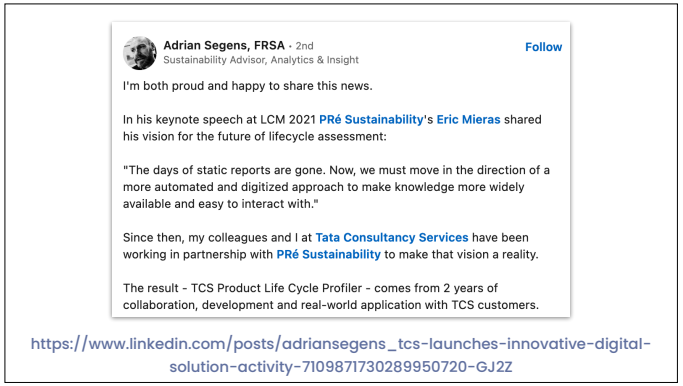
Beyond static inventory datasets

Chris Mutel

LCA DF 84, 21.09.2023

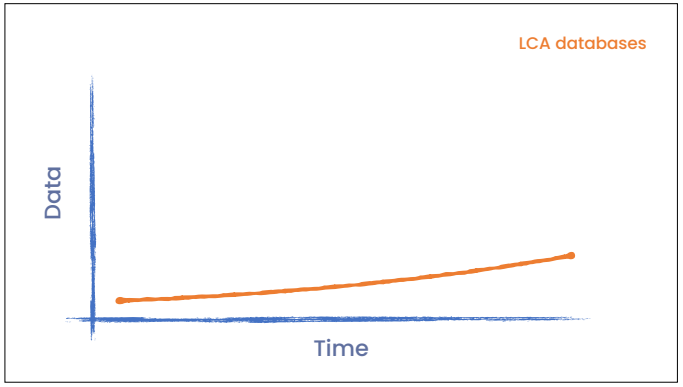
Nic and Emilia, the CEO and CTO of ecoinvent
*It's funny because Nic is from Australia
**Jokes that have to be explained are the funniest kind of jokes

Last year I had a conversation that changed my life. I met with Nic and Emilia, and they **believed** in LCA, in a way that many people in this space don't. Nic really believes in the mission - high quality environmental decision support for **everyone** - and already realised that we need to think about **scale**. Our current processes don't scale.

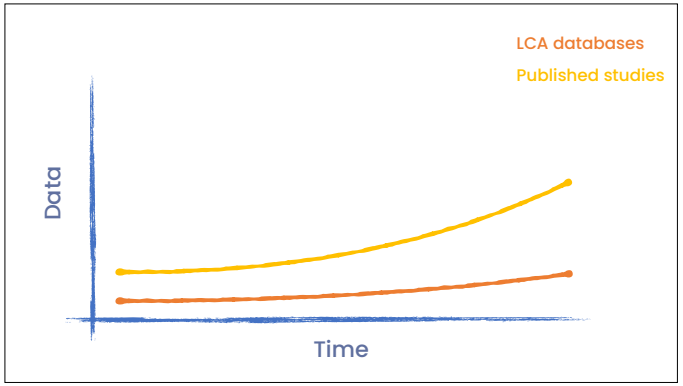


This showed up on my LinkedIn page a day before the discussion forum, and Eric is in the audience. This is a solution for businesses to calculate their corporate and product footprints at scale, with automated data processing and the availability to easily adjust modules (automated LCI datasets) for specific customers. There is a lot of development in this space, and companies like Watershed and Sweep are pushing similar ideas.

