Fundamentals of Business Intelligence

Univ.-Prof. Dr. Wilfried Grossmann
Univ. Prof. Dr. Stefanie Rinderle-Ma

Exercise Chapter 3: Data Transformation and Integration

The goal is to create a uniquely structured MXML log based on two differently structured data sources.

1. Data source: accounting data from hospital stays and medications
2. Data source: patient data (dermatology)

One source is already provided as database. The second one is a spreadsheet (excel). Both sources are to be integrated by means of given data integration model. The integration is designed as log format (MXML). Hence the integration is to be done at the database level. From the integration model the target MXML structure can be exported.

The integration should be repeatable – given the same data structures.

Tasks:

a) Create a list of problems and assumptions when/for doing the integration.
b) Integrate the data into the given data integration model from Figure 1.
c) Export the data from the data integration model into MXML and submit the resulting MXML file. One option to do so is using the SQLXML standard.

Important: Provide the sources as well as the description of your integration process.

Recommendations for integration:

1. Look up which data from both sources are *Activities* and which are *Parameters.*
2. Fill in the tables *Activity*, *Parameter* and *PossibleValue* using SQL.
3. Integrate the data from both sources into the integration model using SQL.

Data provided on the web site:

- Script for creating the data integration tables
- Script for creating the first data source
- Excel spreadsheet with data for the second data source

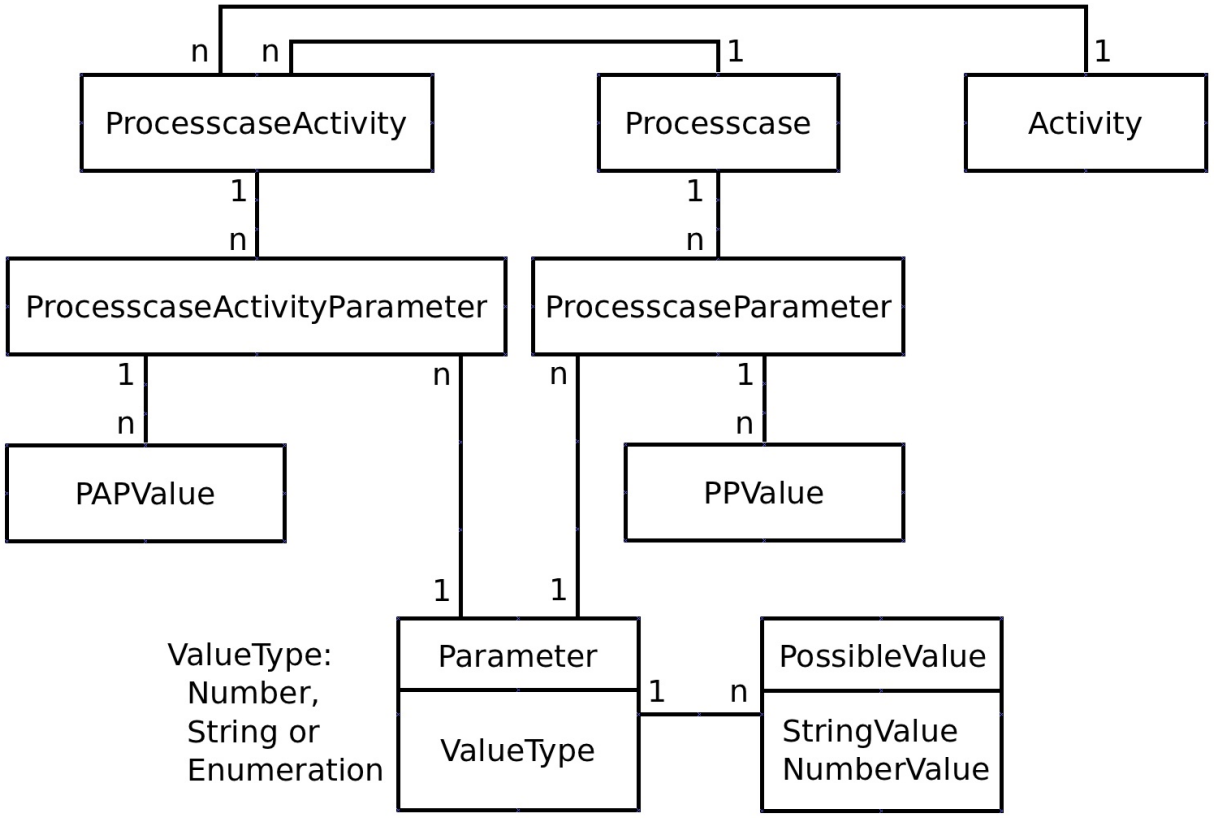

Figure 1: Data Integration Model

Additional information second data source:

- Localisation is determined at primary excision
- AJCC State is determined at Histological Examination
- MRI Diagnosis is determined at MRI
- CT Diagnosis is determined at CT
- Localization Distant Metastasis is derived from the two imaging examinations (MRI & CT)
- Tumor marker LDH is a result of lab test
- AJCC State Therapy is derived from the imaging examinations (MRI & CT) and the lab
- Each therapy is a separate activity

Hints:

- Insert statements can be created from Excel.
- A description of the MXML format can be found here: http://www.processmining.org/logs/mxml
- Use the options of bulk inserts and subselects
- For dissolving Strings using SQL in DB2: http://www.ibm.com/developerworks/data/library/techarticle/0303stolze/0303stolze1.html
- Some introduction slides to SQLXML are provided on the web site.