Generating Web Services for Statistical Survey Packages from Domain-specific Visual Languages

22 May 2007

THE UNIVERSITY OF AUCKLAND

NEW ZEALAND

Te Whare Wānanga o Tāmaki Makaurau

Chul Hwee Kim, John Hosking & John Grundy
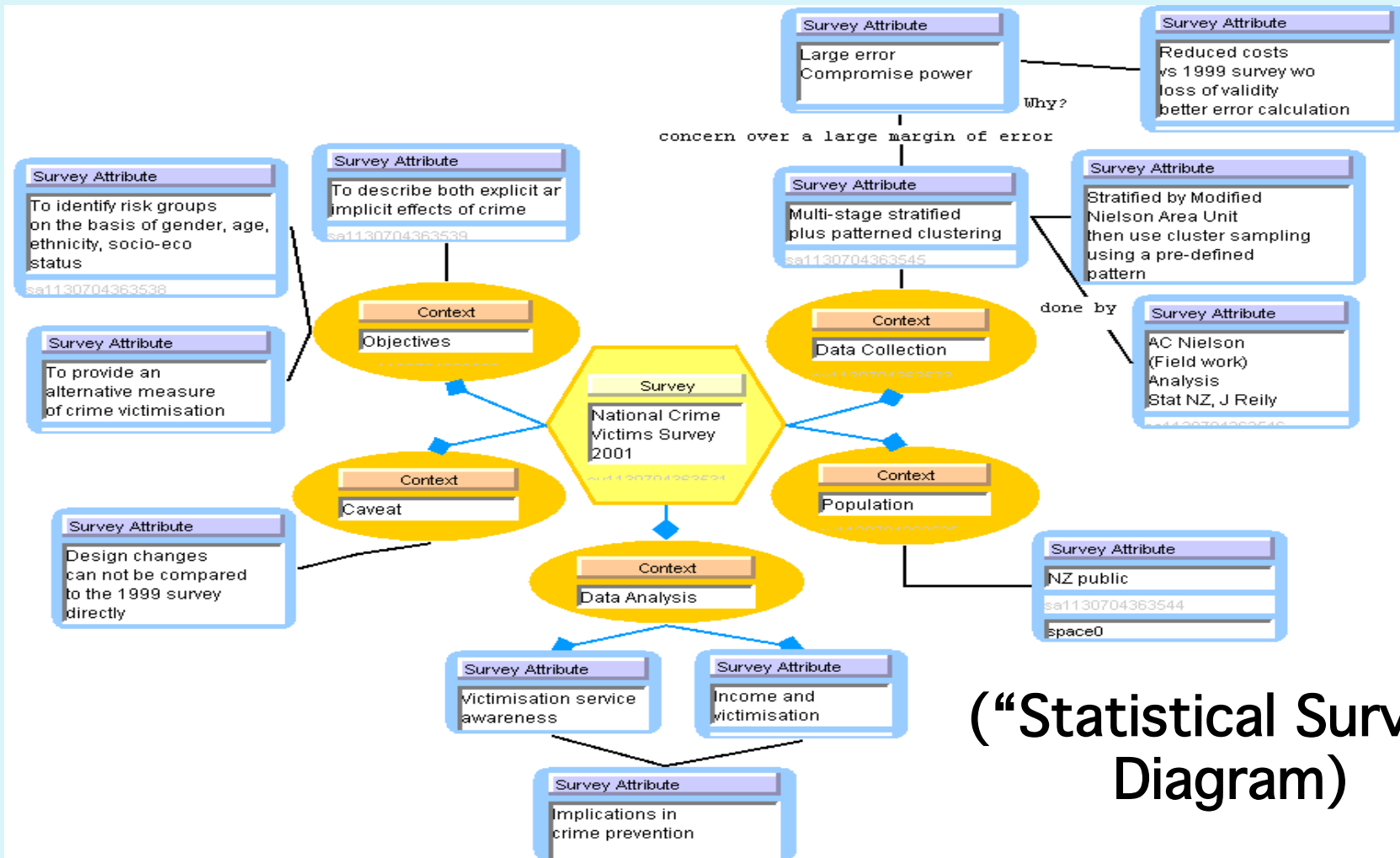
University of Auckland, New Zealand

Outline

- Motivation
- Current approach to COTS statistical tool usage
- Out Approach: SDL (Survey Design Language) & SDLTool modelling & generating:
  - COTS statistical analysis technique implementations
  - Web services as interfaces to technique implementations
- Example usage
- Architecture
- Discussion
- Conclusions

