Sustainability Science (2020) 15:955–974
https://doi.org/10.1007/s11625-019-00759-9

REVIEW ARTICLE

A review of data-intensive approaches for sustainability: methodology, epistemology, normativity, and ontology

Vivek Anand Asokan[1] · Masaru Yarime[2,3,4] · Motoharu Onuki[5]

Received: 12 April 2018 / Accepted: 12 November 2019 / Published online: 2 December 2019

Abstract

With the growth of data, data-intensive approaches for sustainability are becoming widespread and have been endorsed by various stakeholders. To understand their implications, in this paper, data-intensive approaches for sustainability will be explored by conducting an extensive review. The current data-intensive approaches are defined as an amalgamation of traditional data-collection methods, such as surveys and data from monitoring networks, with new data-collection methods that involve new information communication technology. Based on a comprehensive review of the current data-intensive approaches for sustainability, key challenges are identified: the lack of data availability, diverse indicators developed from a narrowly viewed base, diverse definitions and values, skewed global representation, and the lack of social and economic information collected, especially among the business indicators. To clarify the implications of these trends, four major research assumptions regarding data-intensive approaches are elaborated: the methodology, epistemology, normativity, and ontology. Caution is required when data-intensive approaches are masked as "objective". Overcoming this issue requires interdisciplinary and community-based approaches that can offer ways to address the subjectivities of data-intensive approaches. The current challenges to interdisciplinarity and community-based approaches are also identified, and possible solutions are explored, so that researchers can employ them to make the best use of data-intensive approaches.

Keywords Data-intensive approaches · Sustainability · Sustainability indicators · SDGs · Planetary boundary · Open data · Big data

Handled by Daniel J. Lang, Leuphana University of Lueneburg, Germany.

✉ Vivek Anand Asokan
viv.asok@gmail.com

Masaru Yarime
yarime@ust.hk

Motoharu Onuki
onuki@edu.k.u-tokyo.ac.jp; onuki@k.u-tokyo.ac.jp

[1] Graduate Program in Sustainability Science, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Japan

[2] Division of Public Policy, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR, China

[3] Department of Science, Technology, Engineering and Public Policy, University College London, London, UK

[4] Graduate School of Public Policy, The University of Tokyo, Tokyo, Japan

[5] Graduate Program in Sustainability Science (GPSS), Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Japan

Introduction

Trade interlinkages and global supply chains have led to the extraction of multiple resources (Williams et al. 2016): groundwater (Dalin 2017), human net primary production (Imhoff et al. 2004), and mineral wealth (Wiedmann et al. 2015). Subsequently, the extraction of resources has to led to transboundary problems: air pollution (Zhang et al. 2017), biodiversity impact (Moran and Kanemoto 2017), and climate change [Intergovernmental Panel on Climate Change (IPCC) 2018], with the potential to cross the "planetary boundaries"[1] (Rockström 2013). Combatting these varied and diverse challenges on such a scale requires data about trade linkages delineating different actors—consumers, intermediaries, and extractors—across different countries and consumption categories, together with tools to interpret the data.

[1] The planetary boundaries frame a safe operating space as a condition for sustainable development which is applicable to governments, business, and researchers (Rockström 2009).

According to data proponents, data from multiple fields and evidence from experimentation and machine learning are the best approaches to provide solutions in sustainability science research. Evidence has been emphasized as an important aspect in the sustainability science scholarship (Clark et al. 2016; Miller et al. 2014). Evidence can be provided from data obtained through experiments and this is finding increasing support from governments, which have provided the impetus for evidence-based policy making (House of Commons 2011; Rutter 2012). The traditional research on sustainability focuses on problem solving from a particular perspective (Komiyama and Takeuchi 2006; Jerneck et al. 2011; Lang et al. 2012) and sustainability science is positioned as a transdisciplinary research that provides solutions to sustainability incorporating multiple perspectives (Spangenber 2011; Kates 2011; Clark 2007). It is proposed that data-intensive approaches will facilitate experimentation, promote social learning, and enhance decision-making (Yarime 2017, 2018).

Yet, the impacts of data-intensive approaches are poorly understood. The current data infrastructure, as well as the data-collection strategies and their interpretation have not yet been carefully studied in regard to the impact on decision-making for sustainable development (IEAG 2014; O'Niel 2017). The current status of data-intensive approaches and the implications of gaining knowledge from these approaches need to be studied further. In this paper, we explore how data-intensive approaches are used to gathered evidence, collaborate across disciplines, and complement participatory and transdisciplinary approaches. We question the assertion made by data proponents that data are all that is needed. The implications of data-intensive approaches for sustainability are elaborated in terms of the research assumptions made with such approaches (O'Niel 2017; Derman 2011; Mittelstadt and Floridi 2016).

We capture various facets and define data-intensive approaches for the purpose of this paper. To further explore the broad implications of data-intensive approaches for sustainability, we establish the following research questions to conduct a critical literature review:

1. What are old, new, and data-driven revolutions?
2. What is the current usage of these data approaches given the high hopes surrounding the use of data-intensive approaches in sustainable development?
3. What are the research assumptions of data-intensive approaches and their implications for sustainable development?

In spite of the exuberance of data proponents, information on many sustainability related issues is still lacking. Furthermore, caution is required when using data-intensive approaches, as they are reductive and inductive, and the subjectivities involved in data-intensive approaches, which are masked as "objective", can skew the outcomes of research, as Ronald Coase cautions when he states: "If you torture the data long enough, it will confess"(Tullock 2001). The current data-intensive approaches are plagued with ontological relativism and scientific imperialism, and therefore, these data-intensive approaches cannot be considered objective. Overcoming these structural issues requires interdisciplinary and community-based approaches that will offer ways to address sustainability by highlighting subjectivities to make the use of data-intensive approaches objective. We will emphasize the challenges faced by researchers in transitioning to interdisciplinarity and community approaches, and recommend new approaches based on the review. Data can open up new directions in providing causality-based knowledge to provide a fertile approach to bridge data from the multiple fields available. Furthermore, data-intensive approaches such as experimental approaches, normative contextual approaches and quantitative storytelling approaches can provide new community-based data-intensive approaches. In addition, the author's flag up issues that are often neglected, e.g., the environmental impact of ICT-based data-intensive approaches.

This paper is structured as follows: "Methodology" lists the methodology used in the paper; "Understanding data approaches in decision-making and research" describes the data-intensive approaches; "Challenges of data-intensive approaches: issues and implications" lists the data-intensive approaches focused on sustainability; "Implications for future research: challenges and way forward" lists the assumptions made in data-intensive research approaches—methodology, epistemology, normativity, and ontology; "Conclusion" discusses the implications of data-intensive approaches, focusing on sustainability for academia and policy making; and the last section summarizes the conclusion.

Methodology

Although data-driven approaches have been specified recently, data have also been part of researchers' everyday practice. Furthermore, different characteristics and features of data-driven approaches have been articulated by different proponents, including researchers, policy makers, and businessmen. More recently, their use in sustainable development has been actively proposed and encouraged. As pointed out earlier, the implications of the use of data-intensive approaches in sustainable development have not been studied or discussed. At the same time, understanding the implications of data-intensive approaches, which has a varied historical context and widespread usage, is fundamental. Such a diverse topics requires

Table 1 Data source used in this paper to describe various data-intensive approaches

Data-intensive approaches	Data source used in the paper
Planetary boundaries	Global scientists including scientists from the Stockholm Resilience Centre have pioneered the concept of Planetary boundaries. Information on planetary boundaries is sourced from documents and papers published by researchers at Stockholm Resilience Centre (Cornell and Downing 2004)
Sustainable development goals	The Inter-agency and Expert Group on SDG Indicators (IAEG-SDGs) was created by United Nations Statistical Commission with the task of developing the global indicator framework for SDGs and targets. Information on SDGs is sourced from the publication of IAEG-SDGs (IAEG-SDG 2015)
Corporate data	The Global Initiative for Sustainability Ratings (GISR) has been working to create a rating for raters. GISR collects information on company-level sustainability ratings. Data on corporate indicators is sourced from GISR hub (Ceres 2016)
Open data	The Open Data Impact Map created by the Open Data for Development Network (OD4D) is a public database of organizations that use open government data from around the world. Data on open data is sourced from Open Data Impact Map (Center for Open Data Enterprise 2016)
Big data	The Global Pulse is an initiative of the United Nation Secretary General on big data, which has the mission of accelerating its adoption for sustainable development and humanitarian action. Information on big data approaches is sourced from the repository of the Global Pulse's project listing on the website (Global Pulse 2018)

covering a comprehensive body of literature to understand the past, present, and future prospects.

The subject matter is captured based on a literature review in four stages: (1) covering the past impacts of data-intensive approaches by studying the diverse data-collection methods, data interpretation techniques, and uses of data in decisions to define data-intensive approaches ("Understanding data approaches in decision-making and research"); (2) mapping the present data-intensive approaches used in sustainable development to elucidate the current trends ("Challenges of data-intensive approaches: issues and implications"); (3) studying the implications of the current data-intensive approaches by delineating the various research assumptions to understand the objective nature of data-intensive approaches ("Implications for future research: challenges and way forward"); and (4) mapping possible solutions that may be employed to enhance the future use of data-intensive approaches ("Implications for future research: challenges and way forward").

As well as the literature review, secondary data from the comprehensive database are used to map the present data-intensive approaches used in sustainable development. The data sources of various of data-intensive approaches is presented in Table 1. These approaches are varied, and we have captured wide-ranging information based on the concept (planetary boundary), use [Sustainable Development Goals (SDGs) and corporate data], and characteristics (Open and Big Data).

Understanding data approaches in decision-making and research

There are many different data-intensive approaches for sustainability that are promoted by different actors—corporate actors, development agencies, and researchers—and some of the practical applications are listed here. Data-intensive approaches in sustainability frameworks and corporate industries are promoted by multinational organizations to aid decision-making in regard to sustainability issues (CSIS and JICA 2017; Petrov et al. 2016). On the corporate side, in an attempt to be "more sustainable", businesses and corporations have created Key Performance Indicators (KPIs), sustainability indicators, and reporting mechanisms. Researchers are divided; Marland et al. (2015) and Eccles et al. (2014) found these developments to be adequate, while Trexler and Schendler (2015), Ceres and Sustainalytics (2012), and Leisinger and Bakker (2013) found these developments to be inadequate in terms of reflecting sustainability.

Similarly, at the local, national, and international government levels, the United Nations has set 17 aspirational Sustainable Development Goals (SDGs) to combat social, economic, and environmental challenges (United Nations 2015). Researchers have also emphasized the role of data in sustainable development studies, for example, its application in disaster management and Life Cycle Analysis (LCA) studies (Zhongming et al. 2014; Xu et al. 2015). Data-intensive approaches are claimed to foster the transparency of governments and companies, making them accountable for their actions through the use of open data (Rockström et al. 2013). Data-intensive approaches based on planetary boundaries, Sustainable Development Goals, and corporate-based approaches are seen to enhance decision-making by reflecting dimensions of sustainability (OECD 2002).

