Privacy and data confidentiality for Official Statistics: new challenges and new tools

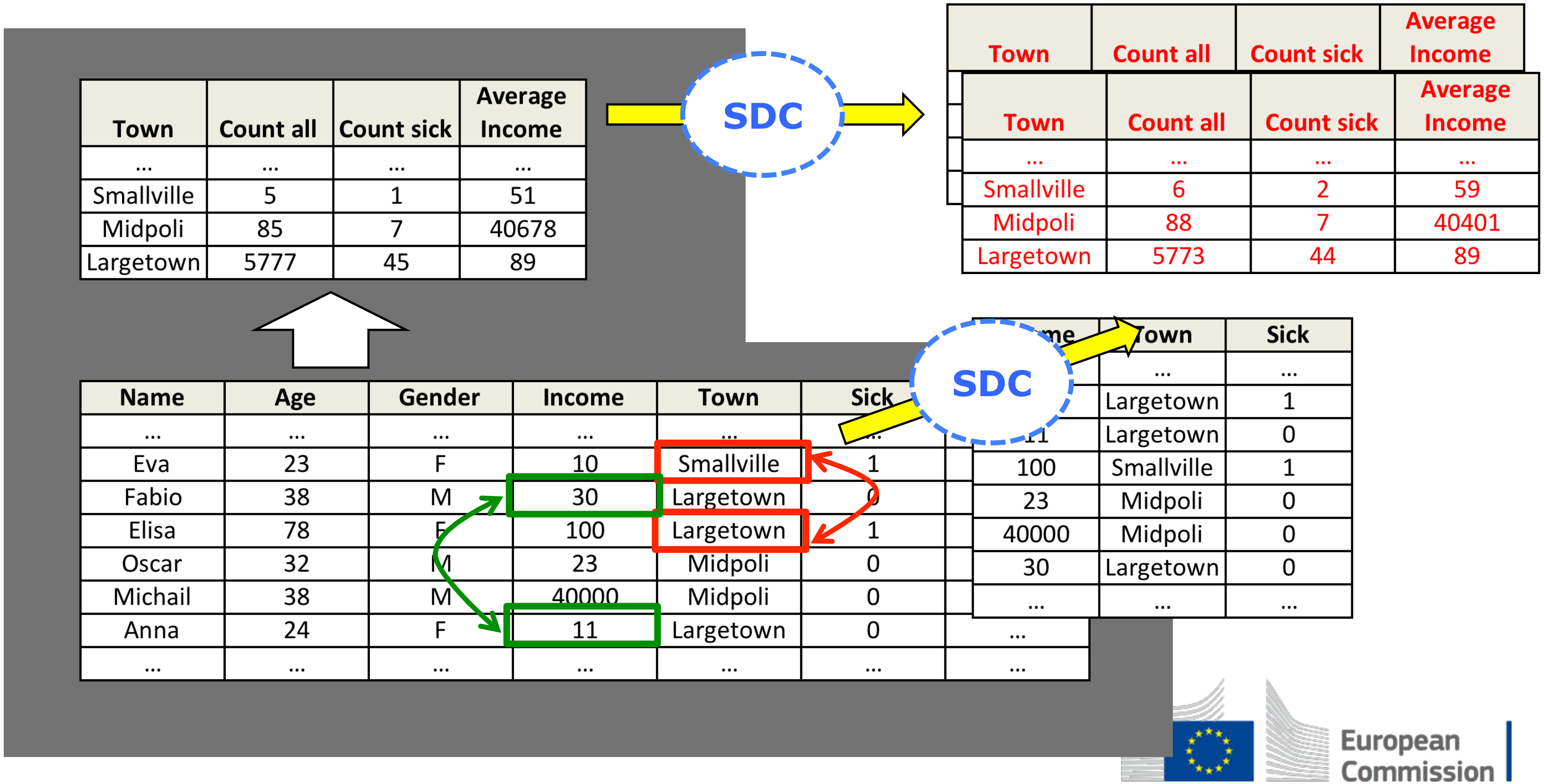
Fabio Ricciato and Aleksandra Bujnowska

EUROSTAT – Unit B1

NTTS 2019, Brussels 14.3.2019

Statistical Disclosure Control (SDC)

- *Suppression (e.g. cell deletion, column removal)*
- *Add noise, perturbation, rounding*



Town	Count all	Count sick	Average Income
…	…	…	…
Smallville	5	1	51
Midpoli	85	7	40678
Largetown	5777	45	89

