



The data dilemma: a risk or an asset?

10 November 2017
The Westin Zagreb Hotel
Zagreb, Croatia

Data and control – risks and possibilities in the financial domain

Tove Engvall (Mid Sweden University)

Data as a strategic asset for central banks policies

Per Nymand (ECB)

Does more data make better economic policy? A view from HM Treasury, the UK's economics and finance ministry

Mario Pisani (HM Treasury) & **Jan Booth** (DEFRA – former HM Treasury)

The 'Big Data Revolution' in banking and financial history. Some French experiments.

Angelo Riva (Paris School of Economics)

Secure long-term preservation of banking and financial records

Djordje Hinic (PiqI AS)

Privacy, confidentiality, security and consumer protection

Kertész Ákos (Central Bank of Hungary)

What happened in the Daily Gold Fixings Auctions: 1919-1968 - The missing data

Fergal O'Connor (University of York)

A tale of rain and bank runs. From small to big data and back

Anton Comanescu (National Bank of Romania)

Calling all archivists - the five grand challenges of the digital environment

Michael Moss (Northumbria University)



The data dilemma: who is in control?

Tove Engvall, Mid Sweden University,
eabh workshop

2017-11-10, Zagreb



”The data dilemma: a risk or an asset?”

- The amount of data about the finance sector is growing exponentially and storing it is becoming easier. Businesses are excited about the commercial possibilities of ‘Big Data’; academics are relishing the research potential of deep data archives and regulators are hoping for a fuller view of systemic risk and stability.
- Will it all turn out well though? The current reality of massive data stores is often no more than massive cost and complexity. The workshop will explore how we got here with data and where we go next. Ultimately, can a meeting of business, academics and regulators resolve the data dilemma and find a way to turn a risk into an asset?
- ???





Challenge of departure

Digitalization challenge traditional institutional practices of accountability and citizens rights, with risks at a wider scale and higher speed

The online environment make it easy to perform fraudulent activities, cyber crime is growing, and escape in the online environment is quite easy.

Individuals and businesses are vulnerable. Who can be trusted?

Can Big Data analysis be a way to meet these challenges?





Challenge of departure

- Lack of control was part of the reason for the financial crisis 2008 (internal governance, insufficient records management, liberalized regulation & market patterns) (Coleman, Lemieux, Stone & Yeo, 2011).
- Records provide evidence of activities.
Much information are in hte form of data.
Challenges to ensure trustworthiness, long-term preservation and usability
- New regulation of the financial market (MiFID2 & MiFIR) requires of financial firms to provide transaction data.
Can this be a way to manage risks and market abuse?



Research objective & Method

- As digitalization challenge traditional institutional structures for accountability, what are the possible use of Big Data?

What means for democratic control can it be, and what risks are there with increased means for control?

Method

- Literature
- Semistructured interviews;
the European Systemic Risk Board
National Financial Supervisory Authorities in 3 EU countries
National Company Registration Office



Literature

- Big data, data mining, machine learning and Visual Analytics
- Computational Archival Science
- eDiscovery and Digital Records Forensic





Big data, data mining, machine learning & VA

- **Big data:** Big volume, variety and velocity, which requires more than commonly used tools to capture, manage and analyse the data (Lemieux, Gormly, Rowledge, 2014)
- **Data mining:** automated extraction of useful information from data, seeking regularities and patterns (Witten, Frank, Hall & Pal, 2017)
- **Machine learning:** computers ability to answer questions, can also include prediction for decision making (Witten, Frank, Hall & Pal, 2017)

“the capacity of computers to learn without being explicitly programmed” (Humphries, 2017)
- **Visual Analytics (VA):** Combines computational capabilities with graphical representations and interactive analysis

Key findings

- Data mining techniques can be used to extract useful information from data & recognize patterns
- Cyber security, fraud detection, tax evasion, credit assessments, prediction of bankruptcy, financial market monitoring, risk management, facilitate informed decisions
- Challenges of quality, interpretation, management and heterogeneity of data
- Monitoring systems; address the entirety, frequent change, interrelations and unpredictable behaviors. Both economic and behavioral analysis.
- Data is biased, risk for discrimination
- Transparency





CAS (Computational Archival Science)

- Long-term preservation, quality, trustworthiness, means for analysis and interpretation
- Transdiscipline; Computational Archival Science
Computer processing techniques for data management
Archival methods to ensure authenticity & reliability and long term preservation
Ethical, security and privacy issues and organizational and societal concerns
Conceptual knowledge field
- Different machine learning techniques can/have been used in the archival field
classification & disposal,
arrangement & description, search facilities
- Mindset





eDiscovery & Digital Records Forensics

- Readiness for risk, as well as to facilitate investigations
- Cybercrime increases, fraud is common, Western Europe second worst (PwC, Global Economic Crime Survey 2016)
- Big data analytics can be useful, Important is also: valuebased organizational culture, ethical alignment of decision making, governance etc
- Big Data can be useful in police work, but risks of discrimination. Advanced technologies are also used by criminals
- Digital Records Forensic; combine archival knowledge and means for authenticity of records with digital forensics methods and concepts. Means for investigation of digital material, as well as proactive design



Interviews - ESRB



- **the European Systemic Risk Board**
- “oversee the financial system of the European Union (EU) and prevent and mitigate systemic risk” (ESRB, 2017).
- Collaborate with ESMA (European Securities and Markets Authority) which works for protection of investors and to promote stability of EU financial markets, avoid market abuse and market manipulation
- EMIR & AIFMD datasets will be analysed with Big Data tools
EMIR: data about transactions in derivative markets
AIFMD: regulation of speculative funds. Market operations
- Not anymore a black box
- Concerns of data quality, completeness of data & correction of bad reporting



Interviews – National Financial Supervisory Authority

- **Authority country A:** detect violations of regulations
Use Big Data analytics for transaction data and other purposes.
Maybe they will use machine learning

Challenges: System performance, developments at the market, complex information management
- **Authority country B:** detect and prevent market abuse, manipulation and crime, properly working market.
Investments in data accuracy
New regulations (MiFID2 & MiFIR) will provide more data, which they will keep up to 15 years
Challenges: trustworthy long term preservation and management of data, several old databases

Tools: data mining for different analysis and alerts
- **Authority country C:** focus on banking sector
Banks provide big volumes of data that is analysed related to indicators, Identification of risks
Different validation steps
Follow trends



Interviews – National Company Registration Office

- Registration of companies and annual reports
- Electronic reporting and standardized format (ixbrl) – will improve transparency and control, and enable exchange of information.
- Fundamental data and annual report data – possibilities for Big Data analysis
Would make it more difficult for criminals
- Look at possibilities to increase control of the accuracy of reports, to prevent crime
- Possibilities for more accurate credit scoring & lower interest rate.
- Improved means for analysis. Suspicious activities and risks can be detected earlier & increased transparency to the public
- Exchange of information between public authorities can improve crime prevention
Important to consider risks



Reflections

- Big data analysis can increase democratic control, accountability, crime prevention, risk management and more.
- Collaboration can improve information management and promote transparency
- New EU-regulation will improve means for Big Data analysis, but also challenges in management of data

Challenges

- Fast technological development
- Capture, management, quality, control and longterm preservation of data
- Democratic innovation & implementation of regulations
- Actors outside regulation



Discussion

Ethical considerations:

- Surveillance, concentration of power, control, privacy and vulnerabilities,
- Algorithmic discrimination, AI society, propaganda, values & intentions?

Possible further research:

- Implementation of MiFID2 & MiFIR,
- CAS & eDiscovery,
- means for user assessment of authenticity,
- further interviews with stakeholders,
- trustworthiness online

Contact:
Tove.engvall@miun.se

What else is possible?



Per Nymand-Andersen
Adviser, European Central Bank

Data: A strategic asset for central banks policies



Source:  prompt cloud

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

1 Pretty Big Data - Reflections for policy purposes

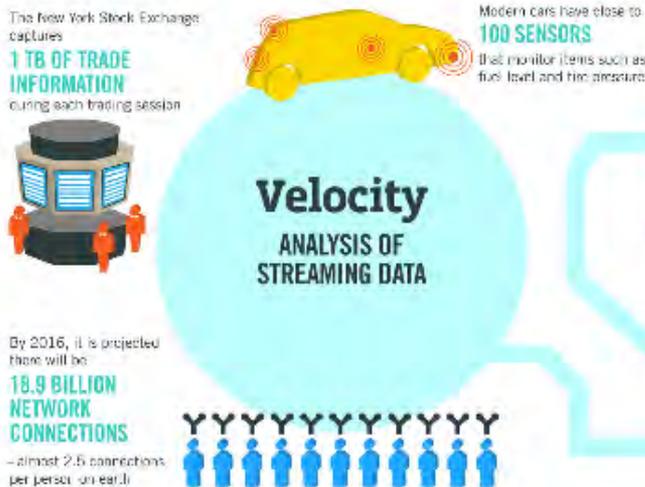
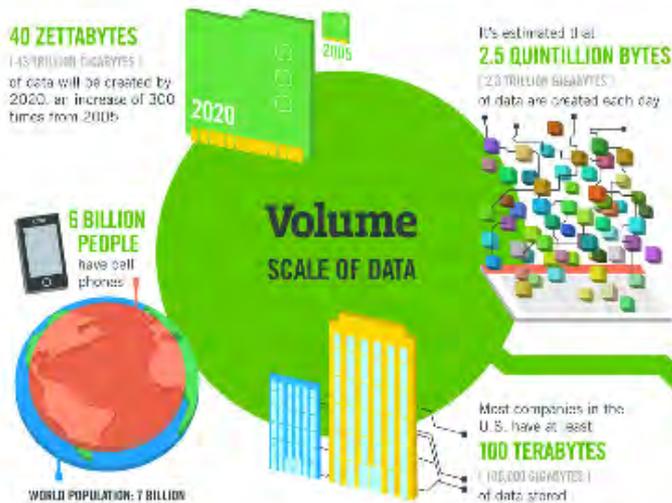


2 Paradigm shift - Moving to micro level and granular data

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)



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 solutions that the world relies on every day. But, when it comes to 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 their infrastructure, and find new sources of revenue.

By 2015, **4.4 MILLION IT JOBS** will be created globally to support big data with 2.9 million in the United States.



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

150 EXABYTES
(150 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



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



27% OF RESPONDENTS

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

Veracity UNCERTAINTY OF DATA

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



Data never sleeps – Challenges for historians

2017 *This Is What Happens In An Internet Minute*



Digital exploration

- Storage capacity
- Linking data sets
- Accessing
- Perform querying
- Slice & Dice sources across time and datasets

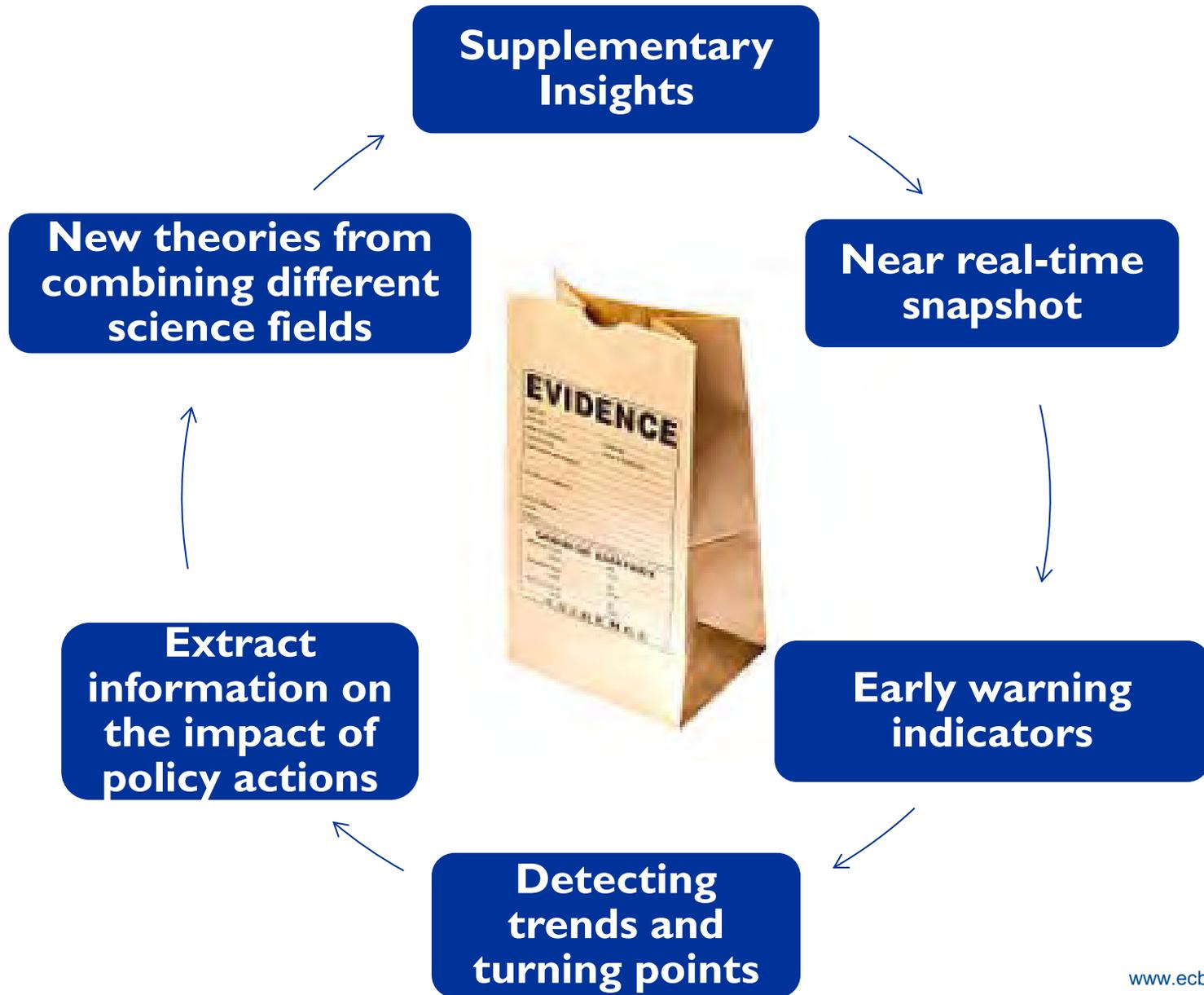
Fit for the future

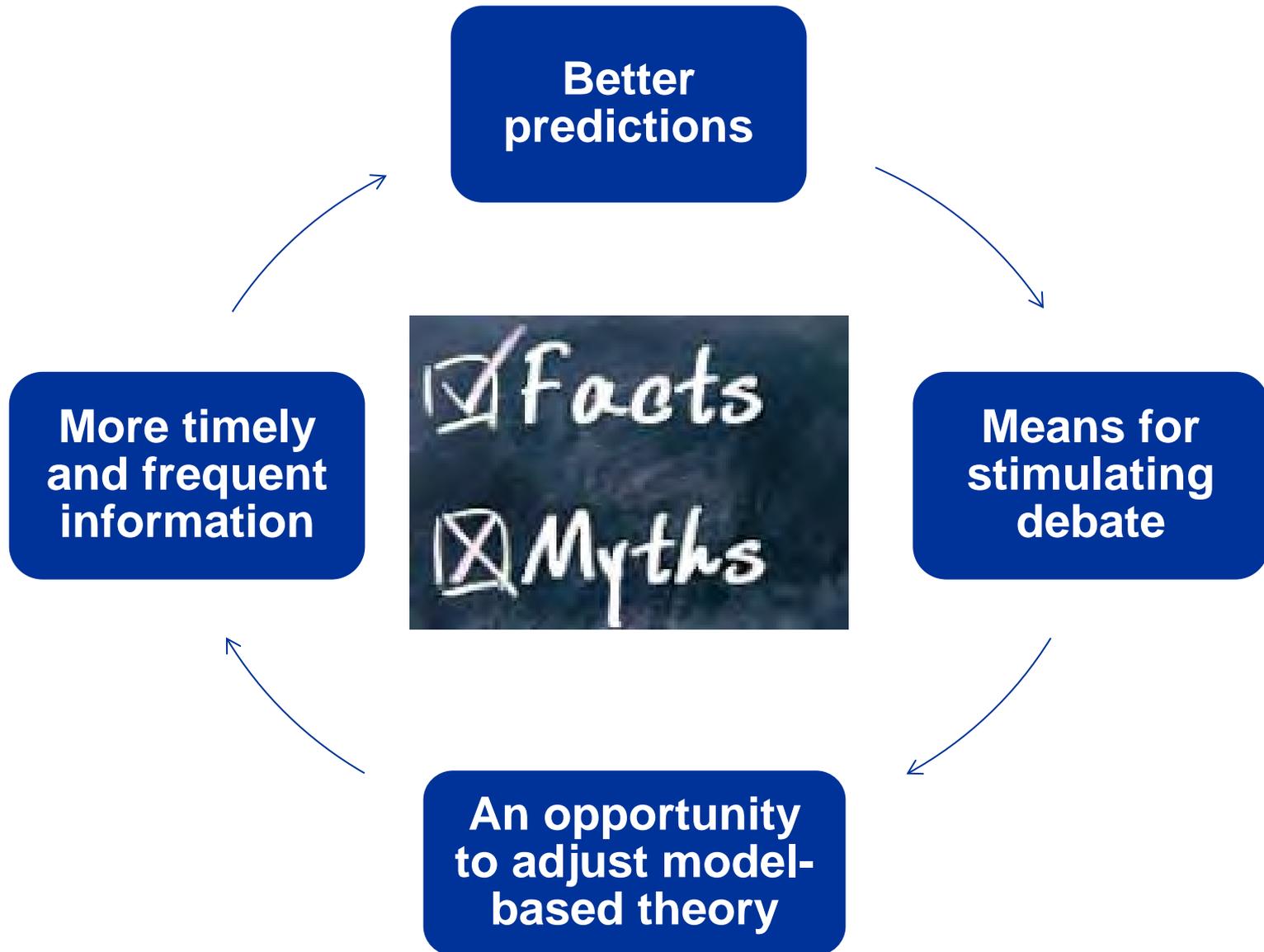
Which preparations are needed today to have the capacity and functionality needed in 10 years time?

- Managing and exploring datasets
- Linking current and past datasets
- Querying variety of formats

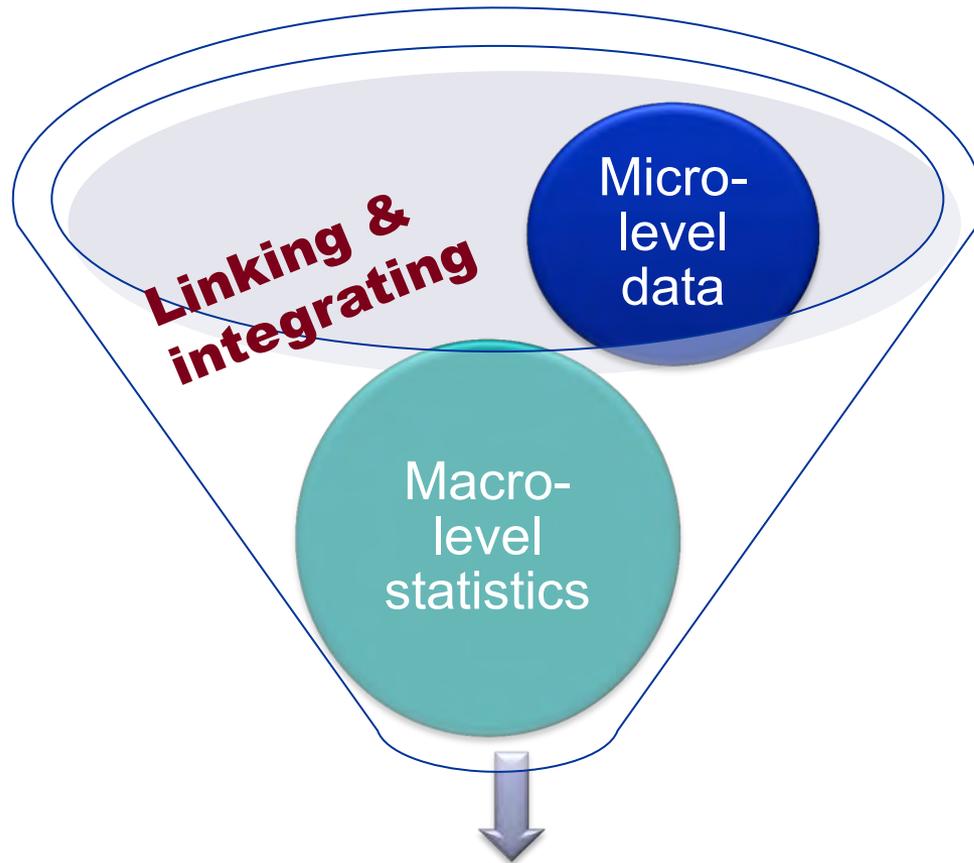
Source: Cumulus Media,

<http://www.businessinsider.de/everything-that-happens-in-one-minute-on-the-internet-2017-9?r=US&IR=T>





Paradigm shift - Moving to micro level/granular data



Statistics & Analytics

Micro-level statistics

- Security-by-security statistics
- Holdings of individual securities
- Money market transactions (MMSR)
- Loans by loans register (Ana Credit)
- Register of Financial Institutions
- Individual bank supervisory data

Macro-level statistics

- Balance sheet statistics
- Monetary aggregates (M1 – M3)
- Securities issues
- Banks interest rates
- Government finance
- Euro area financial accounts

New challenges - Micro-level data (5th “V”!)

