Nullius in Verba[1]

Advancing Data Transparency in Industrial Ecology

Edgar Hertwich,[1] Niko Heeren,[2] Brandon Kuczenski,[3] Guillaume Majeau-Bettez,[4,5] Rupert J. Myers,[6] Stefan Pauliuk,[7] Konstantin Stadler,[5] and Reid Lifset[1]

[1]School of Forestry and Environmental Studies, Yale University, New Haven, CT, USA
[2]Institute of Environmental Engineering, ETH Zurich, Zurich, Switzerland
[3]Institute for Social, Behavioral, and Economic Research, University of California, Santa Barbara, CA, USA
[4]Ecole Polytechnique de Montreal, CIRAIG, Montreal, Quebec, Canada
[5]Industrial Ecology Programme, Norwegian University of Science & Technology, Trondheim, Norway
[6]School of Engineering, University of Edinburgh, Edinburgh, United Kingdom
[7]Faculty of Environment and Natural Resources, University of Freiburg, Freiburg, Germany

Keywords:

academic publishing
data accessibility
data transparency
industrial ecology
open science
publication requirements

Supporting information is linked to this article on the *JIE* website

Summary

With the growth of the field of industrial ecology (IE), research and results have increased significantly leading to a desire for better utilization of the accumulated data in more sophisticated analyses. This implies the need for greater transparency, accessibility, and reusability of IE data, paralleling the considerable momentum throughout the sciences. 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 this article, the members of the DTTF present an overview of developments toward transparent and accessible data within the IE community and more broadly. We argue that increased transparency, accessibility, and reusability of IE data will enhance IE research by enabling more detailed and reproducible research, and also facilitate meta-analyses. These benefits will make the results of IE work more timely. They will enable independent verification of results, thus increasing their credibility and quality. They will also make the uptake of IE research results easier within IE and in other fields as well as by decision makers and sustainability practitioners, thus increasing the overall relevance and impact of the field. Here, we present two initial actions intended 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 data openness badges rewarding journal articles that contain transparent and accessible data. These actions will help the IE community to move toward data transparency and accessibility. We close with a discussion of potential future initiatives that could build on the minimum requirements and the data openness badge system.

Introduction

A core mission of industrial ecology (IE) 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 qualitative and quantitative assessments of: life cycle impacts of products, shifting of environmental burdens

Conflict of interest statement: The authors declare no conflict of interest.

[This article was modified after initial publication to update the copyright line.]

Address correspondence to: Edgar Hertwich, School of Forestry and Environmental Studies, Yale University, 195 Prospect Street, New Haven, CT 06511, USA. *Email:* edgar.hertwich@yale.edu.edu; *Web*: http://environment.yale.edu/profile/hertwich/

DOI: 10.1111/jiec.12738

Volume 22, Number 1

between economic sectors, factors that shape the development of industrial symbiosis, and the opportunities for and effects of closing resource loops. IE research is often data intensive and characterized by ongoing improvements to its analytical tools.

The issue of data transparency was identified by the council of the International Society for Industrial Ecology (ISIE) as an important concern. The Society convened the Data Transparency Task Force (DTTF) in late 2016 to develop guidance on best practices and incentives for sharing IE research data and documenting research workflow (see section SI1 in the supporting information available on the Journal's website for 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 was presented to the Society at the 2017 meeting in Chicago, and feedback was solicited through a survey, a special session, from the *Journal of Industrial Ecology* (*JIE*) editorial board, and the ISIE council. The proposal focused on quantitative research. The DTTF recognizes the diversity of the IE research community—in tools, topics, domains, and expertise. The Society is eager to receive feedback on data transparency for qualitative research as well. The present document summarizes the findings of the DTTF, incorporating the feedback received from the IE community, and presents new, draft editorial guidelines for data transparency.

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 also showing examples of good data management practices. We then present the DTTF proposals for best-practice guidelines for transparent, accessible, and reproducible IE research, and a minimum requirement for IE publications that we have suggested be adopted by the *Journal of Industrial Ecology* (*JIE*). The JIE has agreed to implement these suggestions. We close with a discussion of future efforts to progress toward our goal of fully transparent IE research.

Current Trends in Data Openness

The scientific method builds upon reporting and sharing of research results. Sharing allows scientific 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. With much of IE research being computer based, the ability to acquire or develop, store, and utilize large data sets invariably is having a significant and increasing influence on our field.

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 in the collection of scientific data, its storage in electronic format, and the inclusion of electronic supplementary materials with many articles (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. For example, 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 2017). Pfenninger and colleagues (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 upon 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 (Costello et al. 2013; Michener 2015);
- 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).

Cross-Disciplinary Open Data Guidelines

Numerous guidelines around the provision and management of open data have emerged in recent years, across many fields of research and a broad spectrum of organizations ranging from international (OECD 2007) to local organizations. We focus on what we consider to be two exemplary guidelines for scientific research, namely the TOP and the FAIR guiding principles (Wilkinson et al. 2016), and the Center for Open Science (COS)-led Transparency and Openness Promotion (TOP) guidelines (Nosek et al. 2015) discussing perspectives of research funders, data repositories, and academic journals below.