30 December 2014 | New Zealand | The University of Auckland
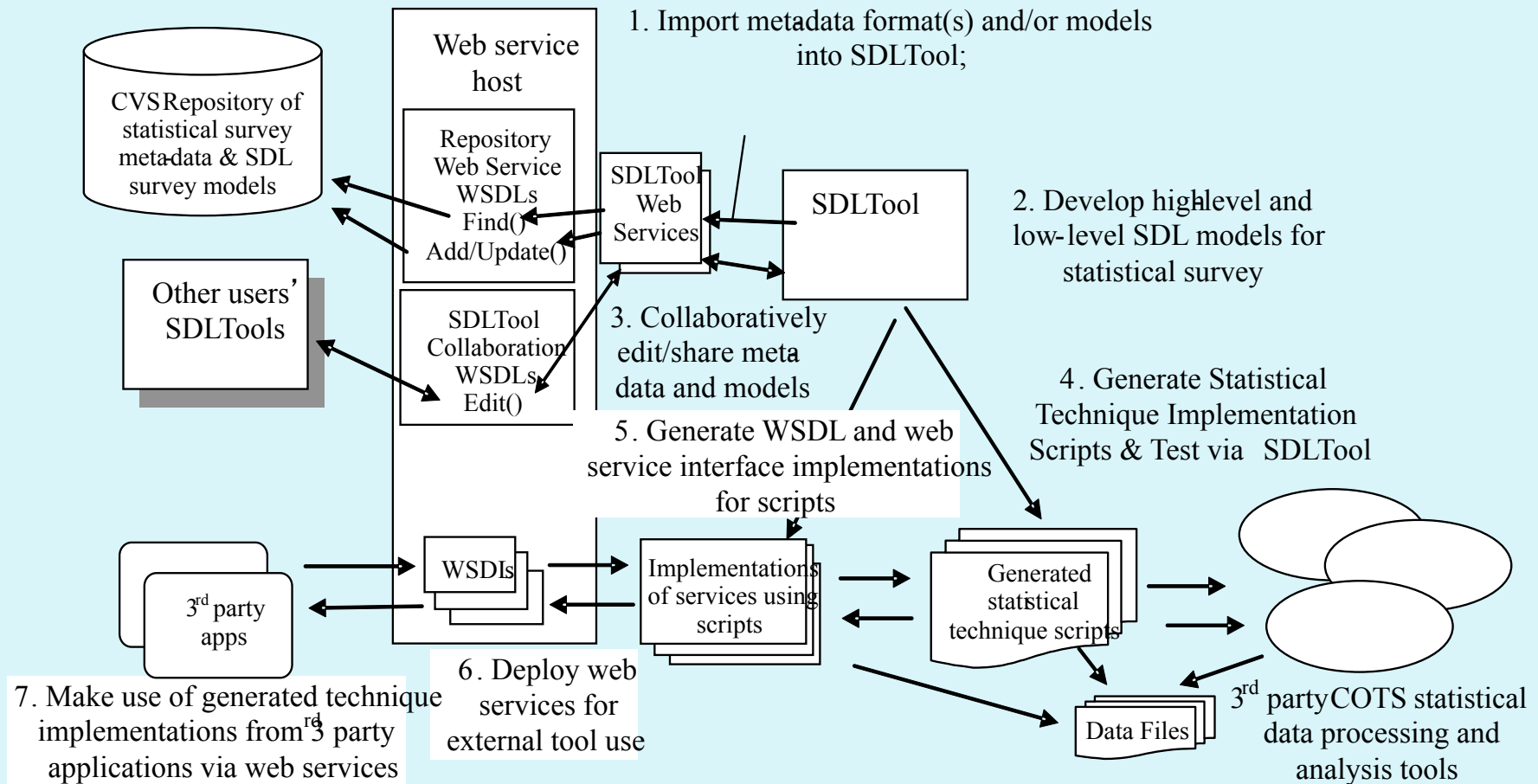

("Statistical Survey" Diagram)