Town	Count all	Count sick	Average Income

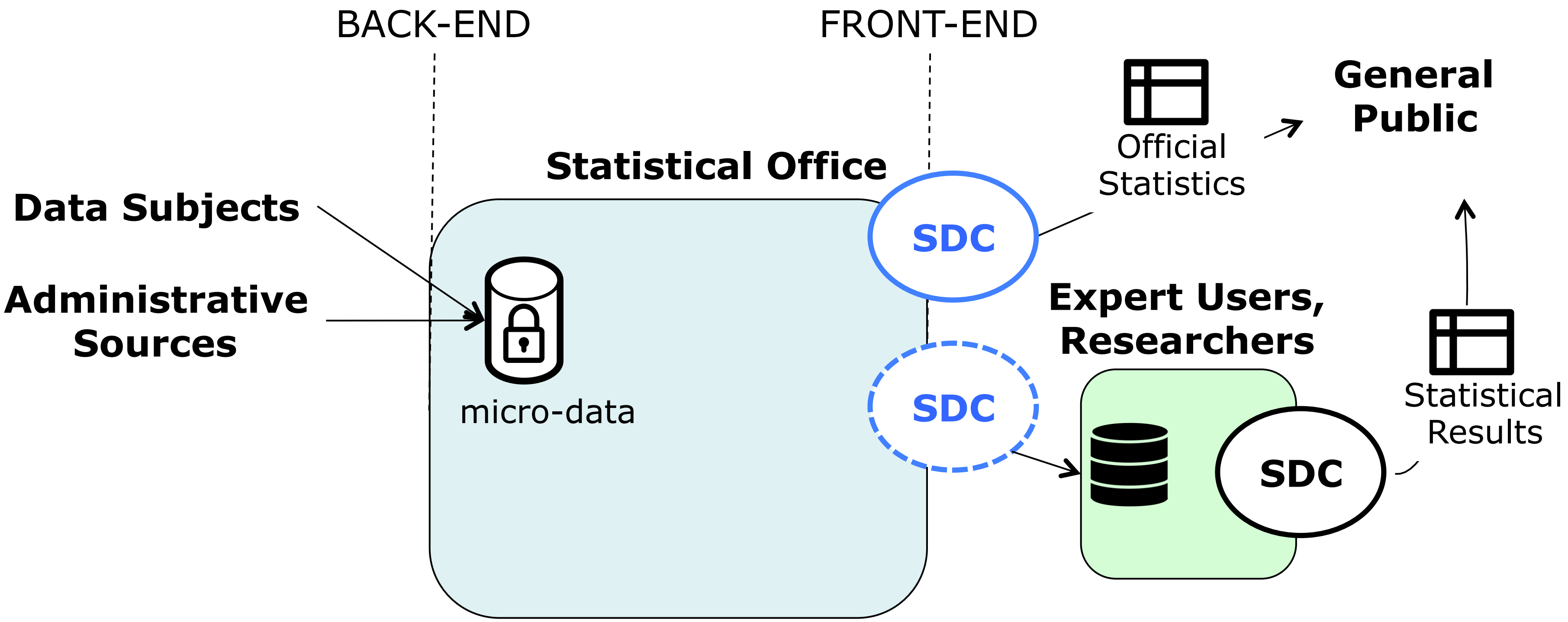
Town	Count all	Count sick	Average Income
…	…	…	…
Smallville	6	2	59
Midpoli	88	7	40401
Largetown	5773	44	89

Name	Age	Gender	Income	Town	Sick	
…	…	…	…	…	…	
Eva	23	F	10	Smallville	1	
Fabio	38	M	30	Largetown	0	
Elisa	78	F	100	Largetown	1	
Oscar	32	M	23	Midpoli	0	
Michail	38	M	40000	Midpoli	0	
Anna	24	F	11	Largetown	0	…
…	…	…	…	…	…	…

…me	Town	Sick
	…	…
	Largetown	1
11	Largetown	0
100	Smallville	1
23	Midpoli	0
40000	Midpoli	0
30	Largetown	0
…	…	…

