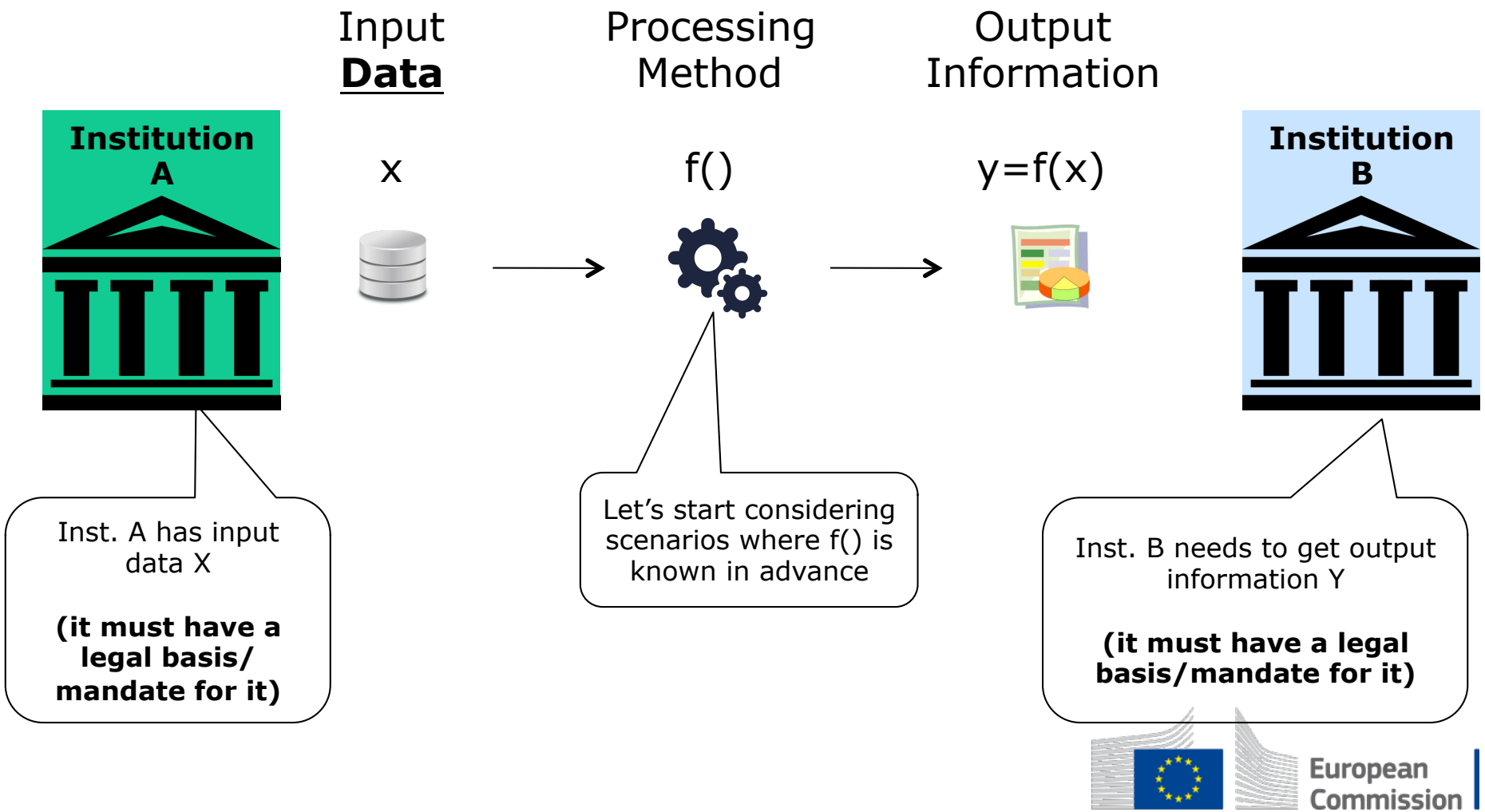
What role for privacy-preserving computation for inter-institutional data sharing?

