TRUSTED SLA-GUIDED DATA INTEGRATION ON MULTI-CLOUD ENVIRONMENTS
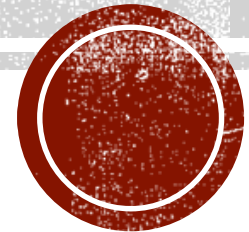
Daniel Aguiar da Silva Carvalho, Magellan, IAE, Université Jean Moulin Lyon3

ADVISORS

Chirine Ghedira Guegan, Magellan, IAE, Université Jean Moulin Lyon3

Genoveva Vargas-Solar, CNRS, LIG-LAFMIA, France
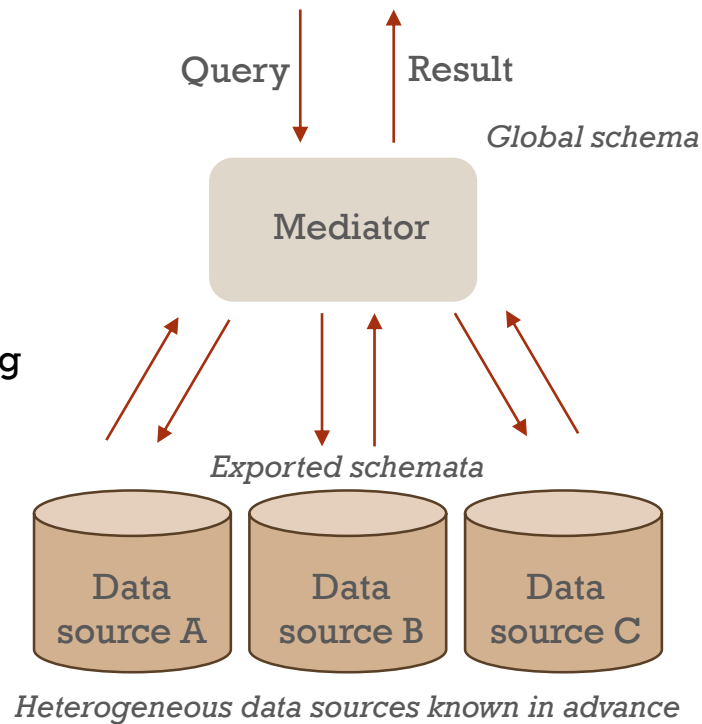
Nadia Bennani, CNRS INSA-Lyon, LIRIS, UMR5205 - France

DATA INTEGRATION: EXISTING WORK

Data integration: the teenage years. Halevy, A., Rajaraman, A., & Ordille, J. (VLDB 2006, September)

Query rewriting
MiniCon algorithm for query rewriting (Pottinger and Halevy, 2001)

Query | Result

Global schema

Mediator

Exported schemata

Data source A | Data source B | Data source C

Heterogeneous data sources known in advance

Schema integration: Past, present, and future (Ram, S., & Ramesh, V. 1999)

Data integration architectures:
Multi-databases, federations, DW, …
*(Domenig & Dittrich 1999 Sigmod Record)*

DATA INTEGRATION: EXISTING WORK

Services lookup and matching

[1] Paolucci, M., Kawamura, T., Payne, T. R., & Sycara, K. (2002, June). Semantic matching of web services capabilities. In International Semantic Web Conference (pp. 333-347). Springer Berlin Heidelberg.

[2} Bramantoro, A., Krishnaswamy, S., & Indrawan, M. (2005, November). A semantic distance measure for matching web services. In International Conference on Web Information Systems Engineering (pp. 217-226). Springer Berlin Heidelberg.

[3} APA   Maximilien, E. M., & Singh, M. P. (2004, November). Toward autonomic web services trust and selection. In Proceedings of the 2nd international conference on Service oriented computing (pp. 212-221). ACM.

Query rewriting techniques *adapted* to *service composition*

[4] Barhamgi, M., Benslimane, D., and Medjahed, B. (2010). A query rewriting approach for web service composition. *IEEE T. Services Computing*, 3(3):206–222.

[5] da Costa, U. S., Alves, M. H. F., Musicante, M. A., and Robert, S. (2013). Automatic refinement of service compositions. In Daniel, F., Dolog, P., and Li, Q., editors, ICWE, volume 7977 of Lecture Notes in Com- puter Science, pages 400–407. Springer.

