Big Data and Industrial Ecology

Ming Xu, Hua Cai, and Sai Liang

Keywords:

big data
collaborative data sharing
complexity
industrial ecology
information and communications technology (ICT)
input-output analysis (IOA)

Summary

Many have witnessed the increasing popularity of "big data" in the past couple of years. Indeed, big data has been transforming how business is done in many industries. For example, online advertisements are increasingly customized for individual consumers based on their purchase history. Big data has also instigated many new areas of investigation, mostly in fields such as computer sciences and statistics. There are many areas in industrial ecology (IE) that can potentially benefit from big data. In this article, we try to explore what big data could bring to IE.

What is Big Data?

Although the term "big data" is ubiquitously accepted and used in many areas, there is no consensus of defining big data. The uprising of big data seemed to take place around 2011, indicated by the increasing search interests of the keyword big data, as shown in figure 1. Although the increase of number of searches for big data can be the result of changing search algorithms (Lazer et al. 2014), it is somewhat correlated with the rapid development and deployment of information and communications technology (ICT), especially the usage of mobile devices (KPCB 2014). In general, ICT makes available two types of information in the form of big data: new information created from ICT applications (e.g., social media, online documents, and phone records) and existing information that is previously unavailable (e.g., business transaction data collected at a large scale with detailed records or daily travel trajectories of individual vehicles equipped with a global positioning system [GPS]).

Diving deeper into what exactly big data means, Ward and Barker (2013) recently surveyed existing definitions of big data in academia, industry, and media. They identified three features differentiating big data from other data we commonly encounter:

- Size: Big data is often large in volume (e.g., in terabytes or larger);
- Complexity: Big data often contains highly complex sets of information that are not easy to understand; and
- Technology: Tools and techniques that are used for processing and analyzing traditional data are not suitable for big data.

Among those three factors, the "technology" factor is actually dependent on the other two factors. New tools and techniques are obviously needed if the data are too large or too complex to surpass the capacity of existing methods to process and analyze. Therefore, it seems that "size" and "complexity" are the fundamental features that characterize big data.

Big data enabled by ICT naturally grows rapidly and becomes massive. That is why big data is often large in volume. However, size is not necessarily a defining factor for big data. By flipping a coin trillions or quadrillions of times, one can generate a large set of data in terabytes or petabytes, which meets the size criterion in big data definitions. But, the only challenge of processing and analyzing this "large" data set would relate to hardware (i.e., storing and retrieving data in such large volume). Existing analytical techniques, such as statistical analysis, are sufficient enough to comprehend it. Therefore, a large data is not necessarily a big data (MIKE 2.0 2014). The size factor is then not necessarily a determining factor for big data. Big data needs to be "big" in complexity, not necessarily in size. In other words, big data essentially characterizes the behavior of complex adaptive systems that are often constituted by a large number of

Address correspondence to: Ming Xu, School of Natural Resources and Environment, University of Michigan, 440 Church Street, Ann Arbor, MI 48109-1041, USA. *Email:* mingxu@umich.edu

DOI: 10.1111/jiec.12241

Editor managing review: Sybil Derrible

Volume 00, Number 0

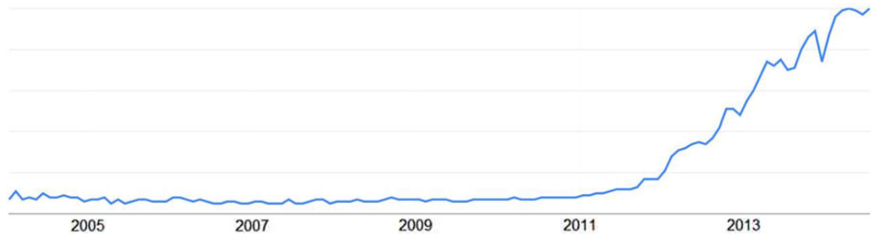

Figure 1 Search interests of big data from Google Trend; data represent search interest relative to the highest point on the chart.

heterogeneous components interacting with one another and evolving (Dijkema and Basson 2009).[1]

