

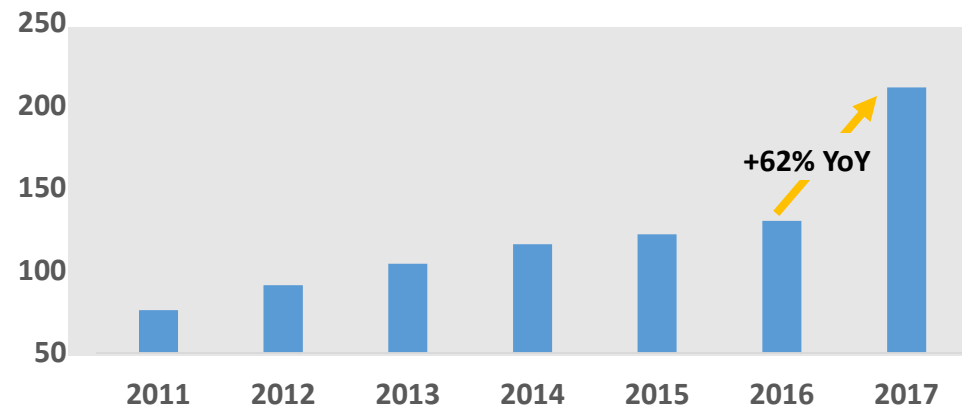
# Alpha from Alt Data

**Philip Higson**  
**Marius Müller**  
December 2017

Carlyon AG

# Growing volume of alt data for investment alpha capture

Number of providers

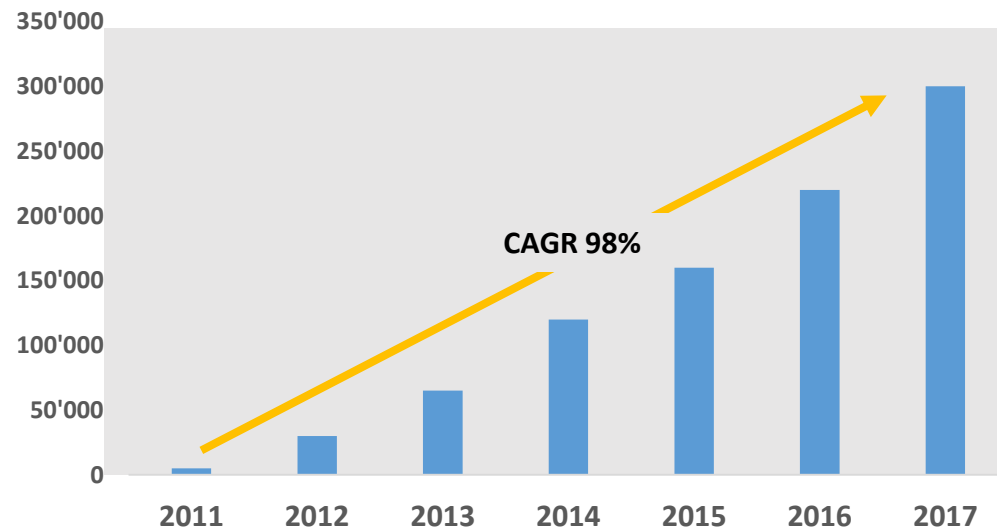


Significant increase in alternative data vendors due to improved natural language processing\*, increased demand from financial firms for new datasets and over \$1bn venture capital funding to start-up data providers. (see Appendix)

\*natural language processing (NLP): a field of Artificial Intelligence which enables computers to analyze and understand the human language. Natural Language Processing (NLP) was formulated to build software that generates and understand natural languages so that a user can have natural conversations with his computer instead of through programming or artificial languages like Java or C.

