



U.S.E. Research Institute  
Working Paper Series 19-22

# From Dirty Data to Tidy Facts: Clustering Practices in Plant Phenomics and Business Cycle Analysis

Marcel Boumans  
Utrecht School of Economics  
Utrecht University  
Sabina Leonelli

December 2019

## Abstract

This chapter considers and compares the ways in which two types of data, economic observations and phenotypic data in plant science, are prepared for use as evidence for claims about phenomena such as business cycles and gene-environment interactions. We focus on what we call “cleaning by clustering” procedures, and investigate the principles underpinning this kind of cleaning. These cases illustrate the epistemic significance of preparing data for use as evidence in both the social and natural sciences. At the same time, the comparison points to differences and similarities between data cleaning practices, which are grounded in the characteristics of the objects of interests as well as the conceptual commitments, community standards and research tools used by economics and plant science towards producing and validating claims.