Statistics Netherlands Application Strategy

CSPA Workshop Wiesbaden
Content

- Strategic Drivers
- Application Strategy
- Implementation
- Three flavours of CSPA
New ambitions..

Progressing towards a data-driven society
..new partnerships..
...and new (but also old) challenges
IV Vision 2020

IV: capabilities CBS needs to realise strategic goals (process, methodology, applications, infrastructure, people)
Application strategy principles

– The automation of work as much as possible done by statistical employee (increased flexibility, less support costs)

– Differentiation of application architecture and governance for different change dynamics (fit for purpose IT)

– Sharing of ICT skills in communities and pooling of scarce specific IT skills in centers of excellence (less risk of single point of knowledge)
Application Strategy Scenarios

Innovative | Ad hoc | Simple | Frequently changing & complex | Evolutionary changing & complex | Corporate

Agile methodologies
Multi-discipline teams
Continuous delivery
Design Board
What is the CSPA?

- A template architecture for official statistics
- A set of standard specifications for new statistical components (services) that can be used in a modular way
- A new way of developing statistical tools, with sharability as a design feature, not an afterthought
Generic application development
Service-Oriented Architecture
Designed for reuse - Blaise 5

- Multi Mobile Devices
- Multi Platform
- Complex and Large Surveys
- Intricate Routing Rules
- Open Architecture (APIs)
- Device Adaptive
- Intelligent Screen Generation
- Multi-Lingual (incl. Non-Western)
- Fast Compilation
- Mixed Mode
- WYSIWYG
- Paradata – Audit Trail
- DDI repository (Colectica)

sharability as a design feature
CSPA “classic” (top-down)

Challenges 😞

– Low number of available services

– Requirement to do design for reuse from start

– Complexity (SOA, GSBPM/GSIM, LIM..)

Opportunities 😊

– ESS/ESSNet projects

– Use in other international SW collaboration (.stat, Blaise..)

– In projects where sharing is primary goal CSPA is state-of-the-art approach
Reuse of core functionality in R packages

A set of standard specifications for new statistical components (services) that can be used in a modular way.
R tools for data validation, correction and imputation

Available Packages

Simputation, rspace, dcmodify, Validate, errorlocate, deductive, lin tools

Chainable

DATA → Analytics → Data Cleaning → Validation
An awesome list of statistical software packages useful for creating official statistics.

**Awesome official statistics software**

An awesome list of open source statistical software packages useful for creating and accessing official statistics.

An item on this list is awesome because

1. it is free, open source, and available for download;
2. it is confirmed to be used in the production of official statistics by at least one institute, or
3. it provides access to official statistics publications.

We prefer packages that are reasonably easy to install and use and that are actively maintained.
## CSPA “light” (bottom-up)

<table>
<thead>
<tr>
<th>Challenges 😞</th>
<th>Opportunities 😊</th>
</tr>
</thead>
<tbody>
<tr>
<td>– Integration not completely plug and play</td>
<td>– Lots of stuff already out there (CRAN, Awesome SW..)</td>
</tr>
<tr>
<td>– Visibility</td>
<td>– de-facto open-source standards (machine learning..)</td>
</tr>
<tr>
<td>– Non-functionals (performance..)</td>
<td>– Reuse of core code in various scenarios (R programs, BI tools, big data platforms, cloud..)</td>
</tr>
</tbody>
</table>
Distributed Processes and Secure Data Sharing (Multi-Party Computation)

- Query (CBS)
- Translate Query
- Metadata (CBS)
- Match Metadata
- Data (CBS)
- Bring Data together:
  - Encryption
  - Oblivious Transfer
  - Data virtualization
- Query (Other)
- Metadata (Other)
- Data (Other)
## CSPA “future” (algorithms, models, MPC..)

<table>
<thead>
<tr>
<th>Challenges 😞</th>
<th>Opportunities 😊</th>
</tr>
</thead>
<tbody>
<tr>
<td>- What is it? Do we need it?</td>
<td>- Important if we want to harvest alternative data sources</td>
</tr>
<tr>
<td>- Variability - can't rely on single standard</td>
<td>- New area (no legacy)</td>
</tr>
<tr>
<td>- Evolving - things are changing as we speak</td>
<td>- New partnerships (beyond NSO community)</td>
</tr>
</tbody>
</table>