Source: Investopedia

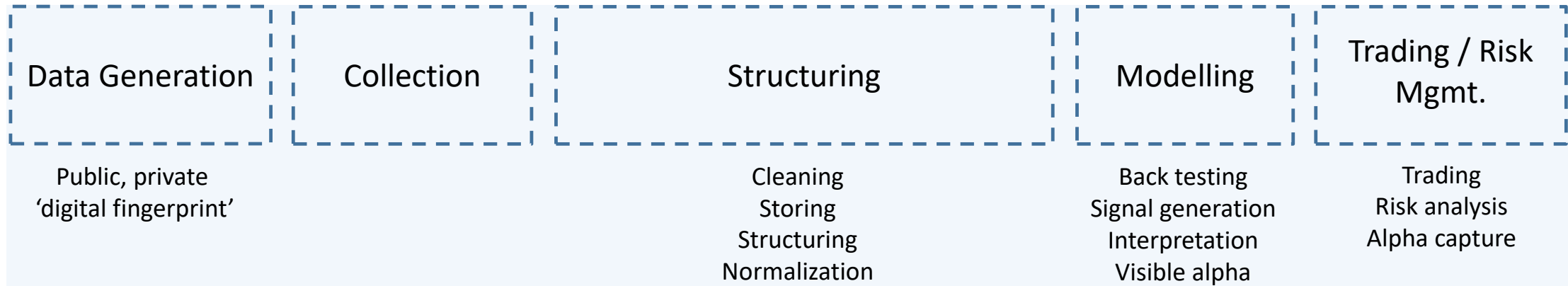
Number of Quandl users



Quandl data	2016	2017	change
Alt data publishers	5	30	500%
Data requests / day (mn)	3.2	4	25%
Number of users	220'000	300'000	36%

