

Reviewer: Mike Maksimov

Paper Title: Mega-modeling for Big Data Analytics

Author(s): S. Ceri, E. Valle, D. Pedreschi, R. Trasarti.

1) Is the paper technically correct?

- Yes
- Mostly (minor flaws, but mostly solid)
- No

2) Originality

- Very good (very novel, trailblazing work)
- Good
- Marginal (very incremental)
- Poor (little or nothing that is new)

3) Technical Depth

- Very good (comparable to best conference papers)
- Good (comparable to typical conference papers)
- Marginal depth
- Little or no depth

4) Impact/Significance

- Very significant
- Significant
- Marginal significance.
- Little or no significance.

5) Presentation

- Very well written
- Generally well written
- Readable
- Needs considerable work
- Unacceptably bad

6) Overall Rating

- Strong accept (award quality)
- Accept (high quality - would argue for acceptance)
- Weak Accept (borderline, but lean towards acceptance)
- Weak Reject (not sure why this paper was published)

7) Summary of the paper's main contribution and rationale for your recommendation. (1-2 paragraphs)

The authors propose a new holistic data and model management approach, calling it "mega-modelling" (in some cases also referred to as mega-module??), for the acquisition, composition, integration and management of data and models. This work was brought on from the need for modelling capabilities to leverage on the power of big data. The paper presents the three main phases of mega-modules, namely, data preparation, data analysis and data evaluation. It then continues to present one of the most interesting parts of the paper in my opinion, which is the ability to combine mega-modules (creating "composite abstractions"). This ability is used to create sophisticated analytical processes, of which the most noteworthy are map-reduce, what-if control and drift control decompositions. Finally, the paper concludes by showcasing a variety of examples, each of which uses a specific combination of composite abstractions in order to display their effectiveness.

The paper's substance is good, but I personally found it to be a little hard to read at times. Not because of the content itself, but because of the way the paper was structured. I feel that it could have been presented in a more clear, tightly coupled and coherent way. After reading the first half of the paper, I still couldn't determine the exact research

question that the authors were looking into, or the problem that they were trying to solve. It only became clear after they started introducing examples. The authors also made it seem as though they coined the term "mega-modelling", which is something that has existed long before this work as far as I know. This, and the fact that the terms mega-modelling and mega-module were used interchangeably at times only added to the confusion.

8) List 1-3 strengths of the paper. (1-2 sentences each, identified as S1, S2, S3.)

S1 - The authors create and propose the use of a mega-module graphical element which captures the 3 stages of information processing (data preparation, data analysis, data evaluation).

S2 - The mega-modules can be combined into what is called a compositional abstraction. These combinations enable different analytical processes by allowing multiple mega-modules to be enclosed (nested) into one enclosing mega-module.

9) List 1-3 weaknesses of the paper (1-2 sentences each, identified as W1, W2, W3.)

W1 - The work is preliminary and therefore lacks formalization. There is no meta-model explaining the model's syntax and behavior.

W2 - All of the example applications that were used in this paper to demonstrate and evaluate mega-modules and their capabilities were previous works of the authors. They referenced themselves a total of 8 times (each a different paper - 11,13,15,16,17,21,22,23). Referencing oneself multiple times is not necessarily a bad thing, but evaluating your work on benchmarks of your own creation is convenient and doesn't reflect neutral results.