Such a diverse range of data-intensive approaches for sustainability, promoted by different actors, demands a better understanding of these approaches. In this section, we define data-intensive approaches as an amalgamation of "old" and "new" approaches of data collection and interpretation. Data-intensive approaches use data made available from the open data movement by providing access to diverse "old data" sources—e.g., surveys, monitoring stations and general public information—and use big "new data" collected from new technologies—e.g., wearable technology, mobile technology, smart card technology, satellite technology, and social media. Collecting data and producing information based on a framework while attempting to eliminate biases are seen as a hallmark of the old approach. The new approaches are based on machine learning and controlled experiments, and have a disregard for theory. The data are understood to have a meaning that can be uncovered without the need for human interpretation (Crawford 2013; Crawford and Schultz 2014; Callebaut 2012; Fairfield and Shtein 2014; Puschmann and Burgess 2014).

Following this description, the section elaborates data-intensive approaches centred on sustainable development frameworks that invoke both old and new data approaches. Based on the analysis, the challenges of diverse and disaggregated data-intensive approaches are listed and explained.

Old approaches

The old approach consists of data from surveys, monitoring stations and general public information, and uses data made available by the increasing movement towards open data. Open data can be seen as a movement towards making these data available for wider use to increase economic growth, transparency, efficiency, service delivery, and information sharing (World Bank 2015). Planetary boundaries, sustainable development, and corporate indicators are currently based on this approach. We further elaborate the processes of data collection and data interpretation and their use in decision-making.

Traditional approaches of data collection used in decision-making include surveys collected by government statistical agencies and researchers from data gathering systems such as census bureaus, national statistical offices, and environmental monitoring networks. Similarly, businesses collect data either to submit sustainability reports or to prepare key performance indictors, which are used to drive their business operations. The data-collection strategies are designed with a framework and are considered robust; however, the frequency of collection is not frequent.

The collected data are further processed into indicators, which are used in policy making, as they are considered "objective" (Head 2010; Parkhurst 2017). Indicators can be defined as "variables that summarize or otherwise simplify relevant information, make visible or perceptible phenomena of interest, and quantify, measure, and communicate relevant information" (Gallopin 1996), based on a conceptual framework, connecting indicators to a broader discourse or ideological position (Gudmundsson 2003; Pintér et al. 2005). Composite indicators are normalized and have weights; that is, they are weighed against other indices/indicators, while simultaneously combining indicators (Böhringer and Jochemc 2016). This allows for the presentation of multiple variables together as a single indicator, for example, the GDP indicator.

Some examples of indicators developed by multilateral agencies and businesses are as follows. The Organization for Economic Cooperation and Development (OECD) pioneered using traditional data interpretation approaches through composite indicators (JRC-EC 2018; OECD 1974, 1993, 2002), which are now popular with the UNEP (2002), the UNDP (2004) and the World Bank (2004), as well as other agencies (WRI 2003; World Council on City Data 2015). Using these old data interpretation approaches, corporate industry consortiums have developed many data-driven approaches: sustainability indicators and reporting mechanisms (Global Reporting Initiative 2015; Integrated Reporting Council 2013; Gilman and Schulschenk 2013; Climate Disclosure Standards Board 2016; Carbon Disclosure Project 2016); specific Key Performance Indicators (KPI) (Fitz-Gibbon 1990; Lydenberg et al. 2010); and business indices (Madnick and Siegel 2002; FTSE4 Good Index Series 2016; DJSI Annual Review 2015) to achieve sustainability-oriented goals.

The quantitative indicators mentioned above are a popular science–policy interface tool (Porter 1995; Boulanger 2014) that is used in decision-making. The composite indices and indicators are associated with an accepted salience, credibility, and legitimacy in the decision-making process, as they are considered objective (Cash et al. 2004; Parris and Kates 2003). The idea underpinning their popularity is that scientific knowledge, from indicators, compels action and has an impact on decision-making, where more research leads to a better understanding, which in turn helps to resolve political disagreements (Beck 2011).

New approaches

Big data refer to data that are large in volume. It is a complex set of information that is difficult to understand, and requires the use of new tools and techniques to analyze it (Ward and Barker 2013). Big data have been made possible due to new technologies and are seen to provide information at a higher frequency. It also provides the opportunity to link data, which is done through ICT technology (Heath and Bizer 2011); for example, the European Union pushed for the publication of government data in machine-readable formats (MELODIES Project 2016). These are seen as the strength of new approaches in providing "objective" data free of bias.

New data are gathered via different types of ICT, e.g., wearable technology, mobile technology, smart card technology, social media, and satellite technology (CSIS and

JICA 2017). It is believed that ICT will play a facilitating role in gathering data in the future (Moir et al. 2014). Big data are data acquired from diverse sources, including and especially from ICT tools, which are exploding in numbers (Keeso 2014; Orts and Spigonardo 2014).

The two major interpretation strategies are predictive and experimental tools. These approaches are anti-theory, and are built explicitly for prediction, classification and to produce evidence through experiments for interventions. Co-relation-based approaches use big computing, mathematical modelling, and algorithms, where the emphasis is on unveiling correlations or patterns. The data are fed and trained with the help of algorithms based on a mathematical model from a sample data set and used for classification. Rapid prototype testing (A/B testing or RCTs) studies particular "interventions" from an existing database or data collected from specific experiments (Narasimhan and Arun 2017). Specific interventions, in comparison with theory, provide the reasons behind the success or failure of a particular intervention in producing "evidence"-based science practice. Both open "old" data and big "new" data can be used for A/B testing and their use has been enthused about by businesses.

Artificial intelligence and big computing tools such as analytics and algorithms are seen to play a major role in this movement (Brown et al. 2011). The big data movement focuses on approximate analytics, scalability, and pattern identification with data visualization as a key strategy (Lazar 2012). The use of these tools is seen to uncover hidden patterns, unexpected relationships, and market trends, and reveal preferences by providing real-time, data striking an informed balance between accuracy and timeliness (World Bank 2018). Furthermore, business leaders have stressed that data can be used for experimentation to provide evidence-based knowledge (Davenport 2009; Thomke and Manzi 2014; McAfee et al. 2012; Brown et al. 2011). While its use in sustainability-based research is not known, its use is prevalent in college rankings, advertising, predictive policing, the selection of job applicants, evaluation of teachers, and the determination of creditworthiness (O'Niel 2017).

Defining data-intensive approaches—an amalgamation

The "old" and "new" data approaches appear to be similar in their claim to provide quantitative information that is considered "objective". For the purpose of this paper, we define data-intensive approaches as quantitative measures made available from the opening of "old data" such as surveys, environmental monitoring data, and big "new data approaches" from new technology, where information is processed either by: interpreting data to produce indicators or indices, machine learning to make predictions, or experiments to measure interventions. For example, in data-intensive approaches, an experimentation-based or predictive analysis can be performed on survey or census data. Concurrently, indices can also be constructed based on big data. The current data-intensive approaches for decision-making in regard to sustainability fall under this same category.

Data-intensive sustainability approaches

The data-intensive approaches selected cover both old data approaches—planetary boundary and sustainable development goals—and mixed approaches—corporate indicators and open data, and new data approaches—big data. The planetary boundaries and SDGs, developed by researchers and policy makers, currently predominantly use "old" data approaches. Furthermore, information on sustainable development from corporate indicators and open data is also chiefly based on old approaches. However, this is changing with data from ICT and satellite technology influencing corporate indicators and open data. The big data approaches promoted by the United Nations are exclusively based on new approaches. Since these approaches are varied, we have captured wide-ranging information about these approaches, which are not comparable. This serves to understand the current usage of these data approaches in sustainable development.

Planetary boundaries

The multiple planetary boundaries are selected and decided on through a top–down scientific process decided by a group of scientists. We list the data availability regarding four major planetary boundaries frameworks in Table 2: climate change, chemical pollution, nitrogen pollution, and ecosystems. Data on climate change were found to be the most comprehensive of the four boundaries, though the data on climate change come primarily from networks in the developed world, making the global data quality poor. The data on ecosystems and biodiversity are sparse and decentralized, since biodiversity issues are seen as local and contextual. The information on nitrogen and phosphorus is patchy, with academic communities making efforts to develop this database. The information on chemical pollution was also found to be lacking at a global level. However, research-driven networks are present and taking shape in Europe, North America, and East Asia.

Sustainable development goals

The SDGs have come to define approaches towards sustainability by influencing policymakers and businesses around the globe. Here, we try to capture the process of

Table 2 Data availability for planetary boundaries framework (Cornell and Downing 2004)

Planetary boundary (PB)	Basic data about PB	Data networks	Global coordination	Knowledge for action	Issue
Climate change	An immense number of sources	Present, however researcher-driven and global distribution restricted to the US and Europe	National Oceanic and Atmospheric Administration (NAOO)-US, ICSU World Data System (WDS)	IPCC	Quality management is poor, quality of emission inventories is a concern
Ecosystem change and biodiversity loss	Descriptive and context-specific	Present, however researcher-driven and global distribution restricted	Creation of new institutions like Group on Earth observations—biodiversity observation network	Many, intergovernmental and non-governmental	Issue is local- and context- specific
Nitrogen	Scarce and patchy	Global emissions initiative, FAOSTAT, US Geological Society	Nitrogen research community moving towards global coordination (international nitrogen initiative), however lacking in phosphorus	Trying to build with UNFCCC, CBD, etc	Issue is local- and context-specific, geo-politics with concentrated reserves in few countries
Chemical pollution	Fractured databases	Lacking	Lacking	Convention on long-range transboundary air pollution, however, researcher-driven and global distribution restricted to the US and Europe	Lack adequate data, monitoring, and enforcement mechanism

how these SDG indicators were decided on and collected and, consequently, the availability, or lack of data. Table 3 demonstrates around 169 indicators that have been highlighted by the IAEG-SDGs committee. However, as of 20 April 2017, 82, 61, and 84 indicators were classified as Tier I, Tier II, and Tier III, respectively. Among the Tier I indicators, only 75 of the 97 had data that were publicly accessible as of April 2017. According to the analysis by the Centre for Global Development (Dunning and Jared 2016), among the Tier 1 indicators for which data are more readily available, only slightly more than half of the countries have information on 60 of 72 indicators. The Tier 1 indicator data are not readily available in developing countries. Even when the indicators are widely available, like the forest indicator, the definition of forests is different in each country and is based on the potential for timber harvesting rather than the qualitative health of the forest.

However, it is important to note that most of the key indicators of the SDGs only received comments from a few organizations during the IAEG-SDG Open Consultation on Green Indicators, held in Bangkok from Wednesday, 4 November to 7 November 2015, as shown in Table 4. This was true for the indicators related to the environment, such as sustainable consumption and production, climate change, oceans, and terrestrial ecosystem. SDG 13 on climate change received the least comments, followed by the goals for oceans and marine resources (Goal 14) and biodiversity, forests, and desertification of terrestrial ecosystems (Goal 15). The SDGs on health, education, and gender received the most comments. It has been noted that the selection process for the SDG indicators has received a few comments from other organizations, and surprisingly, these indicators now influence every major platform. Most policy decisions globally, nationally, and regionally are framed in terms of these indicators.

Corporate data

Major corporate indicators and sustainability reports are profiled by companies for investors. Fewer indicators target business-to-customers (B2C) and business-to-business (B2B) audiences. The majority of corporate indicators focus on integrated Environmental–Social–Governance (ESG) issues, with significant attention being paid to carbon emissions. Amongst the topical focuses, the stress is on the environment, followed by social and governance indicators. These indicators reveal a relative lack of transparency and lack of attention to social issues. In addition, most company data are presented as an index, followed by rankings and ratings. Figures 1, 2, and 3 present these features.

