Automating the process of building flexible Web Warehouses with BPM Systems

Andrea Delgado, Adriana Marotta
Instituto de Computación, Facultad de Ingeniería
Universidad de la República,
Montevideo, Uruguay
Octubre 2015

INCO .::. FING .::. UDELAR

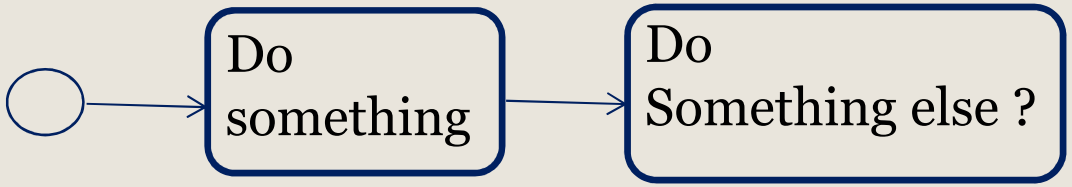
Agenda

- Motivation
- Quality-Aware Web Warehouse (QAWW)
- Business Process Management (BPM) & Systems
- Automated support for the WW configuration
- Case study with open data
- Conclusions and future work

Motivation

- In last years the amount of data generated on the web has grown considerably
  - Need to collect and analyze it to obtain useful information

    Data Warehouse with data sources from the Web = WW
- Extraction, Transformation and Loading (ETL) process
  - carried out in every DW construction without guidelines or standardized process structures



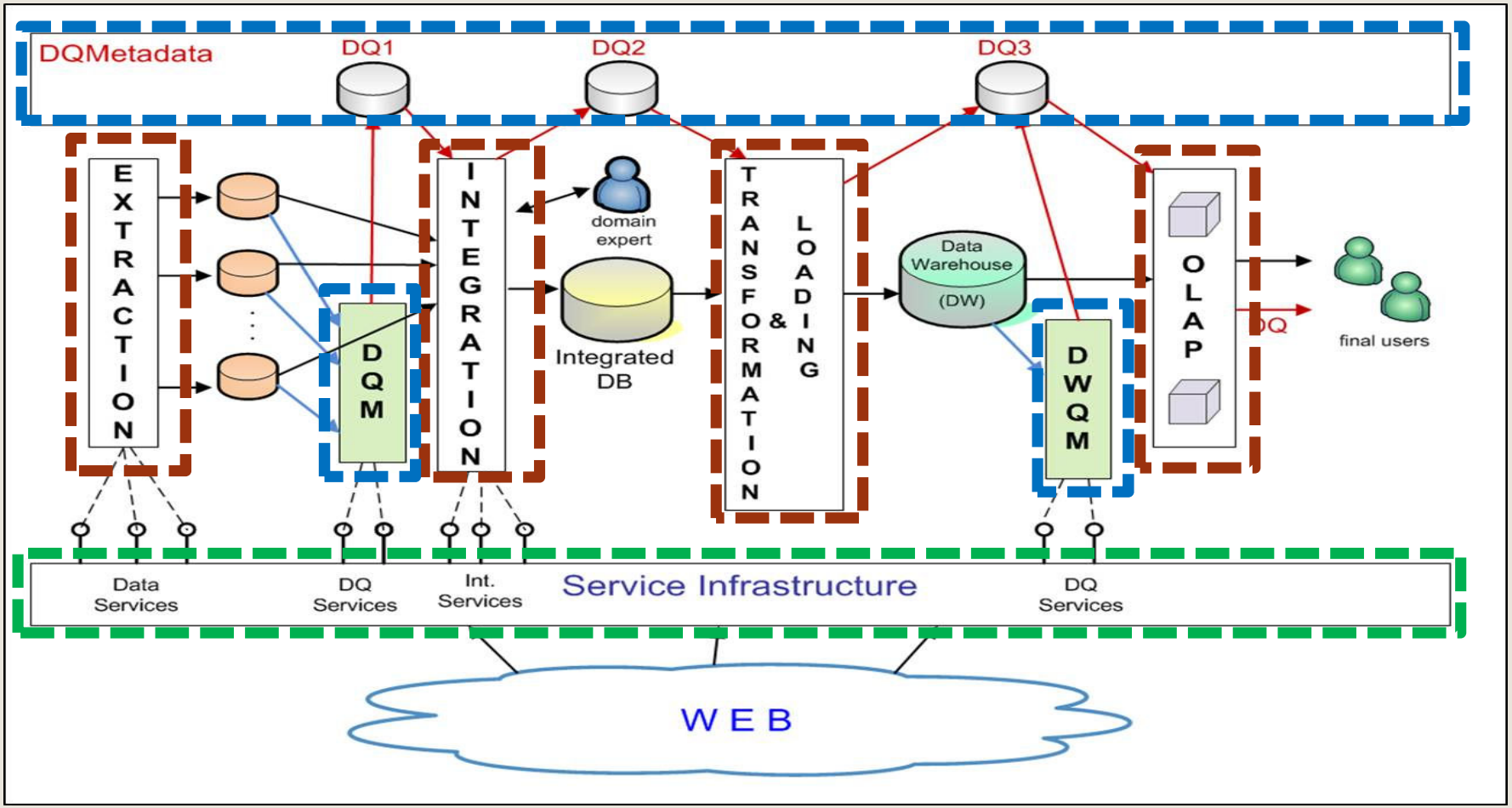
??????