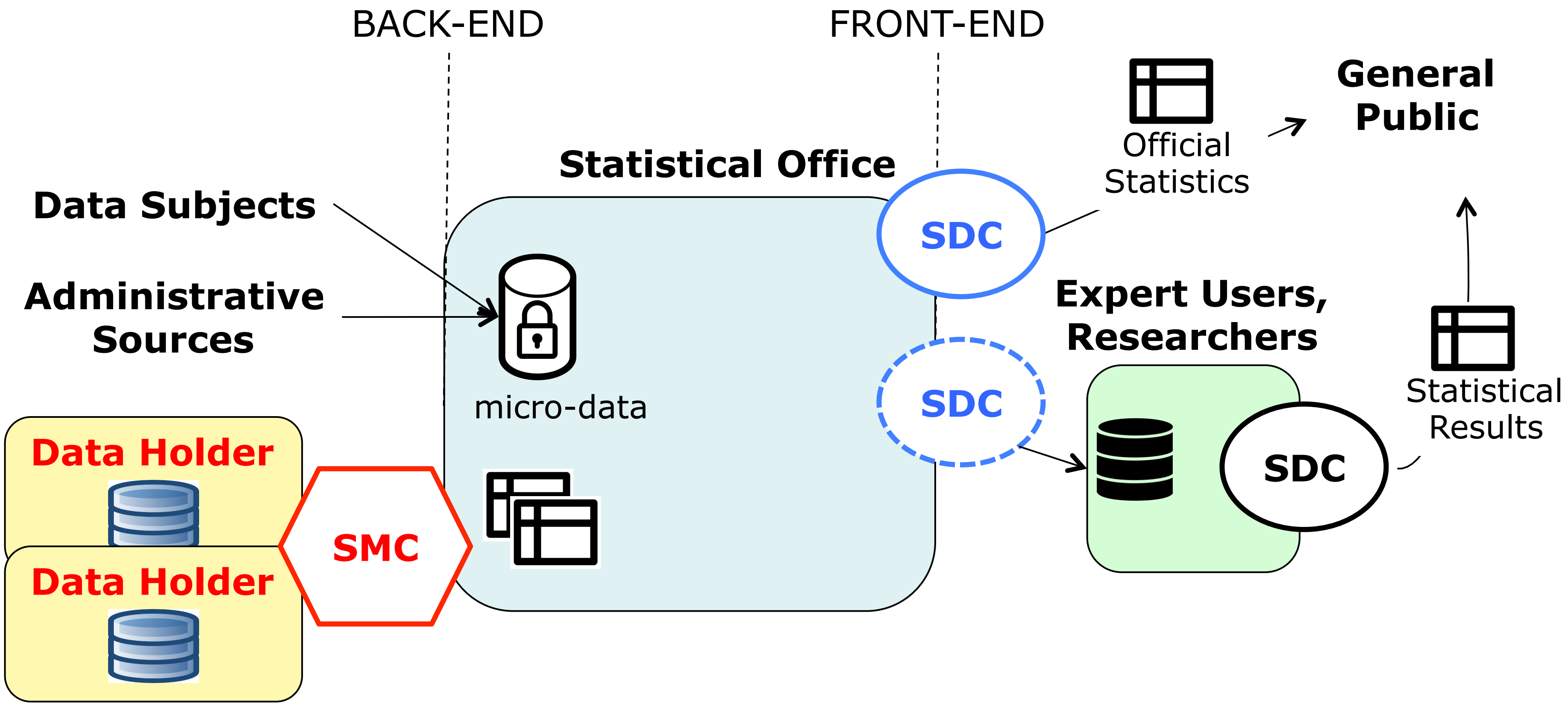
European Commission

SDC on the front-end



SDC: Statistical Disclosure Control

SMC on the back-end



SDC: Statistical Disclosure Control

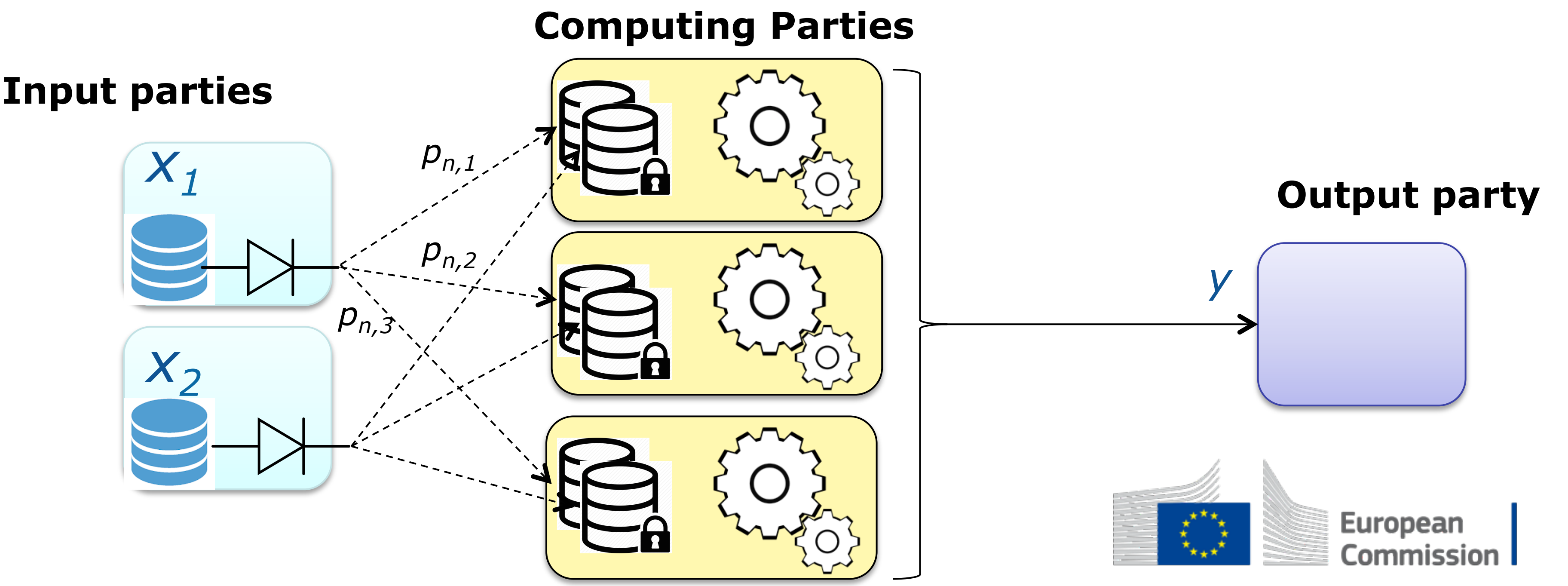
SMC: Secure Multi-Party Computation

Secure Multi-Party Computation (SMC)

- *Each element of secret input x_n is transformed into K “shares” $p_{n,1}$, $p_{n,2}$ … $p_{n,k}$ that are distributed to different computing parties.*

