Framework for big data usage in risk management process in banking institutions

Ivana Dvorski Lacković, Vladimir Kovšca, Zrinka Lacković Vincek
Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2, 42 000 Varaždin, Croatia
{idvorski, vkovsca, zlackovi}@foi.hr

Abstract. *Nowadays banks operate in changing environment influenced by regulatory requirements, emerging risk types and competition on the market. At the same time banks have available large datasets arising from internal and external sources. The potential for this data usage in risk management has only recently been discovered and has not been the subject of extensive scientific research. There are two goals of this paper. Firstly, authors give an overview of available scientific literature and practical research related to big data usage in risk management in banks. Secondly, based on the literature review authors are presenting framework with specified detailed use of big data in specific key risk management areas. Expected contribution of this paper is in presenting framework that can be used for practical purposes in banking industry as a matrix for using big data in certain risk management area. Second expected contribution is in increasing scientific public awareness on the topic and on the potential of research in the field of big data technologies usage in risk management in banks.*

Keywords. Big data, banks, risk management

1 Introduction

Big data technology has been an important topic in the information technology research area and has been used in banking for some time, mainly in marketing domain and as a tool for fraud detection but its potential in banking regarding risk management has only recently been discovered. Consulting companies that operate in financial domain are doing extensive research and disseminating knowledge about importance and potential of big data in risk management. On the other hand, scientific literature on the subject is rather scarce. The purpose of this paper is to synthetize existing knowledge on the topic and to propose general framework for big data usage in risk management in banks. The paper is focused on one specific narrowed area of big data application in banking – risk management - due to several reasons. Firstly, big data usage in some other areas of banking business activity (such as marketing, compliance or fraud detection and prevention) has already been a subject of scientific research, while application on risk management activities has not been so extensively covered. Secondly, due to Basel III regulatory requirements risk management in banking is considered one of the three key control functions in banks (together with compliance and internal audit function) whose main task is insurance that the bank is conducting its business activity in accordance with defined strategy, rules and policies on risk management. The findings of risk management control function are subject of internal and external reporting and basis for strategic business decisions. The authors therefore find the research of big data usage on one specific segment of banking business activity – risk management – scientifically interesting and justified. In first section authors discuss general terms related to big data. Second section is related to big data usage in banks with special emphasis on scientific literature and on practical implications for risk management. In third section we present impediments to big data usage in organizations. Fourth section is devoted to proposing framework for big data usage in risk management in banks based on identification of key risk management areas. Finally, we present conclusions and proposals for future research.

2 Big Data Basics

According to Gartner IT Glossary (2014) big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making. Originally big data was a term used to describe large datasets that could not be captured, stored, managed nor analysed using traditional databases. With years the term

broadened and now encompasses also the set of technologies that perform all the mentioned functions, varied collections that solve complex problems and make value from that data more economical (Deutsche Bank, 2015).

IBM (2014) defines the 'four V's' behind big data mechanism:

- volume (quantity of data. i.e. massive data sets),
- velocity (accelerating speed at which data is being generated today),
- variety (increasing diversity in the types and sources of data requiring management and analysis),
- veracity (reliability of particular data).

Özköse et al. (2015) argue that there are 'five V's' behind the big data term, adding 'value' defined as the result generated from all the procedures and enriching the process to the before mentioned 'four V's'.

Gandomi and Haider (2015) divide the process for extracting insight from big data into following sequences:

- Data Management (acquisition and recording; extraction, cleaning and annotation; integration, aggregation and representation) and
- Analytics (modelling and analysis; interpretation).

3 The potential of big data usage in risk management in banking

Mohanty et al. (2013, pp. 54) argue that big data are applicable on several levels in the banking industry, including retail (branch and mortgage banking, credit card and private banking), commercial (credit risk analysis, non-credit products, client management and sales, middle market lending), capital markets (trading and sales, underwriting and structured finance, non-depository credit institutions) and asset management (wealth management, asset investment management, asset issuer services, global asset reporting and investment deposit analytics).

According to Deutsche Bank research (2015) drivers of big data technology adoption in the financial industry are: (1) *explosive data growth* (increased number of banking instruments and transactions), (2) *regulation* (requirements regarding the real-time view on risk management and related financial transactions; enhanced risk reporting; risk simulations), (3) *fraud detection and security* (operational risk management issues, such as spotting fraud risk among its own employees or identifying rogue traders), (4) *customer insight and marketing analytics* (understanding consumer behaviour and interests and consequently shaping new products in accordance with customer needs).

