

# REFRAME Project

March 2013–September 2016



<http://www.reframe-d2k.org/>

Rethinking the **E**ssence, **F**lexibility and **R**eusability of **A**dvanced **M**odel **E**xploitation

Nicolas Lachiche  
(Université de Strasbourg, France)

Bern, 28 April 2016

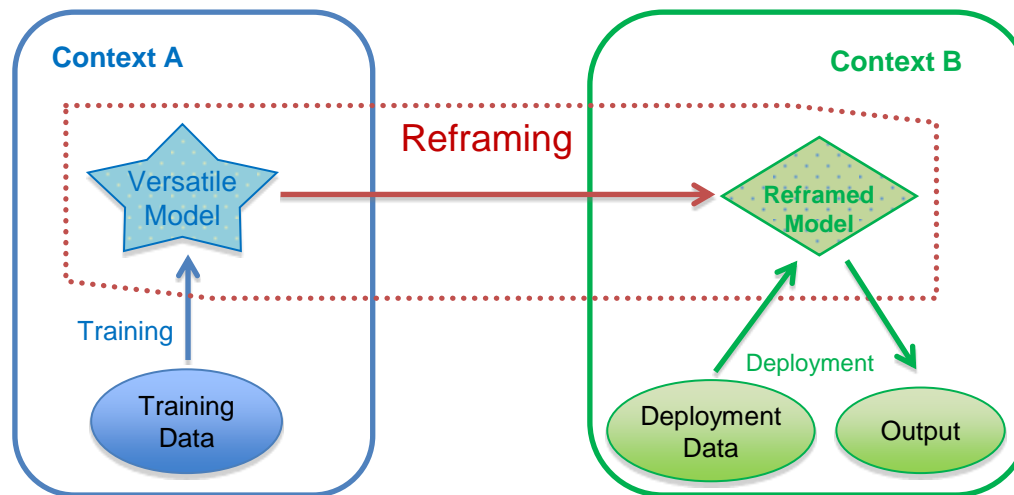
# Outline

---

- ▶ Scientific background
- ▶ Key challenges and potential impact
- ▶ Partners
- ▶ Scientific results and dissemination
- ▶ Sustainability/Valorisation

# Scientific background (1)

- ▶ What is “reframing”?
  - “process of applying an existing [machine learning/data mining] model to the new operating context by the proper transformation of inputs, outputs and patterns”.



# Scientific background (2)

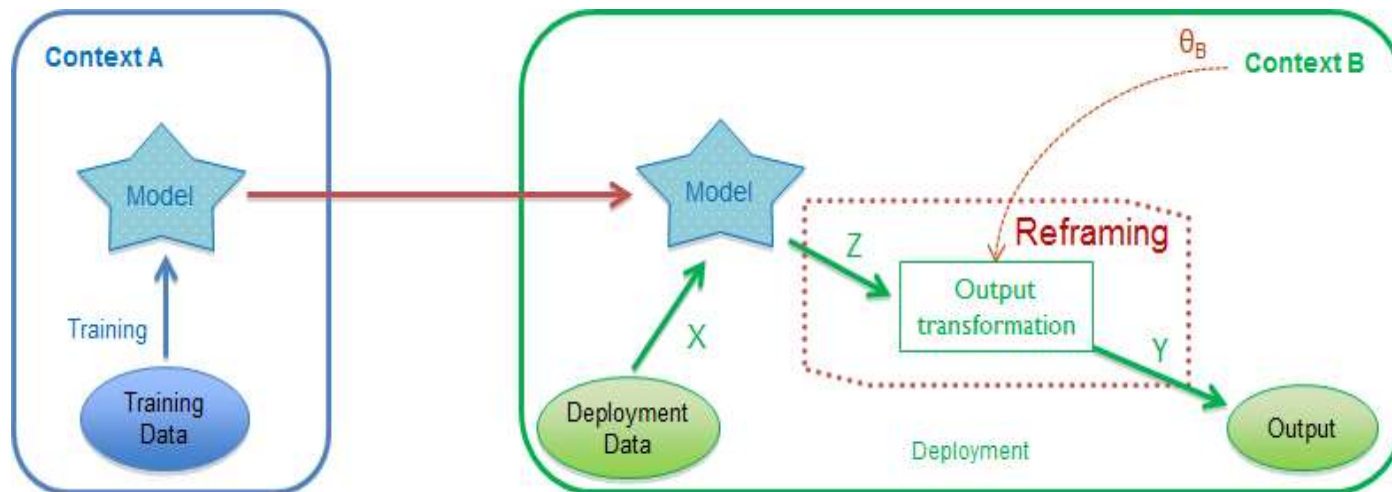
- ▶ What is a “context”?
  - Changes in data distribution.
  - Costs (output, inputs, labelling, etc.).
  - Data quality (noise, missing data)
  - Representation change (granularity, different attributes, background knowledge)
  - Task change
- One of the goals of the project is to generalise the notion of “operating context”.

