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 (Piql 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 challenges traditional institutional practices of accountability and citizens' rights, with risks at a wider scale and higher speed.

The online environment makes 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 the 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 evation, 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
What else is possible?

Contact:
Tove.engvall@miun.se
Data: A strategic asset for central banks policies
“Progress lies not in enhancing what is, but in advancing towards what will be” (Khalil Gibran)

Overview

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)
Pretty Big Data – 5 Vs (Value → knowledge)

**Volume**
- Scale of Data
  - 40 Zettabytes (1 Zettabyte = 1,000 Petabytes)
  - 6 billion people have cell phones
  - World population: 7 billion

**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

**Variety**
- Different Forms of Data
  - 30 billion pieces of content are shared on Facebook every month
  - 400 million tweets are sent per day by about 200 million Twitter accounts
  - 4 billion hours of video are watched on YouTube each month

**Veracity**
- Uncertainty of Data
  - 27% of respondents in one survey were unsure of how much of their data was inaccurate

**The FOUR V’s of Big Data**
- Value, Volume, Velocity, Variety

**The FOUR V’s of Big Data**
Data never sleeps – Challenges for historians

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,
Reflections for central banking policy purposes

Supplementary Insights

New theories from combining different science fields

Near real-time snapshot

Extract information on the impact of policy actions

Early warning indicators

Detecting trends and turning points

New theories from combining different science fields

Extract information on the impact of policy actions

Detecting trends and turning points

Near real-time snapshot

Early warning indicators

Supplementary Insights
Better predictions

More timely and frequent information

Means for stimulating debate

An opportunity to adjust model-based theory

Reflections for central banking policy purposes
Paradigm shift - Moving to micro level/granular data

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”!)
Data sources for Monetary Policy & Financial Stability

Transmitin of central banking policies, managing expectations, feedback 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

Other systems
- Financial systems
- Real economy

Other sources
- Policy events, fiscal policies, Global economy
- External shocks
- Sector failures

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

Data Scientist

Data Analytics

S-media

Price scans

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

Data lab

Systematic acquire, Process, summarize

Statistical algorithm and data explorations

Packaging data for Insights & business
1. 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.
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 $R = 0.95$

“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

- Transparency in methodology
- Representativeness and robustness
- Accessibility
- Confidentiality
- Micro-aggregation methods

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

- **Financial Market**: Frequency trading, Prices, bid-offer spreads, Volumes and liquidity, Loans and Holdings, Transactions
- **IFC Pilots**
- **Commercial**
- **Administrative**

Fiscal, trade and corporate balance sheet data
Collaboration and partnership

*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
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.

- Spending control
- Tax policies
- Accounting standards across the public sector
- etc

- Macroeconomic policies
- Productivity and growth policies
- Financial stability and financial services
- Etc
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

March 2016
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.

**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.
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
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.

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.

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.

**Traditional price collection**
Monthly manual collection of prices across 140 locations

**Web scraping**
Daily automatic collection of price quotes from 3 online supermarkets, 6,500 price quotes a day
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

Increased used 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

- Objectives
- Evidence
- Recommendations
- Implementation
New data and data sources have made viable a range of different approaches to policy-making.

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

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

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

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...

Put your suggestion in the box
Have your say on what should be in the Autumn Budget

11:15 pm - 18 Sep 2017

153 Retweets 61 Likes

© 2017 Twitter About Help Centre Terms Privacy policy Cookies Ads info
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’s information archives

A really great story that’s really hard to tell

Jan Booth (DEFRA – former HM Treasury)
# Treasury information (my type of ‘data’)

<table>
<thead>
<tr>
<th>TYPES</th>
<th>THEMES</th>
<th>WHERE’S THE STUFF?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Ministers’ files</td>
<td>• Domestic economy</td>
<td>• Core series up to 1990 – National Archives</td>
</tr>
<tr>
<td>• Policy files</td>
<td>• Institutions</td>
<td>• 1990-1998 paper files</td>
</tr>
<tr>
<td>• Budget records</td>
<td>• Tax and spending</td>
<td>• 1998 on electronic records</td>
</tr>
<tr>
<td>• Secondary sources (stats, reports, analyses...)</td>
<td>• Shocks and crises</td>
<td></td>
</tr>
</tbody>
</table>

