AnoMed

Finanziert von der Europäischen Union
NextGenerationEU

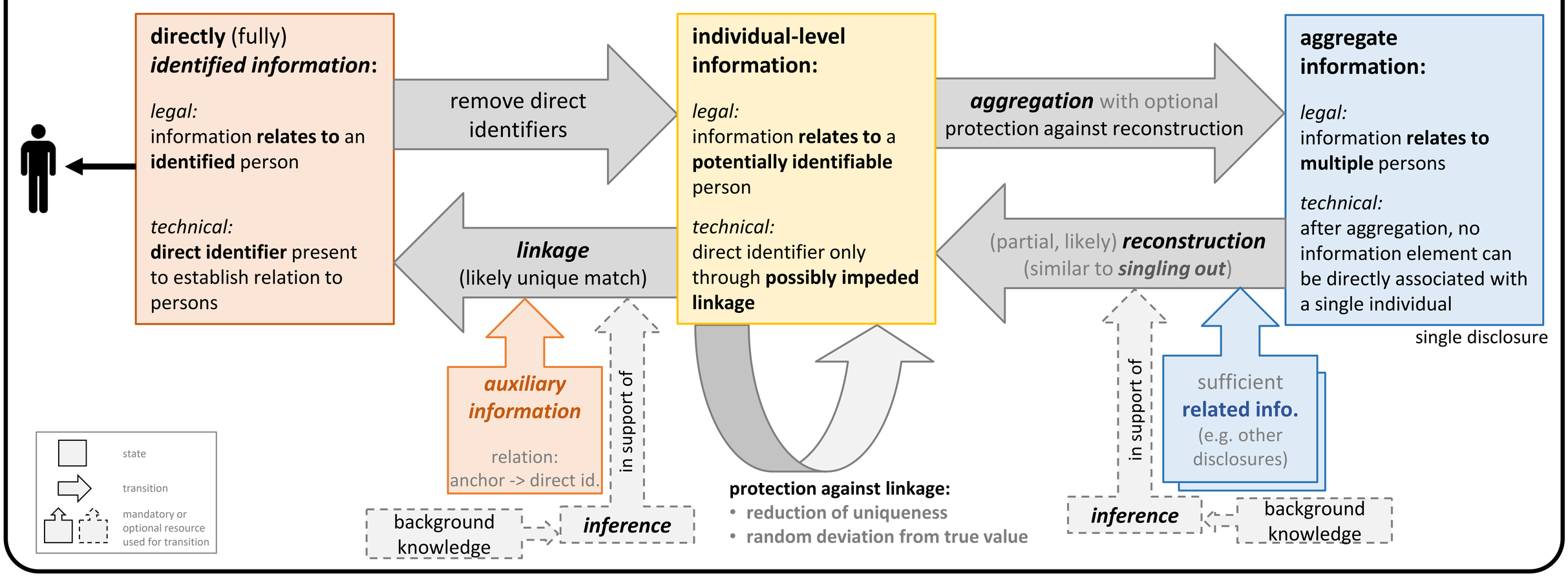
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Analysis of Data Protection Risks

Bud P. Bruegger, Moritz Kirschte, Niklas Zapatka, Harald Zwingelberg, Hannes Federrath, Esfandiar Mohammadi, Sebastian Meiser

