Spreadsheet Auditing and Change Analysis

John Nash

(jcnash@uottawa.ca)

Neil Smith

(neil.f.smith@sympatico.ca)

Andy Adler

(adler@site.uottawa.ca)

Initial Motivations

- Course mark management:
  - Multiple teaching assistants submit marks as spreadsheet files after each assignment
  - Merge into master
  - Re-submissions after the merge
- Easier methods to review and check spreadsheet models for "bugs"

Spreadsheet Auditing
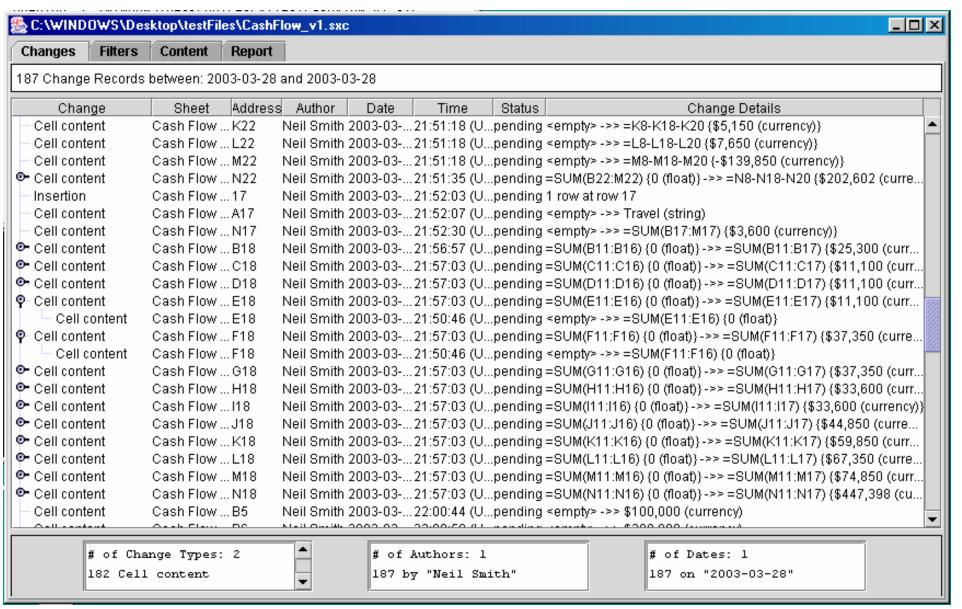# - a 2-part task

- A tool and technique that presents the status and changes in a spreadsheet file

- A Server-based system to ensures overall integrity and continuous record of changes in all relevant files

SSScan – the audit tool

- Tool for auditing spreadsheet files
  - Change tracking + content analysis
- Reads OpenOffice.org **Calc** file format
  - Underlying XML structure
  - *Calc* Reads/writes Excel files (up to 97/2000)
  - Open Source taking hold in various industry sectors
  - Java stand-alone program
    - Multi-platform operation (Windows/Mac/Unix)
    - Auditor / user not required to have *Calc*