*HM Treasury*
Treasury readers

- Treasury people
- Professional analysts
- Academics
- Research students
- Journalists

Organisational boundary
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
• 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)
• Collecting balansheets of all French banks from archival sources to study French banks failures in the Great Depression.
• Main source: Crédit Lyonnais (CL)’s collection of banks balance sheets (about 450 banks from 1910 to 1938): The “Album”.
• The ‘Album’ was built by the Economic & Financial Research Dpt of the Crédit Lyonnais, whose archives have already been used extensively by historians
• Hitherto unseen material
• Full digitization. Connect the Album with other sources.
• Other sources from Crédit Lyonnais and Bank of France, on the number and causes of bankruptcies
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
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
1880

Restaurant et Caves du Café Anglais

Société anonyme formée par acte sous-seing privé du 22 juillet 1878 ; définitivement constituée le 31 juillet ; statuts déposés à M. Delphande, notaire à Paris, le 20 du même mois.

Objet. — 1° Exploitation du Café-Restaurant connu sous la dénomination de Café Anglais.
2° Commerce des vins en France et à l'étranger ;
3° Acquisition, fondation et revente de tout établissement de café ou de restaurant, en France ou à l'étranger ;
4° Acquisition et revente d'immeubles servant à l'exploitation desdits établissements.

Dénomination. — Société anonyme du Restaurant et des Caves du Café Anglais.

1907

Société générale des Eaux de Calais (Scriveur et C°)

Société d'origine en 1850, a été réorganisée en société anonyme par acte sous-seing privé du 8 février 1905, sur la motion de MM. Scriveur et C°.

Siège social : à Calais.

Directeur-gérant : MM. Scriveur et C°.

Bilan au 31 décembre 1905

1974

SOCIETE DES ECLAIS DE PRODUITS CHIMIQUES


1980

Société d'énergie et de produits chimiques


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)
<table>
<thead>
<tr>
<th>Name</th>
<th>Montpellier a Gette (Chemin de Fer)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sector</td>
<td>Railroads</td>
</tr>
<tr>
<td>Type</td>
<td>action de capital</td>
</tr>
<tr>
<td>Market</td>
<td>Spot (Au Comptant)</td>
</tr>
<tr>
<td>Price 1</td>
<td></td>
</tr>
<tr>
<td>Price 2</td>
<td></td>
</tr>
<tr>
<td>Price 3</td>
<td></td>
</tr>
<tr>
<td>Price 4</td>
<td></td>
</tr>
<tr>
<td>Price 5</td>
<td></td>
</tr>
<tr>
<td>Price 6</td>
<td></td>
</tr>
<tr>
<td>Price 7</td>
<td></td>
</tr>
<tr>
<td>Price 8</td>
<td></td>
</tr>
<tr>
<td>Price 9</td>
<td></td>
</tr>
<tr>
<td>Price 10</td>
<td></td>
</tr>
<tr>
<td>Higher (Plus haut)</td>
<td></td>
</tr>
<tr>
<td>Lower (Plus bas)</td>
<td></td>
</tr>
<tr>
<td>Previous (Dernier cours coté)</td>
<td>At dd/mm/yyyy</td>
</tr>
<tr>
<td>Price yesterday (Cours de la veille)</td>
<td></td>
</tr>
<tr>
<td>Demand(e)</td>
<td></td>
</tr>
<tr>
<td>Supply (Offre)</td>
<td></td>
</tr>
<tr>
<td>Volume</td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>Official List</td>
</tr>
<tr>
<td>Comments</td>
<td></td>
</tr>
<tr>
<td>Nominal</td>
<td>No value found</td>
</tr>
<tr>
<td>Dividend</td>
<td>No dividend information available</td>
</tr>
<tr>
<td>Amount</td>
<td>No quantities found for this date</td>
</tr>
</tbody>
</table>

**Source:** Official List

**Comments:**
- No value found
- No dividend information available
- No quantities found for this date
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
<table>
<thead>
<tr>
<th>Région</th>
<th>Nom</th>
<th>Comptant</th>
<th>Cour de terme</th>
<th>Coût de l'achat</th>
<th>Coût de l'abattage</th>
<th>Coût de l'entraînement</th>
<th>Coût de l'atelier</th>
<th>Coût de l'exploitation</th>
<th>Coût de l'entretien</th>
<th>Coût de l'administration</th>
</tr>
</thead>
<tbody>
<tr>
<td>CANAUX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHEMINS DE FER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EAU, ELECTRICITÉ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FORGES, FONDERIES BOUILLERES, MINES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
- Les coûts de l'achat, de l'abattage, de l'entraînement, de l'atelier, de l'exploitation, de l'entretien, et de l'administration sont mentionnés pour chaque région et catégorie spécifiée.
<table>
<thead>
<tr>
<th>Numéro de la page</th>
<th>Dénomination des valeurs</th>
<th>Comptant</th>
<th>Chambre de commerce</th>
<th>Banque</th>
<th>Crédit</th>
<th>Comptant</th>
<th>Chambre de commerce</th>
<th>Banque</th>
<th>Crédit</th>
</tr>
</thead>
<tbody>
<tr>
<td>400 000</td>
<td>Réal Foncières, act. 400 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
</tr>
<tr>
<td>600 000</td>
<td>Réal Foncières, act. 600 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
</tr>
<tr>
<td>800 000</td>
<td>Réal Foncières, act. 800 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
</tr>
</tbody>
</table>

**CANAUX**

<table>
<thead>
<tr>
<th>Numéro de la page</th>
<th>Dénomination des valeurs</th>
<th>Comptant</th>
<th>Chambre de commerce</th>
<th>Banque</th>
<th>Crédit</th>
<th>Comptant</th>
<th>Chambre de commerce</th>
<th>Banque</th>
<th>Crédit</th>
</tr>
</thead>
<tbody>
<tr>
<td>400 000</td>
<td>Réal Foncières, act. 400 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
</tr>
<tr>
<td>600 000</td>
<td>Réal Foncières, act. 600 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
</tr>
<tr>
<td>800 000</td>
<td>Réal Foncières, act. 800 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
</tr>
</tbody>
</table>

**CHEMINS DE FER**

<table>
<thead>
<tr>
<th>Numéro de la page</th>
<th>Dénomination des valeurs</th>
<th>Comptant</th>
<th>Chambre de commerce</th>
<th>Banque</th>
<th>Crédit</th>
<th>Comptant</th>
<th>Chambre de commerce</th>
<th>Banque</th>
<th>Crédit</th>
</tr>
</thead>
<tbody>
<tr>
<td>400 000</td>
<td>Réal Foncières, act. 400 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
</tr>
<tr>
<td>600 000</td>
<td>Réal Foncières, act. 600 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
</tr>
<tr>
<td>800 000</td>
<td>Réal Foncières, act. 800 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
</tr>
</tbody>
</table>

**EAUX, ÉLECTRICITÉ**

<table>
<thead>
<tr>
<th>Numéro de la page</th>
<th>Dénomination des valeurs</th>
<th>Comptant</th>
<th>Chambre de commerce</th>
<th>Banque</th>
<th>Crédit</th>
<th>Comptant</th>
<th>Chambre de commerce</th>
<th>Banque</th>
<th>Crédit</th>
</tr>
</thead>
<tbody>
<tr>
<td>400 000</td>
<td>Réal Foncières, act. 400 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
</tr>
<tr>
<td>600 000</td>
<td>Réal Foncières, act. 600 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
</tr>
<tr>
<td>800 000</td>
<td>Réal Foncières, act. 800 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
</tr>
</tbody>
</table>

