

# End-User-Oriented Tool Support for Modeling Data Analytics Requirements

Hourieh Khalajzadeh, Andrew Simmons, Mohamed Abdelrazek, John Grundy, John Hosking, Qiang He

HumanISE Lab, Faculty of IT, Monash University

## Introduction

Data analytics application development introduces many challenges including: new roles not in traditional software engineering practices – e.g. data scientists and data engineers; use of sophisticated machine learning (ML) model-based approaches replacing many programming tasks; uncertainty inherent in the models; interfacing with models to fulfill software functionalities; as well as deploying models at scale and undergo rapid evolution, as business goals change and new data sources become available. BiDaML toolset brings all stakeholders around one tool to specify, model and document big data applications.

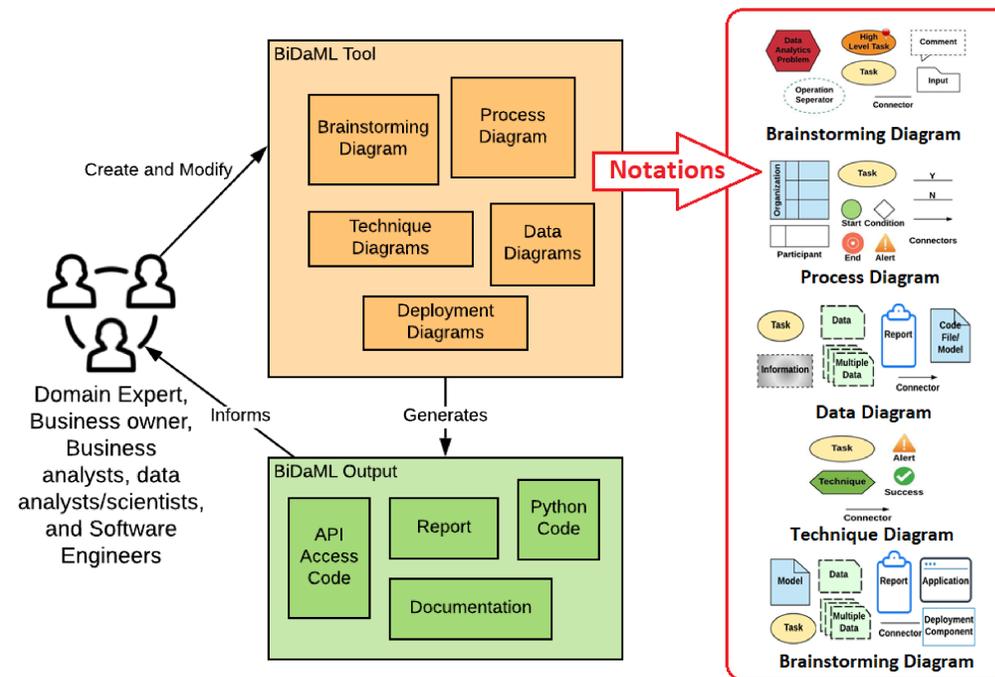


Figure 1: BiDaML Notations

## Example Usage

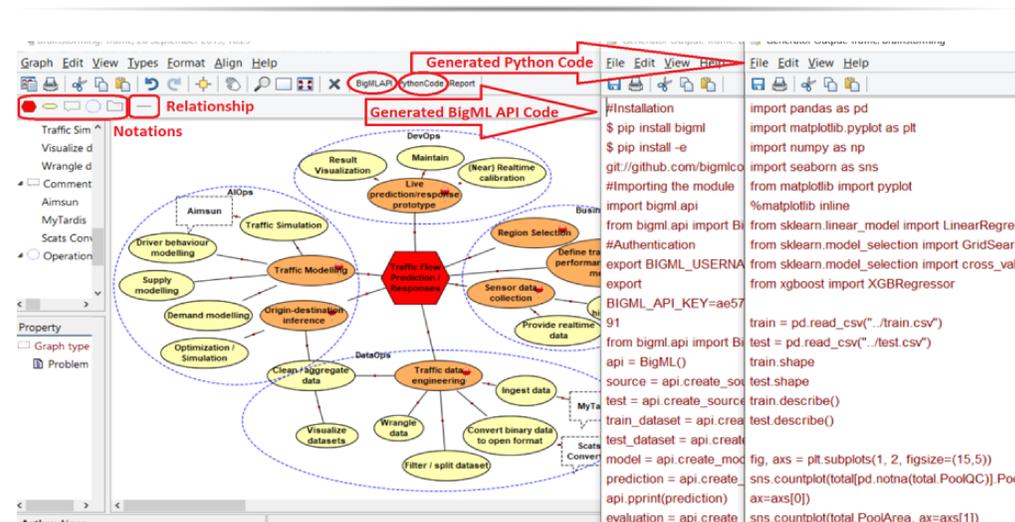


Figure 2: Brainstorming Diagram Created in BiDaML Tool for the Traffic Analysis Example and Snippets of the Generated Python Code.

