Brightcon 2020, Oct. 20, 2020

A General Data Model for Socioeconomic Metabolism and its Implementation in an Industrial Ecology Data Commons

Stefan Pauliuk, Uni Freiburg, Germany

Content:

- Data sharing in industrial ecology: State and vision
- A general data model for socioeconomic metabolism
- Applications of the data model

Image taken from: https://highered.colorado.gov/Data/Sharing.html

Data sharing in industrial ecology

Established databases for process- LCA, IO, MRIO, and MFA (material flow accounting at the national level).

However, many data are not shared and those that are shared, typically via the supplementary material of journal papers or via reports, are difficult to access and retrieve

No community-wide data model and format

Lacking incentive for sharing and reusing data.

→ As a community, we need to establish a data sharing culture to make our research more cumulative, timely, and relevant!

Data vs. methods

Industrial ecology (IE) are commonly seen as existing within the domain of particular methods or models, such as input–output, life cycle assessment, urban metabolism, or material flow analysis data.

This artificial division of data into methods contradicts the common phenomena described by those data: the objects and processes in the industrial system, or socioeconomic metabolism (SEM).

Pauliuk et al. 2019, DOI 10.1111/jiec.12890

Data integration

- “Longitudinal”: Append more data of the same type
  + product material composition, lifetimes, MFA cycles

- “Lateral”: Link across different data types
  + link total mass flows to material composition data
  + link consumption to product lifetime data
  + link waste/EoL flow data to recycling coefficients

MFA/IE/SEM data integration: At what level?

Highly integrated database: (like ecoinvent)
+ common classifications
+ consistency
Broad (RDF?) or limited scope (ecoinvent)

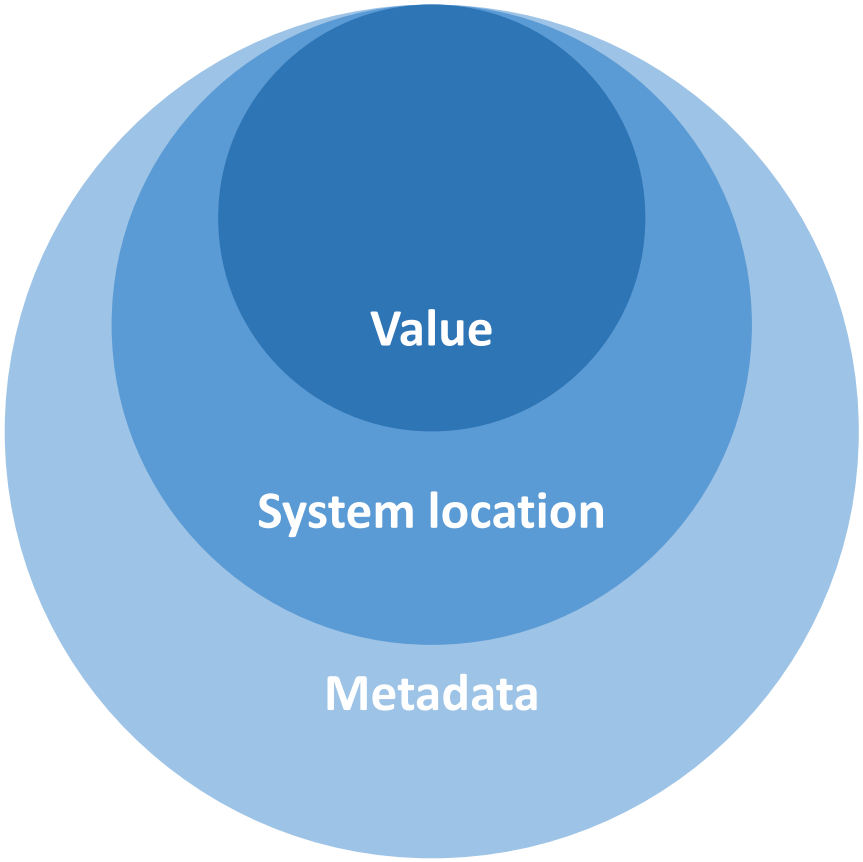
Structured database
Relational db model
Machine-readable
Multible and local classifications,
Little data integration