Change display panel

Filter panel

Spreadsheet content panel

Spreadsheet Report panel

Screenshot of application window titled "C:\WINDOWS\Desktop\testFiles\CashFlow_v1.sxc" with tabs: Changes, Filters, Content, Report.
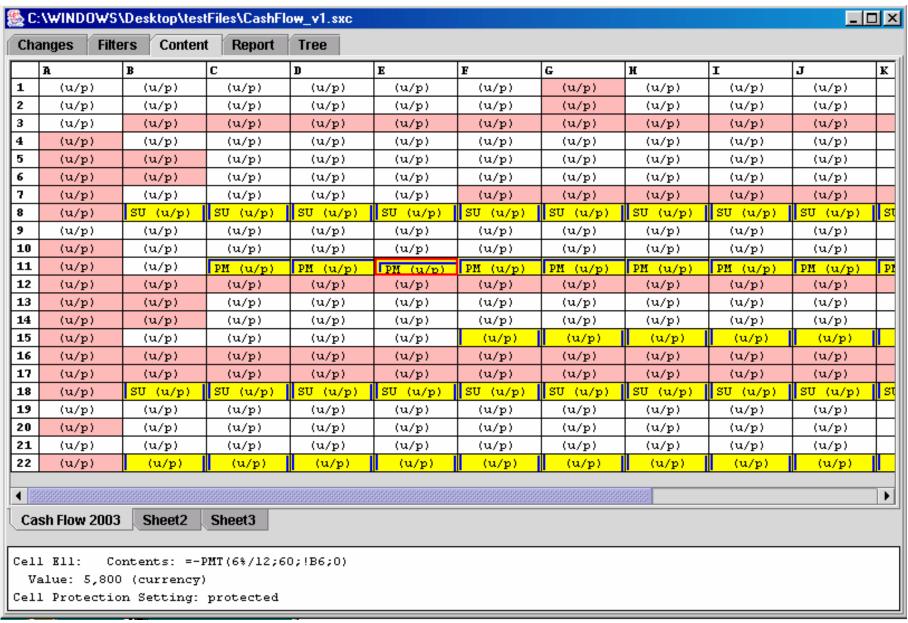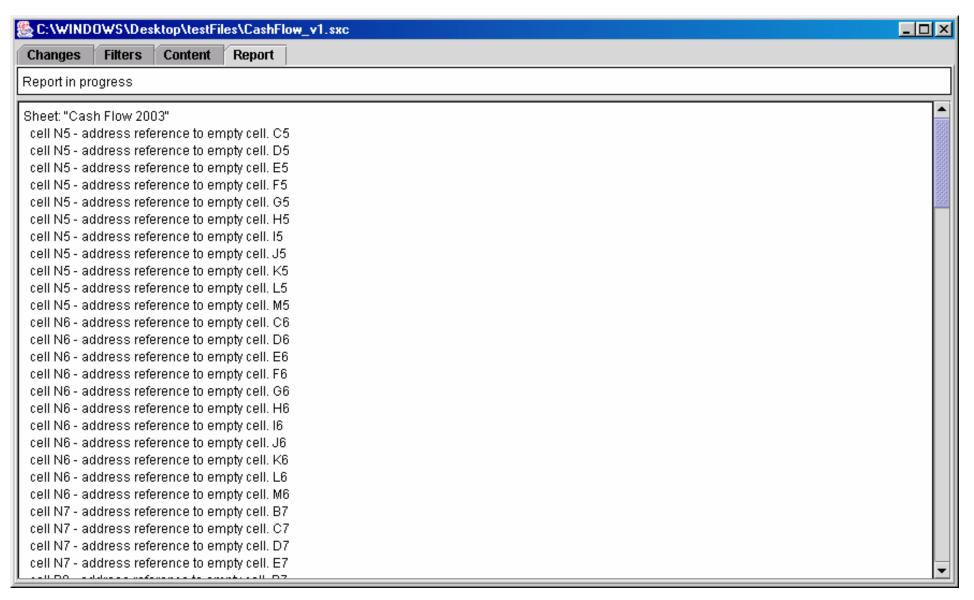
Report in progress

```
Sheet: "Cash Flow 2003"
  cell N5 - address reference to empty cell. C5
  cell N5 - address reference to empty cell. D5
  cell N5 - address reference to empty cell. E5
  cell N5 - address reference to empty cell. F5
  cell N5 - address reference to empty cell. G5
  cell N5 - address reference to empty cell. H5
  cell N5 - address reference to empty cell. I5
  cell N5 - address reference to empty cell. J5
  cell N5 - address reference to empty cell. K5
  cell N5 - address reference to empty cell. L5
  cell N5 - address reference to empty cell. M5
  cell N6 - address reference to empty cell. C6
  cell N6 - address reference to empty cell. D6
  cell N6 - address reference to empty cell. E6
  cell N6 - address reference to empty cell. F6
  cell N6 - address reference to empty cell. G6
  cell N6 - address reference to empty cell. H6
  cell N6 - address reference to empty cell. I6
  cell N6 - address reference to empty cell. J6
  cell N6 - address reference to empty cell. K6
  cell N6 - address reference to empty cell. L6
  cell N6 - address reference to empty cell. M6
  cell N7 - address reference to empty cell. B7
  cell N7 - address reference to empty cell. C7
  cell N7 - address reference to empty cell. D7
  cell N7 - address reference to empty cell. E7
```