Systematic approach for sustainable provision of statistics

Semantics and standardisation

- Definition and methodology
- Data Dictionary & data Model
- **Standardisation** of identifiers
- Mapping and linking datasets

Quality & streamlined IT

- Flexible reception and data views
- Linking internal & external data-sets
- **Data Quality** (automatic machine readable checks & probabilities)
- Managing data cubes

Statistics analysis

- Slice & dice micro-data
- Data Discovery and summary statistics
- **Data science**
- Statistics and analytical capacities

Communication outreach

- Visualisation & Presentations
- **Communication**
- Explainers
- Outreach to frequent users

Data sources for Monetary Policy & Financial Stability

Other systems

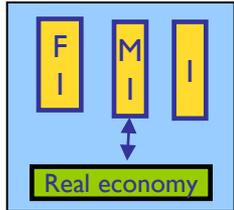


Financial system

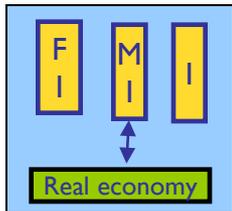


Real economy

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

Data mania versus phobia – a 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



Systematic acquire,
Process, summarize

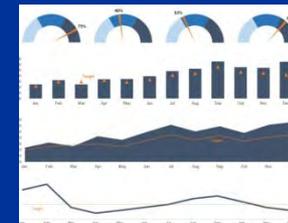


Data lab



Statistical algorithm and
data explorations

Packaging data for
Insights & business



3 remarks on Big Data 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

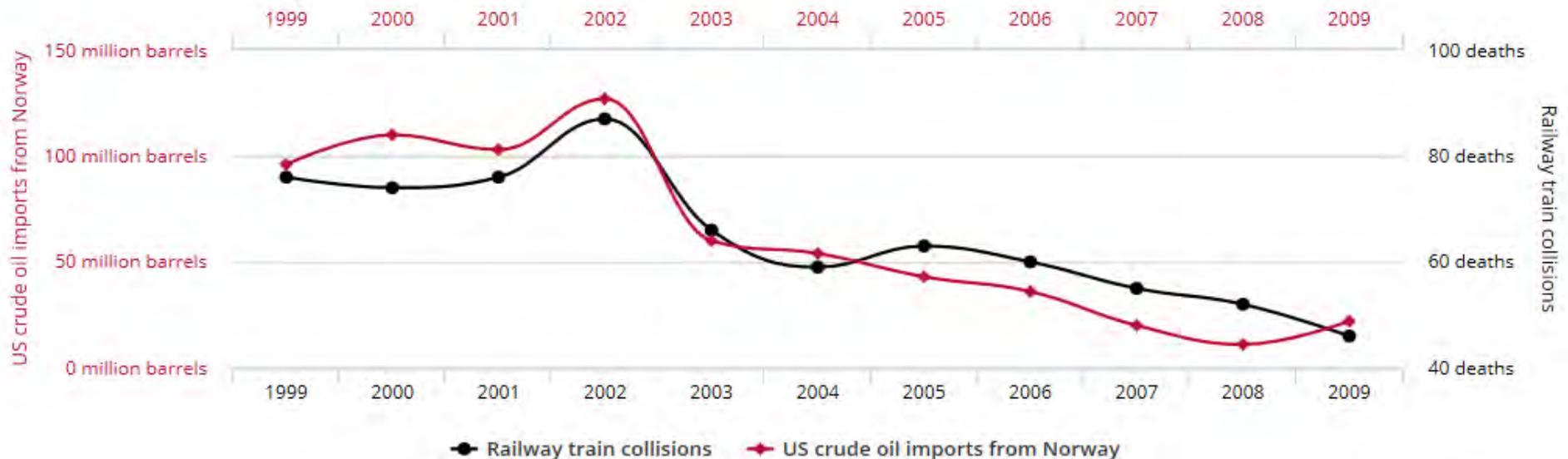
3 remarks on Big Data analytics

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: $R=0.95$



byleniger.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



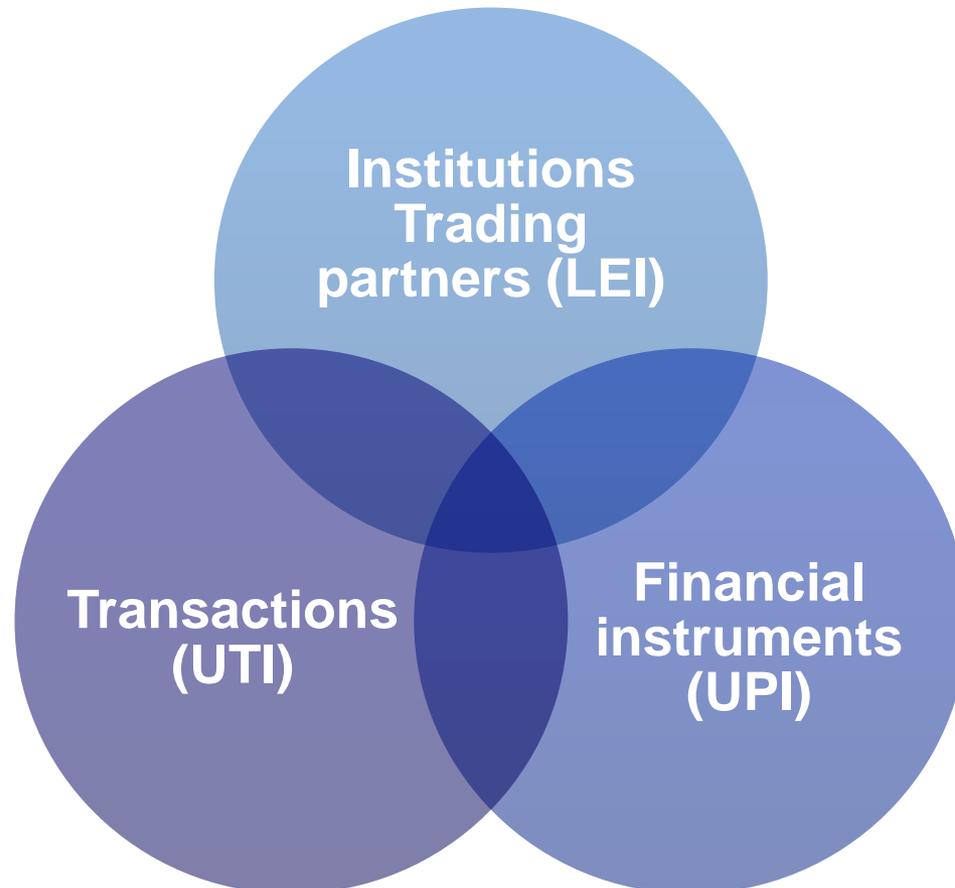
3 remarks on Big Data analytics

3. Any source is subject to statistical quality standards



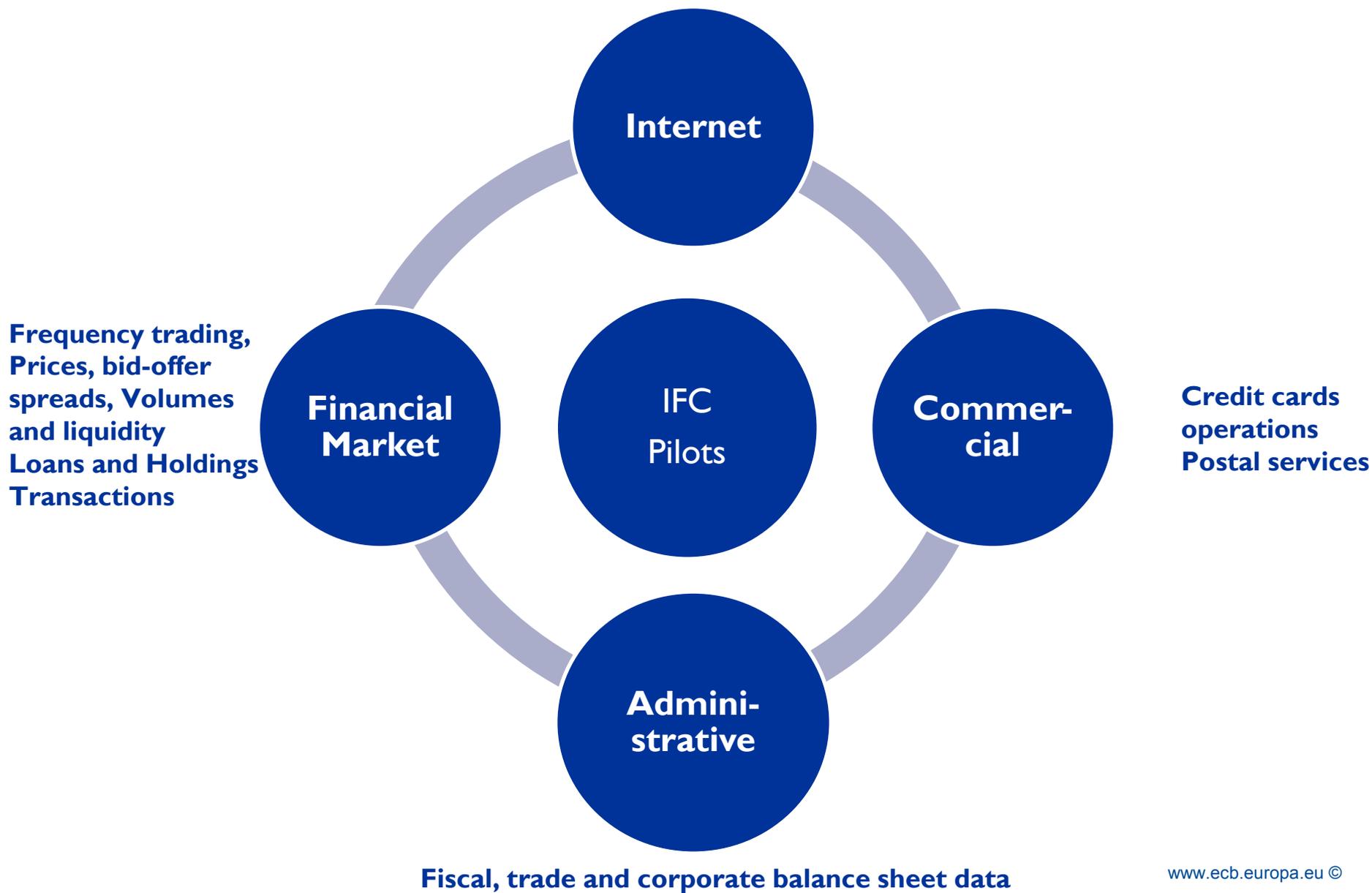
The call for standardisation on Big Data

- Big data is mainly a *by-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**



49 Central banks collaborating - The Irving Fisher Committee

Search machines, Web-prices and properties, Tweets, internet media,



Five take!

1 *There is a business case for exploring “big data” in economics & finance*

2 *Call for applying common standards in financial markets*

3 *Data scientists for ensuring quality and generating insights*

4 *Collaborate for piloting big data and use public/private partnerships*

5 *Preserve the digital footprint – A rich (future) source for historians*

Thank you for your attention



Any questions?

Annex: [ECB Paper Series](#) as an outlet for big data research

➤ An outlet for big data 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



HM Treasury

Data and economic policy in the UK: opportunities and challenges

Mario Pisani

Deputy Director, Fiscal Group

10 November 2017

Contents of my talk today

Section 1: quick overview of HM Treasury

Section 2: understanding the economy

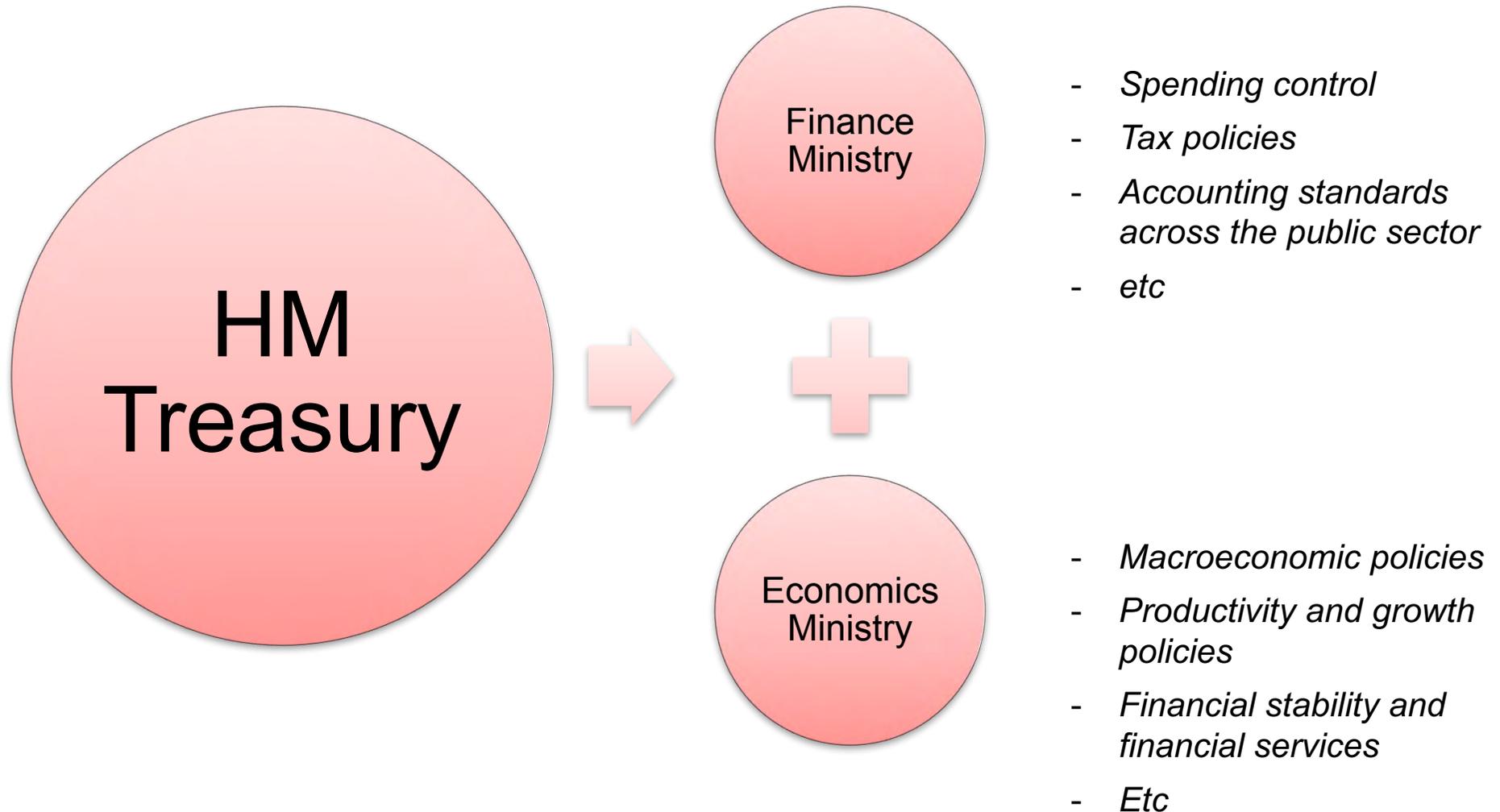
Section 3: policy-making process

Section 4: challenges

Her Majesty's Treasury is one of the oldest government departments in the United Kingdom



In the UK, the Treasury is both the finance ministry and the economics ministry



Section 1: quick overview of HM Treasury

Section 2: understanding the economy

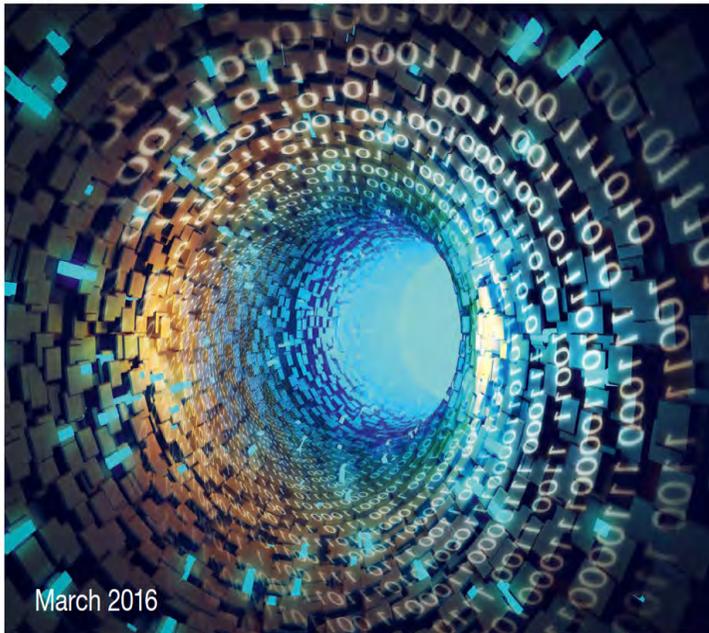
Section 3: policy-making process

Section 4: challenges

2016: HM Treasury commissions Professor Charlie Bean to do independent review of economic stats

Independent Review of UK Economic Statistics

Professor Sir Charles Bean



Digital revolution means the economy is changing fast – policy-makers need to understand this

Digital economy

81%

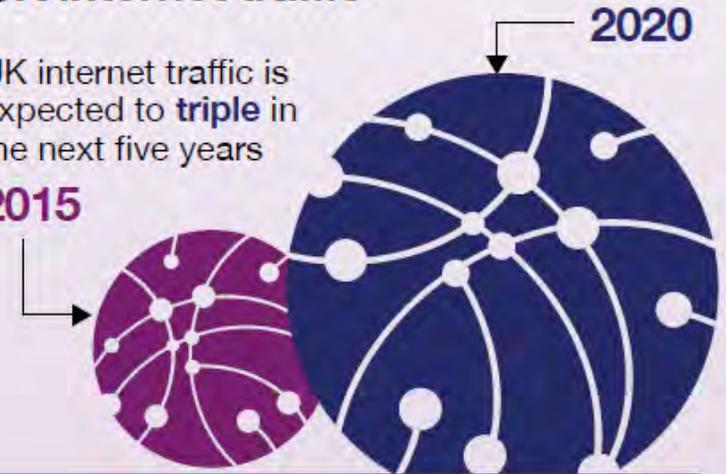
The UK has the highest share of consumers **purchasing online** out of all EU countries



UK internet traffic

UK internet traffic is expected to **triple** in the next five years

2015



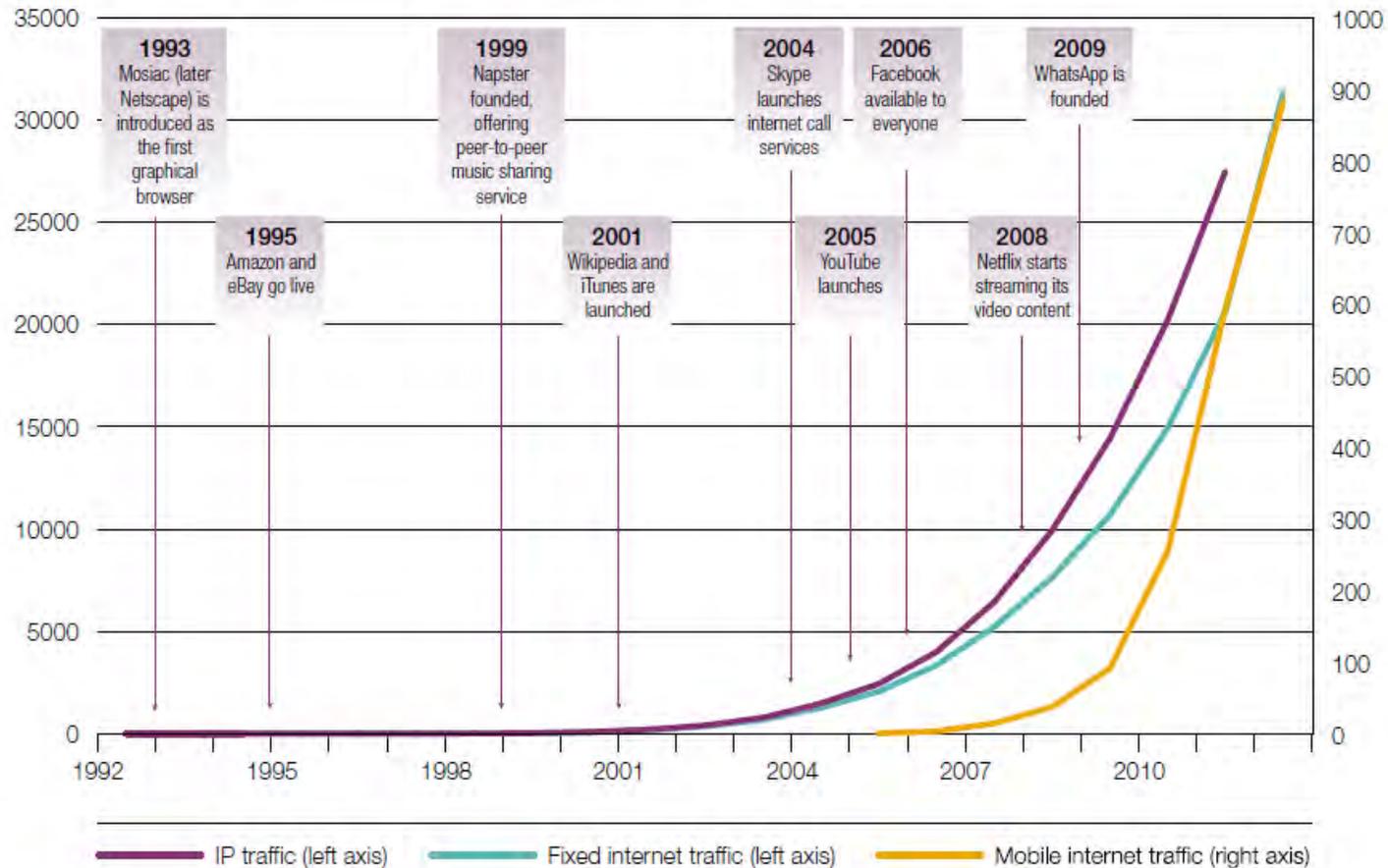
London has produced the third largest number of **sharing economy** start-ups in the world

SEVENTY
TWO

San Francisco 131
New York 89
Paris 24
Boston 20



We've seen a very rapid increase in online activity – this and other digital processes create data



Notes: Petabytes per month. Aggregating from multiple sources and applying usage and bitrate assumptions, Cisco Systems, a major network systems company, has published the following historical Internet Protocol (IP) and internet traffic figures.

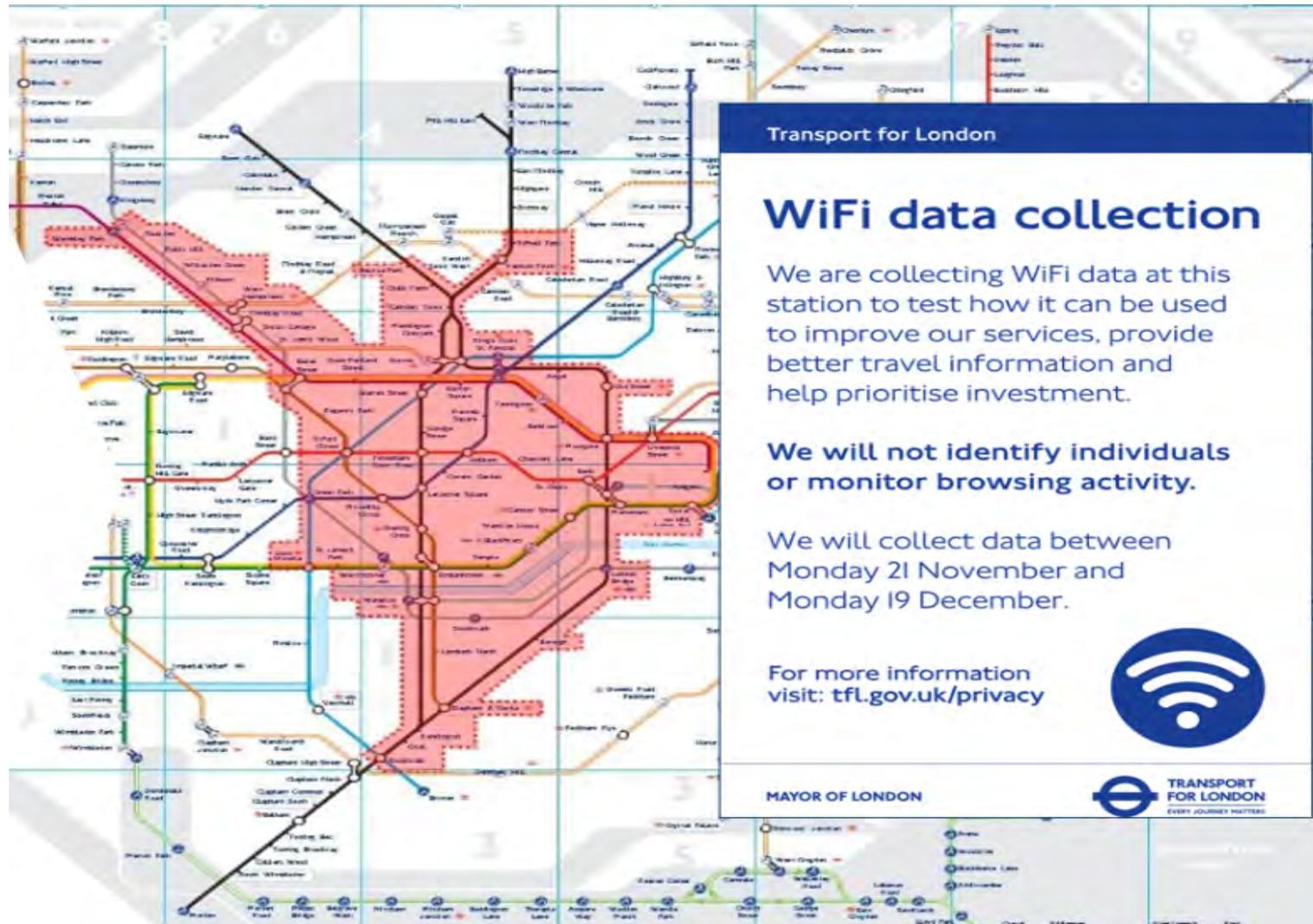
Source: Wikipedia.