The degree of complexity of big data often relates to the amount and type of unstructured data it contains. These unstructured data are not organized in predefined fashions, may come from multiple sources and in various formats, are masked by noises with little obvious usefulness, and need to be processed and analyzed—often involving significant computational work—for hidden, but useful, information. Processing and harmonizing such data may not be difficult methodologically. However, when the size of such unstructured data increases, the computational challenge to automate data processing and harmonization may become significant. An example related to industrial ecology (IE) is from Cooper and colleagues (2013) discussing the role of big data in life cycle assessment (LCA). Data needed in LCA come from multiple sources in different formats. Significant efforts are required to integrate and harmonize these data before they can be used in LCA. It is also worth noting that, although unstructured data are a more prevalent form of big data, semistructured and structured data can also be complex and big, requiring similar computational resources and techniques as unstructured big data does.

Another type of big data complexity comes from the complexity of the underlying system behavior it describes. The data set itself in this case might not be as complex as expected (semistructured or structured data); but it can be used to obtain useful information to characterize the complex behavior of the underlying system. The online recommendation system is an example that recommends products to consumers based on purchase history of all users. The data describing the purchase history are relatively simple: a binary matrix corresponding users with products using 1 representing purchased and 0 representing not purchased. By mining this simple, but often large, matrix, one can characterize a particular user's preference based on his or her purchase history and other users' purchase history. This is a complex behavior that cannot be directly observed from the purchase history data, but can be extracted by analyzing it.

What can Big Data Offer?

What big data can offer includes not only rich information from large-scale data sets, but also an alternative modeling approach in many fields, including IE.

The mainstream modeling approaches in IE and many disciplines are generally based on three rational steps: developing a model based on theories explaining a phenomenon; validating the model (and the underlying theories) using limited observations; and using the validated model to forecast or predict future phenomena without observations. This is related to what Immanuel Kant described in his *Critique of Pure Reason* as a priori knowledge: One can understand the system through logic reasoning; experiences are used only to validate the logic model. Therefore, theories are critical in this a priori modeling approach, whereas data only play a supporting role for theory development and validation. Moreover, theories explaining target phenomena are often developed using a reductionism approach, in which a real-world phenomenon is decomposed to a set of variables based on certain assumptions. After validation, these variables are composed back to mimic the real-world phenomena.

The classic theory-driven, reductionism, a priori approach has been successful in simple, relatively static systems. However, when it comes to complex systems, which are constituted by a large number of heterogeneous components interacting with one another and constantly evolving, the a priori modeling approach reaches a limit commonly known as the *curse of dimensionality* (Bellman 1961). In particular, a complex system often needs to be decomposed to a large number of variables owing to its complexity. The number of variables represents the number of dimensions of the space characterizing the system. Limited observations used for validation in the reductionism approach then become sparse in the high-dimensional space, which brings the difficulty of characterizing the high-dimensional space using sparse, low-dimensional data. On the other hand, a posteriori modeling approaches are data driven, aiming to learn from experiences (e.g., Bayesian analysis in many fields of engineering). With big data, such data-driven approaches can be significantly improved with significantly more data to experience with.

First, big data generated from large-scale ICT implementations can potentially provide additional information regarding the behavior of a system from many different aspects for a relatively long period of time or in real time. With appropriate configurations (e.g., what data to collect and how to collect), big data can potentially help address the curse of dimensionality by increasing the number of observations for model validation.

Second, big data could provide an alternative way of modeling by allowing direct encapsulation of rich information into operational models without the theorization step in the first place. For example, in figure 2, limited observation of the target system may imply that the system is either a triangle or a quadrangle.

