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

September 2009

Welcome to the end-of-summer issue of our Newsletter. It contains the latest news about standards for data exchange in clinical research, our involvement in them, and how these are implemented in the industry.

SDTM-ETL™ v.1.3-beta now available

In our previous newsletter, we announced the availability of v.1.2 of our popular SDTM-ETL software for providing and executing mappings between operational clinical data and the SDTM standard.
But we were not completely satisfied yet. So, during summer, we worked hard on a new version, for which the beta is now available[1].
The two major new features of the new version are:

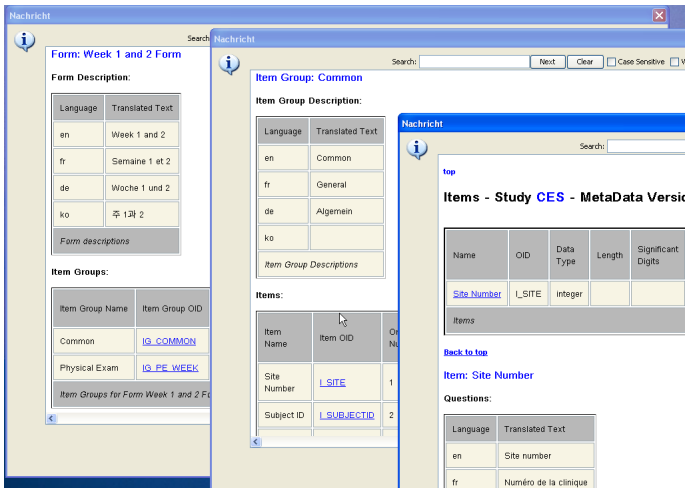
- Use of non-standard SDTM variables: the user can add (sponsor-defined) non-standard SDTM variables. These are treated like any other SDTM variable, but at creation of the SDTM datasets (in SAS Transport 5 or as XML), the data for the non-standard variables is automatically moved to SUPP--datasets, and the exported define.xml is automatically adapted (use of ValueLists).
- Full support for SDTM 1.2 – SDTM-IG 3.1.2. When starting with a new mapping, the user has the choice between SDTM 1.1 (SDTM-IG 3.1.1) and SDTM 1.2 (SDTM-IG 3.1.2). The templates for both the standards, as well as the "CDISC Notes" from both the implementation guides have been implemented in the software. Also the guidance from the latest IG has now been implemented in the software.

As far as we know, our SDTM-ETL software is the first software system implementing the newest version of the SDTM, which was published at the end of last year.

New ODMViewer version available

We also released the newest version of the ODMViewer (version 1.3), allowing to inspect ODM files in a very user-friendly way (inter-hyperlinked HTML documents).

New in this version is that the software uses "deep" hyperlinks, allowing to deeply dive into the details of any data or metadata point. For example, when inspecting the list of used forms in a visit, clicking the hyperlink of a specific form opens a new window with the details of the form, including the used subforms (ItemGroups). When clicking the hyperlink of one of the subforms (ItemGroup), a new window opens with the details of that ItemGroup, including all questions (Items), etc..



This means that it becomes much easier to "dive" into the structure and the details of the ODM.

Similarly, when inspecting clinical data points, a simple click on a hyperlink allows to inspect its metadata (like type, length, associated codelists), and/or to inspect its administrative data (investigator and site details).

A whole new feature is the detailed inspection of **sets of audit trails**. Using this new feature, the whole set of audit trails for a single subject (or group of subjects) or of an individual site or investigator (or groups of them) can be generated and inspected. Also simple statistics can be generated, e.g. in order to detect unusual high amounts of audit records for a site, subject or investigator.

1 Existing customers who purchased the software after June 2007 automatically receive a free upgrade.

Audit Record counts by Subject

Total number of Audit Records: 630

Subject	Count	Percentage
ZBI-00009	40	6.35
ZBI-00001	34	5.40
ZBI-00010	31	4.92
ZBI-00015	31	4.92
ZBI-00023	28	4.44
ZBI-00017	27	4.29

Of course, the earlier developed functionality to generate and inspect subsets of clinical data, and to generate PDF reports is still available and has even been extended.

