

# FAIR-ification of LCA data using Semantic Web

AGNETA GHOSE



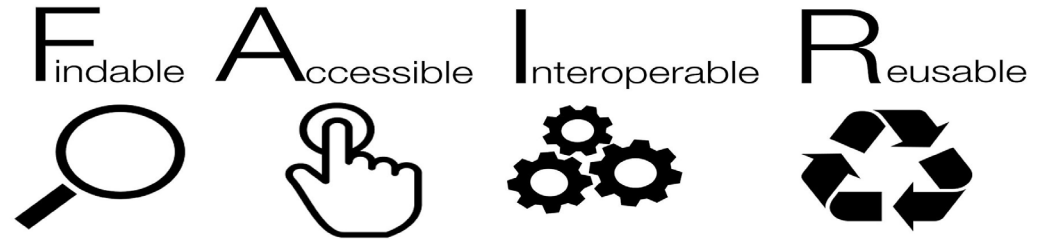


**Challenge** - Research data sits in different systems- hard to discover, obtain and integrate - eventually building new insight takes time

## Why is it important to share research data?

- ▶ Makes research more open- transparency and reproducibility
- ▶ Makes your research more useful and citable by other researchers
- ▶ Allows data to be used to answer new questions
- ▶ Increasingly required by high impact journals
- ▶ Required by government funding agencies

# FAIR guiding principles



## Findable:

- Data are assigned with a unique identifier (e.g. doi).
- Registered and indexed in a searchable resource.
- Described with rich metadata.

## Interoperable:

- Knowledge representation - Use a formal, accessible, shared and broadly acceptable language
- Data include qualified references to other data

## Accessible:

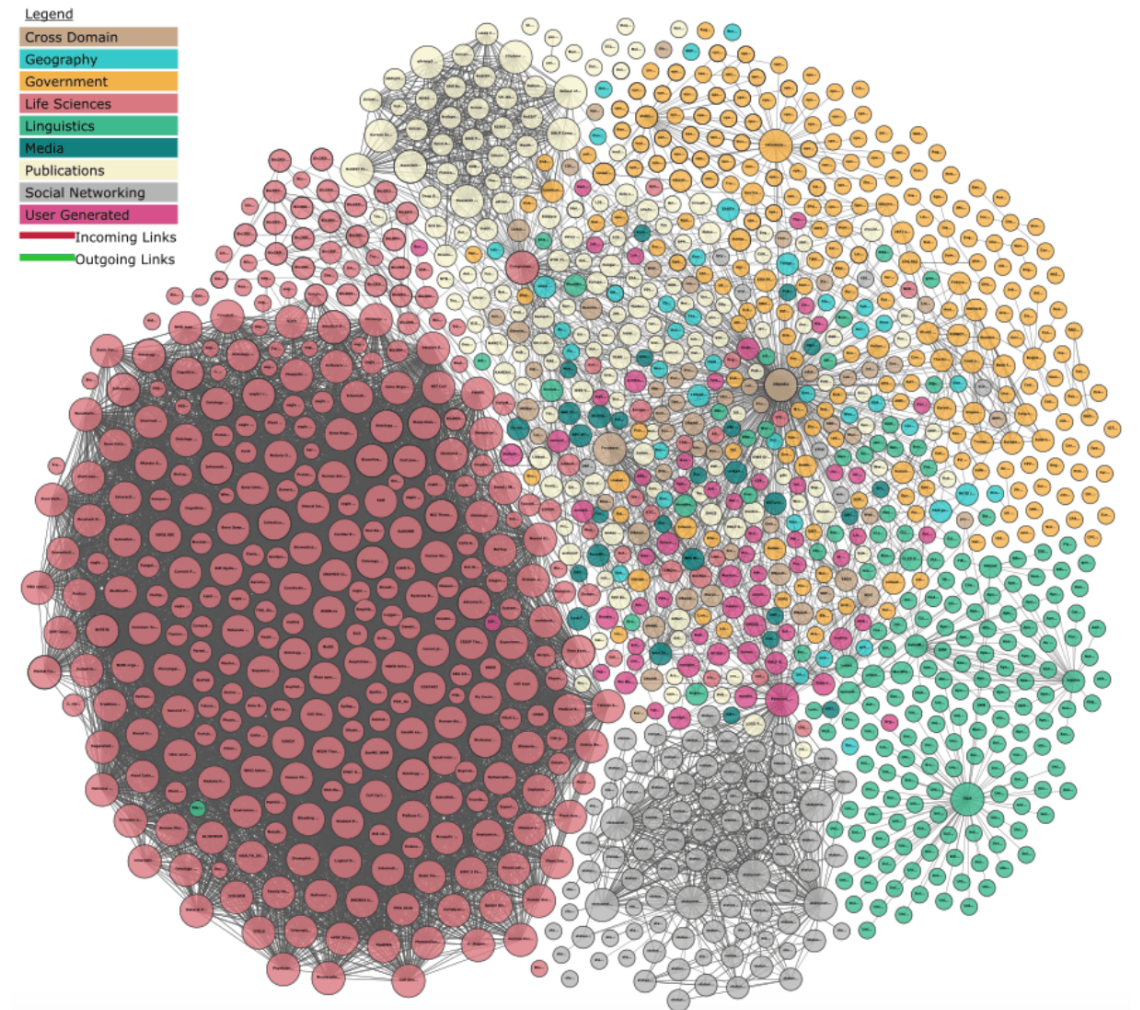
- Data are retrievable by their identifier using a standardized protocol.
- Metadata should be accessible even if the data is proprietary.

