March 14, 2017

Re: Comments to the ESAs' discussion paper on big data in financial institutions

Dear Sir or Madam

On behalf of the Center for Data Innovation (datainnovation.org), we are pleased to submit these answers in response to a request for comments from the Joint Committee of the European Supervisory Authorities (ESAs) on its discussion paper on big data in financial institutions.

The Center for Data Innovation is the leading think tank studying the intersection of data, technology, and public policy. With staff in Brussels and Washington, the Center formulates and promotes pragmatic public policies designed to maximize the benefits of data-driven innovation in the public and private sectors. It educates policymakers and the public about the opportunities and challenges associated with data, as well as technology trends such as artificial intelligence, predictive analytics, open data, cloud computing, and the Internet of Things. The Center is a nonprofit, nonpartisan research institute affiliated with the Information Technology and Innovation Foundation.

The Center welcomes the discussion paper, which presents a largely accurate view of the benefits of big data in the financial sector. However, the Center is concerned that some basic economic tradeoffs (such as acceptances and rejections for loans and insurance service) as well as some benefits (such as fairer and more accurate distribution acceptances and rejections) are mistakenly characterized as risks, threats, or costs. The Center also believes that the so-called "right to be forgotten" of the incoming General Data Protection Regulation (GDPR) is neither a feasible nor effective means for ensuring fair and accountable use of algorithms in finance. These concerns, and recommendations for addressing them, are set out in our answers to the ESA's questions below.

Yours sincerely,

Daniel Castro
Director
Center for Data Innovation
dcastro@datainnovation.org

Nicholas Wallace
Senior Policy Analyst
Center for Data Innovation
nwallace@datainnovation.org

SUMMARY

The Joint Committee of the European Supervisory Authorities (ESAs) published a discussion paper on the use of big data by financial institutions, and posed a set of questions for respondents to answer.[1] The Center welcomes the discussion paper, which presents a largely accurate view of the benefits of big data in the financial sector. More effective use of data could make financial services lower cost, fairer, more flexible, more convenient, and more accessible to a greater number of people.

However, the Center is concerned that the paper mistakenly portrays some straightforward economic tradeoffs (such as acceptances and rejections for loans and insurance service) as well as some benefits (such as fairer and more accurate distribution of those acceptances and rejections) as risks in need of a response. EU policy also threatens some of the benefits of big data. Restrictions on the repurposing of personal data undermines the basic premise of big data, which is that vast quantities of data generated in a variety of different ways can be harnessed for new purposes. Furthermore, the so-called "right to explanation" of the incoming General Data Protection Regulation (GDPR), is neither a feasible nor effective means for ensuring fair and accountable use of algorithms in finance.

QUESTIONS ON "DESCRIPTION OF THE PHENOMENON"

1. Do you agree with the above description of the Big Data phenomenon? If not, please explain why. Please also mention whether you consider that other characteristics are relevant to understanding the use of Big Data.

The description is for the most part accurate, but incomplete. Big data will improve productivity, but the benefits of big data are not merely economic (page 8). The same methods and technologies can be used for social purposes too, as demonstrated in a recent report published by the Center for Data Innovation.[2]

[1] Joint Committee of the European Supervisory Authorities, 'Joint Committee Discussion Paper on the Use of Big Data by Financial Institutions', December 19, 2016, https://esas-joint-committee.europa.eu/Publications/Discussion%20Paper/jc-2016-86_discussion_paper_big_data.pdf

[2] Alexander Kostura and Daniel Castro, 'Europe Should Promote Data for Social Good', Center for Data Innovation, October 3, 2016. http://www2.datainnovation.org/2016-data-social-good.pdf

2. Which financial products/activities are (likely to be) the most impacted by the use of Big Data and which type of entities (e.g. large, small, traditional financial institutions, Fintechs, etc.) are making more use of Big Data technologies? In light of ESAs' objective to contribute to the stability and effectiveness of the financial system, to prevent regulatory arbitrage, do you consider that there is a level playing field between financial institutions using Big Data processes and those not using them (e.g. because they do not have access to data or the (IT) resources needed to implement Big Data processes) or between established financial institutions and potential new entrants (e.g. Fintechs) using Big Data processes? Please explain.