  - in general it is left to the developer's criteria how to do it

with not much tool support

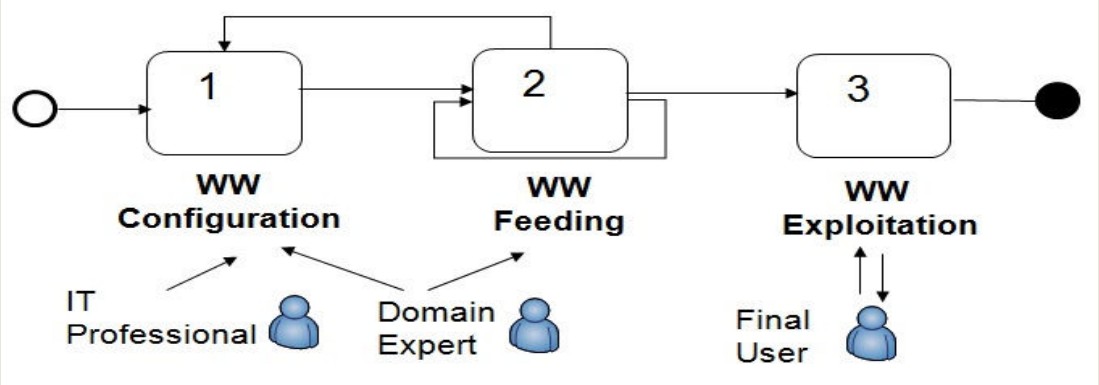
Quality-Aware Web Warehouse

- General Architecture of the QAWW



Quality-Aware Web Warehouse

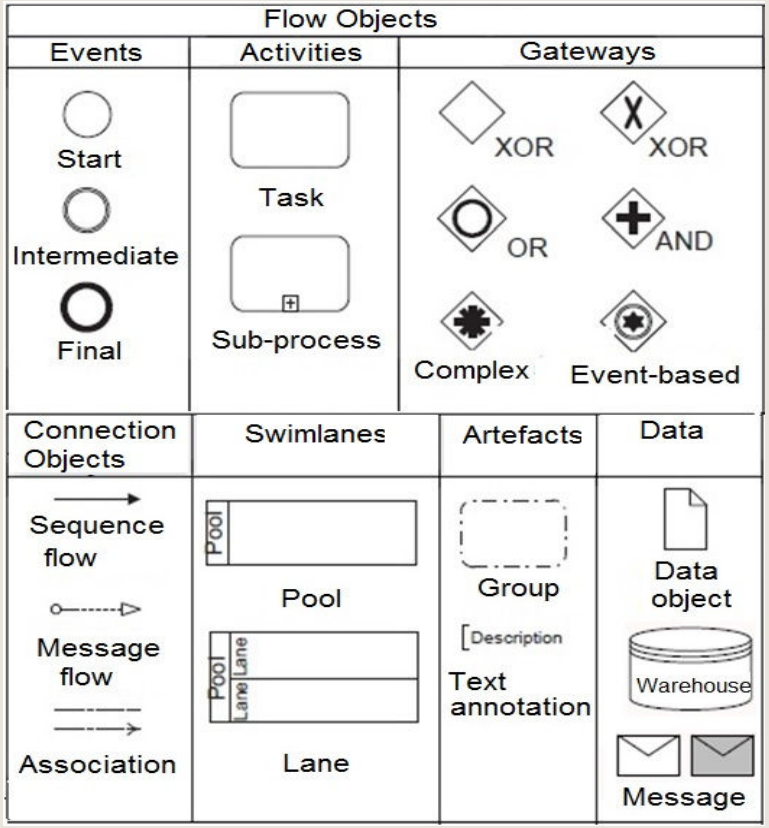
- QAWW phases and roles involved



- Two processes defined for WW construction
  - the configuration process to put light on
    - the activities to be performed to be able to configure the WW
    - the data needed to be able to construct the WW
  - the feeding process which uses these data
    - to automate as much as possible the WW construction

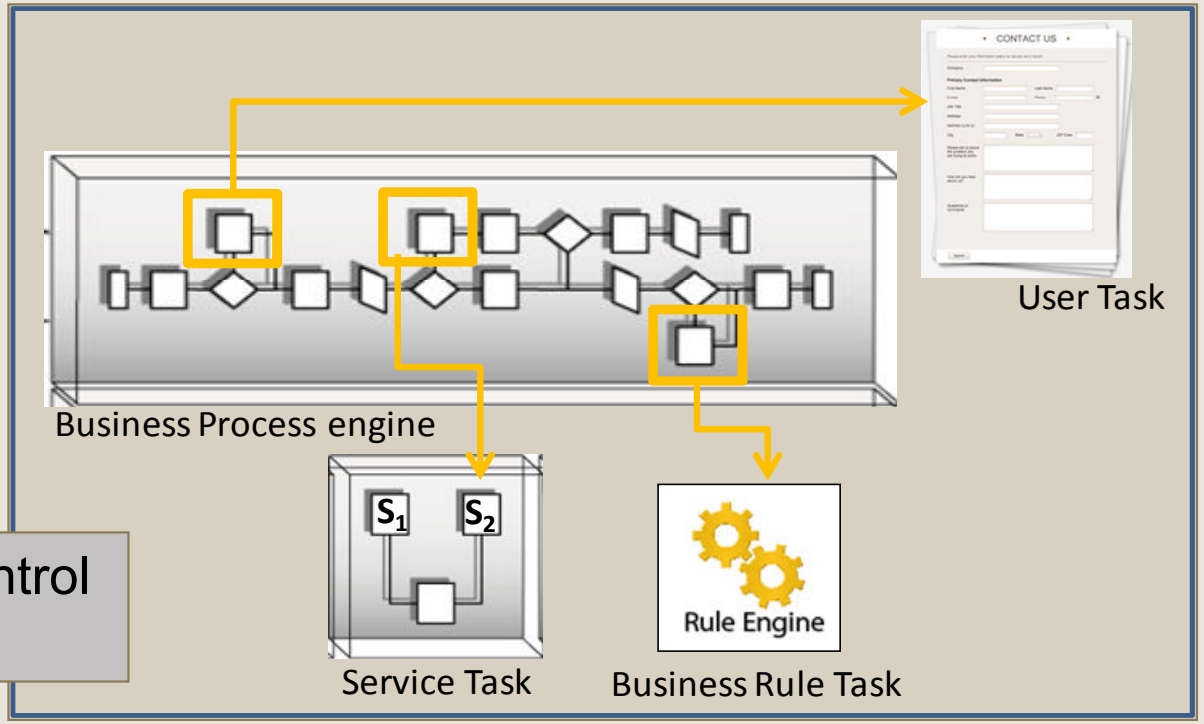
Business Process Management (BPM) & Systems

- Business Process Management (BPM)
  - provides an horizontal vision based on business process supporting their lifecycle
  - BPs are modeled, implemented, executed and evaluated in a continuous improvement cycle
    (Weske, 2007-Delgado et. al 2011-14)
    - Business Process Model and Notation (BPMN 2.0)
      - defines specific elements to model and execute business processes



Business Process Management (BPM) & Systems

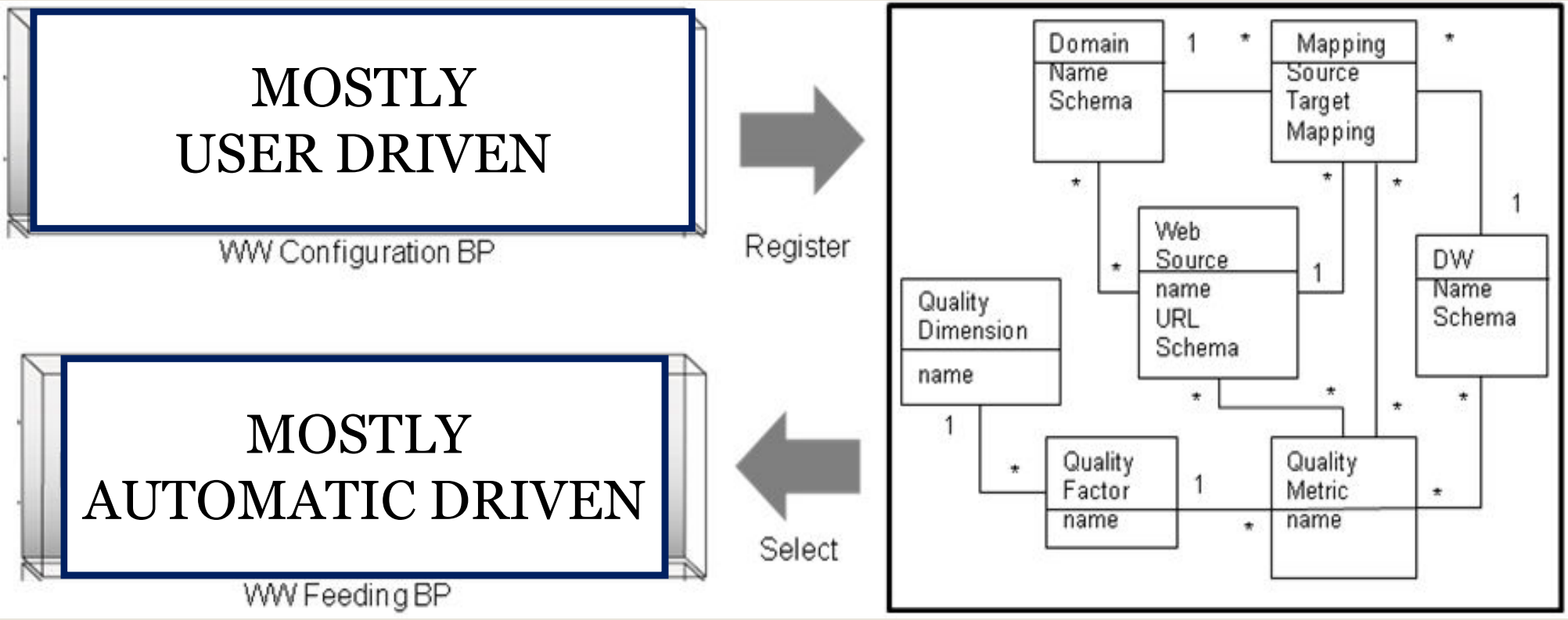
- Business Process Management Systems (BPMS)
  - provides specific platforms for executing business process models providing tools for each lifecycle phase
  - What we construct are Process Aware Information Systems (PAIS)



