



Session 2- Recent developments

Data updates :  
milestones achieved and EM Y<sub>13</sub> data  
preparation

EUROMOD annual meeting  
4 – 5 October 2021



# Content

- Achievements since last meeting
- Next steps to prepare EM Y<sub>13</sub>
- Country plans for EM Y<sub>13</sub>
- EMSD data preparation timeline
- Feedback from the FE (Aura)

# Milestones achieved –

## **New process to prepare EM input data (reminder)**

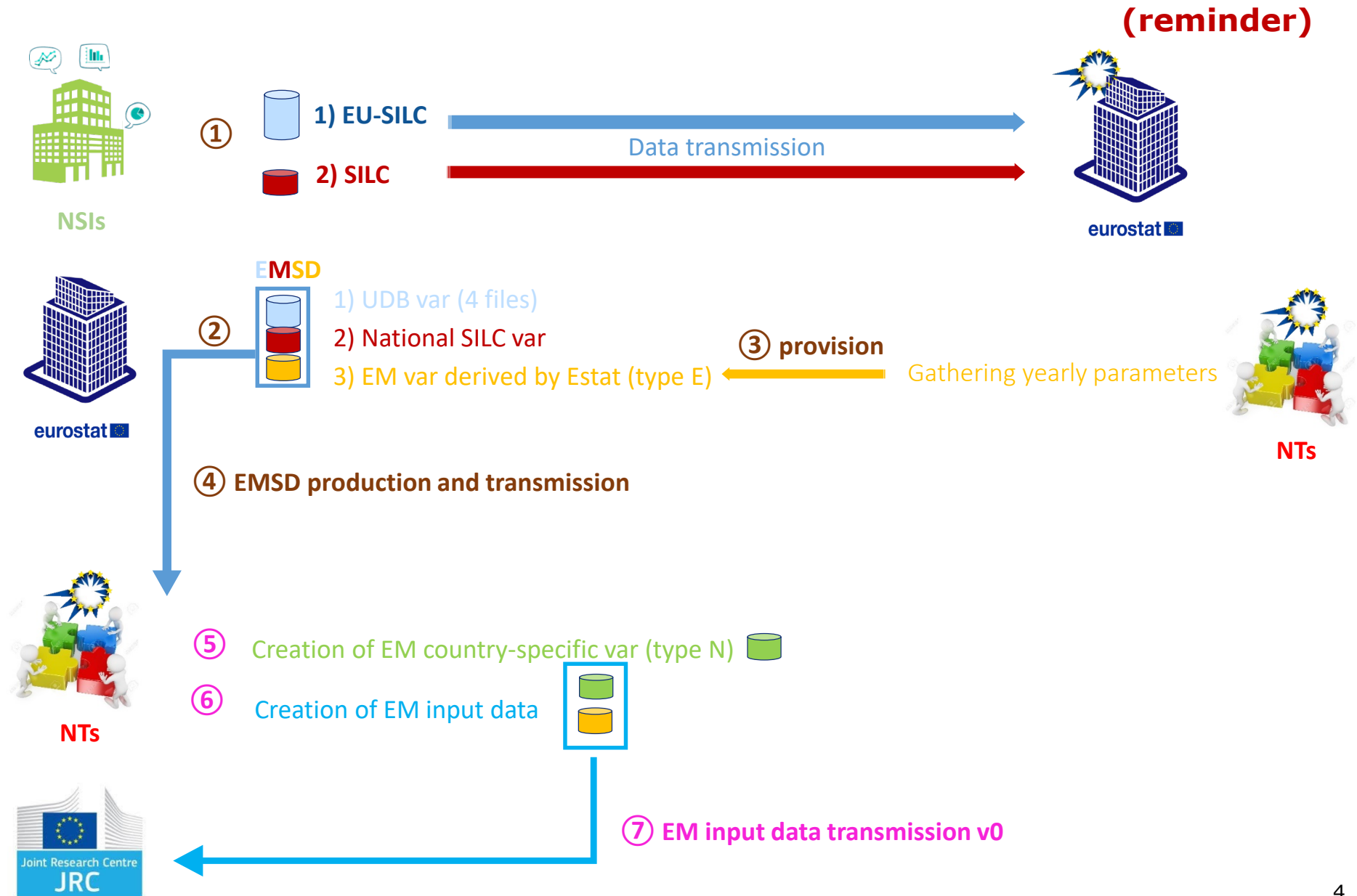
### Key elements to prepare the EM input data

- Provision of a new dataset – EMSD (no need to merge ≠ datasets, no ID maps issues, data consistency already implemented)
- Additional info from non anonymised EU-SILC dataset (PDB var.)
- Utmost use of national information for all countries
- Classification of EM var. to take on board NT's expertise and info available inside Estat



More details in Eurostat presentation (2020 EM annual meeting)

# Milestones achieved-New data workflow



# EM Y<sub>12</sub> - Milestones achieved – Delivering 10 EMSDs

Y12 EM data preparation subject to legal aspects →  
legal basis (PA) signed before 2020 Xmas break

- List of national variables transmitted by NSIs
- Non anonymised var. (PDB var.) used by Estat (solely)
- Conditions to access the EMSD (access form) and the EM input var.