Financial institutions that make use of data and analytics are going to be more competitive than those that do not. Better use of data enables financial institutions to achieve many goals, including more efficiently acquiring and retaining customers, identifying and preventing fraud, and assessing credit risk, and lower costs. For example, some lenders are using data analytics to better predict risk and calculate interest rates.[3] These IT resources necessary to use data analytics are available to firms of all sizes. Established financial institutions operating globally will be less competitive if they do not operate in a regulatory environment that facilitates competition from new entrants and others investing in data-driven innovation. Therefore, the priority for regulators should be to ensure that rules support data-driven innovation, rather than trying to create a level playing field for firms who do not make use of this capability.

3. Do you offer/are you considering using Big Data tools as part of your business model? If so, please briefly describe: i) what type of entity you are, e.g., long established, start-up, a product provider, an intermediary; ii) the service you provide; iii) the nature of your clients; iv) your business model; v) whether the Big Data tools/strategy were developed by an external company or internally and whether you have related agreements with other entities (including non-financial entities)23; vi) what are the types of data used (personal, anonymized, user data, statistical data etc.) sources of data; and vii) the size of your Big Data related activity and/or forecast activity (e.g. to what extent are business decisions already taken on the basis of Big Data analysis; what other business actions could be based on Big Data in the future)?

N/A

[3] Shunpo Chang, Simon Dae-oong Kim, and Genki Kondo, 'Predicting Default Risk of Lending Club Loans' CS229: Machine Learning, Standford University, Autumn 2015-2016 http://cs229.stanford.edu/proj2015/199_report.pdf

4. If you are a consumer or a consumer organization, do you witness any of the uses of Big Data? In what fields?

Companies that use data create many benefits for consumers, including designing better products and services, enabling more personalization, and offering lower prices

First, firms use data to better understand and predict what their customers want, which in turn helps them to develop better products. For example, usage data from connected devices, such as fitness trackers or smartphones, or from computer software, can help developers to identify opportunities for improvement in the design and do better in the next update or iteration, without having to rely on sparse feedback from customers themselves. This mechanism is as appropriate for financial products and services as any other. For example, financial institutions may use data-driven insights to identify an underserved group of customers and design a new product or service to meet their needs.

Second, data enables firms to offer personalization, such as recommendations for online shopping or music and movies. Banks also make use of this method, such as by offering savings accounts or personal loans to customers based on their circumstances.

Third, data-driven services reduce costs which mean lower prices for consumers. For example, more accurate insurance assessments can reduce premiums for low-risk customers.

5. Do you consider there are (non-regulatory) barriers preventing you (or which could prevent you in the future) from collecting and processing data? Are there barriers preventing you from offering/developing Big Data tools in the banking, insurance and securities sectors? If so, which barriers?

One important non-regulatory barrier to collecting and processing data is skills shortages.[4] The policy problem of the skills gap is complex and cannot be solved with a single policy measure. There is also no specific solution for the financial sector: this is a labor market problem. Policymakers should strengthen support for data-related skills across a variety of academic disciplines in the educational sector. Policymakers should also develop labor market policies that incentivize investment by employers in training inexperienced candidates, and tackle negative externalities—such as competitive pressure on

[4] Paul MacDonnell and Daniel Castro, 'Europe Should Embrace the Data Revolution' Center for Data Innovation, February 29, 2016, http://www2.datainnovation.org/2016-europe-embrace-data-revolution.pdf

young people to undertake unpaid work in order to get experience, which freezes-out individuals from low-income families, regardless of their abilities.

