Quality-driven Information Integration Evolution

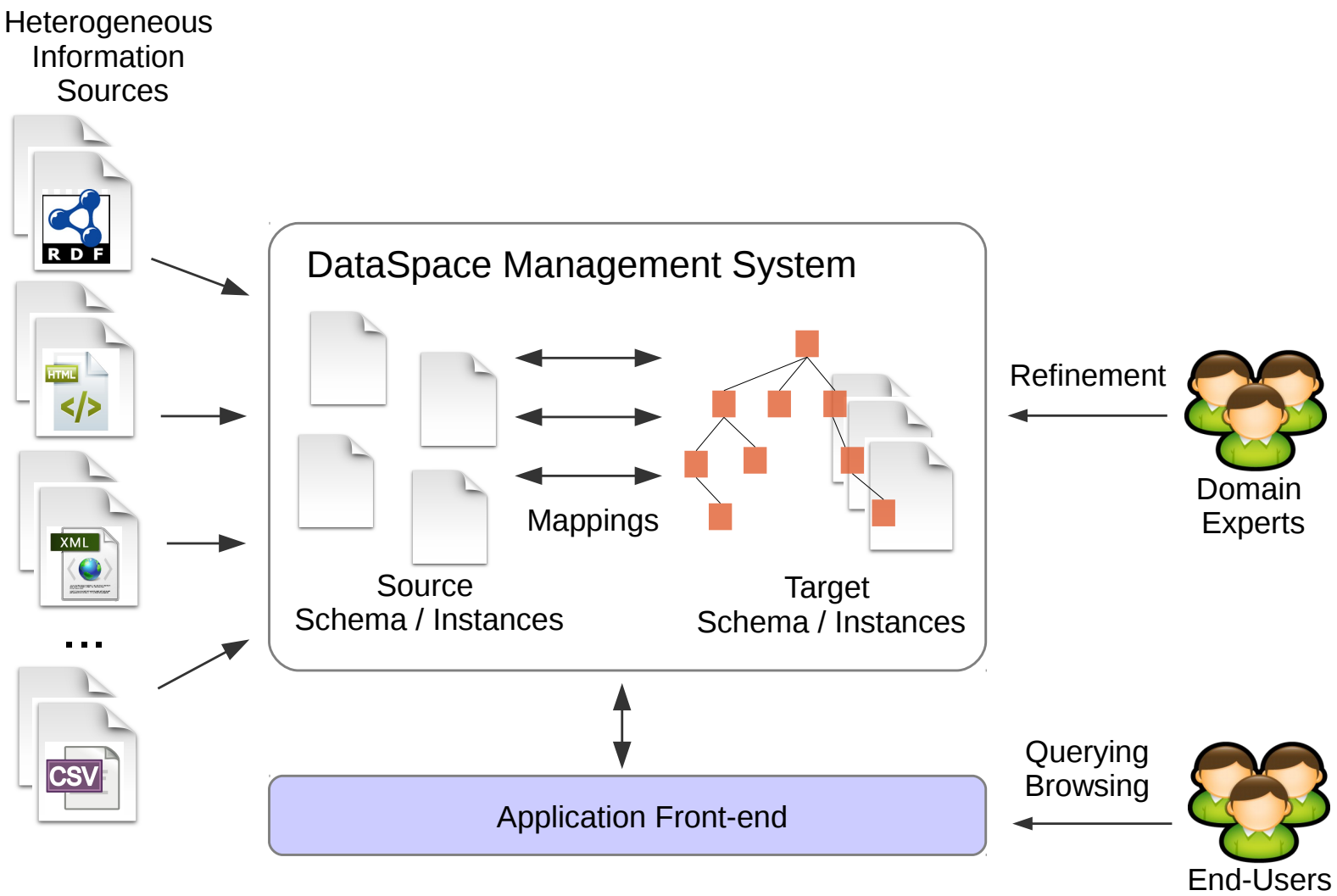
Ph.D. in Computer Science Course - Series XXVIII