1939: there has always been an interest in using data to understand patterns of economic activity...



... now there is more data, which allows quicker and more precise analysis of economic behaviour



Transport for London

WiFi data collection

We are collecting WiFi data at this station to test how it can be used to improve our services, provide better travel information and help prioritise investment.

We will not identify individuals or monitor browsing activity.

We will collect data between Monday 21 November and Monday 19 December.

For more information visit: tfl.gov.uk/privacy

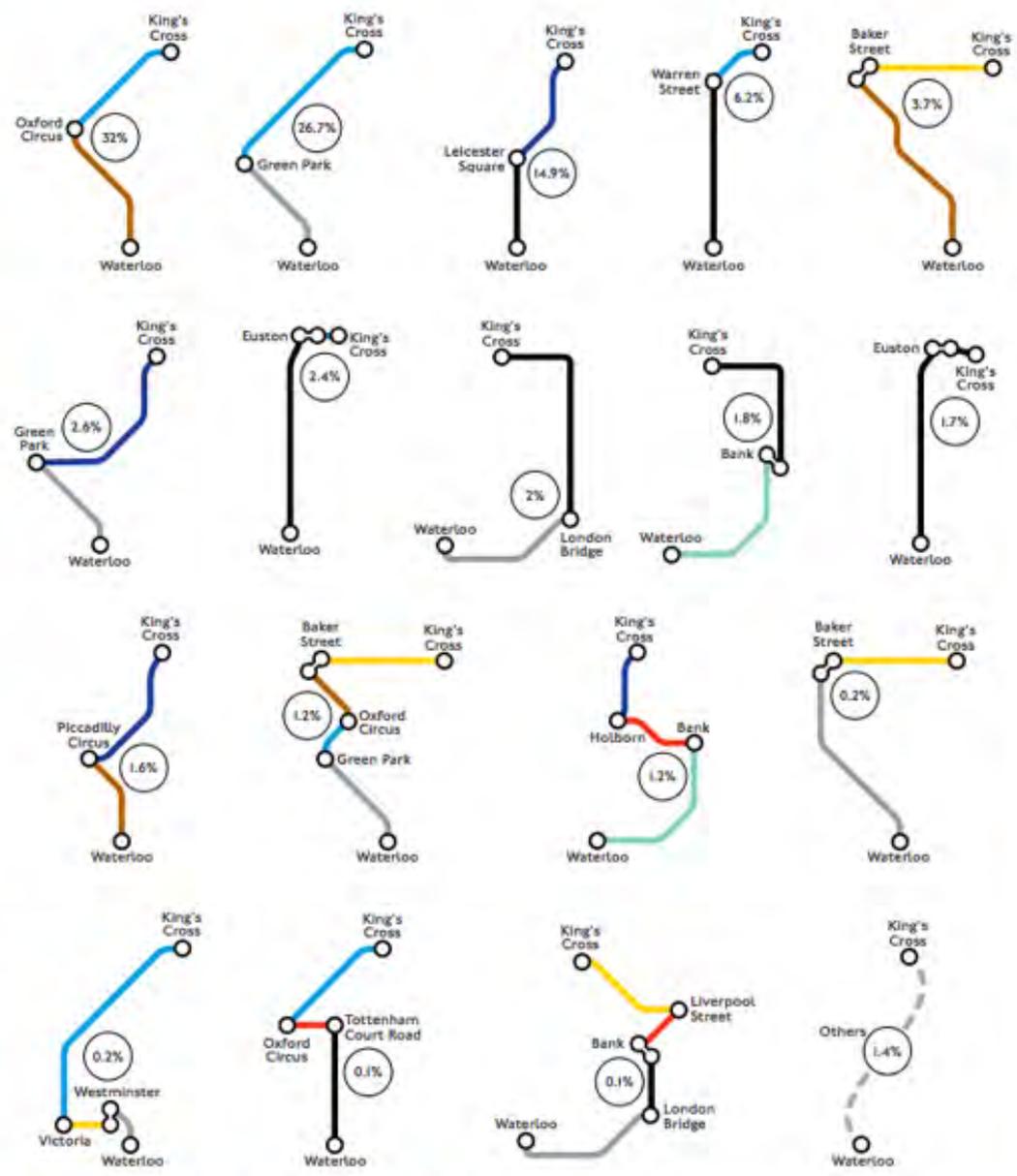
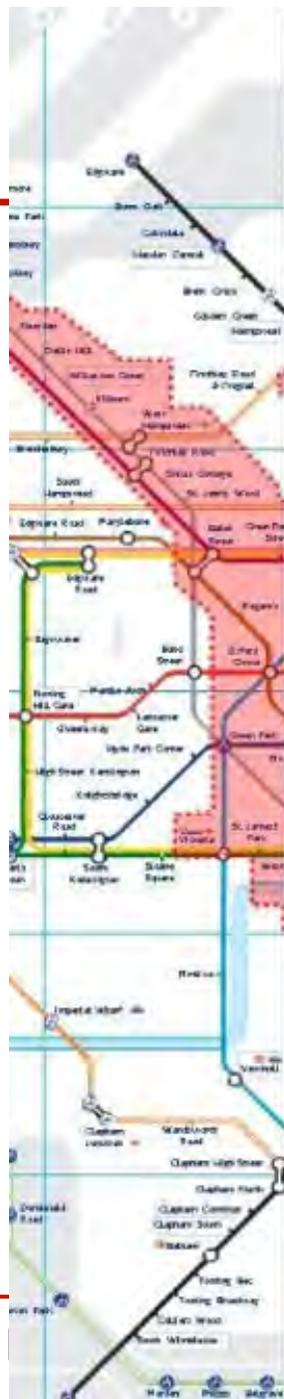


MAYOR OF LONDON



TRANSPORT FOR LONDON
EVERY JOURNEY MATTERS

Figure 14: Route options between King's Cross St. Pancras and Waterloo, and the proportion of devices on each one



Government itself sits on lots of administrative data which can help us measure the economy

Businesses sampled by the Annual Business Survey

63,000

The HMRC VAT database has information for



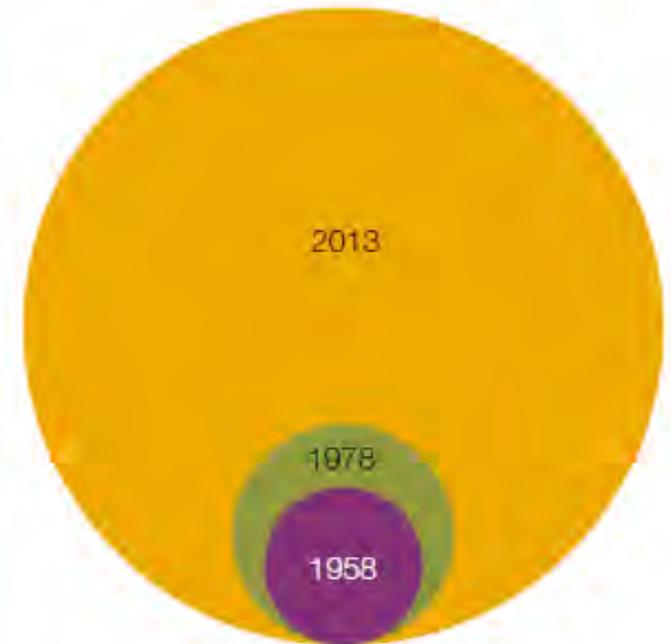
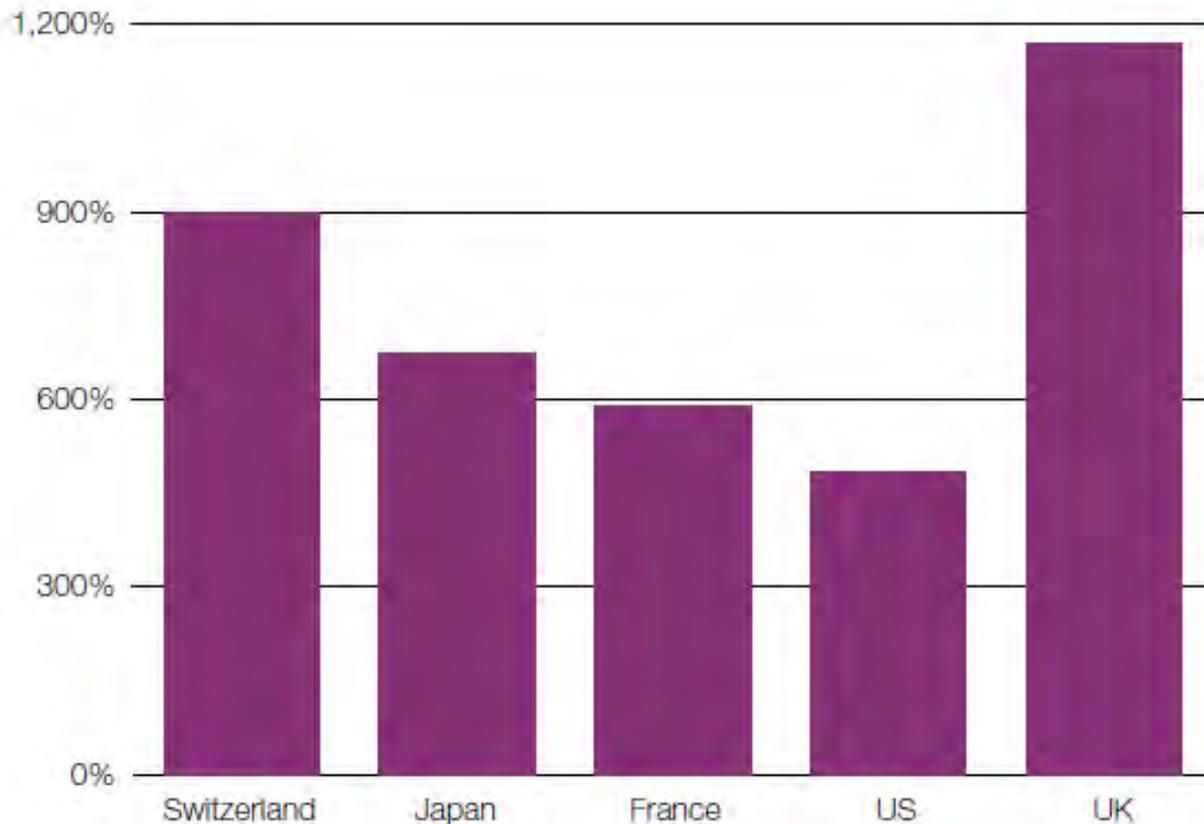
**1.8 million
businesses**

Employees sampled as part of wages and salaries survey:
9,000

Income tax records on HMRC pay-as-you-earn system:
1.5 million

The UK has a very large financial system – its regulation and supervision also generates data

Size of the financial system as a % of GDP



At the moment our understanding of financial flows is limited to sectors and subsectors only

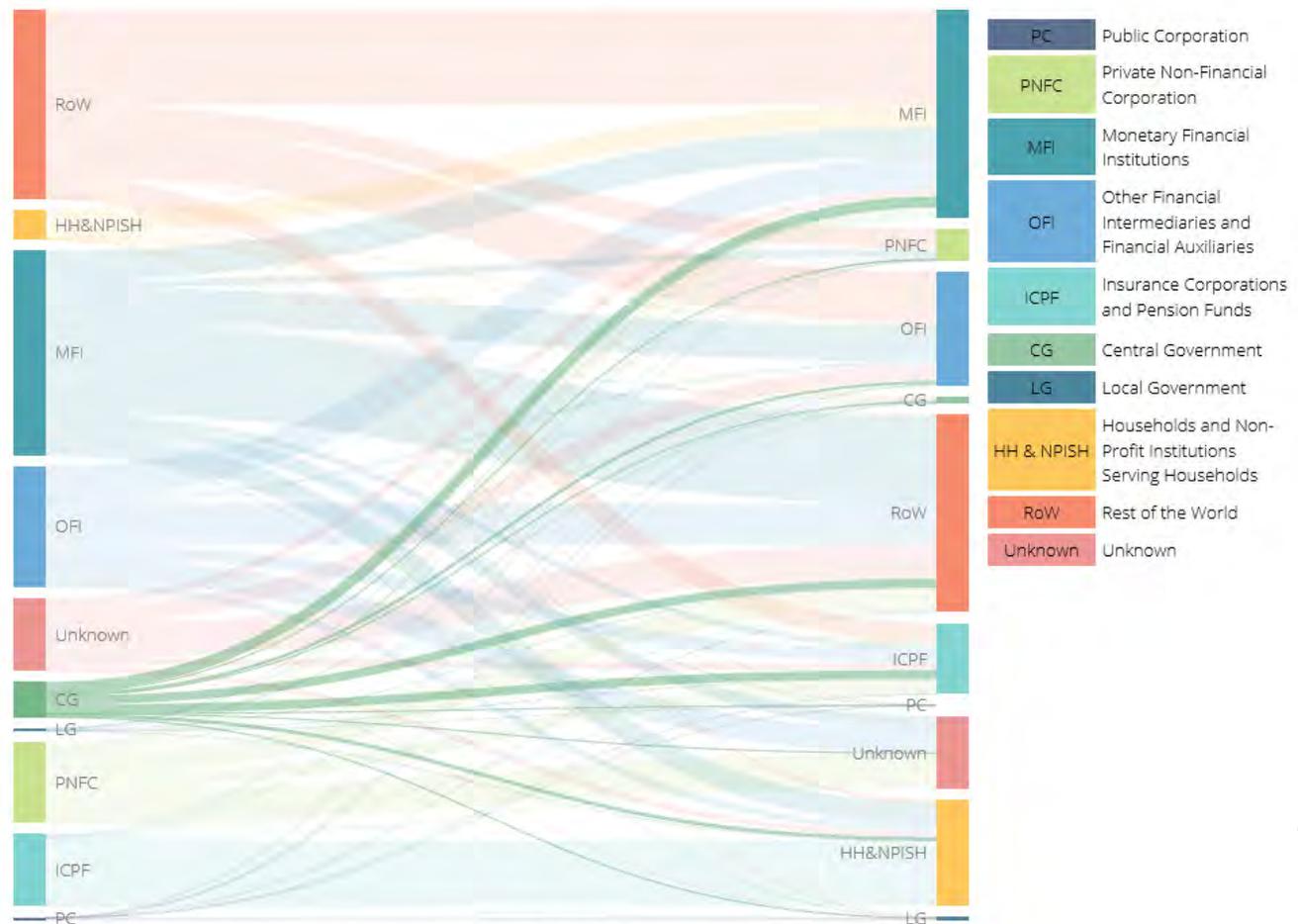
Sector-to-sector interactions for financial balance sheets

Selected year: 2014

Liability

Asset

Central Government liability: £1,906 billion

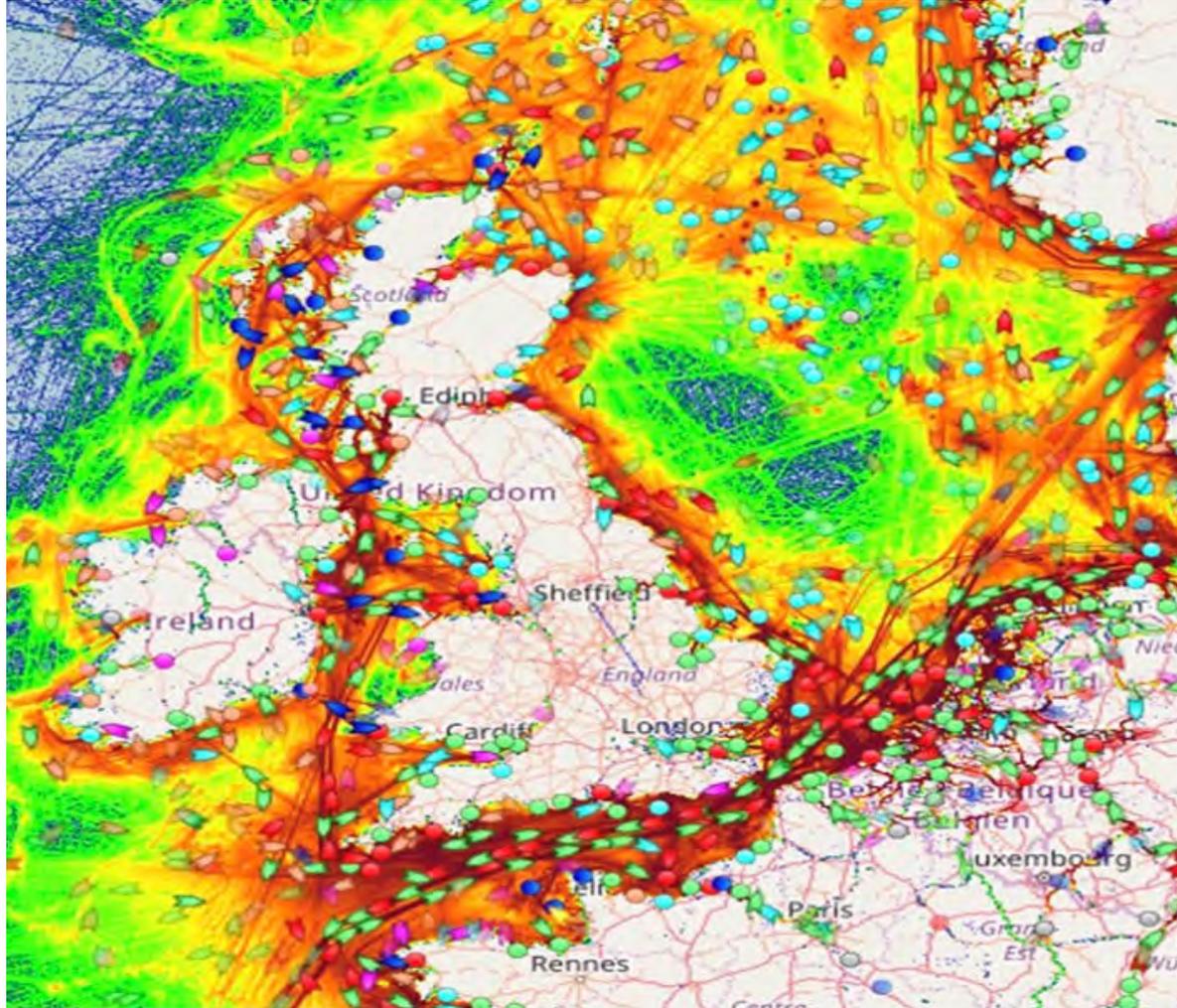


Link to Flow of Funds Sankey diagrams [here](#) and [here](#)

Lots of useful data outside the public sector – for example web scraping and scanner data for prices



Can shipping and ports data provide a better estimate of international trade flows?



Section summary - more data is an opportunity for improving our understanding of the economy

We offer three kinds of service:

GOOD - CHEAP - FAST

You can pick any two

GOOD service CHEAP won't be FAST

GOOD service FAST won't be CHEAP

FAST service CHEAP won't be GOOD

Increased use of administrative and alternative data sources could:

- Improve timeliness and allow us to obtain information about the economy quicker = FASTER
- Greater sample sizes could improve granularity and accuracy of economic statistics = BETTER
- Reduce the need for expensive surveys and sampling = CHEAPER



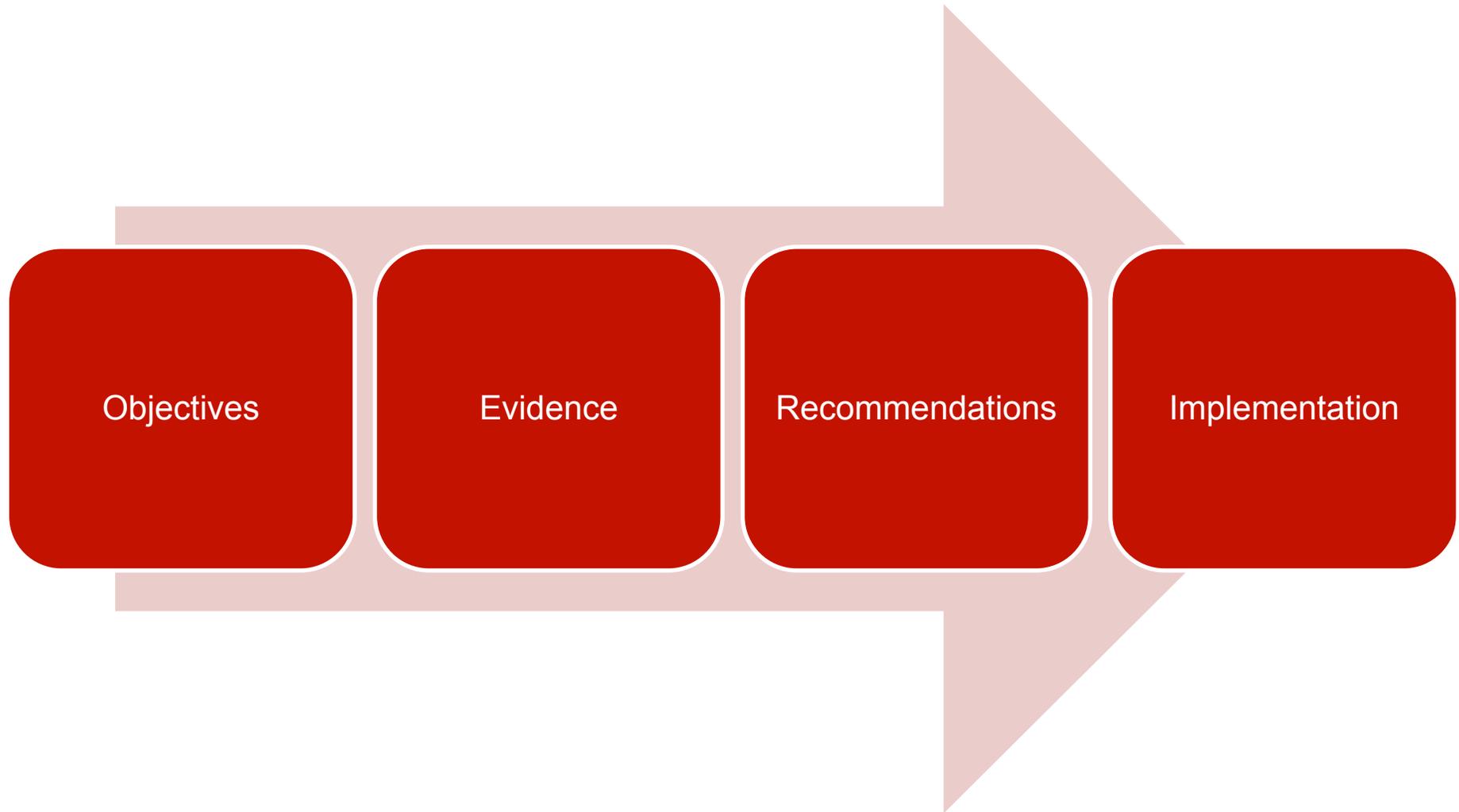
Section 1: quick overview of HM Treasury

Section 2: understanding the economy

Section 3: policy-making process

Section 4: challenges

The traditional approach to policy development



New data and data sources have made viable a range of different approaches to policy-making

Acting

Challenge setting

Challenge panels

Change cards

Crowdsourcing

Data and social media analysis

Data tool cards

Guerrilla Testing

Hope and fear cards

Interviews

Journey mapping

Evidence safari

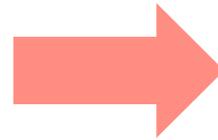
Personas

Open Data

An introduction to prototyping

Prototyping in a workshop: Tabletop prototyping

Touchpoint prototyping



Some examples:

Social media engagement

Crowdsourcing

Open data

User consultation

Example 1: HM Treasury crowd-sourcing ideas for efficiencies in public spending

Policy paper

Public Sector Efficiency Challenge summary of responses and results

Published 25 November 2015

Contents

1. Who participated in the survey?
2. What were the most common suggestions for change?
3. Which ideas have departments decided to take forward?
4. What will happen to the remaining ideas?

