unimc UNIVERSITÀ DI MACERATA
l'umanesimo che innova

DIPARTIMENTO DI
ECONOMIA E DIRITTO

BIG DATA

LITERATURE REVIEW AND RESEARCH DIRECTIONS

Nicola Castellano

University of Macerata – Department of Economics and Finance

nicola.castellano@unimc.it

international
open data day
italia 2016

Big Data. A definition

Douglas Laney (Gartner group) 2001. "3D Data Management: Controlling Data Volume, Velocity, and Variety" The hallmark of attempting to characterize and visualize the changes that are likely to emerge in the future.
After a decade Laney updated its work (2012) "The importance of 'Big Data': A Definition". It has become Gartner's updated definition:

*Big Data is high **V**olume, high **V**elocity, and/or high **V**ariety information assets that require new forms of processing to enable enhanced decision-making, insight discovery and process optimization.*

Big Data characteristics are often summarized by **V**s:
huge Volume, high Velocity, huge Variety, and uncertain Veracity (Laney 2001; IBM 2012)

Management Disciplines involved in Research on Big Data

(Goes, 2014)

1. *Information sciences.*
2. *All disciplines that have realized the potential of working with diverse and large datasets (health care, public health, education, public policy, government studies, marketing and retail, finance, etc).*

Research streams on Big Data

1. *Big Data Infrastructure*
   - Data integration (Dey et al, 1998; Zhao and Ram, 2005)
   - Data validation (Fuller et al. 2011; Jiang et al. 2005; Nunamaker et al. 2011)
   - Designing and managing platform of services
   - Security
2. *Analytics and Decision Support Systems (Data → Information → Knowledge → Intelligence) Interdisciplinary applications. Particularly in the marketing field Dellarocas 2012; Ghose et al. 2012; Yang and Ghose 2010)*

Evolution of BI&A (Chen et al. 2012)

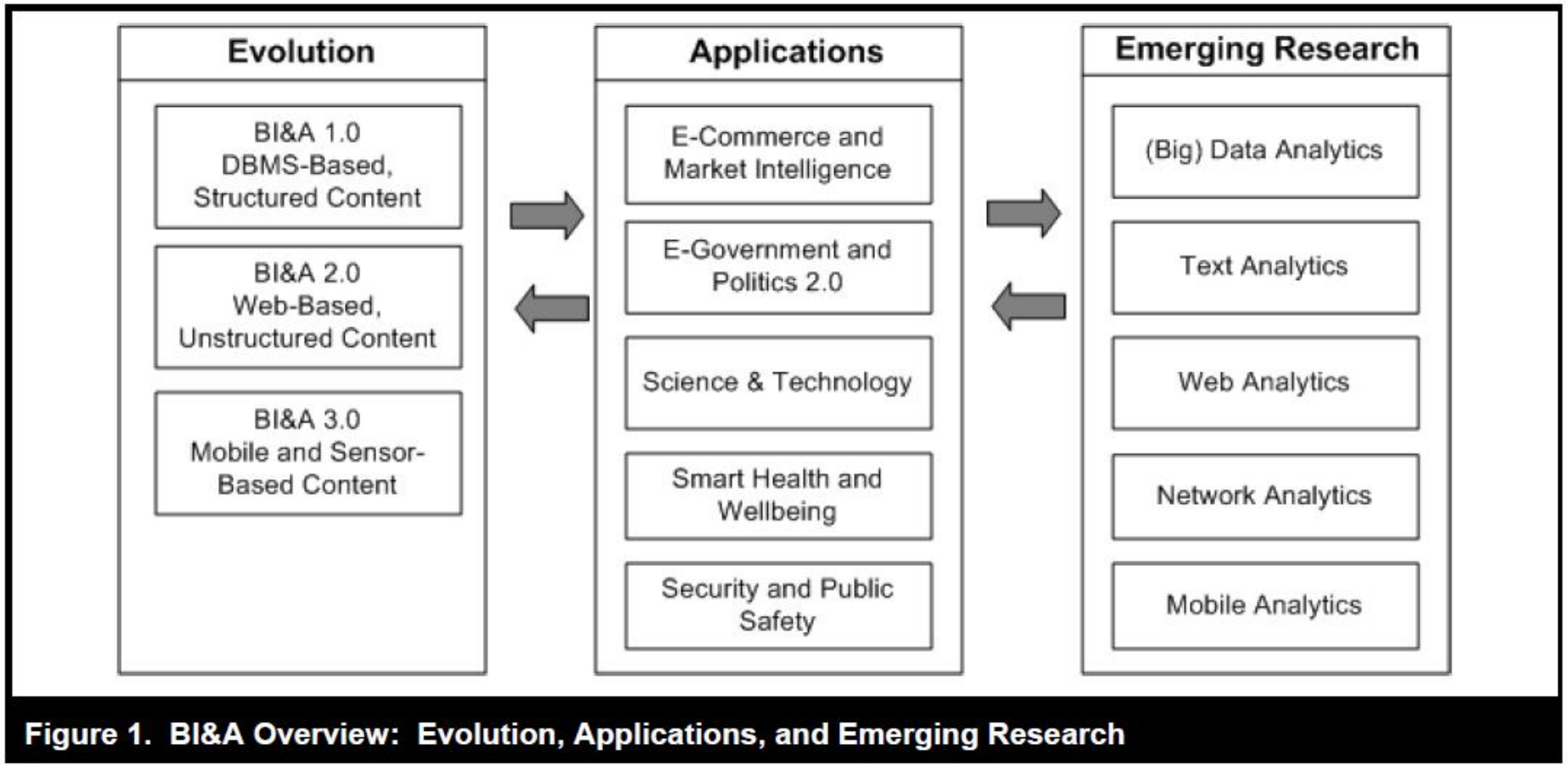

Figure 1. BI&A Overview: Evolution, Applications, and Emerging Research

Evolution of BI&A (Chen et al. 2012)

BI&A 1.0

data are mostly structured, collected by companies through various legacy systems, and often stored in commercial relational database management systems (RDBMS). The analytical techniques commonly used in these systems, popularized in the 1990s, are grounded mainly in statistical methods developed in the 1970s and data mining techniques developed in the 1980s.

