Bimonthly newsletter of XML4Pharma,
Schlossbergstrasse 20, DE-78224 Singen, Germany
Phone: +49 7731 975044
Web : www.XML4Pharma.com
Mail: Info@XML4Pharma.com
April-May 2009

SPECIAL EDITION: CDISC EUROPEAN INTERCHANGE

This edition has a lot of news from the CDISC European Interchange which took place in Budapest (Hungary) from April 20 to 24.
As usual, there were training courses before and after the two-day conference, to which we contributed. I gave a presentation in the main session of the conference on integration of electronic health records with CDISC standards, and we also had a booth at the commercial exhibition.
But first, the other news!

CDISC published SDTM 1.2 and SDTM-IG 3.1.2

For those who haven't noticed yet: CDISC has publised the SDTM v.1.2 model and the corresponding Implementation Guide (SDTM-IG v.3.1.2). This is the good news.
The less good news is the announcement that the FDA cannot immediately accept submissions using this updated version of the standard, as they still have problems with the "availability of updated software". The FDA has already retarded the publication of this new version by more than a year (Yes, the SDTM team had the new version ready already more than a year ago), and knowing the skills in software development at the FDA (or the lack of it), is letting us fear that it may still take a good time before the FDA will be ready for it.
Personally, I think this is not an ideal situation, as there is no drive for "early implementers". The question is "who will be investing in the new version as long as it is uncertain when the FDA will accept submissions in it?".
The new model is mostly downwards compatible with the previous version, though there are some important changes: the "Subject Visits" and "Subject Elements" domains have been moved to the "Special Purpose Domains" class from the "Study Design" class (as these are not design, but observation). Furthermore, a new "Clinical Events" has been added to the "Findings" class (for clinical events that are not adverse events), and a set of "Pharmacokinetics" and "Microbiology" domains has been added. Also a "Findings about Events or Interventions" domain (FA domain) has been added in a separate class.
Altogether the Implementation Guide has grown from 183 to 298 pages, which is a good amount of reading ...
The format for submissions to the FDA remains SAS Transport 5, with all its limitations (SAS Transport 5 originates from the IBM Mainframe time – do you still have one at home?).
Unfortunately, it will still take a good number of years before the FDA has an XML representation ready. This will be an HL7 message (another monster – see our previous newsletters), this although some XML experts have already developed an extension to the ODM model to also carry submission information. However, the FDA decided not to implement that.
The SDTM v.1.2 model and Implementation Guide can be downloaded from the CDISC website.

Protocol Representation Group publishes draft of Protocol Representation Model v.1.0

After a good number of hard-work years, the CDISC "Protocol Representation Group" (PRG) has published the draft of version 1.0 of their model.

The model is essentially a set of UML diagrams with explanations of the different classes and fields.

The website where all information can be found is here – please ensure that you get the publication of May 4th, as the first publication was not complete.

The model is one that has been fully mapped to BRIDG, and is (in my personal opinion) only a first step to come to a machine-readable protocol. This as UML is not XML (nor suitable to come to XML). The next important step will be the publication of an ODM extension which implements this model. Some XML-specialists of the PRG are currently working hard on that ODM-extension, so we hope that it will not take too long before it is published.

The concurrent HL7-v3-XML message, which is only meant for submission to the FDA, and not for execution in computer systems, currently only exists (as far as we know) as "storyboards". As it is an HL7 message, I believe its draft will also not

become available to the CDISC community for revision.

For those working with our "ODM Study Designer": we will implement the ODM-extension in the software as soon as it becomes publicly available.

The CDISC European Interchange – day 1 and 2: SDTM Training Course

Peter van Reusel (from Business and Decision) and I gave the SDTM training on day 1 and 2 of the Interchange. We had 24 participants from all over Europe and as well from sponsors, CROs as from technology companies.

Especially interesting about such a public training are the discussions between the participants, reflecting the different ways of working with clinical data in different (company) cultures. This is pretty different from training at companies, where the discussions are usually between the data management people and the statisticians.

