# Conference: The Ninth International Conference on Data Analytics - 2020

## Papers: Big Data Monetization: Discoveries from a Systematic Literature Review

Center of Advanced Studies and Systems of Recife

Authors: Domingos Monteiro - dsmpm@cesar.school
Luciano Monteiro – lam@cesar.school
Silvio Meira – silvio@cesar.school
Felipe Ferraz - fsf@cesar.school

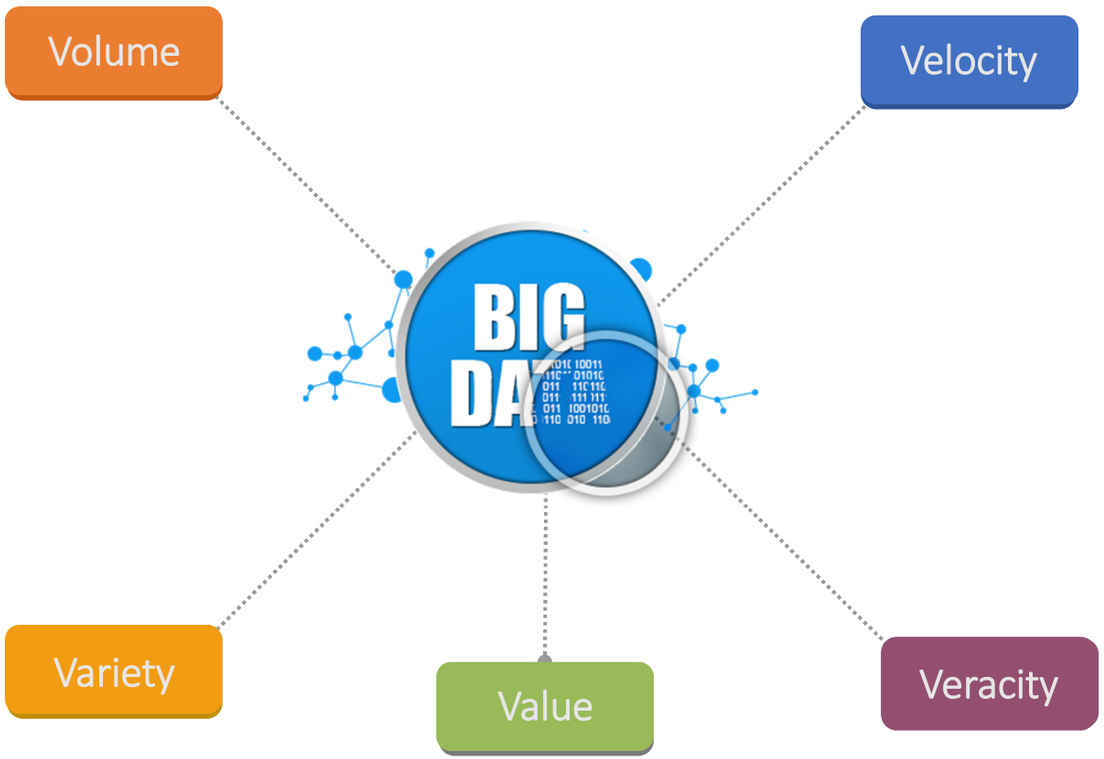
# About the Authors

“Domingos graduated in Computer Science from Federal University of Pernambuco (UFPE) in 1995 and Master in Artificial Intelligence from UFPE (1999). Strong entrepreneurship professional with a EPGC focused in Executive Education from Stanford University Graduate School of Business. He is currently CEO at Neurotech SA and PhD candidate at Cesar School. Domingos has experience in Computer Science, with an emphasis on Big Data and AI, working for different business industries in Brazil helping them to connect data with intelligence making the future more predictable.”

# Purpose of Papers

Our goal was to find out what methods have been applied to determine the relevance and value of data in a Big Data environment and if these methods are based, in any way, on information theory.

