THE DATA REVOLUTION: BIG DATA, OPEN DATA, DATA INFRASTRUCTURES AND THEIR CONSEQUENCES. By Rob Kitchin, London: Sage, 2014.

Can we discuss the data revolution without touching the data? Starting from the mid-1980s, experimental natural science, financial institutes and enterprises have been generating huge arrays of data. For the first two decades, Western society was either excited with the constantly growing collections of satellite images and weather observations that make the earth a better place for the human race, or just tolerated banks and companies' databases on goods transactions and customers, as the inevitable next step in the technological development.

The explosion of social networking in the 2000s is changing our attitude to the mounting information on everything and everywhere. Hardware was never a problem, and from Gigabytes (2^{30}) and Terabytes (2^{40}) we are advancing to Petabytes (2^{50}) and Exabytes (2^{60}) having in mind Zettabytes (2^{80}) and Yottabytes (2^{90}). Street video surveillance, smartphones, internet search engines and social networks capture and instantaneously process information on our habits and preferences. Google, Facebook, Whatsapp, Twitter are convincing us that it is all for the good of humanity in general, and each of us personally. A part of me is excited, but the other part is shocked by the idea that the information I disclose voluntarily, may work against me.

Society enters a new era of DATA and social science dares to take responsibility for the necessary arrangements. Rob Kitchin is on the frontline. Not all attempts of this kind are successful - "much talk of big data is a big hype…" says David Lyon on the back cover of the book and continues that Kitchin did succeed in offering "… a pointer to the crucial social, political and ethical issues…" of a data revolution. The focus on the social aspects of the data revolution has its drawbacks – Kitchin is quite indifferent to the new knowledge that the big data brings to society.

The book consists of eleven short chapters. Chapter 1 – "Conceptualizing Data" is quite technical. It defines basic notions of a "data science" – data types, databases, tools for data processing, and continues with the ethical, political and philosophical views of the recent data revolution. Chapter 2 – "Small Data, Data Infrastructure and Data Brokers" continues this line and formulates operational criteria for recognizing when the data are yet small. The chapter describes the development of the data infrastructure and the formation of multi-billion dollar data market and provides examples of the data that people are ready to pay for. Chapter 3 – "Open and Linked Data" presents to the reader the major dilemma of data providers and data users. A provider is always unsure whether to sell the data or services that are based on the data, while the users are always unsure whether to purchase cleaned and reliable data and services produced by commercial supplier or to save their budget and stick to open data sources. The open data have their own problems and the chapter exposes them.

The three introductory chapters inspire the reader's curiosity, but Chapter 4 – "Big Data" is not that convincing and has a flavor of scholasticism. It explores the ontology of big data, and discusses in depth their basic characteristics - volume, velocity of accumulation, variety of features, spatial and temporal resolution, scalability, etc. Chapter 5 – "Enablers and Sources of Big Data" presents the sources of the big data: directed surveillance, automated data production by sensors of different kinds; volunteered data generation; social media; crowdsourcing and citizen science. Chapter 6 – "Data Analytics" starts with a bold claim that "Data … only have utility if meaning and value can be extracted from them", and promises to tell us about new analytics that include artificial intelligence, expert systems, machine learning and predictive modeling. However, the author limits himself to taxation of methods into four categories: Data mining and pattern recognition; data visualization; statistical analysis; and prediction, simulation, and optimization. This may be of interest to those who know a number of methods representing each group, but I doubt this is the audience of the book. I especially liked Chapter 7 – "The Governmental and Business Rationale for Big Data" that lively presents lines of discourse caused by the big data in the society, business community and government. Chapter 8 – "The Reframing of Science, Social Science and Humanities Research" discusses, at a general level, the popular topic of 'the end of theory', and data-driven versus knowledge-driven science with application to digital humanities and computational social sciences. Chapter 9 – "Technical and Organizational Issues" presents concerns of the big data manager - datasets, data quality, metadata, data integration, access limitations, interoperability, necessary skills and organizational capabilities. Chapter 10 – "Ethical, Political, Social and Legal Concerns" examines the issues of privacy, data security, profiling, social sorting and links these aspects to the public policy via anticipatory governance, technocratic and corporate governance data ownership and intellectual property. The major claim of the author is that any solution consists of compromises. The book is concluded by Chapter 11 – "Making Sense of the Data Revolution" that proposes a conceptual view of the data revolution.

I finished reading the book with mixed feelings. Does all this big data hype bring any new knowledge about society? Is this just another big problem to add to the already existing pile of big problems of humanity? Do the big social data bring any new insight on the behavior and decision making of individuals' groups and the entire society? Kitchin's book guides the reader through the complexity of contemporary views on the data revolution and makes us aware of its dangerous sides. To the other side of a coin – the positive one – we will have to wait for another book.

Itzhak Benenson
Tel Aviv University