Server-based system

- Need to protect the file(s), control access
- Many issues
  - Efficient operation
  - Minimal fuss and bother for user
  - Maximal security and freedom from errors
  - Version history and checkpointing
  - Cross-platform usage
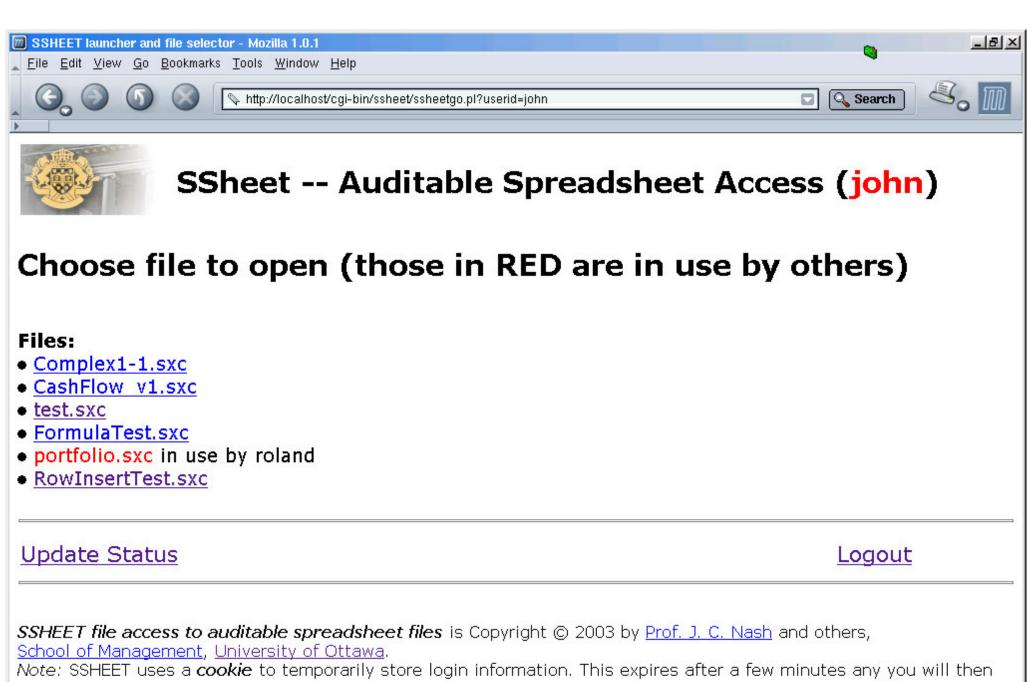
Server architecture

- Graphics in VNC (virtual networked computing)
- Apache server
- Highly customized configuration of oocalc
- Perl scripts to configure and glue together components
- Web based interface to allow cross-platform use

Server architecture (cont.)

- Runs own password/user base
- CVS back-end to provide for versioning and checkpointing
- DB_File::Lock file locking
- Many "fixes" needed to components
- Advantages and obstacles with each choice made in the implementation…

SSHEET User Login

Enter your identification information

Userid?   john

Password?   ****

[Login]

Note: Your browser must allow a cookie to be set to permit the login.

SSHEET file access to auditable spreadsheet files is Copyright © 2003 by Prof. J. C. Nash and others,
School of Management, University of Ottawa.
Note: SSHEET uses a **cookie** to temporarily store login information. This expires after a few minutes any you will then have to log in again.