Our class working hard on one of the many excercises

CDISC European Interchange – the German User Group Meeting

On Tuesday evening there was also a short user group meeting of the German CDISC community. After a few announcements from Daniel Rehn (Roche), the coordinator of the user group, we split up in different "workstreams" where we discussed about implementations of the CDISC models.

The meeting was then followed by a nice reception where we also met our colleagues from the other user groups. Most of us then used the evening for sightseeing in Budapest – I rushed off to build up my booth at the commercial exhibition.

CDISC European Interchange – the main conference

Wednesday and Thursday were the days of the main conference. The first day had four sessions.

In the first session (Welcome and Keynote) Tim Jaeger (chairman of the European CDISC Committee) gave us a warm welcome, as well as Becky Kush (CDISC president) who also gave us a short update on what is going on within CDISC.

The first keynote presentation was given by Charlie Mead (HL7) who gave a very inspiring presentation (some people told me that his presentations are usually even more inspiring) on HL7-CDISC intersections. Very interesing was his SWOT[1] analysis of HL7 and CDISC. For me, some eye-openers (I only list the ones in the category of "weaknesses") were:

- HL7 is not good in XML development
- HL7 has a legacy "messaging" focus[2]
- V3 has not reached its goals (very small penetration)
- CDISC is not good (yet) in modelling

I agree on all of these points, also on the last point: we developed our models as silos. If we had started with something like BRIDG ten years ago, the integration of all the standards would have been much more easy. However, my observation is that the current emphasis on modeling considerably retards the development of our standards. A good example is the Protocol standard, which has taken 5 years to develop.

In the second session (European Perspectives of EHR Integration), Pierry-Yves Lactic and Isabelle de Zegher made it clear that technology is not the limiting factor, but that local laws, different interpretation of European rules (e.g. about privacy), and uncertainty about the position of the FDA on data from EHRs are the factors that retard implementation.

I then gave a presentation about a pilot project that was done by an IHE-CDISC group implementing the IHE "Clinical Research Data Capture" (CRD) profile, where we developed a technical solution (using ODM and XSLT) for retrieving information from EHRs into CDASH forms. A download of the presentation is available **here**.

1 **S**trengths, **W**eaknesses, **O**pportunities and **T**hreats

2 Seen in the view of their time, V2 messages were very good for what they meant for, but the concept should never have been taken to V3, as at that time, messaging standards such as SOAP (web services) had already been developed (my personal opinion!)

David Iberson-Hurst (CDISC VP Technical Strategy) then gave a presentation on "Single Source", and especially about the position of the FDA and EMEA. This was very interesting, as David showed us the "general rules" under which EHRs and CDISC standards can interact so that the resulting (submission) data that can be accepted by FDA and EMEA..

The first afternoon session was devoted to CDASH. We heard a lot about how people implement CDASH in their company. Some discussions took place whether it is allowed to change/adapt some CDASH variables for local usage. My answer to that is "Yes, that is allowed - as I consider CDASH more as a recommendation than as a standard. CDASH is not a goal in itself, it is a means allowing to more easily come to submission data".
Altogether, this session clearly showed that CDASH is "hot", and extremely welcomed by the industry.

The second afternoon session was devoted to experiences of SDTM submissions to the FDA. As I have followed the discussion forum on the CDISC website on this topic, I was well prepared. Some findings from this session and from the forum are:

- about half of the SDTM checks in WebSDM is wrong or unclear
- Many FDA reviewers still do not have (sufficient) knowledge of SDTM and need training
- Some (fully correct) submissions require the FDA to "adapt" or "tweak" the Janus data warehouse.
- Although define.xml is meant to replace "paper", many reviewers require a PDF of the define.xml for printing (falling back in old habits)
- The FDA requires you to submit a stylesheet to view your define.xml, this although the XML format is just there to allow the reviewers to have a view independent from that of the submitter[3].
- Even if your submission does not pass the WebSDM checks (sometimes WebSDM is wrong, not the submitter), this does not mean that the FDA is rejecting your submission.