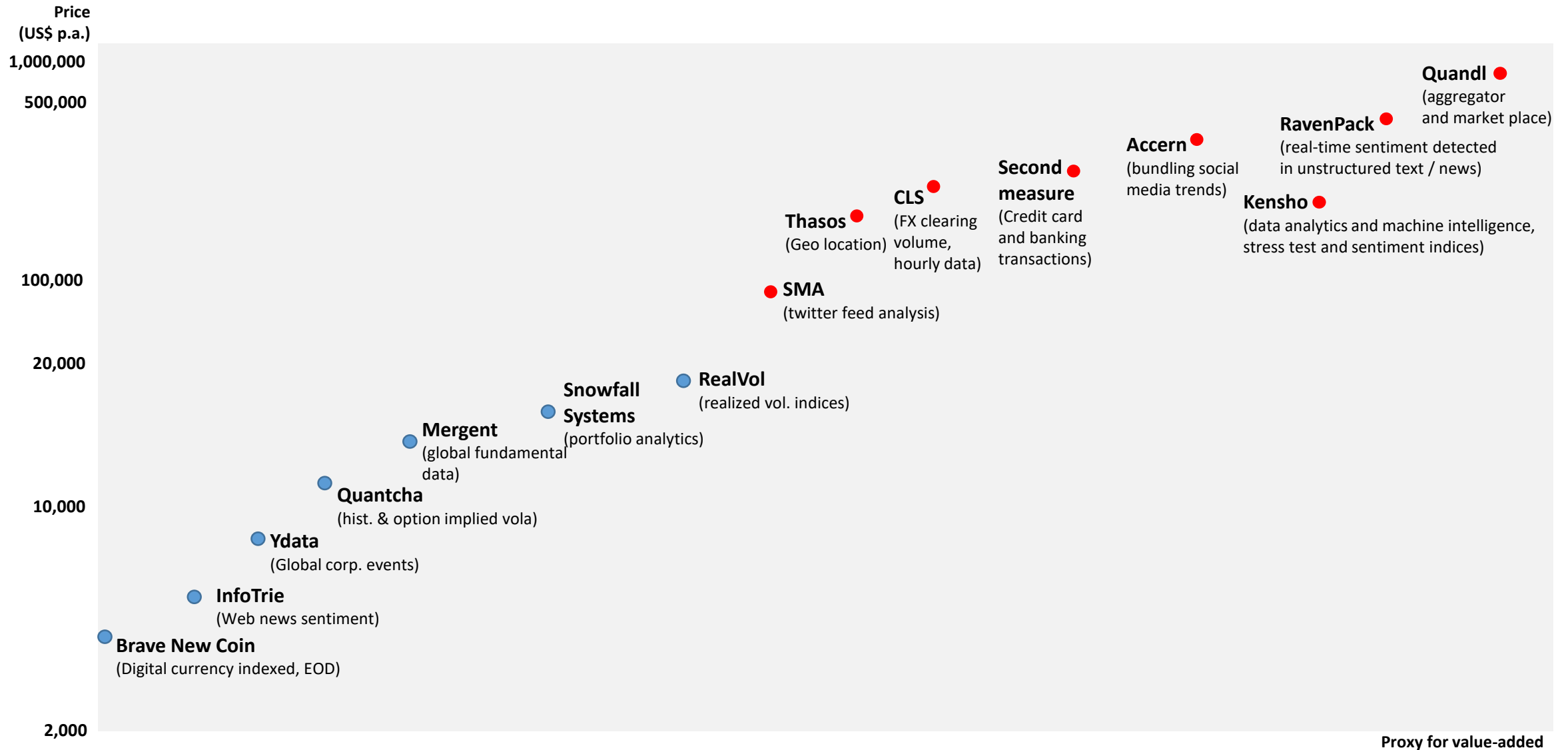
Source: Quandl

# Value chain














# Subscription fees as proxy for value-added

Variables: frequency, granularity, alpha capture, degree of customisation, number of subscriptions, level of efficiency / cost savings












● Conducted telephone interview with representatives. Additional enquiry should include Dataminr, Orbital Insights, Trifacta, Spire, Descartes Labs, Affinity Solutions

# Equity/Credit strategies in Consumer sector



Category	Tool set	Use cases
Geo-location	 Location data is being derived from mobile app owners, Bluetooth connections and sensors (e.g. retailers, restaurants, hotel, travel, transportation)	Dataset by Advan, derived from apps that track geo-location, can be used to proxy the top-line revenue of publicly traded firms. Advan noted traffic at Lululemon (LULU) stores were up 7% and predicted sales to beat the street consensus. Sales surprise few days later boosted stock from \$58 to \$67 intraday
Pricing of goods and services	 Crawling online shops, collecting data via panel, etc.	In Q4-2014 the China Auto Insight data (collected using a large panel of dealerships throughout China) was more accurate than consensus at predicting revenue growth
Consumer Credit marketplaces	 Amount of loan issuance, loan pricing, borrower quality and default levels	Testing of a data basket (crowd-sourced data from vendor's subscribers, credit managers and supply chain professionals) on credit risk data showed that the probability of bankruptcy within 12 month ranges from 10% to 50% when a company gets a "stress score" of 1
Store Locations	 Track store hours and product promotions	Correlation with retail sales data
(Point of sales) consumer transactions	 Payment processing data (credit card, PayPal, Square) provides insights to merchant level transaction data (retailer, airline) and product level purchase data (beverages, automobiles)	Cross-sectional analysis of anonymized purchase data set of 2mn active shoppers scanned from email purchase receipts generated annualized return of 16.2% and a SR of 1.13
Advertising (proxy of popularity)	 Consumer's internet browser habits	Correlation with retail sales
Online search	 Usually, data derived from a basket of terms, are more efficient than single term searches. Google, Bing, etc.	A search indicator (derived from Google search) predicted weakness for Finish Line (FL) in same store sales ahead of Dec. 2016 earnings.
Sentiment and consumer trends from news feeds and social media	 Various Social media channels, mainly twitter; May be applied to investor commentary, consumer attitudes towards products, brands	A social media dataset (enriched with web queries analytics on over 90mn web sources) was used to highlight that the Overwatch game was well positioned to set new sales record for Activision Blizzard
Web crawled company strategy insights	 Mine the deep web, track website changes, e-commerce activity, public commentary	Correlation with company revenues
Business Insights	 Changes of corporate websites, new services, changes in mid-level management	Correlation with company revenues
Event Detection	 Breaking news from news wires or social media, government filings, weather and changes in residence allow traders to react ahead the broad public community and before news are fully incorporated in asset prices	Strong benefit from combining Raven Pack Analytics (RPA) and Barra risk model. Cleaner, less volatile return profiles resulting in improvement in risk-adjusted returns