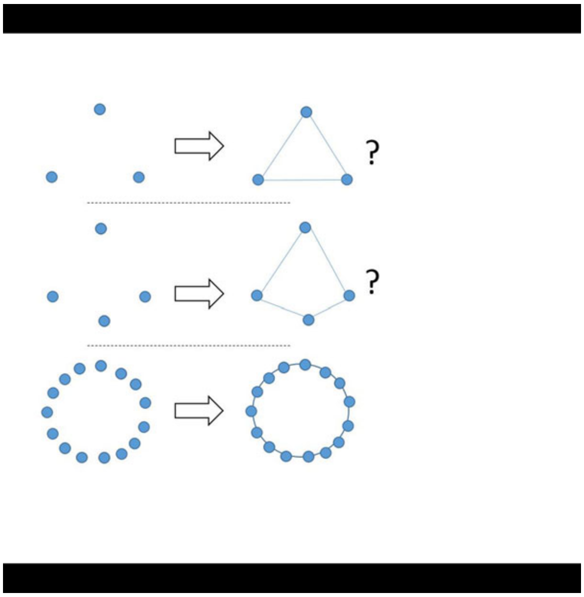

Figure 2 Increasing the size of observations could help to better understand the system.

However, sufficient amount of data "speak for themselves" without significant theorization that the system is actually a circle. This is described by Kant as a posteriori knowledge: One needs to understand a system through experiences. In a data-driven, a posteriori modeling approach, the causality of the behavior of a system is relatively less important to know. Characteristics of the data reflect the behavioral features of the underlying system; thus, the data-driven model can mimic the behavior of the system without exhaustively decomposing the system to variables, provided that the model is rigorously developed and the data are large enough to characterize the system. Note that more data does not always mean better representation of the real world. Correctly collecting appropriate data with the consideration of noises is the key for enabling big data speaking for themselves (more in the *Critiques* section of this article).

An example that illustrates the utility of this data-driven, a posteriori modeling approach is provided by González and colleagues (2008). In this study, individual human mobility patterns are characterized using a big data of the travel trajectory of 100,000 mobile phone users during a 6-month period. The study found strong temporal and spatial regularities in individual human travels that are different from what previous theory-driven models (i.e., Lévy flight and random walk models) predict. The modeling approach applied in this study is purely data driven, in the way that no theory is used to describe how individual users travel as theory-driven models usually start with. The patterns of individual mobility are entirely derived from analyzing the travel trajectory data.

What can Big Data Bring to Industrial Ecology?

Several studies in IE have involved big data and techniques dealing with big data. The main focus thus far has been using communication and computational infrastructure to facilitate virtual collaboration among researchers. For example, Kraines and colleagues (2001) developed a distributed object-based modeling environment (a cloud-based computational infrastructure, in modern terms) to facilitate virtual collaboration among researchers. Kraines and colleagues (2005) presented an Internet-based knowledge integration and collaboration platform for integrated environment assessment. Davis and colleagues (2010) proposed an Industrial Ecology 2.0 agenda to utilize the World Wide Web for collecting, processing, curating, and sharing data as a community, rather than individuals. More recently, Lenzen and colleagues (2014) used a cloud-based computational infrastructure to develop and apply multiregional input-output (I-O) models by allowing researchers at different sites to collaborate on a virtual platform.

In addition to virtual collaboration, big data and big data concepts have been discussed in the IE literature, such as for better measuring environmental footprints of human consumption (Hubacek et al. 2014) and data challenges in LCA (Cooper et al. 2013). Nevertheless, big data and the suite of analytical tools available for analyzing big data offer opportunities to address some long-standing issues in IE or provide alternative approaches.

Despite the increasing level of sophistication, many areas of IE still rely on aggregated data to represent averages of industrial systems at various scales. For example, LCA mainly uses industrial average data to measure life cycle environmental impacts of product or service systems. With the increasing availability of big data, it is possible to bring IE to the next phase focusing more on the spatial, temporal, and demographic heterogeneity of industrial systems, through new data, new data analysis techniques, or the combination of both. Note that industrial ecologists do not necessarily need to be experts in sourcing and collecting data themselves. Instead, research focuses should be on utilizing newly available data and newly developed data analytics tools to complement existing data and methods through interdisciplinary collaborations.