FAIR: Findable, Accessible, Interoperable, Reusable

The FAIR Guiding Principles define a fundamental set of four attributes of open data; they should be: (1) *findable*;

(2) *accessible*; (3) *interoperable*; and (4) *reusable*. It is envisaged that these attributes can be achieved if authors publish appropriate sets of metadata alongside their specific datasets. Here, we summarize the specific guidance provided in Wilkinson and colleagues (2016) for data to meet FAIR criteria:

(1) findable: indexing or archiving (meta)data with unique identifiers (e.g., digital object identifiers [DOIs]) at a searchable resource;
(2) accessible: (meta)data use an open standard for machine readability and are made permanently available.
(3) Interoperable: (meta)data use standard data vocabularies, in a formal, open, and broadly applicable language, and include references to connected data.
(4) reuseable: (meta)data are defined with relevant attributes for reuse such as a clearly defined license statement.

Importantly, FAIR Guiding Principles can usually be fulfilled by archiving data together with an appropriate metadata set in most of the public repositories indexed in re3data.org, a recently established catalogue of such repositories (see the *Data Repositories* section below). For example, the general-purpose data repository Zenodo is fully compliant with FAIR (Zenodo 2017).

TOP: Transparency and Openness Promotion

The COS recently published a set of TOP guidelines to promote open science practices (Nosek et al. 2015). Although this overarching aim is similar in principle to the goals of FAIR, TOP guidelines focus on journals rather than authors. The TOP guidelines describe eight open science standards for journals. These include:

(1) Citation standards: how are data cited?
(2) Data transparency: how are data reported?
(3) Analytic methods (code) transparency: how is code reported?
(4) Research materials transparency: are all materials used in the research methodology reported?
(5) Design and analysis transparency: how transparent are the reported research procedures?
(6) Preregistration of studies: ensuring the existence of the study in a public registry.
(7) Preregistration of analysis plans: certification of *hypothesis-testing* versus *hypothesis-generating* research.
(8) Replication: how will replication studies be published?

As our community moves forward with data transparency incentives tailored to the specific data characteristics of our field, care should be taken to ensure maximum long-term compatibility with major cross-disciplinary guidelines such as the FAIR and TOP principles listed above.

It is worth noting that the notions of preregistration and of replication studies are most applicable to fields that rely primarily on statistical confidence tests.

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 (EC) asks Member States to ensure that "research data that result from publicly funded research become publicly accessible, usable and re-usable" (EC 2012, 6). Similarly, the U.S. National Science Foundation (NSF) now requests that all data resulting from NSF-funded research be deposited in appropriate data repositories (NSF 2015). *Harnessing data* is also one of the ten major strategic directions for future NSF investment.

Therefore, an increasing number of (inter)national funding agencies require data management plans (DMPs) to be defined in submitted proposals, which often closely follow the FAIR Guiding Principles (Wilkinson et al. 2016). This includes: the Swiss National Science Foundation (SNSF) (SNSF 2017); the Europen Union (EU) Commission's Horizon 2020 scheme, which requires authors to establish a DMP in the initial phase of a project—their guidance explicitly referencing FAIR (EC 2016); and the NSF (NSF 2017).

Data Repositories

A number of repositories have been developed in response to the growing need for data storage. The data repository registry re3data now lists over 1,500 individual data repositories from multiple scientific fields ranging from general-purpose ones such as Figshare, Zenodo, and Dryad to subject specific ones such as GenBank for genetic sequence data, PANGAEA for Earth and Environmental Science, or the Interdisciplinary Earth Data Alliance.[2]

Persistent and indexed repositories provide three major advantages relative to the practice of scientists publishing their work individually or through their institutions. First, data hosted at these repositories usually receive a DOI, which creates a static and persistent reference to a specific data set including version information. Second, an analysis of research data availability on journal homepages has shown that availability declines (i.e., links become broken) at a rate of about 17% per year (Vines et al. 2014); archiving data in dedicated repositories prolongs their use. Third, data sets in repositories are often independently indexed by search engines and are thus more readily findable. Therefore, we support the use of persistent and indexed repositories.

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 data availability and the requirements of funding agencies (McNutt 2014). Such actions include increasingly stringent requirements for data sets to be published alongside journal articles. In early 2017, *Nature* implemented a requirement for a data availability statement at the end of articles summarizing whether/how the data necessary to replicate, interpret, and build upon the findings of the paper are available to readers (Nature 2016). Nature also recently established *Scientific Data*, a new journal dedicated to publishing and describing openly accessible data sets.[3] *Science* requires authors to deposit large data sets at an official repository prior to publication,[4] and similar policies exist, for example, for PLOS journals[5] and biomedical journals as with, for example, *Cell* and *Neuron* from Cell Press.[6] Additionally, as mentioned above, there are also ongoing efforts to characterize and promote journal publication policies that increasingly facilitate open science, for example, through implementation of the TOP guidelines (Nosek et al. 2015).

