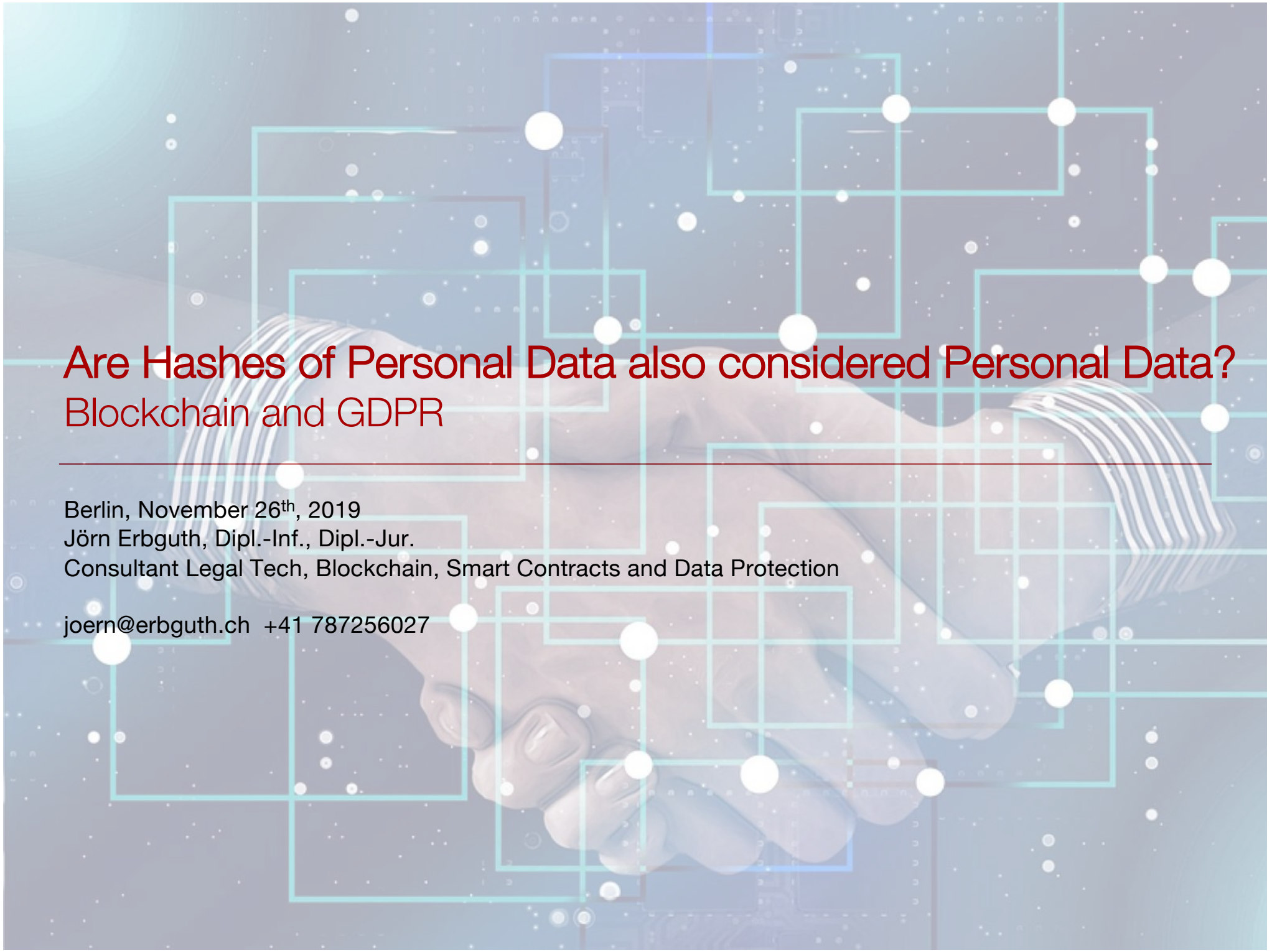


Are Hashes of Personal Data also considered Personal Data?