How its done currently…

- Informally design the survey - currently very limited tools for statistical survey *process* definition & enactment
- Plan survey – sampling, data cleansing, analysis, results presentation and dissemination
- Use various COTS statistical packages, other applications to do:
    - Survey data modelling, planning, co-ordination
    - Design survey data capture e.g. forms (web or paper)
    - Survey sampling e.g. choose people to answer survey
    - Data cleansing and processing – outliers, testing of captured data, pre-processing of captured data, translation of raw data to formats for other stats packages
    - Apply various analysis techniques e.g. regression analysis
    - Summarise, visualise, publish results
- Need to exchange various kinds of data between tools
- Limited "standards" – for processes, plans, methods, techniques, data formats, ontologies, …
- Basically statistician *has to have* & *know how to use* each tool both individually & together…
- [Situation not unlike IDEs for software engineering before Eclipse ☺]

Our Approach
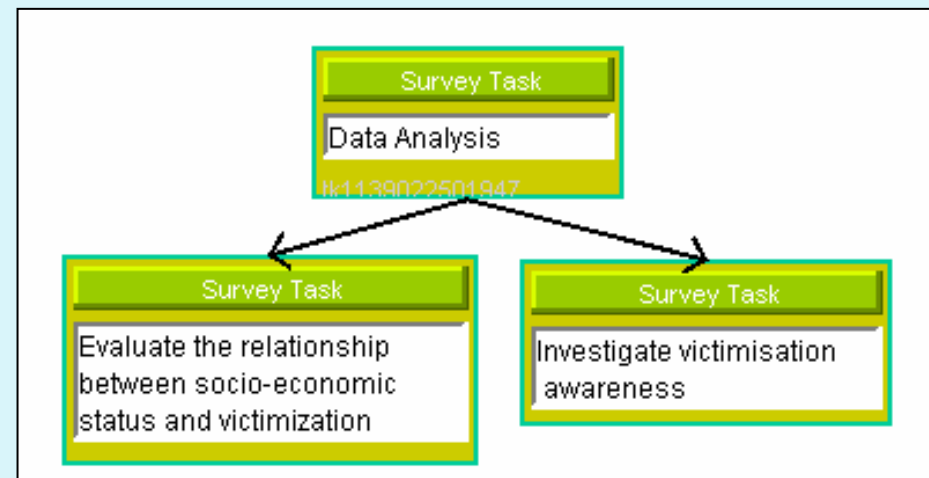
30 December 2014

New Zealand

The University of Auckland

CVS Repository of statistical survey meta-data & SDL survey models

Web service host

Repository Web Service WSDLs
Find()
Add/Update()

SDLTool Web Services

SDLTool

1. Import metadata format(s) and/or models into SDLTool;

2. Develop high-level and low-level SDL models for statistical survey

Other users' SDLTools

SDLTool Collaboration WSDLs
Edit()

3. Collaboratively edit/share meta data and models

4. Generate Statistical Technique Implementation Scripts & Test via SDLTool

5. Generate WSDL and web service interface implementations for scripts

WSDIs

3rd party apps

Implementations of services using scripts

Generated statistical technique scripts

Data Files

3rd party COTS statistical data processing and analysis tools

6. Deploy web services for external tool use

7. Make use of generated technique implementations from 3rd party applications via web services

Example Usage of SDLTool

- Survey process diagram (as previous) – "big picture" for the planned survey

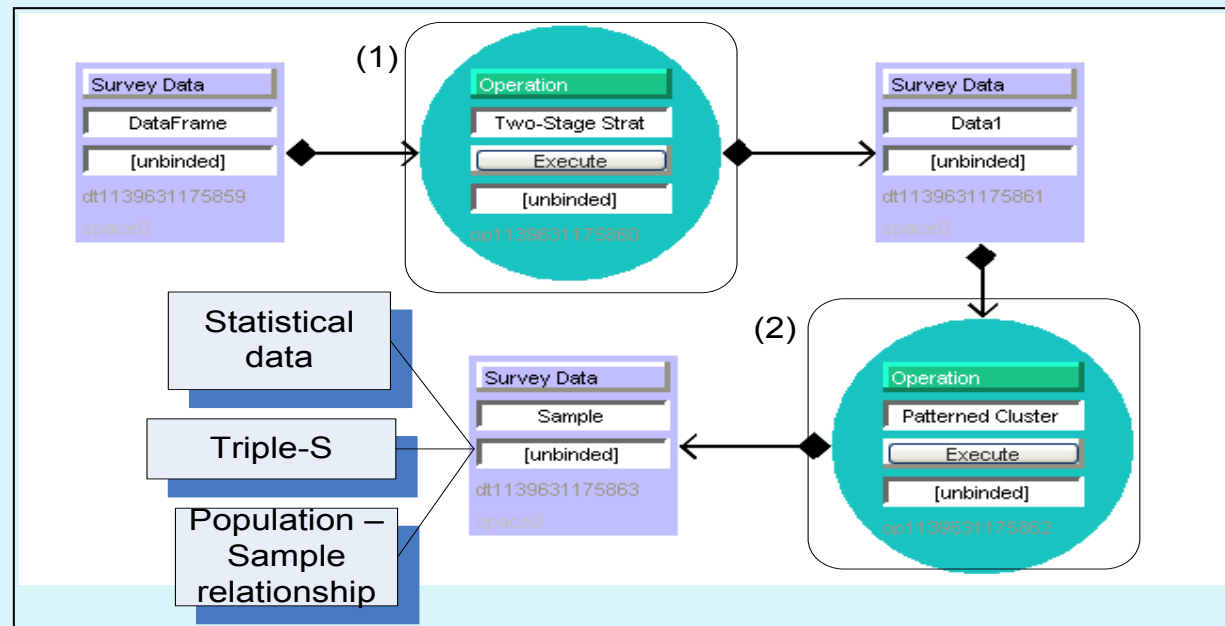- Survey task diagram – basically a work breakdown schedule with associated data, tools to use etc

