The relevance of data:
Is all the data equally important?

Dr. Lorena Parra, Universitat Politècnica de València and Instituto Madrileño de Investigación y Desarrollo Rural, Agrario y Alimentario, Spain.

In Wireless Sensor Networks (WSNs), Internet of Things (IoT), or other data gathering systems generate colossal volumes of data. In many cases, data is gathered every few minutes in order to control the ongoing monitored processes or activities. If we have ten sensors monitoring a process every five minutes, we generate 84,000 data registers each month. Considering the cost of data storage in terms of energy and the economic cost of renting the virtual space, we must evaluate the rentability of data storage. Data sending is one of the most depleting activities that nodes carried out. In addition, when sending sensors' data, there is a huge inefficiency in the communication since the portion of the packet occupied by the data is minimum compared with the heathers. Thus, we need to ask ourselves, is all the data equally important? Do we need to send every gathered data? In a series of data, we have to identify which values can be considered useful information to decide if the data should be sent or not. In this keynote, we analyse what we need to consider in order to define if gathered data must be sent. We will focus on data variability in terms of variables and in terms of monitored scenarios. The same variable might have different variability in different environments. Moreover, the particular case of images will be explained with real examples of green area monitoring data. Finally, we will focus on how we can take advantage of managing data variability properly to define algorithms. This data management will be illustrated with different algorithms that can determine if data is essential and must be sent or not will be. The examples will show three different algorithms for indoor and outdoor monitoring, which have been evolving. The first is integrated into a single-hop network in which each node determine if the data should be sent or not according to the previously transmitted data. In the second case, the algorithm is implemented in a multi-hop network. In this case, the sensors will send the data according to the gathered data, or if the node receives a message from the preceding node is received. In the last case, the algorithm includes storing the data to be sent once a new value is considered valuable information and must be sent.