Riccardo Porrini

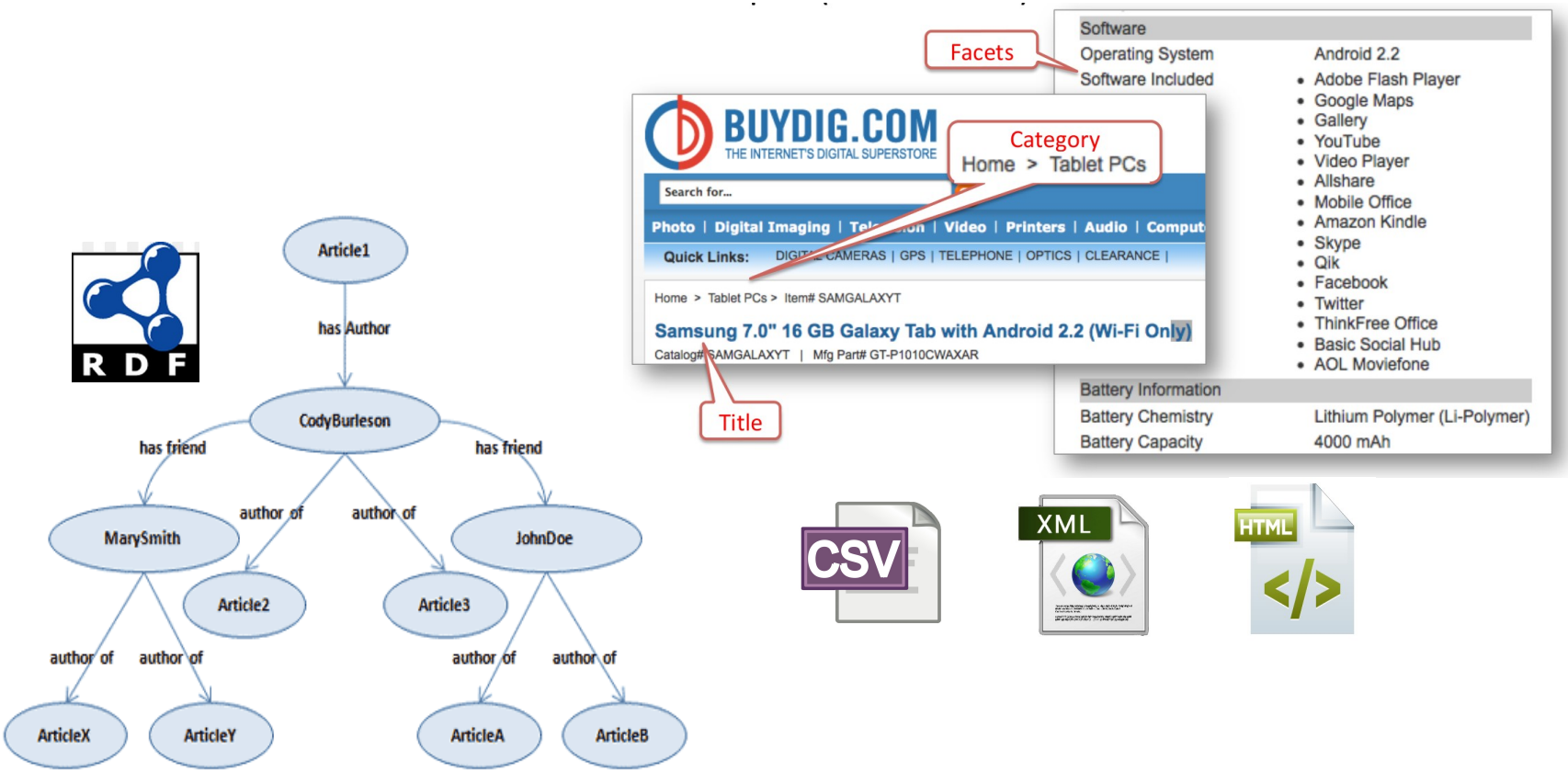
Supervisor: Dott. Palmonari

Tutor: Prof. Messina

Web Information Integration

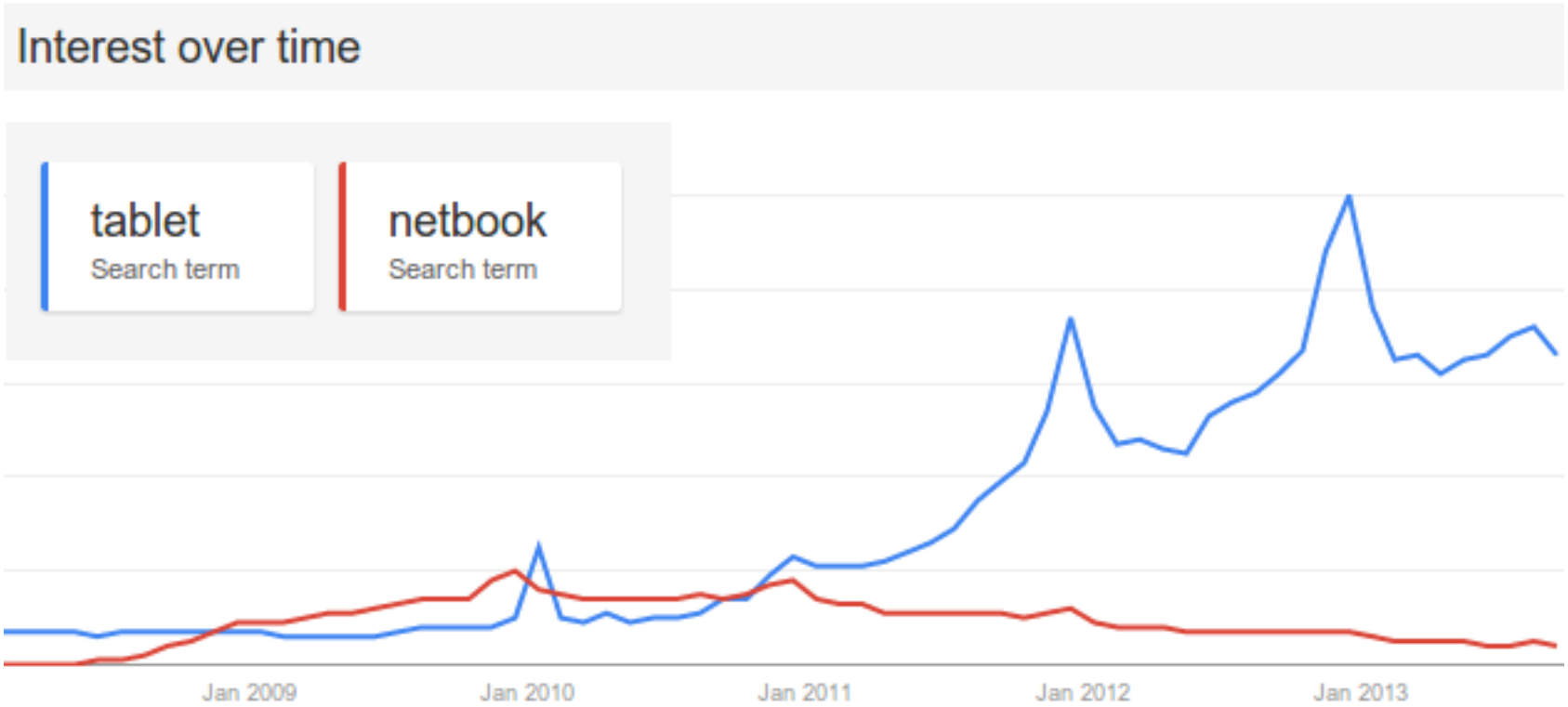


Schema and Instances Changing Over Time



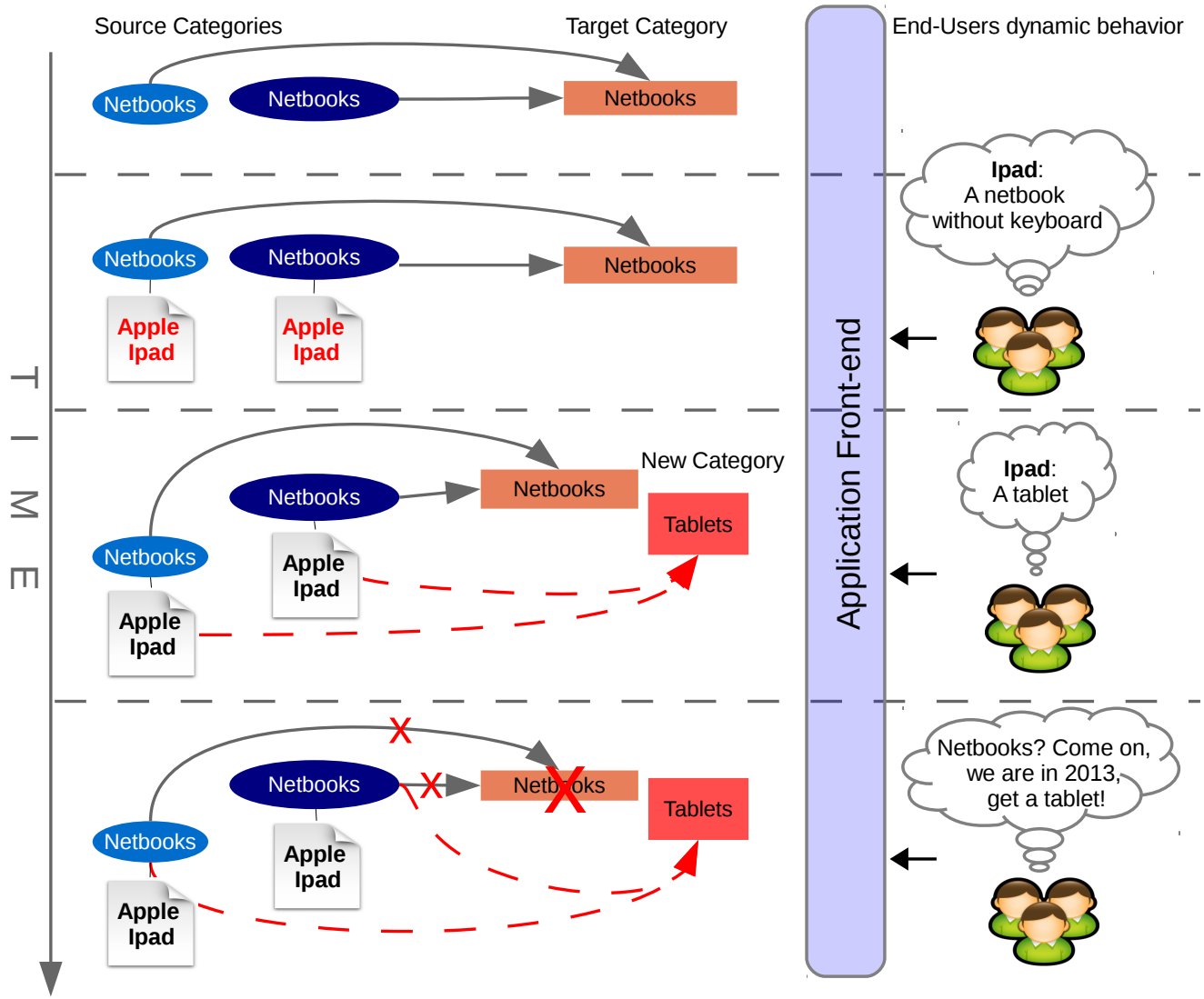
Web Dynamicity

End-Users' Information Needs Changing Over Time

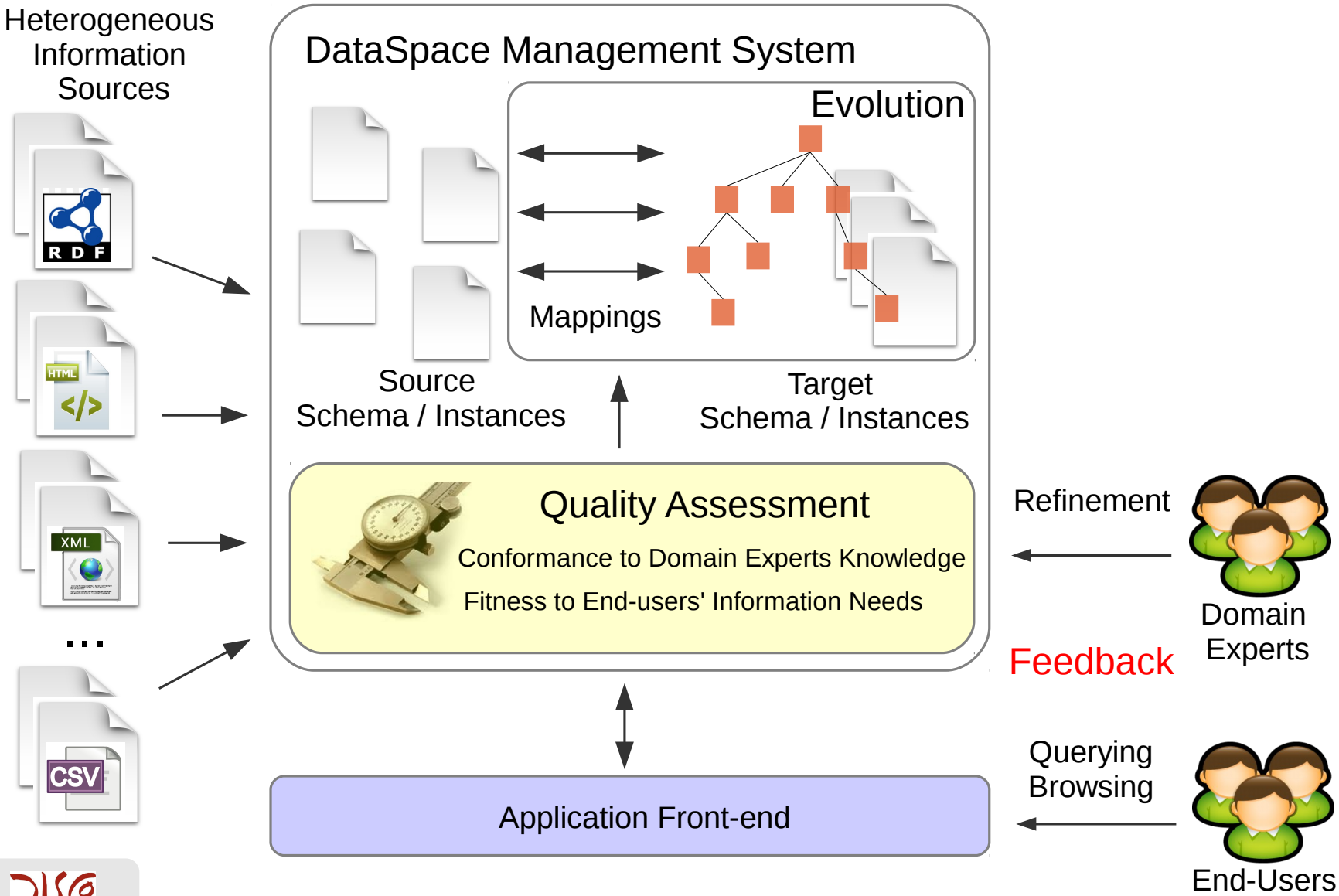


Example in the eCommerce Domain

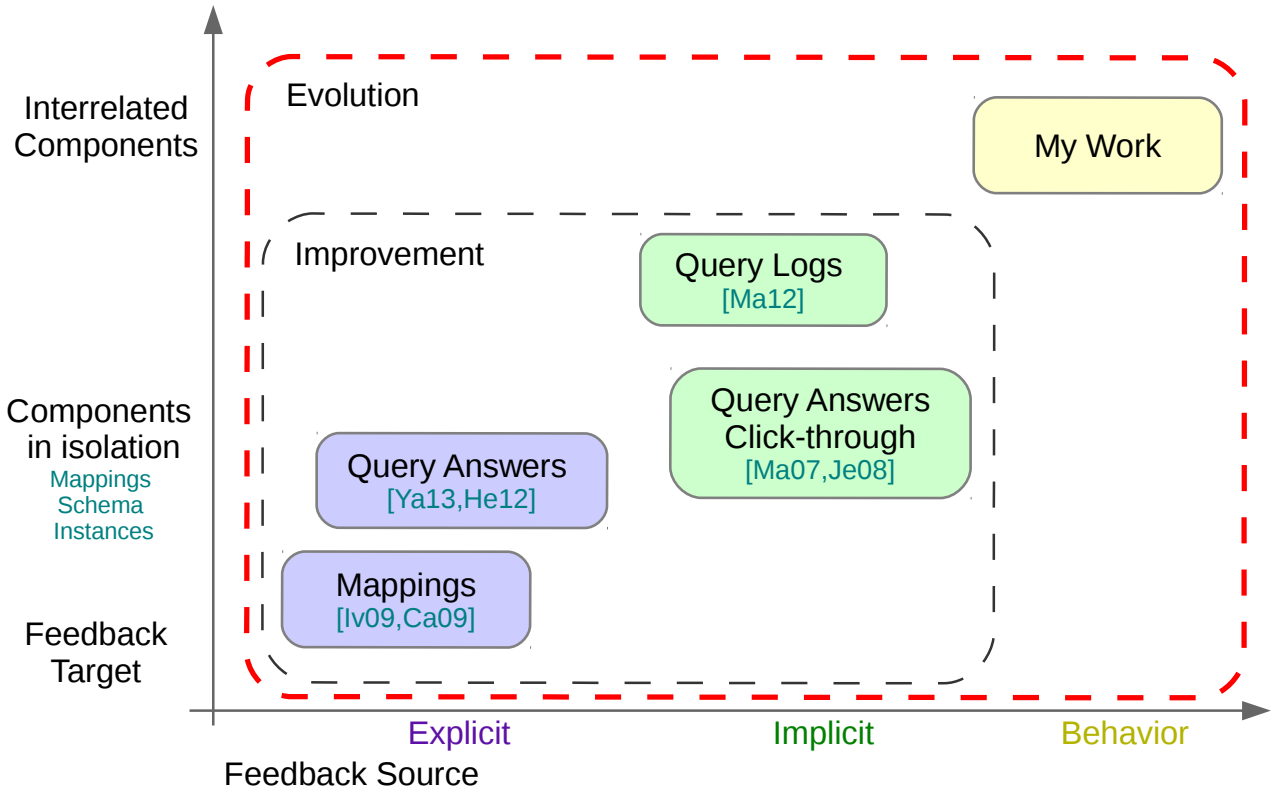
Taxonomy Integration and Evolution



Quality-driven Information Integration Evolution



Related Works



[Ma12] Maskat et al. Pay-as-you-go ranking of schema mappings using query logs. In *DILS*, 2012

[Ma07] Madhavan et al. Web-scale data integration: You can afford to pay as you go. In *CIDR*, 2007

[Je08] Jeffery et al. Pay-as-you-go user feedback for dataspace systems. In *SIGMOD*, 2008

[Iv09] Ives et al. Interactive data integration through smart copy & paste. In *CIDR*, 2009

[Ca09] Cafarella et al. Data integration for the relational web. In *PVLDB*, 2009

[Ya13] Yan et al. Actively soliciting feedback for query answers in keyword search-based data integration. In *PVLDB*, 2013

[He12] Hedeler et al. Dstoolkit: An architecture for flexible dataspace management. In *T-LSD-KCS*, 2012

Goals

Hypothesis

Quality defined in terms of **fitness** to end-users information needs and **conformance** to domain experts knowledge can be used to evolve integration over time

Goals

Hypothesis

Quality defined in terms of **fitness** to end-users information needs and **conformance** to domain experts knowledge can be used to evolve integration over time

