Nullius in verba*
A guideline proposal for data accessibility and transparency in industrial ecology research

Draft proposal by the ISIE data transparency task force
Version 1.1

Edgar Hertwich[1], Niko Heeren[2], Brandon Kuczenski[3], Guillaume Majeau-Bettez[4,5], Rupert J. Myers[1], Stefan Pauliuk[6], Konstantin Stadler[4], and Reid Lifset[1]

[1]School of Forestry & Environmental Studies, Yale University, New Haven, Connecticut, USA
[2]Institute of Environmental Engineering, ETH Zurich, Switzerland
[3]Bren School of Environmental Science and Management, University of California, Santa Barbara, USA
[4]Industrial Ecology Programme, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
[5]CIRAIG, Polytechnique Montreal, Montreal, Canada
[6]Faculty of Environment and Natural Resources, University of Freiburg, Germany

June 26, 2017

Abstract

The scope and size of industrial ecology (IE) is growing, leading to an accumulation of available IE results and data, and a desire for more data and sophisticated analysis tools. With growth, the need for greater transparency, accessibility, and reusability of IE data has increased. Parallel to this, there is

* "Nullius in verba" is the motto of the UK Royal Society. It roughly translates as 'take nobody's word for it'. It is an expression of the determination of Fellows to withstand the domination of authority and to verify all statements by an appeal to facts determined by experiment (Royal Society 2017).

considerable momentum throughout the sciences–in many fields, in journals, and among research funders–for greater data transparency and accessibility.

The Data Transparency Task Force (DTTF) was convened by the governing council of the International Society for Industrial Ecology in late 2016 to propose best-practice guidelines and incentives for sharing data. In developing proposals for the industrial ecology community, the DTTF recognizes the diversity of the IE research community–in tools, topics, domains and expertise–and seeks to integrate feedback from the many segments of the IE community.

In this article, the members of the DTTF present an overview of developments toward transparent and accessible data broadly and within industrial ecology. We argue that increased transparency, accessibility, and reusability of IE data will enhance IE research by enabling more detailed and reproducible research and facilitating meta-analysis. These benefits will make the results of IE work more timely. They will enable independent verification of results, thus increasing their credibility and of higher quality. And they will make the uptake of IE research results easier within IE and in other fields and by decision makers and sustainability practitioners, thus increasing the relevance and impact of the field.

Here, we propose two preliminary actions to advance these goals: (1) a minimum publication requirement for IE research to be adopted by the *Journal of Industrial Ecology*; and (2) a system of optional badges awarded to journal articles that contain increasing levels of transparent and accessible data. Our intention is that these actions start an inclusive discussion for all within the IE community to engage with and respond to these key concerns of data transparency and accessibility; therefore, they should be treated as being under an active community-led development program. We close with a discussion of potential future initiatives that could build on the minimum requirements and the badge system.

1 Introduction

A core mission of the industrial ecology (IE) field is to contribute to the scientific basis of sustainable development. The value of this contribution depends on the quality, timeliness, and relevance of the scientific insights discovered in IE research. These insights include quantitative assessments of life cycle impacts, shifting of environmental burdens between economic sectors, factors that shape the development of industrial symbiosis, and the future recycling potential. IE research is often data-intensive and characterized by an ongoing improvement of its analytical tools.

The issue of data transparency has been identified by the council of the International Society for Industrial Ecology as an important concern and it convened the Data Transparency Task Force (DTTF) in late 2016 to obtain guidance on objective to best-practice and incentives for sharing IE research data and documenting research workflow (see SI 1 for a copy of the mandate). The goal of the DTTF is to develop guidelines and incentives that encompass the whole research process, ranging from documenting input data and assumptions, to methodological aspects such as accessible software code to providing access to generated output data. A proposal for such guidelines is presented here. This discussion is primarily focused on quantitative research. The DTTF, however, is eager to receive feedback from IE researchers involved in qualitative research as application of efforts toward data transparency may be relevant for their work as well.

We start by reviewing data sharing in other fields, academic journals, and funding agencies. Current use, provision, and sharing of data within IE is then summarized, mapping out possible improvements and benefits and showing examples of good data handling practices. We then present some initial proposals for best-practice guidelines for transparent, accessible, and reproducible IE research, and a minimum requirement for IE publications that we suggest be adopted by the *Journal of Industrial Ecology* (*JIE*). We close with a call for feedback on our proposal and discuss future efforts to progress towards our goal of fully transparent IE research.

2 Current trends in data openness

The scientific method builds upon reporting and sharing of research results. Sharing allows scientific investigations and results to be independently tested and scrutinized; it enables the accumulation of data, findings, and insights and thus leads to an advancement of research over time. The advent of big data leads some (but not all) to call for a new scientific paradigm in which a new empiricism replaces theory (Kitchin 2014). Independent of this paradigmatic debate, the ability to acquire or develop, store, and utilize large data sets invariably is having a significant an increasing influence on science.

2.1 Trends across scientific fields

Data sharing requirements and practice vary across research fields; they also change with progress in data processing and storage opportunities. Recent decades have seen a massive increase of scientific data stored in electronic format, and many journal articles now contain electronic supplementary material (Kenyon and Sprague 2014). Despite the ease of storing and exchanging data, which was brought about by computers and the Internet, there is an increasing concern within the scientific community regarding the accessibility of research data. In a 2010 survey across scientific fields, a majority of researchers indicated that a lack of access to research data hinders progress in science and almost half of the respondents stated that this lack of data access limits their ability to answer scientific questions (Tenopir et al. 2011). According to a survey conducted by *Nature*, more than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments (Baker 2016), often due to problems with accessing data from the original studies (Van Noorden 2015).

Pfenninger et al. (2017) give a detailed overview of the barriers to and benefits of data and model sharing. They also suggest that institutional and personal inertia play a role in maintaining the attractiveness of closed models and data.

Without sufficient access to data, scientific analyses cannot be replicated and subsequent research cannot build on previous results, both of which undermine the foundation of scientific progress. As a consequence, data archiving and sharing have become a cross-cutting issue across all scientific fields, including:

- physics (Hey and Payne 2015);
- political science (Gherghina and Katsanidou 2013);

- bioinformatics (Hothorn and Leisch 2011);
- ecology and biodiversity research (Michener 2015; Costello et al. 2013);
- medical research, neuroimaging, and genomics (Walport and Brest 2011; Warren 2016; Poline et al. 2012; Kaye and Hawkins 2014; Farber 2017); and
- materials science (Graulis et al. 2009; Lafuente et al. 2016).

2.2 The perspective of funders

Driven by a debate about the role of publicly-funded science in society, funding agencies are increasingly requesting open access to data generated and used in sponsored research. For example, the European Commission asks Member States to ensure that "research data that result from publicly funded research become publicly accessible, usable and re-usable" (European Commission 2012). Similarly, the U.S. National Science Foundation (NSF) now requests that all data resulting from NSF-funded research be deposited in appropriate data repositories (National Science Foundation 2015). 'Harnessing data' is one of the ten major strategic directions for future NSF investment.

2.3 Data repositories

A number of repositories have been developed in response to the growing need for data storage. The recently established data repository registry re3data (*http://www.re3data.org/*) now lists over 1 500 individual data repositories from multiple scientific fields ranging from general-purpose ones such as Figshare (*http://www.figshare.org*), Zenodo (*https://www.zenodo.org/*), and Dryad (*http://datadryad.org/*) to subject-specific ones such as GenBank for genetic sequence data (*https://www.ncbi.nlm.nih.gov/genbank/*), PANGEA for Earth and Environmental Science (*https://pangaea.de/*), or the Interdisciplinary Earth Data Alliance (IEDA) (*http://www.iedadata.org/*). The integrated assessment modelling (IAM) community is currently establishing a database that will serve as a central data access hub for IAM input data and modelling output and will thus also be relevant to the IE community (IIASA 2017).

