Turning to the enthusiasts, we see the ﬂip side of the same coin. If a “bad” cor-

relation is one for which we have not yet been able to provide a story accounting

for it, a supposedly “good” correlation cannot prove its goodness by just lining

up its numbers. The correlation must be shown to manifest a lawlike regularity;

there mu st be a theoretical account of it; the numbers are the beginning of the sto-

ry that needs to be told, not (by far) the end of it. More than that, this theoretical

account (this law) must cohere with all the other laws we also acknowledge: as

soon as a contradiction arises, we know that we cannot be speaking about the

world.

Turning to the enthusiasts, we see the ﬂip side of the same coin. If a “bad” cor-

relation is one for which we have not yet been able to provide a story accounting

for it, a supposedly “good” correlation cannot prove its goodness by just lining

up its numbers. The correlation must be shown to manifest a lawlike regularity;

there mu st be a theoretical account of it; the numbers are the beginning of the sto-

ry that needs to be told, not (by far) the end of it. More than that, this theoretical

account (this law) must cohere with all the other laws we also acknowledge: as

soon as a contradiction arises, we know that we cannot be speaking about the

world.

Turning to the enthusiasts, we see the ﬂip side of the same coin. If a “bad” cor-

relation is one for which we have not yet been able to provide a story accounting

for it, a supposedly “good” correlation cannot prove its goodness by just lining

up its numbers. The correlation must be shown to manifest a lawlike regularity;

there mu st be a theoretical account of it; the numbers are the beginning of the sto-

ry that needs to be told, not (by far) the end of it. More than that, this theoretical

account (this law) must cohere with all the other laws we also acknowledge: as

soon as a contradiction arises, we know that we cannot be speaking about the

world.

Turning to the enthusiasts, we see the ﬂip side of the same coin. If a “bad” cor-

relation is one for which we have not yet been able to provide a story accounting

for it, a supposedly “good” correlation cannot prove its goodness by just lining

up its numbers. The correlation must be shown to manifest a lawlike regularity;

there mu st be a theoretical account of it; the numbers are the beginning of the sto-

ry that needs to be told, not (by far) the end of it. More than that, this theoretical

account (this law) must cohere with all the other laws we also acknowledge: as

soon as a contradiction arises, we know that we cannot be speaking about the

world.

23 May, 2019

Fulvio Mazzocchi

CNR – Institute for the Conservation and Valorization of Cultural Heritage

I-00015 Monterotondo (RM), Italy

fulvio.mazzocchi@cnr.it

Dear Dr. Mateu,

I am writing to submit a new version of the manuscript entitled “On Big Data: How should we make sense of them?”, to be taken into consideration for publication in *Mètode Science Studies Journal.*

The article has been revised and extended. Now it has a more original content, especially in its second part.

The English of the manuscript has been checked again.

Thanking you for your attention and looking forward to your response, please accept my best regards.

Sincerely,

Fulvio Mazzocchi

**On Big Data:**

**How should we make sense of them?**

Fulvio Mazzocchi

CNR – Institute for the Conservation and Valorization of Cultural Heritage

Via Salaria Km 29,300

I-00015 Monterotondo (RM), Italy

fulvio.mazzocchi@cnr.it

**Abstract**

The topic of Big Data is today extensively discussed, not only on the technical ground. This also depends on the fact that Big Data are frequently presented as allowing an epistemological paradigm shift in scientific research, which would be able to supersede the traditional hypothesis-driven method. In this piece, I critically scrutinize two key claims that are usually associated with this approach, namely, the fact that data speak for themselves, deflating the role of theories and models, and the primacy of correlation over causation. In so doing, I will also refer to a recent case history of data mining projects in the field of biomedicine, i.e. EXPOsOMICS. My intention is both to acknowledge the value of Big Data analytics as innovative heuristics, and to provide a balanced account of what could be expected and what not from it. Besides, I also focus on one aspect that today is subject to growing attention, i.e. the opacity that surrounds the algorithms underlying Big Data.

**Keywords**: Big Data, data-driven science, epistemology, end of theory, causality, opacity of algorithm.

**Author's bionote**

Fulvio Mazzocchi, biologist and philosopher, is a researcher at the Institute for Institute for the Conservation and Valorization of Cultural Heritage of the CNR (Italy). His activity is focused on epistemology (e.g. the relationship between Western science and other knowledge systems; the epistemology of complexity), philosophical issues of scientific research (e.g. the reductionism-holism debate in biology; climate model validation and robustness), and knowledge organization.