**FOUDRES, FONDERIES, HOUILLÈRES, MINES**

<table>
<thead>
<tr>
<th>Numéro de la page</th>
<th>Dénomination des valeurs</th>
<th>Comptant</th>
<th>Chambre de commerce</th>
<th>Banque</th>
<th>Crédit</th>
<th>Comptant</th>
<th>Chambre de commerce</th>
<th>Banque</th>
<th>Crédit</th>
</tr>
</thead>
<tbody>
<tr>
<td>400 000</td>
<td>Réal Foncières, act. 400 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
</tr>
<tr>
<td>600 000</td>
<td>Réal Foncières, act. 600 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
</tr>
<tr>
<td>800 000</td>
<td>Réal Foncières, act. 800 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
<td>1 500 000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date</th>
<th>Nom</th>
<th>Adresse</th>
<th>Téléphone</th>
<th>Télégraphe</th>
<th>Téléscripteur</th>
<th>Ventes</th>
<th>Ventes</th>
<th>Ventes</th>
<th>Ventes</th>
<th>Ventes</th>
<th>Ventes</th>
<th>Ventes</th>
<th>Ventes</th>
<th>Ventes</th>
<th>Ventes</th>
<th>Ventes</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/01</td>
<td>CANAUX</td>
<td>89 rue de la Poste, 75017 Paris</td>
<td>12345678</td>
<td>ABCD</td>
<td>EFGH</td>
<td>9999</td>
<td>9999</td>
<td>9999</td>
<td>9999</td>
<td>9999</td>
<td>9999</td>
<td>9999</td>
<td>9999</td>
<td>9999</td>
<td>9999</td>
<td>9999</td>
</tr>
<tr>
<td>11/01</td>
<td>CHEMINS DE FER</td>
<td>Résidence de la Gare, 123 Route de Lyon, 75018 Paris</td>
<td>87654321</td>
<td>IJKL</td>
<td>MNOP</td>
<td>7777</td>
<td>7777</td>
<td>7777</td>
<td>7777</td>
<td>7777</td>
<td>7777</td>
<td>7777</td>
<td>7777</td>
<td>7777</td>
<td>7777</td>
<td></td>
</tr>
<tr>
<td>12/01</td>
<td>EAU, ÉLECTRICITÉ</td>
<td>54 Rue des Moulins, 75019 Paris</td>
<td>22222222</td>
<td>QRST</td>
<td>UVWX</td>
<td>1111</td>
<td>1111</td>
<td>1111</td>
<td>1111</td>
<td>1111</td>
<td>1111</td>
<td>1111</td>
<td>1111</td>
<td>1111</td>
<td>1111</td>
<td></td>
</tr>
<tr>
<td>13/01</td>
<td>FORGES, FONDERIES, BOUILLES, MINES</td>
<td>32 Rue de la Mine, 75020 Paris</td>
<td>33333333</td>
<td>ZYXW</td>
<td>PQRS</td>
<td>6666</td>
<td>6666</td>
<td>6666</td>
<td>6666</td>
<td>6666</td>
<td>6666</td>
<td>6666</td>
<td>6666</td>
<td>6666</td>
<td>6666</td>
<td></td>
</tr>
</tbody>
</table>

**Remarque:** Les détails spécifiques des ventes et des services pour chaque secteur sont regroupés dans les colonnes correspondantes.
<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
<th>Value</th>
<th>Category</th>
<th>Note</th>
<th>Comptant</th>
<th>Action</th>
<th>Coupon</th>
<th>Rate</th>
<th>1st Class</th>
<th>Flat Rate</th>
<th>Flat Rate</th>
<th>General Rate</th>
<th>2nd Class</th>
<th>3rd Class</th>
<th>4th Class</th>
<th>5th Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>2022-05-15</td>
<td>Description 1</td>
<td>123.45</td>
<td>Category 1</td>
<td>Note 1</td>
<td>Comptant 1</td>
<td>Action 1</td>
<td>Coupon 1</td>
<td>Rate 1</td>
<td>1st Class 1</td>
<td>Flat Rate 1</td>
<td>Flat Rate 1</td>
<td>General Rate 1</td>
<td>2nd Class 1</td>
<td>3rd Class 1</td>
<td>4th Class 1</td>
<td>5th Class 1</td>
</tr>
<tr>
<td>2022-05-16</td>
<td>Description 2</td>
<td>234.56</td>
<td>Category 2</td>
<td>Note 2</td>
<td>Comptant 2</td>
<td>Action 2</td>
<td>Coupon 2</td>
<td>Rate 2</td>
<td>1st Class 2</td>
<td>Flat Rate 2</td>
<td>Flat Rate 2</td>
<td>General Rate 2</td>
<td>2nd Class 2</td>
<td>3rd Class 2</td>
<td>4th Class 2</td>
<td>5th Class 2</td>
</tr>
</tbody>
</table>
At the end of the day