Data Sharing in Industrial Ecology

IE research builds on a substantial amount of secondary data, which are often based on prior IE research, such as life cycle inventories (LCIs) and (multiregional) input-output (I-O) databases, or external sources, such as official government 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, geographical information systems (GIS) data analysis, statistical analysis, and optimization. IE research encompasses a wide range of approaches and, as a result, different data formats and system representations have developed. The establishment of a common data format would thus be desirable, but is currently challenging. There is, however, a growing convergence or at least hybridization of the approaches of material flow analysis (MFA), I-O analysis, and life cycle assessment (LCA). Additionally, as IE researchers often work closely with companies, 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 (Hertwich 2007) and use of interlinked data (Davis et al. 2010), and improved programming practices and data sharing (Pauliuk et al. 2015). The proposals developed by the DTTF represent an effort to find feasible solutions to these challenges.

Benefits of Data Transparency and Long-Term Costs of Business as Usual

The inaccessibility of detailed results in IE research presents a significant lost opportunity. Inaccessible details cannot be used or fully understood by others, nor cross-checked, replicated, and verified, nor become part of larger meta-analyses. The lack of standards for the consistency and comparability of data disclosures is particularly evident in LCA. For example, in the application of LCA to climate-change mitigation, 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 *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's 5th Assessment Report, concerns by participating scientists about the quality of the LCA results negatively impacted the degree to which the results were trusted by the IPCC, and thus also the way that they were communicated in the report and used to support the subsequent policy-making process (Hertwich 2014).

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 access data sets may widen their use, thus enhancing their 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 or reuse it in their own work. 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 reconcile 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 (Lawrence et al. 2016; Tavoni et al. 2014; Riahi et al. 2017). Although model and data comparison occurs within IE, its strategic potential is by no means exhausted (Owen 2017; Moran and Wood 2014; Speck et al. 2016).

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.

Examples of Data Sharing in the Industrial Ecology Community

We have compiled a list of good examples of data and procedural transparency within IE building on community input solicited via e-mail exchanges and forum posts. This list is now too long be shown in full in the paper; it is available in section SI2 in the supporting information on the Web.

Defining Data Transparency for Industrial Ecology

Transparency is key for fostering collaborative science. The task force is convinced that a change toward new data management practices and data transparency are required in the current publication practice worldwide. Incentives are necessary to create an environment that facilitates data contributions and processing. Here, we describe practical recommendations of the DTTF that the *JIE*—the society's journal—will apply to increase data transparency in its publication and review process. We expect discussion of ways to improve data transparency to continue and anticipate a further evolution of the publication standards and practice as the IE community learns from these first efforts. We also hope that other journals in the field follow suit. Our recommendations include modest mandatory requirements to ensure all *JIE* publications meet basic data transparency requirements and propose a series of data openness badges to reward authors that supply well-documented 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 some IE research and associated confidentiality issues; and
2. the variety of data types and analysis methods used, which stems from IE's interdisciplinary and broad-ranging scope, and which, in turn, leads to questions of methodological consistency, interoperability of data, 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 devised to reflect the characteristics of IE research.

We follow a multilayered 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, that is, the observations reported in the paper. 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 for published articles to reward higher levels of data availability and accessibility.

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.

As discussed in the section *Community Engagement*, the DTTF pursued an inclusive approach soliciting feedback from participants of the 2017 ISIE Conference, the ISIE sections, the *JIE* editorial board, and the IE community at large. Survey results are reported in in section SI3 in the supporting information on the Web. The *JIE* editors have agreed to implement the DTTF proposal as part of the *JIE* publication process.

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 inspection 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.

Minimum Publication 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 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.

Minimum Publication Requirement 2: Enumerate primary results: The data that are represented in each graph or figure in an article must be published in numerical form, clearly referenced in the text, and labeled. For example, a simple spreadsheet containing the quantitative data shown in figures and tables in an article fulfills this requirement; such data can be provided in supporting information or in a publicly accessible repository. This requirement 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 visually estimate them from figures or manually copy them from tables and thus avoiding any uncertainties or errors introduced from this process. This requirement aims to 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

