Privacy Concerns and the Regulatory Future of Big Data

William Barker
Scott Breitbach
Peyton Mosbaugh
Bellevue University
Bellevue, NE 68005, USA
wbarker@my365.bellevue.edu
sbreitbach@my365.bellevue.edu
mpmosbaugh@my365.bellevue.edu

Abstract
Data science can be a useful tool in solving problems by using data to create predictive models, but as data collection becomes both easier and more ubiquitous, we've seen an explosion in the volumes of data collected and stored on individual citizens. This data is used by corporations and governments for a wide variety of purposes. Often these tools can benefit both those collecting the data as well as the individuals themselves, but sometimes they can have negative consequences, whether the intent was malicious or not. In order to protect the privacy of the individuals, we need clear guidelines and regulations around data collection and use. Many organizations and governments have already started down this path, but as with any rapidly developing technology, some are lagging behind. Here we will look at some current guidelines and regulations and where it looks like the future of data privacy regulation is headed.

Author Keywords
Data privacy; data security; data collection; data regulations; big data.

Introduction
As technology has progressed in the last 50 years, computers have shrunk, processing power has increased, and computers have integrated themselves into nearly every aspect of our lives, from smart homes and cars, to social networks, online shopping, and even our Internet searches. This increase in computer technology has led to an explosion in data due to the ease of collection and the ever-decreasing cost of processing and storing large data sets.

These data are used primarily by corporations, governments, and researchers and for a wide variety of purposes. Ideally, the information generated from this data will benefit both those doing the data collection and analysis and the individuals or groups from whom the data is being collected [13].

Generally speaking, data used for research is good for everyone because the end goal is typically the pursuit of knowledge, though there can still be unintended biases, both in the data and in the algorithms we build, which we need to constantly be on the lookout for. Oftentimes we might be mistrustful of corporations [6] and governments collecting our data because corporations are often concerned primarily with their bottom line and governments might want to control their populace, however they can be good too. Targeted marketing from corporations can help a customer find a useful product that they didn't know exists and governments often use traffic data to plan roads in order to relieve congestion.

With all this data being collected, security and privacy are major concerns that arise [10]. Problems can arise both intentionally (e.g. through those using the data for unethical purposes or through third-

party data hacking), and unintentionally (e.g. when inherent data biases aren't accounted for or when sufficient security measures aren't taken).

In order to protect against these privacy and security risks [19], we need to utilize a multi-pronged approach. We will need regulations to maintain checks on the industry as well as standards for the industry to follow. For those working with the data, training around ethics [13], biases, and security as well as building security into the software architecture itself [1].

Many have already started down this path, but as with any rapidly developing technology, some are lagging behind. Here we will look at the current state of privacy and security guidelines and regulations around big data as well as where it looks like the future of data privacy regulation is headed [14].

Building Security and Privacy Protection into Data Architecture

With the sheer quantity of data being collected, it has to be stored somewhere. Initially organizations would store their data in their own databases and data warehouses, but as the Internet has become all-encompassing, most data storage and management, as well as computing power itself, has moved to the cloud. This move to the cloud has come with its own privacy and security risks [20] in the storage and in the constant exchange of data [21]. For instance, each time a user or computer makes a request to grant access to data on a server there is an opportunity for a breach of security [22].

In light of these risks, the data science community may need to rethink their practices and methodologies, from how queries are processed [23] to how data is discovered [24], even how machine learning is used [25]. To this end, there are several approaches currently being pursued within the big data framework, such as incorporating security and privacy controls directly within the code during processing [26] and utilizing encryption in order to ensure security of the data [27] or even user-facing software approaches such as showing the usage of the data directly and requiring authentication [28]. Trust models can also be used to measure security strength of cloud services [29].

Data brokers are those that collect personal data from numerous sources and then sell that data to business worldwide. Oftentimes this happens without consumer knowledge and can be harmful and discriminatory. While US policy makers have been urged to regulate these brokers, they remain unregulated, whereas the European Union has introduced broad data privacy guarantees for consumers and a data protection regime that limits data brokers' activities if consumers' consent has not been obtained for specific purposes [18].