[6] Zhao, W., Liu, C., and Chen, J. (2011). Automatic composition of information-providing web services based on query rewriting. Science China Information Sciences, pages 1–17.

Query   Result

Mediator

Exported API

Distributed data services
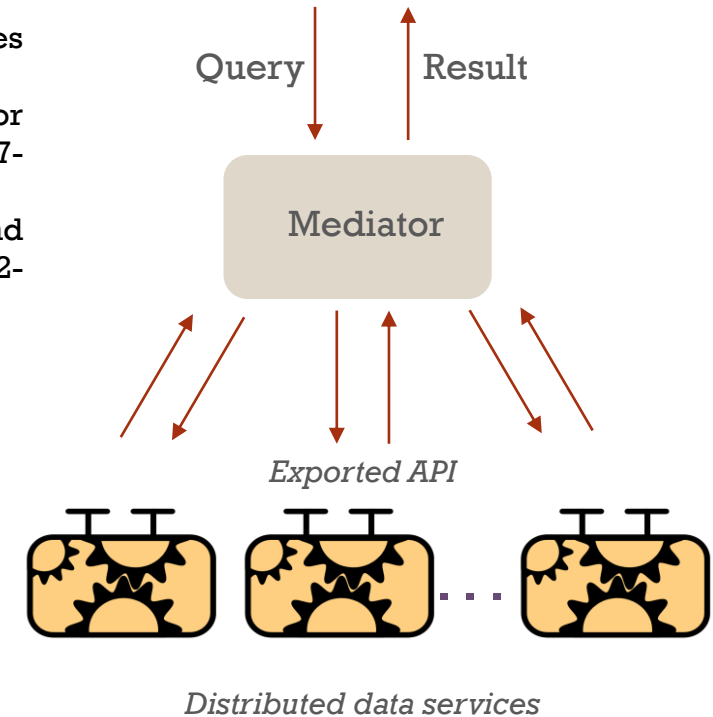
CHALLENGES

- Which services should I select ? Are the requirements being respected?

- How to be sure that all SLAs are being respected?

- How to integrate different SLAs associated to services involved with user's requirements?