Inversion of control to the user

Automated support for the WW configuration

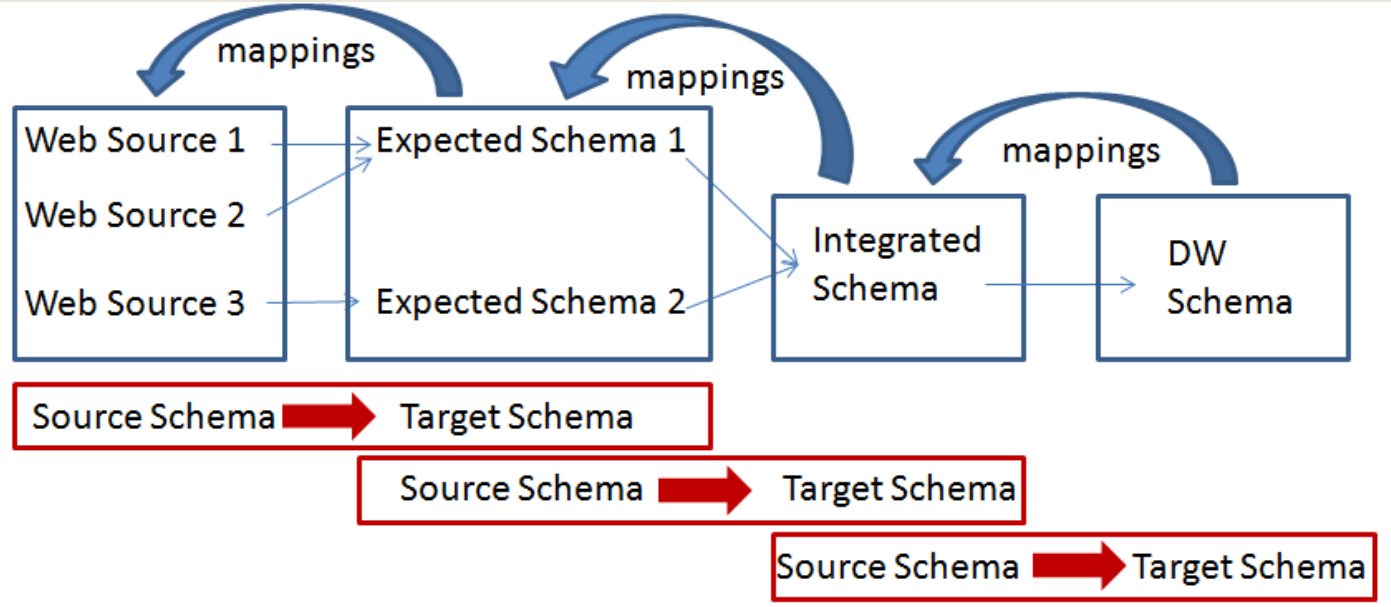
- Configuration and Feeding processes & metadata



Configuration Database

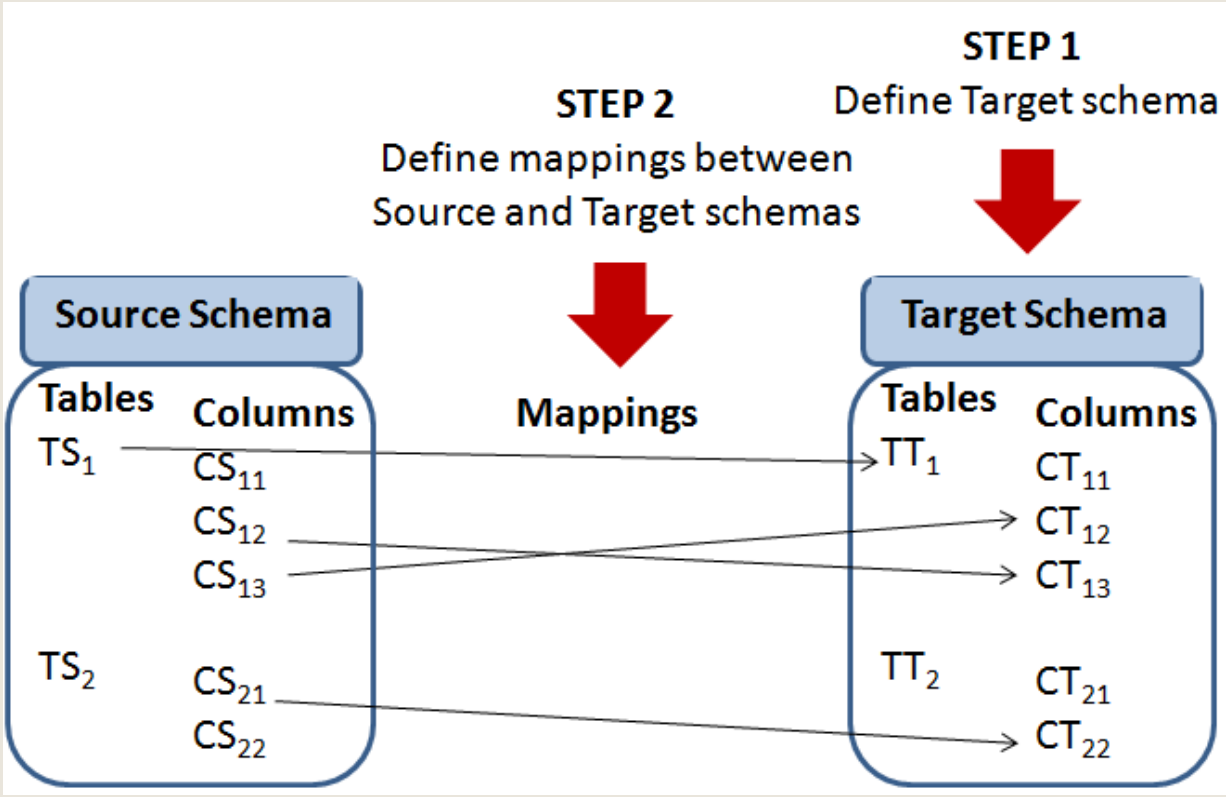
Automated support for the WW configuration

- Configuration process
  - Main objective is to gather from the user and register it in the configuration data base for the feeding process to read it
  - Things to register
    - Web sources, Schemas definition, Mappings between schemas



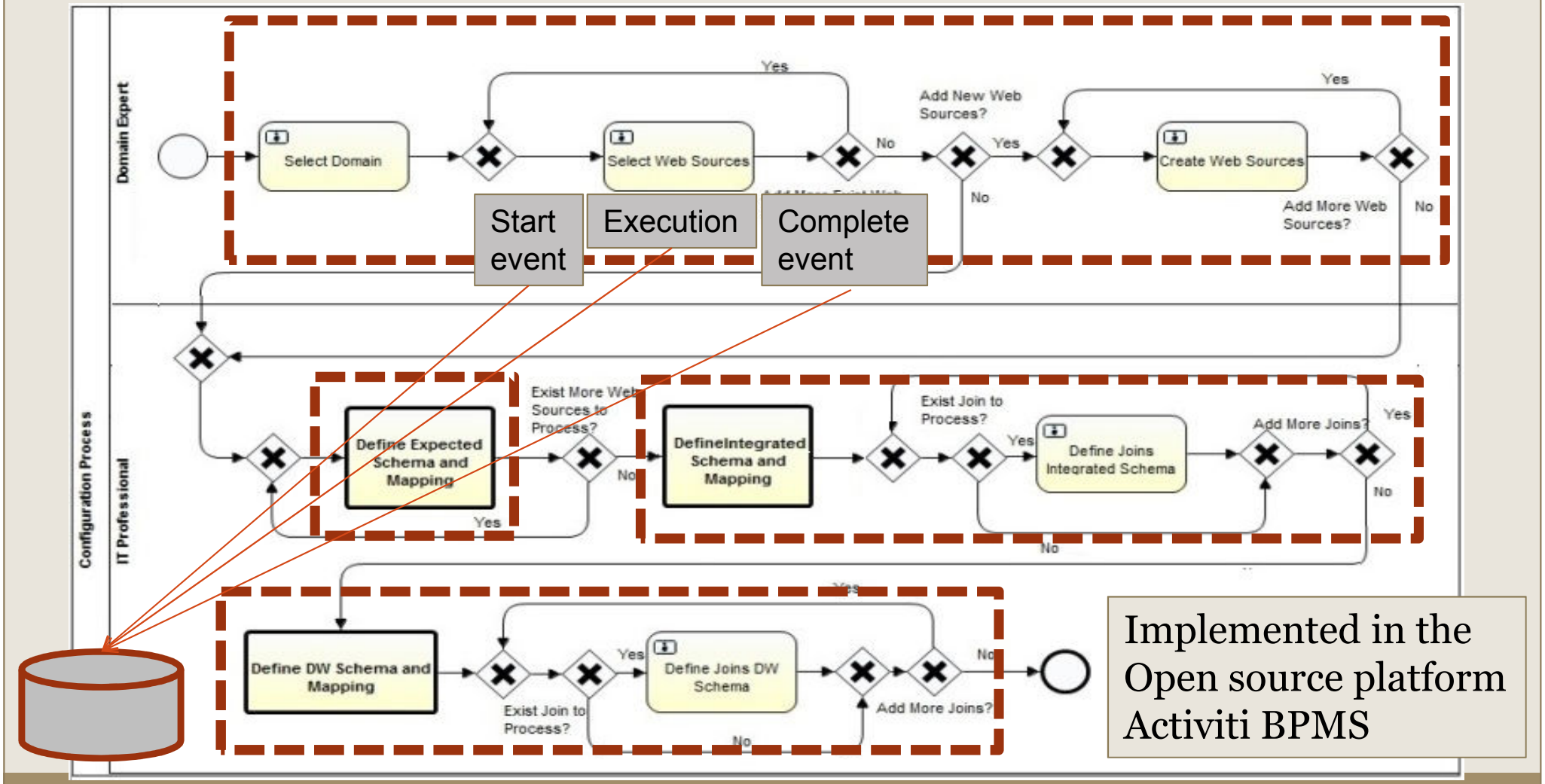
Automated support for the WW configuration