BI&A 2.0

Centered on text and web analytics for unstructured web contents. The new gold mine for understanding customers' needs and identifying new business opportunities. BI&A 2.0 systems will require the integration of mature and scalable techniques in text mining, web mining, social network analysis, and spatial-temporal analysis with existing DBMS-based BI&A 1.0 systems.

BI&A 3.0

«Mobile-BI». Mobile and sensor-based content; Location-aware analysis; Person-centered analysis; Context-relevant analysis; Mobile visualization & HCI (Human-Computer Interaction).

Application of BI&A (Chen et al. 2012)

BI&A is highly applied and can leverage opportunities presented by the abundant data and domain-specific analytics needed in many critical and high-impact application areas:

(1) Ecommerce and market intelligence. Social media analytics of customer opinions, text analysis; sentiment analysis techniques; product recommender systems (such as association rule mining, database segmentation and clustering, anomaly detection, and graph mining
(2) e-government and politics 2.0. Online political participation (selected opinion mining, social network analysis, social media analytics techniques, e-democracy, political blogs and forums analysis, e-government service delivery, process transparency and accountability).
(3) science and technology,
(4) smart health and well-being, genomics-driven big data (genotyping, gene expression, sequencing data) and payer–provider big data (electronic health records, insurance records, pharmacy prescription, patient feedback and responses)
(5) security and public safety.

Major topics in BI&A Literature (Chen et al. 2012)

Big Data and Business Analytics

Big data forces you to wrestle with three key strategic and operational challenges:

Information Strategy – You need to harness the power of information assets. Big data is causing enterprises to find new ways to leverage information sources to drive growth.

Data Analytics – You need to draw more insight from your big data analytics or large and complex datasets. You need to predict future customer behaviors, trends and outcomes.

Enterprise Information Management – Information is everywhere - volume, variety, velocity - and it keeps growing. You need to manage access to growing extreme information management requirements and drive innovation in rapid information processing.

(Source: Gartner Group)

Data Analytics

Predictive Analytics – How are you using data for predictive and real-time analysis across various business domains? How can you use unstructured enterprise information and data to drive a better client experience? How are you leveraging new data types: sentiment data, clickstream data, video, images and text data?

Behavioral Analytics – How will you tap into complex data sets to create new models to drive business outcomes, decrease costs, drive innovation or improve customer satisfaction?

Data Interpretation – What new business analysis can be drawn from your data? How will IT help support insight discovery and new information trends? You need to know which data to integrate for new product innovation

(Source: Gartner Group)

Big Data in accounting (Vasarhelyi et al. 2015)

Accounting measurement has progressively lost its informational value with a substantial reduction of market value explanation by accounting variables; in particular for new knowledge intensive firms having higher intangible intensity.
Internal reporting has become very rich with enormous sets of information of a non-directly financial nature such as human resources, production, marketing, supply chain, and R&D. Although these ‘‘structured data stores’’ in ERPs are large, they can be dwarfed by expanded sets of less structured data.

Research Questions:

- What new measurements better represent an increasingly virtual set of variables that determine the value of the enterprise?
- What guidelines (e.g., GAAP) are needed to guide, inform, and standardize the above measurements?
- Can these measurements be extracted from extant ERPs, or are other data sources required?
- How can current ERP structures be enriched to provide a richer set of measurements?

Big Data «bridges» (Vasarhelyi et al. 2015)

Business is finding greater value when it links automatically captured and outside data sources. For example:

B2B (Business-to-Business) and B2C (Business-to-Consumer) databases can be used to replace historic values in Balance Sheets (Warren et al. 2015) (in line with fair value).

Extractions from video data can serve to support marketing and identify bottlenecks in the supply chain (Warren et al. 2015).

Textual data streams from social media can support marketing, give an early warning to customer service of product flaws, and serve to predict volume of sales (Warren et al. 2015).

Special devices with embedded RFID and GPS chips can automatically collect enriched electronic transaction information for a plethora of different uses.

Research Questions:

- Can Big Data bridges be used to enrich the above new measurements without adding volatility?
- How intrusive are these bridges in personal privacy?
- How to quantify the above-mentioned relationships into operationally usable functions?
- Should these functions be included into accounting standards to facilitate comparability?

Practical applications of Big Data (Vasarhelyi et al. 2015)

Practical applications of the expansion of accounting data sets may include: (1) the ability to process and analyze detailed rather than summary transaction data; (2) the ability to integrate into analysis a variety of both internal and external data with financial data; (3) the ability to do ‘‘soft integration’’ of environmental Big Data (e.g., social networks and news pieces) with accounting measurement and audit assurance processes; and (4) the ability to transform accounting, business, and audit processes based on (1), (2), and (3).