

Aggregating flags

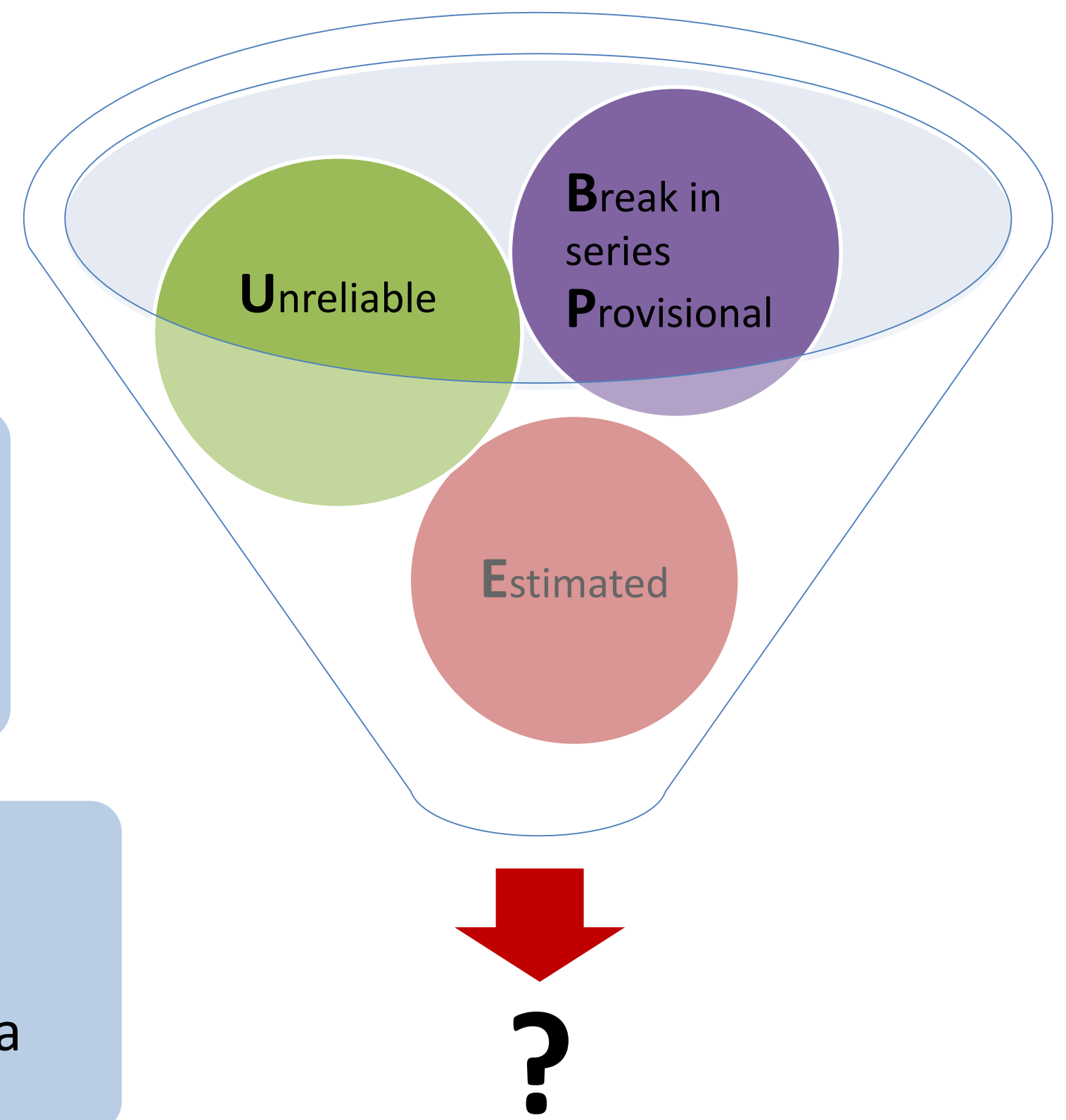
– a standardised and rational approach –

Mészáros Mátyás Tamás, Eurostat

A **flag** is an attribute of a cell in a data set that **provides additional qualitative information** about the statistical value of that cell.