Goal 1

Model the **quality** of a DataSpace Management System considering **end-user behavior**

Goals

Hypothesis

Quality defined in terms of **fitness** to end-users information needs and **conformance** to domain experts knowledge can be used to evolve integration over time

Goal 1

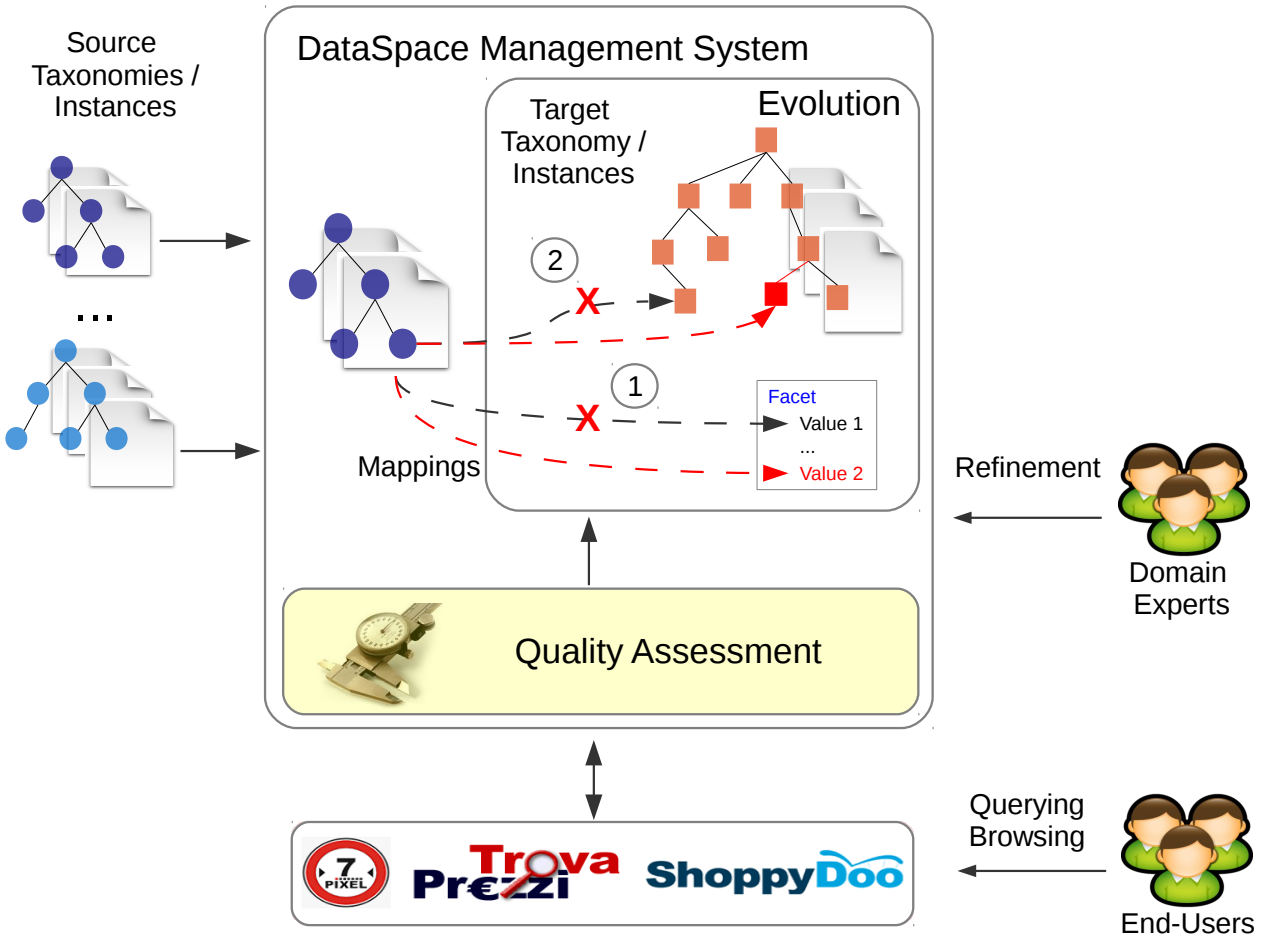
Model the **quality** of a DataSpace Management System considering **end-user behavior**