Blockchain and GDPR

Berlin, November 26th, 2019

Jörn Erbguth, Dipl.-Inf., Dipl.-Jur.
Consultant Legal Tech, Blockchain, Smart Contracts and Data Protection

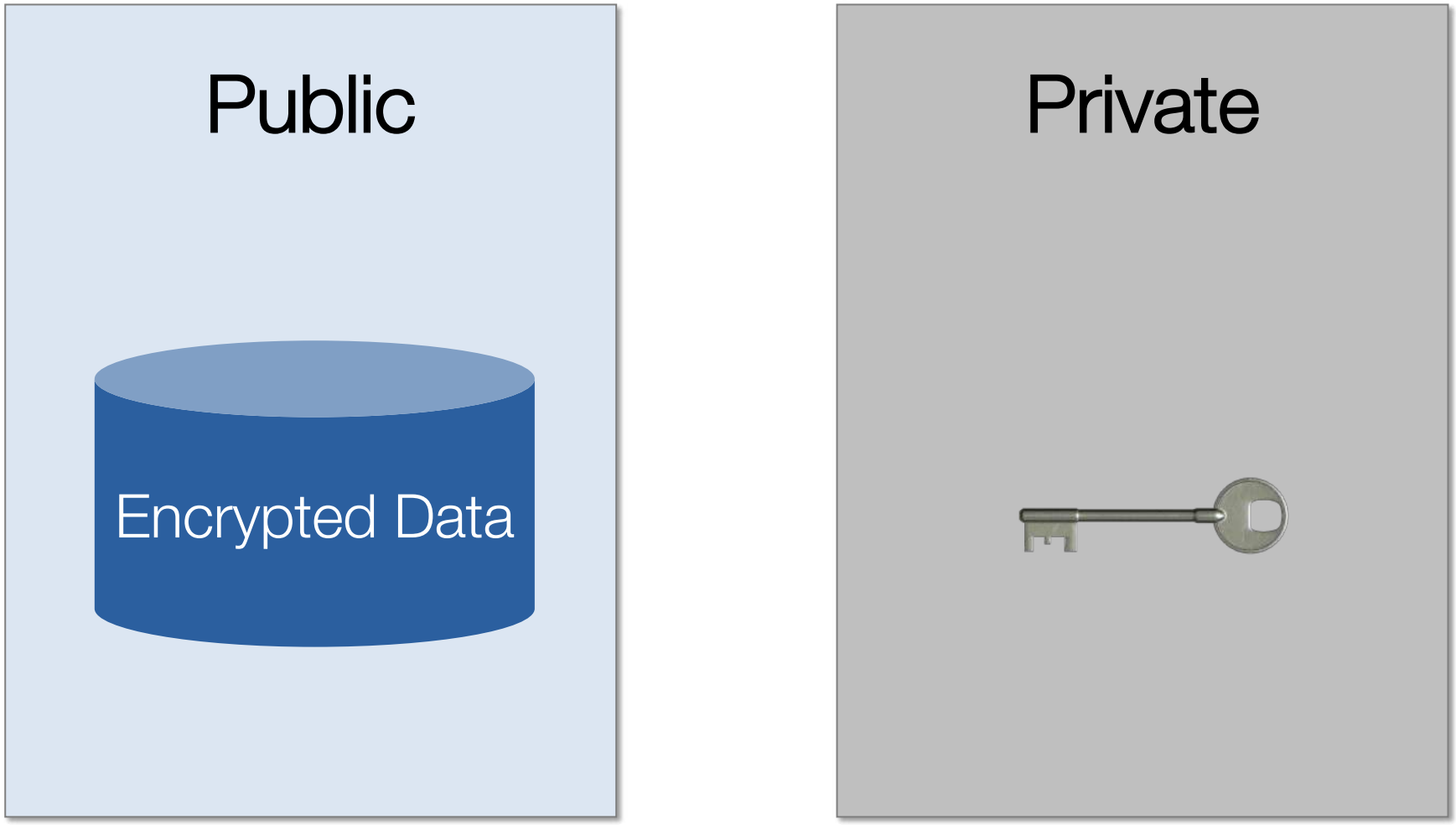
