Validation, Shared Services and Enterprise Architecture: How it fits

Agenda

- Motivation
- Enterprise Architecture: Different views
  - Business Architecture
  - Information Architecture
  - Application Architecture (focus!)
- Technical Approaches to Validation in the ESS
  - Types
  - Existing Validation Tools
  - Comparison
- Conclusions & Invitation
Motivation

Validation matters

- In most statistical organisations still the most time consuming (and therefore expensive) activity in the production process
- And probably the most important activity in terms of data quality

Often it is not really well understood and organised

- Methodology is heterogeneous across institutions and domain
- Tools and Services are not harmonized or shared
Motivation

Why this presentation?

- Showing how Enterprise Architecture can help
- Enhance international collaboration
Motivation

Why this presentation?

- Get you interested in two workshops
  - ESSnet SCFE on CSPA/SERV July 2017

- Regional workshops on Validation in the ESS from November 2017 till February 2018
Enterprise Architecture

How do we align business needs with IT?

- Talk to each other!
  - Do not assume you know the other side well
  - We have similar but not identical goals

- Use a common framework to organise collaboration
  - Common language
  - Common methodology
  - Common standards
Enterprise Architecture

Layers and Bricks

- The different perspectives are ordered in layers
- Understanding of the
- We have similar but not identical goals
Business Architecture

Remember GSBPM?
- Applicable to all organisations doing official statistics
- Usable for all statistical domains
- But maybe a bit too abstract

What about GSDEM?
- More detailed
- Layer 3 in GSBPM terms (activities)
Business Architecture

Deep blue / National initiatives

- Advantages
  - Even deeper insight
  - Smaller building blocks

- Disadvantages
  - Blurs the picture

Example from Slovenia
Information Architecture

Sharing of services requires an Information Model:

- **Conceptual**: User view, independent of physical implementation
- **Logical**: Meets requirements, CSA Logical Information Model
- **Physical**: Implemented in CSPA Services

Level of Detail

- **Under Construction!**

Human Oriented

Computer Oriented

Experiences in context of CSPA: It is too coarse.

Intermediate level between GSIM and physical model is required: Logical Information Model (LIM) under construction.
Application Architecture

Coming to the Core: Choices

1. Monoliths („One size fits all“)
   - Known from the mainframe age
   - Goal: All statistical methods and functions are included into one application
   - For many institutions still very attractive
   - Often promoted by IT
   - Reduces the complexity of the real world

2. Standard application(s) („Do it yourself“)
   - Using general purpose tools (Excel, SQL-based systems, SAS, R) for statistical production
   - Program or configure applications to domain specific solutions
   - Often promoted by domain specialists or methodologists
   - Production is done individually (with all its advantages and drawbacks)
Application Architecture

3. Specialised applications („Haute couture“)
   - Applications developed for each domain specifically
   - Usually done in a general purpose programing language (Java, C#)
   - Often promoted by SW-Developers and asked for by Domain specialist
   - Slow and costly!

4. Tool chains („Cross-cutting approach“)
   - Applications for phases or sub-processes (GSBPM driven)
   - Often started by data collection platforms
   - Quickly becoming „networks“ instead of chains
   - Promoted by „standardisation people“ (management)
   - Can become a coupling „nightmare“
Application Architecture

Coming to the Core: Choices

5. Service oriented architecture („the future“?)
   - Functionalities are encapsulated in (web-)services
   - Modularity and decoupling are „built-in“
   - Promoted by the UNECE and Eurostat
   - Integration platform needed
   - Standard information model needed
   - Not widely distributed yet

6. Hybrids („real world“)
   - All kind of mixtures of the former..
Application Architecture

Approaches to data editing

- Results from a survey 2015 (ESS-Member States)
- All application architecture models are present
- Standard application (SQL, SAS, Excel) are most prominent
- Many NSIs are working on the Monolithic approach
- All others are of minor importance
- SOA-like approaches (CSPA, ESS.SERV) were near to no existent
Tools and CSPA-compliance

Three examples

- Dutch Validation Package
- German eStatistik
- Eurostats Validation Services
Tools and CSPA-compliance

Validation Package (by Mark van der Loo)

- Based on R
- Easy to use
- Stand alone

```
# define rules
v <- validator(turnover >= 0, other.rev >= 0
, turnover + other.rev == total.rev)
# confront with data
cf <- confront(dat, v)
# analyze results
summary(cf)
```

<table>
<thead>
<tr>
<th>rule items</th>
<th>passes</th>
<th>fails</th>
<th>nNA</th>
<th>error</th>
<th>warning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

expression

- turnover >= 0
- other.rev >= 0
Tools and CSPA-compliance

eStatistik

- Production chain
- XML based
- Validation just one functionality

GSBPM 4.0 (2009)
Tools and CSPA-compliance (eStatistik)
Tools and CSPA-compliance

StruVal and ConVal:
- True Services
- Not yet CSPA compliant
- Limited in scope
- Wrapping already existing tools (SDMX-converter & EDIT)
Tools and CSPA-compliance

ToDos

- Dutch Validation Package
  - Wrap it as a service
- German eStatistik
  - Isolate parts and wrap them
- Eurostats Validation Services
  - Make them more generic
  - Integrate VTL
Conclusions

- Enterprise Architecture is a useful framework for communication and as a guideline for action
- As methodologist or subject matter specialist, Business Architecture needs to be elaborated
- As IT-Expert, make yourself familiar with CSPA
- Work on Information Models
Motivation (revisited)

Remember
CSPA
Wiesbaden
3-7 July 2017
CSPA: Service Granularity and Integration

A service does not make production

- Service are autonomous (by definition)
- They need to be integrated
- Different integration patterns have been identified

**Point to Point Integration**

**Centralized orchestration**

**Smart Services**

© Results from a hackathon in Geneva 2016