

Chapter 101

The Disruptive Impact of Emerging Technology

Gordon J. Murray
Kent State University, USA

ABSTRACT

In this chapter, context for understanding the phenomenon of “big data” and disruptive innovation is introduced relative to current changes affecting the future of the journalism industry. Perspective is provided on market forces and emerging technologies that now shape the demand for data journalism. Current best practices and strategies to analyze, scrape, personalize, visualize and map data are presented. Trends and resources to access data and effectively analyze information are outlined for journalists to use when researching and reporting online. Three contemporary case studies explore the day-to-day operations and decision-making processes of media organizations struggling to remain profitable; adapt to changing consumer demands and try to serve a new demographic that is increasingly global, wireless, mobile and socially networked.

BIG DATA AND DIGITAL DISRUPTION

Data matures like wine, applications like fish. —James Governor, Principal Analyst and founder of RedMonk

Humans have been producing volumes of data—collecting, storing and aggregating them in one form or another for millennia. Researchers studying clay balls from Mesopotamia have discovered clues to a lost code used for record-keeping about 200 years before writing was invented. The clay balls may represent the world’s “very first data storage system,” said Christopher Woods, a professor at the University of Chicago’s Oriental Institute (Jarus, 2013).

Data are plain facts and alone, are fairly useless. One of the clay balls found in Mesopotamia contained 49 pebbles and a cuneiform text. When data are processed, analyzed and interpreted to determine true meaning, they are transformed into information. The cuneiform text is now thought to be a contract commanding a shepherd to care for 49 sheep and goats. Four thousand kilometers northwest of Mesopo-

DOI: 10.4018/978-1-4666-9840-6.ch101

tamia in Wiltshire, England, there is an astronomical calculator that can calculate times in the solar year and the lunar month. We now know ancient worshippers may have used the calculator—Stonehenge, to collect the data required to predict eclipses on a 56-year cycle.

The capacity of our data storage devices has improved remarkably since the days of clay balls and stone monoliths. According to the *Oxford English Dictionary*, the first attempts to acknowledge the volume of data, or what has popularly been known as the “information explosion,” is encountered in 1941, when the term first appeared in the common vernacular. In *A Very Short History Of Big Data*, author Gil Press (2013) marks many of the significant milestones on the continuum of how our digital data—now stored microscopically on silicon—became so big. Highlights from the timeline include:

- 1944:** *The Scholar and the Future of the Research Library*, (Rider, 1944) estimates American university libraries were doubling in size every sixteen years and speculates the Yale Library in 2040 will have “approximately 200,000,000 volumes, which will occupy over 6,000 miles of shelves... [requiring] a cataloging staff of over six thousand persons.
- 1961:** *Science Since Babylon*, (Price, 1961) charts the growth of scientific knowledge and concludes the number of scientific journals and papers has grown exponentially rather than linearly, doubling every fifteen years and increasing by a factor of ten during every half-century.
- 1971:** In *The Assault on Privacy*, Miller (1971) writes, “Too many information handlers seem to measure a man by the number of bits of storage capacity his dossier will occupy.”
- 1980:** In *Where Do We Go From Here?* Tjomsland, (1980) says, “Those associated with storage devices long ago realized that Parkinson’s First Law may be paraphrased to describe our industry—‘Data expands to fill the space available’.... I believe that large amounts of data are being retained because users have no way of identifying obsolete data; the penalties for storing obsolete data are less apparent than are the penalties for discarding potentially useful data.”
- 1996:** Digital storage becomes more cost-effective for storing data than paper (Morris & Truskowski, 2003).
- 2000:** Lyman & Varian (2000) published, *How Much Information?* the first comprehensive study to quantify, in computer storage terms, the total amount of new and original information created in the world annually. The study finds in 1999, the world produced about 1.5 exabytes of unique information, or about 250 megabytes for every man, woman, and child on earth. It also finds “a vast amount of unique information is created and stored by individuals” (what it calls the “democratization of data”) and “not only is digital information production the largest in total, it is also the most rapidly growing.” Calling this finding “dominance of digital,” Lyman and Varian say that “even today, most textual information is ‘born digital,’ and within a few years this will be true for images as well.”

WE ARE DATA STORAGE

Two men were examining the output of the new computer in their department. After an hour or so of analyzing the data, one of them remarked: Do you realize it would take 400 men at least 250 years to make a mistake this big? —Nicholas Negroponte Chairman Emeritus of Massachusetts Institute of Technology’s Media Lab

21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/the-disruptive-impact-of-emerging-technology/150263

Related Content

Combining kNN Imputation and Bootstrap Calibrated: Empirical Likelihood for Incomplete Data Analysis

Yongsong Qin, Shichao Zhang and Chengqi Zhang (2010). *International Journal of Data Warehousing and Mining* (pp. 61-73).

www.irma-international.org/article/combining-knn-imputation-bootstrap-calibrated/46943

When Spatial Analysis Meets OLAP: Multidimensional Model and Operators

Sandro Bimonte, Anne Tchounikine, Maryvonne Miquel and François Pinet (2010). *International Journal of Data Warehousing and Mining* (pp. 33-60).

www.irma-international.org/article/when-spatial-analysis-meets-olap/46942

Routing Attribute Data Mining Based on Rough Set Theory

Yanbing Liu, Shixin Sun, Menghao Wang and Hong Tang (2006). *International Journal of Data Warehousing and Mining* (pp. 27-41).

www.irma-international.org/article/routing-attribute-data-mining-based/1769

Privacy Preserving Data Mining, Concepts, Techniques, and Evaluation Methodologies

Igor Nai Fovino (2008). *Successes and New Directions in Data Mining* (pp. 277-301).

www.irma-international.org/chapter/privacy-preserving-data-mining-concepts/29963

SQL-Based Fuzzy Query Mechanism Over Encrypted Database

Zheli Liu, Jingwei Li, Jin Li, Chunfu Jia, Jun Yang and Ke Yuan (2014). *International Journal of Data Warehousing and Mining* (pp. 71-87).

www.irma-international.org/article/sql-based-fuzzy-query-mechanism-over-encrypted-database/117159