joern@erbguth.ch +41 787256027

Cryptographic hash functions

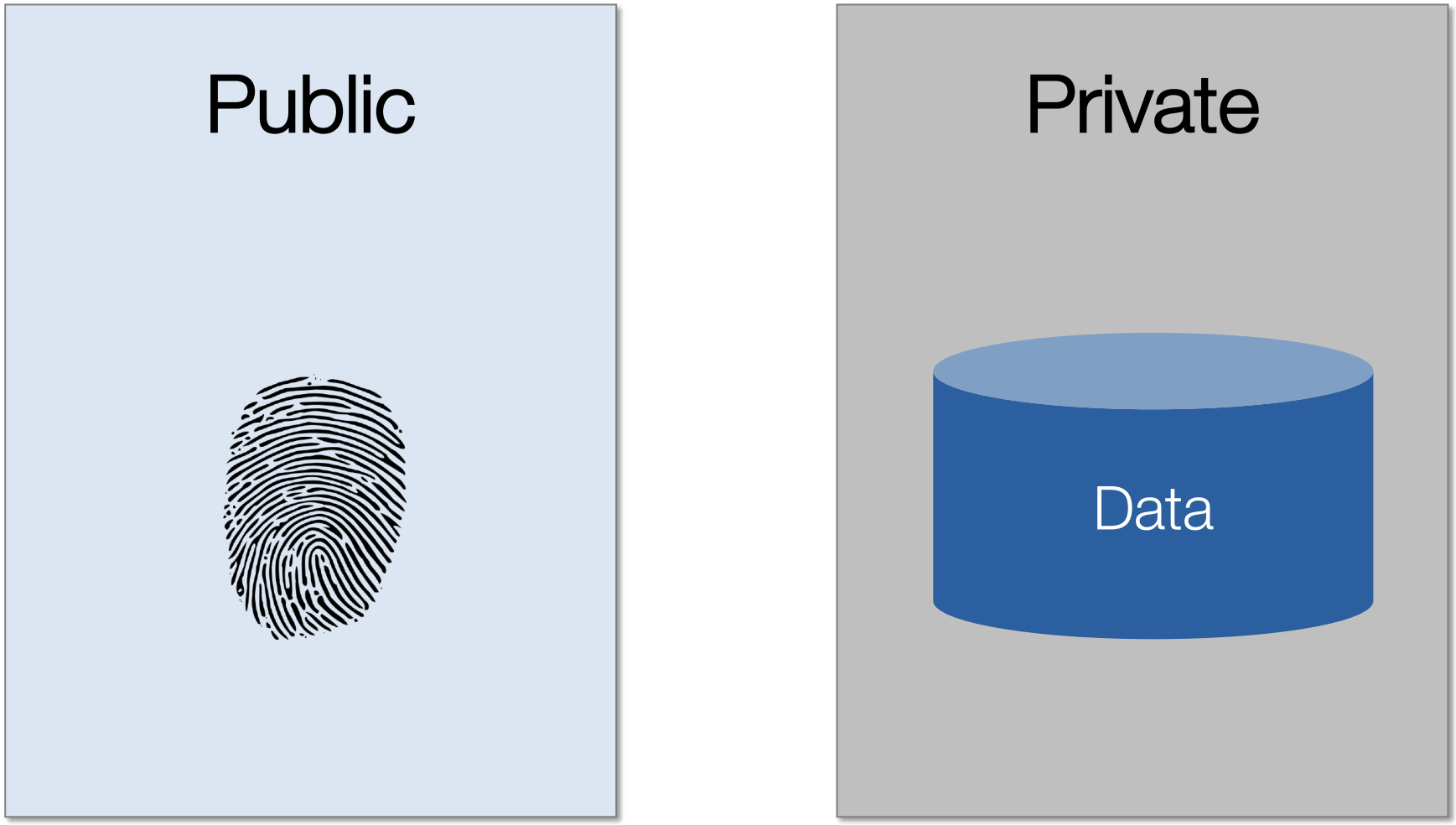
- Serve as digital fingerprints
- Virtually unique
- Fixed length (e.g. 32 bytes)
- For digital objects of any size