Example usage (cont.)

■ Survey data diagram – data capture, processing operations

- Survey technique diagram – low-level data manipulation (statistical techniques)
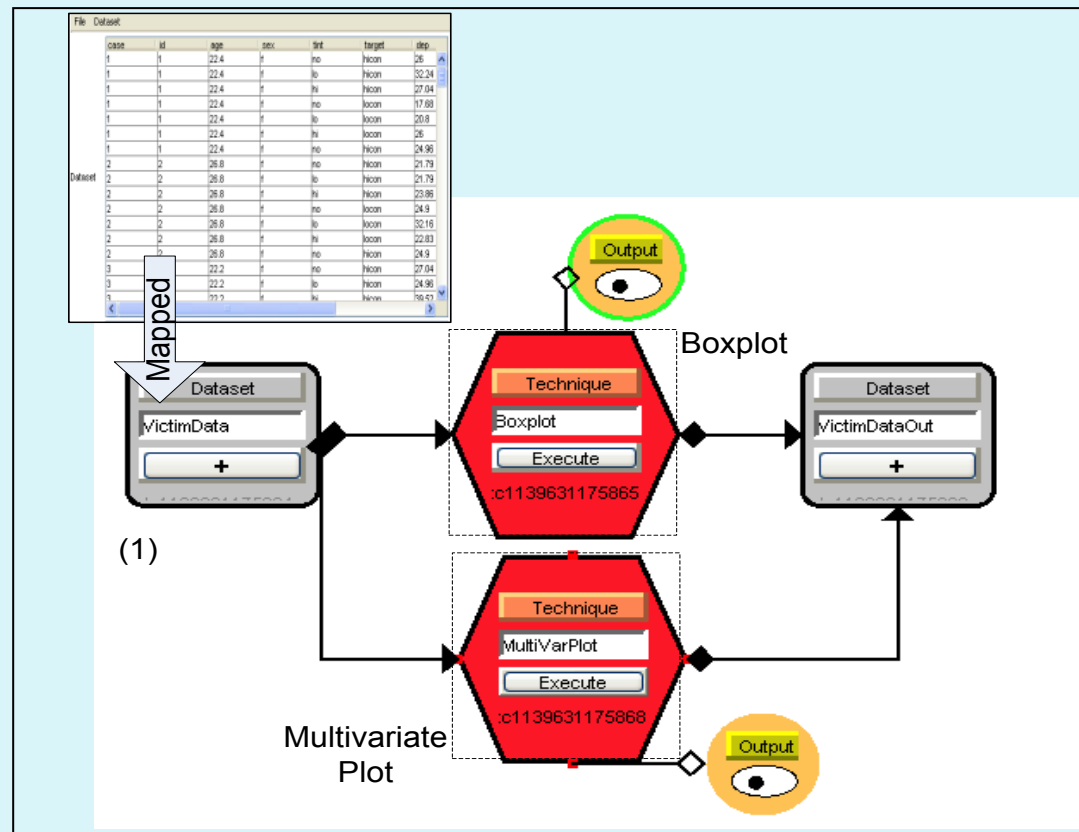
Example usage (cont)