# Equity/Credit Strategies in Info Tech / Telco / Energy / Industrials / Materials / Financials

Sector	Category	Tool set	Use cases
Information Technology & Telco	App usage, app reviews & web traffic (as proxy for product adoption) 	Number of users, downloads	Data on installations of apps showed an early indicator of positive inflection in revenue growth for HBO, one of Time Warner's largest division
Information Technology & Telco	Reviews & Ratings 	Product and service Reviews & Ratings, B2B opinions	Correlation with product/service revenues
Energy	B2B 	B2B/enterprise data (Alibaba's B2B trade index, databases of oil contract and drilling concessions)	Correlation with company revenues
Energy	ESG profile of companies 	business complaints, business reputation	Correlation with company revenues
Industrials / Materials	Open data 	e.g. e-vehicle charging stations, vehicle tracking	Correlation with company revenues
Industrials / Materials	Satellite 	Satellite images (Economic activity, agricultural productivity analysis, mining, construction and real estate, shipping, oil & gas)	Using a RS metrics dataset, a traffic growth indicator (fill rate across all parking lots for a given company) generated a negative outlook 3 times for Ross Stores in Q1-16. In line with predictions, ROST reported lower than expected revenues
Financials	Subject matter experts 	Expert tweets, commentary and views, which might differ significantly from trends observed in general population	Additional source of news/changing sentiment for individual corporate products etc.
All sectors	Employment (to evaluate corp. strategy) 	Job platforms, corporate websites	Facilitates analysis of corporate growth strategy
All sectors	Internet of things (Connected devices) 	Data which tracks digital footprint of IoT activity	Raw data for correlation across all sectors and fundamental strategies





# Macro strategies and additional use cases

## Macro strategies

Category	Tool set	Use cases
Public sector and trade datasets 	Data sets published by government agencies (e.g. customer complaints relating to financial services companies, database for tender calls, import/export statistics, central banks conference transcripts)	Statistical back testing on an international trade and industrial production dataset has proven that indicators are frequently a better predictor than street-mean estimates, while also having the advantage of substantial lead time
Online search 	Usually, data derived from a basket of terms, are more efficient than single term searches. Google, Bing, etc.	US Unemployment index derived from google search has a 5-year correlation of 0.9 with the US Unemployment rate, with an out-of-sample prediction improvement of 14% compared with baseline ARIMA (autoregressive moving average) model







Source: eagle alpha, Alternative Data: Applications & Case Studies

## Additional use cases derived from conducted telephone interviews

Category	Tool set	Use cases
GPS systems installed in tractors 	GPS-based augmentation systems, geo-location, other technology	Devices in John Deere's tractor puts the company into the position to be one of the best informed sources on the accurate state of the US agriculture harvesting situation
Global ship tracking intelligence 	AIS (Automatic identification system) transmitter installed in vessels	Tracking system allows to derive intelligence from monitoring vessel movements
Intelligence derived from smart home equipment 	Logitech produces a range of accessories, which are plugged into computers or other devices (TV, etc.)	The information, Logitech receives from its client base, allows to derive intelligence on market shares in television, computers, etc.
Information contained in insurance datasets generates lead-time 	Derive insights from insurance policies data	Insurance company data allows to extrapolate new car sales from car-policies ahead of release of sales figures by car industry







