



Data Federation and Preservation in Food Composition Data

Karl Presser (Premotec GmbH & EuroFIR AISBL)

EOSC Symposium 2023

21.09.2023, Madrid, Spain



**Co-funded by
the European Union**

What is Food Composition Data?



Food Composition Data Levels

Data Source

Source Data

Aggregated Data



| |
|--------------------|
| Apple |
| Protein 0.3 g/100g |
| Protein 200 mg/50g |
| ... |

| |
|--------------------|
| Pear |
| Protein 0.4 g/100g |
| Fat 0.3 g/100g |
| ... |

Unit conversion

Matrix unit conversion

$$\sum_{i=1}^n w_i * value_i$$

| |
|-----------------------|
| Apple |
| Protein 0.35 g/100g |
| Fat 0.3 g/100g |
| ... |
| Energy 231.9 kJ/100 g |

Borrow

$$ENERCJ[kJ] = 17*PROT[g] + 17*CHO[g] + 37*FAT[g] + 29*ALC[g] + 13*OA[g] + 10*POLYL[g]$$

| | |
|---|-----------------------------|
| Whole Milk | |
| Serving Size 8 fl oz (240mL) | |
| Servings Per Container 2 | |
| Amount Per Serving | |
| Calories 150 | Calories from Fat 70 |
| % Daily Value* | |
| Total Fat 8g | 12% |
| Saturated Fat 5g | 25% |
| Cholesterol 35mg | 12% |
| Sodium 125mg | 5% |
| Total Carbohydrate 12g | 4% |
| Dietary Fiber 0g | 0% |
| Sugars 11g | |
| Protein 8g | |
| Vitamin A 6% | • Vitamin C 4% |
| Calcium 30% | • Iron 0% |
| • Vitamin D 25% | |
| * Percent Daily Values are based on a 2,000 calorie diet. Your daily values may be higher or lower depending on your calorie needs. | |
| Calories: | 2,000 2,500 |
| Total Fat | Less than 65g 80g |
| Sat Fat | Less than 20g 25g |
| Cholesterol | Less than 300mg 300mg |
| Sodium | Less than 2,400mg 2,400mg |
| Total Carbohydrate | 300g 375g |
| Dietary Fiber | 25g 30g |

Dataset Versions



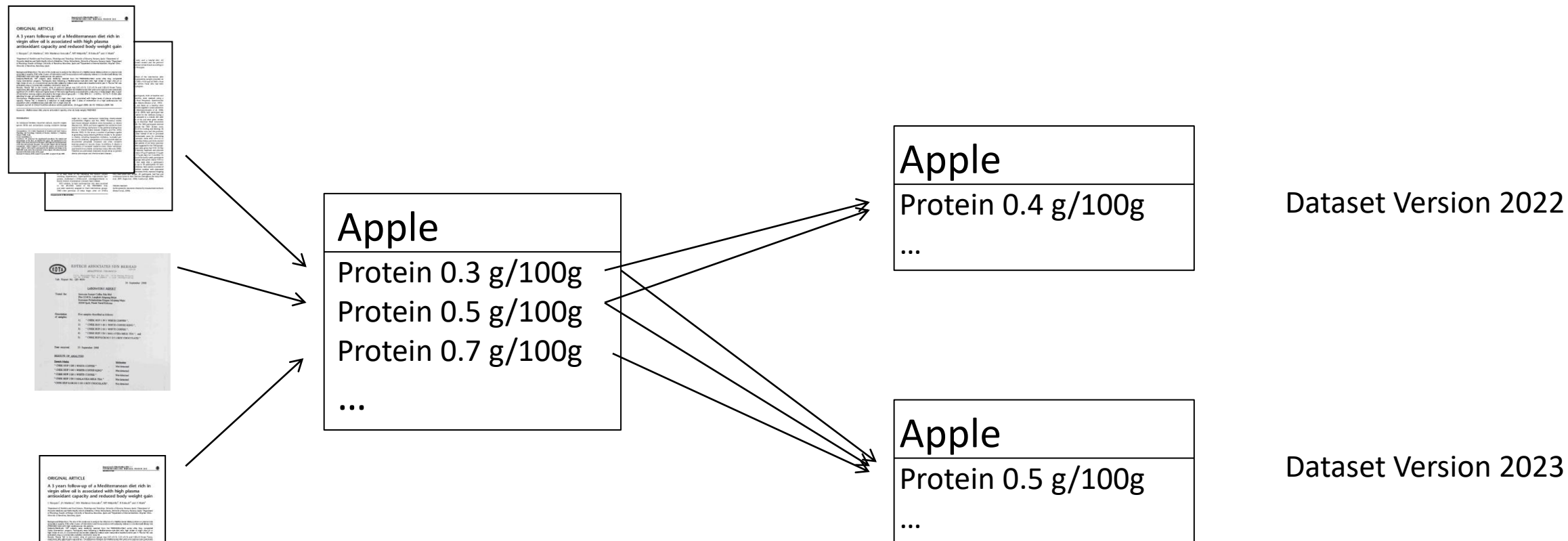
| |
|--------------------|
| Apple |
| Protein 0.3 g/100g |
| Protein 0.5 g/100g |
| ... |