# Challenges and potential impact (1)

- ▶ Usual D2K process
  - Models are trained in a context: overfitting, retraining
    - Inefficient and unreliable process
  - Either a set of models for a range of operating contexts, or a very general, but inflexible model.
    - We aim at building **versatile models** that can be properly deployed in a range of operating contexts.
- ▶ The purpose is to ensure “model reuse”.
  - Appropriate reframing procedures to apply it to any possible context.

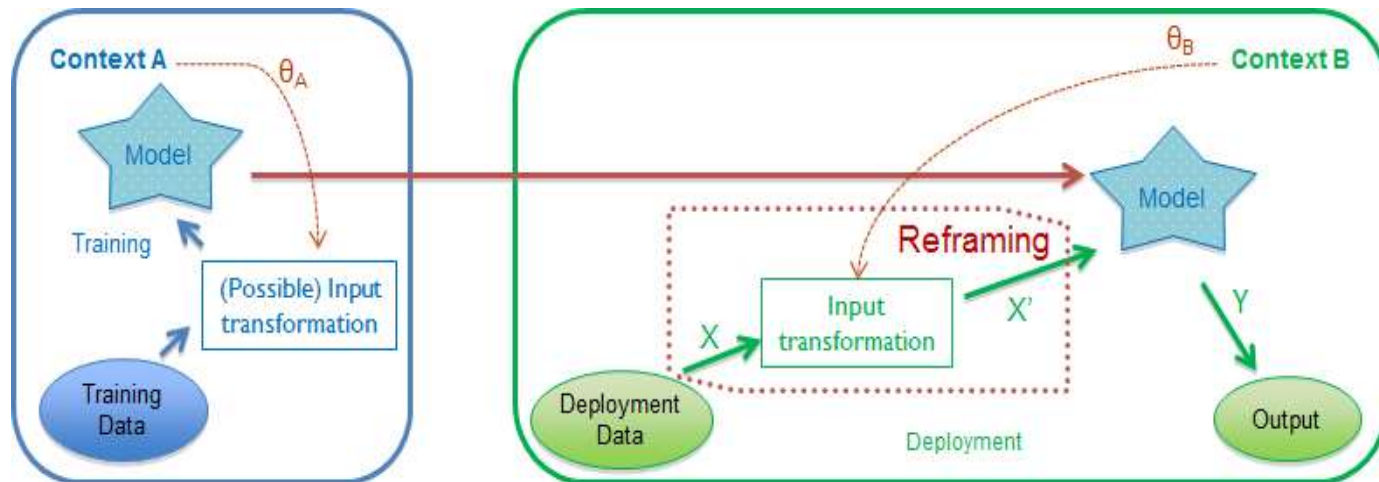
# Challenges and potential impact (2)