New Regulations and Standards

The EU, or European Union, has really led the way in terms of putting regulations into place to help protect people's personal data. In May of 2016, the EU put together the GDPR, the General Data Protection Regulation. Under the GDPR, both individuals and companies benefit as people have more control over their personal data and businesses get put into a level playing field. The first regulation is a person's right to receive clear and understandable information regarding their data. This includes knowing who is process their data, what kind of data they are collecting, and why their data is even being collected. An individual also has the right to request access to this data at any point in time. The right for an individual to request one company to send his or her personal data to another company [3]. One example of this is when an individual links up their social networking accounts, such as sending photographs from Instagram to Facebook [19]. Initial consent must be granted by the individual whose data is being collected, and at any point after, the individual has the right to be able to have their data forgotten. If a company experiences a data breach with important data from the individual that may put them at risk, the company must alert the individual immediately. The GDPR also outlines stricter guidelines to protect the data of children. Those who are under the age of 16 must also have parental consent to sharing the data, as children often do not understand the consequences of this sharing. More specific details were outlined for situations of data collection being used in cases of legally binding agreements like loans. The individual must first be informed that their data is being used in their consideration, and if the

consideration results in a refusal of a loan or the like, a person, rather than a computer, must check the processing and the decision. In the case that data is used in a situation like this, the individual also has the right to contest the decision. The GDPR lays out guidelines for the use of data in marketing as well. Individuals have the right to opt out of direct marketing based on their data storage and collection at any time [7]. These are just the main, over-arching policies that are outlines in the GDPR.

All of these policies are great on paper and seem to have the ability to protect an individual's data, but how are they being enforced? The European Data Protection Board is an independent European body that is made up of representatives of National Data Protection authorities from countries belonging to the European Union. The EDPB is in charge of ensuring the consistent application of these rules protecting data privacy throughout the EU, as well as reprimanding businesses who are not following these guidelines and policies [3]. Since the GDPR requires vast changes in the way that some companies process data, there are different stages of consequences. At first a company will get a warning that they are not following the GDPR, followed by the company being officially reprimanded. If after this, they are caught again not following the guidelines, they will have to suspend and stop all data processing for their entire company. A company that continues to disregard the policies to protect privacy along with previous consequences will have to pay a fine. This fine can be up to 20 million Euros or be as high as 4% of all global annual turnover for the company. If individuals do not feel as though a company is properly collecting and using their data, individuals are able to call National DPA hotlines that are set in place. Having this direct point of contact for individuals is another great way to help make the individuals feel safe and secure with their data privacy [6].

How effective have the GDPR and the EDPB been in protecting the rights of individuals and their data privacy? Approximately only 20% of businesses believe that they are now GDPR compliant, and even worse, is the fact that about one in every four companies have not even began to work on making their organization complaint. Businesses in the EU countries were given one year to get their organizations up to date with the regulations, but almost half of businesses are still in the implementation stage. Although this does not seem very promising, some bigger companies have already been caught and are now facing fines by failing to comply. British Airways is facing up to 200 million Euros in fines due to a data breach that occurred at the company in 2018 because the company did not properly inform individuals [7].

It seems as though the countries of the European Union have a very long way to go in terms of getting companies up to speed with the new policies, but at least they have started the journey to help make individuals feel safe with their data and privacy. The EU countries are the first to initiate a big action in policy change to protect data, but with growing concerns from the public on the use and collection of their data, they are likely not the last to put legal policies into play to help make individuals feel secure.