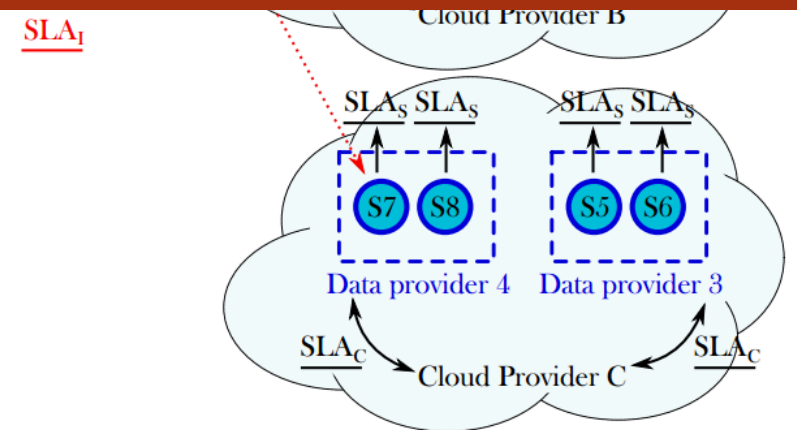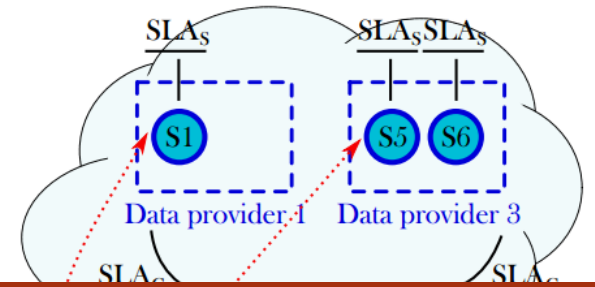- How results can be reused for a next query?
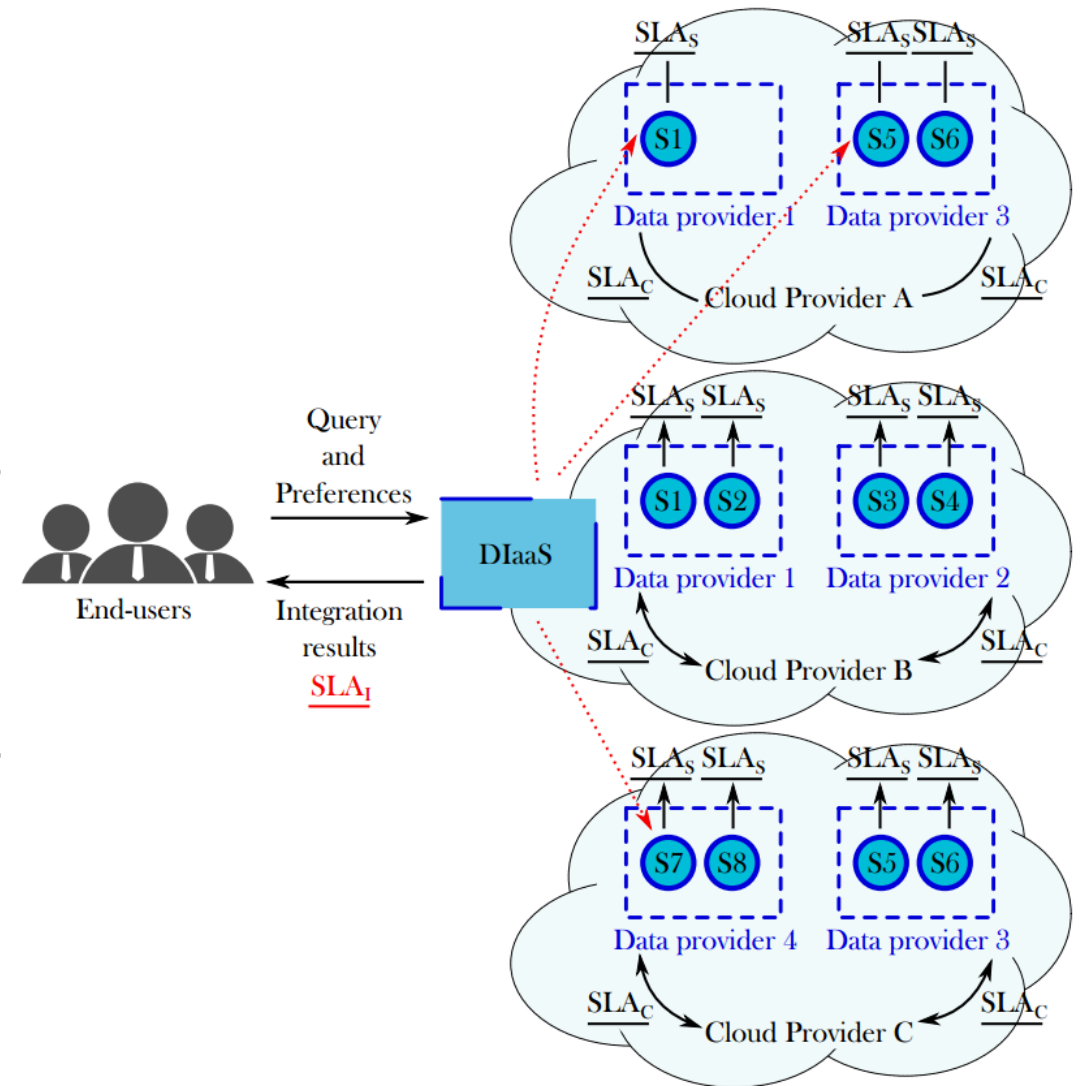
CHALLENGES

- Which services should I select ? Are the

Propose a data integration solution in a multi-cloud environment guided by user requirements and SLA exported by different clouds

- How to integrate different SLAs associated to services involved with user's requirements?

- How results can be reused for a next query?

MAIN CONTRIBUTIONS

- Systematic mapping[1]
  - build the corpus of the state of the art and identify new trends and open issues around our research topic.

- Rhone query rewriting algorithm[2] for data integration
  - Considers user integration preferences and services' quality aspects expressed in SLAs.