# Dimensions Big Data



# THEORETICAL FRAMEWORK – BIG DATA

“Big Data is the information asset characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value”.

De Mauro, Greco and Grimaldi [8]

# THEORETICAL FRAMEWORK – *The Value of Data as Digital Asset*

- The advent of computers brought the world a new category of assets, digital ones, represented in a discrete numerical way and used in digital devices with computational processing.

- These digital assets are capable of delivering a new category of products and services: better decisions, increased performance, competitive advantages and they can even be sold directly as a Product.

# Methodology Literature Review

# Research Problem

*How can we monetize new data generated and available in the Market in a Big Data environment?*

# Research Questions

- **RQ1**

  What methods have been suggested or applied to determine the relevance of the datum?

- **RQ2**

  What methodologies have been applied to identify the value of a datum?

- **RQ3**

  Does the methodology to identify the value use the information theory in its formulation? How?

# Search Protocol

| Applied Search String | | |
|---|---|---|
| | **Search Terms** | **Rational** |
| Population | (“Big Data”) | Studies that approach themes related to Big Data |
| **AND** | | |
| Intervention | data assets" OR "value evaluation" OR "data monetization" OR "data marketplace" OR "information value" OR "data value" OR "business value" | Studies that must be related to monetization of data, that is, the extraction of financial value. |