EMSD preparation (from Nov 2020 to Feb 2021)

- Kick-off meetings with NTs (Nov to January)
  - 1<sup>st</sup> discussion on EM var. categorisation (E/E+p/N)
  - Yearly parameter for EM Y<sub>12</sub> (available in CR)
  - Estat data processing
  - Presentation of informative documents accompanying EMSD dataset

# EM Y<sub>12</sub> - Milestones achieved – Delivering 10 EMSDs

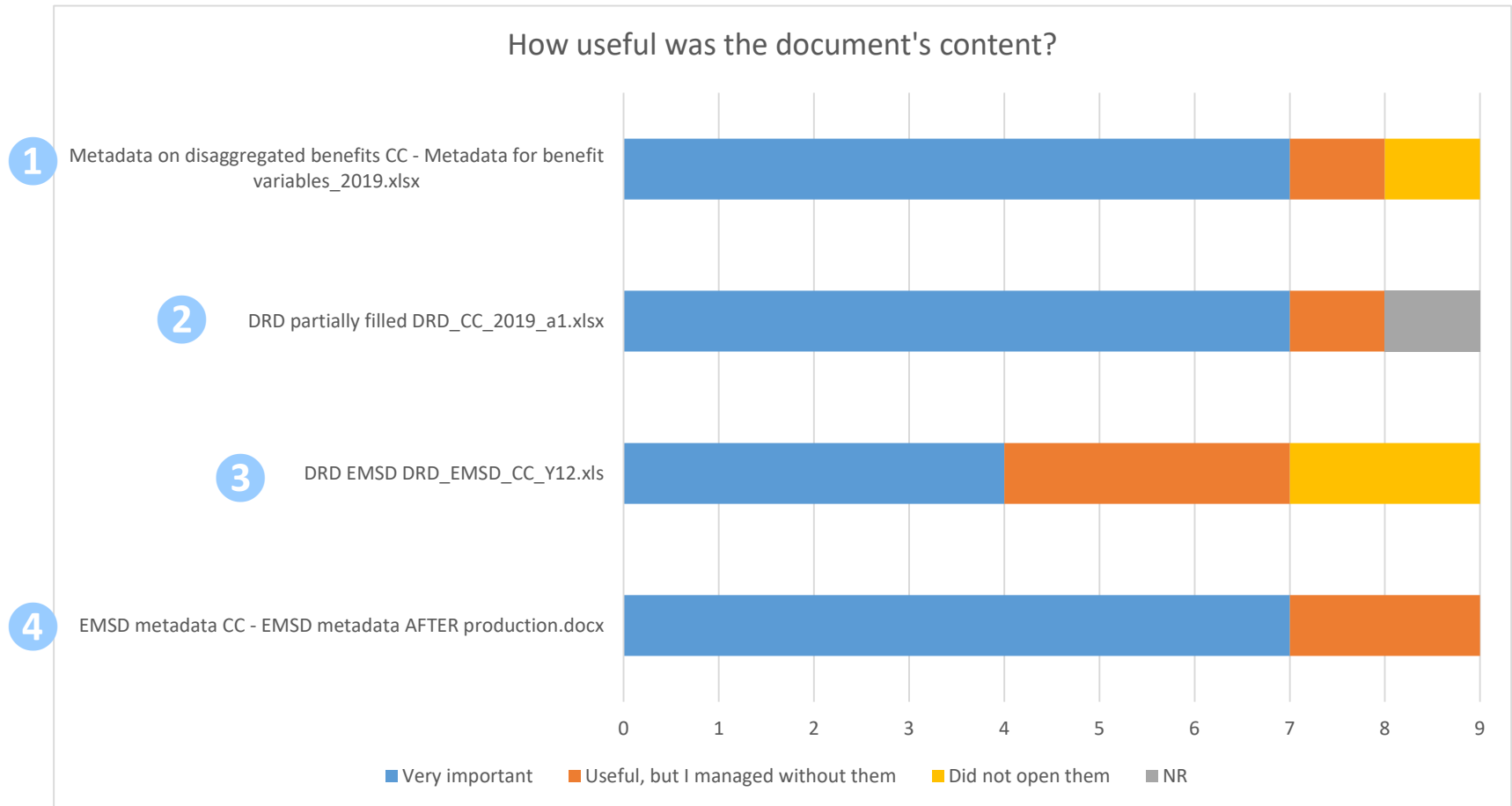
## EMSD preparation

- EMSDs production
  - EMSD<sub>0</sub>: EM var. computed by Estat (on UDB information - Estat internal use)
  - EMSD<sub>2</sub>: UDB + national var. + EM var. computed by Estat (with selected PDB var.)
- Summary statistics on EMSD<sub>0</sub> and EMSD<sub>2</sub> → impact of PDB var. use
- Support notes on EMSDs contents, data handlings, data quality

