Classification of Types of Big Data

The following classification was developed by the Task Team on Big Data, in June 2013. Comments and feedback are welcome.

1. **Social Networks (human-sourced information)**: this information is the record of human experiences, previously recorded in books and works of art, and later in photographs, audio and video. Human-sourced information is now almost entirely digitized and stored everywhere from personal computers to social networks. Data are loosely structured and often ungoverned.
   - 1100. Social Networks: Facebook, Twitter, Tumblr etc.
   - 1200. Blogs and comments
   - 1300. Personal documents
   - 1400. Pictures: Instagram, Flickr, Picasa etc.
   - 1500. Videos: Youtube etc.
   - 1600. Internet searches
   - 1700. Mobile data content: text messages
   - 1800. User-generated maps
   - 1900. E-Mail

2. **Traditional Business systems (process-mediated data)**: these processes record and monitor business events of interest, such as registering a customer, manufacturing a product, taking an order, etc. The process-mediated data thus collected is highly structured and includes transactions, reference tables and relationships, as well as the metadata that sets its context. Traditional business data is the vast majority of what IT managed and processed, in both operational and BI systems. Usually structured and stored in relational database systems. (Some sources belonging to this class may fall into the category of “Administrative data”).
   - 21. Data produced by Public Agencies
     - 2110. Medical records
   - 22. Data produced by businesses
     - 2210. Commercial transactions
     - 2220. Banking/stock records
     - 2230. E-commerce
     - 2240. Credit cards

3. **Internet of Things (machine-generated data)**: derived from the phenomenal growth in the number of sensors and machines used to measure and record the events and situations in the physical world. The output of these sensors is machine-generated data, and from simple sensor records to complex computer logs, it is well structured. As sensors proliferate and data volumes grow, it is becoming an increasingly important component of the information stored and processed by many businesses. Its well-structured nature is suitable for computer processing, but its size and speed is beyond traditional approaches.
   - 31. Data from sensors
     - 311. Fixed sensors
       - 3111. Home automation
     - 3112. Weather/pollution sensors
     - 3113. Traffic sensors/webcam
     - 3114. Scientific sensors
     - 3115. Security/surveillance videos/images
   - 312. Mobile sensors (tracking)
     - 3121. Mobile phone location
     - 3122. Cars
3123. Satellite images
32. Data from computer systems
   3210. Logs
   3220. Web logs