# THE XML DDI: SOME QUESTIONS AND REMARKS

Thinking this spring about the coming expert-seminar about the DDI, I collected some questions I would like to see addressed to in presentations or discussions. I did send them to the organisers and the probable keynote speakers. Asked by Sami to present theses questions in one of the Friday sessions, I proposed him to hand out a written version I could comment, taking into account what the speakers will already have elaborated on.

These remarks can be read as expectations to the seminar. They have benefited from discussions with Steinmetz colleagues, in particular Marion Wittenberg and I know they have met with the interests of Cor van der Meer.

The DDI is a very interesting matter. Not only because it opens new horizons as to the data distribution techniques, but also because it fosters new thoughts about metadata and metadata structures. For that reason, some **related topics** could also be reviewed (see below). To make good documentation according to the DDI, you have to understand the broader context the DDI is in. A glance into the think tank would be great.

I prefer to let the finnish colleages send out these remarks before the seminar rather than to present them formally in the workshop, in order to give the keynote speakers the opportunity to include some responses - or further questions - in their introducing presentations.

## DATA DOCUMENTATION INITIATIVE

In addition to the basics, which are always welcome to some participants, the following questions are of interest:

a) The operation of the links to specific identifiers inside the codebook.

Example: a varGroup pointing to the variables belonging to it.

I understand the logic of relations; I wonder how they operate. Do you need an appropriate software? Can a style sheet using Xpointer do it? How?

b) The operation of that special DTD "Exchange Tables model" and the type of information it is used for.

c) Which software using DDI are presently under development? Which of the make use of the identifiers? Which of the xml:lang attribute - I mean, to really differentiate between multiple language versions?

In my view, the priorities are not only on making a lot of these XML files. I see them also on programming more applications using the XML-DDI. People interested by the DDI should be able to download an XML-style sheet showing more clearly some functionalities of the new codebook format, for example these internal links cited above or the tables integrated in one of the examples on the DDI Web-page. It should be possible to browse the XML codebooks in a printable format or in some alternative formats, so people get a more concrete feeling of what they can do with it... besides browsing codebooks with Nesstar. If such a style sheet could be presented at Tampere, it would be great.

## RELATED TOPIC I: THE APPROPRIATE RELATIONAL DATA STRUCTURE OR OBJECT MODEL

XML was certainly a good choice as an exchange format, but it is not a basic data management format. Archives tend to operate databases, some relational databases. They will produce XML files as outputs from these databases, confirming the XML DDI in its role as an exchange format. Some archives want even to take information from XML files to put them in relational databases (Steinmetz, see Marion). Hence the following questions:

- What would be the appropriate database structure to serve the DDI? Of course, I don't expect a definitive answer on that question but at least a discussion of the opportunity asking it.
- *Where does the DDI make it eventually difficult to work with a relational model?*
- Now and then, the necessity to develop an object data model for the metadata and data structures pops up in conversations. Could somebody explain what is at stake in these developments and what directions are currently being explored?

## RELATED TOPIC II: XLINK AND XPOINTER

At the IASSIST Conference in Chicago, somebody asked during the DDI workshop which model is more general, the XML DDI or some relational model; Peter Joftis did answer the question with a very short, sibylline phrase: "In general, the XML-DTD is more general than the relational one". I find the point very interesting. Before knowing anything about X-Link and X-Pointer, I was convinced that the relational model is more general. With some hints on these extensions to XML, I am no more such sure. I expect one of the Petes to say more about it.

It would be important to learn somewhat more about what developments these two extensions (X-Link/Pointer) will make possible. I dream of a demo.

Subsidiary question: would the DDI have developed as one unique hierarchy per codebook if these extensions had been there at the start? Or would the XML DDI have come out more... relational?

At the start, the SGML-DDI was presented as an exchange format, the information being kept and managed in various structures at various archives. If the XML DDI represents a more general model, it could be sensible to handle the XML DDI as the basic metadata management structure, not only as an exchange format. Till now, I did consider the relational model as more general but I am ready to change my opinion about it.

Well, to put it on a funny way: I feel our situation in front of XML is somewhat similar the following: you are presented a wonderful new format, say some *.dbf and *.dbt file types, you are told you can make anything out of it, but dBASE III is not yet there, nor Clipper…

## RELATED TOPIC III: NESSTAR USE OF DDI

Most of the participants in the seminar will use the XML DDI primarily in connection with Nesstar. Without making of that DDI seminar a Nesstar seminar, I would welcome an extension of the discussion to the **strategies** the various archives will use to integrate their metadata into the Nesstar system.

This is not only a matter of the structure of the DDI, but also of how Nesstar uses the DDI. An example. There is an attribute xml:lang in the DDI, but Nesstar makes obviously no use of it, for example to allow browsing of the information in various languages or for restricting research to specific languages. I heard there is no way to allow searching on some fields which were translated in English and showing the information in some other national language, as in the IDC.

Archives already have a lot of metadata in various formats. But the structure of that information will not always match the structure of the DDI DTD. What should be done in those cases? To make extensive use of less structured elements, like notes, adapting the originating structure if it is a database, or editing by hand on feeding the DDI?

Certainly, cases will be discussed. I would welcome some coordination effort, making the Nesstar catalogue more coherent than the IDC was. In the IDC, the strategies of the various archives were very diverse. I think it would be useful to organise some level of exchange between the archives concerning the contents, not the techniques of DDI. There is a need for a somewhat more formal organisation of the participation in the archives' Nesstar network. That organisation goes beyond the organisation of the technical work brilliantly done by the Nesstar Consortium.

## RELATED TOPIC IV: BROAD ACCESS TO METADATA THROUGH DUBLIN CORE ELEMENTS

The aim of DC is to make more selective searches possible on the Internet at large. DDI and Nesstar stress the importance of filling out the DDI elements that are explicitly mapped to the DC. All right. But in my view, this is of no use if these metadata elements are concealed in specific applications like Nesstar, where you have much more detailed metadata anyway. The DC elements become useful if they are accessible to non-specific applications like spiders.

As a consequence, besides making metadata for Nesstar or the Virtual Data Centre, you have to make the DC mapped elements available at large. Is there some model about how to do it?

- At least, you can describe the data catalogue on your web server to give crawlers chances to find you.
- You could make all your dataset descriptions available for crawlers, duly meta-tagged for DC, giving maximum publicity to the researchers and institutes that contribute to the archive.

See you soon in Tampere

Reto

Reto Hadorn

SIDOS
Ruelle Vaucher 13
CH - 2000 Neuchâtel

reto.hadorn@sidos.unine.ch
www.sidos.ch