[Big Data for Development: From Information to Knowledge Societies](https://irevolutions.org/2013/02/04/big-data-for-development-2/) by Martin Hilbert, 2013

Hilbert uses a three-dimensional conceptual framework to describe the opportunities and threats of Big Data used in international development. According to [Castells](https://deterritorialinvestigations.files.wordpress.com/2015/03/manuel_castells_the_rise_of_the_network_societybookfi-org.pdf), ‘the ability to cope with the uncertainty caused by the fast pace of change in the economic, institutional and technological environment’ is a key goal in the Big Data age (p. 165). Since each decision has a certain probability of being wrong, information can be used to make estimates if our prior knowledge is improved. This component of information theory leads to what Hilbert describes as the Big Data (BD) paradigm.

Hilbert describes the BD paradigm as a phenomenon that delivers a cost-effective way to improve decision-making in domains such as health care, natural disaster management and economic productivity in the developing world. He states the key to the BD paradigm is not the quantity of data but ‘its analysis for intelligent decision-making.’ This paradigm creates a framework where different components of BD intersect.

The main point of Hilbert’s research is describing successful digitization as having three dimensions: technology, social change and policy (p. 5). Prerequisites to making Big Data work for development are appropriate policies and a robust infrastructure, software services and human skills to interpret the data. Hilbert develops a persuasive argument that the 3D framework is important in using Big Data for development. His evidence is examples in data intensive sectors such as health care, natural disaster management and economics. He predicts BD will have a similar positive effect on efficiency as Information and Communication Technology (i.e. - the internet and mobile phones) had in the 1990s and 2000s.

Three development sectors that are benefitting from BD are health, natural disaster management and economics. Twitter data revealed important health information about how cholera was spreading in 2010 in Haiti and was available up to two weeks earlier than official statistics ([Chunara](http://www.icahn.org/files/Emergency_Preparedness_Library/Haiti-Cholera.pdf)). Another example of BD use in the health sector is Global Viral Forecasting as described by [Wolfe](http://www.economist.com/node/10717931), which tracks pandemics by using text data along with other data sources. The potential for BD to track and even prevent disease is just in its infant stages in the developing world.

In addition to the health sector, BD can be used for natural disaster management. Nature is one of the biggest sources of uncertainty in the developing world. Data collection and analysis can reduce this uncertainty which will lead to big impacts. One example is when the United Nations University used weather and climate data to analyze where rain would fall to improve food security in the [Where the Rain Falls](http://unu.edu/publications/articles/how-rainfall-variability-food-security-and-migration-interact.html) project. Another example is the [Open Data for Resilience Initiative](https://opendri.org/) which uses climate and local government data to decide where and how to build safer schools, how to protect farmers from drought, and how to minimize the impact of natural disasters on coastal cities.

Hilbert describes the economic sector as benefitting from Big Data. Natural resource extraction is a big source of income in many developing countries (p 13). [Sanchez](http://transparencyinitiative.theideabureau.netdna-cdn.com/wp-content/uploads/2012/12/TAIBrazil1.pdf) argues that data publication of the natural resource extraction industry could make the industry better without hurting the economic competitiveness. He says with the publication of data came increased transparency in Brazil. Hilbert uses this example to persuade about how important the 3D framework is development.

Hilbert describes three dimensions of the BD framework as evidence for his argument: technology (p. 15-20), social change (20) and policy. The technology dimension includes infrastructure, software and technical skills to process BD. Hilbert states there is a need to create these capacities in developing countries. Hilbert gives examples of modular and decentralized approach to solving infrastructure problems and building technical BD skills. Regarding the social change component, developing countries don’t have as much awareness about the importance of analytical capabilities. He cites [Noormohammad](https://pdfs.semanticscholar.org/d4a6/f1126ffcf7d464554bc4e5f26939db94ab98.pdf) as an example of how adequate training prevents program failure due to social views.

Making Big Data work for development requires carefully designed policies (5). Public and private policies can be categorized as positive feedback and negative feedback. An example of positive feedback would be initiatives to encourage governments to open more data portals (p 24). The argument Hilbert makes is persuasive because just because the data is freely available in a portal does not mean it’s being used to make improvements. Regulation around who owns the data, how it is collected and maintaining privacy would be an example of necessary negative feedback policies.

Hilbert presents two main critiques to BD in development. The first critique is that too much power is given to algorithms. Should automated machine decision-making be encouraged? Should all power be given to the algorithms? (p 29). By definition algorithms only can do processes that are programmed into them by a human. Algorithms are imperfect since they ‘are mainly informed by the world as it was or, at best, as it is’ (p 30). The assumptions behind algorithms will only be accurate if future events are similar to past events. BD can give us correlations (the what) between events but not the causation or the why.

The second critique is that since algorithms are based on information from the past, they reinforce past behavior, which may have unintentional biases. According to [Rissanen,](http://www.springer.com/gp/book/9780387366104) the objective should be to extract all meaningful and useful information from the data to give the most accurate predictions possible since perceiving reality is challenging. [Garland](http://www.theatlantic.com/international/archive/2012/04/peak-intel-how-so-called-strategic-intelligence-actually-makes-us-dumber/255413/) argues that making decisions based on past data patterns makes it harder to adapt to a changing world.

Like [Kshetri](http://bds.sagepub.com/content/1/2/2053951714564227), Hilbert argues that there is a digital divide, or a gap between developing countries and developed countries with respect to their digital capabilities. If a developing country doesn’t have the technological, social and policy mechanisms to support digitization, a Big Data ecosystem will be slower to be adopted. Hilbert says that ‘data itself becomes a commodity’ and is subject to existing economic divides in the developing world (p. 19). There is an unevenness in the ability to get and use data for decision making.

Hilbert’s work is significant for data science since it shows that the Big Data paradigm can be used for the development if certain barriers are addressed. [Rogers](http://www.indiana.edu/~t581qual/Assignments/Diffusion_of_Innovations.pdf) thinks that Big Data is a technology innovation that creates this digital divide as knowledge is diffused. Although Big Data won’t solve all development problems, it has a promising future if approached with careful thought. In conclusion, I agree with [Bateson](http://www.edtechpost.ca/readings/Gregory%20Bateson%20-%20Ecology%20of%20Mind.pdf) who says ‘information is a difference which makes a difference.’ There are a countless number of data that can lead to transformative insights and actions in the developing world.