Demo 2

Use of Hash Values



Use of Hash Values

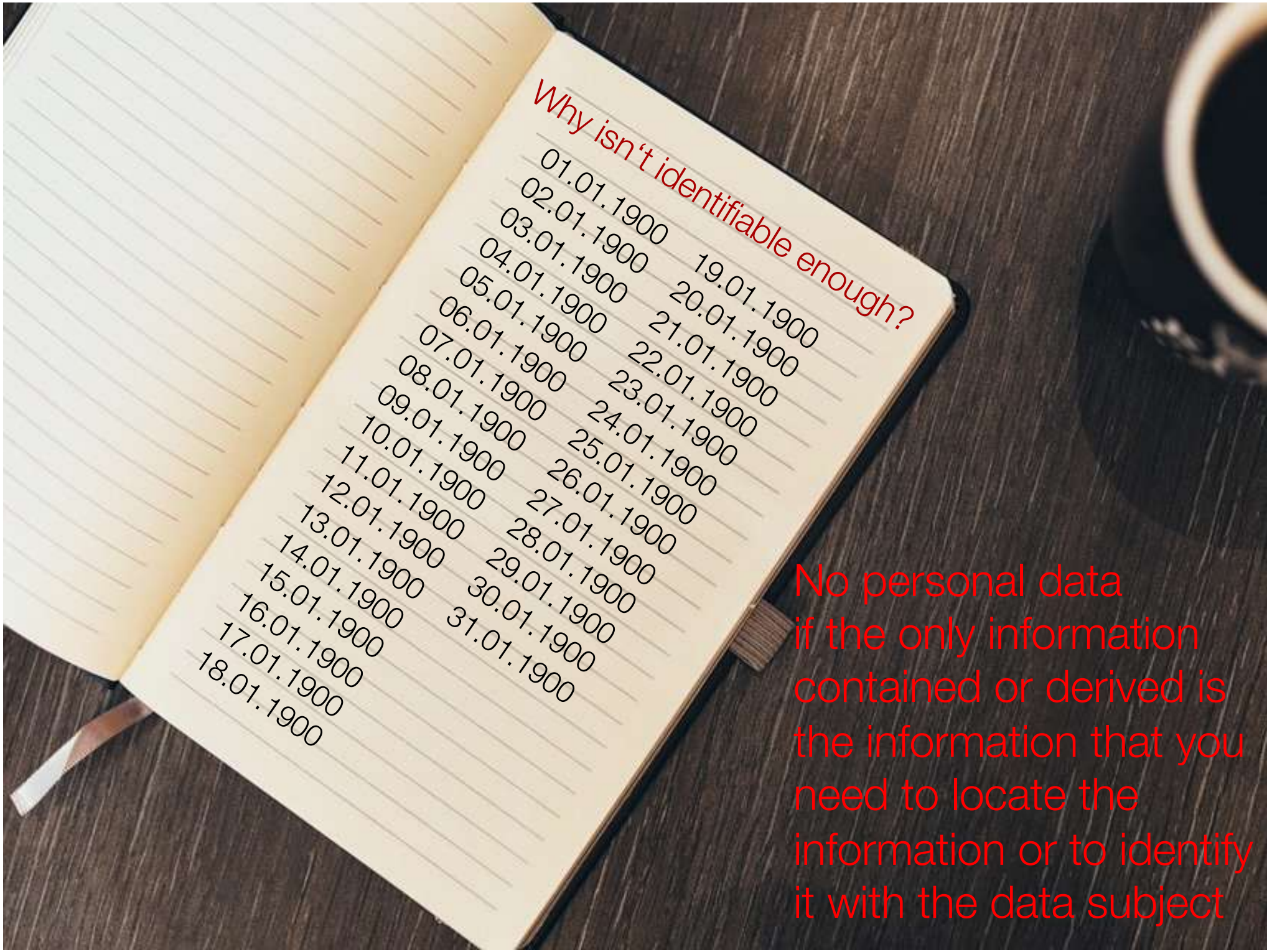


Personal Data

Art. 4 GDPR
Definitions

For the purposes of this Regulation:

(1) 'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;


Why isn‘t identifiable enough?
01.01.1900
02.01.1900
03.01.1900
04.01.1900
05.01.1900
06.01.1900
07.01.1900
08.01.1900
09.01.1900
10.01.1900
11.01.1900
12.01.1900
13.01.1900
14.01.1900
15.01.1900
16.01.1900
17.01.1900
18.01.1900
19.01.1900
20.01.1900
21.01.1900
22.01.1900
23.01.1900
24.01.1900
25.01.1900
26.01.1900
27.01.1900
28.01.1900
29.01.1900
30.01.1900
31.01.1900
No personal data
if the only information
contained or derived is
the information that you
need to locate the
information or to identify
it with the data subject

When is a person identifiable?

