CHAPTER 1.6

Rebalancing Socioeconomic Asymmetry in a Data-Driven Economy

PETER HAYNES, Atlantic Council
M-H. CAROLYN NGUYEN, Microsoft

It is more than half a century since economist Fritz Machlup, in his book *The Production and Distribution of Knowledge in the United States,*[1] developed the concept of the *knowledge economy,* a term later popularized by management theorist Peter Drucker.[2] Both used the phrase to differentiate between an economy based largely on goods and services produced by manual labor and an economy based on the production of new ideas, information, and knowledge.

The wide availability of low-cost information and communication technologies (ICTs)—which started in the early 1980s but gained real traction in the mid-1990s with the advent of the public Internet—drove the economic transformation that Machlup and Drucker envisioned. Research by the Organisation for Economic Co-operation and Development (OECD) shows how rapidly the combination of personal computers, digital telecommunication devices, and the Internet impacted economic growth in the early years of that troika. A 2004 OECD paper estimates that inflation-adjusted investment in ICTs accounted for an average of 0.5 percentage points of annual growth in real GDP in OECD countries between 1995 and 2001.[3] This represents about 20 percent of total growth in real GDP—a significant impact. And a 2002 OECD study demonstrated that corporate investment in networked computer systems is consistently associated with increased labor productivity—for example, productivity was found to be 11 percent higher in US manufacturing plants that have computer networks.[4] It took about 80 years for steam engines to increase labor productivity by approximately that amount, about 40 years for electricity, and more than 20 years for pre-Internet ICTs.[5]

Yet it could be argued that what we saw in the years from the early to mid-1990s was less the emergence of a *knowledge* economy than an *information* economy. Early Internet-era ICTs enabled more efficient and effective processing and use of data, resulting in information that was, for the most part, used to improve the performance of *existing* processes, businesses, and industries. Although in some cases that information transformed businesses completely (e.g., Amazon), these are still traditional businesses. To describe what was created during this period as a *knowledge* economy would be an exaggeration. Much information has been generated from the raw data collection made possible by advances in ICTs, and significant new knowledge has been created, but neither is yet the *foundation* on which the global economy is built. That honor still belongs to physical goods and services.

The recent emergence of big data, along with what is being called the "data-driven economy," may finally make possible a *true* knowledge economy—by which we mean entirely new classes of economic activity predicated on insights and value derived from analyzing, contextualizing, and commingling vast datasets in ways that were previously either unknown or impossible. For companies (and entire economies) to keep growing, the development of such an economy is an imperative: battered by global competition, commoditization, and shrinking product/service cycles, corporations seeking to maintain or grow their profit margins will increasingly rely on creating new (and hard-to-emulate) products and services based on insights derived from the datasets that they own or can gain access to, especially those pertaining to their customers.

Two decades after the emergence of the consumer Internet, the world is awash in data. By one estimate, almost 3 zettabytes (3 billion terabytes) of information had been created by 2012, a digital deluge that is growing at around 50 percent a year.[6] By the end of 2013, the number of wirelessly connected devices, at an estimated 8 billion, will have exceeded the 7.2 billion people on the planet.[7] By 2020, as many as 50 billion devices could be wirelessly connected to the Internet,[8] while the world population is forecast to be fewer than 8 billion people.[9] At the same time, from 2012 to 2017 machine-to-machine data traffic is set to grow an estimated 24 times, to reach 6×10^{17} bytes per month—an astonishing compound annual growth rate of 89 percent.[10] Indeed, the majority of big data will be collected passively and automatically, via machine-to-machine transactions, and users will not be actively involved in the majority of those transactions.

Big data, analytics, and machine learning promise new solutions to previously intractable problems (e.g., in healthcare, disaster response, the environment, and transportation); new businesses will be able to create innovative services by selecting, combining, and parsing data in groundbreaking ways; and individuals will be empowered because they will be able to draw on a wide range of yet-to-be-invented data-based services and tools to improve the quality of their lives. Big data, then, truly does promise to create new knowledge—and indeed new *kinds* of knowledge—on which an entirely new economy can be founded.