The most obvious and direct application of big data in IE is to help develop more-realistic complex systems models (e.g., agent-based models) to better capture essential features of human behavioral dynamics (Axtell et al. 2008). Based on sufficient data directly reflecting human behavior (e.g., social media data), it is possible to derive more-realistic characterization of human behavioral dynamics, instead of solely relying on assumptions or loosely tested theories. For example, Amazon's Mechanical Turk[2] provides a crowdsourcing platform for conducting quick, inexpensive surveys that can help characterize human behaviors in agent-based models. Mining social media data can help understand consumer preferences toward products with different environmental implications. Incorporating with geospatial information, big data can help design spatially explicit models to reflect the spatial heterogeneity of human behaviors. On the other hand, increasingly available big data directly recording human activities, such as GPS-enabled travel trajectories, can be used to validate and calibrate models that were previously difficult to do (Windrum et al. 2007).

Material/substance flow analysis (MFA/SFA) usually models stock changes of materials or products using average or assumed product lifespan (e.g., Yu et al. 2010; Bollinger et al. 2011). Many forms of big data can help better characterize human consumption behavior, such as purchase history from online retailers or calorie intake recorded by smartphone apps. The results will be useful to model stock changes of materials or products with higher spatial and temporal resolutions in MFA/SFA.

Urban metabolism, which has long relied on sector-aggregated or industrial average data to measure resource and energy demand in urban areas, is another area that can benefit greatly from big data. With big data that can better characterize the consumption activity of urban residents, one may be able to develop better urban metabolism models and provide more-effective policy interventions. For example, location-based big data, such as GPS-enabled travel trajectories and geotagged social media data, is increasingly used to characterize human travel behaviors at the individual level (e.g., González et al. 2008). This potentially can help refine urban metabolism models to reach finer spatial and temporal resolutions in transportation-related areas. Hubacek and colleagues (2014) discussed the potential of using geodemographics data and social media data to measure the environmental footprints of human consumption at high spatial resolution. Similar data can also be applied to urban metabolism and quantify urban teleconnetions (Liu et al. 2013). Last, big data on urban consumption activities can also be used to validate and calibrate some of the microsimulation efforts seen in the urban metabolism literature to better model human activities (e.g., Keirstead and Sivakumar 2012).

One of the key tasks for an LCA is to characterize the supply chain of a product or service system. Static, aggregated characterization of supply-chain relationships works fine for established systems, but faces challenges for emerging, dynamic systems (McKone et al. 2011; Miller et al. 2012). Consumer behavior patterns informed by big data (e.g., social media data) can also be used to add a new human dimension to LCA. In addition to the general application to LCA, big data, particularly those data characterizing human mobility dynamics (e.g., geotagged social media data), can be useful to improve LCA studies for transportation systems (e.g., Hawkins et al. 2013) by incorporating the dynamics of user travel patterns. This is useful to facilitate the development of emerging transportation technologies, such as electric vehicles. For example, Cai and Xu (2013) used travel trajectory data of more than 10,000 taxi cabs to study the impacts of electric vehicle design, cost, and government subsidy to the adoption and utilization of electric vehicles and the consequent greenhouse gas emissions.

I-O analysis (IOA) in IE has long been challenged by the limited availability of I-O tables (IOTs). Government statistics are the main source of IOTs. However, government-sponsored IOTs usually are available for every couple of years and do not cover all countries. Developing time-series IOTs for the global economy with high sector and country resolutions has been an impossible mission until recently, owing to increased capacity of dealing with big data. The University of Sydney has developed a time-series high-resolution world multiregional I-O model, Eora, using sophisticated computational infrastructure and techniques to harmonize and reconcile large amounts of raw data from many different sources (Lenzen et al. 2012). Cloud computing is also used for virtual laboratories for collaboratively developing large-scale I-O models with high resolutions (Lenzen 2014). Although the data involved in these efforts are large in volume, they are less complex in the sense that the underlying I-O methodology is still based on linear relationships between sectors. However, computational infrastructure and techniques such as cloud computing that were developed partially for analyzing big data played a critical role in these endeavors. A relevant effort enabled by big data is the Billion Prices Project,[3] which uses high-frequency price data from hundreds of global online retailers on a variety of topics related to price in macroeconomics and international economics, such as developing inflation data on a daily basis. I-O research can potentially also use online retailer price data to improve I-O models with respect to prices.