Table 3 Indicators divided by IAEG-SDG committee

	Tier 1	Tier 2	Tier 3
Definition	The indicator is conceptually clear, with at least 50% of countries having them	The indicator is conceptually clear, but data are not regularly produced by countries	No internationally established methodology
October 2016	159 "green" (generally agreed)	159 "green" (generally agreed)	64 "grey" (needs further discussion)
April 2017	82 (75 available online)	61	84

Table 4 Comments received for each goal at IAEG-SDG open consultation on green indicators

SDG	Comments from organization
Poverty (Goal 1)	121
Hunger (Goal 2)	117
Health (Goal 3)	206
Education (Goal 4)	240
Gender (Goal 5)	189
Sustainable consumption and production (Goal 12)	40
Climate change (Goal 13)	19
Ocean, seas, and marine resources (Goal 14)	26
Terrestrial ecosystem (Goal 15)	40

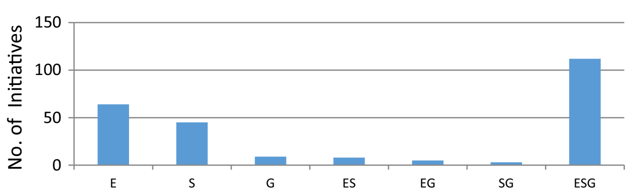


Fig. 1 Corporate data-intensive approaches by focus (figure by the author source—GISR)

Open data

Various aspects of open data-intensive approaches are covered in Figs. 4, 5, 6, 7 and 8. Most open data are currently used by for-profit organizations for operational optimization and for products and services. They are also used, to a lesser extent, by non-profit organizations, for advocacy and research purposes. Most organizations using open data work on ICT, governance, professional geospatial services, and consultancy projects. Other organizations work on thematic areas such as health, finance, education, energy, logistics, etc. Most of the organizations that utilize open data work at the national level, followed by the regional and global levels. However, the majority of open data are generated in North America, Europe, and East Asia.

Big data

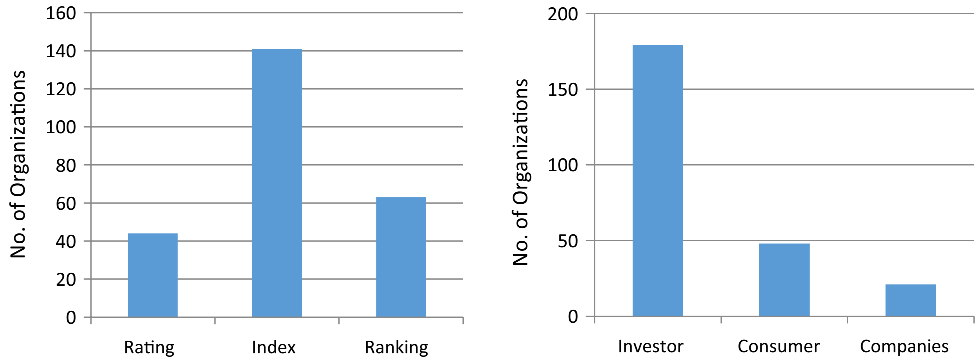
The UN Global Pulse is a United Nations project on the use of big data-intensive approaches for sustainable development. The topical and geographical focuses of these approaches are presented in Figs. 9 and 10, respectively. The geographical focus of these projects is on the developing world, with major projects in Latin America, Asia, and Africa. The major projects focus on the economy, food and agriculture, real-time monitoring, and other areas concerning the SDGs. Some activities of the UN Global Pulse are more streamlined towards monitoring the SDGs than creating more primary data or revealing data on non-economic aspects of the SDGs.

Diverse and disaggregated data-intensive approaches

The current data-intensive approaches for sustainability are diverse in terms of their target audience and focus, and the value base upon which they have been built. For example, the planetary boundary framework has a strong sustainability approach, focuses on environmental issues (Rockström 2009) and currently targets researchers and policy specialists. Information for many boundaries is lacking, and in some cases, even the mechanisms and networks regarding how the various kinds of data should be used and how the information is to be collected are missing.

The SDGs cover environmental, social, and economic themes to be used by business and governments. The SDGs, as a political process, originate from different value sets, as they have emerged from a global discourse on global environmental management (Independent Expert Advisory Group on a Data Revolution for Sustainable Development 2014). Diverse indicators and policy processes stem from

Fig. 2, 3 Corporate data-intensive approaches by type and target audience (figure by the author source—GISR)



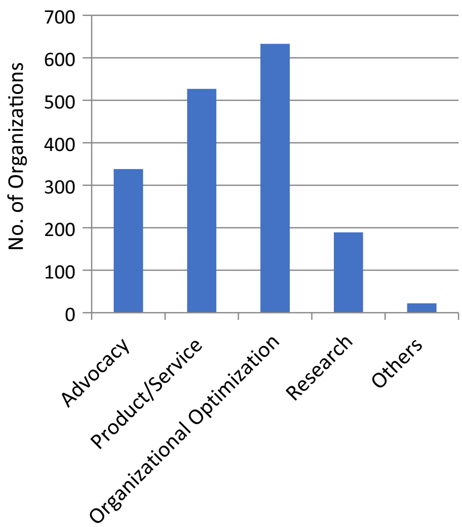


Fig. 4 Open data use by companies (figure by author source—Center for Open Data Enterprise)

the SDGs; however, the SDGs were formed from a narrow base of comments during the consultation process, as highlighted earlier. However, the impacts of the SDGs are felt extensively or are experienced in society, as major policies by both businesses and governments are framed based on SDGs.

The corporate reporting strategies are influenced by the triple bottom approach and by the idea that businesses should move beyond non-financial data and towards incorporating those non-financial data into their corporate decision-making (Milne and Gray 2013). This has influenced the broader corporate data-intensive approaches, albeit with a small difference, as the focus is on investors and environmental issues—mostly carbon, with water coming a distant second—followed by social and then governance issues. Similarly, company-based studies show that big data are not used in companies' sustainability reporting but, rather, is used to increase energy efficiency with its focus being on carbon emissions (Zawistowska 2015). Furthermore, most corporate indicators are generated in the developed world.

Major open data-intensive approaches are seen in the developed world and focus on governance and ICT. On the other hand, corporations focus on using information available from national databases. The dominant focus of big data-intensive approaches stems from ideas on smart governance, which include the use of ICT in governance. The big data project of the United Nations, the UN Global Pulse, focuses on the use of big data for sustainable development. Most projects at the current stage focus on economic growth, monitoring SDGs, and collecting real-time information.

Data-intensive approaches have varying focuses, as summarized in Table 5. The challenges with data-intensive approaches are: the lack of data availability, especially for planetary boundaries; diverse SDG indicators developed from a narrowly viewed base; skewed global representation of open data and big data; the lack of social and economic information collected, especially among the corporate indicators; and the narrow focus on issues related to governance in the current open data and big data approaches. The development of these approaches shows that their current biases are based on the diverse definitions and values that have shaped them. Furthermore,

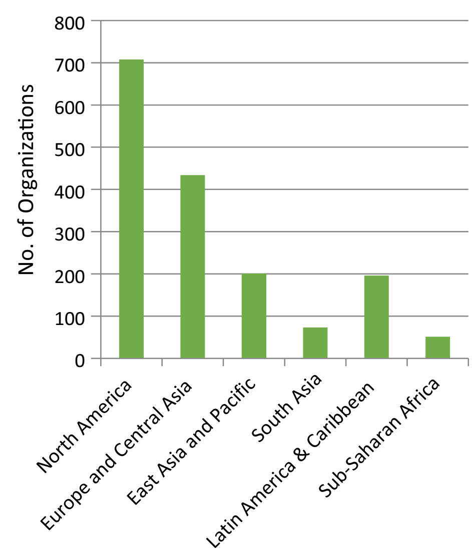


Fig. 5 Geographical distributions of companies using open data (figure by the author source—Center for Open Data Enterprise)

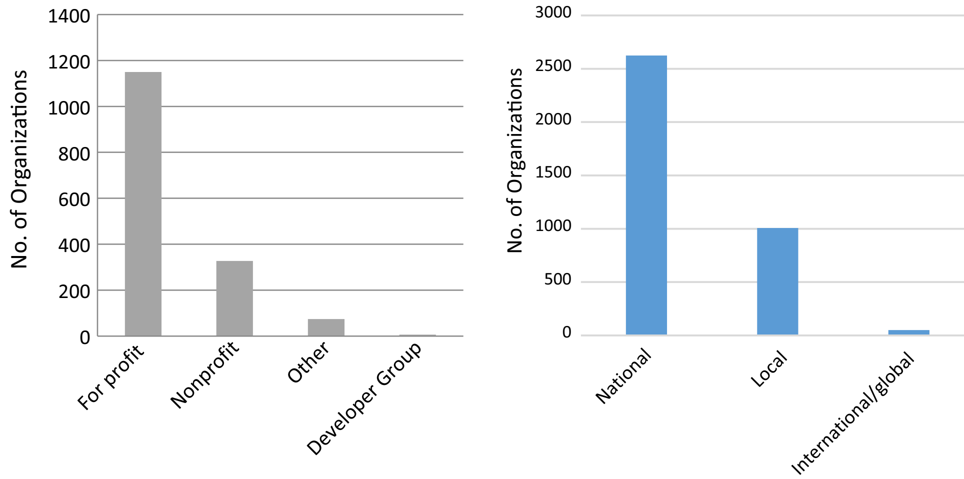


Fig. 6, 7 Characteristics of companies using open data (figure by the author source—Center for Open Data Enterprise)

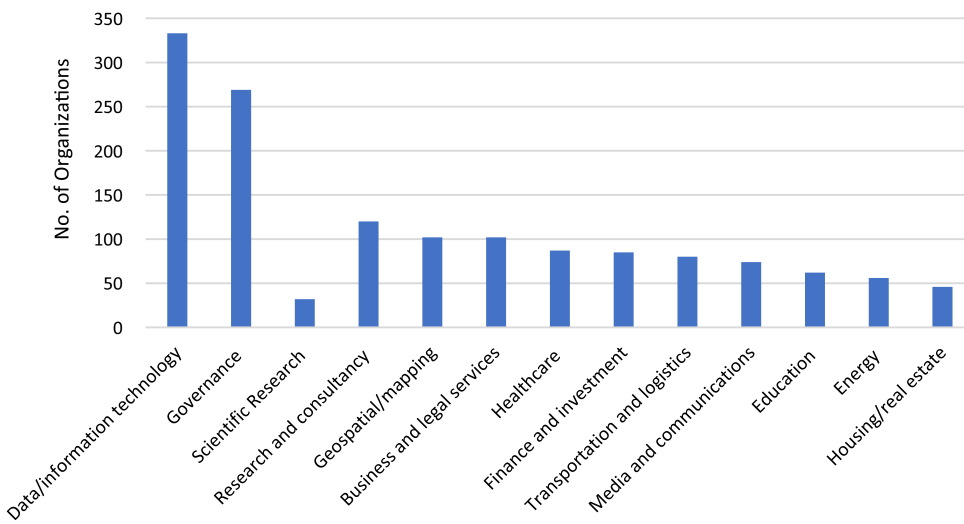


Fig. 8 Fields of companies using open data (figure by the author source—Center for Open Data Enterprise)

structural issues with data-intensive approaches such as data availability, data collection, data interpretation, and data presentation can further compound the issues of data-intensive approaches towards sustainable development. An analysis of the research assumptions of data-intensive approaches is used to deconstruct the impacts towards sustainable development.

