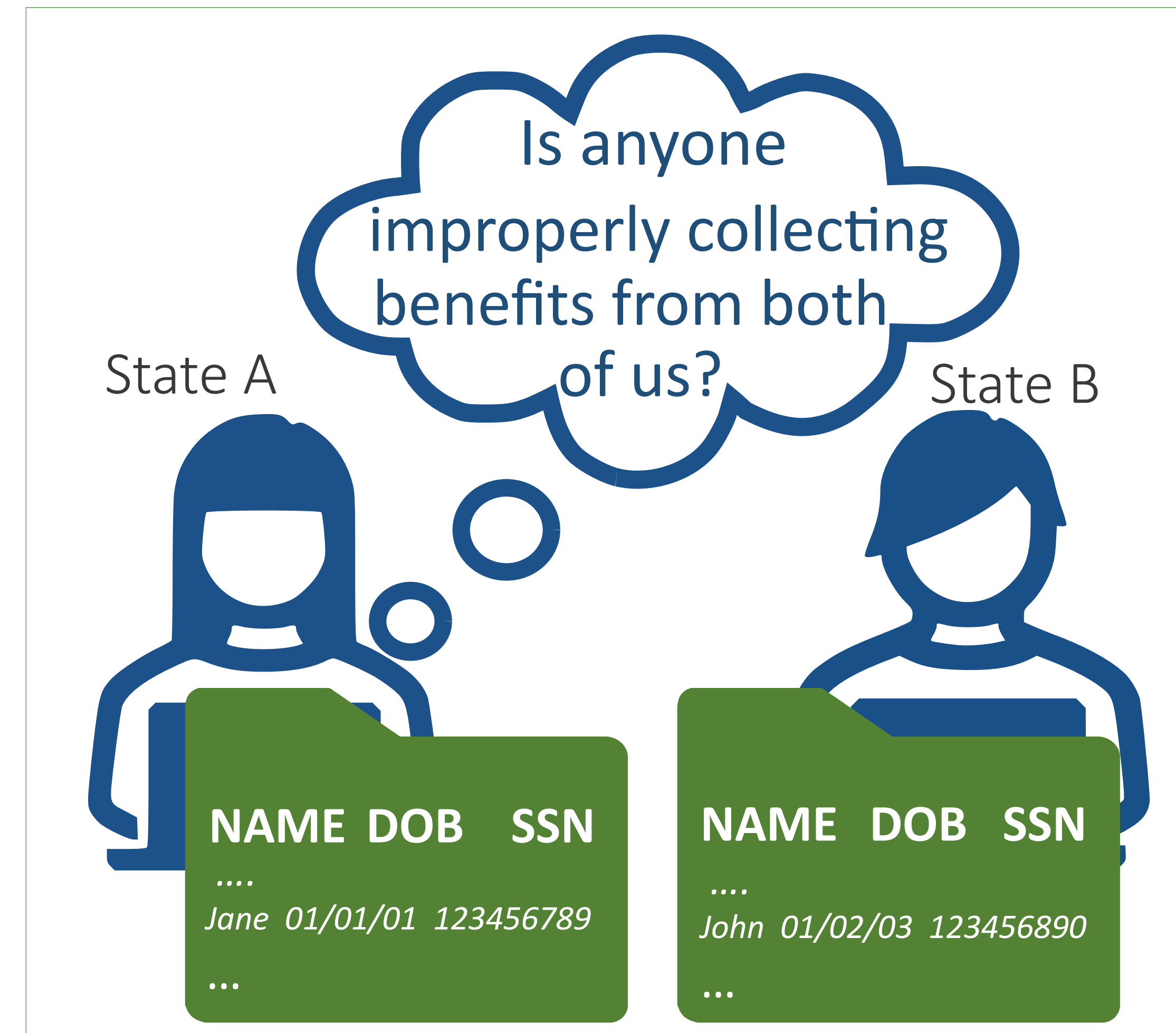


## Utility of Data Intersection

Determining the overlap of sensitive information between sets of private data is widely beneficial:

- discovery of common contacts
- customers using leaked secret passwords
- persons improperly claiming benefits in two states

One approach involves granting dataset access to a trusted third party. Privacy-enhancing crypto offers a better solution; private data is not shared.