A **free trial version**, as well as more information about this new version of the ODMViewer is available from our website at: http://www.xml4pharma.com/CDISC_Products/CDISC_ODMViewer.html

RPE – Retrieve Protocol for Execution: a new IHE profile

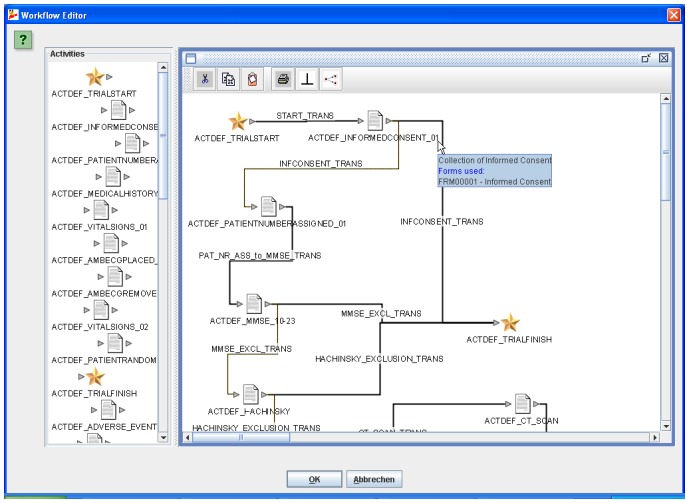
The CDISC Protocol Presentation Group (PRG) is working hard on an ODM extension to describe more of the protocol (or trial design) information in a machine-readable format in the ODM than has been possible until now. The basis for this work has recently been published on the CDISC website.

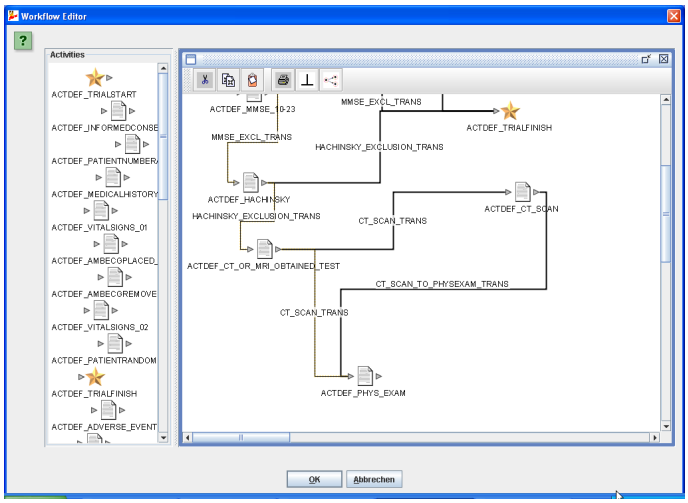
In parallel, an HL7-v3 message is being developed by a third party for the purpose of electronic submissions to the FDA.

IHE "Integrating the Healthcare Enterprise", has recently started a new profile[2] "Retrieve Protocol for Execution" (RPE), investigating how a machine-readable protocol (either as ODM+extension, or as HL7-v3 message) can be used to automate a number of tasks in an hospital information system.

Due to our knowledge of ODM and XML technologies, we have been asked to join this team and have accepted. One of our tasks (not an easy one) will be to investigate whether a mapping between the ODM-extension and the HL7 message can be developed, and to provide a good sample file of a non-trivial protocol in the form of an ODM file with the trial design extension elements[3].

For this, we also extended our CDISC-certified ODM Study Designer software, which was already capable to handle any ODM-extension. Especially, a lot of effort is currently going into the development of a graphical editor for designing activity workflows (one of the main features of the trial design extension).





