



# The Data Law is coming...

short essay

8th Eastern European e|Gov Days  
Praha 22.4.2010

[vladimir.kyjonka@cze.sas.com](mailto:vladimir.kyjonka@cze.sas.com)

**THE  
POWER  
TO KNOW<sup>®</sup>**

# What we will talk about

- Oh, quality, it makes troubles ...
- How is it in a government sector?
- Data Centric Organization
- Two real life examples
- The Law is coming (to data)...

# Impact of poor data quality

It turned out that 1/5 of addressees of a mail campaign of a British retailer are deceased. Nevertheless they were sent a letter with a greeting „**Dear Mr. Deceased**“ . <sup>1)</sup>

A certain insurance company found out that the majority of their clients have profession of **Astronauts** - further investigation showed that „Astronaut“ is the first selection in their CRM system. <sup>1)</sup>

44 000-98 000 of Americans yearly die due to a avoidable medical error, such as typing mistakes on medicine prescriptions, wrongly described blood test results, illegible information at the patients' files. It is the eighth most common cause of death in the USA. <sup>2)</sup>

On the 7<sup>th</sup> May 1999 the US air forces struck against the Chinese Embassy in Yugoslavia. Investigation showed: CIA used out-dated maps; moreover, an officer provided with a **wrong address data** – „He literally put the X in the wrong spot“<sup>3)</sup>

1) Peel, M: Letters to the dead and other data dereliction. © 2007 Financial Times Deutschland. <http://www.ftd.de>, vydání z 2.10.2007

2) Oash, J. (1999): IT Can Reduce Medical Errors. Obsaženo v: Wang, Pierce, Madnick: Information Quality, 2005

3) BBC: Americas chinese embassy warning ignored. © 1999 BBC. <http://news/bbc.co.uk/1/hi/world/americas/37775.stm>, vydání z 2.10.2007

# How about data quality

- There are errors in the data (and there have always been)
- Today: more and more activities are covered by computer processing
  - More and more activities are tightly dependent on computer processing
- High rate of non-quality data – significantly reduces its usability
  - Incorrect results of the processing
  - Malfunction of enterprise business processes → financial loss
  - Processing, management and archiving of wrong, meaningless and redundant data → direct financial loss – **GARBAGE IN/GARBAGE OUT**
- Increase in complexity of information systems → even bigger impact of non-quality data

# Where non-quality data arise from?

- Human factor
  - People make errors
    - Typos
    - Entering data into wrong fields
    - Ignorance
    - „Creativity“ and folklore
    - ..
- Information system inconsistency
  - Many autonomous subsystems
  - Independent organization units
  - Various subject areas
  - Various technologies and applications
  - Different suppliers and vendors changing in time
  - Product oriented systems (not subject-oriented)
  - Implementation of new systems

# Poor data quality costs money

- Mail not delivered - money is gone
- Inability to identify the notorious defaulter – money is gone
- Wrongly targeted marketing campaign – money is gone
- Large volumes of redundant data processed – money is wasted
- Inability to create reliable reports we can trust – money is gone (or sbdy goes behind bars)
- ...
- Thus it is worth the money and time to invest into data quality
  - Put some money into it in order to save or get more money

# Typical questions

- How many clients do we have?
- What losses are caused by low quality data?
- How many clients we could lose due to an incorrect identification of target groups during a mkt campaign?
- How much money we lose, if we do not precisely identify the risks related with poor quality data?
- ...
- How much to invest to data quality so that it pays off?

# How it is in public sector

- May we consider the following question:  
*How much shall we invest into data quality to gain a profit?*



- Answer:  
*Not really! We cannot ask this!*



- **Why?**  
Institution of state and public administration **must**  
**ALLWAYS** fulfill all mandatory tasks by law,  
not only if *it pays off!*

# Government intuitions are DCO!

- What is DCO: **Data Centric Organization**
  - Organization intentionally takes care of its data in order to gain benefits from it
  - Based on this fact, organization builds its strategy
  - Organization guards its data as the most valuable treasure
  - Simply: Organization is **DATA CENTRIC**
  
- In government sector information/data is essential for everything
  - They are DCO whether they want or not and whether they know or not
  
- The more they understand this, the better they work
  - Lets look at two examples from real world

# Case 1

- Top government organization
- Task: Cleansing and consolidation of private individuals data filed in an official registry
- Beginning: 2 institutions were arguing:
  - The first was the owner of the data, the second utilized them
  - Who had the right to provide the data?
  - Who should provide resources for collaboration?
- Reactions of consultants after the project completion :
 

„...*We have not seen anything like this before...*“  
 (The volume and type of errors and defects in data have significantly exceeded their expectations based on experience from commercial sector...)

## Case 2

- Important institution of state administration
- Task: Data quality assesment of data administered by central registries
- Beginning: Instant perfect collaboration
  - Necessary data samples provided immediately
  - Clearly specified processes and competencies
  - From the very beginning it was obvious that the organization (whithout knowing that) already was a DCO
- Reactions of consultants after the project completion :
 

„...*We have not seen anything like this before...*“  
 (In comparison to their experience from the commercial sector, data quality exceeded the practical and possibly even theoretical limits – often more than 99,9% !!! ...)

# What we have learned

- For government organisations a *certain* data quality is not sufficient but **ONLY THE BEST!**
- Government organizations DCO based on their nature
- Data quality terrain is not yet fully explored and the biggest discoveries are still ahead
- „The disadvantage“ for government organizations regarding data quality: **THEY MUST** – they are allways under pressure of the law
- „The advantage“ for government organizations regarding data quality : **THEY CAN**, they can employ a power of the law





# The Data Law is coming...

short essay

8th Eastern European e|Gov Days  
Praha 22.4.2010

[vladimir.kyjonka@cze.sas.com](mailto:vladimir.kyjonka@cze.sas.com)

**THE  
POWER  
TO KNOW®**