## Private Set Intersection

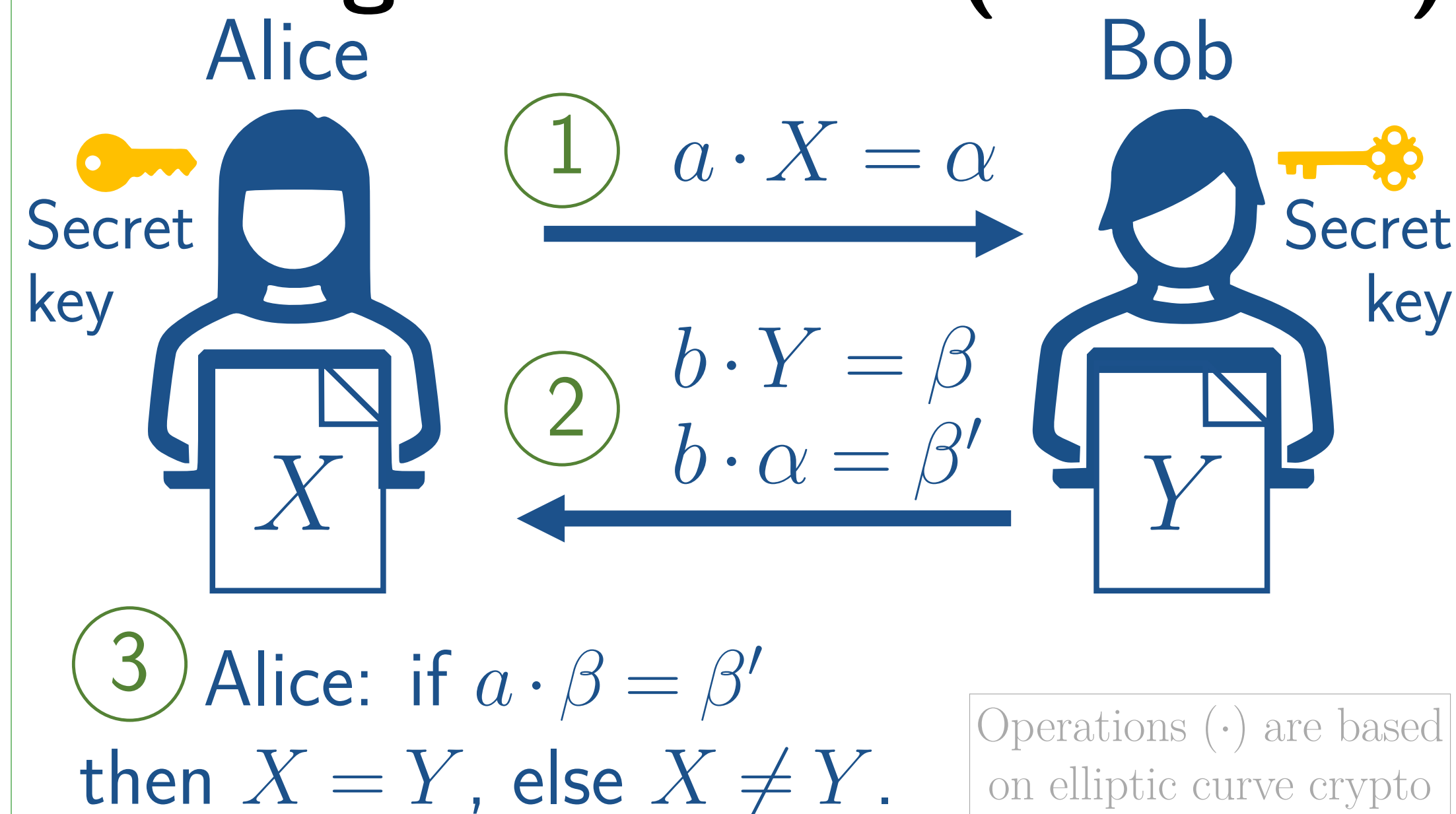
With **Private Set Intersection (PSI)**, two parties compute the intersection of their sets, without revealing or disclosing the non-intersecting elements.

If Alice has the set  $\{p, r, i, v, a, t, e\}$  and Bob has  $\{s, e, c, r, t\}$ , then they get the set  $\{r, t, e\}$ . Instead of letters, each party may have PII.