Persistent and indexed repositories provide two major advantages over self- or institute-hosted solutions: First, data hosted at these repositories usually receive a digital object identifier (DOI), which allows for easy referencing of a specific dataset including version information. Second, an analysis of research data availability on journal homepages has shown that availability declines at a rate of about 17% per year (Vines et al. 2014); therefore, archiving data in dedicated repositories prolongs their use. Therefore, we support the use of persistent and indexed repositories.

2.4 Academic journals

In parallel with the development of common data repositories, academic journals have also sought to increase data availability, recognizing the basic scientific need for high data availability and the requirements of funding agencies (Mcnutt 2014). Such actions include increasingly stringent requirements for datasets to be published alongside journal articles. In early 2017, *Nature* implemented a requirement for a data availability statement at the end of articles summarizing the availability of the data that is necessary to replicate, interpret, and build upon the findings of the paper (Nature 2016). *Nature* also recently established *Scientific Data*, which is a new journal dedicated to publishing and describing openly accessible datasets (*https://www.nature.com/sdata/about*). *Science* requires authors to deposit large datasets at an official repository prior to publication (*http://www.sciencemag.org/authors/science-editorial-policies*) (SpringerNature Group 2016), and similar policies exist, for example, for PLOS journals (*http://journals.plos.org/plosone/s/data-availability*) and bio-medical journals like *Cerebral Cortex* (*http://www.oxfordjournals.org/our_journals/cercor/for_authors/general.html*), *Cell* (*http://www.cell.com/cell/authors*), and *Neuron* (*http://www.cell.com/neuron/authors*).

3 Data sharing in industrial ecology

IE research requires a substantial amount of secondary data, which is often based on existing IE research, and which includes life cycle inventories (LCIs), (multi-regional) input-output (IO) databases, official statistics, and market surveys. Physical-chemical properties of materials and processes or the manipulation of such secondary data are often used and combined to create new data and insights. Quantitative research methods used in IE include various computational techniques such as linear algebra, geographic information systems (GIS) data analysis, statistical analysis, and linear programming. IE research has specific data requirements, as it covers a wide range of approaches, data formats, and system representations which makes data harmonization and the establishment of common data formats challenging. There is, however, a growing convergence or at least hybridization of the approaches of material flow analysis (MFA), IO analysis, and life cycle assessment (LCA). Additionally, as IE researchers often work closely with industry, the issue of data confidentiality is important and may restrict data sharing.

While there has been some success in compiling databases for aggregated data such as country-level material flows, generic LCIs, and impact assessment characterization factors, a pervasive culture of sharing case-specific data along with the publication of new research results appears to be lacking. The problems of inadequate data transparency and accessibility within the IE community has triggered calls for more reproducibility (Frischknecht 2004), better digital communication (E. Hertwich 2007) and use of interlinked data (Davis et al. 2010), and improved code standards and data sharing (Pauliuk et al. 2015). The proposals developed by the Data Transparency Task Force (DTTF) represent an effort to find feasible solutions to these challenges.

3.1 Benefits of data transparency and Long-term costs of business as usual

The inaccessibility of IE detailed research results presents a significant lost opportunity. Inaccessible details cannot be used or fully understood by others, nor cross-checked, replicated, and verified, and become part of larger meta-analyses. The lack of properly formatted, documented, and comparable data is particularly evident in the field of life cycle assessment. For example, in the application of LCA to climate change mitigation (E. G. Hertwich 2014), the U.S. National Renewable Energy Laboratory (NREL) undertook a project to produce a comparable set of LCI data for the Intergovernmental Panel on Climate Change (IPCC) special report on renewable energy. However, collecting, extracting, and harmonizing these data to provide a broader assessment required a painstaking amount of effort. The *Journal of Industrial*

Ecology (*JIE*) special issue on harmonization of LCA which documented the NREL work indicated that approximately half of the LCA studies reviewed had to be discarded because the LCI data were not published or so poorly documented that they could not be unambiguously interpreted (Heath and Mann 2012). A recent review of data quality of electricity LCAs discusses this lack of consistency and transparency and suggests that it adversely impacts not only the usefulness, but also the quality of LCA results (Astudillo et al. 2016). In the case of the IPCC 5th Assessment, concerns by participating scientists about the quality negatively impacted the degree to which the LCA results were trusted by the IPCC and thus employed in the policy making process.

Inaccessibility of data and poor harmonization of data and procedures results in a wasteful duplication of effort with little benefit to the community as a whole. There is also significant advantage if data are available in a harmonized, machine-readable data format, as the effort needed to parse data that others have produced is reduced. Reduced cost of comparing and reusing results and conducting harmonization studies allows the whole field to speed up its progress.

Although sharing and documentation of data require additional effort, it is an effort that advances the field and can offer rewards and immediate benefits for the individual researcher. Recent studies suggest that publishing open datasets may lead to a wider use, thus enhancing its status and increasing citations (Piwowar and Vision 2013; Drachen et al. 2016). Furthermore, supply of the underlying data and intermediate results contributes to validation and quality control. Other researchers may add to the data, and as the accessible knowledge base grows, it provides the opportunity for follow-up work, such as meta-analyses, resolving potential disagreements, and providing more robust insights. It may also provide the opportunity for researchers to join together in larger efforts that lead to more high-profile publications. Some fields, such as climate sciences, earth systems modeling, and energy scenario modeling, have a tradition of carrying out projects to compare individual model results which help to provide common benchmarks, create acceptance for new research questions and model-oriented papers, and tend to result in joint high-level publications by the whole community. Model comparison is now also occurring within IE (Owen 2017; Moran and Wood 2014; Speck et al. 2015).

We therefore corroborate the need to improve data transparency as identified by the council of the ISIE, which we believe can:

- improve research communication;
- enhance accumulation of IE knowledge;
- speed up scientific progress within IE;
- enable independent verification of results, thus increasing credibility, reliability, and quality; and
- increase the significance of IE research by facilitating uptake of IE research results by other fields and decision makers.

3.2 Examples of data sharing in the IE community

We have compiled a list of good examples of data and procedural transparency within IE building on community input solicited via email exchanges and forum posts. This list is now too long be shown in full in the paper; therefore we have decided to publish this list of examples in the supporting information accompanying this paper.

4 Defining data transparency for industrial ecology

Transparency is key for fostering collaborative science. The task force is convinced that a change towards new data management practices and data transparency are required in the current publication practice worldwide and in the *Journal of Industrial Ecology (JIE)*. Incentives are necessary to create an environment that facilitates data contributions and processing. Here we propose how the *JIE* could change its publication and review regime in the future, which we submit for debate among the IE community. Our recommendations include modest mandatory requirements to ensure all *JIE* publications meet basic data transparency requirements and propose a *data openness badge* to incentivize authors to supply additional data.

The IE community faces two fairly unique key challenges with regard to data openness and reuse:

1. the central role of industry data in IE research and associated confidentiality issues; and

2. the variety of data types used, which stems from IE's multidisciplinary and broad-ranging scope, and which in turn leads to questions of interoperability and ease of data reuse

These issues make it challenging to develop general guidelines on data formatting and documentation. The proposal for publication requirements and incentives was were devised to reflect the characteristics of IE research.

We follow a multi-layered strategy: First, we propose minimum publication criteria, which focus on clear citation of secondary data and reusability of results (labeled with an asterisk (*) in Figure 1, as neither are restricted by confidentiality of primary data. For the publication and reuse of intermediate models and detailed system descriptions (labeled with two asterisks (**) in Figure 1, we propose a progressive badge system to incentivize both higher levels of transparency and accessibility. Furthermore, we aim to promote data transparency and accessibility through other optional means, such as recommending use of data repositories.

As further clarified below, these criteria and incentives for data openness are planned as an initial step, to be complemented by additional incentives for higher procedural openness at a later date (labeled with three asterisks (***) in Figure 1. Our vision is that the implementation of each of these aspects will progressively lead to high levels of transparency, accessibility, and reproducibility in all research steps, from raw data to final results.

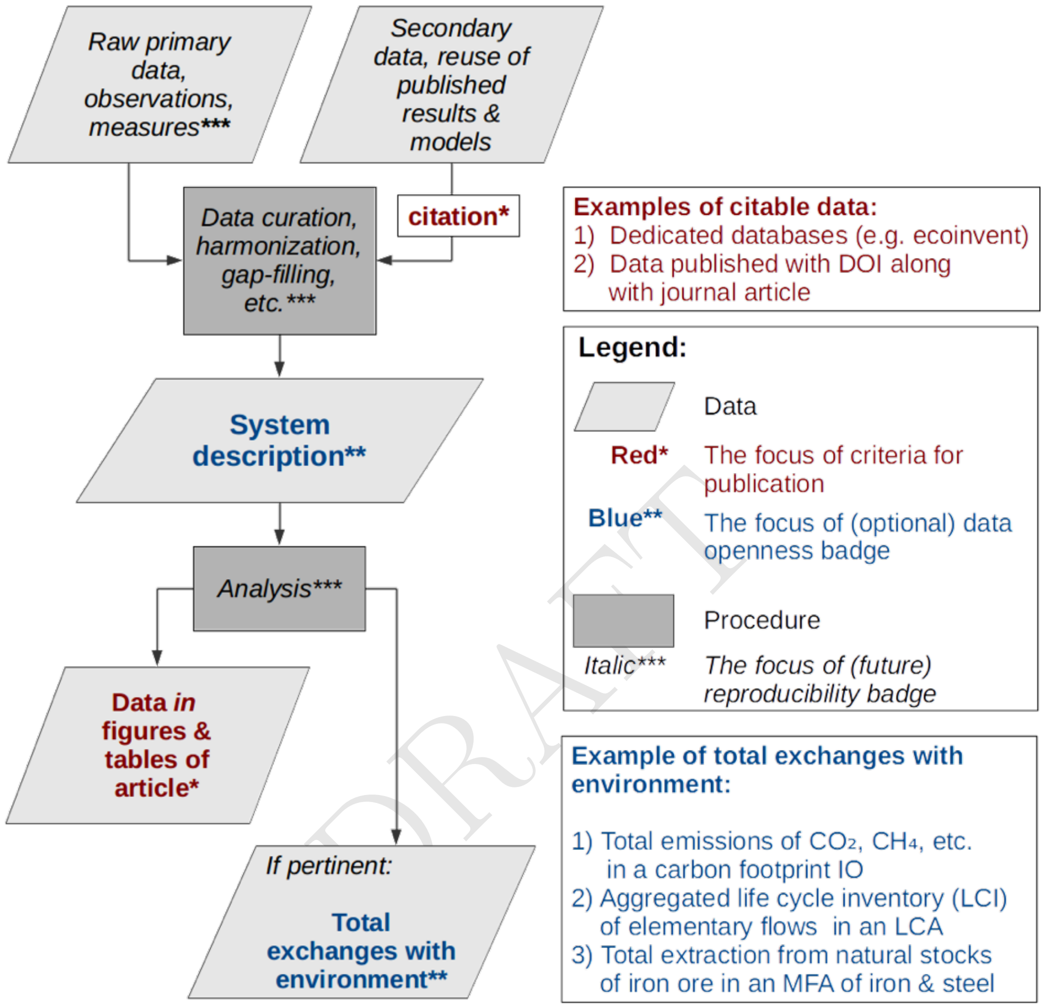

Figure 1: The scope of the proposed minimum transparency criteria (red*) and the proposed data openness badges (blue**) within a conceptual representation of IE research and publication process (flow chart), with data and manipulations respectively in pale parallelograms and dark gray rectangles. The scope of a future iteration of the badge system is also outlined (Italic***).

The DTTF proposes to incorporate these data accessibility criteria into the *JIE* publication process in the future. As discussed in the section "Community engagement", we propose an inclusive approach, with the results of feedback from participants of the 2017 ISIE Conference, the ISIE sections, the *JIE* editorial board and the IE community at large to be used in a planned, subsequent re-iteration of this proposal.

4.1 Minimum data transparency criteria

We propose minimum data publication requirements for IE research. These aim to be applicable to all IE research regardless of the confidential nature of the system description or its underlying data, and therefore aim at facilitating the inspection and reuse of results (rather than the more demanding process of replication of the analysis). We identify two key issues that often make replication of IE research difficult: 1) digital data are typically inadequately identified; and 2) data extraction is often more difficult than necessary. The requirements below are intended to address these two issues.

Requirement 1 – Data citation: All secondary data and databases used in the analysis must be cited in accordance with the journal's citation style. This information can include database version, database settings (e.g., allocation), date accessed, and digital object identifier (DOI), if pertinent. This requirement both clarifies data sources and provides incentives for publication of reusable and citable data. Data may be cited in the main section of the paper or in the supporting information.

Requirement 2 – Enumerate primary results: The data that are represented in each graph or figure in an article must be published, clearly cited, and labeled. For example, a simple spreadsheet containing the quantitative data that are the basis of figures and tables in an article fulfills this requirement; such data can be provided in supporting information. This should facilitate the unambiguous inspection and usage of quantitative information contained in all key results presented as figures and graphs. The underlying quantitative data would become directly accessible, avoiding the need to manually extract them from figures and avoiding any uncertainties introduced from this process. This may facilitate increased citation, reuse, and meta-analyses of published work.

In all cases, the data supplied should be published in the supporting information or archived in a trusted repository, preferably an official repository which assigns DOIs, and cited accordingly in the original article. We expect practices in this regard to evolve as scientific publishing continues to address data transparency and accessibility.

We believe that these two simple criteria will greatly improve the transparency and usefulness of IE publications while avoiding confidentiality issues or cumbersome alterations to the workflow of IE researchers. Overall, we consider these criteria to be relatively modest and to reflect good practice of scientific publishing in general. Nevertheless, we have explicitly stipulated them here to provide a first step towards full data transparency of IE research.

4.2 Data openness badges

To reward authors whose articles achieve higher levels of data transparency, accessibility, and interoperability between data formats beyond the strict minimum for publication, we propose that an optional *data openness badge* system be introduced into the *JIE* publication process. The *data openness badge* would be requested by the authors upon submission of their manuscript and reviewers would be asked to verify its applicability. Once a badge is granted, it will be visible on the publication (see Figure 2) to reward and showcase author efforts towards data openness. The badge system aims to be progressive and flexible, with two dimensions and two levels to accommodate the diversity of research in the IE community. The first dimension addresses *data transparency*, while the second concerns *data accessibility*, the latter meaning the interoperability and reusability of the data supplied. Although further criteria have been identified by the DTTF and suggested in the literature (Nosek et al. 2015), these were not adopted in this initial proposal, as it was our goal to focus on those criteria that would facilitate IE research the most while being the least disruptive to the established practice in the community.

Defining incentives that are both broadly applicable and flexible as well as clear and practical constitutes a major challenge, and we intend to continually engage the community in the development process (see section titled "Community engagement" below). The following definitions should therefore be considered as initial proposals to initiate a broader debate within the IE community.

Criteria for data transparency badges

Level 1: Transparent plus

The entire system description is published at the same level of resolution and completeness as was used by the authors to calculate their results.