- Configuration process
  - Reusable conceptual sub-process to define mappings between source and target schemas



Automated support for the WW configuration

- Configuration Business Process specification in BPMN2.0



Implemented in the Open source platform Activiti BPMS

Automated support for the WW configuration

- Example form for user task in the Configuration process

Task Select Domain and Web Sources

Domain

Name	Selection
Web Source 2	☐
Web Source 3	☐
Web Source 4	☐

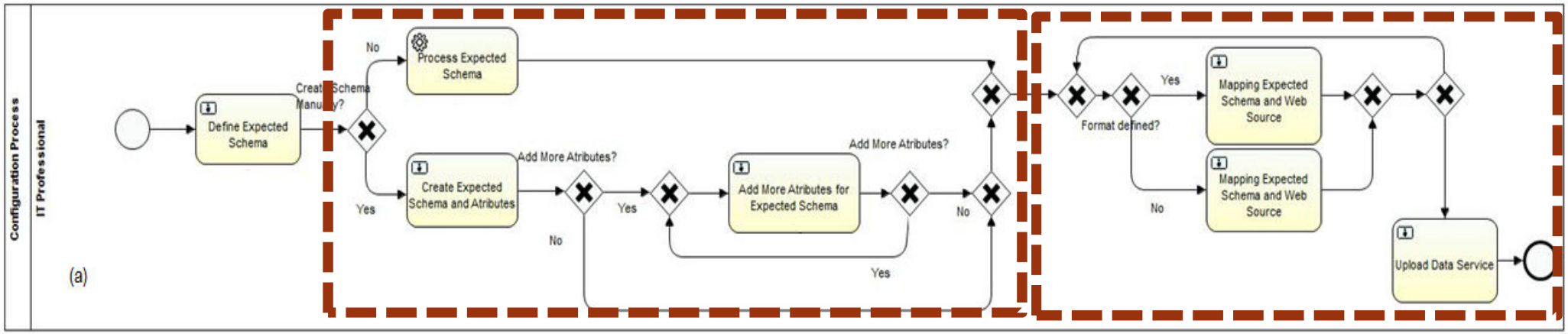
Register new Web Source? ☐

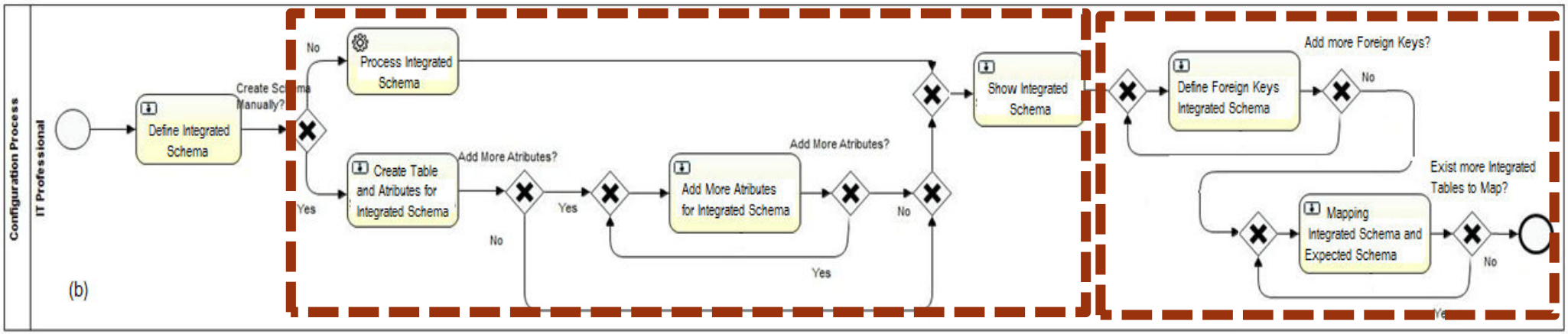
SubTask Register new Web Source

Name	URL

Automated support for the WW configuration

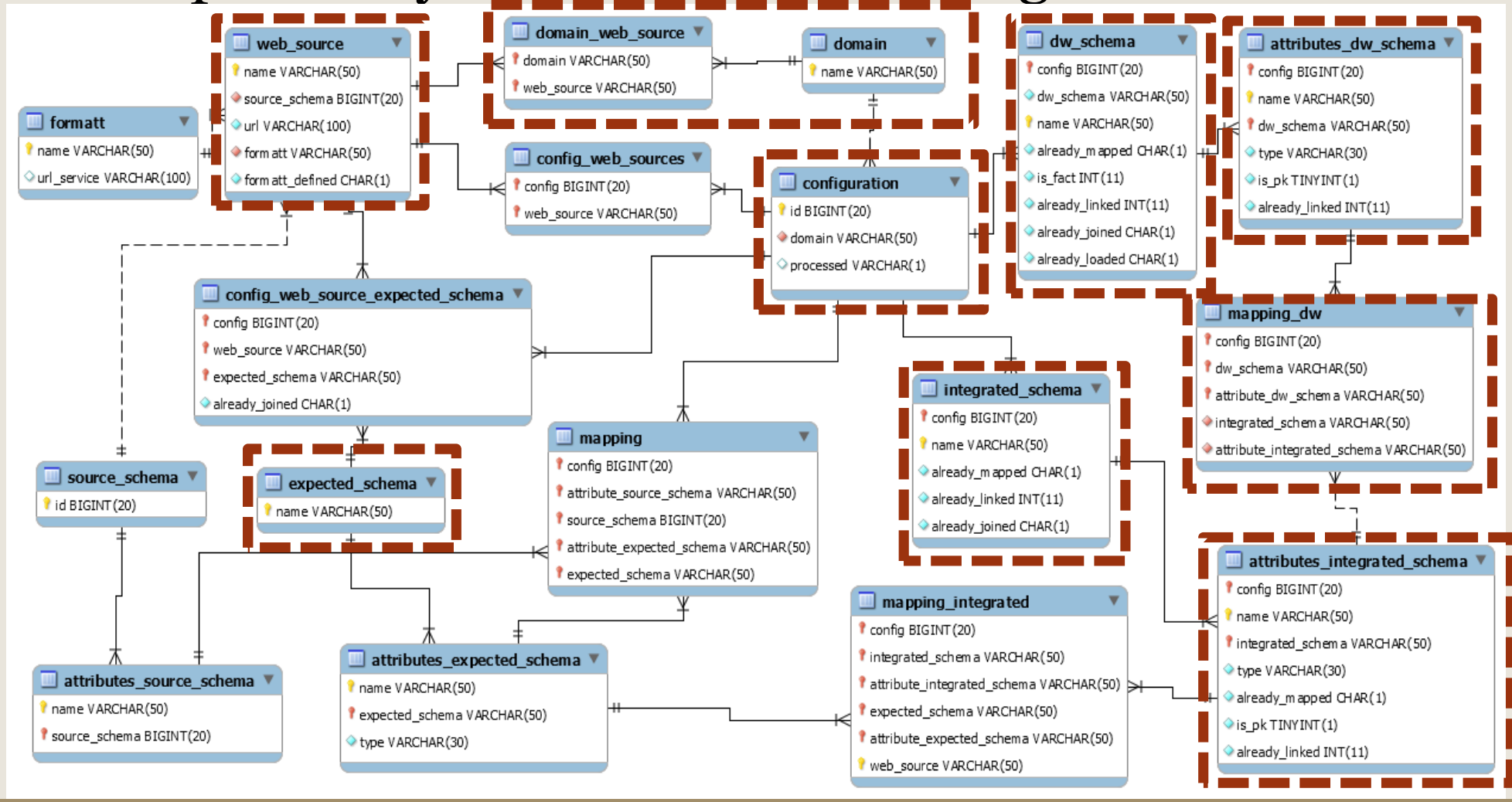
- Example of Expected and Integrated schema definition





