

# Ethical Expedients of Big Data: Questions that confront us

Meghana V. Aruru

**Abstract** - In the current digital era, big data presents unique opportunities while also posing ethical challenges regarding individual privacy and freedom. Ensuing debates over "society" versus "individuals" will provide interesting opportunities to explore and think about issues that might call for new ethical norms as data becomes freely available across the globe.

---

◆

## 1 INTRODUCTION

*"Technology is neither good nor bad; nor is it neutral...technology's interaction with the social ecology is such that technical developments frequently have environmental, social, and human consequences that go far beyond the immediate purposes of the technical devices and practices themselves."* - Melvin Kranzberg (1985), historian of technology.

The era of big data is upon us, although what constitutes big data is yet to be precisely understood. While such ambiguity could allow a multiplicity of perspectives to come together and enrich the field in its initial stages, it also leaves room for certain vagueness that may give rise to matters of serious concern. For instance, today the scientists, data analysts, sociologists and scholars of many other disciplines are increasingly demanding (and feeling comfortable while doing so) access to large amounts of data resulting from daily social interactions among people, often including those that may originate in individuals' personal lives (including, of course, public figures). Is it always clear as to what the different ethical implications of such research and analyses are?

Ethical norms that were established after the horrors of World War II are rapidly outpaced while ethicists strive to keep up with the world as it continues to evolve around them.(1) The situation is not unlike that faced by lawmakers and law enforcers who may be found trying to catch up with the times during the myriad cyber invasions and security threats that happen across the globe. With an increasing societal ob-

session to quantify human behavior and interactions, it is not always easy to comprehend the full range of costs and benefits of quantitative analysis of big data to change or effect individual behaviors (say, customer recommendation systems).(2)

The excitement of gathering and churning massive amounts of data brings critical issues of quality (of knowledge) over quantity (of data) to the fore. Data generated from un-designed experiments or surveys may not be representative of the underlying population structures, much less of groups and individuals. The ethical dilemma about the overall utility of analyzing such data - whether freely available or obtained for a price - stems from the concern that, in the wrong hands, it could potentially compromise individual privacy and freedom, which are much-cherished ideals for the free societies that have developed post WWII. The question therefore is, for any individual, is personalized recommendation more important than one's privacy or consent (in the form of freedom)?

Ethical risks of big data can be broadly categorized as: (i) risk of breach of privacy and confidentiality, (ii) risk of integrative analysis, and (iii) risk of predictive analytics.

## 2 RISK OF BREACH OF PRIVACY AND CONFIDENTIALITY

Privacy and freedom constitute individual rights in traditional biomedical ethics with measures like "informed consent" of patients supposed to prevent unnecessary harm. Even with traditional research, the

question of whether informed consent is carried out well, and whether it hinders or fosters research, remains unanswered.(3–5) Now, in instances where newer data sources, such as social network datasets, are being merged and freely distributed, the question of informed consent must extend to: where exactly does the “individual” end and the “social” begin?

Various stakeholders – government, industry, and academia among others – may use big data for specific purposes with the best interests of the population in mind but possibly at the expense of the individual. For example, educational data about individual students could be used to preferentially admit students who are “predicted” to perform better in the future, as opposed to providing a wholesome environment for all students to fulfill their respective potential, and thereby, risking the creation of a class divide by design. Even if one might assume that a prediction is correct at the time it is made, still that should not present an excuse to prevent any individual to strive and shape one’s own destiny.

Such risks are well known in healthcare, where insurance companies may use data about populations to prevent access to healthcare to, or even penalize, individuals based on larger group or population characteristics regardless of whether a particular individual conforms to the group norms or not.(6–8)

### **3 RISK OF INTEGRATIVE ANALYSIS**

Free and transparent discussions of organizational and governmental interests with regards to individual privacy, freedom, reputation or ownership of data are absolutely vital toward preserving integrity of the analytical process. Concerns about data ownership, legacy, sharing and usage require collective thinking, and should lead to conscious efforts in policy-making. This is because in the digital era, data never die. The World Economic Forum has recently defined personal data as a ‘new economic asset’ in their work on ‘Re-thinking personal data’.(8)

In this context, big data is a real game-changer. Personal data obtained through social media can open up issues of veracity and privacy when it comes to their analysis. Worse still, given the myriad different sources of information – some public and some not quite so, and many of which might have originally been there for purposes that are very different from

the intended analytics – all of these now have the potential to be combined or triangulated to extract unanticipated insights (say, into an individual’s current socio-economic or health status).

While, real world information about the individual may not be readily available, an approximation based on individual interests, their social circle, employment, education may very well come close. For example, an analyst may be able to determine an individual’s age, habits and preferences, financial worth, and more, based on one’s social media interactions, peer networks and details such as location or marital status. While data volume can help improve the accuracy of the automated learning algorithms, concerns regarding the privacy of individual information will remain a moot question for a considerable amount of time into the future.

