Prototyping in R, Operationalisation in SAS

or why we teach both SAS and R in our study courses

Prof. Gerold René Baudinot
Head of Institute of Applied Information Technology
Zurich University of Applied Sciences
Mission

- Our research is on smart information technologies.
- We make information available.
- To the right people anytime anyplace and securely.
Availability of Data

NRI vs GNI per capita

WEF Global Report 2013
How can we cope with all the Data

ride the information wave
is to search the relevant data, augment, enrich analyze and package the data to become useful Information aka Dataproducts
What is a Dataproduct

**Definition and examples**

**Characterization**
It is not one-off analysis but rather a repeatable, reproducible Analysis or Report

**Definition**
A piece of software at the intersection of Domain expertise, Statistical expertise, Software Engineering and Data

**Examples**
LinkedIn’s People you may know
Optimization of truck routes
Google Trends

Source: Drew Conway
25.11.2015
© Zurich University of applied Sciences, Prof. Gerold Baudinot
What is Data Science?

Enables Data Products
⇒ Applied Science
⇒ Interdisciplinary

Data Science := “Unique blend of skills from analytics, engineering & communication aiming at generating value from the data itself [...]”

(ZHAW Datalab*)
### The Link between Data an ICT Strategy

<table>
<thead>
<tr>
<th>IT Impact on core activities</th>
<th>High</th>
<th>Low</th>
<th>IT Impact on enterprise strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factorymode</td>
<td><strong>Goal:</strong> Improve performance of core processes</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Leadership:</strong> Businessunit</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Here is where we find ERP Systems</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strategymode</td>
<td><strong>Goal:</strong> Transformation of Organisation or Industry</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Leadership:</strong> Executive Management and Board</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Here is where we make power out of knowledge with «Dataproducts»</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supportmode</td>
<td><strong>Goal:</strong> Local / partial optimisation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Leadership:</strong> Team</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turnaroundmode</td>
<td><strong>Goal:</strong> Development and introduction of a new venture</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Leadership:</strong> Businessarea/new venture</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Here is where we find Prototypes</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Corporate Information Strategy and Management, Applegate, Austin, McFarlan

25.11.2015 © Zurich University of applied Sciences, Prof. Gerold Baudinot
 ICT Strategy

- Business Perspective
  - Efficiency
  - Reliability
  - Resilience

- Fundamentals
  - Highly Structured Processes
  - Highly Standardized Architecture

Processes

Embedded Architecture

Application Architecture

Software Architecture

System Architecture

Operations

25.11.2015 © Zurich University of applied Sciences, Prof. Gerold Baudinot
Iterative development style -> Agile

- **Agility**
  - Having the characteristics of speed, coordination
  - The ability to react quickly and appropriate to change

- **Agile development**
  - requirements and solutions evolve through collaboration between cross-functional teams.
  - It promotes adaptive planning, evolutionary development, early delivery, continuous improvement,
  - encourages rapid and flexible response to change.
Service Development paradigm

High

**Factorymode**
- **Projektmanagement:**
  - Process Reengineering
  - Highly Structured Processes
  - Highly Standardized Architecture

**Strategymode**
- **Projektmanagement:**
  - Change Management
  - Integration of two worlds methods agile and factory

Supportmode
- **Projektmanagement:**
  - Experiment,
  - KISS

Turnaroundmode
- **Projektmanagement:**
  - New Venture development
  - Agile development

Low

IT Impact on enterprise strategy

High

25.11.2015 © Zurich University of applied Sciences, Prof. Gerold Baudinot
3 Loops – fusion of 2 worlds

Deliver Business Success

Loop Awareness

Intake Standardized DataProducts

Intake Results of new DataProducts

Improve Standard DP & Proc

Analyze non Standard Dataproduts

Create new non Standard Dataproduts

Standardize Operating Processes

Standardize DataProducts

Detect new Threats & Opportunities

Intake Standardized DataProducts

Deliver Business Success

Loop Balance

Loop Agility

Quality

Agility

25.11.2015 © Zurich University of applied Sciences, Prof. Gerold Baudinot
Implications concerning Tools