| |
|--------------------|
| Apple |
| Protein 0.4 g/100g |
| ... |

Version 2022



Dataset Versions



Co-funded by
the European Union

Preserve all version for long-term studies



PREMOTEC



The Community



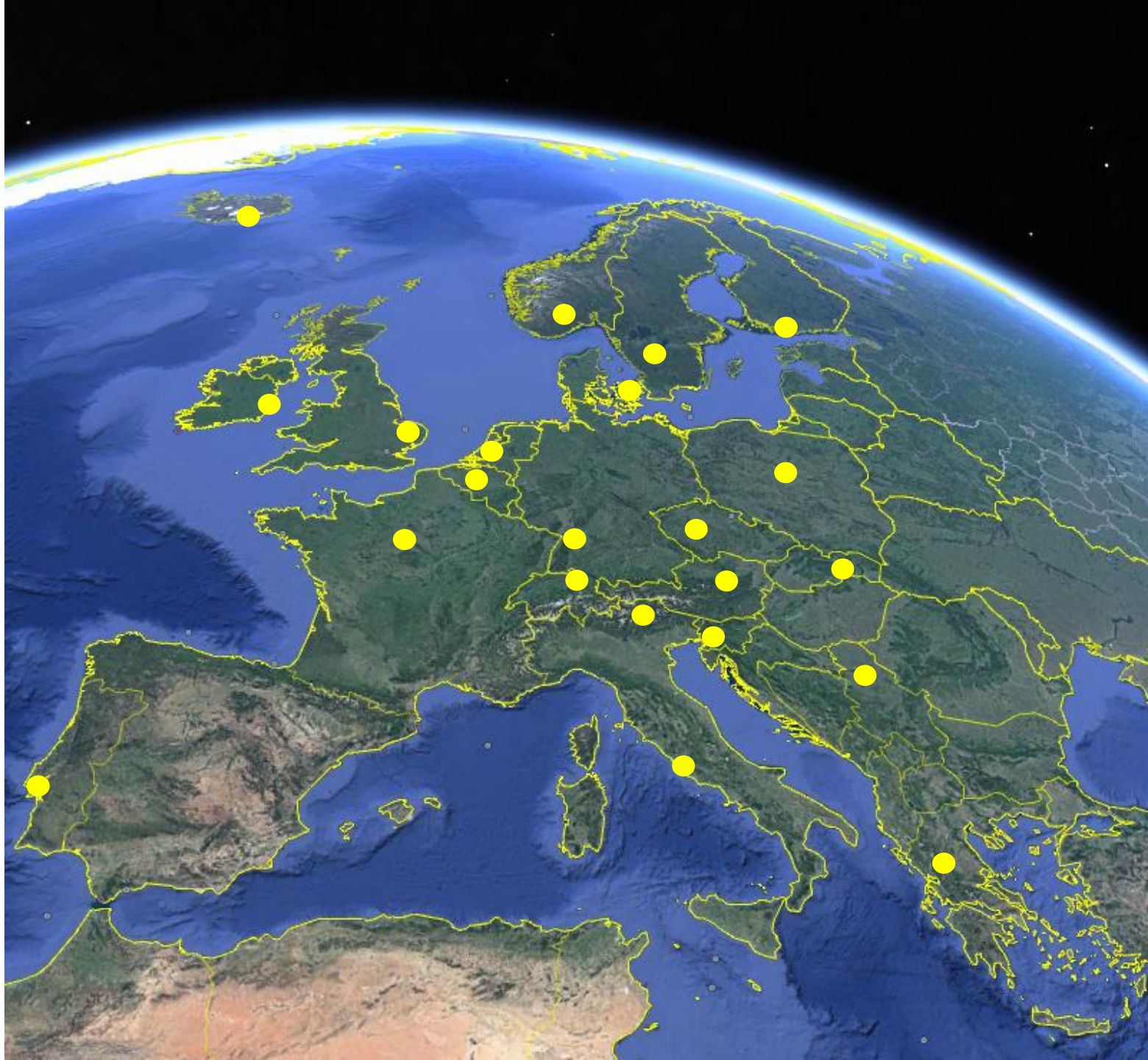
Co-funded by
the European Union

Collaboration

- Each country has to provide such a database
 - Data is very similar
 - A lot of borrowing is needed (data exchange)
 - A central European database is not an option
- ➔ EuroFIR AISBL
Non-profit organisation
with members



