PREFACE

There is a long history of governments, businesses, science and citizens producing and utilising data in order to monitor, regulate, profit from, and make sense of the world. Data have traditionally been time-consuming and costly to generate, analyse and interpret, and generally provided static, often coarse, snapshots of phenomena. Given their relative paucity, good-quality data were a valuable commodity, either jealously guarded or expensively traded. Recently, this state of affairs has started to change quite radically. Data have lost none of their value, but in other respects their production and nature is being transformed through a set of what Christensen (1997) terms disruptive innovations that challenge the status quo as to how data are produced, managed, analysed, stored and utilised. Rather than being scarce and limited in access, the production of data is increasingly becoming a deluge; a wide, deep torrent of timely, varied, resolute and relational data that are relatively low in cost and, outside of business, increasingly open and accessible. A data revolution is underway, one that is already reshaping how knowledge is produced, business conducted, and governance enacted.

This revolution is founded on the latest wave of information and communication technologies (ICTs), such as the plethora of digital devices encountered in homes, workplaces and public spaces; mobile, distributed and cloud computing; social media; and the internet of things (internetworked sensors and devices). These new technical media and platforms are leading to ever more aspects of everyday life – work, consumption, travel, communication, leisure – and the worlds we inhabit to be captured as data and mediated through data-driven technologies. Moreover, they are materially and discursively reconfiguring the production, circulation and interpretation of data, producing what has been termed 'big data' – vast quantities of dynamic, varied digital data that are easily conjoined, shared and distributed across ICT networks, and analysed by a new generation of data analytics designed to cope with data abundance as opposed to data scarcity. The scale of the emerging data deluge is illustrated by the claim that '[b]etween the dawn of civilisation and 2003, we only created five exabytes of information; now we're creating that amount every two days' (Hal Varian, chief economist with Google, cited in Smolan and Erwitt 2012).

Big data are not the only components of the data revolution. Rather, there are related initiatives such as the digitisation, linking together, and scaling-up of traditionally produced datasets (small data) into networked data infrastructures; the open data movement that seeks to make as much data as possible openly available for all to use; and new institutional structures that seek to secure common

guidelines and policies with respect to data formats, structures, standards, meta-data, intellectual property rights, licensing and sharing protocols. Together, these constitute a set of new data assemblages – amalgams of systems of thought, forms of knowledge, finance, political economies, governmentalities and legalities, mat-erialities and infrastructures, practices, organisations and institutions, subjectivities and communities, places, and marketplaces – that frame how data are produced and to what ends they are employed.

The impact of big data, open data and data infrastructures is already visible in science, business, government and civil society. Used to operating in data deserts, seeking to extract information and draw conclusions from relatively small num-bers of observations, established disciplines are now starting to grapple with a data avalanche (H.J. Miller 2010). They are accompanied by new fields, such as data science, social computing, digital humanities, and computational social sci-ences, that are explicitly concerned with building data infrastructures and finding innovative ways to analyse and make sense of scaled and big data. In business, big data are providing a new means to dynamically and efficiently man-age all facets of a company's activities and to leverage additional profit through enhanced productivity, competitiveness, and market knowledge. And data them-selves have become an important commodity, actively bought and sold within a global, multi-billion dollar market. For governments, widespread, dynamic data are providing new insights about their own operations, as well as reshaping the means to govern and regulate society. Through examining open datasets, citizens and non-governmental organisations (NGOs) are drawing their own conclu-sions, challenging corporate and government agendas, and forwarding alternative visions of how society should be organised and managed.

These new opportunities have sparked a veritable boom in what might be termed 'data boosterism'; rallying calls as to the benefits and prospects of big, open and scaled small data, some of it justified, some pure hype and buzz. In turn, the terms big data and open data have become powerful memes, not just a way of describing data but symbolic of a wider rhetoric and imaginary that is used to garner support and spread their roll-out and adoption. Such boosterism and memes can make it easy to drift into uncritically hyping the changes taking place, many of which raise numerous ethical, political and legal concerns. History, though, does reveal earlier precedents of disruptive information-related innovations – the radical transformation of knowledge production in the wake of the printing press, for example. Indeed, every new era of science has had at its inception new technologies that lead to an information overload and spark a transition to new ways of generating, organising, storing, analysing and interpreting data (Darnton 2000). For example, Strasser (2012) notes, the explorations of the Renaissance, enabled by better navigation, mapping and scientific instruments, yielded vast quantities of new discoveries that led to new methods of categorisa-tion, new technologies of analysis and storage, and new scientific insights.

Given the relatively early point in the present data revolution, it is not at all certain how the present transformations will unfold and settle, and what will be the broader consequences of changes taking place. What is clear is that there is an urgent need to try and make sense of what is happening. Thus, the aim of this book is to provide a synoptic, conceptual and critical analysis of data and the data revolution underway. It seeks, on the one hand, to chart the various ways in which the generation, processing, analysis and sharing of data is being reconfigured, and what this means for how we produce and use information and knowledge; and, on the other, to open up debate and critical reflection about data: their nature, how they are framed technically, philosophically, ethically and economically, and the technological and institutional assemblages that surround them. Rather than setting out a passionate case for the benefits of big data, open data and data infrastructures, or an entrenched critique decrying their more negative consequences, the book provides a contextual, critical appraisal of the changes taking place.

The analysis presented is based on an extensive engagement with the literature from across humanities, social sciences and the sciences, and from popular culture, journalism, and industry publications, and on first-hand experience of working on large-scale data archiving/infrastructure and data analytics projects. The book is divided into eleven chapters. The first provides an overview and critical reflection on the concept of data and how to make sense of databases and data infrastructures. The second examines the continued role of small data and how they are being scaled up into digital archives and infrastructures, and sold through data brokers. Chapter 3 discusses the drive towards creating open and linked data that are more widely shared and reused. Chapters 4 and 5 detail the nature of big data and its enablers and sources. Chapter 6 provides an overview of a new set of data analytics designed to make sense of scaled small data and big data. The next two chapters examine the arguments used to promote big data and their impact on governance and business, and the ways in which the data revolution is reshaping how research is conceptualised and practised. Chapters 9 and 10 discuss the technical, organisational, ethical, political and legal challenges of the data revolution. The final chapter sets out some overarching conclusions and provides a road map for further research and reflection.