In August 2015 the Chancellor and Chief Secretary wrote to all public sector workers asking for their ideas on how the government could do more for less. The Spending Review and Autumn Statement 2015 announces the results of the exercise.

1. Who participated in the survey?

Just over 22,000 suggestions were submitted as part of the challenge, from a wide range of organisations in the public sector:

- 5,000 who work in healthcare and the NHS
- 3,200 who work in local government
- 2,500 who work in education and schools
- 1,400 who work in defence and the military
- 1,300 from those who work in policing

The remainder were submitted by civil servants, including those working in agencies and public bodies, such as Jobcentre Plus.

Example 2: engaging with businesses to understand how to change or improve regulation

Focus on Enforcement

Help us to identify where enforcement can be improved, reduced or done differently



[Home](#) [Business Focus on Enforcement](#) [Focus on Enforcement Reviews](#) [Benefits of FoE reviews](#) [Regulator Information](#) [Contact us](#)

Welcome to the Focus on Enforcement website

Three new BFoE reviews published

On Monday 23 March we published the first three *Business Focus on Enforcement* reviews alongside responses from the relevant regulators. The reviews examined enforcement arrangements in respect of [livestock farm inspections](#) (led by the *National Farmers Union*), [imports of fresh produce](#) (led by the *Fresh Produce Consortium*) and [the classification of electronics exports](#) (led by *techUK*).



Focus on Enforcement

We know ill thought out and unnecessary regulations cost business time and money. So the Government is tackling this through the Red Tape Challenge.

But sometimes the regulation itself is fine – it is inconsistent or inappropriate enforcement that causes problems or could just be so much better.

Search the site

Search

Regulator Information

Click below for more background information on:

- [Information on National Regulators](#)
- [Local authority regulatory functions and their remit](#)

Links

- [The Red Tape Challenge](#)
The website for you to tell Government which regulations are working and which are not.
- [Better Regulation Executive](#)
More on the Government's Better Regulation Strategy
- Scotland: [Give us your view](#)

Example 3: HM Treasury using social media to engage with citizens about the upcoming budget



HM Treasury 
@hmtreasury Follow

This year's Budget will be on 22 November – there's still time to have your say on what should be included [gov.uk/government/pub...](https://gov.uk/government/publications)



11:15 pm - 18 Sep 2017

153 Retweets 61 Likes



30 153 61

The Trading Twit @TheTradingTwit · Sep 19

Section summary: new technology means the policy-making process itself generates new data

Alternative approaches to policy-making have the potential to:

- Make it easier to obtain feedback direct from users and those affected by policy changes
- Harness the creativity and insight of a wider group of people, and exploit a range of experiences and approaches

These approaches can both generate new data and facilitate the analysis of data

Section 1: quick overview of HM Treasury

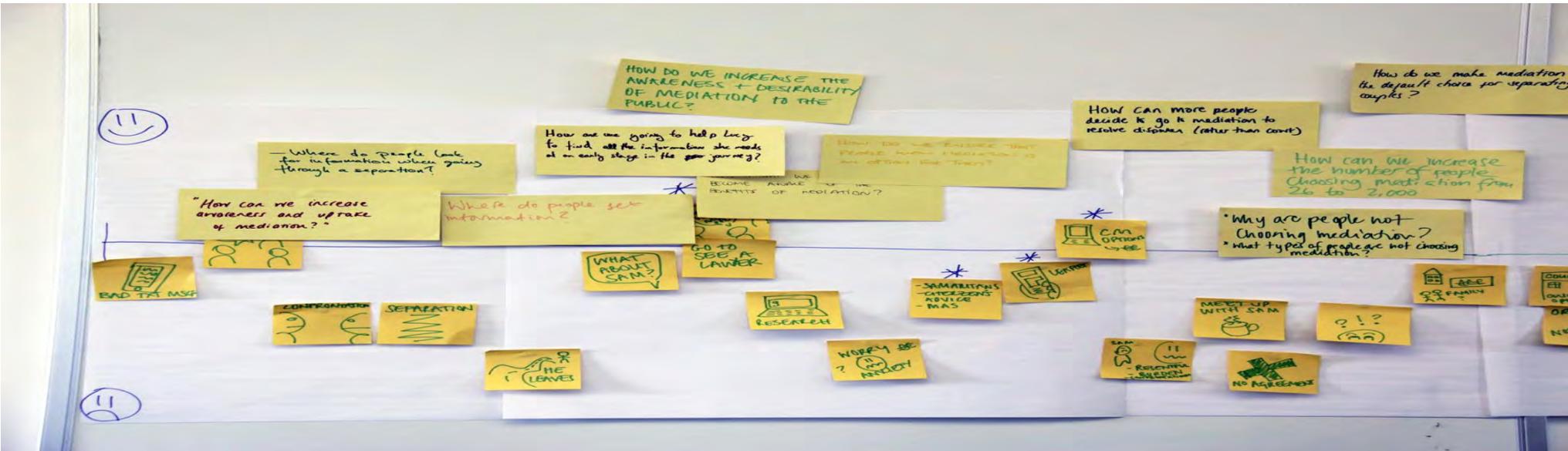
Section 2: understanding the economy

Section 3: the policy-making process

Section 4: challenges

Some examples of challenges with data (there are likely many others...)

- Volume: too much data – deciding what to ignore
- Access: different data cannot be matched due to limited access
- Costs: cost of managing and analysing data can be very high
- Continuity: some data sources may be discontinued



Links

[Charlie Bean Review into big data and economic statistics](#)

[ONS article on flow of funds](#)

[Open policy-making](#)

[Public spending challenge](#)

Mario.Pisani@hmtreasury.gsi.gov.uk



HM Treasury

HM Treasury's information archives

***A really great story that's really
hard to tell***

Jan Booth (DEFRA – former HM Treasury)

HM TREASURY



HM Treasury

Treasury information (my type of 'data')

TYPES	THEMES	WHERE'S THE STUFF?
-------	--------	--------------------

- **Ministers' files**
- **Policy files**
- **Budget records**
- **Secondary sources (stats, reports, analyses...)**

- **Domestic economy**
- **Institutions**
- **Tax and spending**
- **Shocks and crises**

- **Core series up to 1990 – National Archives**
- **1990-1998 paper files**
- **1998 on electronic records**



Treasury readers

Treasury people

Organisational boundary

Professional analysts

Academics

Research students

Journalists



My existential crisis

What's the point of archives?

What is an archivist for?



What about public service?

What am I doing here?



The challenges

Getting behind the wire

I can fix

Poor metadata

I can help

The digital heap

I need help

Resources

I need help



Some questions...

**Can researchers
add archival
value?**

**Can we share
digital tools and
know-how?**

**Can researchers
act as Sherpas for
Treasury users?**

**Do private
collections play
their part?**

Is there a dimension beyond mere moral obligation?

The 'Big Data Revolution' in banking and financial history. Some French experiments.

Angelo Riva
European Business School-Paris
&
Paris School of Economics

Data dilemma: risk or asset?

- “The current reality of massive data stores is often no more than a massive cost and complexity”: Yes
- “Academics are relishing the research potential of deep data archives and regulators are hoping for a fuller view of systemic risk and stability”: should we give up?
 - Maybe...

No. Why not?

- If growth is to be strong and spread fairly, the EU needs a sound, academic evidence base with data about the long-run performance of European finance
 - History is a boundless laboratory for real-size natural experiments
 - The weak empirical foundations of the models used to analyze structural and cyclical changes have become obvious (crises... what ?)
- Crucial historical understanding of our society remains totally inadequate, because we lack the requisite empirical basis
- The EU Strategy Report on Research identifies Big Data in the social sciences and the humanities as the first science driver for these fields.

Innovation

- Investing time and money in developing new technologies to capture and connect FAIR data
 - Findable, Accessible, Interoperable and Re-usable => High quality data
 - Reading writings, not just getting numbers ... to understand the numbers
- Some French Experiments at the Paris School of Economics



Données Financières Historiques

- Project designed to develop a comprehensive database on the French stock markets since 1796, to be extended to other kinds of data.
- Fortnightly spot, forward, options prices of all the assets listed from 1796 to 1976 on the Paris stock market
 - Assets: securities (stock and bonds, French and foreign, private and public), gold and silver materials (bullions, various coins), exchange rates, bills of exchange...
 - 1 asset => several prices per day
- Securities events:
 - coupons/dividends; new issues, split, reverse split, M&A..
- Data on issuers:
 - Juridical statutes from the foundation (dates)
 - Equity capital and subsequent issues of bonds/short terms notes
 - Localization of headquarters and factories
 - Balance sheets
 - Administrators
 - Governance (distribution of profit, specific rules)

Parisian Financiers'
Belle Epoque



Collecting data from deeds of partnerships of Parisian bankers and partners (1783-1913)

- map of financial operators in Paris and of their partners around the world
- data on their social and professional characters
- GIS “Old Paris” at the Ecole des hautes Etudes en Sciences Sociales

Cooperation with archivists

- Archives... where all of it began
- Paris Stock Exchange sources at the Centre des Archives Economiques et Financières (CAEF)
 - Archives' organizational setting to host (many) RAs
 - Partnership to scan sources
- Deeds of bankers at the Archives de Paris
 - Facilities for RAs and sources digitalization
- Archives of the Crédit Agricole for SYSRI-30
 - Locating and scanning « The Album »



Pascal Penot, Crédit Agricole

Watch the video at <https://www.youtube.com/watch?v=AMcqSvZlvOE>



Données Financières Historiques

DFIH Sources

- Two main (serial) sources:
 - Lists of the exchanges
 - Yearbooks of the exchanges
- Additional sources
 - Additional printed sources (from exchanges, from other bodies)
 - Archives

Paris Stock Exchange Official Lists

BOURSE DE PARIS du 19 Juin 1915				1870 COURS AUTHENTIQUE SEUL OFFICIEL (1) Bourse de Paris du 19 Juin 1915												1960 Mardi 21 Juin 1960			
Cours Jours		Paris Mars		Cours		Paris		Cours		Paris		Cours		Paris		Cours		Paris	
Paris	Paris	Paris	Paris	Paris	Paris	Paris	Paris	Paris	Paris	Paris	Paris	Paris	Paris	Paris	Paris	Paris	Paris	Paris	Paris
<p>Amsterdam 147 Londres 147 Cote d'Estim 147 Lyon 147 Marseille 147 Nantes 147 Paris 147 Rouen 147 Strasbourg 147 Toulouse 147 Bordeaux 147 Nîmes 147 Montpellier 147</p>				<p>10 - CARBURANTS, SALINES, CHARBONS, PHOSPHATES</p> <p>11 - CIMENT, MASTIQUE, ATTRACTIONS</p>												<p>19 - PETROLES ET CARBURANTS</p> <p>METROPOLE</p> <p>ANTAR-PETROLES DE L'ATLANTIQUE</p> <p>AQUITAINE</p> <p>BERRE</p> <p>COPIREX</p> <p>FINAREP</p>			

Data Entry for the Official List (1)

IT organized manual data entry

- Reproduction of the structure of the Official list into the DB
- Sources' digitalization
- Creation of a data entry mask (java program)
- Training of data entry firm's operators (outsourcing for the most part of)
- Data delivery
- Data quality checks (IT validation, check of outliers; complete check on a representative sample)

SCOR Input module for securities

31/12/1845 Stockexchange Paris (Official List) Manual

Paris (Official List)

- NOT QUOTED STOCKS or BONDS
- RENTES FRANCAISES
- FONDS ETRANGERS
- CHEMINS DE FER (actions)
 - Bordeaux a la Teste (Chemin de Fer)
 - Marseille a Avignon (Chemin de Fer)
 - Montpellier a Gette (Chemin de Fer)
 - Orleans (Chemin de Fer)
 - Paris a Rouen (Chemin de Fer)
 - Rouen au Havre (Chemin de Fer)
 - Saint Etenne a Lyon (Chemin de Fer)
 - Saint Germain (Chemin de Fer)
 - Strasbourg a Bale (Chemin de Fer)
 - Versailles (Rive Droite)(Chemin de Fer)
 - Versailles (Rive Gauche) (Chemin de Fer)
- OBLIGATIONS (Chemins de Fer)
- FONDS DE LA VILLE
- COMPAGNIES DES CANAUX
- COMPAGNIES ANONYMES
- FONDS ETRANGERS
- VALEURS DIVERSES

Name: Montpellier a Gette (Chemin de Fer)

Sector: Railroads

Type: action de capital

Market: Spot (Au Comptant)

Price 1

Price 2

Price 3

Price 4

Price 5

Price 6

Price 7

Price 8

Price 9

Price 10

Higher(Plus haut)

Lower(Plus bas)

Previous(Dernier cours coté)

At dd/mm/yyyy

Price yesterday (Cours de la veille)

Demand(e)

Supply(Offre)

Volume

Source: Official List

Comments

Nominal: No value found

Dividend: No dividend information available

Amount: No quantities found for this date

Outside exchanges(Hors parquet)

Price 1

Price 2

Price 3

Price 4

Price 5

Repo Market(Marché des réports)

Emission

Save

Legend(e)

New

Amount

New Currency

New code

No: 48 - St: 35 - Sh: 1 - QF: - Lan: France - Zet: France - Code: - Corp: Montpellier a Gette

Start

14:53

Data Entry for the Official List (2)

Specific OCR based software: some experiment
(the same for Yearbooks)

- Sources digitalization
- Elaboration of a lexical dictionary and specifications rules to instruct the OCR software
- Design and creation of an interface human – software
- Design and creation of a workflow management system

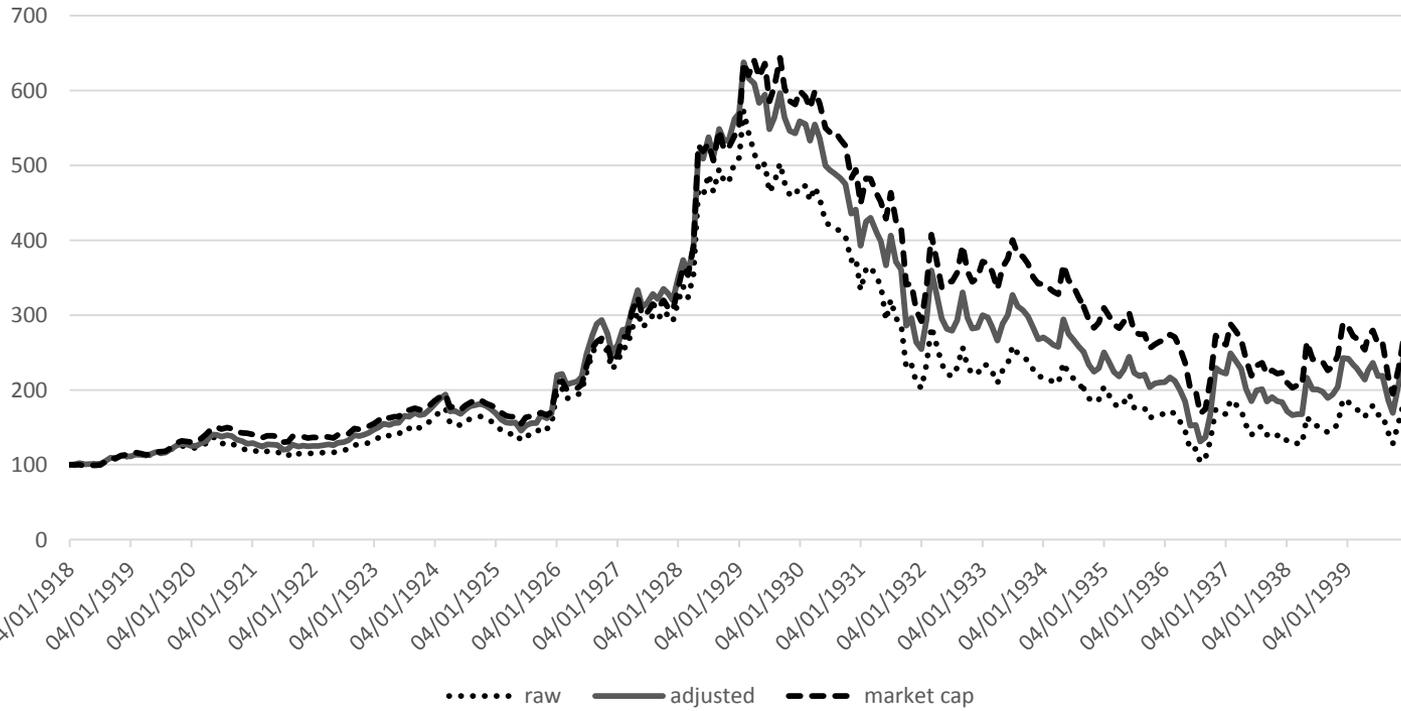
23 522	1580	936	1240	Thomson-Houston (Cie Francaise des Procédés), act. 500 fr., t. p. (ex-c. 41)..... 15 juillet 29										1250	1255	1260	1261	1265	3 10	1260	1235	en liquid. au 31....	1258 ..	1272 ..	1255 ..	1262 ..						
				FORGES, FONDERIES HOUILLÈRES, MINES																			Pr. dem..	1275 ..	1280 ..	1310 ..	1315 ..						
																													Pr. au 31.	1300 ..	1305 ..	1310 ..	1315 ..
																													Pr. au 15.	1300 ..	1305 ..	1310 ..	1315 ..
																													Pr. au 28.	1300 ..	1305 ..	1310 ..	1315 ..
																													Pr. au 15.	1300 ..	1305 ..	1310 ..	1315 ..
23 522			1240	Thomson-Houston (Cie Francaise des Procédés), act. 500 fr., t. p. (ex-c. 41)										1250	1255	1260	1261	1265		3 10	1260		En Liquid.	1258 11									
				(Id=4434) Thomson-Houston (Cie Francaise des Procédés), act. 500 fr., t. p. (ex-c. 37)																			Au 31 1265	1272 11	1255 11	1262 11							
																						(1) Pr. dem	1275 11	1280 11	1310 11	1315 11							
																						(2) Pr. au 31	1300 11	1305 11	1310 11	1315 11							
																						(3) Pr. au 15	1300 11	1305 11	1310 11	1315 11							
																						(4) Pr. au 28	1300 11	1305 11	1310 11	1315 11							
																						(5) Pr. au 15	1300 11	1305 11	1310 11	1315 11							

300 000	juillet ...	701 ..	70 ..	54 55	1069	685	800	Peugeot (Automobiles), act. 500 fr., t. p. (ex-c. 21)..... 29 mai 29										808	809	810	811	814	815	817	2 ..	820	814	en liquid. au 31....	810 ..	815 ..	825 ..	814 ..	820 ..								
										Peugeot (Automobiles), act. 500 fr., t. p. (ex-c. 21)																						Pr. dem..	830 ..	835 ..	840 ..	840 ..	840 ..							
																														Pr. au 31.	830 ..	835 ..	840 ..	840 ..	840 ..		Pr. au 15.	830 ..	835 ..	840 ..	840 ..	840 ..		
																														Pr. au 28.	830 ..	835 ..	840 ..	840 ..	840 ..		Pr. au 15.	830 ..	835 ..	840 ..	840 ..	840 ..		
300 000			7	70	54 55		1069	685	800	Peugeot (Automobiles), act. 500 fr., t. p. (ex-c. 21)										808	809	810	811	814	815	817		2	820		En Liquid.	810												
										(Id=1150 81 82%) Peugeot (Automobiles), act. 500 fr., t. p. (ex-c. 35)																				Au 31 815	825 11	814 11	820 11											
										(Id=2281 61 11%) Saint-Denis Automobiles, act. 100 fr., t. p. (ex-c. 8)																				(1) Pr. dem	830 11	835 11	840 11	840 11	840 11									
										(Id=8008 57 5%) Entreprises Automobiles, act. 500 fr., t. p. (ex-c. 34)																				(2) Pr. au 30	830 11	835 11	840 11	840 11	840 11									
																												(3) Pr. au 15	830 11	835 11	840 11	840 11	840 11	840 11										
																												(4) Pr. au 28	830 11	835 11	840 11	840 11	840 11	840 11										
																												(5) Pr. au 15	830 11	835 11	840 11	840 11	840 11	840 11										

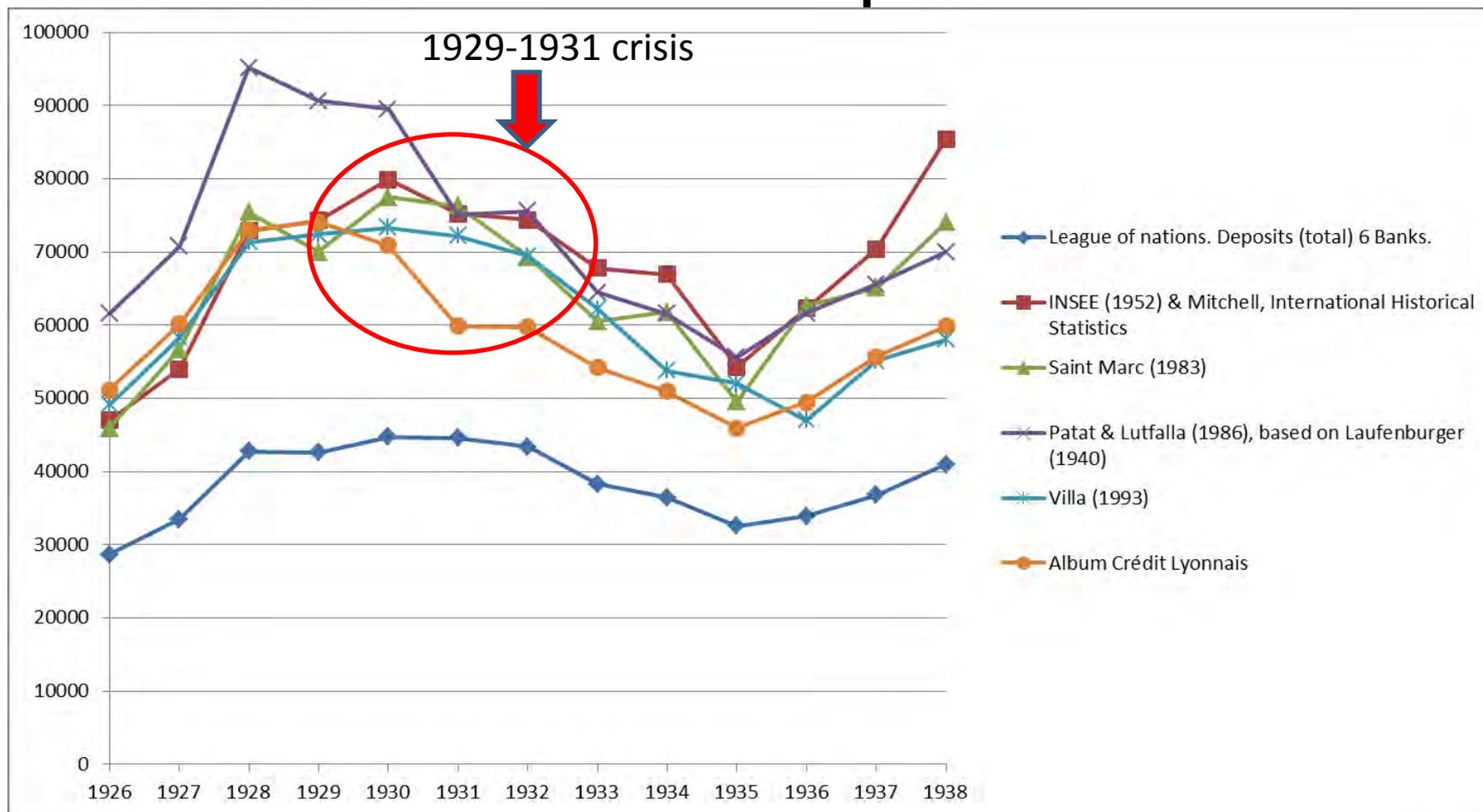
At the end of the day

- Many experiments
- Many failures
- Some success....

