

From: "Thomas White" <Coevtmw@omh.state.ny.us>
To: <cmcdonald@regenstrief.org>
Date: 11/9/2003 7:53:01 AM
Subject: Re: Re Survey Instrument Content

Clem-

Thanks for the vote of confidence.

I have not been able to make the changes as quickly as desired. I had planned to send you the Excel file I'd prepared for the Salt Lake meeting, but after reviewing your comments, I wanted to re-organize the content to reflect your proposals. Unfortunately, I haven't had the time for that yet. So, I'm attaching that original Excel file so that you can see the original thinking. One worksheet is Orig-LOINC-Nursing-Instruments * this is the all of the original content from LOINC on assessment instruments. The next worksheet is LOINC-to-Dialogix, which re-organizes the content from Orig-LOINC-Nursing-Instruments into a form usable by Dialogix. The final worksheet is Dialogix-to-LOINC, which takes four sample Dialogix-encoded instruments and adds the content necessary for LOINC. The four instruments are the Hamilton Depression Inventory (HAM-D), the Dr. Ware's SF-36®, McLean Hospital's Basis-32®, and several modules from the Diagnostic Interview Schedule for Children (DISC). Row 291 on the Dialogix-to-LOINC sheet is highlighted yellow since that is where I stopped when manually specifying the Property, Time_Aspect, and Scale_Type, which don't natively exist in Dialogix.

I'm also attaching two Dialogix instruments for your reference, so that you see how they could be pulled out of the LOINC tables. The .htm versions are a view of the logic file, and the .txt files are the actual Dialogix syntax files (as a tab separated table). SF-36 is an auto-scoring implementation of Dr. Ware's Short-Form-36®. LIV-HIV is conversion of one of the instruments currently included in LOINC into a Dialogix-ready instrument.

So, I'm still working on making those changes, but I hope these documents will let us continue our discussions in the interim.

I'm looking forward to seeing you here at AMIA.

/Tom

>>> Clement McDonald <cmcdonald@regenstrief.org> 10/14/03 12:22PM >>>
I like everything you proposed and we will accommodate all of that content.

In the discussion, you emphasize the encoding of this content as XML. That is OK, but you should realize that we store everything in a relational data base (Access) and our tools are not centered on XML. I assume that the actual encoding of the content should not matter to the end result.

The general principle that we have worked on so far, is that most questions (or variable or measurement) stand on their own. So that they can be understood at least on the surface if moved to a different questionnaire.

In that context we would think about all of the scoping being part of

the question--though that would not have to be operationalized in how you ask the question. So we would mostly - not think of the question as having two parts- one part that is the introduction that included scoping and one part that would be the specialization of that scoping.

So the easiest way to think of the questions would be

In the last week have you had any feelings about sadness?

One could consider a slight modification via embedded delimiter (or tags) within the question that indicated the part that might be parsed out.

E.g.

<Fixed_part>In the last week have you had</Fixed_part>
<variable_part>Feelings about sadness<variable_part>. Alternatively, we could have two fields for the question, one that carried what I call the fixed part and one the variable part. Again the principle would be the same in a particular instrument you could look at the order, find fixed parts that were the same, ask the fixed part first and then ask the variable parts for the individual questions. This gets trickier if there is nesting of the fixed parts that you would have to represent. By the way, you should not pay attention to the tags I used. There is probably a much better way to say this.

You could in theory include the fixed part (or introduction) in the questionnaire part--but then the questions could not be mixed and matched at all. That is, you could not invent a new question by taking one question from one survey and another from another without a lot of twisted poly hierarchy relationships.

Am not sure if you had been thinking of the LOINC level questions as I described them above. We sure would want to think of any clinical questionnaires (about smoking, amounts of pain, etc, mini mental status), that way.

The panel data structure has separate relationship table that defines the relationship between the questions and the survey instrument. We can represent the presentation order in this relationship table (so the order in a particular survey can be fixed), could also store simple skip logic as an extra field--but would prefer that skip logic would be easy to implement and very general.). My assumption has been that if the answer list changes that defines a new question. If that is correct, then the answer list would be anchored to the LOINC term it self (not the relationship table). If we decided that the answers could change within a specific LOINC code, per survey instrument then we would link the answer lists to the relationship table. However that would again mean that a survey question might have a quite different meaning depending on the context of the survey instrument and we reduce the option for question sharing. The panel structure is already hierarchical-

Would like to be clear about what additional attributes you would want attached to the question in relationship to the survey instrument and the options for expressing skip logic. The XML schema seemed pretty general - and I am not the best schema reader.