- *The computation on secret shares*
  - is distributed (shared) among the computing parties
  - returns the same output value that would be obtained from the input data (homomorfism)

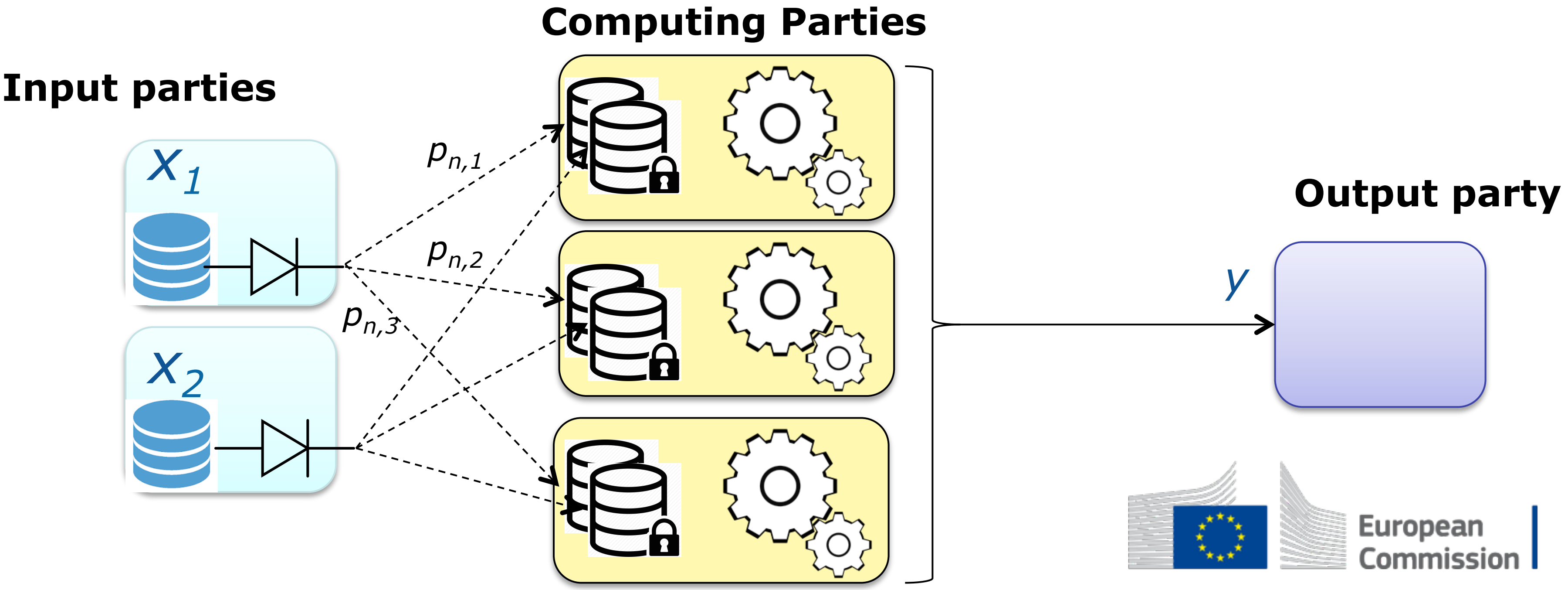
$$y = f_s\left(\left\langle p_{1,1}, p_{1,2}, p_{1,3}\right\rangle, \left\langle p_{2,1}, p_{2,2}, p_{2,3}\right\rangle\right) = f\left(x_1, x_2\right)$$



