Harmonized sub-national crop statistics of the European Union

Dataset and Flagging System

The current version of the *Harmonized European Union subnational crop statistics* dataset has **13 fields**, including:

- REGION: The code of region based on the Eurostat classification of NUTS 2016.
- CROP_NAME: The name of crop as in the Eurostat definition.
- YEAR: Year of data publication.
- VARIABLE: The indicator measured in the year (i.e., Area, Yield, Production).
- VALUE: The value of the indicator reported in the db.
- UoM: Unit of measure on the variables: Area [ha], Production [t], Yield [t/ha].
- SOURCE: The data source of value (i.e., National Statistical Institutes, Eurostat or Mixed).
- CALCULATED_R (flagging system): If a value has been derived from a NUTS version different from NUTS 2016.
- CALCULATED_C (flagging system): If a value has been derived from different crop codes from the ones in use.
- CALCULATED_V (flagging system): If a value is missing or null or 0 and has been calculated after the data collection, according to some defined rules.
- ZERO_AS_NULL (flagging system): If a value of zero has been turned into a Null because of any inconsistencies.
- COHERENCE_APY (flagging system): If there is coherence among values of Area, Production and Yield.
- COHERENCE_CROP (flagging system): If there is coherence among values of Total wheat (Total barley), Soft wheat (Winter barley) and Durum wheat (Spring barley).

Flagging system:

A total of six flags were reported together with the data, representing additional information on data processing and data quality. In general, all the flags are set as *Yes* or left *blank*, depending on whether the specific condition occurs or not. Fields about coherence are an exception, because for those cases flags are set as *Yes* or *No*, depending on whether coherence is verified or not, and left *blank* when it is not possible to evaluate coherence, due to missing or null values.

All the rules implemented in the data post-processing procedure are summarized in the following Tables:

Table 1. Rules implemented for calculating new values and convert zero as null.

Available variables:		**Yield to be calculated:**	
AREA	*PRODUCTION*	*0*	*NA*
0	*0*	0	0**
	NA	0	NA
	Value	NA*	NA
NA	*0*	0	NA
	NA	NA*	NA
	Value	NA*	NA
Value	*0*	NA	NA
	NA	NA	NA
	Value	P/A**	P/A**

Available variables:		**Production to be calculated:**	
AREA	*YIELD*	*0*	*NA*
0	*0*	0	0**
	NA	0	NA
	Value	NA*	NA
NA	*0*	0	NA
	NA	NA*	NA
	Value	NA*	NA
Value	*0*	NA	NA
	NA	NA	NA
	Value	AxY**	AxY**

Available variables:		**Area to be calculated:**	
YIELD	*PRODUCTION*	*0*	*NA*
0	*0*	0	0**
	NA	0	NA
	Value	NA*	NA
NA	*0*	0	NA
	NA	NA*	NA
	Value	NA*	NA
Value	*0*	NA	NA
	NA	NA	NA
	Value	P/Y**	P/Y**

* ZERO_AS_NULL = Yes; ** CALCULATED_V = Yes; NA = Not Available

Table 2. Rules implemented for calculating new values for total wheat or total barley.

AREA		
Crop1	*Crop2*	*Total*
0	*0*	0*
	NA	NA
	Value	$A_{crop1} + A_{crop2}$*
NA	*0*	NA
	NA	NA
	Value	NA
Value	*0*	$A_{crop1} + A_{crop2}$*
	NA	NA
	Value	$A_{crop1} + A_{crop2}$*

PRODUCTION		
Crop1	*Crop2*	*Total*
0	*0*	0*
	NA	NA
	Value	$P_{crop1} + P_{crop2}$*
NA	*0*	NA
	NA	NA
	Value	NA
Value	*0*	$P_{crop1} + P_{crop2}$*
	NA	NA
	Value	$P_{crop1} + P_{crop2}$*

YIELD		
Area $_{total}$	*Production* $_{total}$	*Yield* $_{total}$
0	*0*	0*
	NA	NA
	Value	NA
NA	*0*	NA
	NA	NA
	Value	NA
Value	*0*	P_{total}/A_{total}*
	NA	$[(A_{crop1} x Y_{crop1})+(A_{crop2} x Y_{crop2})]/A_{total}$*
	Value	P_{total}/A_{total}*

* CALCULATED_V = Yes; NA = Not Available; Crop1=Soft wheat (winter barley); Crop2=Durum wheat (spring barley)

Table 3. Rules implemented for checking agreement among values of Area, Production and Yield.

COHERENCE_APY		
Yes	*No*	*Blank*
abs(P-(AxY)) <= 0.01xP	abs(P-(AxY)) > 0.01xP	Any variable is NA*

* NA = Not Available

Table 4. Rules implemented for checking agreement among values of Total wheat (Total barley), Soft wheat (Winter barley), and Durum wheat (Spring barley).

COHERENCE_CROP - Area		
Yes	*No*	*Blank*
$abs(A_{total} - (A_{crop1} + A_{crop2}) <= 0.01xA_{total}$	$abs(A_{total} - (A_{crop1} + A_{crop2})) > 0.01xA_{total}$	Any crop is NA*

COHERENCE_CROP - Production		
Yes	*No*	*Blank*
$abs(P_{total} - (P_{crop1} + P_{crop2})) <= 0.01xP_{total}$	$abs(P_{total} - (P_{crop1} + P_{crop2})) > 0.01xP_{total}$	Any crop is NA*

COHERENCE_CROP - Yield		
Yes	*No*	*NA*
$abs(Y_{total} - [(Y_{crop1}xA_{crop1} + Y_{crop2}xA_{crop2})/A_{total}]) <= 0.01xY_{total}$	$abs(Y_{total} - [(Y_{crop1}xA_{crop1} + Y_{crop2}xA_{crop2})/A_{total}]) > 0.01xY_{total}$	Any crop is NA*

* NA = Not Available