The ease of acquiring the data itself depends on the data in question. Customer data is obtained by having customers, and therefore cannot reasonably be treated as an unfair barrier to entry. Incumbents in the credit scoring industry may have some advantage due to their possession of significant historical data, which they often use to calculate investment risks. However, this does not necessarily constitute an insurmountable barrier, so it does not follow that this is a market failure in need of regulation. Startups such as Transferwise, Funding Circle, and Zopa have built successful financial services business without access to incumbents' data.

Another non-regulatory barrier is that some societal groups may be excluded from certain types of public and private sector services because there is simply not enough information collected about them.[5] In the European context, the social and economic exclusion of the Roma people are a clear example of this. This exclusion could become worse as the importance of data grows, and the solution is to address this "data divide" so that all populations benefit from data-driven innovation rather than limit the benefits of this technology to everyone.

[5] Daniel Castro, 'The Rise of Data Poverty in America' Center for Data Innovation, September 10, 2014, http://www2.datainnovation.org/2014-data-poverty.pdf

CENTER FOR DATA INNOVATION

QUESTIONS ON "REGULATORY FRAMEWORK APPLICABLE TO BIG DATA"

6. Do you agree with the above short, non-exhaustive, presentation of some of the main applicable requirements? If not, please explain why. Please also mention whether you consider that other legal requirements are essential and should be mentioned.

The section accurately describes the European regulatory environment, but it overlooks factors such as the so-called "right to explanation" of GDPR Articles 13, 14 & 22. This rule has implications for finance, where algorithms may be used to assess loan eligibility. If a decision to refuse a loan is taken by a human, customers have a legal right to know what data was used, but not to an explanation of the logic involved in the decision. But when the GDPR comes into force, if the same decision is made by an algorithm, the customer has the right to an explanation of the logic involved, which may be both unintelligible to the customer and very costly to provide. This may deter some companies from using algorithmic decision-making, which in turn would diminish opportunities to make lending fairer and more accessible.

The law should ensure that there is human and corporate accountability in the use of artificial intelligence. However, the right to explanation introduces new requirements for algorithmic decision making that were not deemed necessary to achieve accountability in human decisions. Rights to information and explanation should be based on the nature of the decision, and should be independent of the technology used (or not used) to make it.

The report also does not address the problem of barriers to data flows across borders: EU policymakers should outlaw national restrictions on data flows to other EU member states, and should reduce barriers to data-flows to non-EU countries.[6] This is important because machine learning systems are improved with larger pools of data to access.

7. Do you consider any of these regulatory requirements as unjustified barriers preventing you from using Big Data technologies? If so, please explain why. Please also explain whether you consider that further regulation (including soft law/guidance, etc. and insofar as it falls within the scope/remit of the ESAs) should be introduced to facilitate the use of Big Data technologies.

---

[6] Nick Wallace ' European Commission Should Stand Firm on Free Data Flows' Center for Data Innovation, March 8th, 2017, https://www.datainnovation.org/2017/03/european-commission-should-stand-firm-on-free-data-flows/

The regulations are excessive. In addition to the aforementioned problems with the right to explanation, the prohibitions on data reuse are an unnecessary barrier to the use of big data. On page 14, the discussion paper says that: "while the nature of Big Data encourages the use of as much data as possible to see what connections can be made, financial institutions should be able to justify the use of certain data categories… as any further processing should not be incompatible with the purposes originally specified when the data was collected, financial institutions should assess the extent to which they could use, for instance, purchased data for own purposes [sic]."

If all repurposing of data is illegal, then the effective use of data is not possible, because data processing tends to involve analyzing data generated and accumulated by various different activities and giving it new purposes. For example, evidence from California shows fitness tracker data can be repurposed to track the impact of earthquakes.[7] Data protection law should require transparency in how personal data is used, and should restrict uses that are harmful, but there should not be a blanket ban on repurposing data, particularly because the goal in this case would be to improve the provision of financial services.