Goal 2

Model mappings evolution considering **continuous quality evaluation**

Case Studies from Price Comparison Engines

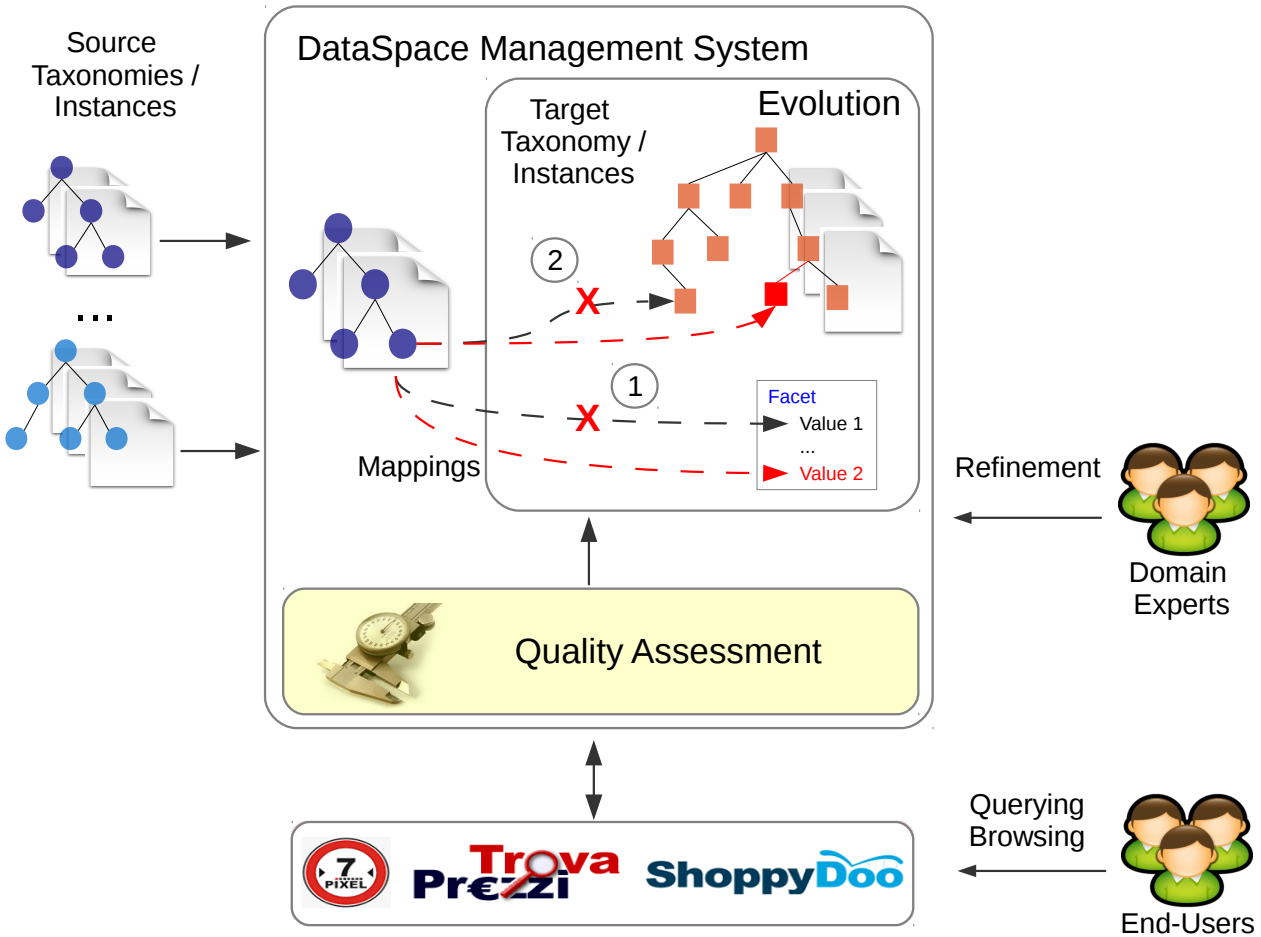
Case Study ① Adaptive Mappings and Facets Management
Case Study ② Adaptive Mappings and Taxonomy Refinement



- **8M** product offers
- **2000** sources
- **500** target categories
- **10M** users/month

Case Studies from Price Comparison Engines

Case Study ① Adaptive Mappings and Facets Management
Case Study ② Adaptive Mappings and Taxonomy Refinement

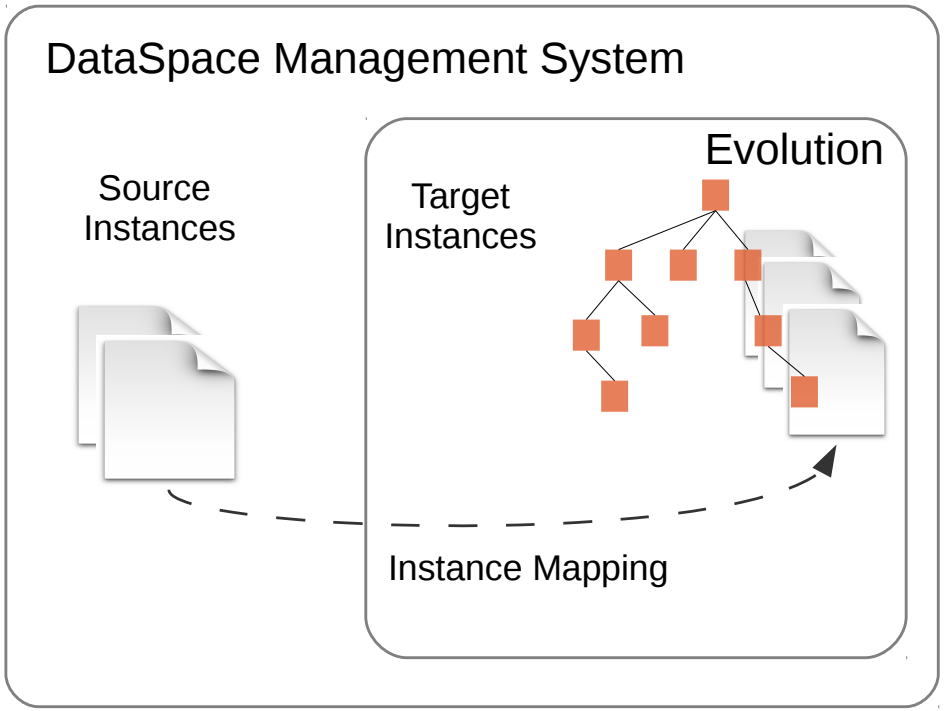


- **8M** product offers
- **2000** sources
- **500** target categories
- **10M** users/month