Document libraries:
Figshare
Zenodo
Researchgate

- Level of data integration and interoperability
- Effort to extract, format, and reclassify data
- Value added to data
- Data infrastructure complexity

Data in a systems context

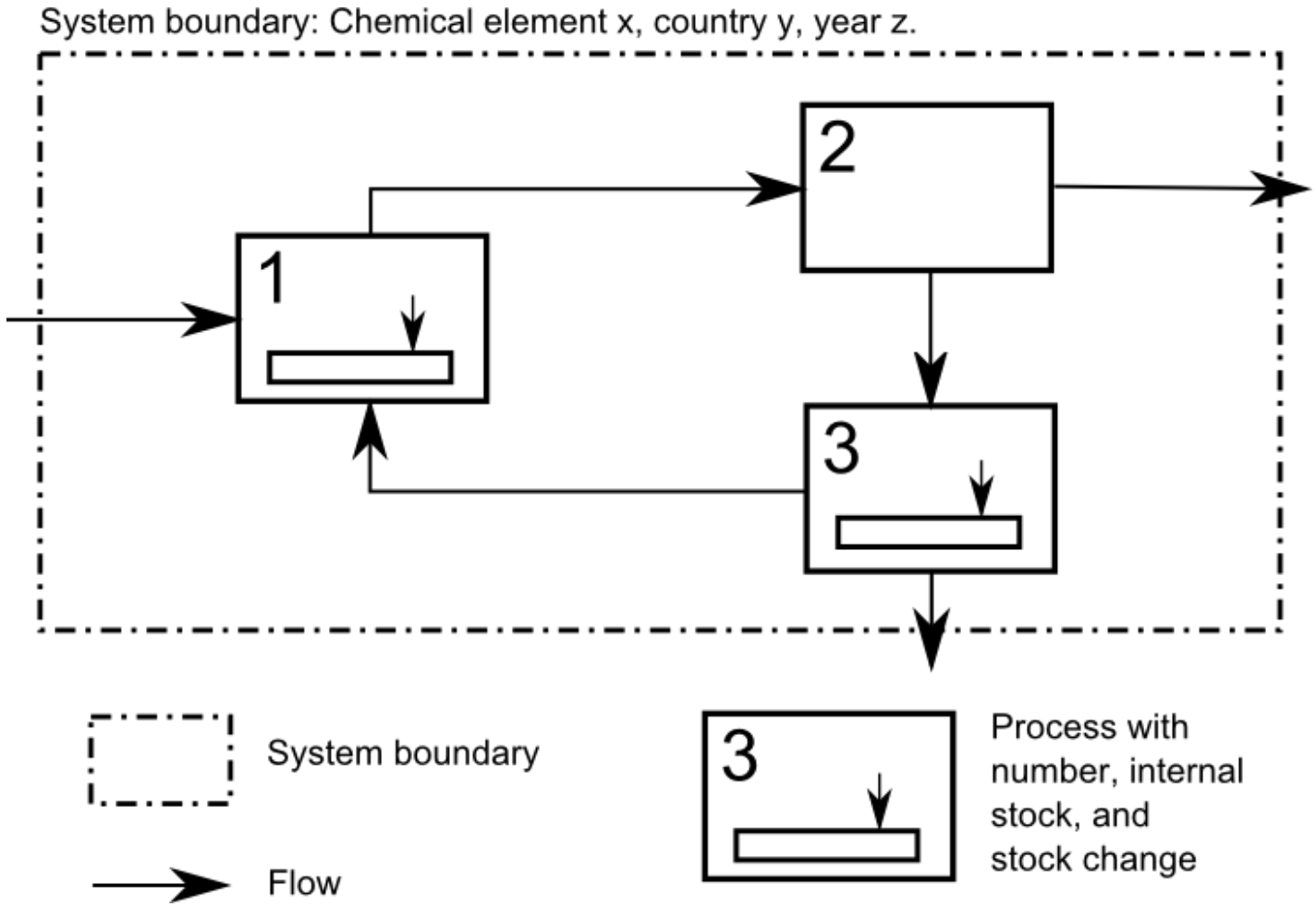
Quantitative information on processes, stocks, flows, etc., in socioeconomic metabolism has three components:



1) **Value.** The actual numerical information, including unit and uncertainty
2) **System location.** The information needed to locate the data in the systems context, i.e., the link between data and the system dimensions (process, time, region, material, …)
3) **Metadata.** Information like provenance, source document, author and version information, and license.