The Factory World

- SW Requirements
  - State-of-the-Art
  - Extensive API support
  - Maturity & Legacy inertia
  - Compliance & Certifications
  - Performance leader
  - Realiability & Resilience
  - Portability
- Ecosystem Requirements
  - Vendor stability
  - Partner Stability
  - Global Presence
  - Core IT References
  - Industry References

The new Venture World

- SW Requirements
  - cutting edge
  - Latest technology
  - Basic API Support
  - light weigth IDE
  - free documentation
  - free Software (OSS)
- Ecosystem Requirements
  - Close proximity of SW development to science
  - Active development community
  - Active user community
  - ….
Integrated development Requirements

Tool Integration Requirements

- Variation of integration methods
  - Code Wrapping
  - Intertool API’s
  - Variation of Data Integration
  - (Enduser) prototyping Tools
- Batch processing
- Documentation
  - Interfaces
    - extensive Documentation
    - samples
  - Methods/Algorithms
    - extensive documentation
    - samples

Ecosystem Requirements

- Professional Network
- Education offering for the specific toolset
- R&D Projects with use of the specific toolset
- University curricula which embeds the specific toolset
- Vendor support / sponsoring for Universities
- Longterm commitment by SW Vendors
Proof of concept

Project: KoLe
Cost and performance statistics
Authority: FOPH
Federal Office of Public Health
Teams

Analytical Team
Prof. Marc Wildi, IDP
Prof. Simon Wieser, WIG
Andreas Holenweger, IDP
Carmello Iantosca, SAS

ICT Team
Prof. Gerold Baudinot, InIT
Barbara Flaad, InIT
Simon Gadient, InIT
Jonas Looser, InIT
Remo Lehmann, InIT
Fasika Daksa, InIT

25.11.2015 © Zurich University of applied Sciences, Prof. Gerold Baudinot
Objectives

- Provisioning of the database with different groups of drugs
- Reporting of figures relating to the prescription of narcotics
- “Extrapolation” at the federal level of the figures by Canton, Region, Age and Gender

Deliverables

- Data Model
- OLAP Cubes
- Semi-automatic process for the “extrapolation”

Challenges

- Data Delivery chain delivery standard, Cleansing, sheer size/load time
- Anonymization of the data of the insured
- Integration of R-Prototypes into ETL Processes
- Size of OLAP Cubes
Approach to the delivery object

Kole DWH

Reports, Charts

Extrapolation

Datamart

Load

Transform

Sternmodell Cube(s)

Create Cube

health insurer
Process for Data Adjustment (Extrapolation)

1. KoLe Datawarehouse
2. Selection of data to adjust
3. Conversion to CSV
4. R-Program
5. Conversion to SAS Table
6. Load into Data Warehouse
Extrapolation Prototype/Pilot

Biased Data

Auxiliary data

R-Program
Extrapolation Implementation

**Input**
- datamart_abdeckungsgrad
- Abdeckungsgrad_KoLeDWH_2005.xls
- Abdeckungsgrad_KoLeDWH_2006.xls

**SAS Enterprise Guide Process**
- Process A
  - Extract
  - Transpose
  - Build star schema

  **Output**
  - kantonkategorie_fact
  - alterskategorie_fact
  - jahre_dim
  - kanton_dim
  - alterskategorie_dim

- KantonKategorie UserInput
- AltersKategorie UserInput

- Process B
  - Extract
  - Merge
  - Run Hochrechnung function

**V2 Proc Summary**

**V1 SAS-Macro**
Conclusion

❖ The Toolset enabled us to Run this 3 Loop Process
❖ Education of Customer was insufficient
❖ Educational Requirements
  ● Endusers
    o Application Level Education in Enterprise Guide
    o Recall and Understand Level Education in R
  ● Developers
    o Synthesis Level Education in Enterprise Guide, SAS BASE, SAS Macro, DIStudio, SAS Graph
    o Evaluation Level SAS Stat, SAS IML
    o Recall and Understand Level Education in R
  ● Data Scientist
    o Recall and Understand Level Education in Enterprise Guide, SAS BASE, SAS Macro, DIStudio, SAS Graph
    o Evaluation Level Education in SAS Stat, SAS IML
    o Synthesis Level Education in R

• Note: The competence levels are based on Blooms Taxonomy