Data Quality at EuropeAid

Data Quality monitoring and control

SAS Forum, 9th October 2014

Anne-Catherine Gridelet (Head of Sector Business Management & Reporting - EuropeAid)
Hugues Paillé (Senior Lead - Deloitte)
Agenda

1. Presentation of EuropeAid
2. Context of the project
3. The Data Quality framework
4. Lessons learned and Going forward
Directorate-General for Cooperation and Development

Design of EU Development policies

Deliver aid throughout the world

4500 people
140 EU Delegations + Headquarters Bxl

2013
BUDGET: 4,609 mil. €
EDF: 3,735 mil. €
Agenda

1. Presentation of EuropeAid
2. Context of the project
3. The Data Quality framework
4. Lessons learned and Going forward
Context and objectives of the project

Context

- Issues with the quality of the data encoded in CRIS
- Significant impact on publications

- Program defined to tackle the issues
- Pilot project to assess the Data Quality level of the contract data

Objectives of the project

- Identify the business rules
- Implement the BR to automate the verification
- Implement a re-usable framework to measure and monitor the data quality
- Provide a tool for the visualization of the results by the business data owners
EuropeAid’s Data Quality Challenges
Main data challenges at EuropeAid

- Transparency
- Reputation
- Compliance
- Decision making
- Cost Reduction
Project approach
A 4 step approach starting with the BR definition

Top-down approach:
- Workshops with business users
- Study of the existing documentation of the Business Rules
- Review of reports on Data Quality
- Own knowledge of the environment
- Classification of the BR according to Data Quality dimensions

Bottom-up approach:
- Data analyses on the contract data
- Identification of the key business rules based on results

Define prioritization criteria:
- Agree on prioritization criteria for the implementation of the business rules
Top-down approach
Approach at EuropeAid

Reports & Business knowledge on known issues

Data Identification & Business rules identification

Questions to identify business rules:
• How do you control the figures?
• What are the most important contract fields for your daily activities?
• What are the business rules specific to your activities?
**Bottom-up approach**

Screening of the data (data profiling)

**Data profiling**

Key analysis step to understand the data at field level (e.g.: uniqueness, range of values, list of values) and datasets relationship

**Input**

- Data set extracted from CRIS

**Key Activities**

- Data pattern matching and processing
- Develop and execute business rule assessment
- Relationship between data

**Output**

- Data statistics identifying data anomalies, interpretability, completeness, and integrity issues
- Data remediation strategy
Define prioritization criteria

The DQ controls are prioritized based on the following criteria:

1. **Feasibility**
   - What is the complexity of implementing a business rule to control the data?
   - How long will it take to develop the control?

2. **Materiality**
   - What is the impact of not respecting the business rules in man days? What is the efficiency loss in man days due to bad data quality?
   - Because of this error, I need to spend 3 more days to correct my report.

3. **Reputation**
   - What is the impact on the DG reputation?
   - What will happen if I send wrong figures to external parties?

4. **Compliance**
   - Are we compliant with the European regulation?
   - Are we compliant with the Financial regulation, with the accounting rules...?
Project approach

Step 2: Implementation

**Business rules identification**

**Implement Business rules**

**Monitor Results**

**Feedback loop**

**Design:**
- Document functionalities of framework (use case and activity diagrams)
- Define template for manual data quality checks results

**Build:**
- Build the framework using SAS Data Management Studio
- Apply the business rules to a set data
- Produce the results of the quality assessment through a dashboard
- Validate the implementation

Data Owner: IT
Project approach
Step 3: Visualization of the results

**Business rules identification**
- Implement business rules
- Monitor results
- Feedback loop

**Design:**
- Design the dashboard, contents and layout
- Data quality results per dimension
- Evolution of the results across time

**Build:**
- Build the dashboard
- Validate the implementation
Project approach

Step 4: Results usage

Feedback Loop:

- Identification of the issue by the data owner
- Evaluation of the issue(s)
- Prioritization
- Communication to the data encoder for correction
- Communication to CRIS team for potential bug/defect correction
- Communication to the IT DQ team for post processing
- Communication to USMs for acceptance of the RFCs

Data Owner role at the DG EuropeAid

Person who is accountable for a data or a set of data.
Specifies the DQ standards even though the data may have been collected or disseminated by another party.
Agenda

1. Presentation of EuropeAid
2. Context of the project
3. The Data Quality framework
4. Lessons learned and Going forward
Data Quality framework

High level architecture

DQ Framework key components

1. **Profiling**: Screening of all tables in scope
2. **Metadata Comparator**: Performs the technical controls on the data model
3. **Business Rules**: Execute all business controls defined with the data owners
4. **DQ Data Mart**: Consolidate results from technical and business controls for reporting purpose
Data Quality framework

Key strengths

**Profiling**
Check up of your database by screening all tables and columns in scope through several criteria

**Metadata Comparator**
Compare technical constraints on database with the profiling results

**Business validation**
Meetings organized with data owners and data stewards to identify pain points and categorize DQ checks by priority

**50 Business Rules**
Implemented in SAS DMS
- Logging and error handling
- Integration of manual checks
- DQ issue exclusion
Data Quality dashboard

- Key strengths:
  - Conformity
  - Reliability
  - Accuracy
  - Completeness
  - Consistency
  - Duplication

- Dataset view:
  - Data Quality Score = 97.29%

- Results by Error Types:
  - Evolution by Error Types:
    - Error Types: Accuracy, Completeness, Conformity, Reliability
    - Evolution from 2014 to Q3 2017

- Error Types:
  - Accuracy: 99.1%
  - Completeness: 99.1%
  - Conformity: 99.1%
  - Reliability: 99.1%

- Number of Business Rules controlled:
  - Number of Business Rules = 0

- Error Percentage:
  - Deviation
  - Error Percentage
  - Error Percentage

- Error Types:
  - Accuracy: 99.1%
  - Completeness: 99.1%

- Error Details:
  - Conformity
  - Reliability
  - Accuracy
  - Completeness
  - Consistency
  - Duplication

The dashboard provides drill down capabilities from the directorate general to the delegation in charge.

A dedicated tab called 'Error Details' provides detailed information on the contract(s) in error, including the contract number and title, person in charge, contract status, nature and submission, as well as technical information such as the table where the data is stored.

The detailed information can be carried from the dashboard to export to Excel.
Agenda

1. Presentation of EuropeAid
2. Context of the project
3. The Data Quality framework
4. Lessons learned and Going forward
Lessons learned of the project

Skills and Capabilities
- Functional knowledge of the DG, its role and functioning
- Facilitators’ skills
- Experienced SAS DMS developers

Approach and technology
- Need a structured approach, as the methodology specifically developed by Deloitte
- Data Quality dedicated tools to support the exercise

Accelerators
- Data Quality project amongst a data quality program
- Deloitte SAS DMS framework to accelerate development hence help focusing on solving the data quality issues
Going forward

Further develop the BR

• Use the framework to extend the business rules to other domains

Governance

• Support and enforce the governance using the data quality dashboard for communication and monitoring
2. How many EU delegations are supported by EuropAid?

A. 140  
B. 150  
C. 1234

Tweet your answer:

Example: @spicyanalytics 2C

Prizes to win:

1st prize: a ticket for Analytics 2015
2nd prize: a book of Prof Bart Baesens: “Analytics in a big data world”
3rd to 30th prize: chocolates with pepper

Winners will be contacted post-Forum!
Thank you