According to research on big data and credit risk conducted by McKinsey&Company (2015) there are six key trends that are expected to change bank risk management. These are: (1) expanded regulatory requirements, (2) use of technology and mathematics in risk management, (3) changing customer expectations, (4) de-biasing judgmental decision-making, (5) emerging risk types and (6) growing cost discipline.

When discussing risk management, regulatory changes are the key factor that has influence on business practice and are starting point for changes in business process. Regulatory requirements do not only define general risk management framework, but also have significant impact on risk appetite definition, strategic decision making with special emphasis on financial transactions, stress tests and scenario analysis and risk reporting that encompasses all types and sources of risk to which bank is exposed. Krishna (2016) states that requirements imposed on banks by regulators have brought transformational change to banks in regards to organisational structure, reporting hierarchies and on underlying data and technology infrastructure of the institution required to monitor risk and report it to the regulator. The reporting requirements are imposed not only on global-systematically important banks but also on banks of all sizes. One of the prerequisites for bank to be complied with regulatory requirements is banks' ability to apply a bank-wide risk governance system in order to achieve one single and truthful source of required information.

When having all of the mentioned requirements in mind it is evident that banks need to use all the available data in order to predict risk, manage it and report it.

3.1 Overview of scientific literature

In this section we will give a short overview of scientific research on big data application in risk management. Cerchiello et al. (2016) used big data technology to propose a framework for bank risk contagion. The framework is based on graphical models that can estimate systemic risks with models based on two different sources: financial markets and financial tweets and suggest a way to combine them using Bayesian approach. This is the first systemic risk model based on big data and it sheds further light on the interrelationships between financial institutions. Ngai et al. (2011) discuss application of data mining techniques in financial fraud detection, Ravisankar et al. (2011) use big data technologies in detection of financial statement fraud and Hu et al. (2012) use big data for the development of a network approach to risk management for modelling and analysing systemic risk in banking systems.

Tian and al. (2015) specify several challenges regarding big data analytics in financial organizations, namely: (1) dealing with the *problem of massive data storage and organization* due to fact that many organizations need to keep historic data for many years for trend prediction and other complex analytics what poses a great challenge for reliability of the storage system, (2) dealing with *various data types* since most of big data today is generated from different sources

and thus unstructured, (3) *designing highly efficient computing system* to process the distributed historical and incoming data, (4) *optimizing the usage of memory* to keep the temporal data in memory is of special importance for risk management activities in financial organization due to requirements for fetching historical data in order to analyse new coming data.

The respondents of the Irving Fisher Committee Report on *Central banks' use of and interest in big data* (2015) identified credit and market risk identification as an additional expected use of big data analysis.

3.2 Overview of research done by financial consultants

According to research conducted by The Economist Intelligence Unit Limited (2014) over a half of risk managers stated that they lack sufficient data to support robust risk management. The results of Global Forensic Data Analytics Survey conducted by Ernst and Young (2014) suggest that 72% of respondents believe that big data can play a key role in fraud detection and prevention, 7% are aware of any specific big data technologies and only 2% are actually using them. When having in mind the growing complexity and number of financial transactions, volatility of markets, multiplication of devices, multiplication of channels and ever increasing demands from regulators, it is evident that what banks need is to make a connection between data acquisition, analysis and action (The Economist Intelligence Unit Limited, 2014).

McKinsey research (2016) states that in near future some banks will lag behind because they are unable or unwilling to make investments into systems and infrastructure that will help them use big data for complex data investigation and analysis.

The Ernst&Young Global Forensic Data Analytics Survey (2014) gives an overview of current big data usage in compliance monitoring:

- Payment stream and accounts payable analysis (altered invoices, duplicate or fake invoices, suspicious payment descriptions),
- Vendor master/employee master analysis and comparison (vendor risk ranking, conflicts of interest),
- Employee expenses/travel and entertainment (over limits, split or duplicate expenses),
- Payroll (ghost employees, falsified wages, commission schemes),
- Financial misstatement (fictitious revenues, concealed liabilities, overstated assets),
- Bribery and corruption (conflict of interest, bid rigging),
- Capital projects (contract non-compliance, project abuses and overcharges).

