

Data plausibility checks under time pressure

Workshop Hackathon / Setting

Armin Bürgi

Schiffbau, 20 March 2015

Classification: public



Agenda



Goal



Background



Challenge



Hackathon Setting



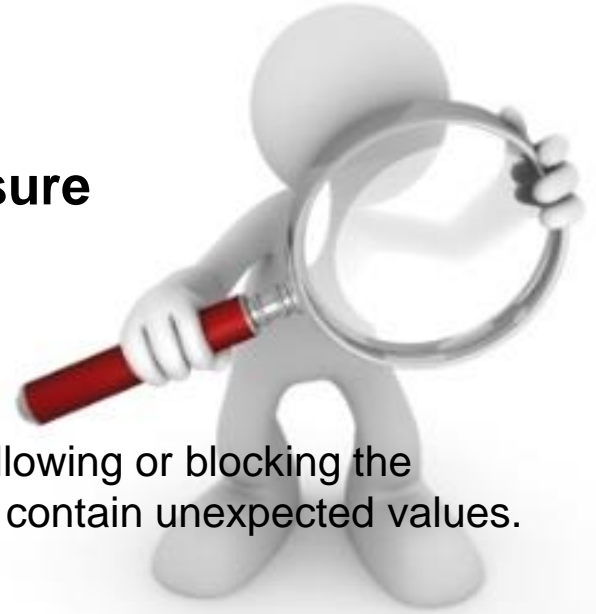
Possible results

Goal



Data plausibility checks under time pressure

Achieving an optimal human-computer interaction for allowing or blocking the publication of real-time financial market data feeds that contain unexpected values.

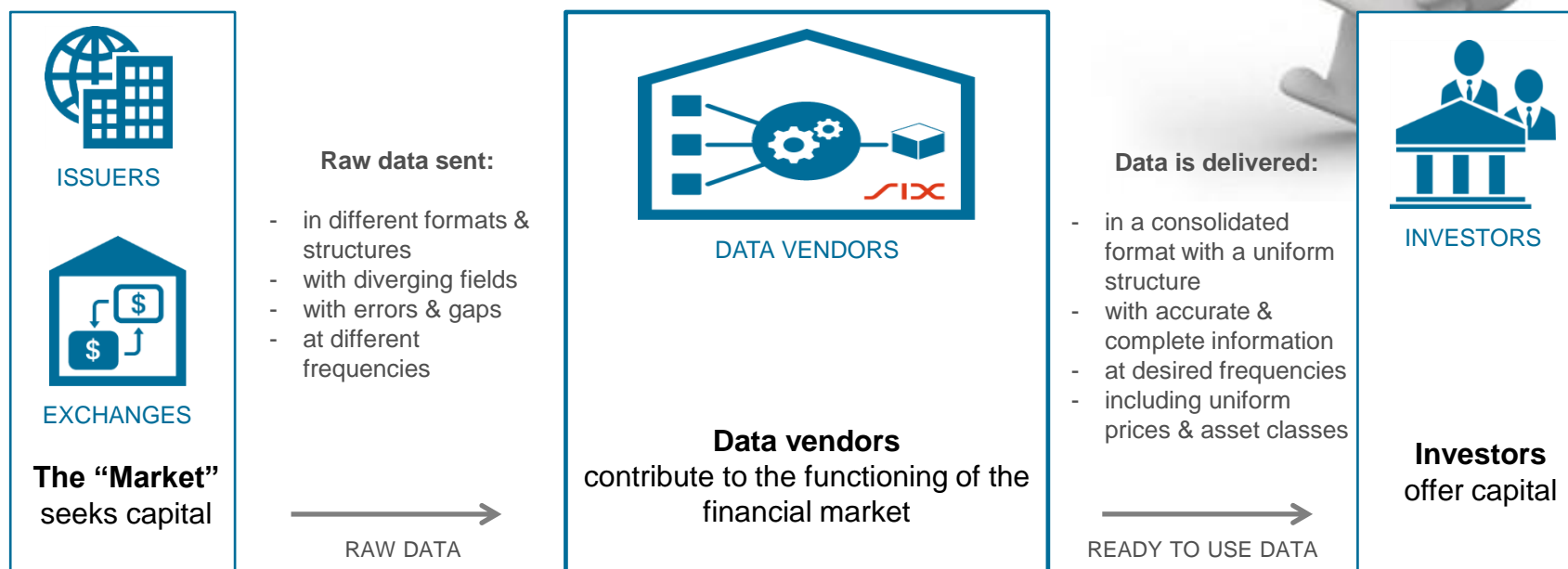


Data plausibility checks under time pressure



SIX Financial Information is a leading global provider of data services and solutions for financial and insurance firms, corporations and the media.