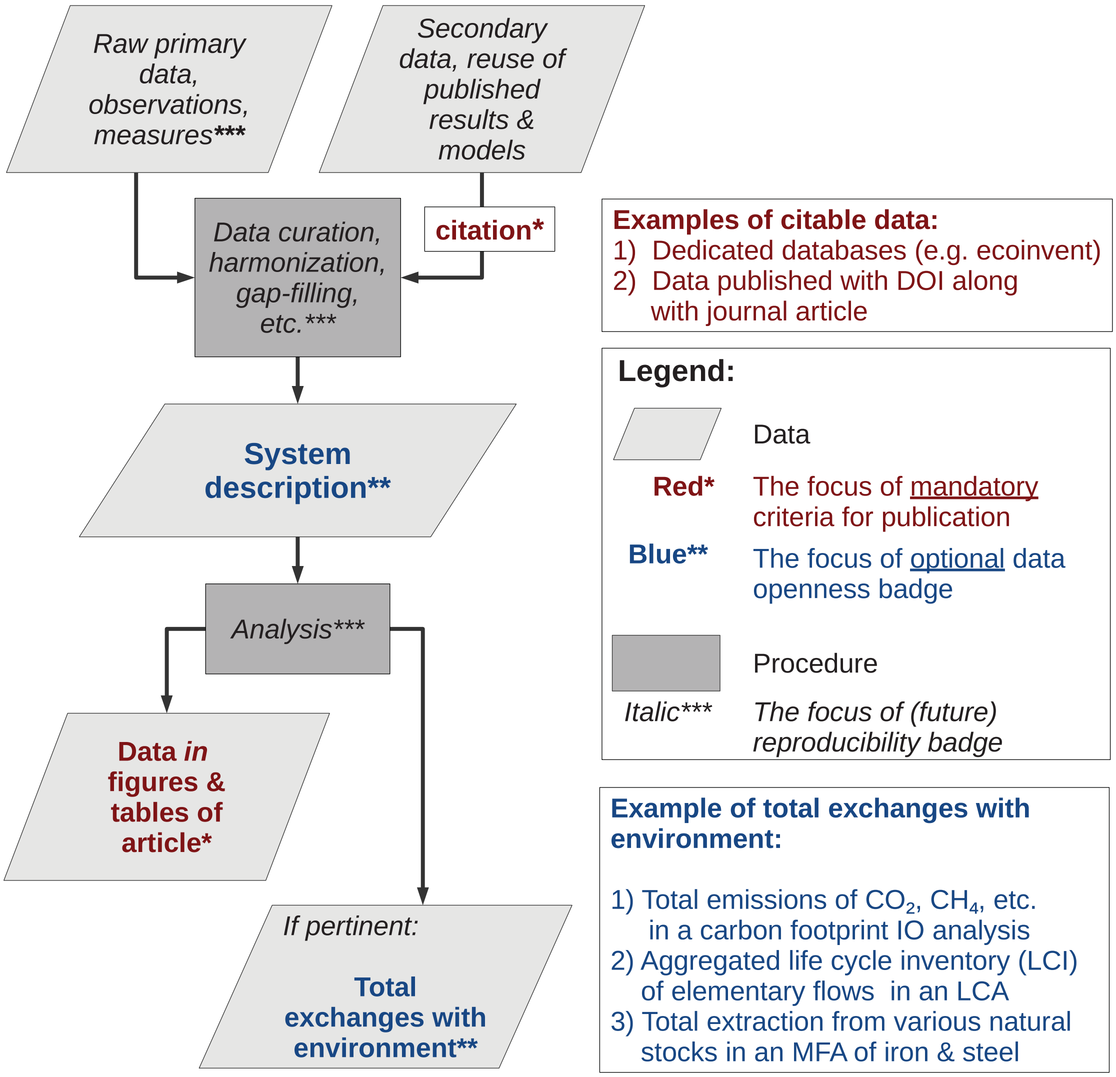

Figure 1 The scope of the proposed minimum transparency criteria (red*) and the proposed data openness badges (blue**) within a conceptual representation of the IE research and publication process (flow chart), with data and manipulations respectively in pale parallelograms and dark gray rectangles. The scope of a future iteration of the badge system is also outlined (*italic****). CH_4 = methane; CO_2 = carbon dioxide; DOI = digital object identifier; IE = industrial ecology; IO = input-output; LCA = life cycle assessment; MFA = material flow analysis.

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 requirements 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 toward full data transparency of IE research.

Data Openness Badges

To reward authors whose articles exhibit higher levels of data openness, accessibility, and interoperability between data formats beyond the strict minimum for publication, we have proposed that an optional data openness badge system be introduced into the *JIE* publication process. It addresses the primary data that underlie the analysis and modeling, rather than the derived results, which are covered by the minimum publication requirements. Eventually, we envision more badges to recognize other contributions such as methods development, harmonization, or the development of free software tools.

Authors will be able to request a data openness badge upon submission of their manuscript and reviewers will 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 toward 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.

Figure 2 The four possible combinations of the Data Openness Badge.

The first dimension addresses data contribution, while the second concerns data accessibility, the latter meaning the interoperability and reusability of the data supplied. The criteria for both dimensions need to be met at least at the silver level to be eligible for a badge. Although further criteria have been suggested, for example, as part of the TOP guidelines discussed above (Nosek et al. 2015), these were not adopted in this initial proposal as it was our goal to focus on those criteria most relevant to IE research while being the least disruptive to the established practice in the community.