The sources of big data in banks can be shown by Figure 1.

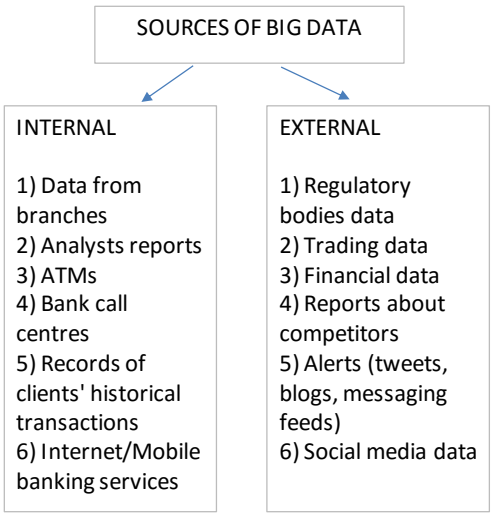

Figure 1. Sources of big data in banks, illustration by authors based on Oracle Enterprise research (2015)

4 Impediments to big data usage in risk management

Capgemini Consulting (2014) defines three levels of building analytics maturity in organization and their characteristics are:

- Beginner:
  - There is no defined data structure/Informal, conflicting and dispersed data,
  - Poor data governance/Basic data reporting using mainly spreadsheet based tools,
  - Pockets of reporting and analysis capability/Dispersed talent,
  - Preliminary analytics strategy.
- Proficient:
  - Data available for existing and potential customers/Most data is still unstructured and internal,
  - Use of some statistical and forecasting tools/Coherent procedures for data management,
  - Well-defined recruitment process to attract analytics talent/Budget for analytics training,
  - Analytics is used to understand issues and develop data-based options across the business.
- Expert:
  - Internal, external and social media data is merged to build and integrated and structured dataset,
  - Established robust master data management framework for structured and unstructured data sets,
  - Existence of analytics centre of excellence for promotion of best practice/strategic partnership for supplementary analytics skills,
  - Full executive sponsorship of analytics.

Evidently organisations on lower level of analytics maturity will have more difficulty in extracting value from big data for their internal processes. According to survey conducted by Capgemini and the Economist Intelligence Unit (2012) biggest impediments for big data usage in effective decision-making process in organizations are:

- Too many "silos" – data is not pooled for the benefit of the entire organization;
- Time taken to analyse large data sets;
- Shortage of skilled people for data analysis;
- Big data is not viewed sufficiently strategically by senior management;
- Unstructured content in big data is too difficult to interpret;
- The high cost of storing and analysing large data sets;
- Big data sets are too complex to collect and store.

Duan & Xiong (2015) state that big data contain more unstructured than structured data (namely text data, graph data and time-series data) and that biggest challenges for unstructured data are to: (1) transfer unstructured data into a structured format and (2) develop a new method to handle unstructured data.

According to Irving Fisher Committee on Central Banking Statistics report (2015) regular production of big data-based information takes time, especially because of resource issues.

Krishna (2016) defines following enhanced data capabilities necessary for banks to be complied with regulatory reporting requirements:

- *Data sourcing* in timely, accurate and complete manner from catalogued sources,
- *Data processing and retention* should be efficient and able to support historical analysis,
- *Data analytics and reporting* should allow standardised and repeatable analytics with flexible drilldown and visual reporting capabilities,
- *Data management* with regard to access, retention, distribution and quality of data, data quality monitoring, error handling and reconciliation capabilities,
- *Data governance and control* relate to clear ownership, accountability and organisational standards combined with a robust data control framework.

5 Framework for big data usage in risk management in banking

According to Basel III regulatory requirements risk management control function, together with the compliance function and internal audit function, presents one of the three bank's control functions. Main tasks of control functions are insurance that the bank is conducting its business activity in accordance with defined strategy, rules and policies on risk management. The findings of risk management control function are subject of internal and external reporting. Internal reports are related to management board and supervisory board reports, while external reports are subject of central banking authority monitoring. Therefore, the findings of control functions incorporated in reporting form basis for bank's management strategic business decisions, but also a tool for external supervisory monitoring. Having on mind the role and importance of risk management in individual bank on impact on stability of financial sector in whole, it is evident that banks need to adequately manage its risks and exploit all the possibilities, tools and technology that they have on their disposal in order to adequately manage its risks.