Two snapshots of our workflow editor for clinical activities defined by a protocol

This extended functionality will become available to our customers as soon as the ODM-extension for trial design is public and final.

2 An IHE profile can be regarded as a "pilot" or "feasibility" project, trying to demonstrate that a technical implementation is possible.

3 It is our believe that the ODM extension will be the first choice for implementation in EDC systems and connected hospital systems, whereas the FDA will require that electronic submissions will be made in the HL7-v3 format.

The success of OpenClinica

I have been following OpenClinica since its start in 2004-2005. OpenClinica is not the first open source project in the clinical research world, so I was pretty skeptical at that time. Other projects, like Visitrial never made it to a great success, and Phosco seems to have disappeared from the "market".

But Akaza Research, the company behind OpenClinica, really made OpenClinica a success. Akaza Research now has a team of 22 people. Their strategy is simple but seems to be working: to provide a free, open source system for everyone, and to offer paid, professional support to "Enterprise" customers, with services including implementation, hosting, training, validation support, study setup etc..

One of the pillars of their success is surely the adoption of OpenClinica by NCI's caBIG™ initiative. In 2005, caBIG was looking for an affordable, if possible open-source system for all their centers, and Akaza won the contract.

Another pillar is probably the interoperability with other systems. Although OpenClinica does not (yet) allow for full study-setup via CDISC-ODM files with metadata (MS-Excel is used for CRF-setup), it has not only excellent ODM export capabilities but also allows for the import of clinical data from ODM files coming from other sources. Furthermore, OpenClinica is, in its architecture, increasingly moving towards being based on ODM itself, rather than just providing import and export capabilities.

OpenClinica is a big hit in the academic world, with over 7,000 community members. Some of them also contribute to the further development of the system, for example by providing translations ("localization") into different languages of the graphical user interface and the documentation, or by extending the software capabilities. The "localization" is surely one of the major contributors to the success outside the USA.

Standards: can we learn from OpenTravel?

In the discussion about how complex/abstract XML-based standards should be, I am currently taking a look at the OpenTravel standard.

I am doing so, as in clinical research, we are currently working with as well the CDISC ODM standard, which is non-abstract at all (i.e. it gives specific, meaningful names to XML-elements) as (still seldomly) with the HL7-v3-XML set of standards, which are very abstract (trying to generalize as much as possible). So in ODM, a form is represented by the "FormDef" and "FormData" elements, a subject by "SubjectData" whereas in HL7-v3-XML we have constructs like "NonPersonLivingObject", which can mean an animal, a plant, a virus, a bacteria (why not call it like that?). HL7-v3 works with Entities, Roles, Participations, Act and Act Relationships, and every XML element needs to be derived from one of these.

The argument for this abstraction has always been that the healthcare world is too big and too complex to be captured in a non-abstract XML format[4], and thus, that an overall, abstract model (the RIM) is necessary.

Big and complex. So is the traveling world.
So I was wondering how OpenTravel deals with these issues.

First of all, let me say something about the scope of OpenTravel. OpenTravel is, just like CDISC and HL7 a non-profit organization. Stakeholders are airlines, hoteliers, car rental companies, banks, creditcard organizations, technology providers, etc..

Inspection of the Implementation Guide (IG) of the 2009A release (773 pages!) reveals that OpenTravel-XML is based on messages (sounds familiar): there is always a request ("---RQ") message and a response ("---RS") message. So there are no attempts to deliver things like "documents" or "archives".

OpenTravel does not attempt to put everything in a single, overall model like the HL7-RIM. They start from use cases, and for each of them build a model and an XML-Schema. Of course there are common schemas, defining reusable data types (like creditcard-type, airport- and airline codes). So in the 2009A release, we found 276 schema files (128 ---RQ.xsd files, 130 ---RS.xsd files, 18 common schema files). So essentially, there are about 130 "pairs of messages".