- Traffic Data
- Domain Experts
- In-vivo experiments

Preliminary Work

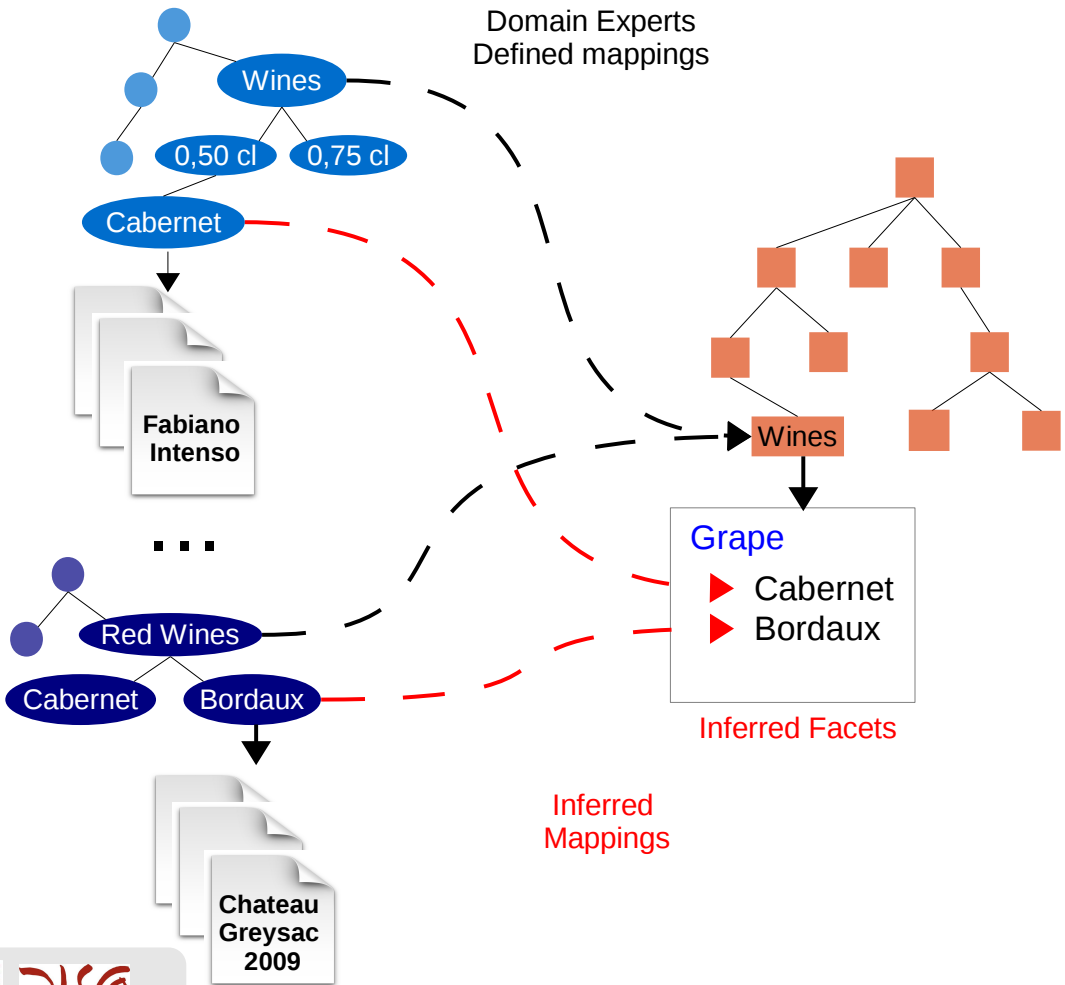
Lightweight matching algorithm for instance mapping tasks [1]



[1] Porrini et al. COMMA: A Result-Oriented Composite Autocompletion Method for E-marketplaces. In *Web Intelligence*, 2012 *(extended work to be published on Web Intelligence and Agent Systems Journal)*

