

# Semantic description of linear mixed model analysis

**Hanna Ćwiek-Kupczyńska<sup>1</sup>, Katarzyna Filipiak<sup>2</sup>,  
Augustyn Markiewicz<sup>3</sup>, Philippe Rocca-Serra<sup>4</sup>,  
Alejandra N. Gonzalez-Beltran<sup>4</sup>, Susanna-Assunta  
Sansone<sup>4</sup>, Emilie J. Millet<sup>5</sup>, Fred van Eeuwijk<sup>5</sup>,  
Agnieszka Ławrynowicz<sup>6</sup>, and Paweł Krajewski<sup>1</sup>**

<sup>1</sup>*Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland*

<sup>2</sup>*Institute of Mathematics, Poznań University of Technology, Poland*

<sup>3</sup>*Department of Mathematical and Statistical Methods, Poznań University  
of Life Sciences, Poznań, Poland*

<sup>4</sup>*Oxford e-Research Center, University of Oxford, UK*

<sup>5</sup>*Biometris, Wageningen University and Research Centre, The Netherlands*

<sup>6</sup>*Faculty of Computing, Poznań University of Technology, Poland*

## Abstract

In this talk, we discuss the application of results of linear mixed model (LMM) analysis for the purpose of experimental data management.

Intensive production of data in many research domains requires that the obtained results are efficiently processed and managed. As data reuse is anticipated, especially in experimental life sciences, it becomes increasingly important to share research datasets in public repositories. The availability of comprehensive statistical data summaries can be useful for efficient exploration of those data resources.

To improve data management practices, FAIR data principles [1] have been proposed: scientific data should be findable, accessible, interoperable and reusable. Semantic web technologies [2] provide methods to exchange data in a machine readable and interoperable way. The application of public ontologies to classify individual pieces of information allows to explicitly define their meaning and context, and to make them available as Linked Data [3].

In the present work, we describe a semantic model for the statistical analysis of datasets by linear mixed models [4]. We use ontologies, in particular Statistics Ontology (STATO) [5] to annotate the conclusions from the LMM analysis. We demonstrate the use of our approach for automated processing of datasets obtained from plant phenotyping experiments.

## Keywords

experimental data analysis, linear mixed model, FAIR data, scientific data interoperability, Statistics Ontology, semantic web, Linked Data

## Acknowledgements

This research is partially supported by the National Science Centre (NCN), project no. 2016/21/N/ST6/02358.

## References

- [1] Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 2016;3.
- [2] Berners-Lee, T., Hendler, J. (2001). Publishing on the semantic web. *Nature* 26;410(6832):1023-4.
- [3] Linked Data: <https://www.w3.org/standards/semanticweb/data>
- [4] Ćwiek-Kupczyńska, H., Filipiak, K., Markiewicz, A., Rocca-Serra, P., Gonzalez-Beltran, A.N., Sansone, S-A., Millet, E.J., van Eeuwijk, F., Lawrynowicz, A., Krajewski, P. (2019). A semantic concept schema of the linear mixed model of experimental observations. Submitted.
- [5] STATO 2012, statistics ontology: <http://stato-ontology.org>