SIX Financial Information maintains a real-time database of global market data sourced from more than 1'500 securities exchanges and contributors. From individual wealth managers to major media organizations, our customer base use this information to gain a comprehensive view of global markets.





What is the background ?

The financial markets are never standing still. SIX Financial Information gathers traded prices, order volumes, and other market data about trading activities worldwide, around the clock and from 1'500 sources like exchanges, banks and others. In a highly automated process, SIX Financial Information streams this data on a real-time basis to market participants (banks, investors, media), that use this real-time data for trading decisions, market monitoring, investment valuations, risk management and much more.

But occasionally data is received that does not match range of expected values. There can be sudden market disruptions, like market crashes, central bank decisions with large market influences, currency devaluations, but also system failures, erroneous data entries, data format changes and other technical issues.

But of course the data receivers expect correct, complete, consistent data, and they expect it real-time, continuously streamed to them.

Challenge (1)



What is the challenge (1) ?

The SIX data monitoring team is quality-checking this continuous data stream for unexpected values (like a suddenly negative interest rate) and data outliers (like the drop of the EUR/CHF exchange rate from 1.20 to 1.02). Once such an event or questionable data series is detected, the financial market experts are under time pressure to make decisions if this is plausible data (an actual market movement) and the data needs to be processed without interruption; or if the data must be considered wrong (due to some technical failure or human error) and needs to be blocked from further processing.

Two goals are in conflict in this situation:

- A quick decision is needed to minimise the interruption of the real-time data streaming
- A correct decision is needed to protect SIX Financial Information' reputation for high quality data



Challenge (2)



What is the challenge (2) ?

SIX data monitoring experts check the plausibility of such suspicious data using their knowledge about the financial markets and the continuously incoming information like:

- Directly related market data, e.g. if the EUR/CHF exchange rate looks suspicious, compare with the movement of the USD/CHF exchange rate
- Related news, e.g. are there related messages on the international news ticker that can confirm the plausibility of a large market movement, like a Swiss National Bank announcement
- Related meta data, e.g. did the trading activity increase, did the bid-ask spread increase, is the suspicious data coming from one source, or from multiple sources with a good quality track record
- Various further data and information (will be ignored for this hackathon), e.g. known possible causes of market data disruptions.



Challenge (3)



What is the challenge (3) ?

The plausible vs. erroneous data decision can be done quicker and with higher accuracy, the faster the user is presented with the related information, the more relevant and complete it is, the easier it is made for him to understand the displayed information (e.g. graphs), how easy it is for him to look into further data, and finally how easy it is made to allow or block data with minimal effort.



Hackathon Setting



For this Hackathon, we will focus on FOREX (Foreign exchange rates) only. Also, only a subset of a few instruments is given. Please imagine, that SIX Financial Information receives about 8 billion quote telegrams daily. Therefore, about 50'000 quote telegrams daily are suspicious.

You will be given

- A CSV file of time-stamped data feed values
- Time-stamped News feed messages

allowing to simulate a real-time financial data streaming situation. Imagine yourself tasked with making the allow/block decisions on the suspicious data (as defined by SIX), as the data is continuously flowing in and you receive data exception alerts.



Possible Hackathon results (1)



A) User Interface

Design a user interface (working prototype or mock-up, incl. Excel) suitable for data monitoring experts,

- informing them of the (pre-defined) data exceptions
- showing them the relevant data, charts and news to resolve each exception alert
- minimising their time to understand and evaluate the displayed information
- giving them the possibility to make further data drill-downs (but without losing the overview in this stressful situation) and,
- giving them the possibility to manually allow or block single data point or whole data sets from publication.

Consider how your user interface is continually optimised as new data is continually coming in, potentially new exceptions appear, and the user makes his plausibility decisions under time pressure.

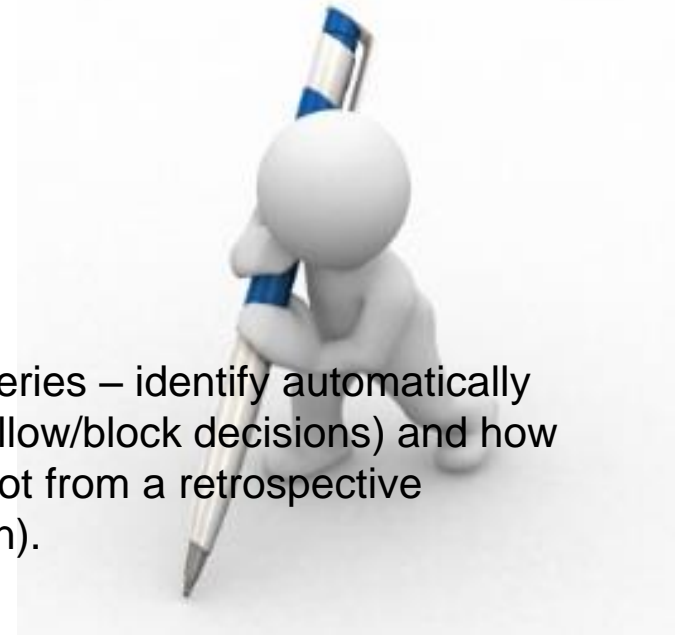


Possible Hackathon results (2)