However, the knowledge economy relies on the availability of an adequate supply of data to enable the discovery of new knowledge. This requires policy frameworks that permit data—including personal data—to be collected, analyzed, and exchanged freely, across geopolitical boundaries, while minimizing risks and harms to individuals and enterprises globally. Existing regulatory approaches that are based on the principles of notice and consent to restrict the collection of data pre-designated as *personal* may overly restrict the supply of data available, hampering the foundation for the new economy. Furthermore, what is considered *personal* and *acceptable use* are individual decisions, subject to context, perceived value, and social and cultural norms—all of which are in a constant state of flux.

In reality, it is not the *collection* of data that is the source of potential harm, but its unconstrained use. Moreover, in the world of big data, it would be impractical, if not impossible, for individuals to give express consent for all the data that may be generated about them.

Together, the above factors necessitate a change in policy approach from a collection-based model toward a use-based model, where individuals give permission for the *use* of data related to them.

What is increasingly clear about an economy based on the collection, use, and analytics of big data is how little we actually know about it—its potential risks and rewards, as well as its implications for individuals, organizations, policy development, and growth. The rest of this chapter will focus on some core challenges that the authors believe could be particularly problematic—and that may threaten to impede the development of a promising 21st-century knowledge economy.

THE DEMISE OF FAIR VALUE EXCHANGE

All previous economic revolutions have been based on the idea of an explicit (i.e., transparent) fair value exchange. For example, in return for $850, early 20th-century consumers could obtain a 1909 Ford Model T; $1,565 bought a base-model 1981 IBM PC; and today, a hardback edition of Adam Smith's *The Wealth of Nations* is available for a price of about $17. The costs and benefits to those on both sides of this value equation (usually an individual and a corporation) are both clear and easily discoverable. The process by which the transaction is executed is well established in modern economics: rational, self-interested economic actors determine the price they are willing to pay for a good or service based on their subjective perception of its utility—something that is usually quite simple for them to determine. Assuming that the market in which the economic actors are engaging is not subject to monopolistic or other distortions, prices tend to settle at the point where supply equals demand.

Research sponsored by Microsoft and published last year by the International Institute of Communications found that, among other things, users *do* consider fair value exchange in allowing the use of their data.[11] They have some expectation of what they will receive in return—for example, discounts, better service, an improved product, or potential benefits to the larger community in which they live.

Retailer loyalty cards are an example of this type of transaction, and they also illustrate one of the challenges of the data-driven economy. Most consumers understand that the discounts they receive via a loyalty card are provided in exchange for data they supply to the retailer. But very few realize that the *primary* value to the retailer is the ability to track and analyze the spending patterns of both individuals and aggregated datasets of groups of consumers. In other words, significant information asymmetries are embedded in the transaction, and the average consumer lacks all the information required to make a rational decision about whether he or she should participate in it.

As the global economy becomes increasingly grounded in the exchange of data, the ways in which those data are collected and analyzed will become even more opaque to the consumer and the value exchange even harder to discern; trust will decrease correspondingly. An individual may have only a vague idea of what data exist about him or her and what is being done with these data. Some will have been actively volunteered by the consumer; some will have been obtained passively, with or without his or her explicit knowledge; and yet more may have been inferred by commingling a range of public and private, personal and non-personal datasets in ways that might expose new information or knowledge about each consumer's habits, lifestyle, health, or financial situation. Although the individual may receive something in return for this information, the real values of both the data provided and the service returned (in other words, the underlying exchange of value) may be almost impossible to determine.

Today little agreement exists about how best to value online data. The most comprehensive survey of valuation methodologies was presented in a recent OECD study (on which the authors of this chapter consulted) that identified numerous ways in which data *might* be valued in the market (refer to Box 1).[12] However, each of these methods has significant flaws, and none addresses the potential social and economic benefits of personal data. For example, corporate revenues per record/user are problematic because revenues contribute to economic growth only insofar as they generate added value (or surplus). Revenue of $4 per record/user with near-zero profitability is very different from $4 per record/user with 40 percent net profit. Similarly, the vast amount of personal data on Facebook have a relatively low per-person value because the company, while making significant profits from the sheer scale of its data holdings, has yet to find the Holy Grail of social-media data monetization. Amazon, by contrast, collects far less personal information from individuals, but its business model is predicated on advanced purchase analytics. Thus, on a per-user level, its inferred personal data (which are at present mostly outside the user's control) are more valuable than Facebook's volunteered personal data (which the user has painstakingly assembled, and over which she or he has at least nominal control).

