



**BANK NEGARA MALAYSIA**  
CENTRAL BANK OF MALAYSIA



**EUROPEAN CENTRAL BANK**  
EUROSYSTEM

**Per Nymand-Andersen**  
Adviser, European Central Bank

# Technology transformation in central banking



## Big Data: Strategies and Advanced Analytics

*“Progress lies not in enhancing what is, but in advancing towards what will be” (Khalin Gibran)*

## Overview

**1 Pretty Big Data - Reflections for policy purposes**

**2 Embracing technology and Advanced Analytics**

**3 Discovery, collaboration and Partnership**

Ref.: **“Big data: The hunt for timely insights and decision certainty - Central banking reflections on the use of big data for policy purposes**, IFC working Paper No 14, 2016, Per Nymand-Andersen

**Disclaimer:** The opinions expressed in this presentation are not necessarily those of the European Central Bank (ECB) or the European System of Central Banks (ESCB)

# Pretty Big Data - The "4V"

**40 ZETTABYTES**  
[43 TRILLION GIGABYTES]  
of data will be created by 2020, an increase of 300 times from 2005.

**6 BILLION PEOPLE**  
have cell phones



## Volume SCALE OF DATA

It's estimated that **2.5 QUINTILLION BYTES** [2.3 TRILLION GIGABYTES] of data are created each day



Most companies in the U.S. have at least **100 TERABYTES** [100,000 GIGABYTES] of data stored

## The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what, exactly, is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: Volume, Velocity, Variety and Veracity.

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015, **4.4 MILLION IT JOBS** will be created globally to support big data, with 1.9 million in the United States.



As of 2011, the global size of data in healthcare was estimated to be

**150 EXABYTES**  
[161 BILLION GIGABYTES]



**30 BILLION PIECES OF CONTENT** are shared on Facebook every month



By 2014, it's anticipated there will be **420 MILLION WEARABLE, WIRELESS HEALTH MONITORS**

**4 BILLION+ HOURS OF VIDEO** are watched on YouTube each month



**400 MILLION TWEETS** are sent per day by about 200 million monthly active users

## Variety DIFFERENT FORMS OF DATA



The New York Stock Exchange captures **1 TB OF TRADE INFORMATION** during each trading session



## Velocity ANALYSIS OF STREAMING DATA

By 2016, it is projected there will be **18.9 BILLION NETWORK CONNECTIONS** - almost 2.5 connections per person on earth



Modern cars have close to **100 SENSORS** that monitor items such as fuel level and tire pressure



**1 IN 3 BUSINESS LEADERS** don't trust the information they use to make decisions



Poor data quality costs the US economy around **\$3.1 TRILLION A YEAR**



**27% OF RESPONDENTS**

in one survey were unsure of how much of their data was inaccurate

## Veracity UNCERTAINTY OF DATA

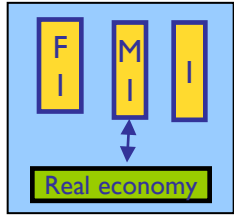
Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTec, SAS



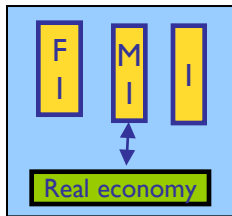
# Data sources for Monetary Policy & Financial Stability

Other systems ↔ Financial system ↔ Real economy

## Financial systems



## Financial systems



## Other sources

Policy events,  
fiscal policies,  
Global economy  
External shocks  
Sector failures

Transmission of central banking policies,  
managing expectations, feed-back loops

## Regulatory and oversight framework

### Financial Intermediaries (FI)

Banks  
Insurance,  
pension  
hedge funds,  
Rating agents,  
others

### Markets and instruments (MI)

Securities,  
Derivatives  
and other  
products

### Financial infrastructures (I)

Pre-trading,  
Trading,  
post-trading  
Infrastructures  
Clearing,  
settlements

### Raw sources:

Pre-trading, trading and post trading activities, actors, instruments, platforms, volumes, prices, frequencies, maturities, regulators and overseers alike

### Derived indicators:

Yield curves, density, liquidity, price measures,  
Functioning of markets, instruments and post trading activities  
Risks measures and contagions indicators  
Imbalances and concentrations measures  
Effects on performance (Collective behaviour, Interconnections networks, transmission of instability measures)