[1] D. A. S. Carvalho, P. A. Souza Neto, G. Vargas-Solar, N. Bennani, C. Ghedira, **Can Data Integration Quality be Enhanced on Multi-cloud using SLA?**, In 26th Int. Conf. on Database and Expert Systems Applications, Spain, 2015.

[2] D. A. S. Carvalho, P. A. S. Neto, C. Ghedira, G. Vargas-Solar, N. Bennani. **Rhone: a quality-based query rewriting algorithm for data integration**. East-European Conference on Advances in Databases and Information Systems, Aug 2016, Prague, France. ADBIS East-European Conference on Advances in Databases and Information Systems, 2016.
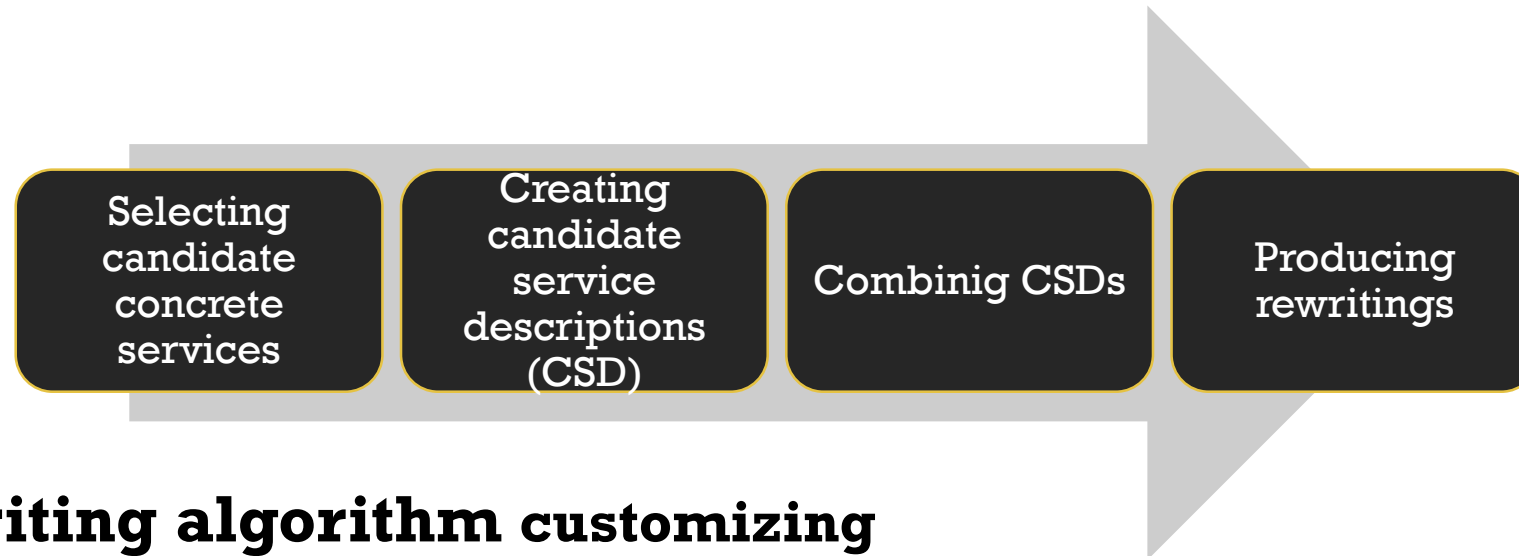
APPROACH