Challenges of data-intensive approaches: issues and implications

To understand the subjective nature of data-intensive approaches, we deploy the use of four research approaches to describe various research assumptions in data-intensive approaches:

1. Methodology—the research design, methods, approaches, and procedures used in an investigation that is well planned to find out something (Keeves 1997).
2. Epistemology—as Cooksey and McDonald (2011) put it, looking into what counts as knowledge.
3. Normativity—the values associated with the approach.
4. Ontology—the assumptions we make to believe that something makes sense or is real (Scotland 2012).

The discussion is based on a literature review of general data-intensive approaches; and the examples are specific to sustainability sciences. We delve into the challenges of methodology, epistemology, ontology, and normativity with regard to data-driven approaches, mapping the related issues and implications for data-intensive approaches. In comparison with the purported objectivity of data-intensive approaches, we show the subjectivity of data-intensive approaches in: the data collection and data interpretation; the

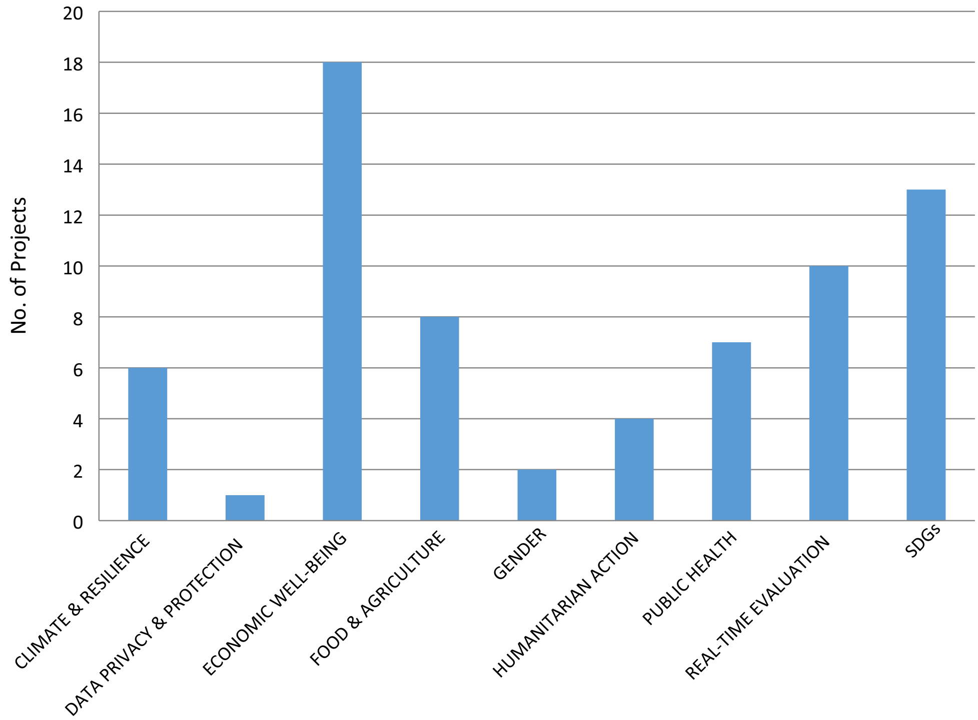


Fig. 9 Big data projects by topical focus (figure by the author source—Global Pulse 2018)

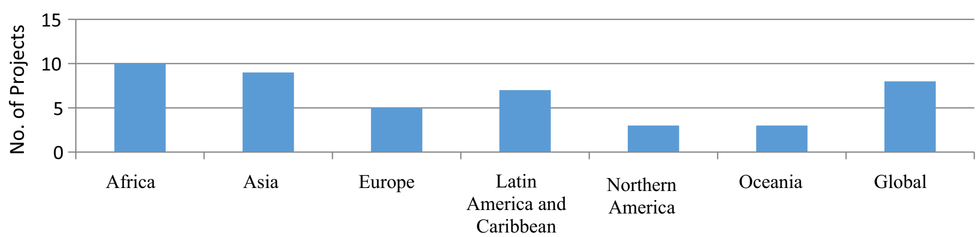


Fig. 10 Big data projects by geographical focus (figure by the author source—Global Pulse 2018)

framework employed; the normative values utilized; and the ontological frames used.

Methodology

In this section, we cover issues surrounding data collection and data interpretation. Collecting data that bring about better results is necessary for monitoring performance and decision-making (Shepherd et al. 2015). However, unlike governments, firms do not collect regular data on crucial uses. Evidence of this can be seen when they are not able to present data when there is regulatory pressure to do so (Davenport 2015). Furthermore, selecting the relevant metrics and learning the necessary skills to handle new forms of data are a major concern (Paddison 2013; Association of Chartered Certified Accountants 2013). In general, although companies claim to use voluntary guidelines, no major criteria are used to make any commitments towards sustainability to influence strategic decision-making (Cort 2015). Shah et al. (2012) have stressed that executives do not manage information well and focus on technology rather than on the information itself. It has been emphasized, however, that technology is not the solution in terms of corporate sustainability (Redman 2013).

Even minor issues with data storage in Excel can change the results of an analysis (Ziemann et al. 2016), although these details regarding data collection remain hidden (Edwards et al. 2011). The non-transparent aspects of data collection need to be revealed (Weller and Kinder-Kurlanda 2015; Olteanu et al. 2016). Human cognitive biases intentionally or unintentionally result in biases with data suiting the researcher's needs being chosen (Parkhurst 2017). It is necessary to address data issues considering the confirmation bias while using data in the context of sustainability science (Samson 2014). For example, unintentional human biases can result in a "sliding baseline". One study on fisheries was evaluated by experts who had fixed their baselines to the times when they first started working, rather than when the fishery was in its untouched state (Pauly 1995). This greatly impacted the way in which the data was collected and how the analysis was carried out. Results that do not

Table 5 Data-intensive approaches and their various characteristics

Data-intensive approaches	Target audience	Geographical area	Focus	Values
Planetary boundaries	Predominantly researchers and policy makers	Developed world	Climate change, biodiversity loss, biogeochemical, ocean acidification, land use, freshwater, ozone depletion, atmospheric aerosols, and chemical pollution	Strong sustainability, limits to growth
SDG	Business and government	World	Greater emphasis on social and economic issues, than the environment	Sustainable development, Rio declaration
Corporate indicators	Focusing on investors	Developed world	Environment followed by social. Not much information on governance	The triple bottom approach, UNEP FI—business should move beyond financial data
Open data (excluding SDGs and PBs)	Public	Developed world	IT and governance, with a focus on national data	Facilitate government transparency, accountability and public participation
Big data—UN Global Pulse	United Nations	Developing world	Economic, SDGs, and real-time information	Information technology, smart city, smart governments

incorporate historical information can be unreliable as the wrong baselines are extrapolated (Guzman 2015).

Data interpretation plays an important role in academic research and data from the same source can be interpreted in various ways. One single peer-reviewed publication exemplified that 29 analyses and interpretations were derived from one dataset (Silberzahn and Uhlmann 2015). The method used in the study highlighted that players play an important role in the interpretation of data and can lead to multiple conclusions. Cases of statistical and objective illusion reveal a need to expose partisanship associated with data interpretation in statistical analysis (Berger and Berry 1988).

"Raw data" are an oxymoron and there is a need to focus on the weakness of data recording practices and data interpretation methods, and the cognitive biases at play during the methodological process (Gitelman 2013). The detailed processes of data collection and interpretation are fraught with these hidden frameworks, or lack thereof, which can lead to subjectivity.

Even among the data-intensive approaches, there is a tendency to handle the methodological issues regarding old and new approaches with confusion; for example, James et al. (2013) and Varian (2014) emphasize the tension between traditional causal inferences (econometrics) and correlation-based prediction (machine learning and RCTs). Other scholars have emphasized experiments with data as an alternative to prediction and the statistical analysis of data (Brown et al. 2011). However, the danger with these approaches is misperceiving correlation as causation and finding "misleading" patterns in the data (McAfee et al. 2012). The subjective process of data collection and data interpretation is not adequately emphasized.

It is necessary to use data-intensive approach analysis tools that focus on novel measurement, research designs, and statistical learning to embrace heterogeneity in research. Understanding the challenges of data access, data management, and computation, and asking the right questions are critical aspects of better research methodologies. Data are not to be considered sacrosanct, as a narrow focus on data can lead to the development of diverse conclusions depending on the interpretive framework that is implicitly or explicitly used, thus leading to relativism. Proponents of data-intensive approaches have to emphasize the importance of subjectivity in the process of interpretation to be more proactive towards attaining objective scientific results.

Epistemology

Frameworks, concepts, and tools need to be used to overcome the diverse framings of different datasets. However, there are a multitude of sustainability assessments that have been developed based on varying subjects and specific framings (Bond et al. 2012; Bond and Morrison-Saunders 2011). Several ecosystem processes have not yet been identified and efforts to link individual, regional, national, and corporate levels with these planetary boundaries' concepts are ongoing (Nykvis 2013; Dao et al. 2015; Cole et al. 2014; Rockstrom et al. 2009). Similarly, understanding the footprint approaches—water footprint and carbon footprint—together and applying them to the national, corporate, or product levels is still elusive (Galli et al. 2011). These issues present an unsurmountable challenge to the current data-intensive approaches.

The varying framings need to be integrated to present a well-rounded view of sustainability and this poses a significant challenge (Gasparatos et al. 2012). There has not been much research on how to integrate the diverse tools and their different values (Gasparatos 2010; TEEB 2010). There needs to be a tool-based integration of the LCA framework—an absolute measure—with the planetary boundaries concept—a relative measure (Bjørn et al. 2015; Bjorn and Hauschild 2015). The integration of the LCA and planetary boundaries framing of a firm's environmental sustainability would require a decoupling of the environmental impacts from their growth in production by not increasing their portion of the safe operating space (Kim and Kara 2014). Companies explicitly do not make any reference to safe operating space in their frameworks, and as they do not refer to any ecological limits, information about "sustainability" continues to remain hidden (Bjørn et al. 2017).

In comparison with a framework-based indicator conceptualization, experiment-based approaches founded on interventions are preferred for their ease-of-use, as they lack the use of theory. It has been suggested that evidence-based knowledge production should use a hierarchy of evidence (placed in descending order): system reviews of RCTS, RCTs, observational studies, mechanistic explanations, expert opinions, and anecdotal experience (Blunt 2015; Greenhalgh 1997). This hierarchy privileges certain modes of knowledge production, which is not beneficial for sustainability science research, as it has ambitions to be transdisciplinary.

Data "without a theory" can be taken to work as a control theory when the certainty of the research process is well established and when the research involves measurable outcomes or a simple policy, where decision makers have enough reaction time. Furthermore, both "evidence" from controlled experiments and indicators pertain to measuring or counting, and not exploring or knowing the meaning of that measure. The issues related to sustainable development are complex, however, and the solutions need to be participatory and encompass many stakeholders.

Data-intensive approaches will not work well with complex issues that have long-term horizons, as is the case with most sustainability challenges. Sustainability challenges have long time horizons, and are complex and uncertain (Small and Sampson 2014). Furthermore, social scientist Justin Grimmer (2015) suggests that big data should be complemented with the knowledge of social scientists when dealing with humans. When this is done in data-intensive research, it can bring scientifically valuable knowledge to data-based approaches (Starr 2014). The sense-making skills of humans are crucial in the era of "human–machine partnerships", one application of big data (Davenport 2016). Theory, sense-making skills, and contextual approaches bring depth to understanding context-based awareness of research (Crawford 2013). This calls for frameworks when using and interpreting data (Mazzocchi 2015).

