

4 things to keep in mind about big data

Privacy and children's rights

The main criticism about big data and analysis dealing with such content is the issue of privacy. Much of the data generated is produced passively as a by-product of browsing, social media or other online activities. Therefore, people might not be aware of that private companies hold administrative data, purchase history, search hits and location data in their repositories. The problem with such information is that it can easily go into wrong heads and violate our right to privacy and personal freedom. In the case of children, it is even more important that governments ensure safeguarding processes by installing policy checks and encouraging transparency on how big data is collected, stored and analysed by technology firms. And while big data offers many new approaches and benefits to child protection, it is important to keep in mind that data collected from children should be protected and anonymised according to privacy rights. A paper produced by the UN raises concerns about uncertainty of future use, sale of data for market research and the lack of control over digital identities as the main issues for data ethics and children's rights. They recommend ensuring that children and young adults have the means to communicate their opinions about these matters and are treated as equal partners in the discussion about privacy rights and data sharing. One step towards data protection and children's rights has been the [new data rules passed by the EU](#) require parental consent for data collection and usage by companies among children under the age of 16, which member countries can lower to 13 if decided. However, even with regulations, it is essential that companies exhibit tighter control over third-party organisations when selling raw data and that children and parents are aware of the existence of such data in the first place.

Lack of context

Data gathered from online sources is often taken out of context and reduced to mere text for analytical purposes which presents both risks and challenges. First, without the context in which it was written, it is difficult to uncover the true meaning of the words or what the individual intended to communicate. For example, it can be hard to separate perceptions from facts, indicated best by the Google Flu study. Originally aimed at predicting flu rates, the software wrongly picked up on text that described the same symptoms but was indeed related to a common cold. It could have resulted in doctors' making decisions about diagnosis or stocks of medication in light of the wrongly predicted rates of flu. Accents and slang also pose barriers to text based analysis. Furthermore, it can often be hard to verify the validity of online information and distinguish between real opinions or facts and false information. Second, removing data from its original context can lead to discrimination on grounds of sexual orientation, health conditions or other private matters. Similarly, companies or governments can use data gathered through social media and the internet to identify individuals as high-risk and remove them from the platform, thus discriminating on the basis of their online presence.

Data and biases

Much of the criticism regarding algorithms that predict future outcomes based on identified risk factors such as domestic violence in the home or substance misuse, have centred around potential biases. Critics have argued that computer analyses might pick up on minority communities more often as they tend to possess more of the identified risk factors on average. In analytics using maps or location based identification of risk factors, more disadvantaged areas can be selected by the system which can lead to unnecessary intervention and removal of children. It is therefore important, that systematic analysis is

combined with case assessment by social workers or managers. One expert commenting on the family screening tool in Pennsylvania, has emphasized the importance of transparency for data based predictions and that the aim of the systems should be to minimise existing biases by systematic analysis of the information.

Unequal availability of digital technologies

While big data and data analysis in general has grown rapidly in recent years in most parts of the world, there are still large segments - mainly in developing countries - where citizens do not own mobile phone or have access to the internet. It is therefore important to bear in mind that data gathered through such mediums could not be representative of the overall population in many cases. While research using big data has many benefits, especially when combined with development goals, research and policy need to ensure that they are not excluding segments of the population and their respective needs.

Sources

UN Global Pulse (2012) Big Data for Development

<http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseMay2012.pdf>

UNICEF (2017) Children and the Data Cycle: Rights and Ethics in a Big Data World

https://www.unicef-irc.org/publications/pdf/IWP_2017_05.pdf

The Guardian (2015) Why we shouldn't get too excited about using big data for development

<https://www.theguardian.com/global-development-professionals-network/2015/jun/22/why-we-shouldnt-get-too-excited-about-using-big-data-for-development>

PBS (2016) Can big data save these children? <https://www.pbs.org/newshour/nation/can-big-data-save-these-children>

The Guardian (2018) What children services teams need to know about new EU data rules

<https://www.theguardian.com/social-care-network/2018/jan/25/childrens-services-eu-data-rules-gdpr>

Written by Fanni Kovcsdi, 2018.