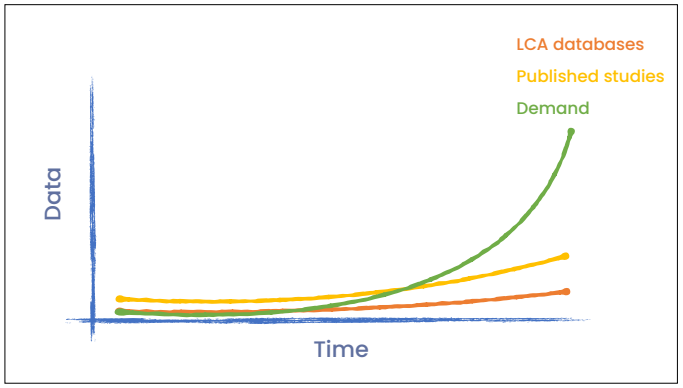
Today I want to talk about the data needed to make these initiatives work - what we sometimes call background data. Here we don't have movement on anywhere near the same scale.



The number of LCI datasets available in background databases like the Federal LCA Commons, ecoinvent, etc. are growing semi-linearly.



The amount of LCI data being published is growing much more rapidly. Most of this data is not usable by the broader community.



But both of these trends are completely swamped by the demand for high quality background data.

International organisations and funders have more or less given up on LCA and set up initiatives like Climate Trace to satisfy the massive need for high quality climate data. Industry has also given up on many of the fundamental axioms of LCA and focused on providing B2B carbon foot printing at scale, see Catena-X, TFS, PACT, etc.

We risk becoming irrelevant in the long term because we can’t deliver what the customers need.

We need to see differently

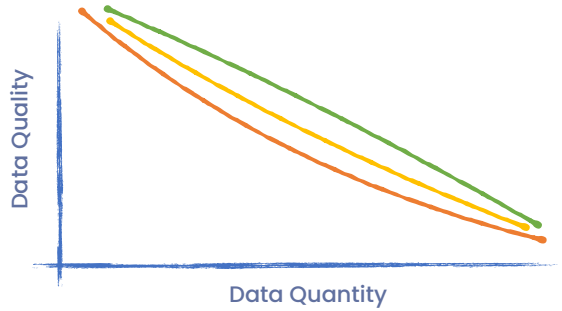
Make hard choices
Be honest about ourselves
Focus on the mission

Can't find the right data
The data found doesn't match our operations
The data found is out of date

The first step is to admit we have a problem.

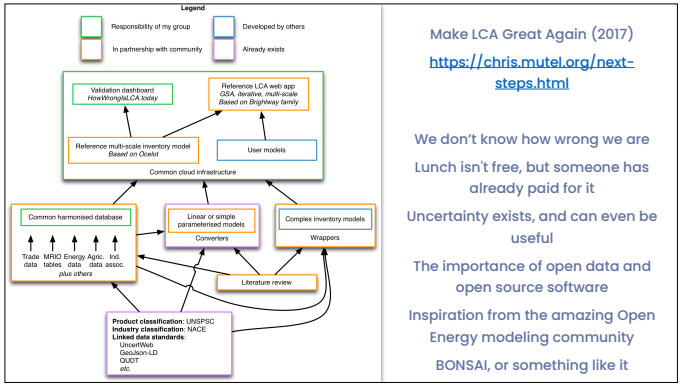
We, as an LCA community, should be honest with ourselves about what we can do well, and what we need to hand over to other experts. The goal is **the mission**, not any particular person's reputation or the success of any research group or business. Take a long, hard look in the mirror, and be honest about what you can and can't do.

Personally, I have now had enough experience with good data scientists (including at Brightcon, it's awesome, you should go next year: https://brightcon.link/), to know that what I can best do is support them with good partnerships, funding, and connections to other cool initiatives who are open to cooperation.



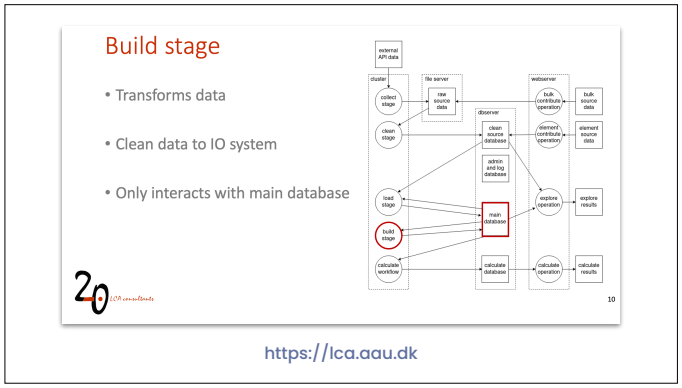
The shape of this curve is not fixed, and indeed there isn't one curve - it varies depending on the question and can improve over time.