QUESTIONS ON "POTENTIAL BENEFITS AND RISKS FOR CUSTOMERS AND FINANCIAL INSTITUTIONS"

8. Do you consider the potential benefits for consumers and respectively financial institutions to be accurately described? Have you observed any of them in practice? If so, please provide examples. If not, please explain whether you are aware of any barriers that may prevent the above potential benefits from materializing?

The description of the potential benefits is largely accurate.

9. Do you agree with the description of the risks identified for consumers and respectively financial institutions? Have you observed any of these risks (including other risks that you are aware of) causing detriment to consumers and respectively financial institutions? If so, in what way? If not, please explain why. Please also mention whether certain risks for consumers and financial institutions have not manifested yet but have the potential of developing in the future and hence need to be closely monitored by Supervisory Authorities.

[7] Nick Wallace 'Norwegian Watchdog Turns Fire on Fitness Trackers and Misses the Mark Entirely' Center for Data Innovation, January 10th, 2017, https://www.datainnovation.org/2017/01/norwegian-watchdog-turns-fire-on-fitness-trackers-and-misses-the-mark-entirely/

Some of the so-called risks suggest something of a "have your cake and eat it" attitude to the distribution of finite resources. Paragraphs 34 and 38 claim as a risk the fact that in addition to granting previously excluded customers access to insurance and loans, financial services that use big data may exclude some customers who may otherwise have been accepted. This completely ignores the point that the role of a greater supply of data in such decisions is to make them more accurate, which inevitably leads to redistribution of outcomes. Customers that are incorrectly accepted today for insurance and loans raise premiums for everyone else. If data analysis can lead to more accurate decisions in the future, then that is in society's interest. More sophisticated data-driven pricing mechanisms could make it financially viable for firms to insure or lend to a greater number of people overall, but that some people may be found unsuitable upon analysis of new information is not a risk, it is a benefit.

Paragraph 38 also asserts that customers in high-risk areas may struggle to get house insurance: this is already the case without better use of data. Better use of data, however, may narrow these zones of exclusion by using more sophisticated details--such as analysis of floodplains based on GPS coordinates, rather than postcodes.

Paragraph 39 expresses the fear that behavioral analysis may be unfair or discriminatory, and that such discrimination may not be transparent. Behavioral monitoring could in fact make things fairer, as characteristics previously regarded as risky—such as certain professions, or youth—may be dismissed in favor of demonstrably low-risk behavior, such as driving habits. Similarly, high-risk behavior may reduce the negative externalities created by free riders in the system, such as reckless individuals with supposedly low-risk characteristics, who previously benefited at the expense of others.

Concerns about bias and transparency are justified, as myriad innocuous characteristics may have subtle relationships to criteria on which it would be prejudicial and unfair to discriminate, such as ethnicity, leading to unwittingly biased decisions. Moreover, these characteristics are impossible to predict, as they will vary greatly throughout Europe. The solution to this problem is to monitor outcomes relative to protected characteristics (like ethnicity) and, controlling for factors (such as income) to identify bias and the sources of it. Moreover, financial institutions have a financial interest in reducing bias, which can affect decisions without the use of big data, and big data tools can help them do that.

Unfortunately, the current regulation on this matter—the GDPR's "right to explanation"—fails to achieve this level of accountability, as it requires the explanation of individual decisions, which in

addition to being unworkable and costly, will be far less effective in identifying bias than the method described above.

Finally, the need for transparency in these decisions is independent of whether data or algorithms are used. Humans are fully capable of both deliberate and unintended bias, and existing regulation does not impose the same standards of transparency on human decision-makers as on machines.

10. Is the regulatory framework adequately addressing the risks mentioned above? Bearing in mind the constant evolution of technologies/IT developments and that some of the above mentioned regulatory requirements are not specific to the financial services sector (e.g. GDPR), do you think further regulation is needed to preserve the rights of consumers of financial services in a Big Data context? Please explain why.