Would almost assume that we would wan fields for required and optional

and conditional as we have for panels.

I think we should bite the bullet and store answers as structured things rather than as a list of phrases in a memo field. That way we can represent the text of the answer and the code of the answer and/or describe the answers in terms of a well defined external code system (like ICD9 or DSM 3). That way we could also provide the structure needed by a tool that gathered the survey instrument. We already have such a structure in the Attachment table and could not do without it.

I am not clear about the kinds of content that one would see in survey questions source. We will provide a place for it, but I would expect it would have some structure and we might want another table to store the details once for many questions. This is also something that comes up in Attachment questionnaires. Wouldn't we want similar attributes for the survey instruments source? (Would they be different sources?) Could we assume that the source of the questions could be inherited from the source of the survey instrument?

I don't understand your questions about method and class when a bunch of instruments use the same questions. Class is not a structural part of the name. So we can do with it what we want. For class we would either assign the question to the survey instrument where it first appeared, or to a class that was high enough in the tree of classes that there was no confusion. If we build these as hierarchies within we don't need the class to show the relationships for which we had used it in the past.

Regarding method, what we need is some simple rule to disambiguate questions (IF WE DONT USE THE WHOLE QUESTION AS PART OF THE DEFINING KEYS OF THE FIELD). So again, if we assigned it a name that corresponded to the first instrument in which it was found and the year published- that might do.

Am not sure what you mean by scoping groups within an instrument (Is this same as a hierarchy of questionnaires)?

Again for clinical questions we would really like to encourage re-use of questions for different instruments. And discourage the idea that every study is an excuse to define a whole new set of questions.

We have not discussed where to store the calculations or how to represent the formula (there is a field for formula in LOINC)

We do have some cases in LONIC where one question refers to another. So one question will identify a tooth and the next question will express some issue about a tooth. Not certain that this capability would work in your context. If it did, it would require more care in the automatic definition of questions.

Now to return to some more basic questions

We have two choices on the question text.

- a) include it as the component verbatim
- This approach has the advantage of automatically

distinguishing questions that are different.

- It has the dis-advantage of not including any sense of the general kind of information the query is able to capture.

b) include it as a separate field (or fields) but then includes some meaningful descriptor (or descriptors) in the component.

c) In either case, we have the opportunity to code the time scope (regardless of how we include the statement of that scope in the question).

I think having a good meaningful but short descriptive concept in the component would provide more utility than the full text. But I thought I heard you say it would not be feasible to keep up with the invention of such meaningful concepts. In that case we should go with the question as the component because we could actually keep up with that and avoid collisions. So the answer boils down to what you can really absorb.

We could in principle use the time aspect of the LOINC code to represent the time scoping even though we would probably have to include it the statement of that time scope in the question to assure that it is asked right. There was some negative discussion about this choice at the meeting, but think there are some strong parallels to the clinical case. In the clinical case we report an 8 hour shift and then indicate the total urine output that occurred during that shift or the frequency of some event. So if the questing is asking how often over a period of time- this would be very parallel.

The one difference might be do we count forward or back ward from the observation time. In the clinical lab we usual record the observation time as the start time and then note that it was 24 hour urine. (But I may be wrong- it might more often be the end time- which would be more parallel to the questionnaire case. Would need to find out how clinical measures about urine output and max temperatures are recorded. Is the observation time the end of the shift when nurses are looking backward for X hours or is it recorded as the beginning of the shift...? If there is a difference in the relationship of the time extent to the recorded observation time as compared to usual clinical reports we could find a way to encode that difference too.

Would like to get some examples of some of the content you want to store as soon as possible and then get some decisions about exactly what to store so that we can create the needed fields in our data base.

Thanks

Clem
(ay)U

IMPORTANT NOTICE: This e-mail is meant only for the use of the intended recipient.

It may contain confidential information which is legally privileged or otherwise protected by law. If you received this e-mail in error or from someone who was not authorized to send it to you, you are strictly prohibited from reviewing, using, disseminating, distributing or copying the e-mail. PLEASE NOTIFY US

IMMEDIATELY OF THE ERROR BY RETURN E-MAIL AND DELETE THIS MESSAGE FROM YOUR SYSTEM.
Thank you for your cooperation.

CC: <coshuff@ihc.com>, <cosmatne@ihc.com>, "Thomas White"
<Coevtmw@omh.state.ny.us>