Svn Cleanup 2008

Motivation:	Cleanup SVN repository
Contact:	Adrian Custer
Tagline:	SVN CLEANUP

This proposal is now complete!

The SVN has been cleaned out and refactored as of revision 30258. The new SVN repo went live on 2008-05-14. All commits should appear sequentially as of revision 30262.

The geotools *repository is now: http://svn.geotools.org/

The main branch of development, "trunk", is available as http://svn.geotools.org/trunk/.

Description

The svn repo is huge to the point of being uweildy, mixes both uDig and Geotools, contains big image files which have been deleted but remain in the repository, contains files that have been added multiple times rather than being svn copied, and have other miscellaneous errors. This proposal brings love to the repository, cutting its size in half (3 GB -> 1.5 GB) and getting things on a better footing going forward.

The **user visible** changes will be:

1. the repo will be shutdown briefly while we run the cleanup, and
2. the repo will have moved name and up a directory and may have moved server.

Following this, everyone will have to do a new checkout. Because we are doing both a change of location and a change of structure, it looks like fighting through using 'svn switch' will be more of a struggle than it's worth. Probably it's easier to save any non-committable changes in an 'svn diff' generated patch file and apply that to the new checkout but to each their own.

The old repository will not be touched so uDig can keep going and so we have a clean fall back option if anything goes wrong.

Going forward, we expect uDig to undergo a similar cleanup and also move off to a new server at which point geotools can reclaim the svn.geotools.org domain (although by then we may also move on to OSGeo).

Proposed work

We have developed a multi-step process to clean the repo. We start with an 'svn dump /path/to/repo > dumpfile' of the repo to get a dumpfile. (These files are mostly text but with binary blobs inside them.) We then:

1. run eight svndumpfilter scripts to exclude uDig (gain 1.2 GB), drop huge image files (gain 0.2 GB) and clean other bits and pieces (gain a further 0.1 GB) (These are attached.)
2. run a java cleanup file which detects multiple adds of identical files and replaces those with an svn copy instruction (this saves a further 80 megs). (Also attached)
3. hand edit the result to add instructions to add parent directories which are now missing and to drop the BZR commit mess.
   The result is an svn dumpfile which we can reload into an svn repository with the 'svn create /path/to/repo'

and 'svn load /path/to/repo < editeddump' commands.

We propose the following:

1. shutdown the repo (sometime in early May)
2. have Refractions post a copy of the repo or a dump file
3. grab the file, run our cleanup scripts, test the result, and ship the clened dump back
4. have Refractions admins load the file into a new repo in the same location.
5. announce and have everyone 'svn switch' to the new repo.

The developer guide will have to be updated, notably the source code page.

This proposal does not address the desire to move to a newer version of the svn server---our host refractions has internal constraints that prevent it. That can be our incentive to move to OSGeo infrastructure.

Needed coordination.

We need consensus from the major participants in the Geotools community and then need to coordinate with refractions, with the uDig community and with Geotools developers and users to schedule the actual work.

Status

The work is tentatively scheduled for the week of the 12th of May 2008, subject to Refractions' ability.

DONE

The svn repo has been split from the common geotools/uDig/geovista repo. The current SVN is about half the size of the old.

Resources

Name	Location	Date
3	Mar 02, 2008	
0	Mar 02, 2008	
6	Mar 02, 2008	
1	Mar 02, 2008	
0	Mar 02, 2008	
2	Mar 02, 2008	
0	Mar 02, 2008	
6	Mar 02, 2008	

0 Apr 02, 2008
Apr 02, 2008
step01_keepGtnotUdig.sh

Children: