

# Transparent Integration and Sharing of Life Cycle Sustainability Data with Provenance

Emil Riis Hansen<sup>✉1</sup>[0000-0003-4103-1244], Matteo Lissandrini<sup>1</sup>[0000-0001-7922-5998], Agneta Ghose<sup>2</sup>[0000-0003-1972-1433], Søren Løkke<sup>2</sup>[0000-0002-4544-1788], Christian Thomsen<sup>1</sup>[0000-0002-0862-0509], and Katja Hose<sup>1</sup>[0000-0001-7025-8099]

<sup>1</sup> Department of Computer Science, Aalborg University

<sup>2</sup> Department of Planning, Aalborg University

{emilrh,matteo,chr,khose}@cs.aau.dk, {agneta,loekke}@plan.aau.dk

**Abstract.** Life Cycle Sustainability Analysis (LCSA) studies the complex processes describing product life cycles and their impact on the environment, economy, and society. Effective and transparent sustainability assessment requires access to data from a variety of heterogeneous sources across countries, scientific and economic sectors, and institutions. Moreover, given their important role for governments and policymakers, the results of many different steps of this analysis should be made freely available, alongside the information about how they have been computed in order to ensure accountability. In this paper, we describe how Semantic Web technologies in general and PROV-O in particular, are used to enable transparent sharing and integration of datasets for LCSA. We describe the challenges we encountered in helping a community of domain experts with no prior expertise in Semantic Web technologies to fully overcome the limitations of their current practice in integrating and sharing open data. This resulted in the first nucleus of an open data repository of information about global production. Furthermore, we describe how we enable domain experts to track the provenance of particular pieces of information that are crucial in higher-level analysis.

**Keywords:** Open Data · Provenance · Sustainability Analysis.

## 1 Reproducibility

### 1.1 Github Repo

The following link provides access to an open github repository, containing all documentation and steps needed to replicate results from the paper:

<https://github.com/IKnowLogic/BONSAI-reproducibility>

### 1.2 Datasets

The data for the *Exiobase* dataset can be obtained from the exiobase website: <https://www.exiobase.eu/>. We use the hybrid Exiobase dataset version

3.3.17. Data access needs a free user. For convenience, we also provide a copy of the dataset freely available at the following url:  
<https://silol1.sciencedata.dk/shared/20ee45e130a37e87c5b19e07b81b61ec>

The data for the *YSTAFDB* dataset can be downloaded from the following url:  
[https://www.sciencebase.gov/catalog/file/get/5b9a7c28e4b0d966b485d915?f=\\_\\_disk\\_\\_0f%2F58%2Fa7%2F0f58a74db669ee5418f36a698bc85781e867e0ab](https://www.sciencebase.gov/catalog/file/get/5b9a7c28e4b0d966b485d915?f=__disk__0f%2F58%2Fa7%2F0f58a74db669ee5418f36a698bc85781e867e0ab).

### 1.3 License

**BONSAI Software.** Using the BONSAI software tools (provided on <https://github.com/BONSAMURAI>) is bound by the BSD 3-Clause License.

**Exiobase Dataset.** Using the exiobase dataset is bound by the Creative Commons license: <https://creativecommons.org/licenses/by-sa/4.0/>.

**YSTAFDB Dataset.** Using the ystafdb dataset is bound by the Creative Commons license: <https://creativecommons.org/publicdomain/zero/1.0/>.