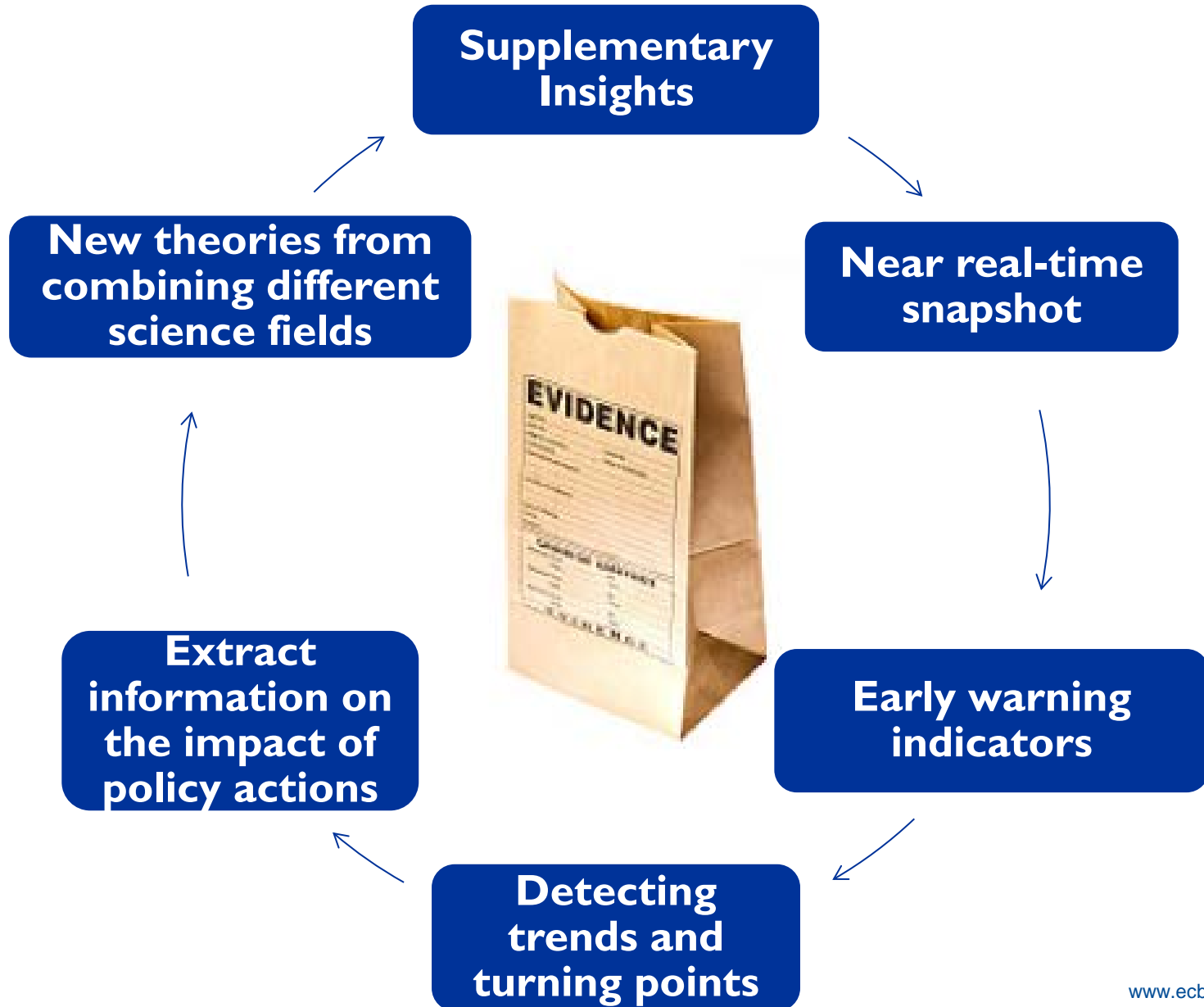
CORPORATE  
SECTOR

HOUSEHOLD  
SECTOR

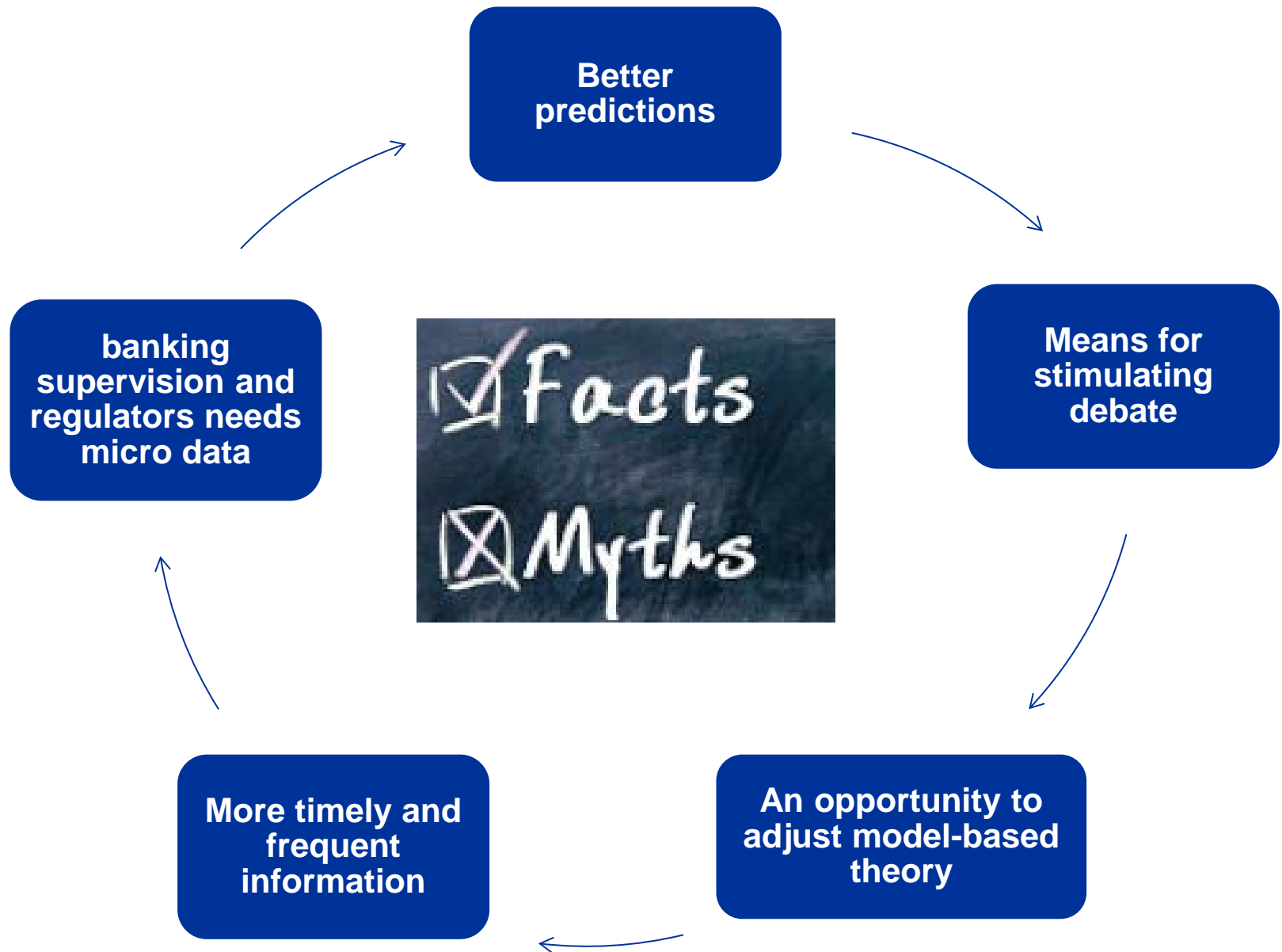
GOVERNMENT  
SECTOR

FOREIGN  
SECTOR

# Reflections on use for central banking policy purposes



# Reflections for central banking policy purposes



# Data mania versus phobia – a new paradigm of records

Digital transformation in finance and economics



E- trade



Settlement systems



Credit cards



Mobile trans



Lending & financing



Big data



- Fintech
- DLT
- D-coins
- S-contracts



Price scans



S- media



Data Scientist



Data Analytics



## Data lab

Systematic acquire,  
Process, summarize



Statistical algorithm and  
data explorations

Packaging data for  
Insights & business



# 4 building blocks of Data Analytics

Digital transformation in finance and economics

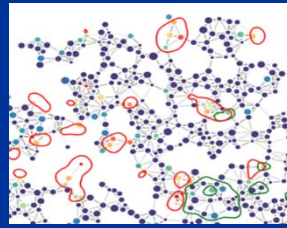


Sense



**Descriptive statistics analytics**

Respond



**Statistics pattern analytics**

Predict



**Now casting and forecasting analytics**  
**Statistical modelling**

Act



**Prescriptive statistics analytics**

# Conceptual framework for showcasing big data

## 5 steps on the way for exploring Big Data's relevance

Input

Quality

Production

Results

Assessment



- Source
- What info
- Coverage
- Access
- Formats
- Language
- Standards
- Free or cost



- Methodology
- Units
- Sample
- Representative
