The Problem

1. Determine how data will be used
2. Build wrangling strategy
3. Profiling

Data obtained from other sources was not prepared for our purpose; needs wrangling.

50 – 80% of the time is spent here.

Complex & important process and yet not everyone knows how to do it.

Proposed Solution

1. Recurring problems
2. Recurring solutions
3. Reusable high quality solutions

Recurring solutions

Data obtained from other sources was not prepared for our purpose; needs wrangling.

50 – 80% of the time is spent here.

Complex & important process and yet not everyone knows how to do it.

Approach

Data Wrangling

Integration

Transformation

Visualisation

Cleaning

Design Patterns (DP)

Design Patterns (DP)

Extracting Design Patterns

A. Literature Reviews / Online Repositories

B. Extract from wrangling recipes

C. Get original dataset

Try to derive the recipe used

Get wrangled dataset

Evaluation of the DSL

DP as language construct in a new Domain-Specific Language for DW.