## Reusable:

- Data are released with a clear, appropriate and accessible license.
- Data are associated with detailed provenance.

# Semantic Web and Linked Data

- World Wide Web – linked content/documents
- The key technology of the original web —was the hyperlink (from the end user's point of view)
- Semantic web is a cloud of linked data.
- Data gets an unique identifier
- Applications, that receive large amount of data from many different sources, benefit enormously from this feature.





## Semantic Web (contd.)

Semantic Web consists primarily of three technical standards:

1. **RDF (Resource Description Framework):** Data modeling language for the Semantic Web. Machine readable data format. All Semantic Web information is stored and represented in the RDF.
2. **OWL (Web Ontology Language):** The knowledge representation language of the Semantic Web. OWL enables to define concepts so that these concepts can be reused as much and as often as possible.
3. **SPARQL (SPARQL Protocol and RDF Query Language):** Query language of the Semantic Web. It is specifically designed to query data across various systems.





# Why go semantic?

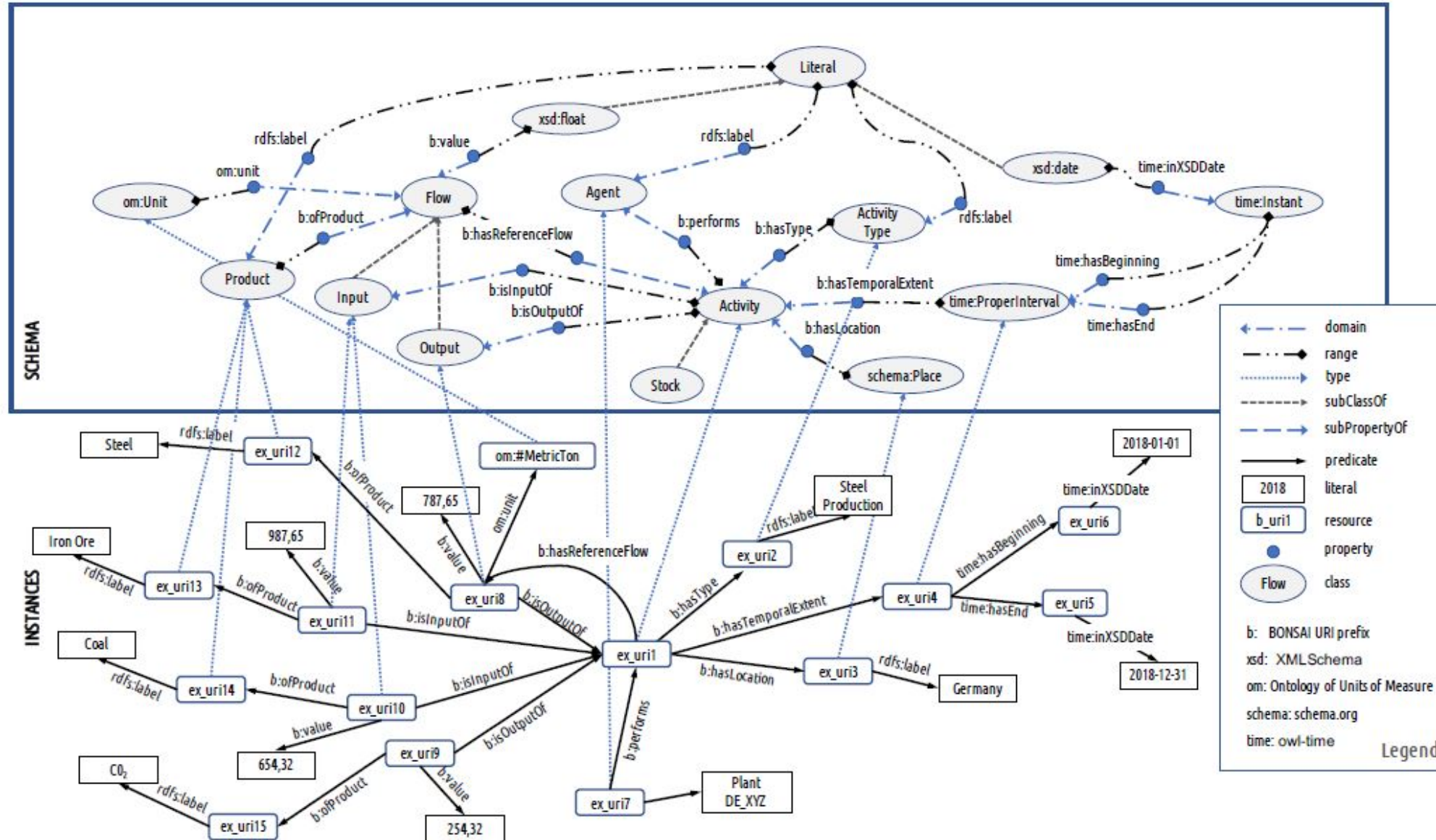
- To present knowledge about your data
- To allow data integration

## What is the difference between RDF and OWL?

- Data formatted using these languages are machine readable and enhance automation.
- RDF is used to represent data in triple format, give it some structure and unique identifiers.
- OWL provides a rich vocabulary to allow reasoning and inference



# Step 1 - Conceptualizing LCA on the Semantic Web – BONSAI ontology



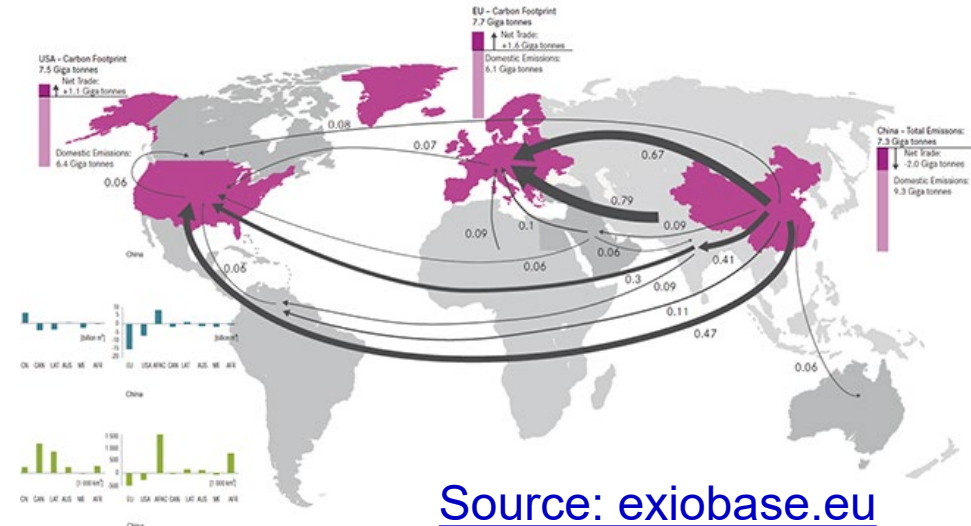
# Step 2 – Convert existing open certified databases into machine readable RDF format