Address data integration considering
data quality (freshness, provenance, cost, availability) properties &
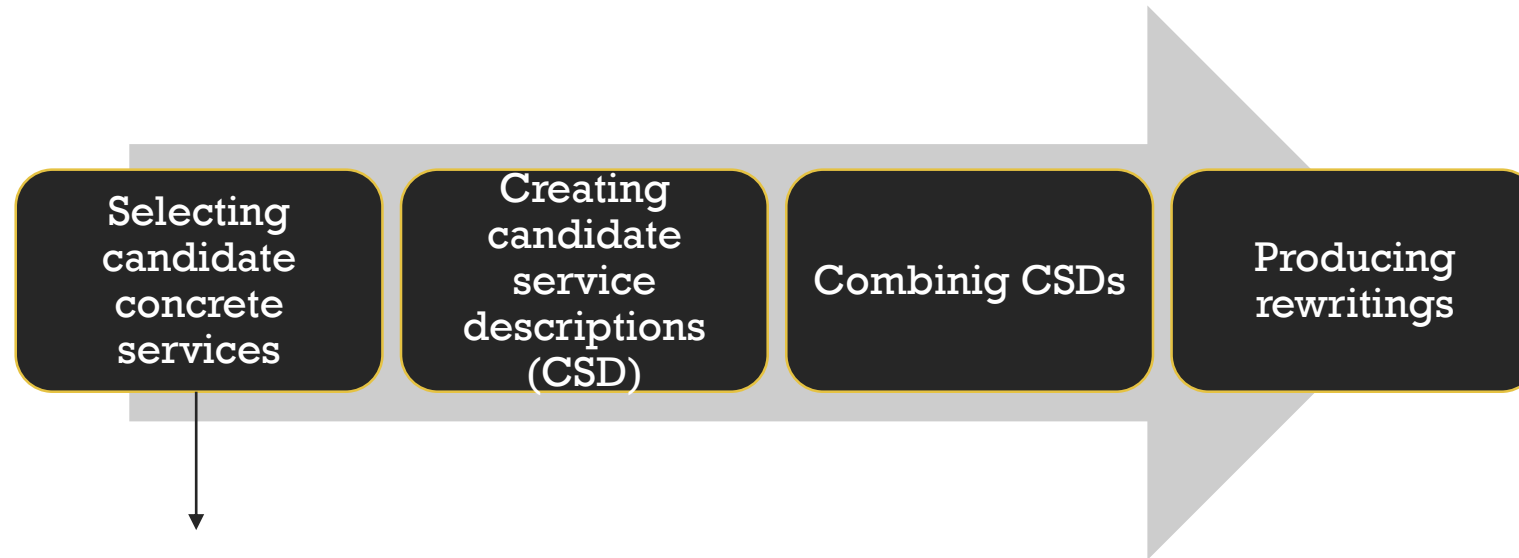**service level agreements (SLA)**

- **Hypothesis**
  - the data integration process is totally or partially externalized on different clouds that provide necessary resources under different conditions (SLA)
  - data can be retrieved from several data providers (i.e., services) with different quality properties

RHONE SERVICE-BASED QUERY REWRITING ALGORITHM

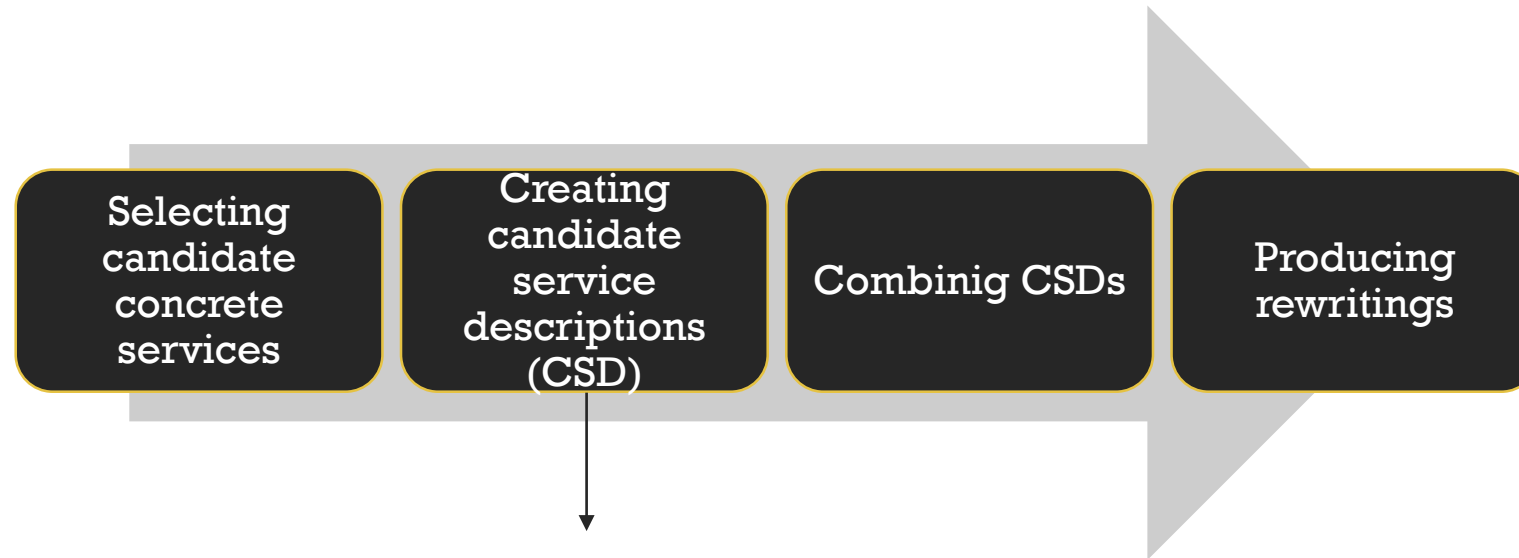| Selecting candidate concrete services | Creating candidate service descriptions (CSD) | Combinig CSDs | Producing rewritings |

