Usage of external software tools at SURS - experiences and lessons learned so far

Rudi Seljak, Andreja Smukavec, Igor Kuzma
Statistical Office of the Republic of Slovenia
Summary of presentation

- Introduction
- Usage of external applications at SURS
- Experiences with usage of the shared application
- Sharing of SURS products
- Conclusions
Introduction

• What we all want to obtain:
  – Modernised and harmonised statistical production
  – Transition from domain specific, stove-pipe oriented to generic, process oriented production
  – Maximal automated procedures – ‘push the button principle’
  – All that with minimal costs !!!

• Facts that are fostering the principle of sharing of software applications:
  – Insufficient resources for own development
  – Common sense argument: „Why to invent the wheel if somebody already did it“
  – Better and more successful international cooperation
Examples of external applications at SURS
Banff

• Collection of SAS procedures, developed at Statistics Canada
• Developed to support editing and imputation procedures
• Commercial product – can be purchased for certain fee
• Main characteristics:
  – For SAS oriented institutions very flexible and easy to use tool
  – It has a very good documentation
  – External users can also get the support of the Stat Canada team
  – Designed only for numeric (continuous) variables – sometimes shortcoming
Banff cont‘d

• Usage at SURS
  – Mostly used inside the generic application for editing and imputation (SOP)
  – Mostly used for imputations and outlier detection
  – Was also tested to be used as analytical tool for statisticians – never really came to life
Calmar

- SAS macro program developed by INSEE
- Free software - can be downloaded from the INSEE website
  - Not open source
- Also only for SAS users
  - Additional pre-condition: IML module required
- Lack of good English documentation is a shortcoming
- Usage at SURS:
  - Main tool when the calibration procedures are in question
  - Most commonly used in social surveys – advanced calibration techniques are more common practice
  - So far only in ad-hoc programs – also planned to be included in general solution for weighting procedures
Tau Argus

- Software application designed to protect statistical tables
- Result of several European projects
  - The main part of Tau Argus has been developed at Statistics Netherlands
- A project was started in 2012 to rewrite τ-ARGUS to an open source version.
- Usage at SURS:
  - Already used for more than 10 years
  - Mainly used by statistical disclosure control experts
    - In few cases also by survey statisticians
  - We are in final phase of development of the module of general metadata driven application for data protection – Tau Argus very important part of it
Demetra

• Software tool developed to support the seasonal adjustment procedures.
• First version developed by Eurostat at late 90s
• Later upgraded with Demetra+ and the last version JDemetra+, which is already open source
• Usage at SURS:
  – Survey statisticians are responsible for running Demetra in regular production
  – General instructions and guidelines for working with Demetra+ are in place
  – Annual reviews of the models are still done by experts for seasonal adjustment
External softwares - some general experiences

• There is quite a large input needed in the first phase of introduction of new software.
  – Usually one person appointed to become a specialist
• Input very much dependent on documentation and available support
• Very diverse practice of usage of external applications:
  – End-users applications (e.g. Demetra)
  – Tools used mostly by specialists (e.g. Tau Argus)
  – Tools for ad-hoc programming (e.g. Calmar)
  – Tools that are incorporated in generic solutions (e.g. Banff)
• General trend and future perspective: open source solutions
Sharing of SURS product

• SURS is a small institution → possibilities of own developments are limited

• There are few cases of products that could be shared.
  – **STAGE.** Web-GIS application for cartographic visualisation.
    • The application enables the user to present statistical data while selecting the spatial level
    • Geospatial statistical data for over 300 statistical variables in variously available time series are presented for up to 10 levels of geospatial data
    • Development of Stage of co-financed by Eurostat and it was developed in the cooperation with Geodetic Institute of Slovenia.
    • STAGE can already be shared under the European Union Public License.
    • This makes the STAGE an open source product
Sharing of SURS product cont‘d

- **SOP.** General application for statistical data processing.
  - Developed inside SURS with certain financial support of Eurostat
  - Generic, modular, metadata driven application. Each module covers one sub-process (e.g. data validation, data editing, aggregation, standard error estimation, data protection,…)
  - SAS macros, that are run on the basis of the metadata driven principle and could be through the set of process metadata easily parameterised for the needs of different surveys.
  - SOP could be shared on different levels:
    - Sharing only the conceptual design and conceptual architecture of the application
    - Sharing only the SAS macros
    - Sharing the whole application
Conclusions

• Sharing of already developed IT tools is for SURS a very tempting option, mainly due to:
  – Significant reduction of the development resources
  – Usage of more advanced statistical methods
  – Usage of harmonised procedures and methods

• SURS is also in favor to share its own products but several issues are always considered:
  – Significant amount of work that the product will be adjusted for the “external users”
  – Installation of the software application to another environment would certainly require certain support
  – Is it possible to ensure sufficient support for the external users
Thank you for your attention