

Praktijk-case Europees Patentbureau

**Migratie zonder migraine
de aanleiding voor datakwaliteit**

Bart Heinsius
3 april 2008

Agenda

- Over het Europees Patentbureau (EPO)
- Over EOM
- Het project *NPL Rebuild - Cleaning*

Over EPO



- Rijswijk, München
- Patentbescherming in 38 landen
- 3.500 examiners beoordelen 135.000 aanvragen (2006)

Over EOM



- Levert SAS-consultancy- en implementatiediensten
- 8 medewerkers
- SAS Partner sinds oprichting in 1998
- Eerste Data Quality-project met Dataflux

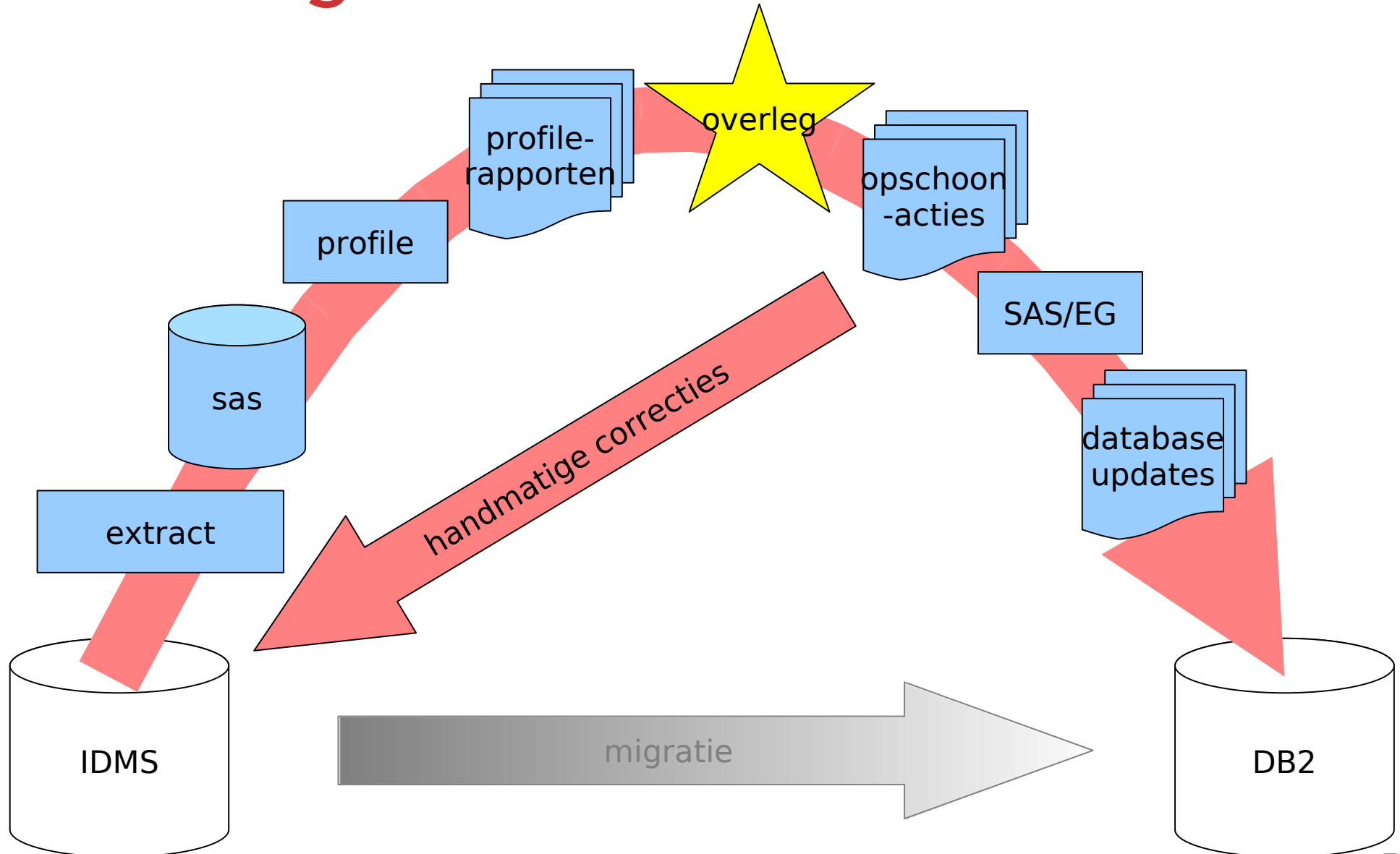
Het project *NPL Rebuild*

- Migratie bibliografische data van IDMS naar DB2
 - Boeken
 - ISBN, Titel, Auteur, Publicatiedatum, Uitgever, ...
 - Tijdschriften
 - ISSN, Titel, Jaargang en Nummer, ...
 - Conference Proceedings
 - Conferentiestartdatum, -einddatum, ...

Het project *NPL Rebuild*

- Migratie bibliografische data van IDMS naar DB2
 - Internet
 - URL
 - Artikelen
 - Op welke pagina's, Titel, Auteur, ...

Cleaning



Overleg

- Kennis van zaken
- Historie
- Standaarden
- Gebruikers
- Beslissingen
- Doorlooptijd

Bevindingen

- Ca. 1% ongeldige ISBN's en ISSN's
- Controlekarakters, spaties, punten, komma's
- Boeken: 16% foutieve landcode Uitgever
- 34% URL's bestaan niet (meer)
- Misgelopen inlees-batches
- 20080000, 19980582, 00002001, 0, NULL
- ...

Vervolg

- Corrigeren
- Blijven controleren en corrigeren
- Interfaces aanpassen
- Matching

VRAGEN