All element and attribute names are very specific. For example, a seat in an airplane is represented by the element <airSeat>, and a cabin on a cruise ship by the element <selectedCabin>. No attempt was made to define a more abstract type that spans both.

What surprised me the most is that, althoug the use cases can be quite complex, no use of UML[5] has been made at all. I haven't even found the word "UML" in any of the documentation. One of my

4 Rather surprisingly, some HL7-v3-XML elements are named very specific, like <streetName>, <city>, <postalCode>

5 Universal Modeling Language, a modeling method often used in software design – defines classes, methods, etc..

critiques to HL7-v3 is on the blind believe in UML for constructing models, with fully automated creation of XML-Schemas from the UMLs[6]. So it seems to be possible after all to construct models and XML-schemas for big and complex worlds like the traveling world without needing to use any UML[7].

Another observation I made in the OpenTravel XML-Schemas is that they consequently try to use the primitive XML types as much as possible. A float is a float, an integer an integer, a boolean a boolean (so no Yes/No) and a date an ISO-8601 date. No such things as non-ISO dates like in HL7[8], or "flavors of null". Also the CDISC-ODM standard is evolving in the direction of using primitive XML data types as much as possible, but is not completely there yet.

Also very interesting is that OpenTravel has a document "XML Schema Design Best Practices". We do not have such a document within CDISC. Instead, we have the newly erected "XML Governance Team" who's task it is to support the development of new XML-based standards or versions. Maybe the switch to XML-Schema 1.1 (within 2-3 years?)[9] is a big opportunity to develop and publish such a "Schema Best Practices" document for use within CDISC.

A snippet of an OpenTravel XML Response document for the request of fares for a flight between Zürich and Londen on the fifth of March 2005 can be found at the end of this newsletter.

The future of the ODM Checker

One of the software tools that we make available for free (at least for CDISC members) is our ODM Checker. Unfortunatley, although the ODM 1.3 standard is now out for more than 2 years, the Checker only has full support for versions 1.1 and 1.2 of the standard.

We have long doubted whether we should release a new version of the ODM Checker, also supporting ODM 1.3. Reasons where the appearance of other validation tools on the market, and the upcome of new standards like Schematron which may be an alternative to validation software.

In 2007, as his thesis work, a student of the University of Heidelberg developed the basis of a major upgrade of the ODMChecker (essentially a full redesign) but some refinement and the development of a GUI were still necessary.
So recently, we decided to start working on these, and we can now report that we are making good progress, and that we envisage to make a major update of the ODMChecker, with support for ODM 1.3, available before the end of the year.

Of course, the software will be freely available to CDISC members, as well as any companies or institutes with their seat in Germany, as the project was supported by the German government.

EHRs, CDISC ODM and Veterans Affairs

For those who do not know Veterans Affairs (VA), it is a US Government-run system for former military personnel. One of its tasks is to provide health care services. As such, VA has hundreds of facilities and hospitals, and employs about 280,000 people.

What is probably less known, is that VA has one of the largest active Electronic Health Record (EHR) systems in the world. Their VistA system has EHRs of estimated 5,3 million patients. Furthermore, as VistA is open source (!), many non-VA hospitals in the US have implemented VistA as their EHR system. Although interfaces for HL7 CRD/CCD exist, VistA is not based on HL7-v3 XML (it is much older).

VA not only provides healthcare for veterans but spends millions each year researching diseases common to its veteran population. They are using some of our tools to create study designs in CDISC ODM format.

Very recently, one of their visionary people demonstrated us how they generate InfoPath forms out of an ODM study design[10]. These are then optimized, and used in a SharePoint portal that is used as the CDMS/CTMS. These intriguing technologies are now being commercialized. Preliminary information is available at www.infopathedc.com and at www.sharepointctms.com.

6 This often leads to bizarre XML-Schemas

7 It must be remarked here that UML was never designed to be used for the generation of XML-Schemas. The latter has become some kind of fashion in some organizations – probably due to lack of XML knowledge and skills within these organizations.