The success of industrial symbiosis (IS) is partially facilitated by information sharing (Chertow 2007; Ashton 2008). Grant and colleagues (2010) evaluated ICT-based information-sharing tools for IS development. The Delft University of Technology has set up an online platform for documenting eco-industrial parks around the world.[4] Data generated from these information-sharing platforms are big data in the sense that they come from different sources characterizing different systems. Analyzing and mining the ensemble of those data can potentially help to better understand IS.

Last, but not least, data visualization has been an important research approach in many fields to observe patterns from raw data, guide further analysis, and communicate the results with stakeholders. Big data brings challenges of how to effectively visualize large amounts and various types of data in meaningful ways. This is particularly relevant for industrial ecologists to communicate with decision makers given that the ultimate goal of many IE studies is to shape policy making. There are several projects visualizing data generated from environmental IOA. For example, Economy Map 2.0[5] turns the results of a typical environmental IOA, environmental impacts induced by the consumption of goods or services, in particular, industries in the U.S. economy, into an interactive visual map. Another example is the interactive visualization of the 2007 IOT for the United States,[6] which allows users to see the direct and indirect economic relationship between industries.

Critiques

While having many promises, big data is also criticized for potentially creating wrong, misleading results and unethical practices. Boyd and Crawford (2012) raised six issues with big data:

- "Big data changes the definition of knowledge," which comes with certain limitations with these new knowledge systems;

- "Claims to objectivity and accuracy are misleading," given that big data may not necessarily represent the objective truth of the underlying system;
- "Bigger data are not always better data," because better data come from better approach of data collection and analysis, not necessarily from the size of the data;
- "Taken out of context, big data loses its meaning," given that each data set represents a particular aspect of a system and thus is suitable for particular lines of inquiry;
- "Just because it is accessible does not make it ethical," because many big data available today are about human behaviors at the individual level;
- "Limited access to big data creates new digital divides," which might create new injustice between groups that have easy access to data and groups that do not.

It is worth noting that the second and third issues are particularly relevant to IE research involving modeling. Concretely, big data's ability of being "theory free" relies on the assumption that the data can represent the entire system under observation without significant sampling bias (Harford 2014). However, this is not always the case. For example, U.S. Twitter users in 2013 only represent approximately 18% of all Internet users in the United States and are disproportionately young (i.e., 31% between ages 18 and 29) (Pew Research Center 2013). Therefore, it is important to understand the nature of the data and decide whether to use it or where to use it.

When it comes to big data with significant sampling bias, one can either try to find more data to improve representation or use appropriate data analytic tools to examine the data with caution. In addition to more data, what big data could offer is new, advanced analytical tools that can be applied to many fields. For example, matrix completion is an approach to estimate missing data in a matrix solely based on a few entries (Candès and Recht 2009). It has been used in image processing (Bruckstein et al. 2009) and online recommendations (Bell et al. 2008). Similar methods are potentially useful for IE by estimating data for I-O matrices.

Outlook

Big data seems to be a mega trend. Media attentions on big data might fall eventually. But our dependence on ICT perhaps will be inevitably intensified. Massive amounts of data describing how the industrial system works may become increasingly available to industrial ecologists. Such data can function to industrial ecologists the same as data obtained from experiments in other fields: direct observation of the research target.

Looking forward, ICT is helping develop an "Internet of Things" that connects the real and virtual worlds with increasingly widely deployed sensors and communication networks. The German government has started an Industrial 4.0 initiative[7] to prepare for the upcoming "fourth industrial revolution," which is facilitated by the Internet of Things. If the Internet of Things is successful, our ability of using data to describe the industrial system and its socioeconomic surroundings will be significantly improved. Many unresolved questions in IE may be addressed with the newly available cyberinfrastructure and data. For example, using radiofrequency identification to track waste electrical and electronic equipment has been discussed extensively in IE literature (O'Connell et al. 2013). This could be realized soon with the Internet of Things. More broadly, imagine a world where all goods are sensored and tracked using ICT devices. The cost of obtaining data would become marginal. We might easily know, in real time, how much energy and materials are used by particular buildings, communities, cities, regions, countries, and even the entire world. LCA practitioners might not need to spend too much time on sampling and collecting life cycle inventory data; instead, data specific to their LCA projects might be readily available at their fingertips. However, more, better data do not necessarily mean better understanding of the real world. Instead, much work has to be done to transform data into knowledge and knowledge into action. At that time, IE might become a field of study focusing on how to make sense of big data and how to facilitate decision making based on knowledge gained from analyzing big data.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1132581. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the view of the National Science Foundation. Hua Cai and Sai Liang also thank the support provided by the Dow Sustainability Fellowship Program.