- ▶ Three kinds of reframing
  - Output reframing



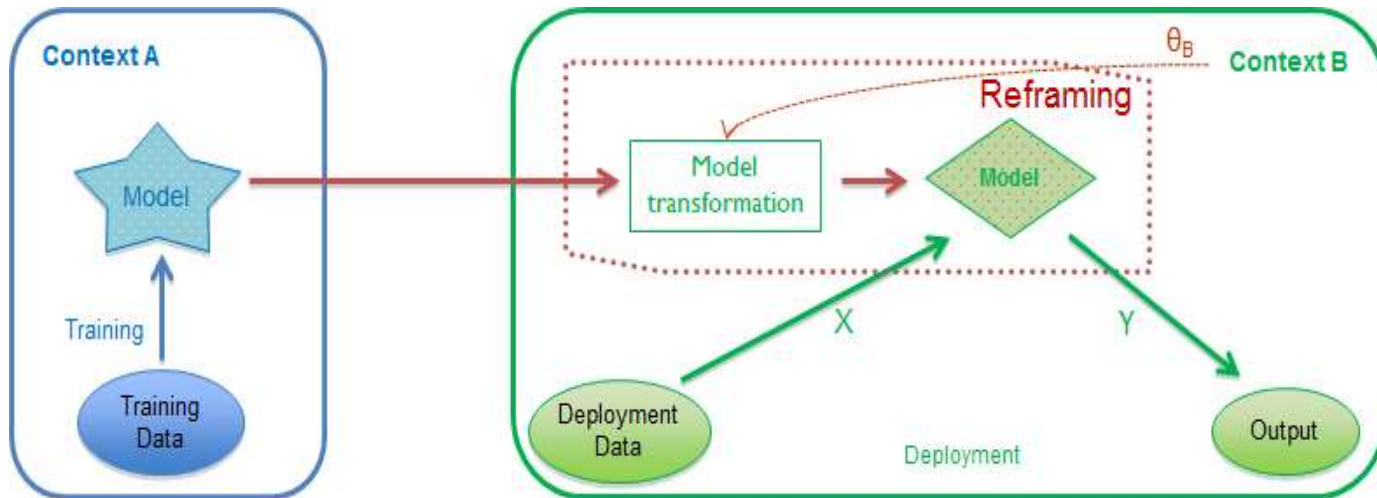
# Challenges and potential impact (3)

- ▶ Three kinds of reframing
  - Input reframing



# Challenges and potential impact (4)

- ▶ Three kinds of reframing
  - Structural reframing





# Consortium

- ▶ BRIS: University of Bristol / Intelligent Systems Laboratory
- ▶ STRAS: University of Strasbourg / ICube
- ▶ VAL: Universitat Politècnica de València / DSIC
  
- ▶ Most tasks involved two or even all partners.

# University of Bristol

- ▶ Senior investigators
  - Peter Flach, Ricardo Prudencio (Brazil)
- ▶ Specificity and added value
  - Machine learning, in particular metrics: ROC, precision–recall, etc.
    - ECML PKDD 2014 Best paper award!
  - Classifier calibration
  - Abduction and background knowledge
  - Multilabel versatile decision trees

# Université de Strasbourg

- ▶ Senior investigators
  - Nicolas Lachiche, Agnès Braud
- ▶ Specificity and added value
  - Relational data mining, complex aggregates
  - Practice of data mining: chemistry, bioinformatics, geography and environment
  - Input and output reframing
  - Integration of tools into the online data mining server [clowdflows.unistra.fr](http://clowdflows.unistra.fr) (cf. demo)

- ▶ Senior investigators
  - Jose Hernandez–Orallo, Cesar Ferri, Maria–Jose Ramirez–Quintana
- ▶ Specificity and added value
  - Data mining, from theory to applications
  - Quantification, cost–sensitive regression
  - Multidimensional prediction models
  - Various other changes of contexts: learning task, noise, attribute costs
  - Item response theory

# Results and dissemination (1)

- ▶ Characterisation of reframing
  - Input reframing
  - Output reframing
  - Structural reframing
- ▶ Several learning tasks
  - Classification
  - Regression
  - Frequent itemsets
- ▶ Various changes of contexts
  - Distribution shift
  - Costs
  - Data quality
  - Representation change
  - Task change

# Results and dissemination (2)

## ▶ Publications

	Y1	Y2	Y3
Int. Journal	3	3	2
Int. Conference	2	6	7
Int. Workshop	0	6	10
Total	5	15	19

## ▶ ECMLPKDD 2014 best paper award

# Results and dissemination (3)

- ▶ **Invited talks**
  - AAAI 2015 Fall Symposium on Embedded ML
  - PAPIs 2016 connect: AI for Business and Society
- ▶ **Workshops and challenge**
  - LMCE 2014 and LMCE 2015: Learning over Multiple Models @ ECMLPKDD 2014 and 2015
  - MoReBikeS: Model Reuse with bike rental station data @ ECMLPKDD 2015
- ▶ **Under review**
  - A systematic approach for model reuse in machine learning, AI Communications
  - Context-Aware Knowledge Discovery, ECMLPKDD 2016

# Sustainability/Valorisation

- ▶ Context-aware data mining methodology
  - From CRISP-DM to CASP-DM
- ▶ And tools
  - All the code available online
  - Progressively integrated into an online data mining network, [clowdflo.ws.unistra.fr](http://clowdflo.ws.unistra.fr)
- ▶ At the era of big data, open data and data science, learning reusable models is important for the sustainability of data mining itself.