Evolution of Big Data

Big data and technology can be regarded in stages of evolution. The first is Big Data 1.0, which is the stage at which many organizations are now. The next stage is Big Data 2.0. Some organizations are already at this stage. There will certainly be future stages as the technology continues to evolve.

Big Data 1.0

In Big Data 1.0, organizations began using the Internet to conduct business and compile data about their customers. Insurers developed online insurance applications. They also used the data from applications, as well as claims history, to improve underwriting efficiency and to provide customer information for product development and marketing.

Other industries were performing similar functions in Big Data 1.0. For example, retail investment firms, such as E*Trade, could use the Internet to provide new products to customers. Computers could place stock and bond trades much more efficiently than previous methods.

Big Data 2.0

Because of rapid advances in data science, we have reached the stage of Big Data 2.0. This stage allows organizations to obtain and aggregate vast amounts of data very quickly and extract useful knowledge from it.

For example, investment firms have computers that can scan the Internet instantly for news and information about products, prices, economic and geopolitical developments, and consumer trends. This data is then provided to computer trading algorithms that conduct automated trading of stocks and bonds at high speeds. Many people in the financial industry believe that computer trading has completely altered the world's major stock markets. There are questions regarding whether traditional forms of investing that individual investors use, such as selecting particular stocks, can be effective in the new trading environment of Big Data 2.0.

Big Data 2.0 also allows organizations to process and analyze data from sources such as vehicles, homes, and wearable technology. Sensors can be built into products, and data science provides methods to process and analyze the data that comes from those sensors. Big Data 2.0 is evolving quickly. Some insurers are already actively involved in this phase, while others are lagging. It is not an exaggeration to say that insurers that can adapt quickly and effectively to the convergence of big data and technology will have significant competitive advantages.

BIG DATA CHARACTERISTICS AND SOURCES

Insurers' traditional data is organized into databases with defined fields. Insurance professionals can produce reports that show results from this data.

Risk managers can usually access their insurers' data to produce reports about claims they are handling. Big data has introduced different types and sources of data than those traditionally used.

It is important for risk management and insurance professionals to understand the different types of big data. The varieties, volume, and sources of data are rapidly increasing. For example, risk managers can obtain new types of data on safety from sources such as sensors on employees. Underwriters can obtain new data on risks from drones and social media. Claims adjusters can better identify fraud by identifying patterns in internal and external data. Risk management and insurance professionals increasingly work with data scientists to determine the types of data that are useful to make business decisions.

To better understand big data, these categories will be discussed:

- Data characteristics
- Internal and external data
- Structured and unstructured data

Data Characteristics

The term "big data" implies large quantities of data. Although this is true, big data is also different from traditional data in other respects.

These are characteristics that differentiate big data from traditional data:

- Volume—There is an enormous amount of data that is now available, and the amount continues to increase. To use an analogy, traditional data is like the planet earth, and big data is like the solar system. Eventually, big data could increase to the infinite size of the universe.
- Variety—Traditionally, insurers used **structured data**. Big data also includes structured data in larger volume than traditional data. However, because big data comes from multiple sources, much of it is **unstructured data**.
- Velocity—This is the constantly increasing speed at which data arrives at an insurer. Velocity also includes the growing rate of change in the types of data.
- Veracity—This refers to the completeness and accuracy of data. Unstructured big data is more likely to have less veracity than structured data. However, even traditional structured data will not be perfect. Also, insurers can often gain useful information from big data, even if it has lower veracity.
- Value—Value is derived from the results of data analysis to help insurers make better business decisions. Big data has great potential to add value, but it must be obtained and analyzed with techniques that provide meaningful results. This is the goal of data science.

Structured data

Data organized into databases with defined fields, including links between databases.

Unstructured data

Data that is not organized into predetermined formats, such as databases, and often consists of text, images, or other nontraditional media.

Internal and External Data

Data science allows insurers access to increasingly larger and more varied data, referred to as big data. Some of this data is received directly by the insurer or the risk manager's organization. Other data is obtained from outside sources.

Internal Data

Insurers have always relied on data, and they possess large quantities of it. Certain types of data, such as risk factors, losses, premium, rating factors, rates, and customer information, have been traditionally used to make business decisions. Much of this data is also reported to state rating bureaus. Similarly, risk managers also have relied on data about losses and premium, in addition to safety statistics that are reported to the Occupational Safety and Health Administration (OSHA).

Internal data
Data that is owned by an organization.

However, risk managers and insurers have vast quantities of **internal data** that they have not used because of its volume and/or a lack of techniques to access it. For example, an insurer may not analyze computer claims codes about claimants' preexisting medical conditions—traditional internal data that could be useful if data science techniques are applied to it.

Data science also provides techniques to use nontraditional internal data. For example, voice analysis can be applied to recorded statements of claimants to identify vocal characteristics associated with fraud, such as gaps in the claimant's version of the accident or a defensive tone. Underwriters often obtain photos of property, and claims adjusters often obtain photos of an accident scene. Artificial intelligence can analyze these photos to find information that may be missed by a human analysis, such as potential environmental hazards around a property or, through facial recognition, the identity of a witness to an accident.

External Data

External data
Data that belongs to an entity other than the organization that wishes to acquire and use it.

There is sometimes a blurring of the boundary between internal and **external data**. For example, data from telematics is obtained from a device that is installed on a customer's vehicle. The insurer owns the device, but the customer owns the vehicle. Because of the customer's ownership of the vehicle and the nature of the data provided—the customer's personal driving habits—telematics should probably be categorized as external data.

Another source of data with a blurred boundary is a wearable sensor used by an employee while working. The employer provides the sensor, but it is placed on the employee's person and transmits information about the individual. This information could be considered internal data because it is used in the workplace and because the sensor is owned by the employer. However, it could also be considered external data because it is obtained from and about an individual.