B) Identify the relevant data

Define algorithms that – based on the provided data series – identify automatically the possibly most relevant related data (to make the allow/block decisions) and how the user is informed or can interact with it (of course not from a retrospective viewpoint, but within the real-time streaming simulation).

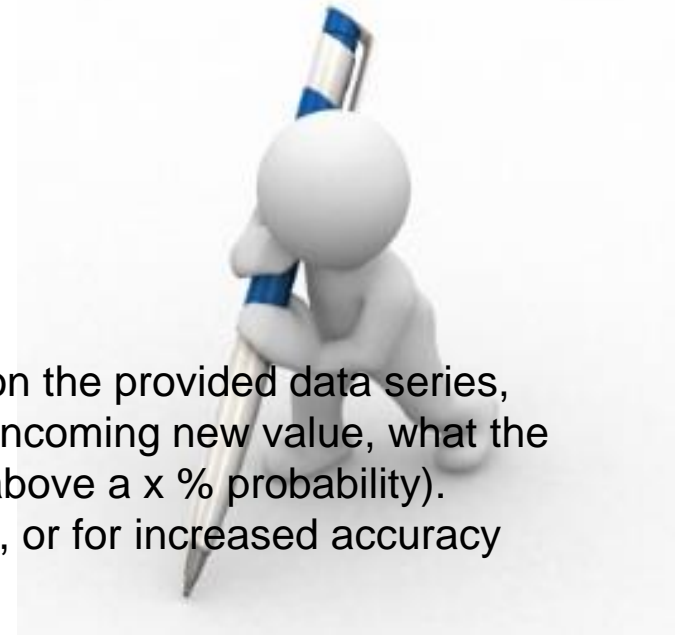


Possible Hackathon results (3)



C) Identify the exceptions your way

Ignore the “exception” flags that SIX has pre-defined on the provided data series, and program your own algorithms to identify for each incoming new value, what the probability is that the value is correct (i.e. plausible if above a x % probability). Consider either the data of each individual data series, or for increased accuracy consider the possibility related & relevant data series.

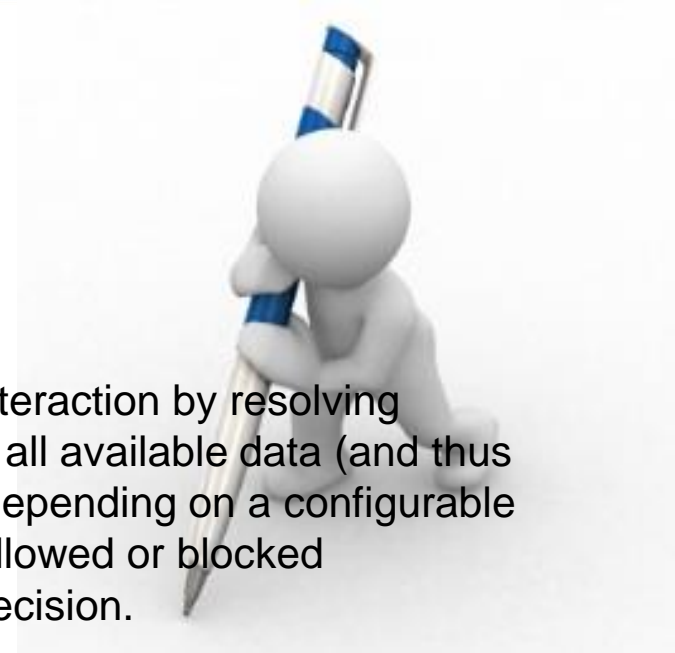


Possible Hackathon results (4)



D) Automate the allow/block decisions

Reduce the number of exceptions that need human interaction by resolving exceptions automatically by your algorithms based on all available data (and thus reduce the time delay). Your algorithm might decide, depending on a configurable probability, if incoming exceptional values should be allowed or blocked automatically, or presented to the user for a manual decision.



E) Any crazy idea regarding Data plausibility checks is welcome



Financial Information

Thank you for your attention.