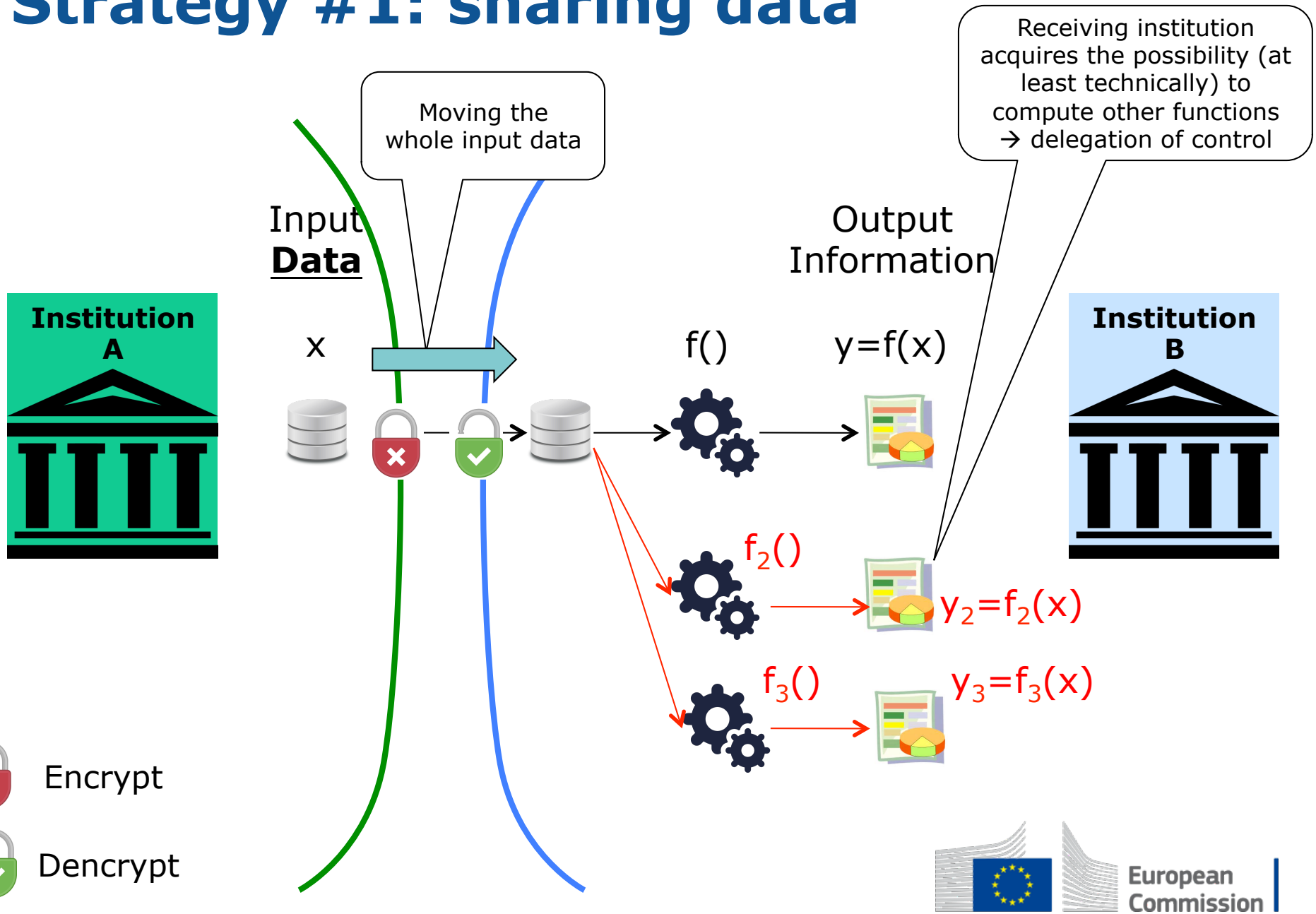
Fabio Ricciato
Unit A5 – Methodology; Innovation in Official Statistics
Eurostat - European Commission

G20 DGI-2 Workshop on Recommendation II.20 “Promotion of Data Sharing”
24-25 March 2021

How to deliver information in B from data in A?

