



Marketing Minds Specialising in Research™

NFO EX▪A▪MINE Centre

Incorporating CRISP-DM in the Data Matching Process

York, 8th November 2002

Ursula Becker

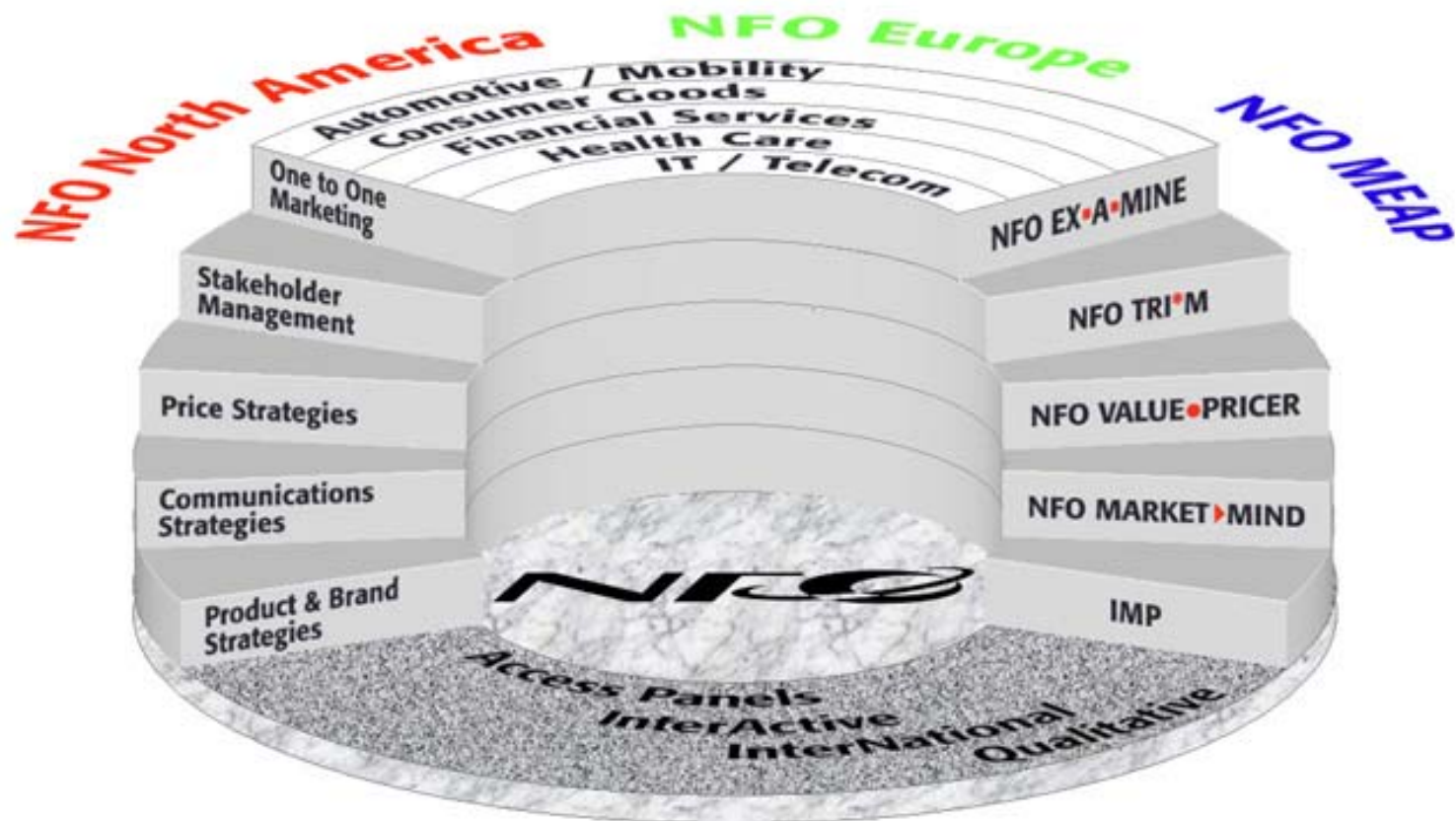
Agenda

- NFO – who we are
- Introduction EX▪A▪MINE and Data Matching
- Project challenges and project documentation
- Summary

NFO – who we are

- NFO WorldGroup is the world's premier provider of custom, research-based marketing information and counsel.
- NFO Infratest is part of NFO WorldGroup (Greenwich, CT / USA),
- thus belonging to one of the leading market research organisations with 65 companies on all continents –
- thereof NFO Infratest in 15 European countries.