Distinguishing personally beneficial uses of data from socially beneficial uses is a further challenge because each creates separate and significant value. For example, the *personal* value of using an electronic health record is improved treatment for the patient—and this undoubtedly has direct monetary value in the form of reduced costs, better outcomes, and so on. But *socially beneficial* uses also create (or could create) value—for example, by facilitating research into new drugs, identifying new epidemiological trends, or improving medical protocols. However, because the value created does not involve explicit market transactions, attributing this benefit directly to data involves some inspired approximation. And even though one estimate puts the savings in this case at up to US$300 billion,[13] most of the ways in which data are valued today would consider such benefits an externality to be ignored.

For many people, however, the various ways in which data *might* be valued are largely irrelevant today, because they have already given away their digital crown jewels for free. Individuals are passing massive amounts of personal and other data to large corporations with little or no thought to its potential monetary value—and those corporations are making significant profits as a result, because their cost of materials is essentially zero. The concept of fair value exchange no longer exists, at least not in any conventional sense. Facebook users, for example, provide it with data that have the potential to generate immense long-term value for the company; in return they receive a "free" service, but the transaction is wholly asymmetrical. As the computer scientist Jaron Lanier has observed, "[T]he dominant principle of the new economy, the information economy, has lately been to conceal the value of information.… We've decided not to pay most people for performing the new roles that are valuable in relation to the latest technologies. Ordinary people 'share,' while elite network presences generate unprecedented fortunes." And if an individual's information is not valued in economic terms, Lanier adds, "a massive disenfranchisement will take place."[14]

Box 1: Potential approaches for estimating the value of personal data

The following methods for valuing personal data have been identified, but each has important drawbacks. Possible approaches include:

- determining the market capitalizations of firms with business models predicated on personal data;
- ascertaining the revenues or net income per data record;
- establishing the market prices at which personal data are offered or sold;
- establishing the economic cost of a data breach;
- determining prices for personal data in illegal markets;
- reviewing economic experiments and surveys that attempt to establish the price companies would need to pay for individuals to give up some of their personal information; and
- ascertaining how much individuals would be willing to pay to protect their data.

Source: OECD 2013.

In other words, under the current model, the greater the role that data play in the global economy, the less the majority of individuals will be worth. This could mean that a data-driven economy may become a contracting economy. Like Lanier, we believe that if a truly sustainable data-driven economy is to be established, the way in which data are traded between individuals and corporations will require a major reset. For a data-driven economy to thrive, individuals would have to receive fair/appropriate monetary compensation for each specific datum they provide, perhaps with additional payments whenever that datum produces incremental profits for the entity to which it has been given (a concept popularized by Lanier). Such an arrangement would be complex: a specific datum might gain value only when commingled with other data, for example, and any payment/micropayment system would have to be capable of keeping track of such complexities (assuming the individual has given permission for this to happen). And a sustainable data-driven economy might also entail individuals paying fees (likely modest) for services they now consider (erroneously) to be "free."

Such systems, or similar approaches that address these concerns, will be essential to establish the concept of fair value exchange in the world of big data. The importance to our economic future—to the entire concept of a data-driven economy—of undergoing this evolution cannot be overstated. Without it, the consumers who today are the engine of economic growth will increasingly lose their ability to participate in the economy. Without fair value exchange for data along with inherent trust in the data ecosystem, everyone will ultimately lose—consumers, corporations, and countries alike. Establishing a system of fair value exchange will require new thinking on how technology *and* policy can work in parallel.