- Task/technique diagrams – linking high-level survey tasks to data processing and statistical analysis techniques
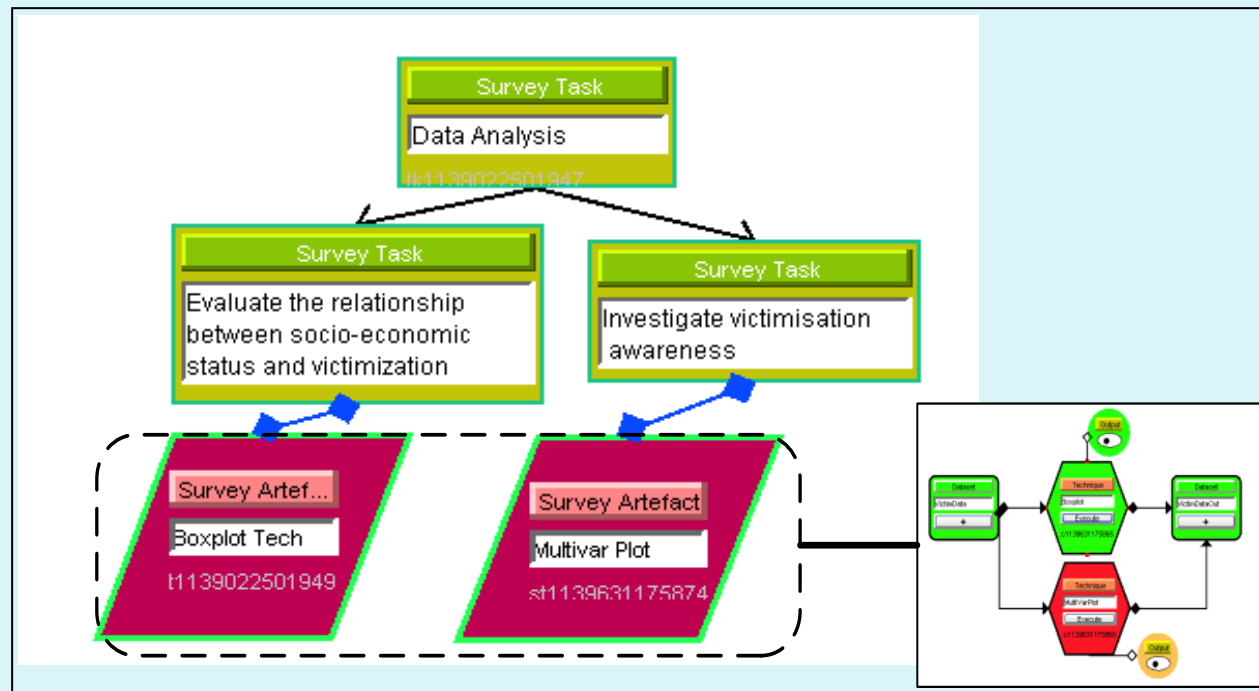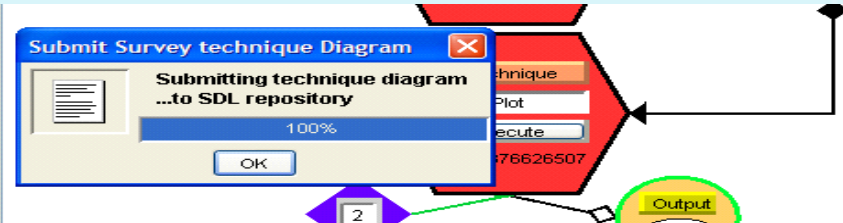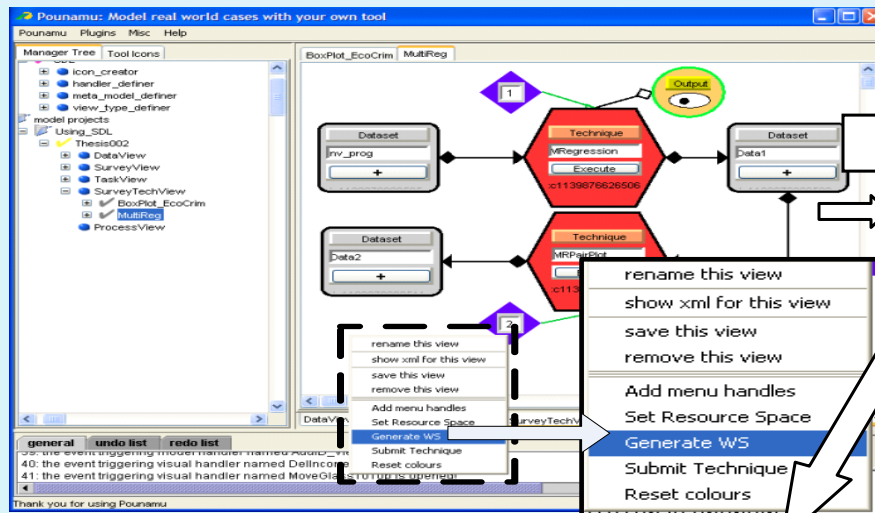
Key steps in survey data processing

