CDER will be able to implement HL7-XML before somewhere between 2018 and 2020. So making the switch to SDTM (even though it uses the prehistorical SAS Transport files) makes complete sense.

An alternative proposal

The mandatory use of the SAS Transport 5 is a serious handicap when preparing data sets with SDTM content, costing the industry a lot of time and especially money.

As the new envisaged format will not be there in the next few years, and as it still remains to be proven that CDA can port SDTM data, we (together with a few other CDISC volunteers) have formulated a counterproposal.

CDISC has a number of people that at the same time are XML experts (some even say "XML-gurus") **and** have an extremely good knowledge of SDTM and of define.xml. Some of them are currently developing a define.xml extension for ADaM.
The proposal is that these people sit together and develop a transport format for the SDTM data themselves, based on define.xml (which is based on ODM). As much is already in place, this can be accomplished in less than a year. The results can then be used in a pilot with the FDA. If successful, the FDA can then implement this transport format to replace SAS Transport 5. This does however not touch SDTM itself at all, except for the current length limitations.

As ODM and define.xml are much less complex than HL7-XML, this can also be an excellent transition stage for CBER and CDER to move into HL7-XML, if ever proven to be implementable.

Furthermore, the companies that these XML gurus represent can develop tools that allow sponsors to transform existing SDTM-SAS datasets into the new format. The technology for this already exists, so that the tools can come at low cost or even for free.

The full text of the proposal can be found on our website, and has been referenced in a comment to Dave's blog on the CDISC website.

Transforming ODM into SAS Transport 5

It is something we have already been thinking about for a longer time ... But now that it is (unfortunately) clear that SAS Transport 5 (XPT files) will be there for many many years, we make this an opportunity.

We are currently working on a transformation tool that allows to create SAS XPT files from ODM files. The user has the choice between creating one dataset per Form, or one dataset per ItemGroup.
Furthermore, there are many more options whether additional context information (such as the Study OID) should be added to the records.

The software is currently in the late development phase, and we are thinking about a formal release in

February or March. The name of the package is still undecided, maybe something like "ODM2XPT"?

The software will be especially interesting for those who do not wish to use PROC CDISC (which does not work in all cases[4]), or for those who desire a low cost solution for importing clinical data in OMD format into SAS or any other statistical software package.

A screenshot of the graphical user interface can be found below.

CDISC publishes new CDISC Glossary v.8.0

The most recent news is that CDISC just published version 8.0 of the "CDISC Glossary". The full set of documents can be found here.

The Glossary itself consists of 35 pages with definitions of terms. What I found very good is that in many cases, the definitions also contain a reference when the definition was copied from another source, or was based on a definition from another document ("after ...").

Also very useful is the separate document with abbreviations and acronyms. Though I missed some important ones (e.g. CCD: Continuity of Care Document, eCDT) this document allows to quickly look up an abbreviation used in other CDISC documents.

The Glossary is also available as a MS Word document (open standard?), but unfortunately not as an XML document. Maybe the next step should be to apply a semantic web approach and develop an XML presentation in a language like RDF or OWL.

Altogether, a great piece of work again by Steve Raymond and his team!

4 For example, PROC CDISC does not seem to work with ODM 1.3.