BIEN 3 - Task #289

look for formal mapping mechanism

12/01/2011 04:32 PM - Aaron Marcuse-Kubitza

Status:	Resolved		**Start date:**	12/01/2011
Priority:	Normal		**Due date:**	
Assignee:	Aaron Marcuse-Kubitza		**% Done:**	100%
Category:			**Estimated time:**	0.00 hour
Target version:				
Activity type:				

Description

Conference call:

- look into VegBranch's way of capturing mappings and metadata
- ~~look into Altova XMLSpy's graphical generation of XPaths~~
- ~~look into NVS mapping tool~~
- **~~determine if XQuery's superset of XPath will do the queries we want~~**: no
- ~~research~~ [Bourret's XML-ER mapping](#)
  - [MapManager class](#)/org.xmlmiddleware.xmldbms.tools.MapManager.html-
- ~~research XQuery pointer dereferencing with higher-level operators~~
- read ~~CLIO~~ articles and look up relevant references: found site w/ screenshots of mapping tool
- ~~look into RDF querying with SparQL~~

History

#1 - 12/01/2011 04:33 PM - Aaron Marcuse-Kubitza

- Description updated

#2 - 12/01/2011 04:36 PM - Aaron Marcuse-Kubitza

- Description updated

#3 - 12/01/2011 04:40 PM - Aaron Marcuse-Kubitza

- Description updated

#4 - 12/05/2011 12:09 PM - Aaron Marcuse-Kubitza

Mike Lee's explanation of the VegBank XML serialization format: (e-mail on 2011-12-2)

My recollection is that our initial developer had developed something really simple like:

<plot>
<latitute>35</latitude>
<longitude>-77</longitude>
</plot>
<observation> ...

But it didn't work well because it didn't contain the entire data model and it didn't link the elements together very well. So we settled on embedding the foreign key elements within the foreign keys themselves. Some foreign keys are "inverted" keys in that instead of representing the foreign element, we include all related entities in the parent element, so for example all taxonObservation records go within the observation element.

To allow schema declaration of fields that might have the same name, but appear in different tables, we use table.field structure for the field names, not unlike RDF.

<table>
<table.field1>value</table.field1>
<table.foreignKeyName><foreignTable> ... </foreignTable></table.foreignKeyName>
<invertedElement1> ... </invertedElement1>
<invertedElement1> ... </invertedElement1>
<invertedElement2> ... </invertedElement2>
</table>

From there, it wrote itself just about. I wrote an XSL stylesheet that takes our data model XML and creates a schema document. So whenever we change the data model, we can autogenerate XML schema to validate VegBank XML files.

We have a simplified schema for adding data from VegBranch because VegBranch knows what data already reside on VegBank. When it encounters such data, it doesn't export the full data for that element, just a reference to the extant data in VegBank, via the accessionCode.

Here's the page on the xml
vegbank.org/xml

Hope that helps. We didn't do a lot of deciding about it. This just seemed the right way to do it. The downside since it's all nested, is that it can get quite large as repeated elements get repeated.

#5 - 12/07/2011 12:38 PM - Aaron Marcuse-Kubitza

- % Done changed from 0 to 10

Altova XMLSpy's graphical generation of XPaths:

- **summary: XMLSpy and Oxygen XML both have Copy XPath commands (Oxygen just for data), but of course don't handle VegX's custom pointers and thus are of limited use for our pointer-heavy VegX mappings**
- XMLSpy has a Copy XPath right-click command for XML *schemas*
- Note that Oxygen XML also has a Copy XPath right-click command, but only for XML *data*
- use XPath Analyzer
  - "XPath 1.0 / 2.0 Builder: The XMLSpy® 2012 XPath builder helps you define XPath 1.0 and 2.0 expressions with a simple point-and-click interface. You simply select an element or attribute in your XML data file, and the "Copy XPath" command will automatically copy the corresponding XPath expression to the clipboard."
  - "Intelligent XPath Auto-completion: As you're composing an XPath expression in Text View, Grid View, or in the XPath Analyzer window, XMLSpy® 2012 provides you with valid XPath functions, as well as element and attribute names from the associated schema and XML instance(s)."

#6 - 12/07/2011 12:45 PM - Aaron Marcuse-Kubitza

- % Done changed from 10 to 20

XQuery:

- XQuery Tutorial
  - XQuery iterates over XML documents stored in database text fields
  - XPath is only used within each XML document; a SQL variant is used to search the database itself

#7 - 12/07/2011 01:11 PM - Aaron Marcuse-Kubitza

- % Done changed from 20 to 30

Bourret's XML-ER mapping**:**

- **summary: his various mapping methods are already used by VegBank and VegX**
- simple and complex XML to database mappings are basically identical to the two versions of VegBank XML described by Mike Lee above
- choices and optional children are mapped to nullable fields
- repeated children are mapped to child tables with foreign keys to their parent
- if XML node order is significant, need to store it in a separate table
- IDREF attributes are mapped using id attributes and fields, exactly the way VegX does it
  - note that there are two options for representing IDREFs (pointer targets): VegX-style as described by Bourret or VegBank-style as a child of the pointer field

#8 - 12/07/2011 01:23 PM - Aaron Marcuse-Kubitza

- Description updated

- % Done changed from 30 to 50

updated to do list

#9 - 12/07/2011 02:01 PM - Aaron Marcuse-Kubitza

- Description updated

- % Done changed from 50 to 60

IBM Clio**:**

- "Clio then also interprets these mappings to construct a set of database queries that transform and integrate source data to conform to the target schema"
- "For a demo or information on code availability please contact Howard Ho (lastname @ almaden.ibm.com)"
- last updated 2007

#10 - 12/07/2011 02:18 PM - Aaron Marcuse-Kubitza

- Description updated

- % Done changed from 60 to 70

RDF SPARQL**:**

- SELECT-style queries for RDF data
- uses concise Turtle syntax for WHERE conditions

#11 - 12/14/2011 01:59 PM - Aaron Marcuse-Kubitza

- Description updated

- % Done changed from 70 to 80

Got NVS mapping tool from Nick Spencer, which is on nimoy in /home/bien_shared/raw_data/nvs/VegX/

#12 - 12/14/2011 02:02 PM - Aaron Marcuse-Kubitza

- Description updated

#13 - 04/02/2012 08:55 AM - Aaron Marcuse-Kubitza

- Status changed from New to Resolved

- Assignee set to Aaron Marcuse-Kubitza

- % Done changed from 80 to 100