- Frequency
- Quality and updates
- Reliability
- Sustainable



- Hardware
- software
- Process
- Aggregations
- Testing
- Models
- Algorithm



- What can be produced
- Frequency and timelines
- Behaviour
- Purpose
- Feasibility



- Summary
- Purpose
- Production
- + / -
- Business model
- Costs
- Benefits
- Timetable

## 2 remarks & 2 lessons on Big Data and Analytics

- I. One misperception of big data is that we **do not need** to worry about **sample bias and representativeness**, as large volumes of information supersede standard sampling theory, since big data provide census-type information



Studies done on Corporate, consumer, household analysis & indexing, text mining, pulse of the economy, consumer emotions, stock market correlations.



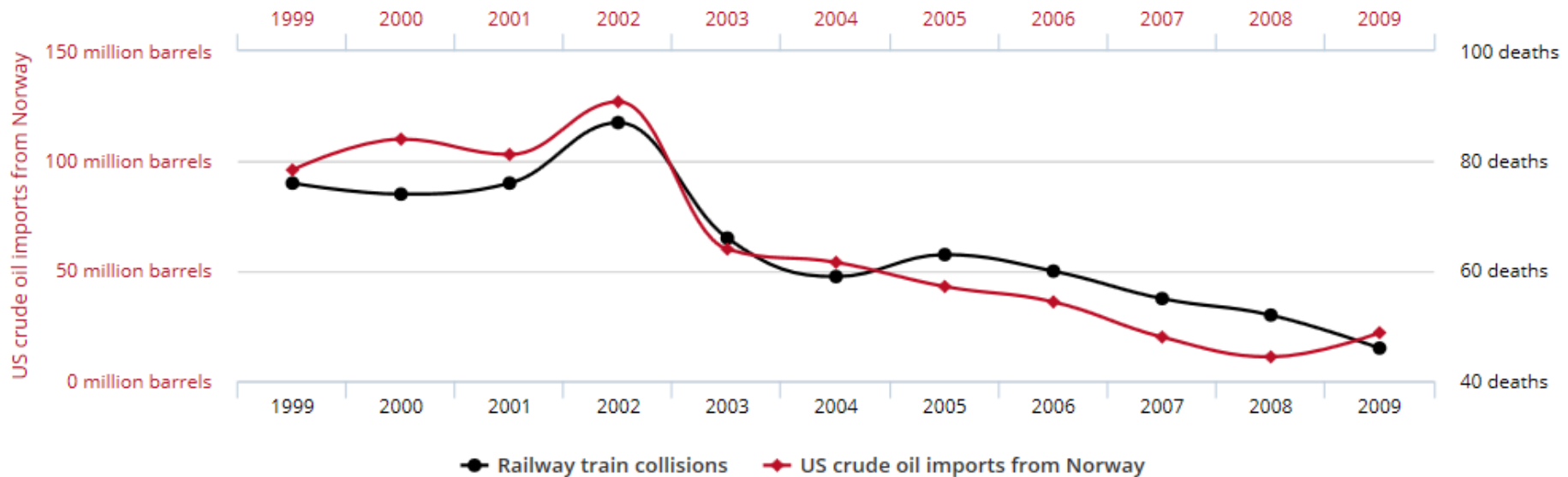
- Access to all tweets *means access to* **the characteristics** of the “tweeting” population, which **may differ** from those people/corporates who do not tweet
- Not all groups are represented. 21% of online adults use Twitter, **varies according** to age, gender, income, education, ethnic origin and country;
- Tweets vs **unit measurement**, double counting (tweeting and re-tweeting), over-representativeness, statistical corrections are needed
- An **event driven** source – **volume changes** do not necessary refer to reporting units nor to changes in demand

