

Case study

Europees Patent Office

**Migrating without migraine
the perfect opportunity to improve data
quality**

Bart Heinsius
April, 3 2008

Content

- About the European Patent Office (EPO)
- About EOM
- The project *NPL Rebuild - Cleansing*

About EPO



- Rijswijk, Munich
- Patent protection in 38 countries
- 3,500 examiners assessing 135,000 requests (2006)

About EOM



- Provides SAS consultancy and implementation services
- 8 employees
- Partner of SAS since establishment in 1998
- First Data Quality project with Dataflux

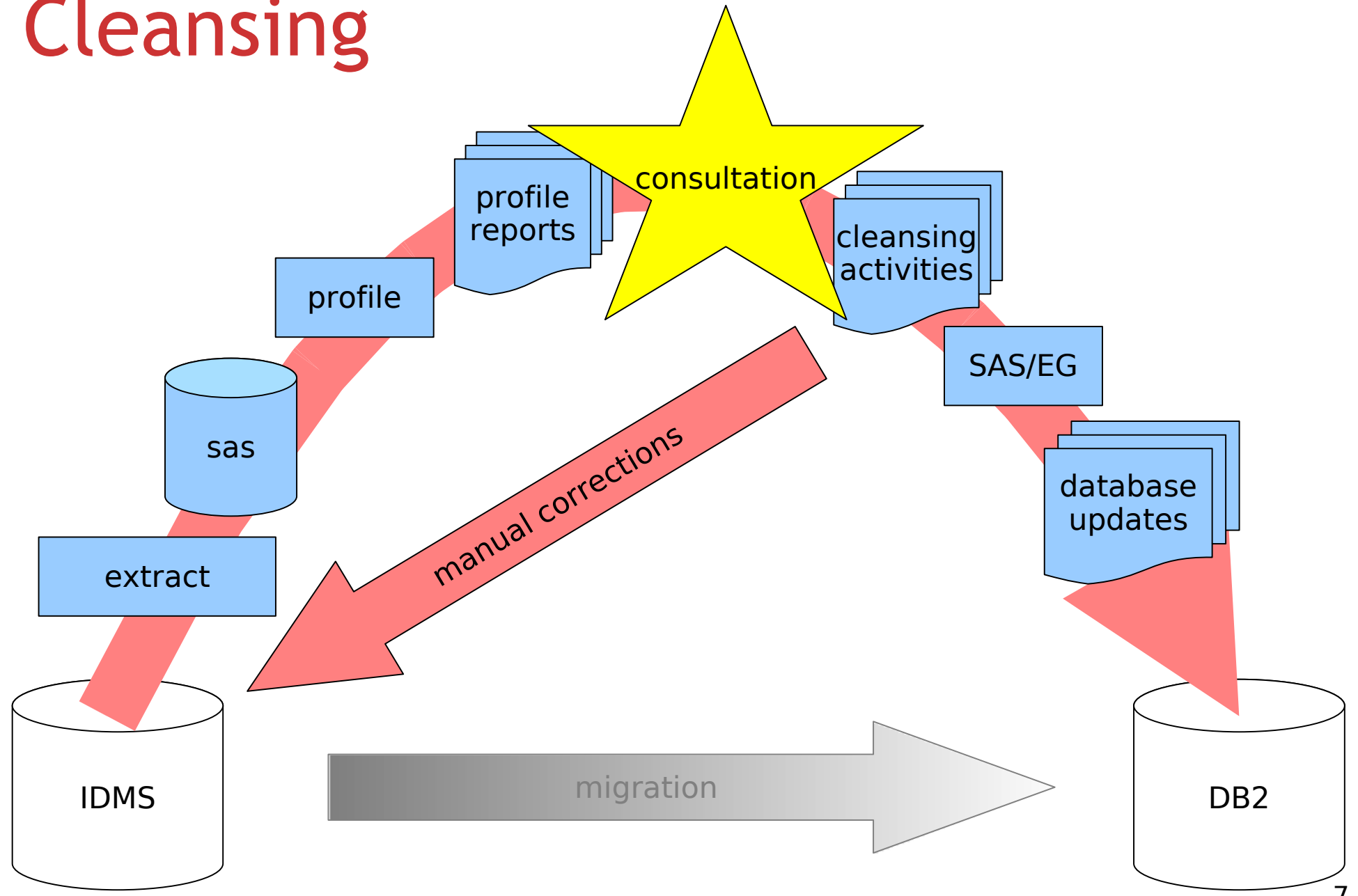
The project *NPL Rebuild*

- Migration bibliographical data from IDMS to DB2
 - Books
 - ISBN, Title, Author, Publication Date, Publisher, ...
 - Magazines
 - ISSN, Title, Volume and Issue, ...
 - Conference Proceedings
 - Conference start date, conference end date, ...

The project *NPL Rebuild*

- Migrating bibliographical data from IDMS to DB2
 - Internet
 - URL
 - Articles
 - Page Number, Title, Author, ...

Cleansing



Consultation

- Know-how
- History
- Standards
- Users
- Decisions
- Project duration

Results

- Approx. 1% invalid ISBNs and ISSNs
- Control characters, spaces, dots, commas
- Books: 16% incorrect Country Code Publisher
- 34% URLs don't exist (anymore)
- Missed input batches
- 20080000, 19980582, 00002001, 0, NULL
- ...

Future requirements

- Correcting
- Continuous monitoring and correcting
- Adapt interfaces
- Matching

QUESTIONS