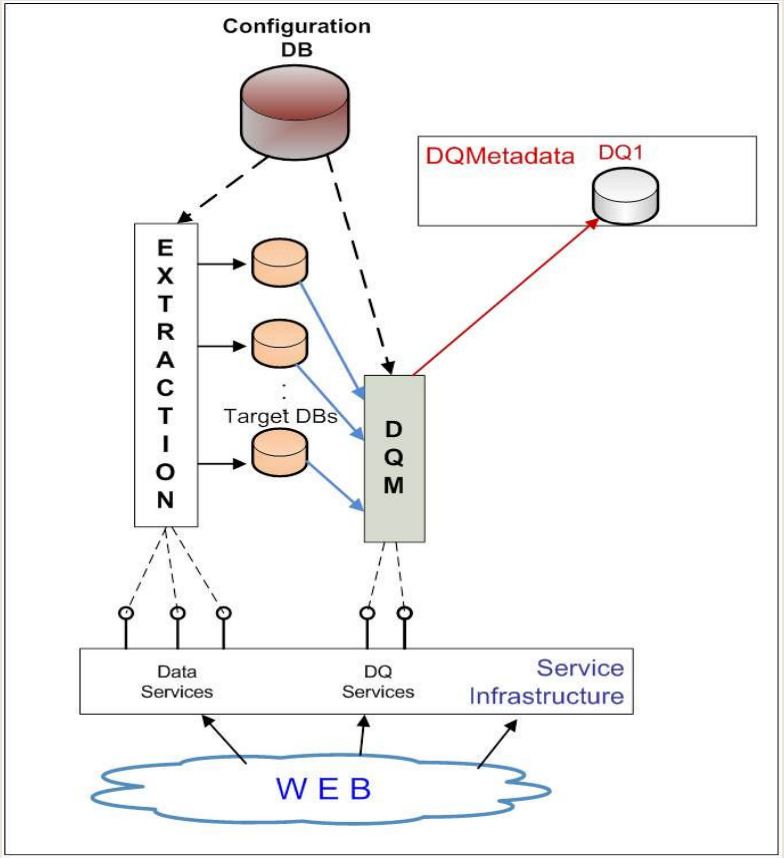
Automated support for the WW configuration

- Example of key tables from the configuration database



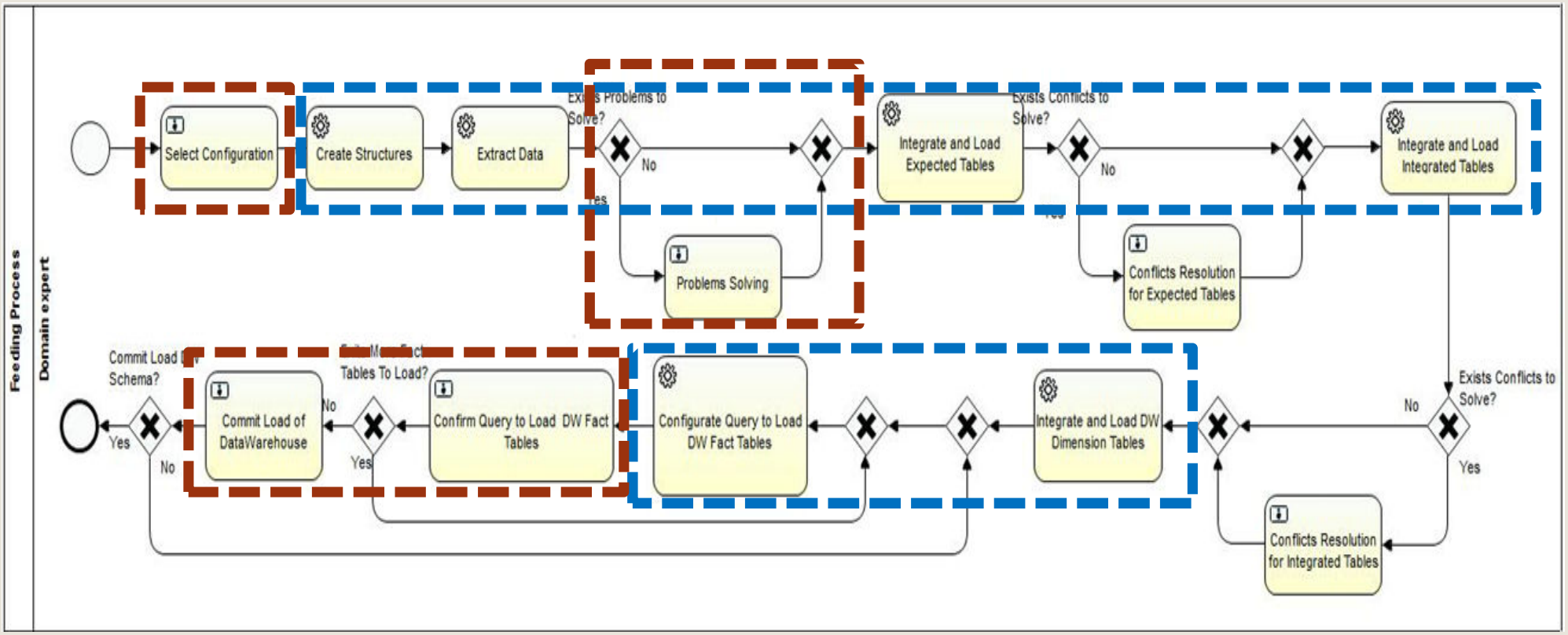
QAWW construction with BPMN 2.0

- Feeding Business Process definition
  - the activities corresponds mainly to the ETL traditional process of constructing a DW
  - it is automated as much as possible based on reading the data registered in the Configuration process
    - Extraction process using data from the configuration DB and executing DS for each Web source using defined quality information



