



# **EUROMOD input data preparation - role of Estat**

**Albane Gourdol**

EUROMOD Annual Meeting  
Milan, 26 September 2019

# Estat's role



- Responsibility for the
  - Creation (start and end up the process)
  - Documentation (useful info to all users involved in the data preparation)
  - Validation of the EM input datasets
    - Final say as EC statistics but interaction with NTs/JRC when doubts, discrepancies
    - Monitor consistency of EM input dataset as a whole

# Work carried out so far

European  
Commission

## Detailed info

- “EM model relies on detailed info at ind. level”
  - 36 disaggregated benefits mandatory from 2021 on
  - New metadata on income benefits from 2018 onwards

List of national benefits to be pasted in this column	Benefit name in English	Click on the cell and select the variable from the drop-down list		Summary statistics on the disaggregated variables						Notes	...
		Contained in Harmonised Variable (G)	Contained in Disaggregated Variable (G)	NON-WEIGHTED				WEIGHTED			
				Number of beneficiaries	Total amount (national currency)	Minimum value (national currency)	Maximum value (national currency)	Number of beneficiaries	Total amount (national currency)		
Peněžité pomoci v mateřství	Income maintenance in the event of childbirth	HY050	HY052	150	9567703	8050	203000	98079	6213634862		
Porodné	Birth grant	HY050	HY053	13	158000	10000	23000	8272	100045923		
Přídavky na děti	Family or child allowance	HY050	HY053	291	4228360	3000	120000	185630	2540532227		
Rodičovský příspěvek	Parental leave benefit	HY050	HY054	534	35552520	2100	138000	351564	23425823564		
Dávky péče o děti (opakované)	Other cash periodic benefits non means-tested	HY050	HY054	22	3602394	54000	480000	10418	1609950568		
Dávky péče o děti (jednorázové)	Other cash lump sum benefits non means-tested	HY050	HY054							included in Other cash periodic benefits non means-tested	

- Pass on the list of needed detailed var to NSIs

# Work carried out so far



## Data quality

- Consistency checks done earlier in the process (UDB production step)
  - Between ID person and ID household
  - Father/mother ID are the same as Partner ID (*new in 2018*)
  - Age difference between parents and children (*broaden 2018*)
  - Disaggregated benefits (*new 2019*)
    - Consistency between the data and the flag var
    - Aggregate filled and at least one component also
    - Aggregate = sum (disaggregated components)

# Work carried out so far



## Data quality

- Imputation of non income var.
  - EUROMOD model features: run with no missing values
  - Missing values imputed according to various methods with a research oriented view
  - MS opinion: improve data quality for statistical purposes by transmitting data without missing values for these variables (when possible) (written consultation: 10Y – 4N - 6 ?)
  - Variables concerned: 3 domains – 40 var



## Premises behind Estat's workplans

- Involvement of the EM NT experts → role in the creation of the EM input file when country-specific knowledge required
- Ensure the quality of the datasets created (EMSD + EM input file)
  - Compliance with the ESS quality framework
- Ensure data access to:
  - EMSD for all actors involved in the preparation of EM input file (NT/Essex, JRC)
  - EM input file for all EM users