Criteria for the Data Contribution Badges

Data Contribution: Gold

This badge indicates that the entire system description is published at the same level of resolution and completeness as was used by the authors to calculate their results.[7]

- These system descriptions notably include, as applicable, the descriptions of all processes, activities, agents, objects, flows, stocks, exchanges with the biophysical environment, system boundaries, and behaviors and actions, *along with links to external or secondary data sets (including licensed databases)*.
- All the primary data and the necessary data citations 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 I-O 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 data set (e.g., ecoinvent) for all secondary data used.
- The data are published under a free, open-content license that explicitly allows use, distribution, and production of derivative work, such as Creative Commons' Public Domain (CC0), Attribution (CC-BY), and Attribution-ShareAlike (CC-BY-SA) licenses.[8]

Data Contribution: Silver

Option 1: In situations where authors cannot share their entire system description, for example, when facing confidentiality issues, they should be nonetheless commended for sharing the detailed description of the nonsensitive parts of the system under a free, open-content license.

- Published data sets 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 the silver level applies to studies that link a technological system to a damage or an impact (e.g., global warming) through multiple types of interactions with the environment (emissions and resource use, e.g., releases of carbon dioxide [CO_2], methane [CH_4], and nitrous oxide [N_2O]). Because of the diversity of characterization methods to translate interactions into impacts, the badge recognizes the benefits for the community of publishing the total interactions of the technological system with the environment in a readily reusable and *uncharacterized* format.

- Example 1: An I-O analysis calculating the carbon footprint of nations would provide the results not only in terms of characterized CO_2-equivalents, but also in terms of the total emissions of the different greenhouse gases (CO_2, CH_4, N_2O, etc.)

- Example 2: In the case of an LCA study, a complete LCI of elementary flows would be published at the systems process level. That means the study would contain the cumulative total for the whole life cycle of each type of emission flow and each type of resource use.

Criteria for Data Accessibility Badges

Accessibility Gold

For this badge, the system description must be 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 read and *understood by humans* in plain text files. Examples of such file formats include plain text, csv, json, and xml files, but compressed versions of these formats are also accepted, such as xlsx and ods spreadsheet formats, but not the proprietary xlsb or xls formats. The system description should also be *machine readable* in the sense that a relevant software can readily distinguish words from numbers, recognize table structures, etc. For example, a system description in a spreadsheet is machine readable, whereas a system description in PDF or word processing formats (.docx, odt, etc.) is not.
- *Direct imports in relevant free 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 nonproprietary 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. Both the data and the calculations of the study are fully embedded in a spreadsheet (e.g., ods, xlsx file). If this spreadsheet can be opened in a free office suite (e.g., LibreOffice) without loss of functionality, it 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).

Accessibility Silver

For this badge, the system description must be formatted such that it can fulfill at least one of the two criteria of the *Accessibility Gold* badge: It must either be directly readable by humans and machines, or be directly importable in a relevant free analysis software.

In order to obtain a badge, it is necessary that the authors respect at least the silver level of both dimensions: data publication and accessibility. Our proposed design for the badges is presented in table 1 and figure 2.

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

	Data contribution	*Data accessibility*
Gold	Entire system description is contributed	Human & machine readable, *and* directly importable into free analysis software
Silver	*Option 1*: Detailed, useful, and self-contained descriptions of significant parts of the system *Option 2*: Total exchanges of the technological system with the environment published in an uncharacterized form	Human & machine readable, *or* directly importable into free analysis software

By proposing the data openness badges, we seek to incentivize the publication of any data that are formatted such that it can be directly imported into a free analysis tool. We expect that more exchange of data 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, the data openness 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 rewarding badges.

In addition to promoting the aforementioned publication of system descriptions, the data openness badges also aim to incentivize another type of useful data disclosure: the provision of uncharacterized elementary flows. 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* to translate them into impacts (global warming, resource depletion, consequences on human health, etc.). However, as science generates 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. Without the publication of the uncharacterized emissions data, however, these IE results are incommensurable with those that rely on older or different characterization methods. We thus seek to reward any study that publishes the total of each exchange of the technological system with the environment in an uncharacterized manner with a data contribution badge.

The badges signal to the reader that the detailed data and system descriptions underlying the published work are open and accessible. We believe that our gold-level badges will allow for recognition of the effort it takes to contribute data sets in an accessible manner. For example, consistent use of the badge system may guide meta-analyses to select studies with levels of transparency sufficient to harmonize system boundaries and assumptions. The inventories supplied by those studies may also be consolidated in IE databases at a later stage. We emphasize that our proposal is to make the badges optional, not a requirement for publication.

The data openness badge system 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 and are therefore not covered by the badges. If the badge system is found to be effective, it could be extended to recognize other contributions to research transparency.

Implementation and Next Steps

With this proposal, the DTTF presents to the IE community its recommendations for greater data transparency. Implementing the suggestions made here affects journal policies, the reviewing process, and, not the least, the workflow of IE researchers. We used a community engagement and expert consultation process, described in section SI3 in the supporting information on the Web to refine our proposal. Further feedback will be required in the implementation and for addressing additional aspects of moving toward open science. To facilitate communication, we created the permanent e-mail address data@is4ie.org as well as the webpage www.is4ie.org/opendata, where additional information and material on data openness is supplied. Here, we describe the first implementation of the badge system in the *JIE*, the future development of data referencing, data formats, and databases, and the link between data and procedural transparency (in section SI4 in the supporting information on the Web). We sincerely hope that these actions and this proposal engage a constructive and sustained dialogue within the IE community on the important topic of data availability and open science.