Recital 26

[3]To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. [4]To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.

- Only if not „practically impossible on account of the fact that it requires a disproportionate effort in terms of time, cost and man-power“ (ECJ, Breyer)
- But under consideration of future developments like Quantum Computing

Are hash-values of personal data personal data?

WP29 (05/2014, WP216)

Each technique described in this paper fails to meet with certainty the criteria of effective anonymisation (i.e. no singling out of an individual; no linkability between records relating to an individual; and no inference concerning an individual). However as some of these risks may be met in whole or in part by a given technique, careful engineering is necessary in devising the application of an individual technique to the specific situation and in applying a combination of those techniques as a way to enhance the robustness of the outcome.

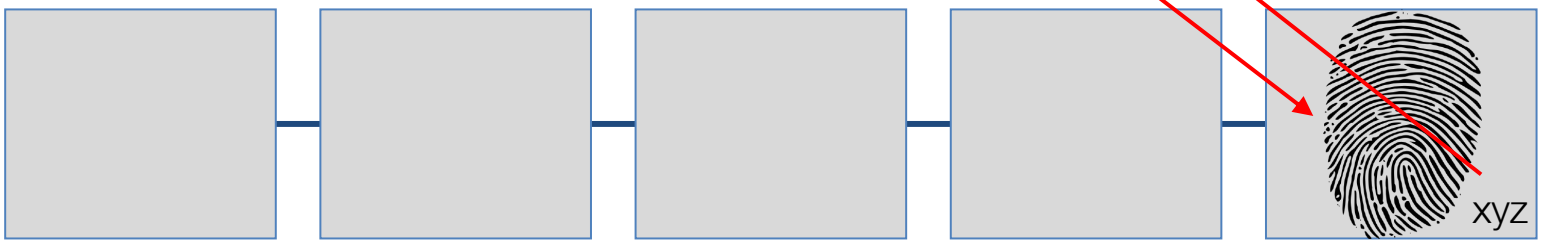
4. Pseudonymisation

Pseudonymisation consists of replacing one attribute (typically a unique attribute) in a record by another. The natural person is therefore still likely to be identified indirectly; accordingly, pseudonymisation when used alone will not result in an anonymous dataset. Nevertheless, it is discussed in this opinion because of the many misconceptions and mistakes surrounding its use.

Risks to watch out when using hash-functions

Singling out?

xyz



Risks to watch out when using hash-functions

Linkability?


Universität Passau
Juristische Fakultät

Diplomurkunde

Die Juristische Fakultät der Universität Passau verleiht mit dieser Urkunde

Herrn **Jörn Horst Erbguth**

geboren am 5. Dezember 1967 in Marl

den akademischen Grad

Diplom-Jurist Univ.

(Dipl.-Jur. Univ.)

