Tax and Revenue Service scenario for Ontology Matching

Stefano Brida[1], Marco Combetto, Silvano Frasson[2] and Paolo Giorgini[3]

[1] Trentino Riscossioni S.p.A., Trento, Italy
[2] Informatica Trentina S.p.A., Trento, Italy
[3] D.I.S.I., University of Trento, Italy

Abstract. In this paper we present a scenario for ontology matching posed by the Trentino Riscossioni S.p.A data integration system focusing the opportunity to enhance the level of data integration over a large set of Tax and Revenue industry-specific data sources.

Introduction. The mission of Trentino Riscossioni S.p.A[1], a company owned by the Autonomous Province of Trento, is to promote simplification processes and harmonize the activity of more than 250 public entities in the province, creating policies for fair taxation and for operating costs reduction. The need for consistent and contextual use of the heterogeneous information sources between its offices, the municipalities and the other public bodies is a fundamental requirement for the implementation of an accurate and balanced taxation system. In this paper we want to focus on the possibility offered by matching technology [1] to enhance the in the present day data integration architecture and increase its flexibility in managing hundreds of new data sources with reduced software development for each new sources added. Besides, even if the data integration has been extensively studied in the database community, according to some recent research works [2,3,4,5], the issue to improve the automatic schema matching in a data integration scenario for the Tax and Revenue market is a relative new ground of application. The contribution of this paper includes a specific scenario focusing several of the basic requirements that have to be considered in order to build a data integration system capable to support dynamically hundreds of data sources.

Scenario. The scenario is to make possible the insertion, management and deletion of new data sources (e.g., new data source from a new provincial database). The inclusion of a new data source would result in the census of syntactic and semantic information related to the attributes of the source and in an automatic mapping of these attributes over the proper attributes of the destination database schema. If the attributes are not present in the destination schema, the system must support the design of a schema extension. The source information is collected in a knowledge base. The search results will be available for at least 2 types of applications: (i) the business intelligence application that enables the monitoring, tracking and management of the data quality [6] of the integrated database and four (ii) mission-critical applications focusing specific business-strategic tasks: assessment revenue, territory mapping, planning support, final users services. As depicted hereafter in Figure 1, the information coming from the external data sources is processed through the SSMB (Semantic Schema/data Matching Box). The SSMB must be able to calculate the new system status $n + 1$ through a function based on the previous states

---

[1] http://www.trentinoriscossionispa.it

(*n*, *n-1*) in order to support a GUI tool that will provide the interface to the required information to the Information Engineer and to the calculated matching suggestions enabling to integrate the sources more rapidly than currently. There are about 10 different data sources for each municipality and 7-8 for each provincial data source. In the next 2 years, the plan is to integrate about 200 municipalities and other significant sources.
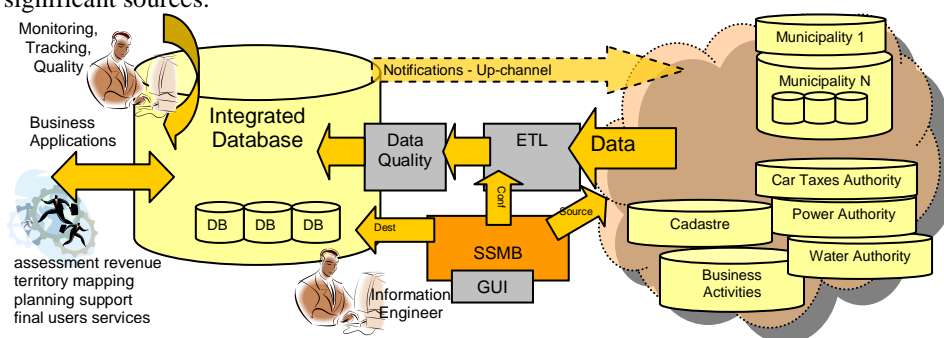

Figure 1 – The scenario description

The process analysis and breakdown provides confidence to motivate an implementation based on the use of a schema matching workbench like the HARMONY[7] integration workbench. In fact, beside the other advantages this approach enables the interoperation and the selection between different and various prototypes and commercial tools for schema matching and enables the sharing a common knowledge repository.

Conclusions and future works. We presented the business scenario for a solution that leverages on matching technology in order to scale-out over hundreds of data sources. Future works proceed in the following directions: (i) formalization of the scenario, (ii) evaluation and test of the HARMONY workbench features, and (iii) development of a specific working prototype for Trentino Riscossioni S.p.A.

References

[1] J. Euzenat and P. Shvaiko. Ontology matching. Springer, 2007.

[2] A. Bitterer, M. A. Beyer, and T. Friedman. Magic Quadrant for Data Integration Tools. Gartner, 2008.

[3] P. Shvaiko and J. Euzenat. Ten Challenges for Ontology Matching. In Proc. of ODBASE, 2008.

[4] K. Smith, P. Mork, L. Seligman, A. Rosenthal, M. Morse, C. Wolf, D. Allen, M. Li: The Role of Schema Matching in Large Enterprises. In Proc. of CIDR, 2009.

[5] Y. Asnar, P. Giorgini, P. Ciancarini, R. Moretti, M. Sebastianis, N. Zannone. Evaluation of Business Solutions in Manufacturing Enterprises. In International Journal on Business Intelligence and Data Mining, Inderscience, 2008

[6] Jeffery G. Watson. Data Quality Essentials. Uni. of Wisconsin-Madison, 2006

[7] P. Mork, L. Seligman, A. Rosenthal, J. Korb, and C. Wolf. The Harmony Integration Workbench. Journal on Data Semantics, 2008.