## Feedback from NTs on the use of EMSD data

- E-survey launched after the EM input data preparation (RR=90%)
- Synthesis available, various topics covered:
  - Organisation (i.e. EMSD delivery date - 100% satisfy 😊)
  - Support documents (usefulness, contents, clarity, improvements)
  - Eurostat data processing
  - Request for additional national var. needed → discussion with the concerned NSIs

# Support documents - 4 files sent with the EMSDs (usefulness, contents, user-friendliness, improvements)



- Majority of NTs found the documents to be user-friendly in terms of their clarity, layout and useful regarding their contents
- Doc 2 and 4 most important files
- Doc 3 can be removed from Estat's delivery

## Support documents – Improvements of EMSD metadata doc

- No need to remove or add tables / information
- Add a dictionary of var. and their modalities for UDB and national SILC datasets



- include link to the DOC65, which is available **on-line**
- ask NSIs the National SILC codebook

- For each EM var. computed by Estat, # obs. for which the missing values were imputed



- implemented in Estat Y<sub>13</sub> SAS programs but only total # of obs. imputed available (software not suitable to compute the figure at each step of the imputation process)

## Estat data processing – really useful

- Data consistency checks
- In particular corrections of ID (personal ID  $\neq$  to Household ID; parent ID refers to a non-existing person)



## Overall opinion – workflow easier

- 😊 time to focus on country-specific var.
- 😊 access to national datasets
- ☹️ if already access to full national info, process longer with an additional actor

## Overall opinion – suggestions for improving the whole process

- Organization of the do-files (“a do-files” not relevant anymore)
  - 1<sup>st</sup> year: check Estat var. so run the code lines for type E var. as well
  - Afterwards: automatically replace the EM var. with those computed by Eurostat
- Summary statistics in DRD file to be computed with one software (in Y12 - type E with SAS and type N with Stata)
  - Eurostat will not compute them anymore, NTs will do it for all EM var.
- ? Access to the do-files used by Eurostat
  - Eurostat works in SAS but code available in the DRD partially filled document

# EM Y<sub>12</sub> - Milestones achieved – Template for the CR

Draft paragraph for the CR (harmonization purpose)

- Short description of EMSD
- Use EMSD as input source
- Resort to PDB var. for some imputations
- Disclosure rules applied and consequences if end users replicate imputations based on the UDB dataset

Variable name	Description	Rules in UDB	Disclosure rules in the EMSD	Disclosure rules in the Euromod input file
DB040	Region	NUTS2 grouped for 1 region	none	none
PE020	ISCED level currently attended	Top coded	none	Same top coding rule as in the UDB for the EUROMOD variable “decde”
RB070	Month of birth	Re-grouped	none	Same top coding rule as in the UDB for the EUROMOD variable “dmb”
RB080	Year of birth	Top coded	none	not disseminated
age (PDB) / RX020, PX020 (UDB)	Age at the end of the income reference period	Top coded	none	Same top coding rule as in the UDB for the EUROMOD variable “dag”

# Next steps to prepare EM Y<sub>13</sub> (2020 SILC data)

## Legal work (practical arrangements)

- Edit some PAs ready in 2020 but not signed (list of national var – LV; signatory names) and amendments to include additional var. (MT BG NL)
- Follow up the signatures for the remaining countries → deadline with NSIs: November 2021

## Collect required information

- Optional COVID<sub>19</sub> var. included in UDB
- Collect info from NTs to prepare EMSDs Y<sub>13</sub>  
10 NTs already contacted, deadline 20.10.2021
- Kick-off meeting with other NTs between October-December  
1<sup>st</sup> discussion on EM var. categorisation (E/E+p/N)  
Yearly parameter for EM Y<sub>13</sub> (available in CR)  
Data handling practices requested  
Presentation of support documents accompanying EMSD dataset  
→ Doodle poll to find a date

# Next steps to prepare EM Y<sub>13</sub> (2020 SILC data)

## Documentation

- Improve support documents with NTs feedback
- Upgrade the DRD with new information so it remains the main source of information when it comes to data-related issues

\* = mandatory  
! = var derived or imputed with PDB var.