Source: „A General Data Model for Socioeconomic Metabolism and its Implementation in an Industrial Ecology Data Commons Prototype “, by Pauliuk et al. (2019), JIE, DOI: 10.1111/jiec.12890

The basic elements of a system definition



System boundary:
Shows what processes are part of the system studied. System needs to be specified in space and time.

Process: Element of a system were material or energy is transformed, stored, or distributed.

Stock: Storage of material, products, or energy. A stock is always associated with a process. It is measured at a given point of time t. ‚snapshot'

Flow: Transport of material products, or energy across the system. A flow always connects two processes or one process with the environment. It is measured over an interval $[t_1,t_2]$.

Data aspects and system dimensions I

- **System definition prescribes a number of *dimensions* along which the system content is described:**

- the *time dimension* is used to order events by the time of their occurrence
- the *location dimension* is used to order objects by their location
- the *process dimension* is used to identify balance volumes or to group events
- the *object dimension* is used to identify different goods or substances
- the *layer dimension* is used to indicate the unit in which the data are measured

Source: „A General Data Model for Socioeconomic Metabolism and its Implementation in an Industrial Ecology Data Commons Prototype “, by Pauliuk et al. (2019), JIE, DOI: 10.1111/jiec.12890

Data aspects and system dimensions II

System dimension	Description	Related data aspects (example)
Layer	Unit of measurement	Mass, volume, economic value
Process	Transformation, distribution, storage events	Process of residence (stock), process of origin (flow), process of destination (flow)
Location	Location in space	Region of residence (stock), region of origin (flow), region of destination (flow)
Object	Objects of interest (goods, substances, commodities, waste, products, …)	Commodity, good, product group, product type (sub-product), substance, chemical element, waste type, environmental extension
Time	Location in time	Historic time, model time, age-cohort, time point (stock), time interval (flow)
Scenario	Describing different “realities” or manifestations	Scenario for model drivers, scenario for process parameters

Table: The common system dimensions for socioeconomic metabolism and the related data aspects. The list of dimensions and aspects is not exhaustive.

Source: „A General Data Model for Socioeconomic Metabolism and its Implementation in an Industrial Ecology Data Commons Prototype “, by Pauliuk et al. (2019), JIE, DOI: 10.1111/jiec.12890

Data aspects and system dimensions III

(D1) Each data item (number quantifying a fact in a system) requires a minimum number of aspects to be meaningfully located in the system dimensions.

(D2) Each data type (stock, flow, material content, product lifetime, …) has a specific data model that prescribes which aspects are required and which aspects are optional for the meaningful location of this data type in the system definition.

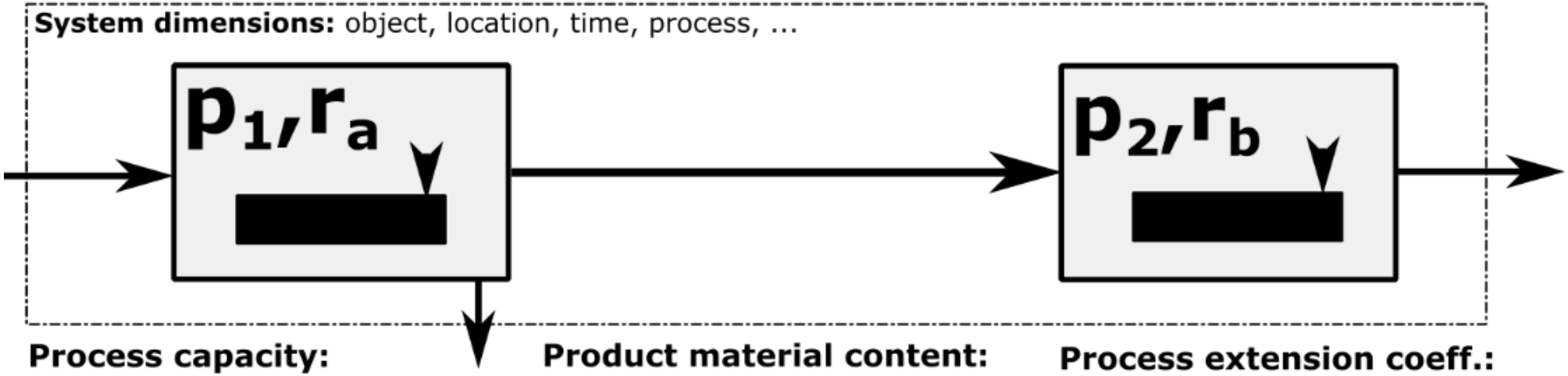
Source: „A General Data Model for Socioeconomic Metabolism and its Implementation in an Industrial Ecology Data Commons Prototype “, by Pauliuk et al. (2019), JIE, DOI: 10.1111/jiec.12890