Furthermore, the use of evidence has been found to be inadequate by philosophers of science, who have pointed out that such studies can devalue other forms of evidence or knowledge such as causality-based basic sciences and the tacit knowledge that accumulates with practice and experience (Anjum et al. 2018; Greenhalgh et al. 2014). It is crucial for sustainable development that information from diverse knowledge bases are integrated. Excessive reliance on data-intensive approaches, however, can create a hierarchy of evidence, which can be detrimental to the practice of sustainability and the integration of knowledge. Thus, researchers should also focus on how this knowledge is analyzed, integrated, and used in decision-making by emphasizing frameworks and theory, which are the building blocks of scientific knowledge production. Integration frameworks that can represent ecological boundaries and human values, and incorporate participatory processes are necessary to legitimize knowledge production and make the data-intensive approaches relevant for decision-making.

Normativity

The problems of sustainability are fundamentally value laden, and hence political. However, not all data-intensive approaches reflect these normative assumptions. It may be summarized that eco-efficiency, eco-effectiveness, and social justice are three major elements of sustainability (Gray and Bebbington 2005). Planetary boundaries provide the safe operating space concept that can also be further broken down into eco-efficiency and eco-effectiveness. Mihelcic et al. (2003) noted:

> "(eco-efficiencies) alone are not sufficient to achieve sustainability, because even systems with efficient material and energy use can overwhelm the carrying capacity of a region [eco-effectiveness] or lead to other socially unacceptable outcomes".

Though eco-efficiency is often discussed, the concept of eco-effectiveness should be added to data-intensive approaches, as it relates directly to ecological capacity (Bjørn and Hauschild 2015). For example, while cost–benefit analysis (CBA) is popular in sustainability science research, it carries the notions of a "utilitarian approach" espoused by Bentham (Sandel 2010). It is viewed as a representation of monetary values, claiming that higher utility is achieved when there is a higher monetary output, such that the monetary focus disregards normative assumptions of eco-efficiency, eco-effectiveness, and social justice.

Furthermore, these approaches do not adequately emphasize the importance of inter-generational equity, intra-generational equity, and the safe operating space concept. Issues of social justice and ethics are also not well studied within sustainability science (Nelson and Vucetich

2012). For example, concerns for social equity are missing from public policy measures of cities' sustainability indices (Wachsmuth et al. 2016). Social justice issues are relatively difficult to engage in given the claimed objective stance of data-intensive approaches (Newell and Frynas 2007). This corroborates what we see in our findings, that the importance of social justice has not gained much emphasis in either enterprise reports or policy making. Research on eco-effectiveness and social justice should, therefore, be further improved.

Both Campbell's law (Campbell 1979) and Goodhart's law (Strathern 1997) stress the pitfalls of primarily using indicators and data in policy making and research. Quantitative measures are important for decision-making; however, excessive reliance can allow them to be hijacked for political purposes and can lead to the control of knowledge by certain actors. As stressed earlier, data-intensive approaches should be not understood as value neutral (Redman 2013; Goldenberg 2005; Parkhurst 2017), as they are derived from data collection, data interpretation, and framework methodologies that have specific values and assumptions that are not adequately revealed. Furthermore, data-intensive approaches capture tactical issues, but fall short of capturing issues of strategy. These strategies to address the challenges of sustainable development involve questions like, "What is to be sustained?", and "What is to be developed?" Asking the right questions requires input from local communities and experts, and data-intensive approaches focusing on service delivery and monitoring might not provide tactical operation knowledge.

Even if the results of data-intensive approaches are extrapolated, as Davidian and Louis (2012) put it, there is a need for society to "be able to interpret increasingly complex information and recognize both the benefits and pitfalls of statistical analysis" (Busch 2014). However, political deliberation of knowledge (Parkhurst 2017; Pearce 2014) is undergone when data-intensive approaches imply consensus for a public policy move. Data-intensive approaches should provide room for diverse evidence from different knowledge bases, which can allow for "disagreement and consensus" to explore the tension between technocracy and democracy (Horton 2015; Parkhurst 2017), as the centralized use of data-intensive approaches may inhibit innovation and the democratic needs of society (Ratti and Helbing 2016). Data-intensive approaches should be open and participative; this can lead to insight and action through contestation and discussion in society.

An over-reliance on data-intensive approaches stalls the participation of diverse stakeholders in decision-making, as "evidence" and "indicators" from data-intensive approaches are presented as objective tools, which are regarded as better than subjective consensus building or democratic processes. Under these circumstances, data-intensive approaches do not provide much space for discussion and deliberation.

Ontology

Ontological frames dictate how conceptual frameworks are chosen for indicator development, data collection, and data interpretation. Ontology relates to a template or a frame that stratifies the nature of reality and includes the object, process, particular, individual, whole, part, event, property, quality, state, etc. (Poli 1999). Indicators, indices, and ratings are based on frames regarding reality and tied to ontological assumptions. Indicators should be able to measure the impact of the progress made in achieving sustainability, and additionally tie information to the local context, as frames and indicators are co-produced. This allows for their subjective and objective information to be critically understood (Boulanger 2014).

Machine learning-based approaches work well in revealing patterns from a large sample. However, they lack the ability to reveal information about a singular event or object. The results of controlled experiments with statistically significant results still produce marginal advantages in practice. The results from these approaches overemphasize rules while having a poor fit with issues of multiple causation (Greenhalgh et al. 2014). This requires understanding causal structures or having supporting information on intervening variables, which require theory or other empirical means (Cartwright and Hardie 2012).

The emphasis on "causation" is a hallmark of scientific knowledge production. The ontological focus of sustainability science as system-based research (Clark 2007; Spangenber 2011), where sustainability challenges are wicked problems (Rittel and Webber 1972), requires knowledge and evidence that reflect the holism and complexity of the real world, and causality between variables, which might be captured from studies other than data-intensive approaches.

Controlled experiments and indicator driven approaches are reductive, and machine learning-based approaches are inductive. The ontological underpinnings of an evidence-based approach rest on a narrow definition of "evidence", reflecting reductionism, universalizing the results, capturing frequency, and representing probability, or a homogenization of the unit of analysis. Machine-based approaches are based on Baconian induction, moving from specific to general. The ontological assumptions of sustainability science and data-intensive approaches are divergent, and their use without reflection can lead to scientific imperialism of data-intensive approaches or ontological relativism, which is against scientific knowledge production (Persson et al. 2018a, b). Data-intensive approaches capture a part of reality, while the challenges of sustainability require a larger knowledge base.

Implications for future research: challenges and way forward

The major challenges in the data-intensive sustainability approaches identified were: the lack of data availability, diverse indicators developed from a narrow base, and skewed global representation. Furthermore, the subjectivity in the structures of data-intensive approaches were discussed in regard to: data collection and data interpretation, the framework employed, the normative values, and the ontological frames employed.

However, if data-intensive approaches can overcome their structural issues using scientific approaches and community-based approaches, they will offer ways to address sustainability. In this section, we first map the current challenges to interdisciplinarity and community-based approaches and then elucidate possible solutions that researchers may employ. In addition, we also raise the important issues pertaining to the long-term sustainability of ICT approaches, i.e., the environmental impact of ICT-based data-intensive approaches.

Data and academia—interdisciplinarity

We emphasize the challenges that researchers face in transitioning to interdisciplinarity and recommend focusing on the potential of mechanisms and explanations to complement data-intensive research.

There is a longstanding discussion among philosophers on unification and pluralism relating to how knowledge is produced. Grantham (2004) called for either the theoretical or practical unification of fields, stressing that the "theories and/or methods of a field can guide the generation of new hypotheses in a neighbouring field". In its efforts, sustainability science has been called upon to take on interdisciplinarity, which Jantsch (1970) characterizes as "coordination by the higher level concept". A broader vision of integration was articulated by the National Research Council (2004) as "information, data, techniques, tools, perspectives, concepts, and/or theories from two or more disciplines or bodies of specialized knowledge".

A current challenge in academic research, as Smaldino and McElreath (2016) stress, is that bibliometric data make it easy for academics to focus on their productivity and impact factor. This forces researchers to fit their research into the bounds of high-impact journals and thus weakens research outcomes (Werner 2015; Sarewitz 2016). Richard Horton, editor of The Lancet, states, "Part of the problem is that no one is incentivised to be right" (Horton 2015). Furthermore, not all disciplines are open to interdisciplinary research. Economists, for example, are not keen on interdisciplinary research, while sociologists, political scientists, psychologists, and historians are more open to it (Colander 2005; Fourcade et al. 2015). The abundance of data in one field should not divert the attention of researchers to data rich topics, leading to the neglect of pertinent interdisciplinary research questions. Though data rich topics might create new information and expand a knowledge base, and the knowledge created may not have a significant impact in addressing sustainability.

Forscher (1963) worried that the academic culture was interested in pieces of knowledge, which he termed "bricks", but was not as concerned in placing the bricks as a connected whole. This lack of a coherent whole has also been recently articulated as an issue in sustainability research (Hoffman 2015).

Thoren and Persson (2013) proposed the use of problem-feeding and defined it as when "…questions and problems are transferred to sciences better capable of researching them". One major obstacle in problem-feeding, however, is the ontology of terminologies used in different fields. For example, the term "cause" has different meanings in law and science. The use of shared terminologies or interactional expertise can facilitate discussion between researchers and practioners to link data through frameworks and methods, leading to interdisciplinarity.

Terms such as "flexibility" and "resilience" have been proposed for their ability in allowing "conceptual coordination" between disciplines or "interactional expertise" to connect themes with sustainability science (Asokan et al. 2017). The use of mechanisms is an interesting area of research, especially for sustainability, as they have been used in biology, neuroscience, and psychology (Craver and Tabery 2017). Their functionality in social science has also been emphasized (Lucas 1988; Elster 1989; Hedström and Ylikoski 2010; Deaton 2010). Furthermore, available data can open new directions in providing causality-based knowledge derived from data-intensive mechanism experiments (Ludwig et al. 2011). A scale-based approach has also been suggested, which provides a set of scales for structuring sustainability science data across disciplinary boundaries and a range of sustainability topics (Asokan et al. 2019). Mechanism experiments, explanations, mechanisms, and concepts for coordination provide a fertile approach to bridge data from the multiple fields available, reflecting a large base of knowledge, and clarifying the subjectivities to make data-intensive approaches "objective". The data proponents should reflect on and critically analyze the philosophical issues raised in the previous section, to make all of the subjectivities of diverse data-intensive approaches explicit to promote a better understanding.

Enterprises and policy making—community-based approach

We propose that a community-based data-intensive approach needs to reflect the local context, global ecosystem, and diverse values. Contextual approaches, experimental approaches, and quantitative storytelling approaches can provide a community-based approach orientation to the existing data-intensive approaches.

A community-based approach would incorporate a diverse knowledge base that includes the scientific knowledge of an ecosystem with accounted-for personal values, power dynamics, institutional play, and philosophical values. The "hierarchy of evidence" places data and data-intensive approaches on a higher pedestal, where politics, personal values, power dynamics, institutional play, and philosophical values are sidestepped. In comparison with knowledge that is formed from a narrow base, such as the current data-intensive approaches, a community-based approach would progressively facilitate decision-making towards sustainability by incorporating the human values, local contexts, and global ecosystem processes.