When analysing the state of current research that is available on the topic of big data usage in risk management in banks, we found that most of the existing research can be categorized in one of the following categories: 1) drivers and challenges of big data usage in risk management in banks, 2) practical application of big data in certain risk management processes (mostly modelling and fraud detection). When having in mind the importance that risk management has for banks we propose a wider strategic look at the potential of big data in this process by defining a framework on how big data can be used in the risk management process. There are several steps that we conducted while defining this framework for all-encompassing big data usage in banks risk management:

1. Selection of appropriate all-encompassing risk management system definition,
2. Using this definition as basis for extraction of key risk management activities,
3. Allocating the potential usage of big data to key risk management activities defined in step two.

Since the banking domain is subject to strict regulation the authors selected following all-encompassing definition of risk management system from Basel III legislative, i.e. Regulation EU No 575/2013 (2013) as starting point for framework definition: "risk management system is made of organisational structure, rules, processes, actions and resources for identification, measurement or assessment, management and reporting on risk exposures, i.e. risk management in whole and it encompasses adequate corporate governance and risk culture." In continuation on the definition, following key risk management activities have been identified: (I) identification, (II) assessment, (III) management and control and (IV) reporting.

Final step in framework definition has been to allocate known potential usage of big data to key risk management activities and it resulted in the formulation of the proposed framework that we are proposing in continuation. The framework is shown by Figure 2. As can be seen from the Figure, big data can be used in banking in all key risk management activities in order to enhance the process of managing various types of risks banks are exposed to.