Co-funded by
the European Union

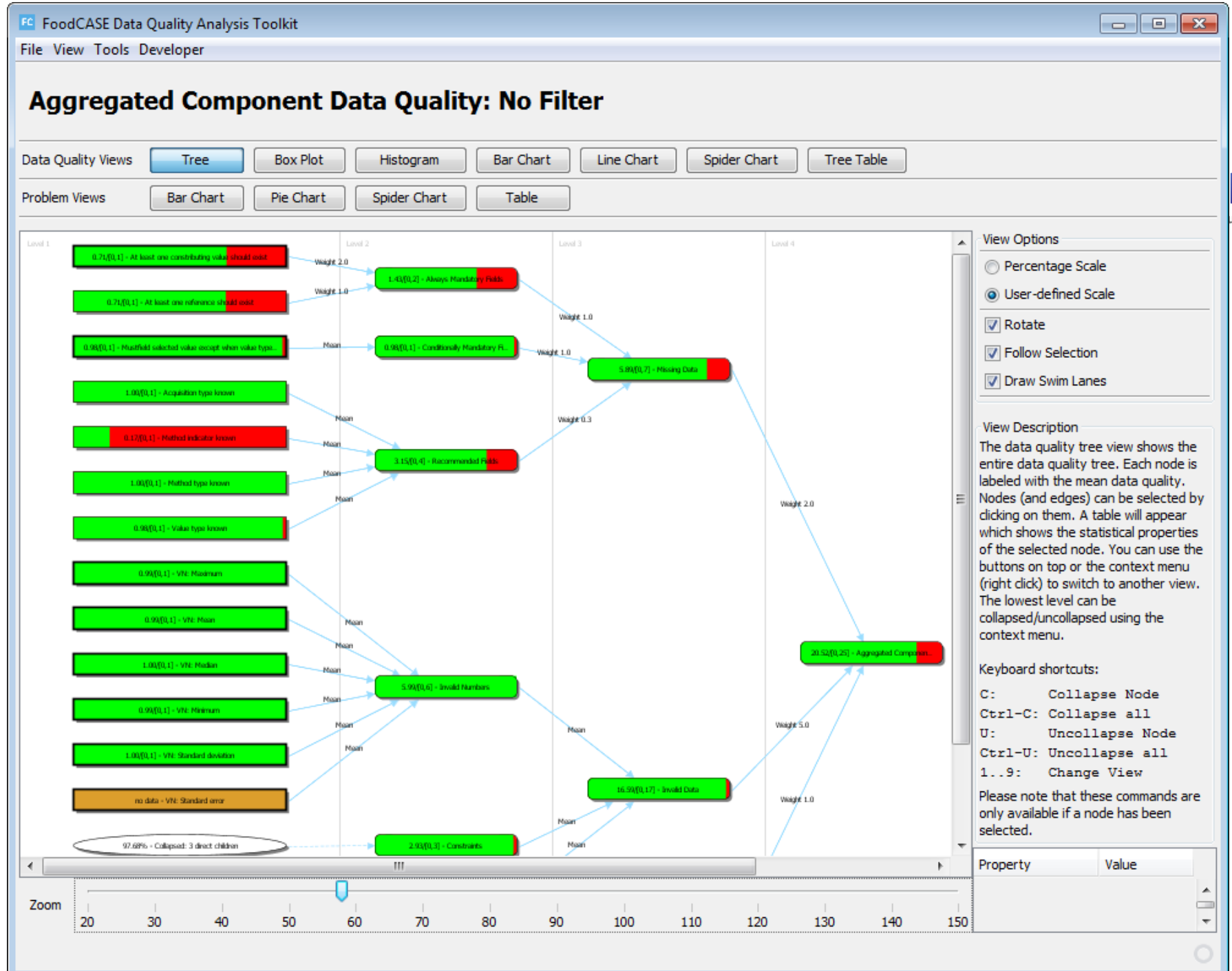


EuroFIR AISBL

- Common thesauri for some attributes
- Data exchange format with entities and attributes (XML and EXCEL)
- Discussion groups on specific topics
- Evaluation visits
- Data quality index
- SOP (Standard Operation Procedures)
- FoodEXplorer

| Action | Thesaurus | Version |
|--------|------------------------------------|---------|
| V R | EuroFIR Acquisition Type Thesaurus | 1.0 |
| V R | EuroFIR Acquisition Type Thesaurus | 1.1 |
| V R | EuroFIR Acquisition Type Thesaurus | 1.2 |
| V R | EuroFIR Component Thesaurus | 1.0 |
| V R | EuroFIR Component Thesaurus | 1.1 |
| V R | EuroFIR Component Thesaurus | 1.2 |
| V R | EuroFIR Component Thesaurus | 1.3 |
| V R | EuroFIR Component Thesaurus | 1.4 |
| V R | EuroFIR Component Thesaurus | 1.5 |

Data Quality Analysis



Co-funded by
the European Union

Long-Term Data Preservation



Data Preservation

- Usage of relational DBMS
- Data model follow the EuroFIR data exchange format
 - Metadata about data: Extensive value documentation (sampling, analytical method, aggregation, year, ...)
 - Metadata about dataset: Producer, contact, year, ...
- Usage of data classification and data description system (FoodEx2, LanguaL -> FoodON)



Data Preservation

- Same tables are used to store versions
-> if a new attribute is added, also old versions get this attribute

- Usage of EuroFIR thesauri
-> Dataset version is connected with thesauri version
-> Curation and preservation level: Between A and B

V2022 – Nut Thes 1.2
V2023 – Nut Thes 1.4

- All databases are trustworthy repositories, given my participating in the community



- future