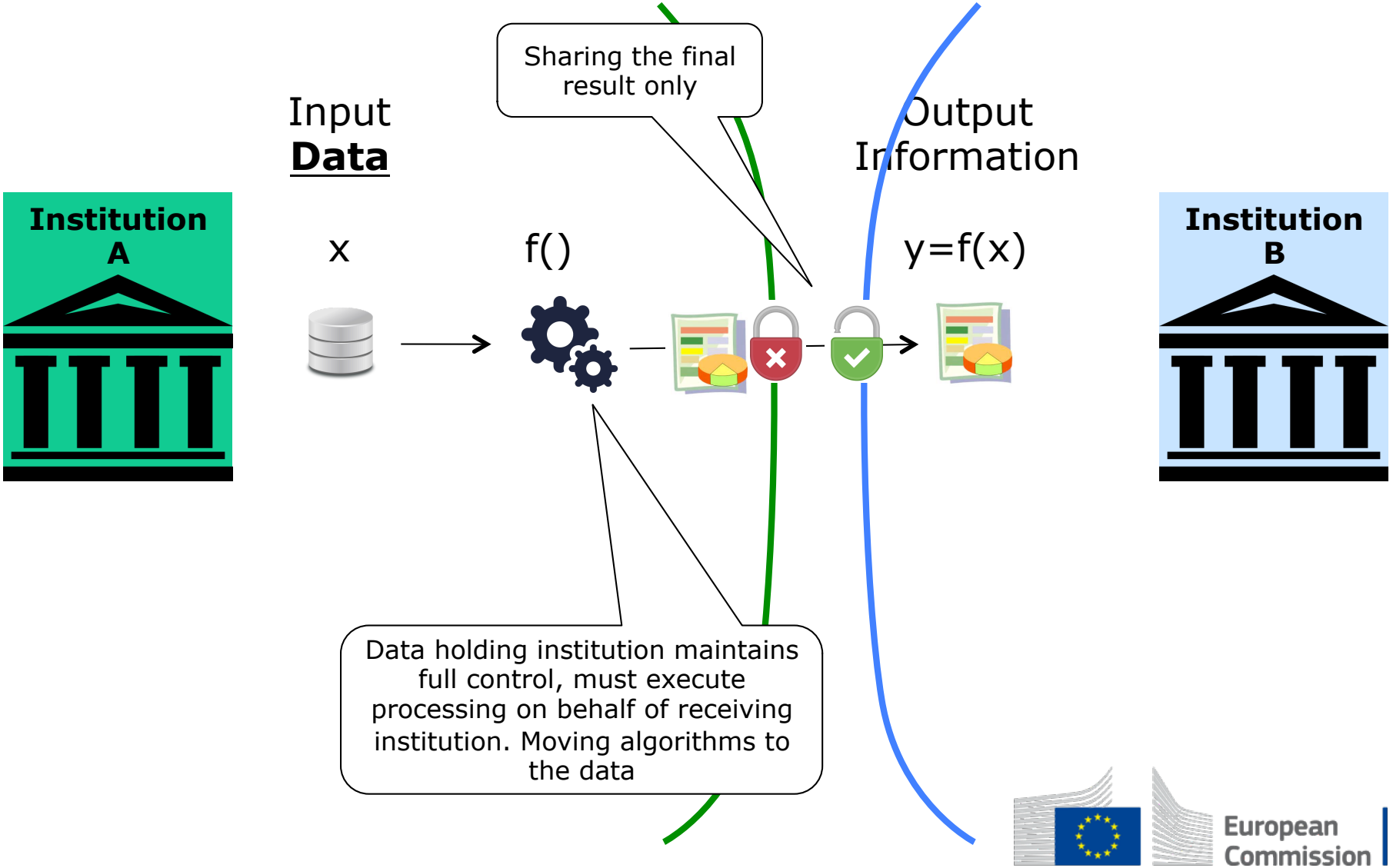


Strategy #1: sharing data



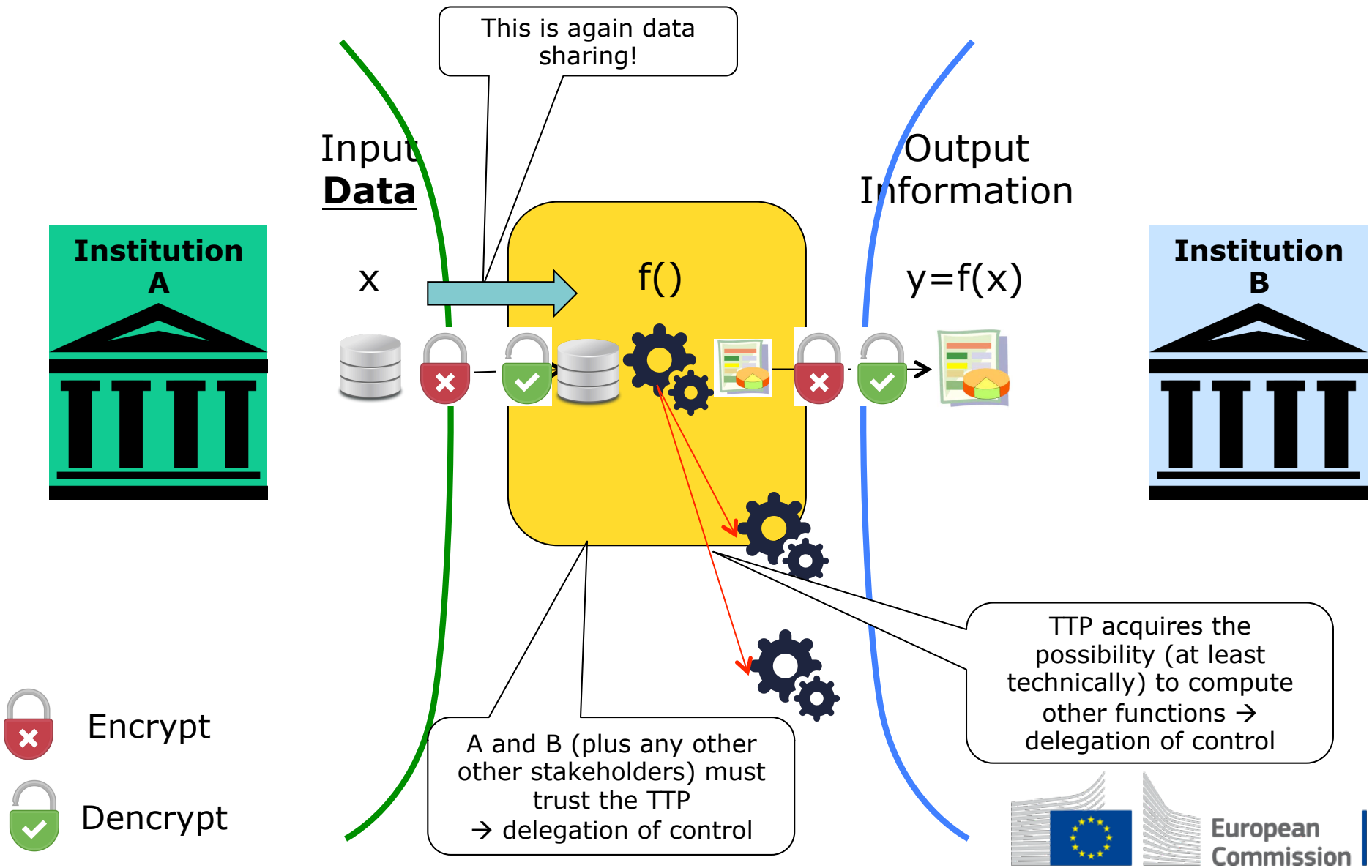
European Commission

