Nullius in Verba

Citation for published version:

Digital Object Identifier (DOI):
10.1111/jiec.12738

Link:
Link to publication record in Edinburgh Research Explorer

Published In:
Journal of Industrial Ecology

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
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 Kuczynski (UCSB), Guillaume Majeau-Bettez (CIRAIG), Rupert Myers (University of Edinbugh), 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 Kenned et al. (2015) for the metabolism of megacities by Eygen et al. (2017) for an MFA of plastics, by Zoboli et al. (2015) for an MFA study on phosphorus, 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/sozialeoekologie/data-download/. A new material flow accounting dataset (http://uneplive.org/country/res 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. Survey results

An introduction of minimum publication requirements and the data openness badges requires acceptance by the IE community; therefore, the DTTF and JIE have undertaken a community engagement process.

In a first step, we solicited feedback on our proposal and on other options to enhance data transparency and accessibility from the entire IE community during the summer of 2017. To that end we created the permanent email address data@is4ie.org and posted the topic on IE social media channels and announced it during the 2017 ISIE-ISSST conference in Chicago. Furthermore, we invited all interested parties to participate in an anonymous online survey posted on www.is4ie.org/opendata. JIE editors and section board members were consulted as well. We were particularly interested in hearing about whether the level of ambition of the badge system was reasonable, how the measures proposed would affect the 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 received 33 mostly positive responses, with around 15 respondents of the MFA, IO, and LCA communities, respectively (double-affiliations were allowed). We list the opinions polled in Table S3.1 and the main items of critique and our response to them in table y.

The survey results may not be representative of our entire community but they gave us the confidence to move forward and the specific feedback to refine our suggestions. In particular, it became clear that more specific guidance and examples are needed to specify the exact changes that are needed in publishing process, to make the whole procedure predictable and void of surprises.

Based on the feedback obtained, we submitted a refined set of recommendations to the ISIE council in December 2017, including a proposal for the amendment of the JIE publication requirements regarding the introduction of the data openness badge system.
**Table S1:** General opinion towards the data openness guidelines among the survey participants

<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you agree that there should be a minimum data openness requirement for all JIE articles publishing quantitative results?</td>
<td>Yes: 31&lt;br&gt;No: 1&lt;br&gt;No answer: 4</td>
</tr>
<tr>
<td>If yes: do you agree with the minimum requirements proposed by the DTTF (proper citation of all secondary data and databases and delivery of all primary results in spreadsheets or the like) or should there be fewer or more requirements?</td>
<td>More: 6&lt;br_OK: 22&lt;br_fewer: 3&lt;br_No answer: 4</td>
</tr>
<tr>
<td>Do you agree that the JIE should establish an optional data openness badge to distinguish particularly open data-driven research articles?</td>
<td>Yes: 24&lt;br_No: 6&lt;br_No answer: 5</td>
</tr>
<tr>
<td>If Yes: Do you agree with the proposed structure of the badges (two dimensions: data availability/transparency and data accessibility, two levels for each dimension) and with the requirements for the different levels?</td>
<td>Yes: 19&lt;br_No: 6&lt;br_No answer: 10</td>
</tr>
<tr>
<td>Are the requirements for the minimum standard and the different badge grades clear and unambiguous?</td>
<td>Yes: 18&lt;br_No: 9&lt;br_No answer: 8</td>
</tr>
<tr>
<td>Do you see any potential conflicts between the requirements for data openness and your current workflow (funder, software, ...)?</td>
<td>Yes: 21&lt;br_No: 9&lt;br_No answer: 5</td>
</tr>
<tr>
<td>Do you see a benefit in applying for a data openness badge for your own research published in JIE?</td>
<td>Yes: 20&lt;br_No: 10&lt;br_No answer: 5</td>
</tr>
<tr>
<td>What is your preferred way of making data stemming from your published research available? (multiple answers were possible)</td>
<td>Public repositories: 15&lt;br_Community repos.: 8, Journal repos.: 12, Institution or private: 10</td>
</tr>
<tr>
<td>How would you describe the current state of data openness within your specialisation of industrial ecology?</td>
<td>Problematic: 15&lt;br_Some improvement needed: 13, sufficient: 2</td>
</tr>
<tr>
<td>How will industrial ecology, according to your opinion, benefit most from data openness?</td>
<td>Faster progress: 25&lt;br_Higher impact: 16&lt;br_Cumulative Research easier: 22</td>
</tr>
<tr>
<td>Where do you see the main responsibility for data openness?</td>
<td>Authors: 25; Editors and reviewers: 13; Funders: 10; Community: 5</td>
</tr>
</tbody>
</table>
### Table S2: Overview of core feedback items on the data openness guidelines proposal and our response.

<table>
<thead>
<tr>
<th>Feedback (ordered by no. of times mentioned)</th>
<th>#</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Need to resolve IN DETAIL the distinction between proprietary or licensed data that cannot be published and data that can be published with proper citation</td>
<td>11</td>
<td>We checked the pertinent copyright laws and the end user license agreements of a major database of the field to refine our guidelines regarding what constitutes a fair use of data from a licensed database. We discussed the results with the database providers and have adjusted the draft author guidelines. Also, our proposed framework promotes citing and linking to data from external sources, rather than reproducing or “re-publishing” these data.</td>
</tr>
<tr>
<td>Assessing compatibility with badge requirements may be too much work for reviewers</td>
<td>9</td>
<td>We agree that reviewers will have an extra effort when becoming familiar with the badge criteria. We formulated a set of detailed guidelines to make the work of the reviewers as easy as possible.</td>
</tr>
<tr>
<td>There is a problem of subjectivity when applying badge requirements</td>
<td>5</td>
<td>We agree and suggest the following mitigations of that risk: a) Detailed guidelines with specific examples are provided for author, reviewers, and editors. b) A <em>JIE</em> data editor has been appointed to address and manage these issues.</td>
</tr>
<tr>
<td>Data openness may be hindered through data formatting problems, software constraints, lack of metadata etc.</td>
<td>5</td>
<td>That is true and there will be cases where a data openness badge cannot be obtained because certain data required cannot be extracted from proprietary software. For the majority of all cases the relatively flexible formatting requirements posed by the badge system should not be a significant obstacle to data supply.</td>
</tr>
<tr>
<td>Good scientific practice requires researchers to make data available upon request, that interaction also is valuable feedback on who is interested in research. How does that relate to the badge system?</td>
<td>5</td>
<td>The requirement to publish certain results and data along with each paper makes it easier for readers with no personal connections to the authors to access this information, and we therefore see the minimum requirements and the badge system as a fairer solution than personal contact. Of course we encourage readers to personally contact the authors to clarify further questions and provide feedback on the work.</td>
</tr>
<tr>
<td>Need to provide detailed examples of good practice and detailed guidelines.</td>
<td>4</td>
<td>During revision of this document and author guidelines, we made a significant effort to provide good examples of data citation, data formatting, and a description of the requirements of the different badge levels.</td>
</tr>
<tr>
<td>The badge system may be too complicated.</td>
<td>4</td>
<td>We agree that data openness has many dimensions and aspects, and covering most of them could stymie the publication process. We are confident that our two-dimensional approach, encompassing</td>
</tr>
</tbody>
</table>
### SI4. 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 (Ludäscher et al. 2006). Derived from transactional workflow management developed in the business world (Singh and Vouk 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**


S-9