Altogether, this does not give the best impression of the IT-technical capabilities of the FDA.

The day was closed with the traditional conferece event which took place in a typical hungarian restaurant (in a brewery), with great food and even better beer and wine.

Somewhere, somehow, all these people seem to have some connection with Holland (the Dutch connection?)

Day 2 of the conference had parallel sessions. For me, this is always difficult, as it often leads to conflicts in choosing which presentation to attend.

A personal highlight this day was the presentation of Bill Rosen (Pfizer) about Adverse Event Reporting and how integration between hospital information systems and CDISC do help. Another highlight (for me) was the presentation of Peter van Reusel (Business and Decision) about his experiences with submissions to the FDA.

Peter van Reusel explaining why WebSDM reports full with WebSDM errors from the FDA is not a disaster

Also very interesting was the presentation of Ingo Beinlich (Cidar) about designing SDTM domains for efficacy data. It once again showed us that transforming operational data to submission data is an "interpretation and categorization" step, so not exact science.

The session "Metadata Management and Model Extensions" of course had my special interest (as it was mainly about ODM). I liked the presentation of Doug Bain from Medidata) about how they apply CDASH, web services and ODM extensions in their

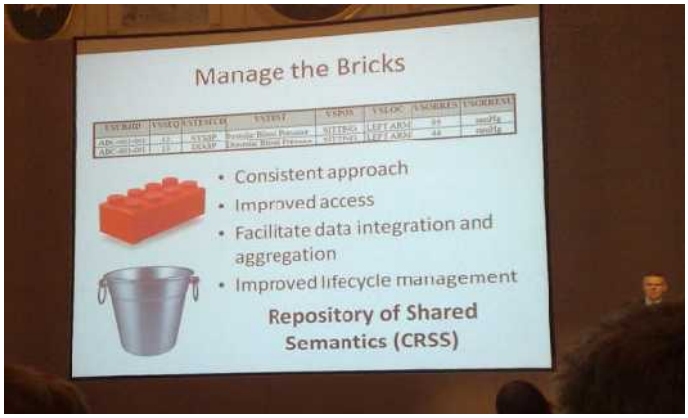
3 Ideally, the FDA should have their own stylesheet(s), so that they can look at the metadata with their own 'eyes', and not with the 'eyes' of the sponsor.

products. Personally, I would say that a representative of Medidata in our ODM Development Team would be very welcome.

Last but not least in this parallel sessions was the presentation of dr. Carsten Heil of Cardinal Health Research in Würzburg. Cardinal Health is one of my customers, and we developed a large extension to the ODM for use in complex multi-device clinical trials. A good amount of these extensions will also be presented to the CDISC ODM development team for consideration, or as a base for future standardized extensions in the field of devices and ePRO.

dr. Carsten Heil explaining why the extension to the ODM was necessary

A very important presentation came at the end[4] of the conference. David Iberson-Hurst (CDISC VP for Technical Strategy) explained the plans for the "Repository of Shared Semantics" (earlier designated as CDISC metadata repository). He presented his view of how in future CDISC and CDISC implementers will maintain and use metadata and data together, as "Lego bricks", and how these will be submitted to the FDA.



David Iberson-Hurst explaining the concept for the "Repository of Shared Semantics"

As the concept is similar to that used in "Object Orientation" (OO) I like it very much[5], but I am a bit sceptical about the implementation, for the following reasons:

- I have doubts about the capability of the FDA to work with such concepts – reviewers are used to think in terms of "tables"
- I have strong doubts about the capabilities of the FDA to develop any software based on OO concepts (or have any software developed by third parties).
- I still believe an HL7-v3 message is NOT the intermediate solution for submission data[6] for replacement of SAS Transport 5. I also do not see how the currently planned HL7-v3 message is compatible with this concept
- Neither do I believe that a future HL7-v3 message is the right format for implementation of the concept. For example, the HL7 datatype "date" is even not compatible with the SDTM datatype "date".
- What is the timeframe? Do we need to wait another 10 years for replacement of SAS Transport? Even when it is there – how much time will the FDA need to implement?