8 In HL7-v3, February 30 2009 (which does not exist) is represented by 20090230, and is a valid value for a date. In normal XML (datatype xsd:date) the corresponding date 2009-02-30 will be flagged as an invalid date by any validating parser.

9 See the previous issue of our newsletter

10 We did something similar in the past, using XForms however. See http://www.xml4pharma.com/ODMinEDC/Samples.html

Another very interesting fact is that VA does EHR-Clinical Research integration already a long time before this was a discussion point at CDISC! Therefore it is surprising that VA is not strongly involved in the discussions and pilots of CDISC and IHE about integration between healthcare and clinical research. I strongly believe there is so much we can learn from VA, instead of trying to reinvent the wheel.

German CDISC User Group Meeting

There will be a very interesting User Group meeting of the German-speaking CDISC community on September 22. The invitation can be found below.

OpenTravel XML example

```
<OTA_AirFareDisplayRS xmlns="http://www.opentravel.org/OTA/2003/05" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.opentravel.org/OTA/2003/05 OTA_AirFareDisplayRS.xsd" TimeStamp="2005-05-17T09:30:50"
Target="Production" Version="1.000" SequenceNmbr="1" TransactionStatusCode="Start" PrimaryLangID="en-us">
    <Success/>
    <FareDisplayInfos>
        <FareDisplayInfo FareRPH="1" NotAvailableForFareQuotation="false" FareApplicationType="OneWay" ResBookDesigCode="C"
MileageIndicator="true" LastFareUpdate="2005-02-02T04:20:00" FareIndication="public" FareStatus="published">
            <TravelDates DepartureDate="2005-03-05"></TravelDates>
            <FareReference>COW</FareReference>
            <RuleInfo/>
            <FilingAirline CompanyShortName="Swiss" Code="LX"/>
            <MarketingAirline Code="YY"/>
            <DepartureLocation LocationCode="ZRH">Zurich</DepartureLocation>
            <ArrivalLocation LocationCode="LON">London</ArrivalLocation>
            <Restrictions>
                <Restriction>
                    <GlobalDirection GlobalIndicatorCode="EH" MaximumPermittedMileage="579"/>
                </Restriction>
            </Restrictions>
            <PricingInfo NegotiatedFare="false" TicketDesignatorCode="C" PassengerTypeCode="ADT">
                <BaseFare Amount="110800" CurrencyCode="CHF" DecimalPlaces="2"/>
            </PricingInfo>
        </FareDisplayInfo>
        <FareDisplayInfo FareRPH="2" NotAvailableForFareQuotation="false" FareApplicationType="Return" ResBookDesigCode="C"
MileageIndicator="true" LastFareUpdate="2005-02-02T04:20:00" FareIndication="public" FareStatus="published">
            <TravelDates DepartureDate="2005-03-05"></TravelDates>
            <FareReference>CRT</FareReference>
            <RuleInfo/>
            <FilingAirline CompanyShortName="Swiss" Code="LX"/>
            <MarketingAirline Code="YY"/>
            <DepartureLocation LocationCode="ZRH">Zurich</DepartureLocation>
            <ArrivalLocation LocationCode="LON">London</ArrivalLocation>
```

Invitation to the German CDISC User Group Meeting

We invite all our German-speaking CDISC users to our user group meeting that will take place on September 22 at the IBM Forum Stuttgart. The morning will be devoted to User Group Workstreams (ODM, SDTM, ADaM, define.xml). After lunch, a presentation with discussions about the difference between the SDTM 1.1 and the new SDTM 1.2 (with SDTM-IG 3.1.2), is planned, followed by a LOINC tutorial by Sebastian Semler from TMF e.V..

The User Group Meeting is an excellent opportunity to learn about how others in Germany, Switzerland and Austria implement the CDISC set of standards, and to build a network with colleagues in the clinical research world.

The complete invitation and program can be found here.