Two datasets:

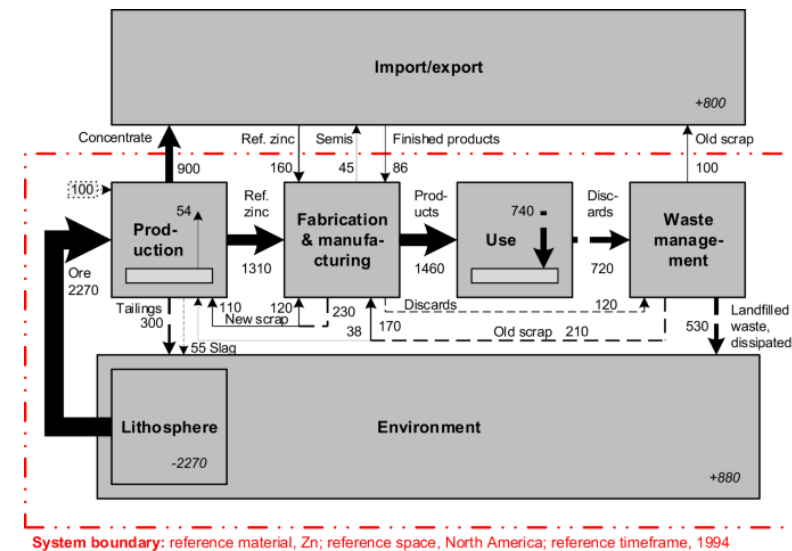
1. [Exiobase – multi-regional Input Output data base](#)
  2. [Yale stocks and flow Database \(YSTAFDB\)](#)
- Datasets developed using different modelling methods
    - Multi-regional Supply and Use Tables
    - Material Flow Analysis
  - Different sources
    - Exiobase – specific website (needs sign up)
    - YSTAFDB – Zenodo.org
  - Different data storage formats (csv, xlsx)
  - Certified as CC-BY 4.0



Attribution 4.0 International (CC BY 4.0)



[Source: exiobase.eu](https://exiobase.eu)



AALBORG UNIVERSITY  
DENMARK

[Source: Myers \(2019\)](#)



## Step 3: Develop SPARQL endpoint to query RDF data

### BONSAI Sparql Endpoint

Select a query, or write own 

```
1 PREFIX bont: <http://ontology.bonsai.uno/core#>
2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3 PREFIX om2: <http://www.ontology-of-units-of-measure.org/resource/om-2/>
4
5 SELECT ?activityType ?flowObject (xsd:string(?value) as ?value) ?unit
6 FROM <http://rdf.bonsai.uno/data/exiobase3_3_17/hsup>
7 FROM <http://rdf.bonsai.uno/data/exiobase3_3_17/huse>
8 FROM <http://rdf.bonsai.uno/data/exiobase3_3_17/emission>
9 FROM <http://rdf.bonsai.uno/location/exiobase3_3_17>
10 FROM <http://rdf.bonsai.uno/flowobject/exiobase3_3_17>
11 FROM <http://rdf.bonsai.uno/activitytype/exiobase3_3_17>
12 FROM <http://rdf.bonsai.uno/unit>
13 FROM <http://rdf.bonsai.uno/time>
14 WHERE
```