LCA puts itself too much in the upper left have corner. This is one of the reasons we aren't expanding fast enough - we simply don't have the skills to incorporate large corpuses of lower data quantity.

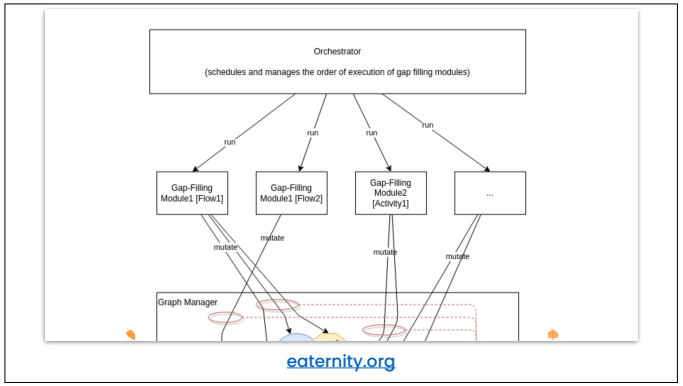


I put together this diagram in 2017, and wonder if I have had any real new thoughts since then. I think most ideas here still are valid.

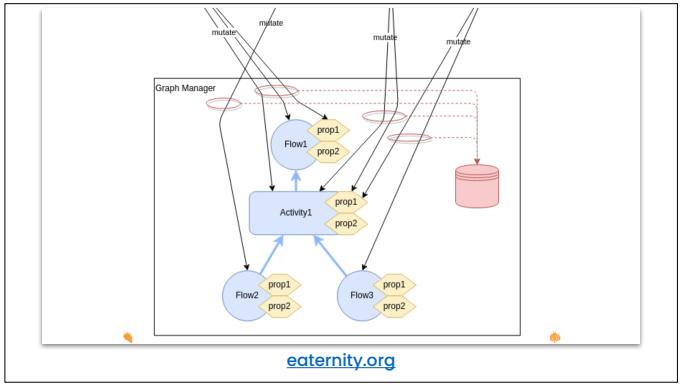
You can read the blog post for more detail. One place where I have changed my mind is that the previous conception of BONSAI - a decentralised network of volunteers - can't do what is needed to get to scale. We need coordinated resources to bring everyone together for long-term projects. The current version of BONSAI, building on the GTDR project, could be successful, I don't know enough to say.



At Brightcon the GTDR project presented their data engineering pipeline for building on and publishing the next iteration after the mixed-unit EXIOBASE tables. They are using Big Data technologies, though we aren't yet at true Big Data.



The company eaternity also presented at Brightcon. They have some impressive data engineering, with an extensive set of gap-filling and data-processing models. Their models can examine and modify the whole graph - to do this efficiently, they keep the graph in memory and constantly available. This presents some interesting challenges :)



Gap-Filling-Modules

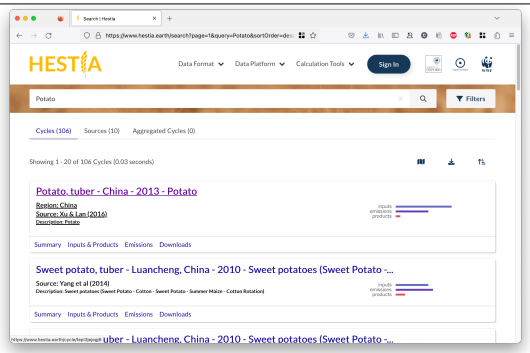
Ingredient Splitter
Glossary Matching
Subdivision (dry, wet)
Ingredient Amount Estimation
Import / Export / Production Statistics
Transport (Location, Distance & Mean)
Base Inventory Data
> Impact Assessment <

eaternity.org

This is one of the first true cases of trained ML models applied to LCA that I have seen (though of course there must be more) - by working with food producers, caterers, and restaurants for over a decade, eaternity has built up enough of a training dataset that they can guess the amounts of various ingredients given the list you get on the nutritional information section of food packaging.