# Two remarks on Big Data and Analytics

## 2. Correlation is *not* (necessary) causation

No conclusion can be drawn simply on the basis of correlations between two variables. The similarity is a coincidence. We say that there is no causation

US crude oil imports from Norway  
correlates with  
Drivers killed in collision with railway train  
Correlation  $R= 0.95$



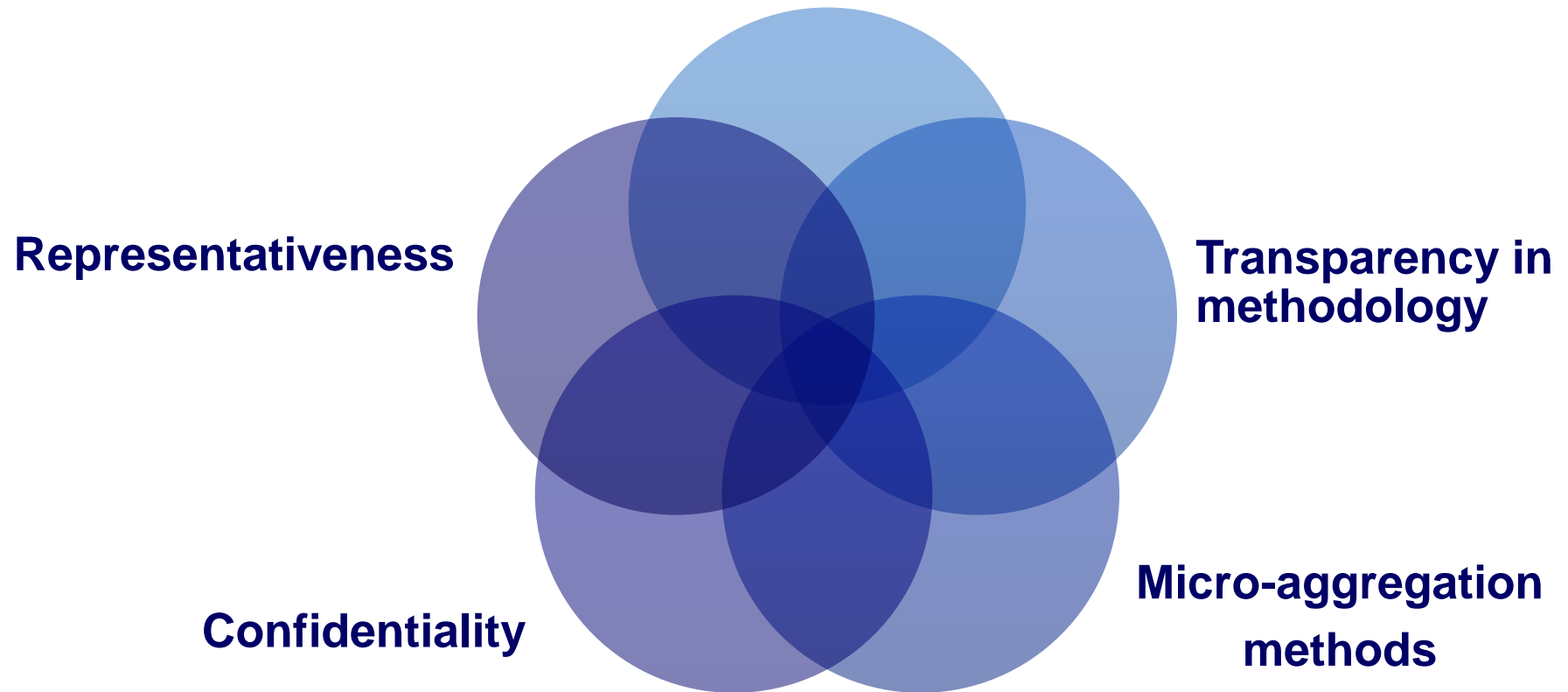
tylervigen.com



“The invalid assumption that correlation implies cause is probably among the two or three most serious and common errors of human reasoning”  
**Stephen Jay Gould, American evolutionary biologist and author, 1981**