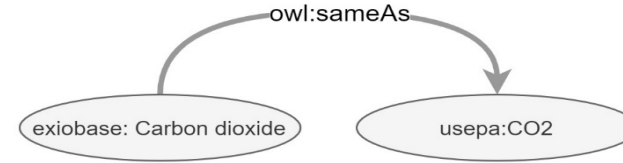
 

 Table  Response

## FAIR Data for Humans and machines

Country	cc	Product	cc	Product	cc	Unit	A_PARI	A_WHEA	A_OCR	A_FVEG	A_OILS	A_SUGB	A_FIBR	A_OTCR	A_CATL	A_PIGS	A_PLTR	A_OMEA	A_OANP	A_MIL
AU		Paddy rice	p01.a	C_PARI	tonnes		0	0	0	0	0	0	0	0	0	0	0	0	0	0
AU		Wheat	p01.b	C_WHEA	tonnes		0	0	0	0	0	0	0	0	679096.2	156069.8	1238991	0.007474	0	36341
AU		Cereal gra	p01.c	C_OCR	tonnes		0	80548.11	0	0	0	0	0	0	22762.89	823039.1	1948149	0.125796	0	29534
AU		Vegetable	p01.d	C_FVEG	tonnes		0	0	24614.16	0	0	0	0	0	1546.483	254.4394	107948.9	82.65126	0	1455
AU		Oil seeds	p01.e	C_OILS	tonnes		0	0	0	0	0	0	0	0	0	0	0	0	0	0
AU		Sugar cane	p01.f	C_SUGB	tonnes		0	0	0	109.6995	0	0	0	0	0	0	0	0	0	0
AU		Plant base	p01.g	C_FIBR	tonnes		0	0	0	0	0	0	0	0	0	0	0	0	0	0
AU		Crops nec	p01.h	C_OTCR	tonnes		0	0	0	0	0	0	0	0	0.004781	0	0	0	0	0
AU		Cattle	p01.i	C_CATL	tonnes		0	0	0	0	0	0	0	0	0	19354.6	0	0	0	0
AU		Pigs	p01.j	C_PIGS	tonnes		0	0	0	0	0	0	0	0	0	0	0	0	0	0
AU		Poultry	p01.k	C_PLTR	tonnes		0	0	0	0	0	0	0	0	0	0	0	0	0	0
AU		Meat untr	p01.l	C_OMEA	tonnes		0	0	0	0	0	0	0	0	0	0	0	0	0	0
AU		Animal pro	p01.m	C_OANP	tonnes		0	0	0	0	0	0	0	0	0	0	0	0	0	0.007218
AU		Raw milk	p01.n	C_MILK	tonnes		0	0	0	0	0	0	0	0	0	0	0	0	0	0
AU		Wool, alk	p01.o	C_WOOL	tonnes		0	0	0	0	0	0	0	0	0	0	0	0	0	0
AU		Manure (b)	p01.v.1	C_MANB	tonnes		0	0	0	0	0	0	0	0	0	3432461	507184.8	1858370	330930.4	0
AU		Manure (b)	p01.v.2	C_MANB	tonnes		0	0	0	0	0	0	0	0	0	0	0	0	0	0
AU		Products ntr	p01.w	C_FCHP	tonnes		0	0	0	0	0	0	0	0	0	0	0	0	0	0



### Interoperability

Semantic mapping - gives meaning to data  
Link ontology and data to existing ontologies and datasets

Step 1

### Findability and Accessibility

Provides an unique identifier to each data instance

[http://rdf.bonsai.uno/flowobject/exiobase3\\_3\\_17#C\\_1](http://rdf.bonsai.uno/flowobject/exiobase3_3_17#C_1)

Carbon dioxide, fossil

Step 2

### Re-Usability

Creation and querying of RDF Linked Data  
Human and Machine understandable data

Step 3



# Are semantic web the future for data sharing?

## Opportunities

- Develop web applications and visualizations to automate the work process for LCSA researchers to integrate, query and explore data
- Use queried data directly with open LCA softwares
- Interoperability with major datasets such as data from FAOSTAT, Dbpedia, GIS sources etc can be a huge advantage in expanding the data

## Challenges

- **Usability** - Researchers not accustomed to these models. Need for extensive programming knowledge
- **Infrastructure sustainability** – Setting up and maintaining servers. Project dependent
- **Data quality and validity**



**THANK YOU!**  
**BONSAI WIKI: [GITHUB.COM/BONSAMURAI](https://github.com/Bonsamurais/Bonsai/wiki)/BONSAI/WIKI**  
**GITHUB: [GITHUB.COM/BONSAMURAI](https://github.com/Bonsamurais)**  
**SPARQL ENDPOINT: [HTTP://ODAS.AAU.DK/](http://odas.aau.dk/)**