Regulatory Impact on Data Collection

As regulatory framework around data collection is laid out in a variety of ways around the world, how companies and government bodies are able to collect and use citizen personal data will begin to change. In countries like the US, where a "self-regulation" approach to regulation is followed and no general federal data protection laws are enforced, the way data is collected is clearly unaffected by regulation, and the responsibility of respecting consumer privacy falls on companies through either a code of conduct or legal contracts. [13] The Asia-Pacific Economic Cooperation (APEC, a forum of 21 Asia-Pacific economies) has developed a self-regulatory framework setting out the principles that economies should implement and companies then follow to ensure a common, minimum level of data protection across member economies. The aim is to enable the easier transfer of data among economies where the level of data protection regulation varies greatly. [13] As long as all member countries follow these minimal guidelines when collecting data, countries with no data collection laws can still send data to those with heavy restrictions. The regulatory frameworks set up in the European Union not only works to protect data and privacy in its member countries, but also defines what is considered personal data and how such data can or cannot be used, as well as well as sets organizational and

technological requirements. [13] These regulations will have a greater impact on data collection than anywhere else worldwide, with companies being required to implement technological and organizational measures to protect the data gathered, and European countries not being able to send personal data at all to any country that it deems does not have an appropriate level of protection. [13].

The Future of Regulation

So far it seems like if the US and other countries around the globe decide to increase data regulation, they will most likely follow the European Union's lead. How long it will take for those countries to finally implement those regulations is still up in the air, and how they go about enforcing it will vary. Future potential of enabling "open data" (data that can be freely used, re-used, and redistributed by anyone) could generate more than $3 trillion in additional value every year, if enabled over education, transportation, consumer products, electricity, oil and gas, healthcare, and consumer finance domains. However, there are several obstacles in the way of generating this value. Increased uptake of big data will require the adoption of next-generation telecommunications infrastructure, which is still in its early development in many parts of the world. There also needs to be a large enough pool of talent with the advanced analytical skills needed to put the data to good use, and this workforce will need to be trained. Big data uptake will also depend on whether ways can be found to protect information technology infrastructures and the data they carry from cyberattacks. It'll also be important to build the trust of citizens as they are growing increasingly suspicious about how information about them is being used. [13]

Governments and regulators will need to frame data protection policies that safeguard the privacy of both customers and citizens, but at the same time these policies must not stifle the innovation that big data can deliver, or its attendant economic and social benefits. Data protection laws are evolving not only in an attempt to keep pace with technological developments and new ways of using, collecting, and sharing personal data, but also to keep pace with attitudes toward privacy. [13]