- **A rewriting algorithm** customizing
  - data providers (services) **look up**
  - data integration considering different data consumers requirements and expectations
  - requirements & expectations depend on the context in which they consume data (e.g., mobile devices with few physical capacities, critical decision making)

RHONE SERVICE-BASED QUERY REWRITING ALGORITHM

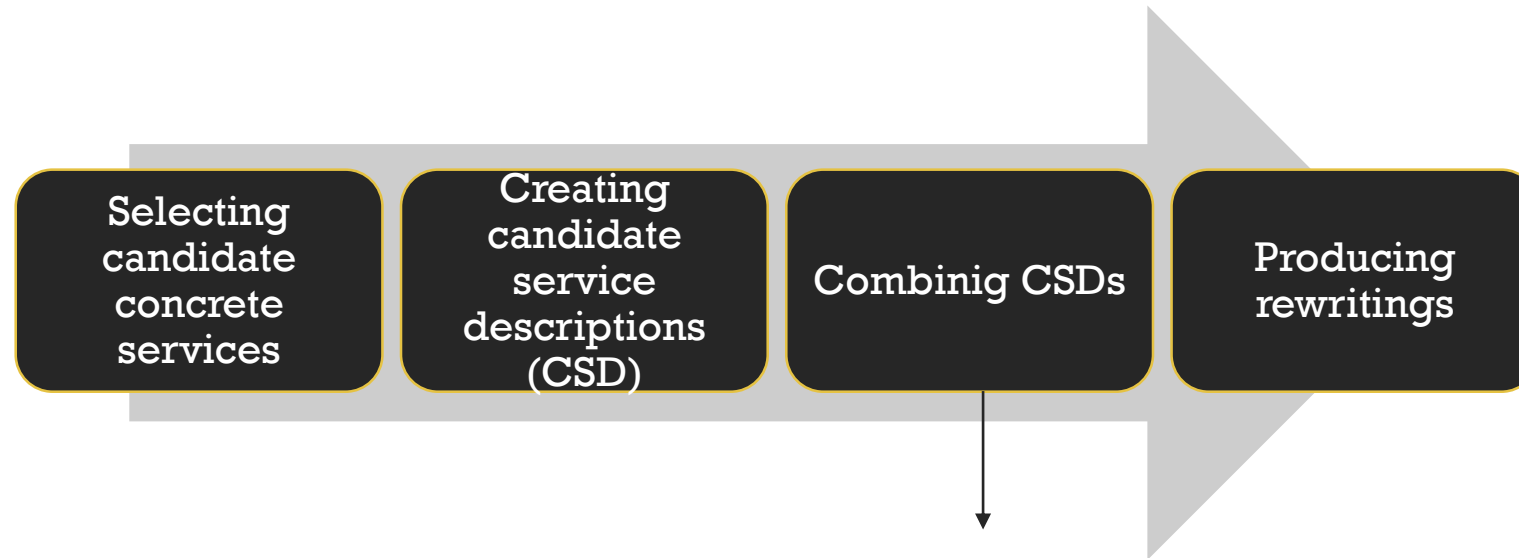| Selecting candidate concrete services	Creating candidate service descriptions (CSD)	Combinig CSDs	Producing rewritings