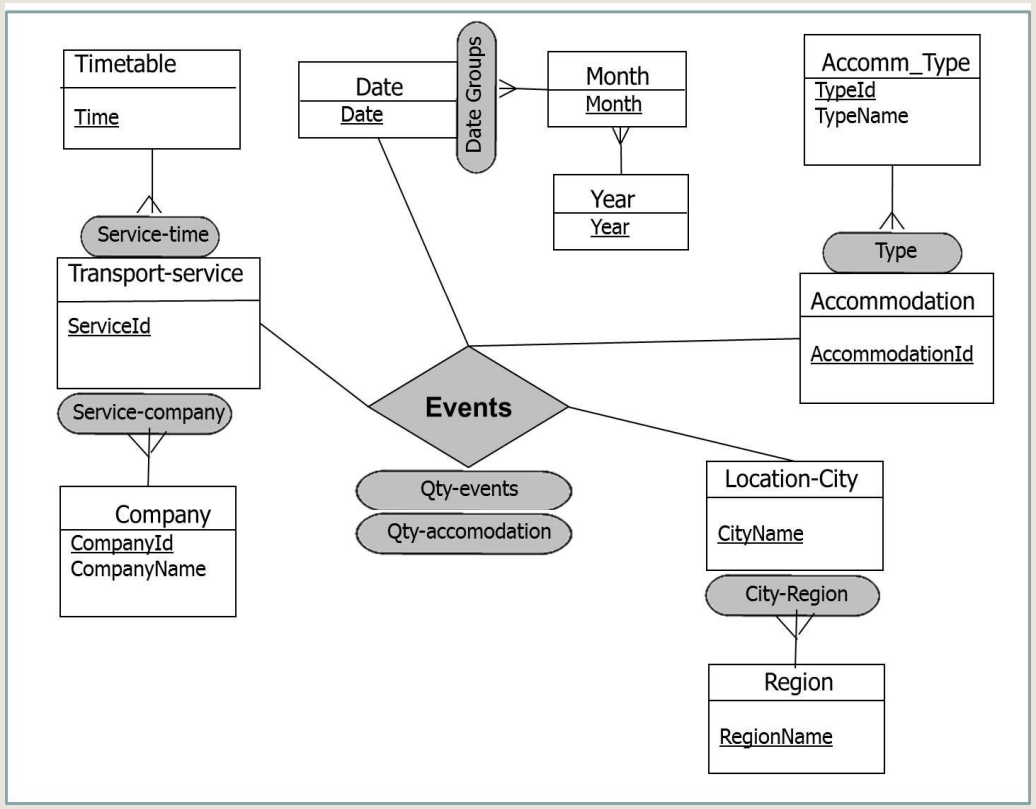
QAWW construction with BPMN 2.0

- Feeding Business Process initial definition
  - as a prototype to validate the configuration process

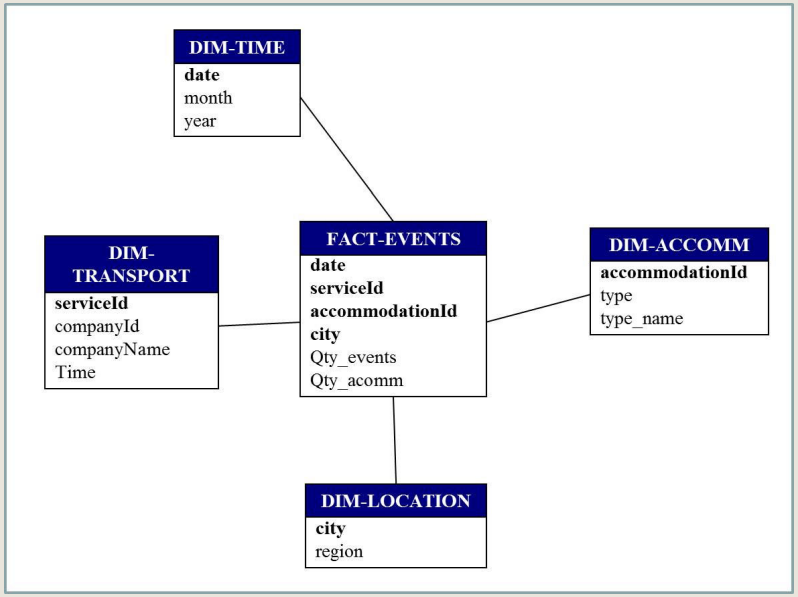


Case study with open data

- From the Tourism domain in Uruguay



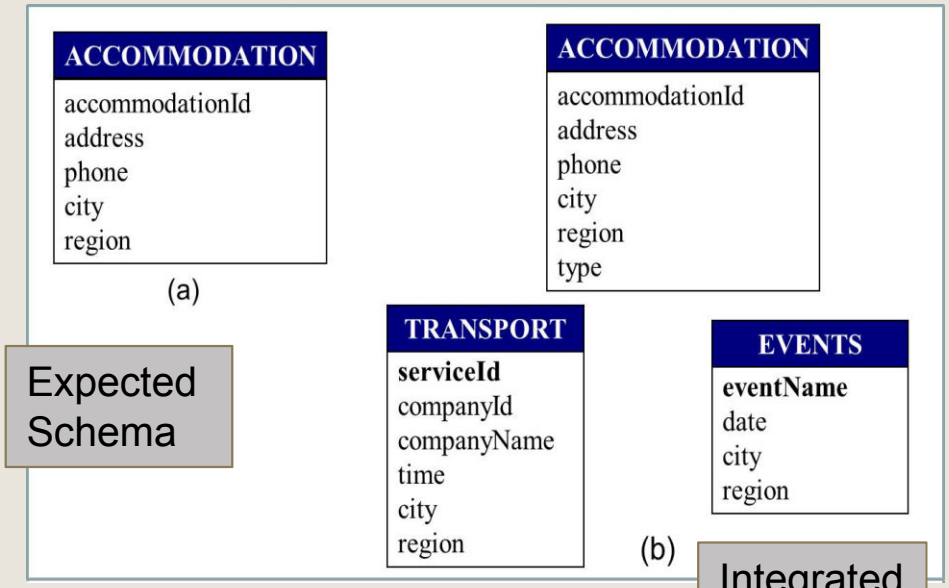
Conceptual multidimensional model
(Zimanyi and Malinowski, 2008)