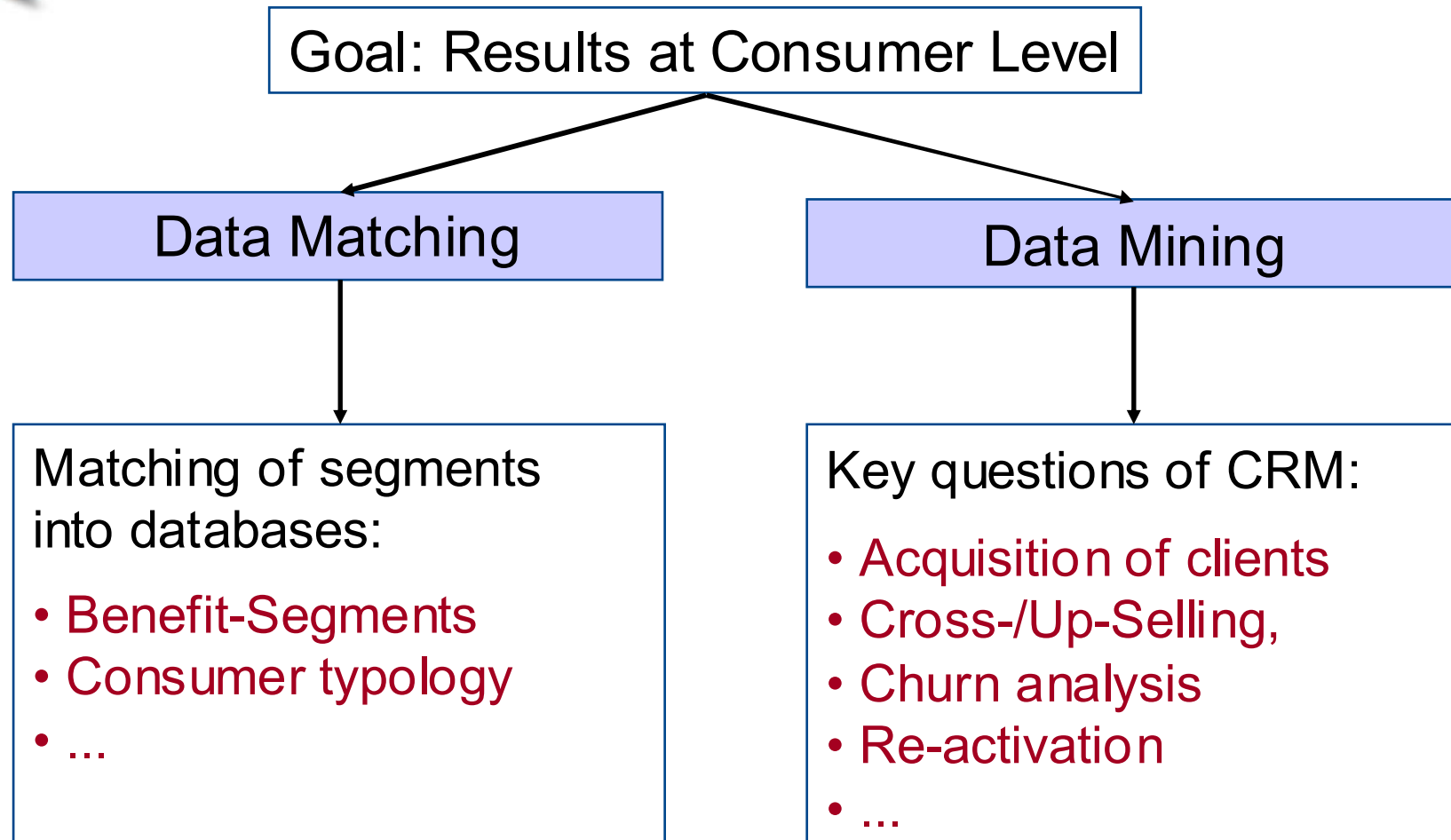
NFO – who we are



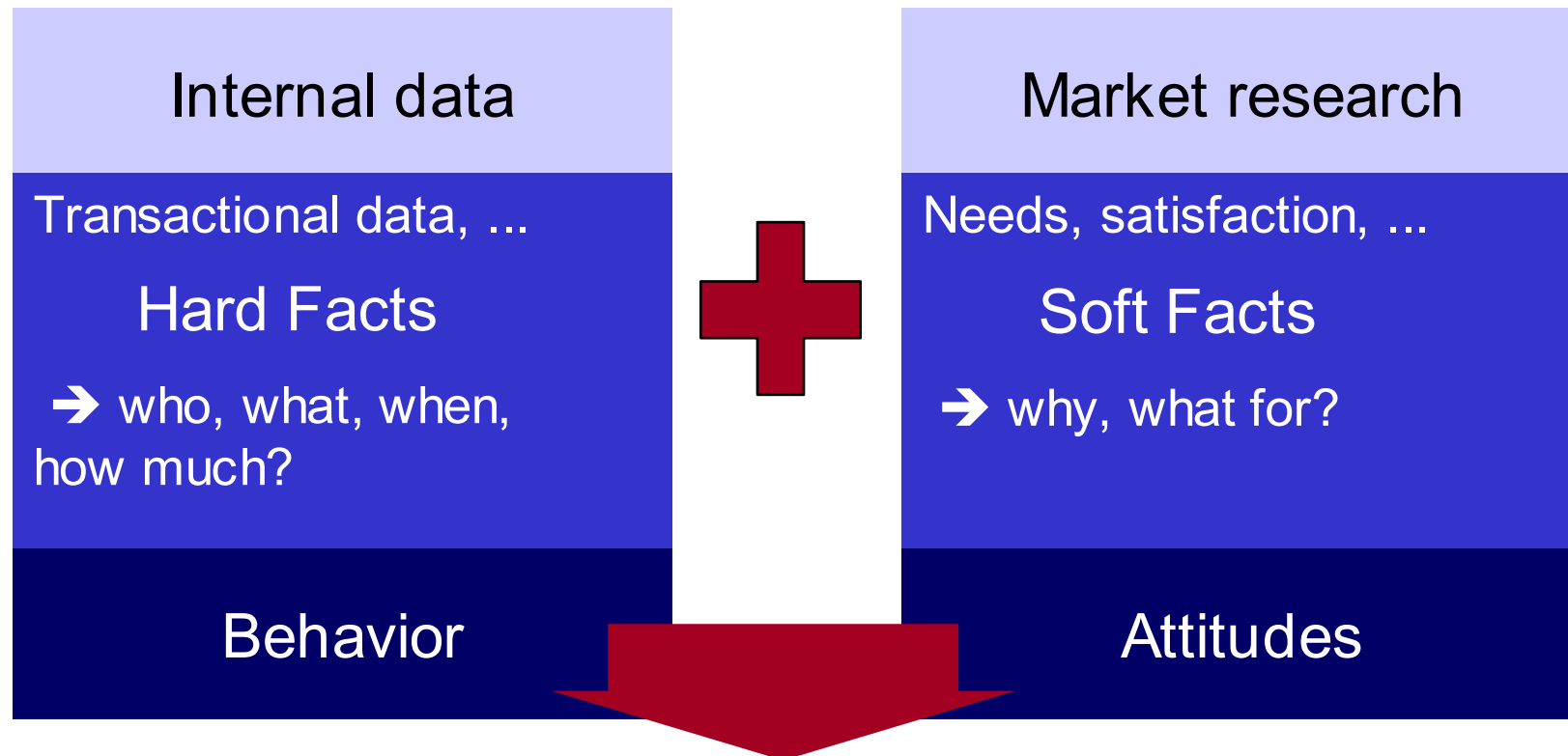
Agenda

- NFO – who we are
- Introduction EX▪A▪MINE and Data Matching
- Project challenges and project documentation
- Summary

Goals of EX▪A▪MINE



External data brings additional value to internal data



Attitudinal database enrichment

The NFO modules for a holistic consumer understanding

| INTERNAL | EXTERNAL | | |
|---|--|---|---|
| <p>Client's data</p> <p>Level: Person</p> | <p>Attitudes</p> <p>Market research</p> <p>Level: Microsegment</p> | <p>Competitors / Market</p> <p>Market research e.g., NFO Access Panels</p> <p>Level: Microsegment</p> | <p>Psycho-social structural data</p> <p>Geo-demo- graphics & Lifestyle</p> <p>Level: Street/Block</p> |

NFO EX·A·MINE
Database Enrichment - Data Matching

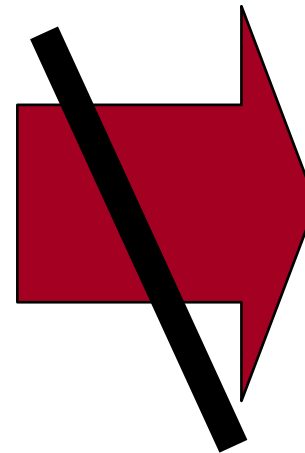
Holistic Consumer Understanding (for e.g., CRM)

