# Increasing life science resources re-usability using Semantic Web technologies

Marine LOUARN\*†, Fabrice CHATONNET†, Xavier GARNIER\*, Thierry FEST†, Anne SIEGEL\* and Olivier DAMERON\*

\* Dyliss Team, Univ Rennes, INRIA, CNRS, IRISA, Rennes, FRANCE † U1236, INSERM, Univ Rennes, CHU Rennes, EFS, Rennes, FRANCE

#### Introduction

There is a need for standardization and integration in life science, to increase reusability and reproducibility. Currently, data-set are not published in standard format and pipeline are poorly documented. We hypothesize that Semantic Web technologies are efficient for facilitating the integration and reuse of reference data and knowledge bases.

Method: Integration of data with Semantic technologies and formalization of analysis pipeline as **SPARQL** queries.

Case study: Regulatory Circuits (RC) datasets and pipeline [1]

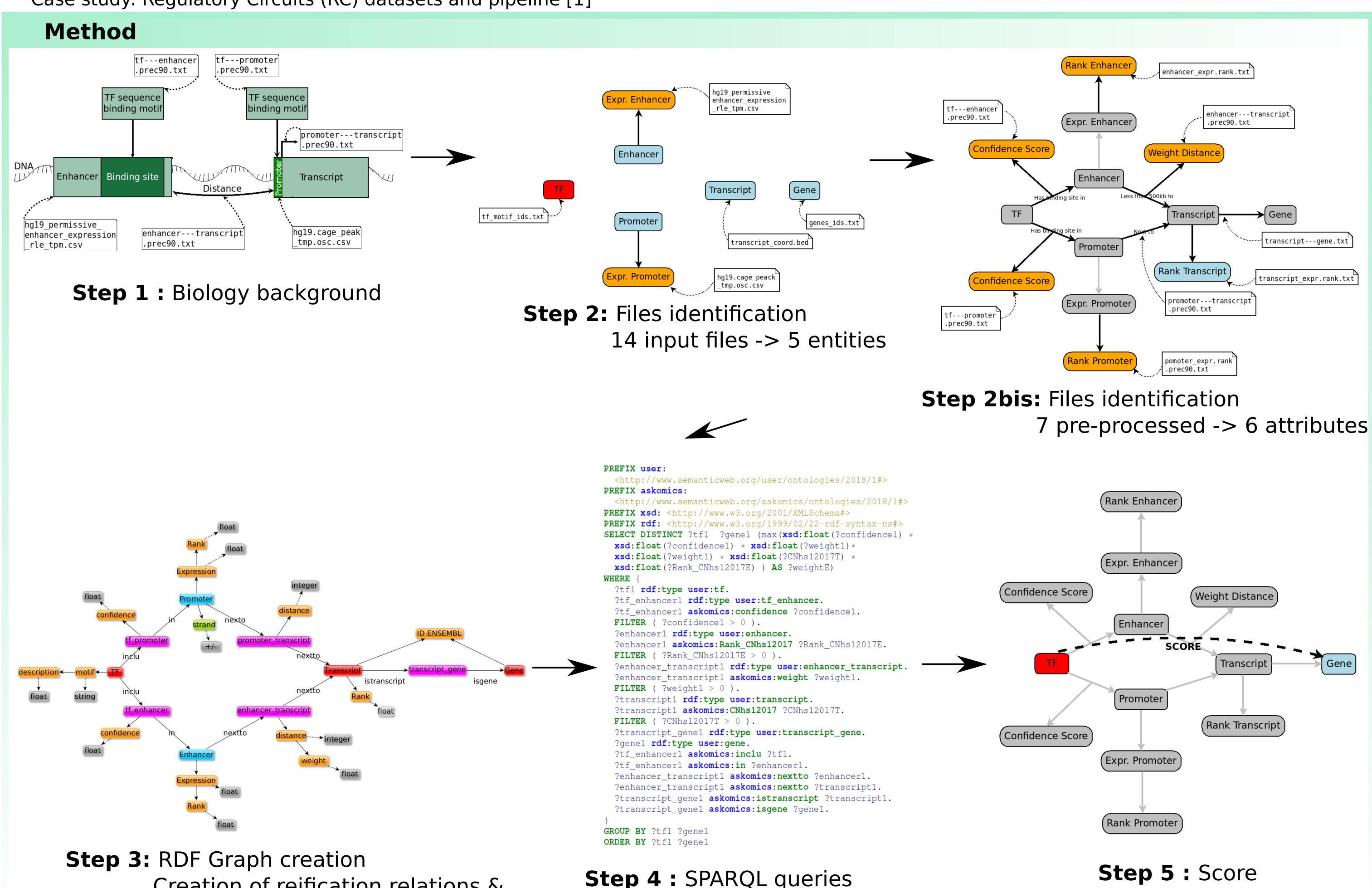
# Regulatory Circuits (RC) data[1]

http://regulatorycircuits.org -> 394 tissue-specific regulatory interaction networks.

21 files, 14 input and 7 pre-processed

**Issues**: - Poorly described files

- Link between entities (& files) are hidden
- No name standardization across files
- No header or missing header cases
- Static networks that can't be updated



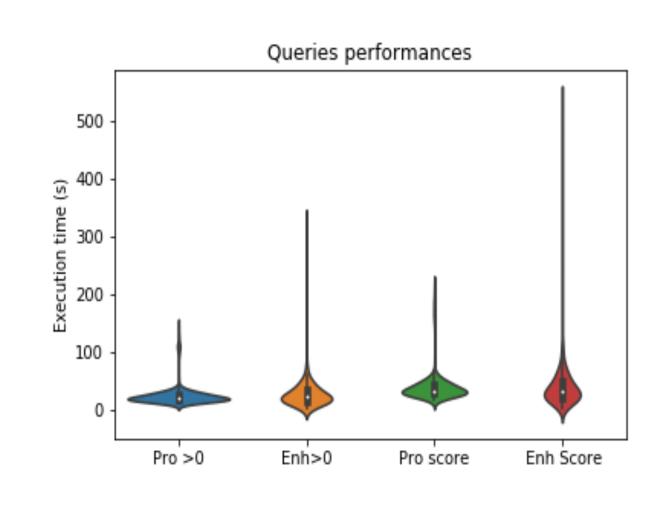
#### **Triple-Store data**

identifiers normalization

## **Triple store content:**

	Number of elements
Triples	335 429 988
Entities	3 226 341
Classes	10
Datasets	53

Queries' execution time: SPARQL queries to extract all tissue-specific regulatory networks



For all tissues using RDF = 4h30For 4 tissues using bash pipeline = 4h

Data set available at: https://regulatorycircuits-rdf.genouest.org/dump Queries can be executed at: https://regulatorycircuits-rdf.genouest.org/sparql

Creation of reification relations &

# Conclusion

Through Enhancer & Promoter

The RDF version of RC allows:

- To clarify link between entities
- A fine-grained exploration of the relations involved.
- To be **expended** with **user-specific data** and be linked to external databases.

We showed that the Regulatory Circuits analysis pipeline can be formalized as 2 SPARQL queries and that the performances were acceptable.

### **Bibliography & Acknowledgment**

- [1] D. Marbach, D. Lamparter, G. Quon, M. Kellis, Z. Kutalik, and Bergmann, "Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases," Nature methods, 2016.
- We acknowledge the GenOuest bioinformatics core facility www.genouest.org) for providing the computing infrastructure.









