BIEN 3 - Task #286

CSV-XML-database mapping script

11/17/2011 10:07 AM - Aaron Marcuse-Kubitza

Status:	New	**Start date:**	11/17/2011
Priority:	Normal	**Due date:**	
Assignee:	Aaron Marcuse-Kubitza	**% Done:**	90%
Category:		**Estimated time:**	0.00 hour
Target version:			
Activity type:			

Description

Python script to map CSV, XML, and database datasources to each other, using a map spreadsheet when needed

---

History

#1 - 11/17/2011 10:48 AM - Aaron Marcuse-Kubitza

- *Description updated*

#2 - 11/21/2011 05:57 PM - Aaron Marcuse-Kubitza

- *Subject changed from CSV to XML conversion script to CSV/XML to XML conversion script*

- *Description updated*

- *% Done changed from 80 to 50*

#3 - 11/23/2011 05:39 PM - Aaron Marcuse-Kubitza

- *% Done changed from 50 to 60*

I updated the data2xml script (demo at nimoy:/home/bien_shared/svn/scripts/data2xml/test) to support the pointer format in the updated mappings.

#4 - 11/30/2011 01:35 PM - Aaron Marcuse-Kubitza

- *Subject changed from CSV/XML to XML conversion script to CSV-XML-database mapping script*

- *Description updated*

Merged in XML to database conversion script

#5 - 12/01/2011 10:53 AM - Aaron Marcuse-Kubitza

I've merged data2xml and xml2db into one script called map, which can be run on nimoy at /home/bien_shared/svn/scripts/map . The tester can be run at /home/bien_shared/svn/scripts/util/test_map and outputs VegX to /home/bien_shared/svn/scripts/util/NYSpecimenDataAmericas.test.xml . There is also a wrapper, map2vegbank, which accepts VegBank XML and outputs to an (empty) VegBank database on nimoy.

The map usage is similar to data2xml and xml2db, and can be obtained by running map with no arguments. In general, you set database access info in environment variables, pass any mapping spreadsheet as a single command-line argument, and the CSV/XML data is input from STDIN and output to STDOUT or a database.

#6 - 12/06/2011 04:26 PM - Aaron Marcuse-Kubitza

- *% Done changed from 60 to 70*

We are now able to import NYBG data directly into VegBank, using the map2vegbank script on nimoy at /home/bien_shared/svn/scripts/map2vegbank . To demo it, run /home/bien_shared/svn/scripts/test/map . I ran it in commit mode, so you can browse the data it loaded by running /home/bien_shared/svn/scripts/util/psql_vegbank .

(Note that test/map currently produces errors about timestamp input syntax, because SALVIAS uses unconventional data formats that I am in the process of writing a conversion function for. Also note that it will say Inserted 0 rows because the data has already been loaded.)

#7 - 12/14/2011 01:54 PM - Aaron Marcuse-Kubitza

- *% Done changed from 70 to 80*

Added support for database and XML inputs

#8 - 12/15/2011 01:03 PM - Aaron Marcuse-Kubitza

- % Done changed from 80 to 90

I added support for mapping XML to XML, which will enable us to process NVS's VegX data, and eventually also their internal XML data. I added tests to verify that mapping data straight through to VegBank produces the same results as outputting to VegX first and then mapping the VegX, which it now does after some bug fixes.

You can run the tester on nimoy with the command: make test --directory=/home/bien_shared/svn/scripts/