Strategy #2: compute locally at source (only for single source)

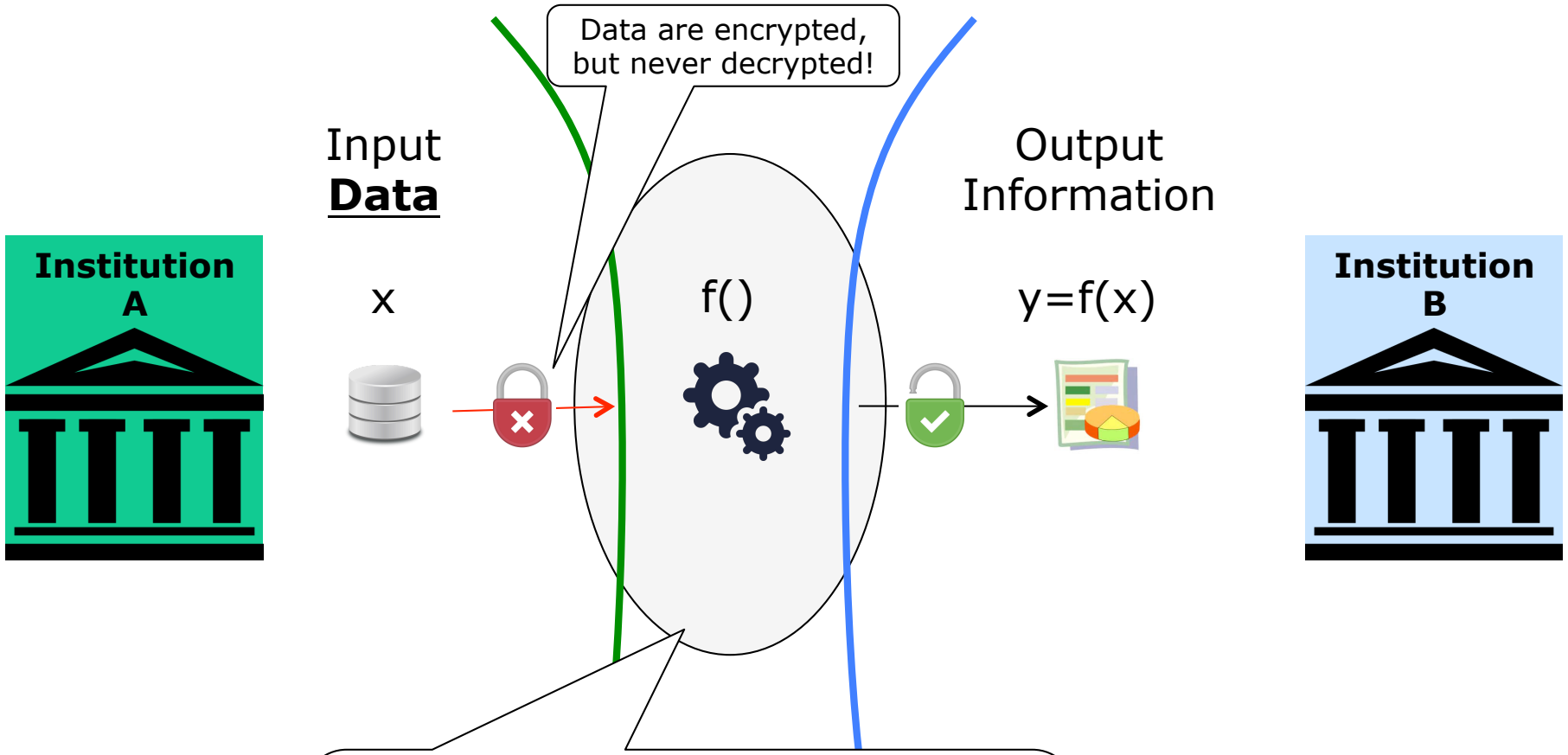


European Commission

Strategy #3: Trusted Third Party (TTP)