Notes

1. Also see the Special Issue on Complex Adaptive Systems of *Journal of Industrial Ecology* (Volume 13, Number 2, 2009).
2. www.mturk.com/.
3. http://bpp.mit.edu/.
4. http://ie.tudelft.nl/.
5. http://economymap.org/.
6. www-personal.umich.edu/~mingxu/usio2007/.
7. www.bmbf.de/en/19955.php/.

References

Ashton, W. 2008. Understanding the organization of industrial ecosystems. *Journal of Industrial Ecology* 12(1): 34–51.

Axtell, R. L., C. J. Andrews, and M. J. Small. 2008. Agent-based modeling and industrial ecology. *Journal of Industrial Ecology* 5(4): 10–13.

Bell, R. M., Y. Koren, and C. Volinsky. 2008. The BellKor 2008 solution to the Netflix Prize. *AT&T Research*. http://public.research.att.com/~volinsky/netflix/Bellkor2008.pdf/. Accessed 30 July 2014.

Bellman, R. E. 1961. *Adaptive control processes: A guided tour*. Princeton, NJ, USA: Princeton University Press.

Bollinger, L. A., C. Davis, I. Nikolić, and G. P. J. Dijkema. 2011. Modeling metal flow systems: Agents vs. equations. *Journal of Industrial Ecology* 16(2): 176–190.

Boyd, D. and K. Crawford. 2012. Critical questions for big data. *Information, Communication & Society* 15(5): 662–679.

Bruckstein, A. M., D. L. Donoho, and M. Elad. 2009. From sparse solutions of systems of equations to sparse modeling of signals and images. *Society for Industrial and Applied Mathematics (SIAM) Review* 51(1): 34–81.

Cai, H. and M. Xu. 2013. Greenhouse gas implications of fleet electrification based on big data-informed individual travel patterns. *Environmental Science & Technology* 47(16): 9035–9043.

Candès, E. J. and B. Recht. 2009. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics* 9(6): 717–772.

Chertow, M. R. 2007. "Uncovering" industrial symbiosis. *Journal of Industrial Ecology* 11(1): 11–30.

Cooper, J., M. Noon, C. Jones, E. Kahn, and P. Arbuckle. 2013. Big data in life cycle assessment. *Journal of Industrial Ecology* 17(6): 796–799.

Davis, C., I. Nikolic, and G. P. J. Dijkema. 2010. Industrial ecology 2.0. *Journal of Industrial Ecology* 14(5): 707–726.

Dijkema, G. P. J. and L Basson. 2009. Complexity and industrial ecology. *Journal of Industrial Ecology* 13(2): 157–164.

González, M. C., C. A. Hidalgo, and A.-L. Barabási. 2008. Understanding individual human mobility patterns. *Nature* 453(7196): 779–782.

Grant, B. G., T. P. Seager, G. Massard, and L. Nies. 2010. Information and communication technology for industrial symbiosis. *Journal of Industrial Ecology* 14(5): 740–753.

Harford, T. 2014. Big data: Are we making a big mistake? *The Financial Times*, 28 March. www.ft.com/intl/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html#axzz2yNfkYQbQ/. Accessed 30 July 2014.

Hawkins, T. R., B. Singh, G. Majeau-Bettez, and A. H. Strømman. 2013. Comparative environmental life cycle assessment of conventional and electric vehicles. *Journal of Industrial Ecology* 17(1): 53–64.

Hubacek, K., K. Feng, J. C. Minx, S. Pfister, and N. Zhou. 2014. Teleconnecting consumption to environmental impacts at multiple spatial scales. *Journal of Industrial Ecology* 18(1): 7–9.