There is a tendency among some researchers to legitimize certain values as playing a "positive or negative" role in the environment (Grenier 1998; Sundar 2000; Agrawal 1995). For example, indigenous knowledge is claimed to have positive value, though this is not always the case. However, the knowledge representing values from experience should be empirically tested, as it can reflect confirmation bias that cannot be generalized to a larger population and provide an incorrect representation of a dynamic fast-changing ecosystem (Munro 2014). However, local values and power dynamics could be utilized to inform research agendas for data-intensive approaches based on practical experience rather than wholesale integration of indigenous knowledge (Persson et al. 2018b).

In addition to values, the sustainability performance of enterprises and entities need to be "systematic, structured, and integrated" (Searcy 2016), such that they are connected to the local context and global ecosystem process. To do so, enterprise sustainability should ensure "scientifically informed standards" that incorporate the ecosystems' "land use and land conversion, clean air (including greenhouse gas emissions), availability and quality of freshwater, degradation of coastal and marine habitats, and sustainable use of renewable resources such as soil, timber, and fisheries" (Kareiva et al. 2015). There is a crucial need to understand these interactions and their implications for and between enterprises and policy making (Ruggie 2017).

We suggest three community-based data-intensive approaches to capture the local context, global ecosystem, and diverse value sets to inform subjectivities and thereby make data-intensive approaches "objective".

Context-based indicators should capture normative aspects such as social justice, eco-efficiency, and eco-effectiveness while also capturing local contexts. Context-based approaches contrast with the absolute and relative indicators in use today. Context-based indicators are defined as valid data placed appropriately in context (Slater and Zwat 2015). Some enterprises are developing context-based greenhouse gas emission indicators such as the CSI metric and C-FACT (Autodesk Inc 2016; BTplc 2016; Sustainability Context Group 2012). The chief discussions revolve around how appropriate references to and allocations of global "safe operating space" should be developed for these contextualized indicators (Raupach et al. 2014).

New experimental approaches based on local data, e.g., smart city projects or satellite-based community projects, can help in steering new forms of data-intensive approaches. Living labs or niche experiments, for example, allow experimentation at the local governance scale (Caniglia et al. 2017). These experiments can be either Darwinian experiments or generative experiments and can reveal knowledge for innovation through variations and solutions through iterative and adaptive processes, respectively (Ansell and Bartenberger 2016). These experimental approaches, in comparison with controlled experiments, could directly focus on sustainability problems or sustainability solutions (Bulkeley and Castán Broto 2013; Evans and Karvonen 2011).

The results from data-intensive approaches can be communicated through quantitative storytelling (QST), emphasizing diverse quantitative narratives to the public (Giampietro et al. 2014). QST emphasizes a rough quantitative appraisal of multiple frames rather than the quantification of a single frame, such that the frames are tested for their feasibility, viability, and desirability. Importantly, the data-intensive approaches should look at presenting the information and data to the public in an easily accessible manner. Adopting these steps could create a community-based data-intensive approach that incorporates values, local contexts, and relevance for governance.

Environmental impact of data-intensive approaches

The nature of "internet use" is evolving and forms of data growth are emerging that are more disconnected from human activity and time use than ever before. This suggests that although there may well be limits, in principle, to some forms of growth, the total data traffic seems likely to continue expanding. This calls for careful attention to the nature of the trends involved as a basis for intentionally building limits to internet use, data storage, and data size before the considerable levels of internet electricity demand become problematic (Hazas et al. 2016).

Conclusion

In this paper, the current data-intensive approaches are defined as an amalgamation of the old data approaches, which were dependent on surveys and environmental monitoring networks, with new data approaches from new ICT and satellite technologies. An extensive literature review was conducted, and the implications of data-intensive approaches were discussed with a particular emphasis on sustainability. Four research approaches—methodology, epistemology, normativity, and ontology—were deployed to describe various research assumptions in data-intensive approaches. The subjectivities of data-intensive approaches during data collection, as well as the frameworks, normative values and ontological frames employed were also elaborated. Possible solutions to these challenges were explored, which researchers could employ to enhance the future use of data-intensive approaches by tapping into interdisciplinarity and improving community-based approaches.

Acknowledgements We acknowledge Monbukagakusho (MEXT), the Government of Japan, for their financial assistance during the doctoral study. We would like to thank the anonymous reviewers and editors for providing new insights and broadening the scope of the paper. We would like to especially thank Nikole Roland and Clare Sandford for providing editing and proofreading assistance.

References

Agrawal A (1995) Dismantling the divide between indigenous and scientific knowledge. Dev Change 26(3):413–439
Anjum RL, Copeland S, Rocca E (2018) Medical scientists and philosophers worldwide appeal to EBM to expand the notion of 'evidence'. BMJ Evid Based Med **(bmjebm-2018)**
Ansell CK, Bartenberger M (2016) Varieties of experimentalism. Ecol Econ 130:64–73
Asokan VA, Yarime M, Esteban M (2017) Introducing flexibility to complex, resilient socio-ecological systems: a comparative analysis of economics, flexible manufacturing systems, evolutionary biology, and supply chain management. Sustainability 9(7):1091
Asokan VA, Yarime M, Onuki M (2019) Bridging practices, institutions, and landscapes through a scale-based approach for research and practice: a case study of a business association in South India. Ecol Econ 160:240–250
Association of Chartered Certified Accountants (2013) Big data: its power and perils. http://www.accaglobal.com/bigdata
Autodesk Inc (2016) Autodesk Inc. http://sustainability.autodesk.com/available-solutions/c-fact/. Retrieved 10 Jan 2016, from Corporate and city sustainability —C-FACT
Beck S (2011) Moving beyond the linear model of expertise? IPCC and the test of adaptation. Reg Environ Change 11(2):297–306
Berger JO, Berry DA (1988) Statistical analysis and the illusion of objectivity. Am Sci 76(2):159–165
Bjørn A, Hauschild MZ (2015) Introducing carrying capacity-based normalisation in LCA: framework and development of references at midpoint level. Int J Life Cycle Assess 20(7):1005–1018
Bjørn A, Bey N, Georg S, Røpke I, Hauschild MZ (2017) Is Earth recognized as a finite system incorporate responsibility reporting? J Clean Prod 163:106–117. https://doi.org/10.1016/j.jclepro.2015.12.095
Blunt C (2015) Hierarchies of evidence in evidence-based medicine. PhD thesis, The London School of Economics and Political Science (LSE)
Böhringer C, Jochemc PE (2016) Measuring the immeasurable—a survey of sustainability indices. ftp://ftp.zew.de/pub/zew-docs/dp/dp06073.pdf. Retrieved Feb 2016 from ZEW—discussion paper no. 06-073
Bond AJ, Morrison-Saunders A (2011) Re-evaluating sustainability assessment: aligning the vision and the practice. Environ Impact Assess Rev 31(1):1–7
Bond AJ, Morrison-Saunders A, Pope J (2012) Sustainability assessment: the state of the art. Impact Assess Project Appraisal 30(1):53–62
Boulanger PM (2014) Elements for a comprehensive assessment of public indicators. In: Report procured by the European Commission-Joint Research Centre, Econometrics and Applied Statistics (DDG 01)
Brown B, Chui M, Manyika J (2011) Are you ready for the era of 'big data'. McKinsey Q 4(1):24–35
BTplc (2016) Climate stabilisation intensity targets. https://www.btplc.com/Betterfuture/NetGood/OurNetGoodgoal/OurCSIMethodology/CSI_Methodology.pdf. Retrieved Jan 2016 from BTplc
Bulkeley H, Castán Broto V (2013) Government by experiment? Global cities and the governing of climate change. Trans Inst Br Geogr 38(3):361–375
Busch L (2014) Big data, big questions| a dozen ways to get lost in translation: inherent challenges in large scale data sets. Int J Commun 8:18
Bjørn A, Diamond M, Owsianiak M, Verzat B, Hauschild, MZ (2015) Strengthening the link between life cycle assessment and indicators for absolute sustainability to support development within planetary boundaries
Callebaut W (2012) Scientific perspectivism: a philosopher of science's response to the challenge of big data biology. Stud Hist Philos Sci Part C Stud Hist Philos Biol Biomed Sci 43(1):69–80
Campbell DT (1979) Assessing the impact of planned social change. Eval Program Plan 2(1):67–90
Caniglia G, Schäpke N, Lang DJ, Abson DJ, Luederitz C, Wiek A, Laubichler MD, Gralla F, von Wehrden H (2017) Experiments and evidence in sustainability science: a typology. J Clean Prod 169:39–47
Carbon Disclosure Project (2016) Reporting to CDP. https://www.cdp.net/en-US/Respond/Pages/companies.aspx. Retrieved Jan 2016 from Carbon Disclosure Project
Cartwright N, Hardie J (2012) Evidence-based policy: a practical guide to doing it better. Oxford University Press, Oxford
Cash DW, Clark W, Alcock F, Dickson NM, Eckley N, Jaeger J (2004) Salience, credibility, legitimacy and boundaries: linking research, assessment and decision making. Global Environmental Assessment Project, Harvard University (http://www.ksg.harvard.edu/gea)
Center for Open Data Enterprise (2016) The open data impact map. http://opendataenterprise.org/index.html
Center for Strategic and International Studies (CSIS), Japan International Cooperation Agency (JICA) Research Institute (2017) Harnessing the data revolution to achieve the sustainable development goals: enabling frogs to leap. Rowman & Littlefield, Lanham
Ceres (2016) Global initiative for sustainability ratings (GISR). http://ratesustainability.org/hub/index.php/search/
Ceres and Sustainalytics (2012) The road to 2020: corporate progress on the Ceres roadmap for sustainability. Ceres and Sustainalytics, Boston

