# 1 Defining Big Data

*Small data are slow and sampled. Big Data are quick and n=all.*

Kitchin & McArdle (2016)

*This chapter searches for defining properties of big data, focusing on characteristics with possible implications for cartographic practice. Review of related works outlines the main attitudes towards grasping the concept.*

## 1.1 Ontological characteristics

Despite the lively interest triggered by the subject, the explanation of the term *big data*[[1]](#footnote-2) remains hazy and there is no widely accepted definition to the date. Perhaps the most systematic effort in this matter by Kitchin (2014) (refined in Kitchin & McArdle (2016)) summarizes the key properties attributed to big data. Kitchin critically evaluates these properties and goes on to assign them a relative importance in distinguishing big from “small” data. He also takes care to separate the concept in itself from accompanying social phenomena, hence he speaks of *ontological* characteristics.

# 2 Making sense of spatial big data

*Technology is the answer, but what was the question?*

Cedric Price

*This chapter looks more closely on the properties of data with point spatial reference that count for the majority of spatial big data. Then we will outline the tendencies in spatio-temporal knowledge discovery, discuss general ways how cartography can support understanding the world trough the lens of big data. We will also discuss some objections*

## 2.1 Spatial big data classification: stations, events, and agents

The vast majority of what is presently understood as spatial big data has point spatial reference. This prevalence comes naturally if we realize that the “data point” location is described basically as a coordinate pair – two digits that can be easily stored in standard database systems without the need to observe topological rules and other constraints that GIS vector data model enforces on line and polygon geometries. Point data are spatial data that are easily created and handled by non-spatial (meaning not GIS-enabled) systems that account for majority of data production. For this reason, and due to the scope limits of this thesis, we will

1. Throughout the text we will treat the term as plural, without capitalization. Although there are strong arguments for “data” as singular (Widman (2014), Nunberg (2013), for counterargument emphasizing the plurality of big data see Wilson, Thompson, Watson, Drew, & Doyle (2017)) and some authors do capitalize, we chose to match with the majority of big data related literature. This does not apply to direct citations where we preserve the original author’s formulation. [↑](#footnote-ref-2)