As described in the previous answer, the GDPR's solution to algorithmic bias will fail because it requires transparency in individual decisions when the data subject demands it, which is an unreliable and inadequate method for detecting bias systematically, and for establishing accountability in the use of algorithms and big data. The behavior of algorithms and the outcomes for protected groups has to be carefully monitored over time in order for bias and the sources of it to be identified, and to ensure algorithms and big data are being used responsibly.

11. Do you agree that Big Data will have implications on the availability and affordability of financial products and services for some consumers? How could regulatory/supervisory authorities assist those consumers having difficulties to access financial services products?

The answer to the first question is yes; it will make products more accessible. In answer to the second question, regulators should monitor algorithmic behavior and outcomes in the manner described above in order to identify whether exclusions are fair. Regulators should also take steps to free-up more data, including under certain conditions privately held data, to improve transparency in markets and support further data driven services. For example, choice engines draw on publicly available information—such as data on restaurant hygiene inspections—to help customers make more informed choices.[8] If transparency rules in the financial sector were coupled with standards on machine-readable data, online

[8] Daniel Castro, 'States Should Use Open Data to Empower Customers', Government Technology, July 9, 2015 http://www.govtech.com/opinion/States-Should-Use-Open-Data-to-Empower-Consumers.html

tools and services could draw on this data to make it easier for consumers to compare financial services, thus creating more competition.

12. Do you believe that Big Data processes may enable financial institutions to predict more accurately (and act accordingly) the behavior of consumers (e.g. predicting which consumers are more likely to shop around, or to lodge a complaint or to accept claims settlement offers) and could therefore compromise the overarching obligations of financial institutions to treat their customers in a fair manner? Please explain your response.

Financial institutions can already shirk their responsibilities to treat people fairly without the help of data. It goes without saying that some could—-and probably will—use behavioral analytics (and any other tools they have available to them) unfairly, but it would be a fallacy to conclude from this that behavioral analytics is intrinsically unfair. Behavioral analytics allows companies to better predict what customers might want and need. This enhances business opportunities for companies (if it did not, they would not bother) but it also improves customer satisfaction. Rules already exist regarding false advertising and hard-selling to customers with diminished responsibility: provided these rules are enforced in data-driven services and advertising, there is little cause for concern. Competent adults of sound mind are always responsible for their own actions and decisions, no matter how well-timed or well-targeted a given offer is. As mentioned above, proper oversight of algorithmic decision making will also help to enforce accountability.

13. Do you agree that Big Data increases the exposure of financial institutions to cyber risks? If yes, what type of measures has your institution adopted or is going to adopt to prevent such risks? What could supervisory/regulatory authorities do in this area?

The quantity of data a firm possess has no impact on the quality of its cybersecurity. Having more data does not weaken cybersecurity. However, it does raise the costs of a successful cyberattack. Existing regulation already manages liability for such data breaches well. But a sensible additional measure would be to force companies to publish cybersecurity policies. Currently, bad cybersecurity practices are only exposed by hackers. The regulation would create a strong incentive for firms to invest in better protections.

Data-driven technologies can also help firms improve their cybersecurity practices, such as by developing firewalls and antivirus systems that rely on predictive analytics. Unlike traditional cybersecurity systems, which rely on lists of predefined attacks, these methods are effective against “zero-day” threats that have not yet been identified. Companies that use big data for their products and

marketing are more likely to have the in-house skills necessary to support these new kinds of cybersecurity measures.

14. Would you see merit in prohibiting the use of Big Data for certain types of financial products and or services, or certain types of customers, or in any other circumstances?

No. Even if there were any justification for that—and there is not—it is hard to imagine how such a prohibition could work. One might just as well say banks should not use anything more sophisticated than an abacus. A better option is for regulators might be to consider appropriate protections for vulnerable consumers—such as people with learning difficulties—but without impairing these people's ability to access data-driven financial services. For example, a complaint from the next-of-kin of a person with diminished responsibility for their actions should void have the power to void a transaction by that person that resulted from data-driven advertising or cross-selling. If anything the focus in Europe should be on spurring the transformation of financial services through better use of data in order to boost consumer welfare.