Implementation in the Journal of Industrial Ecology

Based on the feedback obtained, we submitted a refined set of recommendations to the ISIE council in December 2017, including a proposal for the minimum requirements and the badge system to become part of the regular *JIE* review process. We also created and submitted to the editorial board suggested text for the *JIE* author, editor, and reviewer guidelines regarding the introduction of the minimum publication requirement and data openness badge system into the publication process of the *JIE*. It is our goal that authors who wish to obtain a data openness badge will be able to indicate so in the *JIE* manuscript submission process, and editors and reviewers will receive clear instructions for the review of the different data contribution and accessibility criteria. The status of the implementation of our proposals can be followed on the *JIE* website.

We encourage the editors-in-chiefs of other journals of the field to consider implementing the requirements and incentives proposed here or elsewhere, for example, the COS-led TOP guidelines.

Data Referencing

During our work, we realized that practices for *data referencing*, that is, the reference to or citation of individual data sets or parts of larger databases, are diverse and not formalized. Clearly, there must be an evolution of practices in data referencing, and a development of procedures for giving credit to underlying contributors. The process of data referencing should be considered as parallel to the evolution of data structures and databases for IE.

Data Structure and Database Development in Industrial Ecology

One of the most important opportunities arising from data openness is that it facilitates data processing in follow-up studies, for example, for meta-analysis. Our current proposal aims at making data available in some convenient form, but does not specify the data format itself. Data providers are aware of the benefit of supplying their and using other researchers' data in widely accepted and standardized formats, and there is a clear need in the different subfields of IE to continue the process of data format development. We see the method-specific data format development as being the responsibility of individual ISIE sections or other, related organizations.

Ultimately, structured data from a wide spectrum of studies could be integrated into a common database so that researchers have the opportunity to query multiple relevant data sets at the same time. However, such an approach requires that data are sufficiently harmonized, for example, by using compatible classifications, across several major IE techniques (e.g., I-O, LCA, and MFA). Care must be taken, however, when developing harmonization processes, as authors may perceive a data harmonization step in their work as an undue burden without tangible benefit. Moreover, as analysis methods evolve (e.g., from LCA to hybrid IO-LCA), strict data formats may become inadequate for studies presenting novel methodology, which may therefore become an obstacle for innovation.

We believe that the further development of data structures and their harmonization across the different subfields of IE presents a clear opportunity to substantially advance IE research and that should play a central in the process toward data openness.

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. Open results alone are often insufficient to allow researchers to infer the underlying procedures and assess their correctness and validity. Therefore, open data need to be accompanied by a transparent description of procedures in order to achieve full reproducibility of study results.

In practice, the steps required to organize data sources, process data, and extract results are study dependent and nuanced. As a consequence, even studies with high degrees of data openness can be difficult to compare. To progress beyond pure data contribution and toward validation, verification, and reproducibility of results, a higher transparency of methodology documentation is needed.

We ultimately believe that the IE community needs to move beyond the summary descriptions often presented in methods sections of papers and toward the publication of detailed research procedures and computational scripts that fully *reproducible research* requires. This shift in documentation and publication practice, however, comes at a cost. 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 fully reproducible IE research can only be achieved through a step-by-step process, and that this document provides an important preliminary step toward realizing this aim.

The Future of the International Society for Industrial Ecology Data Transparency Task Force

As part of their current mandate, the members of the DTTF will continue to develop and improve guidelines for achieving higher levels of data openness based on the experience gained during the coming implementation process. The focus of our work for 2018 is to facilitate the implementation of data openness in the *JIE*. The ISIE council will review the mandate of the task force in regular intervals and adjust it to keep pace with the development of data openness in the field and the requirements set by the Society.

Final Thoughts and Conclusions

The effort for higher data transparency and accessibility has just begun, and our proposal for minimum requirements and the data openness badge system will continue to be open for debate and revision. We believe that the contribution of IE research cannot be wholly realized until results become more readily comparable, integrated, citable, and reusable. In order to achieve fully reproducible IE research, the data contribution and accessibility standards suggested here would further require data openness to be linked to procedural transparency and harmonization of data structures and—to some extent—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 implications related to 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 IE research therefore also has profound ethical consequences, including the valuation of our own work, our role as recipients of public funds, and the contribution of IE research to grand challenges such as sustainable development and improved social, economic, and cultural well-being.

Acknowledgments

We thank Aristide Athanassiadis, Jesse Ausubel, Oliver Cencic, Nina Eisenmenger, Stefan Giljum, Paul Hoekman, Daniel B. Müller, Shinichiro Nakamura, Raymond Paquin, Oliver Schwab, Valerie Thomas, Gregor Wernet, and 33 anonymous survey participants for sharing their experiences with data transparency and/or for providing input to the guideline development.