- Services are selected considering their characteristics (expressed in the SLAs)
- Services that can produce results that are useless to the user query are discarded in the first step

RHONE SERVICE-BASED QUERY REWRITING ALGORITHM

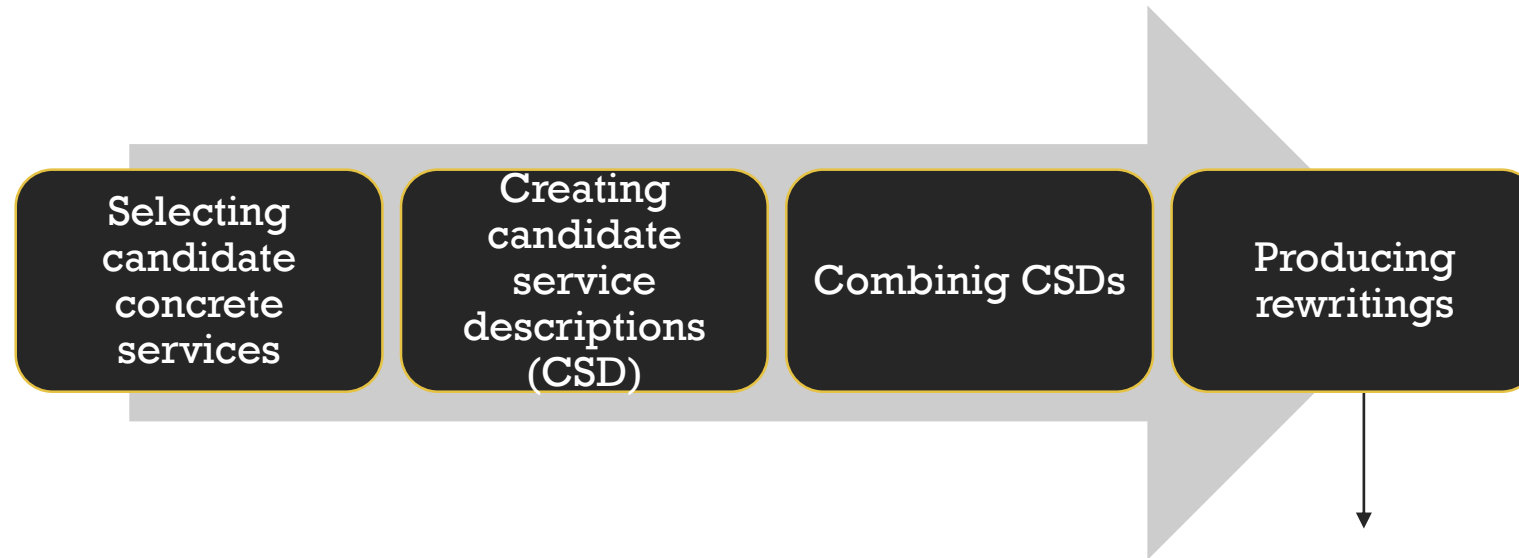| Selecting candidate concrete services | Creating candidate service descriptions (CSD) | Combinig CSDs | Producing rewritings |

▪ Differently from the other algorithms, the mapping are created considering the concrete service definition and not each abstract service that compose it

RHONE SERVICE-BASED QUERY REWRITING ALGORITHM

Selecting candidate concrete services	Creating candidate service descriptions (CSD)	Combinig CSDs	Producing rewritings

- Combinations are produced according to the part of the query that a given concrete service covers like the combinations in the Bucket algorithm

RHONE SERVICE-BASED QUERY REWRITING ALGORITHM

| Selecting candidate concrete services | Creating candidate service descriptions (CSD) | Combinig CSDs | Producing rewritings |