SSheet -- Auditable Spreadsheet Access (john)

Choose file to open (those in RED are in use by others)

Files:
- Complex1-1.sxc
- CashFlow_v1.sxc
- test.sxc
- FormulaTest.sxc
- portfolio.sxc in use by roland
- RowInsertTest.sxc

Update Status   Logout

File Edit View Go Bookmarks Tools Window Help

http://localhost/cgi-bin/ssheet/oolauncher.pl?filename=test.sxc&userid=john

Search

USER: **john**, FILE: **test.sxc**, After File->Exit from Spreadsheet click here to exit

File Edit View Insert Format Tools Data Window Help

Tahoma 10 **B** *i* U A ≡ ≡ ≡ ≡

A7 Σ = Date

	A	B	C	D	E	F	G	H
1			Internet Institute – Ottawa					
2			**Instructor Schedule**					
3			**John Nash**					
4	18 Nov 02							
5								
6								
7	**Date**		**Appointment**		**Status**	**Start Time**	**End Time**	**Facilities**
8	**Nash, John**							
9	03/18/2003		Practical Forecasting for Managers		Pending	9:00 AM	4:30 PM	Classroom A, Internet Ins
10	06/11/2003		Practical Forecasting for Managers		Pending	9:00 AM	4:30 PM	Classroom A, Internet Ins
11								
12								
13								
14								
15								
16								

Schedule / Sheet2 / Sheet3 /

Applet started.

Launch | Shell No 2 - Kon... | SSHEET User L... | SSHEET - Mozill... | Screen Capture | 1 2 3 4 | 02:36 pm

files/

Current directory: [ssheet] / files

File	Rev.	Age	Last log entry
Parent Directory			
CVS/			
john/			
quon/			
roland/			
CashFlow_v1.sxc	1.1	47 hours	Initial Version: setup
Complex1-1.sxc	1.1	47 hours	Initial Version: setup
FormulaTest.sxc	1.1	47 hours	Initial Version: setup
RowInsertTest.sxc	1.5	46 hours	Modified by quon
portfolio.sxc	1.1	47 hours	Initial Version: setup
test.sxc	1.1	47 hours	Initial Version: setup

Show only files with tag: All tags / default branch Module path or alias: files/   Go

Status and prognosis

- SSScan is working and quite stable
  - But filter customization is important to make audit efficient for particular situations

- Server configuration and setup is tricky
  - Large number of disparate components
  - Many customizations for efficiency
  - Interaction with OS and network
  - Some weaknesses, but with known "fixes"

Status (cont.)

- Need a viable business model
- Current view:
  - "Package" customization of SSScan and configured server hardware
  - Release each customization as Open Source, but delay installer until experience is gained

Structure: Apache – SSL

- Config
  - SSL
  - Directories
  - httpd.conf
- Runs as www-data, member of group ssheet
- Cookies
- Login --> database of users/passwords & files/permissions

Apache-SSL cont.

- File locking & versioning
  - » --> ssheet userdb {users & shuids & files}
  - » ssheet-userdb.lock
  - – Copy file to indir
  - – Grow user name
  - – (select new random password)
  - – Create web page
    - » Active content for vnc
    - » Custom vncviewer.jar
  - – Check in file in outdir (CVS manages "real" changes

SHUIDs: a "farm" of pseudo-users

- Runs as ssheet-uid**** (start-ssheet.sh opens m sshuid's)
- Member of www-data
- Start vncserver --> xstartup (perl script)
- Wait on indir (group read)
- Copy to procdir (shuid only)
- Edit user
- (insert new password for vnc)
- Oocalc initiated with special profile
- (mechanism to kill oocalc)
- Move file to outdir
- (kill password)

CVS Server

- /var/www/ssheet/repository
- INIT:
  - » Take each file and check in
- CHECKIN by shuids
  - » Comment string from www-data
- Web interface to versioning
  - » Modified CVSWeb accesses repository

Admin Interface (to do)

- Add new files to repository
  - » Verify change recording is "on"
- Sophisticated import
  - » Set user name
  - » Set previous history
- Handling of "duplicate" files
- Handling of different permissions for different users
- User creation, modification, deletion