DESIGNING A TRUSTWORTHY—AND ECONOMICALLY VIABLE—DATA ECOSYSTEM

We believe that an essential element of the foundation that can enable user trust and fair value exchange is an interoperable metadata-based architecture. In such an architecture, data are logically accompanied by a "metadata tag" that contains references to the permissions and policies associated with the data, along with related provenance information, specified in an extensible and interoperable markup language. The metadata is logically *bound* to the data and cannot legally be unbound or modified for the entire data lifecycle by any parties other than the user or as specified by, for example, a related policy or rules of a "trust framework." More comprehensive consideration of these issues can be found in *Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: The Path Forward* and the *Digital Enlightenment Yearbook 2013*.[15]

But the use of metadata does not stop at enabling the enforcement of user permissions and related policies. It can also be utilized to track and capture the *monetary value* produced by personal data, over time, in a decentralized data ecosystem—and consequently provides a foundation for both trustworthy data *and* fair value exchange. Consider: metadata enables individuals to change their personal data preferences and permissions over time, prevent undesirable use of previously collected data, address unanticipated uses, and adjust to changing norms. Thus, if we consider personal data to be the product of an individual's online "labors," and if we further consider that, in order to introduce the concept of fair value exchange (and sustainability) into a data-driven economy, those data must be assigned monetary value, then metadata is the mechanism that will enable individuals to "direct" their labors and reap the related benefits for the duration of its existence in the data ecosystem—enabling a more enlightened society in the digital space. How such an approach would incorporate machine-to-machine generation and use of data remains unanswered, however, and requires considerably more research.

Such an approach is technologically non-trivial. A primary challenge is security: although metadata can be logically *bound* to data, it can also be *unbound* by bad actors (a situation similar to the vulnerability of today's financial systems to hackers). Thus a strong legal and policy framework will be required to ensure that criminals are discouraged from doing this (again analogous to the laws governing today's financial systems). Another challenge lies in specifying the user permissions and policies that would govern how data can be used within—and shared across—trust boundaries, and how those permissions and policies would be negotiated among the multiple parties with claims on the data or claims to its monetary value.[16] Yet another, highly significant challenge is developing the appropriate interfaces that will enable individuals to specify their permissions, either directly or through other means (such as recommender systems or data intermediaries).

Achieving all this will require the specification of an interoperable metadata-based architecture that can function at Internet scale. The development of such an architecture needs to be a collaboration between multiple data stakeholders to ensure its feasibility and inherent security, as well as its ability to enable alternative policy frameworks.

A metadata-based architecture offers value to all stakeholders in the data ecosystem, not only users. Data controllers and processors can more easily understand and comply with permissions and policies defined for specific data. They can also establish a dynamic, economically viable and sustainable "marketplace" in data that would ideally mirror the way in which fair value exchange is established in the physical world. Solution providers can create applications and services that produce new business value and track the associated value chain, yet still use data in privacy-preserving ways. Companies can develop metadata schemas that fully describe data use, codes of conduct, and relevant policies to meet industry and regulatory requirements. And regulators can take advantage of greatly improved auditability of data, along with a stronger and better-defined connection between the data and those policies that govern its use.

Although metadata can help facilitate a data-driven economy, it cannot *guarantee* that entities handling the data will honor the permissions and policies associated with them. However, when implemented as part of a principles-based policy framework that provides guidance on trustworthy data practices—supplemented by voluntary but enforceable codes of conduct and underpinned by legal redress—this is a flexible approach that holds the promise of satisfying the interests of regulators, individuals, and industry. In addition, as noted above, the authors believe that metadata could also be a key to establishing a viable and sustainable *economic* ecosystem in a data-driven economy, enabling the monetary value generated by data to be tracked, captured, and realized as payments to and from the ecosystem's participants.

CONCLUSION AND WAYS FORWARD

There are many challenges here, and today we have more questions than answers. But what is clear is that, in order to create a sustainable data-driven ecosystem, technology and policy must work symbiotically. For that to happen, governments and their regulatory representatives need to partner closely with industry, academic researchers, and consumer groups to gain a better understanding of the issues and to jointly develop innovative and evidence-based approaches to policy frameworks that address the above needs. Similar to evolution of new technologies, such approaches will need to be iterative.