Secure Multi-Party Computation (SMC)

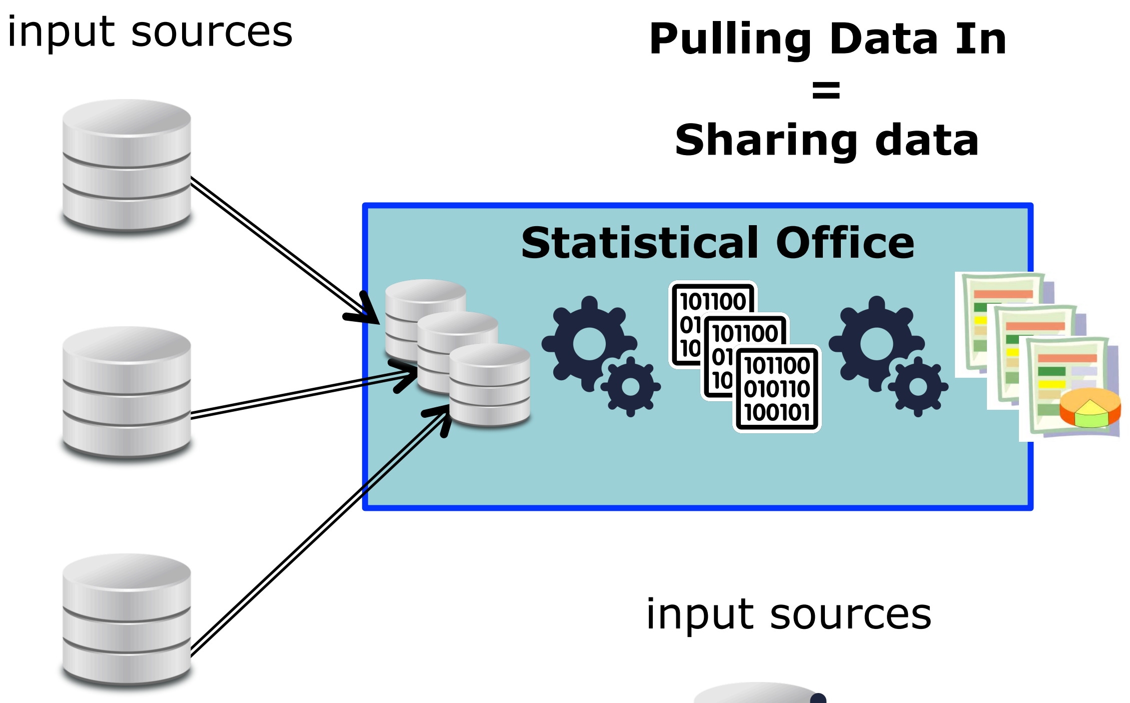
- *Individual shares do not reveal nothing about the secret input*
  - → no single party holds "data"
    → "passing shares" ≠ "sharing data"