Top ten French Banks



Album Comparison with previous series of deposits



EURHISFIRM European Long-Run Firm Data

H2020 – Infrastructure Development Project



EURHISFIRM

- **RESEARCH INFRASTRUCTURE** that collects, connects, and shares **LONG-TERM HIGH-QUALITY DATA** on **EU COMPANIES**
- Innovative technologies to spark the “**BIG DATA REVOLUTION in HISTORICAL SOCIAL SCIENCES**”: Scaling up quantity, quality and variety of available “born-on-paper” data on EU companies
- “Flexible” OCR technologies + data warehouse + friendly browsers and data visualisation

Thank you!



Safeguarding irreplaceable financial information

Djordje Hinic, Business Development Manager



**Understanding
Forecasting
Better decisions**





1. authenticity

2. cost-saving

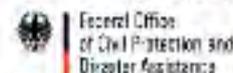
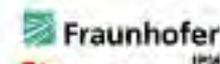
Data security & storage problem



€ 28 mil



ultimate data
solution



VISITECH

P+S TECHNIK
Pulsed Power Technology

IR=VISION

Oops, your files have been encrypted! English



What Happened to My Computer?

Your important files are encrypted. Many of your documents, photos, videos, databases and other files are no longer accessible because they have been encrypted. Maybe you are busy looking for a way to recover your files, but do not waste your time. Nobody can recover your files without our decryption service.

Can I Recover My Files?

Sure. We guarantee that you can recover all your files safely and easily. But you have not so enough time. You can decrypt some of your files for free. Try now by clicking "Decrypt". But if you want to decrypt all your files, you need to pay. You only have 3 days to submit the payment. After that the price will be doubled. Also, if you don't pay in 7 days, you won't be able to recover your files forever. We will have free events for users who are so poor that they couldn't pay in 6 months.

How Do I Pay?

Payment is accepted in Bitcoin only. For more information, click "About Bitcoin". Please check the current price of Bitcoin and buy some bitcoins. For more information, click "How to buy bitcoins". And send the correct amount to the address specified in this window. After your payment, click "Check Payment". Best time to check: 9:00am - 11:00am [Bitcoin](#) [Blockchain](#) [Bitcoin](#)

Payment will be raised on 5/15/2017 14:57:41

Time Left

02: 23: 59: 02

Your files will be lost on 5/15/2017 14:57:41

Time Left

06: 23: 59: 02

[Show Offer](#)

[Check our Bitcoin](#)

[Contact Us](#)

Send 1300 worth of bitcoin to this address:

115p7UMMngoj1pMvkpHjcrDRJNXj6LrLn

Copy

[Check Payment](#)

biol


Home
Pages
Blog
Contact

Cyber-Safe
0

Ransomware attack: Who's been hit

by [Julien Vautier](#) and [Jill Ople](#) #OWTech
03/15/2017 10:00 AM EDT

Cyber-Safe


Massive ransomware at countries

Analysis: Global ransomware attack why Apple wouldn't hack terror

[More on ransomware read it](#)

Telefónica hack: Ransomware attack on internal network forces computer shut down

[How the ransomware attack on Telefónica leaked online](#)

[Ransomware: World's largest ransomware attack](#)

The world's biggest cyberattack has hit at least 150 countries and infected 300,000 machines since it started spreading last Friday.

The victims include hospitals, universities, manufacturers and government agencies in countries like Britain, China, Russia, Germany and Spain.



“best practice” myth



Lessons



Ultra-secure data storage:



Unalterable

It must be impossible to modify or delete.



Secure

It must be protected from cyber attacks, logical threats, EMP and physical threats.

Common



Flexible

It should be able to store any kind of files; digital and visual.



Searchable

It must be searchable.

Long-term digital preservation



Migration-free:

It must be migration-free to avoid the risk of data loss and the migration cost.



Future-proof

Data retrieval process must be independent of technological obsolescence.



Permanent

It must be scientifically tested for 500 + years longevity.

“Unplug”

Virgil Gligor



An integrated turnkey solution

piqlWriter

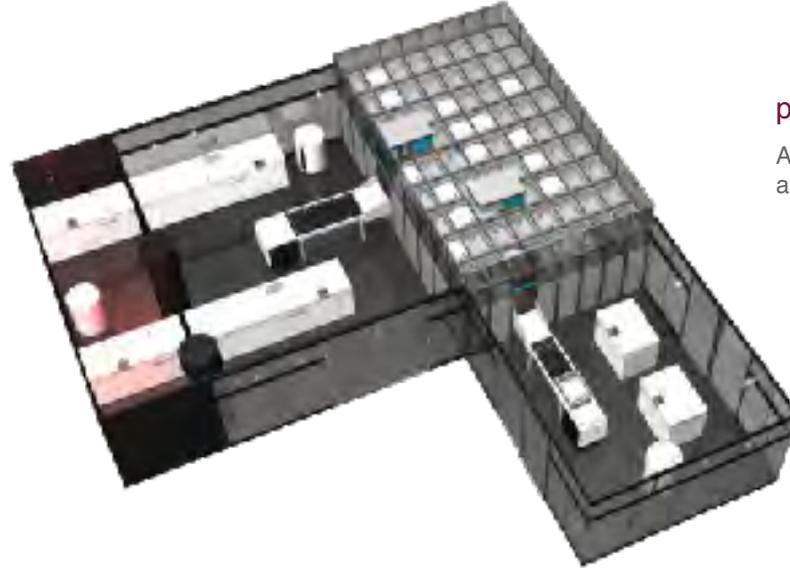
Writes data onto the piqlFilm. The piqlWriter is a high-speed industrial grade data writer utilizing Piql's proprietary, sophisticated software

piqlProcessor

Develops the piqlFilm and makes the data readable and permanent

piqlBox

A box/cartridge developed to protect the piqlFilm. The piqlBox constitutes newly developed polymers with 500 years + longevity



piqlVault

A robotic vault for safe, space efficient and automated storage of piqlBoxes

piqlReader

Reads data off the piqlFilm using open source software

piqlFilm

A newly developed nano-technology 35mm ultra-high resolution film optimized for digital storage, with documented 500 years lifespan





and many more...



Instability





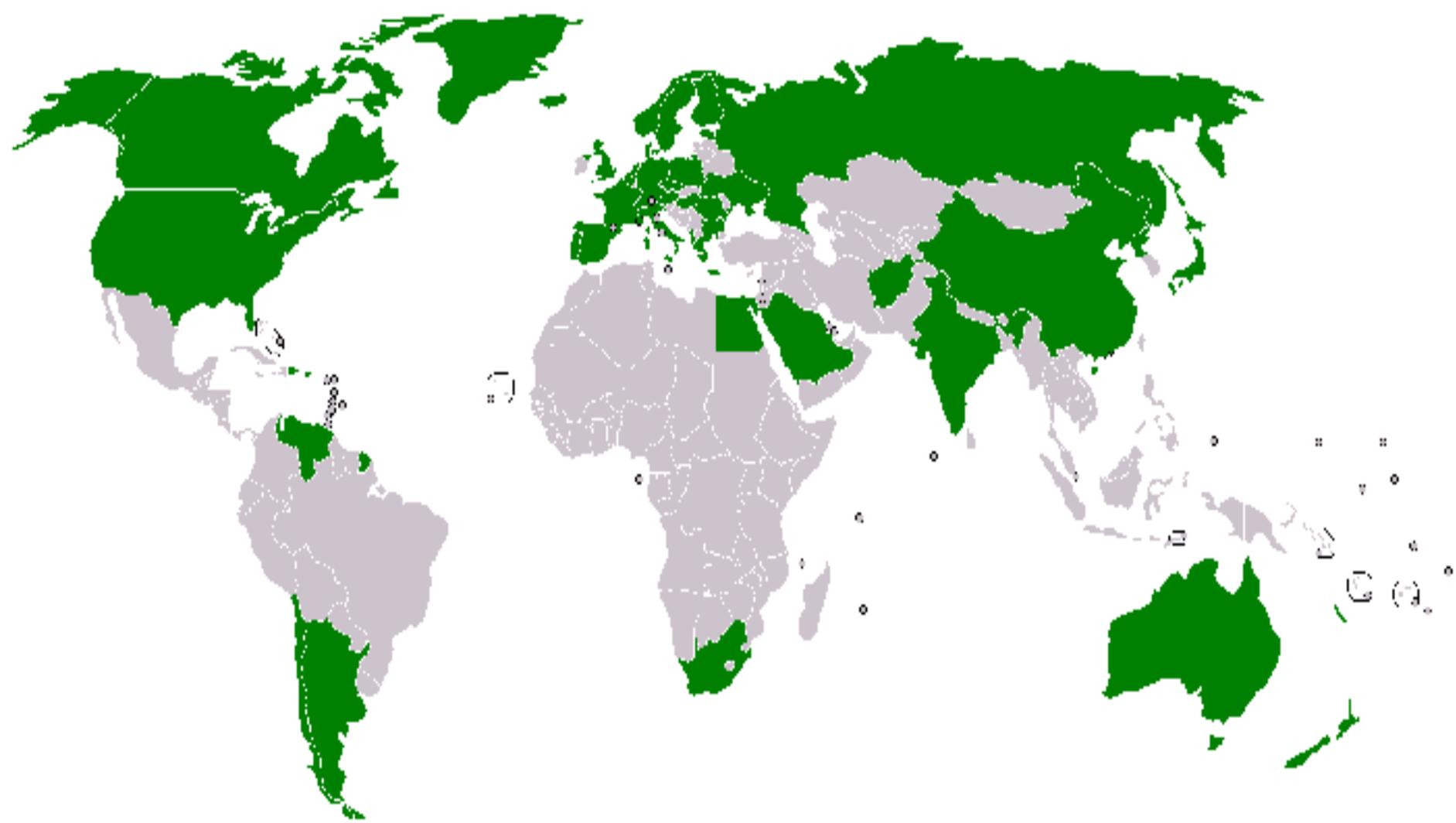
Safest place?

Seed vault

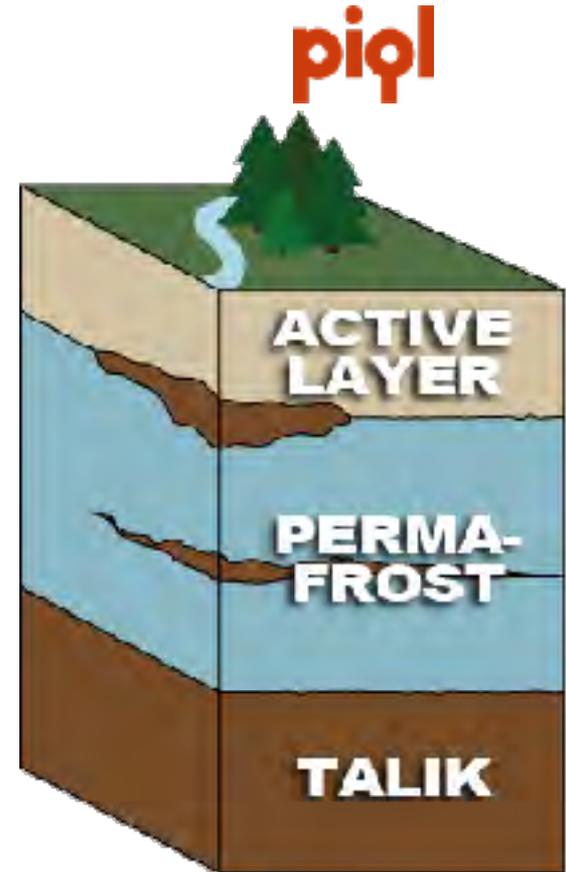




**Arctic
World
Archive**



Protected by polar bears & permafrost



Arctic 'doomsday' vault seeks to protect world's most precious data

By Heidi Ledford, CNN
@heidledford



Just the entry tunnel (and vault), the seed vault is located underground on a remote, high island.

Story highlights
CNN — Deep-frozen and abandoned mine on the Arctic island of Svalbard soon to hold 1.1 million seeds from...



News & Life
Arctic 'doomsday' vault seeks to protect world's most precious data
Arctic 'doomsday' vault seeks to protect world's most precious data

Viral



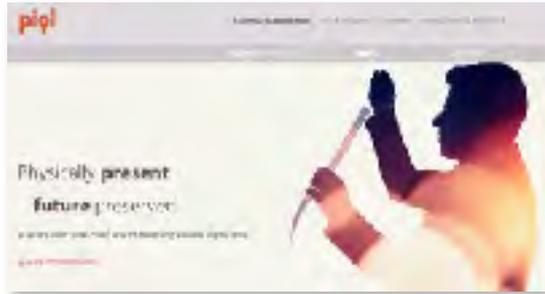
Svalbard Will Soon Have Another "Doomsday Vault" For Storing Precious Literature

28.0K SHARES

Share on Facebook | Share on Twitter | +



Interested in learning more?



www.piql.com



<https://vimeo.com/106280961>



“What we do
- behind the scenes”
<http://cld.bz/8JK9G0y>



“When quality matters”
<http://cld.bz/h38aCa>



“Alternative storage
technologies” <http://cld.bz/2zaEa>



Follow our updates on
[LinkedIn](#) | [Facebook](#)



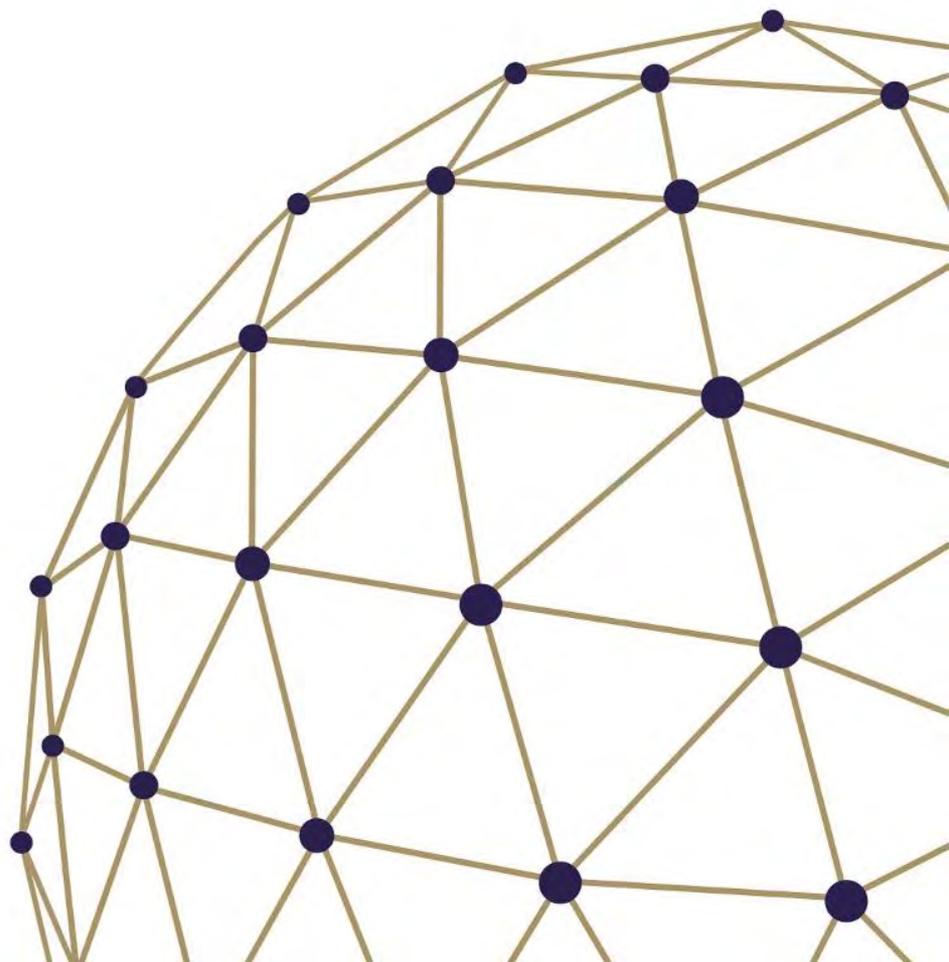
The data dilemma: a risk or an asset ?

Privacy, confidentiality, security and consumer protection

Kertész Ákos

Senior Supervisor

Zagreb, 10.11. 2017





Introduction to basics

What is data ?

Definition

Data is a set of values of qualitative or quantitative variables. Pieces of data are individual pieces of information. Data becomes information by interpretation. **Data is a series of symbols**, while information occurs when the symbols are used to **refer to something**.

Usage

Data is measured, collected and reported, and analyzed, whereupon it can be visualized using graphs, images or other analysis tools. Data as a general concept refers to the fact that **some existing information or knowledge is represented or coded in some form suitable for better usage or processing**.

What kind of data?

Financial data

Financial data consists of pieces or sets of **information related to the financial health of a business, or a person**.

People and organizations outside a business will also use financial data reported by the business to judge its credit worthiness, decide whether to invest in the business, and determine whether the business is complying with government regulations.

Big Data

The term 'Big Data' is used to describe the collection and analysis of data on a scale or of a complexity that makes the use of such data challenging.



Big Data

Financial usage

Example

Credit scoring

Traditional credit scoring use information provided by customer (payment history, types of credit used, length of credit history,..). In case the customer has insufficient credit history, Banks **can not calculate the risk**.

In the era of 'Big data', credit decisions may be based not only on credit information but also on a wide variety of **non-traditional data** that are not directly related to creditworthiness. Following an approach that "all data is credit data", Banks can identify patterns and habits of customers, which may drive to creditworthiness, so new customers can access to credit.





Which are the risks related to data?

Gathering, storing, processing

Cost of processing



Bad Analytics

Data Security

Data Privacy

Bad Data



Risks

Cost of data processing

Huge amount of data

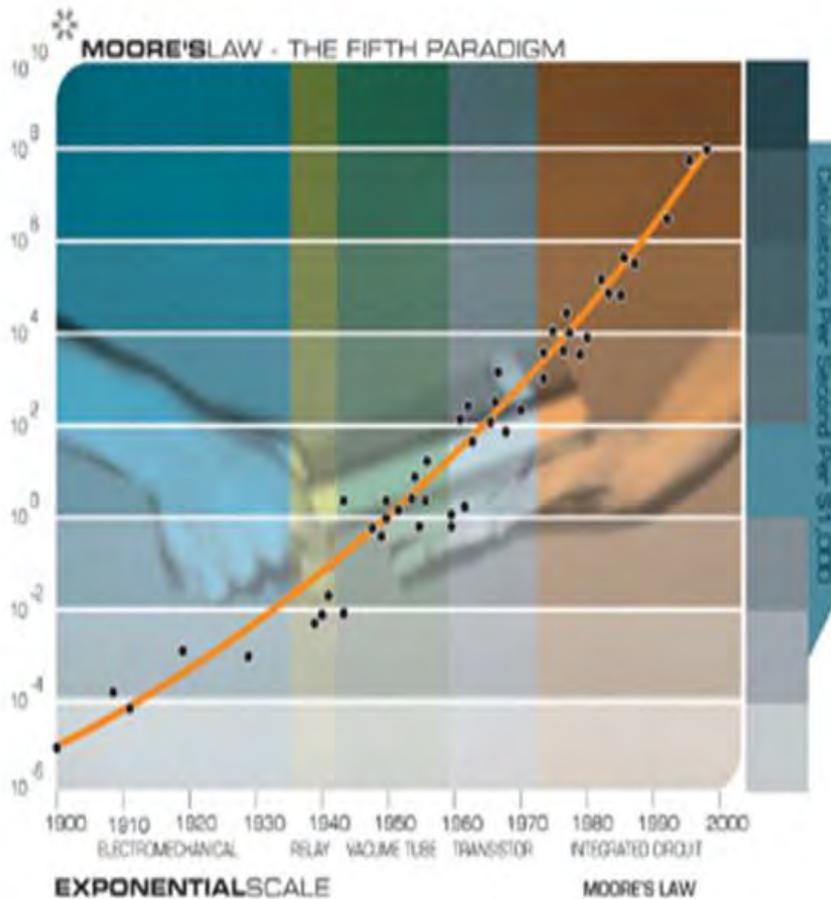
- The data volumes are exploding, more data has been created in the **past two years** than in the **entire previous history** of the human race.
- Every second we create new data. For example, we perform 40,000 search queries every second, which makes it 3.5 billion searches per day and 1.2 trillion searches per year.
- Social Network users send on average 31.25 million messages and view 2.77 million videos **every minute**.
- Over 1.4 billion smart phones were shipped in 2015 - all packed with sensors capable of collecting all kinds of data, not to mention the data the users create themselves.
- Within five years there will be over 50 billion **smart connected devices** in the world, all developed to collect, analyze and share data.
- At the moment **less than 0.5%** of all data is ever **analyzed** and used.



Risks

Cost of data processing

Increased computational power



The Law of Accelerating Returns

Technological change is exponential, contrary to the common-sense “intuitive linear” view.

Computer speed (per unit cost) doubled every three years between 1910 and 1950, doubled every two years between 1950 and 1966, and is now doubling every year. Chip speed and cost-effectiveness, also increase exponentially.



Risks

Data Security

The financial sector is under fire

In general

- Financial data is a popular target for criminals and **cyber-attacks**, however the need of **sharing information**, and the potential benefits of analyzing data have growing tendency.
- Cyber criminals can **easily monetize** the wealth of data financial institutions collect – either by selling that data on the dark web or using the data to conduct fraud.

In details

- The financial services industry is the most breached industry, accounting for 35% of data breaches.
- 68% of financial services firms experienced multiple successful attacks in 2016.

Accessing systems to fraudulently transfer money or using personal information of customers for identity theft are two examples of financially motivated misuse.