Personal Information			Description
Variable	Label	Notes: derivation from original data, and comments	Description (name in original language in brackets)
* idhh	IDENTIFIER : hh	idhh = db030	idhh contains the id number of the household.
* idperson	IDENTIFIER : person	idperson = rb030	id number of the individual
* idfather	IDENTIFIER : father	idfather = rb220	idfather includes the id of the step/adoptive/foster father
* idmother	IDENTIFIER : mother	idmother = rb230	idmother includes the id of the step/adoptive/foster mother
* idpartner	IDENTIFIER : partner	idpartner = rb240	idpartner contains the id number of the partner. It includes the id of the partner's household.
* idorighh	IDENTIFIER : orighh	idorighh = db030	id number of the household.
* idorigperson	IDENTIFIER : origperson	idorigperson = rb030	id number of the individual
! dag	DEMOGRAPHIC : Age	dag = rx020	information on the age of the individual at the end of the survey year
! * dncsy	DEMOGRAPHIC : Number of children born	Children born after income reference period are identified with	number of children born in survey year (i.e. children born in survey year)
! * dec	DEMOGRAPHIC : Education - currently attended	gen temp_pe020 = pe020	ISCED level currently attended
! * decde	DEMOGRAPHIC : Education - currently attended (detailed)	gen decde = pe020	ISCED level currently attended (detailed)
! * deh	DEMOGRAPHIC : Education - highest level attended	gen temp_pe040 = pe040	highest ISCED level attended
! * dehde	DEMOGRAPHIC : Education - highest level attended (detailed)	gen dehde = pe040	highest ISCED level attended (detailed)
! * dew	DEMOGRAPHIC : Education - year when highest level of education was attained	dew = pe030	the year when the highest level of education was attained
! * dmb	DEMOGRAPHIC : Month of Birth (proxied by the quarter of the year)	dmb = 2 if rb070 == 1 dmb = 5 if rb070 == 2 dmb = 8 if rb070 == 3	month of birth (proxied by the quarter of the year)
* dgn	DEMOGRAPHIC : Gender 0: Female 1: Male	dgn = rb090	gender
* dms	DEMOGRAPHIC : Marital Status 1: Single 2: Married 3: Separated 4: Divorced 5: Widowed	dms = pb190	marital status of the individual
* dcz	DEMOGRAPHIC : Citizenship 1: This country 2: other EU 3: Other	dcz = 1 if pb220a == "LOC" dcz = 2 if pb220a == "EU" dcz = 3 if pb220a == "OTH"	information on citizenship
* ddi	DEMOGRAPHIC : Disability	ddi = 1 if pl031 == 8	ddi contains basic information on whether a person is disabled

IT LU IE EE = 4

### EMSD Y<sub>13</sub> - new data workflow

- PA signed in 2021

AT BE BG CZ ES FR HR HU MT  
NL=10

### EMSD Y<sub>13</sub> - routine process

EM input data prepared with  
EMSD Y<sub>12</sub>

Condition: PA between JRC-Estat-  
NSI signed in 2020

## Country plans for EM Y<sub>13</sub>

### EMSD Y<sub>13</sub> - likely new process

PA not yet signed but meeting  
with NSIs and regular contacts

Deadline: November 2021

### No EMSD in Y<sub>13</sub>

PL : legal issue

PT : no need of extra national var.

RO : NSI wants to start with 2021  
survey year data

CY DE DK EL FI LT LV SE SI SK = 10

PL PT RO = 3

# EMSD data preparation timeline

Oct – Nov

- UDB available inside Eurostat
- **National data in Estat**

Eurostat - **NSIs**

Nov-December

- Kick-off meeting EM  $Y_{N+1}$
- Changes in EM var. derivations
- EMSD preparation

NTs-JRC-Estat

Jan - Feb

EMSD and support documents preparation

**EMSD preparation = 3,5 months**

Eurostat

Feb-March

- Launch of EM  $Y_{N+1}$
- EMSD available to NTs + documentation

NTs-JRC-Estat

April

- EM release for the FE
- EM input files
- Model (policies)

NTs-JRC-Estat (disclosure rules)

August

EUROMOD Beta release

NTs-JRC-Estat

December-January

EUROMOD public release ( $\approx 200$  user requests  $Y_{12}$ )

JRC-Estat

# Before moving to the FE

Contact: [ESTAT-EUROMOD@ec.europa.eu](mailto:ESTAT-EUROMOD@ec.europa.eu)

Or 🤝 Olga – Albane - 🙋🏻 Henri

Thanks to your collaboration, commitment and support:

- 2021: new data workflow implemented in 1/3 NTs, >0 feedback
- 2022: expect the new flow for all NTs (a few exceptions) → *conditions: PA signed in 2021*