# Academic Bases

01 IEEE Xplorer

02 ACM Digital Library

03 ScienceDirect

C.E.S.A.R school

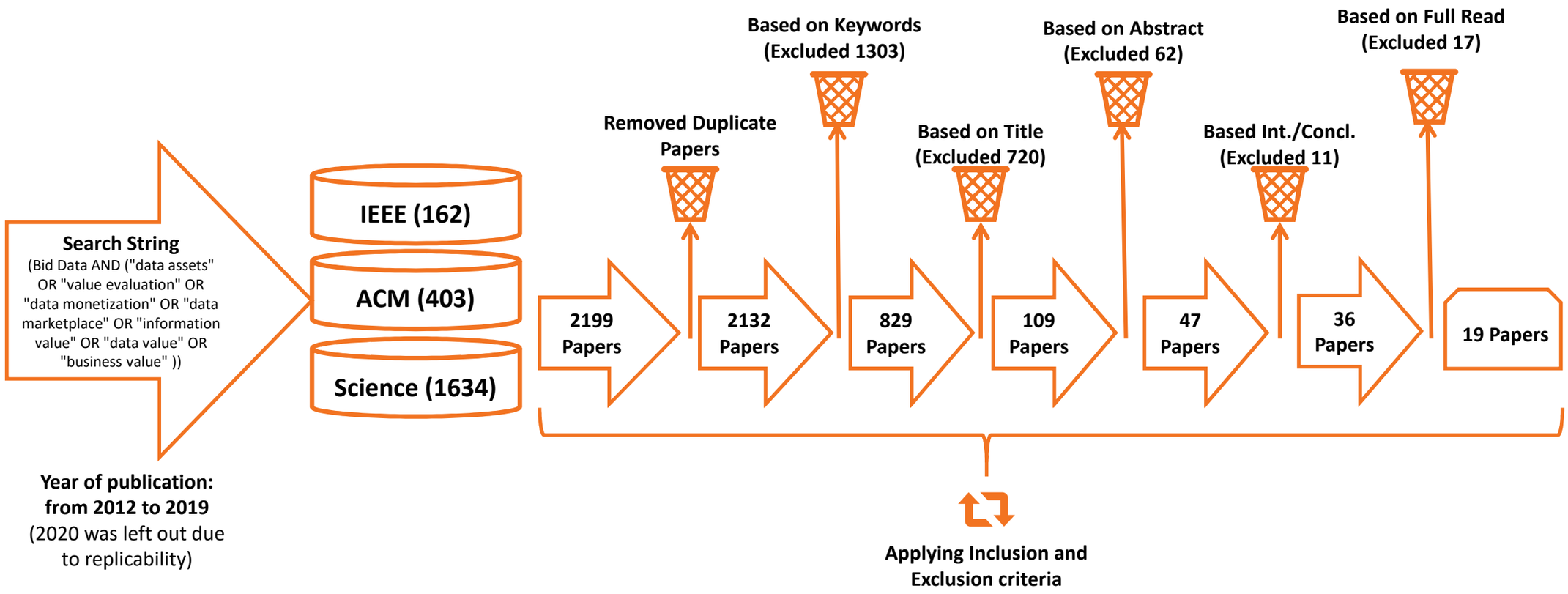
# Criteria for Inclusion

| CI | CRITERIA FOR INCLUSION |
|---|---|
| 1 | Primary studies published in English |
| 2 | Studies that approach the theme of Big Data and monetization of data |
| 3 | Studies that answer at least one of the research questions. |

# Criteria for Exclusion

| CI | CRITERIA FOR EXCLUSION |
|---|---|
| 1 | Academic or teaching-oriented articles |
| 2 | Short articles, courses, tutorials and secondary and tertiary studies (reviews) |
| 3 | Duplicated studies (in these cases, only one version was considered) |

# Results


Search String
(Bid Data AND ("data assets" OR "value evaluation" OR "data monetization" OR "data marketplace" OR "information value" OR "data value" OR "business value" ))
Year of publication: from 2012 to 2019
(2020 was left out due to replicability)
IEEE (162)
ACM (403)
Science (1634)
2199 Papers
Removed Duplicate Papers
2132 Papers
Based on Keywords (Excluded 1303)
829 Papers
Based on Title (Excluded 720)
109 Papers
Based on Abstract (Excluded 62)
47 Papers
Based Int./Concl. (Excluded 11)
36 Papers
Based on Full Read (Excluded 17)
19 Papers
Applying Inclusion and Exclusion criteria

# Results – Studies Answering Research Question by Database

| Bases | Papers |
|---|---|
| ACM Digital Library | 2 |
| IEEE xPlore | 12 |
| ScienceDirect | 5 |
| Total | 19 |

# Results - Answers to the Research Questions

RQ1: What methods have been suggested or applied to determine the relevance of the datum?

**Answer:**

a) In this review, the most applied methods to determine the relevance of data were the analytical techniques (50% of the findings, 7 cases);

b) Methods that evaluated the relevance of data in a business context were found in 29% of the cases (4 cases);

c) In 2 of the cases (14%) there were methods based in intrinsic characteristics on the data (kinds of data, volume, velocity,…);

d) Finally, in 1 of the articles, the use of the datum was applied as a method to determine the relevance of the datum (defined as the number of hits or references to a certain piece of information).

Even among the most used methods, it was not possible to find uniformity, for instance, in the case of the use of analytical techniques to determine the relevance of a datum, many distinct techniques were applied without the predominance of a specific one.

C.E.S.A.R school

# Results - Answers to the Research Questions

RQ2: What methodologies have been applied to identify the value of a datum? (14 answers)
?

| Way of determining value | % |
|---|---|
| SUPPLY X DEMAND | 35.71% |
| ANALYTICAL TECHNIQUES | 21.43% |
| USE OF THE DATUM (USE/HITS) | 14.29% |
| BOOK VALUE | 14.29% |
| RISK X RETURN | 7.14% |
| INFORMATION THEORY | 7.14% |

# Conclusions

The results showed that despite the progress in the theme of Big Data and in the application of analytical methodologies over the last decades, there is not yet a widely used data-based method to determine the value of a datum as a digital asset.

Based on the studies we found, it is possible to conclude that very little attention has been given to the dimension Value in academic research when compared to the three classical dimensions (Volume, Velocity and Variety) of the original definition of Big Data. Research is even more scarce in terms of financial value, and the existing studies do not share a common view on the right way of measuring and defining this value.