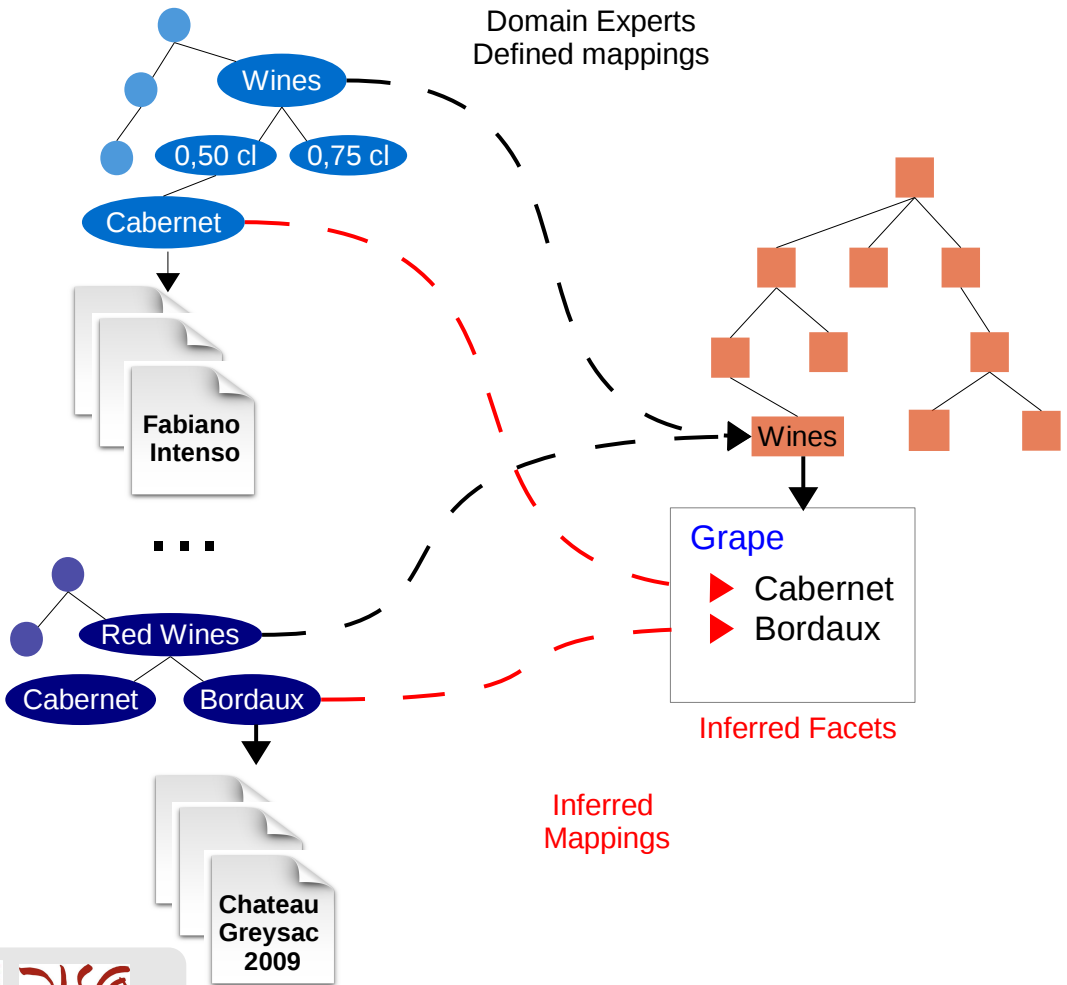
Ongoing Work

Case Study ① Adaptive Mappings and Facets Management



Ongoing Work

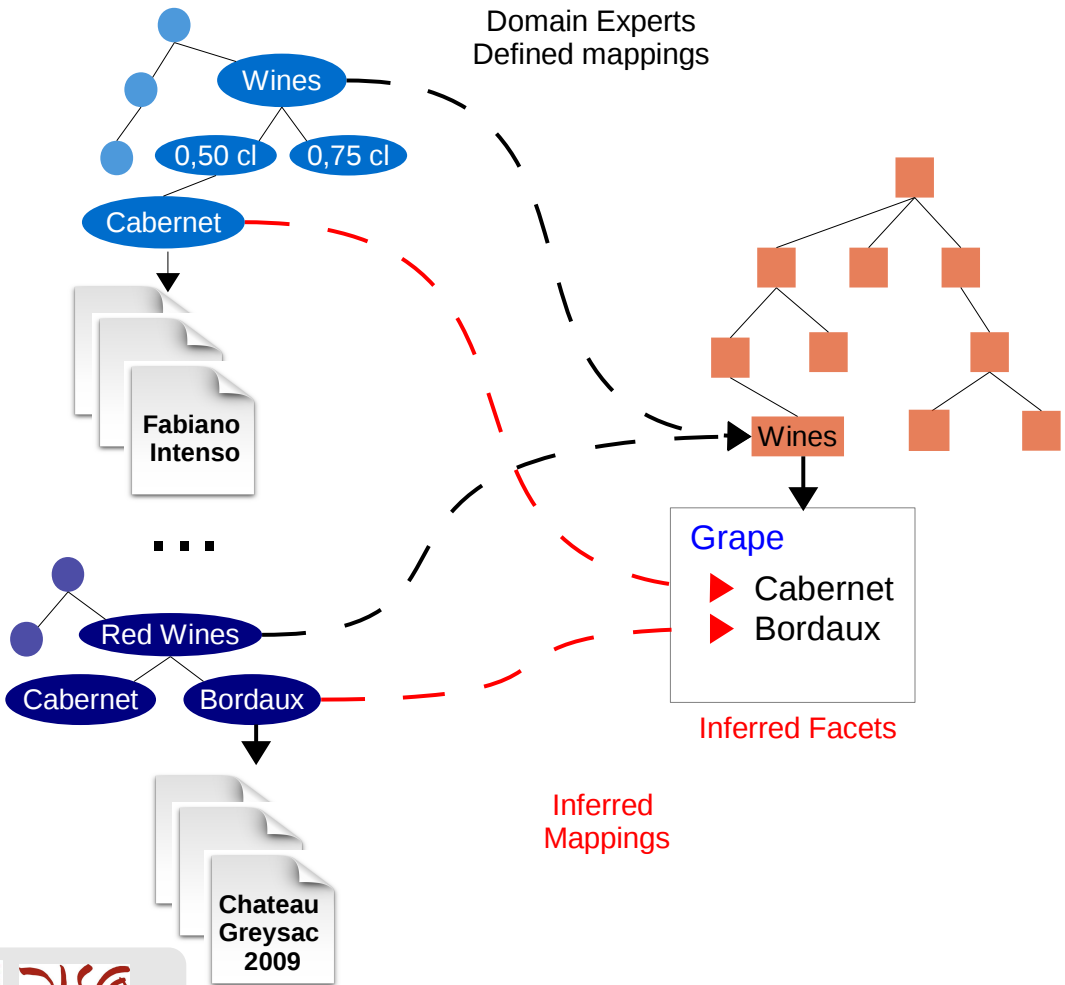
Case Study ① Adaptive Mappings and Facets Management



Deployed on TrovaPrezzi

Ongoing Work

Case Study ① Adaptive Mappings and Facets Management



Currently collecting end-users usage statistics

Next Steps

Case Study (1) Adaptive Mappings and Facets Management

- Qualitative analysis of usage statistics
- Include usage statistics analysis into extraction and evolution phase

Next Steps

Case Study (2) Adaptive Mappings and Taxonomy Refinement

- Model mappings and taxonomy management operations
- Include temporal information on mapping definition
- Model user behavior
- Link mappings and taxonomy management model to user behavior model

Past Activities

Publications

[1] Palmonari et al. Comma: A Result-Oriented Composite Autocompletion Method for e-Marketplaces. In *Web Intelligence*, 2012

[2] Porrini et al. Composite Match Autocompletion (COMMA): a Semantic Result-Oriented Autocompletion Technique for e-Marketplaces. In *Web Intelligence and Agent Systems Journal*, accepted for publication

Attended Courses and Schools

- Advanced Analytics and Behavior Informatics - DISCo
- Foundations of Data Exchange and Integration - Politecnico di Milano
- Third ESWC Summer School - Kalamaky - Crete (GR)

Attended Workshops and Seminars

- WOA 2012: 13^{th} National Workshop "Dagli Oggetti agli Agenti" - DISCo
- Big Data e la forza degli eventi - DISCo
- Semantic Constraints for Data Quality Assessment and Cleaning - DISCo
- Progettare e Fare Open Data. Metodologia e tools sviluppati in Evodevo a partire dallesperienza Open Data INPS - DISCo
- Phase Transitions in Social and Economic Systems - DISCo
- Optimum Hyperpaths in Directed Hypergraphs - DISCo

Teaching

- Lecturer (1 lecture) for the *"Artificial Intelligence"* course (MSc, 2^{nd} year) - DISCo
- Tutor for the *"Distributed Systems"* course (BSc, 2^{nd} year) - DISCo
- Co-Advisor of a student BSc thesis

Thank You

Questions?