PSI on numerical data may be used for useful computations on intersections: averages, sums, etc.

PII = private identifiable information (e.g., names DOB, SSN)

### Toy example: One-sided PSI for single elements ( $X$ and $Y$ )



## PSI Properties and Variants

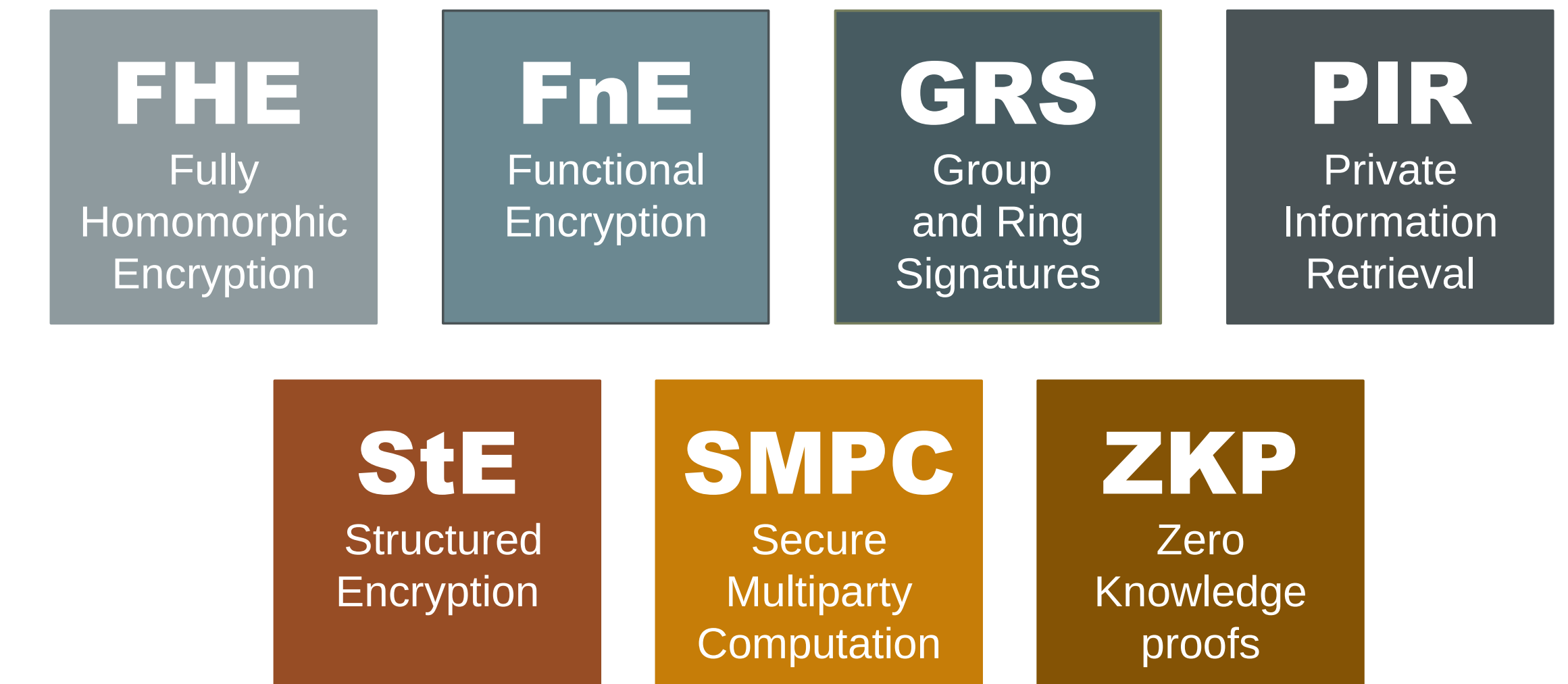
**Caution:** Exchanging hashes is not good! Anyone can compute the hashes of all possible SSNs and birthdates.

**Good solution:** Reveals nothing, even predictable elements.

**Variants:** Real scenarios may involve more parties, larger datasets, input consistency checks, statistics of the intersection.

**NIST role:** Future guidance may facilitate secure deployments.

### Other PEC tools



Can enhance privacy in many applications