Indeed, our view is that governments are the *only* entities with the ability to convene the broad societal coalition that will be required if the promise of a data-driven knowledge economy is to be fully realized. Such dialogues will, of course, need to be conducted on a global level.

NOTES

1 Machlup 1962.

2 Drucker 1969.

3 Ahmad et al. 2004.

4 Astrotic et al. 2002. Data are for value-added labor productivity.

5 Bughin and Manyika 2013.

6 Gens 2011.

7 Cisco 2013.

8 Ericsson 2011.

9 United Nations, Department of Economic and Social Affairs 2013.

10 Cisco 2013.

11 International Institute of Communications 2012.

12 OECD 2013.

13 Manyika et al. 2011.

14 Lanier 2013, p. 15.

15 See PCAST 2010 and Nguyen et al. 2013. When two or more entities agree to abide by a common set of legal rules, codes of conduct, other business and technical rules, and operational rules, they are generally referred to as belonging to the same trust framework.

16 In some respects, this is no more complicated than the agreements that have existed for years among telecommunications carriers—the scale and scope are different, the principles similar.

REFERENCES

Ahmad, N., P. Schreyer, and A. Wölfl. 2004. "ICT Investment in OECD Countries and its Economic Impacts." In *The Economic Impact of ICT: Measurement, Evidence and Implications.* Paris: OECD Publishing. 61–83.

Astrotic, B. K., P. Boegh-Nielsen, K. Motohashi, and S. Nguyen. 2002. *IT, Productivity and Growth in Enterprises: New Results from International Micro Data.* Paper presented at the OECD Workshop on ICT and Business Performance, OECD, Paris, December 9. Available at http://www.rieti.go.jp/it/productivity/docs/DJUS_OECD_.pdf.

Bughin, J. and J. Manyika. 2013. "Measuring the Full Impact of Digital Capital." *McKinsey Quarterly,* July. Available at http://www.mckinsey.com/insights/high_tech_telecoms_internet/measuring_the_full_impact_of_digital_capital.

Cisco. 2013. *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012–2017,* February 6. Cisco. Available at http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf.

Drucker, P. F. 1969. *The Age of Discontinuity: Guidelines to Our Changing Society.* New York: Harper and Row.

Ericsson. 2011. *More Than 50 Billion Connected Devices.* Ericsson White Paper 284 23-3149 Uen, February. Available at http://www.ericsson.com/res/docs/whitepapers/wp-50-billions.pdf.

Gens, F. 2011. *IDC Predictions 2012: Competing for 2020.* IDC #231720, Volume 1. Framingham, MA: IDC. Available at http://cdn.idc.com/research/Predictions12/Main/downloads/IDCTOP10Predictions2012.pdf.

International Institute of Communications. 2012. *Personal Data Management: The User's Perspective.* International Institute of Communications. Available at http://www.iicom.org/open-access-resources/doc_details/226-personal-data-management-the-users-perspective.

Lanier, J. 2013. *Who Owns the Future?* New York: Simon & Schuster.

Machlup, F. 1962. *The Production and Distribution of Knowledge in the United States.* Princeton, NJ: Princeton University Press.

Manyika, J. M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Hung Byers. 2011. *Big Data: The Next Frontier for Innovation, Competition and Productivity.* McKinsey Global Institute Report, May. Available at http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

Nguyen, M.-H. C., P. Haynes, S. Maguire, and J. Friedberg. 2013. "A User-Centred Approach to the Data Dilemma: Context, Architecture, and Policy." In *Digital Enlightenment Yearbook 2013.* IOS Press Ebooks. 227–42. Available at http://ebooks.iospress.nl/publication/35146.

OECD (Organisation for Economic Co-operation and Development). 2013. "Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value." *OECD Digital Economy Papers,* No. 220. OECD Publishing. Available at http://www.oecd-ilibrary.org/science-and-technology/exploring-the-economics-of-personal-data_5k486qtxldmq-en.

PCAST (President's Council of Advisors on Science and Technology). 2010. *Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: The Path Forward,* December. Available at http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf.

United Nations, Department of Economic and Social Affairs. 2013. *World Population Prospects: The 2012 Revision.* Available at http://esa.un.org/wpp/.