Keirstead, J. and A. Sivakumar. 2012. Using activity-based modeling to simulate urban resource demands at high spatial and temporal resolutions. *Journal of Industrial Ecology* 16(6): 889–900.

Kraines, S., R. Batres, M. Koyama, D. Wallace, and H. Komiyama. 2005. Internet-based integrated environmental assessment. *Journal of Industrial Ecology* 9(3): 31–50.

Kraines, S., D. R. Wallace, Y. Iwafune, Y. Yoshida, T. Aramaki, K. Kato, H. Hanaki, et al. 2001. An integrated computational infrastructure for a virtual Tokyo. *Journal of Industrial Ecology* 5(1): 35–54.

KPCB (Kleiner Perkins Caufield & Byers). 2014. Internet trends 2014. www.kpcb.com/internet-trends/. Accessed 30 July 2014.

Lazer, D., R. Kennedy, G. King, and A. Vespignani. 2014. The parable of Google flu: Traps in big data analysis. *Science* 434(6176): 1203–1205.

Lenzen, M. 2014. An outlook into a possible future of footprint research. *Journal of Industrial Ecology* 18(1): 4–6.

Lenzen, M., A. Geschke, T. Wiedmann, J. Lane, N. Anderson, T. Baynes, J. Boland, et al. 2014. Compiling and using input-output frameworks through collaborative virtual laboratories. *Science of the Total Environment* 485–486(1): 241–251.

Lenzen, M., K. Kanemoto, D. Moran, and A. Geschke. 2012. Mapping the structure of the world economy. *Environmental Science & Technology* 46(15): 8374–8381

Liu, J., V. Hull, M. Batistella, R. DeFries, T. Dietz, F. Fu, T. W. Hertel, et al. 2013. Framing sustainability in a teleconnected world. *Ecology and Society* 18(2): 26.

McKone, T. E., W. W. Nazaroff, P. Berck, M. Auffhammer, T. Lipman, M. S. Torn, E. Masanet, et al. 2011. Grand challenges for life-cycle assessment of biofuels. *Environmental Science & Technology* 45(5): 1751–1756.

MIKE 2.0. 2014. The open source standard for information management: Big data definition. http://mike2.openmethodology.org/wiki/Big_Data_Definition/. Accessed 28 January 2014.

Miller, S., S. Moysey, B. Sharp, and J. Alfaro. 2012. A stochastic approach to model dynamic systems in life cycle assessment. *Journal of Industrial Ecology* 17(3): 352–362.

O'Connell, M., S. Hickey, M. Besiou, C. Fitzpatrick, and L. N. van Wassenhove. 2013. Feasibility of using radio frequency identification to facilitate individual producer responsibility for waste electrical and electronic equipment. *Journal of Industrial Ecology* 17(2): 213–223.

Pew Research Center. 2013. Social media update 2013. www.pewinternet.org/files/2013/12/PIP_Social-Networking-2013.pdf/. Accessed 30 July 2014.

Ward, J. S. and A. Barker. 2013. Undefined by data: A survey of big data definitions. arXiv: 1309.5821v1 [cs.DB]. Accessed 28 January 2014.

Windrum, P., G. Fagiolo, and A. Moneta. 2007. Empirical validation of agent-based models: Alternative and prospects. *Journal of Artificial Societies and Social Simulation* 10(2): 8.

Yu, J., E. Williams, M. Ju, and Y. Yang. 2010. Forecasting global generation of obsolete personal computers. *Environmental Science & Technology* 44(9): 3232–3237.

About the Authors

Ming Xu is an assistant professor in the School of Natural Resources and Environment and also in the Department of Civil and Environmental Engineering at the University of Michigan, Ann Arbor, MI, USA. **Hua Cai** is a Dow Sustainability Doctoral Fellow pursuing a joint interdisciplinary PhD between the School of Natural Resources and Environment and the Department of Civil and Environmental Engineering at the University of Michigan. **Sai Liang** is a Dow Sustainability Postdoctoral Fellow in the School of Natural Resources and Environment at the University of Michigan.