Data aspects and system dimensions IV

Stock:
- good/substance* (object)
- residence process* (process)
- residence region* (location)
- time point* (time)
- age-cohort (time)
- component (object)

Flow:
- good/substance* (object)
- origin region (location)
- origin process* (process)
- destination region (location)
- destination process* (process)
- time interval* (time)

Product lifetime:
- good/substance* (object)
- residence region* (location)
- age-cohort (time)
- scenario (scenario)



Process capacity:
- process* (process)
- time point* (time)
- good/substance* (object)
- residence region* (location)
- age-cohort (time)

Product material content:
- substance* (object)
- good* (object)
- age-cohort (time)
- production region (location)

Process extension coeff.:
- extension* (object)
- reference output* (object)
- process* (process)
- residence region* (location)
- age-cohort (time)

Source: „A General Data Model for Socioeconomic Metabolism and its Implementation in an Industrial Ecology Data Commons Prototype “, by Pauliuk et al. (2019), JIE, DOI: 10.1111/jiec.12890

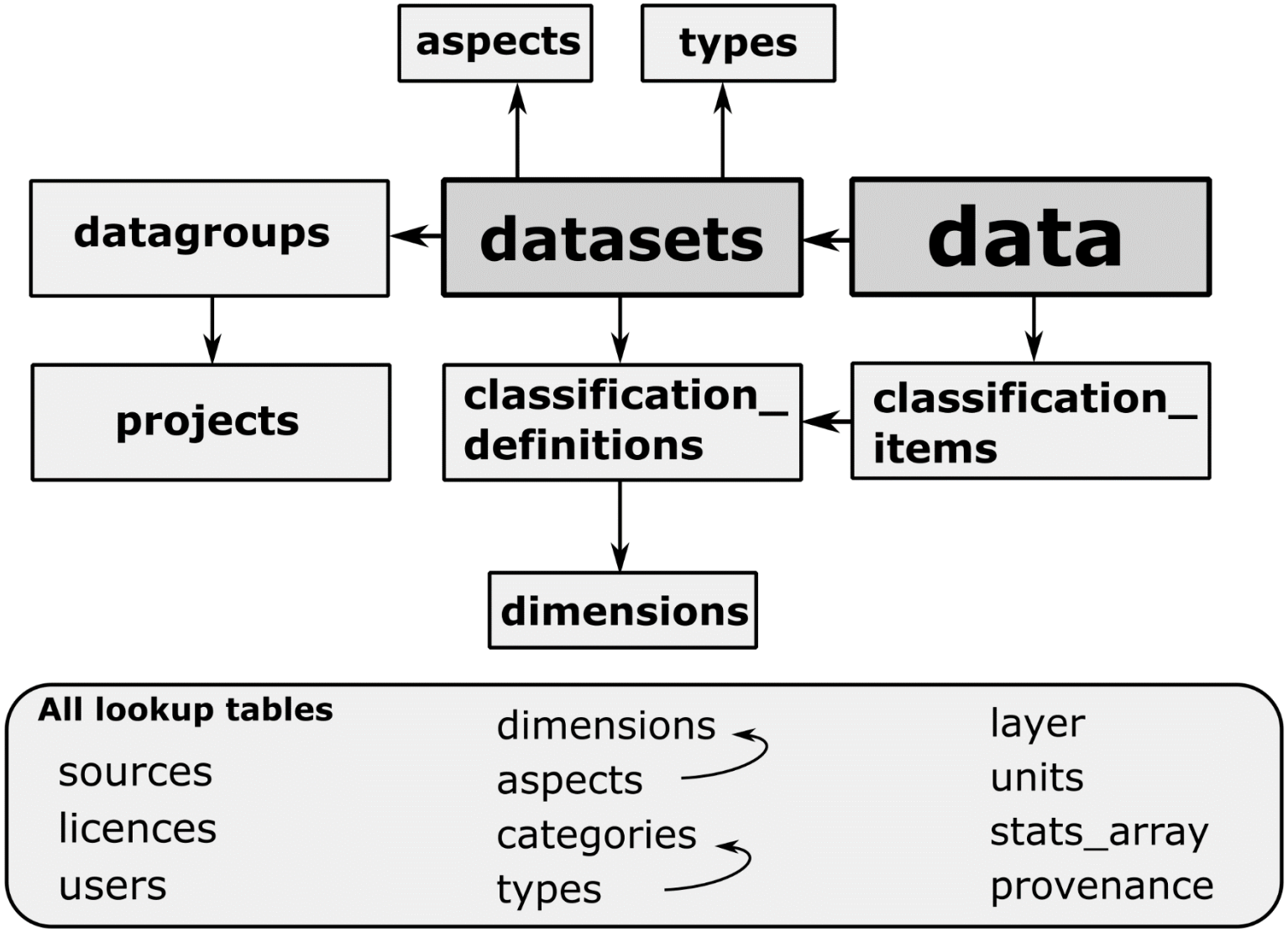
Data aspects and system dimensions V

	Objects of interest (goods, products, substances, commodities, waste, …)	Processes (industries, markets, end-use sectors, use-phase)
Extensive (at scale)	**Flows (1)** • Flow • Process inventory • Unit process inventory • Births/Deaths	**Extensive process properties (5)** • Process output capacity
	Stocks (2) • Stock • In-use stock • Population	
Intensive (per unit)	**Intensive object properties (3)** • Product material composition • Product lifetime • Price of products • Specific energy consumption	**Intensive process properties (4)** • Process yield factors • Process environmental extensions per output • Process operating costs per output
	General ratios (6) • Per capita stock • Per capita flow • Material substitution coefficient	

7) Correspondence tables

Source: „A General Data Model for Socioeconomic Metabolism and its Implementation in an Industrial Ecology Data Commons Prototype “, by Pauliuk et al. (2019), JIE, DOI: 10.1111/jiec.12890