Risks

Data Security

	Incidents				Breaches			
	Total	Small	Large	Unk	Total	Small	Large	Unk
Total	42,068	606	22,273	19,189	1,935	433	278	1,224
Accommodation (72)	215	131	17	67	201	128	12	61
Administrative (56)	42	6	5	31	27	3	3	21
Agriculture (11)	11	1	1	9	1	0	1	0
Construction (23)	6	3	1	2	2	1	0	1
Education (61)	455	37	41	377	73	15	15	43
Entertainment (71)	5,534	7	3	5,524	11	5	3	3
Finance (52)	998	58	97	843	471	39	30	402
Healthcare (62)	458	92	108	258	296	57	68	171
Information (51)	717	57	44	616	113	42	21	50
Management (55)	8	2	3	3	3	2	1	0
Manufacturing (31-33)	620	6	24	590	124	3	11	110
Mining (21)	6	1	1	4	3	0	1	2
Other Services (61)	69	22	5	42	50	14	5	31
Professional (54)	3,016	51	21	2,944	109	37	8	64
Public (92)	21,239	46	20,751	442	239	30	59	150
Real Estate (53)	13	2	0	11	11	2	0	9
Retail (44-45)	326	70	36	220	93	46	14	33
Trade (42)	20	4	10	6	10	3	6	1
Transportation (48-49)	63	5	11	47	14	3	4	7
Utilities (22)	32	2	5	25	16	1	1	14
Unknown	8,220	3	1,069	7,128	68	2	15	51
Total	42,068	606	22,273	19,189	1,935	433	278	1,224

Table 1: Number of security incidents by victim industry and organization size, 2016 dataset.



Risks

Data Security

Cyber threats

Financial services firms fight an escalating and asymmetric war against cyber-attacks and internal threats. To effectively prioritize cyber defenses, financial institutions must understand the cyber threats they are up against.

The most frequent cyber security threats financial services firms must address:

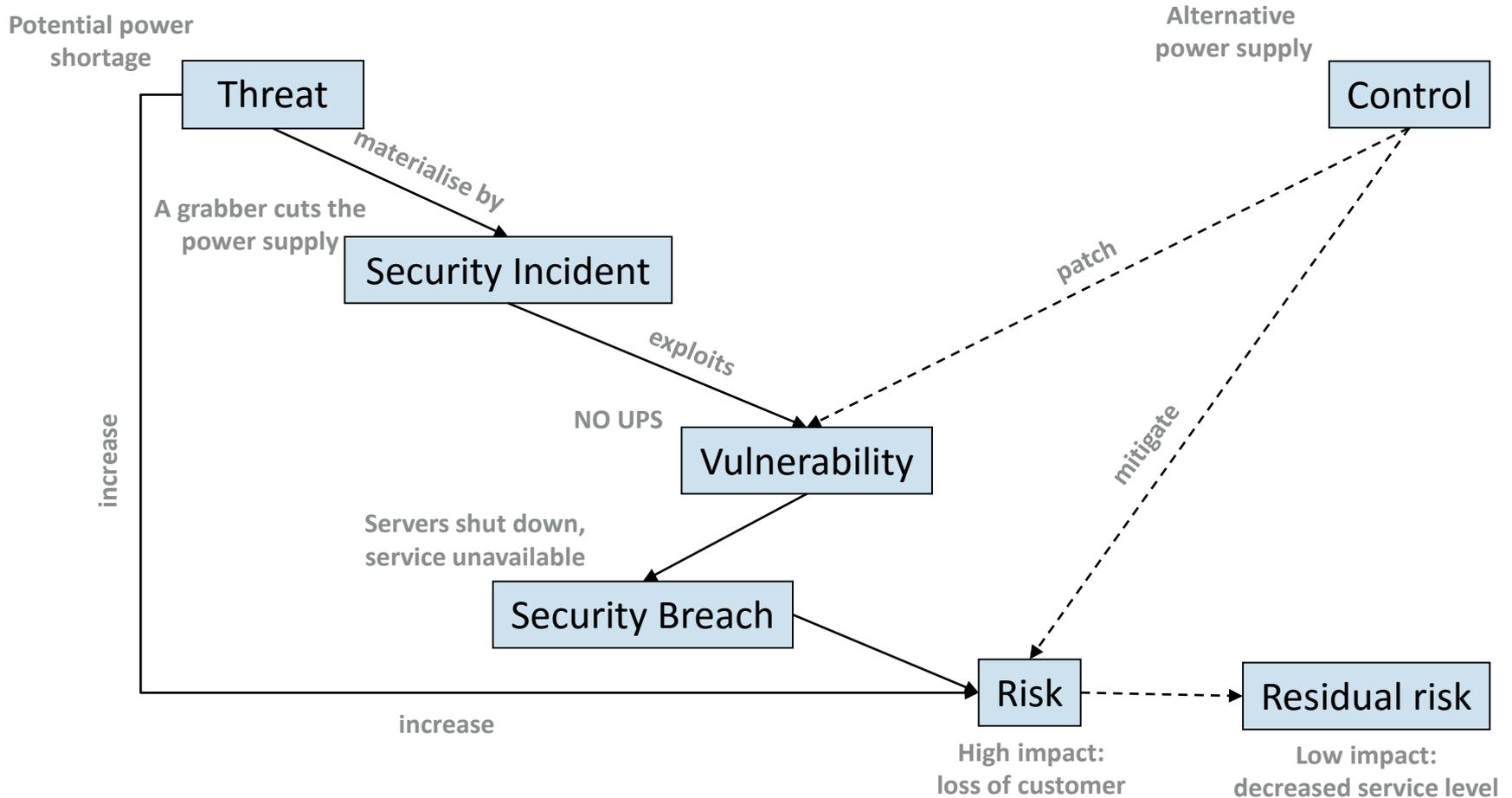
1. Distributed Denial of Services (DDoS) attacks
2. Web application attacks
3. Data Theft or Breach
 - In 86% of cases where data was stolen, financial sector systems were compromised in minutes or less.
 - In 69% of cases, financial services victims didn't discover a security incident for weeks or months.
4. Insider Threats
 - 30% of phishing messages were opened
 - 12% of targets clicked to open the malicious attachment



Risks

Data Security

Correlation

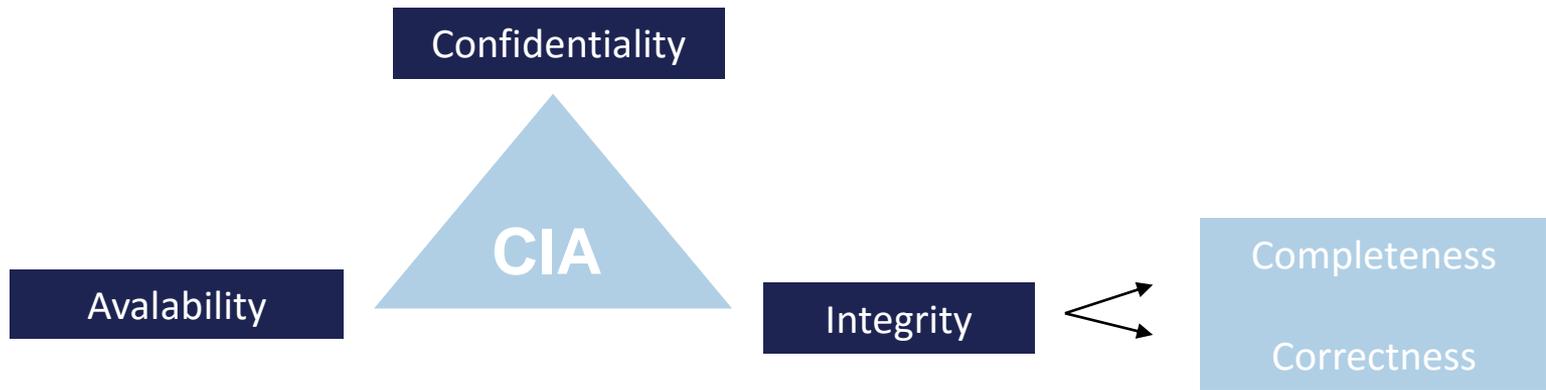




Risks

Data Security

Fundamentals of security and control measures



Preventative	Detective	Corrective	Compensatory
Security Awareness Training	System Monitoring	OS Upgrade	Backup Generator
Firewall	IDS	Backup Data Restoral	Hot Site
User Right Management	Anti-Virus	CSIRT	Encryption
Security Guard	Motion Detector	Vulnerability Mitigation	...



Which are the risks related to data?

Gathering, storing, processing

Cost of processing



Bad Analytics

Data Security

Data Privacy

Bad Data



Risks

Bad data, bad analytics

Data Scientist

Data scientists combine statistics, mathematics, programming, problem-solving, capturing data in ingenious ways, the ability to look at things differently to find patterns, along with the activities of cleansing, preparing, and aligning the data.



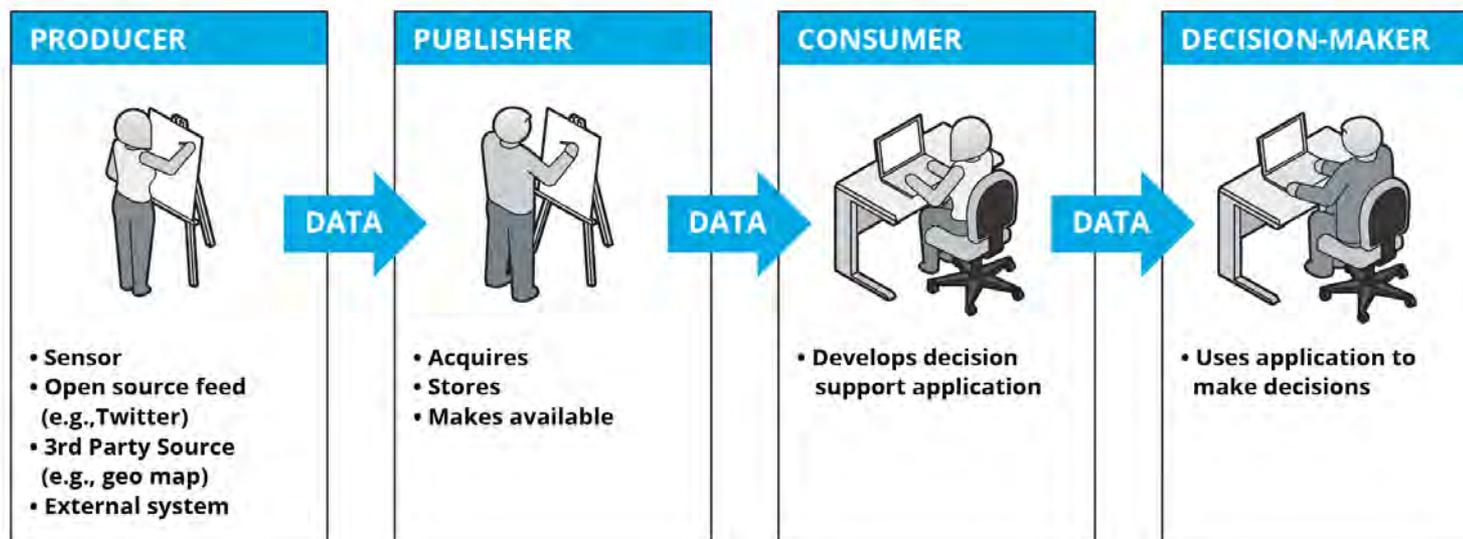
- New channels and new data
- Complexity of interactions
- Data quality and consistency
- Extracting business value
- Lack of Big Data skills



Risks

How to use it, and for what?

Data has a value chain



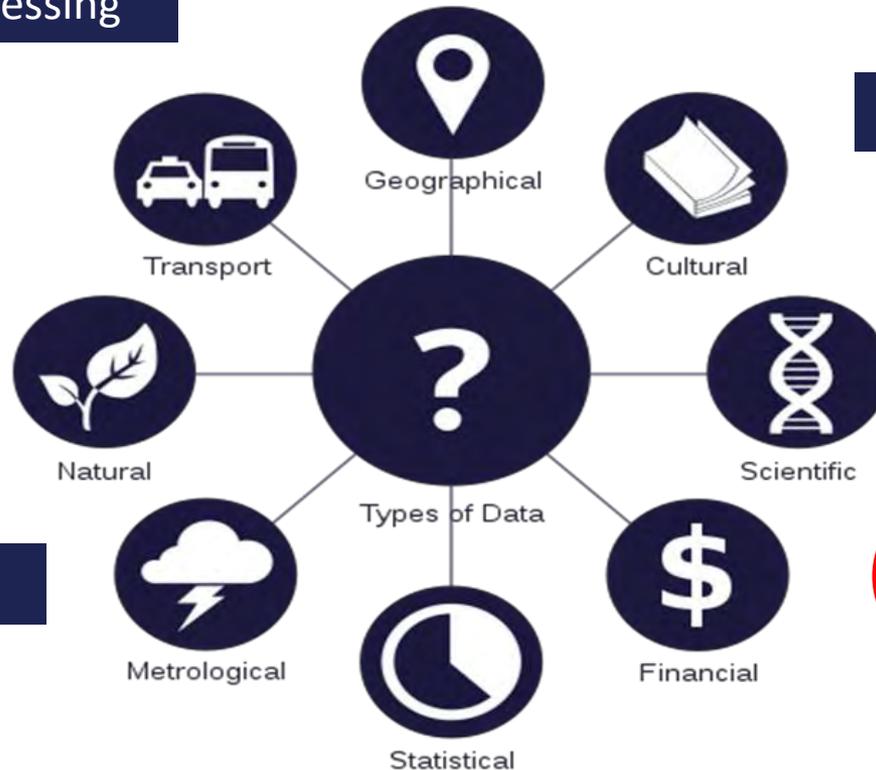
Customer – Fincial Institution – Fintech Company– Added Value



Which are the risks related to data?

Gathering, storing, processing

Cost of processing



Bad Analytics

Data Security

Data Privacy

Bad Data



Risks

Data Privacy

Data Privacy



- Data vulnerability and fraud ?
- Transparency and trust ?
- Understanding of data usage ?

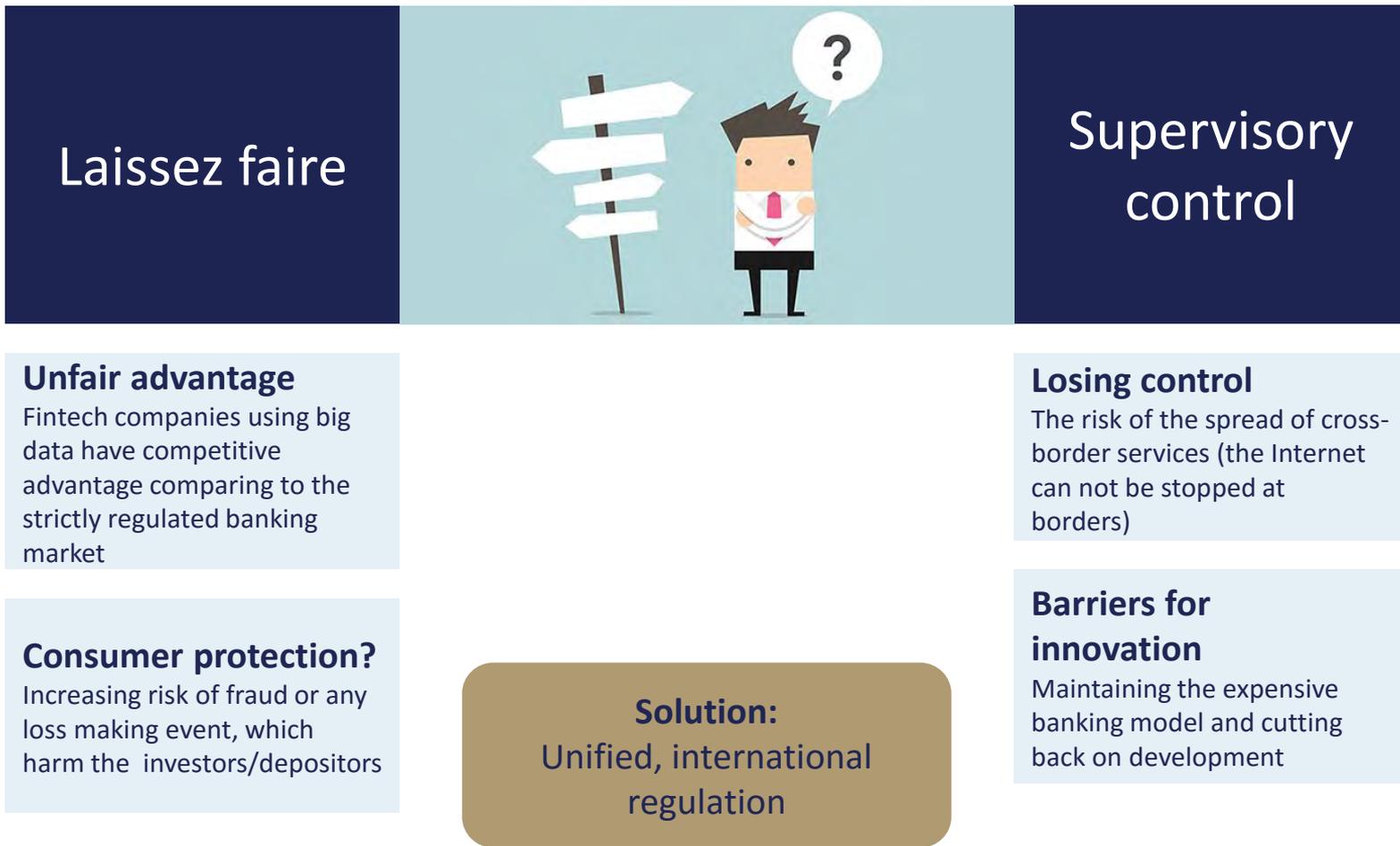


- Data security
- Outsourcing, partnership
- Need to know, need to do
- Consumer education



Supervisor's dilemma

Consumer, data and market protection





Regulations

Consumer, data and market protection

Regulations

Unfair Commercial Practices Directive

E-Commerce Directive

Banking secrecy rules impose restrictions to the use of consumer data by financial institutions

Directive on Electronic Communications and Privacy

Unfair Contract Terms

Directive on Distance Marketing of Financial Services

Payment Services Directive

recommendation on the use of community and public **cloud** services (HU)

GDPR
the 'right to be forgotten'
access to one's own data
the right of data portability

Consumer Credit Directive

Mortgage Credit Directive

Payment Accounts Directive



Summary

Conclusion

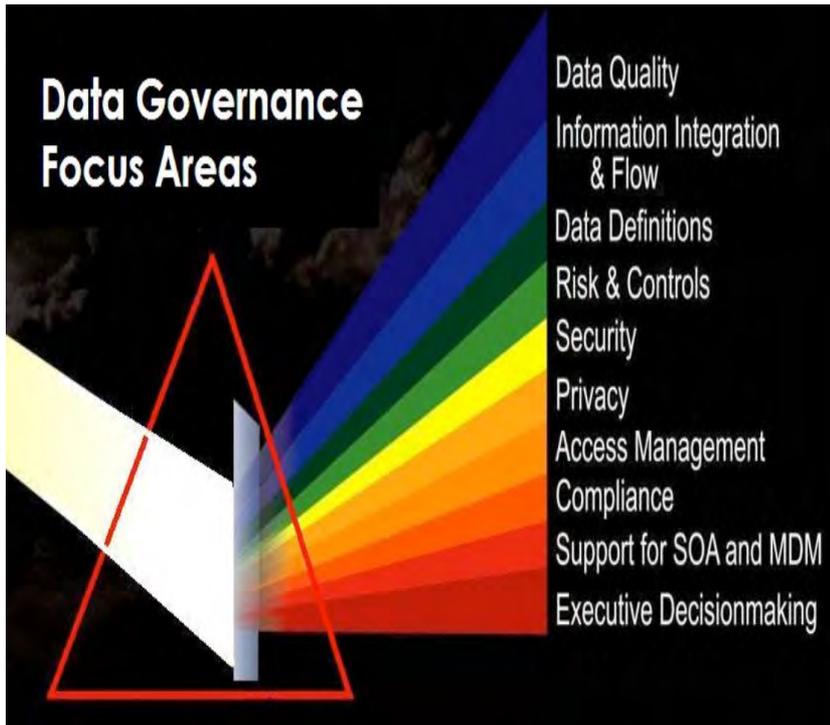
To be successful on these topics is not only prepare more strict regulations and provide an even more secure infrastructure, but institutions should handle data in a different way.

Collecting, storing and using data is not enough, we should provide a structured **Data Governance** approach.

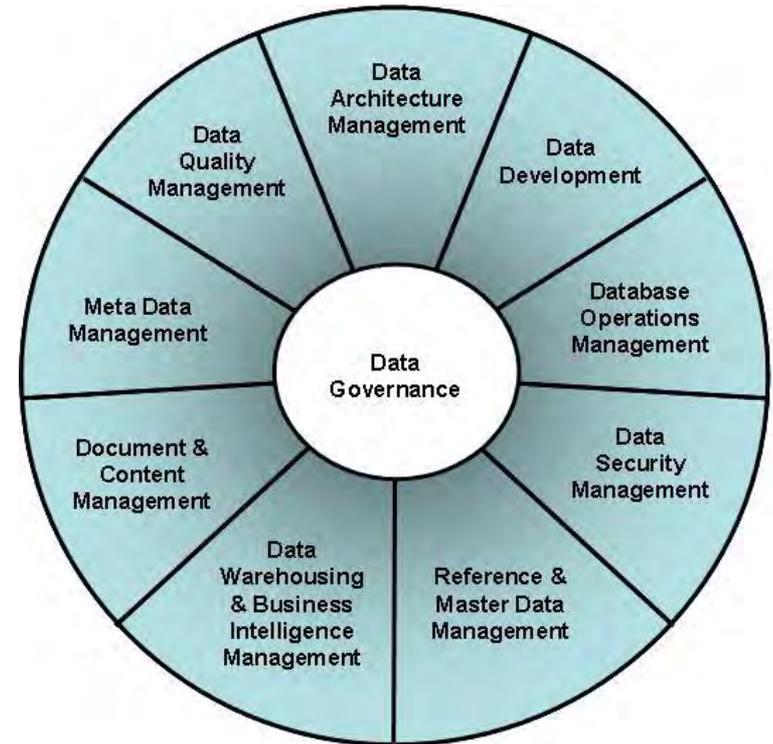




Data Governance Methodology



Source: The Data Governance Institute (2014)



Source: DAMA International (2009)



Thank you for your attention!

Kertész Ákos CISA, CISM
kertesza@mnb.hu



Appendix

Wikipedia (2017): Data

<https://en.wikipedia.org/wiki/Data>

Forbes (2015): Big Data: 20 Mind-Boggling Facts Everyone Must Read

<https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#5acafb0f17b1>

Ray Kurzweil (2001): The Law of Accelerating Returns

<http://www.kurzweilai.net/the-law-of-accelerating-returns>

Imperva (2016): Top 4 cyber threats facing the financial services industry

<https://www.imperva.com/blog/2016/07/top-4-cyber-threats-facing-the-financial-services-industry/>

Verizon (2017): Data Breach Investigations Report 10th Edition

<http://www.verizonenterprise.com/verizon-insights-lab/data-breach-digest/2017/>

SEI (2017): Six Things You Need to Know About Data Governance

https://insights.sei.cmu.edu/sei_blog/2017/06/six-things-you-need-to-know-about-data-governance.html

EBA (2016): Innovative uses of consumer data by financial institutions

<http://www.eba.europa.eu/-/eba-publishes-report-on-consumer-data-and-identifies-a-number-of-applicable-requirements-under-eu-law>

The Data Governance Institute (2014): The DGI Data Governance Framework

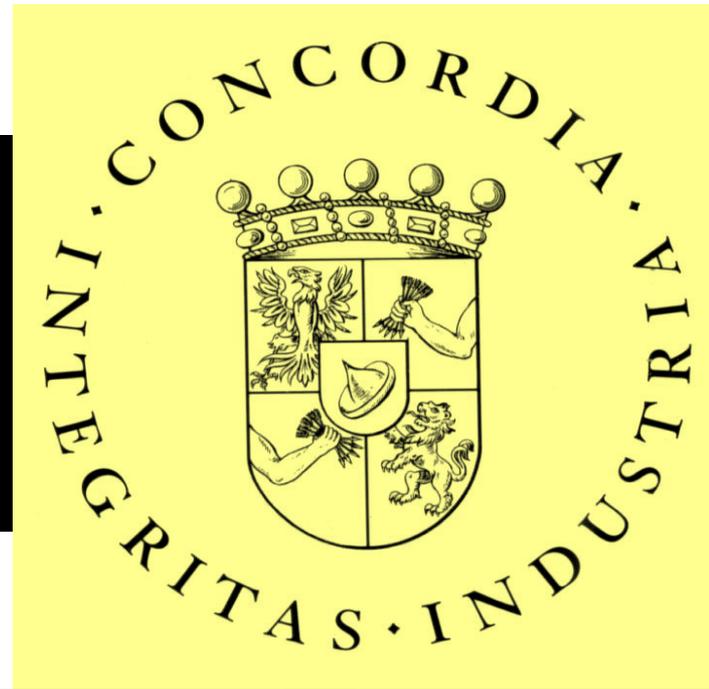
http://www.datagovernance.com/wp-content/uploads/2014/11/dgi_framework.pdf

DAMA International (2009): Guide to the Data management Body of Knowledge

<https://www.dama.org/content/body-knowledge>

What happened in the Daily Gold Fixings Auctions 1919 to 1968 - The Missing Data

Dr. Fergal O'Connor
Associate Professor of Finance
The York Management School



The London Gold Market

- Roots in trade between the East India Company and Moses Mocatta in late 17th Century
- Historically the Worlds Gold Market
- Still the Largest Market by Volume
- Significant recent changes

What was the Gold Fixing?