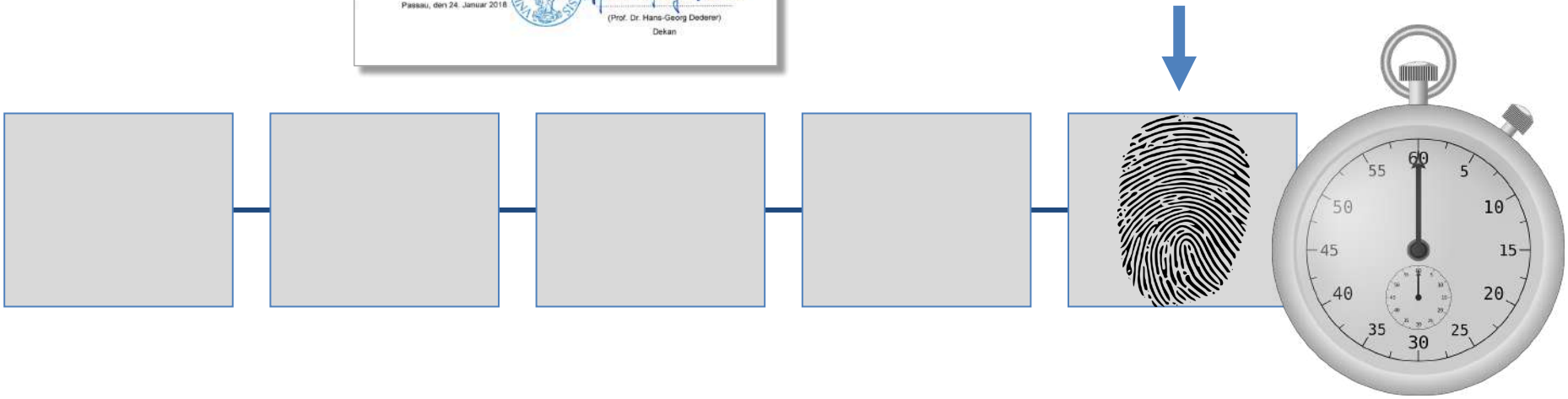
aufgrund der gemäß Ausbildungs- und Prüfungsordnung für Juristen (JAPO) in der jeweils gültigen Fassung am 23. Juli 1992 bestandenen Ersten Juristischen Prüfung.

Passau, den 24. Januar 2018

(Prof. Dr. Hans-Georg Dederer)
Dekan

Risks to watch out when using hash-functions

Inference?



Risks to watch out when using hash-functions

Reversibility?

Universität Passau
Juristische Fakultät

Diplomurkunde

Die Juristische Fakultät der Universität Passau verleiht mit dieser Urkunde

Herrn **Jörn Horst Erbguth**

geboren am 5. Dezember 1967 in Marl

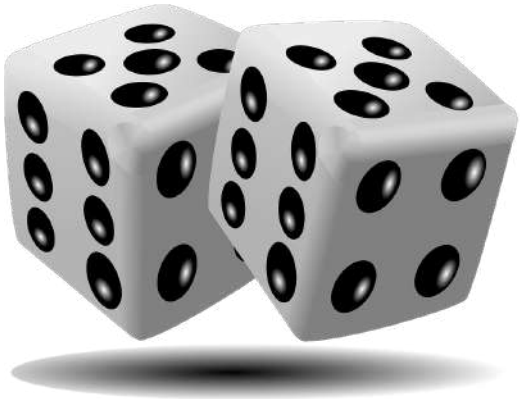
den akademischen Grad

Diplom-Jurist Univ.
(Dipl.-Jur. Univ.)

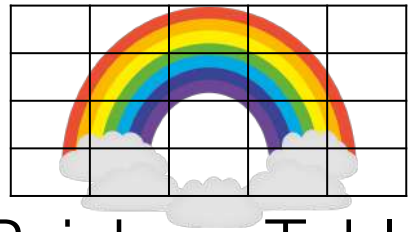
aufgrund der gemäß Ausbildungs- und Prüfungsordnung für Juristen (JAPO) in der jeweils gültigen Fassung am 23. Juli 1992 bestandenen Ersten Juristischen Prüfung.