- These system descriptions notably include (as applicable) the definitions of all processes, activities, agents, objects, flows, stocks, exchanges with the environment, system boundaries, and behaviors and actions, along with links to external/secondary data.

- All the necessary data are made available such that the results can be reproduced; although the authors are not required to share all detailed calculation and analysis steps that were performed using the system description.

    - Example 1: A global input-output footprint analysis links to an open and accessible system description including the matrix of technical requirements, exchanges with the environment, final consumption, and value added.
    - Example 2: An LCA study makes available its foreground (all process descriptions based on own research and primary data) and also publishes all the links to a published dataset (e.g., ecoinvent) for all secondary data used.

Level 2: Transparent

Option 1: In situations where authors cannot share their entire system description, for example, when facing confidentiality issues, they can nonetheless share the detailed description of the non-sensitive parts of the system.

- Published datasets would include, for example, complete process descriptions, extensive descriptions of stocks and flows, and tabulated product compositions.

- The intent is that *a significant portion of the system is described in a self-contained and useful manner* with clear metadata allowing for unambiguous interpretation of each data point within this part of the system.

- Example 1: An LCA study of Li-ion battery use may be unable to fully describe the assembly of battery cells because the data on energy requirements to do this are commercially sensitive. This analysis may nonetheless usefully characterize unit processes describing at full resolution the production of the anodes, cathodes, and electrolytes, thereby contributing useful primary data to the community.

- Example 2: The publication of an extensive MFA model may similarly be unable to include the whole system description. Nonetheless, the authors are able to share an extensive table of the mass and elemental composition for many of the stocks and flows in the model, which will likely prove useful to other research.

Option 2: The second approach to fulfilling the objective of level 2 applies to studies of interactions of technological systems with the environment, for example, through multiple types of emissions and resource use. In such a case, the total exchanges of the technological system with the environment should be published in a readily reusable and uncharacterized format.

- Example 1: In the case of an LCA study, a complete LCI of elementary flows would be published, that is, the cumulative total for the whole life cycle of each type of emission flow and each type of resource use.

- Example 2: An input-output analysis calculating the carbon footprint of nations would include publication of the results not only in terms of characterized CO_2-equivalents, but also the actual emissions of CO_2, CH_2, N_2O, etc. This means, that the full stressor (satellite accounts) matrix necessary for calculating the results should be made available alongside the publication.

Criteria for data accessibility badges

Level 1: Accessible plus

The system description is formatted and archived such that it is both human readable *and* directly importable into free software capable of completing the relevant IE analysis.

- *Human and machine readability:* The system is described such that it can be naturally read and understood by humans and easily imported into standard software. Examples of such formats include plain text, csv, json, and xml files, along with zipped xml files such as spreadsheet formats xlsx and ods.

- *Direct imports in relevant software:* The relevant analyses can be directly performed on the system description without requiring payment for software. Many situations fulfill this objective, for example:

1. A system description is exported in a non-proprietary structured format (e.g., ecospold XML files) that can be imported directly into free software (e.g., openLCA and brightway2) which can perform the relevant analysis (e.g., LCA calculations).

2. A study is fully performed in a spreadsheet and does not require other software to complete the analysis, thus ensuring that the spreadsheet can be opened without loss of functionality in a free office suite (e.g., LibreOffice or Google Sheets) fulfills the requirement.

3. A study publishes not only the data but also the (free) software to parse and analyze it (e.g., a Python script).

Level 2: Accessible

The complete system description is provided in another suitable format.

- This badge requires the same human readability as in level 1, but without the requirement that the data be directly importable into free analysis software. It should, however, be machine readable in the sense that a software can readily distinguish words from numbers, recognize table structures, etc. For example, an MS Excel spreadsheet would be considered machine readable, whereas PDF and word processing formats (.docx, odt, etc.) would not.

In order to obtain a data transparency badge, it is necessary that the authors respect at least the minimum level of both dimensions (i.e., data and format). The level of the badge is represented with a color code: black for full compliance with level 1 requirements, and white for compliance with level 2 requirements. Our proposed design for the badges is presented in Table 1 and Figure 2. The DTTF welcomes suggestions on graphic design in addition to general feedback on the badge system.

Table 1: Summary table of the two dimensions and two levels of the proposed Data Openness Badge system

	Data transparency	**Data accessibility**
Level 1	**Transparent plus** Entire system description	**Accessible plus** Human & machine readable (directly importable into free software)
Level 2	**Transparent** *Option 1:* Detailed and fully transparent description of non-proprietary parts of the system *Option2*: Total exchanges of the technological system with the environment published in a fully transparent format	**Accessible** Human & machine readable (any format)

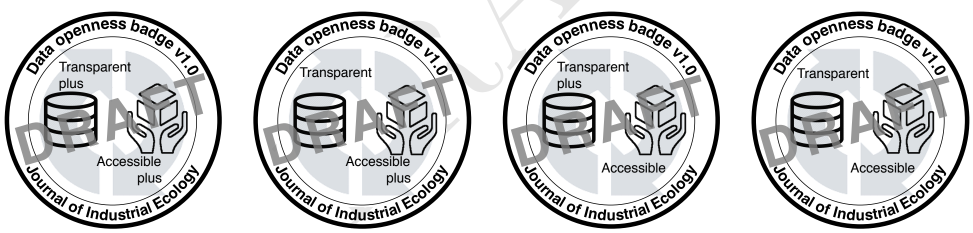

Figure 2: The four possible combinations of the proposed Data Openness Badge

As a "standard" data format for IE research has yet to be defined, we propose to incentivize the publication of any data that is formatted such that it can be directly imported into a free analysis tool. The aim is that this will promote convergence in terms of data formatting and a greater interoperability with free analysis tools.

Similarly, as it is challenging to describe systems in IE research in a standard way, we refrain from prescribing a specific manner of describing the studied system. Rather, our proposed badge system aims to reward disclosure of the system description (or part of it) as it was used by the authors to generate their results. We hope that this specification leads to maximum flexibility and applicability in the use of badges.

We have envisaged two ways of meeting of achieving the data transparency and accessibility sought through the badge system We hope this flexibility will incentivize another type of data disclosure and address the following challenge. IE studies typically describe technological (sub)systems that interact with the environment or society. These interactions are often represented as exchanges of substances, energy, money, etc. These exchanges are then typically *characterized* in LCA studies to translate them into impacts (global warming, resource depletion, consequences on human health, etc.). However, as science gains knowledge on the response of systems (e.g., natural and social) to these exchanges, characterization factors are continuously updated (e.g., ReCiPe2008, ReCiPe2016, CML2001, ImpactWorld+, etc.). The choice of characterization method has a significant influence on the results of the study. For example, following the publication of the IPCC 2013 5$^{\text{th}}$ Assessment Report, our understanding of the climate impacts associated with different greenhouse gases has been refined, which is reflected in some recent IE studies. Without the publication of the uncharacterized emissions data, however, these IE results are incommensurable with those that rely on older or different characterization methods. Therefore, *uncharacterized* flows to and from the environment are essential to update and build knowledge of impacts across studies. We therefore propose to reward any study that publishes the total of each exchange of the technological system with the environment in an uncharacterized manner with a "Transparent, level 2" badge.

These badges all signal to the reader that detailed data and system descriptions underlying the published work are transparent and accessible. We believe that our proposed grading system will allow recognition of the effort it takes to make datasets transparent and accessible. Consistent use of the badge system may guide meta-analyses to select studies with levels of transparency sufficient to harmonize system boundaries and assumptions. These inventories may also be consolidated in IE databases at a later stage. We emphasize that our proposal is to make the badges optional, i.e., not essential for publication.

Furthermore, this badge system for data openness is clearly not applicable to all IE research. Qualitative research and methodological research without detailed case studies, for example, may contain no quantitative system description and have no relevant data to share. We have developed this badge system in a way that we believe will not deter from their usefulness and publication.

4.3 Implementation

In order to implement these changes the DTTF proposes that the minimum requirements become part of the regular *JIE* review process. Concerning the data badges, we propose that authors who wish to obtain a data transparency badge have to indicate so in the *JIE* manuscript submission process. In such a case the editor asks the reviewers to score data transparency and data availability, for example, on two 1-5 scales like other questions currently asked during review, and the data badges can be assigned based on numerical score (level 2 = 3-4, level 1 = 4-5). The handling editor will then be able to judge the claim of the authors that their manuscript is indeed compliant with the requirements.

5 Next steps

5.1 Community engagement

With this piece the DTTF submits to the IE community its recommendations for greater data transparency. An introduction of minimum publication requirements and the data transparency badges requires a political process to build consensus in the community, for which we see the following next steps:

- We **solicit feedback on our proposal** and on other options to enhance data transparency and accessibility from the entire IE community. To that end we created the permanent email address ***data@is4ie.org*** and posted the topic on IE social media channels. Furthermore, we invite all interested parties to participate in an online survey, which can be found on ***www.is4ie.org/opendata***. This survey site also provides the possibility to leave an anonymous feedback. *JIE* editors and section board members have already been contacted. We are particularly interested in hearing whether the level of ambition of the badge system is reasonable, how the measures proposed would affect your workflow, preferred licensing and storage places, and whether there are ongoing data or procedural transparency and accessibility efforts in other communities to which we should link our efforts.

- We **ask the ISIE sections for their comments** on our proposal, and for a statement of the future role of the respective sections in implementing data transparency and accessibility.

- Based on the feedback obtained, we plan to **submit a refined set of recommendations to the ISIE council** at latest by the end of this year, including a proposal for the amendment of the *JIE* publication requirements regarding the introduction of the transparency badge system.

We now offer some reflections on the further development of data sharing, its relation to procedural transparency, to the development of software tools, and trends in other research communities.

5.2 Data structure harmonization and an IE database

One of the most important opportunities arising from data transparency is that it facilitates data processing in follow-up studies, e.g. for meta-analysis. Our current proposal aims at making data available in a convenient format. In the long term, we envisage that it would be ideal to develop a central database for IE research that contains data from a wide spectrum of studies and that can be queried by researchers. However, such an approach requires that data are comprehensively harmonized across at least the major IE techniques (e.g., IO, LCA, and MFA), which will naturally lead to stricter requirements regarding a harmonized nomenclature and formatting of data. However, there are additional barriers to harmonizing IE data, for example:

- Authors may perceive a data harmonization step in their work as an undueburden.

- As analysis methods evolve (e.g., from LCA to hybrid IO-LCA), data formats may become inadequate for studies presenting novel methodology, which may therefore become an obstacle for innovation.

At this point, we believe that the challenges of establishing a harmonized data format are too large at present to propose a solution here. However, we believe that a harmonized data structure presents a clear opportunity to substantially advance IE research and that should not be missed in future efforts towards full data transparency and accessibility.

5.3 Linking data transparency with procedural transparency

Another major barrier to achieving transparency of results is the absence of clear documentation of a method or procedure by which the results were generated. While transparent procedures can be used to reproduce results, open results alone cannot be used to infer the underlying procedures.

In practice, the steps required to organize data sources, process data, and extract results are study-dependent and nuanced. It is common for authors to devote a significant portion of their manuscripts to documenting their methodology, and in many cases, there are opportunities for legitimate disagreements about methodological choices. As a consequence, even highly 'transparent', well-documented studies can be difficult to reconcile with one another. To progress beyond data transparency and toward validation, verification, and reproducibility of results, a higher transparency of methodology documentation is needed.

We believe that the IE community needs to move beyond the summary descriptions often present in methods sections of papers, and towards the publication of detailed research procedures and computational scripts that fully *reproducible research* requires. Apart from concerns about disclosing information to competitors, IE researchers may object to the workload that would be associated with the relatively high level of documentation required attain this goal. Here, we note that data and procedural transparency go hand in hand, and so the latter is inevitably needed to attain fully reproducible research. We believe that full reproducibility can only be reached through a step-by-step process, and that this document provides an important preliminary step towards realizing this aim.

6 Final thoughts and conclusions

The endeavor for higher data transparency and accessibility has just begun, and our proposal for minimum requirements and the data transparency badge system is up for debate. We believe that the contribution of IE research cannot be achieved until results become more readily comparable, integrated, citable, and reusable. In order to achieve fully reproducible IE research, the data transparency and accessibility standards suggested here would further require data transparency to be linked to procedural transparency and harmonization of data structures and computational methods. The upcoming changes will affect the workflow of each of us as IE researchers. They will likely also have consequences for data ownership, which may entail legal and institutional considerations, and for competitiveness, which requires careful evaluation of the disadvantage of sharing data versus the advantage of access to other researchers' data. Free-riding on the willingness of others to share their data should be frowned upon; conversely, developing a highly collaborative and integrated IE community should be viewed as the gold standard in our collective ability to deliver high impact research that provides tangible and valuable scientific contributions to society. A more reproducible scientific workflow in industrial ecology research therefore also has profound ethical consequences, including the valuation of our own work, our role as spenders of public funding, and the contribution of IE research to grand challenges such as sustainability and improved social, economic, and cultural well-being. How we achieve these ideals, however, is completely up to us.

Acknowledgments

We thank Daniel B Müller, Oliver Cencic, Paul Hoekman, Oliver Schwab, Aristide Athanassiadis, Shinichiro Nakamura, and Stefan Giljum for sharing their experiences with data transparency and for providing input to the guideline development.

Supplementary Information

SI1. Data Transparency Task Force mandate

This is the text of the mandate given by the governing council of the International Society for Industrial Ecology for the Data Transparency Task Force (DTTF).

Proposal ISIE Task Force – Open Access Industrial Ecology

Documentation and publication of industrial ecology data

Within the ISIE there is a need for better documentation and accessibility of the work of industrial ecologists, to be able to aggregate, validate, and contribute it to the public, policy makers and companies. IE currently lacks harmonized procedures, standards, and a platform to share open access data, as well as a tradition of publishing the data along with research results. These deficiencies represent some important missed opportunities:

i. It hinders the systematic exploitation of IE results for the greater good of society.

ii. The contribution of IE to international assessment efforts, such as those of the IRP, IPCC, and IPBES, are hampered.

iii. Collaboration within the community is made difficult.

iv. Research results of the different members are 'incompatible' to one another, limiting comparability and building upon previous work.

This lack of properly formatted, documented, and comparable data is nowhere more evident than in the most detailed and specifically focused on IE methods, life cycle assessment, where longstanding efforts have not lead to work that can easily be contributed to the IPCC assessment process.

It is therefore suggested that ISIE sets down a task force charged with coming up with a set of guidelines and propose or develop a data repository for the publication of data in industrial ecology that could become part of the policy of JIE and would be recommended to other journals. It should address life cycle inventories , but also of material stock and flow data, supply and use tables, and other quantitative information about socioeconomic metabolism.

The policy should address following issues:

– Requirement of publishing and giving access to underlying data for relevant papers where admissible in a community-wide data repository.

– Document and publish the code on a repository (such as Github) in a form that makes results reproducible.

– Encourage the use of ISIE tools & code, fostering its continued development.

– Encourage the use of open source tools formats (e.g. R / Python instead of Matlab) and open data (csv instead of Excel), to avoid copyright issues, facilitate reproducibility and offer interfaces to other tools.

– Options to improve transparency, citation of data, providing credit for making data accessible.

– Suggestions, where appropriate, for data formats and nomenclatures.