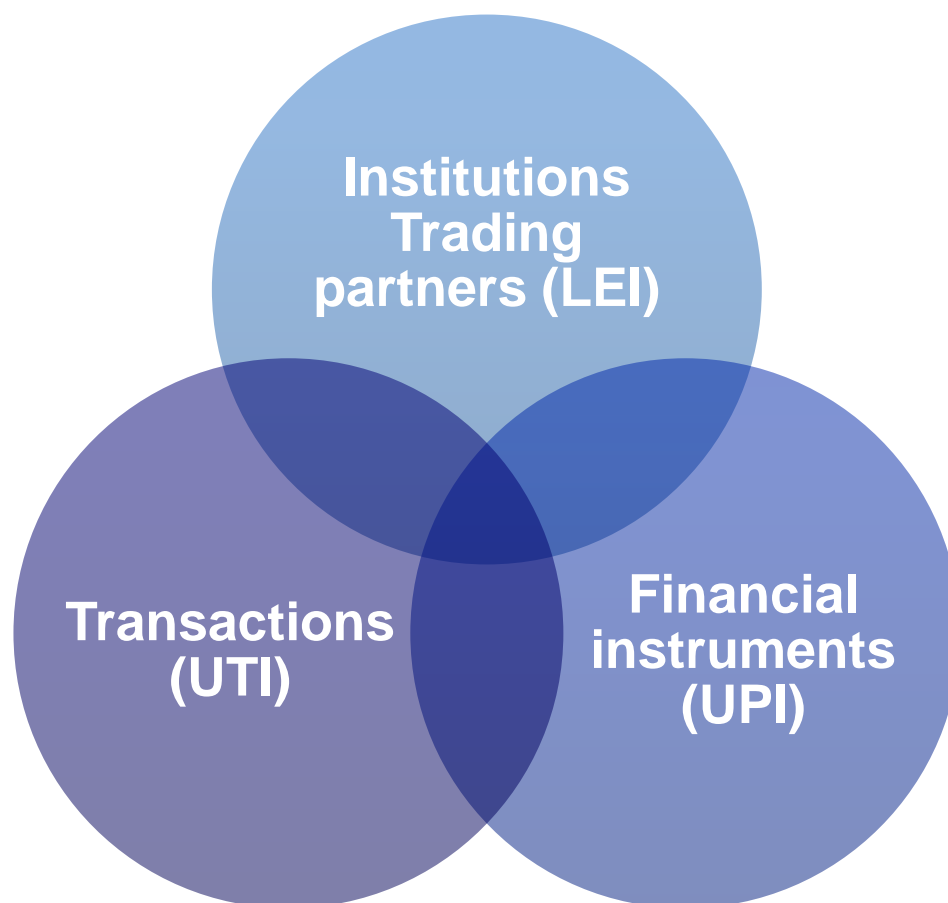
# A source is subject to statistical quality standards

## Quality standards



# Standardisation and harmonisation ! The call for Big Data

- Big data is mainly a bi-product and not the **core business** of the sources
- **Standardisation is a pre-condition for managing & combining large datasets**
- Pre-trading, trading and post trading activities - **ISO20022** and **SEPA**
- **Money Market Reporting & Banks' loans to corporates and households**





## The Irving Fisher Committee

**49 central banks collaborating in showcasing Big Data**

- i. Administrative dataset** (e.g. Fiscal and corporate balance sheet data)
- ii. Web search dataset** (e.g. Google type search info, web-prices)
- iii. Commercial dataset** (e.g. credit cards, operations)
- iv. Financial market dataset** (e.g. frequency trading, price and bid-offer spreads)

**Preliminary results expected still in 2017**

# Showcasing “big data” for macroeconomic purposes

- **Source: Google search data**
- **Objective: Nowcasting euro area unemployment**
  
- ❑ **Dataset:** Google search data (google search machines)
- ❑ Data from **14 countries**
- ❑ Google’s taxonomy of categorising search terms, includes **26** main categories and **269** sub-categories (Finance and Banking)
- ❑ Google search data is **an index** of weekly **volume changes**
- ❑ The volumes are normalised starting at 1.00 and next week value shows the relative change of Google searches within the category (no absolute volumes)

# Showcasing “big data” for macroeconomic purposes

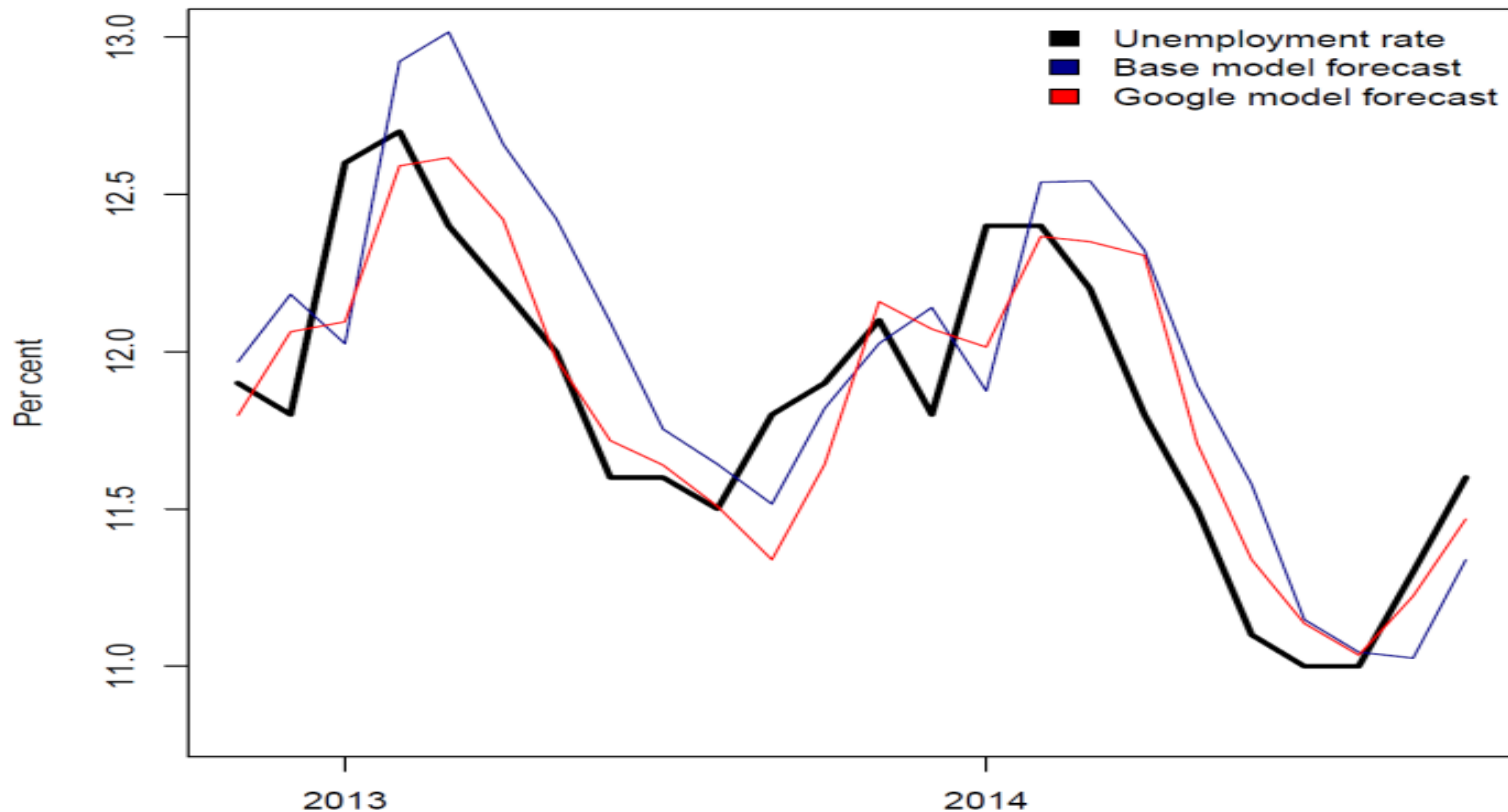
*Two autoregressive models are used to nowcast euro area unemployment rate*