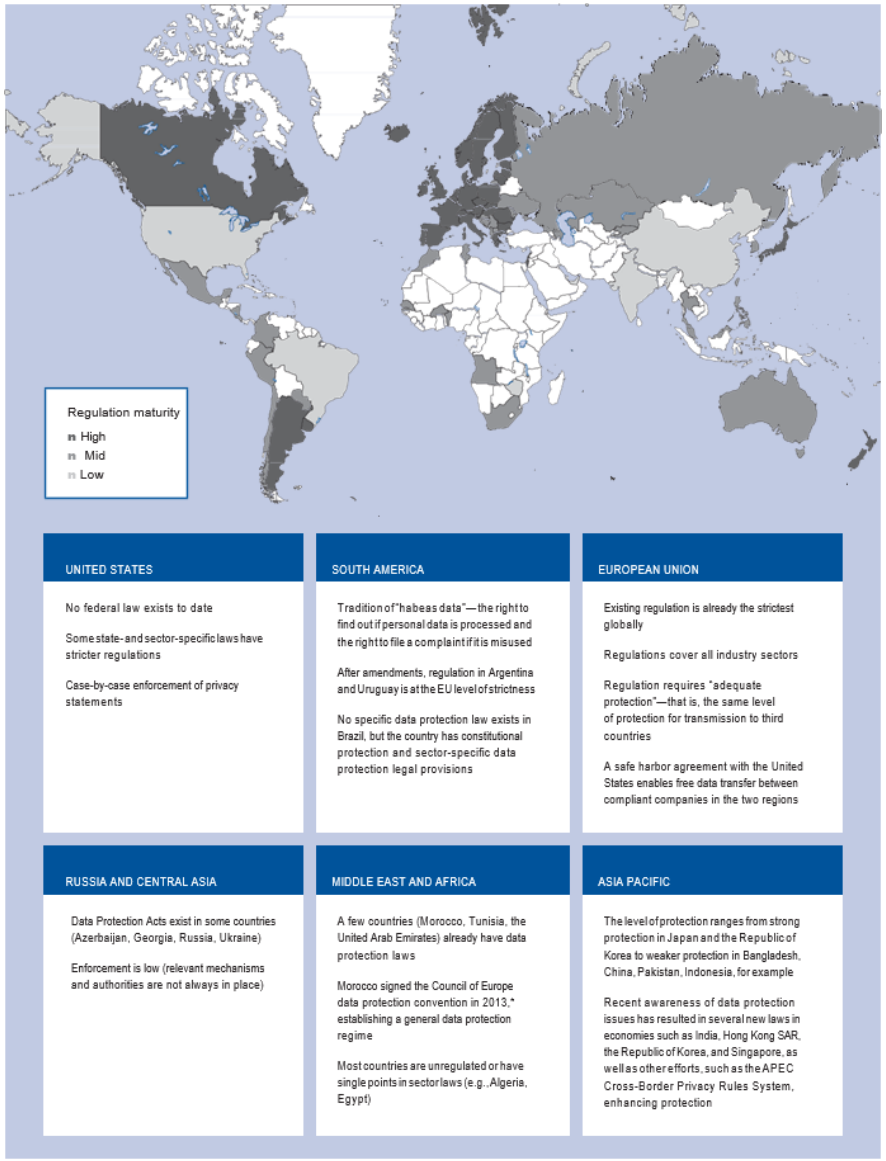

Figure 1: Current state of data regulations in the world. Source: http://www3.weforum.org/docs/GITR/2014/GITR_Chapter1.7_2014.pdf

Acknowledgements

Thank you to Professor Shankar Parajulee and our DSC500 classmates for their support.