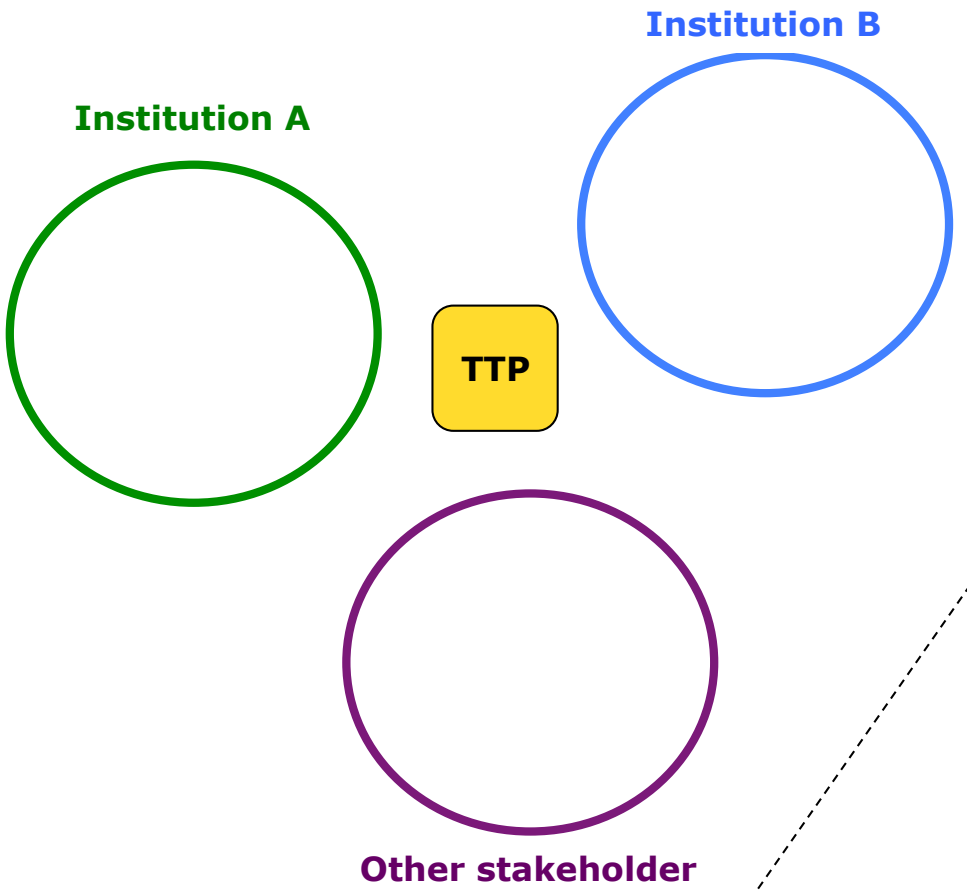


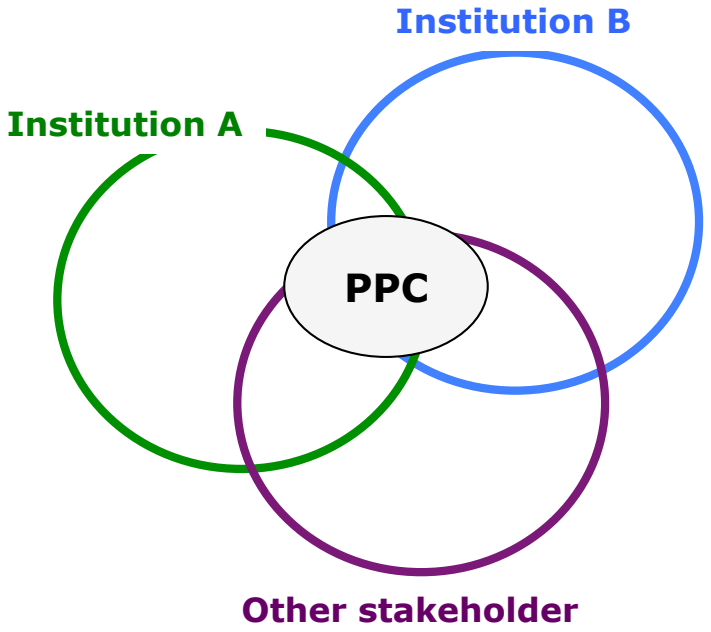
Strategy #4: privacy-preserving computation (PPC) infrastructure