Market research results are anonymous and aggregated

Conventional market research:
45% of all customers in region A
belong to segment 2



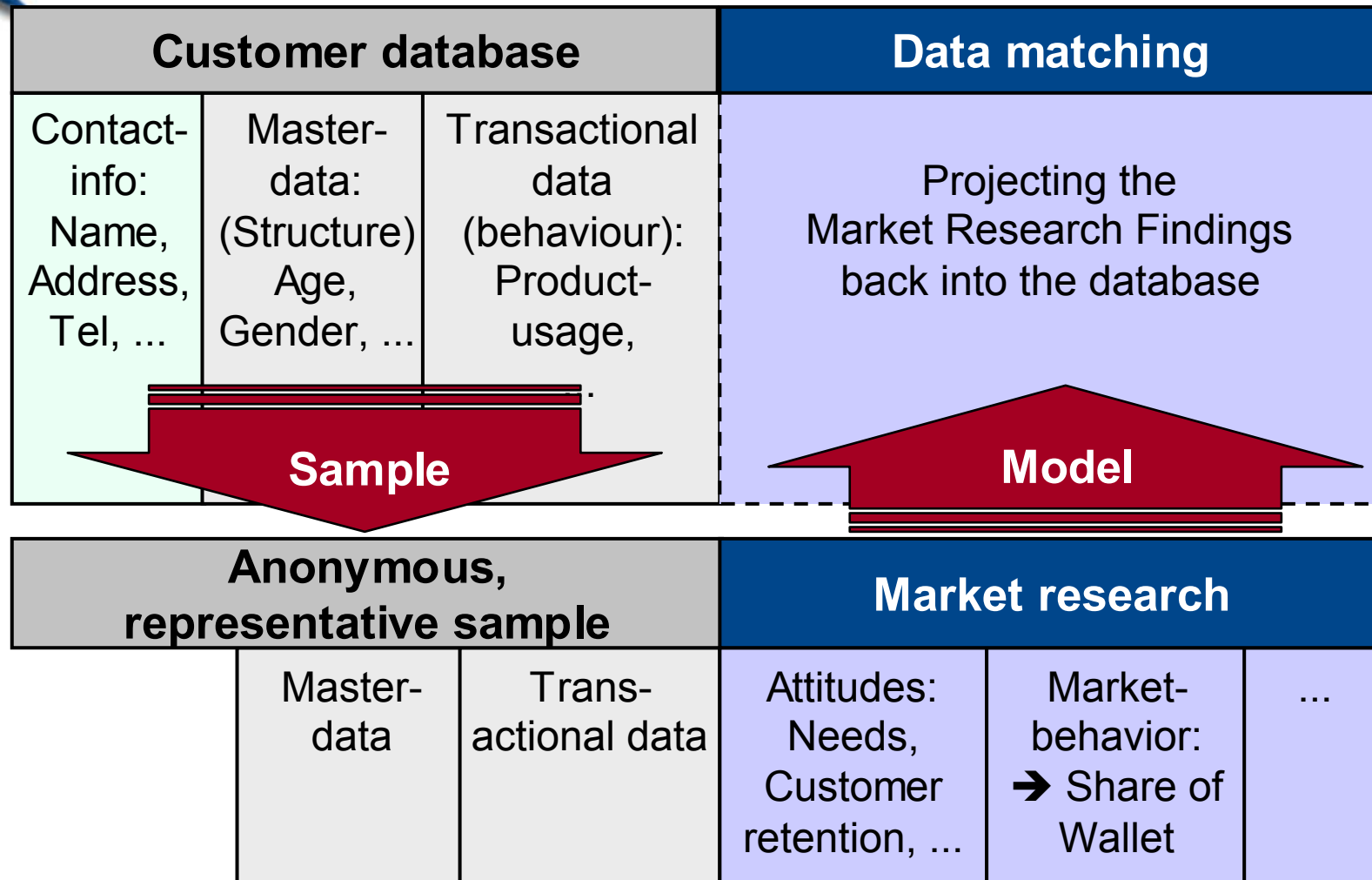
Not-interviewed:
Segment 1?
Segment 2?



MR results have to be anonymous!