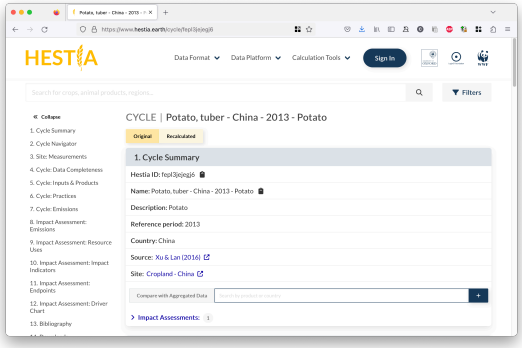
https://github.com/Depart-de-Sentier/brightcon-2023-talks

You can see the full context for both of the previous projects, and more, at the Brightcon talks repository.

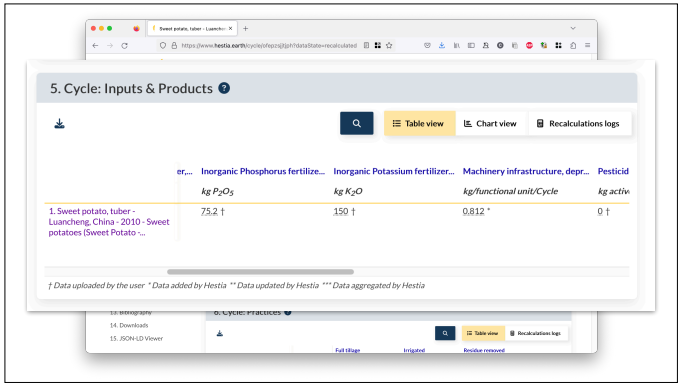


hestia.earth

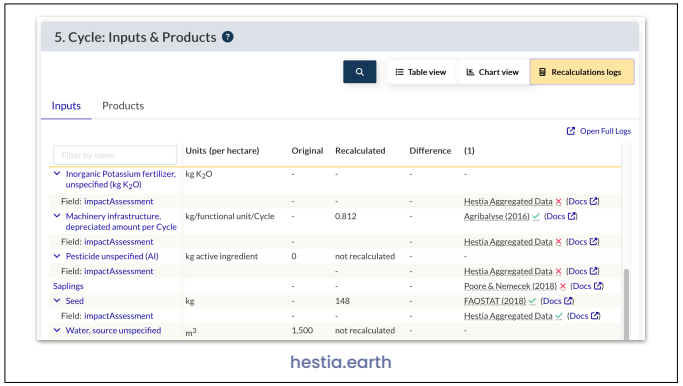
Hestia is a large repository of agricultural data that can be used for LCA, and they have also developed a large set of gap-filling models. I think they have done an awesome job of communicating what they have done, and why.

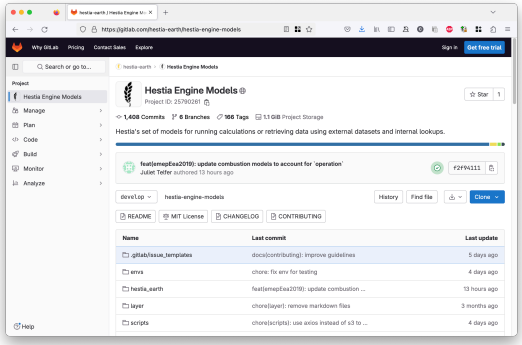


Let’s look at this idea in practice.