• Many experiments
• Many failures
• Some success....
Album Comparison with previous series of deposits

1929-1931 crisis

- League of nations. Deposits (total) 6 Banks.
- INSEE (1952) & Mitchell, International Historical Statistics
- Saint Marc (1983)
- Patat & Lutfalla (1986), based on Laufenburger (1940)
- Villa (1993)
- Album Crédit Lyonnais
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
1. authenticity
2. cost-saving
Data security & storage problem

€ 28 mil

ultimate data solution

European Union
The Research Council of Norway
Fraunhofer IPM
Texas Instruments
Federal Office of Civil Protection and Disaster Assistance
VISITECH
P+S Technik
Infision
Ransomware attack: Who's been hit

Cyber-Safe

Massive ransomware attack on countries

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

The world's biggest cyberattack has hit at least 150 countries and infected 300,000 machines since it started spreading last Friday.
“best practice” myth
It must be impossible to modify or delete. It must be protected from cyber attacks, logical threats, EMP and physical threats. It should be able to store any kind of files; digital and visual. It must be searchable. It must be migration-free to avoid the risk of data loss and the migration cost. Data retrieval process must be independent of technological obsolescence. 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.

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

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

**piqlReader**
Reads data off the piqlFilm using open source software.
and many more…
Instability
Safest place?
Seed vault
Protected by polar bears & permafrost
Arctic 'doomsday' vault seeks to protect world's most precious data

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

Viral
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
Getting values from „shared” information

Why Big Data?

• This world is moving fast as data becomes more accessible, attributable and analytical.
• Sitting behind the surge in digital services is data and therefore data is of increasing importance to all organizations.
• It can help businesses and entrepreneurs to identify areas of opportunity for innovation in new products, processes and services; improve customer engagement; identify inefficiencies; improve productivity, identify market trends.
Big Data

Financial usage

<table>
<thead>
<tr>
<th>Example</th>
</tr>
</thead>
</table>

**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
Data Security
Bad Analytics
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.

Source: Forbes (2015)
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.

Source: Ray Kurzweil (2001)
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.

Source: Verizon (2016)
# Risks

## Data Security

<table>
<thead>
<tr>
<th>Industry</th>
<th>Incidents</th>
<th>Breaches</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Small</td>
</tr>
<tr>
<td>Total</td>
<td>42,068</td>
<td>606</td>
</tr>
<tr>
<td>Accommodation (72)</td>
<td>215</td>
<td>131</td>
</tr>
<tr>
<td>Administrative (56)</td>
<td>42</td>
<td>6</td>
</tr>
<tr>
<td>Agriculture (11)</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>Construction (23)</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Education (61)</td>
<td>455</td>
<td>37</td>
</tr>
<tr>
<td>Entertainment (71)</td>
<td>5,534</td>
<td>7</td>
</tr>
<tr>
<td>Finance (52)</td>
<td>996</td>
<td>58</td>
</tr>
<tr>
<td>Healthcare (62)</td>
<td>458</td>
<td>92</td>
</tr>
<tr>
<td>Information (51)</td>
<td>717</td>
<td>57</td>
</tr>
<tr>
<td>Management (55)</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Manufacturing (31-33)</td>
<td>620</td>
<td>6</td>
</tr>
<tr>
<td>Mining (21)</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Other Services (81)</td>
<td>69</td>
<td>22</td>
</tr>
<tr>
<td>Professional (54)</td>
<td>3,016</td>
<td>51</td>
</tr>
<tr>
<td>Public (92)</td>
<td>21,239</td>
<td>46</td>
</tr>
<tr>
<td>Real Estate (53)</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>Retail (44-45)</td>
<td>326</td>
<td>70</td>
</tr>
<tr>
<td>Trade (42)</td>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td>Transportation (48-49)</td>
<td>63</td>
<td>5</td>
</tr>
<tr>
<td>Utilities (22)</td>
<td>32</td>
<td>2</td>
</tr>
<tr>
<td>Unknown</td>
<td>8,220</td>
<td>3</td>
</tr>
</tbody>
</table>