Following questions should be considered in this work:

– What is the current state of documenting IE studies and making accessible data, considering the entire universe of academic and corporate/-consultancy work?

– What do available databases or repositories contain and how are they assembled?

– Are available data formats widely used and sufficient?

– What can we learn from open access or subscription-based repositories used in other fields?

– What are opportunities offered by big data approaches?

– What degree of documentation and standardization of published data is desirable?

– What copyright and legal issues need to be solved when distributing data to the community?

– How can published data be critiqued and a learning process implemented?

- Can and should we still give room to publish case studies which do not reveal the underlying data? Under what circumstances is this desirable? How can we work with confidential data?

- What incentives can we provide academic and corporate members to contribute?

- Do ISIE member have data from previous work available that could be gifted to the initial efforts?

The task force should come up with a proposal or a set of recommendations to be presented at the ISIE meeting in Chicago, June 2017 and an editorial piece or column in JIE that goes along with it.

Founding members of the task force: Niko Heeren (ETH), Brandon Kuczenski (UCSB), Guillaume Majeau-Bettez (CIRAIG), Rupert Myers (Yale), Stefan Pauliuk (Freiburg), Konstantin Stadler (NTNU).

Niko Heeren and Edgar Hertwich
Zurich/New Haven, September 2016

SI2. Examples for transparent publications in IE

Bulk data

Many IE research projects would be futile without the use of bulk data for industrial processes, material flows, and multi-regional input-output tables (MRIO). Transparency and availability varies greatly across the bulk IE databases. Process and life cycle inventories are made available in life cycle databases, such as ecoinvent or GaBi, and the most complete collection of both free and proprietary databases can be found at *https://nexus.openlca.org/*. Ecoinvent, the most widely used database for scientific LCA research, is a proprietary process inventory database. The compilation and processing of ecoinvent data, however, is documented in detail.

On the other side of the spectrum, most of the six currently available MRIO databases Tukker and Dietzenbacher (2013) provide open access but the transparency of data harmonization steps is often insufficient (Lutter and Giljum 2014).

A database for national material flow accounting is available in an aggregated free version and a proprietary high-resolution section (*http://www.materialflows.net/*). A bulk database of elemental and substance flows and stocks does currently not exist. Other examples of large open datasets/inventories in ISIE community include enipedia, a semantic data store of energy production and flows (*http://enipedia.tudelft.nl/wiki/Main_Page*) and openei, which gathers energy related data (*http://en.openei.org/wiki/Main_Page*).

Journal publications and technical reports

Despite the difficulties in providing data, a number of good examples of partial or complete supply of research data exist in our community.

Detailed life cycle and process inventories have been published along with a number of recent articles, including a battery manufacturing inventory (Ellingsen et al. 2014) and several inventories for passenger vehicles (Hawkins et al. 2013).

Complete datasets for material and energy flow analysis and accounting were published, amongst others, by Kennedy et al. (2015) for the metabolism of megacities by van Eygen Van Eygen et al. (2017) for an MFA of plastics, by Zoboli et al. (2015) for an MFA study on phosphorous, by Northey et al. (2017) for an overview of copper, nickel, and lead-/zinc mines, by Hoekman and Blottnitz (2016) on the urban metabolism of Cape Town , and for product lifetimes by Murakami et al. (2010) (*http://www.nies.go.jp/lifespan/isic_search_e.php*)

The Social Ecology Group at Alpen-Adria University frequently provide whole datasets as additional information at *https://www.aau.at/soziale-oekologie/data-download/*. A new material flow accounting dataset (*http://uneplive.org/country/resources/AT#more-tab1_7*) was published in support of a new UNEP report on global material flows and resource productivity (Schandl et al. 2016).

Within IO, a transparent IO model for the US, USEEIO, is now available including the model builder software (Yang et al. 2017). Lenzen et al. (2017) compiled and provide a time-series (2008-2015) of balanced sub-national, multi-regional supply-and-use tables (MR-SUTs), integrated with a set of socio-economic and environmental accounts, for Australia.

SI3. Procedural transparency and workflow automation in IE

While documentation of laboratory procedures is a core part of science, the era of data-intensive science has brought about a new approach to "digital" methods: the scientific workflow (Lud\"ascher 2006). Derived from transactional workflow management developed in the business world (Singh 1996), scientific workflows provide a way to repeatedly and consistently apply a sequence of processing steps to input data in order to generate scientific results. A signal characteristic of a scientific workflow is the generation of provenance information–in simple terms, indicating how the data were processed–which provides structured documentation of how a scientific result was generated (Davidson and Freire 2008). A number of scientific workflow management systems have been developed, which enable users to perform repetitive tasks by constructing step-by-step procedures.

Ultimately these tools will help to automatically document and reproduce results. Many procedures in IE can be described precisely enough to automate. For instance:

In LCA:

- *Inventory lookup*: Lookup exchanges, LCI results, or LCIA scores for specific processes
- *Emission Characterization:* Lookup the characterization factor for an emission into a given environmental compartment

In MFA:

- *Mass balance*: Given a set of known flows and one unknown flow, into and out of a particular node and including accumulation (i.e., net additions to stock), compute a mass balance and assign it to the unknown flow
- *Stocks In use:* Given a time series of flows and a set of parameters for a lifetime distribution, estimate stocks in use
- *Assumption:* apply an assumption to estimate the magnitude of one flow from another.

- *Aggregation by region*: Given a set of material flows over small spatial scales, compute the total over a larger region

In IOA:

- *Data reconciliation* of partial or conflicting information in national IO tables or trade data.

- *IO model building* by applying a construct to a supply and use table.

- *Footprint calculation and structural path analysis* with IO tables.

Moreover, as observed above, major data sources are increasingly available online, so there is no reason why data retrieval must remain a manual task.

More research is needed to determine whether IE studies can be described consistently enough to be automated. If so, then it will change the meaning of authoring a study. Instead of manipulating spreadsheets and data columns in specialized software, a researcher may spend time precisely identifying data sources and specifying rigorously how data points are combined to compute results. It becomes possible to imagine studies that can be (a) instantly reproduced by another party, and (b) automatically updated when background data are changed.

References

Astudillo, Miguel F., Karin Treyer, Christian Bauer, Pierre Olivier Pineau, and Mourad Ben Amor. 2016. "Life cycle inventories of electricity supply through the lens of data quality: exploring challenges and opportunities." *International Journal of Life Cycle Assessment:* 1–13. ISSN: 16147502. doi:10.1007/s11367-016-1163-0.

Baker, Monya. 2016. "Is there a reproducibility crisis?" *Nature* 533:452–454. ISSN: 0028-0836. doi:10.1038/533452a.

Costello, Mark J., William K. Michener, Mark Gahegan, Zhi-Qiang Zhang, and Philip E. Bourne. 2013. "Biodiversity data should be published, cited, and peer reviewed." *Trends in Ecology & Evolution* 28, no. 8 (August): 454–461. ISSN: 01695347. doi:10.1016/j.tree.2013.05.002. http://linkinghub.elsevier.com/retrieve/pii/S0169534713001092.

Davidson, Susan B, and Juliana Freire. 2008. "Provenance and scientific workflows." In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data - SIGMOD '08,* 1345. New York, New York, USA: ACM Press. ISBN: 9781605581026. doi:10.1145/1376616.1376772. http://www.scopus.com/inward/record.url?eid=2-s2.0-57149126952&partnerID=tZOtx3y1.

Davis, Chris, Igor Nikolic, and Gerard P.J. Dijkema. 2010. "Industrial Ecology 2.0." *Journal of Industrial Ecology* 14, no. 5 (October): 707–726. ISSN: 10881980. doi:10.1111/j.1530-9290.2010.00281.x. http://doi.wiley.com/10.1111/j.1530-9290.2010.00281.x.