Using the data model I



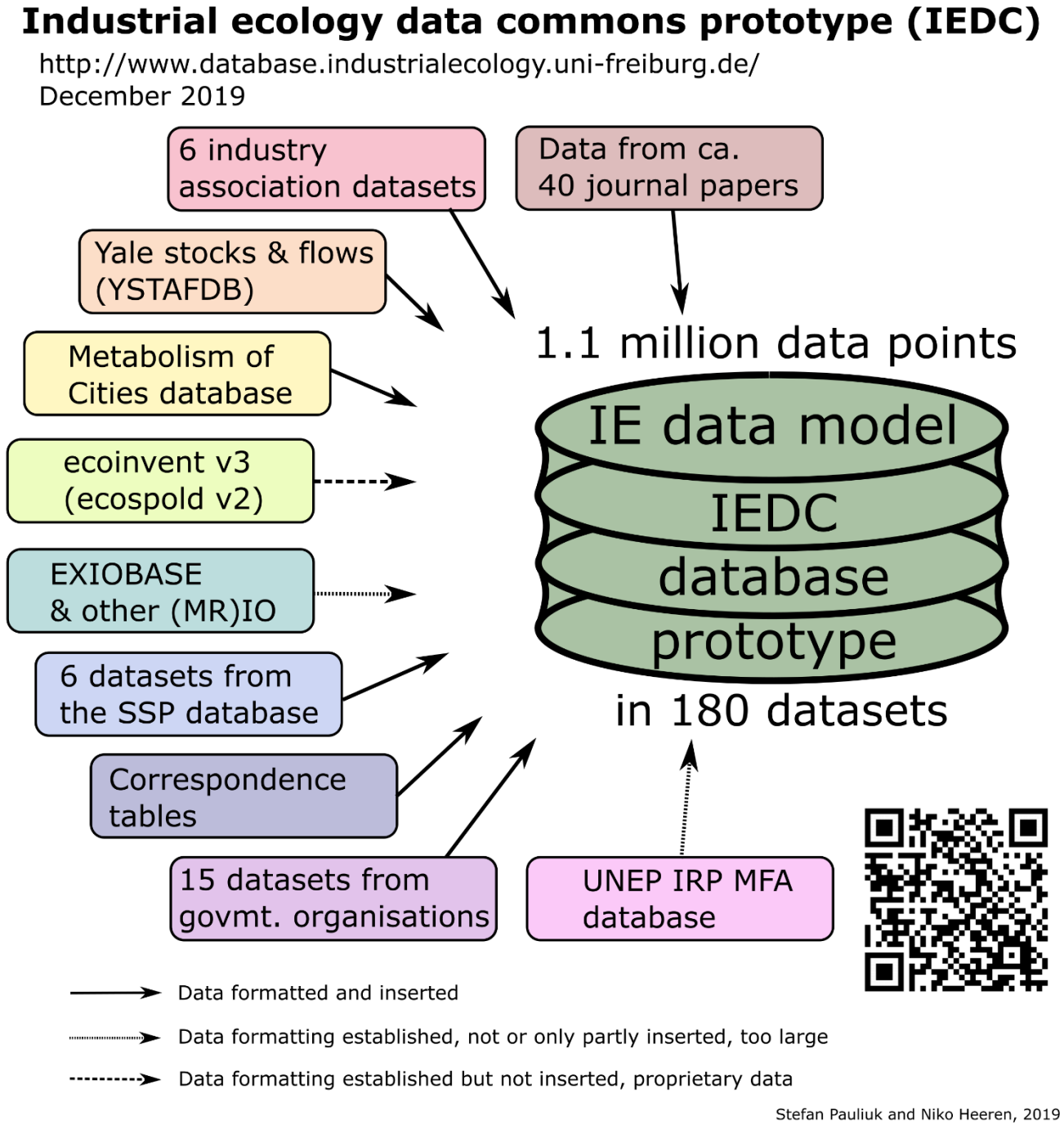
Source: „A General Data Model for Socioeconomic Metabolism and its Implementation in an Industrial Ecology Data Commons Prototype “, by Pauliuk et al. (2019), JIE, DOI: 10.1111/jiec.12890

Using the data model II

The IEDC prototype contains more than 100 IE-related datasets from the literature, including stocks, flows, process descriptions, IO tables, material composition of products, and many more.

Our mission is to show how data storage and exchange in sustainability science could look like in the future and to develop some of the building blocks of the required infrastructure.

www.database.industrialecology.uni-freiburg.de



MFA/IE/SEM data infrastructure /data commons: Where to go?

Application	Infrastructure	Incentives/Business models
Link related data across datasets Automize computations and data Interpretation How exactly?	**Highly integrated database:** + common classifications + consistency Broad (RDF?) or limited scope (ecoinvent)	Data access & analys. much easier Subscription? Initial funding? Community support? Is data model/toolchain ready?