Clark WC (2007) Sustainability science: a room of its own. PNAS 104(6):1737–1738. https://doi.org/10.1073/pnas.0611291104
Clark WC, Van Kerkhoff L, Lebel L, Gallopin GC (2016) Crafting usable knowledge for sustainable development. Proc Natl Acad Sci 113(17):4570–4578
Climate Disclosure Standards Board (2016) CDSB framework for reporting environmental information & natural capital. http://www.cdsb.net/what-we-do/reporting-frameworks/environmental-information-natural-capital. Retrieved 10 Jan 2016 from Climate Disclosure Standards Board
Colander D (2005) The making of an economist redux. J Econ Perspect 19(1):175–198
Cole MJ, Bailey RM, New MG (2014) Tracking sustainable development with a national barometer for South Africa using a downscaled "safe and just space" framework. Proc Natl Acad Sci 111(42):E4399–E4408
Cooksey R, McDonald G (2011) Surviving and thriving in postgraduate research. VIC, Tilde University Press, Prahran
Cornell S, Downing A (2004) Environment, absolute? The quality infrastructure of the planetary boundaries. Discussion paper. Physikalisch-Technische Bundesanstalt, Braunschweig
Cort T (2015) Vanishing materiality in sustainability reporting. Environmental leader. https://www.environmentalleader.com/2015/09/vanishing-materiality-in-sustainability-reporting/
Craver C, Tabery J (2017) Mechanisms in Science. In: Zalta EN (ed) The Stanford encyclopedia of philosophy. Springer, New York. https://plato.stanford.edu/archives/spr2017/entries/science-mechanisms/
Crawford K (2013) The hidden biases in big data. Harv Bus Rev 1
Crawford K, Schultz J (2014) Big data and due process: toward a framework to redress predictive privacy harms. BCL Rev 55:93
Dalin C, Wada Y, Kastner T, Puma MJ (2017) Groundwater depletion embedded in international food trade. Nature 543(7647):700
Dao H, Peduzzi P, Chatenoux B, Bono AD, Schwarzer S, Friot D (2015) Environmental limits and Swiss footprints based on planetary boundaries. Swiss Federal Office for the Environment (FOEN)
Davenport TH (2009) How to design smart business experiments. Strateg Direct 25(8)
Davenport TH (2015) It's way too late not to know where your data is. Wall Street J. http://www.tomdavenport.com/wp-content/uploads/Its-Way-Too-Late-Not-to-Know-Where-Your-Data-Is.pdf
Davenport TH (2016) Rise of the strategy machines. MIT Sloan Manag Rev. Magazine: Fall 2016 issue. http://sloanreview.mit.edu/article/rise-of-the-strategy-machines/?utm_content=buffer20039&utm_medium=social&utm_source=facebook.com&utm_campaign=buffer
Davidian M, Louis TA (2012) Why statistics? Science 336(6077):1–12. https://doi.org/10.1126/science.1218685
Deaton A (2010) Understanding the mechanisms of economic development. J Econ Perspect 24(3):3–16
Derman E (2011) Models behaving badly: why confusing illusion with reality can lead to disaster, on wall street and in life. Simon and Schuster, New York
DJSI annual review 2015 (2015) http://www.sustainability-indices.com/. Retrieved 7 Jan 2016 from ROBECOSAM
Dunning C, Jared K (2016) 'What SDGs can we track now?' Center for Global Development. https://www.cgdev.org/blog/what-sdgs-can-we-track-now. Accessed 16 Sep 2018
Eccles RG, Ioannou I, Serafeim G (2014) The impact of corporate sustainability on organizational processes and performance. Manag Sci 60(11):2835–2857
Edwards PN, Mayernik MS, Batcheller AL, Bowker GC, Borgman CL (2011) Science friction: data, metadata, and collaboration. Soc Stud Sci 41(5):667–690
Elster J (1989) Nuts and bolts for the social sciences. Cambridge University Press, Cambridge
Fairfield J, Shtein H (2014) Big data, big problems: emerging issues in the ethics of data science and journalism. J Mass Media Ethics 29(1):38–51
Fitz-Gibbon CT (1990) Performance indicators, vol 2. Multilingual Matters, Bristol
Forscher BK (1963) Chaos in the brickyard. Science 142(3590):339
Fourcade M, Ollion E, Algan Y (2015) The superiority of economists. J Econ Perspect 29(1):89–114. https://doi.org/10.1257/jep.29.1.89
FTSE4Good Index Series (2016) http://www.ftse.com/products/indices/FTSE4Good. Retrieved 8 Jan 2016 from FTSE Russell
Galli A, Wiedmann TO, Ercin E, Knoblauch D, Ewing BR, Giljum S (2011) Integrating ecological, carbon and water footprint: defining the footprint family and its application in tracking human pressure on the planet. Open-EU
Gallopin GC (1996) Environmental and sustainability indicators and the concept of situational indicators. A systems approach. Environ Model Assess 1(3):101–117
Gasparatos A (2010) Embedded value systems in sustainability assessment tools and their implications. J Environ Manag 91(8):1613–1622
Gasparatos A, Scolobig A (2012) Choosing the most appropriate sustainability assessment tool. Ecol Econ 80:1–7
Giampietro M, Aspinall RJ, Ramos-Martin J, Bukkens SGF (2014) Resource accounting for sustainability assessment: the nexus between energy, food, water and land use. Taylor & Francis (Routledge explorations in sustainability and governance)
Gilman K, Schulschenk J (2013) Sustainability accounting standards board by Ernst & Young
Gitelman L (2013) Raw data is an oxymoron. MIT Press, New York
Global Pulse (2018) Global pulse. https://www.unglobalpulse.org/jakarta
Global Reporting Initiative (2015) G4-GRI sustainability reporting guidelines. https://www.globalreporting.org/resourcelibrary/GRIG4-Part1-Reporting-Principles-and-Standard-Disclosures.pdf. Retrieved Jan 2016 from Global Reporting Initiative
Goldenberg MJ (2005) Evidence-based ethics? On evidence-based practice and the" empirical turn" from normative bioethics. BMC Med Ethics 6(1):11
Grantham TA (2004) Conceptualizing the (dis) unity of science. Philos Sci 71(2):133–155
Gray R, Bebbington J (2005) Corporate sustainability, accountability and the pursuit of the impossible dream. In: Handbook of sustainable development, pp 376–394
Greenhalgh T (1997) How to read a paper: getting your bearings (deciding what the paper is about). BMJ 315(7102):243–246
Greenhalgh T, Howick J, Maskrey N (2014) Evidence based medicine: a movement in crisis? BMJ 348:g3725
Grenier L (1998) Working with indigenous knowledge: a guide for researchers. IDRC, Ottawa
Grimmer J (2015) We are all social scientists now: how big data, machine learning, and causal inference work together. PS Polit Sci Polit 48(1):80–83
Gudmundsson H (2003) The policy use of environmental indicators-learning from evaluation research. J Transdiscip Environ Stud 2(2):1–12
Guzman K (2015) A good socially responsible investment (SRI) fund is hard to find. Yale School of Management. https://som.yale.edu/blog/2015/09/good-socially-responsible-investment-sri-fund-hard-find. Accessed 14 Sep 2015
Hazas M, Morley J, Bates O, Friday A (2016) Are there limits to growth in data traffic?: on time use, data generation and speed. In: Proceedings of the second workshop on computing within limits. ACM, New York, p. 14

Head BW (2010) Reconsidering evidence-based policy: key issues and challenges
Heath T, Bizer C (2011) Linked data: evolving the web into a global data space. Synth Lect Semant Web Theory Technol 1(1):1–136
Hedström P, Ylikoski P (2010) Causal mechanisms in the social sciences. Annu Rev Sociol 36
Hoffman AJ (2015) How culture shapes the climate change debate. Stanford University Press, Stanford
Horton R (2015) Offline: what is medicine's 5 sigma. Lancet 385(9976):1380
House of Commons, Parliamentary Office of Science and Technology (2011) Evidence-based conservation: Post Note 379. Parliamentary Office of Science and Technology, Houses of Parliament, London
IAEG-SDG (2015) IAEG-SDG open consultation on green indicators. Bangkok. https://unstats.un.org/sdgs/iaeg-sdgs/open-consultation-2
Imhoff ML, Bounoua L, Ricketts T, Loucks C, Harriss R, Lawrence WT (2004) Global patterns in human consumption of net primary production. Nature 429(6994):870
Independent Expert Advisory Group on a Data Revolution for Sustainable Development (2014) A world that counts-mobilising the data revolution for sustainable development. United Nations, New York
Integrated Reporting Council ('The IIRC') (2013) The International <IR> Framework. http://integratedreporting.org/wp-content/uploads/2015/03/13-12-08-THE-INTERNATIONAL-IR-FRAMEWORK-2-1.pdf. Retrieved 10 Jan 2016
Intergovernmental Panel on Climate Change (2018) Summary for policymakers. In: Masson-Delmotte V, Zhai P, Pörtner HO, Roberts D, Skea J, Shukla PR, Pirani A, Moufouma-Okia W, Péan C, Pidcock R, Connors S, Matthews JBR, Chen Y, Zhou X, Gomis MI, Lonnoy E, Maycock T, Tignor M, Waterfield T (eds) Global warming of 1.5 °C. An IPCC special report on the impacts of global warming of 1.5 °C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. World Meteorological Organization, Geneva
James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning, vol 112. Springer, New York
Jantsch E (1970) Inter- and transdisciplinary university: a systems approach to education and innovation. High Educ 1:7–37
Jerneck A, Olsson L, Ness B, Anderberg S, Baier M, Clark E, Hickler T, Hornborg A, Kronsell A, Lövbrand E, Persson J (2011) Structuring sustainability science. Sustain Sci 6(1):69–82
Joint Research Centre-European Commission (JRC-EC) (2008) Handbook on constructing composite indicators: methodology and user guide. OECD Publishing, Paris
Kareiva PM, McNally BW, McCormick S, Miller T, Ruckelshaus M (2015) Improving global environmental management with standard corporate reporting. Proc Natl Acad Sci 112(24):7375–7382
Evans J, Karvonen, A (2011) Living laboratories for sustainability: exploring the politics and epistemology of urban transition. In: Cities and low carbon transitions, pp 126–141
Kates RW (2011) What kind of a science is sustainability science? Proc Natl Acad Sci 108(49):19449–19450. https://doi.org/10.1073/pnas.1116097108
Keeso A (2014) Big data and environmental sustainability: a conversation starter. Smith School of Enterprise and the Environment. In: Working paper series (14-04)
Keeves JP (1997) Educational research methodology and measurement. Cambridge University Press, Cambridge
Kim SJ, Kara S (2014) Predicting the total environmental impact of product technologies. CIRP Ann Manuf Technol 63(1):25–28
Komiyama H, Takeuchi K (2006) Sustainability science: building a new discipline
Lang DJ, Wiek A, Bergmann M, Stauffacher M, Martens P, Moll P, Swilling M, Thomas CJ (2012) Transdisciplinary research in sustainability science: practice, principles, and challenges. Sustain Sci 7(1):25–43
Lazar N (2012) The big picture: big data hits the big time. Chance 25(3):47–49
Leisinger KM, Bakker MP (2013) The key challenges to 2030/2050: mapping out long-term pathways to sustainability and highlighting solutions that should be scaled up. United Nation Sustainable Development Sustainability Network
Lucas RE Jr (1988) On the mechanics of economic development. J Monet Econ 22(1):3–42
Ludwig J, Kling JR, Mullainathan S (2011) Mechanism experiments and policy evaluations. J Econ Perspect 25(3):17–38
Lydenberg SD, Rogers J, Wood D (2010) From transparency to performance: industry-based sustainability reporting on key issues. Hauser Center for Nonprofit Organizations, Cambridge
Madnick S, Siegel M (2002) Seizing the opportunity: exploiting. Web aggregators. MIS Q Exec 1(1):35–46
Marland G, Kowalczyk T, Cherry TL (2015) Green fluff"? The role of corporate sustainability initiatives in effective climate policy: Comment on "Science-based carbon targets for the corporate world: the ultimate sustainability commitment, or a costly distraction? J Ind Ecol 19(6):934–936
Mazzocchi F (2015) Could Big Data be the end of theory in science?: a few remarks on the epistemology of data-driven science. EMBO Rep 16(10):1250–1255
McAfee A, Brynjolfsson E, Davenport TH, Patil DJ, Barton D (2012) Big data: the management revolution. Harv Bus Rev 90(10):60–68
MELODIES Project (2016) MELODIES project. http://www.melodiesproject.eu/. Retrieved 10 Jan 2016 from MELODIES project
Mihelcic JR, Crittenden JC, Small MJ, Shonnard DR, Hokanson DR, Zhang Q, Chen H, Sorby SA, James VU, Sutherland JW, Schnoor JL (2003) Sustainability science and engineering: the emergence of a new metadiscipline. Environ Sci Technol 37(23):5314–5324
Miller TR, Wiek A, Sarewitz D, Robinson J, Olsson L, Kriebel D, Loorbach D (2014) The future of sustainability science: a solutions-oriented research agenda. Sustain Sci 9(2):239–246
Milne MJ, Gray R (2013) W (h) ither ecology? The triple bottom line, the global reporting initiative, and corporate sustainability reporting. J Bus Ethics 118(1):13–29
Mittelstadt BD, Floridi L (2016) The ethics of big data: current and foreseeable issues in biomedical contexts. Sci Eng Ethics 22(2):303–341
Moir E, Moonen T, Clark G (2014) What are future cities? Origins, meanings and uses (PDF). Foresight future of cities project and future cities Catapult
Moran D, Kanemoto K (2017) Identifying species threat hotspots from global supply chains. Nat Ecol Evol 1(1):0023
Narasimhan M, Arun A (2017) Shifting gears: randomised control trials and the future of development evaluation. South Asia@LSE, London
National Research Council (2004) Facilitating interdisciplinary research. The National Academies Press, Washington, DC. Accessed 27 Feb 2015
Nelson MP, Vucetich JA (2012) Sustainability science: ethical foundations and emerging challenges. Nat Educ Knowl 3(10):12
Newell P, Frynas JG (2007) Beyond CSR? Business, poverty and social justice: an introduction. Third World Q 28(4):669–681