- Daily Meeting at NM Rothschild and Son
- 5 participants in the beginning
- Auction where the price and quantity was/is allowed to vary



Academic Gold Research

- Lots of research on the Macroeconomic aspects of the gold standard
 - Relies on monthly and annual gold price data
 - Assumes gold price fixed by CB's
- Lots of research on the Financial Economics of precious metals markets post 1968 (O'Connor et al. 2015)
 - Twice daily data available from the London Market

Where was the data?

- Royal Mint Report 1919-1925
- Quins Metal Handbook and statistics 1919-1965

Established 1913

THE METAL BULLETIN

THE LEADING METAL TRADES PAPER

PUBLISHED EVERY

TUESDAY & FRIDAY

Indispensable to all seeking reliable, concise and up-to-date market reports, prices and statistics, etc. Specimen copy free on demand from the Publishers:—

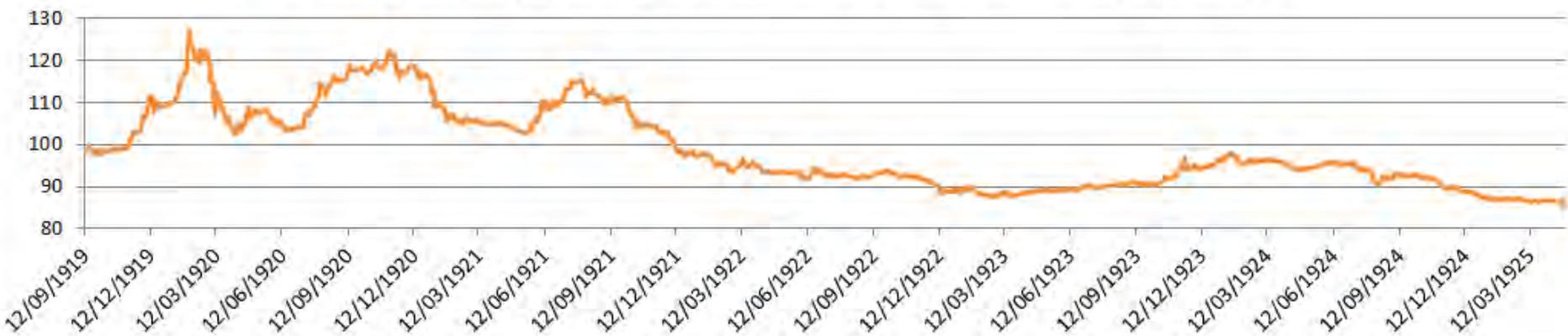
METAL INFORMATION BUREAU LTD
IBEX HOUSE • MINORIES • E.C.3

197

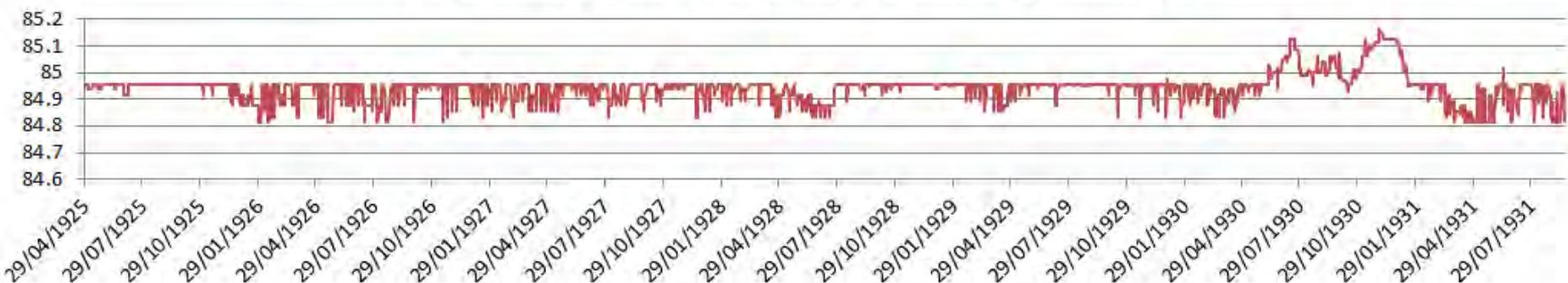
GOLD—DAILY PRICES, 1925 (Fine per oz.)

1925	Gold	1925	Gold	1925	Gold
Jan. 1	87/9	Mar. 5	86/9	May 9	84/11 $\frac{1}{2}$
2	87/9	6	86/8	11	84/11 $\frac{1}{2}$ nom.
5	87/4 nom.	9	86/8	12	84/11 $\frac{1}{2}$
6	87/7	10	86/9	13	84/11 $\frac{1}{2}$
7	87/5	11	86/8	14	84/11 $\frac{1}{2}$
8	87/4	12	86/7	15	84/11 $\frac{1}{2}$
9	87/2	13	86/4	16	84/11 $\frac{1}{2}$
12	86/11	16	86/4	18	84/11 $\frac{1}{2}$
13	87/1 nom.	17	86/7	19	84/11 $\frac{1}{2}$
14	87/8 nom.	18	86/8	20	84/11 $\frac{1}{2}$
15	87/1 nom.	19	86/8	21	84/11 $\frac{1}{2}$
16	87/2 nom.	20	86/8	22	84/11 $\frac{1}{2}$
19	87/1 nom.	23	86/7	23	84/11 $\frac{1}{2}$
20	87/2	24	86/6	25	84/11 $\frac{1}{2}$ nom.
21	87/3	25	86/4	26	84/11 $\frac{1}{2}$
22	87/-	26	86/5	27	84/11 $\frac{1}{2}$
23	86/10	27	86/5	28	84/11 $\frac{1}{2}$
26	86/9	30	86/6	29	84/11 $\frac{1}{2}$
27	86/11	31	86/7	30	84/11 $\frac{1}{2}$
28	86/11	Apl. 1	86/6	June 2	84/11 $\frac{1}{2}$
29	87/-	2	86/6	3	84/11 $\frac{1}{2}$
30	87/1	3	86/6	4	84/11 $\frac{1}{2}$ nom.
Feb. 2	86/11	6	86/5	5	84/11 $\frac{1}{2}$
3	87/-	7	86/5	6	84/11 $\frac{1}{2}$
4	87/1	8	86/5	8	84/11 $\frac{1}{2}$
5	87/1 nom.	9	no price.	9	84/11 $\frac{1}{2}$
6	87/2	14	86/5	10	84/11 $\frac{1}{2}$
9	87/4	15	86/7	11	84/11 $\frac{1}{2}$
10	87/2	16	86/7	12	84/11 $\frac{1}{2}$
11	86/10	17	86/7	13	84/11 $\frac{1}{2}$
12	86/8	20	86/6 nom.	15	84/11 $\frac{1}{2}$
13	86/8	21	86/6	16	84/11 $\frac{1}{2}$
16	86/10	22	86/4	17	84/11 $\frac{1}{2}$
17	86/10	23	86/5	18	84/11 $\frac{1}{2}$
18	86/11	24	86/3	19	84/11 $\frac{1}{2}$
19	87/1	27	85/9	20	84/11 $\frac{1}{2}$
20	86/11	28	86/-	22	84/11 $\frac{1}{2}$
23	87/1	29	84/11 $\frac{1}{2}$	23	84/11 $\frac{1}{2}$
24	86/11	30	84/11 $\frac{1}{2}$	24	84/11 $\frac{1}{2}$
25	87/-	May 1	84/11 $\frac{1}{2}$	25	84/11 $\frac{1}{2}$
26	87/1	4	84/11 $\frac{1}{2}$	26	84/11 $\frac{1}{2}$
27	87/3	5	84/11 $\frac{1}{2}$	27	84/11 $\frac{1}{2}$
Mar. 2	86/10	6	84/11 $\frac{1}{2}$	29	84/11 $\frac{1}{2}$
3	86/10	7	84/11 $\frac{1}{2}$	30	84/11
4	86/9	8	84/11 $\frac{1}{2}$		

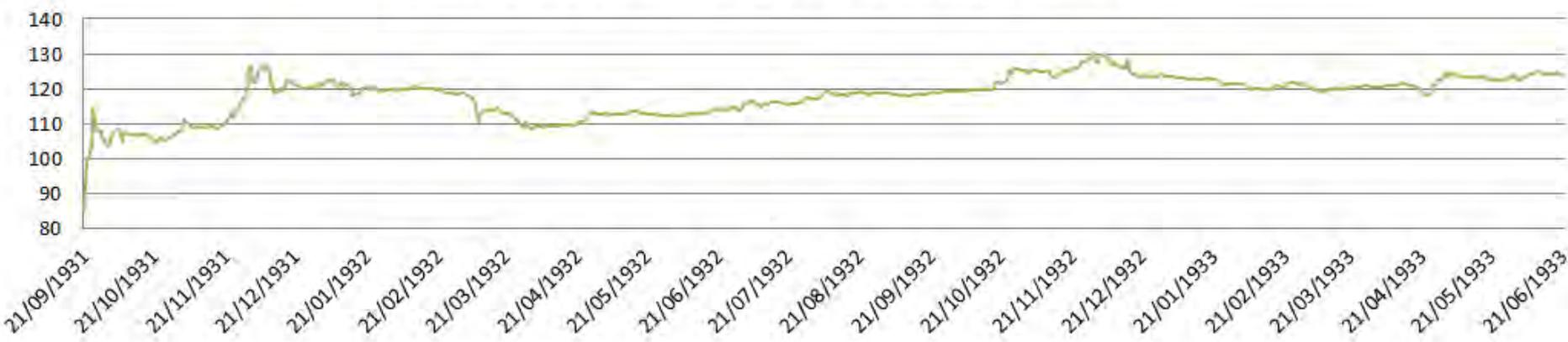
Gold price up to the Return to the Gold Standard, Shillings



Gold Price during the Gold Standard, Shillings



Gold Price after the Gold Standard, Shillings

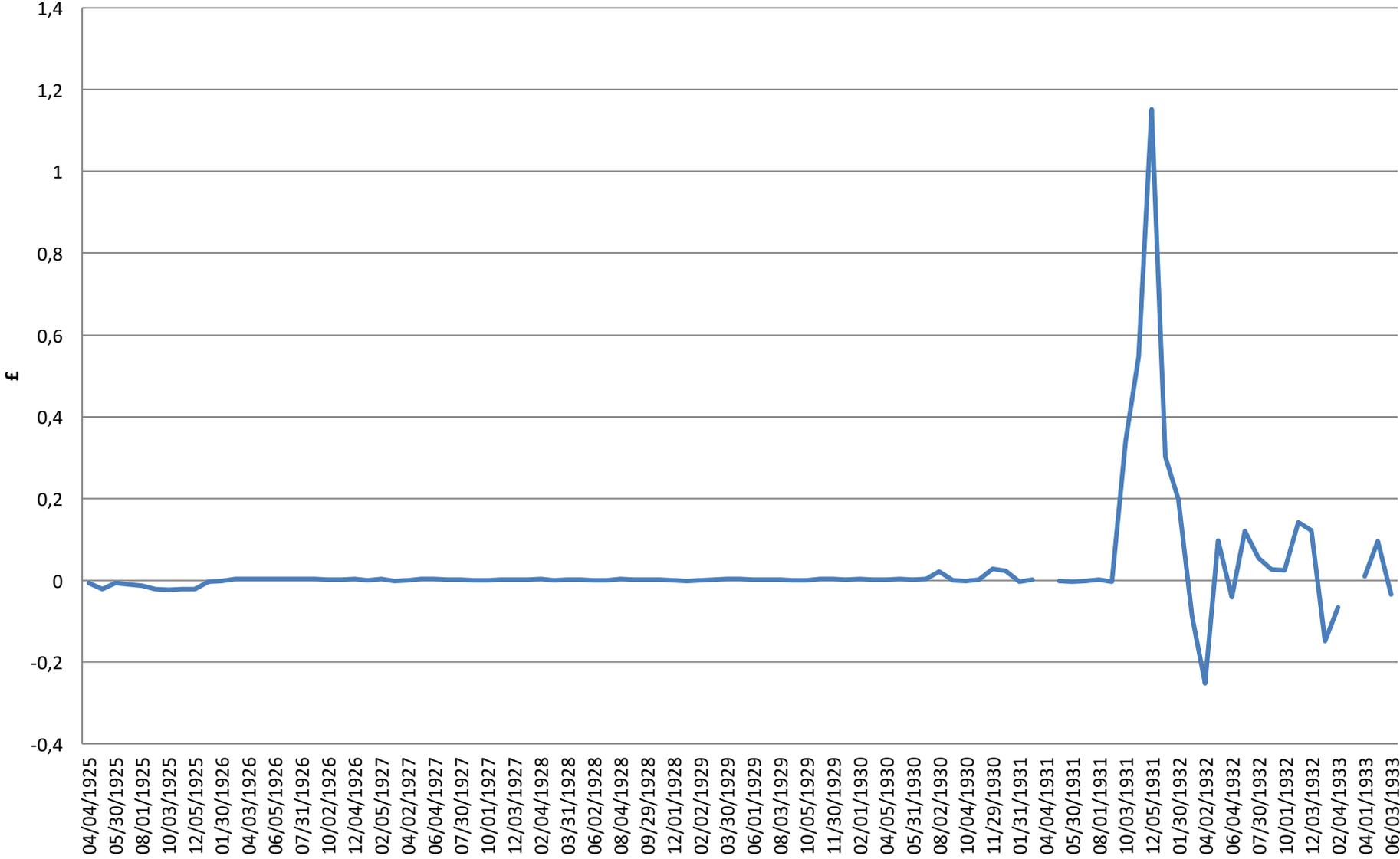


Other Sources of Gold Price Data

- **Annual** gold prices from 1257 freely available from MeasuringWorth.com
- **Monthly** gold prices are available from 1723 to 1968 and daily thereafter from the Global Financial Database
- **Daily** gold prices are freely available from the London Bullion Market Association website from 1968

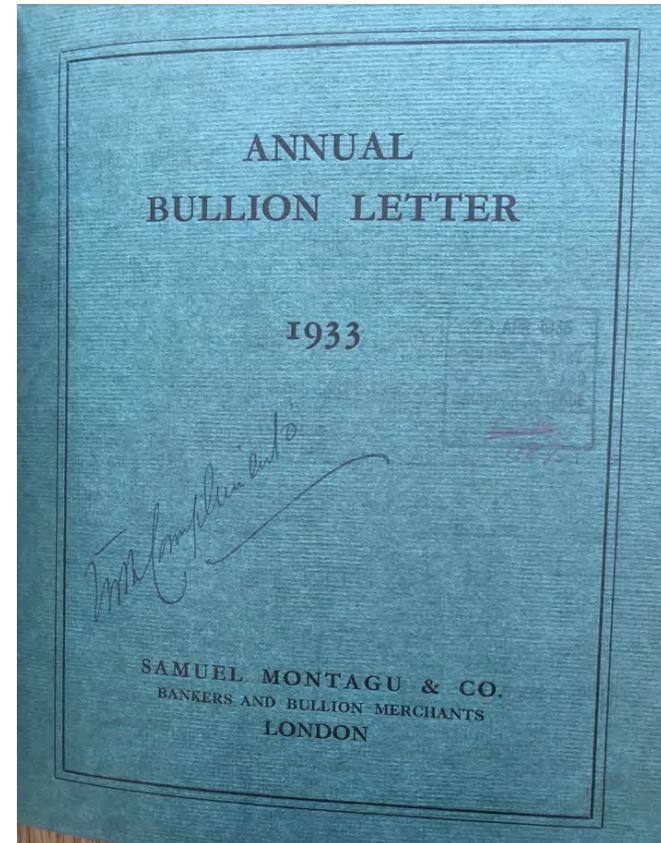


Errors in Current Data Global Financial Database Vs. Quin's



Other aspects to the project

- Qualitative History of the London Gold Market
 - Based on sources such as Samuel Montagu & Co.'s Annual Bullion Letter
- Daily Silver Fixings Data 1913-1970
- Daily LME data for Copper, Lead and Zinc 1913-1970



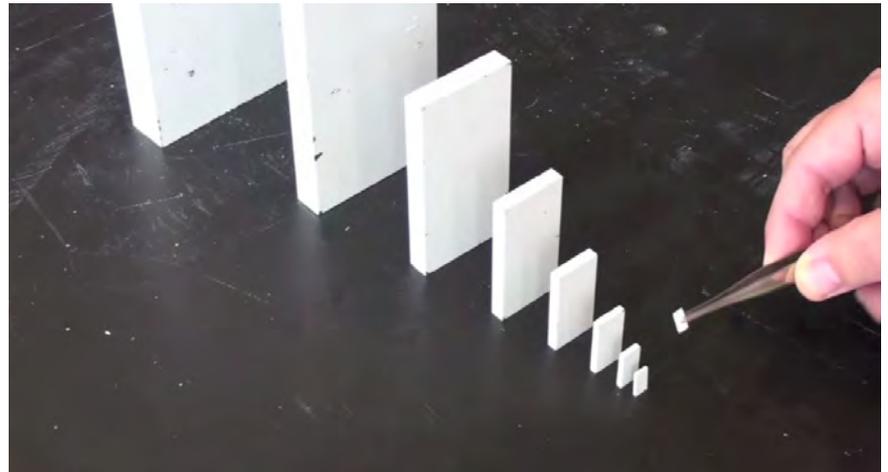
- **Thank you for Listening**
- **Questions and/or Suggestions
Welcome**

A Tale of Rain and Bank Runs

From Small To Big Data and Back

ANTON COMANESCU
National Bank of Romania

Zagreb, Croatia
10 November 2017



“Knowledge is nothing else than perception.”

PLATO

“Perceptions can make or brake policies, even the best ones”

Jacques Santer, Report to the EU Parliament

Structure:

- Tales about the Devil in the details
- From demand and supply of information to Perceptions

The Iranian revolution of 1979 and the diplomats with dirty shoes



- In the aftermath of the Iranian revolution the UK Foreign Office commissioned a secret inquiry into the failure of British diplomats to predict the events
- The inquiry found out that one problem was that the embassy in Tehran had little contact with the world outside the Shah's entourage
- Subsequent generations of diplomats learned the importance of **"ground truth"**
- One ambassador to Iran used then to check if staff's shoes were dirty

CIA, Kremlinology and the failure to predict the fall of USSR

- Churchill once characterized Russia as a "riddle wrapped in a mystery inside an enigma"
- During the Cold War, the West became obsessed with gathering data about the Soviet Union
- CIA had a Department of Kremlinology; The Soviet Studies, based on data about USSR, tended to exaggerate its strength and underestimate its flaws
- A CIA report and a National Security Directive of September 1989 are two of many documents famously failing to predict the withdrawal of USSR from its engagement in Eastern Europe and ultimately the fall of communism

Document No. 1: CIA Intelligence Assessment, "Gorbachev's

Domestic Gambles and Instability in the USSR"

September 1989

Gorbachev's rule at the top. For the time being, his power looks secure. If, somehow, a successful challenge were mounted against him over the next year or so, the most likely outcome would be a traditionalist restoration that would attempt to "draw the line" in various areas--

The character of the changes taking place in the Soviet Union leads to the possibility that a new era may be now upon us. We may be able to move beyond containment to a U.S. policy that actively promotes the integration of the Soviet Union into the existing international system. The

<https://nsarchive.gwu.edu/>

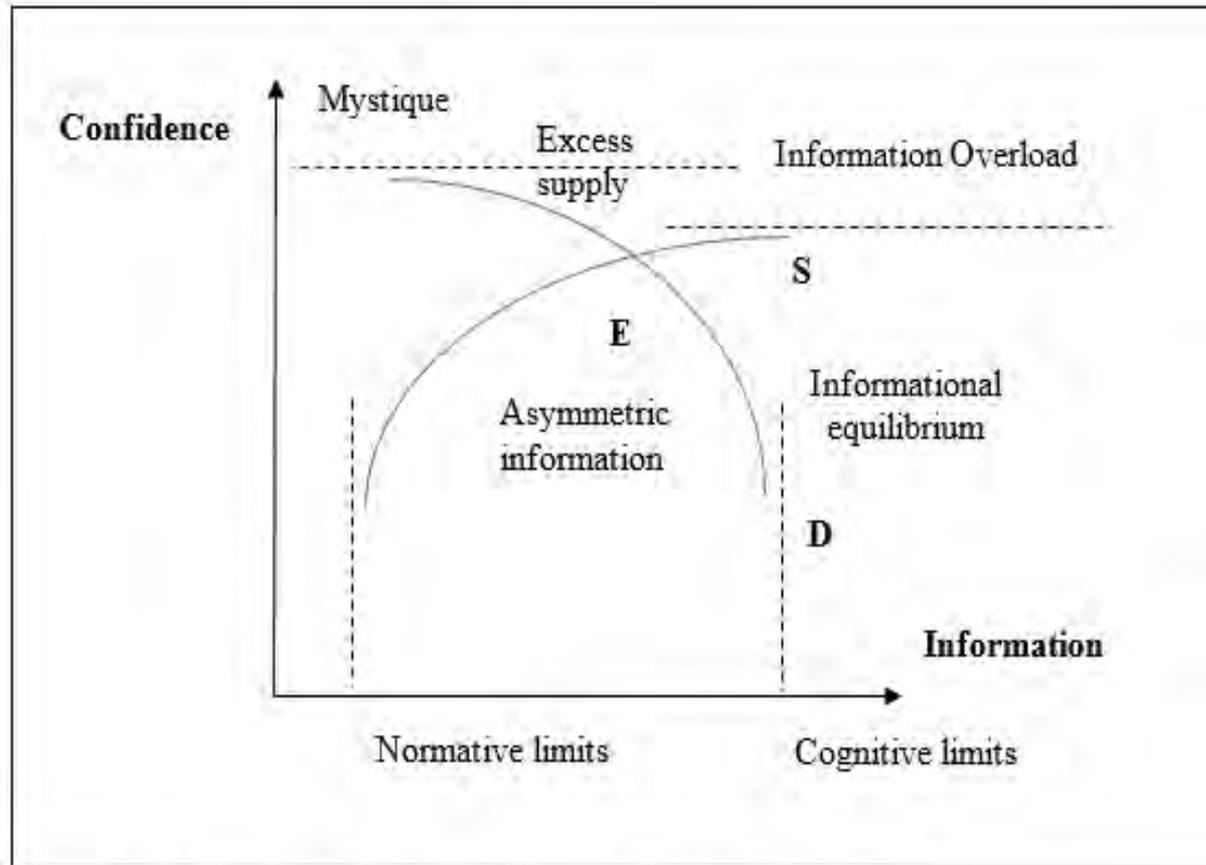


Make sense of big data but also dig for small data and “kick the tires” of the economy

- **Alan Greenspan** used his own private tennis club to check the mood of politicians and business leaders; he then used this first-hand information for his assessment of the economic outlook
- The members of the **ECB** Executive Board learned from the morning newspapers about the peak in the US residential mortgages delinquency rate in **August 2007**
- With unemployment at historical lows and reduced income inequality, 52 percent of UK voters have chosen, against all odds, to leave the EU
- **BBRD** Resolution algorithm to assess bank risks
- **Banking Union**: if we got it wrong, you all did
- **The Juncker Plan**: failing to pay attention to local culture



A model of demand and supply of information



**There is a limit of how much you
can learn about the economy
from data**

“We have never successfully modeled the transition from euphoria to fear.”