Drachen, Thea Marie, Ole Ellegaard, Asger Væring Larsen, and Søren Bertil Fabricius Dorch. 2016. "Sharing Data Increases Citations." *LIBER QUARTERLY* 26, no. 2 (August): 67–82. ISSN: 2213-056X. doi:10.18352/lq.10149. https://www.liberquarterly.eu/article/10.18352/lq.10149.

Ellingsen, Linda Ager Wick, Guillaume Majeau-Bettez, Bhawna Singh, Akhilesh Kumar Srivastava, Lars Ole Val??en, and Anders Hammer Str??mman. 2014. "Life Cycle Assessment of a Lithium-Ion Battery Vehicle Pack." *Journal of Industrial Ecology* 18 (1): 113–124. ISSN: 10881980. doi:10.1111/jiec.12072.

European Commission. 2012. *COMMISSION RECOMMENDATION of 17.7.2012 on access to and preservation of scientific information.* Technical report. Brussels, Belgium: European Commission. doi:10.4403/jlis.it-8649. http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf.

Farber, Gregory K. 2017. "Can data repositories help find effective treatments for complex diseases?" Developing drugs for neurological and psychiatric disorders: challenges and opportunities, *Progress in Neurobiology* 152:200–212. ISSN: 0301-0082. doi:10.1016/j.pneurobio.2016.03.008. http://www.sciencedirect.com/science/article/pii/S0301008215300253.

Frischknecht, Rolf. 2004. "Transparency in LCA-a heretical request?" *The International Journal of Life Cycle Assessment* 9, no. 4 (July): 211–213. ISSN: 0948-3349. doi:10.1007/BF02978595. http://link.springer.com/10.1007/BF02978595.

Gherghina, Sergiu, and Alexia Katsanidou. 2013. "Data Availability in Political Science Journals." *European Political Science* 12 (3): 333–349. ISSN: 1682-0983. doi:10.1057/eps.2013.8.

Graulis, Saulius, Daniel Chateigner, Robert T. Downs, A. F.T. Yokochi, Miguel Quirós, Luca Lutterotti, Elena Manakova, Justas Butkus, Peter Moeck, and Armel Le Bail. 2009. "Crystallography Open Database - An open-access collection of crystal structures." *Journal of Applied Crystallography* 42 (4): 726–729. ISSN: 00218898. doi:10.1107/S0021889809016690.

Hawkins, Troy R., Bhawna Singh, Guillaume Majeau-Bettez, and Anders Hammer Strømman. 2013. "Comparative Environmental Life Cycle Assessment of Conventional and Electric Vehicles." *Journal of Industrial Ecology* 17 (1): 53–64. ISSN: 10881980. doi:10.1111/j.1530-9290.2012.00532.x.

Heath, Garvin A., and Margaret K. Mann. 2012. "Background and Reflections on the Life Cycle Assessment Harmonization Project." *Journal of Industrial Ecology* 16, no. SUPPL.1 (April): S8–S11. ISSN: 10881980. doi:10.1111/j.1530-9290.2012.00478.x. http://doi.wiley.com/10.1111/j.1530-9290.2012.00478.x.

Hertwich, Edgar. 2007. *ISIE - An Offline Community?* Technical report 2. http://is4ie.net/publications/pdf/ISIE%20News%207(2).pdf.

Hertwich, Edgar G. 2014. "ES&T and the IPCC." *Environmental Science & Technology* 48, no. 5 (March): 2517–2518. ISSN: 0013-936X. doi:10.1021/es500491j. http://pubs.acs.org/doi/abs/10.1021/es500491j.

Hey, Tony, and Mike C Payne. 2015. "Open science decoded." *Nature Physics* 11, no. 5 (April): 367–369. ISSN: 1745-2473. doi:10.1038/nphys3313. http://www.nature.com/doifinder/10.1038/nphys3313.

Hoekman, Paul, and Harro von Blottnitz. 2016. "Cape Town's Metabolism: Insights from a Material Flow Analysis." *Journal of Industrial Ecology* (December). ISSN: 1530-9290. doi:10.1111/jiec.12508.

Hothorn, Torsten, and Friedrich Leisch. 2011. "Case studies in reproducibility." *Briefings in Bioinformatics* 12, no. 3 (May): 288–300. ISSN: 1467-5463. doi:10.1093/bib/bbq084. https://academic.oup.com/bib/article-lookup/doi/10.1093/bib/bbq084.

IIASA. 2017. *Databases.* Accessed June 13, 2017. http://www.iiasa.ac.at/web/home/research/researchPrograms/Energy/Databases.en.html.

Kaye, Jane, and Naomi Hawkins. 2014. "Data sharing policy design for consortia: challenges for sustainability." *Genome Medicine* 6 (1): 4. ISSN: 1756-994X. doi:10.1186/gm523. http://genomemedicine.biomedcentral.com/articles/10.1186/gm523.

Kennedy, Christopher A, Iain Stewart, Angelo Facchini, Igor Cersosimo, Renata Mele, Bin Chen, Mariko Uda, et al. 2015. "Energy and material flows of megacities." *Proceedings of the National Academy of Sciences of the United States of America* 112, no. 19 (April). ISSN: 1091-6490. doi:10.1073/pnas.1504315112. http://www.ncbi.nlm.nih.gov/pubmed/25918371.

Kenyon, Jeremy, and Nancy R. Sprague. 2014. "Trends in the use of supplementary materials in environmental science journals." *Issues in Science and Technology Librarianship* 75. ISSN: 10921206. doi:10.5062/F40Z717Z.

Kitchin, Rob. 2014. "Big Data, new epistemologies and paradigm shifts." *Big Data & Society* 1, no. 1 (July): 205395171452848. ISSN: 2053-9517. doi:10.1177/2053951714528481.

Lafuente, Barbara, Robert T. Downs, Hexiong Yang, and Nate Stone. 2016. "1. The power of databases: The RRUFF project." In *Highlights in Mineralogical Crystallography,* 1–30. Berlin, München, Boston: DE GRUYTER. ISBN: 9783110417104. doi:10.1515/9783110417104-003. https://www.degruyter.com/view/books/9783110417104/9783110417104-003/9783110417104-003.xml.

Lenzen, Manfred, Arne Geschke, Arunima Malik, Jacob Fry, Joe Lane, Thomas Wiedmann, Steven Kenway, Khanh Hoang, and Andrew Cadogan-Cowper. 2017. "New multi-regional input–output databases for Australia – enabling timely and flexible regional analysis." *Economic Systems Research* 29, no. 2 (April): 275–295. ISSN: 0953-5314. doi:10.1080/09535314.2017.1315331.

Lutter, Stephan, and S. Giljum. 2014. *Demand-based measures of material flows. A review and comparative assessment of existing calculation methods and data options.* Technical report. Paris, France: OECD / Working Party on Environmental Information.

Mcnutt, Marcia. 2014. "Journals unite for reproducibility." *Science* 346 (6210): 679. doi:10.1038/515007a.

Michener, William K. 2015. "Ecological data sharing." *Ecological Informatics* 29 (P1): 33–44. ISSN: 15749541. doi:10.1016/j.ecoinf.2015.06.010.

Moran, Daniel D., and Richard Wood. 2014. "Convergence Between the Eora, Wiod, Exiobase, and Openeu'S Consumption-Based Carbon Accounts." *Economic Systems Research* 26, no. 3 (July): 245–261. ISSN: 0953-5314. doi:10.1080/09535314.2014.935298. http://www.tandfonline.com/doi/abs/10.1080/09535314.2014.935298.