European Commission

TTP vs PPC:
delegating control vs. sharing control





See *Trusted smart statistics: Motivations and principles*
https://ec.europa.eu/eurostat/cros/system/files/sji190584.pdf

European Commission

Privacy-preserving computation technologies

- *Privacy-preserving computation technologies*
  - Secure Multi-Party Computing, Secret Sharing (software)
  - Homomorphic Encryption (software)
  - Trusted Execution Environment (hardware)
  - Different combinations of the above …
  - … possibly integrated with distributed ledger technologies

- *PPC technologies have been maturing quickly in the last decade, now ready for deployment*

- *PPC infrastructure := combination of technological and organisational measures*

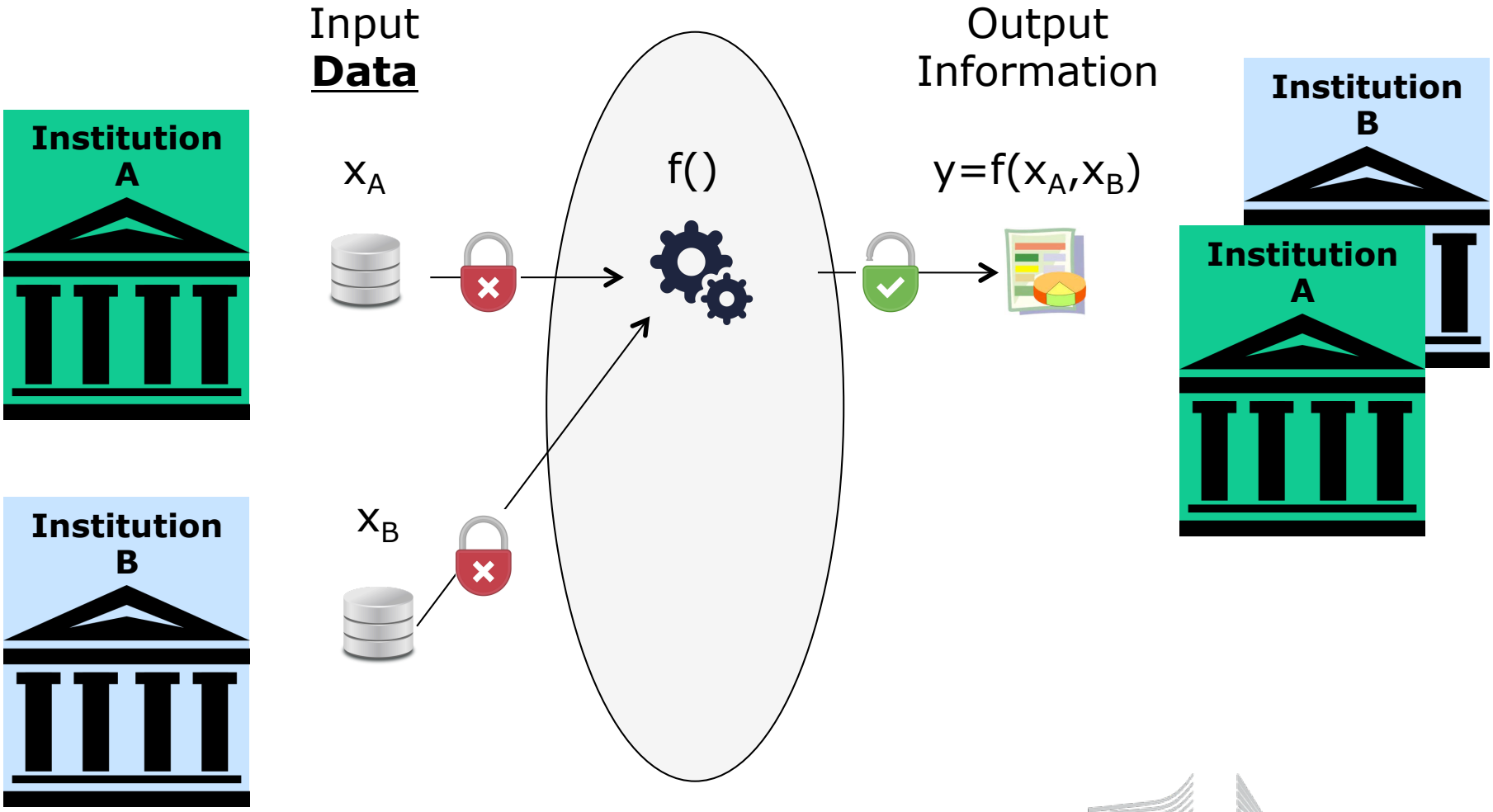
- *Privacy-preserving computation-as-a-service?*

Which strategy to prefer?