- Obtain data

- Data processing and sampling

- Re-using techniques:

  - SDLTool generates moderately complex statistical technique implementations – scripts that are run by COTS stats packages

  - Need COTS stats package that runs it

  - Need to transform input data to process it

  - Need to transform output data to visualise it

- Make results available

- Visualising analysis results

Technique WS generation

SDLTool Architecture

30 December 2014

New Zealand

The University of Auckland

Pounamu Meta-tool (a)

SDL Diagrams

SDL Models

Event Handlers: (c)
-model/view consistency
-external stats tool interaction
-import meta-data formats
-export meta-data formats
-collaborative editing support
-generation of web services

(b) Local SDLTool Repository

(e)

Stats tool script generator (d)

Web service generator (f)

3rd party COTS statistical analysis tools (k)

Generated Scripts (j)

Web Service Implementation (i)

Local datasets

(l)

3rd Party Stats WS client applications (h)

Web Service host
Generated Statistical technique Web services & WSDLs (g)

a. Model SDL diagrams
b. Save/load locally
c. Event handlers: constraint checks, import/export, collab. Editing, WS generation
d. Generate scripts for COTS stats packages
e. (test scripts locally…)
f. WS generator given SDL
g. Generate WSDL, WS implementation, register in Ajax server
h. 3rd party apps call WS
i. WS interface calls WS implementation
j. WS impl calls COTS tool/ script/program
k. Data stored/processed by COTS stats apps
l. Result returned (XML)

Discussion

- Used SDLTool for modelling and implementing several statistical surveys
- User evaluation (professional statisticians and statistical teachers and researchers)
- Cognitive dimensions evaluation
- Reused generated statistical technique web services from custom .NET applications
- Web services wrapped technique implementations from three COTS statistical analysis packages
- Currently web services use proprietary message and data formats
- Current deployment to single web server; very limited COTS tool management (threading, local data management etc) i.e. it won't scale to very many clients

Conclusions

- Statistical COTS packages are very sophisticated but often use proprietary data formats, scripting languages, algorithms, libraries…

- SDLTool allows us to model statistical analysis techniques and generate COTS tool scripts – basically domain-specific visual language for COTS stats package scripting/usage

- Want to allow statisticians (and non-statisticians!) to use these "technique implementations" without needing COTS packages, detailed knowledge of techniques

- Generated web service interfaces/implementations from SDL models

- Allows COTS stats package capabilities to be made available remotely and by non-experts for complex statistical data processing and analysis in statistical surveys

References

- Kim, C.H., Grundy, J.C., Hosking, J.G. A suite of visual languages for model-driven development of statistical surveys and services, to appear in Journal of Visual Languages and Computing, Elsevier, 2015

- Kim, C., Hosking, J.G. and Grundy, J.C. Model Driven Design and Implementation of Statistical Surveys, In Proceedings of HICSS40 Mini-track on Model-driven development, Hawaii, January 2007, IEEE CS Press.

- Zhu, N., Grundy, J.C., Hosking, J.G., Liu, N., Cao, S. and Mehra, A. Pounamu: a meta-tool for exploratory domain-specific visual language tool development, Journal of Systems and Software, Elsevier, vol. 80, no. 8, pp 1390-1407.

- Kim, C.H. Hosking, J.G. and Grundy, J.C. Generating Web Services for Statistical Survey Packages from Domain-specific Visual Languages, In Proceedings of the ICSE2007 Workshop on Incorporating COTS Software into Software Systems: Tools and Techniques, Minnesota, USA, May 22nd 2007, IEEE CS Press.

- Kim, C. Hosking, J.G., Grundy, J.C. A Suite of Visual Languages for Statistical Survey Specification, In Proceedings of the 2005 IEEE Conference on Visual Languages/Human-Centric Computing, Dallas, Texas, 20-24 September 2005, IEEE CS Press.

THE UNIVERSITY OF AUCKLAND
NEW ZEALAND
Te Whare Wānanga o Tāmaki Makaurau