Geophysical Research Abstracts
Vol. 19, EGU2017-12116, 2017
EGU General Assembly 2017

Towards a Big Data Infrastructure for DataLab experiments

Jan Willem Noteboom, Andrea Pagani, and Raymond Sluiter
Royal Netherlands Meteorological Institute (KNMI), De Bilt, the Netherlands (janwillem.noteboom@knmi.nl)

KNMI has established a DataLab to serve data-driven innovations through the execution of collaborative experiments with public and private sector partners. These experiments are proof of concepts that combine traditional meteorological data with data from other domains (e.g. justice, transportation, economics) and emerging data (e.g. crowed sourced data, IoT). Big data analytics including machine learning are playing an important role in these experiments. This paper presents the road towards a Big Data Infrastructure that facilitates these KNMI DataLab experiments most effectively and include lessons learned in this journey.

Key challenge for an effective Big Data Infrastructure is to find the right balance between customer needs, resources, technologies and processes to facilitate big data research experiments. A set of principles has been used to guide the realization of the Infrastructure taking into account related complexities. The principles include both managerial and technical elements and cover data and scientific aspects such as: i) reproducibility of results, ii) scalability of computation by design, iii) collaboration on code and result interpretation, iv) open data, results, and code. Examples of the implementation of principles by KNMI DataLab experiments using the Big Data Infrastructure are presented as part of this paper.