Source: Carlyon AG, Primary Research

# Large hedge funds investigating new sources of alpha (1/2)

 <p><b>BRIDGEWATER</b> <b>Greg Jensen</b> Co-CIO</p>	<p>"everything (including social media data, real-time internet price data and search engine data) that is available (online) to track the economy on a day-to-day basis and to be rally on the pulse of what's going on from social media to real-time Internet prices, to model economic activity in what is effectively real-time." (Greg Jensen, Co-CIO)</p>
 <p><b>TWO SIGMA</b> <b>Andrew Janian</b>, CIO <b>Alfred Spector</b>, CTO</p>	<p>"We've been pushing the frontiers of Big Data since before it was a buzzword. To us, it's just data. We work with vast sets of it, from over a thousand diverse sources." Two Sigma will take over "The Bridge," a space on the new Cornell Tech campus on Roosevelt Island in New York, where engineers and entrepreneurs will work. Already hired hundreds of PhDs with machine learning background</p>
 <p><b>DE Shaw &amp; Co</b> <b>Alexander Wong</b> Head of Venture Capital</p>	<p>(In 2017) invested in Triton, a developer of financial data and analysis platform that uses tech to power data-driven investment decisions</p>
 <p><b>WINTON</b> <b>David Hand</b> Chief Scientific Adviser <b>Geoff Cross</b> Head of Data</p>	<p>Winton stated that it trawls "through credit card and social network data to forecast sales, analyzing rail car networks to measure the pulse of the US economy or...weather data to forecast crop yields"</p>
 <p><b>WORLDQUANT</b> <b>Glenn Dooley</b> Co-Head Data Strategy</p>	<p>Hiring at least 15 teams of quant managers for its Accelerator platform, offering them the right to keep their intellectual property. Has also opened more than 20 offices in 15 countries, including emerging markets like Russia and Romania, to find engineers. "everything will become more and more predictable from the perspective of data" (Igor Tulchinsky, Founder/CEO)</p>
 <p><b>ACADIAN</b> <b>Ryan Stever</b> Director of quant. global macro research</p>	<p>In 2017 signed a partnership with Microsoft using the company's search engine technology "Bing Predicts" to forecast macroeconomic data "With Bing search history, we know what consumers are searching for, and how that will relate to the future earnings of companies. When we think about what Bing has to offer, it's an enormous source of potential in terms of what it can predict in the financial markets"</p>



# Large hedge funds investigating new sources of alpha (2/2)

 <p>Point72 <b>Dave Loaiza</b> Chief Data Scientist <b>Matthew Granade</b> Head of Ventures</p>	<p>“It is a real change from how investing used to work...if you want to understand what is going on with McDonald’s, you are going to have to look at credit card transactions data, you are going to look at geo-location data, at app downloads and a handful of other things. And suddenly you are going to have a very robust picture of how McDonald’s is doing and you are not going to have to talk to McDonald’s about that.” (Matthew Granade, Head of Ventures)</p> <p>“It’s a big enough deal that we have a couple of full-time tech people wash the data [from personal information] ourselves”</p>
 <p><b>Scott Rickard</b> Chief Data Scientist</p>	<p>The firm is building a large centralised data organisation.</p> <p>“Our ability to leverage big data effectively in our investment process is critical to our success as a firm” (Ken Griffin, founder &amp; CEO)</p> <p>“We need strong, analytical thinkers who can work through big, complex data sets” (Justin Pinchback, head of talent strategy)</p>
 <p><b>Jared Lynn</b> Senior Social Media Analyst</p>	<p>We have an analytical approach, driven by in-depth fundamental research. Our multi-disciplinary team includes substantial experience in equity analysis combined with forensic accounting, behavioral intelligence, interview techniques and social media research.</p>
 <p><b>Eli Bernstein</b> Head of Data Architecture</p>	<p>In 2016, Tudor has invested several million dollars late last year in a pair of startups led by entrepreneur Gil Syswerda and has partnered with them to use machine learning in investment decisions (FeatureX et al.)</p> <p>In 2017, Tudor has invested in CargoMetrics, a Boston-based fund backed by Google chairman Eric Schmidt that crunches global shipping data and satellite images to find tradeable patterns.</p>
 <p><b>Anthony Ledford</b> Chief Scientist Man AHL</p>	<p>Man Group, which started the Oxford-Man research center on the university’s campus in 2007 and revamped it in August 2017 to focus on artificial intelligence, creating a hunting ground for recruits</p>
 <p><b>Matt Ober</b> Chief Data Scientist</p>	<p>“We have added data science to our toolkit for identifying interesting, uncorrelated opportunities.</p> <p>Over the past year [2016], we have developed an internal data science team and hired a full time practitioner of the dismal science of market strategy” (Q3-2017 statement)</p>

### **Bloomberg: Hedge Funds, Beware: You Might Be Relying on Illegal Data Feeds**

By Nico Grant

(Bloomberg) - 2017-12-07 18:14:03.54 GMT - Hedge funds are always looking for an edge. To get one, they're increasingly poring through big data that details consumer habits to try to spot trends.

But a former U.S. prosecutor who was involved in the government's landmark crackdown on Wall Street insider trading is raising a caution flag.

