

Combining big data sources consistently

LCA Discussion Forum 64

March 30th, 2017

Gregor Wernet

Combining big data sources



- Overview of a few recent projects that combined big data sources for LCA work
- Experiences and recommendations for practitioners

Not all big data is Big data

- Life cycle background data systems are large and complex
- Research and efforts on “**Big data**” are not transferable
 - Big data is a moving target
 - Currently tens of Terabytes to many Petabytes
 - ecoinvent 3.3 matrix is about 35 Megabytes
 - Common features are unfiltered, diverse, live access
- This presentation is on combining large sets of data for LCA purposes

- Several databases exist for Social LCA
 - Currently sector-based
- ecoinvent contains both a model of human-driven processes and their interactions with the environment
- Existing unit process structure of human activities can be the foundation for process-based Social LCA work
 - soca, work with SHDB

- Developed a matching
 - Value generated, region and sector
 - Data for a basic matching introduced with version 3.3
- Example: 12\$ generated in the textiles sector in India is associated with a specific set of Social LCA flows
- The result is an adapted unit process database that contains both environmental and social elementary flows
 - An algorithm merges the social flows into an existing v3.3 copy

- What is the value of combined data sources?
 - Unit process-based Social LCA
- Goal and scope:
 - ecoinvent data was developed for environmental assessments
 - Social assessments may require spacial resolutions or technology differentiations that ecoinvent does not offer
 - Activities may be missing
 - Geographical disaggregation may be needed
 - ...?

- ecoinvent is involved in the PEF in several tenders
- Consortium lead for the “chemicals” data call
 - 261 products
- One requirement is the use of existing PEF data for energy and transport
 - Available in ILCD format, supplied by thinkstep
 - Some issues with the data, e.g. naming inconsistencies
 - Matching of elementary flows and compartments
 - Format conversion to ILCD
 - Unit conversions (sometimes require assumptions)

Merging PEF energy and transport data



- Mapping table between thinkstep data and the ecoinvent data they shall replace
- Generated a matrix with consistent relinking
- Several ite data
- Reviews

activityName	geography	name	Climate change midpoint_3.3	Climate change midpoint_thinkstep-ecoinvent	Climate change midpoint_thinkstep-ecoinvent /3.3	Climate change midpoint_chemicals	Climate change midpoint_chemicals /thinkstep-ecoinvent
selenium production	RER	selenium	2.355	2.323	0.986	50.683	21.815
lime production, milled, loose	CH	lime	0.005	0.005	1.166	0.025	4.670
citric acid production	RER	citric acid	2.732	2.981	1.091	6.100	2.046
sodium hypochlorite production,	RER	sodium hypochlorit	0.891	0.796	0.894	1.510	1.896
activated silica production	GLO	activated silica	1.757	1.621	0.923	2.591	1.598
lubricating oil production	RER	lubricating oil	1.013	0.957	0.945	1.384	1.447

- First, know your data sources and their scope
- Specify your procedures, iterate and log progress
 - Eliminate manual steps due to effort and error potential
 - Use or develop fast, simple tools
- Review and assess the resulting data from many different angles
- Prepare maintenance for further iterations

The world's most
consistent and
transparent Life
Cycle Inventory



Thank you