Another concern is about who will further develop the concept to a standard. David told us that "CDISC will not develop any technology". If he meant that CDISC will not develop software (as was shown on one of his slides) I would agree. So who will develop the XML-based standard format for future SDTM submissions? Personally, I think it should not be HL7 ("HL7 is not good in developing XML" - sic). CDISC has a number of volunteers that are world-class XML developers, so should they do it?

Personally, I believe the best solution is to replace the SAS Transport 5 format immediately (it can be done within months) by an ODM-extension, keeping the concept of two-dimensional tables. This would also allow to publish the SDTM-FDA rules (as currently implemented - but not always correctly - in WebSDM) as a Schematron[7], so that:

- the rules are 100% transparent to everyone
- everyone uses exactly the same rules
- no expensive software is necessary

This would immediately eliminate the limitations of SAS Transport, and at the same time make the submission format compatible with define.xml, also allowing validation of submission data agains their metadata (which currently seems not to be possible).

On the longer term the concept of the "Repository of Shared Semantics" can then be implemented using a new XML format, to be developed by XML specialists, using a framework that is not extremely over-generalizing (which HL7-v3 does), and which

4 In my personal opinion, it should have come at the start of the conference, so that people would be able to discuss it.

5 Not to say that I am very enthusiastic about it.

6 Also see the articles in our previous newsletters

7 Another XML technology – see the wiki.

is not a "message" (we have worldwide standards for that, e.g. "SOAP" for webservices). Mapping to BRIDG should not be a problem, as the new format transports SDTM, which is already being mapped to the BRIDG.

So far my personal opinion.

CDISC European Interchange – the XML4Pharma booth

Of course we also had a booth at the commercial exhibition. This allowed the conference participants to obtain information about our services and our software offerings.

CDISC board members using our booth as a mini-meeting place after the conference

The booth was well-visited and we obtained many request for information, especially about our ODM-to-SDTM mapping tool (SDTM-ETL™), our define.xml Checker and our ODM Study Designer software.

Some afterthoughts triggered by the CDISC European Interchange

For me, the European CDISC Interchange was again a great event. Many of the presentations and discussions during the conference have initiated new thoughts about how CDISC can evolve in future. Though the liason with HL7 is very important, it should not lead us to making the same mistakes as HL7 did during the last 5-10 years. What we can learn from HL7 however is to better model. But even when we do so more and more, I currently see the tendency within CDISC to stop at the (UML) model, and not invest sufficiently in implementing the model. In the earlier days, we just developed XML-based standards without really developing a model first, which later gave us difficulties when integrating one standard with another. Maybe the current tendency is too much to try to model everything, without thinking about how it should be implemented technically.

At the same time, I wonder whether the modeling that we do should be UML-modeling: UML is meant for software development, and the CDISC goal is not to develop software. UML is currently being "abused" for XML-development, and I have my hesitations whether it may be used for that.

Personally, I think the liason with IHE[8] may even be more important than the one with HL7, as IHE is SDO[9]-neutral, and really tries to break through barriers, and accomplishes real-world achievements. Our standards developers (especially our technical people) should get more involved in IHE profiles and pilots, but I also recognize that very often, they just do not have the time, as it is all unpaid volunteer work ...

CDISC as an organization is still rapidly growing: CDISC hired several people again last year and this year – so CDISC is also professionalizing.
And I think this is very good.