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

Source: Verizon (2017)
**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

Source: Imperva (2016)
Risks

Data Security

Potential power shortage

Threat

A grabber cuts the power supply

materialise by

Security Incident

exploits

NO UPS

Vulnerability

Servers shut down, service unavailable

Security Breach

Risk

High impact: loss of customer

Residual risk

Low impact: decreased service level

Alternative power supply

Control

patch

mitigate

National Bank of Hungary
Risks
Data Security

Fundamentals of security and control measures

Confidentiality
Availability
Integrity

Completeness
Correctness

<table>
<thead>
<tr>
<th>Preventative</th>
<th>Detective</th>
<th>Corrective</th>
<th>Compensatory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Security Awareness Training</td>
<td>System Monitoring</td>
<td>OS Upgrade</td>
<td>Backup Generator</td>
</tr>
<tr>
<td>Firewall</td>
<td>IDS</td>
<td>Backup Data Restoral</td>
<td>Hot Site</td>
</tr>
<tr>
<td>User Right Management</td>
<td>Anti-Virus</td>
<td>CSIRT</td>
<td>Encryption</td>
</tr>
<tr>
<td>Security Guard</td>
<td>Motion Detector</td>
<td>Vulnerability Mitigation</td>
<td>...</td>
</tr>
</tbody>
</table>

National Bank of Hungary
Which are the risks related to data?
Gathering, storing, processing

- Cost of processing
- Data Security
- Data Privacy
- Bad Analytics
- 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

Source: Imperva (2016)
Risks
Bad data, bad analytics

Data management strategy

1. Learn how to collect the right data the right way
2. Inventory your data
3. Fill gaps in your data and analytics capabilities
4. Hire experts who can make data tell a story
5. Make data and statistics visual
6. Offer data-driven thought leadership
Risks
How to use it, and for what?

Data has a value chain

Customer – Financial Institution – Fintech Company – Added Value

Source: SEI (2017)
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 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

Laissez faire

Unfair advantage
Fintech companies using big data have competitive advantage comparing to the strictly regulated banking market

Consumer protection?
Increasing risk of fraud or any loss making event, which harm the investors/depositors

Supervisory control

Losing control
The risk of the spread of cross-border services (the Internet can not be stopped at borders)

Barriers for innovation
Maintaining the expensive banking model and cutting back on development

Solution:
Unified, international regulation
Regulations
Consumer, data and market protection

- Unfair Commercial Practices Directive
- E-Commerce Directive
- Directive on Electronic Communications and Privacy
- Unfair Contract Terms
- Directive on Distance Marketing of Financial Services
- Payment Services Directive
- Mortgage Credit Directive
- Payment Accounts Directive
- GDPR: the ‘right to be forgotten’ access to one’s own data, the right of data portability
- Banking secrecy rules impose restrictions to the use of consumer data by financial institutions
- recommendation on the use of community and public cloud services (HU)

Source: EBA (2016)
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

https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#5acaf0f17b1

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/


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

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

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

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 17\textsuperscript{th} Century
- Historically the World's 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
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 date 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
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)
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.
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 as Data Science

We turn documents into data and develop tools to explore history.
Countries Mentioned in Documents

What countries are most often discussed in these documents? To select them for search you can zoom into the map or select from the list below.
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.
I hope you have been listening – this is not going away.