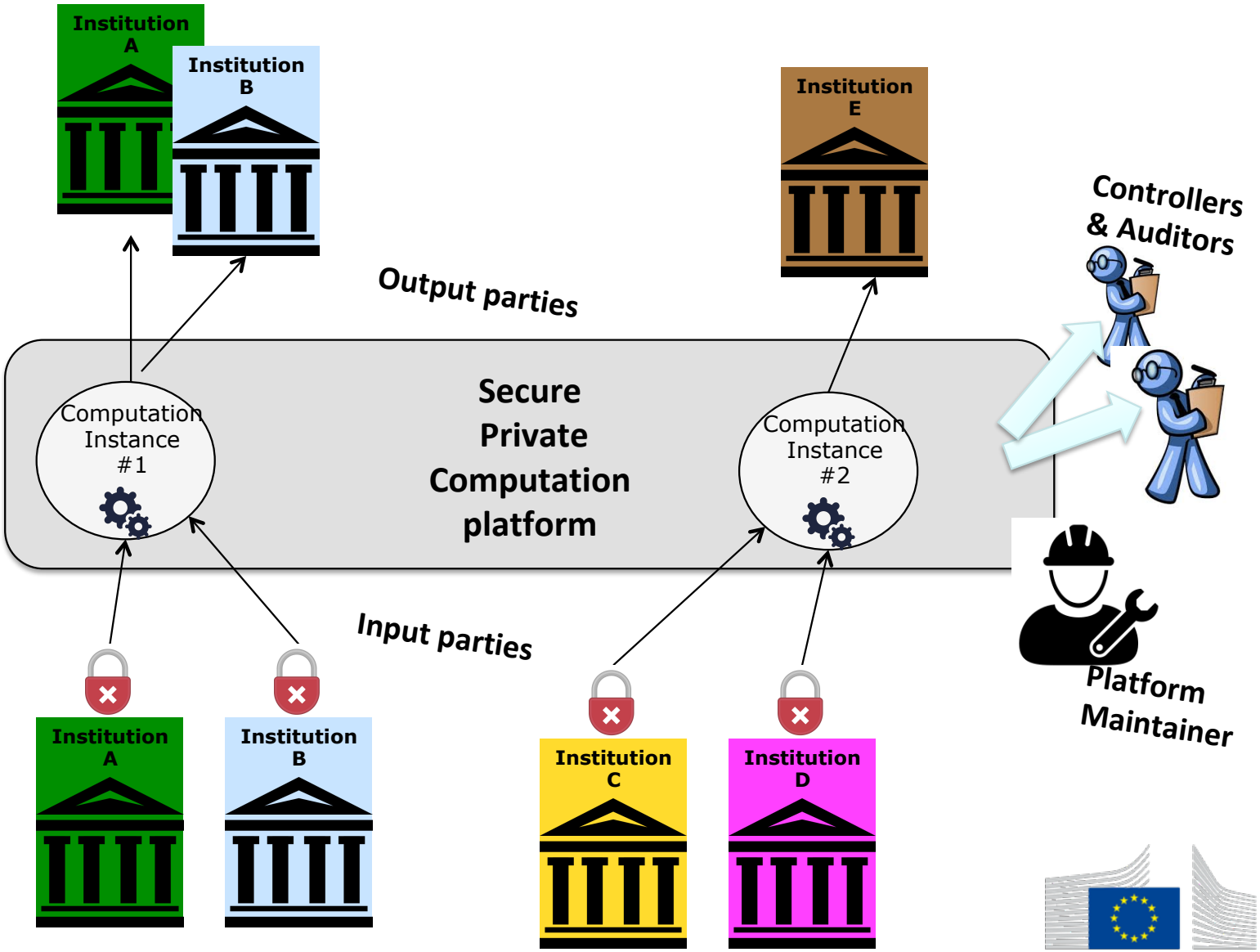
- *Costs vs benefits*
  - All four strategies have different benefits and costs, strengths and limitations, … each strategy entails a different *trust model*
  - Preferred strategy (legally, technically) depends on scenario

- *Advantages of PPC*
  - Flexible configurations of ex-ante and ex-post controls for different stakeholders
  - Allows each participating institution to **stay in control** of each computation instance (shared, non-exclusive control)
  - Extends naturally to multiple input parties (next slide)
- *Potential limitations*
  - Computational scalability (depending on technology)
  - Interactive analysis / data exploration may be difficult

PPC with two multiple input parties



PPC-as-a-service?