References

1. Alfredo Cuzzocrea. 2014. Privacy and Security of Big Data: Current Challenges and Future Research Perspectives. In Proceedings of the First International Workshop on Privacy and Secuirty of Big Data (PSBD '14). Association for Computing Machinery, New York, NY, USA, 45–47. DOI:https://doi.org/10.1145/2663715.2669614
2. Cho, Do-Eun, er al. "Double Privacy Layer Architecture for Big Data Framework." International Journal of Software Engineering and It's Applications, vol. 10, no. 2, 2016, pp. 271-278., doi:10.14257/ijseia.2016.10.2.22.
3. Covert, Quentin, et al. "Towards a Triad for Data Privacy." Proceedings of the 53rd Hawaii International Conference on System Sciences, 7 Jan. 2020, doi:10.24251/hicss.2020.535
4. Kambatla, K., Kollias, G., Kumar, V.,& Grama,A.(2014). Trends in big data analytics. Journal of Parallel and Distributed Computing, 74(7), 2561-2573.
5. Lei Xu, Chunxiao Jiang, Jian Wang, Jian Yuan and Yong Ren, "Information Security in Big Data: Privacy and Data Mining," in IEEE Access, vol. 2, pp. 1149-1176, 2014. https://ieeexplore.ieee.org/abstract/document/6919256
6. Lui, Yang, and Connor Greene. "The Dark Side of Big Data: Personal Privacy, Data Security, and Price Discrimination." Digital Transformation in Business and Society, May 2019, pp145-153., doi:10.1007/978-3-030-08277-2_9.
7. Sharma, Anil, and Gurwinder Singh. "A Review of Big Data Challenges and Preserving Privacy in Big Data." Advances in Data and Information Sciences, vol. 94, 3 Jan. 2020, pp. 57-65., doi:10.1007/978-981-15-0694-9_7.
8. Shui Yu, "Big Privacy: Challenges and Opportunities of Privacy Study in the Age of Big Data," in IEEE Access, vol. 4, pp. 2751-2763, 2016. https://ieeexplore.ieee.org/abstract/document/7485855
9. Strang, Kenneth David, and Zhaohao Sun. "Big Data Paradigm: What Is the Status if Privacy and Security?" Annals it Data Science, vol. 4, no. 1,2017, pp. 1-17., doi:10.1007/s40745-016-0096-6.
10. Tene, Omer, and Jules Polonetsky. "Privacy in the Age of Big Data: A Time for Big Decisions." Vol. 63, 2 Feb. 2012.
11. Jacob Metcalf and Kate Crawford. 2016. Where are human subjects in Big Data research? The emerging ethics divide - Jacob Metcalf, Kate Crawford, 2016. (June 2016). Retrieved May 1, 2020 from https://journals.sagepub.com/doi/full/10.1177/2053951716650211
12. Mandy Chessell. Ethics for big data and analytics. 2014. PDF. (2014).
13. Scott Beardsley, Luis Eriquez, ferry Grijpink, Sergio Sandoval, Steven Spittaels, and Malin Strandell-Jansson. 2014. Building Trust: The Role of Regulation in Unlocking the Value of Big Data. The Global Information Technology Report (2014), 73–80.
14. Daniel Gozman, Wendy Currie, and Jonathan Seddon. 2015. The Role of Big Data in Governance: A Regulatory and Legal Perspective of Analytics in Global Financial Services. SSRN Electronic Journal (2015), 1–54. DOI:http://dx.doi.org/10.2139/ssrn.2752561
15. Sharyn Ohalloran, Sameer Maskey, Geraldine Mcallister, David K. Park, and Kaiping Chen. 2016. Data Science and Political Economy: Application to Financial Regulatory Structure. RSF: The Russell Sage Foundation Journal of the Social Sciences 2, 7 (2016), 87–109. DOI:http://dx.doi.org/10.7758/rsf.2016.2.7.06
16. D. Mendelson and D. Mendelson. 2017. Legal protections for personal health information in the age of Big Data – a proposal for regulatory framework. Ethics, Medicine and Public Health 3, 1 (2017), 37–55. DOI:http://dx.doi.org/10.1016/j.jemep.2017.02.005
17. Chih-Liang Yeh. 2018. Pursuing consumer empowerment in the age of big data: A comprehensive regulatory framework for data brokers. Telecommunications Policy 42, 4 (2018), 282–292. DOI:http://dx.doi.org/10.1016/j.telpol.2017.12.001

18. Jensen, M. Challenges of Privacy Protection in Big Data Analytics. Proc. of BigData Congress, 2013.
19. Betgé-Brezetz, S., Kamga, G.-B., Dupont, M.-P., and Guesmi, A. End-To-End Privacy Policy Enforcement in Cloud Infrastructure. Proc. of CLOUDNET, 2013.
20. Machanavajjhala, A., and Reiter, J.P. Big Privacy: Protecting Confidentiality in Big Data. ACM Crossroads 19(1), 2012.
21. Jang, M., Yoon, M., Chang, J.-W. A Privacy-Aware Query Authentication Index for Database Outsourcing. Proc. of BigComp, 2014.
22. Agrawal, R., and Srikant, R. Privacy-Preserving Data Mining. Proc. of SIGMOD, 2000.
23. Ishibuchi, H., Yamane, M., and Nojima, Y. Learning from Multiple Data Sets with Different Missing Attributes and Privacy Policies: Parallel Distributed Fuzzy Genetics-based Machine Learning Approach. Proc. of BigData Conference, 2013.
24. Agrawal, D., El Abbadi, A., and Wang, S. Secure and Privacy-Preserving Database Services in the Cloud. Proc. of ICDE, 2013.
25. Arasu, A., Eguro, K., Kaushik, R., and Ramamurthy, R. Querying Encrypted Data. Proc. of SIGMOD, 2014
26. Wu, C., and Guo, Y. Enhanced User Data Privacy with Pay-By-Data Model. Proc. of BigData Conference, 2013.
27. Rizvi, S., Ryoo, J., Liu, Y., Zazworsky, D., and Cappeta, A. A Centralized Trust Model Approach for Cloud Computing. Proc. of WOCC, 2014