- *Computing parties need to be trusted collectively, not individually*

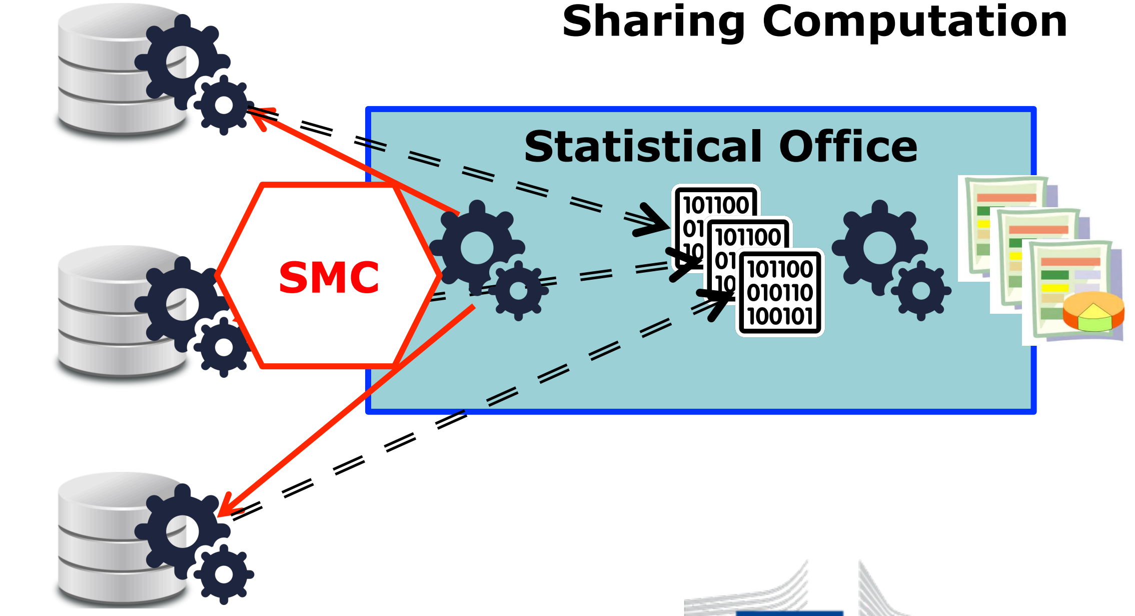


European Commission

Pulling Data In
=
Sharing data

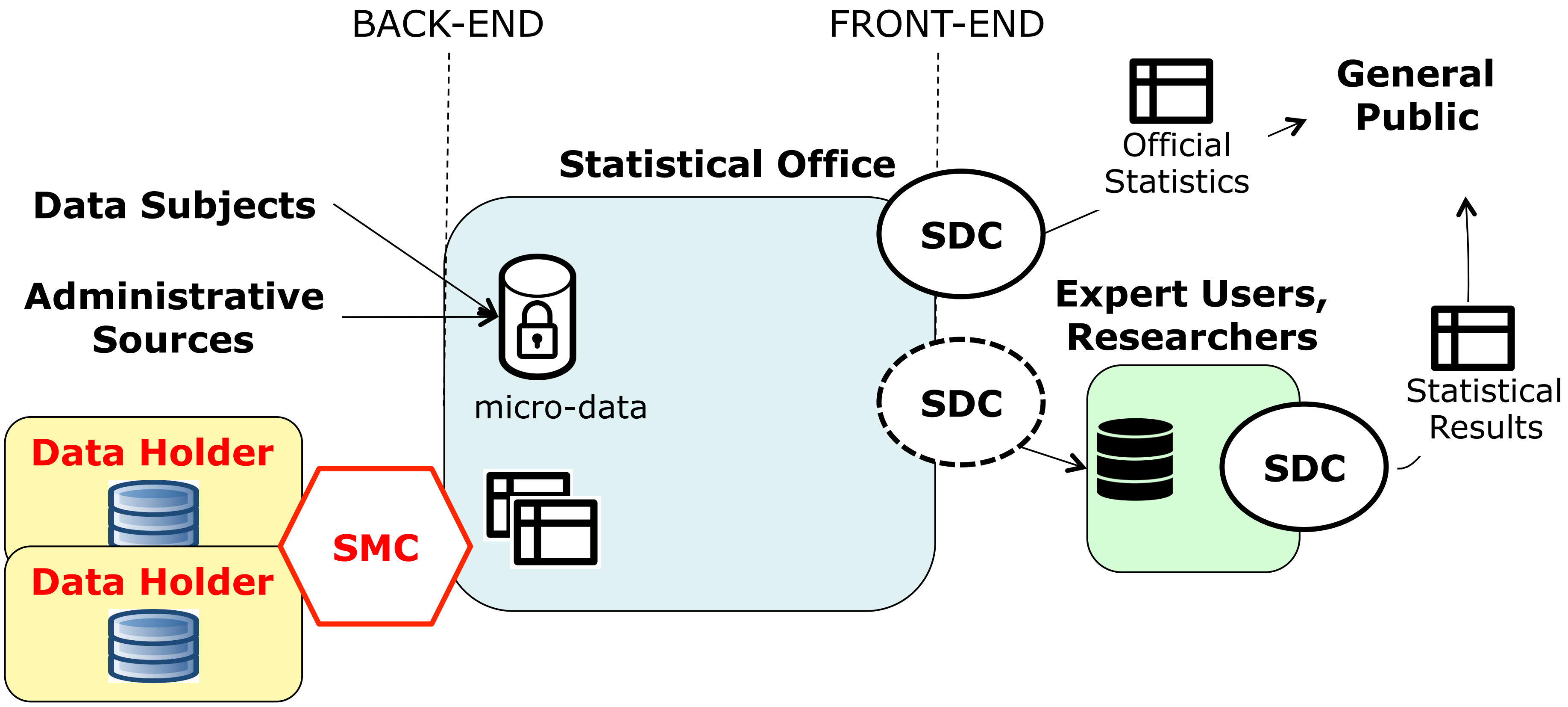


Pushing Computation out
=
Sharing Computation



SMC: Secure Multi-Party Computation

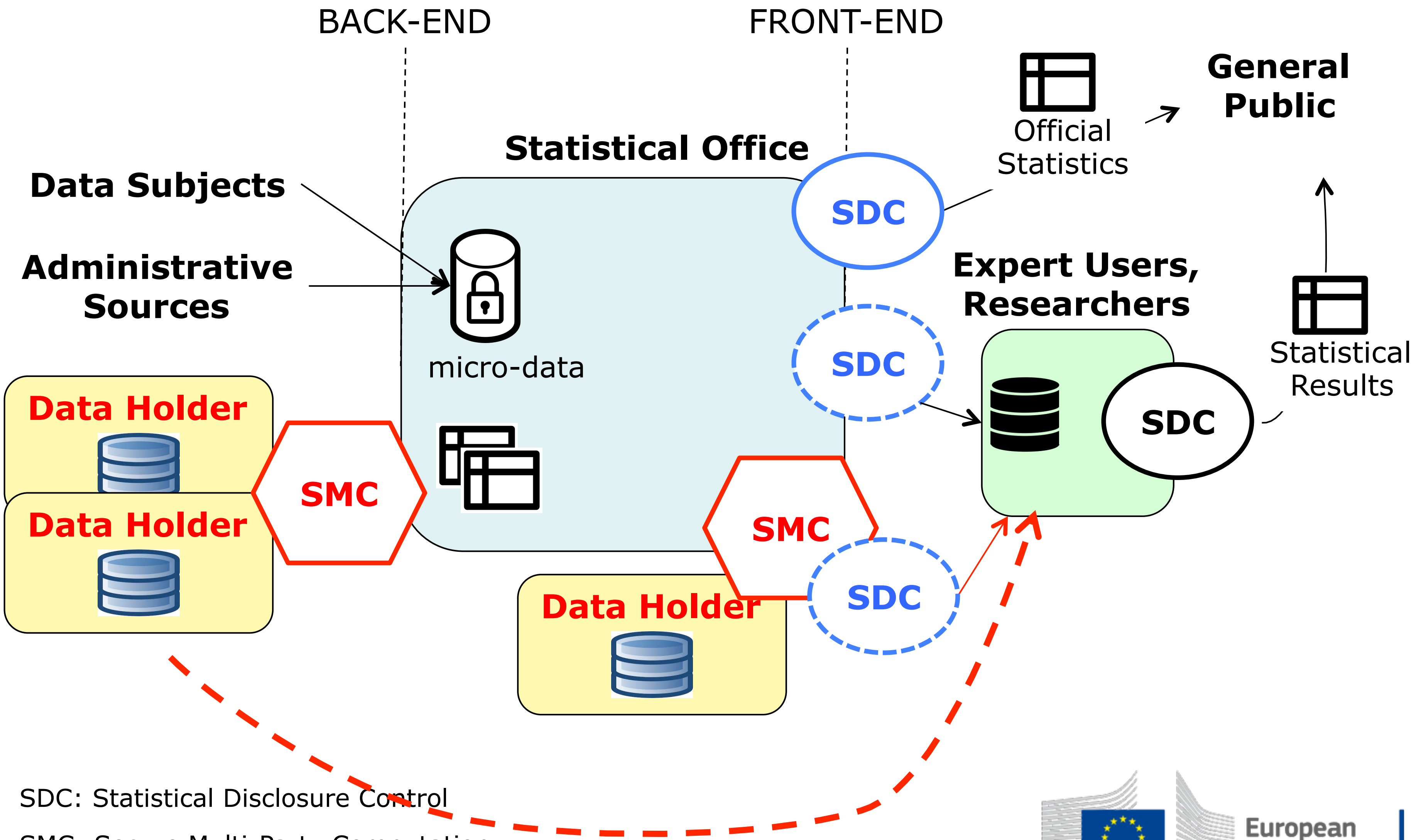
SMC on the back-end



SDC: Statistical Disclosure Control

SMC: Secure Multi-Party Computation