European Commission

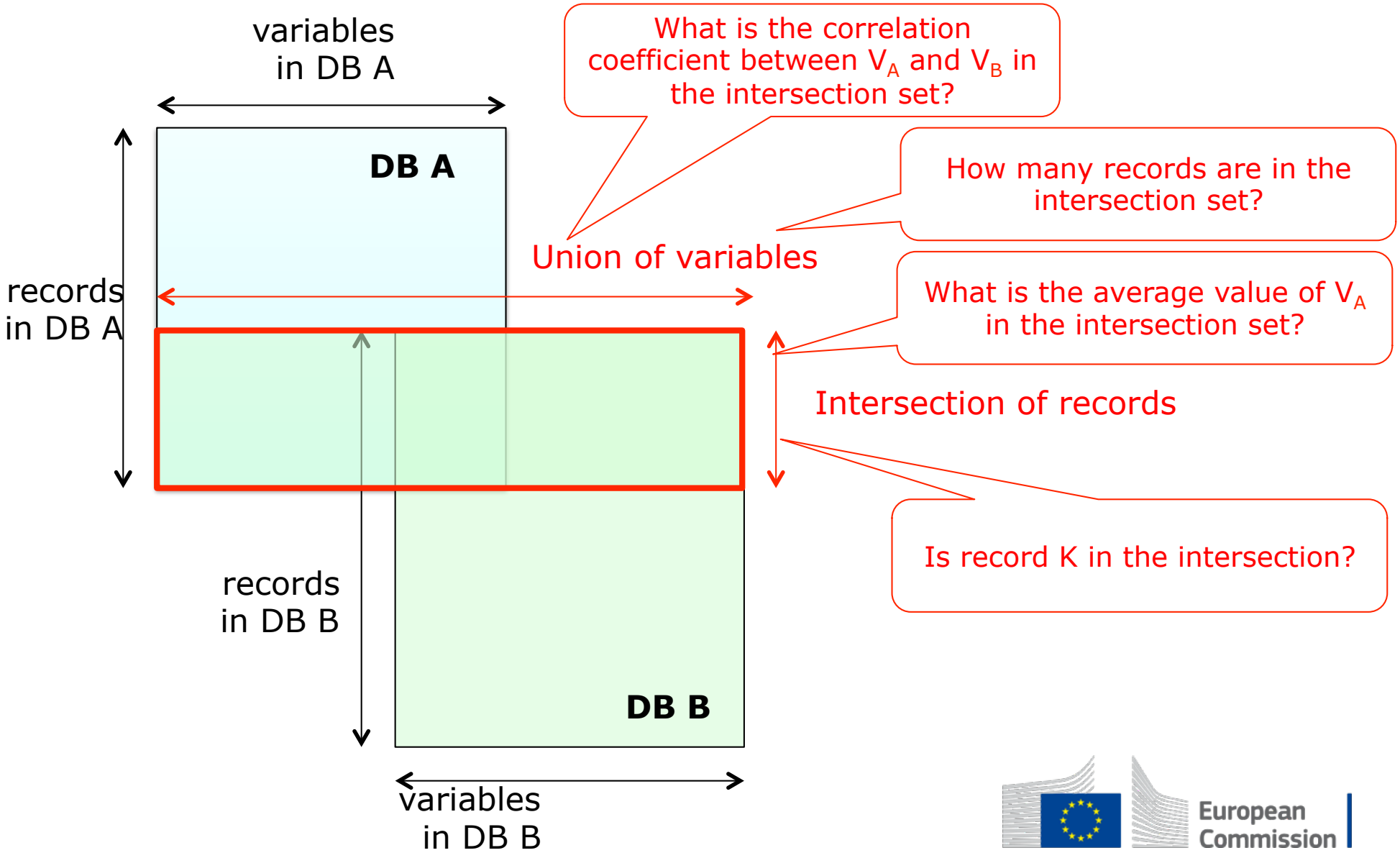
Legal aspects

- *Receiving institutions need anyway a legal basis/mandate to acquire final information y*
  - may be less critical than legal basis/mandate to acquire full input data x

- *PPC technologies enable a paradigm shift*
  - Let only the desired (output) information y flow, not all (input) data x
  - From "**sharing data**" to "**sharing control**" on computation
  - Data processing gets strictly bound to specific method f()
    → the object of discourse shifts from "*access to data x*"
    to "***processing of data x with method f()***"

- *PPC and GDPR*
  - Enable tight application of GDPR principles "**data minimization**", "**purpose limitation**", "storage limitation", "integrity and confidentiality"
  - Related open issue: encryption/secret sharing qualify as anonymization or pseudonymization ???

Examples of queries for microdata



European Commission

For follow-up

- *Trusted Smart Statistics: Motivations and Principles*
  https://ec.europa.eu/eurostat/cros/system/files/sji190584.pdf

- *Trusted Smart Statistics: How new data will change official statistics*
  https://doi.org/10.1017/dap.2020.7

- *Trusted Smart Surveys: a possible application of Privacy Enhancing Technologies in Official Statistics*
  https://ec.europa.eu/eurostat/cros/system/files/main_ricciato_sis2020_v4c_noita.pdf

- *Towards a reference architecture for Trusted Smart Surveys*
  https://ec.europa.eu/eurostat/cros/system/files/tssurveys_ipp_ricciato_v4.pdf

European Commission

Thanks for your attention

Fabio.Ricciato@ec.europa.eu