
Financial Market Data

We now turn our attention to the topic of financial market data. This term can refer to any type of data generated by the systems described earlier but generally, the focus is on publicly available data or data available for educational and research purposes.

Introduction

Financial market data is a source of knowledge about trading behaviour, strategies and patterns; market design, operations and efficiency; products being traded; and financial news. Financial market is a source of information about a number of entities which can be from the Finance or Electronic Market domains (Fig. 3).

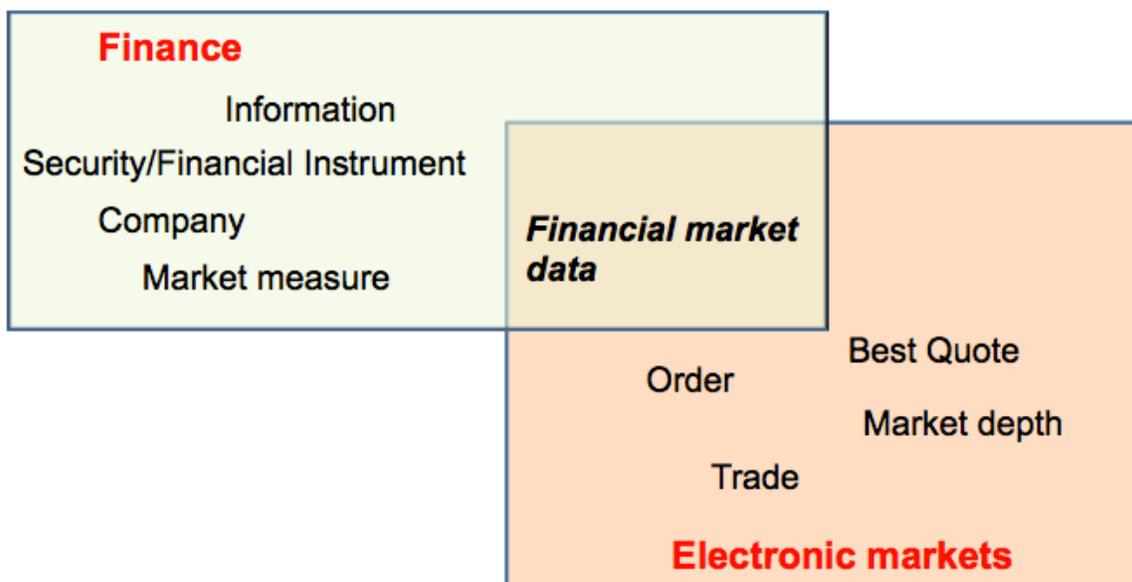


Fig. 3 Financial Market Data

Examples of finance entities include:

- Company: this entity can be modelled using many attributes such as the Name, Country, Date listed etc. There are many ways companies can be structured and an adequate representation is a challenge particularly when dealing with multiple jurisdictions.
- Security/Financial Instrument: depending on the type of financial instrument (e.g. equity, option, future, index, interest rate...), there are many attributes such as underlying company,

expiry date etc. One example of a standard dedicated to the representation of derivative financial products is FpML (<http://www.fpml.org/>).

- Market Measure: this entity represents the metrics used to measure market activities such as returns, volatility and liquidity.
- Information: this entity represents any textual information that affects the market activities in some way such as corporate announcements and general news.

Market entities consist of the following entities:

- Order: this represents both Buy/Sell orders sent to the market. Buy orders are referred to bids and Sell orders are referred to asks. All orders have common attributes such as market participant ID, product, price and quantity.
- Trade: this represents the occurrence of a trade with attributes such as buyer ID, seller ID, product, price and quantity
- Best quote: this represents the best price (on both sides of the orderbook) for a particular product and the available quantity
- Market depth: this is similar to the previous one but all quotes are given up to a specific depth (e.g. 10 best bids and best asks)

In addition, market data has a time dimension as it is used to record the evolution to these entities through time (creation, modification, deletion). We choose to represent such evolutions via the event concept. For example, consider the following records from a financial market dataset:

BHP trade at \$2.00 at 11:56:09.876: this is a record that an event occurred, which is a trade for BHP shares at \$2.0 at a specific time

Dividends for BHP issued on 15 March 2009: this is a record of an event, which is a corporate action (dividends issuing) at a specific date

There are many standards for representing financial market data which differ a lot depending on the entities involved. For example, the FIX protocol (<http://www.fixtradingcommunity.org/>) focuses mainly on the message format around market activities. Also, the Market Data Definition Language (<http://www.mddl.org/>) has been proposed as a way to standardise financial market data formats.