- Privacy protection guidelines
- ESOMAR and national associations' conventions
- Cost saving by interviewing only sample of customers
- Different response behavior:
personalised ↔ anonymous interviews
- Misleading to work with personalised attitudinal statements

Database enrichment through data matching

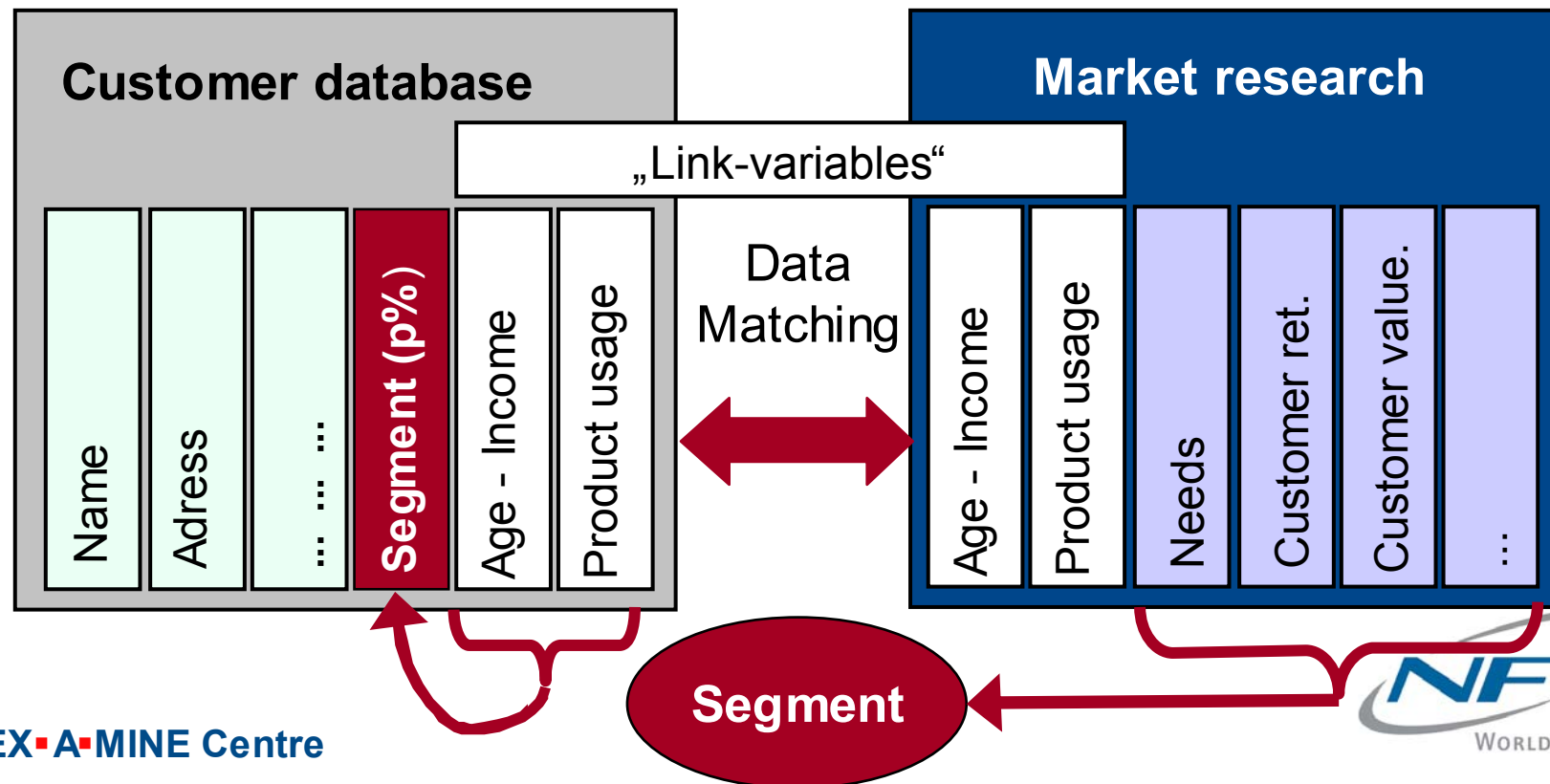


Data matching: modelling of homogeneous segments

Men between 25 and 35 have a **probability** of 76% of belonging to segment 1



In the group of men between 25 and 35 years, the **share** of segment 1 is 76%.