**Hedge funds might get in trouble if they're buying data feeds from companies that didn't get consumers' approval to pass on information like credit card transactions, online browsing and emailed receipts,** said Jonathan Streeter, who worked at the U.S. Attorney's Office for the Southern District of New York under Preet Bharara.

The U.S. Securities and Exchange Commission, in particular, could pursue enforcement cases, he said Wednesday at Newsweek's artificial and data science conference in New York.

#### **Violating Rights**

**"If the SEC finds you didn't do diligence and someone's privacy rights have been violated, they can bring an action against you,"** said Streeter, who's now a partner at Dechert LLP. **"They would love to make a case related to this because it's a new, growing space."**

He added that the SEC could take issue with hedge funds trading on non-public consumer information that isn't widely accessible. Streeter was part of a team that won a legal case that shook Wall Street: The 2011 conviction of former Galleon Group hedge fund manager Raj Rajaratnam.

SEC spokeswoman Judy Burns did not respond to a request for comment.

### **FT - knowledge boundaries set to blur insider trading laws**

Investors are already crunching big data to find so-called life patterns, for example in billions of credit card transactions or social media posts. The latest trend, however, is the pursuit of observational truths. A satellite supported with clever software can potentially see what high street shops people are going into. Are they leaving with one bag or three? How many cars is that locomotive pulling to the port in Karratha, Western Australia? What blend of iron ore does its colour indicate? **Soon it is likely that investors will know more about the fortunes of a business than a line manager relying on weekly or monthly reports. This changes everything.** For a start, the definition of insiders and outsiders becomes redundant. So, therefore, does the regulation.

**If knowledge boundaries are set to blur, insider trading laws need to be updated.** Reporting requirements have to change too. One answer is real-time reporting. Companies have to improve the data they have on themselves to a point where it matches the information gathered by outsiders. They must then publish it instantly. The nuclear option would be for regulators to make a clear distinction between data and analysis, then insisting that every scrap of information about a company is published as soon as it comes in — from insiders and outsiders alike.

Source: FT, Nov. 29<sup>th</sup> 2017 (see appendix)

# Disclaimer

---

The information in this presentation was compiled from sources believed to be reliable for informational purposes only.

The information contained herein is not intended to be a source of credit or investment advice with respect to the material presented, and the information and/or documents contained do not constitute investment advice by Carlyon AG.

All information herein should serve as a guideline, which you can use to create your own judgement.

We trust that you will review the information to reflect your own conclusion and believe that these may serve as a helpful platform for this endeavor.

Any and all information contained herein is not intended to constitute legal or financial services advice. You should not take, or refrain from taking action based on its content.

We do not guarantee the accuracy of this information or any results and further assume no liability in connection with this publication including any information contained herein. Moreover, this presentation cannot be further distributed to third parties without the accord of Carlyon AG.

This presentation is provided on a strictly private and confidential basis for information purposes only.

By attending or reading this presentation, you will be deemed to have agreed to the obligations and restrictions set out below:

Without the express prior written consent of Carlyon AG, the presentation and any information contained within it may not be (i) reproduced (in whole or in part), (ii) copied at any time, (iii) used for any purpose other than your own evaluation or (iv) provided to any other person.

This presentation does not constitute or form part of, and should not be construed as, an offer, invitation or inducement to purchase or subscribe for securities nor shall it or any part of it form the basis of, or be relied on in connection with, any contract or commitment whatsoever.

This presentation does not constitute either advice or a recommendation regarding any securities.

No representations or warranties, express or implied are given in, or in respect of, this presentation. To the fullest extent permitted by law in no circumstances will Carlyon AG, or any of its respective subsidiaries, shareholders, affiliates, representatives, partners, directors, officers, employees, advisers or agents be responsible or liable for any direct, indirect or consequential loss or loss of profit arising from the use of this presentation, its contents, its omissions, reliance on the information contained within it, or on opinions communicated in relation thereto or otherwise arising in connection therewith.

The information contained in this presentation has not been independently verified.

Recipients of this presentation are not to construe its contents, or any prior or subsequent communications from or with Carlyon AG or its representatives as investment, legal or tax advice.