$$\log(y_t) = a + b * \log(y_{t-1}) + c * \log(y_{t-12}) + e_t$$

$$\log(y_t) = a + b * \log(y_{t-1}) + c * \log(y_{t-12}) + G + e_t$$

Where  $Y_{(t)}$  is the unemployment rate at month( $t$ ) &  $G$  is the google search index

**Euro area 19 unemployment rate and predictions**



➤ *The improvement (reduction in the errors (MAE)) range from 18.1% to 28,7%*

## *Lessons learned and way forward*



### **Usability**

- nowcasting retail consumption
- macro-economic indicators
- conjunctural analysis
- consumer behaviour
- price indexes

### **Availability**

- public and free, easy to use
- one system for all countries
- comparability & timeliness
- large taxonomy of searches

### **Innovation**

- trends in communications
- product loyalty
- advertisement
- social patterns in retail markets
- households & business surveys

## *Lessons learned and way forward*



### **Robustness**

- stability of search terms
- volatility in analytical results
- based on one search engine

### **Methodology**

- coverage, weights, normalisations
- aggregation methods
- price information
- short time series

### **Quality**

- differ across regions
- no quality measurements
- no unit tracking
- rebasing and time lag
- home and host concept

# Take 5: Technology transformation and data analytics

1

There is a business case for exploring “big data” for central banking using digital sources from the financial system and beyond

2

The set-up, *competencies are diverse* and source specific

3

Use a conceptual framework & standardising financial markets

4

Join forces in private-public-university partnerships

5

Pretty Data Analytics will change behaviour, business models, knowledge & theories and regulators – Explore digital footprints !