Notes

1. "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 the Society to withstand the domination of authority and to verify all statements by an appeal to facts determined by experiment (Royal Society 2017).
2. See re3data: www.re3data.org/, Figshare: www.re3data.org/, Zenodo: www.zenodo.org/, Dryad: datadryad.org/, GenBank: www.ncbi.nlm.nih.gov/genbank/, PANGEA: https://pangaea.de/, IEdata: www.iedadata.org/.
3. See www.nature.com/sdata/about.
4. See www.sciencemag.org/authors/scie editorial-policies.
5. See http://journals.plos.org/plosone/s/data-availability.
6. See www.cell.com/cell/authors.
7. It should be noted that the gold level for data contribution in our progressive badge system is well aligned with the "Open Data" badge of the COS (Open Science Framework 2017). In our badge framework, we rather opted for a more progressive approach to data contribution, with two levels to reflect IE's focus on industrial processes that can prove proprietary. We also stress accessibility as a second dimension because of our community's reliance on databases with diverse formats and focus.
8. See https://creativecommons.org/share-your-work/public-domain/cc0/, https://creativecommons.org/licenses/by/4.0/, and https://creativecommons.org/licenses/by-sa/4.0/.

References

Astudillo, M. F., K. Treyer, C. Bauer, P. O. Pineau, and M. B. Amor. 2016. Life cycle inventories of electricity supply through the lens of data quality: Exploring challenges and opportunities. *The International Journal of Life Cycle Assessment* 22(3): 374–386.

Baker, M. 2016. Is there a reproducibility crisis? *Nature* 533(7604): 452–454.

Costello, M. J., W. K. Michener, M. Gahegan, Z.-Q. Zhang, and P. E. Bourne. 2013. Biodiversity data should be published, cited, and peer reviewed. *Trends in Ecology & Evolution* 28(8): 454–461.

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

Drachen, T. M., O. Ellegaard, A. V. Larsen, and S. B. F. Dorch. 2016. Sharing data increases citations. *Liber Quarterly* 26(2): 67–82.

EC (European Commission). 2012. *Commission Recommendation of 17.7.2012 on access to and preservation of scientific information*. Brussels: EC. http://ec.europa.eu/research/science-society/document%7B%5C_%7Dlibrary/pdf%7B%5C_%7D06/recommendation-access-and-preservation-scientific-information%7B%5C_%7Den.pdf. Accessed 25 November 2017.

EC (European Commission). 2016. *Guidelines on FAIR Data Management in Horizon 2020 v3.0*. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf. Accessed 12 October 2017.

Farber, G. K. 2017. Can data repositories help find effective treatments for complex diseases? *Progress in Neurobiology* 152: 200–212.

Frischknecht, R. 2004. Transparency in LCA—A heretical request? *The International Journal of Life Cycle Assessment* 9(4): 211–213.

Gherghina, S. and A. Katsanidou. 2013. Data availability in political science journals. *European Political Science* 12(3): 333–349.

Graulis, S., D. Chateigner, R. T. Downs, A. F. T. Yokochi, M. Quirós, L. Lutterotti, E. Manakova, J. Butkus, P. Moeck, and A. L. Bail. 2009. Crystallography open database—An open-access collection of crystal structures. *Journal of Applied Crystallography* 42(4): 726–729.

Heath, G. A. and M. K. Mann. 2012. Background and reflections on the life cycle assessment harmonization project. *Journal of Industrial Ecology* 16(s1): S8–S11.

Hertwich, E. 2007. ISIE—An offline community? *ISIE News*. Vol. 7. http://is4ie.org/publications/pdf/ISIE. Accessed 14 April 2017.

Hertwich, E. G. 2014. ES&T and the IPCC. *Environmental Science & Technology* 48(5): 2517–2518.

Hey, T. and M. C. Payne. 2015. Open science decoded. *Nature Physics* 11(5): 367–369.

Hothorn, T. and F. Leisch. 2011. Case studies in reproducibility. *Briefings in Bioinformatics* 12(3): 288–300.

Kaye, J. and N. Hawkins. 2014. Data sharing policy design for consortia: Challenges for sustainability. *Genome Medicine* 6(1): 4.

Kenyon, J. and N. R. Sprague. 2014. Trends in the use of supplementary materials in environmental science journals. *Issues in Science and Technology Librarianship* 75. www.istl.org/14-winter/refereed5.html. Accessed 25 November 2017.

Lafuente, B., R. T. Downs, H. Yang, and N. Stone. 2016. 1. The power of databases: The RRUFF project. In *Highlights in mineralogical crystallography*, 1–30. Berlin; München, Germany; Boston, MA, USA: De Gruyter. www.degruyter.com/view/books/9783110417104/9783110417104-003/9783110417104-003.xml. Accessed 25 November 2017.