Nykvist B (2013) National environmental performance on planetary boundaries: a study for the Swedish Environmental Protection Agency. Swedish Environmental Protection Agency, Stockholm

OECD (1974) "Core set of indicators"

OECD (1993) OECD core set of indicators for environmental performance reviews: a synthesis report by the Group on the State of the Environment, vol 1(18)(83). OECD, Paris

OECD (2002) Indicators to measure decoupling of environmental pressure from economic growth. SG/SD(2002) 1/final

Olteanu A, Castillo C, Diaz F, Kiciman E (2016) Social data: biases, methodological pitfalls, and ethical boundaries. Available at SSRN: https://ssrn.com/abstract=2886526

O'Neil C (2017) Weapons of math destruction: how big data increases inequality and threatens democracy. Broadway Books, New York

Orts E, Spigonardo J (2014) Sustainability in the age of big data. IGEL/Wharton, University of Pennsylvania, Pennsylvania, p 16

Paddison L (2013) The effectiveness of non-financial reporting—live chat round up. Guardian. https://www.theguardian.com/sustainable-business/effectiveness-non-financial-reporting-live-chat

Parkhurst J (2017) The politics of evidence: from evidence-based policy to the good governance of evidence. Taylor & Francis, Routledge

Parris TM, Kates RW (2003) Characterizing and measuring sustainable development. Annu Rev Environ Resour 28(1):559–586

Pauly D (1995) Anecdotes and the shifting baseline syndrome of fisheries. Trends Ecol Evol 10(10):430

Pearce W (2014) Scientific data and its limits: rethinking the use of evidence in local climate change policy. Evid Policy J Res Debate Pract 10(2):187–203

Persson J, Johansson EL, Olsson L (2018a) Harnessing local knowledge for scientific knowledge production: challenges and pitfalls within evidence-based sustainability studies. Ecol Soc 23(4):38

Persson J, Thorén H, Olsson L (2018b) The interdisciplinary decision problem: Popperian optimism and Kuhnian pessimism in forestry. Ecol Soc 23(3):40

Petrov O, Gurin J, Manley L (2016) Open data for sustainable development (no. 24017). The World Bank, Washington

Pintér L, Hardi P, Bartelmus P (2005) Indicators of sustainable development: proposals for a way forward. In: Expert group meeting on indicators of sustainable development, New York, pp 13–15

Poli R (1999) Framing ontology, vol 23, pp 19–26. http://www.formalontology.it/essays/Framing.pdf. Available online from the Ontology resource guide for philosophers

Porter T (1995) Trust in numbers: the pursuit of objectivity in science and public life. Princenton University Press, Princenton

Puschmann C, Burgess J (2014) Big data, big questions metaphors of big data. Int J Commun 8:20

Ratti C, Helbing D (2016) The hidden danger of big data. Project syndicate. https://www.project-syndicate.org/commentary/data-optimization-danger-by-carlo-ratti-and-dirk-helbing-2016-08#comments

Raupach MR, Davis SJ, Peters GP, Andrew RM, Canadell JG, Ciais P, Friedlingstein P, Jotzo F, Van Vuuren DP, Le Quere C (2014) Sharing a quota on cumulative carbon emissions. Nat Clim Change 4(10):873

Redman TC (2013) Data's credibility problem. Harv Bus Rev 91(12):84–88

Rittel HWJ, Webber MM (1972) Dilemmas in a general theory of planning. Policy Sci 4:155–169

Rockström J, Steffen W, Noone K, Persson Å, Chapin III FS, Lambin E, Lenton TM, Scheffer M, Folke C, Schellnhuber HJ, Nykvist B (2009) Planetary boundaries: exploring the safe operating space for humanity. Ecol Soc 14(2):32. http://www.ecologyandsociety.org/vol14/iss2/art32/ **(online)**

Rockström J, Sachs JD, Öhman MC, Schmidt-Traub G (2013) Sustainable development and planetary boundaries. In: Background research paper submitted to the high level panel on the post-2015 development agenda. Sustainable Development Solutions Network, Paris

Ruggie JG (2017) Multinationals as global institution: power, authority and relative autonomy. Regul Govern. https://doi.org/10.1111/rego.12154

Rutter J (2012) Evidence and evaluation in policy making. Institute for Government, London, pp 1–30

Samson A (2014) The behavioral economics guide 2014. Behavioral Science Solutions Ltd., Dostupnona

Sandel MJ (2010) Justice: what's the right thing to do?. Macmillan, New York

Sarewitz D (2016) The pressure to publish pushes down quality. Nature 533(7602):147

Scotland J (2012) Exploring the philosophical underpinnings of research: relating ontology and epistemology to the methodology and methods of the scientific, interpretive, and critical research paradigms. English Lang Teach 5(9):9–16. https://doi.org/10.5539/elt.v5n9p9

Searcy C (2016) Measuring enterprise sustainability. Bus Strateg Environ 25(2):120–133

Shah S, Horne A, Capellá J (2012) Good data won't guarantee good decisions. Harv Bus Rev 90(4):23–25

Shepherd K, Hubbard D, Fenton N, Claxton K, Luedeling E, Leeuw J (2015) Development goals should enable decision-making. Nature 523(7559):152

Silberzahn R, Uhlmann EL (2015) Many hands make tight work. Nature 526(7572):189. https://doi.org/10.1038/526189a

Slater A, Zwat G (2015) Informing decisions, driving change: the role of data in sustainable future. GRI and Oxfam Novib, Amsterdam

Smaldino PE, McElreath R (2016) The natural selection of bad science. R Soc Open Sci 3(9):160384

Small ML, Sampson RTJ (2014) Executive summary of bringing social science back. In: The 'big data' revolution and urban theory. https://www.radcliffe.harvard.edu/executive-summary-bringing-social-science-back-in-%E2%80%98big-data%E2%80%99-revolution-and-urban-theory

Spangenberg JH (2011) Sustainability science: a review, an analysis and some empirical lessons. Environ Conserv 38(3):275–287

Starr MA (2014) Qualitative and mixed-methods research in economics: surprising growth, promising future. J Econ Surv 28(2):238–264

Strathern M (1997) 'Improving ratings': audit in the British University system. Eur Rev 5(3):305–321

Sundar N (2000) The construction and destruction of indigenous knowledge in India's Joint Forest Management Programme. Indigenous environmental knowledge and its transformations. Harwood Academic Publishers, Amsterdam, pp 79–100

Sustainability Context Group (2012) Statement to GRI on the need to enhance treatment of the sustainability context principle in G4, 24 Sep 2012

TEEB (2010) The economics of ecosystems and biodiversity ecological and economic foundations. In: Kumar P (ed) Earthscan, London. http://www.teebweb.org/our-publications/teeb-study-reports/ecological-and-economic-foundations/

Thomke S, Manzi J (2014) The discipline of business experimentation. Harv Bus Rev 92(12):17

Thorén H, Persson J (2013) The philosophy of interdisciplinarity: sustainability science and problem-feeding. J Gen Philos Sci 44(2):337–355

Trexler M, Schendler A (2015) Science-based carbon targets for the corporate world: the ultimate sustainability commitment, or a costly distraction? J Ind Ecol 19(6):931–933

Tullock G (2001) A comment on Daniel Klein's 'a plea to economists who favor liberty. East Econ J 27(2):203–207

UNDP (2004) Human development report 2004. UNDP, New York

UNEP (2002) Global environment outlook 3. UNEP, New York
United Nations (2015) Sustainable development. https://sustainabledevelopment.un.org/?menu=1300. Retrieved from United Nations
Varian HR (2014) Big data: new tricks for econometrics. J Econ Perspect 28(2):3–28
Wachsmuth D, Cohen DA, Angelo H (2016) Expand the frontiers of urban sustainability. Nat News 536(7617):391
Ward JS, Barker A (2013) Undefined by data: a survey of big data definitions. arXiv preprint arXiv:1309.5821
WB (2004) World development report 2005: a better investment climate for everyone. WB, Washington
Weller K, Kinder-Kurlanda KE (2015) Uncovering the challenges in collection, sharing and documentation: the hidden data of social media research. In: Standards and practices in large-scale social media research. International Conference on Web and Social Media, Oxford
Werner R (2015) The focus on bibliometrics makes papers less useful. Nature 517(7534):245–246
Wiedmann TO, Schandl H, Lenzen M, Moran D, Suh S, West J, Kanemoto K (2015) The material footprint of nations. Proc Natl Acad Sci 112(20):6271–6276
Williams M, Zalasiewicz J, Waters CN, Edgeworth M, Bennett C, Barnosky AD, Ellis EC, Ellis MA, Cearreta A, Haff PK, Ivar do Sul JA (2016) The anthropocene: a conspicuous stratigraphical signal of anthropogenic changes in production and consumption across the biosphere. Earth's Future 4(3):34–53
World Bank (2015) Open data for sustainable development. World Bank, Washington
World Bank (2018) Information and communications for development 2018: data-driven development. World Bank, Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/30437. License: CC BY 3.0 IGO
World Council on City Data (2015) About. http://www.dataforcities.org/about. Retrieved from World Council on City Data
WRI (2003) World resources 2002–2004. Decisions for the Earth
Xu M, Cai H, Liang S (2015) Big data and industrial ecology. J Ind Ecol 19(2):205–210
Yarime M (2017) Facilitating data-intensive approaches to innovation for sustainability: opportunities and challenges in building smart cities. Sustain Sci 12(6):881–885
Yarime M (2018) Learning and open data in sustainability transitions: evolutionary implications of the theory of probabilistic functionalism. Environ Syst Decis 38(1):88–91
Zawistowska S (2015) Big data and corporate sustainability—what is the role of big data in MCNS with corporate sustainability agenda and what is the practical use they can make of it? Masters thesis, Copenhagen Business School
Zhang Q, Jiang X, Tong D, Davis SJ, Zhao H, Geng G, Feng T, Zheng B, Lu Z, Streets DG, Ni R (2017) Transboundary health impacts of transported global air pollution and international trade. Nature 543(7647):705
Zhongming Z, Linong L, Xiaona Y, Wangqiang Z, Wei L (2014) Gap analysis on linking open data for global disaster risk research
Ziemann M, Eren Y, El-Osta A (2016) Gene name errors are widespread in the scientific literature. Genome Biol 17(1):177