15. Do you agree that Big Data may reduce the capacity of consumers to compare between financial products/services? Please explain your response.

On the contrary, it might make it easier. One of the difficulties of comparing financial products today is translating their offers and terms to one's own individual circumstances. Better data would allow that to happen automatically: for example, price comparison websites could provide customers personalized comparisons of various financial services based on their unique circumstances. One way to enable these types of services is by encouraging data portability, as per GDPR Article 20 so that consumers can download their data from their financial service provider in a standardized format. Consumers could then ask third-parties to analyze their data to help them identify opportunities for savings. Another way to enable this type of competition is by encouraging financial institutions to make their fees and terms available in standardized, machine-readable formats so that offers can be more easily compared. Both of these can be achieved if financial institutions make better use of open application programming interfaces (APIs)—sets of functions that allow any developer to build software that interacts directly with other digital services.

16. How do you believe that Big Data could impact the provision of advice to consumers of financial products? Please explain your response.

As highlighted above, information on financial products given by their providers is typically given on a non-advice basis. Third-party financial advisors—accountants, for example—would need to keep up-to-date with new technologies and practices, and may well have to use data-driven tools of their own. Data portability will help customers ensure the people charged with advising them have access to the necessary data, and this will most likely need to be enforced through regulation. Furthermore, some kinds of financial advice can be automated using data-driven systems, called "robo-advisors," which lower the cost of basic financial advice, thereby making it more accessible to more consumers. As mentioned above, companies can also use data to personalize price comparison services.

17. How do you believe Big Data tools will impact the implementation of product governance requirements? Please explain your response.

N/A

18. How do you believe Big Data tools will impact know-your-customer processes? Please explain your response.

N/A

QUESTIONS ON "POSSIBLE EVOLUTION OF THE MARKET?" (P. 35-36)

19. What are key success factors for a Big Data strategy (i.e. the adaptation of the business model/plan towards Big data driven technologies and methods)?

N/A

20. What are the greatest future challenges in the development and implementation of Big Data strategies?

N/A

21. The discussion paper refers to a number of measures and tools meant to ensure compliance with conduct and organizational regulatory requirements as well as data and consumer protection rules in

the context of big data analytics. Are other measures and tools needed? If so, what are they and what they should cover?

As mentioned above, monitoring outcomes in applications of artificial intelligence are important to maintain accountability, and are more effective and more feasible than the current attempts at algorithmic transparency set out in the GDPR.

22. How do you see the development of artificial intelligence or blockchain technology in connection with Big Data processes?

AI is necessary to keep up with the pace of data collection. It is not feasible to have humans monitor all of the data that can be collected today, so some of this work must be handled by algorithms. In addition, one important development in the field of artificial intelligence are machine learning algorithms. Many of these machine learning techniques (i.e. supervised learning) require vast amounts of curated data in order to train the system to perform a particular task.

Blockchain—or, more correctly, distributed ledgers—are a means to transparently and verifiably log information. Ergo, besides the well-known uses in supporting digital currency transactions, the technology could also support the use of big data in more conventional financial services—such as loans and insurance—by creating an automatic and accessible record of transactions, decisions, and credit-relevant occurrences.[9]

ADDITIONAL COMMENTS:

23. Are there any other comments you would like to convey on the topic of use of Big Data by financial institutions? In particular, are there other relevant issues that are not covered by this Discussion Paper?

N/A

[9] Alan McQuinn, Weining Guo, and Daniel Castro, 'Policy Principles for Fintech', Information Technology and Innovation Foundation, October 2016, http://www2.itif.org/2016-policy-principles-fintech.pdf?_ga=1.117079046.1075727067.1489163431