Passau, den 24. Januar 2018

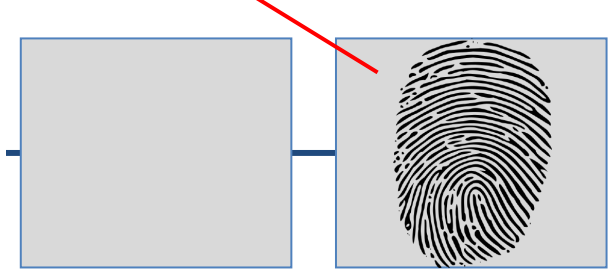
(Prof. Dr. Hans-Georg Dederer)
Dekan



Low Entropy



Rainbow Table

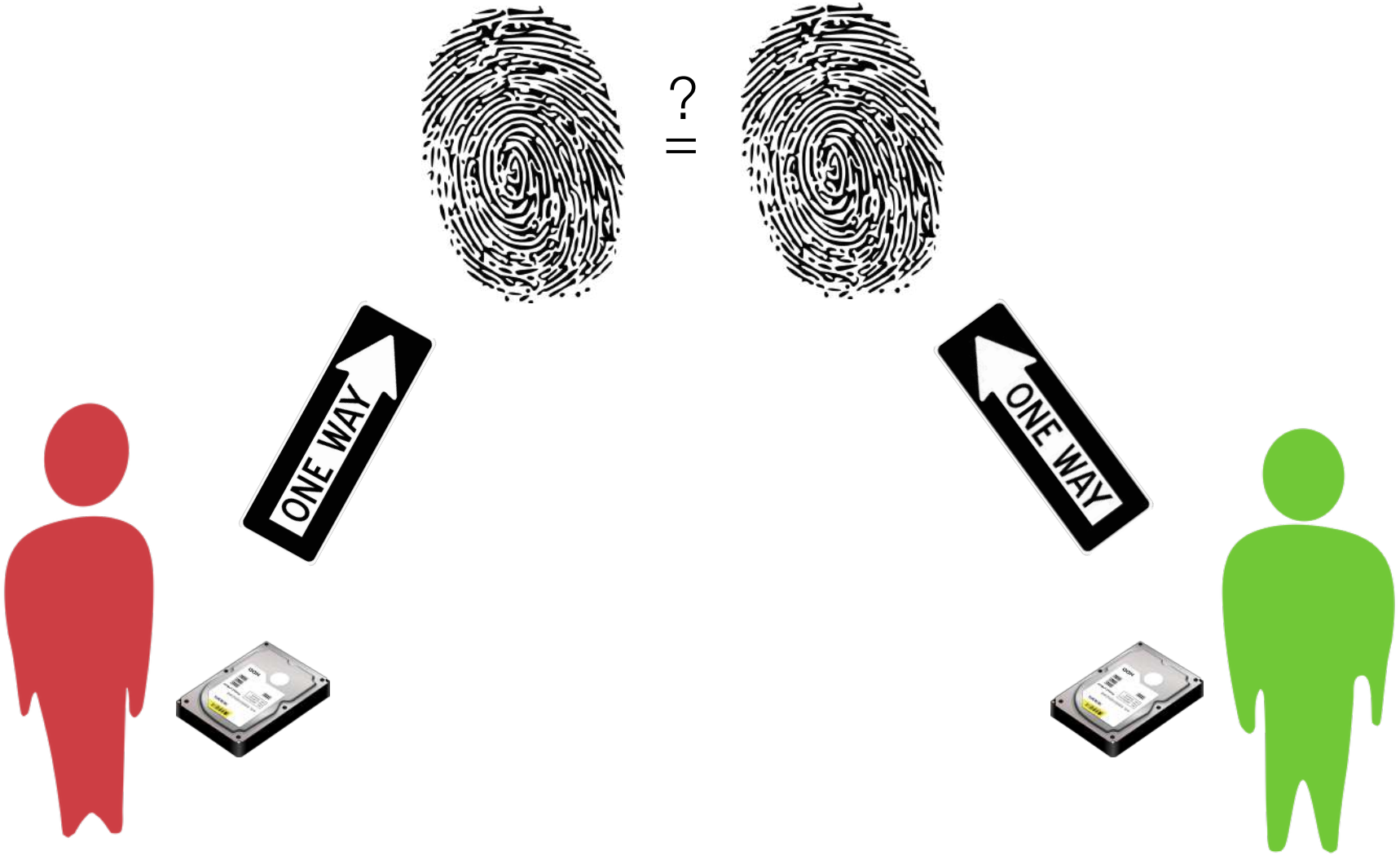


agencia española protección datos

EDPS

European Data Protection Supervisor

Compare data through hash-values



Facebook Custom Audience

facebook

customers

Display adds only to customers that are Facebook-users

users

john.smith@gmail.com

john.smith@gmail.com

Comparison of hash-values only

How to Hash Data

Data

First Name	Last Name	Article	Quantity	Price
John	Smith	1984 by George Orwell	1	10
Lisa	Doe	Ulysses by James Joyce	1	20
John	Smith	Inside Wikileaks by Domscheit-Berg	1	15

Wrong solution

Off-chain

First Name	Last Name	Salt		Hash
John	Smith	87683746776923452362	→	87627648267459265308697
Lisa	Doe	987936034857436363 65	→	98796983579348569273643