Searchable and ready-to-use datasets Pre-defined data models -> full systems context and metadata Some common classifications Little dataset linkage	**Structured database** Relational db model Machine-readable Multible and local classifications, Little to no data integration	Contribute vs. full access? Subscription fee/pay per dataset Assistance with data formatting/ classifation required Does ISIE need/want that? Infrastructure costs?
Make indiv. Datasets searchable by keywords; Catalogue of datasets No common data model/format pdf, png & xlsx collection Manual extraction/formatting	**Document libraries:** Figshare Zenodo Researchgate	Journal policy External actors (data supply) Authors: good keywords & document struct.

A general data model and a prototype for an Industrial Ecology data commons (general database) is now available!

http://www.database.industrialecology.uni-freiburg.de/

- Database to record small to medium size datasets all across industrial ecology
- Includes stocks, flows, product properties, process coefficients, impact indicators, socioeocomic data, etc.
- Built on newly develo **Thank you!** strial ecology
- Data stored in relatio l templates and other
- Medium level of data integration, between Figshare etc. and ecoinvent etc.
- Shows how data archiving in industrial ecology can look like

DOI: 10.1111/jiec.12890

<- Check out the prototype of the Industrial Ecology Data Commons!

Check out the paper on the Industrial Ecology Data Model! ->