# Q&A

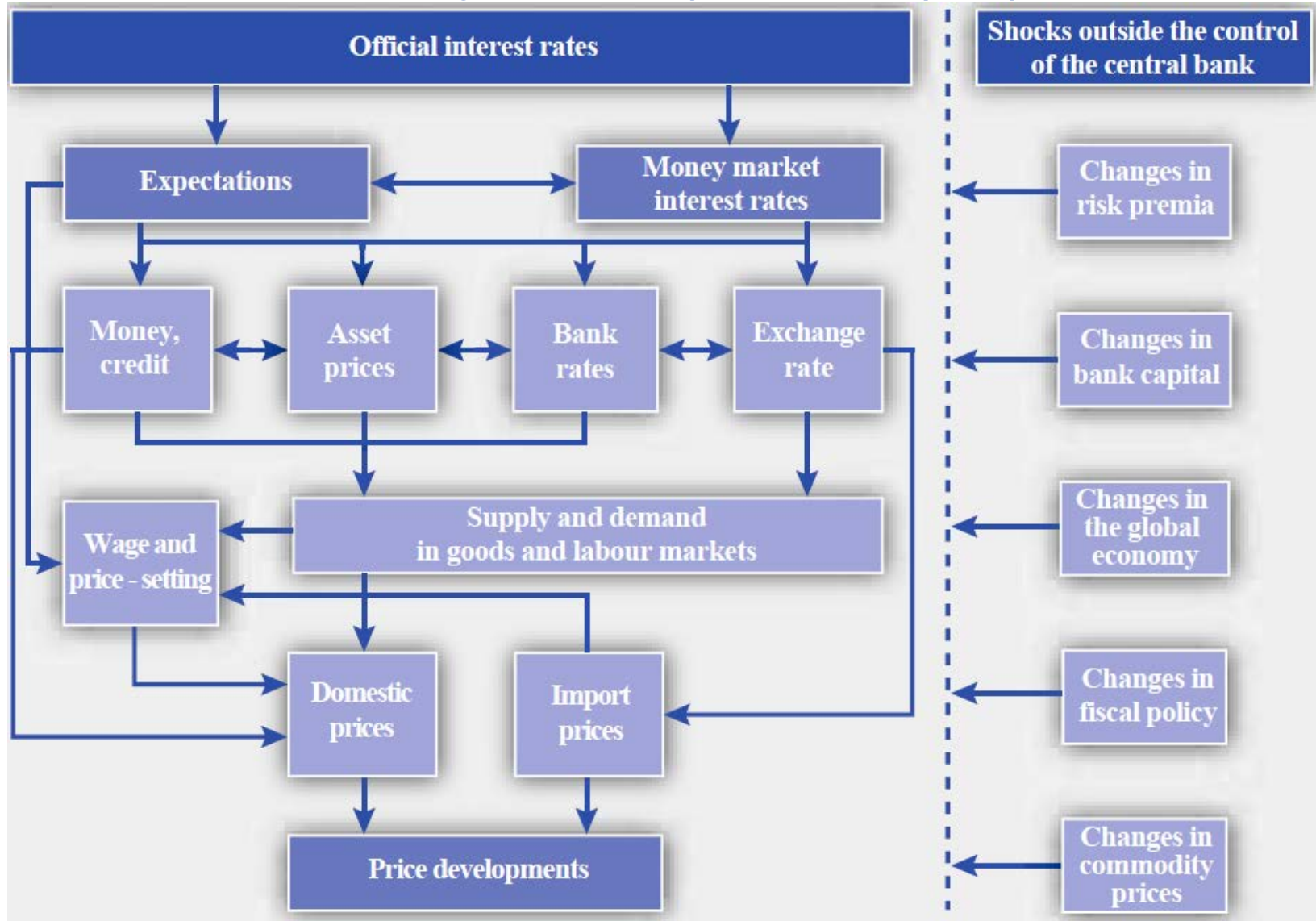


Annex: Examples of Fintech outreach in finance and Fintech companies  
European Central Bank Paper Series - An outlet for big data statistics and research

# The monetary policy transmission process

## How does monetary policy influence price levels?

“Transmission process”, i.e. the process through which actions of the central bank are transmitted through the economy and, ultimately, to prices



# Annex : FinTech in financial IT services – growing field

## BLOCKTECH in FINANCIAL SERVICES VIRTUALscape

by William Mougayar

### APPLICATIONS & SOLUTIONS

<b>Brokerage</b> coinbase, BIT Pagos, UnoCoin, BTCC, BITFINEX, CIRCLE, COINJAI, QUADRIGACX, safello, volobit, coinfloor, coins.ph	<b>Exchanges</b> BTER.com, coinbase, KRIBON, BITSTAMP, POLONIEX, BITCOIN.DE, mexbt, BITSO, CAMPEX, BITMEX, Coinfeine, BitOasis, PAYMIUM, CEX.IO, SHARPE SHIFT, BTC EXPRESS, coinsecure, coinsetter	<b>Soft Wallets</b> BLOCKCHAIN, airBitz, ARMORY, xapo, Mycelium, bread wallet, Coinkite, MultiBIT HD, coinprism	<b>Hard Wallets</b> TREZOR, Ledger Wallet, case, keep key	<b>Investments</b> Grayscale, magnr, loanbase, string, Yuanbao, KOIBANK, Bitbond, WeiFund, WEALTHCOIN, lighthouse, BSAVE.IO, dangpu.com, BTCjam, CHROMA FUND
<b>Merchants</b> bitpay, Bitnet, Coinkite, PEY, Coinify, CoinPayments, coinsnap, coinbase, CoinSimple, BIT Pagos	<b>Compliance</b> third key solutions, PROTUS, ELLIPTIC, CHAINALYSIS, Sig, BLACKSEER, CryptoCorp, IdentityMind, VOGOGO, COINALYTICS, COINRETRY, Merkle Tree	<b>Trading Platforms</b> COINIGY, HEDGY, OrderBook, tradewave, COINUT, AltOptions, MAKER, BITNOMIAL, TERA EXCHANGE, BIRMEX, Mirror, CRYEX, 1 Broker, TABTRADER, dxmarkets, NOBLE MARKETS, AlphaPoint, HitFin	<b>Capital Markets</b> Chain, symbiont, NASDAQ Private Market, Digital Asset Holdings, clearmatics, itBit, TradeBlock, t0, R, epiphyte	<b>Money Services</b> CRYPTO PAY, cashila, ABRA, Fuzo, tether, OBitwala, coins.ph, Simplex, ATLAS, BITX, coinx, R4BIT, Uphold, DUO, BITEXD, CoinPip, LocalBitcoins.com, BitPesa, BlinkTrade, COINAPULT, MELOTIC, Glidera, bridge21
<b>Financial Data</b> bitcoinity, CoinMarketCap, CryptoCoin, BRAVE NEW COIN, BlockJockey, KYF TRADER, BitcoinWisdom, TradeBlock, CoinGecko, Coinhills	<b>Payments</b> Align Commerce, About Payments, COIN, BLADE, GAZEBO.IO, GemPay, cuber, SETL.io, safe cash	<b>Payroll &amp; Insurance</b> paybits, bitWAGE, DYNAMIS	<b>ATMs</b> LocalBitcoins.com, Robocoin, bitxatm, bitaccess, Project Skyhook, btcpoint, LAMASSU, GB, SERY, genesiscoin, COINOUTLET, Modenero Concierge	<b>Banks</b> BBVA, UBS, LHV, London Stock Exchange, secco, BNY MELLON, BARCLAYS, moni, fidor BANK, citibank
<b>Trade Finance</b> GAZEBO.IO, everledger, CHRONICLED, WAVE, skuzon, digi, PROVENANCE, thingchain				