- Differently from the other approaches, the rewritings are produced considering the user preferences and constraints, and the SLAs exported by the different data services.
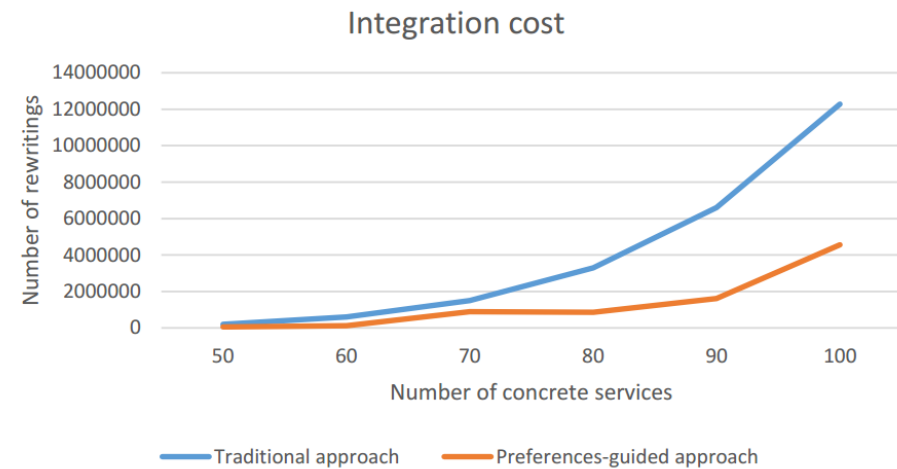
EXPERIMENTATION: LESSONS LEARNED

- **Complexity**
  - Selecting candidate concrete services: $O(n^2)$
  - Creating candidate service descriptions: $O(n^3)$
  - Combining CSDs: $O(n^k)^m$
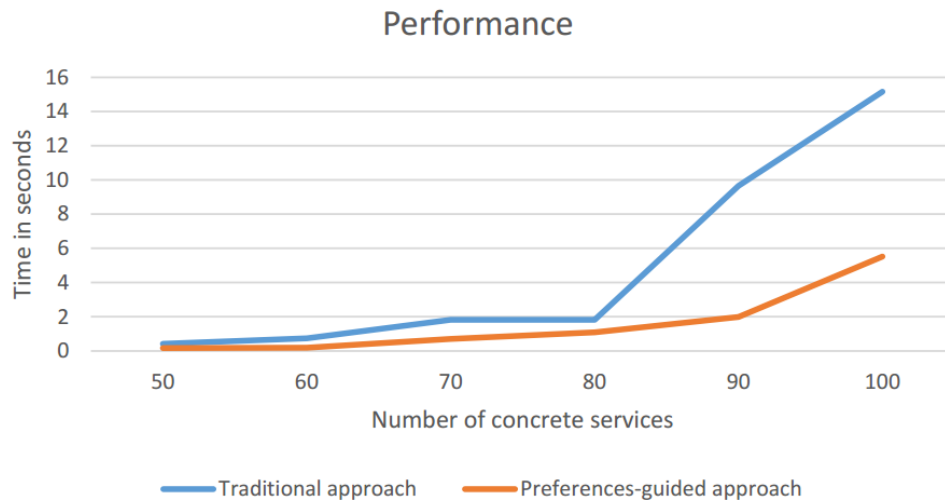
- **Performance** increased reducing
- The number of rewriting solutions
- Integration execution time

- **Rewriting solutions quality** enhanced

Performance

Time in seconds vs Number of concrete services
Traditional approach — Preferences-guided approach

Integration cost

Number of rewritings vs Number of concrete services
Traditional approach — Preferences-guided approach

ONGOING WORK

- Data integration metamodel[1]: a metaprocess and process for data integration adapted to the multi-cloud context

- Cloud SLA, service SLA and integration SLA models

- Reduce overhead caused by quey rewriting
  - Taxonomy of query variations for promoting the reusability of rewriting results
  - Heuristics for optmizing the rewriting approach adapted to the multi-cloud context

- Evaluation of the overall data integration approach adapted to the multi-cloud context

[1] D. A. S. Carvalho, P. A. S. Neto, C. Ghedira, G. Vargas-Solar, N. Bennani. **Towards Quality Guided Data Integration on Multi-Cloud Settings**. 14th international conference on service oriented computing (ICSOC), Oct 2016, Banff, Alberta, Canada.

Daniel Aguiar da Silva Carvalho, Magellan, IAE, Université Jean Moulin Lyon3

Advisors:

Chirine Ghedira Guegan, Magellan, IAE, Université Jean Moulin Lyon3

Genoveva Vargas-Solar, CNRS, LIG-LAFMIA, France

Nadia Bennani, CNRS INSA-Lyon, LIRIS, UMR5205 - France