ALAN GREENSPAN

From data to Greenspan's Weltanschauung

“Well, remember what an ideology is. It's a conceptual framework with the way people deal with reality. Everyone has one. You have to. To exist you need an ideology. The question is whether it is accurate or not. And what I'm saying to you is yes, I have found a flaw. I don't know how significant or permanent it is. But I have been very distressed by the fact.”

The congressman questioning him asked: “In other words, you found that your view of the world, your ideology, was not right. It was not working?”

Greenspan replied: “Absolutely. Precisely. You know that's precisely the reason I was shocked. Because I have been going for forty years or more with very considerable evidence that it was working exceptionally well.”

October 23, 2008, Alan Greenspan's testimony to the Congress about his failure to predict Lehman

Thank you! and here is how to topple a skyscraper!



**Calling all archivists – the
five grand challenges of the
digital environment.**

Michael Moss
(Northumbria University)

A bar chart with four vertical bars of increasing height from left to right. The background is a sunset with a bright sun on the left and a gradient of orange and yellow. The bars are dark, almost black, and have a slightly irregular, hand-drawn appearance. The text 'Big Data' is centered in the upper half of the image.

Big Data

velocity, volume, veracity and variety



**Corpus of 1.7
million messages**

The five grand challenges:

- appraisal, what to keep to meet demand and expectations on a new scale.
- how to identify content that cannot safely be released - termed sensitivity review,
- long term preservation of digital objects, very different from paper,
- how users are going to explore and analyse content whose bulk precludes conventional cataloguing,
- finally who is to pay for all the new services?





What to keep?

No agreement

- No discernible order.
- Penalties for destruction.
- Capstone?
- Digital forensics approach is to capture all email content for analysis with no filtering, essential for public enquiries.
- Extinguishing duplicates and trivia – there is a great deal left.
- Need to experiment before it is too late.



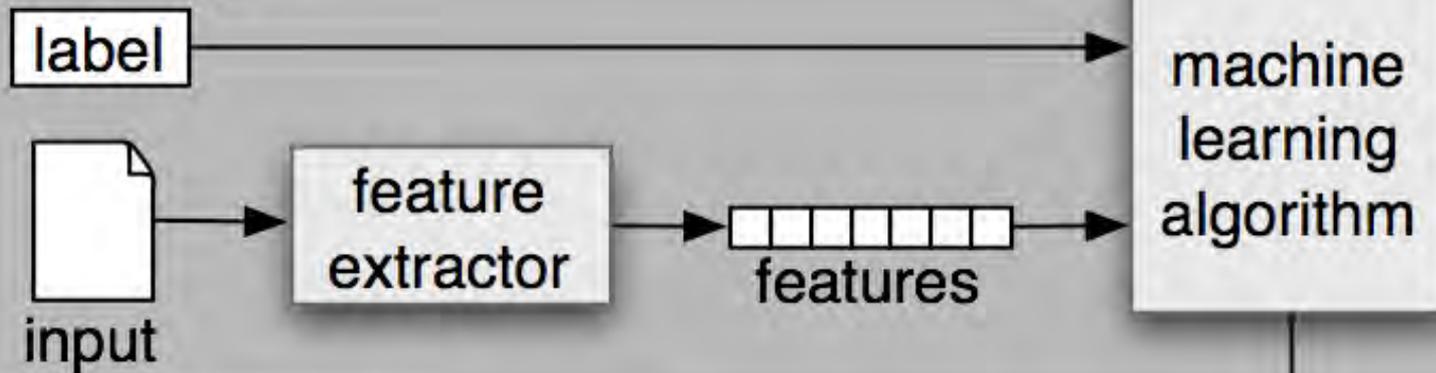
The need to keep PIPs for a long time.

Sensitivity review.
Data protection.
Appendices of
FOIAs.
Little case law.
Time consuming.
Proactive/reactive.
A great muddle.
Be cautious.

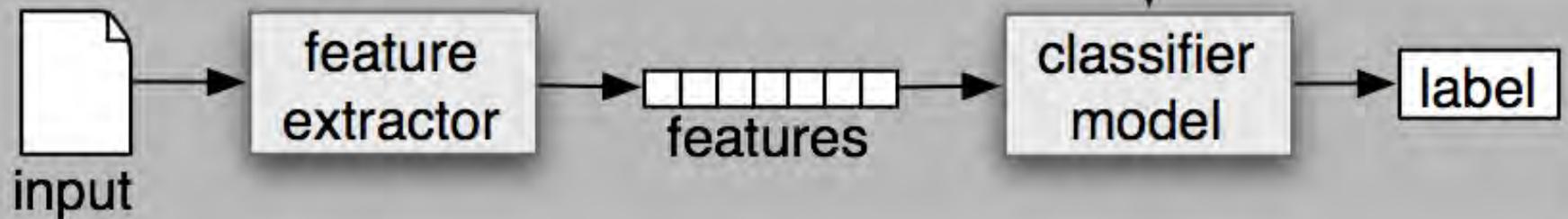




(a) Training



(b) Prediction





You are here: Home

Our digital memory accessible tomorrow

The Digital Preservation Coalition (DPC) is an advocate and catalyst for digital preservation, enabling our members to deliver resilient long-term access to content and services, and helping them derive enduring value from digital collections. We raise awareness of the importance of the preservation of digital material and the attendant strategic, cultural and technological issues. We are a not-for-profit membership organisation and we support our members through knowledge exchange, capacity building, assurance, advocacy and partnership. Our vision is to make our digital memory accessible tomorrow.

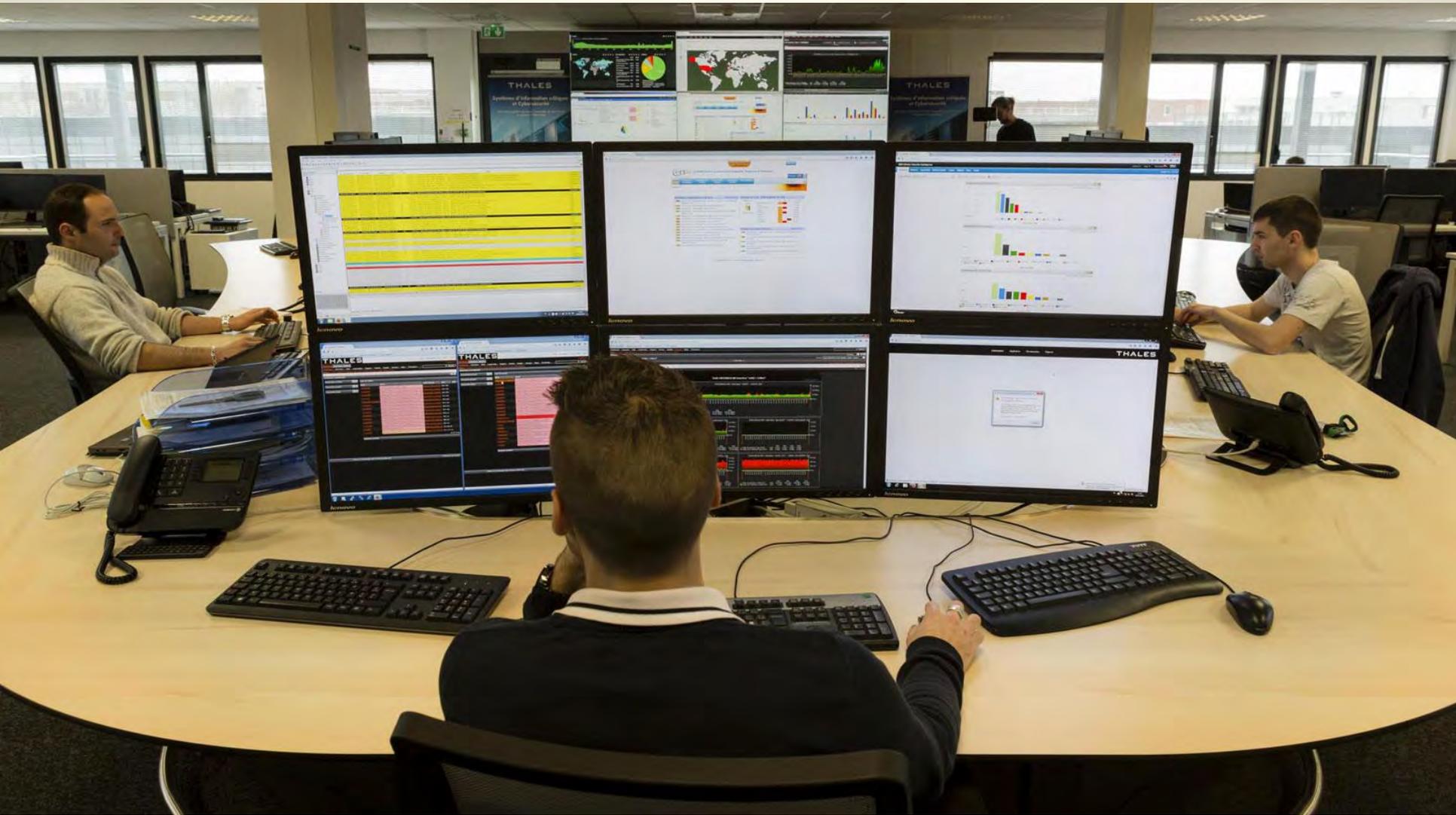


Tim Gollins, 'Parsimonious preservation: preventing pointless processes! (The small simple steps that take digital preservation a long way forward)', London, TNA, 2009)



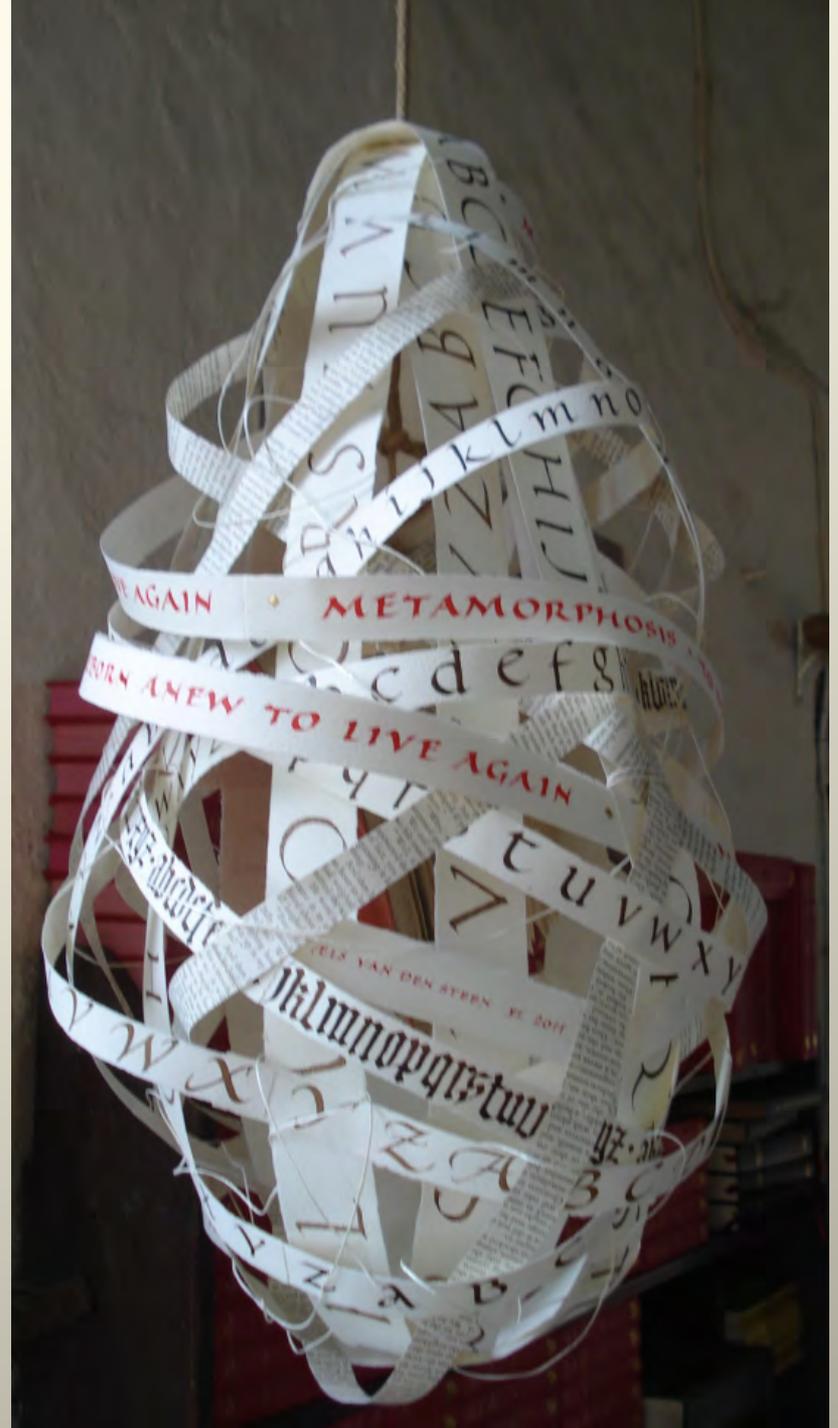
We have three pillars, volume, variety and veracity.





Throughout history the written word has taken on different forms. From clay and wax tablets to papyrus scrolls . . . Incised monumental lettering . . . Medieval manuscripts and eventually the printed book. The metamorphosis of the printed book is taking place right now. And just like before, the written word is evolving into a new form.

Els van den Steen



Exploring Big Historical Data: The Historian's Macroscope

We are not implying that this is the way historians will 'do' history when it comes to big data; rather, it is but one piece of the toolkit, one more way of dealing with 'big' amounts of data that historians are now having to grapple with. What is more, a 'macroscope', a tool for looking at the very big, deliberately suggests a scientist's workbench, where the investigator moves between different tools for exploring different scales, keeping notes in a lab notebook. Similarly, an approach to big data for the historian (we argue) needs to be a public approach, with the historian keeping an open notebook so that others may explore the same paths through the information, while possibly reaching very different conclusions. This is a generative approach: big data for the humanities is not only about justifying a story about the past, but generating new stories, new perspectives, given our new vantage points and tools



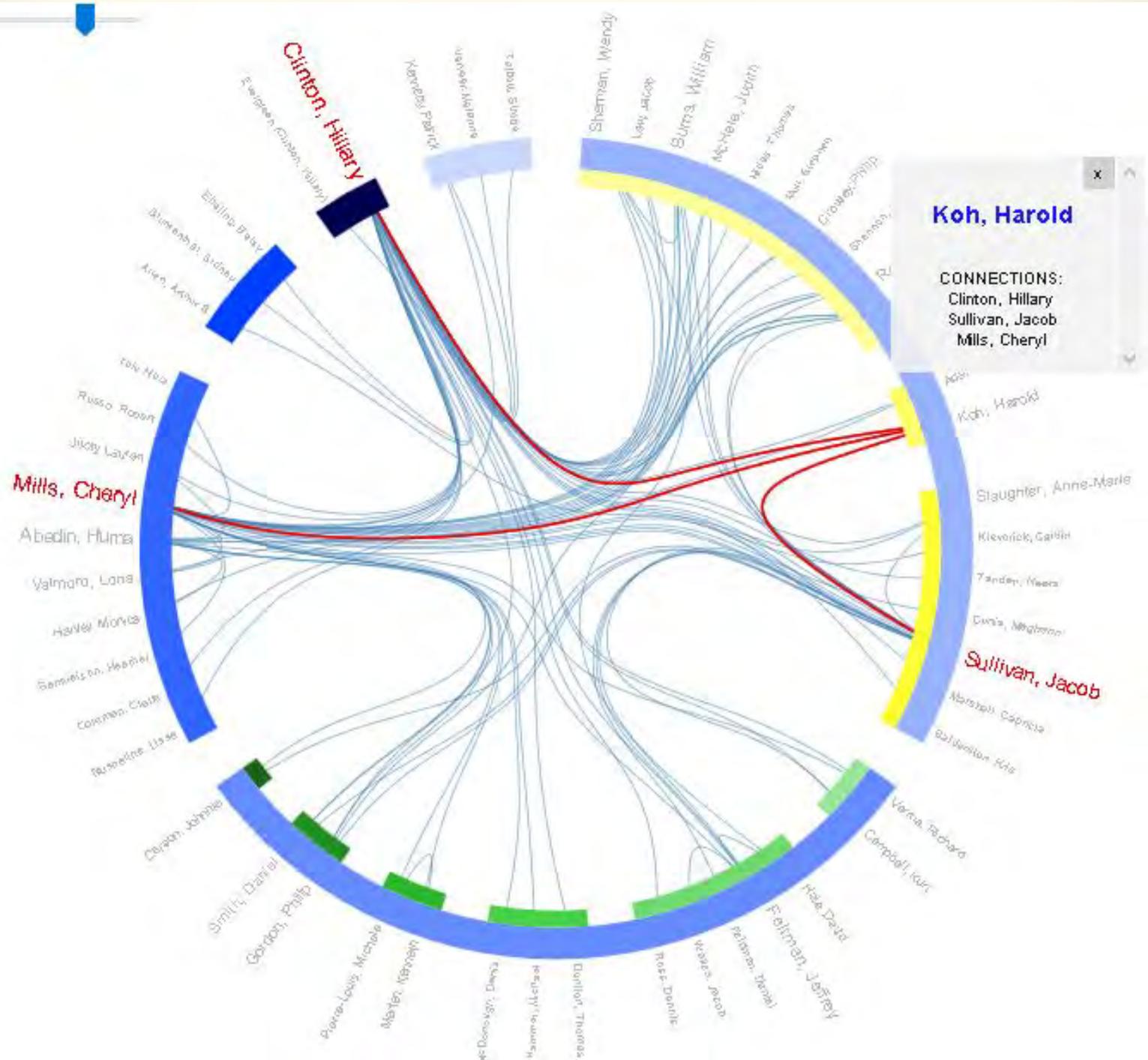
HISTORY
LAB



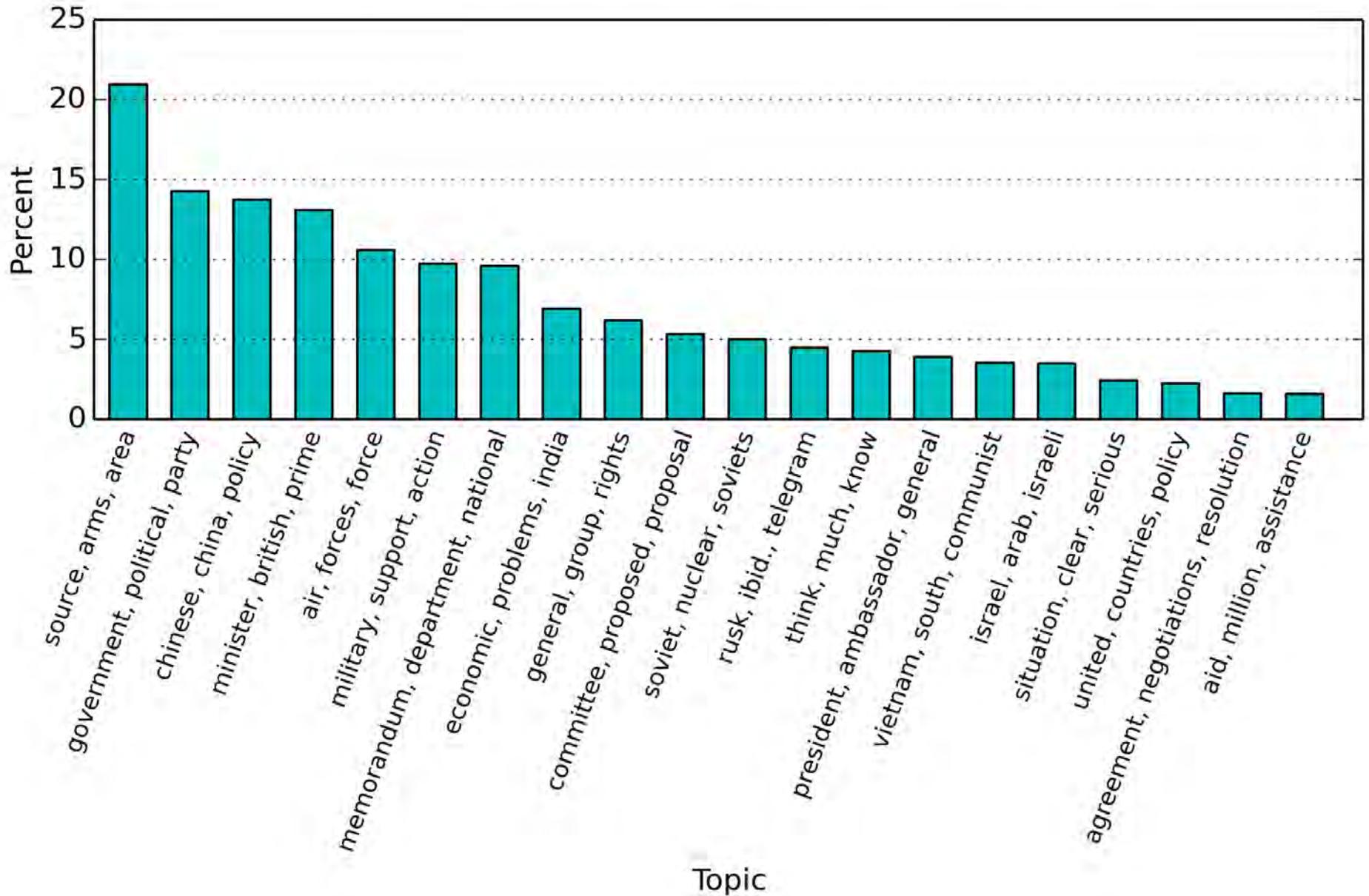
History as Data Science

We turn documents into data and develop
tools to explore history.

Tension:



Percentage of documents with redactions by topic (1961-1968)



Who is to pay?



A Good Question

- Evidential value – the state or organisation.
- Cultural value – tricky.
- Tools – project based, but need to become services.
- Commercial providers.
- Will archives or customers provide tools?
- Raises questions of security and sensitivity.

Understanding the value
of arts & culture

The AHRC Cultural Value Project

Geoffrey Crossick & Patrycja Kaczmarska



**I hope you have
been listening –
this is not going
away.**



Arts & Humanities
Research Council