### MIDDLEWARE & SERVICES

<b>Services</b> CRYPTONOMEX, B9, CONSENSYS, SolidX, appliedBlockchain, RUBIX	<b>Software Development</b> chainscript, HydraChain, Blockstack.io, PEERNOVA, CREDITS, eris, MultiChain, Manifold, Blockchain	<b>General APIs</b> BitGo, neuroware, coinbase, bitcore, Gem, BLOCKCYPER, Coinkite	<b>Special APIs</b> TIERION, Open Assets, bitbind.io, coloredcoins, colu, factom	<b>Platforms</b> Counterparty, Monetas, Omni, ethercoin, blockstack, HYPERLEDGER, BLOCKAPPS, Tendermint, appliedBlockchain	<b>Smart Contracts</b> SmartContract, CoinSpark, ROOTSTOCK, bitshares, Tembusu Systems
---	--	---	---	---	---

### INFRASTRUCTURE & BASE PROTOCOLS

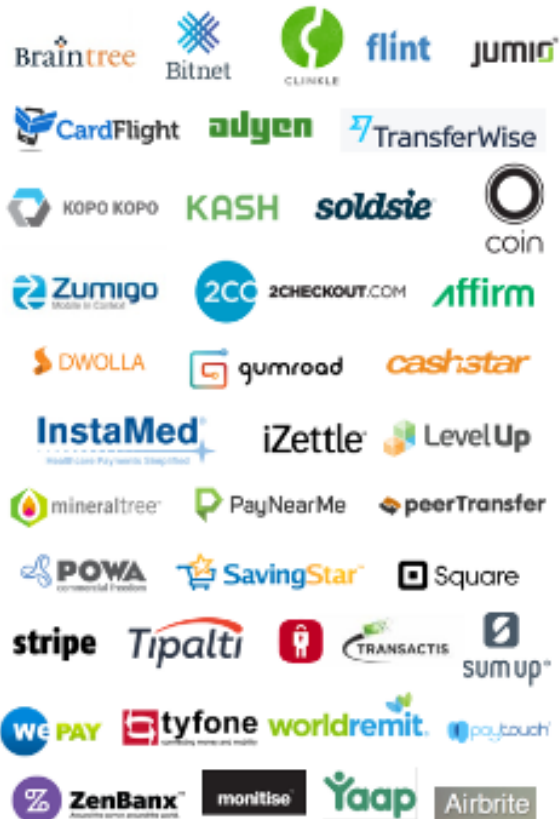
<b>Public</b> bitcoin, bitshares, ethereum	<b>Special</b> ripple, stellar	<b>Payment</b> Lightning Network, MONERO	<b>Miners</b> ANTPOOL, BitFury, 21 INC, BTCC, BITCOIN.CZ
---	-----------------------------------	---	---

# Annex : FinTech outreach in finance – growing field

## FINTECH | LANDSCAPE

everisDigital

### Digital & Mobile payments



### Bitcoin & Cryptocurrency



### Capital Markets & Investing



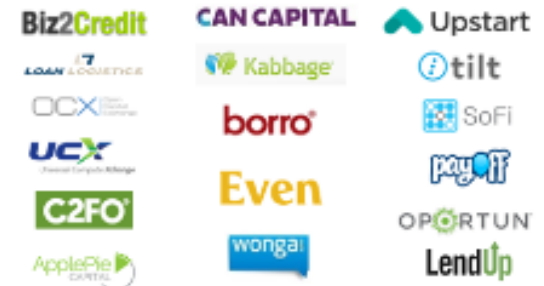
### Banking & Corporate Finance



### Big Data & Analytics



### Financial platforms



### Crowdfunding & peer-to-peer lending



### Personal financial management



### Blockchain technology



# Annex : List of Fintech companies – growing field

## Lending (277 Companies)



## Personal Finance (136 Companies)



## Payments (282 Companies)



## Equity Financing (122 Companies)



## Remittances (42 Companies)



## Retail Investments (99 Companies)



# FinTech

Contact

[info@venturescanner.com](mailto:info@venturescanner.com)  
to see all 1161 companies

## Institutional Investments (63 Companies)



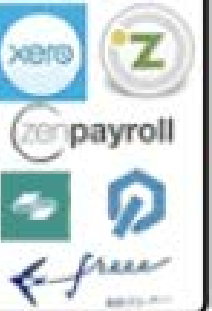
## Security, Auth, Fraud (54)



## Banking Infrastructure (68)



## Business Tools (83)



## Crowd-funding (54)



## Consumer Banking (33)



## Financial Research (35)



## FinTech Investors (34)



- **An outlet for big data statistics and research:**
  - **“Nowcasting GDP with electronic payments data”** by Galbraith J & Tkacz G.
    - Electronic payment transactions can be used in nowcasting current gross domestic product growth
    - finds that debit card transactions contribute most to forecast accuracy
  - **“Social media sentiment and consumer confidence”** by Daas P & Puts M
    - Relationships between the changes in consumer confidence and Dutch public social media?
    - Could be used as an indicator for changes in consumer confidence and as an early indicator
  - **“Quantifying the effects of online bullishness on international financial markets”** by Mao H & Counts S, Bollen J.
    - Develops a measure of investor sentiment based on Twitter and Google search queries
    - Twitter and Google bullishness are positively correlated to investor sentiment