Lawrence, D. M., G. C. Hurtt, A. Arneth, V. Brovkin, K. V. Calvin, A. D. Jones, C. D. Jones, et al. 2016. The land use model intercomparison project (LUMIP) contribution to CMIP6: Rationale and experimental design. *Geoscientific Model Development* 9(9): 2973–2998.

McNutt, M. 2014. Journals unite for reproducibility. *Science* 346(6210): 679.

Michener, W. K. 2015. Ecological data sharing. *Ecological Informatics* 29(p1): 33–44.

Moran, D. D. and R. Wood. 2014. Convergence between the Eora, Wiod, Exiobase, and Openeu's consumption-based carbon accounts. *Economic Systems Research* 26(3): 245–261.

Nature. 2016. Where are the data? *Nature* 537(7619): 137–138.

Noorden, R. Van. 2017. Sluggish data sharing hampers reproducibility effort. *Nature*. https://doi.org/10.1038/nature.2015.17694.

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(6242): 1422–1425.

NSF (National Science Foundation). 2015. *Today's data, tomorrow's discoveries. Increasing access to the results of research funded by the National Science Foundation*. Arlington, VA, USA: NSF. www.nsf.gov/pubs/2015/nsf15052/nsf15052.pdf. Accessed 15 October 2017.

NSF (National Science Foundation). 2017. NSF ENG data management plan requirements. www.nsf.gov/eng/general/dmp.jsp. Accessed 21 October 2017.

OECD (Organization for Economic Cooperation and Development). 2007. *OECD Principles and Guidelines for Access to Research Data from Public Funding*. *Data Science Journal*. www.oecd.org/sti/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm. Accessed 20 October 2017.

Open Science Framework. 2017. Badges to Acknowledge Open Practices. https://osf.io/tvyxz/wiki/home/. Accessed 25 November 2017.

Owen, A. 2017. *Techniques for evaluating the differences in multiregional input-output databases: A comparative evaluation of CO_2 consumption-based accounts calculated using Eora, GTAP and WIOD*. Developments in input-output analysis, edited by E. Dietzenbacher and T. Wiedmann. Cham, Switzerland: Springer International AG.

Pauliuk, S., G. Majeau-Bettez, C. L. Mutel, B. Steubing, and K. 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* 19(6): 937–949.

Pfenninger, S., J. DeCarolis, L. Hirth, S. Quoilin, and I. Staffell. 2017. The importance of open data and software: Is energy research lagging behind? *Energy Policy* 101: 211–215.

Piwowar, H. A. and T. J. Vision. 2013. Data reuse and the open data citation advantage. *PeerJ* 1: e175.

Poline, J.-B., J. L. Breeze, S. Ghosh, K. Gorgolewski, Y. O. Halchenko, M. Hanke, C. Haselgrove, et al. 2012. Data sharing in neuroimaging research. *Frontiers in Neuroinformatics* 6: 1–13.

Riahi, K., D. P. van Vuuren, E. Kriegler, J. Edmonds, B. C. O'Neill, S. Fujimori, N. Bauer, et al. 2017. The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change* 42: 153–168.

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

SNSF (Swiss National Science Foundation). 2017. Data Management Plan (DMP)—Guidelines for researchers. www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/data-management-plan-dmp-guidelines-for-researchers.aspx. Accessed 21 October 2017.

Speck, R., S. Selke, R. Auras, and J. Fitzsimmons. 2016. Life cycle assessment software: Selection can impact results. *Journal of Industrial Ecology* 20(1): 18–28.

Tavoni, M., E. Kriegler, K. Riahi, D. P. van Vuuren, T. Aboumahboub, A. Bowen, K. Calvin, et al. 2014. Post-2020 climate agreements in the major economies assessed in the light of global models. *Nature Climate Change* 5(2): 119–126.

Tenopir, C., S. Allard, K. Douglass, A. U. Aydinoglu, L. Wu, E. Read, M. Manoff, and M. Frame. 2011. Data sharing by scientists: Practices and perceptions, edited by C. Neylon. *PLoS One* 6(6): e21101.

Vines, T. H., A. Y. K. Albert, R. L. Andrew, F. Débarre, D. G. Bock, M. T. Franklin, K. J. Gilbert, J.-S. Moore, S. Renaut, and D. J. Rennison. 2014. The availability of research data declines rapidly with article age. *Current Biology* 24(1): 94–97.

Walport, M. and P. Brest. 2011. Sharing research data to improve public health. *The Lancet* 377(9765): 537–539.

Warren, E. 2016. Strengthening research through data sharing. *New England Journal of Medicine* 375(5): 401–403.

Wilkinson, M. D., M. Dumontier, Ij. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, et al. 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3: 160018.

Zenodo. 2017. Zenodo. https://zenodo.org. Accessed 21 October 2017.

Supporting Information

Supporting information is linked to this article on the *JIE* website:

Supporting Information S1: This supporting information includes information on the Data Transparency Task Force mandate (SI1), examples for transparent publications in IE (SI2), results from a survey about data openness conducted by the JIE and DTTF (S13), and a discussion of procedural transparency and workflow automation in IE (SI4).