Murakami, Shinsuke, Masahiro Oguchi, Tomohiro Tasaki, Ichiro Daigo, and Seiji Hashimoto. 2010. "Lifespan of commodities, part I: The creation of a database and its review." *Journal of Industrial Ecology* 14 (4): 598–612. ISSN: 10881980. doi:10.1111/j.1530-9290.2010.00250.x.

National Science Foundation. 2015. *Today's Data, Tomorrow's Discoveries. Increasing Access to the Results of Research Funded by the National Science Foundation.* Technical report. Arlington, VA.: National Science Foundation. https://www.nsf.gov/pubs/2015/nsf15052/nsf15052.pdf.

Nature. 2016. "Where are the data?" *Nature* 537:137–138. ISSN: 0028- 0836. doi:10.1038/537138a. http://www.nature.com/news/announcement-where-are-the-data-1.20541.

Northey, Stephen A., Gavin M. Mudd, Timothy T. Werner, Simon M. Jowitt, Nawshad Haque, Mohan Yellishetty, and Zhehan Weng. 2017. "The exposure of global base metal resources to water criticality, scarcity and climate change." *Global Environmental Change* 44 (October 2016): 109–124. ISSN: 09593780. doi:10.1016/j.gloenvcha.2017.04.004. http://www.sciencedirect.com/science/article/pii/S0959378016303545.

Nosek, B. A., G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, et al. 2015. "Promoting an open research culture." *Science* 348, no. 6242 (June): 1422–1425. ISSN: 0036-8075. doi:10.1126/science.aab2374. http://www.sciencemag.org/cgi/doi/10.1126/science.aab2374.

Owen, Anne. 2017. *Techniques for Evaluating the Differences in Multiregional Input-Output Databases: A Comparative Evaluation of CO2 Consumption-Based Accounts Calculated Using Eora, GTAP and WIOD.*

Pauliuk, Stefan, Guillaume Majeau-Bettez, Christopher L. Mutel, Bernhard Steubing, and Konstantin Stadler. 2015. "Lifting Industrial Ecology Modeling to a New Level of Quality and Transparency: A Call for More Transparent Publications and a Collaborative Open Source Software Framework." *Journal of Industrial Ecology* 00 (June): n/a–n/a. ISSN: 10881980. doi:10.1111/jiec.12316. http://doi.wiley.com/10.1111/jiec.12316.

Pfenninger, Stefan, Joseph DeCarolis, Lion Hirth, Sylvain Quoilin, and Iain Staffell. 2017. "The importance of open data and software: Is energy research lagging behind?" *Energy Policy* 101, no. November 2016 (February): 211–215. ISSN: 03014215. doi:10.1016/j.enpol.2016.11.046. http://linkinghub.elsevier.com/retrieve/pii/S0301421516306516.

Piwowar, Heather A, and Todd J Vision. 2013. "Data reuse and the open data citation advantage." *PeerJ* 1 (October): e175. ISSN: 2167-8359. doi:10.7717/peerj.175. arXiv: peerj.preprints.270v1 [10.7287]. http://www.ncbi.nlm.nih.gov/pubmed/24109559.

Poline, Jean-Baptiste, Janis L. Breeze, Satrajit Ghosh, Krzysztof Gorgolewski, Yaroslav O. Halchenko, Michael Hanke, Christian Haselgrove, et al. 2012. "Data sharing in neuroimaging research." *Frontiers in Neuroinformatics* 6 (April): 1–13. ISSN: 1662-5196. doi:10.3389/fninf.2012.00009. http://journal.frontiersin.org/article/10.3389/fninf.2012.00009/abstract.

Royal Society. 2017. *History of the Royal Society.* Accessed June 13, 2017. http://web.archive.org/web/20130325210916/http://royalsociety.org/about-us/history/.

Schandl, Heinz, Marina Fischer-kowalski, James West, Stefan Giljum, Monika Dittrich, Arne Geschke, Mirko Lieber, Anke Schaffartzik, and Fridolin Krausmann. 2016. *Global Material Flows and Resource Productivity. An Assessment Study of the UNEP International Resource Panel.* Paris. ISBN: 9789280735543.

Speck, Ricky, Susan Selke, Rafael Auras, and James Fitzsimmons. 2015. "Life Cycle Assessment Software: Selection Can Impact Results." *Journal of Industrial Ecology* 00 (March): n/a–n/a. ISSN: 10881980. doi:10.1111/jiec.12245. http://doi.wiley.com/10.1111/jiec.12245.

SpringerNature Group. 2016. *Over 600 Springer Nature journals commit to new data sharing policies.* Accessed June 13, 2017. http://www.springernature.com/br/group/media/press-releases/over-600-springer-nature-journals-commit-to-new-data-sharing-policies/11111248.

Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame. 2011. "Data Sharing by Scientists: Practices and Perceptions." Edited by Cameron Neylon. *PLoS ONE* 6, no. 6 (June): e21101. ISSN: 1932-6203. doi:10.1371/journal.pone.0021101. arXiv: 0803973233. http://dx.plos.org/10.1371/journal.pone.0021101.

Tukker, Arnold, and Erik Dietzenbacher. 2013. "Global multiregional input-output frameworks: An introduction and outlook." *Economic Systems Research* 25, no. 1 (March): 1–19. ISSN: 0953-5314. doi:10.1080/09535314.2012.761179. http://www.tandfonline.com/doi/abs/10.1080/09535314.2012.761179.

Van Eygen, Emile, Julia Feketitsch, David Laner, Helmut Rechberger, and Johann Fellner. 2017. "Comprehensive analysis and quantification of national plastic flows: The case of Austria." *Resources, Conservation and Recycling* 117, Part B (February): 183–194. ISSN: 0921-3449. doi:10.1016/j.resconrec.2016.10.017. http://www.sciencedirect.com/science/article/pii/S0921344916302956.

Van Noorden, Richard. 2015. "Sluggish data sharing hampers reproducibility effort." *Nature* (June). ISSN: 1476-4687. doi:10.1038/nature.2015.17694.

Vines, Timothy H., Arianne Y.K. Albert, Rose L. Andrew, Florence Débarre, Dan G. Bock, Michelle T. Franklin, Kimberly J. Gilbert, Jean-Sébastien Moore, Sébastien Renaut, and Diana J. Rennison. 2014. "The Availability of Research Data Declines Rapidly with Article Age." *Current Biology* 24, no. 1 (January): 94–97. ISSN: 09609822. doi:10.1016/j.cub.2013.11.014. arXiv: 1312.5670. http://linkinghub.elsevier.com/retrieve/pii/S0960982213014000.

Walport, Mark, and Paul Brest. 2011. "Sharing research data to improve public health." *The Lancet* 377, no. 9765 (February): 537–539. ISSN: 01406736. doi:10.1016/S0140-6736(10)62234-9. http://linkinghub.elsevier.com/retrieve/pii/S0140673610622349.

Warren, Elizabeth. 2016. "Strengthening Research through Data Sharing." *New England Journal of Medicine* 375, no. 5 (August): 401–403. ISSN: 0028-4793. doi:10.1056/NEJMp1607282. http://www.nejm.org/doi/10.1056/NEJMp1607282.

Yang, Yi, Wesley W Ingwersen, Troy R Hawkins, Michael Srocka, and David E Meyer. 2017. "USEEIO: A new and transparent United States environmentally-extended input-output model." *Journal of Cleaner Production* 158 (August): 308–318. ISSN: 0959-6526. doi:10.1016/j.jclepro.2017.04.150. http://www.sciencedirect.com/science/article/pii/S0959652617308806.

Zoboli, Ottavia, David Laner, Matthias Zessner, and Helmut Rechberger. 2015. "Added values of time series in material flow analysis: the Austrian phosphorus budget from 1990 to 2011." *Journal of Industrial Ecology.* ISSN: 1530-9290. doi:10.1111/jiec.12381.