Terminology with Relations between Concepts: State-Transition-Diagram

I



Taxonomy of Possible Claims that Data are Anonymous

II

2: The data is aggregated, and thus direct or indirect identification are only possible after successful (possibly partial) reconstruction	**2.2:** *Fact:* Data **with mathematically guaranteed** reconstruction protection	2.2.2: *Claim*: The privacy budget is managed for both, **own and external disclosures**	
		2.2.1: *Fact*: The privacy budget is managed **for own disclosures,** ..	2.2.1.2: *Claim*: ..**no** significant number of **external disclosures exists**
			2.2.1.1: Claim: ..**no** significant number of **external disclosures** are **accessible** to attackers
	2.1: *Fact:* Data **without mathematically guaranteed** reconstruction protection	2.1.3: *Claim*: Reconstruction is assumed to be impossible based on current **state of the art**, ..	2.1.3.2: *Claim*: Known attacks fail as **verified with own data**
			2.1.3.1: *Claim*: Known attacks fail **based on assumptions** about state of the art
		2.1.2: *Claim*: Reconstruction is assumed to be impossible based on **assumptions about additional disclosures**, ..	2.1.2.2: *Claim*: ..Significant additional disclosures **don't exist**
			2.1.2.1: *Claim*: **..addl. disclosures exist but are not accessible** by potential attackers
		2.1.1: *Claim*: Reconstruction is assumed to be impossible based on **assumptions about potential attackers**:	2.1.1.2: *Claim:* Attackers **lack capability** (skill, resources, ..)
			2.1.1.1: *Claim:* Attackers **lack motivation** (cost/benefit)
1: The **data does not contain direct identifiers** and has been **protected against linkage**	(i) *Facts:* protection of **quasi-identifiers**	1.3: *Claim:* Linkage impossible since **data** provides **no unambiguous link anchors** (any unique combination of attributes; arbitrary auxiliary information) !inference!	1.3.2: *Claim:* **No unique records** contained **in data** (all attributes treated as quasi-identifiers: classes of equal values or clusters of close values) !inference!
			1.3.1: *Claim:* **Modification of** potential **anchors** renders **matches uncertain** and **deniable** (noise, swapping, ..)
		1.2: *Claim:* Linkage impossible based on **assumptions about suitable auxiliary information** !inference!	1.2.2: *Claim:* Suitable auxiliary information **does not exist**
			1.2.1: *Claim:* Suitable auxiliary information exists but is **not accessible to potential attackers**
	(ii) *Facts:* protection of **other attributes:**	1.1: *Claim:* Linkage not possible based on **assumptions about potential attackers**	1.1.2: *Claim:* Attackers **lack capability** (resources, skill)
			1.1.1: *Claim:* Attackers **lack motivation** (cost benefit)

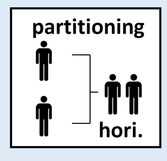
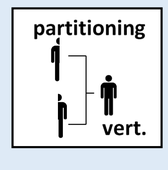
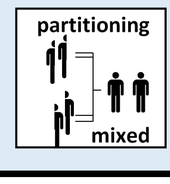
Use Cases for EHDS: required level of identification

III

Relation to GDPR	GDPR terminology	*Type of Data*	Necessary Data Transformation	Effort/Difficulty	*Identification*
Personal data (inside GDPR)		**Directly identified**	none	none	100%
	Pseudonymous data	**Reversibly pseudonymous**	**Separate** & protect direct identifiers	Medium (TOMs to control identification)	Certain but controlled
		Irreversibly pseudonymous	**Remove** direct identifiers	Minimal (fewer TOMs)	Can happen unintentionally
		Aggregated pseudonymous	**Aggregate values** over several persons	Minimal	Requires intentional re-identification
Outside of GDPR	Anonymous data	**Truly anonymous**	Successfully **anonymize**	**Substantial**	Very unlikely
	--	**No data**	**Delete**	Minimal	Impossible

Use Cases for EHDS: partitioning of data

IV

type of partitioning	definition	example	visualization
none	data comes from a **single source**	A single source provides a data set of person's height. The analysis computes and average height.	
horizontal	**Multiple data sources** provide the **same attributes** about **different persons**	A multitude of sources provide data sets of height of persons living in their geographic area of operation. The analysis computes the overall average height over the combined geographic area.	
vertical	**Multiple data sources** provide **different attributes** about the **same persons**	To compute the average Body Mass Index (BMI), two data sources must be combined: one providing height data of a given population; the other providing weight data of the same population.	
mixed	**Multiple data sources** provide **different attributes** about **different persons**	An analysis requires to logically combine data sources both, vertically and horizontally.	

UNIVERSITÄT ZU LÜBECK

ULD

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

IM FOCUS DAS LEBEN