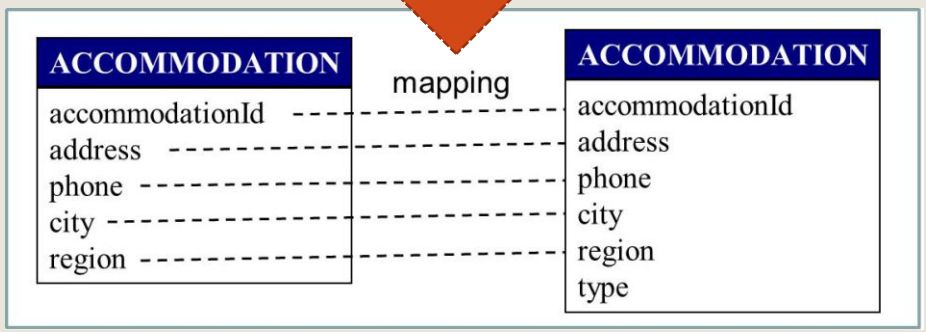


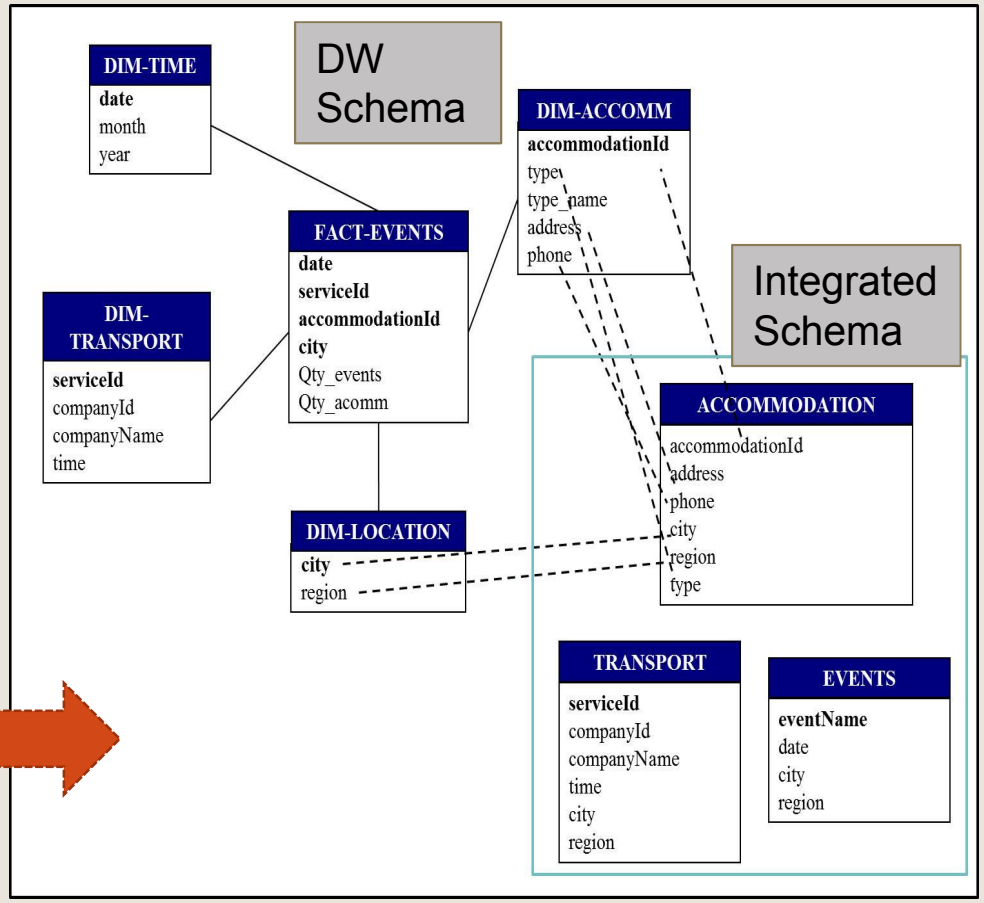
Logical DW Schema

Case study with open data

- Example of defined schemas and mappings

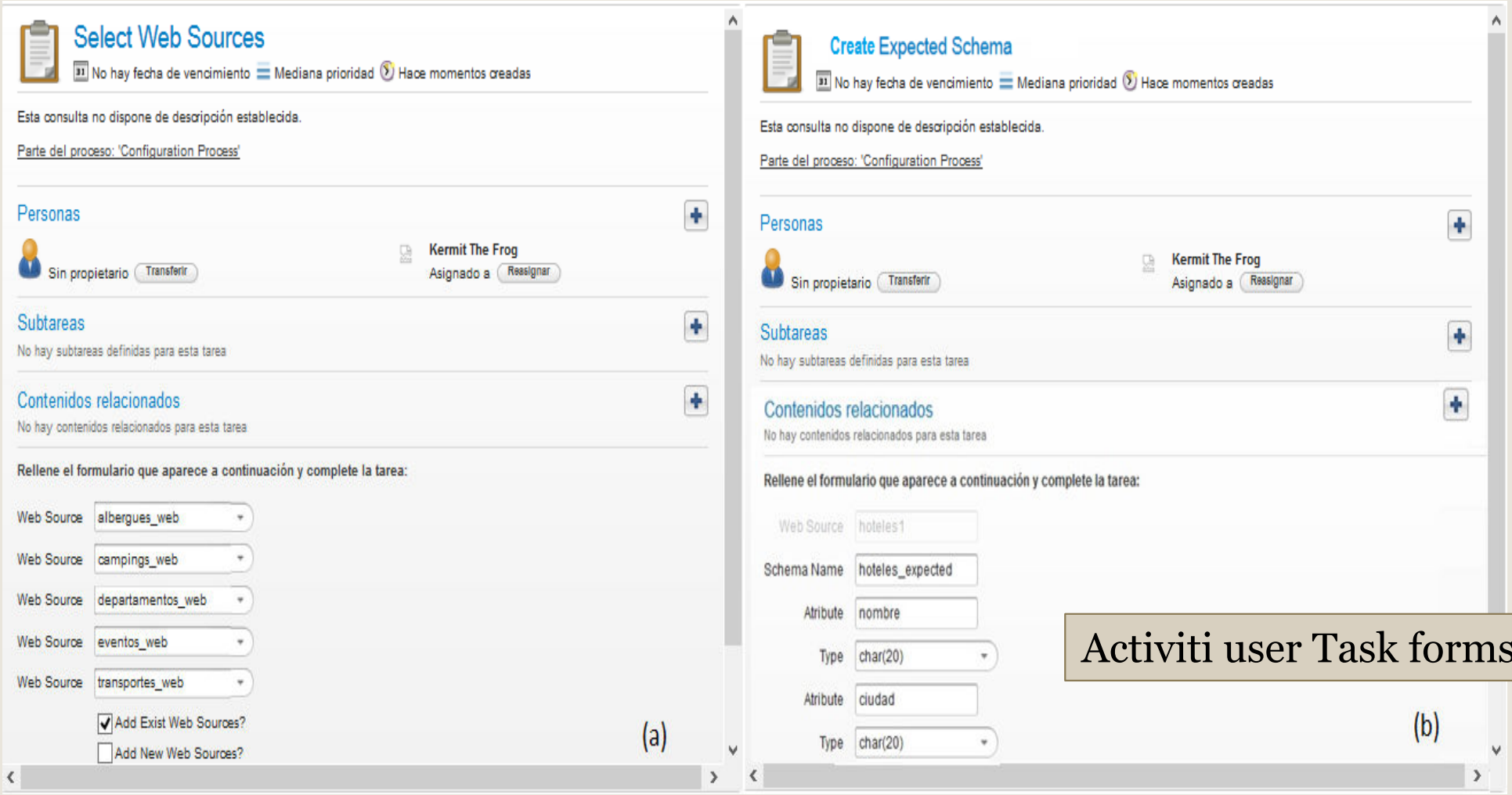






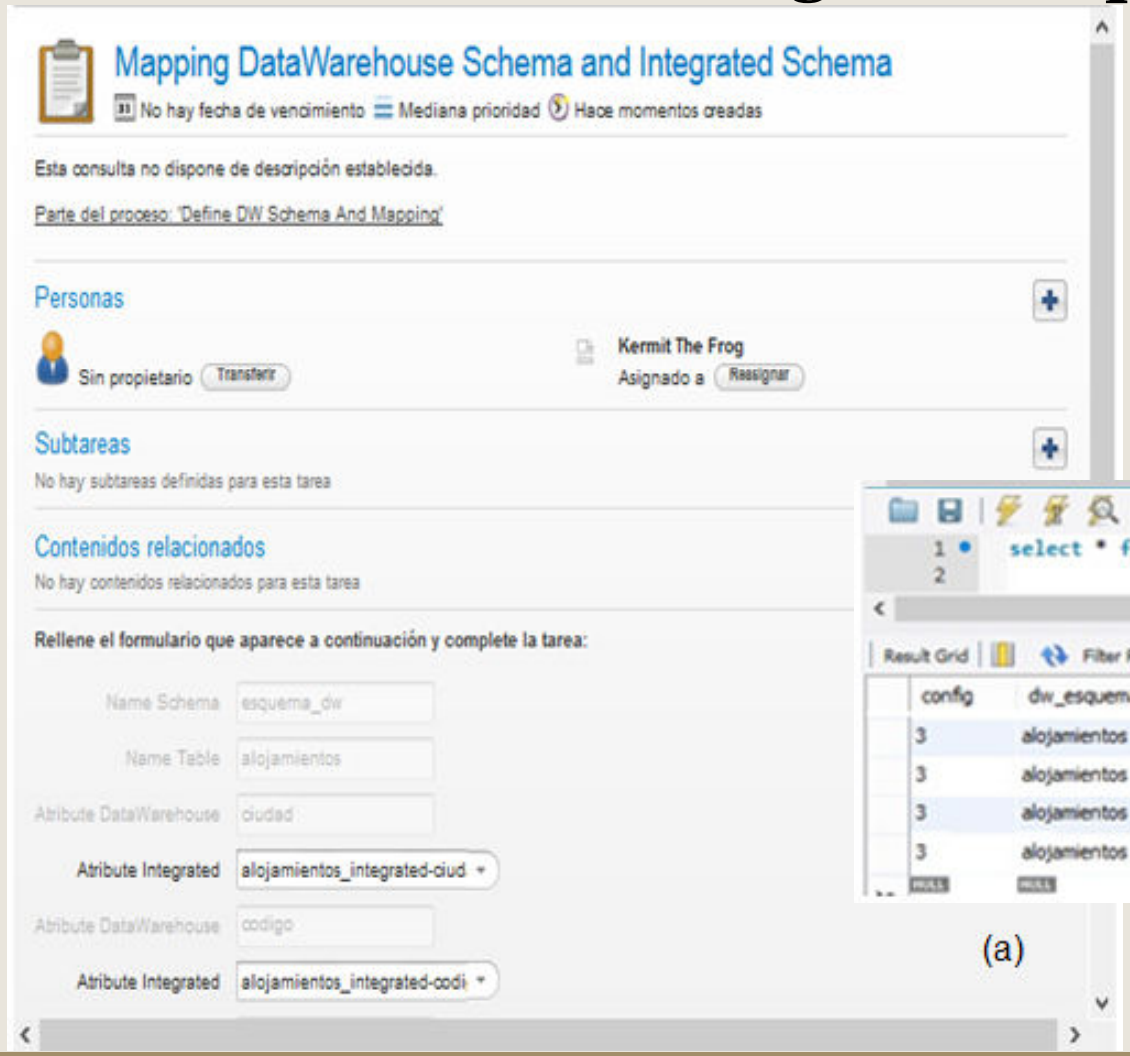
Case study with open data

- Execution of the configuration process

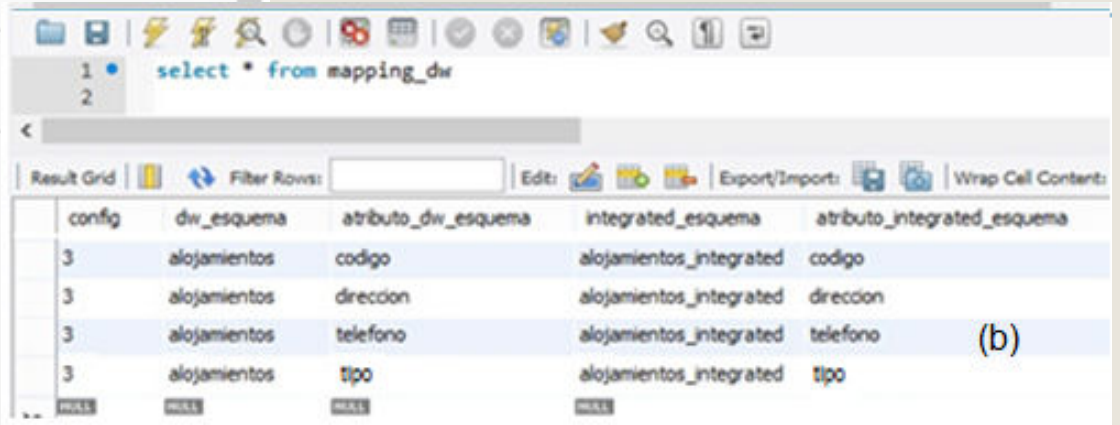


Case study with open data

- Execution of the configuration process

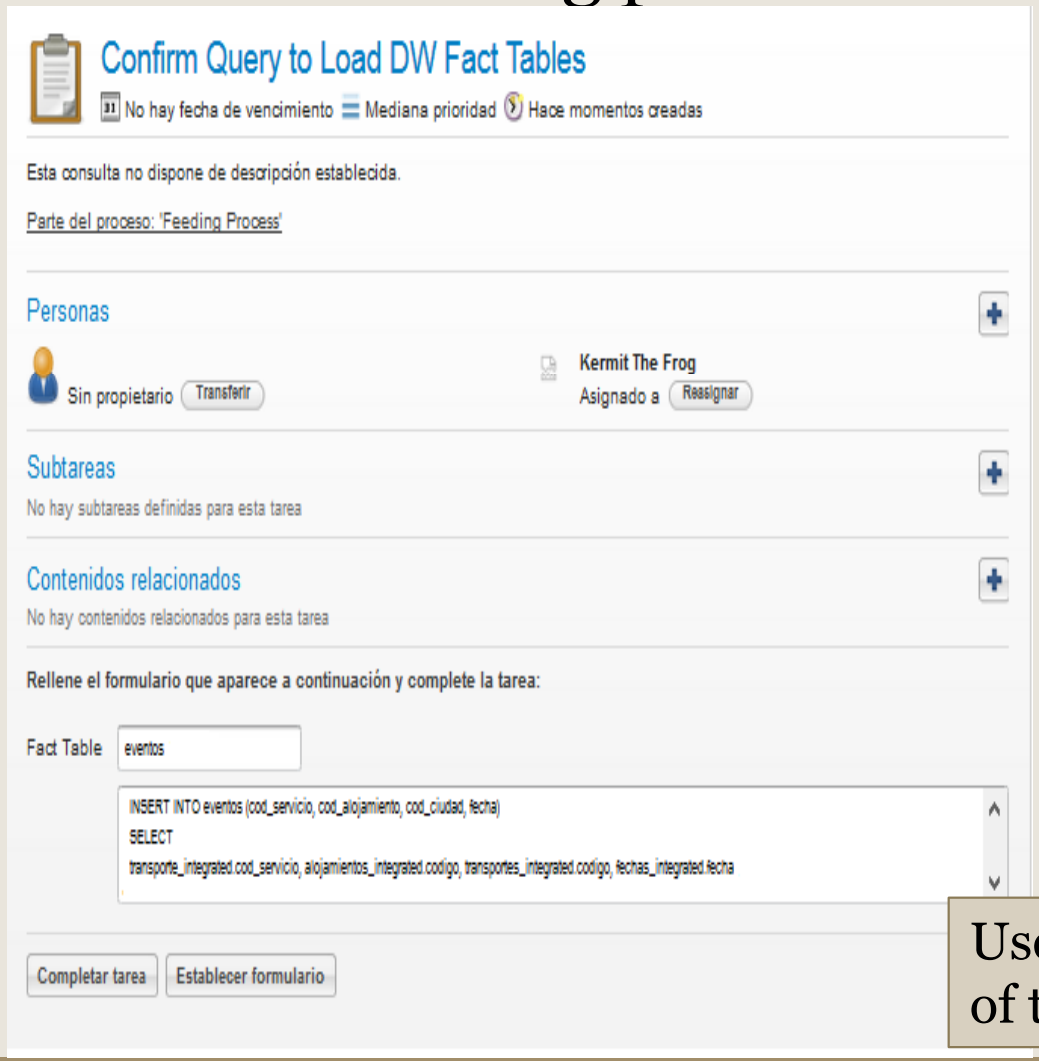


corresponding data in the "mapping_dw" table



Case study with open data

- Execution of the feeding process



User task for confirmation of the DW query

Conclusions and future work

- Our approach takes a two level Business Process vision to support the process of building flexible WW
  - a Configuration BP user oriented to define things
  - a Feeding BP to generate the WW using the data defined
- We have modeled, executed and validated the Configuration process using BPMN 2.0 and Activiti
  - Defining the user tasks needed to gather the data and the metadata database to register it
  - Implementing the model in the Activiti BPMS and validating the proposal in a case study with open data

Conclusions and future work (2)

- The automated support for building flexible WW we propose will provide organizations with many benefits
  - Bridging the gap between domain experts and IT professionals based on the BPMN 2.0 model
  - Providing easier to use guides to the user as the execution presents the tasks accordingly to the defined control flow
  - providing automated execution as much as possible, whenever something can be automated it will be automated
  - providing data about the definitions taken by the user i.e. who defined what in what task, automatically registered in the BPMS process engine as BPs are executed.
- As future work
  - We are adding management of quality information and defining the feeding process automating the integration of data

Thank you very much for your attention !

Questions ?

InCo, FING, UdelaR

http://www.fing.edu.uy/inco/

adelgado@fing.edu.uy