### **4 RISK OF PREDICTIVE ANALYTICS**

Big data analytics in healthcare uses predictive algorithms to forecast health events in real time. From improving health of patients to lowering the cost of care to increasing accessibility to services to enabling precision medicine – all constitute merits of analytics in healthcare. Patient engagement outside the healthcare system can be harnessed for big data analytics to identify support networks, develop key messaging, improve clinical trial recruitment and retention among others. While for some types of data, patient consent is sought, for data that exist in the public domain, such consent is often not deemed necessary.(5,8–10)

To consider another example of predictive analytics, PREDPOL is a predictive policing program that uses just 3 pieces of information – past type of crime, place of crime and time of crime – to make its predictions about crime events. PREDPOL tries to help law enforcement gather information and make inference about a potential event’s location and time – one which has not yet taken place – supposedly with the aim of preventing this future crime.(11)

Knowledge can serve as a double-edged sword. The promise and lure of big data to reveal patterns and new knowledge from unexamined troves of data make application of traditional ethics and law enforcement challenging and retrospective. Most organizations and governments understand the dynamic nature of data in response to changing environments.

Yet, ethical questions and concerns seem to emerge in a retrospective manner and those responsible to address such issues are generally found to be "looking over the shoulder" in the context of some violation or disruption as opposed to being proactive and seeking adequate protections well in advance. (1,12)

## 5 CONCLUSION

Big data is ultimately about money and power that result from knowledge and insights. Omniscience and omnipotence are two sides of the same coin. Ethics and best practices have key roles to play in maintaining a balance between the quest for omniscience and the rights of every individual who contributes to or is affected by decisions based on such knowledge. Individuals should indeed have maximal ability to manage the flow of their own information across third party systems. The enforcement of rules for transparency in data acquisition, provenance, retention, usage, and sale should rather be the norm than the exception. Open and lively participation in ethical dialogues must continue on determining ways to live efficiently and safely in an increasingly data-driven world while being on the guard against the ever more innovative threats to personal information.

## REFERENCES

- [1] Metcalf J. Ethics codes: History, Context, and Challenges. 2014. Available from: <http://www.datasociety.net/>
- [2] Adali S, Escrivá R, Goldberg MK, Hayvanovych M, Magdon-Ismaíl M, Szymanski BK, et al. Measuring Behavioral Trust in Social Networks. In: Intelligence and Security Informatics (ISI), 2010 IEEE International Conference on IEEE, 2010. 2010. p. 150–2.
- [3] Grady C. Do IRBs Protect Human Research Participants? JAMA. 2010;304(10):1122. Available from: <http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2010.1304>
- [4] Steinbrook R. Improving Protection for Research Subjects. N Engl J Med. 2002;346(18):1425–30. Available from: <http://www.nejm.org/doi/abs/10.1056/NEJM200205023461828>
- [5] Metcalf J, Crawford K. Where are Human Subjects in Big Data Research? The Emerging Ethics Divide. Big Data Soc. 2016;3(1):2053951716650211.
- [6] The Opportunities and Ethics of Big Data. Available from:

<http://www.rss.org.uk/Images/PDF/influencing-change/2016/rss-report-ops-and-ethics-of-big-data-feb-2016.pdf>

- [7] Hermon R, Williams PAH. Big data in healthcare: What is it used for? Available from: <http://ro.ecu.edu.au/aeis/22>
- [8] Rethinking Personal Data: Strengthening Trust. 2012. Available from: [http://www3.weforum.org/docs/WEF\\_IT\\_RethinkingPersonalData\\_Report\\_2012.pdf](http://www3.weforum.org/docs/WEF_IT_RethinkingPersonalData_Report_2012.pdf)
- [9] Narayanan A, Huey J, Felten EW. A Precautionary Approach to Big Data Privacy. 2016. p. 357–85. Available from: [http://link.springer.com/10.1007/978-94-017-7376-8\\_13](http://link.springer.com/10.1007/978-94-017-7376-8_13)
- [10] Boyd D. Networked Privacy. New York, New York, USA: Personal Democracy Forum; 2011. Available from: papers3://publication/uuid/36B968CE-B823-4D0B-849A-2BDFE55A253D
- [11] PREDPOL. Available from: <http://www.predpol.com/about/>
- [12] Big Data Ethics: 8 Key Facts To Ponder - InformationWeek. Available from: [http://www.informationweek.com/big-data/big-data-analytics/big-data-ethics-8-key-facts-to-ponder/d/d-id/1322143?image\\_number=3](http://www.informationweek.com/big-data/big-data-analytics/big-data-ethics-8-key-facts-to-ponder/d/d-id/1322143?image_number=3)



**Dr. Meghana V. Aruru, PhD, MBA**, is Associate Professor in the Indian Institute of Public Health, Hyderabad. Her research interests include Health Economics and Outcomes Research, Health Policy and Ethics, Health Promotion and Communication, and Translational Science. She received her PhD from the University of Illinois at Chicago, USA.