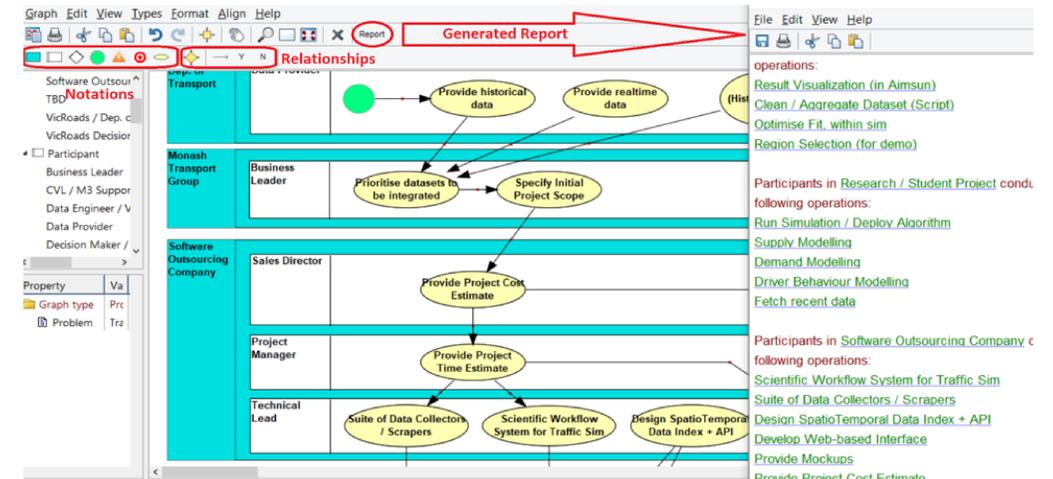


Figure 3: Process Diagram Created in BiDaML Tool and a Report Generated from the Process Diagram.

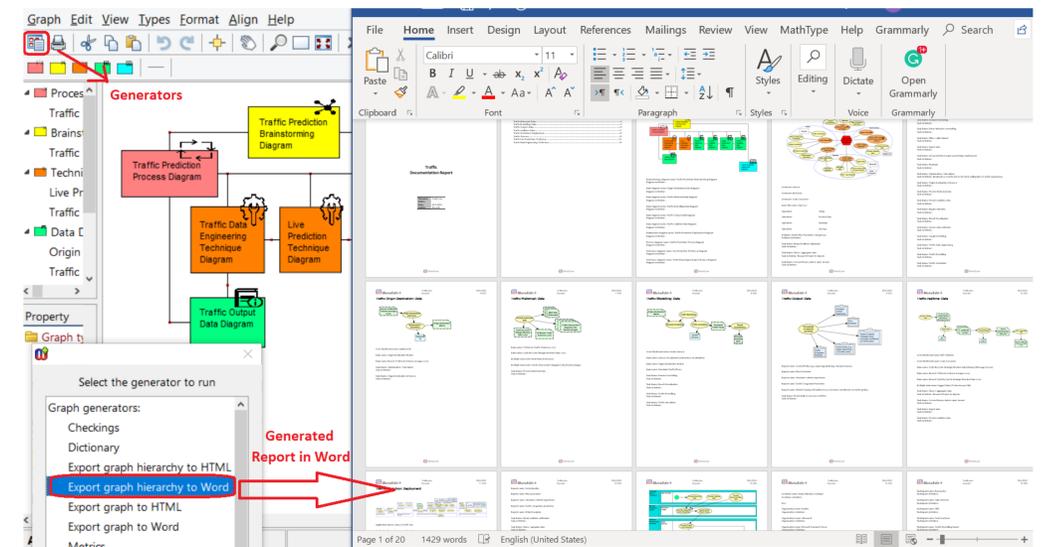


Figure 4: Overview of all the Diagrams Created in BiDaML Tool and the Final Report in Word Generated from the Overview Diagram.

## Challenges in developing Big Data Applications

- Domain experts, business analysts and business managers do not have a **background in data science and programming**;
- Data analysts, data scientists and software engineers do not have **domain knowledge**;
- Data scientists **lack software engineering expertise**;
- Lack of a **common language** between team members;
- Poorly support of **solution evolution**;
- Simply re-using existing solutions is not feasible. It is not easy to choose one language that is **understandable among such diverse teams**.

## Our BiDaML Approach

BiDaML [1] is a suite of domain-specific visual languages to support interdisciplinary teams through development of data analytics systems.

We worked with transport researchers and used our method to specify the intended VicRoads software solution workflow. We used our tool to document the entire data analytics workflow. As BiDaML forces the user to consider all phases of the project, the modeling process helped in revealing the gaps. Different diagrams created in BiDaML tool and the generated outputs are shown in Figs 2-4.

## References

- [1] H. Khalajzadeh, A. Simmons, M. Abdelrazek, J. Grundy, J. Hosking, and Q. He. An end-to-end model-based approach to support big data analytics development. *Journal of Computer Languages*, page 100964, 2020.

## Contact Information

- Web: <https://www.monash.edu/it/humanise-lab>
- Email: [hourieh.khalajzadeh@monash.edu](mailto:hourieh.khalajzadeh@monash.edu)