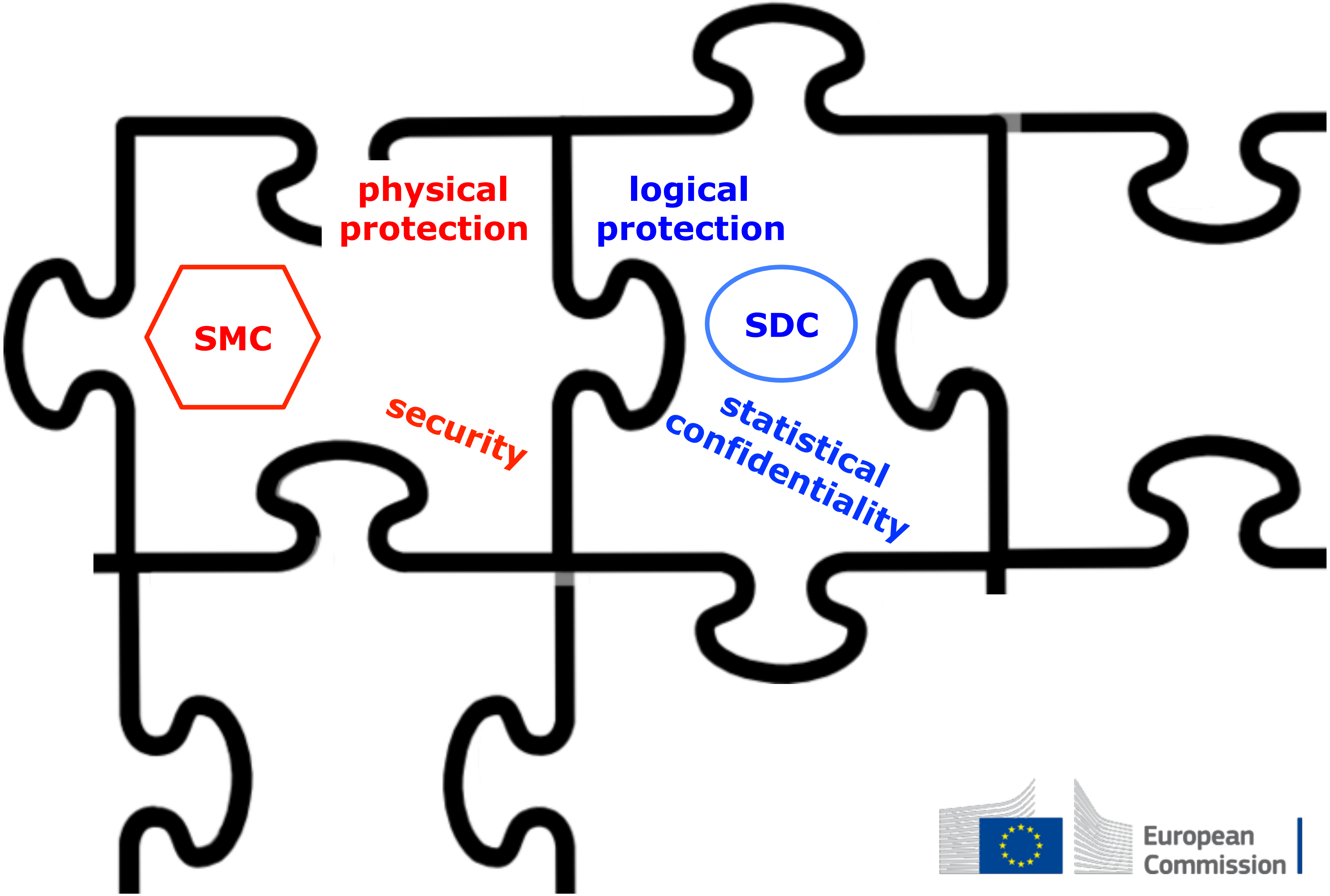
Combining SMC+SDC on the front-end?



SDC: Statistical Disclosure Control

SMC: Secure Multi-Party Computation

SMC & SDC as complementary but distinct components



Take-home messages

Confidentiality in Official Statistics need to evolve towards more articulated solutions

*Evolution of SDC solutions from traditional **static tools** solutions towards **dynamic SDC** Table Builder, on-the-fly anonymization is part of the story*

SMC can complement (not replace!) SDC in multi-source scenarios

*Towards a system-level view of "**confidentiality engineering**"*

- *learn to compose multiple elements/layers/components in a consistent design (technology, legal, organizational)*
- *centrality of feasible attack models, analysis & minimization of risks*

Thanks for your attention

For follow-up :

fabio.ricciato@ec.europa.eu

aleksandra.bujnowska@ec.europa.eu