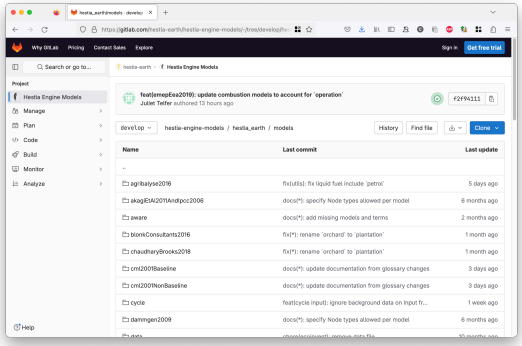


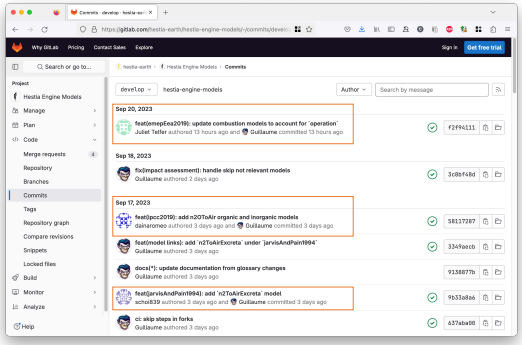
You can see here the provenance of the data points. One can turn the gap-filling models on and off.



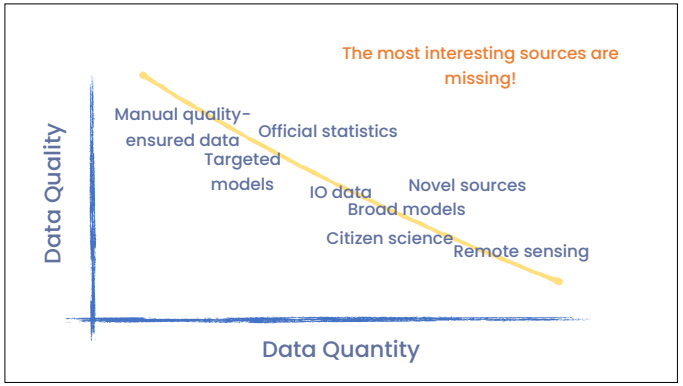


All the models are open source and easily accessible.





In the Brightcon hackathon, three models were added or improved, by people completely new to the Hestia project, and its data formats and workflows. To do everything right at scale will take a lot, but there are some low-hanging fruit out there.



If we build a broad community, with many (most?) of the people coming from outside traditional LCA, we have a chance to combine these data sources and break off of this Pareto front, iteratively achieving higher and higher quantity **and** quality.

Conclusions

Let data scientists do their job

LCA practitioners should focus on their specific expertise

(Arguing about allocation)

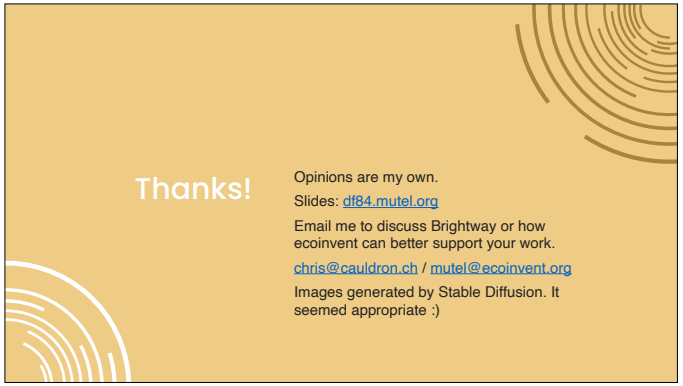
Auditable and reviewed models FTW

We need everyone working together

Biodiversity?

The era of Big Data, or at least bigger data, is coming to LCA. Can we as a community help guide it, using our experience, our knowledge of specific systems, and our networks? We need it, so let's be honest about what we can and can't do, and make decisions based on the mission.

Both startups and enterprise software makers are moving towards this goal rapidly. If we don't participate we will end up on the sidelines, and the systems thinking behind LCA will be lost.


Thanks!
Opinions are my own.
Slides: df84.mutel.org
Email me to discuss Brightway or how ecoinvent can better support your work.
chris@cauldron.ch / mutel@ecoinvent.org
Images generated by Stable Diffusion. It seemed appropriate :)