Case Studies in Mining Text for Plant Pests and Pathogens

Dr. Laura Tateosian
North Carolina State University

Pest observation and distribution data for biosecurity measures and forecasting models commonly come from field observations, official reports, and genetic records consolidated through published literature and databases. While vital, these data can suffer from latency and spatial and temporal sparseness, due to the cost of collecting and collating these data at scale. These limitations have implications for the predictive capabilities of models and the success of control and eradication programs. At the same time, there is a wealth of historical text data as well as growing evidence that Web media could be a valuable source of supplementary data to tap into information for understanding pest encroachments. Gathering data from these sources presents its own challenges. We discuss practical considerations for processing historical text records and collecting online media in order to extract data from text to support pest biosurveillance. We document the presence of mentions of invasive pests and pathogens in these sources through contemporary case studies regarding *Lycorma delicatula*, *Tuta absoluta*, and *Phytophthora infestans* and an exploration of the early history of *P. infestans* through historical documents. Our results using past data demonstrate that these sources provide valuable spatial information describing pest presence and spread.