Agenda

- NFO – who we are
- Introduction EX▪A▪MINE and Data Matching
- Project challenges and project documentation
- Summary

Our project characteristics

- Several people involved in one project
- Joint analysis of several data sources
- Evaluation of different algorithms

The EX▪A▪MINE analytical toolbox

- **Multivariate Statistics**

- Logistic, Categorical, Linear Regression,...
- Multivariate Adaptive Regression Splines (MARS)
- Latent Class analysis, Hierarchical Bayes Regression

- **Decision Trees/Rule Induction**

- CART, C5.0, QUEST, CHAID, Association Rules, ...

- **Artificial Neural Networks**

- Cascade Correlation, MLP, SOM, ...

- **Hybrid Methods**

- Genetic Algorithms, Neuro Fuzzy Algorithms, ...
- Interactive data structure visualisation methods

Our project characteristics

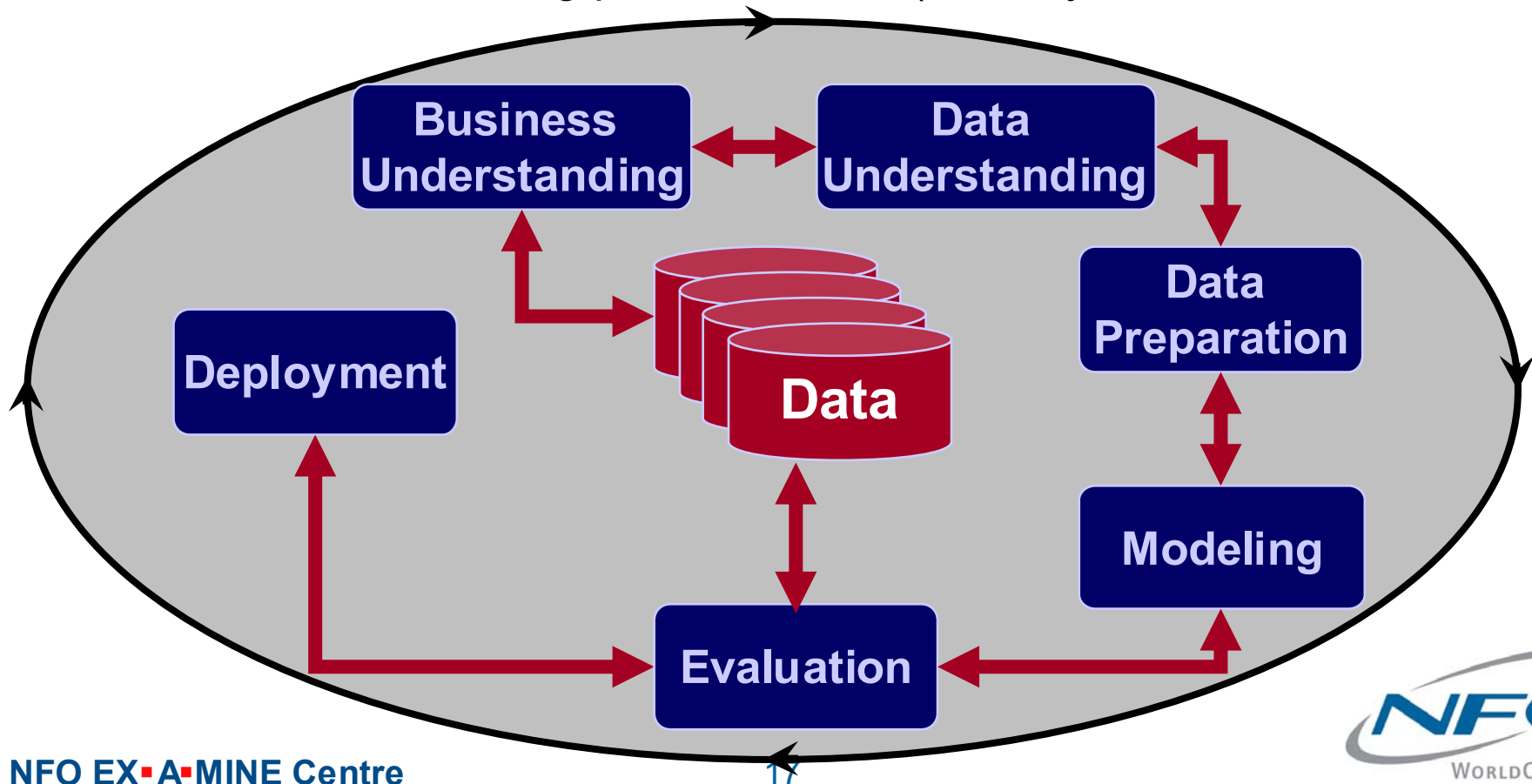
- Several people involved in one project
- Joint analysis of several data sources
- Evaluation of different algorithms
- Iterative procedure with client workshops
- Inquiries after a long time and re-analysis of old data with new data added

→ Documentation of project and analyses is indispensable!