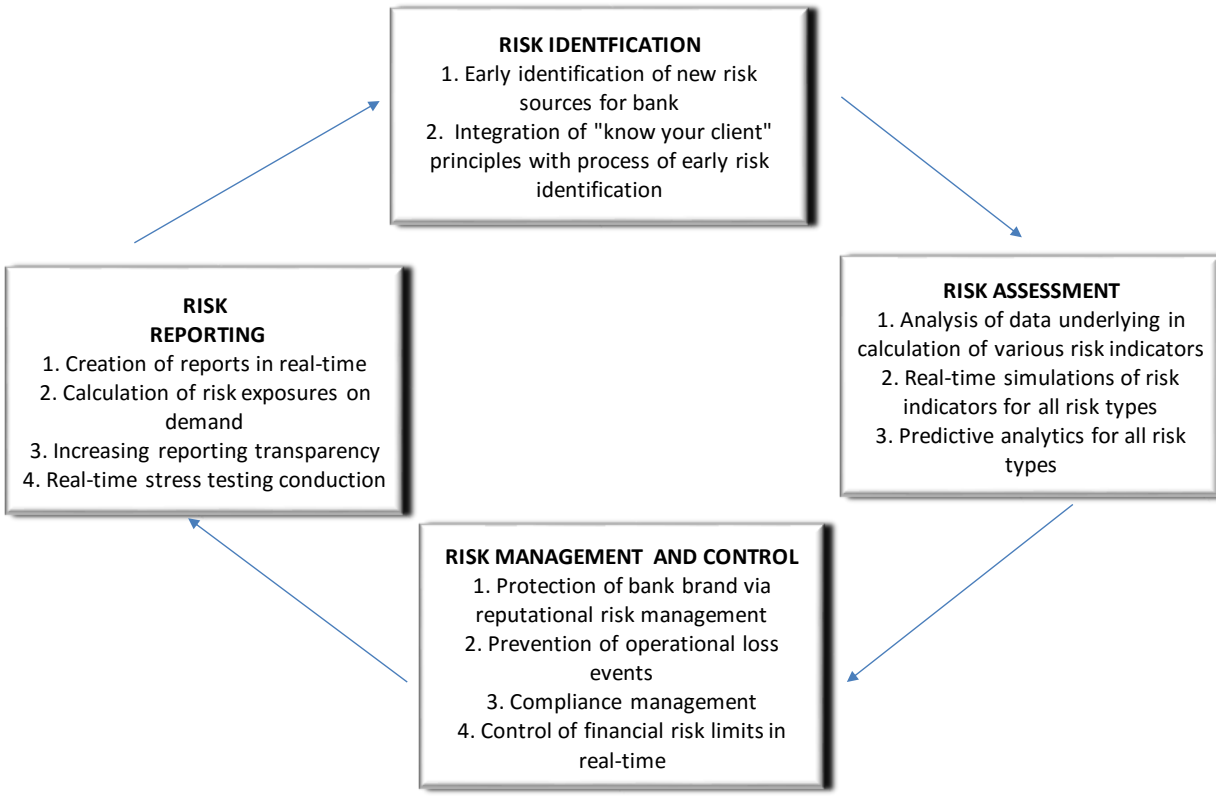

Figure 2. Framework for big data usage in risk management

6 Conclusion

In this paper authors tried to synthetize knowledge available from literature on big data usage in risk management. Literature is divided in scientific resources and practical research conducted by consulting companies. Further to this authors proposed framework for allocation of potential big data usage to key risk management activities. The expected contribution of big data usage in risk management process according to this framework is in minimization of time required to respond to new risks banks are faced with through direct risk identification, quantification of risks in real-time, daily control of risk exposures and increasing reporting transparency.

Conclusion of the paper is that big data technologies can have practical value in bank's risk management, especially when having in mind ever increasing regulatory requirements banks are faced with. Proposals for further research on this topic include "on the field" research of big data technologies practices employed in Croatian banks and their maturity levels according to big data usage, identification of gap between proposed framework and technologies, research on costs of big data technologies implementation and evaluation of cost/benefit ratio for big data technologies usage in risk management.

References

Capgemini Consulting: *Big Data Alchemy: How can Banks Maximize the Value of their Customer Data?* (2014)

Capgemini Consulting and the Economist Intelligence Unit: *The Deciding Factor: Big Data and Decision-making* (2012)

Cerchiello, P., Giudici, P. & Nicola, G. (2016). Big data models of bank risk contagion. DEM Working Paper Series, 117.

Deutsche Bank: *Big Data: How it can become a differentiator* (2015)

Duan, L.& Xiong, Y. (2015). Big data analytics and business analytics. *Journal of Management Analytics*, 2 (1), 1-21.

Ernst&Young: *Big risks require big data thinking, Global Forensic Data Analytics Survey* (2014)

Gandomi, A., Haider, M. (2015). Beyond the hype: Big data concepts, methods and analytics. *International Journal of Information Management*, 35, 137-144.

Gartner IT Glossary, What is Big Data? (2014), Retrieved from http://www.gartner.com/it-glossary/big-data/ (17.05.2016.)

Hu, D., Zhao, J.L., Hua, Z. & Wong, M.C. (2012). Network-based modelling and analysis of systemic risk in banking systems. *MIS Quarterly*, 36 (4), 1269-1291.

IBM: *Operational risk management in the world of big data* (2014)

Irving Fisher Committee on Central Bank Statistics: *Central banks' use of and interest in "big data"* (2015)

Krishna, D. (2016). Big Data in risk management. *Journal of Risk Management in Financial Institutions*, 9 (1), 46-52.

McKinsey&Company: *Digital and Big Data Opportunities in Credit Risk* (2015)

McKinsey&Copany: *The future of bank risk management* (2016)

Mohanty, S., Jagadeesh, M. & Srivatsa, H. (2013). *Big Data Imperatives: Enterprise Big Data Warehouse, BI Implementations and Analytics*. Apress.

Ngai, E.W.T., Hu, Y., Wong, Y.H., Chen, Y. & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50 (3), 559-569.

Oracle Enterprise: *Big Data in Financial Services and Banking* (2015)

Özköse, H., Sertac Ari, E., Gencer, C. (2015). Yesterday, Today and Tomorrow of Big Data. *Procedia – Social and Behavioural Sciences*, 1042-1050.

Ravisankar, P., Ravi, V., Raghava Rao, G. & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems*, 50 (2), 491-500.

Regulation EU No 575/2013 of The European Parliament and of The Council on prudential requirements for credit institutions and investment firms and amending Regulation EU No 648/2012 (2013)

The Economist Intelligence Unit Limited: *Banks and big data: Risk and compliance executives weigh in* (2015)

The Economist Intelligence Unit Limited: *Retail banks and big data: Big data as the key to better risk management* (2014)

Tian, X., Han, R., Wang, L. Lu, G. & Zhan, J. (2015). Latency critical big data computing in finance. *The Journal of Finance and Data Science*, 1, 33-41.