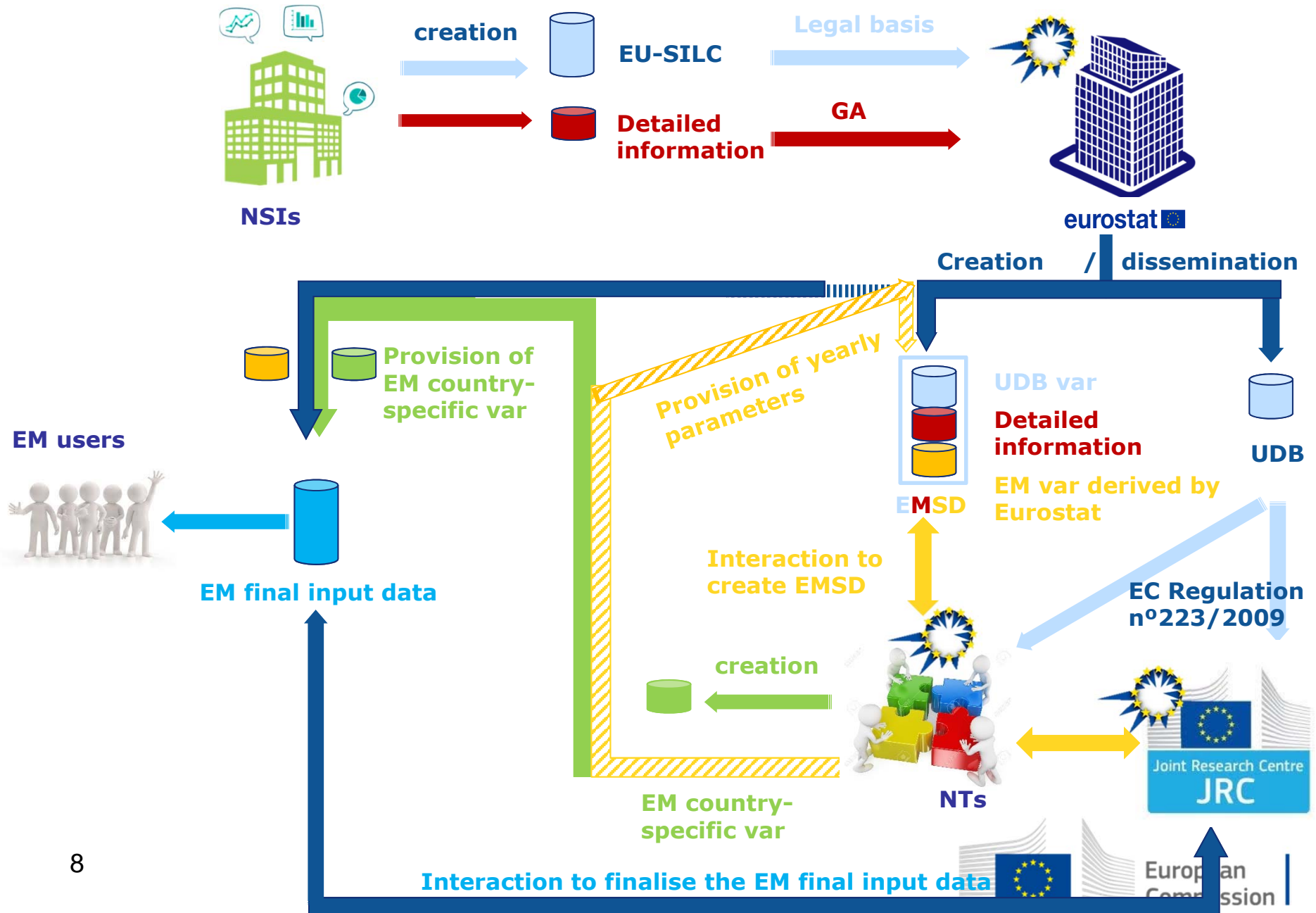


## Changes in data flow

- Data preparation based on EMSD country-specific dataset
  - UDB variables common among MS
  - detailed data from national SILC national list  
ex. LV (municipality of residency, State Social Security benefits (old age, survivors, disability), civil servant information)
  - imputations of missing values/inconsistencies in EM var. performed in a harmonised way across countries  
ex. marital status (dms) var. common EM var. among MS
  - imputations of EM missing values based on variables from the production database (PDB) to avoid top/bottom codings  
country-specific EM var.
- EMSD provided beginning of EM year

**EUROMOD data preparation (foreseen) from 2021 onwards**

GA stands for gentleman's agreement  
EM stands for Euromod







## EM input file preparation

### EM input file

Personal information	EM var. created by Eurostat in the EMSD
	Countric-specific EM var. generated by the NT
Labour market	EM var. created by Eurostat in the EMSD
	Countric-specific EM var. generated by the NT
Income, Benefits and Taxes information	EM var. created by Eurostat in the EMSD
	Countric-specific EM var. generated by the NT
Value of assets	EM var. created by Eurostat in the EMSD
	Countric-specific EM var. generated by the NT
Expenditures	EM var. created by Eurostat in the EMSD
	Countric-specific EM var. generated by the NT



## Changes in data flow

- EMSD advantages
  - No need to merge  $\neq$  datasets
  - No ID maps issues
  - Sustainable and consistent access to data sources
  - all EM var. where Estat is in charge of their generation
  - Saving time to dvp further blocks in the model
  - Benefit from NTs country-specific knowledge for more complex EM var.
- EMSD provided with
  - Metadata to explain the EMSD
  - Metadata on the disaggregated benefits
  - DRD updated
  - Information to draft chap 3 in the CR



## In Eurostat

- EMSD Y10 for 6 countries (LV, BG, EE, ES, CZ, FI)
    - comprehensive panorama of EMSD structure
      - EM var. straightforward derived from SILC (national var, combination UDB-PDB var)
      - EM var. generated with imputation methods
      - EM var. where a country-specific knowledge is required (NT)
    - split EM var. into 3 types ( $\neq$  for each MS, based on Y7, Y9)
      - Type "E" – Estat ex. bed (BG)
      - Type "E" with parameters ex. dec (BG)
      - Type "N" – NTs ex. yivwg (BG)
- Estat's portfolio
- NTs responsibility

# From theory to



# practice

## In Eurostat




- Method to create EMSD Y10
  - Derivation of types "E" var. according Y9 methodology
  - UDB-PDB-national datasets
  - Assessment of our results based on summary stat. on both collection channels
  - A step further on the disaggregated benefits (type "N" var.)
- Bilateral discussions with NSIs for the EMSD dissemination



## Transition

	Official flow	Pilot test
2020	Y11 EUROMOD input files (provided by national teams)	Y10 EUROMOD input file on 2017 data for 9 MS (AT, EL, HR, LT, LU, RO, SE, SI, UK) Y11 EMSD on some MS Y10 EUROMOD input file on 2017 data for a subset of MS according to the availability of the national teams

	EUROMOD input file provided by a joint work from Eurostat and the national teams		current EUROMOD input data files created by national teams		EMSD created by Eurostat
--	--	--	--	--	--------------------------

- 2020: test the proposed data flow with some of NTs on EM Y10



## After the transition period

- No change in the starting/ending dates of EM year (Feb-Dec)
- Start of EM year: bilateral meeting with each NT+JRC+Estat
  - Agreement on EM types "E" and "N" var.
    - fuzziness for some var: type "E" or "N" ?
  - ex1. afc – financial capital*  
easy to compute but data source for "avir" country-specific
  - ex2. in BG dcz var.*  
change in Y10 do-files compared to Y9 on a yearly basis or one shot?
    - Provision of national parameters needed
      - Nov of EM  $Y_N$  for EM  $Y_{N+1}$  (available in CR)?
- EM input file: back and forth as today

# Summary



- In 2019
  - Important work on the SILC data quality and consistency
  - In-depth analysis on detailed var., relevant for EM
  - Preparation inside Estat of the EMSD
  - Reflection on changes in current data flow and agreement with JRC and Essex
- In 2020
  - Test the new data flow with some NTs
  - Continue bilateral discussions with NSIs

- Now : time for questions

