

Contributed Talk

Splinter E-Science

REPRODUCIBILITY IN AN ERA OF DATA DRIVEN SCIENCE

Polsterer, Kai

Heidelberger Institut für Theoretische Studien

Reproducibility of scientific research results is of tremendous importance, to enable other researchers to validate, to check and to build on published results. In data-driven research this requirement is more than publishing research results as a plain paper. We have to start sharing and publishing code as well as referencing the software packages that had been utilized. Data-sets used to train and/or derive models have to be published alongside with the code. The provenance of the data is as important as providing uncertainties. The use of proper scores to evaluate the performances and the publication of reference data-sets have to become standard in astronomy. When using deep learning schemes the derived weights, biases and hyper-parameters have to be published, too. This talk will focus on some of these important aspects.