CRISP-DM:

The framework of our project documentation

CRISP-DM: Cross Industry Standard Process for Data Mining
a universal Data Mining process model (industry neutral and tool neutral)




The project directory


 0_Correspondence


 1_Business Understanding

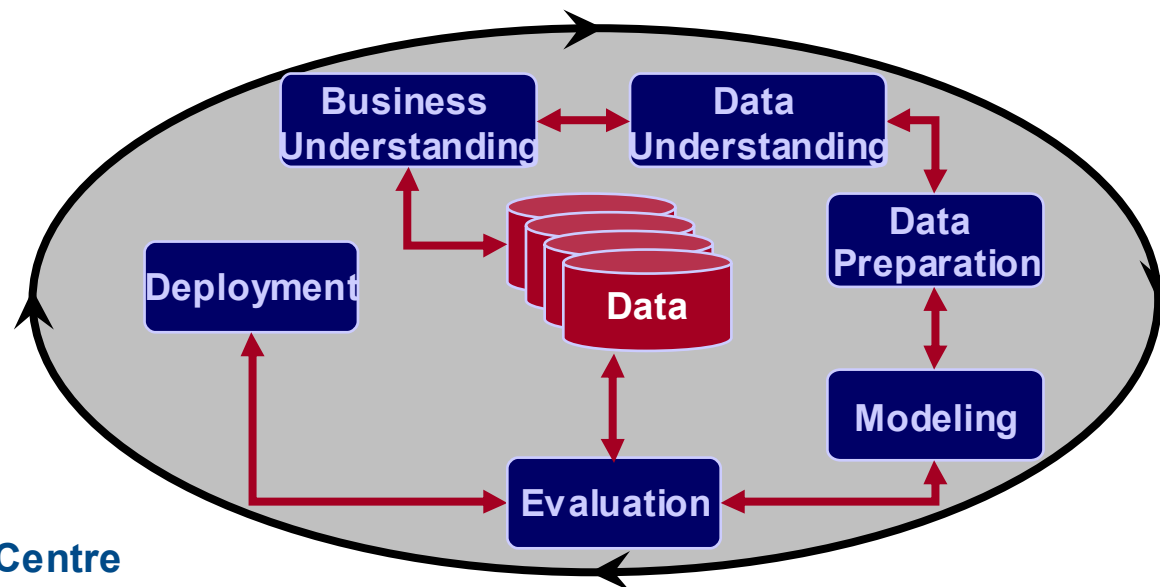
 2_Data

 3_Data Preparation

 4_Modelling

 5_Evaluation

 6_Deployment

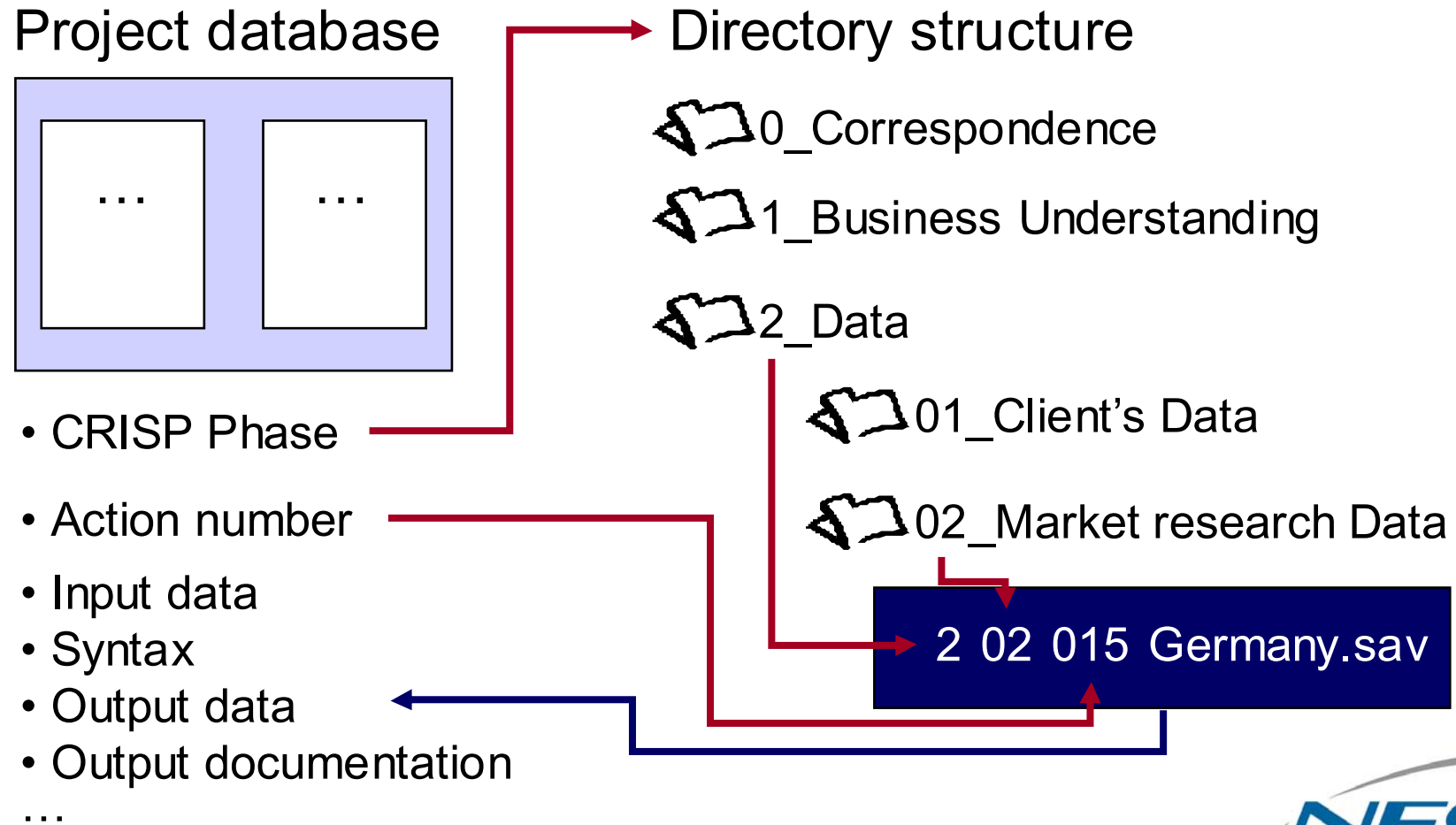


The EX▪A▪MINE project database allows a fast and easily understandable whilst comprehensive project documentation

1. Action number
2. Goal and description of action
3. CRISP-phase and CRISP-task
4. Status
5. Responsible
6. Starting date
7. End Date

8. Input data
9. Syntax
10. Output documentation
11. Output data
12. Open questions
13. Solution

Our project documentation



SPSS – the backbone of our analyses

Reasons, why we use SPSS

- Quick overview of the data
- Use of Syntax
 - Crucial for the documentation of the analyses
 - PASTE function to generate syntax
 - Altering of syntax for standard jobs
 - Easy use of annotations
 - Import and Export of data easily possible
- Ability to deal with large data sets
- Reasonable cost/performance ratio
- Labelling of variables

SPSS – the backbone of our analyses

Reason, why we don't use SPSS alone:

- unfortunately not all algorithms we need are implemented, especially novel developments are not included:
 - MARS
 - Latent Class Analysis
 - Neural Networks
 - Specialised software offers a wider variety of options and variants for algorithms (e.g., Cluster analysis)
 - ...

Reason, why we don't use a Data Mining Suite:

- so far, no Suite covers everything we would like to have
- cost/performance ratio of suite vs. our stand-alone solution does not justify the purchase of such a suite

Agenda

- NFO – who we are
- Introduction EX▪A▪MINE and Data Matching
- Project challenges and project documentation
- Summary

Summary

- Data Matching is an iterative process
- CRISP-DM is a very useful and well established standard for data analysis
- A CRISP-DM project database is indispensable
- SPSS is a very good tool for data handling and certain analyses
- Several new algorithms are missing in SPSS

→ Use CRISP-DM
not just as theoretical framework but also in practice!

The EX▪A▪MINE Centre

- Dr. Thomas Liehr (Head of the EX▪A▪MINE Centre)



++ 49 - 89 - 5600 520

mail: thomas.liehr@nfoeurope.com

- Ursula Becker



++ 49 - 89 - 5600 423

mail: ursula.becker@nfoeurope.com

- Gernot Hennig



++ 49 - 89 - 5600 542

mail: gernot.hennig@nfoeurope.com

ex-a-mine@nfoeurope.com

<http://www.ex-a-mine.com>