Many new concepts are developed, new liasons and cooperations are being established, but my impression is that we are becoming slower and slower in implementation. My personal opinion is that the reason for this is that we do not sufficiently invest in technology. With technology I do not mean software development (I agree with David I.-H. that CDISC should not develop software[10]), I mean skills like XML-Schema, Schematron etc..
CDISC still hasn't hired any professionals in this field, and I think CDISC should. If CDISC does not want to go into that direction, I see a good alternative:
If we look at other organizations that develop XML-based standards (for example MathML or OpenEHR), I see that many academic groups are involved in the (technical) development of the standard. In the case of CDISC, I haven't heard about any academic group actively contributing to technical development of CDISC standards. Though we have "HL7-professors", "OpenEHR professors", "MathML professors", there is not a single "CDISC professor" worldwide.
Academic groups have the advantage that they can

8 "Integrating the Healthcare Enterprise" - also see our previous articles on CDISC – electronic health record integration.

9 SDO: Standardization Development Organization

10 Though I think CDISC should better overlook software development. For example, the FDA uses software that claims to implement CDISC standards, but it doesn't do that correctly. So, I think CDISC should certify every software package which is intended to be used by the FDA before it is allowed to be used by the FDA

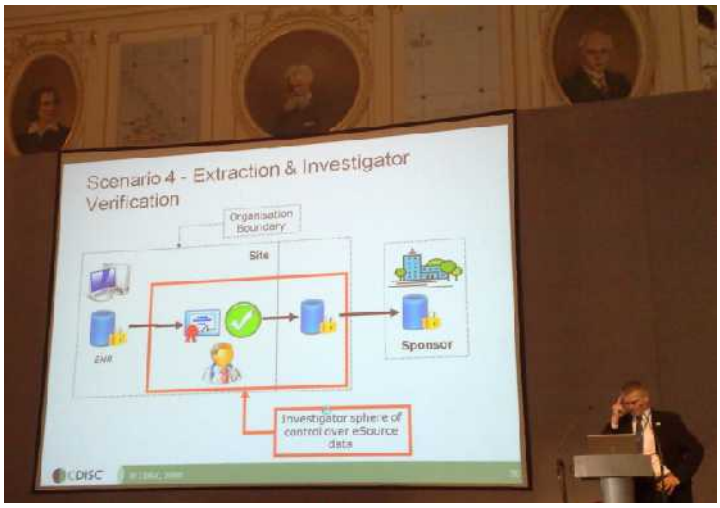
think free, independent from any pressures from e.g. the FDA, that they have resources (students, time and skills), and that they ... just do not need to make money.

Maybe the time has come that academic groups (e.g. in medical informatics) start actively contributing to the CDISC set of standards, or that new academic groups are started for this purpose.

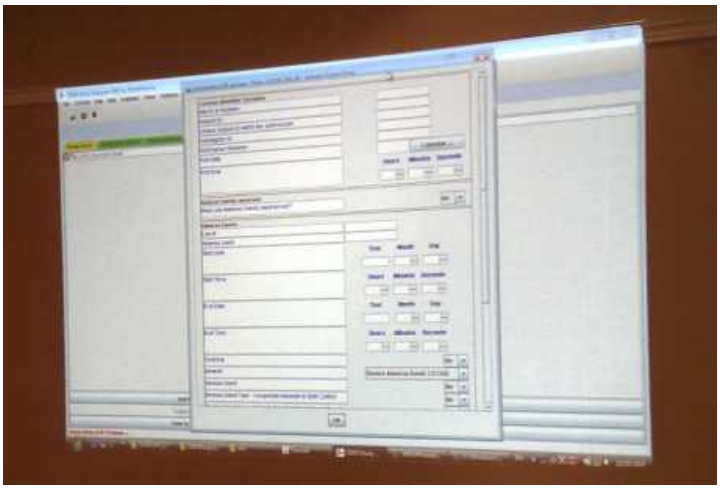
There are many academic groups in drug research (supported or not by the pharma industry), contributing to a faster development of new drugs, so why not have academic groups contributing to the development of IT standards in clinical research?

Other pictures from the European Interchange

Another picture from the SDTM training course



David Iberson-Hurst (on the right) on eSource and EHR Integration



Presenting the ODM Study Designer software at our booth

Doug Bain (Medidata) on ODM extensions for EDC