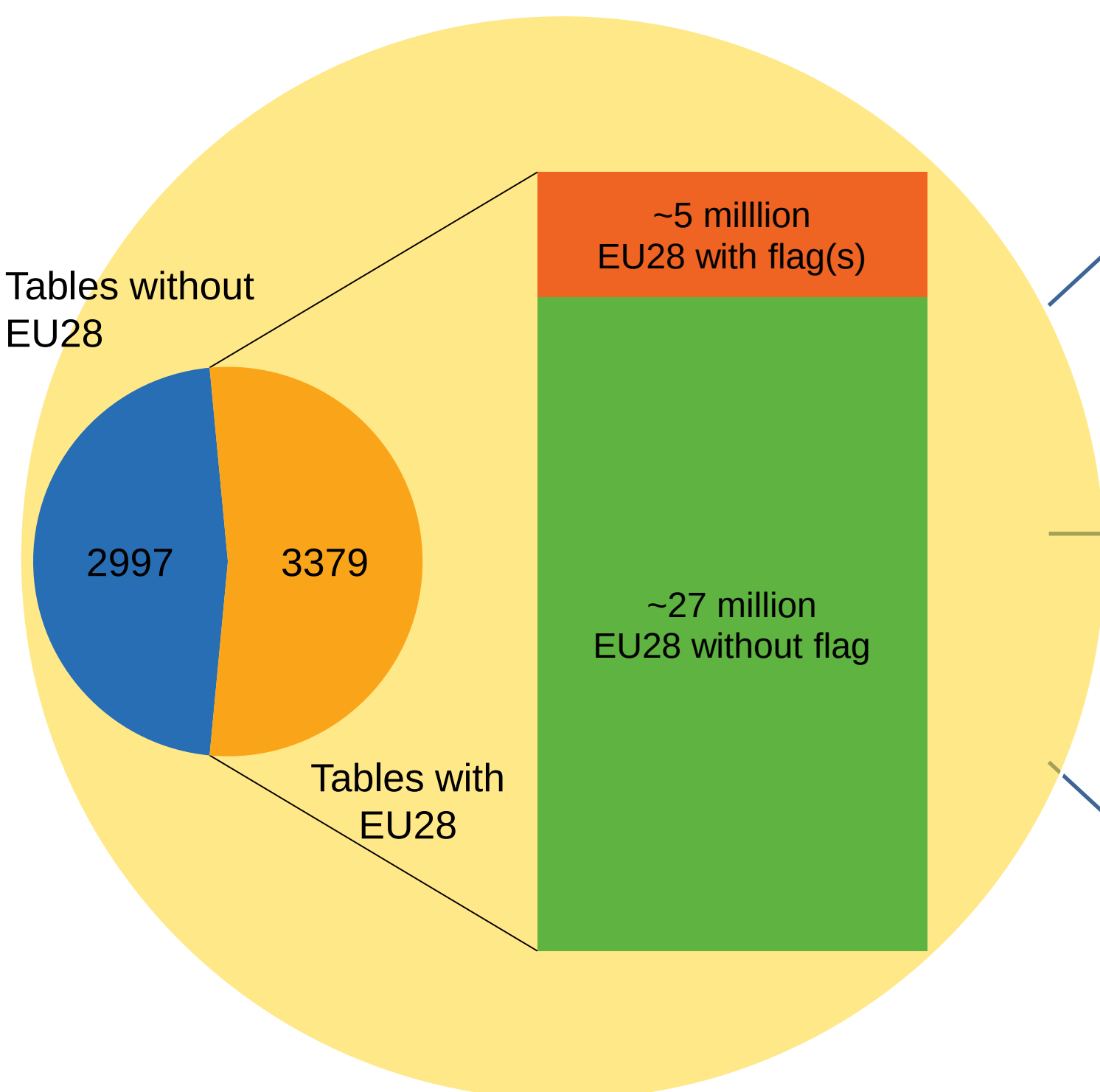
Wide variety of flags: - SDMX
- Eurobase
- OECD
- Regulations, handbooks, etc.

- Different approaches for aggregation
- Important for Eurostat to calculate the EU aggregates
- Relevant when NSIs derive national aggregate from regional data



Standardised methods

EU28 flags in Eurostat database*



* The chart reflects the status on 2019/02/21

Hierarchical e.g. SDMX

✓ Easy to implement

✗ Needs a pre-agreed hierarchy

✗ One flag can overrule the majority

Example: if the predefined hierarchy is: 1. B, 2. D, 3. F, 4. E, 5. P, 6. U
 $f(P, BP, U, PE, D, U) \rightarrow B$

Frequency

✓ Can take into account multiple flags per value

✗ Assumes same weight for each flag

Example: $f(P, BP, U, PE, D, U) \rightarrow P$

Weighted Frequency

✓ Takes into account the relative contribution to the total

✗ Needs defined/known weights

Example: $f(flags = \{P, BP, U, PE, D, U\}; weights = \{1, 3, 5, 4, 1, 6\}) \rightarrow U$

The methods are implemented in an R package ('*flagr*'), available at:

<https://github.com/eurostat/flagr>

> devtools::install_github("eurostat/flagr")