On-chain

Hash	Article	Quantity	Price
87627648267459265308697	1984 by George Orwell	1	10
98796983579348569273643	Ulysses by James Joyce	1	20
87627648267459265308697	Inside Wikileaks by Domscheit-Berg	1	15

How to Hash Data

Data

First Name	Last Name	Article	Quantity	Price
John	Smith	1984 by George Orwell	1	10
Lisa	Doe	Ulysses by James Joyce	1	20

Still problematic solution

Off-chain

First Name	Last Name	Article	Quantity	Salt	Hash
John	Smith	1984 by George Orwell	1	87683746776923452362	→ 7648265467265308697 4532
Lisa	Doe	Ulysses by James Joyce	1	98793603485743636365	→ 35793485692736433524132
John	Smith	Inside Wikileaks by Domscheit-Berg	1	29749850385739857395	→ 86786876868594939653656

On-chain

Hash	Price
7648265467265308697 4532	10
35793485692736433524132	20
86786876868594939653656	15

How to Hash Data

Data

First Name	Last Name	Article	Quantity	Price
John	Smith	1984 by George Orwell	1	10
Lisa	Doe	Ulysses by James Joyce	1	20

Better solution

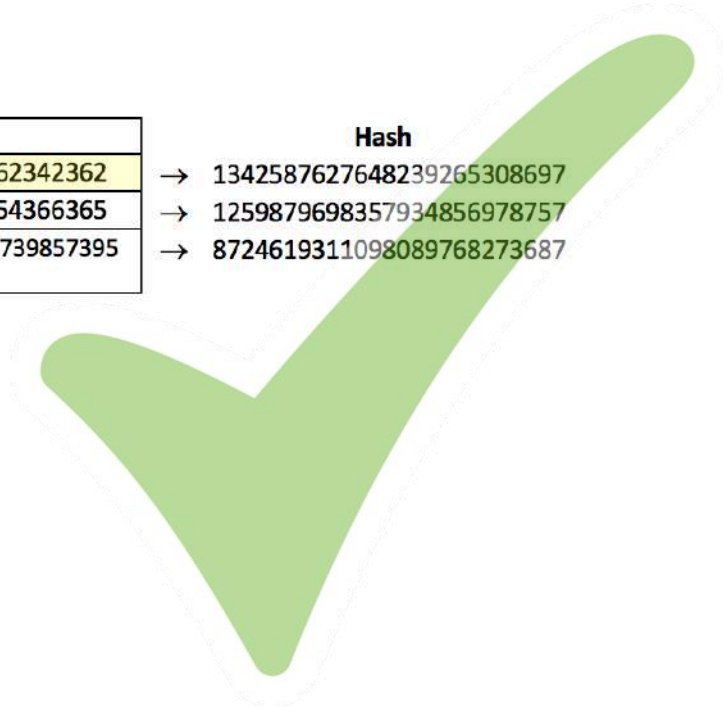
Off-chain

First Name	Last Name	Article	Quantity	Price	Salt
John	Smith	1984 by George Orwell	1	10	876837467762342362
Lisa	Doe	Ulysses by James Joyce	1	20	987936034854366365
John	Smith	Inside Wikileaks by Domscheit-Berg	1	15	29749850385739857395

	Hash
→	134258762764823926530869 7
→	125987969835793485697875 7
→	872461931109808976827368 7

On-chain

Hash
134258762764823926530869 7
125987969835793485697875 7
872461931109808976827368 7



Use Cases for Cryptographic Hash Functions

- Validate external documents
- Time-stamping
- Proof of Existence
- Basic functionality for cryptography and DLT
- Compare documents without disclosing them

Using hash functions the wrong way can lead to
the identification of data subjects and exposure of personal data!

How to check if remaining data on a blockchain constitutes personal data?

Is it possible to identify some information through the remaining data on a blockchain with natural persons?

What if

- somebody knows one transaction, can she see further transactions of the same person?
- somebody knows part of a transaction, can she see further details?
- somebody knows personal details of a person, can she discover information about the person's activity?

If there is remaining personal data on a blockchain, is there proper justification for continuous storage of this data?