Financial Market Data Providers

Financial market data is available from a number of providers. There are two main types of providers depending on whether real-time or historical data is distributed to users. Pricing of market data is usually depending on how fast it is distributed. Market prices are the core focus for most market data providers but increasingly, other type of data like news are gaining in popularity. Each provider can maintain multiple types of datasets. Here are some examples of financial market data providers:

Thomson Reuters

Formed by the merger between Thomson Corporation and Reuters Group PLC in 2008, they recently commanded 33.4% of the global market share for market data providers. Thomson Corporation can trace its history back to 1799 with the establishment of the Sweet & Maxwell legal publishing firm in London. Reuters Group PLC began in London in 1851 when they delivered stock prices and news to clients using a combination of telegraph cables and carrier pigeons. They quickly achieved a reputation for the quality and impartiality of their service. In 1964 Reuters began transmitting market quotations by computer, and in 1971 their journalists began writing and transmitting news with computers. In 1987 Reuters launched their Integrated Data Network (IDN) to deliver prices and news at high speed around the world.

Their business strategy is to act as a wholesaler and allow their customers to purchase licenses to use their software and access data from a variety of different sources. They currently employ roughly 2,300 journalists across the world that generate news content for their services, and maintain a reputation for unbiased journalism. Furthermore, they deliver content generated by other newswires including PR Newswire since 1997, Business Wire and Dow Jones since 1999. They also transmit regulatory announcements from the regulators covered in this document, including the RNS and ASX since at least 1996 and Edgar Online since 2001.

Thomson Reuters' main core business is in the delivery of real-time data via its terminals (e.g. Eikon) although the hardware part is playing a lesser role than previously. Thomson Reuters also maintains a number of databases covering different aspects of financial market data.

Bloomberg

Founded in New York in 1981 they recently commanded a 27.7% of the global market share for market data providers. As the main competitor for Thomson Reuters, their business strategy is to lease terminals to their clients which receive content transmitted by Bloomberg. Their news service was launched in 1990, and currently employs over 2,300 news professionals in 152 locations worldwide. They also deliver content generated by other newswires, including PR Newswire and Business Wire.

Sirca

Sirca (www.sirca.org.au) which is an example of a historical market data provider was founded in 1997 by a group of collaborating Australian and New Zealand universities as a not for profit company. Sirca was created to support the needs of academic researchers, in a world where data volumes were accelerating dramatically. Sirca provides its members and customers with access to vast and comprehensive online repositories of global news and financial markets data.

WRDS

Wharton Research Data Services (WRDS see <http://wrds-web.wharton.upenn.edu/wrds/>) is a web-based business data research service from The Wharton School at the University of Pennsylvania. Developed in 1993 to support faculty research at Wharton, the service has evolved to become a common tool for research for over 290 institutions around the world. WRDS provides access to mostly US data

(historical), including COMPUSTAT, CRSP, IBES, NYSE-TAQ, Bureau van Dijk, Global Insight, OptionMetrics and other important business research databases.

Characteristics of a Financial Market Dataset

We have seen that there are many providers that offer different types of datasets. In many cases, data is duplicated between datasets from the same providers or across multiple providers. For example, Trades and Quotes data for New York Stock exchange can be obtained from TAQ Database, Thomson Reuters and Bloomberg at different levels of detail. To distinguish between different financial market datasets, we define the following characteristics for a financial market dataset:

Instrument naming scheme

The instrument naming scheme is used to identify tradable instruments in a unique way publicly. It varies among providers (i.e. different data providers have different naming scheme). The scheme can be exchange-specific, vendor-specific or abide by some certain international standard (e.g. ISIN). An example of a vendor specific naming scheme is the Reuters Instrument Code (RIC).

Events

A financial market data is characterized by the type of events it supports. The event type determines the attributes in the data. For instance, in daily stock data, the type "End Of Day" determines that the attributes of the data include open price, close price, high price, low price, volume, etc.

Coverage

For each provider, the data provided is covered for a particular period. Also, the frequency of data varies. In this sense, financial market data can be classified into two types:

- High-frequency (also known as intraday): data collected at a frequency smaller than 24 hours (within a single day), such as bids, asks and trades (often in discrete intervals) observed within a day.
- Low-frequency (daily, weekly, monthly, etc.): data summarised for a long period (e.g. a day, a week, etc.) such as daily stock data.

Instrument classes

Each dataset specialises in providing data about one or different types of instrument classes e.g.

- Equities / Derivatives / Fixed Income
- Indices / Exchange Rates / Interest Rates
- Exchange / OTC data